

IBM InfoSphere Global Name Management

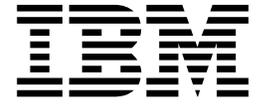


# Guía del desarrollador

*Versión 6 Release 0*



IBM InfoSphere Global Name Management



# Guía del desarrollador

*Versión 6 Release 0*

**Nota**

Antes de utilizar esta información y el producto al que hace referencia, lea la información del apartado Avisos.

**Edición**

Esta edición corresponde a la versión 6.0 de IBM InfoSphere Global Name Recognition (número de producto 5724-Q20) y a todos los releases y modificaciones posteriores hasta que se indique lo contrario en ediciones nuevas.

© Copyright IBM Corporation 2001, 2013.

# Contenido

## Prefacio . . . . . v

## Capítulo 1. Visión general de IBM InfoSphere Global Name Management . 1

Novedades de la versión 6.0 . . . . .	2
Versión 6.0: Notas del release . . . . .	2
Versión 6.0: características y mejoras . . . . .	7
Arquitectura del producto . . . . .	9
Las API de componente . . . . .	9
IBM NameWorks . . . . .	10
Aplicaciones cliente . . . . .	11
Distributed Search . . . . .	11

## Capítulo 2. Visión general de los nombres y la comparación de nombres 13

Procedimientos de comparación de nombres . . . . .	13
Categorías de nombres . . . . .	14
Nombres de persona . . . . .	14
Nombres de organización . . . . .	18
Componentes del nombre . . . . .	19
Árboles de análisis . . . . .	19
Nombres analizados . . . . .	20
Campos de nombre . . . . .	21
Frases de nombre . . . . .	22
Cadenas de caracteres de nombre . . . . .	23
Listas de nombres . . . . .	28
Name data archive . . . . .	28
Lista de cadenas de caracteres externa . . . . .	29
Transliteración de nombres . . . . .	29
Archivos de reglas de transliteración . . . . .	30
Visión general de la transliteración de chino . . . . .	30
Visión general de la transliteración de japonés . . . . .	33

## Capítulo 3. Analizar nombres (desglosar) . . . . . 35

Análisis de nombres utilizando NameWorks . . . . .	35
Desglosar nombres en componentes individuales . . . . .	35
Analizar nombres mediante NameParser . . . . .	36
Tipos de series de entrada . . . . .	36
Funciones de NameParser para el análisis de nombres . . . . .	37
Lista de alteraciones temporales de frases de NameParser . . . . .	38

## Capítulo 4. Analizar nombres . . . . . 41

Análisis de nombres utilizando NameWorks . . . . .	41
Identificación del entorno cultural de un nombre utilizando NameWorks . . . . .	42
Identificación del entorno cultural . . . . .	42
Códigos de entorno cultural . . . . .	43
Identificar el entorno cultural de un nombre completo . . . . .	44
Identificar el entorno cultural de campos de nombre . . . . .	45

Identificación del entorno cultural de un nombre de organización . . . . .	45
Identificación del género de nombres utilizando NameWorks . . . . .	46
Identificar el género de un nombre completo . . . . .	46
Identificación del género de un nombre de pila . . . . .	47
Identificación del país de asociación para nombres mediante NameWorks . . . . .	47
COA (Country of association) . . . . .	47
Identificar el país de asociación para nombres completos . . . . .	48
Identificar el país de asociación para un nombre de pila y apellido . . . . .	48
Generación de las variantes de nombre utilizando NameWorks . . . . .	49
Variantes de nombre . . . . .	49
Generar una lista de variantes para nombres completos . . . . .	50
Generar una lista de variantes para nombres de pila y apellidos . . . . .	51
Analizar nombres con las API de componente . . . . .	52

## Capítulo 5. Buscar nombres . . . . . 53

Gestionar listas de datos en IBM NameWorks . . . . .	53
Listas de datos . . . . .	53
Añadir nombres a listas de datos . . . . .	54
Actualizar nombres contenidos en listas de datos . . . . .	55
Suprimir nombres de listas de datos . . . . .	56
Migración de IBM NameWorks . . . . .	57
Preparar nombres para la búsqueda . . . . .	60
Casos de ejemplo: búsqueda de nombres . . . . .	61
Crear objetos de nombres para la búsqueda de nombres . . . . .	61
Buscar nombres mediante IBM NameWorks . . . . .	64
Gestionar estrategias de búsqueda . . . . .	65
Alteración temporal de parámetros de comparación . . . . .	68
Preparar nombres para la búsqueda . . . . .	72
Categorizar nombres, comparar nombres y comparar fechas mediante IBM NameWorks . . . . .	73
Buscar nombres en una lista de datos . . . . .	80
Recuperar datos complementarios para nombres asociados a un nombre exclusivo . . . . .	81
Buscar nombres mediante NameHunter . . . . .	82
Visión general de NameHunter . . . . .	82
Ejemplos de inicio rápido de la API NameHunter . . . . .	96
Aplicaciones de ejemplo de NameHunter . . . . .	99
Modificar parámetros de comparación . . . . .	100
Configurar conjuntos de reglas de transliteración para NameHunter . . . . .	131
Buscar nombres mediante Distributed Search . . . . .	132
Introducción a Name Preprocessor . . . . .	134
Visión general del rendimiento y la configuración de Distributed Search . . . . .	146

Interfaz XML de NameHunter Distributed Search . . . . .	157
Búsqueda de nombres con Enterprise Name Search	170
Gestión de la seguridad del usuario de Enterprise Name Search . . . . .	170
Gestión de listas de nombres con el programa de utilidad NameLoader . . . . .	175

**Capítulo 6. Configuración de IBM NameWorks . . . . . 189**

Especificación de valores de configuración mediante el archivo de configuración de IBM NameWorks . . . . .	190
Sección General del archivo de configuración	192
Sección Custom tokens (cadenas de caracteres personalizadas) del archivo de configuración . . . . .	193
Sección Datalist del archivo de configuración	194
Especificación de los valores de configuración mediante la clase Configuration . . . . .	203
Actualizar la configuración de IBM NameWorks para utilizar archivos adicionales de reglas de transliteración . . . . .	204

**Capítulo 7. Resolución de problemas y soporte . . . . . 207**

Lista de comprobación para la resolución de problemas de IBM InfoSphere Global Name Management . . . . .	207
Códigos de error de componente de la API para C++ . . . . .	209
Códigos de error de datos de referencia . . . . .	210
Códigos de error global . . . . .	214
Códigos de error de entrada . . . . .	215
Códigos de error interno . . . . .	217
Códigos de error de IBM NameWorks . . . . .	219
Códigos de error de IBM NameWorks para C++	220

Códigos de error de Distributed Search. . . . .	224
Códigos de error de Enterprise Name Search . . . . .	226
Códigos de error de la consola ENS . . . . .	226
Códigos de error de búsqueda de ENS . . . . .	233
Buscar en las bases de conocimientos . . . . .	239
Archivos de registro . . . . .	240
Rastreo . . . . .	240
Cómo ponerse en contacto con el soporte de IBM	241
Suscripción a actualizaciones de soporte . . . . .	242

**Apéndice. Glosario . . . . . 245**

A . . . . .	245
C . . . . .	245
D . . . . .	246
E . . . . .	246
F . . . . .	246
H . . . . .	247
I . . . . .	247
J . . . . .	247
L . . . . .	247
M . . . . .	247
N . . . . .	247
P . . . . .	248
R . . . . .	249
S . . . . .	249
T . . . . .	250
V . . . . .	250

**Avisos . . . . . 251**

**Marcas registradas . . . . . 255**

**Términos y condiciones. . . . . 257**

**Índice. . . . . 259**

---

## Prefacio

IBM InfoSphere Global Name Management utiliza datos de nombre específicos de entorno cultural y crea reglas asociadas con el entorno cultural de los nombres para obtener los mejores resultados de comparación, gestión, análisis y puntuación. Global Name Management es una tecnología puntera que permite buscar, reconocer y gestionar nombres multiculturales, explorar amenazas potenciales y realizar comprobaciones en segundo plano en múltiples entornos geográficos y culturales.

### Acerca de esta publicación

Las publicaciones *Guía del desarrollador* y *Referencia de API* de IBM® InfoSphere Global Name Management se proporcionan para ayudarle que cree aplicaciones que saquen el máximo rendimiento de la tecnología de IBM InfoSphere Global Name Recognition a través de las API proporcionadas.

Para su conveniencia, esta información se suministra en varios formatos. Además del formato PDF (que se encuentra en el DVD de instalación del producto que se puede descargar en [ibm.com](http://ibm.com)), encontrará una versión en línea que incluye toda la información del producto en el Information Center de IBM InfoSphere Global Name Management. También puede instalar el centro informativo del producto basado en navegador en una máquina local.

### Destinatarios

La *Guía del desarrollador* y la *Referencia de API* tienen como finalidad a facilitar el despliegue satisfactorio del producto en el entorno y crear aplicaciones.

### Cómo enviar sus comentarios

Sus comentarios son importantes para ayudarnos a proporcionar información con la mayor exactitud y la más alta calidad. Si desea hacer comentarios sobre esta publicación o cualquier otra documentación de IBM InfoSphere Global Name Management, puede utilizar el enlace *Comentarios* del Information Center o bien este formulario:

<http://www.ibm.com/software/data/rcf/>



---

## Capítulo 1. Visión general de IBM InfoSphere Global Name Management

El producto IBM InfoSphere Global Name Management contiene tecnologías para gestionar, buscar, analizar y comparar conjuntos de datos de nombres multiculturales mediante la utilización de datos de nombres específicos del entorno cultural y reglas lingüísticas asociadas al entorno cultural de los nombres.

Los componentes de IBM InfoSphere Global Name Management permiten:

- Identificar y clasificar el entorno cultural (categoría étnica) más probable de un nombre, incluidos los países en los que el nombre de pila o apellido se encuentra con más frecuencia.
- Reconocer y notificar las frecuencias relativas de género (masculino o femenino) asociadas con nombres de pila.
- Desglosar nombres de persona en componentes de apellido y nombre de pila.
- Generar listas que contengan formas variantes de los componentes (nombre de pila y apellido) de un nombre.
- Buscar y comparar nombres mediante estrategias de búsqueda específicas de entorno cultural.
- Comparar nombres aunque resulten afectados por variaciones típicas ortográficas y culturales, relacionados por fonética pero no por ortografía, o perjudicados por errores ortográficos o tipográficos.
- Comparar nombres aunque falte alguno de sus componentes (nombre de pila o apellido) o no esté en el orden correcto.
- Comparar nombres en cuanto a pronunciación y ortografía, devolviendo primero las coincidencias más exactas.
- Ajustar parámetros de búsqueda para obtener resultados altamente ajustables específicos de aplicación.
- Separar los nombres de persona de los nombres de organización.
- Comparar valores de fecha, lo cual puede ser de utilidad al buscar un nombre que tiene un valor de fecha asociado (por ejemplo una fecha de nacimiento) o calcular diferencias entre valores de fecha.

### Componentes de API y procesos de servidor

Los componentes de API y procesos de servidor de este paquete incluyen:

- IBM NameWorks y API integrada de alto nivel
- NameParser
- NameClassifier
- NameClassifier - Country of Association
- Country of Association
- NameHunter
  - Proceso de Distributed Search
- NameGenderizer
- NameVariationGenerator
- NameSifter
- DateCompare

**Conceptos relacionados:**

“Novedades de la versión 6.0”

Esta versión de IBM InfoSphere Global Name Management contiene numerosas características nuevas y mejoras del producto.

“Versión 6.0: Notas del release”

Estas notas del release contienen información sobre IBM InfoSphere Global Name Management, versión 6.0, como las notas de instalación, problemas conocidos, problemas solucionados y notas de uso.

---

## Novedades de la versión 6.0

Esta versión de IBM InfoSphere Global Name Management contiene numerosas características nuevas y mejoras del producto.

Para obtener la información más reciente sobre IBM InfoSphere Global Name Management Versión 6.0, vaya al sitio web del producto en <http://www-03.ibm.com/software/products/en/infosphere-global-name-management>

### Adiciones y obsolescencias de plataforma

Esta versión de IBM InfoSphere Global Name Management da soporte a versiones más recientes de varias plataformas. Otras plataformas han quedado en desuso en este release. Para obtener un listado de las plataformas y entornos de desarrollo admitidos, consulte los requisitos del sistema en [ibm.com](http://ibm.com).

**Conceptos relacionados:**

Capítulo 1, “Visión general de IBM InfoSphere Global Name Management”, en la página 1

El producto IBM InfoSphere Global Name Management contiene tecnologías para gestionar, buscar, analizar y comparar conjuntos de datos de nombres multiculturales mediante la utilización de datos de nombres específicos del entorno cultural y reglas lingüísticas asociadas al entorno cultural de los nombres.

## Versión 6.0: Notas del release

Estas notas del release contienen información sobre IBM InfoSphere Global Name Management, versión 6.0, como las notas de instalación, problemas conocidos, problemas solucionados y notas de uso.

**Nota:** Si desea obtener la versión más reciente de las notas de release, consulte la versión en línea en el Information Center de [ibm.com](http://ibm.com) o el archivo HTML de las notas de release independiente que acompaña al soporte de instalación y que se puede descargar por separado del portal de soporte del producto.

### Contenido

- “Requisitos del sistema” en la página 3
- “Consideraciones relativas al rendimiento” en la página 3
- “Actualización a la versión 6.0 de IBM Global Name Management” en la página 3
- “Lista de arreglos y correcciones” en la página 3
- “Problemas conocidos a la hora de utilizar el producto” en la página 3
- “Problemas conocidos: Enterprise Name Search” en la página 4
- “Información más reciente sobre los problemas conocidos” en la página 6
- “Documentación del producto” en la página 6

- “Anuncios” en la página 7

## Requisitos del sistema

Para obtener la información más reciente sobre la compatibilidad de hardware y software, consulte el documento de requisitos de sistema en <http://www-01.ibm.com/support/docview.wss?&uid=swg27019150>.

## Consideraciones relativas al rendimiento

Las búsquedas no deben ser de más de cinco cadenas de caracteres cuando se realiza una búsqueda en una lista de datos grandes (al menos 5 millones de nombres) en una máquina que ejecuta Linux para IBM zSeries (s390). Si se incluyen más de cinco cadenas de caracteres en la consulta de búsqueda se pueden obtener resultados que superen el tamaño límite, ocasionando el tiempo de espera excedido de una transacción y, por consiguiente, errores en los servicios web. Para obtener más información de rendimiento, consulte el centro de información del producto.

## Actualización a la versión 6.0 de IBM Global Name Management

El programa de instalación escribe nuevos archivos de configuración en *vía\_acceso\_instalación/data/\*.config.template*, donde *vía\_acceso\_instalación* es la vía de acceso completa del directorio donde ha instalado IBM InfoSphere Global Name Management. Los archivos de configuración existentes se conservan; el programa de instalación no los sobrescribe.

## Lista de arreglos y correcciones

En la lista siguiente se describen algunas de las mejoras funcionales y los problemas resueltos que se han incluido en la versión 6.0:

- Reparación de defecto número 1
- Reparación de defecto número 2
- Reparación de defecto...

## Problemas conocidos a la hora de utilizar el producto

Consulte la siguiente información antes de instalar y utilizar la versión 6.0 de IBM InfoSphere Global Name Management.

### Cambios en la API para C++ de NameParser

La API de NameParser se ha simplificado en la versión 6.0, para estar en consonancia con el comportamiento de la API de análisis de NameWorks. Los programas en C++ que utiliza NameParser se han de modificar directamente para que funcionen con la versión 6.0.

### Cambios y adiciones en las entradas del archivo de configuración de NameWorks

Como apoyo a la comparación de scripts nativos, en la versión 6.0 se ha añadido soporte para los archivos de comparación.

- *NativeTaq=vía de acceso al archivo TAQ de script nativo*  
Especifica los datos TAQ de script nativo.
- *NativePnVar=vía de acceso al archivo de variantes de nombre de persona de script nativo*  
Especifica los datos de variantes de nombre de persona de script nativo

- **NativeOnVar**=vía de acceso del archivo de variantes de nombre de organización de script nativo  
Especifica los datos de variantes de nombre de organización de script nativo
- **NativePnReg**=vía de acceso al módulo de regularización de nombres de persona de script nativo, nombre de script  
Especifica los datos de regularización de los nombres de persona de script nativo para un tipo de script específico
- **NativeOnReg**=vía de acceso al módulo de regularización de nombres de organización de script nativo, nombre de script  
Especifica los datos de regularización de los nombres de organización de script nativo para un tipo de script específico

## Problemas conocidos: Enterprise Name Search

Tenga en cuenta la información adicional y las instrucciones siguientes antes de instalar y utilizar Enterprise Name Search.

### Enterprise Name Search requiere codificación UTF-8

Si está trabajando con datos que incluyen caracteres no ASCII con servicio web de SOAP, asegúrese de que las variables del entorno local se hayan establecido para utilizar la codificación UTF-8. Los servicios web de REST no se ven afectados. Si sus sistemas no toman de forma predeterminada la codificación UTF-8, haga lo siguiente:

#### Entorno Linux/UNIX:

1. Busque o instale la versión UTF-8 del entorno local para su idioma y país. Este entorno local se debe instalar en todas las máquinas utilizadas en la célula de ENS. Utilice el mandato **locale -a** para ver una lista de los entornos locales instalados en la máquina. Por ejemplo:

```
jksmith@din:~$ locale -a
CC.UTF-8
en_AG
en_AG.utf8
en_AU.utf8
en_GB.utf8
en_US.utf8
ja_JP.utf8
POSIX
```

Busque o instale la versión UTF-8 del entorno local para su idioma y país. Por ejemplo, si está en los Estados Unidos, puede seleccionar "en\_US.UTF-8" de la lista.

2. Asegúrese de que el entorno de shell que inicia procesos de ENS WebSphere (ejecuta los scripts **<Inicio instalación de ENS>/bin/start-<nombre de perfil>.sh**) tenga establecido el siguiente entorno local para el entorno local UTF-8 seleccionado ANTES de iniciar ENS. Ejemplo del shell Bourne (sh):

```
export LC_ALL=en_US.UTF-8
export LANG=en_US.UTF-8
```

Ejemplo de un shell Bourne:

```
LC_ALL=en_US.UTF-8
LANG=en_US.UTF-8
export LC_ALL
export LANG
```

#### Entorno Windows:

Windows no da soporte a la posibilidad de que UTF-8 sea la codificación de caracteres predeterminada en todo el sistema o para todas las tareas que lleve a cabo el usuario, por lo que se tendrá que configurar las aplicaciones para que utilicen UTF-8 caso por caso. Para ENS, debe configurar la JVM de Java que ejecuta WebSphere y ENS para que utilice UTF-8. La JVM de Java de IBM J9 viene en el paquete con WebSphere y tiene una variable de entorno para la especificación de opciones JVM denominada "IBM\_JAVA\_OPTIONS". Esta variable se puede establecer en Windows de diversas formas.

- Desde el indicador de mandatos de Windows:
  1. Establezca la variable de entorno: **IBM\_JAVA\_OPTIONS=-Dfile.encoding=UTF-8**
  2. Ejecute el script **start-(Nombre perfil ENS).bat**.
- Con alcance a un usuario específico de Windows:

Este valor se propagará a todas las aplicaciones Java basadas en IBM J9 que ejecute el usuario.

  1. Pulse el botón **VARIABLES DE ENTORNO...** en la pestaña **Avanzado** del Panel de control del sistema. En Windows 2008 Server vaya a **Inicio > Panel de control > Valores del sistema avanzados** para llegar directamente a la pestaña Avanzado del panel de control del sistema. Aparecerá el diálogo **VARIABLES DE ENTORNO**.
  2. En la sección **Variable de usuario para <nombre usuario actual>**, pulse el botón **Nuevo...**
  3. Aparecerá el recuadro **Variable de usuario nuevo**. En el campo "Nombre de variable" escriba IBM\_JAVA\_OPTIONS.
  4. En el campo "Valor de variable" escriba -Dfile.encoding=UTF-8.
  5. Pulse **Aceptar** en el recuadro. A continuación, pulse **Aceptar** en el diálogo Variables de entorno y en el Panel del control del sistema.
  6. Finalice la sesión en Windows y, a continuación, vuelva a iniciarla. Las aplicaciones Java que haya iniciado este usuario que utiliza la JVM de IBM J9 utilizarán ahora la codificación UTF-8 de forma predeterminada.
- Con alcance a todo el sistema Windows

Este valor se propagará a todas las aplicaciones Java basadas en IBM J9 que se hayan iniciado en el servidor de Windows.

  1. Pulse el botón **VARIABLES DE ENTORNO...** en la pestaña **Avanzado** del Panel de control del sistema. En Windows 2008 Server vaya a **Inicio > Panel de control > Valores del sistema avanzados** para llegar directamente a la pestaña Avanzado del panel de control del sistema. Aparecerá el diálogo **VARIABLES DE ENTORNO**.
  2. En la sección **VARIABLES DEL SISTEMA**, pulse **Nuevo...**
  3. Aparecerá el recuadro **Variable de sistema nuevo**. En el campo "Nombre de variable" escriba IBM\_JAVA\_OPTIONS.
  4. En el campo "Valor de variable" escriba -Dfile.encoding=UTF-8.
  5. Pulse **Aceptar** en el recuadro. A continuación, pulse **Aceptar** en el diálogo Variables de entorno y en el Panel del control del sistema.
  6. Reinicie el sistema. Las aplicaciones Java que se hayan iniciado en este sistema y que utilicen la JVM de IBM J9 ahora utilizarán la codificación UTF-8 de forma predeterminada.

**Nota:** La variable de entorno IBM\_JAVA\_OPTIONS anterior también debería funcionar en cualquier otra plataforma que utilice la JVM de IBM J9 para ejecutar WebSphere, incluyendo Linux y AIX. No funcionará en Solaris porque IBM no utiliza la JVM de IBM J9 con ese sistema operativo.

### **Creación de varios perfiles utilizando el programa de utilidad CU de ENS**

Debería crear sólo un perfil cada vez. La creación de varios perfiles utilizando el programa de utilidad CU de ENS puede tardar bastante más tiempo, especialmente con una elevada cantidad de perfiles. La creación de cada perfil adicional es más lenta que la anterior porque el perfil recién creado hace que se genere una clave de cifrado SSL entre esta creación y cada perfil existente. Estas claves son necesarias para dar soporte a la función de comunicación segura SSL/HTTPS ente dos perfiles cualesquiera. Por ejemplo, si crea 35 perfiles, el perfil número 35 tendrá que crear 34 claves de cifrado.

### **Un controlador de base de datos para todos los sistemas en un entorno distribuido de ENS**

Tal como se indicaba en otra parte de la documentación de ENS, el controlador de base de datos debe ser el mismo y hallarse en una ubicación para todas las máquinas host y servidor de ENS en su entorno de ENS (denominado célula de ENS). Puede crear un enlace dinámico a la ubicación del controlador de base de datos. Si todas las máquinas apuntan al enlace dinámico, puede realizar cambios en la ubicación del controlador de base de datos sin tener que realizar cambios en cada máquina.

### **Información más reciente sobre los problemas conocidos**

Los problemas conocidos se documentan en las notas técnicas en el portal de soporte: [http://www.ibm.com/support/entry/portal/Overview/Software/Information\\_Management/InfoSphere\\_Global\\_Name\\_Recognition](http://www.ibm.com/support/entry/portal/Overview/Software/Information_Management/InfoSphere_Global_Name_Recognition):

1. Utilice la función **Buscar soporte** y en el campo **Indicar términos, código de error o número de APAR** escriba la palabra clave, frase, código de error o número de APAR que se va a buscar.
2. Seleccione **Resolver un problema**.
3. Pulse **Buscar**.

A medida que se van detectando y resolviendo los problemas, el equipo de soporte de IBM actualiza el portal de soporte. Al realizar búsquedas en el portal de soporte, encontrará rápidamente soluciones a los problemas.

En el momento de la publicación, no existían problemas de instalación conocidos. Compruebe el portal de soporte para obtener la información más actual.

### **Documentación del producto**

Puede encontrar documentación del producto de la versión 6.0 en los lugares siguientes:

#### **Information Center de la versión 6.0**

- Acceda a [ibm.com](http://pic.dhe.ibm.com/infocenter/gnrngm/v6r0m0/index.jsp): <http://pic.dhe.ibm.com/infocenter/gnrngm/v6r0m0/index.jsp>
- Acceda al servidor local, que se instala como parte de la instalación del producto:

Abra un navegador web y especifique el URL del servidor:  
`http://<nombre_servidor>:<número_puerto_HTTP>/help/index.jsp`

`<número_puerto_HTTP>`

Número de puerto de la documentación que ha especificado durante la instalación.

`<nombre_servidor>`

Nombre de host o dirección IP de WebSphere Application Server.

### **Referencias culturales de Name Analyzer**

La información de referencia del entorno cultural se proporciona en la herramienta Name Analyzer (analizador de nombres) de IBM InfoSphere Global Name Management.

### **Inicio de soporte a productos IBM**

Acceda a [ibm.com](http://www.ibm.com/support/entry/portal/Overview/Software/Information_Management/InfoSphere_Global_Name_Recognition): `http://www.ibm.com/support/entry/portal/Overview/Software/Information_Management/InfoSphere_Global_Name_Recognition`

Además de notas técnicas y otra información relacionada con el soporte, encontrará enlaces al Information Center, versiones en PDF de la información del producto y las últimas actualizaciones de las notas del release.

## **Anuncios**

Encontrará la carta de anuncio más reciente en forma de enlace en la página siguiente en <http://www.ibm.com/software/data/infosphere/global-name-recognition/>. Consulte el anuncio para obtener la información siguiente:

- Descripción detallada del producto, que incluye la descripción de las nuevas funciones
- Declaración de posicionamiento del producto
- Detalles de embalaje y pedido
- Información de compatibilidad internacional

### **Información de copyright y de marcas registradas**

IBM, el logotipo de IBM e [ibm.com](http://www.ibm.com) son marcas registradas de International Business Machines Corp. en muchas jurisdicciones en todo el mundo. Los nombres de otros productos y servicios pueden ser marcas registradas de IBM o de otras compañías. En la página web [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) existe una lista actualizada de marcas registradas de IBM bajo el epígrafe "Información de copyright y marcas registradas".

#### **Conceptos relacionados:**

Capítulo 1, "Visión general de IBM InfoSphere Global Name Management ", en la página 1

El producto IBM InfoSphere Global Name Management contiene tecnologías para gestionar, buscar, analizar y comparar conjuntos de datos de nombres multiculturales mediante la utilización de datos de nombres específicos del entorno cultural y reglas lingüísticas asociadas al entorno cultural de los nombres.

## **Versión 6.0: características y mejoras**

La versión 6.0 de IBM InfoSphere Global Name Management contiene nuevas funciones del producto y mejoras de las funciones existentes.

Tabla 1. Características y mejoras del producto en la versión 6.0

Mejora de la versión 6.0	Descripción
Nuevos entornos culturales estándar: polaco, portugués y turco	Se han añadido como entornos culturales estándar nuevos el polaco, el portugués y el turco. Estos nuevos entornos culturales dan soporte a variantes específicas del entorno cultural y a parámetros de datos de género y de comparación. El portugués también se incluye en el entorno cultural de grupo europeo y el turco se incluye en el entorno cultural de grupo sudoeste asiático.
Soporte a extensiones para scripts nativos	Se ha ampliado el soporte a scripts nativos para gestionar nombres de persona escritos en latín ampliado (hispanico, polaco, portugués) y devanagari (hindú), así como los nombres de organización escritos en cirílico (ruso), hangul (coreano), kanji (japonés), hanzi (chino) y devanagari (hindú).
Entornos culturales de nombres de organizaciones nuevos	Se ha ampliado la puntuación de los nombres de organizaciones para dar soporte a entornos culturales adicionales: chino, hispanico, japonés, coreano, polaco, portugués y ruso.
Puntuación mejorada para nombres abreviados	ahora, los procedimientos de puntuación nuevos y los parámetros de comparación nuevos tienen en cuenta la posibilidad que dos nombres abreviados que difieran de un posible error tipográfico sean realmente nombres distintos. Por ejemplo: "Amir" y "Mair" se diferencian en una única transposición pero son nombres diferentes, mientras que "Amir" y "Aimr" representan un error tipográfico. Para realizar esta determinación se utiliza la información sobre la frecuencia del Name Data Archive (NDA). Seguidamente se utiliza una lógica especial para comparar los nombres.
Información de género mejorada	La cantidad de datos de género disponible para el análisis ha aumentado más de 10 veces la cantidad existente, y ahora hay disponible información de género específica del entorno cultural. Por ejemplo, "Juan" es un nombre masculino en la mayoría del mundo, pero se trata de un nombre especialmente femenino en China.
API para C++ de NameParser simplificada	La API para C++ para NameParser se ha simplificado: una única llamada devolverá todos los análisis de un nombre.
Soporte de ENS para varios nombres con el mismo ID externo	Las versiones anteriores de ENS requerían que los nombres fueran exclusivos en una lista de nombres. La versión 6 elimina esta restricción y permite que haya varios nombres con el mismo ID externo en la misma lista.
Soporte de ENS y utilización de los cambios de script y entorno cultural de GNM	Ahora, ENS da soporte y utiliza los cambios de GNM a scripts y entornos culturales incluidos los entornos culturales que se han añadido, la inferencia del entorno cultural del tipo de script, el reconocimiento de tipo de script añadido y el soporte para los entornos culturales de nombres de organizaciones.
Soporte de NameLoader de ENS para comas en nombres	Los campos de nombre en un archivo de entrada con valores separados por comas (.csv) de NameLoader de ENS ahora pueden contener comas, si esos campos se hallan entre comillas.
Soporte de ENS para DB2 10.5 y Oracle 12c	ENS ahora se ejecuta con DB2 10.5 y Oracle 12c. Saca provecho del soporte de la versión 10.5 de DB2 cuanto a índices funcionales. En Oracle, permite nombres de servicio (o SID) cuando se especifican los detalles de la base de datos.

Tabla 1. Características y mejoras del producto en la versión 6.0 (continuación)

Mejora de la versión 6.0	Descripción
Uso de ENS de WebSphere Liberty	ENS utiliza ahora WebSphere Liberty de forma interna como servidor de aplicaciones para alojar componentes web, en lugar del producto anterior Embedded WebSphere. Como resultado, la implementación de seguridad ha cambiado y la configuración de ENS es más simple.

Para obtener una lista completa de los arreglos, cambios mínimos y mejoras, que incluyen API modificadas y nuevas, consulte el archivo de notas del release en el soporte de instalación del producto o en el portal de soporte del producto IBM InfoSphere Global Name Management .

## Arquitectura del producto

La arquitectura de producto de IBM InfoSphere Global Name Management consta de las API de componente, las API de IBM NameWorks y de las aplicaciones de cliente y servidor que se comunican con estas API.

Las aplicaciones de servidor son aplicaciones del lado del servidor creadas sobre las API de componentes y que suministran las funciones de éstas. IBM InfoSphere Global Name Management incluye las aplicaciones de servidor NameHunter y Distributed Search. También puede desarrollar sus propias aplicaciones de servidor mediante las API de componente y de IBM NameWorks.

Además de las aplicaciones y los componentes siguientes, la versión 6.0 incluye Enterprise Name Search, un conjunto de componentes instalado de forma independiente con una interfaz de aplicación web que facilita la gestión, mantenimiento y realización de búsquedas de nombres.

## Las API de componente

Las API de componente IBM InfoSphere Global Name Management son bibliotecas C++ que se pueden integrar en cualquier aplicación C++.

Todas las API de componente realizan una función analítica de un único nombre, pero NameHunter y DateCompare toman dos o más objetos (nombres y fechas, respectivamente) y los comparan. En la lista siguiente figura cada una de las API de componente:

### NameClassifier

El paquete de NameClassifier (ibmgnr::classifier) determina cómo se asocia probablemente un nombre de persona con uno o varios entornos culturales.

### NameGenderizer

El paquete de NameGenderizer (ibmgnr::genderizer) suministra estadísticas de distribución de género para el componente de nombre de pila de un nombre de persona.

### NameParser

El paquete de NameParser (ibmgnr::parser) desglosa los nombres de persona en sus componentes constituyentes (nombre de pila, apellido, fórmulas de tratamiento y calificadores).

### **NameVariationGenerator**

El paquete de NameVariantGenerator (ibmgmr::nvg) genera una lista de las formas variantes de cada componente de un nombre de persona. Estas ortografías alternativas se basan en patrones de variación ortográfica observados generalmente en los nombres de un mismo origen cultural o étnico.

### **NameSifter**

El paquete de NameSifter (ibmgmr::sifter) separa los nombres de organización de los nombres de persona.

### **NameClassifier-Country of Association (NC\_COA)**

El paquete de NameClassifier-Country of Association (ibmgmr::cc) utiliza COA (Country of Association) junto con NameClassifier para generar resultados muy exactos para un entorno cultural de nombres asociado.

### **Country of Association (COA)**

El paquete de COA (ibmgmr::coa) hace referencia a los datos contenidos en IBM Name Data Archive (NDA) para listar los países en los que se ha observado que aparece cada uno de los componentes de un nombre de persona.

### **NameHunter**

El paquete de NameHunter (ibmgmr::hunter) compara pares de nombres de persona y de organización y también busca en listas con estos tipos de nombres.

### **DateCompare**

DateCompare (ibmgmr::datecompare) compara dos valores de fecha y devuelve una puntuación de similitud. DateCompare sólo puede comparar fechas del calendario gregoriano de 12 meses.

## **IBM NameWorks**

IBM NameWorks combina los componentes individuales de IBM InfoSphere Global Name Management en una única API (interfaz de programación de aplicaciones) unificada y de fácil utilización, y también amplía esta funcionalidad a las aplicaciones Java y como servicio web.

IBM NameWorks consta de dos clases de API distintas:

### **Clase Analytics**

Incluye las funciones necesarias para evaluar un único nombre, que incluyen el análisis del nombre, la clasificación del entorno cultural, el establecimiento del género, la categorización, la generación de variantes e información de país de asociación. Puede utilizar estos procesos lingüísticos individual o conjuntamente. Por ejemplo, el método analyze() realiza todas las operaciones lingüísticas y genera un único resultado combinado que contiene toda la información analítica de un nombre.

### **Clase Scoring**

Incluye las funciones necesarias para comparar dos nombres o buscar un nombre en una o varias listas de datos, junto con tareas complementarias tales como comparación de fechas y categorización de nombres, que pueden utilizarse para refinar los resultados de búsqueda. La preparación de la búsqueda (análisis y clasificación de entorno cultural) puede realizarse por separado o incluirse en una operación de búsqueda.

Puede acceder a IBM NameWorks de tres maneras: mediante las funciones C++, las *funciones Java* o *los servicios*. Las interfaces C++ y Java pueden utilizarse

directamente en cualquiera de las plataformas soportadas, y la interfaz de servicios web puede utilizarse local o remotamente en entornos SOA. Cualquier entorno de programación que pueda utilizar servicios Web puede aprovechar las herramientas de análisis y comparación de nombres suministradas por IBM NameWorks. También puede utilizarse la interfaz Java para crear aplicaciones SOA personalizadas.

## **Aplicaciones cliente**

Las aplicaciones cliente se crean sobre las API de componente o de IBM NameWorks. Estas aplicaciones pueden comunicarse con aplicaciones del lado del servidor creadas sobre la misma infraestructura.

Puede utilizar cualquier paquete de API para crear aplicaciones que visualicen una amplia variedad de arquitecturas físicas, desde soluciones autónomas sencillas que operen en una única plataforma de host hasta soluciones más complejas que operen como procesos independientes en varias plataformas de host conectadas en red, como por ejemplo en un entorno cliente-servidor. Existen dos tipos principales de aplicaciones cliente:

### **aplicaciones de usuario final**

Aplicaciones creadas sobre las API de componente o el paquete IBM NameWorks que se compilan para ejecutarse en la máquina del usuario.

### **aplicaciones del lado del cliente**

Aplicaciones del lado del cliente que se comunican con aplicaciones de servidor creadas sobre las API de componente o el paquete IBM NameWorks. Por ejemplo, un cliente de servidor Web de IBM NameWorks creado desde las API de SOAP.

## **Distributed Search**

Distributed Search expone la funcionalidad de la API de NameHunter en forma de un único proceso servidor que puede dar cabida a requisitos de búsqueda complejos e intensivos en cuanto a rendimiento debidos al tamaño de las listas de datos en las que debe buscarse o al número de transacciones de búsqueda realizadas en un momento determinado.

Distributed Search es adecuada para la carga de listas de datos de grandes dimensiones, que comprenden millones de nombres. Sin embargo, la aplicación no puede cargar varias listas de datos en una sola búsqueda. Esta limitación impide que los clientes busquen en varias listas de datos desde un solo mensaje XML. Si tiene que cargar varias listas de datos pequeños en una sola aplicación de servidor, debe utilizar la aplicación de búsqueda de IBM NameWorks.

Puede interactuar con Distributed Search directamente o por medio de aplicaciones cliente de usuario final y aplicaciones de servidor creadas sobre IBM NameWorks.



---

## Capítulo 2. Visión general de los nombres y la comparación de nombres

La comparación de nombres puede ser especialmente difícil debido a que no existen estándares globales coherentes para los nombres, y porque éstos pueden contener una variedad de información (muchas de la cual es opcional) que puede hacer que los nombres aparezcan de forma muy diferente. Los productos IBM InfoSphere Global Name Management utilizan una base de conocimientos exclusiva de nombres multiculturales e información lingüística que permite las mejores prestaciones de búsqueda y comparación de nombres específicas de entorno cultural.

---

### Procedimientos de comparación de nombres

El software de la comparación de nombres automatizada utiliza generalmente cuatro procedimientos básicos: coincidencia exacta, diccionario de nombres, basada en claves y analítica.

#### **Coincidencia exacta**

Los sistemas de coincidencia exacta exigen que el nombre de consulta coincida exactamente con el nombre de la base de datos para devolver un resultado. Es decir, lo que el sistema devuelve coincide exactamente con la serie de consulta, y nada más.

#### **Diccionario**

Los sistemas de comparación de diccionario buscan el nombre en un directorio para encontrar sus variantes, cualquiera de las cuales puede coincidir con las entradas de la base de datos. El conjunto de variantes ortográficas asociadas con un nombre específico se compila manualmente por adelantado, para que ninguna variante aún no conocida u observada aparezca en el diccionario de nombres. También pueden listarse determinadas formas de nombre como variantes de dos o más nombres diferentes, lo cual complica la lógica de decisión basada en la búsqueda de varias coincidencias de diccionario para el mismo nombre de consulta.

#### **Basada en claves**

Los sistemas basados en claves aplican un algoritmo para reducir un nombre a una forma estandarizada conocida como clave. En teoría, todos los nombres interpretados como formas ortográficamente equivalentes o coincidentes visualizan la misma clave cuando se procesan mediante el algoritmo de generación de claves. El algoritmo de generación de claves más antiguo y conocido es Soundex, patentado por primera vez en 1911.

#### **Analítica**

Los sistemas de comparación de nombres analítica tienen en cuenta simultáneamente información ortográfica, técnicas de filtrado de palabras superfluas y patrones semánticos, culturales y sintácticos para medir la similitud entre dos nombres. Este procedimiento de comparación de nombres por pares depende en gran medida del acceso a información empírica exhaustiva relativa a la utilización de un nombre dentro de su contexto lingüístico y cultural asociado.

Los productos IBM InfoSphere Global Name Management se basan en el procedimiento analítico para la comparación de nombres. Los datos lingüísticos y

culturales utilizados para el análisis de nombres están soportados por un amplio repositorio de datos relativos a los nombres, recogidos de diversos países de todo el mundo. Esta base de conocimientos permite minuciosas distinciones al comparar nombres para que la relativa similitud entre dos nombres pueda medirse con precisión. La comparación de nombres analítica permite listar los resultados de la comparación por orden jerárquico, mostrando en primer lugar las mejores coincidencias. También pueden ajustarse los algoritmos de clasificación y puntuación para que los resultados de comparación se ajusten a diversos valores operativos, preferencias de usuario y reglas empresariales diferentes.

---

## Categorías de nombres

Durante el proceso de nombres, éstos se asocian con una categoría de nombres, que puede ser de persona o de organización. Aunque pueden compartir una utilización similar, los nombres de estas dos categorías están separados por importantes diferencias y, por tanto, se aplican tipos diferentes de recursos lingüísticos y de datos de referencia a los nombres de cada categoría durante el análisis y la comparación.

Al categorizar nombres, los componentes de IBM InfoSphere Global Name Management los sitúan en las categorías siguientes:

- Nombres de persona, que no contienen indicadores que sugieran su pertenencia a ninguna otra categoría (por ejemplo: "Linda K. Smith")
- Nombres de organización, que contienen algún tipo de indicador no personal (por ejemplo: "Smith & Company")
- Nombres desconocidos, que contienen algún elemento que parece estar mal escrito o que contienen alguna otra construcción que no aparece normalmente en los nombres de persona o de organización (por ejemplo: "SMI")
- Ambos, que son nombres que contienen un calificador profesional que puede sugerir que se trata de un nombre de empresa derivado de un nombre de persona (por ejemplo: "Linda Smith Architect")

Si un nombre se categoriza como un nombre no de persona, el componente suministra un código de razón que identifica el indicador o patrón que cualifica al nombre como no de persona.

## Nombres de persona

Un nombre de persona consta de uno o varios nombres de pila, los nombres de familia o grupo (por ejemplo nombres de tribu o clan) u otros elementos de tipo apellido utilizados en el entorno cultural del que procede el nombre y las fórmulas de tratamiento y otros calificadores de nombre asociados con el titular del mismo. Un nombre de persona completo hace referencia a un individuo y puede codificar información que indique la clase social, los antecedentes religiosos y políticos, el nivel de formación, el origen étnico o cultural y la procedencia regional.

### Modelo de nombres de persona de IBM InfoSphere Global Name Management

Para describir nombres de persona y trabajar con ellos independientemente de su formato nativo, es importante utilizar una terminología coherente. También es muy importante poder desglosar los nombres de forma coherente en sus componentes constituyentes, para poder comparar los componentes equivalentes.

El formato del modelo de nombres de persona de IBM InfoSphere Global Name Management está motivado por la necesidad de tratar con los nombres tal como están codificados en conjuntos de datos en la realidad. Es un procedimiento práctico para determinar la estructura de un nombre. Por ejemplo, aunque los nombres en muchas partes del mundo no tienen verdaderos apellidos en el sentido del entorno cultural occidental, sin embargo dichos nombres se colocan en bases de datos que asumen apellidos. Por tanto, con la finalidad de realizar un proceso de nombres coherente, IBM InfoSphere Global Name Management aplica una estructura de dos campos. El campo al que pertenecen los diversos componentes de un nombre está determinado en parte por la frecuencia con la que cada componente del nombre se ha asociado con un campo de nombre de pila o de apellido. Dentro de cada campo, los elementos de nombre individuales se desglosan en unidades más grandes. Por ejemplo, el apellido “de la Salle” se reconoce como una frase de nombre formada por una raíz de nombre principal y dos prefijos, no como tres componentes de nombre separados.

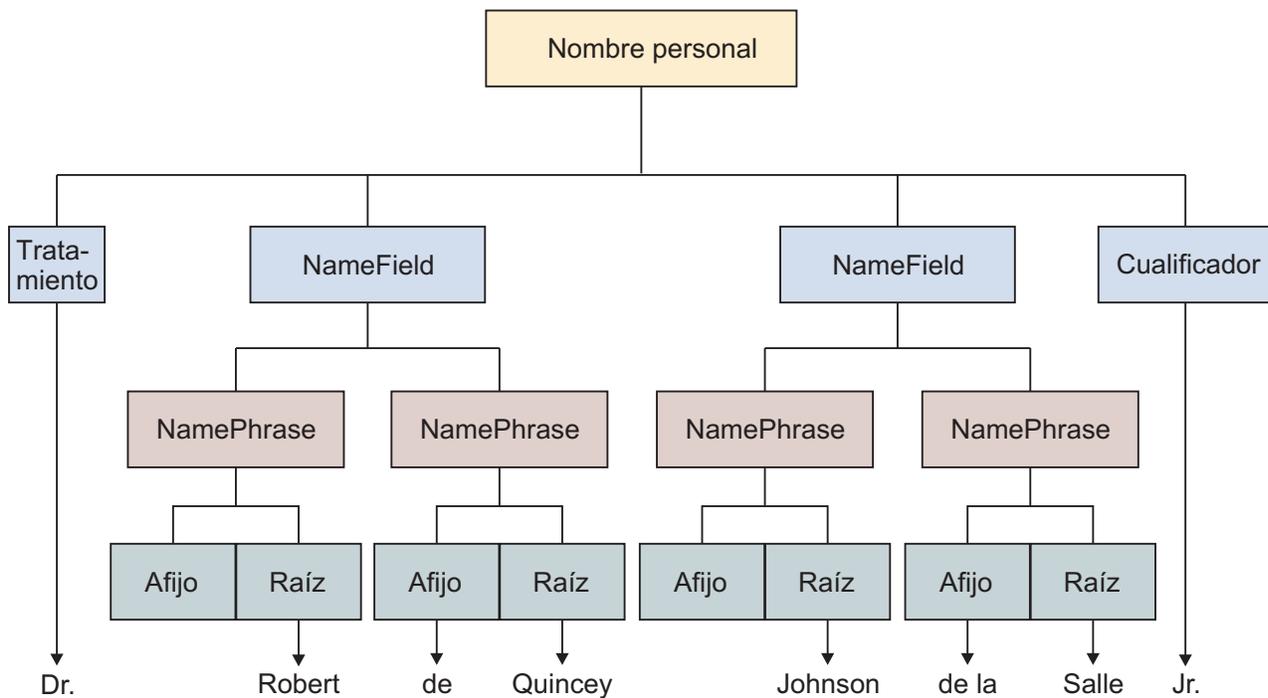


Figura 1. Modelo de nombres de persona de IBM InfoSphere Global Name Management

## Estructura y componentes de los nombres de persona

Los nombres de persona pueden contener componentes muy diferentes. Estos componentes y el modo en que se estructuran difieren en función de los grupos culturales.

A continuación se indican algunos de los componentes que pueden utilizarse en los nombres de persona:

- Nombre de pila
- Apellido
- Nombre de familia
- Nombre de tribu, clan o casta

- Indicadores de parentesco o linaje (por ejemplo, patronímicos (nombres derivados del nombre del padre), matronímicos (nombres derivados del nombre de la madre), tecnonímicos (nombres derivados del nombre de un hijo) e indicadores generacionales)
- Calificadores que indican el orden de nacimiento, el género, la religión o la confesión religiosa
- Fórmulas de tratamiento
- Partículas (como por ejemplo "bin" (hijo) y "al" (el) en árabe o "de" en español y francés)

La estructura de los nombres de persona o el orden de los componentes del nombre también varían según el país o el grupo cultural.

A continuación figuran algunos ejemplos de estructuras de nombre:

#### **Nombre(s) de pila + Apellido**

- Megan Marie Andrews (Europeo)
- Fereshteh Gholamzadeh (Iraní)
- Rattima Nitisaroj (Tailandés)
- Hasan Incirlioglu (Turco)

#### **Apellido + Nombre de pila**

- Lim Yauw Tjin (Chino)
- Pak Mi-Ok (Coreano)
- Suzuki Ichiro (Japonés)

#### **Apellido + Segundo nombre + Nombre de pila**

- Trinh Van Thanh (Vietnamita)

#### **Nombre de pila + Nombre de pila del padre**

- Ahmed bin Eisa (algunas comunidades árabes)
- Abdurrahman Wahid (Indonesia)
- Mahmud bin Haji Basir (Malasia)

#### **Nombre de pila + Patronímico (Nombre del padre) + Apellido**

- Ivan Andreyevich Saratov (Rusia)
- Basimah Ali Al-Qallaf (algunos países árabes)

#### **Nombre tribal + Nombre religioso**

- WOUKO Philomene (Camerún)

#### **Sólo Nombre de pila**

- Sukarno (Indonesia)
- Habibullah (Afganistán)

#### **Referencia al nombre del hijo**

- Abu Hassan (que se traduce literalmente como *padre de Hassan*, países árabes)

### **Nombres conjuntos**

El término *nombre conjunto* hace referencia a dos o más personas dentro de una única estructura de nombre, en la que dos nombres distintos están enlazados por fórmulas de tratamiento (Mr. y Mrs. Smith), nombres de pila (John y Marie Smith), nombres completos (John Smith y Marie Smith) o alguna otra combinación de

elementos de nombre con conjunciones como *y*, *o* o signos de puntuación equivalentes, como el carácter ampersand (&).

Los nombres conjuntos se desglosan primero en nombres individuales, para poder aplicar más funciones de análisis y otras funciones del producto a cada nombre de persona. Los productos IBM InfoSphere Global Name Management (IBM NameWorks y NameParser) reconocen seis tipos de construcciones de nombres conjuntos:

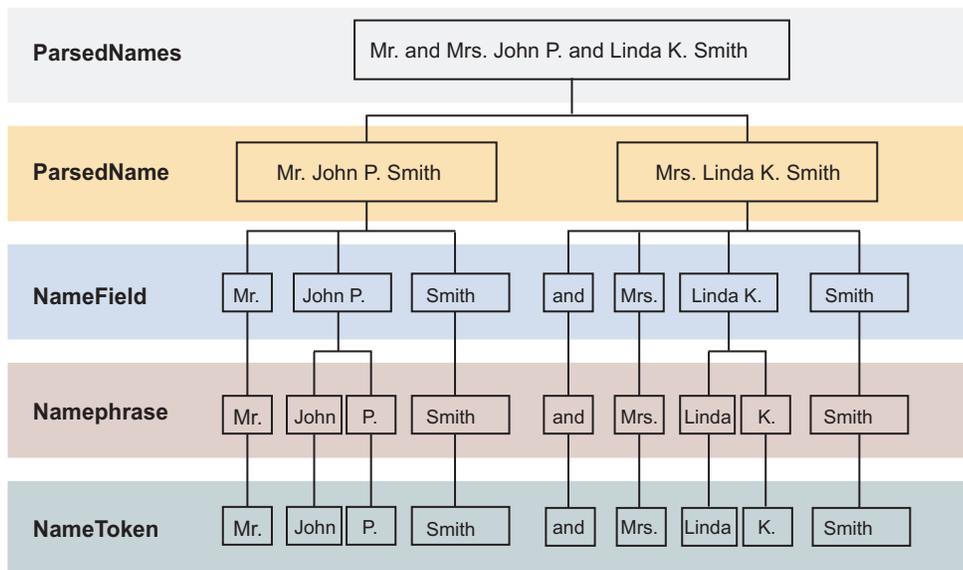
- Fórmulas de tratamiento conjuntas, como por ejemplo *Mr. and Mrs. John Smith*
- Nombres de pila conjuntos, como por ejemplo *John and Linda Smith*
- Pares de fórmula de tratamiento y nombre de pila conjuntos, como por ejemplo *Mr. John and Mrs. Linda Smith*
- Fórmulas de tratamiento y nombres de pila conjuntos en construcción paralela, como por ejemplo *Mr. and Mrs. John and Linda Smith*
- Nombres completos conjuntos, como por ejemplo *John Smith and Maria Jones*
- Combinaciones de las otras cinco construcciones de nombres conjuntos, como por ejemplo *John and Linda Smith and Bob and Maria Jones*

Por omisión, el recurso de análisis de los productos IBM InfoSphere Global Name Management reconoce los nombres conjuntos por la presencia de las conjunciones *and*, *or* o el carácter &. Los nombres que no contienen uno de estos indicadores no se tratan como nombres conjuntos. Por ejemplo, la serie *John Smith Marie Smith* no se desglosará en dos nombres distintos. Además, las listas de nombres delimitados por comas (por ejemplo *Bob, Karen, and David Smith*) no se desglosan en sus partes constituyentes, aunque esté presente una conjunción.

El recurso de análisis acepta lista personalizadas de caracteres y palabras, como por ejemplo la barra inclinada invertida (\), *et* o *y*, que deben tratarse como conjunciones. Puede añadir estos caracteres por medio de la lista de cadenas de caracteres externa.

### **Ejemplo de nombre conjunto**

El ejemplo siguiente muestra el árbol de análisis creado para una construcción de nombre conjunto formada por dos fórmulas de tratamiento conjuntas y dos nombres completos conjuntos. Primero se separan los nombres individuales, y luego cada nombre completo individual se desglosa en sus componentes constituyentes.



### Conceptos relacionados:

"Árboles de análisis" en la página 19

NameParser crea un árbol de análisis a partir del resultado del análisis de los patrones de estructura y distribución de un nombre de entrada. El *árbol de análisis* es una jerarquía que agrupa los elementos de un nombre en unidades estructurales, empezando por las cadenas de caracteres individuales (series delimitadas por espacios o signos de puntuación), que pueden combinarse en frases de nombre, que a su vez se combinan para formar un nombre de persona completo.

## Nombres de organización

Un *nombre de organización* es un nombre no de persona que hace referencia a un cuerpo estructurado formado por una o más personas cuyo objetivo es realizar alguna función común. Las organizaciones pueden ser empresas, clubes, escuelas, agencias gubernamentales, partidos políticos o manifestaciones de la World Wide Web. Los nombres de organización incluyen generalmente algún tipo de indicador, patrón o palabras que ayuden a identificarlos como nombres no de persona.

Por lo general, aunque no siempre, los nombres de organización incluyen alguna palabra o frase que indica su función, como por ejemplo "escuela secundaria", "fontanería", "departamento de policía" o "banco".

Los nombres de organización también contienen un *elemento de denominación* o alguna serie de caracteres, palabras o frases que identifica de forma exclusiva esta organización entre otras del mismo tipo. Por ejemplo, "First Union Bank," "Joe's Italian Restaurant," "AAA Auto Wash." Algunas organizaciones, como por ejemplo las empresas, están reguladas por los gobiernos y tienen elementos de nombre prescritos que indican su estado de registro, como por ejemplo "PTY" o "LTD".

Los tipos de cadenas de caracteres y combinaciones de las mismas que se encuentran en los nombres de organización no se parecen generalmente ni utilizan los mismos patrones que las de los nombres de persona. Estos patrones corresponden a códigos (denominados códigos de razón de categoría de nombres) que identifican la razón por la que un nombre se ha clasificado como nombre de organización, en lugar de como nombre de persona. Estos códigos de razón no

definen un nombre de organización, sino que indican patrones que no se esperarían en un nombre de persona. Por ejemplo, una serie de tres consonantes idénticas en una fila (por ejemplo “DDD”) sería muy inusual en un nombre de persona, pero no es inhabitual en los nombres de organización.

Cuando los componentes de IBM InfoSphere Global Name Management categorizan un nombre, si éste coincide con uno o más códigos de razón de categoría de nombres, se presupone que es un nombre de organización. De lo contrario, es candidato a nombre de persona.

---

## Componentes del nombre

Los nombres de persona están formados por una o varias palabras combinadas en estructuras acordes a reglas específicas de idioma y convenios de entorno cultural.

### Árboles de análisis

NameParser crea un árbol de análisis a partir del resultado del análisis de los patrones de estructura y distribución de un nombre de entrada. El *árbol de análisis* es una jerarquía que agrupa los elementos de un nombre en unidades estructurales, empezando por las cadenas de caracteres individuales (series delimitadas por espacios o signos de puntuación), que pueden combinarse en frases de nombre, que a su vez se combinan para formar un nombre de persona completo.

Durante el proceso, se impone a un nombre una estructura de datos de dos campos que divide las frases de nombre en campos de nombre de pila y apellido, en función de patrones de distribución estadística. El objeto que se encuentra en la raíz del árbol representa toda la serie de entrada, sus hijos representan las subdivisiones mayores del nombre, los hijos de éstas representan subdivisiones de las subdivisiones, y así sucesivamente.

El ejemplo siguiente muestra los elementos estructurales de un nombre bastante complejo, que empieza por la fila inferior con las cadenas de caracteres de nombre individuales, sube un nivel para capturar las frases de nombre y luego sube otro nivel para mostrar la división en campos de nombre. La forma de nombre original se muestra en la parte superior del árbol como **ParsedName**. Este árbol de análisis representa las estructuras reconocidas por NameParser.

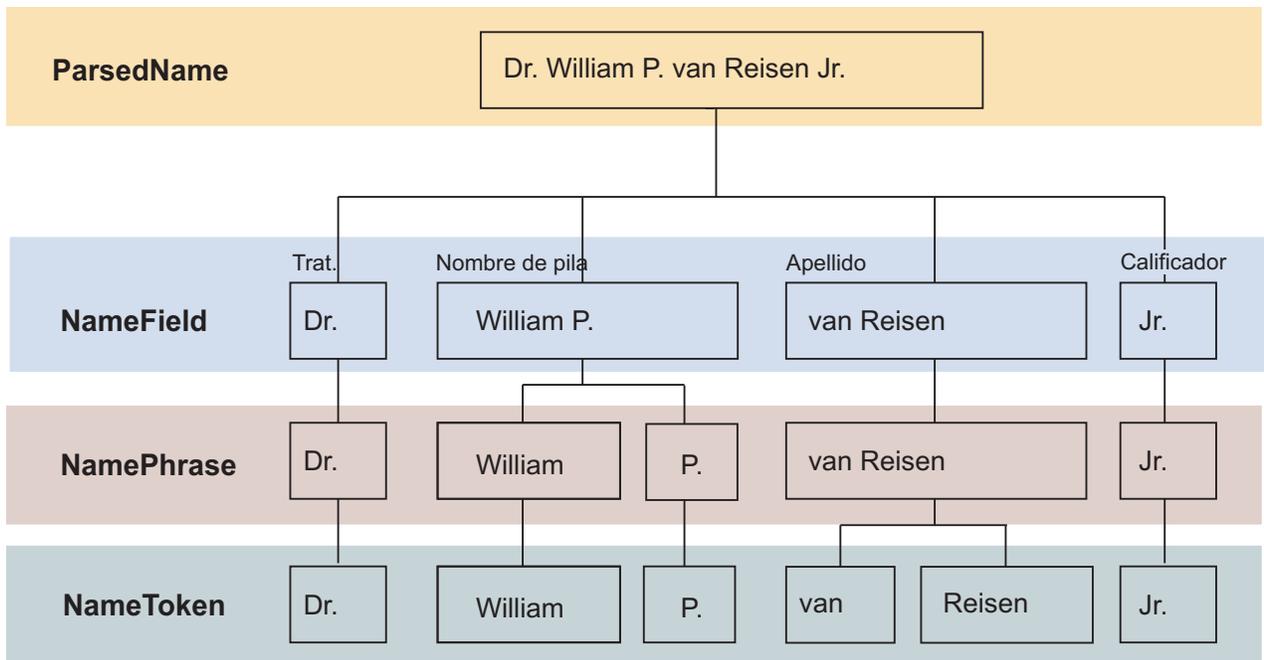


Figura 2. Ejemplo de árbol de análisis de NameParser

#### Conceptos relacionados:

“Nombres conjuntos” en la página 16

El término *nombre conjunto* hace referencia a dos o más personas dentro de una única estructura de nombre, en la que dos nombres distintos están enlazados por fórmulas de tratamiento (Mr. y Mrs. Smith), nombres de pila (John y Marie Smith), nombres completos (John Smith y Marie Smith) o alguna otra combinación de elementos de nombre con conjunciones como *y*, *o* o signos de puntuación equivalentes, como el carácter ampersand (&).

“Nombres analizados”

El análisis de nombres es el proceso de organizar las cadenas de caracteres de un nombre en las unidades estructurales más grandes que las contienen. Algunas de estas estructuras, como las frases de nombre, son estructuras gramaticales naturales que se encuentran en el idioma origen del nombre. El análisis de nombres se realiza mediante la utilización de información estadística y reglas basadas en lingüística para reconocer las estructuras sintácticas de los nombres.

## Nombres analizados

El análisis de nombres es el proceso de organizar las cadenas de caracteres de un nombre en las unidades estructurales más grandes que las contienen. Algunas de estas estructuras, como las frases de nombre, son estructuras gramaticales naturales que se encuentran en el idioma origen del nombre. El análisis de nombres se realiza mediante la utilización de información estadística y reglas basadas en lingüística para reconocer las estructuras sintácticas de los nombres.

Las operaciones de proceso de nombres de nivel superior, como por ejemplo la búsqueda y comparación, devuelven los mejores resultados cuando cada componente de un nombre se maneja según su valor informativo relativo. Es decir, las cadenas de caracteres como, por ejemplo, las raíces de nombre, cuyo valor de contenido es alto, obtienen más peso durante la búsqueda que las partículas gramaticales como los prefijos, sufijos o cadenas de caracteres como las fórmulas de tratamiento, que son externas al nombre en sí. De igual modo, los componentes

de un nombre que representan el nombre de pila o el apellido deben manejarse en paralelo con otros nombres de pila o apellidos. Para poder determinar el papel que juegan los diversos componentes de un nombre, éste debe analizarse.

Otras estructuras del nombre, como los campos de nombre, son estructuras de datos artificiales que pueden (o no) corresponder a estructuras semánticas o sociales reconocidas por una comunidad cultural o lingüística. Por ejemplo, en el entorno cultural norteamericano, muchas personas tienen un nombre de pila, un segundo nombre y un apellido. Un nombre como *Karen Lee van der Meer* consta de los componentes siguientes:

- Una frase de nombre, formada por la raíz *Karen*, que actúa como nombre de pila.
- Una frase de nombre, formada por la raíz de nombre *Lee*, que actúa como segundo nombre.
- Una frase de nombre, *van der Meer*, formada por dos prefijos y la raíz *Meer*, que actúa como apellido.

Este nombre no corresponde claramente a una estructura de datos de dos campos, ya que no se reconoce el estado diferenciador del segundo nombre.

Los productos IBM InfoSphere Global Name Management utilizan una estructura de desglose jerárquico en la que las cadenas de caracteres (series de caracteres delimitados por espacios o signos de puntuación) se agrupan en estructuras que se producen de forma natural conocidas como frases de nombre, que, a su vez, se agrupan en elementos de un campo de nombre de pila y un campo de apellido, en función de la distribución estadística de frases de nombres y patrones de nombres específicos del entorno cultural. Los dos campos de nombre, junto con las fórmulas de tratamiento o calificadores que aparezcan en el nombre de entrada original, forman el nombre completo.

#### **Conceptos relacionados:**

“Árboles de análisis” en la página 19

NameParser crea un árbol de análisis a partir del resultado del análisis de los patrones de estructura y distribución de un nombre de entrada. El *árbol de análisis* es una jerarquía que agrupa los elementos de un nombre en unidades estructurales, empezando por las cadenas de caracteres individuales (series delimitadas por espacios o signos de puntuación), que pueden combinarse en frases de nombre, que a su vez se combinan para formar un nombre de persona completo.

## **Campos de nombre**

Un *campo de nombre* es una estructura de datos artificial impuesta a los nombres para facilitar el proceso de los datos. Muchas bases de datos dividen los nombres en dos campos, generalmente correspondientes al nombre de pila y al apellido, aunque algunas especifican los nombres en un único campo y otras pueden utilizar tres o más campos. Los campos de nombre son la primera rama de un árbol de análisis de nombres de persona.

Los productos IBM InfoSphere Global Name Management utilizan una estructura de nombres de persona de dos campos, denominados nombre de pila y apellido, con dos campos adicionales para fórmulas de tratamiento y calificadores. Cada campo tiene su propio indicador de tipo de campo:

#### **Apellidos**

El apellido (o SN) es el componente del nombre que generalmente, aunque no necesariamente, es común a un grupo de personas, como por ejemplo

una familia, tribu o clan. En ciertas áreas del mundo, algunos apellidos pueden ser exclusivos de un individuo, como por ejemplo los que indican una característica personal o una profesión.

Los apellidos son elementos portadores de contenido clave de un nombre de persona. Sin embargo, no todas las personas utilizan apellidos. En partes de Indonesia, por ejemplo, la mayoría de personas tiene sólo un nombre de pila.

### **Nombres de pila**

Los nombres de pila (o GN) son el componente de un nombre que identifica de forma exclusiva un individuo con respecto a otros miembros de la familia o del grupo. En un nombre anglosajón, el primer y segundo nombre son nombres de pila.

Los nombres de pila son el único elemento de nombre conocido como requisito de denominación universal en todas las culturas del mundo. No todas las culturas tienen elementos de apellido, pero todas asignan un nombre de pila a los individuos.

### **Fórmulas de tratamiento**

Las fórmulas de tratamiento son palabras o frases externas al nombre en sí, pero que indican algún tipo de información relativa al poseedor del nombre. Las fórmulas de tratamiento indican el estado civil, el orden de nacimiento, la cualificación formativa o profesional, la confesión religiosa, el rango social u otra información. Las fórmulas de tratamiento no se procesan igual que los elementos centrales del nombre, ya que tienden a ser opcionales y es posible que no siempre aparezcan en un nombre.

### **Calificadores**

Los calificadores son términos o frases que se añaden a un nombre de persona para distinguirlo mediante la especificación de una posición generacional (como Junior o Senior, o "fils" en francés), un logro o distinción que la persona ha obtenido (por ejemplo, Ph.D o doctorado) o una cualificación de algún tipo (como D.D.S. u odontólogo). Habitualmente, los calificadores se especifican a continuación del nombre. Al igual que las fórmulas de tratamiento, los calificadores no se consideran parte de un nombre de persona.

### **Conjunciones precedentes**

Si un nombre contiene una construcción de nombres conjuntos, el último elemento de la construcción debe ir precedido de una conjunción que una el nombre con el que le precede en la serie de entrada. Por ejemplo, la serie de entrada *John and Mary Smith* es una construcción de nombres conjuntos que representa dos nombres: *John Smith* y *Mary Smith*. El nombre *Mary Smith* tiene la conjunción precedente *and*. NameParser permite conservar esta conjunción en un campo propio.

El motor de búsqueda de NameHunter incorpora las fórmulas de tratamiento al campo de nombre de pila y los calificadores al campo de apellido. Estos elementos que no forman parte del nombre se manejan de forma diferente durante las operaciones de búsqueda y comparación, de acuerdo con la configuración de los parámetros de búsqueda de NameHunter. En cambio, productos similares conservan campos separados para los elementos que no forman parte del nombre.

## **Frases de nombre**

Una frase de nombre es una cadena de caracteres o una secuencia de cadenas de caracteres que forma una única estructura lingüística, análoga a una frase

sustantiva o preposicional en un idioma. Las frases de nombre constan de una o varias raíces de nombre y de prefijos, sufijos, conjunciones u otros elementos gramaticales relacionados con las raíces.

Por ejemplo, la frase de nombre en español *de la Cruz* contiene una raíz de nombre (Cruz) y dos prefijos (de y la). La frase de nombre en chino *Mei-hui* contiene dos raíces, y el nombre inglés *Smith* es una frase de nombre que consta sólo de una raíz. Las frases de nombre forman el nivel intermedio del árbol de análisis y agrupan cadenas de caracteres individuales en estructuras mayores. Sin embargo, no constituyen necesariamente un campo de nombre, que puede contener varias frases de nombre.

IBM NameWorks y NameParser utilizan un repositorio interno de información sobre frases de nombres derivadas de un conjunto de datos originales, de más de 800 millones de nombres de prácticamente todos los países del mundo. Este repositorio suministra información a todas las operaciones de proceso de nombres, incluidas las de análisis, generación de variantes, clasificación y búsqueda.

## Cadenas de caracteres de nombre

*Cadenas de caracteres de nombre* son los elementos indivisibles más pequeños de un nombre que se componen de series de caracteres delimitadas por un "espacio en blanco" o signos de puntuación.

Las cadenas de caracteres son generalmente afijos o raíces, aunque a veces pueden ser frases de nombre completas en los casos en que las unidades gramaticales más pequeñas de un nombre se escriban como una sola palabra. Por ejemplo, la frase de nombre completa *de la Cruz* está formada por los afijos *de* y *la* y la raíz *Cruz*. Las cadenas de caracteres de nombre son el último nivel de un árbol de análisis, a veces conocidas como nodos hoja. La función de una cadena de caracteres de nombre depende de su contenido y de su posición en relación a otros elementos del nombre. Una cadena de caracteres como *de*, por ejemplo, puede ser un prefijo, un nombre de pila o un apellido en función del origen idiomático del nombre en el que aparece y de su posición dentro del nombre.

Las cadenas de caracteres de nombres pueden constar de una única o varias cadenas de caracteres. Por ejemplo, el término de múltiples cadenas de caracteres *Sociedad de responsabilidad limitada*.

Los tipos de cadenas de caracteres de nombre son:

- Prefijos
- Sufijos
- Fórmulas de tratamiento
- Calificadores
- Designador de organización
- Calificador profesional
- Paradas ignoradas
- Afijos de la organización
- Iniciales
- Conjunciones
- Raíces de nombre
- Términos

## **Fórmulas de tratamiento, afijos y calificadores (TAQ)**

Los productos IBM InfoSphere Global Name Management utilizan una lógica especial para otorgar puntuaciones a las fórmulas de tratamiento, afijos y calificadores, conocidos colectivamente como *TAQ*. Las fórmulas de tratamiento y calificadores son elementos opcionales que normalmente no son parte integrante del nombre. Los afijos, aunque gramaticalmente forman parte del nombre, tienen poco valor en cuanto a contenido y, por tanto, se les otorga menos peso en las operaciones de comparación de nombres.

### **Fórmulas de tratamiento**

Las fórmulas de tratamiento son términos destinados a dirigirse a una persona que generalmente preceden al nombre y pueden indicar educación, posición social o cualificación profesional.

Ejemplos de fórmulas de tratamiento:

- Dr. por Doctor
- Mr. por Mister
- Hajj para las personas de credo islámico que han peregrinado a La Meca

### **Afijos**

Los afijos son prefijos o sufijos asociados a un nombre. Aunque gramaticalmente forman parte del nombre, generalmente tienen poco valor en cuanto a contenido y, por tanto, se les otorga menos peso en las operaciones de comparación de nombres.

Ejemplos de afijos de nombres de persona:

- *de la* en "de la Torre"
- *van der* en "van der Meer"
- *Abdul* en "Abdul Rahman"
- *Al Din* en "Nur Al Din"

### **Calificadores**

Los calificadores son términos o frases que se añaden añadidas a un nombre de persona para distinguirlo mediante la especificación de una posición generacional, un logro o distinción que la persona ha obtenido o una cualificación de algún tipo. Habitualmente, los calificadores se especifican a continuación del nombre y por lo general no se consideran parte del nombre real.

Ejemplos de calificadores:

- Jr. por Junior
- Sr. por Senior
- Esq. por Esquire
- PhD por Doctor of Philosophy
- D.D.S por Doctor of Dental Science

### **TAQ de nombres de organización**

Es necesario un proceso especial para los términos que se encuentran únicamente en los nombres de organización, que, para ser coherentes, se denominan *TAQ*. Aunque las fórmulas de tratamiento no estén presentes; si que lo están determinados calificadores y afijos.

Estos tipos de cadenas de caracteres no se consideran elementos significativas de un nombre de organización. Por tanto, la totalidad de los siguientes TAQ se listan en el archivo de TAQ (taq.ibm), por lo que no se tratan como raíces de nombre en las comparaciones realizadas por NameHunter. Estos términos están sujetos a los siguientes factores predeterminados de TAQ asignados en el archivo de TAQ.

Se aceptan varias cadenas de caracteres soportadas para reconocer términos con un espacio en blanco intercalado. Por ejemplo, Sociedad de responsabilidad limitada.

*Tabla 2. Ejemplo de TAQ de nombres de organización y factores predeterminados de TAQ*

Tipo de TAQ	Ejemplo	Factor de TAQ diferente (predeterminado)	Factor de TAQ ausente (predeterminado)
Designador de organización (OD)	COMPANY, CORP, LLC, LIMITED LIABILITY CORPORATION, COUNTRY CLUB	.98	.99
Palabra ignorada (SW)	OF, THE	1.0	1.0
Afijo de organización (OA)	DE, DEL, LA, LAS	.97	.98

### **Designador de organización:**

Un designador de organización (OD) suministra información relativa al registro legal de una organización, pero generalmente no es un componente esencial del nombre de organización. Por esta razón, a menudo el OD se ignora por completo o puede aparecer en diversos formatos.

Por ejemplo, todos los nombres siguientes hacen referencia a la misma organización:

- 20TH CENTURY PRODUCTIONS INCORPORATED
- 20TH CENTURY PRODUCTIONS INCORP
- 20TH CENTURY PRODUCTIONS INC
- 20TH CENTURY PRODUCTIONS

Cuando NameHunter compare los dos últimos nombres de esta lista, se aplicará el factor TAQ ausente, asignando una penalización de .99 a una coincidencia que, por lo demás, es perfecta. Sin embargo, las comparaciones entre los tres primeros nombres no deben estar sujetas al factor de TAQ diferente, ya que estos OD son equivalentes entre sí (INCORPORATED, INCORP e INC tienen el mismo significado). De hecho, la mayoría de OD tienen formas variantes intercambiables. Por ejemplo, los siguientes OD son sinónimos:

- CO/COMPANY
- JSC/JOINT STOCK COMPANY
- LLC/ LIMITED LIABILITY COMPANY

El reconocimiento de variantes de OD se maneja mediante un conjunto de reglas de regularización que se aplican tanto al nombre de la consulta como al nombre de la lista de datos. Cada regla convierte una forma de OD de variante a una forma regularizada predeterminada, y estas formas regularizadas se incluyen todas en el archivo de TAQ. Por ejemplo una regla de regularización convierte INC e INCORP

a INCORPORATED, que luego se incluye en el archivo de TAQ. Los tres primeros nombres de la lista anterior se regularizan luego como 20TH CENTURY PRODUCTIONS INCORPORATED y, por tanto, reciben una puntuación de 1.0 durante la comparación. Esta solución se ha adoptado para reducir la ambigüedad y poder manejar TAQ de varias cadenas de caracteres.

### *Ambigüedad*

Muchos OD también pueden ser elementos de nombre significativos. Por ejemplo, aunque LLC es con frecuencia un OD que indica LIMITED LIABILITY COMPANY, también existen nombres de organización, por ejemplo LLC COUNSELING SERVICES, en los que LLC es una raíz de nombre que distingue esta organización en particular de otras organizaciones del mismo tipo (por ejemplo, FRANKLIN COUNSELING SERVICES).

Si LLC se incluyera simplemente en el archivo de TAQ, cualquier aparición de LLC se manejaría como un elemento de TAQ durante las comparaciones de nombres realizadas por NameHunter. Un nombre como LLC COUNSELING coincidiría así con todos los demás nombres de organización de la lista de datos que contuvieran la frase COUNSELING SERVICES. La regularización ofrece un modo de ser sensible a la posición de términos tales como LLC; es decir, si este término aparece al final de un nombre, posiblemente se trate de un OD, y la regla de regularización indicará: LLC = LIMITEDLIABILITYCOMPANY. Si el término aparece al principio de un nombre, se tratará probablemente de una raíz de nombre.

### *TAQ de varias cadenas de caracteres*

Muchos OD contienen varias cadenas de caracteres, como por ejemplo LIMITED LIABILITY COMPANY. Sin embargo, el motor de búsqueda no da soporte a archivos TAQ y de variantes que contienen entradas de varias cadenas de caracteres. La regularización ofrece un modo de convertir una serie de varias cadenas de caracteres en una única cadena de caracteres que puede incluirse en el archivo de TAQ. Debido a la regularización, los nombres siguientes reciben una puntuación de comparación perfecta de 1.0 del motor de búsqueda cuando se comparan.

Series de caracteres originales	Series de caracteres regularizadas
BISON LEGACY LLC	BISON LEGACY
BISON LEGACY LIMITED LIABILITY COMPANY	LIMITEDLIABILITYCOMPANY

Con respecto a los nombres de organización, NameHunter sólo compara el nombre de consulta regularizado con los nombres de la lista de datos regularizados. Los nombres originales de la consulta y de la lista de datos no se comparan. En el ejemplo siguiente, el nombre de consulta no resulta afectado por las reglas de regularización de OD, porque LLC aparece al principio del nombre. LLC está regularizado en la lista de datos porque aparece al final del nombre.

Nombre de consulta	Nombre de lista de datos
LLC CONSULTING SERVICES	BEVERLY CONSULTING SERVICES
	LIMITEDLIABILITYCOMPANY

Esta comparación devuelve una puntuación de comparación baja, ya que las series LLC y BEVERLY no coinciden, y LIMITEDLIABILITYCOMPANY desencadena la

aplicación de una penalización de TAQ. Este resultado es correcto, ya que estas organizaciones son diferentes. Sin embargo, si también se hubiera mantenido la forma original como buscable, se hubieran comparado los nombres siguientes:

Nombre de consulta	Nombre de lista de datos
LLC CONSULTING SERVICES	BEVERLY CONSULTING SERVICES LLC

El archivo TAQ contiene sólo formas regularizadas, como LIMITEDLIABILITYCOMPANY, y no formas originales como LLC. Por tanto, en este caso ninguna instancia de LLC se reconocería como TAQ, y las dos apariciones de LLC se compararían como raíces. Esta comparación devolvería una puntuación relativamente alta, ya que tres de sus cuatro cadenas de caracteres son idénticas. Evitar comparaciones de nombres de organización no regularizados evita esta problema.

Además de la variación de OD que se maneja mediante la regularización, en ocasiones la variación se maneja mediante variantes. A diferencia de un par como, por ejemplo, INC/INCORPORATED (cuyos elementos son equivalentes y debe recibir una puntuación perfecta), existen otros pares que son más variantes que equivalentes. En estos casos, un par de variantes se incluye en el archivo TAQ con una puntuación y un entorno cultural asignados (que son siempre 0 en los nombres de organización). Por ejemplo, una entrada podría ser la siguiente: COMPANY, COMPANIES, .99, 0. Debido a este conjunto de variantes, un par que incluya las series COMPANY y COMPANIES recibirá una puntuación de comparación más alta que la que hubiera obtenido de estar sujeto al factor de TAQ diferente.

#### Palabra ignorada:

Una palabra ignorada es aquella que no añade ningún significado a un nombre de organización y, por tanto, no se incluye en ninguna comparación ni puntuación de nombres. El release actual de IBM InfoSphere Global Name Management incluye sólo OF y THE como palabras ignoradas.

El valor de factor TAQ predeterminado de las palabras ignoradas es 1.0, que hace que estas palabras se ignoren en las comparaciones de nombre. Por lo tanto, una puntuación perfecta se consigue cuando se compara los siguientes conjuntos de nombre:

*Tabla 3. Conjuntos de nombre que contienen palabras ignoradas*

Nombre de organización	Nombre de organización con palabra ignorada
ACME INKS CANADA INC.	ACME INKS OF CANADA INC.
TRACTOR & ENGINEERING CO.	THE TRACTOR & ENGINEERING CO.
ASSOCIATION OF AMERICAN GEOGRAPHERS	THE ASSOCIATION AMERICAN GEOGRAPHERS

#### Afijo de organización:

Un afijo de organización (OA) es una palabra (como las palabras ignoradas) que no añade en realidad ningún significado a un nombre de organización. Sin embargo, la presencia o ausencia de un OA, especialmente en otros idiomas, tiende a tener

más significado que una palabra ignorada, por lo que los factores de TAQ asignan una pequeña penalización cuando un OA es diferente o está ausente.

Los OA incluyen la conjunción inglesa AND, así como diversos artículos no ingleses que implican información de número y género. Si estas cadenas de caracteres están ausentes; por ejemplo, en los siguientes pares de nombres; se aplica el factor TAQ siguiente. Cuando se comparan OA diferentes, se aplica el factor TAQ diferente.

Tabla 4. Nombres de organización con afijo de organización

Nombre de organización	Nombre de organización con OA
AAI ENGINEERING SALES SERVICE	AAI ENGINEERING SALES AND SERVICE
BUTCHER HAUS	DAS BUTCHER HAUS
BANCO ORO	BANCO DE ORO

### Cadenas de caracteres raíz

Una *cadena de caracteres raíz* (también denominada raíz de nombre) es un elemento de nombre que puede aparecer de forma autónoma o combinado con afijos para formar una frase de nombre. Por ejemplo, la frase de nombre *de la Torre* es un apellido hispánico que consta de la raíz de nombre *Torre* precedida de dos afijos.

Una raíz de nombre sin afijos asociados, por ejemplo, *Robert* o *Gonzales*, es una frase de nombre por sí misma. Generalmente, un campo de nombre de pila o de apellido contiene una o varias frases de nombre. La mayoría de las frases de nombre en el campo de nombre de pila en las culturas coreana y china, y en la mayoría de los nombres de pila comunes de las culturas francesa, alemana e hispánica, se componen de dos raíces de nombre, por ejemplo *Shu Dong*, "*Eun Jung*" o "*Jean Luc*."

---

## Listas de nombres

Los productos IBM InfoSphere Global Name Recognition utilizan varios tipos de listas de nombres para procesar los nombres.

### Name data archive

Name data archive es una colección de aproximadamente un billón de nombres de todo el mundo, junto con el género y el país de asociación de cada nombre. Este gran repositorio de información de nombres potencia la capacidad de los algoritmos y reglas que los productos IBM InfoSphere Global Name Management utilizan para categorizar, clasificar, analizar, establecer el género y comparar nombres.

Las cuentas de frecuencia de cadenas de caracteres de nombres individuales y frases de nombre extraídas del name data archive forman la base de los algoritmos estadísticos y de cálculo que los productos IBM InfoSphere Global Name Management utilizan para analizar nombres. Por ejemplo, el componente de análisis de nombres utiliza estas estadísticas para calcular una puntuación de validez para una combinación determinada de campos de nombre de pila y apellido. Una puntuación de validez baja puede indicar que los nombres se han colocado incorrectamente en los campos, lo cual llevará al motor de análisis a sugerir una o varias combinaciones alternativas más probables.

## Lista de cadenas de caracteres externa

Las listas de cadenas de caracteres externas son archivos a los que puede añadir fórmulas de tratamiento, afijos, calificadores o raíces de nombre para complementar la información de la base de datos interna de IBM InfoSphere Global Name Management . Durante el análisis o la puntuación de nombres, antes que en la base de datos interna se busca primero en las listas de cadenas de caracteres personalizadas.

Por omisión, tanto la lista de cadenas de caracteres externa de NameParser como la lista de cadenas de caracteres personalizada de IBM NameWorks están vacías. Puede acceder a la lista de cadenas de caracteres personalizada de IBM NameWorks por medio de su archivo de configuración.

### Referencia relacionada:

Lista de cadenas de caracteres externa NameParser

Puede añadir cadenas de caracteres a la lista de cadenas de caracteres externa para personalizar el comportamiento de NameParser. La lista de cadenas de caracteres externa es una lista suplementaria de datos de cadena de caracteres en la que NameParser busca cadenas de caracteres en esta lista suplementaria de datos de cadena de caracteres. Si no encuentra ninguna, NameParser busca sus propias tablas internas que incluyen información sobre millones de cadenas de caracteres de nombres.

---

## Transliteración de nombres

La *transliteración de nombres* es el proceso de conversión de un nombre desde un sistema de escritura o convenio de codificación de caracteres determinado a otro. Por ejemplo, la transliteración de nombres permite analizar y comparar un nombre escrito en árabe con un nombre similar escrito en el alfabeto latino.

La *transliteración* suele confundirse con la *traducción*. Traducción es transmitir en otro idioma el significado de algo escrito o hablado en un idioma. Por ejemplo, "a horse" en inglés es "un cheval" en francés. La transliteración es transferir los sonidos representados por una ortografía (sistema de escritura) a la forma en que esos sonidos se representan en otra ortografía diferente. Así, el carácter chino que se utiliza para escribir la forma de pronunciar la palabra caballo en chino mandarín se translitera como "ma" en el alfabeto latino.

Los productos IBM InfoSphere Global Name Management contienen soporte incorporado para la comparación de nombres en diversos sistemas de escritura, incluidos el árabe, el griego, el cirílico, el devanagari, el kana, el hangul, el hanzhi y el latino ampliado, es decir, el alfabeto latino estándar con caracteres adicionales y marcas diacríticas que se utiliza en muchos idiomas europeos y asiáticos.

Los nombres se transliteran como primer paso de los procesos de análisis y puntuación de nombres. NameTransliterator convierte los nombres desde su codificación nativa a la codificación ASCII como paso de preproceso antes de analizar, clasificar u otorgar una puntuación a los nombres. Muchos métodos de IBM NameWorks están diseñados para realizar primero la transliteración, antes de realizar ninguna de las demás funciones del método.

El soporte para la transliteración de Kanji a Kana se proporciona en un paquete Java independiente que se puede utilizar junto con las interfaces de servicio web o Java de IBM NameWorks. El paquete com.ibm.gnr.java está descrito en la sección de referencia de la API de Java.

## Archivos de reglas de transliteración

Los archivos de reglas de transliteración son archivos binarios cifrados que permiten manejar un sistema de escritura o alfabeto determinado. Cada archivo de reglas posibilita que los componentes de IBM InfoSphere Global Name Management manejen la información de entrada de un script determinado. El texto se convierte en caracteres ASCII en mayúsculas en formatos adecuados para el análisis y puntuación de nombres.

Puede utilizar los siguientes archivos de reglas, o módulos, para ampliar las funciones básicas del transcriptor de nombres:

### **arabicTransRule.ibm**

Para nombres de persona escritos en script árabe.

### **chineseTransRule.ibm**

Para nombres de persona escritos en script hanzi.

### **chineseOnTransRule.ibm**

Para nombres de organización escritos en script hanzi.

### **cyrillicTransRule.ibm**

Para nombres de persona escritos en script cirílico.

### **cyrillicOnTransRule.ibm**

Para nombres de organización escritos en script cirílico.

### **greekTransRule.ibm**

Para nombres de persona escritos en script griego.

### **hindiTransRule.ibm**

Para nombres de persona escritos en script devanagari.

### **hindiOnTransRule.ibm**

Para nombres de organización escritos en script devanagari.

### **japaneseTransRule.ibm**

Para nombres de persona escritos en script kana (katakana e hiragana).

### **japaneseOnTransRule.ibm**

Para nombres de organización escritos en script kanji o kana (katakana e hiragana).

### **koreanTransRule.ibm**

Para nombres de persona escritos en script hangul.

### **koreanOnTransRule.ibm**

Para nombres de organización escritos en script hangul.

### **latinTransRule.ibm**

Para nombres de persona y de organización escritos en latín básico o en cualquier ampliación del latín. Este módulo se incorpora en *NameWorks* y no es necesario mencionarlo en los archivos de configuración.

## Visión general de la transliteración de chino

IBM InfoSphere Global Name Management busca coincidencias de nombres de persona del hanzi chino en sus formas hanzi exactas y equivalentes con las equivalencias de nombres latinos transliterados.

Los idiomas chino y japonés escritos comparten muchos caracteres. Los nombres del chino escritos en caracteres hanzi son parecidos a los nombres japoneses

escritos en kanji con muchos caracteres compartidos. Los nombres hanzi en chino presentan las siguientes características, que los diferencian de los nombres kanji:

- Hay pocos apellidos chinos de varios caracteres, que significa que la mayoría de los nombres completos en chino no son ambiguos con sólo un apellido y un nombre de pila para analizar.
- Hay pocos caracteres con múltiples lecturas (sin contar las variaciones tonales) en chino mandarín moderno.
- En caso de caracteres con múltiples lecturas, se presupone la más común para los nombres de persona. Existe un conjunto pequeño de caracteres que tiene una pronunciación específica de apellido. Debido a esto, normalmente no se proporciona ayuda de pronunciación para nombres de persona chinos en el uso normal.

Estas características implican que casi todos los nombres chinos tienen una única pronunciación en chino mandarín. Debido a esto, los nombres chinos pueden transliterarse directamente con el componente NameTransliterator. Esto difiere de los nombres japoneses, que a menudo tienen relaciones de tipo muchas a muchas entre nombres kanji y formas latinizadas.

El proyecto de código abierto ICU (International Components for Unicode) tiene un conjunto de reglas de sistema que transliteran caracteres chinos de uso común en representaciones pinyin mandarín. Cada carácter tiene sólo una forma de salida. En el caso de caracteres con varias pronunciaciones, se selecciona el más común. El proceso de transliteración de IBM InfoSphere Global Name Management utiliza el conjunto de reglas internas ICU para la mayoría de los caracteres chinos. Reglas especiales se encargan de gestionar las excepciones.

El proceso de nombres chinos requiere algo más que añadir reglas de transliteración. Todos los caracteres del chino mandarín son monosilábicos: se pronuncian como una sola sílaba. Hay alrededor de 1.760 de sílabas posibles en mandarín: 1.350 sílabas con tonos y 410 sílabas sin tonos. Sin embargo, hay decenas de miles de caracteres chinos, cada uno con un significado distinto. Esto significa que hay docenas de caracteres distintos con significados distintos que se pueden pronunciar exactamente igual. Por lo tanto, los nombres escritos con caracteres diferentes pueden transliterarse en la misma forma latina puesto que la romanización se basa en la pronunciación y no en el significado. En otras palabras, existe una relación de muchos a uno entre los nombres de caracteres chinos y las formas romanizadas. El problema surge cuando el nombre de la consulta es un nombre de caracteres chinos y la lista de datos contiene nombres de caracteres chinos distintos que se pronuncian del mismo modo y se han transliterado en la misma forma romanizada. Sin procedimientos de filtro adicionales, estos nombres de caracteres chinos diferentes se devuelven en una búsqueda de nombres como coincidencias perfectas.

Observe la siguiente lista que muestra cinco nombres diferentes, cada uno de los cuales tiene como mínimo un carácter distinto de los otros nombres:

1. 黄书东 - nombre escrito con el juego de caracteres simplificado
2. 黃書東 - mismo nombre que (1), pero escrito con el juego de caracteres tradicional
3. 黄书冬 - último carácter diferente en nombre de pila
4. 皇书东 - carácter de apellido diferente

## 5. 皇舒冬 - todos los caracteres diferentes.

Todos estos nombres se transliteran a la misma forma romanizada, principalmente "HUANG SHU DONG" (o HUANG2 SHU1 DONG1 si se incluyen marcas de tono numéricas). Sin embargo, sólo los nombres (1) y (2) son los mismo nombres en chino. Si todas estas formas romanizadas están la lista de datos, la consulta (1) "

黄书东 " también devolverá (3), (4) y (5) en 1.0 aunque todos sean nombres diferentes para un hablante nativo. El proceso de búsqueda de NameHunter se ha mejorado para tratar este tipo de resultado problemático.

### Gestión de los datos de nombres hanzi chinos

La función NameHunter analiza los datos de nombres hanzi chinos con el proceso general siguiente:

- La transliteración de nombres hanzi se realiza fuera de NameHunter. Los nombres hanzi se han de transliterar antes de enviarse a NameHunter.
- NameHunter requiere las equivalencias en nombres latinos y los datos de nombres hanzi originales.
- NameHunter compara primero los nombres de los formatos latinos y, a continuación, elimina los falsos positivos (que se pueden crear en las correlaciones hanzi a alfabeto latino de tipo muchos a uno) comparando los caracteres hanzi originales de las posibles comparaciones.

Algunas de sus funciones son:

- Reconocimiento de elementos de nombre de pila y apellido en nombres de persona escritos en script hanzi y proceso de estos elementos según corresponda.
- Comparación de formas de variantes de los mismos caracteres de hanzi.
- Comparación de equivalentes de nombres de persona hanzi y latinos.

La puntuación china se aplica según el criterio aceptable o erróneo, o ambos nombres contienen los mismos caracteres hanzi (o formas de variantes de los mismo caracteres) o falla la comparación. Si la comparación del script hanzi es correcta, se utiliza la puntuación generada para las correlaciones de nombres romanizados salvo si la puntuación romanizada es de 1,0 y la puntuación hanzi es inferior a 1,0 debido a valores no exactos provocados por variantes de caracteres. En este caso, se aplica una penalización de -0.02 a la puntuación romanizada por lo que se convierte en 0,98. Esto indica que las formas de nombres hanzi no son idénticas pero que, en cambio, contienen formas de variantes de los mismos caracteres.

El algoritmo de puntuación utiliza una tabla de variantes de chino que incluye chino simplificado y chino tradicional junto con otras variantes. La puntuación de variante más alta es .995. La tabla está en un formato similar a otras tablas de variantes de NameHunter y se puede ampliar. Por ejemplo, puede añadir conjuntos de caracteres que no son variables verdaderas pero que están relacionados entre sí (por ejemplo, se pronuncian de la misma forma y se escriben con rasgos similares) para evitar un resultado erróneo de la comparación si fueran caracteres totalmente distintos.

Los apellidos y nombres de pila en chino no están delimitados en el uso normal. Incluso los datos de nombres estructurados, como por ejemplo en un formulario de solicitud de residencia, generalmente tienen un único campo de nombre completo. El archivo de reglas de transliteración incluye un algoritmo de análisis mediante el

cual un nombre completo en caracteres chinos sin analizar se desglosa en un apellido y un nombre de pila antes de transliterarse. Este análisis es esencial para procesar nombres entre idiomas distintos y ayuda a proporcionar formas romanizadas correctas para los pocos caracteres de apellidos excepcionales que no siguen la pronunciación más común.

El análisis de los datos de nombres hanzi chino tiene las limitaciones siguientes:

- La transliteración del chino solamente genera resultados en pinyin mandarín.
- La comparación entre caracteres chinos sólo es posible si sus formas romanizadas coinciden en el umbral predefinido.
- Es poco probable que los nombres con caracteres hanzi similares pero que se pronuncian de forma distinta superen la comparación inicial, puesto que las formas latinas (que se basan en la pronunciación) probablemente no coincidan. La adición de este tipo de caracteres a la tabla de variantes de caracteres no es efectiva porque la búsqueda directa y la comparación de nombres de caracteres chinos no reciben soporte.

La transliteración del chino requiere los archivos `chineseTransRule.ibm` y `chineseOnTransRule.ibm`, el primero para los nombres de persona y el último para los nombres de organización. Es necesario realizar actualizaciones en los archivos de configuración para Distributed Search y NameWorks si se efectúa una migración de un release anterior del producto.

## Visión general de la transliteración de japonés

Los nombres en japonés nativo se escriben en el sistema autónomo han, conocido en Japón como kanji. Por lo general, estos caracteres son los mismos que los que se utilizan para escribir en chino.

Para un solo dialecto chino como, por ejemplo, el mandarín, un único carácter han tiene una pronunciación o lectura. Sin embargo, en japonés no hay una única lectura para un carácter o secuencia de caracteres. Los mismos caracteres kanji se pueden utilizar para escribir nombres completamente diferentes, con pronunciaciones completamente diferentes. Por este motivo, es posible que un registro que contenga un nombre de persona kanji contenga también la ortografía en kana de ese nombre, para aclarar qué lectura debe utilizarse.

Kana es un sistema de escritura basado en la pronunciación que se utiliza en japonés, no así un alfabeto. Cada símbolo Kana representa una sílaba diferente, normalmente una combinación de vocal y consonante. Hay dos "estilos" de kana, hiragana y katakana, ambos admitidos en GNM.

Los nombres extranjeros en el Japón se escriben en kana.

GNM gestiona los nombres de persona y los nombres de organizaciones en kanji de forma distinta. Los nombres de persona en kanji deben procesarse mediante un componente especial sólo para Java, Kanji Transliteration, que genera varias lecturas en kana para la cadena de entrada. A continuación, se puede pasar a GNM para su posterior proceso una lectura en kana de todas las lecturas posibles. Este método permite conseguir una coincidencia lo más amplia posible de los nombres japoneses latinizados.

Los nombres de organización en kanji se pasan directamente a través del módulo de transliteración de GNM. Cada carácter kanji se correlaciona coherentemente con una lectura, representado en el alfabeto latino.

Puesto que el chino y el japonés utilizan el mismo conjunto de caracteres han, los nombres en japonés deben marcarse como entorno cultural japonés para poderlos transcribir con lecturas japonesas. De forma predetermina, se supone que una serie de entrada en caracteres han se trata de un nombre en chino. Cualquier nombre escrito con caracteres han para el que no se haya especificado el entorno cultural pasará al módulo de transliteración chino.

---

## Capítulo 3. Analizar nombres (desglosar)

El análisis de nombres consiste en identificar las partes de los componentes de nombres y en separar aquellos componentes en campos de nombres de pila y campos de apellidos, así como en reconocer elementos que no son nombres como, por ejemplo, las fórmulas de tratamiento (Mr., Hajj, etc.) y calificadores (Jr, Esq., etc.). El análisis de nombres es un elemento clave del análisis y puntuación de nombres. El análisis preciso aumenta la posibilidad de que cada componente se analice correctamente, lo que lleva a obtener resultados de búsqueda más precisos.

---

### Análisis de nombres utilizando NameWorks

El desglose de nombres en componentes constituyentes resulta un paso clave en el análisis y puntuación de nombres. NameWorks integra la transliteración en métodos utilizados para desglosar nombres.

#### Desglosar nombres en componentes individuales

Utilice el método `parse()` para transliterar y desglosar uno o más nombres en componentes. Este método devuelve un árbol de análisis para cada nombre de la serie de entrada que proporciona información sobre los análisis alternativos para cada nombres y una puntuación que representa con qué probabilidad es correcto el análisis en función de las frecuencias heurísticas y relativas específicas del entorno cultural de frases de nombres de componentes como nombres de pila y apellidos.

#### Acerca de esta tarea

Utilice el método `parse()` de la clase `Analytics` de NameWorks para analizar un nombre de persona en los componentes de nombre de pila y apellido, así como para separar las formas de tratamiento y los calificadores del nombre. Si se ha habilitado, NameWorks devuelve información sobre los análisis alternativos cuando el grado de fiabilidad es suficientemente alto. Por ejemplo, si la consulta de búsqueda original es sobre el nombre de David, Robert, NameWorks puede devolver un elemento alternativo de Robert, David porque ambas cadenas de caracteres de nombre pueden servir igualmente como nombre de pila o apellido. Debe alcanzarse `alternateThreshold` para que NameWorks devuelva elementos alternativos. Para empezar a analizar un nombre, pase los valores siguientes al método `parse()`.

- El nombre completo que se va a analizar, representado como un nombre completo. Por ejemplo, ROBERT E JONES. El nombre debe pasarse al método `parse()` como un valor *string*.
- Un entero entre 0 y 100 que representa el valor `alternateThreshold` (0 siempre suprime desgloses alternativos). Este valor es el valor de fiabilidad mínimo que es necesario para un análisis inicial antes de que se reordene del nombre y se tomen en cuenta desgloses alternativos.

#### Resultados

El método `parse()` translitera y analiza la serie de caracteres de entrada y devuelve un conjunto de las siguientes estructuras de datos para cada serie de caracteres de entrada. El método `parse()` devuelve cada uno de los objetos siguientes para un elemento de nombre.

**ParseAlternate**

Contiene datos de análisis para un nombre alternativo (conjunto).

**ParseName**

Contiene información sobre posible desgloses de un nombre.

**ParseField**

Contiene información de análisis sobre un campo de nombre.

**ParsePhrase**

Contiene datos de análisis para una única frase de nombre.

---

## Analizar nombres mediante NameParser

NameParser es un componente de bajo nivel que separa los nombres de persona en sus componentes constituyentes como, por ejemplo, el nombre de pila, el apellido, las fórmulas de tratamiento y los calificadores.

NameParser completa cada una de las acciones siguientes cuando analiza un nombre de persona.

- Determina el límite adecuado entre el nombre de pila y el apellido.
- Separa las fórmulas de tratamiento y los calificadores del nombre.
- Separa el nombre de pila y el apellido en frase (gestionando detalles como, por ejemplo, prefijos y raíces) que pueden funcionar individualmente.
- Desglosa en varios nombres a partir de una serie de caracteres de nombre individual. Por ejemplo, la serie de caracteres "John y Mary Smith" se separa en varias, "John Smith" y "Mary Smith."

### Tipos de series de entrada

La función `parseName()` tiene varias versiones sobrecargadas para facilitar la integración con los métodos que la aplicación utilice actualmente para tratar los nombres. Si los nombres ya se han dividido en los campos de nombre de pila y apellido, puede utilizar NameParser para asegurarse de que se han colocado los elementos en cada campo.

Puede pasar nombres a NameParser a través de los métodos siguientes utilizando la función `parseName()`.

**En una serie única, como un nombre si analizar.**

Todo el nombre que debe analizarse puede pasarse como serie única en su formato original. Existe un distintivo `gnFirst` independiente que la aplicación llamante puede utilizar para especificar el orden de campos de la serie de entrada. Si `gnFirst` es `true`, se espera que el nombre se encuentre ordenado con el nombre de pila en primer lugar. En otras palabras, NameParser supondrá que el nombre de pila precede al apellido en la serie. Si `gnFirst` es `false`, NameParser presupone que el apellido precede al nombre de pila en la serie de entrada. Para los nombres que se ha determinado que son chinos, japoneses o coreanos, NameParser supone de forma predeterminada que el apellido aparece en primer lugar en la cadena, pero el distintivo `gnFirst` influye sobre este algoritmo en uno u otro sentido en casos de ambigüedad. De forma predeterminada, el distintivo `gnFirst` tiene el valor de `true`.

**Nota:** El distintivo `gnFirst` debe utilizarse en situaciones en las que se conoce el orden de campos de las series que se procesan; en situaciones en las que no se conoce, pero se sabe que puede ser ambiguo, deje `gnFirst` establecido en `true`.

**En una serie única, como un nombre analizado.**

Si el nombre se ha desglosado antes como nombre de pila y apellido, puede pasar el nombre como serie única, con el apellido en primer lugar y una coma entre éste y el nombre de pila (por ejemplo, "King, Martin Luther"). `NameParser` convertirá el nombre al orden natural y lo reanalizará. Las comas que separan calificadores (por ejemplo, "Martin Luther King, Jr.") están permitidas y no generarán errores en la división entre el apellido y el nombre de pila.

**En dos series, como un nombre analizado**

Si los nombres se encuentran en un contexto en el que ya están divididos en series separadas de nombre de pila y apellido, puede pasar simplemente esas dos series (por ejemplo, "Martin Luther" y "King") a `NameParser`, que las concatenará automáticamente para formar una serie única con el orden natural y la reanalizará.

**En tres series, como un nombre analizado**

Si los nombres se encuentran en un contexto en el que ya están divididos en series separadas de nombre de pila, segundo nombre y apellido, puede pasar simplemente esas tres series (por ejemplo, "Martin", "Luther" y "King") a `NameParser`, que las concatenará automáticamente para formar una serie única con el orden natural y la reanalizará.

## Funciones de `NameParser` para el análisis de nombres

El método para analizar un nombre utilizando `NameParser` es la función `parseName()`.

**Función `parseName()`**

Realiza un análisis completo del nombre y genera un árbol de análisis. El árbol de análisis es una jerarquía que agrupa los elementos de un nombre en unidades estructurales, empezando por las cadenas de caracteres individuales (series delimitadas por espacios o signos de puntuación), que pueden combinarse en frases de nombre, que a su vez se combinan para formar un nombre de persona completo o varios.

### Análisis alternativos y puntuaciones de validación

`NameParser` busca varios análisis de un nombre, empezando por un análisis que compara el primer parámetro `gn` especificado y el orden en el que se ha proporcionado el nombre. `NameParser` también tiene en cuenta los análisis alternativos que pueden ser más probables que el primer análisis.

Tras haber eliminado todas las fórmulas de tratamiento y calificadores, las frases restantes del nombre pueden rotarse pero debe preservarse su orden relativo. Por ejemplo, si el nombre de entrada original incluía la frase *A B C D*, al volver a realizar el análisis se pueden convertir las frases en *B C D A* o *D A B C*, pero no en *A D B C* o *B A C D*. Siempre que dos o más rotaciones sean posibles, el resultado real que se ha vuelto a analizar es el resultado que proporcione la puntuación de validez más alta.

El objeto `ParseData` devuelto por `parseName()` contiene uno o varios nombres conjuntos y ofrece una lista de los análisis alternativos (si los hay) para cada

nombre conjunto (puede haber más de un análisis por nombre).

## Lista de alteraciones temporales de frases de NameParser

Puede añadir frases a la lista de alteraciones temporales de frases para personalizar el comportamiento de NameParser. La lista de alteraciones temporales de frases es una lista suplementaria de frases de nombres en las que NameParser busca. Si no se halla una frase en la lista suplementaria, NameParser busca en sus propias tablas internas que incluyen información sobre millones de frases de nombres.

Puesto que las necesidades de cada despliegue varían, puede que tenga que modificar la base de datos de cadenas de caracteres y frases de NameParser. Por ejemplo, tal vez necesite añadir fórmulas de tratamiento y calificadores al conjunto incorporado en NameParser, complementar las tablas de recursos internos de NameParser con cadenas de caracteres de raíz de nombres o afijos que aparecen con frecuencia en los datos o bien alterar temporalmente los datos de frecuencia de nombres de pila y apellidos para que se ajusten mejor al uso local en los nombres que se procesan.

La lista de alteraciones temporales de frases es una correlación STL que se correlaciona de un objeto `std::string` con un objeto `PhraseOverrideData`. El objeto `std::string` es la cadena de caracteres o frase de nombre que desea añadir a la lista, representada en letras mayúsculas ASCII. El objeto `PhraseOverrideData` es un registro de información acerca de la cadena de caracteres que proporciona un tipo de cadena de caracteres más los factores de nombre de pila y apellido.

Los factores de nombre de pila y apellido se ignoran en todos los tipos de cadenas de caracteres salvo en las cadenas de caracteres `NAME_STEM`, por lo que para todos los demás tipos de cadenas de caracteres estos valores deben establecerse en 0. Con las cadenas de caracteres `NAME_STEM` los factores de nombre de pila y apellido se utilizan para determinar si NameParser debe asociar la cadena de caracteres al campo de nombre de pila o apellido cuando se encuentra en un nombre. En general, la determinación de la asociación de campos se basa en si la cadena de caracteres tiene un factor de nombre de pila más alto o un factor de apellido más alto. La lógica global considera muchos más elementos pero los factores de nombre de pila y apellido son las medidas clave en la toma de decisiones de cómo se debe asociar una cadena de caracteres de nombres.

La lista de alteraciones temporales de frases reales no está accesible de forma directa. Las API con las que se puede trabajar funcionan copiando los nombres y los registros de `PhraseOverrideData` de una lista proporcionada por el código de cliente. La lista de alteraciones temporales de frases está vacía de forma predeterminada

Cada entrada de la lista de alteraciones temporales de frases está asociada a un objeto `PhraseOverrideData` que especifica el tipo, el factor de apellido y el factor de nombre de pila de esa cadena de caracteres. El tipo de miembro de un objeto `PhraseOverrideData` indica al método `setPhraseOverrides()` cómo procesar los datos asociados. La clase de `PhraseOverrideData` incluye soporte para los caracteres siguientes:

- Las letras en latín en minúsculas son equivalentes a las letras en latín en mayúsculas.
- Las letras en latín ampliado son equivalentes a su equivalente ASCII más próximo. Por ejemplo, los acentos diacríticos se eliminan de la mayoría de los caracteres alfabéticos. Se aceptan todos los caracteres alfabéticos de la siguiente codificación Unicode:

- Latín-1
- Latín ampliado A
- Bloques adicionales de latín ampliado
- Bloque de latín ampliado B (la mayoría de caracteres)
- Los caracteres numéricos 0-9, el espacio y los siguientes caracteres especiales son importantes: @ \$ % & + = /
- Los guiones, puntos, comas, separadores y líneas nuevas son equivalentes a los espacios, y los múltiples caracteres de espacios de equivalentes en una fila son equivalentes a un solo espacio.

Los demás caracteres no incluidos en la lista anterior, incluido el apóstrofo, se ignoran.



---

## Capítulo 4. Analizar nombres

Cuando se analizan nombres, se identifican diversos atributos relativos a los nombres como, por ejemplo, el probable género del nombre, el probable entorno étnico (cultural) del nombre, el país de origen probable del nombre, las posibles variantes del nombre y la categoría del nombre, ya sea un nombre de persona o un nombre de organización.

---

### Análisis de nombres utilizando NameWorks

Utilice el método `analyze()` cuando desee realizar un análisis completo de un nombre. Este método translitera y desglosa el nombre, proporciona información sobre el género, una clasificación del entorno cultural, las variantes de nombres y una lista de países en la que aparece el nombre (información sobre el país de asociación).

#### Acerca de esta tarea

Para realizar un análisis completo de un nombre, utilice el método `analyze()` de la clase `Analytics` de `NameWorks` y pase al método los valores siguientes:

- El nombre completo, como valor de entrada simple, o las series sueltas de apellido y nombre de pila si se desea un análisis específico.
- Un entero entre 0 y 100 representa el valor de `alternateThreshold`, que establece el *valor confianza mínimo aceptable* a la hora de decidir si `NameWorks` debe buscar análisis alternativos. Si el análisis inicial tiene un valor de confianza por encima del umbral no se intentarán análisis alternativos. Un valor de `alternateThreshold` de cero (0) siempre suprime los análisis alternativos.
- Un entero que representa el valor de `maxForms` que limita el número de posibles formas de variantes generadas para cada frase de nombre. Un valor de cero (0) indica que se han de devolver todas las posibles formas de variantes.
- Un entero que representa el valor de `maxElements` que limita el número de posibles elementos de país que se obtienen para cada frase de nombre. Un valor inferior a uno provocará que se devuelvan todos los elementos de país posibles.

#### Resultados

El método `analyze()` translitera y analiza la serie de entrada y, a continuación, devuelve el siguiente conjunto de estructuras de datos anidadas que representan los datos del análisis:

- `AnalysisData`
- `AnalysisAlternate`
- `AnalysisName`
- `AnalysisField`
- `AnalysisPhrase`

---

## Identificación del entorno cultural de un nombre utilizando NameWorks

NameWorks incluye funciones que identifican el entorno cultural de los nombres de persona, utilizando el componente de API NameClassifier—Country of Association (NC\_COA). Las funciones llevan a cabo primer una clasificación del entorno cultural simple, si no se puede seleccionar un entorno cultural simple, se realiza un análisis más complejo de las asociaciones de países con el fin de determinar los posibles entornos culturales. En el caso de los nombres de organizaciones, el entorno cultural se puede deducir automáticamente del tipo de script donde se graban los nombres en scripts no latinos soportados o puede proporcionarlo el usuario. Para los nombres de organizaciones de script latino, el usuario puede proporcionar el entorno cultural, de lo contrario se tomará un entorno cultural Ambiguo.

### Identificación del entorno cultural

Los nombres difieren en función de las áreas geográficas y los grupos culturales. Estas diferencias van desde la fonética utilizada en los nombres hasta la ubicación del nombre de pila en relación a los demás componentes del nombre. Identificar el entorno cultural de un nombre puede mejorar significativamente la comparación de nombres.

Mediante la identificación del entorno cultural de un nombre, los productos IBM InfoSphere Global Name Management pueden aplicar especificidades del entorno cultural, como por ejemplo el reconocimiento de apodos y valores de parámetro preajustados, que aumentan la recuperación de resultados de búsqueda y reducen las falsas identificaciones positivas. Los productos IBM InfoSphere Global Name Management aplican una combinación de técnicas de lingüística, estadística y probabilística para identificar la posible naturaleza cultural de los nombres de persona representados en el alfabeto latino.

Al tiempo que la capacidad para identificar el entorno cultural puede aumentar las prestaciones de comparación de nombres, incluso sin conocimiento específico del entorno cultural, los productos IBM InfoSphere Global Name Management son capaces de analizar, establecer el género y comparar nombres de forma eficaz y competitiva.

Este ejemplo muestra cómo la identificación del entorno cultural de un nombre hispánico puede aumentar las prestaciones de comparación de nombres.

En las comunidades hispánicas, las personas tienen generalmente dos apellidos. El primer apellido (el situado más a la izquierda) es el apellido del padre, y es el nombre utilizado como apellido de la persona. El segundo apellido es el de la madre, y puede omitirse.

Dado que los productos IBM InfoSphere Global Name Management pueden identificar el nombre "Ana Garcia Valdez" como nombre hispano y aplicar parámetros de puntuación específicos del entorno cultural durante el proceso, el nombre "Ana Garcia" es una coincidencia de nombre de nivel superior. Sin embargo, el nombre "Ana Valdez" no se considera una coincidencia de nombre de nivel superior, aunque los dos nombres contienen los mismos componentes de nombre, ya que los nombres no coinciden en el primer componente de apellido.

## Códigos de entorno cultural

Los códigos de entorno cultural identifican uno o más entornos culturales asociados a un nombre de persona durante la clasificación de entornos culturales.

Los productos IBM InfoSphere Global Name Management utilizan los códigos de entorno cultural siguientes:

**Nota:** Un código de entorno cultural global representa un conjunto definido de culturas dentro de una región determinada. Si un nombre devuelve varios códigos de entorno cultural dentro de esa región, el código global es el que mejor representa el entorno cultural de ese nombre.

*Tabla 5. Códigos de entorno cultural y culturas asociadas*

Código	Entorno cultural asociado
0	Ambiguo (No se pudo determinar ningún código global válido)
1	Anglosajón
2	Árabe
3	Chino
4	Hispano
5	Coreano
6	Ruso
7	Francés
8	Alemán
9	Tailandés
10	Indonesio
11	Yoruba
12	Farsi
13	Paquistaní
14	Hindú
15	Japonés
16	Afgano
17	Vietnamita
18	Polaco
19	Portugués
20	Turco
38	Sudoeste asiático (Código de entorno cultural global que representa una combinación de las culturas árabe, farsi, paquistaní, afgana y turca)
39	Europeo (Código de entorno cultural global que representa una combinación de las culturas anglosajona, francesa, alemana, hispánica y portuguesa)
40	Han (Código de entorno cultural global que representa una combinación de las culturas china, coreana y vietnamita)

También se han añadido los siguientes códigos para los entornos culturales personalizados:

Tabla 6. Códigos de entorno cultural personalizado

Código	Entorno cultural asociado
41	CUSTOM_01
42	CUSTOM_02
43	CUSTOM_03
44	CUSTOM_04
45	CUSTOM_05
46	CUSTOM_06
47	CUSTOM_07
48	CUSTOM_08
49	CUSTOM_09
50	CUSTOM_10
51	CUSTOM_11
52	CUSTOM_12
53	CUSTOM_13
54	CUSTOM_14
55	CUSTOM_15
56	CUSTOM_16
57	CUSTOM_17
58	CUSTOM_18
59	CUSTOM_19
60	CUSTOM_20

## Identificar el entorno cultural de un nombre completo

Si puede identificar el entorno cultural de un nombre, y añadir esa clasificación cultural a los datos, podrá asociar nombres con mayor fiabilidad y también mejorar significativamente el rendimiento en la asociación de nombres. Esta tarea es útil para identificar la cultura de una serie de nombre que todavía no se ha desglosado en frases de nombre separados.

### Acerca de esta tarea

Para identificar el entorno cultural de un nombre, utilice el método `classify()` de la clase `Analytics` de `IBM NameWorks`, y pase al método el nombre completo en forma de un valor de tipo `serie`.

Este método realiza una transliteración en el nombre, analiza el nombre, asocia un entorno cultural al nombre analizado, y luego devuelve la clasificación de entorno cultural en forma de uno de los 20 códigos de entorno cultural estándar.

### Resultados

El método `classify()` devuelve el objeto `CultureData`.

## Identificar el entorno cultural de campos de nombre

Si puede identificar el entorno cultural de las partes de un nombre (nombre de pila y apellido), y añadir esa clasificación de entorno cultural a los datos, podrá asociar nombres con mayor fiabilidad y también mejorar significativamente el rendimiento en la asociación de nombres.

### Acerca de esta tarea

Para identificar el entorno cultural de un nombre que ya está desglosado en los campos correspondientes al nombre de pila y apellido, utilice el método `classify()` de la clase `Analytics` de `IBM NameWorks`, y pase al método los campos de nombre (`gn` y `sn`) en forma de un par de valores de tipo serie. Cualquiera de los dos campos de nombre puede ser nulo o contener una serie de caracteres vacía.

Este método analiza el nombre, asocia un entorno cultural a los campos de nombre analizados, y luego devuelve la clasificación de entorno cultural para uno o ambos campos de nombre en forma de uno de los 20 códigos de entorno cultural estándar.

### Resultados

El método `classify()` devuelve el objeto `CultureData`.

## Identificación del entorno cultural de un nombre de organización

Si puede identificar el entorno cultural de un nombre de organización, puede aplicar las reglas específicas del idioma para correlacionar dígitos y símbolos especiales con sus formas representadas. Por ejemplo, la serie de dígitos “62” es equivalente a “sesenta y dos” en español, pero coincide con “육십이” en un nombre hangul coreano.

### Acerca de esta tarea

Para los nombres de organización, hay soporte disponible específicos para los entornos culturales anglosajón, chino mandarín, hispánico, hindú, japonés, coreano, polaco, portugués, ruso y turco.

El entorno cultural de un nombre de organización se puede determinar automáticamente en función del sistema de escritura utilizado para los nombres escritos en unos pocos scripts. Para los demás nombres de organización, el usuario debe proporcionar el entorno cultural para activar el uso de los recursos disponibles específicos del entorno cultural. En la tabla siguiente se muestra de qué manera se correlacionan los scripts con los entornos culturales:

Kanji combinado con kana	Japonés
Kana	Japonés
Hanzi	Chino
Devanagari	Hindú
Cirílico	Ruso

Todos los demás nombres de organizaciones se clasificarán como Ambiguos, a menos que se proporcione un entorno cultural específico.

Fíjese que los nombres codificados en caracteres hanzi/kanji se supondrá que son chinos. Para activar las transliteraciones en japonés de caracteres kanji, se debe especificar el entorno cultural japonés para todos los nombres de organizaciones japonesas escritas en kanji.

El sistema de escritura japonés, kanji, utiliza casi el mismo conjunto de caracteres que el sistema hanzi que se utiliza para el chino. Por lo tanto, no es posible muchas veces indicar si el nombre de una organización en caracteres han es japonesa o china. A pesar de la similitud en las formas escritas de los dos idiomas, tienen pronunciaciones completamente distintas de los caracteres. Para asegurarse de que las lecturas japonesas se consiguen de los caracteres han, en la entrada se debe especificar el entorno cultural de un nombre japonés en kanji.

Los nombres de organizaciones japonesas escritos en kana se pueden detectar automáticamente como japonés en función del tipo de script. Sin embargo, los nombres de organización en japonés suelen estar escritos casi exclusivamente en kanji o en una combinación de kanji y kana.

---

## Identificación del género de nombres utilizando NameWorks

NameWorks incluye funciones para identificar las frecuencias relativas de los géneros asociados a un nombre de pila. Puede recuperar las frecuencias de género para un nombre como el total de frecuencias de todas las culturas en que se da ese nombre en Name Date Archive o puede elegir ver solamente las frecuencias halladas para un entorno cultural concreto.

Las frecuencias se expresan en porcentajes del número total de apariciones de un nombre de pila. Por ejemplo, el nombre "ILHAM" devuelve un 79% para mujeres y un 20% para hombres, lo que significa que el 79% de las entradas "ILHAM" en Name Data Archive se notificaron como nombre femenino y el 20% de las entradas se notificaron como nombre masculino y un 1% no se notificó con ningún género específico.

### Identificar el género de un nombre completo

Utilice esta tarea si desea identificar el campo de nombre de pila de un nombre completo y devolver datos sobre el género del nombre de pila. Puede recuperar las frecuencias de género para un entorno cultural específico indicando dicho entorno cultural. Por ejemplo, puede hallar la distribución de género de nombres de pila en turco para el nombre "ILHAM SAYIL" especificando el entorno cultural turco. Si se deja el valor del entorno cultural sin especificar, se devolverá la distribución de género en todos los entornos culturales.

#### Acerca de esta tarea

Para identificar el género de un nombre que todavía no está desglosado en campos de nombre, utilice el método `genderize()` de la clase `Analytics` de IBM NameWorks, y pase al método el nombre completo en forma de un valor de tipo serie.

Este método analiza el nombre, asocia un género al nombre de pila analizado, y luego devuelve datos sobre el género.

#### Resultados

El método `genderize()` devuelve un objeto `GenderData`. El valor de retorno es una estructura de datos que contiene valores que representan la frecuencia relativa de

aparición del nombre de pila, y porcentajes que representan la probabilidad de que el género asociado al nombre de pila sea masculino, femenino o desconocido.

## Identificación del género de un nombre de pila

Utilice esta tarea cuando desee obtener datos sobre el género para un nombre de pila. Puede recuperar las frecuencias de género para un entorno cultural específico indicando dicho entorno cultural. Por ejemplo, puede hallar la distribución de género de nombres de pila en turco para el nombre "ILHAM" especificando el entorno cultural turco. Si se deja el valor del entorno cultural sin especificar, se devolverá la distribución de género en todos los entornos culturales.

### Acerca de esta tarea

Para identificar el género de un nombre de pila, utilice el método `genderizeField()` de la clase `Analytics` de `NameWorks`, y pase al método el nombre de pila en forma de un valor de tipo serie (`givenName`).

Este método asocia un género al nombre de pila y luego devuelve datos sobre el género.

### Resultados

El método `genderizeField()` devuelve un objeto `GenderData`. El valor de retorno es una estructura de datos que contiene valores que representan la frecuencia relativa de aparición del nombre de pila, y porcentajes que representan la probabilidad de que el género asociado al nombre de pila sea masculino, femenino o desconocido.

---

## Identificación del país de asociación para nombres mediante NameWorks

`NameWorks` incluye funciones que se pueden utilizar para identificar el país de asociación para nombres con el fin de ayudarle en el análisis de nombres.

### COA (Country of association)

Si puede asociar un país con un nombre, podrá mejorar las prestaciones de identificación de entornos culturales y comparación de nombres del producto `IBM InfoSphere Global Name Management`. El hecho de asociar un nombre con un país puede reducir el número de entornos culturales "desconocidos" y aumentar la exactitud de los entornos culturales identificados.

Los productos `IBM InfoSphere Global Name Management` incluyen estadísticas de país de asociación (COA) para aumentar la exactitud de la identificación del entorno cultural que, a su vez, permite a los desarrolladores equilibrar y validar la información cultural con los datos de distribución procedentes de `IBM InfoSphere Global Name Data Archive (NDA)`. La función COA puede proporcionar tres valores diferentes relacionados con la asociación de un nombre con un país:

#### Frecuencia

Indica si es común un nombre en un país en comparación con los otros nombres del país. Los valores de frecuencia dependen del nombre.

#### Seguridad

Indica cuántos datos la NDO contiene de un país determinado en comparación con la cantidad de datos que la NDO contiene de otros países. Los valores de seguridad varían según el país.

### Importancia

Conocida también como importancia entre países, esta función ofrece una indicación del grado de representación de un nombre en relación con un país en comparación del grado de representación en otros países. El factor de importancia es un método más completo de establecer el país de asociación porque se toman en cuenta factores como el número total de nombres en circulación en un país, además de la cantidad de datos que la NDO contiene sobre ese país.

La función COA devuelve la información siguiente en este orden:

1. Importancia, de mayor a menor
2. Frecuencia, de mayor a menor
3. Seguridad, de mayor a menor
4. Código de país de la ISO 3166, en orden alfabético ascendente

## Identificar el país de asociación para nombres completos

Puede añadir otra capa de información al análisis de nombres identificando el país de asociación de un nombre. Utilice esta tarea cuando el nombre ya esté desglosado en campos de nombre.

### Acerca de esta tarea

Para identificar el país de asociación de un nombre de pila y apellido, utilice el método `associate()` de la clase `Analytics` de IBM NameWorks, y pase al método los valores siguientes:

- el nombre completo expresado como valor de tipo serie
- un número entero que representa el valor de `maxElements`, el cual restringe el número de elementos de país devueltos para cada frase de nombres (si este valor es menor que 1, se obtienen todos los elementos de país)

Este método analiza el nombre completo y realiza una transliteración en cada campo de nombre, y luego asocia los países cuyos campos de nombre en el análisis tengan la fiabilidad más alta.

### Resultados

El método `associate()` devuelve el conjunto siguiente de objetos anidados para cada frase de nombres:

- `CountryData`
- `CountryPhrase`
- `CountryElement`

## Identificar el país de asociación para un nombre de pila y apellido

Si puede identificar el país de asociación para el nombre de pila y apellido, puede añadir otra capa de información al análisis de nombres. Utilice esta tarea cuando el nombre ya esté desglosado en campos de nombre.

## Acerca de esta tarea

Para identificar el país de asociación de un nombre de pila y apellido, utilice el método `associate()` de la clase `Analytics` de IBM NameWorks, y pase al método los valores siguientes:

- los campos de nombre (`gn` y `sn`) en forma de un par de valores de tipo serie (cualquiera de los campos de nombre puede ser nulo o contener una serie de caracteres vacía).
- un número entero que representa el valor de `maxElements`, el cual restringe el número de elementos de país devueltos para cada frase de nombres (si este valor es menor que 1, se obtienen todos los elementos de país)

Este método realiza una transliteración en cada campo de nombre, y luego asocia países a los campos de nombre.

## Resultados

El método `associate()` devuelve el conjunto siguiente de objetos anidados para cada frase de nombres:

- `CountryData`
- `CountryPhrase`
- `CountryElement`

---

## Generación de las variantes de nombre utilizando NameWorks

NameWorks incluye funciones que se pueden utilizar para generar listas de formas variantes para los campos de un nombre.

### Variantes de nombre

Una *variante de nombre* es una alternativa de un nombre que se considera equivalente a dicho nombre, pero que difiere de él en su forma externa particular. Dicho de otro modo, los dos nombres se consideran hasta cierto punto equivalentes y pueden sustituirse entre sí en ciertos contextos.

Las variantes de nombre pueden aparecer por muchas razones, incluidas las siguientes:

- Variaciones ortográficas (por ejemplo, *Geoff* y *Jeff*)
- Apodos (por ejemplo, *Bill* por *William*)
- Abreviaturas (por ejemplo, *GPE* por *Guadalupe*)
- Palabras afines o traducciones (por ejemplo, *Peter* por *Pierre*)
- Diferencias culturales
- Variaciones en el orden de los componentes (por ejemplo, al adaptar un nombre a otro entorno cultural, *L.N.S. Gandikota* adaptado a partir de *Gandikota Lakshmi Narayana Sastry*)
- Transliteraciones de un sistema de escritura a otro (por ejemplo, de caracteres chinos logográficos a caracteres latinos).

IBM InfoSphere Global Name Management pueden generar variantes de nombre provocadas por variaciones ortográficas, apodos, diferencias culturales y abreviaturas.

El conocimiento de las posibles variantes de nombre ayuda a ampliar las consultas de nombres para incluir las formas variantes e incluir dichas variantes en las búsquedas para generar listas de candidatos a la coincidencia. Las variantes de nombre también pueden ser de utilidad al analizar nombres, ya que los analistas pueden visualizar las listas de formas variantes de un nombre que es posible que aparezcan.

En los productos IBM InfoSphere Global Name Management , el proceso de generación de una lista de variantes de nombre implica dividir los campos de nombre (nombre de pila y apellido) en frases de nombre, y luego generar la lista de variantes para cada una de las frases de nombre. Es importante conocer qué parte corresponde al nombre de pila y cuál al apellido, ya que existen variantes diferentes para una frase en función del campo en el que figura.

También es importante conocer el entorno cultural subyacente a cada campo de nombre, ya que las variantes de nombre difieren ampliamente entre entornos culturales. Sólo porque un nombre determinado pueda encontrarse en muchos entornos culturales y se escriba del mismo modo, no significa que los nombres sean iguales. En realidad, los nombres son diferentes y generan formas variantes diferentes en función del entorno cultural con el que estén asociados.

Por ejemplo, si comparamos las variantes de nombre del nombre hispano *Juan* y el nombre chino *Juan*, observaremos una gran diferencia en las formas variantes de nombre, debido a que el nombre no es el mismo en cada cultura.

Las variantes de nombre hispano de *Juan* incluyen:

- Juam
- Juanch
- Juancho
- Xuan

Las variantes de nombre chino de *Juan* incluyen:

- Chuan
- Chwan
- Jwan
- Zhuan

Aunque ambas culturas comparten algunas de las mismas variantes, los dos nombres son muy diferentes. La diferencia en el orden de las variantes de nombre, que refleja las diferencias en la frecuencia de las diversas ortografías de una cultura a otra, da soporte a la diferenciación de nombres.

## **Generar una lista de variantes para nombres completos**

Mediante la creación de una lista de variantes de nombres, puede ampliar las consultas de nombres, crear mejores listas de resultados posibles de búsqueda, o analizar mejor un nombre observando las variantes posibles asociadas a él. Debido a que el proceso de crear una lista de variantes de nombres depende del desglose de un nombre en campos de nombre y la identificación de entorno cultural de cada campo de nombre, utilice esta tarea cuando el nombre ya esté desglosado en los campos correspondientes al nombre de pila y apellido.

## Antes de empezar

Debido a que la lista de variantes de nombres difiere ampliamente según el entorno cultural asociado al nombre, el método `getVariants()` utiliza un código de entorno cultural como parámetro de entrada. Puede ser conveniente obtener el código de entorno cultural del nombre de pila y apellido antes de efectuar esta tarea. (Puede obtener el código de entorno cultural utilizando el método `analyze()` o el método `classify()` de la clase `Analytics`).

## Acerca de esta tarea

Para generar una lista de las variantes de un nombre completo, utilice el método `getVariants()` de la clase `Analytics` de IBM NameWorks, y pase al método los valores siguientes:

- El nombre completo expresado como valor de tipo serie
- Un código de entorno cultural que proporciona el entorno cultural del nombre (Si no conoce el entorno cultural del nombre, puede pasar al método el valor `NULL` para Java o `-1` para servicios Web, e IBM NameWorks determinará el entorno cultural probable del nombre).
- Un número entero para limitar el número de modalidades de una variante devueltas para cada frase de nombres (Si se especifica un valor negativo o un cero, se obtienen todas las modalidades de las variantes de nombres).

Este método devuelve una lista de variantes. Si ha pasado al método el valor `-1` o `NULL` como parámetro de entorno cultural para que IBM NameWorks determine el entorno cultural, puede examinar el campo de entorno cultural del objeto `VariantData` para averiguar el entorno cultural que IBM NameWorks asoció al nombre.

## Resultados

El método `getVariants()` devuelve el conjunto siguiente de objetos anidados en forma de estructura arborescente, que representa el desglose de las frases de nombres encontradas en el nombre sometido a análisis:

- `VariantData`
- `VariantPhrase`
- `VariantForm`

## Generar una lista de variantes para nombres de pila y apellidos

Mediante la creación de una lista de variantes de nombres, puede ampliar las consultas de nombres, crear mejores listas de resultados posibles de búsqueda, o analizar mejor un nombre observando las variantes posibles asociadas a él. Debido a que el proceso de crear una lista de variantes de nombres depende del desglose de un nombre en campos de nombre y la identificación del entorno cultural de cada campo de nombre, utilice esta tarea cuando ya disponga de los campos correspondientes al nombre de pila y apellido.

## Antes de empezar

Debido a que la lista de variantes de nombres difiere ampliamente según el entorno cultural asociado al nombre, el método `getVariants()` utiliza un código de entorno cultural como parámetro de entrada. Puede ser conveniente obtener el código de entorno cultural del nombre de pila y apellido antes de efectuar esta

tarea. (Puede obtener el código de entorno cultural utilizando el método `analyze()` o el método `classify()` de la clase `Analytics` de `IBM NameWorks`).

### **Acerca de esta tarea**

Para generar una lista de las variantes de un nombre que ya está desglosado en los campos correspondientes al nombre de pila y apellido, utilice el método `getVariants()` de la clase `Analytics` de `IBM NameWorks`, y pase al método los valores siguientes:

- Los campos correspondientes al nombre de pila y apellido (Cualquiera de los dos campos de nombre puede ser `NULL` o contener una serie de caracteres vacía).
- Un código de entorno cultural que proporciona el entorno cultural del nombre (Si no conoce el entorno cultural del nombre, puede pasar al método el valor `NULL` para Java o `-1` para servicios Web, e `IBM NameWorks` determinará el entorno cultural probable del nombre).
- Un número entero para limitar el número de modalidades de una variante devueltas para cada frase de nombres (Si se especifica un valor negativo o un cero, se obtienen todas las modalidades de las variantes de nombres).

Este método devuelve una lista de variantes. Si ha pasado al método el valor `-1` o `NULL` como parámetro de entorno cultural para que `IBM NameWorks` determine el entorno cultural, puede examinar el campo de entorno cultural del objeto `VariantData` para averiguar el entorno cultural que `IBM NameWorks` asoció al nombre.

### **Resultados**

El método `getVariants()` devuelve el conjunto siguiente de objetos anidados en forma de estructura arborescente, que representa el desglose de las frases de nombres encontradas en el nombre sometido a análisis:

- `VariantData`
- `VariantPhrase`
- `VariantForm`

---

## **Analizar nombres con las API de componente**

Al analizar nombres, se identifican diversos atributos relativos a los mismos, como por ejemplo el probable género del nombre, el probable entorno cultural del nombre, el país de origen probable del nombre, diversas variantes del nombre y la categorización del nombre como nombre de persona o nombre de organización.

Los productos `IBM InfoSphere Global Name Management` se suministran con bibliotecas `C++` a las que el usuario puede enlazarse para integrar directamente la tecnología en las aplicaciones o el flujo de trabajo. Junto con las API se suministran aplicaciones de ejemplo que muestran cómo utilizar las bibliotecas.

---

## Capítulo 5. Buscar nombres

Puede utilizar los productos IBM InfoSphere Global Name Recognition para buscar nombres en varias listas de datos.

---

### Gestionar listas de datos en IBM NameWorks

Las listas de datos son colecciones de nombres basadas en memoria que se llenan a partir de un origen de datos externo (como por ejemplo un archivo plano) cuando se inicializa una aplicación IBM InfoSphere Global Name Management . Cada entrada de una lista de datos contiene información exhaustiva acerca de un único nombre al que se accede y que se tiene en cuenta durante el proceso de búsqueda para poder aplicar diversas mediciones lingüísticas, culturales y de similitud de series de caracteres precisas durante una búsqueda de nombres.

Las listas de datos son las estructuras de datos principales utilizadas por IBM para la búsqueda y comparación automática de nombres.

Generalmente, el administrador del sistema configura, llena y gestiona las listas de datos creando y manteniendo un conjunto de parámetros de configuración para cada lista. Una sola solicitud de búsqueda no está limitada a un número de nombres, ya que IBM NameWorks y NameHunter pueden dar soporte a un número indefinido de listas de nombres configurables. Además, la posibilidad de correlacionar archivos externos diferentes para diferir listas de nombres residentes en memoria permite establecer un ámbito de búsqueda dinámico, en función de la transacción, en el que cada solicitud sólo tiene en cuenta la lista o listas relevantes.

El archivo de configuración de IBM NameWorks contiene la información de correlación entre cada lista de datos y su instancia de motor de búsqueda, así como otra información clave utilizada durante la inicialización del sistema.

El archivo de configuración de IBM NameWorks requiere información de lista de datos que especifique la correlación entre las instancias del motor de búsqueda y las listas de datos y un distintivo que indique el tipo de búsqueda que puede realizarse en la lista de datos (búsqueda completa o búsqueda de nombres exclusivos). Si pueden añadirse nombres a la lista de datos en esta instancia de motor de búsqueda, la sección de lista de datos contendrá un distintivo de adición (add).

Los administradores del sistema y las aplicaciones cliente pueden añadir, suprimir o actualizar un nombre de una lista de datos, aunque haya búsquedas activas que hagan referencia a dicha lista de datos. Las búsquedas activas utilizan información almacenada en memoria caché para finalizar la operación, y las búsquedas subsiguientes utilizan la información de nombres de la lista de datos modificada.

### Listas de datos

Una *lista de datos* es una estructura de datos residente en la memoria que se llena con un conjunto de nombres extraídos de una fuente externa, como por ejemplo un archivo plano. Una vez creada y llenada con nombres, una lista de datos está disponible para utilizarla en las solicitudes de búsqueda subsiguientes. Cada lista

de datos debe denominarse de forma exclusiva y debe estar en un formato específico. Las listas de datos pueden contener desde uno hasta cientos de millones de nombres.

Dado que una lista de datos es una estructura de datos dinámica, IBM NameWorks da soporte a transacciones de manipulación de datos, como por ejemplo a la adición, actualización y supresión del contenido de la lista de datos. La manipulación dinámica permite que el contenido de las lista de datos almacenadas en la memoria permanezca sincronizado con el origen de datos subyacente al que representa, incluso cuando cambia el origen de datos.

Las listas de datos son utilizadas por los motores de búsqueda del producto IBM InfoSphere Global Name Management , mediante los cuales se realizan solicitudes de búsqueda en una o varias listas. Cada solicitud de búsqueda debe indicar los nombres de las listas de datos que deben utilizarse.

Habitualmente, un administrador del sistema es quien gestiona las listas de datos, que pueden encontrarse en uno o varios servidores dependiendo de cómo se hayan configurado los servidores de motor de búsqueda para satisfacer las necesidades de la empresa.

Generalmente, los administradores de sistemas asocian una lista de datos con una única instancia de motor de búsqueda, que en NameWorks es una instancia del proceso de Distributed Search (un proceso de gestión de comunicaciones y uno o varios procesos de búsqueda).

Los administradores de sistemas o las aplicaciones cliente pueden añadir, actualizar o suprimir nombres de las listas de datos según sea necesario. Sin embargo, en cualquier instancia del proceso de Distributed Search, sólo una lista de datos, la *lista de adiciones*, puede designarse como destinataria de los nombres que se añaden durante una sesión. Todas las demás listas de datos configuradas con una instancia de Distributed Search pueden contener sólo los nombres con los que se han llenado durante la inicialización de la sesión, de acuerdo con el archivo de configuración asociado. Por tanto, los nombres que se añadan después de la inicialización se colocarán en la lista de adiciones.

## **Añadir nombres a listas de datos**

Como parte de la gestión de listas de datos, el administrador del sistema puede necesitar añadir nombres a listas de datos como actualización provisional de la lista de datos entre renovaciones periódicas. Debido a que las listas de datos residen en la memoria, puede añadir nombres a una lista de datos en cualquier momento, aunque existan búsquedas activas que estén accediendo a la lista de datos. Si una búsqueda activa está en curso cuando se añaden nombres, se puede acceder a los nombres nuevos durante las búsquedas subsiguientes realizadas en la lista de datos.

### **Antes de empezar**

- La lista de datos debe estar configurada con el distintivo add en el archivo de configuración de IBM NameWorks. Si no está configurada de esa forma, se muestra un mensaje de error al intentar añadir nombres a la lista de datos.
- Debe conocer el nombre de la lista de datos. (Utilice el método `getDatalistNames()` para obtener una lista de todos los nombres de listas de datos existentes).
- El nombre ya debe estar desglosado en los campos correspondientes al nombre de pila y apellido. Puede utilizar el método `analyzeForSearch()` para desglosar el

nombre en campos y clasificarlo, lo cual proporciona la información necesaria para añadir el nombre a una lista de datos.

- Si ya dispone de información sobre el entorno cultural para cada campo de nombre (nombre de pila y apellido), puede también añadir el código de entorno cultural al nombre.

## **Acerca de esta tarea**

Para añadir un nombre a una lista de datos existente, utilice el método `addNameToDatalist()` de la clase `Scoring` de `IBM NameWorks`, y pase al método los valores siguientes:

- El nombre de la lista de datos.
- Los valores de los campos nombre de pila y apellido que se deben añadir a la lista de datos.
- El elemento del nombre original, la escritura original del nombre (si el nombre no está expresado en el alfabeto latino), o ambas cosas. Si se pasan ambas cosas, significa que esta adición se debe marcar como un elemento alternativo del nombre o escritura originales.
- Cualquier dato complementario que se deba asociar al nuevo registro de nombre. (Esta información es un valor de clave para identificar datos adicionales o complementarios referentes a un nombre, tales como una fecha de nacimiento o un número de licencia de conducir. Normalmente estos datos complementarios permiten recuperar información más completa sobre un nombre devuelto por una búsqueda. Se pueden utilizar datos complementarios en tareas de filtrado y ponderación después de una búsqueda. También se utilizan al actualizar o suprimir nombres en listas de datos. Se actualizan o suprimen todos los registros de nombres que tengan los mismos datos complementarios).
- Los códigos de entorno cultural del nombre de pila y apellido. (Si no tiene estos códigos de entorno cultural, pase el valor `-1`. Este valor indica el método para clasificar primero el entorno cultural de los campos de nombre de pila y apellido).

## **Resultados**

El método `addNameToDatalist()` añade el nombre y la información asociada a la lista de datos indicada. Si se pasa el valor `-1` como código de entorno cultural, significa que `IBM NameWorks` debe determinar automáticamente el código de entorno cultural más apropiado para el nombre.

## **Actualizar nombres contenidos en listas de datos**

Entre una renovación y otra de la lista de datos, el administrador del sistema puede necesitar actualizar los nombres contenidos en listas de datos existentes, como parte de sus tareas de gestión de listas de datos. Debido a que las listas de datos residen en la memoria, puede actualizar nombres de una lista de datos en cualquier momento, aunque existan búsquedas activas que estén accediendo a la lista de datos. Si una búsqueda activa está en curso cuando se actualizan nombres, se puede acceder a la información nueva durante las búsquedas subsiguientes realizadas en la lista de datos.

### **Antes de empezar**

- Debe conocer el nombre de la lista de datos. (Utilice el método `getDatalistNames()` para obtener una lista de todos los nombres de listas de datos existentes).

- Debe conocer el valor de datos complementario original (valor *originalData*) asociado al nombre, que habitualmente es una fecha de nacimiento, un número de licencia de conducir o un dato similar. Este valor de datos complementario es la clave para identificar los registros de nombres que se deben actualizar, y normalmente proporciona información más completa sobre el nombre que se puede utilizar en tareas posteriores a la búsqueda, tales como la ponderación y el filtrado. Se actualizarán todos los nombres de la lista de datos que contengan un valor de datos complementario coincidente.
- Si está cambiando el nombre de pila, el apellido o ambos elementos, recuerde que el nombre ya debe estar analizado. (Puede utilizar el método `analyzeForSearch()` para preparar el nombre antes de esta actualización. Este método desglosa el nombre en campos de nombre y clasifica el entorno cultural de cada campo de nombre).

## Acerca de esta tarea

Para actualizar un nombre y su información asociada en una lista de datos existente, utilice el método `updateNameInDatalist()` de la clase `Scoring` de IBM NameWorks, y pase al método los valores siguientes:

- El nombre de la lista de datos.
- El valor de datos complementarios original (*originalData*). Todos los registros con este valor de datos complementarios se actualizarán y el valor de datos complementarios original se sustituirá por este valor de datos).
- Los valores de los campos de nombre de pila y apellido que se deben modificar en la lista de datos.
- El elemento del nombre original, la escritura original del nombre, o ambas cosas. (Esta información es la clave para localizar la entrada original, e indica que esta adición se debe marcar como elemento alternativo para el nombre o escritura originales).
- Cualquier dato complementario con el que desee sustituir el valor *originalData* (información complementaria original). Por ejemplo, si ha cambiado la licencia de conducir asociada a la persona.
- Los códigos de entorno cultural del nombre de pila y apellido. (Si no conoce estos códigos de entorno cultural, pase el valor `-1`. Este valor indica el método para clasificar primero el entorno cultural de los campos de nombre de pila y apellido).

## Resultados

El método `updateNameInDatalist()` actualiza todos los registros de nombres que contengan los mismos datos complementarios en la lista de datos indicada. Si ha pasado al método el valor `-1` como código de entorno cultural para el nombre de pila o apellido (o para ambos), el método primero clasifica los códigos de entorno cultural y actualiza el nombre con los códigos de entorno cultural identificados.

## Suprimir nombres de listas de datos

Cuando realiza el mantenimiento de las listas de datos, el administrador del sistema puede a veces necesitar eliminar nombres en una lista de datos como parte de una actualización provisional de la lista de datos. Debido a que las listas de datos residen en la memoria, puede suprimir nombres en una lista de datos en cualquier momento, aunque existan búsquedas activas que estén accediendo a la lista de datos. Las búsquedas subsiguientes del nombre después de que éste se haya suprimido no producirán ninguna coincidencia de búsqueda para esa lista de datos.

## Antes de empezar

- Debe conocer el nombre de la lista de datos. (Utilice el método `getDatalistNames()` para obtener una lista de todos los nombres de listas de datos existentes).
- Debe conocer el valor de datos complementario original (valor *originalData*) asociado al nombre, que habitualmente es una fecha de nacimiento, un número de licencia de conducir o un dato similar. Este valor de datos complementario es la clave para identificar los registros de nombres que se deben suprimir. Todos los nombres de esta lista de datos que contengan un valor de datos complementario coincidente se suprimirán de la lista de datos.

## Acerca de esta tarea

Para suprimir un nombre y su información asociada en una lista de datos existente, utilice el método `updateNameInDatalist()` de la clase `Scoring` de `IBM NameWorks`, y pase al método los valores siguientes:

- El nombre de la lista de datos.
- Los datos complementario asociados al registro de nombre.

Para suprimir todos los registros de nombres que contengan los mismos datos complementarios en la lista de datos indicada, utilice el método `deleteNameFromDatalist()`.

---

## Migración de IBM NameWorks

Este release de IBM InfoSphere Global Name Management incluye diversos cambios en las API de IBM NameWorks. Utilice esta información para conocer los cambios necesarios a fin de migrar las API existentes.

### Nuevos métodos y objetos Java

La tabla siguiente muestra los métodos y objetos Java nuevos en este release.

*Tabla 7. Nuevos métodos y objetos Java*

Método u objeto Java	Descripción
Método <code>createName()</code>	Los métodos <code>search()</code> , <code>compare()</code> , <code>addName()</code> y <code>updateName()</code> aceptan ahora objetos <code>Name</code> como parámetro. Por tanto, debe crear un objeto <code>Name</code> mediante el método <code>createName()</code> para poder llamar a esta versión de búsqueda.
Objeto <code>Name</code>	En versiones anteriores, se utilizaba una serie de caracteres de nombre como entrada de los parámetros de nombre, apellido o nombre de pila. Este release introduce los objetos <code>Name</code> , que encapsulan los campos de nombre (nombre de pila y apellido), la información de entorno cultural y la categoría ( <code>NameCategory</code> ) del nombre. Esta clase representa los nombres que forman la información de entrada como nombre de consulta o como nombre de lista de datos procedente de un archivo de entrada.

Tabla 7. Nuevos métodos y objetos Java (continuación)

Método u objeto Java	Descripción
Objeto NameCategory	<p>NameCategory se utiliza para describir las categorías de nombres soportadas por los productos IBM InfoSphere Global Name Management . En este release están soportadas las siguientes categorías de nombres:</p> <ul style="list-style-type: none"> <li>• Desconocido</li> <li>• Personal</li> <li>• Organización</li> <li>• Ambos</li> </ul> <p>La categoría de un nombre se especifica al crear un nombre mediante el método createName().</p>
Objeto NameCategorySet	<p>NameCategorySet representa una colección de uno o varios valores de NameCategory. Este tipo de datos identifica la categoría de nombres específicos e indica la categoría de nombres que debe devolverse después de llamar al método search().</p>

## Métodos y objetos Java cambiados

Varias estructuras de datos de API de IBM NameWorks se han convertido en clases de datos. Cada uno de los objetos se deriva de la clase Name y, por tanto, hereda varios métodos que contienen información de nombre y entorno cultural. Ahora debe llamar a los métodos get() para acceder a los datos de miembros de varias clases, cada una de las cuales se describe en la tabla siguiente.

Tabla 8. Métodos y objetos Java cambiados

Método u objeto Java	Descripción de los cambios
Método addName()	<p>Existen dos versiones de este método, y ambas requieren el objeto Name como parámetro. Puede optar por añadir o no a la lista el nombre original y la versión transcrita del nombre, o sólo la versión transcrita.</p>

Tabla 8. Métodos y objetos Java cambiados (continuación)

Método u objeto Java	Descripción de los cambios
método analyzeForSearch()	<p>Existen dos versiones de este método: una versión acepta un parámetro de nombre completo y la segunda versión acepta parámetros de nombre de pila y apellido.</p> <ul style="list-style-type: none"> <li>• Si se llama al método que acepta un parámetro de nombre completo, se utiliza NameSifter para categorizar primero el nombre, y el proceso posterior se determina en función de la categoría del nombre (de persona, de organización o ambos). Los datos resultantes se utilizan para crear los objetos QueryName relacionados.</li> <li>• Si se llama al método que acepta parámetros de nombre de pila y apellido, el nombre se trata como nombre de persona. El nombre se analiza y clasifica y los datos resultantes se utilizan para crear los objetos QueryName relacionados.</li> </ul>
Método compare()	<p>Esta versión requiere objetos Name y da soporte a varios tipos de comparación. Se comparan dos nombres en función de la categoría (NameCategory) de queryName, y los resultados de la comparación se devuelven en el objeto CompareData. En lugar de aceptar parámetros de nombre de pila, apellido y entorno cultural individualmente para un nombre, se utilizan dos objetos Name (uno para el nombre de la izquierda y otro para el nombre de la derecha) durante la comparación.</p>
Método search()	<p>El método search() acepta ahora objetos Name como parámetros. Por tanto, debe crear un objeto Name mediante el método createName() para poder llamar a esta versión de búsqueda.</p>
Método updateName()	<p>Existen dos versiones de este método, y ambas requieren el objeto Name como parámetro. Puede optar por actualizar o no la lista con el nombre original y la versión transcrita del nombre, o sólo la versión transcrita.</p>
Objeto CategorizeData	<p>CategorizeData es el resultado de llamar al método categorize(). El nombre se categoriza como Personal, Organización, Ambos o Desconocido.</p>

Tabla 8. Métodos y objetos Java cambiados (continuación)

Método u objeto Java	Descripción de los cambios
Objeto CompareData	CompareData es el resultado de llamar al método compare(). Este objeto contiene dos objetos Name, uno utilizado como nombre de consulta y otro utilizado como nombre de evaluación, junto con los resultados de la comparación de nombres. Las estructuras de este objeto se han condensado ahora de forma que el entorno cultural del nombre se devuelve por medio del objeto Name. Además, puede especificar que se devuelva un distintivo que indique qué parámetro del nombre (izquierdo o derecho) se ha utilizado para la comparación.
Objeto OriginalName	OriginalName contiene los parámetros de nombre, de nombre de pila y de apellido. Además, este objeto contiene información de entorno cultural para los componentes del nombre, así como información relativa a los estados de regularización y de elementos alternativos.
Objeto QueryName	QueryName se utiliza para determinar el grado de confianza de una coincidencia de nombres. Este objeto acepta series de fórmula de tratamiento y calificador que pueden utilizarse durante la comparación de nombres.
Objeto SearchMatch	SearchMatch contiene un registro de nombre coincidente. Cada objeto SearchMatch devuelto desde una operación de búsqueda se entra en una lista, que se encuentra dentro del objeto SearchResults.

## Preparar nombres para la búsqueda

Utilice el método `analyzeForSearch()` para preparar un nombre para utilizarlo en una transacción de búsqueda. Este método categoriza el nombre determinando si se trata de un nombre de persona o de un nombre de organización, analiza los nombres de persona en campos de nombre de pila y apellido y determina un código de clasificación cultural para cada campo en un nombre de persona. Si `NameSifter` determina que el nombre es un nombre de organización, solamente se devuelve la información sobre la categoría del nombre.

### Acerca de esta tarea

El método `analyzeForSearch()` da soporte a series de caracteres de entrada solamente en codificación UTF-16. Si la serie de caracteres de entrada es una codificación no soportada, IBM NameWorks señala un error o produce resultados no deseados. Para preparar un nombre para la búsqueda, pase los valores siguientes al método `analyzeForSearch()` de la clase `Scoring` de IBM NameWorks.

- El nombre completo que se va a analizar, representado como un nombre completo. Por ejemplo, ROBERT E JONES. El nombre debe pasarse al método `parse()` como un valor *string*.

- Un entero entre 0 y 100 que representa el valor **alternateThreshold** (0 siempre suprime desgloses alternativos). Este valor es el valor de fiabilidad mínimo que es necesario para un análisis inicial antes de que se reordene del nombre y se tomen en cuenta desgloses alternativos.

## Resultados

El método `analyzeForSearch()` primero categoriza y luego translitera el nombre. La transliteración de nombres es el proceso de convertir un nombre de un sistema de escritura o convenio de codificación de caracteres determinado a otro. Por ejemplo, la transliteración de nombres permite analizar y comparar un nombre escrito en árabe con un nombre similar escrito en el alfabeto latino. Luego, el método analiza la serie de caracteres de entrada y devuelve una colección de objetos `QueryName` para cada serie de caracteres de entrada y su versión transcrita.

---

## Casos de ejemplo: búsqueda de nombres

IBM InfoSphere Global Name Management ofrece varias formas de buscar nombres. En los siguientes casos de ejemplo se describen varias implementaciones comunes que puede seguir para buscar nombres.

### Crear objetos de nombres para la búsqueda de nombres

En los casos de ejemplo siguientes se describen los diversos métodos de creación de objetos de nombres para preparar la búsqueda, basándose en cómo tenga previsto interactuar con la información de nombres.

Ambos casos de ejemplo describen cómo se pueden crear nombres para preparar la búsqueda, pero alcanzan este objetivo de maneras diferentes. El primer caso de ejemplo describe cómo crear un objeto `Name` mediante el método `createName()` antes de llamar al método `search()`. El segundo caso de ejemplo emplea el método `analyzeForSearch()` para crear objetos `QueryName` utilizados para la búsqueda de nombres.

Cuando se utiliza el método `search()`, primero debe crear los objetos `Name` para los nombres de persona y de organización mediante el método `createName()`. Los objetos `Name` encapsulan los campos de nombre (nombre de pila y apellido especificados), la información del entorno cultural y la categoría del nombre en lugar de aceptar series de caracteres de nombres como en los releases anteriores a IBM InfoSphere Global Name Management, versión 4.2. La clase `Name` representa los nombres que forman la información de entrada como nombre de consulta o como nombre de lista de datos procedente de un archivo de entrada.

Como alternativa, puede invocar el método `search()` utilizando los resultados del método `analyzeForSearch()`. El método `analyzeForSearch()` devuelve objetos `QueryName` y, dado que el objeto `QueryName` se deriva de la clase `Name`, cada elemento de la lista de objetos `QueryName` puede enviarse al método `search()`.

El método `createName()` genera el análisis de máxima confidencialidad del nombre y los entornos culturales recomendados tanto para los nombres de pila como para los apellidos, pero sólo si la categoría del nombre (`NameCategory`) indica un nombre de persona. Sin embargo, el método `analyzeForSearch()` proporciona hasta seis entornos culturales para cada nombre de pila y apellido, así como todos los análisis posibles para cada nombre conjunto. El método que utilice depende de los resultados que desee obtener.

## Crear objetos QueryName y buscar nombres

Este caso de ejemplo describe cómo crear objetos QueryName con el método analyzeForSearch() para preparar la búsqueda de nombres.

### Antes de empezar

Siga estas directrices para crear objetos QueryName que desee utilizar para funciones de búsqueda. Los objetos QueryName heredan información de la clase Name, pero sólo se crean mediante el método analyzeForSearch(). El método analyzeForSearch() categoriza, analiza, clasifica y translitera el nombre como el método createName(). Además, analyzeForSearch() devuelve hasta seis entornos culturales para cada nombre de pila y apellido y hasta tres posibles análisis para cada nombre compuesto.

### Procedimiento

1. Llame al método analyzeForSearch().
2. Opcional: Si desea alterar temporalmente los códigos de entorno cultural de un objeto QueryName, debe llamar al método createName() para crear un objeto Name con un entorno cultural diferente del suministrado por analyzeForSearch(). Puede utilizar el método createName() para crear varios objetos Name que utilizará el método search().
3. Llame al método search() para realizar las operaciones de búsqueda. La categoría del nombre (NameCategory) determina el conjunto de parámetros de comparación utilizados para la búsqueda.
  - a. Opcional: Invoque el método getDataListNames() para obtener los nombres de todas las listas de datos disponibles en el sistema.
  - b. Opcional: Llame al método getSearchStrategyNames() para obtener los nombres de todas las estrategias de búsqueda disponible en el sistema.
4. Opcional: Invoque el método dataFetch() para obtener los datos del nombre original y datos complementarios para los registros de nombre que están asociados a una coincidencia de nombre exclusivo.

### Resultados

Se devolverá una lista de objetos QueryName desde el método analyzeForSearch().

## Crear un objeto Name y buscar nombres

El siguiente caso de ejemplo de utilización describe cómo crear un objeto Name con el método createName() para preparar la búsqueda de nombres.

### Antes de empezar

Siga estas directrices para crear objetos Name de persona o de organización que desee utilizar para funciones de adición, actualización, búsqueda y comparación. No es necesario que especifique una categoría de nombres (de persona u organization) ya que IBM NameWorks utiliza NameSifter para categorizar el nombre automáticamente. Sin embargo, la categoría del nombre determina cómo se procesa. Por ejemplo, si especifica una NameCategory de persona, el método createName() transliterará, analizará y clasificará el nombre, mientras que en un nombre de tipo Organization sólo se realizará la transliteración.

## Procedimiento

1. Llame al método `createName()` para crear un objeto `Name`.

Opción	Descripción
Si no sabe nada del nombre	Pase una única serie de caracteres de nombre al método <code>createName()</code> . IBM NameWorks transliterará el nombre y determinará la categoría ( <code>NameCategory</code> ): <ul style="list-style-type: none"><li>• Si se determina que la categoría del nombre es de persona, IBM NameWorks lo analizará y clasificará.</li><li>• Si se determina que la categoría del nombre es <code>Organization</code>, recibirá un valor de entorno cultural ambiguo (<code>Ambiguous</code>).</li></ul>
Si sabe que la categoría del nombre es de persona o de organización	Especifique la categoría ( <code>NameCategory</code> ) del nombre, que determina cómo se procesa.
Si conoce el nombre de pila y el apellido	Pase estos valores al método <code>createName()</code> , que determinará que el nombre es de persona. Los campos de nombre se clasifican cuando el objeto <code>Name</code> solicita la información de entorno cultural.
Si conoce el nombre de pila, el apellido y el entorno cultural	Pase estos valores al método <code>createName()</code> , que asignará automáticamente una <code>NameCategory</code> de persona al nombre.

2. Llame al método `search()` para realizar las operaciones de búsqueda. La categoría del nombre (`NameCategory`) determina el conjunto de parámetros de comparación utilizados para la búsqueda.
- Opcional: Llame al método `getDatalistNames()` para obtener los nombres de todas las listas de datos disponibles en el sistema.
  - Opcional: Llame al método `SearchStrategyNames()` para obtener los nombres de todas las estrategias de búsqueda disponibles en el sistema.
3. Especifique el parámetro `searchOpt=` para indicar el tipo de `NameCategory` para el que desea realizar la búsqueda. Puede especificar este valor en la estrategia de búsqueda que se pasa a IBM NameWorks.
- 1 = Nombres de persona
  - 2 = Nombres de organización
  - 3 = Todas las categorías de nombres

## Resultados

Los resultados de la búsqueda se devuelven por medio del objeto `SearchResult`, que contiene una lista de objetos `SearchMatch`. Cada objeto `SearchMatch` contiene la información siguiente:

- Nombre de lista de datos en la que se ha encontrado el nombre coincidente
- Datos complementarios asociados con el nombre
- Puntuación de similitud de nombre completo
- Puntuación de similitud de nombre de pila
- Puntuación de similitud de apellido
- Si el nombre coincidente procede de un elemento alternativo
- Si el nombre coincidente procede de una entrada de nombre regularizado

- El número de registros de nombre coincidentes asociados con este nombre exclusivo

Dado que el objeto SearchMatch se deriva de la clase OriginalName, también devuelve la información siguiente:

- Categoría del nombre
- Nombre completo
- Nombre de pila
- Apellido
- Datos complementarios

---

## Buscar nombres mediante IBM NameWorks

IBM NameWorks da soporte a búsquedas de nombres completos y nombres exclusivos. En las búsquedas de nombres completos, IBM NameWorks devuelve en los resultados de búsqueda todos los registros de nombre coincidentes que aparecen en la lista de datos. En las búsquedas de nombres exclusivos, se devuelve una copia de cada nombre completo exclusivo coincidente, junto con una cuenta del número de veces que el nombre aparece en la lista de datos.

**Nota:** Antes de realizar la búsqueda con IBM NameWorks, el administrador del sistema debe asociar los archivos externos de nombres con las listas de datos de IBM InfoSphere Global Name Management mediante entradas en el archivo de configuración de IBM NameWorks.

La búsqueda de nombres exclusivos resulta de utilidad al buscar en listas de nombres muy grandes, en las que muchos nombres habituales pueden aparecer cientos o miles de veces. Esta repetición puede ser problemática para la lógica de búsqueda de IBM NameWorks y para los usuarios que, de otro modo, deberían revisar grandes volúmenes de resultados de búsqueda que muestran nombres idénticos. Después de realizar una búsqueda de nombres exclusivos, puede elegir qué nombres deben investigarse más a fondo mediante el método dataFetch().

Cada archivo físico externo también puede ser una entrada lógica separada (por ejemplo, un archivo para los clientes, un archivo para los empleados). Sin embargo, cada uno de estos archivos externos es un subconjunto de un archivo mucho mayor que se considera una colección lógica de nombres. La finalidad de dividir un archivo grande en un conjunto de archivos más pequeños es aprovechar el paralelismo de búsqueda, a fin de que IBM NameWorks y su componente de búsqueda, Distributed Search, puedan buscar simultáneamente en varios subconjuntos y luego consolidar todos los resultados en una sola respuesta. Esta asociación de física a lógica se realiza en el archivo de configuración de IBM NameWorks.

Puede utilizar la búsqueda incorporada para realizar búsquedas de nombres completos, en las que los datos de nombre se preprocesan cuando se inicializa IBM NameWorks, en lugar de hacerlo en un paso separado. La combinación de las prestaciones de búsqueda con el preproceso de nombres en un único proceso reduce la carga de trabajo de comunicación y administración que puede estar asociada con las búsquedas de nombres completos en listas de datos de gran tamaño. Los valores de las secciones [DataList] del archivo de configuración de IBM NameWorks determinan si una lista de datos es externa o está incorporada. Hay valores adicionales que controlan el tipo de preproceso de nombres realizado cuando se carga una lista de datos incorporada durante la inicialización.

## Gestionar estrategias de búsqueda

Para ayudar a los usuarios a buscar nombres de forma coherente y sencilla, defina el conjunto de valores de parámetros de comparación de búsqueda en las estrategias de búsqueda. Debe definir estrategias de búsqueda para cada búsqueda antes de utilizarlas durante las búsquedas de nombres.

### Acerca de esta tarea

Las estrategias de búsqueda pueden definirse en el archivo de configuración o en un objeto de estrategia que se crea mediante la clase Strategy. Puede especificar el nombre de la estrategia de búsqueda que desea utilizar al invocar la operación de búsqueda. Los parámetros de comparación que se definen en la estrategia de búsqueda se utilizan para la búsqueda.

### Estrategias de búsqueda

Una *estrategia de búsqueda* es una colección de valores de parámetros de comparación de búsqueda que el método search() de la clase Scoring puede utilizar. Las estrategias de búsqueda se utilizan para definir los conjuntos de valores de parámetros de comparación permitidos para el método search(). Cada estrategia de búsqueda tiene un nombre exclusivo.

El archivo de configuración que se suministra con IBM NameWorks contiene tres estrategias de búsqueda de ejemplo que los administradores del sistema pueden utilizar y personalizar:

- Estándar (Standard, contiene los valores de comparación predeterminados)
- Amplia (Broad, contiene valores de parámetros de comparación que amplían la búsqueda)
- Limitada (Narrow, contiene valores de parámetros de comparación que restringen la búsqueda)

Puede crear tantas estrategias de búsqueda como sea necesario, y la estrategia de búsqueda puede variar en función de las necesidades de la aplicación cliente y del tipo de búsqueda que desea ejecutar.

**Nota:** El parámetro de comparación MinScore corresponde al parámetro de comparación NAME\_THRESH de los releases anteriores, y no es necesario establecerlo por separado en otras secciones del archivo de configuración. Si se establecen ambos parámetros, IBM NameWorks utiliza MinScore y notifica NAME\_THRESH como error.

Los factores de ajuste relativo ajustan las puntuaciones calculadas para los parámetros de comparación. Estos factores aparecen después del nombre de parámetro como \_ADJ en los ejemplos siguientes.

### Ejemplo de una estrategia de búsqueda amplia:

Puede utilizar una estrategia de búsqueda amplia para lograr una recuperación mayor de nombres en una búsqueda, pero con una precisión inferior. Los resultados de la búsqueda incluirán más nombres en general pero devolverá un porcentaje menor de nombres relacionados con la búsqueda. El ejemplo siguiente muestra una estrategia de búsqueda amplia.

```
[Strategy:Broad]
  MinScore=65
```

```
[GNParms:Broad]
```

```
ANCHOR_FACTOR=0.97
COMPRESSED_SCORE_MAX=1.00
DO_COMPRESSED_SCORE=Y
FIELD_THRESH=0.50
FIELD_WEIGHT=0.50
NAME_UNKNOWN_SCORE=0.75
NO_NAME_SCORE=0.75
INITIAL_INITIAL_SCORE=1.00
INITIAL_TOKEN_SCORE=0.85
MATCH_INITIALS=Y
OOPS_FACTOR=0.95
```

```
[SNParams:Broad]
ANCHOR_FACTOR=0.97
COMPRESSED_SCORE_MAX=1.00
DO_COMPRESSED_SCORE=Y
FIELD_THRESH=0.50
FIELD_WEIGHT=0.50
NAME_UNKNOWN_SCORE=0.75
NO_NAME_SCORE=0.75
OOPS_FACTOR=0.95
```

### Estrategia de búsqueda limitada de ejemplo:

Puede utilizar una estrategia de búsqueda limitada para lograr una recuperación menor de nombres en una búsqueda, pero con una precisión superior. Los resultados de la búsqueda incluirán menos nombres en general pero devolverá un porcentaje más alto de nombres relacionados con la búsqueda. El ejemplo siguiente muestra una estrategia de búsqueda limitada.

```
[Strategy:Narrow]
MinScore=80
```

```
[GNParams:Narrow]
ANCHOR_FACTOR=0.85
COMPRESSED_SCORE_MAX=0.95
DO_COMPRESSED_SCORE=Y
FIELD_THRESH=0.70
FIELD_WEIGHT=0.50
INITIAL_INITIAL_SCORE=0.75
INITIAL_TOKEN_SCORE=0.70
NAME_UNKNOWN_SCORE=0.70
NO_NAME_SCORE=0.70
OOPS_FACTOR=0.85
```

```
[SNParams:Narrow]
ANCHOR_FACTOR=0.85
COMPRESSED_SCORE_MAX=0.95
DO_COMPRESSED_SCORE=Y
FIELD_THRESH=0.70
FIELD_WEIGHT=0.50
NAME_UNKNOWN_SCORE=0.70
NO_NAME_SCORE=0.70
OOPS_FACTOR=0.85
```

## Creación y modificación de estrategias de búsqueda mediante el archivo de configuración de IBM NameWorks

Los administradores del sistema crean o modifican las estrategias de búsqueda como precisen para facilitar las búsquedas de nombres o para establecer una correlación con un conjunto determinado de roles de usuario o reglas empresariales. Las estrategias de búsqueda se gestionan en el archivo de configuración de IBM NameWorks, por lo que es necesario volver a cargar el archivo de configuración de IBM NameWorks para que los cambios surtan efecto.

## Procedimiento

1. Abra el archivo de configuración de IBM NameWorks en un editor de texto.
2. Cree o modifique las secciones adecuadas del archivo de configuración.
  - Para crear una estrategia de búsqueda nueva, cree nuevas cabeceras de sección nuevas y proporcione las entradas y valores apropiados para cada cabecera de sección:

```
[Strategy:nombre]
  GNCulture=-1
  SNCulture=-1
  ONCulture=0
  MinScore=-1
  MaxReplies=-1
  SearchOpt=1
  IncludeTAQs=[GNParms:nombre] ...
[SNParms:nombre] ...
[ONParms:nombre] ...
```

*nombre* es el nombre exclusivo de la estrategia de búsqueda nueva. Cada cabecera de sección debe especificar el nombre apropiado de la estrategia de búsqueda, de lo contrario, las secciones se pasarán por alto.

La cabecera [Strategy:] es la única cabecera de sección necesaria para crear o ejecutar una estrategia de búsqueda. Si no incluye las otras cabeceras de sección, la estrategia de búsqueda utiliza los parámetros de búsqueda predeterminados para los parámetros de dichas secciones.

- Para modificar una estrategia de búsqueda existente, realice los cambios en las cabeceras apropiadas de las secciones de estrategia de búsqueda. Cada cabecera de sección debe especificar el nombre apropiado de la estrategia de búsqueda, de lo contrario, las secciones se pasarán por alto.
3. Guarde el archivo de configuración.
  4. Reinicie IBM NameWorks para volver a cargar el archivo de configuración a fin de que los cambios surtan efecto.

## Creación de estrategias de búsqueda mediante la clase Strategy

Puede utilizar la clase Strategy para definir una estrategia de búsqueda en lugar de la sección [Strategy:] del archivo de configuración de IBM NameWorks.

### Acerca de esta tarea

El siguiente procedimiento incluye ejemplos de código que utilizan lenguaje Java™, pero los pasos para crear una estrategia de búsqueda en C++ son los mismos.

## Procedimiento

1. Con la aplicación de desarrollo, cree un objeto Strategy y asígnele un nombre exclusivo.

```
Strategy broad = new Configuration.Strategy();
```

2. Llame a diversos métodos para especificar valores para los parámetros de comparación de búsqueda en la estrategia de búsqueda, como el umbral de puntuación mínima predeterminado, el número máximo de coincidencias que se debe devolver y las categorías de nombres de búsqueda. Por ejemplo,

```
broad.setMinScore(70)
.setMaxReply(1000)
.setSearchOptions(EnumSet.of(NameCategory.PERSONAL));
Configuration configuration = new Configuration();
configuration.addStrategy("Broad", broad);
```

3. Guarde el objeto Strategy.

4. Cree un objeto Configuration utilizando la clase Configuration y añada la estrategia de búsqueda al objeto Configuration.

### Supresión de estrategias de búsqueda del archivo de configuración

Si ya no utiliza una estrategia de búsqueda o la estrategia de búsqueda es redundante, suprima la estrategia de búsqueda del archivo de configuración. Debe volver a cargar el archivo de configuración a fin de que los cambios surtan efecto.

#### Procedimiento

1. Abra el archivo de configuración de IBM NameWorks en un editor de texto.
2. Busque la cabecera de la sección [Strategy:] que contenga el nombre de la estrategia de búsqueda que desee suprimir y suprima toda la información contenida en esa cabecera de sección. Asegúrese de suprimir todas las cabeceras de sección posibles para cada estrategia de búsqueda:

```
[Strategy:nombre]  
[GNParams:nombre] [SNParams:nombre] [ONParams:nombre]
```

*nombre* es el nombre exclusivo de la estrategia de búsqueda que desea suprimir.

3. Guarde el archivo de configuración de IBM NameWorks.
4. Reinicie IBM NameWorks para volver a cargar el archivo de configuración a fin de que los cambios surtan efecto.

## Alteración temporal de parámetros de comparación

IBM NameWorks permite reunir colecciones de parámetros de comparación en estrategias de búsqueda con nombre para proporcionar un manejo simplificado de la información de alteraciones temporales de búsqueda. Puede alterar temporalmente los parámetros de comparación predeterminados en función de la transacción creando estrategias de búsqueda diferentes.

### Opciones de alteraciones temporales de parámetros de comparación

Puede especificar alteraciones temporales de parámetros de comparación mediante el uso de estrategias de búsqueda en el archivo de configuración de IBM NameWorks, mediante la clase CompParamsOverride o la clase Configuration. Cada opción proporciona formas diferentes de especificar alteraciones temporales en función de cómo desea que la aplicación de cliente gestione la búsqueda de nombres.

Por ejemplo, supongamos que el umbral de campo predeterminado para los campos de apellido y nombre de pila de un nombre del entorno cultural anglosajón es 0.49 (FIELD\_THRESHOLD = 0.49). Puede especificar FIELD\_THRESHOLD = 0.60 en una estrategia de búsqueda para modificar el valor predeterminado que reside en el archivo de configuración de IBM NameWorks, aumentando el umbral de nombre en un 22%. Esta alteración temporal es absoluta porque especifica un valor nuevo para el parámetro de comparación.

De forma similar, puede incluir factores de ajuste relativo utilizados para ajustar las puntuaciones calculadas. Si especifica un factor de ajuste para el ejemplo anterior especificando FIELD\_THRESHOLD\_ADJ = 0.60, el nuevo valor se calculará multiplicando el valor predeterminado por el factor de ajuste:

$FIELD\_THRESHOLD = 0.49 * 0.60 = 0.29$

El valor ajustado, 0.29, se utiliza para el tiempo que dura la búsqueda, sin modificar el valor predeterminado en el archivo de configuración.

### **Alteraciones temporales de estrategia de búsqueda**

Puede crear estrategias de búsqueda para alterar temporalmente alguno de los parámetros de comparación predeterminados, todos ellos o ninguno. Las alteraciones temporales sólo se utilizan mientras dura la búsqueda, y vuelven a los valores predeterminados en la transacción siguiente.

Las alteraciones temporales especificadas en la estrategia de búsqueda son tanto absolutas como relativas, mientras que las alteraciones temporales en el lado del servidor son absolutas, como cuando se especifican a través de listas de datos que se utilizan con Embedded Search.

Como los valores de parámetros de comparación en el servidor son fijos, el valor `FIELD_THRESHOLD = 0.49` no se puede modificar a menos que el valor se cambie en el archivo de configuración de IBM NameWorks, que contiene los valores predeterminados de los parámetros de comparación. Al cambiar los valores de los parámetros de comparación en el servidor, todas las transacciones utilizan el valor actualizado a menos que los cambie en el archivo de configuración.

Si desea utilizar los parámetros de comparación predeterminados inherentes a Distributed Search o Embedded Search, pase una estrategia de búsqueda con un nombre vacío. Un nombre de estrategia de búsqueda vacío indica que deben utilizarse los parámetros de comparación predeterminados.

### **Alteraciones temporales según lista de datos**

Las alteraciones temporales según la lista de datos se aplican solamente a las listas de datos incorporadas. Puede especificar alteraciones temporales de parámetros de comparación para listas de datos individuales si añade una o más entradas `CompParmsDefaults=` a la sección `[DataList:]` del archivo de configuración. Si utiliza esta opción, puede establecer los valores predeterminados para listas de datos específicas, mientras que las estrategias de búsqueda y las alteraciones temporales que se indican mediante la clase `CompParmOverrides` sólo alteran temporalmente los valores predeterminados de una consulta individual.

### **Alteraciones temporales de la clase `CompParmOverrides`**

Puede utilizar la clase `CompParmOverrides` para proporcionar alteraciones temporales de consultas de búsqueda individuales. Esta clase contiene información de alteración temporal para una estrategia de búsqueda y no proporciona información sobre los valores predeterminados. Para indicar alteraciones temporales con esta clase, cree un objeto `CompParmOverrides` y utilice uno o varios métodos de alteración temporal (como `addSurnameOverride()`). Indique las alteraciones temporales como pares de nombre y valor de la misma manera que se especifican alteraciones temporales en las estrategias de búsqueda que se almacenan en los archivos de configuración de IBM NameWorks, o que se envían a los servidores de Distributed Search en mensajes XML. La clase `CompParmOverrides` admite las alteraciones temporales absolutas y relativas.

Las alteraciones temporales que especifique mediante esta clase se procesan cuando se envían a los métodos `Scoring.search()` y `Scoring.compare()`. Durante el proceso se informa de los errores en los

nombres o valores de los parámetros de comparación, que es cómo se procesan las alteraciones temporales en las estrategias de búsqueda.

### Alteraciones temporales de la clase Configuration

La clase Configuration sustituye el archivo de configuración de IBM NameWorks y ofrece la posibilidad de almacenar las configuraciones en una base de datos o especificar las configuraciones de forma dinámica. De forma alternativa, puede utilizar un archivo de configuración externo e indicar valores predeterminados en la sección [Strategy:] y después utilizar esta clase para especificar alteraciones temporales para listas de datos específicas. Puede utilizar una de las opciones siguientes para especificar alteraciones temporales con la clase Configuration:

- Crear una estrategia de búsqueda utilizando la clase Strategy y especificar alteraciones temporales utilizando los métodos de alteración temporal (como `addSurnameOverride()`) y luego añadir la estrategia de búsqueda al objeto Configuration con el método `addStrategy()`
- Utilizar un archivo de alteraciones temporales externo y llamar a dicho archivo utilizando el método `setDefaultCompParmOverrides()`

### Especificación de alteraciones temporales de parámetros de comparación mediante estrategias de búsqueda:

Puede alterar temporalmente parámetros de comparación de IBM NameWorks si especifica alteraciones temporales absolutas, alteraciones temporales relativas o alteraciones temporales por lista de datos en el archivo de configuración de IBM NameWorks o mediante la creación de un objeto Configuration utilizando la clase Configuration.

#### Acerca de esta tarea

Los factores de ajuste relativo ajustan las puntuaciones calculadas para los parámetros de comparación. Estos factores aparecen después del nombre de parámetro como `_ADJ`.

#### Procedimiento

1. Abra el archivo de configuración de IBM NameWorks en un editor de texto.
2. En la sección [Strategy:*nombre*] del archivo de configuración, donde *nombre* es el nombre de la estrategia de búsqueda, modifique los parámetros que desea alterar temporalmente.

Opción	Descripción
Para especificar alteraciones temporales absolutas	Sustituya el valor del parámetro de comparación que desea modificar por un valor nuevo.
Para especificar alteraciones temporales relativas	Añada <code>_ADJ</code> al parámetro que desea modificar y facilite un valor con el que desea ajustar el valor original.

3. Guarde la estrategia de búsqueda y reinicie IBM NameWorks.
4. Ejecute una nueva búsqueda para que las alteraciones temporales entren en vigor.

### **Especificación de alteraciones temporales mediante la clase `CompParamsOverrides`:**

Puede especificar alteraciones temporales de parámetros de comparación para IBM NameWorks mediante la creación de un objeto `CompParamsOverrides` e indicar qué alteraciones temporales desea incluir. A continuación, debe enviar las alteraciones temporales a los métodos `Scoring.search()` y `Scoring.compare()` para que se incluyan como parte de la comparación de nombres o búsqueda de nombres.

#### **Procedimiento**

1. Con una aplicación de desarrollo, cree un objeto `CompParamsOverrides`.  

```
CompParamsOverrides overrides = new CompParamsOverrides();
```
2. Llame a varios métodos para especificar alteraciones temporales de parámetros de comparación para el objeto `CompParamsOverrides`, donde *nombre* es el nombre del parámetro de comparación que desea alterar temporalmente y *valor\_nuevo* es el valor que desea utilizar.  

```
CompParamsOverrides overrides = new CompParamsOverrides();
overrides.addSurnameOverride("nombre", "valor_nuevo");
overrides.addSurnameOverride("OOPS_FACTOR", "0.60");
overrides.addSurnameOverride("NAME_UNKNOWN_SCORE", "0.45");
...
```
3. Pase el objeto `CompParamsOverrides` al método `Scoring.search()`, al método `Scoring.compare()`, o a los dos, para que se procesen como parte de la comparación de nombres y búsqueda de nombres.

### **Especificación de alteraciones temporales mediante la clase `Configuration`:**

Puede especificar alteraciones temporales de parámetros de comparación para IBM NameWorks mediante la clase `Configuration`. Puede crear estrategias de búsqueda utilizando la clase `Configuration` de la misma manera que crea secciones `[Strategy:]` en un archivo de configuración externo.

#### **Antes de empezar**

Debe crear un objeto `Configuration` como parte de la aplicación de cliente. Este objeto contiene la información de alteración temporal para los parámetros de comparación. Además, debe especificar valores asociados a la búsqueda de nombres a través de una de las opciones siguientes:

- Utilice un archivo de configuración externo que contenga valores predeterminados para los parámetros de IBM NameWorks
- Utilice una estrategia de búsqueda que puede crear mediante la clase `Strategy`

#### **Procedimiento**

1. Abra el objeto `Configuration` en la aplicación de desarrollo.
2. Indique las alteraciones temporales que desea implementar.

Opción	Descripción
Si utiliza un archivo de configuración externo	<p>Especifique alteraciones temporales en el objeto Datalist mediante el método <code>setDefaultCompParmOverrides()</code>.</p> <p>El ejemplo siguiente crea un objeto <code>Configuration</code> que contiene una lista de datos denominada <code>Customers</code>, con un archivo de alteraciones temporales llamado <code>compparms.config</code>:</p> <pre>Configuration configuration = new Configuration(); Datalist customers = configuration.addDatalist("Customers"); customers.setDefaultCompParmsOverridesFile("compparms.config");</pre>
Si utiliza la clase <code>Strategy</code>	<p>Especifique alteraciones temporales en el objeto <code>Strategy</code> mediante uno o varios métodos de alteración temporal.</p> <p>El ejemplo siguiente especifica una estrategia de búsqueda <code>broad</code> (amplia) e incluye una alteración temporal de apellido:</p> <pre>Configuration configuration = new Configuration(); Strategy broad = configuration.addStrategy("Broad"); broad.addSurnameOverride("OOPS_FACTOR", "0.60");</pre>

3. Guarde la aplicación de desarrollo y reinicie IBM NameWorks.
4. Ejecute una nueva búsqueda para que las alteraciones temporales entren en vigor.

## Preparar nombres para la búsqueda

Utilice el método `analyzeForSearch()` para preparar un nombre para utilizarlo en una transacción de búsqueda. Este método categoriza el nombre determinando si se trata de un nombre de persona o de un nombre de organización, analiza los nombres de persona en campos de nombre de pila y apellido y determina un código de clasificación cultural para cada campo en un nombre de persona. Si `NameSifter` determina que el nombre es un nombre de organización, solamente se devuelve la información sobre la categoría del nombre.

### Acerca de esta tarea

El método `analyzeForSearch()` da soporte a series de caracteres de entrada solamente en codificación UTF-16. Si la serie de caracteres de entrada es una codificación no soportada, IBM NameWorks señala un error o produce resultados no deseados. Para preparar un nombre para la búsqueda, pase los valores siguientes al método `analyzeForSearch()` de la clase `Scoring` de IBM NameWorks.

- El nombre completo que se va a analizar, representado como un nombre completo. Por ejemplo, ROBERT E JONES. El nombre debe pasarse al método `parse()` como un valor *string*.
- Un entero entre 0 y 100 que representa el valor **`alternateThreshold`** (0 siempre suprime desgloses alternativos). Este valor es el valor de fiabilidad mínimo que es necesario para un análisis inicial antes de que se reordene del nombre y se tomen en cuenta desgloses alternativos.

### Resultados

El método `analyzeForSearch()` primero categoriza y luego translitera el nombre. La transliteración de nombres es el proceso de convertir un nombre de un sistema de escritura o convenio de codificación de caracteres determinado a otro. Por ejemplo, la transliteración de nombres permite analizar y comparar un nombre escrito en árabe con un nombre similar escrito en el alfabeto latino. Luego, el método analiza la serie de caracteres de entrada y devuelve una colección de

objetos QueryName para cada serie de caracteres de entrada y su versión transcrita.

## **Categorizar nombres, comparar nombres y comparar fechas mediante IBM NameWorks**

Puede utilizar IBM NameWorks para categorizar nombres como de persona o de organización, comparar dos nombres de persona o comparar fechas.

### **Categorías de nombres**

Durante el proceso de nombres, éstos se asocian con una categoría de nombres, que puede ser de persona o de organización. Aunque pueden compartir una utilización similar, los nombres de estas dos categorías está separados por importantes diferencias y, por tanto, se aplican tipos diferentes de recursos lingüísticos y de datos de referencia a los nombres de cada categoría durante el análisis y la comparación.

Al categorizar nombres, los componentes de IBM InfoSphere Global Name Management los sitúan en las categorías siguientes:

- Nombres de persona, que no contienen indicadores que sugieran su pertenencia a ninguna otra categoría (por ejemplo: "Linda K. Smith")
- Nombres de organización, que contienen algún tipo de indicador no de persona (por ejemplo: "Smith & Company")
- Nombres desconocidos, que contienen algún elemento que parece estar mal escrito o que contienen alguna otra construcción que no aparece normalmente en los nombres de persona o de organización (por ejemplo: "SMI")
- Ambos, que son nombres que contienen un calificador profesional que puede sugerir que se trata de un nombre de empresa derivado de un nombre de persona (por ejemplo: "Linda Smith Architect")

Si un nombre se categoriza como un nombre no de persona, el componente suministra un código de razón que identifica el indicador o patrón que cualifica al nombre como no de persona.

### **Nombres de persona:**

Un nombre de persona consta de uno o varios nombres de pila, los nombres de familia o grupo (por ejemplo nombres de tribu o clan) u otros elementos de tipo apellido utilizados en el entorno cultural del que procede el nombre y las fórmulas de tratamiento y otros calificadores de nombre asociados con el titular del mismo. Un nombre de persona completo hace referencia a un individuo y puede codificar información que indique la clase social, los antecedentes religiosos y políticos, el nivel de formación, el origen étnico o cultural y la procedencia regional.

### **Modelo de nombres de persona de IBM InfoSphere Global Name Management**

Para describir nombres de persona y trabajar con ellos independientemente de su formato nativo, es importante utilizar una terminología coherente. También es muy importante poder desglosar los nombres de forma coherente en sus componentes constituyentes, para poder comparar los componentes equivalentes.

El formato del modelo de nombres de persona de IBM InfoSphere Global Name Management está motivado por la necesidad de tratar con los nombres tal como están codificados en conjuntos de datos en la realidad. Es un procedimiento práctico para determinar la estructura de un nombre. Por ejemplo, aunque los

nombres en muchas partes del mundo no tienen verdaderos apellidos en el sentido del entorno cultural occidental, sin embargo dichos nombres se colocan en bases de datos que asumen apellidos. Por tanto, con la finalidad de realizar un proceso de nombres coherente, IBM InfoSphere Global Name Management aplica una estructura de dos campos. El campo al que pertenecen los diversos componentes de un nombre está determinado en parte por la frecuencia con la que cada componente del nombre se ha asociado con un campo de nombre de pila o de apellido. Dentro de cada campo, los elementos de nombre individuales se desglosan en unidades más grandes. Por ejemplo, el apellido “de la Salle” se reconoce como una frase de nombre formada por una raíz de nombre principal y dos prefijos, no como tres componentes de nombre separados.

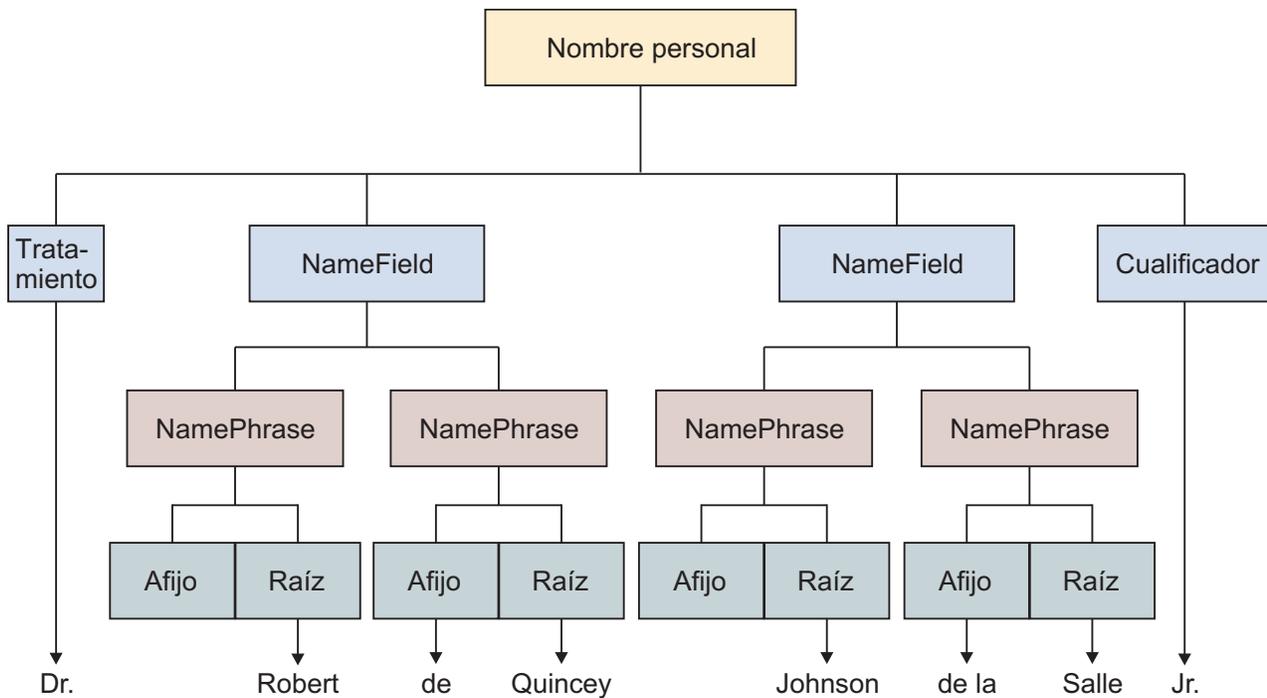


Figura 3. Modelo de nombres de persona de IBM InfoSphere Global Name Management

### Estructura y componentes de los nombres de persona

Los nombres de persona pueden contener componentes muy diferentes. Estos componentes y el modo en que se estructuran difieren en función de los grupos culturales.

A continuación se indican algunos de los componentes que pueden utilizarse en los nombres de persona:

- Nombre de pila
- Apellido
- Nombre de familia
- Nombre de tribu, clan o casta
- Indicadores de parentesco o linaje (por ejemplo, patronímicos (nombres derivados del nombre del padre), matronímicos (nombres derivados del nombre de la madre), tecnonímicos (nombres derivados del nombre de un hijo) e indicadores generacionales)
- Calificadores que indican el orden de nacimiento, el género, la religión o la confesión religiosa

- Fórmulas de tratamiento
- Partículas (como por ejemplo "bin" (hijo) y "al" (el) en árabe o "de" en español y francés)

La estructura de los nombres de persona o el orden de los componentes del nombre también varían según el país o el grupo cultural.

A continuación figuran algunos ejemplos de estructuras de nombre:

**Nombre(s) de pila + Apellido**

- Megan Marie Andrews (Europeo)
- Fereshteh Gholamzadeh (Iraní)
- Rattima Nitisaroj (Tailandés)
- Hasan Incirlioglu (Turco)

**Apellido + Nombre de pila**

- Lim Yauw Tjin (Chino)
- Pak Mi-Ok (Coreano)
- Suzuki Ichiro (Japonés)

**Apellido + Segundo nombre + Nombre de pila**

- Trinh Van Thanh (Vietnamita)

**Nombre de pila + Nombre de pila del padre**

- Ahmed bin Eisa (algunas comunidades árabes)
- Abdurrahman Wahid (Indonesia)
- Mahmud bin Haji Basir (Malasia)

**Nombre de pila + Patronímico (Nombre del padre) + Apellido**

- Ivan Andreyevich Saratov (Rusia)
- Basimah Ali Al-Qallaf (algunos países árabes)

**Nombre tribal + Nombre religioso**

- WOUKO Philomene (Camerún)

**Sólo Nombre de pila**

- Sukarno (Indonesia)
- Habibullah (Afganistán)

**Referencia al nombre del hijo**

- Abu Hassan (que se traduce literalmente como *padre de Hassan*, países árabes)

**Nombres de organización:**

Un *nombre de organización* es un nombre no de persona que hace referencia a un cuerpo estructurado formado por una o más personas cuyo objetivo es realizar alguna función común. Las organizaciones pueden ser empresas, clubes, escuelas, agencias gubernamentales, partidos políticos o manifestaciones de la World Wide Web. Los nombres de organización incluyen generalmente algún tipo de indicador, patrón o palabras que ayuden a identificarlos como nombres no de persona.

Por lo general, aunque no siempre, los nombres de organización incluyen alguna palabra o frase que indica su función, como por ejemplo "escuela secundaria", "fontanería", "departamento de policía" o "banco".

Los nombres de organización también contienen un *elemento de denominación* o alguna serie de caracteres, palabras o frases que identifica de forma exclusiva esta organización entre otras del mismo tipo. Por ejemplo, “*First Union Bank*,” “*Joe’s Italian Restaurant*,” “*AAA Auto Wash*.” Algunas organizaciones, como por ejemplo las empresas, están reguladas por los gobiernos y tienen elementos de nombre prescritos que indican su estado de registro, como por ejemplo “*PTY*” o “*LTD*”.

Los tipos de cadenas de caracteres y combinaciones de las mismas que se encuentran en los nombres de organización no se parecen generalmente ni utilizan los mismos patrones que las de los nombres de persona. Estos patrones corresponden a códigos (denominados códigos de razón de categoría de nombres) que identifican la razón por la que un nombre se ha clasificado como nombre de organización, en lugar de como nombre de persona. Estos códigos de razón no definen un nombre de organización, sino que indican patrones que no se esperarían en un nombre de persona. Por ejemplo, una serie de tres consonantes idénticas en una fila (por ejemplo “*DDD*”) sería muy inusual en un nombre de persona, pero no es inhabitual en los nombres de organización.

Cuando los componentes de IBM InfoSphere Global Name Management categorizan un nombre, si éste coincide con uno o más códigos de razón de categoría de nombres, se presupone que es un nombre de organización. De lo contrario, es candidato a nombre de persona.

### **Categorización de nombres como de persona o no de persona utilizando IBM NameWorks**

Puede categorizar nombres de una lista de datos como nombres de persona o no de persona (nombres de organización) como paso previo antes de analizar, clasificar, analizar o buscar nombres.

#### **Acerca de esta tarea**

Para categorizar un nombre como de persona o no de persona, pase el nombre que se debe categorizar al método `categorize()` de la clase `Scoring` de IBM NameWorks.

#### **Resultados**

El método `categorize()` examina el nombre y devuelve la categoría del nombre que se seleccionó, el código de razón de la categoría del nombre (identifica la razón por la que se seleccionó la categoría), y un grado de confianza entre 0 y 100 que indica el porcentaje probable de que el nombre se categorice correctamente. Cuanto más cercano a 100 sea el porcentaje, mayor será el grado de confianza de la categorización del nombre.

#### **Qué hacer a continuación**

Después de categorizar nombres, puede utilizar los nombres que estén identificados como nombres de persona en IBM NameWorks para analizar, clasificar o buscar.

#### **Códigos de razón de categoría de nombre**

Los códigos de razón de categorización de nombres identifican qué tipo de indicador o patrón no personal se ha encontrado. Proporcionan una explicación de porqué se ha elegido la categoría. Puede utilizar códigos de razón para realizar un análisis más detallado, o para crear sus propias categorizaciones de nombres basadas en los códigos de razón.

Tabla 9. Códigos de razón de categorías de nombres para productos de IBM InfoSphere Global Name Management y descripciones

Código de razón de categoría de nombre	Descripción
BadData	El nombre era demasiado largo.
UrlEnding	El nombre contiene un indicador de URL de Internet habitual, tal como ".COM", ".ORG" o ".NET".
EstateOf	El nombre contiene las palabras "patrimonio de".
KnownOrg	El contiene una frase organizativa conocida.
Phrase	El nombre contiene una frase de organización solamente.
NoTokens	El nombre no contiene ninguna cadena de caracteres de nombre conocido.
AndCompany	El nombre contiene alguna modalidad del indicador "& Company".
MultipleInitials	El nombre contiene varias iniciales.
SingleSequence	El nombre contiene una secuencia de letras individual.
NameAndName	El nombre contiene "nombre & nombre".
LeadingToken	La cadena de caracteres aparece solamente al comienzo del nombre.
Triplet	El nombre contiene un triplete inicial formado por una sola letra.
NforAnd	El nombre contiene "n for and".
SingleHyphen	El nombre contiene un guión entre letras individuales.
MultipleHyphen	El nombre contiene varios guiones.
MultiSlash	El nombre contiene varias barras inclinadas.
Enumeration	El nombre contiene una enumeración, tal como "1 <sup>o</sup> " o "2 <sup>o</sup> ".
Possessive	El nombre contiene un posesivo sajón, tal como "Smith's" o "Jones".
OrgWord	El nombre contiene una palabra de organización conocida.
HyphOrgWord	El nombre contiene una palabra de organización conocida que incluye guiones.
AllSymbols	La cadena de caracteres contiene símbolos solamente.
ConsPlusC	El nombre consta de una cadena de caracteres formada por consonantes solamente más una palabra de tipo C.
CPL	El nombre contiene una palabra de tipo C, una preposición y una ubicación.
TwoTypeC	El nombre contiene dos palabras de tipo C.
LandC	El nombre contiene una ubicación y una palabra de tipo C.
AandC	El nombre contiene un adjetivo y una palabra de tipo C.
TandC	El nombre contiene una palabra de tipo T y una palabra de tipo C.
TwoTypeT	El nombre contiene dos palabras de tipo T.
LandT	El nombre contiene una palabra de tipo T y una ubicación.
TwoPreps	El nombre contiene dos preposiciones.
PrepL	El nombre contiene una preposición y una ubicación.

Tabla 9. Códigos de razón de categorías de nombres para productos de IBM InfoSphere Global Name Management y descripciones (continuación)

Código de razón de categoría de nombre	Descripción
PrepT	El nombre contiene una preposición y una palabra de tipo T.
PrepC	El nombre contiene una preposición y una palabra de tipo C.
Parens	El nombre contiene un paréntesis.
NonAlpha	Una cadena de caracteres del nombre contiene caracteres no alfabéticos.
AllCons	La cadena de caracteres del nombre contiene consonantes solamente.
Name1andName2	El nombre contiene el patrón " <i>nombre1</i> " & " <i>nombre2</i> ".
AndName	El nombre contiene el patrón "... & <i>nombre</i> ".
LLC	El nombre contiene un calificador de profesión y " <i>LLC</i> ".
ProfQual	El nombre contiene un calificador de profesión, tal como " <i>Arquitecto</i> ".
FallThru	El nombre no ha pasado ninguna de las pruebas.

## Comparación de dos nombres mediante IBM NameWorks

Puede utilizar IBM NameWorks para determinar el grado de similitud de dos nombres entre sí.

### Acerca de esta tarea

Utilice el método `compare()` de la clase `Scoring` de IBM NameWorks para comparar dos nombres mediante el pase de dos objetos `Name` al método. Cada objeto `Name` es una serie de caracteres individual (Persona u Organización) o dos series de caracteres que contienen el nombre de pila y apellido de los nombres a comparar.

El método `compare()` translitera los dos nombres y los compara de acuerdo con el valor `NameCategory` del nombre de la consulta. Los parámetros de comparación (`CompParms`) utilizados para la comparación están basados en el valor `NameCategory` de los dos operandos:

- Ambos nombres son de persona: se aplican los parámetros de comparación asociados al primer nombre (lado izquierdo).
- Ambos nombres son de organización: se aplican los parámetros de comparación asociados al primer nombre (lado izquierdo).
- Un nombre es personal y otro es un nombre de organización: los nombres se comparan como nombres de organización y se aplican los parámetros de comparación de nombre de organización.
- Uno de los nombres no tenía una categoría de nombre asociada: se utiliza `NameSifter` para determinar la categoría de nombre. Los parámetros de comparación utilizados están basados en la categoría de nombre elegida.

Si la categoría de nombre es de persona, se analiza y clasifica el nombre. Se clasifica cada campo del nombre de persona y los códigos de entorno cultural del nombre de la consulta se fusionan en un solo código de entorno cultural global, que se utiliza para seleccionar los parámetros predefinidos correctos para la comparación.

## Resultados

El objeto CompareData devuelve puntuaciones de comparación para los dos nombres que están comprendidas entre 0 y 100, donde 0 denota ausencia de similitud y 100 representa una similitud exacta.

### Comparación de dos fechas mediante IBM NameWorks

Muchas tareas de búsqueda de identidades o de verificación dependen de dos elementos esenciales de información: un nombre y una fecha de nacimiento (date of birth, DOB). IBM NameWorks asegura que muchas tareas de búsqueda puedan estar basadas completamente en sus funciones al permitir hacer comparaciones para cada uno de estos dos tipos de datos.

#### Antes de empezar

Las fechas deben ser series de ocho caracteres que contengan dígitos numéricos con el formato AAAAMMDD.

#### Acerca de esta tarea

Para comparar dos fechas, utilice el método dateCompare() de la clase Scoring de IBM NameWorks, y pase al método una serie de caracteres apareada que contenga las dos fechas a comparar.

## Resultados

El método dateCompare() compara las dos fechas y devuelve una puntuación de similitud comprendida dentro del rango de 0 a 100, donde 0 es el grado de similitud menor y 100 es el grado de similitud mayor. Si los argumentos de fecha no se proporcionan con el formato necesario (AAAAMMDD), el método devuelve el valor -1.

### Determinar la diferencia entre dos valores de fecha utilizando IBM NameWorks

Puede utilizar IBM NameWorks para calcular la diferencia entre dos valores de fecha.

#### Antes de empezar

Los valores de fecha deben ser series de ocho caracteres que contengan dígitos numéricos con el formato AAAAMMDD.

#### Acerca de esta tarea

Para determinar la diferencia entre dos valores de fecha, utilice el método dateDifference() de la clase Scoring de IBM NameWorks, y pase al método una serie de caracteres apareada que contenga los dos valores de fecha.

## Resultados

El método dateDifference() calcula y devuelve la diferencia entre dos valores de fecha. Si los argumentos de fecha no se proporcionan con el formato necesario (AAAAMMDD), el método devuelve el valor -1.

## Buscar nombres en una lista de datos

La búsqueda de nombres se realiza en las listas de datos que el administrador del sistema ha configurado en el archivo de configuración de IBM NameWorks para utilizarlas con IBM NameWorks. Cuando especifica el nombre de una o más listas de datos en la solicitud de búsqueda, el sistema devuelve los datos de los nombres coincidentes de acuerdo con el tipo de búsqueda (valor *Type=*) configurado para la lista de datos, ya sea una búsqueda de nombres completos o una búsqueda de nombres exclusivos. Los resultados procedentes de varias listas de datos se combinan para formar una sola respuesta.

### Antes de empezar

- Debe conocer el nombre de la lista de datos en la que se debe buscar. (Utilice el método `getDatalistNames()` para devolver una lista de todos los nombres de lista de datos existentes, si es necesario.)
- Debe conocer el nombre de la estrategia de búsqueda que se debe utilizar como parte de la búsqueda. (Si es necesario, utilice el método `getSearchStrategyNames()` para listar las estrategias de búsquedas disponibles actualmente).
- Prepare el nombre de persona para la búsqueda utilizando el método `analyzeForSearch()`, el cual desglosa el nombre en los campos correspondientes al nombre de pila y apellido, y devuelve un código de entorno cultural individual para cada campo de nombre. Este método también devuelve un valor para el objeto `QueryName`, el cual se utiliza como parte de la búsqueda.

### Acerca de esta tarea

Para buscar un nombre en una lista de datos, utilice el método `search()` de la clase `Scoring` de IBM NameWorks, y pase al método los valores siguientes:

- Un ID de transacción que identifica una solicitud de búsqueda determinada. (Si pasa al método el valor `-1`, IBM NameWorks asigna un valor exclusivo que es devuelto en los resultados de la búsqueda. Este valor se incluye en las entradas del archivo de registro y se puede utilizar para especificar esta operación de búsqueda específica en el archivo de registro).
- El valor de retorno resultante del objeto `QueryName`, que incluye información de campo de nombre y entorno cultural utilizado como criterios de búsqueda
- Los nombres de las listas de datos en las que se debe buscar
- El nombre de la estrategia de búsqueda que se debe utilizar
- Un entero que representa la categoría que desea buscar. Un valor de `-1` indica que debe utilizarse el valor predeterminado de Search Strategy (estrategia de búsqueda).
  - 1 Nombre de persona
  - 2 Nombre de organización
  - 3 Nombres de persona y de organización
- Un número entero que representa el número máximo de respuestas o coincidencias que se deben devolver. Este número limita el número de coincidencias de búsqueda que se deben devolver, seleccionando por filtración las coincidencias de nivel superior. Un valor entre 0 y 1 devuelve todas las coincidencias, y un valor de `-1` indica que debe utilizarse el valor predeterminado de Search Strategy (estrategia de búsqueda).
- Un número entero comprendido entre 0 y 100 que representa la puntuación mínima (valor `nmScore`) que los nombres contenidos en la lista de datos deben

alcanzar o sobrepasar para ser devueltos como coincidencia de búsqueda. (Si se especifica un valor menor que cero, se trata como un 0. Un valor mayor que 100 genera una excepción de parámetro no válido (GODW031E)).

**Nota:** Si el nombre que está buscando es un nombre habitual, es conveniente limitar el número de resultados que se deben devolver y utilizar un valor más alto para la puntuación mínima. De lo contrario, es posible que se excedan los almacenamientos intermedios de mensajes para los resultados devueltos.

## Resultados

El método `search()` realiza el tipo de búsqueda que está configurado para la lista de datos y devuelve la lista de resultados clasificada de acuerdo con la puntuación del nombre completo.

## Qué hacer a continuación

Si la lista de datos está configurada para búsquedas de nombres exclusivos, puede utilizar el método `dataFetch()` para recuperar los datos complementarios para todos los registros de nombres que están asociados a un nombre exclusivo.

## Recuperar datos complementarios para nombres asociados a un nombre exclusivo

Después de realizar una búsqueda en una lista de datos configurada para búsquedas de nombres exclusivos, puede ser conveniente examinar el detalle asociado a los nombres incluidos en la cuenta de nombres exclusivos devuelta. Utilice el método `dataFetch()` para recuperar datos complementarios para todos los registros de nombres que están asociados a un nombre exclusivo.

### Antes de empezar

- Previamente debe haber realizado una búsqueda de nombres exclusivos, utilizando el método `search()`.
- Debe conocer el nombre de la lista de datos donde reside el nombre exclusivo. (Esta información es devuelta por la operación de búsqueda).

### Acerca de esta tarea

Para buscar un nombre en una lista de datos, utilice el método `search()` de la clase `Scoring` de `IBM NameWorks`, y pase al método los valores siguientes:

- El nombre de la lista de datos donde reside el nombre exclusivo
- El valor de clave que sirve para identificar un nombre exclusivo determinado. (Este valor es devuelto por la operación de búsqueda).

## Resultados

El método `dataFetch()` recupera datos complementarios para todos los registros de nombres que están asociados a un nombre exclusivo, basándose en el valor de clave.

---

## Buscar nombres mediante NameHunter

El componente NameHunter puede ampliar la búsqueda de nombres en las aplicaciones de su organización. NameHunter da soporte a solicitudes del tipo "recuperar los 10 nombres más parecidos a X", "recuperar los nombres que coincidan con Y con una puntuación similar al 90% o más" o "¿cuál es el grado de similitud entre el nombre A y el nombre B?".

### Visión general de NameHunter

La biblioteca de programación de NameHunter® incluye funciones y clases que permiten a los desarrolladores agregar búsquedas de nombres de persona y de organizaciones a una aplicación nueva o existente.

Las API de NameHunter ofrecen a la aplicación la posibilidad de dar soporte a peticiones de usuario tales como "Suministrarme los 10 nombres más parecidos a JAMES SLESINGER de mi lista de nombres", "Mostrarme todos los nombres de una base de datos que coinciden con JOHN WONG con una similitud del 90% o más" o "Indicarme el grado de similitud entre PAUL VANESANN y P. VANLESANN".

NameHunter utiliza técnicas integradas de lingüística, probabilística y similitud de series para obtener resultados de búsqueda que van más allá de los suministrados por las mediciones estándar de similitud de series, como por ejemplo la distancia de edición o los algoritmos de agrupación de nombres, como por ejemplo los de Soundex o NYSIIS estándar.

Las bibliotecas de NameHunter están codificadas en C++ estándar y pueden integrarse en cualquier aplicación escrita en C++. Por tanto, la biblioteca de NameHunter puede utilizarse en cualquier plataforma que dé soporte a un compilador C++. NameHunter se ha diseñado para ofrecer simplicidad, facilitar la integración y proporcionar una flexibilidad y ampliabilidad máximas del entorno de ejecución.

**Nota:** NameHunter puede procesar un máximo de seis cadenas de caracteres por campo de nombre. Todas las cadenas de caracteres después de las primeras seis cadenas de caracteres se ignoran.

### Búsquedas de NameHunter configurables y específicas de entorno cultural

Cada búsqueda individual realizada por NameHunter puede configurarse ajustando diversos parámetros de comparación de tiempo de ejecución.

Cada parámetro de búsqueda controla un aspecto determinado de la comparación efectuada entre dos nombres al determinar el grado de similitud. Los parámetros básicos establecen umbrales para determinar el parecido que deben tener dos nombres para ser considerados coincidentes. Otros parámetros controlan las reglas lingüísticas especializadas que se aplican a entornos culturales específicos. Los valores óptimos de estos diversos parámetros de control de búsqueda se determinan con detalle mediante:

- El análisis detallado de la naturaleza de los datos de nombres buscados
- El tipo de consultas realizadas por los usuarios
- Las reglas empresariales que definen la tolerancia de precisión y recuperación de la organización.

En la práctica, el desarrollador o el usuario final pueden empezar con los valores predeterminados que IBM ha establecido para cada parámetro de control de búsqueda de tiempo de ejecución, y luego realizar ajustes según sea necesario. Los parámetros de comparación específicos del entorno cultural se encapsulan en la estructura de datos CompParms.

El producto de API NameHunter incluye determinados paquetes de parámetros predefinidos, cada uno de los cuales está adaptado para búsquedas efectivas de nombres de un entorno cultural o étnico determinado. Por ejemplo, los nombres hispánicos tienen con frecuencia determinadas características (por ejemplo, apellidos compuestos como TORRES DE LA CRUZ), que generalmente causan problemas cuando se procesan con métodos de búsqueda convencionales. Los parámetros de comparación hispánicos contienen valores que resuelven problemas específicos de este tipo de nombres. Los nuevos parámetros de comparación los pueden crear los usuarios para entornos culturales personalizados y los parámetros existentes se pueden modificar para ajustar las búsquedas.

### **Datos de fórmulas de tratamiento, afijos y calificadores (TAQ)**

IBM proporciona una lista de fórmulas de tratamiento, afijos y calificadores (TAGQ) que se utilizan en la coincidencia de nombres. Los TAQ no se descartan al realizar una comparación de nombres, ya que pueden contribuir a la puntuación global de nombre dependiendo de los parámetros de búsqueda.

Puede optar por cargar o no la lista de TAQ (taq.ibm) por medio del archivo de configuración, aunque el valor predeterminado es cargar la lista de TAQ. También puede suministrar una lista de variantes de nombres de pila y apellidos personalizada para sustituir o complementar la lista de IBM. Las listas de TAQ y variantes deben ser archivos de texto delimitado por comas que contengan los parámetros siguientes:

#### **Archivo TAQ:**

El archivo TAQ contiene toda la información necesaria para puntuar los datos TAQ.

Puede proporcionar sus propios archivos TAQ para sustituir o complementar las listas predeterminadas. También tiene la posibilidad de alterar temporalmente la información de los archivos TAQ de valores predeterminados.

La primera entrada del archivo indica la versión del archivo que NameHunter utiliza para determinar si el formato del archivo es correcto. Si falta la cabecera, NameHunter presupone que el archivo está destinado a una versión anterior a la 4.2 y que sólo contiene TAQ de cadenas de caracteres únicas y no variantes, factores o TAQ de texto con varias cadenas de caracteres.

Si realiza migraciones desde una versión anterior a la 4.2, debe ejecutar el programa de utilidad de conversión ConvertTaq.

La versión 4.2 del producto introduce las siguientes consideraciones en el archivo TAQ:

- Dígitos (0-9)
- Caracteres especiales: @ \$ % & + = /
- Los guiones, puntos, comas, separadores y líneas nuevas son equivalentes a los espacios

Este archivo utiliza cabeceras .ini para separar las secciones.

**[version]**

La primera sección del archivo indica la versión del archivo que NameHunter utiliza para determinar si el formato del archivo es correcto. Si falta la cabecera, NameHunter presupone que el archivo está destinado a una versión anterior a la 4.2 y que sólo contiene TAQ de cadenas de caracteres únicas y no variantes, factores o TAQ de términos con varias cadenas de caracteres.

Si está realizando una migración desde una versión anterior a la 4.2, puede utilizar el programa de utilidad de conversión ConvertTaq para actualizar los archivos existentes.

**[taqs]** Esta sección del archivo identifica los TAQ utilizando el formato siguiente:

TaqText, TaqType, TaqPosition, Culture

**TaqText**

El texto de cadena de caracteres.

**TaqType**

Tipo de cadena de caracteres de TAQ. Se admiten los siguientes tipos de TAQ.

**3 PREFIX**

Cadena de caracteres que se incluye en la misma frase de nombre que la cadena de caracteres raíz de nombre subsiguiente. "DE" y "LA" son prefijos.

**4 SUFFIX**

Cadena de caracteres que se incluye en la misma frase de nombre que la cadena de caracteres raíz de nombre anterior. "ALDEEN" es un sufijo (en nombres árabes).

**5 TITLE**

Cadena de caracteres anexo a un nombre que generalmente indica una posición social o profesional. "MR" y "GEN" son fórmulas de tratamiento. Las fórmulas de tratamiento no se incluyen en los campos de nombre de pila ni de apellido, sino que se colocan en un campo propio.

**6 QUALIFIER**

Cadena de caracteres anexo a los nombres que generalmente indica relaciones generacionales o una posición social o profesional. "JR" y "ESQ" son calificadores. Los calificadores no se incluyen en los campos de nombre de pila ni de apellido, sino que se colocan en un campo propio.

**9 ORGANIZATION DESIGNATOR**

Palabras como "Inc.", "SRL" o "Sociedad de responsabilidad limitada" que identifican el tipo de organización.

**10 PROFESSIONAL QUALIFIER**

Palabras como "M.D." o "CPA" que describen las características profesionales de un nombre.

**11 STOPWORD**

Una palabra que no figura en el cálculo. Por ejemplo, una búsqueda de "Tienda de mascotas" coincidirá con la entrada de datos "La Tienda de mascotas", ya que "La" no afecta al cálculo.

## 12 ORGANIZATION AFFIX

Una palabra que no añade ningún significado descriptivo al nombre. Sin embargo, a diferencia de las cadenas de caracteres Stopword, las cadenas de caracteres Organization Affix se contabilizan como factores en los cálculos de nombre.

### TaqPosition

Posición de un TAQ en el nombre. Por norma, primero aparece el tratamiento, los calificadores son los últimos, y los prefijos y sufijos en cualquier posición. Por ejemplo, "Junior" es un calificador, pero solo cuando aparece en la última posición de un nombre.

- E Cualquier posición
- F Primera posición
- L Última posición

### Entorno cultural

La lista delimitada por dos puntos de códigos culturales. Indica el entorno cultural o grupo étnico asociado con este TAQ. Puede especificar varios entornos culturales para una única entrada TAQ personal. Si desea aplicar los TAQ a todas las culturales, especifique A como código **Culture**. Este valor debe ser uno de los entornos culturales soportados actualmente por NameHunter y NameClassifier.

- 0 Ambiguo
- 1 Anglosajón
- 2 Árabe
- 3 Chino
- 4 Hispano
- 5 Coreano
- 6 Ruso
- 7 Francés
- 8 Alemán
- 9 Tailandés
- 10 Indonesio
- 11 Yoruba
- 12 Farsi
- 13 Paquistaní
- 14 Hindú
- 15 Japonés
- 16 Afgana

17	Vietnamita
18	Polaco
19	Portugués
20	Turco
38	Sudoeste asiático
39	Europeo
40	Han
A	Todos

**[variants]**

Esta sección identifica las variantes TAQ utilizando el formato siguiente:  
score,variant:variant,culture:culture

**score** La puntuación que debe asignarse cuando las variantes coinciden. La puntuación debe estar comprendida entre 0,0 y 1,0.

**variant(s)** Lista delimitada por dos puntos de variantes TAQ relacionadas. Por ejemplo, AMBASSADOR:AMB.

**culture(s)** La lista delimitada por dos puntos de códigos culturales. Puede especificar varios entornos culturales para una sola entrada de variante TAQ. Este valor debe ser uno de los entornos culturales soportados actualmente por NameHunter y NameClassifier.

0	Ambiguo
1	Anglosajón
2	Árabe
3	Chino
4	Hispano
5	Coreano
6	Ruso
7	Francés
8	Alemán
9	Tailandés
10	Indonesio
11	Yoruba
12	Farsi
13	Paquistaní
14	Hindú
15	Japonés
16	Afgana
17	Vietnamita
18	Polaco

19	Portugués
20	Turco
38	Sudoeste asiático
39	Europeo
40	Han
A	Todos

El ejemplo siguiente muestra un pequeño subconjunto del archivo TAQ que se suministra con el producto IBM InfoSphere Global Name Management para ayudarle a construir su propio archivo TAQ.

```
[version]
gnr 4.2
```

```
[taqs]
#PERSONAL TAQS
#format - TaqText,TaqType,TaqPosition,Culture1,Culture2,CultureN
#TaqPosition Typically, Titles are F, Qualifiers are L, Prefixes and
#Suffixes are E
ABD,3,E,0:1:18
MAYOR,5,F,10
LETNAN COLONEL,5,F,10
```

```
#ORGANIZATION TAQS
#format - TaqText,TaqType,TaqPosition,Culture(always 0)
#Unambiguous TAQ TaqPosition values are code E (for every position);ambiguous
#taqs are restricted to L (for last position)
&,12,E,0
A B,9,L,0
A PROFESSIONAL CORPORATION,9,L,0
```

```
[variants]
#PERSONAL TAQ VARIANTS
#format -#Factor,Variant1:Variant2:VariantN,Culture1:Culture2:CultureN
1.0,AMBASSADOR:AMB,0:1:3:4:5:6:7:8:9:10:11:12:13:14:15:16:17
1.0,SR:SENIOR,0:1:2:3:5:6:7:8:9:10:11:12:13:14:15:16:17
1.0,DELAS:DE LAS,4:7
```

```
#ORGANIZATION TAQ VARIANTS
#format - Variant1,Variant2,Factor,Culture
COMPANIA,COMPANHIA,.99,0
COMPANY,COMPANIES,.99,0
```

```
#ORGANIZATION TAQ VARIANTS
#format -#Factor,Variant1:Variant2:VariantN,Culture1:Culture2:CultureN
.99,COMPANY:COMPANIES,0
.99,COMPANY:COMPAGNIE:COMPANIA:COMPANHIA,0
```

```
[taqFactors]
#format - TaqType,FactorType,Factor,Culture
3,1,0.97,A
3,2,0.98,A
```

```
#TaqType
# Prefix = 3
# Suffix = 4
# Title = 5
# Qualifier = 6
# OrganizationDesignator = 9
# ProfessionalQualifier = 10
# StopWord = 11
# OrganizationAffix = 12
```

```

#FactorType
# Different = 1
# Missing   = 2

#Culture Codes
# Ambiguous   = 0
# Anglo       = 1
# Arabic      = 2
# Chinese     = 3
# Hispanic    = 4
# Korean      = 5
# Russian     = 6
# French      = 7
# German      = 8
# Thai        = 9
# Indonesian = 10
# Yoruban    = 11
# Farsi       = 12
# Pakistani  = 13
# Indian      = 14
# Japanese   = 15
# Afghan      = 16
# Vietnamese = 17
# Polish      = 18
# Portuguese  = 19
# Turkish     = 20
# SouthwestAsian = 38
# European   = 39
# Han        = 40
# All        = A

```

*Programa de utilidad de conversión converttaq.exe:*

El programa de utilidad de la línea de mandatos **converttaq.exe** convierte los archivos TAQ anteriores a la versión 4.2. Si realiza migraciones desde una versión anterior a la versión 4.2, debe ejecutar el programa de utilidad de conversión **converttaq.exe** en el archivo TAQ existente.

```
converttaq <input_pre_4.2_TAQ_filename> <output_converted_4.2_TAQ_filename>
```

Durante el proceso de conversión, el nuevo valor de parámetro **TaqPosition** adopta el valor predeterminado del valor de E (cada posición).

## Variantes de cadenas de caracteres

Una *variante de cadena de caracteres de nombre* es una alternativa de un nombre especificado que se considera equivalente a ese nombre, pero que difiere de él en su forma externa particular. Normalmente las variantes surgen de variaciones en la escritura. NameHunter suministra funciones para cargar estos archivos.

IBM suministra listas exhaustivas de variantes de cadenas de caracteres de nombre (por ejemplo, Peggy = Margaret) para apellidos y nombres de pila. Las variantes se suministran en el directorio de datos y se denominan gnv.ibm (variantes de nombre de pila), snv.ibm (variantes de apellido) y onv.ibm (variantes de organización).

Puede suministrar sus propias listas de variantes de nombres de pila y apellidos para sustituir o complementar las listas de IBM. También tiene la posibilidad de alterar temporalmente la información de los archivos de variantes existentes para suprimir pares de variantes individuales o modificar la puntuación asignada a un par de variantes. Por ejemplo, considere las siguientes entradas de variantes:

### **variante de nombre de pila**

0.95,ROBERT:ROB:ROBBY:BOB:BOBBY:BERT,0:1

Esta entrada asocia los nombres de variantes de nombre de pila, con una puntuación relacionada de 0,95 para los entornos culturales 0 y 1, que son el genérico y el anglosajón.

### **variante de apellido**

0.95,OYANG:OYEUNG:OYONG:OYOUNG:AU YAN:AU YANG:AU YEUNG:AU YONG:AU YOUNG:  
AW YAN:AW YANG:AW YEUNG:AW YONG:AW YOUNG:OU YAN:OU YANG:OU YEUNG:OU YONG:  
OU YOUNG:OW YAN:OW YANG:OW,3

Esta entrada asocia los nombres de variantes de apellido listados, con una puntuación relacionada de 0.95 para el entorno cultural 3, que es el chino.

### **variante de nombre de organización**

1,WRECKER:WRCKR,0

Esta entrada asocia los nombres de variantes de organización listados, con una puntuación relacionada de 1 para el entorno cultural 0, que es genérico.

### **Archivo de variantes de nombre:**

El archivo de variantes de nombres contiene toda la información que es necesaria para hallar variantes de nombre alternativas y equivalentes.

Puede proporcionar su propio archivo de variantes para sustituir o complementar la lista de valores predeterminados. También tiene la posibilidad de alterar temporalmente la información en el archivo de variantes predeterminado.

Si migra desde una versión anterior a la 4.2, debe ejecutar el programa de utilidad de conversión **convertvar.exe**.

Este archivo utiliza cabeceras .ini para separar las secciones.

#### **[version]**

La primera sección del archivo indica la versión del archivo que NameHunter utiliza para determinar si el formato del archivo es correcto. Si falta la cabecera, NameHunter presupone que el archivo está destinado a una versión anterior a la 4.2 y sólo contiene texto de cadenas de caracteres únicas.

Si está realizando una migración desde una versión anterior a la 4.2, puede utilizar el programa de utilidad de conversión **convertvar** para actualizar los archivos.

#### **[variants]**

Esta sección identifica las variantes utilizando el formato siguiente:

score,variant:variant,culture:culture

**score** La puntuación que debe asignarse cuando las variantes coinciden. La puntuación debe estar comprendida entre 0,0 y 1,0.

#### **variant(s)**

La lista delimitada por dos puntos de variantes de cadenas de caracteres de nombres relacionados (por ejemplo, BOB:ROBERT).

#### **culture(s)**

La lista delimitada por dos puntos de códigos culturales. Puede

especificar varios entornos culturales para una sola entrada de variante. Este valor debe ser uno de los entornos culturales soportados actualmente por NameHunter y NameClassifier.

0	Genérico
1	Anglosajón
2	Árabe
3	Chino
4	Hispano
5	Coreano
6	Ruso
7	Francés
8	Alemán
9	Tailandés
10	Indonesio
11	Yoruba
12	Farsi
13	Paquistaní
14	Hindú
15	Japonés
16	Afgano
17	Vietnamita
18	Polaco
19	Portugués
20	Turco
38	Sudoeste asiático
39	Europeo
40	Han
A	Todos

El ejemplo siguiente muestra el formato para ayudarle a construir su propio archivo de variantes de nombre.

```
[version]
gnr 4.2.1
```

```
[variants]
0.95,ROBERT:ROB:ROBBY:BOB:BOBBY:BERT,0:1
0.95,ELIZABETH:ELIZBETH:LIZ:LIZZY:LIZZIE:LISA:LIBBY:LIBBIE,0:1
0.95,ELISABETH:LIESEL:LIESE:LIESCHEN:LIL,8
```

*Programa de utilidad de conversión convertvar.exe:*

El programa de utilidad de línea de mandatos **convertvar.exe** convierte los archivos de variantes de nombre anteriores a la versión 4.2. Si realiza migraciones

desde una versión anterior a la 4.2, debe ejecutar el programa de utilidad de conversión **convertvar.exe** en el archivo de variantes de nombre existente.

```
convertvar <input_pre4.2_filename> <output_converted4.2_filename>
```

## Términos

Los *términos* son cadenas de caracteres de nombres que se refieren a un concepto. Se utilizan sobre todo para nombres de organización.

El texto del término puede ser una única o varias cadenas de caracteres. La mayoría de las veces los términos se componen de texto de varias cadenas de caracteres. Un término de varias cadenas de caracteres es una secuencia de dos o más palabras (texto con un espacio en blanco incorporado) que juntas se refieren a un solo concepto. El archivo de términos define un conjunto de términos de varias cadenas de caracteres que se utiliza en comparaciones de búsqueda. Los elementos del archivo de términos son tratados por NameHunter como si fueran una sola palabra delimitada por un carácter distinto del espacio. Por ejemplo:

### Término de varias cadenas de caracteres

```
TRACTOR TRAILER,1.0,0
```

"tractor trailer" se trata como si fuera "tractor\_trailer".

### Archivo de términos:

El archivo de términos contiene toda la información necesaria para identificar los términos de organización.

IBM proporciona un archivo de términos de organización predeterminado. Los términos de organización residen en el directorio de datos y se denominan `onterms.ibm`. Puede proporcionar su propio archivo de términos para sustituir o complementar la lista de valores predeterminados. También tiene la posibilidad de alterar temporalmente la información en el archivo de términos predeterminado.

Las entradas del archivo de términos no contienen variantes. Los términos de varias cadenas de caracteres con formas variantes se definen en el archivo de variantes de la organización. El término real `estate` tiene la variante `r| est`. Cada valor se define en un archivo de términos y como variantes entre sí dentro del archivo de variantes de organización.

Este archivo utiliza cabeceras `.ini` para separar las secciones.

#### [version]

La primera sección del archivo indica la versión del archivo que NameHunter utiliza para determinar si el formato del archivo es correcto.

#### [terms]

Esta sección identifica los términos utilizando el formato siguiente:

```
termtext,weight,culture:culture
```

#### termtext

Texto que se debe tratar como término.

El texto del término puede ser una única o varias cadenas de caracteres. La mayoría de las veces los términos se componen de texto de varias cadenas de caracteres. Un término de varias cadenas de caracteres es una secuencia de dos o más palabras (texto con un espacio en blanco incorporado) que juntas se refieren a un solo concepto.

**weight**

Especifica la contribución relativa que un término hace a una puntuación de coincidencia.

El valor debe establecerse en 1.0 .

**culture(s)**

Lista delimitada por dos puntos de los códigos de entorno cultural, que indican el grupo étnico o cultural asociado a este término. Este valor debe ser uno de los entornos culturales que admiten NameHunter y NameClassifier.

0	Genérico
1	Anglosajón
2	Árabe
3	Chino
4	Hispano
5	Coreano
6	Ruso
7	Francés
8	Alemán
9	Tailandés
10	Indonesio
11	Yoruba
12	Farsi
13	Paquistaní
14	Hindú
15	Japonés
16	Afgano
17	Vietnamita
18	Polaco
19	Portugués
20	Turco
38	Sudoeste asiático
39	Europeo
40	Han
A	Todos

El ejemplo siguiente muestra el formato para ayudarle a construir su propio archivo de términos.

```
[version]  
gnr 4.2.1
```

```
[terms]  
ANIMAL CLINIC,1.0,0  
APPLIANCE REPAIR,1.0,0
```

AUTO SALES,1.0,0  
AUTO REPAIR,1.0,0  
AUTO SALES,1.0,0  
BAY AREA,1.0,0  
BEAUTY SALON,1.0,0  
CHILD CARE,1.0,0  
COMMUNITY CHURCH,1.0,0  
ELEMENTARY SCHOOL,1.0,0  
FARM BUREAU,1.0,0  
FIRE DEPT,1.0,0  
HAIR CARE,1.0,0  
HEATING & AIR,1.0,0  
HEATING & AIR COND,1.0,0  
LAWN CARE,1.0,0  
LITTLE LEAGUE,1.0,0  
NAIL CARE,1.0,0  
REAL ESTATE,1.0,0  
SECURITY SYSTEMS,1.0,0

## Regularización de nombres

Para los nombres de persona, la regularización de nombres es una característica que genera escrituras basadas en el sonido de cadenas de caracteres de nombres, para correlacionar diferentes escrituras con la misma pronunciación a la misma forma o a una forma similar. El efecto de estas escrituras regularizadas es permitir que NameHunter identifique nombres que por lo general se consideran relacionados, aunque su escritura pueda ser bastante diferente.

Por ejemplo, los nombres Layton y Leighton se normalizan como Laten. Cuando se comparan estos nombres, se devuelve una coincidencia exacta aunque la puntuación está limitada al parámetro de comparación de puntuación máxima de nombres regularizados (REGULARIZE\_SCORE\_MAX).

Para nombres de organizaciones, la regularización correlaciona dígitos y símbolos para sus formas representadas. En español, por ejemplo, las reglas garantizan que *162, ciento sesenta y dos, uno seis dos y uno sesenta y dos* coincidan todas.

La regularización de nombres la controlan los archivos de reglas de regularización de nombres. Se suministran los siguientes archivos de reglas de regularización de nombres:

### **angloRegRule.ibm**

Para nombres de persona identificados con el entorno cultural anglosajón. También se utiliza para entornos culturales genéricos.

### **angloOnRegRule.ibm**

Para nombres de organización identificados con el entorno cultural anglosajón.

### **chineseRegRule.ibm**

Para nombres de persona identificados con el entorno cultural chino.

### **chineseNativeOnRegRule.ibm**

Para nombre de organización identificados con el entorno cultural chino, incluidas las reglas especiales para números y otro tipo de texto escrito en el script hanzi.

### **chineseOnRegRule.ibm**

Para nombres de organización identificados con el entorno cultural chino.

### **farsiRegRule.ibm**

Para nombres de persona identificados con el entorno cultural farsi.

**frenchRegRule.ibm**

Para nombres de persona identificados con el entorno cultural francés.

**genericOnRegRule.ibm**

Para nombres de organización identificados con un entorno cultural sin especificar (Ambiguo).

**germanRegRule.ibm**

Para nombres de persona identificados con el entorno cultural germano.

**hispanicRegRule.ibm**

Para nombres de persona identificados con el entorno cultural hispánico.

**hispanicOnRegRule.ibm**

Para nombres de organización identificados con el entorno cultural hispánico.

**indianRegRule.ibm**

Para nombres de persona identificados con el entorno cultural hindú.

**indoRegRule.ibm**

Para nombres de persona identificados con el entorno cultural indonesio.

**japaneseRegRule.ibm**

Para nombres de persona identificados con el entorno cultural japonés.

**japaneseNativeOnRegRule.ibm**

Para nombre de organización identificados con el entorno cultural japonés, incluidas las reglas especiales para números y otro tipo de texto escrito en el script kanji.

**koreanRegRule.ibm**

Para nombres de persona identificados con el entorno cultural coreano.

**koreanOnRegRule.ibm**

Para nombres de organización identificados con el entorno cultural coreano.

**polishRegRule.ibm**

Para nombres de persona identificados con el entorno cultural polaco.

**polishOnRegRule.ibm**

Para nombres de organización identificados con el entorno cultural polaco.

**portugueseRegRule.ibm**

Para nombres de persona identificados con el entorno cultural portugués.

**portugueseOnRegRule.ibm**

Para nombres de organización identificados con el entorno cultural portugués.

**russianRegRule.ibm**

Para nombres de persona identificados con el entorno cultural ruso.

**russianOnRegRule.ibm**

Para nombres de organización identificados con el entorno cultural ruso.

**swasianRegRule.ibm**

Para nombres de persona identificados con los entornos culturales del sudeste asiático. Sustituye a `arabicRegRule.ibm`.

**thaiRegRule.ibm**

Para nombres de persona identificados con el entorno cultural thai.

### **turkishRegRule.ibm**

Para nombres de persona identificados con el entorno cultural turco.

Los clientes que elijan crear culturas personalizadas pueden añadir sus propios archivos de reglas de regularización personalizadas. Sin embargo, no se ofrece desarrollo y soporte de los archivos de regularización personalizados como parte del producto Global Name Management.

Si habilita la regularización, el rendimiento resultará afectado. Si se añade un nombre a una lista de datos y la regularización encuentra una forma normalizada, se añadirá una segunda entrada a la lista de datos. Además, si se normaliza un nombre de consulta, ambas formas se compararán con todos los nombres de la lista de datos. Si se habilita esta característica, los tiempos de búsqueda serán aproximadamente del doble.

### **Integrar la API NameHunter en las aplicaciones**

Para integrar la API NameHunter en las aplicaciones, utilice el archivo de cabecera NameHunter.h, que contiene una definición completa de la API.

Existe otro archivo de cabecera en el directorio include, ConfigHandler.h; sin embargo, sólo se utiliza en algunas de las aplicaciones de ejemplo. Puede utilizarlo, pero no es necesario para utilizar NameHunter.

La implementación de NameHunter se encuentra en una biblioteca de objetos; sin embargo, NameHunter utiliza varios componentes compartidos de IBM para dar soporte a la transliteración y regularización de nombres. La mayoría de ellos provienen de los ICU (International Components for Unicode) de IBM. Las bibliotecas son:

#### **NameHunter.lib**

la API NameHunter

#### **NameTransliterator.lib**

una biblioteca utilizada para la transliteración y regularización de nombres

#### **sicudata.lib**

tablas de datos de ICU

#### **sicui18n.lib**

bibliotecas de internacionalización de ICU

#### **sicuuc.lib**

biblioteca Unicode común de ICU

Los nombres de archivo reales variarán en función de la plataforma. En Windows, los nombres serán como los anteriores, y en plataformas Unix serán de tipo "libNameHunter.a". Consulte el directorio bin de su plataforma para obtener los nombres de archivo exactos.

### **Enlace a otros datos**

NameHunter almacena y realiza búsquedas en datos de nombre. La mayoría de sistemas almacenan y utilizan otros tipos de datos además de los nombres, por ejemplo: direcciones, números de cuenta, atributos físicos e imágenes. Por esta razón, NameHunter suministra un campo de ID para cada entrada de la clase SearchList.

Al añadir un registro a una SearchList, puede suministrar un índice/puntero a sus datos en el campo de ID, como por ejemplo un índice de base de datos. El campo

de ID de NameHunter puede tener 256 bytes de longitud como máximo y estar formado por cualquier combinación de caracteres ASCII.

## Ejemplos de inicio rápido de la API NameHunter

Estos ejemplos de inicio rápido ofrecen varios pequeños programas de trabajo que muestran las funciones básicas de NameHunter.

### Ejemplo de inicio rápido de NameHunter: Comparar dos nombres

A continuación figura un ejemplo muy sencillo de un programa en funcionamiento que utiliza la API NameHunter. Compara dos nombres y notifica si coinciden. Compila y crea una instancia de NameHunter y llama a la función NameHunter::nameMatch. Tenga en cuenta que algunas de las características más potentes de NameHunter, como por ejemplo las variantes, no están habilitadas.

```
#include <NameHunter.h>
#include <iostream>

using namespace LAS;
using namespace LAS::NH;
using namespace std;

int main(int argc, char* argv[])
{
    // NameHunter lanzará (principalmente durante el inicio)
    try
    {
        // siempre es necesaria una instancia de NameHunter
        NameHunter nh;

        // llamar a nameMatch utilizando los compParms predeterminados
        if (nh.nameMatch("jack", "johnson jr",
                        "john j", "de la jones" ))
            cout << "los nombres coinciden" << endl;
        else
            cout << "los nombres no coinciden" << endl;
    }
    catch (const exception& e)
    {
        cerr << "Excepción capturada - " << e.what() << endl;
        return 1;
    }

    return 0;
}
```

### Ejemplo de inicio rápido de NameHunter: Puntuar dos nombres

A continuación figura un ejemplo sencillo que informa de la similitud entre dos nombres mediante la función NameHunter::nameScore. No se utilizan características especiales.

```
#include <NameHunter.h>
#include <iostream>

using namespace LAS;
using namespace LAS::NH;
using namespace std;

int main(int argc, char* argv[])
{
    // NameHunter lanzará (principalmente durante el inicio)
    try
    {
        // siempre es necesaria una instancia de NameHunter
```

```

NameHunter nh;

char* gn1 = "jack";
char* sn1 = "johnson jr";
char* gn2 = "john j";
char* sn2 = "de la jones";

// llamar a nameScore con los compParms predeterminados
ScoreInfo score = nh.nameScore(gn1, sn1, gn2, sn2);

cout << "comparing " << gn1 << " " << sn1
      << " to " << gn2 << " " << sn2 << endl;
cout << " name score = " << score.name << endl;
cout << " gn score = " << score.gn << endl;
cout << " sn score = " << score.sn << endl;
}
catch (const exception& e)
{
    cerr << "Excepción capturada - " << e.what() << endl;
    return 1;
}

return 0;
}

```

### Ejemplo de inicio rápido de NameHunter: Puntuar dos nombres utilizando entornos culturales, TAQ y variantes diferentes

A continuación figura un ejemplo mucho más real. Se puntúan dos nombres utilizando todos los entornos culturales posibles de NameHunter. También se utilizan TAQ de NameHunter (fórmulas de tratamiento, afijos y calificadores) y variantes (John = Jack).

```

#include <NameHunter.h>
#include <iostream>

using namespace LAS;
using namespace LAS::NH;
using namespace std;

int main(int argc, char* argv[])
{
    // NameHunter lanzará (principalmente durante el inicio)
    try
    {
        // siempre es necesaria una instancia de NameHunter
        NameHunter nh;

        // suponiendo que estos archivos se encuentren en la vía de acceso.
        // En caso contrario, se lanzará una excepción.
        nh.loadTaq("taq.ibm");
        nh.loadVariants("gnv.ibm", GivenName);
        nh.loadVariants("snv.ibm", SurName);

        char* gn1 = "jack";
        char* sn1 = "johnson jr";
        char* gn2 = "john j";
        char* sn2 = "de la jones";

        for (int i = 0; i < MaxCultureNum; ++i)
        {
            CompParms gnParms((Culture)i, GivenName);
            CompParms snParms((Culture)i, SurName);
            ScoreInfo score = nh.nameScore(gn1, sn1,
                                           gn2, sn2,
                                           &gnParms, &snParms);
            cout << " for culture, " << nh.cultureName((Culture)i) << endl;
        }
    }
}

```

```

        cout << "    name score = " << score.name << endl;
        cout << "    gn score = " << score.gn << endl;
        cout << "    sn score = " << score.sn << endl;
    }
}
catch (const exception& e)
{
    cerr << "Excepción capturada - " << e.what() << endl;
    return 1;
}

return 0;
}

```

## Ejemplo de inicio rápido de NameHunter: Searchlist and Search

Normalmente, deseará almacenar la base de datos de nombres en un objeto SearchList de NameHunter y utilizar un objeto Searcher de NameHunter para buscar en la lista con un nombre de consulta. A continuación figura un ejemplo sencillo del paradigma SearchList, Searcher. Tenga en cuenta que en este ejemplo no se utilizan entornos culturales, ajustes de parámetros de comparación ni ninguna de las demás características especiales de NameHunter.

```

#include <NameHunter.h>
#include <iostream>

using namespace LAS;
using namespace LAS::NH;
using namespace std;

int main(int argc, char* argv[])
{
    // NameHunter lanzará (principalmente durante el inicio)
    try
    {
        // siempre es necesaria una instancia de NameHunter
        NameHunter nh;

        // suponiendo que estos archivos se encuentren en la vía de acceso
        nh.loadTaq("taq.ibm");
        nh.loadVariants("gnv.ibm", GivenName);
        nh.loadVariants("snv.ibm", SurName);

        // crear una lista de búsqueda y añadir algunos nombres
        SearchList searchList(&nh);

        // añadir entradas con el formato GN, SN, ID, donde ID es
        // un campo que puede utilizar para asociar datos propios con
        // este nombre (por ejemplo, fecha de nacimiento, color favorito, etc.).
        searchList.add("john j", "de la jonson", "1");
        searchList.add("jack", "john sr", "2");
        searchList.add("j", "johnson", "3");
        searchList.add("george", "smith", "4");

        // crear un buscador y una lista de resultados
        Searcher searcher(&nh);
        ResultList results;

        char* gn = "jack";
        char* sn = "johnson jr.";

        searcher.search(searchList,
                       results,
                       gn,
                       sn);

        // imprimir el nombre de consulta y todas las coincidencias.
    }
}

```

```

cout << "hits for, " << sn << ", " << gn << endl;
for (size_t i = 0; i < results.size(); ++i)
    cout << " "
        << results[i].sn << ", "
        << results[i].gn << " - "
        << results[i].score.name << endl;
}
catch (const exception& e)
{
    cerr << "Excepción capturada - " << e.what() << endl;
    return 1;
}

return 0;
}

```

## Aplicaciones de ejemplo de NameHunter

Se proporcionan aplicaciones de ejemplo con el paquete de distribución de NameHunter que muestra diversas formas de utilizar la API. Puede utilizar estas aplicaciones de ejemplo como base para empezar a desarrollar sus propias aplicaciones.

### Aplicación de ejemplo search de NameHunter

La aplicación de NameHunter search es un programa de línea de mandatos que compara dos archivos de nombres y graba los resultados en un tercer archivo. Para ejecutar esta aplicación, abra un indicador de mandatos, vaya hasta el directorio /support/bin y escriba search.

Debe especificar valores para parámetros globales, establecer los parámetros de comparación de nombre o apellido y especificar la ubicación de los archivos de datos siguientes en el archivo de configuración search.config. También puede utilizar este archivo de configuración para cargar los archivos TAQ y de variantes y configurar la regularización y la transliteración. Los archivos de datos siguientes son archivos de texto delimitados por comas que el archivo de configuración search.config invoca.

#### queryFile

Incluye los nombres que se buscan.

#### nameFile

Incluye los nombres que se buscan.

#### resultFile

Contiene los registros de nombres coincidentes.

El archivo de configuración search.config se encuentra en el directorio /support/data, junto con diversos otros archivos que la aplicación de ejemplo search utiliza.

### Aplicación de ejemplo why de NameHunter

La aplicación de ejemplo why de NameHunter es un programa de utilidad de línea de mandatos que determina la razón por la que NameHunter ha llegado a una puntuación determinada para dos nombres. Para ejecutar esta aplicación, abra un indicador de mandatos, vaya hasta el directorio /support/bin y escriba why.

Puede modificar el archivo de configuración why.config para cambiar los valores CompParm de NameHunter y especificar los nombres para comprar. Después de modificar el archivo de configuración, guarde el archivo y escriba un punto (.) en el indicador de mandatos. El programa de utilidad why vuelve a cargar el archivo

de configuración y compara los nombres sin tener que reiniciar la aplicación. Los resultados de la comparación de nombres se imprimen en la salida de origen de NameHunter.

## Modificar parámetros de comparación

Los parámetros de comparación o *CompParms* son un conjunto de parámetros ajustables que se utilizan para dirigir y puntuar la comparación entre nombres especificados en una consulta y un nombre contenido en una lista de datos designada.

El personal técnico encargado de la integración y optimización de los resultados de búsqueda de NameHunter debe comprender las consecuencias que están asociadas al ajuste de uno o varios de estos parámetros. Un conocimiento exhaustivo y detallado de los valores de los parámetros de comparación permite configurar NameHunter para que funcione con la máxima eficacia para una aplicación determinada y una lista de datos específica de nombres y datos asociados.

Los valores de los parámetros de comparación se comprueban y establecen mediante un grupo de llamadas de API afines que se proporcionan en el kit de herramientas de desarrollo de NameHunter. Cuando la funcionalidad de NameHunter se accede mediante la aplicación de NameHunter Distributed, los valores de los parámetros de comparación también se pueden ajustar en NameHunter Server (NHServer), ya sea incluyendo valores específicos dentro del mensaje de parámetros con formato XML o proporcionando un código de entorno cultural para utilizar los valores predeterminados de IBM.

La presente información contiene muchos ejemplos de valores de parámetros y resultados de búsqueda de ejemplo. Estos resultados proceden de búsquedas reales con los datos de ejemplo proporcionados con NameHunter Distributed Search. En algunos casos, la lista de resultados se ha reducido a un tamaño manejable.

### Visión general de los parámetros de comparación de NameHunter

Los parámetros de comparación (CompParms) de NameHunter forman una estructura de datos abstracta que proporciona una infraestructura de control persistente, de tiempo de ejecución, para comparaciones entre un nombre especificado en una consulta y cada nombre sucesivo de una lista de datos (base de datos o archivo de registros) que el usuario define como objetivo de la búsqueda basada en NameHunter.

Los controles de búsqueda de los parámetros de comparación están basados en un modelo de datos básico para nombres de persona. Se ha creado un modelo de nombre formado por dos componentes para nombres procedentes de una gran diversidad de orígenes lingüísticos y culturales. Gracias a este modelo de nombre subyacente, se pueden utilizar diferentes parámetros de comparación de NameHunter para el nombre de pila y apellido.

La API NameHunter incluye varios paquetes de valores predefinidos de parámetros de comparación que se han probado que funcionan eficazmente con nombres de grupos importantes. Estos paquetes de parámetros culturales representan valores predeterminados alternativos de CompParms, e incluyen un paquete genérico para ser utilizado con nombres que no están claramente asociados a un entorno cultural específico.

El archivo `compParms.config` contiene las alteraciones temporales de los parámetros de comparación para uno o varios entornos culturales. El archivo de configuración

contiene los valores culturales del nombre de pila y apellido para nombres de persona y un conjunto único de valores para los nombres de organización. Distributed Search lee el archivo de configuración, compila el conjunto correspondiente a un entorno cultural e invoca el método `overrideDefaultParms()` de NameHunter para establecer los nuevos valores predeterminados. Puede recuperar los valores predeterminados de los parámetros de comparación actuales llamando al método `getDefaultParms()`.

NameWorks también puede consultar el archivo `compparms.config`, en el que aplica un nivel de instancia y ello afecta tanto a la comparación Embedded Search como a la comparación por pares (`Scoring.compare()`).

El proceso de puntuación y evaluación se produce en los siguientes niveles consecutivos de abstracción:

- Nivel de nombre completo
- Nivel de campo de nombre (nombre de pila o apellido)
- Nivel de segmento (signo o frase de nombres)

Esta agrupación está pensada para facilitar el diseño y mantenimiento del software de aplicaciones que hacen uso de funciones de NameHunter. La arquitectura en capas de `CompParm` simplifica el acceso administrativo o de usuario en tiempo de ejecución a la secuencia de proceso de nombres de NameHunter, por lo que las aplicaciones habilitadas para NameHunter pueden responder de forma flexible y efectiva a los cambios en las necesidades del usuario y a los cambios (cualitativos o cuantitativos) en la base de datos de nombres en la que se debe buscar.

#### **Factor de puntuación de análisis alternativo:**

Los usuarios pueden penalizar las comparaciones realizadas en función de los análisis alternativos de nombres estableciendo el factor de puntuación de análisis alternativo (`altScoreFactor`) en un valor inferior a 1,0.

El nombre "JOHN ELTON" puede coincidir con el nombre "ELTON JOHN" porque la última genera un análisis alternativo en que "JOHN" se ha marcado como nombre de pila y "ELTON" como apellido. El usuario podría querer que la puntuación de la comparación reflejara la diferencia en el orden original de los nombres, de forma que la comparación no reciba una puntuación perfecta de 1,0. El factor de puntuación alternativo se aplica a todas las puntuaciones de nombres completos que NameHunter calcula.

El factor de puntuación alternativo debe ser un valor entre 0,00 y 1,00. Se puede establecer de cualquiera de estas tres formas:

- A través del componente de bajo nivel de NameHunter (`ibmgmr::hunter::Searcher::altScoreFactor()`)
- Como un valor *General* de NameWorks, que afecta a las puntuaciones de todas las listas de datos de Embedded Search asociadas a un objeto de *Puntuación* específico (consulte los archivos de configuración de NameWorks):

```
[General]
DefaultAltScoreFactor=0.99
```

- Como parte de una definición de estrategia de búsqueda de NameWorks (consulte las definiciones de estrategia de búsqueda de NameWorks):

```
[Strategy:Somename]
AltScoreFactor=0.99"
```

### **Lógica de puntuación de nombres abreviados:**

Utilice los controles de nombres abreviados para conseguir los resultados deseados con nombres abreviados con muchas diferencias en relación a la longitud de los nombres.

IBM InfoSphere Global Name Management incluye ahora una lógica especial para comparar nombres abreviados que tiene en cuenta si un supuesto error tipográfico es en realidad un nombre de persona legítimo. Por ejemplo, "BRET" y "BETR" son diferentes de "BERT" en base a una transposición simple, pero "BRET" podría ser un nombre legítimo mientras que "BETR" es probablemente un error topográfico. IBM InfoSphere Global Name Management utiliza la información de frecuencia de NDA Name Data Archive (NDA) para determinar si dos nombres abreviados que difieren en una letra adicional (por ejemplo, "STEVE" y "STEEVE"), difieren en una letra (por ejemplo, "STEVE" y "STEBE") o en un par de letras intercambiadas (por ejemplo, "STEVE" y "SETVE") son nombres legítimos o errores tipográficos. La lógica especial se aplica entonces para gestionar la puntuación de los nombres. La lógica de puntuación la puede modificar la aplicación de parámetros de comparación. Cuando los nombres son más largos, la diferencia de puntuación entre ellos que se puede atribuir a errores tipográficos es menor. Por este motivo, la nueva lógica de puntuación solamente se aplica a nombres abreviados, considerándose como abreviado un nombre de entre 2 y 7 letras. Los nombres que contengan más de ocho letras pueden estar sujetos a la lógica si están relacionados con un nombre abreviado como, por ejemplo, "MARJORIE" y "MARORIE". Los nombres de dos letras se puntúan con la lógica nueva solamente cuando se comparan nombres con dos o tres letras puesto que se da por supuesto que las cadenas de caracteres de un solo carácter son iniciales y están sujetas a una lógica de puntuación de iniciales especial. Los nombres con más de una diferencia tipográfica entre ellos tampoco están sujetos a la lógica nueva, como "MARJORIE" y "MAROREI" que se puntuarían utilizando la lógica de puntuación normal.

Puede especificar los controles de nombres abreviados mediante la API de NameHunter o los atributos XML de NameHunter Distributed Search.

#### API de NameHunter

- CompParms::shortMinLength
- CompParms::shortMaxLength
- CompParms::shortNameScore
- CompParms::shortValidFactor
- CompParms::shortValidRatio
- CompParms::shortValidMax

#### Atributos XML de NameHunter Distributed Search

- COMP\_PARMS\_GN SHORT\_MIN\_LENGTH
- COMP\_PARMS\_GN SHORT\_MAX\_LENGTH
- COMP\_PARMS\_GN SHORT\_NAME\_SCORE
- COMP\_PARMS\_GN SHORT\_VALID\_FACTOR
- COMP\_PARMS\_GN SHORT\_VALID\_RATIO
- COMP\_PARMS\_GN SHORT\_VALID\_MAX
- COMP\_PARMS\_SN SHORT\_MIN\_LENGTH
- COMP\_PARMS\_SN SHORT\_MAX\_LENGTH
- COMP\_PARMS\_SN SHORT\_NAME\_SCORE
- COMP\_PARMS\_SN SHORT\_VALID\_FACTOR
- COMP\_PARMS\_SN SHORT\_VALID\_RATIO
- COMP\_PARMS\_SN SHORT\_VALID\_MAX

Los parámetros **shortMinLength** y **shortMaxLength** controlan los nombres que corresponden a la lógica de nombres abreviados. Los valores deben hallarse entre 2 y 7, ambos incluidos.

El parámetro **shortNameScore** proporciona la puntuación utilizada para dos nombres abreviados coincidentes que difieren en un solo error tipográfico como, por ejemplo, "ANNA" y "ANNQ", "IVAN" e "IVANM" o "KLAUS" y "KLUAS". Los valores válidos son [0,00, 1,00], ambos incluidos. El valor predeterminado para todas las culturas es 0,97.

El parámetro **shortValidFactor** proporciona un factor aplicado a **shortNameScore** cuando los nombres que tienen una sola diferencia se consideran legítimos, nombres distintos, como, por ejemplo, "DORA" y "CORA", "JUAN" y "JUANA" o "AMIR" y "MAIR". El parámetro **shortNameScore** se multiplica por este factor para determinar la puntuación coincidente en tales casos. Los valores válidos son [0,00, 1,00], ambos incluidos. El valor predeterminado para todas las culturas es 0,75. (La puntuación para coincidencias entre dos nombres legítimos y distintos con una diferencia ortográfica es, por lo tanto, de 0,72.)

Los nombres abreviados se consideran nombres legítimos y distintos si el recuento de la frecuencia de uno de ellos en NDA es mayor que un porcentaje configurable del otro, o si los recuentos de la frecuencia de ambos son mayores que un umbral configurable. Estas dos condiciones las controlas los parámetros **shortValidRatio** y **shortValidMax**, respectivamente. Por ejemplo, con el valor predeterminado de **shortValidRatio** establecido en 0,1, "JAMAM" se puntuará como un error tipográfico cuando se compara con "JAMAL", puesto que el número de apariciones de "JAMAM" en NDA es inferior a 0,01 veces el número de apariciones de "JAMAL". Sin embargo, cuando se compara "KAMAL" con "JAMAL", ambos se tratarán como nombres legítimos y distintos puesto que el número de apariciones

de "KAMAL" en NDA es mayor que la proporción válida. De hecho, tanto "JAMAL" como "KAMAL" aparecen en NDA con más frecuencia que el valor predeterminado del parámetro **shortValidMax**. Por ese motivo también se tratarán ambos como nombres legítimos y distintos.

### Umbral de nombre:

El control de umbral de nombre es el parámetro de comparación de nivel superior que NameHunter consulta durante el proceso de búsqueda de nombres. Este valor define la puntuación general que cada nombre admisible de la base de datos debe alcanzar o sobrepasar para que se considere que es una coincidencia.

Este control es un valor comprendido entre 0.00 y 1.00. Cuando el valor es 1.00, todas las coincidencias deben compararse exactamente con un nombre de consulta, por lo que está previsto que los resultados de la búsqueda contengan un número relativamente más pequeño de registros coincidentes. Las menores diferencias entre el nombre especificado en la consulta y un nombre de la base de datos pueden hacer que el proceso de comparación rechace el registro admisible.

Puede especificar el umbral de nombre mediante la API NameHunter o los atributos XML de Distributed Search para activar o desactivar esta función.

<b>API de NameHunter</b>	<b>Atributos XML de NameHunter Distributed Search</b>
• CompParms::nameThreshold	• GENERAL_PARMS NAME_THRESH

Los ejemplos siguientes muestran los resultados de búsqueda producidos por NameHunter Distributed para los resultados de búsqueda emitidos para la base de datos de ejemplo, cuando el control de umbral de nombre tiene tres niveles diferentes.

Los resultados de búsqueda siguientes se han obtenido con el control de umbral de nombre establecido en 0.80. Tres nombres de la lista han producido coincidencias en este nivel de umbral.

Query name = Johnson, Robert

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Bob	Johnson	0.98	1.00	0.95
Robert	Johnston	0.90	0.82	1.00
Robert Adamson	Johnston	0.90	0.82	0.99
Michael Robert	Johnson	0.82	1.00	0.59

Cuando se cambia el umbral de nombre a 0.70 se devuelven los resultados siguientes.

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Bob	Johnson	0.98	1.00	0.95
Robert	Johnston	0.90	0.82	1.00
Robert Adamson	Johnston	0.90	0.82	0.99

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Michael Robert	Johnson	0.82	1.00	0.59
Craig Robert	Johnston	0.74	0.82	0.63
Robert	Jackson	0.72	0.50	1.00
Robert	Swanson	0.72	0.50	1.00

Cuando se cambia el umbral de nombre a 0.50 se obtienen los resultados siguientes.

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Bob	Johnson	0.98	1.00	0.95
Robert	Johnston	0.90	0.82	1.00
Robert Adamson	Johnston	0.90	0.82	0.99
Michael Robert	Johnson	0.82	1.00	0.59
Craig Robert	Johnston	0.74	0.82	0.63
Robert	Jackson	0.72	0.50	1.00
Robert	Swanson	0.72	0.50	1.00
Henry R	Rokeby Johnson	0.58	0.64	0.50
Albert Leslie	Swanson	0.53	0.50	0.57

### Controles de campo:

Los parámetros de comparación básicos incluyen dos controles que determinan cómo NameHunter procesa y puntúa los componentes de nombre de pila (GN) y apellido (SN) contenido en el nombre de la consulta y en los nombres de la base de datos.

Debido a que las propiedades lingüísticas, culturales, estadísticas y de cálculo de los datos del nombre de pila difieren significativamente respecto a los datos del apellido, estos controles permiten ajustar y refinar el proceso de nombres de pila y apellidos de forma que NameHunter puede dar cabida a una amplia variedad de contenido de base de datos, diseños de aplicación y preferencias de usuario.

Existen dos factores principales que determinan la forma en que la información de los campos interviene en el proceso de comparación de NameHunter:

#### Umbral

Es el grado mínimo aceptable de variación entre el nombre especificado en la consulta y el nombre que se está comparando, tal como está expresado por la puntuación de similitud de la comparación.

#### Ponderación

Es la importancia del campo de nombre de pila o de apellido respecto a la comparación del nombre completo, tal como está expresado por la contribución relativa de la puntuación del campo de nombre de pila o apellido con respecto a la puntuación total correspondiente a la comparación del nombre completo.

### Umbral de campo:

Utilice el umbral de campo para establecer el nivel mínimo de similitud que se debe encontrar entre el campo de la consulta y el componente de campo de un registro de la base de datos para que prosiga el proceso de búsqueda de NameHunter.

Al igual que ocurre con el control de umbral de nombre, este control es un valor comprendido entre 0.00 y 1.00. Cuando su valor es 1.00, cada resultado de una búsqueda en la base de datos debe coincidir exactamente con un nombre de consulta. Es de esperar que los resultados de la búsqueda contengan un número relativamente menor de registros coincidentes. Incluso las menores diferencias entre el nombre especificado en la consulta y un nombre admisible contenido en la base de datos harán que el proceso de comparación rechace ese registro.

Puede especificar el umbral de campo mediante la API de NameHunter o los atributos XML de NameHunter Distributed Search. Cuanto más cerca está de 1.00 el umbral de campo, más parecidos son los nombres comparados. Cuanto menor es el valor, más diferentes son los dos nombres. El control de ponderación de nombres de pila y el control de ponderación de apellidos se combinan para determinar las ponderaciones relativas del nombre de pila y del apellido.

#### API de NameHunter

- CompParms::threshold

#### Atributos XML de NameHunter Distributed Search

- COMP\_PARMS\_GN FIELD\_THRESH
- COMP\_PARMS\_SN FIELD\_THRESH

**Nota:** El parámetro FIELD\_THRESH se ignora en la lógica de puntuación de los nombres de organización. Utilice el parámetro de comparación NAME\_THRESH para nombres de organización para establecer el valor que define la puntuación global que cada nombre admisible contenido en la base de datos debe alcanzar o sobrepasar para que se considere una coincidencia.

Se obtienen los resultados siguientes cuando el umbral de GN se establece en 0.80 y el umbral de SN se establece en 0.30. Tal como se muestra en la tabla, existe una estrecha correspondencia entre los datos de GN en la consulta y los nombres en el resultado.

Query name = Johnson, Robert

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Bob	Johnson	0.98	1.00	0.95
Robert	Johnston	0.90	0.82	1.00
Robert Adamson	Johnston	0.90	0.82	0.99
Robert	Jackson	0.82	1.00	0.59
Robert	Swanson	0.72	0.50	1.00
Robert	Nelson	0.67	0.40	1.00

Si el umbral de GN se establece en 0.30 y el umbral de SN se establece en 0.80 se obtiene una mayor variedad en el campo del nombre de pila.

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Bob	Johnson	0.98	1.00	0.95
Robert	Johnston	0.90	0.82	1.00
Robert Adamson	Johnston	0.90	0.82	0.99
Michael Robert	Johnson	0.82	1.00	0.59
Craig Robert	Johnston	0.74	0.82	0.63
Robert	Jackson	0.72	0.50	1.00
Robert	Swanson	0.72	0.50	1.00
Roy	Johnson	0.72	1.00	0.36

*Valor de ponderación del campo:*

Utilice el control de ponderación de campo para determinar la importancia relativa de la puntuación del nombre de pila o del apellido, con respecto a la otra puntuación de campo, cuando se calcula la puntuación del nombre completo.

Por ejemplo, cuando la ponderación del campo de GN y del campo de SN se establece en 1.00, las puntuaciones de campo contribuyen por igual al cálculo de la puntuación del nombre completo. Cuando la ponderación de GN se establece en 0.00 y la ponderación de SN se establece en 1.00, no se tiene en cuenta la puntuación de GN cuando se calcula la puntuación del nombre completo.

En muchas culturas, el nombre de pila admite formas variantes (tales como apodos y diminutivos), mientras que el apellido mantiene una forma fija. Un cambio en la representación del apellido es mucho más significativo para diferenciar dos personas que un cambio en el nombre de pila. Esta diferencia en el valor distintivo del nombre de pila y del apellido se expresa asignando una ponderación menor al campo del nombre de pila que al campo del apellido.

Puede especificar la ponderación de campo mediante la API NameHunter o los atributos XML de NameHunter Distributed Search para activar o desactivar esta función. En la API NameHunter, el control de ponderación de GN y el control de ponderación de SN se combinan para poder determinar las ponderaciones relativas del nombre de pila y del apellido.

API de NameHunter	Atributos XML de NameHunter Distributed Search
<ul style="list-style-type: none"> <li>• CompParms::weight</li> </ul>	<ul style="list-style-type: none"> <li>• COMP_PARMS_GN FIELD_WEIGHT</li> <li>• COMP_PARMS_SN FIELD_WEIGHT</li> </ul>

Se obtienen los resultados siguientes cuando la ponderación de GN se establece en 0.20 (nombre de pila poco importante) y la ponderación de SN se establece en 1.00.

Query name = Johnson, Robert

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Bob	Johnson	0.98	1.00	0.95
Michael Robert	Johnston	0.93	1.00	0.59

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Roy	Johnson	0.89	1.00	0.36
Robert	Johnston	0.85	0.82	1.00
Robert Adamson	Johnston	0.90	0.82	0.99
Craig Robert	Johnston	0.82	1.00	0.59
Henry R	Rokeby Johnson	0.61	0.64	0.50

Se obtienen los resultados siguientes cuando la ponderación de GN se establece en 0.90 (nombre de pila importante). El efecto sobre la puntuación del nombre completo es muy diferente con respecto al ejemplo anterior.

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Bob	Johnson	0.98	1.00	0.95
Robert	Johnston	0.91	0.82	1.00
Robert Adamson	Johnston	0.90	0.82	0.99
Michael Robert	Johnson	0.82	1.00	0.59
Robert	Jackson	0.72	0.50	1.00
Robert	Swanson	0.72	0.50	1.00
Craig Robert	Johnston	0.74	0.82	0.63
Roy	Johnson	0.72	1.00	0.36

#### Factor de raíz faltante:

Cuando compara dos campos de nombres, NameHunter comprueba a ver si los campos tienen un número diferente de frases de nombre.

En caso de que sean distintos (por ejemplo, un campo es 'de la Cruz Beltran' y el otro es 'de la Cruz'), se aplica una penalización. El importe de la penalización está en función del valor establecido para el parámetro de Factor raíz faltante, que se aplica como factor en la puntuación del campo.

Normalmente, el valor de este parámetro se establece en 0.98, aunque se permite cualquier valor entre 0.00 y 1.00.

Puede especificar el factor raíz faltante mediante la API NameHunter o los atributos XML de Distributed Search para activar o desactivar esta función.

API de NameHunter	Atributos XML de NameHunter Distributed Search
<ul style="list-style-type: none"> <li>CompParms::missingStemFactor</li> </ul>	<ul style="list-style-type: none"> <li>COMP_PARMS_GN MISSING_STEM_FACTOR</li> <li>COMP_PARMS_SN MISSING_STEM_FACTOR</li> </ul>

Se devuelven los resultados siguientes con el factor raíz faltante establecido en 1.00 para el nombre de consulta siguiente. No se asigna ninguna penalización, por lo tanto los nombres coinciden exactamente.

Query name = Johnson, Robert

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Robert	Johnston	0.90	0.82	1.00
Robert Adamson	Johnston	0.90	0.82	1.00

Se devuelven los resultados siguientes cuando el factor raíz faltante tiene el valor 0.90. Se asigna una penalización mayor porque Robert Adamson no es una buena coincidencia como Robert.

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Robert	Johnston	0.90	0.82	1.00
Robert Adamson	Johnston	0.86	0.82	0.90

### Controles de iniciales:

Las iniciales son normalmente menos distintivas que los nombres que representan, pues una inicial puede producir una coincidencia de búsqueda con varios nombres diferentes. Puede utilizar controles específicos de NameHunter para determinar el valor de las coincidencias de búsqueda producidas en iniciales.

Tres parámetros de comparación de NameHunter controlan la forma en que se procesan las iniciales. Cada parámetro se puede aplicar al nombre de pila y al apellido.

#### Parámetro Comparar iniciales

Este valor lógico (true/false) determina si al nombre de campo indicado se le debe aplicar una lógica especial para el manejo de iniciales en NameHunter. El campo de la API para C++ es `CompParm::matchInitials`. Los atributos de XML correspondientes son: `COMP_PARM_GN_MATCH_INITIAL` y `COMP_PARM_SN_MATCH_INITIAL`.

#### Puntuación de coincidencia de inicial con inicial

Puntuación de coincidencia que se debe asignar cuando dos segmentos son ambas letras iniciales y son idénticos. Su valor está comprendido entre 0.00 y 1.00. El campo de la API para C++ es `CompParm::initialOnInitialScore`. Los atributos de XML correspondientes son: `COMP_PARM_GN_INITIAL_ON_INITIAL_SCORE` y `COMP_PARM_SN_INITIAL_ON_INITIAL_SCORE`.

#### Puntuación de coincidencia de inicial

Puntuación de coincidencia que se debe asignar cuando un segmento es una inicial y el otro es un signo formado por varios caracteres cuyo primer carácter es el mismo que la inicial (por ejemplo, H y Harold). Su valor está comprendido entre 0.00 y 1.00. El campo de la API para C++ es `CompParm::initialOnTokenScore`. Los atributos XML correspondientes son: `COMP_PARM_GN_INITIAL_ON_TOKEN_SCORE` y `COMP_PARM_SN_INITIAL_ON_TOKEN_SCORE`.

Se obtienen los resultados de las tablas siguientes cuando se aplican los valores siguientes:

- Initial matching = On
- InitialOnInitialScore = 0.95

- InitialOnTokenScore = 0.85

Query name = Mathers, X

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
X	Mathews	0.81	0.71	0.95
Xavier	Mathews	0.77	0.71	0.85

Se obtienen los resultados siguientes si la comparación de iniciales está desactivada:

Query name = Mathers, X

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
X	Mathews	0.84	0.71	1.00

El valor predeterminado de NameHunter paraInitialOnInitialScore es 1.00.

#### Controles de nombre faltante:

NameHunter incluye parámetros que determinan cómo se debe asignar una puntuación a campos que no contienen ningún dato. Puede utilizar los controles de nombre faltante para regular las comparaciones entre campos correspondientes de dos nombres (nombre de pila o apellido) cuando uno (o ambos) no contiene datos para comparar.

Un problema habitual en muchas colecciones de nombres es que uno de los campos de datos del nombre puede estar vacío. Este problema puede ser debido a que los datos se han colocado en campos incorrectos (por ejemplo, el nombre completo puede estar situado en un campo de apellido), falta parte del nombre en el registro (por ejemplo, solamente se ha registrado el apellido) o la persona tiene un solo nombre únicamente (por ejemplo, Suharto, nombre del ex-presidente de Indonesia). Se utilizan dos pares de nombres para tratar con los campos de nombres inexistentes o vacíos en NameHunter:

- Controles de nombre parcial
- Controles de nombres no existentes

#### Controles de nombre parcial:

Puntuación del nombre parcial (valor comprendido entre 0.00 y 1.00) que se asigna cuando un campo de nombre se designa como desconocido.

Los valores simbólicos FNU (Nombre de pila desconocido) y LNU (Apellido desconocido) son códigos que informan a NameHunter de que un falta un nombre de pila o apellido o bien que son desconocidos. NameHunter trata al valor NFN como un nombre de pila en blanco y al valor NLN como un apellido en blanco. Si el campo de nombre de pila o el campo de apellido está en blanco, o si se designan NFN o NLN, NameHunter utiliza la puntuación del nombre parcial correspondiente cuando el campo de nombre se compara contra un campo de nombre que no falta en el registro de la lista de datos.

Si un campo de nombre se designa como NFN, NLN, o vacío, y se compara con un campo que es desconocido (FNU o LNU), la puntuación resultante viene determinada mediante la ecuación siguiente:

$$(\text{UnknownScore} + 1)/2$$

Puede especificar los controles de nombre parcial mediante a la API NameHunter o los atributos XML de NameHunter Distributed Search XML cuando esta función está activada o desactivada.

<b>API de NameHunter</b>	<b>Atributos XML de NameHunter Distributed Search</b>
• CompParms::noNameScore	• COMP_PARMS_GN NO_NAME_SCORE
• CompParms::nameUnknownScore	• COMP_PARMS_GN NAME_UNKNOWN_SCORE
	• COMP_PARMS_SN NO_NAME_SCORE
	• COMP_PARMS_SN NAME_UNKNOWN_SCORE

Se obtienen los resultados siguientes cuando noNameScore = 0.80 y UnknownScore = 0.85. Una entrada en blanco en la tabla indica que el campo está vacío.

Query name = Farris, Travis

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
FNU	Farris	0.93	1.00	0.85
	Farris	0.91	1.00	0.80
NFN	Farris	0.91	1.00	0.80

*Controles de nombres no existentes:*

Los controles de nombres no existentes ayudan en situaciones en las que se desconoce un nombre o no existe un campo de nombre así.

En algunas culturas se puede identificar una persona solamente con un único nombre, que puede funcionar como nombre de pila o apellido. Por ello puede ser necesario distinguir entre los casos en los que no se conoce un nombre y los casos en que ese campo de nombre no existe. La puntuación de nombre no existente, cuyo valor está comprendido entre 0.00 y 1.00. permite establecer la puntuación de coincidencia cuando un nombre está especificado explícitamente como carente del nombre de pila (GN) o del apellido (SN) mediante los valores simbólicos especiales NFN (sin nombre de pila) y NLN (sin apellido), respectivamente.

El control de nombre no existente también se aplica en los casos en que ambos nombres carecen de valores para un campo de nombres; por ejemplo, cuando NameHunter compara dos nombres que sólo contienen apellidos.

Puede especificar estos parámetros mediante la API de NameHunter para C++ o los atributos XML de NameHunter Distributed Search:

#### API de NameHunter

- CompParms::noNameScore
- CompParms::nameUnknownScore

#### Atributos XML de NameHunter Distributed Search

- COMP\_PARMS\_GN NO\_NAME\_SCORE
- COMP\_PARMS\_SN NO\_NAME\_SCORE
- COMP\_PARMS\_GN  
NAME\_UNKNOWN\_SCORE
- COMP\_PARMS\_SN  
NAME\_UNKNOWN\_SCORE

Se obtienen los resultados siguientes cuando noNameScore = 0.80 y unknownNameScore = 0.85. Los resultados de este tipo de consulta pueden ser extensos. En este caso, cada nombre de pila coincidiría con el nombre en blanco contenido en la consulta.

Query name = Farris,

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
FNU	Farris	0.97	1.00	0.93
	Farris	1.00	1.00	1.00
NFN	Farris	0.96	1.00	0.80
James	Farris	0.91	1.00	0.80

**Nota:** En estos ejemplos se han aplicado los parámetros anglosajones. Tomando el valor promedio de la puntuación del apellido (1.00) y de la puntuación del nombre de pila (0.80) se obtendría un promedio igual a 0.90 para James Farris. Pero las puntuaciones del apellido tienen un valor de ponderación ligeramente mayor que las puntuaciones del nombre de pila cuando se aplican parámetros anglosajones, lo cual produce un puntuación global de 0.91.

#### Método de puntuación de segmentos:

A menudo los nombres contienen más de un solo nombre de pila o apellido en un campo, como ocurre en Kate Marie Smith o Ana Ramos Sánchez. Los parámetros de puntuación de segmentos de NameHunter controlan cómo se debe determinar la puntuación para el campo completo (campo de nombre de pila o de apellido) a partir de las puntuaciones de cada nombre individual (segmento) contenido en el campo.

Puede especificar tres modalidades de puntuación diferentes que determinan cómo NameHunter combina las puntuaciones de frases de nombres de pila y apellidos (segmentos) y forma una puntuación compuesta para el correspondiente campo de nombre de pila o apellido. Puede especificar estas opciones mediante la API de NameHunter para C++ o los atributos XML de NameHunter Distributed Search.

#### API de NameHunter

- CompParm::scoreMode

#### Atributos XML de NameHunter Distributed Search

- COMP\_PARMS\_GN SCORE\_MODE
- COMP\_PARMS\_SN SCORE\_MODE

Existen tres modalidades de puntuación. El valor Highest seleccionado para un campo de nombre, permite obtener muchas más coincidencias, mientras que el valor Lowest tiene el efecto opuesto.

#### Lowest

NameHunter calcula las puntuaciones de comparación para todos los nombres individuales contenidos en el campo. La puntuación de comparación más baja asignada a un nombre se asigna al campo completo. Por ejemplo, si Gina Marie se compara con Ginny Marie, la puntuación para la comparación Marie/Marie sería 1.0, mientras que la comparación Gina/Ginny tendría una puntuación menor (por ejemplo, 0.67). La puntuación que se asigna a este campo de nombre de pila sería la puntuación menor, 0.67. El efecto de esta modalidad de puntuación es que cada segmento individual del campo debe tener una puntuación de comparación lo suficiente alta para pasar el umbral de campo. Por lo tanto, la modalidad de puntuación es la más estricta y requiere el nivel de similitud más alto entre la consulta y el nombre que se está comparando.

#### Average

Se obtiene un valor promedio simple de todas las puntuaciones de segmentos en el campo de nombre para calcular una puntuación compuesta para el campo de nombre.

#### Highest

NameHunter calcula las puntuaciones de comparación para todos los nombres individuales contenidos en el campo. La puntuación de comparación más alta asignada a un nombre se asigna al campo completo. En el ejemplo anterior, (Gina Marie/Ginny Marie), la puntuación para el campo de nombre de pila sería 1.0 si se utilizara esta modalidad de puntuación. Únicamente es necesario que la puntuación de una sola comparación de segmentos sea lo suficiente alta para pasar el umbral de campo. Esta modalidad es la más flexible pues es la que permite el mayor grado de variabilidad entre el nombre de la consulta y el nombre que se está comparando.

Si la modalidad de puntuación de segmentos tuviera el valor Highest se producirían los resultados siguientes para el nombre de pila y el apellido.

Query name = Hamilton Connerly, Lucinda Anna

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Lucinda Anna	Hamilton Connor	1.00	1.00	1.00
Patricia Ann	Hamilton	0.99	0.98	1.00
Hubert A	Hamilton	0.92	0.98	0.85
Wade A	Hamilton	0.92	0.98	0.85
Linda	Charlton	0.62	0.54	0.71

Si la modalidad de puntuación de segmentos tuviera el valor Average se producirían los resultados siguientes para el nombre de pila y el apellido.

Las puntuaciones de las mejores coincidencias para los nombres de pila (Lucinda~Lucinda) y de las mejores coincidencias para los apellidos (Hamilton~Hamilton) se utilizan respectivamente como puntuación del campo de nombre de pila y como puntuación del campo de apellido. El parámetro

missingStemFactor es la causa de las ligeras diferencias de puntuación que se producen cuando se compara un número diferente de signos.

Query name = Hamilton Connerly, Lucinda Anna

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Lucinda Anna	Hamilton Connor	1.00	1.00	1.00
Patricia Ann	Hamilton	0.99	0.98	1.00
Linda	Charlton	0.62	0.54	0.71

Algunas coincidencias se eliminaron porque todas las comparaciones de cadenas de caracteres contribuyen a la puntuación de nombres. Por tanto, Hubert~Lucinda, que recibe una puntuación muy baja, hace que Hubert A Hamilton se elimine de la lista.

### Controles de segmento de ancla:

El parámetro de segmento de ancla determina si una frase de nombre coincidente debe estar situada en el extremo izquierdo (primer lugar) o el extremo derecho (último) en un campo de múltiples nombres para poderse considerar una coincidencia óptima o si la posición dentro del campo no importa. Este parámetro permite tener en cuenta en la puntuación la secuencia de un campo.

La secuencia es importante en los apellidos hispanos; por ejemplo, el apellido situado más a la izquierda de dos apellidos es el apellido de una persona, mientras que el situado más a la derecha muestra simplemente el apellido materno y suele omitirse en el nombre.

Las opciones disponibles son elegir si el nombre más importante es el situado más a la izquierda, el situado más a la derecha o ninguno de los dos.

En muchas partes del mundo, las personas tienen más de un nombre de pila (GN) o apellido (SN). Las costumbres por las que se rige la función de estos nombres o que determinan qué nombre se utiliza en cada circunstancia difieren de un grupo humano a otro.

Por ejemplo, entre las personas de habla inglesa es habitual tener como mínimo dos nombres de pila: un "primer nombre" y un "segundo nombre". El segundo nombre puede estar omitido o estar representado por una inicial. Por tanto, ROBERT WILSON, ROBERT JAMES WILSON y ROBERT J. WILSON podrían aludir todos ellos a una misma persona. Patrones similares de inclusión, omisión y uso sintáctico se pueden encontrar en otras culturas. Se proporciona un control de Segmento de ancla para ambos campos de nombre, SN y GN, para permitir que NameHunter centre la atención en la porción correcta de un campo de nombre de varios segmentos. Este control define si el segmento del campo que se debe considerar que es el segmento de ancla es el situado en primer lugar (más a la izquierda) o el situado en último lugar (más a la derecha) o si ningún segmento se debe considerar más esencial que otros (ninguno).

Entre las personas de habla española, el segmento de ancla del apellido ocupa normalmente la primera posición, pues el apellido situado más a la izquierda es el apellido patronímico del individuo, y el segundo apellido (el apellido materno) a

menudo se omite. En cambio, ocurre lo contrario entre las culturas de habla portuguesa, tales como Brasil, Portugal y determinadas naciones africanas.

Puede especificar los controles de segmento de ancla mediante la API de NameHunter o los atributos XML de NameHunter Distributed Search.

<b>API de NameHunter</b>	<b>Atributos XML de NameHunter Distributed Search</b>
<ul style="list-style-type: none"> <li>• CompParms::anchorType</li> </ul>	<ul style="list-style-type: none"> <li>• COMP_PARMS_GN ANCHOR_MODE</li> <li>• COMP_PARMS_SN ANCHOR_MODE</li> </ul>

La magnitud del efecto del segmento de ancla sobre la puntuación está determinada por el valor del factor de ancla, que se puede establecer mediante la API de NameHunter de C++ o los atributos XML de NameHunter Distributed Search

<b>API de NameHunter</b>	<b>Atributos XML de NameHunter Distributed Search</b>
<ul style="list-style-type: none"> <li>• CompParm::anchorFactor</li> </ul>	<ul style="list-style-type: none"> <li>• COMP_PARMS_GN ANCHOR_FACTOR</li> <li>• COMP_PARMS_SN ANCHOR_FACTOR</li> </ul>

Este factor acepta un valor comprendido entre 0.00 y 1.00, y se aplica a cualquier puntuación de coincidencia de NameParser cuando una de las frases de nombre coincidentes no se encuentra en la posición del Segmento de ancla.

Pongamos el caso del ejemplo siguiente, que muestra una búsqueda de un nombre hispano típico que contiene dos frases de nombre en el campo del nombre de pila (GN) y dos frases de nombre en el campo del apellido (SN):

Query name = Figueroa Martín, Ana María

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Ana	Figueroa	0.98	0.98	0.99
María	Figueroa	0.90	0.98	0.80
Juana	Figueredo	0.61	0.62	0.59

Para esta búsqueda, los Segmentos de ancla de GN y SN se establecieron en la posición primera y el Factor de ancla de GN se estableció en 0.90.

En el primer nombre coincidente, Ana Figueroa, tanto el nombre de pila (Ana) como el apellido (Figueroa) están en la posición más a la izquierda (posición primera del segmento de ancla). Por tanto, el factor de segmento de ancla no se aplica a ninguno de estos dos nombres. En el segundo nombre coincidente, María Figueroa, el nombre de pila (María) coincide con el segundo nombre del nombre de pila utilizado en la búsqueda (Ana María). Sin embargo, el nombre María utilizado en la búsqueda no está en la posición de ancla más a la izquierda. Por tanto, la coincidencia de búsqueda producida en María no es una coincidencia favorecida, y es penalizada por la aplicación del factor de ancla. La coincidencia de búsqueda producida en María recibe pues una puntuación de 0.90, aunque la comparación de María con María es una coincidencia exacta. Con este valor inferior y otras penalizaciones la puntuación del nombre de pila baja a 0.80.

Ahora invertiremos las frases del nombre de pila y mantendremos los mismos valores.

Query name = Figueroa Martín, María Ana

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
María	Figueroa	0.98	0.98	0.99
Ana	Figueroa	0.90	0.98	0.80

El registro coincidente con la puntuación más alta contiene “María”, cuya puntuación de nombre de pila coincidente es 0.99. La puntuación de coincidencia es 0.99 porque la frase de nombre de pila (GN) situado en la posición del Segmento de ancla (la posición más a la izquierda en este caso) coincidió con una frase de nombre de pila que también estaba en la posición del Segmento de ancla. Por tanto, no se aplicó el ancla de GN, y la puntuación preliminar 0.99 correspondiente a la coincidencia exacta de la frase de nombre para la comparación María/María permanece inalterada como puntuación final del campo de nombre de pila.

Las coincidencias de búsqueda producidas con registros de la lista de datos que contienen la palabra Ana en el campo de GN están sujetas ahora a las mismas reducciones de puntuación que las aplicadas a María en el ejemplo anterior, pues Ana ya no está en la posición del Segmento de ancla y no aparece en la correspondiente posición sintáctica en ambos campos del nombre de pila.

#### Controles de segmento fuera de lugar:

Utilice los controles de segmento fuera de lugar (OOPS) en NameHunter para regular la asignación de puntuaciones a nivel de campo de nombre cuando se encuentra una coincidencia entre segmentos de nombre que no ocupan la misma posición sintáctica en el campo de nombre.

El factor OOPS es útil cuando un campo de GN (nombre de pila) que contiene dos frases de nombre (tales como JAMES ROBERT) se compara con un campo de GN que contiene una o más de las mismas frases de nombre, pero en posiciones diferentes (por ejemplo, ROBERT JOSEPH o JOSEPH ROBERT).

Cuando NameHunter determina la mejor forma de encontrar frases de nombre coincidentes en los campos de GN o SN de dos nombres sometidos a comparación, a menudo NameHunter identifica una coincidencia óptima entre dos frases de nombre que no ocupan la misma posición sintáctica dentro del campo de nombre de GN o SN. Esto ocurre a menudo al comparar campos de nombre de pila (GN), puesto que en la mayoría de culturas el nombre de pila consta de varios elementos, y muchos nombres contienen uno o más frases de nombre de pila muy habituales.

El factor OOPS es un valor comprendido entre 0.00 y 1.00. Las puntuaciones de coincidencia más cercanas a 1.00 se penalizan menos para una coincidencia de fuera de posición. Las puntuaciones de coincidencia más cercanas a 0.00 se penalizan para una coincidencia de fuera de posición. Por ejemplo, cuando el factor OOPS está establecido en 1.00, una coincidencia de frase de nombre conserva su puntuación de coincidencia preliminar, incluso aunque se haya emparejado con una frase de nombre que está en una posición diferente dentro del campo de nombre.

Puede especificar el factor OOPS mediante la API de NameHunter o los atributos de XML de NameHunter Distributed Search y establecer la puntuación que se obtiene cuando se encuentra una coincidencia de variante de campo.

API de NameHunter	Atributos XML de NameHunter Distributed Search
• CompParms::oopsFactor	• COMP_PARMS_GN OOPS_FACTOR
	• COMP_PARMS_SN OOPS_FACTOR

En los siguientes resultados de búsqueda de NameHunter el factor OOPS se ha establecido en el valor 0.80 para el nombre de pila:

Query name = Duval, James Robert

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
James	Duval	1.00	1.00	0.99
Jim	Duval	0.97	1.00	0.94
Robert James	Duval	0.91	1.00	0.80
Robert	Duval	0.91	1.00	0.79
Bob	Duval	0.89	1.00	0.75

Cuando las frases de nombre coincidentes para Joseph se encuentran en la misma posición, el factor OOPS no se aplica. La puntuación de elemento de nombre preliminar de 1.00 correspondiente a una coincidencia exacta (James~James – en realidad 0.99 debido a `missingStemFactor`) permanece inalterada en la puntuación del nombre de pila final.

En cambio, cuando las frases de nombre coincidentes correspondientes a Robert se encuentran en posiciones diferentes (no en el extremo izquierdo en el nombre especificado en la consulta; en el extremo izquierdo en el nombre coincidente contenido en la lista de datos), se aplica el factor OOPS de nombre de pila, con lo que la puntuación preliminar de coincidencia del elemento de nombre se reduce de 1.00 a 0.80.

### Controles de nombres comprimidos:

Utilice los controles de nombre comprimidos para mitigar las diferencias accidentales en la segmentación y la posición de los espacios en blanco cuando se comparan dos nombres.

Un problema frecuente que surge cuando se manejan conjuntos grandes de nombres de personas es la posición variable de los espacios en blanco. A menudo los espacios en blanco se eliminan manualmente para dar cabida a más caracteres en un formulario de entrada de datos. Además, muchos sistemas automatizados de proceso de datos eliminan los espacios en blanco, lo cual hace que signos diferenciados de un campo de nombre se fusionen en uno solo.

Otra causa principal de la existencia de espacios en blanco irregulares en los nombres es la gran variedad de normas que se aplican cuando un nombre se convierte desde un sistema de escritura no latino a una forma latinizada. Cuando la forma original del nombre se expresa en un sistema de escritura no alfabético (tal como el árabe, el chino o el coreano), la colocación de los espacios en blanco en

el nombre latinizado a menudo se deja al criterio de la persona o proceso automatizado que efectúa la latinización. Por tanto, dos nombres que son idénticos en su forma original pueden dar lugar a dos representaciones muy diferentes después de la latinización.

Los controles de NameHunter CompParms proporcionan un mecanismo para mitigar y superar estas diferencias accidentales en la segmentación y la posición de los espacios en blanco cuando se comparan dos nombres. Los controles de nombres comprimidos permiten que NameHunter trate todas las posiciones de un campo de nombre (GN o SN) como si constituyeran desde el punto de vista lógico un solo valor, al que luego se le asigna una puntuación con técnicas estándar de similitud de nombres de NameHunter. Si se activa el proceso de nombres comprimido, NameHunter calcula una puntuación para los campos de nombre GN y SN como si se hubieran eliminado todos los espacios en blanco, y utiliza esta puntuación si es mayor que la puntuación de campo de nombre calculada por las mediciones estándar de puntuación de NameHunter.

Puede especificar los controles de nombres comprimidos mediante la API de NameHunter o los atributos XML de NameHunter Distributed Search.

<b>API de NameHunter</b>	<b>Atributos XML de NameHunter Distributed Search</b>
<ul style="list-style-type: none"><li>• CompParms::doCompressedName</li><li>• CompParms::compressedScoreMax</li><li>• CompParms::compressedScoreFactor</li></ul>	<ul style="list-style-type: none"><li>• COMP_PARMS_GN DO_COMPRESSED_NAME</li><li>• COMP_PARMS_GN COMPRESSED_SCORE_MAX</li><li>• COMP_PARMS_SN DO_COMPRESSED_NAME</li><li>• COMP_PARMS_SN COMPRESSED_SCORE_MAX</li></ul>

Cuando el distintivo doCompressedName está establecido en true para el campo de SN o GN, la puntuación se obtiene primero con la forma estándar del campo de nombre y luego se repite con la forma comprimida del valor de ese campo. Si la puntuación para la forma comprimida es mayor que la puntuación para el campo normal, compressedScoreMax se utiliza como puntuación del campo. El valor predeterminado de compressedScoreMax es 0.95.

El factor de puntuación comprimida, CompParms::compressedScoreFactor se utiliza principalmente con nombres de organización en los que los nombres comprimidos pueden devolver coincidencias no deseadas. Este factor descuenta la puntuación del nombre comprimido para evitar coincidencias en las que un único término relativamente insignificante genere una puntuación alta. Por ejemplo, la puntuación comprimida para ABC CONSTRUCTION y XYZ CONSTRUCTION es bastante alta debido al término CONSTRUCTION. El factor de puntuación comprimida siempre se aplica cuando se obtiene una puntuación comprimida. La puntuación resultante se comprueba con respecto a compressedScoreMax y la menor de los dos se considera que es la puntuación de coincidencia. El valor predeterminado para el factor de puntuación comprimida es 0.90 para los nombres de organización y 1.00 para los nombres de persona.

Los controles de nombres comprimidos NameHunter proporcionan un mecanismo eficaz para establecer coincidencias entre instancias de nombres que tienen una posición diferente de los espacios en blanco en idiomas en los que las convenciones

de lengua escrita no latina varían para los nombres. Por ejemplo, el espacio en blanco es diferente en los nombres chinos latinizados como Li Ping, Liping y Li-Ping.

Considere la siguiente consulta cuando `doCompressedName=true`. Si este distintivo no estuviera establecido en `true`, no se habría devuelto ninguno de los resultados siguientes.

Query name = AbdulSalah, Mohamed

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Mohamed	Abdul Salah	0.97	0.95	1.00
Mohamed	Abdel Salah	0.90	0.82	1.00
Mohamed	Abd El Salah	0.90	0.82	1.00
Mohamed	Abdel Salam	0.90	0.64	1.00

### Consideraciones adicionales

Cuando se implementa la puntuación de nombres comprimidos tenga en cuenta lo siguiente:

- NameHunter incluye mecanismos para evitar que nombres de organización, en especial nombres de organización con varias cadenas de caracteres, generen puntuaciones altas erróneas y falsas positivas.
- Se eliminan de la puntuación varios TAQ consecutivos. Además, los TAQ específicos, como por ejemplo, las fórmulas de tratamiento, calificadores, designadores de organización y calificadores de profesión no se tienen en cuenta durante la puntuación.
- Se aplica el parámetro de comparación `compressedScoreFactor` para descontar todas las puntuaciones de nombres comprimidos. Los valores predeterminados de `compressedScoreFactor` son 0.90 para los nombres de organización y 1.0 para los nombres de persona.

Los efectos de los controles de nombre comprimido se pueden suprimir cuando NameHunter utiliza una *opción de salida anticipada* para una lista de datos a fin de acelerar la velocidad del proceso de búsqueda. Esta opción realiza varios cálculos rápidos previos en cada comparación entre un nombre de consulta y un nombre de lista de datos para eliminar nombres de listas de datos que no es probable que produzcan coincidencias de búsqueda. La utilización de la opción de salida anticipada puede ocasionalmente impedir que una coincidencia válida de nombres comprimidos se reconozca, aplique y puntúe. Si los resultados de búsqueda de NameHunter no incluyen coincidencias probables de nombres comprimidos, compruebe si se está utilizando una opción de salida anticipada para acelerar el proceso de búsqueda. Elimine la opción de salida anticipada y repita la misma búsqueda.

### Controles de sesgo a la izquierda:

NameHunter proporciona controles de sesgo a la izquierda que se pueden utilizar para mitigar el efecto de las terminaciones habituales en los cálculos de similitud entre dos nombres.

Los nombres procedentes de muchas culturas europeas occidentales y norteamericanas comparten determinadas características, algunas de las cuales derivan de rasgos comunes de las lenguas románicas, germánicas y eslavas habladas por sus antepasados. Una característica de estas culturas que es importante para los algoritmos de comparación de nombres es que muchos nombres comparten las mismas terminaciones. Por esta razón, las letras situadas en el extremo izquierdo de estos nombres pueden ser más distintivas que las letras situadas más a la derecha.

Por ejemplo, muchos apellidos de personas de habla inglesa reflejan información por línea paterna, que es información que se remonta a los antepasados paternos de una persona. La terminación -SON es habitual en los apellidos de personas de habla inglesa: JOHNSON, STEVENSON, ROBERTSON, JEFFERSON.

Un fenómeno similar se puede observar en otras culturas europeas, tales como la rusa, donde muchas terminaciones típicas tales como -OV, -OVA, -SKI y -SKY aparecen en un alto porcentaje de nombres cuando se convierten a la forma latinizada.

Cuando NameHunter determina si dos nombres coinciden o no, la letras situadas en el extremo derecho del nombre pueden ser de menos valor que las situadas en el extremo izquierdo. Este fenómeno se denomina *sesgo a la izquierda* y está regulado por el control de sesgo a la izquierda. Este control es un distintivo que, cuando se establece en true aplica un cálculo de similitud predeterminado que favorece las coincidencias entre segmentos de los correspondientes campos de nombre (GN o SN) que tienen más letras en común al principio de la raíz del nombre.

Puede especificar los controles de sesgo a la izquierda mediante la API de NameHunter o los atributos XML de NameHunter Distributed Search.

<b>API de NameHunter</b>	<b>Atributos XML de NameHunter Distributed Search</b>
<ul style="list-style-type: none"> <li>• CompParms::leftBias</li> </ul>	<ul style="list-style-type: none"> <li>• COMP_PARMS_GN LEFT_BIAS</li> <li>• COMP_PARMS_SN LEFT_BIAS</li> </ul>

En el ejemplo siguiente, el control de sesgo a la izquierda está habilitado, lo que reduce el impacto de las diferencias encontradas a la derecha de cada nombre. Con el control de sesgo izquierdo habilitado, el valor de las letras similares en la mitad derecha del nombre tiene un impacto reducido por lo que las puntuaciones son más bajas cuando este control está inhabilitado.

Query name = Tarkovsky,Andrei

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Andrei	Tankovsky	0.83	0.70	1.00
Andrei	Tchiakovsky	0.69	0.44	1.00

Cuando el control de sesgo izquierdo está inhabilitado, todas las letras contribuyen de forma igual a la puntuación del nombre. Por consiguiente, la secuencia de la letra compartida a la derecha de cada nombre, OVSKY mejora la puntuación de similitud de la comparación y se obtiene una puntuación de nombre más alta.

Query name = Tarkovsky,Andrei

GN	SN	Puntuación del nombre	Puntuación del apellido	Puntuación del nombre de pila
Andrei	Tankovsky	0.88	0.80	1.00
Andrei	Tchiakovsky	0.79	0.63	1.00

## Parámetros de comparación predeterminados

NameHunter incluye parámetros de comparación predeterminados incorporados que se utilizan al comparar nombres. Estos parámetros de comparación predeterminados se pueden alterar temporalmente a través de un archivo de configuración externo. En el archivo de configuración solamente se alteran temporalmente los valores proporcionados; no se modifican los parámetros que faltan o los que no se han especificado.

Utilice el método `CompParms::overrideDefaultParms()` para establecer los parámetros de comparación predeterminados para un tipo de campo y de entorno cultural determinado.

El mensaje de parámetro de Distributed Search acepta factores que se pueden aplicar a valores umbrales y a otros campos. Por ejemplo, puede modificar el umbral de nombres para definir la puntuación general que cada nombre admisible de la base de datos debe alcanzar o sobrepasar para considerarse una coincidencia. Si especifica `nameThreshold=1.00` en el archivo de alteraciones temporales de parámetros de la comparación, el umbral del nombre cambia a 1.00. Cuando el valor es 1.00, todas las coincidencias deben compararse exactamente con un nombre de consulta, por lo que está previsto que los resultados de la búsqueda contengan un número relativamente más pequeño de registros coincidentes.

El rango válido de factores y de umbrales es de 0,00-1,00 en NameHunter. Los cambios se aplican a parámetros que se envían por separado o que están incluidos en una solicitud de búsqueda.

Los parámetros que controlan una consulta se aplican en un orden específico:

- alteraciones temporales de parámetros proporcionadas con la consulta
- parámetros de alteraciones temporales predeterminados leídos de un archivo `compparms.config`
- parámetros predeterminados incorporados para el tipo de campo y de entorno cultural

## Parámetros de comparación de NameHunter (nombres de organización):

La tabla siguiente proporciona los parámetros de configuración predeterminados para nombres de organización.

	Nombres de organización
Umbral de nombre	0.55
Ponderación	1.00
Sesgo a la izquierda	FALSE
Comparar iniciales	TRUE
Inicial en cadena de caracteres	0.00
Inicial en inicial	1.00

<b>Nombres de organización</b>	
Comparar variantes	TRUE
Nombre desconocido	0.40
Sin nombre	0.40
Tipo de ancla	Ninguna ancla
Factor de ancla	1.00
Factor Oops	0.97
Realizar puntuación comprimida	TRUE
Puntuación comprimida máxima	0.95
Factor de puntuación comprimida	0.90
Modalidad de puntuación	ScoreModeAverage
Factor de raíz faltante	0.95
Comparar variantes de campo	TRUE

Los parámetros de puntuación de nombre abreviado están inhabilitados para los nombres de organización.

**Parámetros de comparación de NameHunter (ambiguo-árabe):**

La tabla siguiente proporciona los parámetros de comparación predeterminados de los campos correspondientes al nombre de pila (GN) y apellido (SN) para los códigos de entorno cultural ambiguo, afgano, anglosajón y árabe.

	<b>Ambiguo - 0</b>		<b>Afgano - 16</b>		<b>Anglosajón - 1</b>		<b>Árabe - 2</b>	
	GN	SN	GN	SN	GN	SN	GN	SN
Umbral de nombre	0.60	0.60	0.65	0.65	0.65	0.65	0.65	0.65
Umbral de campo	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
Valor de ponderación del campo	0.80	1.00	1.00	1.00	0.80	1.00	1.00	1.00
Raíz faltante	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.98
Comparar variantes	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Comparar iniciales	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
Inicial en inicial	0.90	0.00	0.85	0.00	0.90	0.00	0.85	0.00
Inicial en cadena de caracteres	0.85	0.00	0.75	0.00	0.85	0.00	0.75	0.00
Nombre desconocido	0.60	0.60	0.60	0.75	0.60	0.60	0.75	0.75
Sin nombre	0.60	0.60	0.60	0.75	0.60	0.60	0.75	0.75

	Ambiguo - 0		Afgano - 16		Anglosajón - 1		Árabe - 2	
Modalidad de puntuación	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG
Tipo de ancla	NONE	NONE	FIRST	NONE	NONE	LAST	FIRST	NONE
Factor de ancla	1.00	1.00	0.85	1.00	1.00	0.70	0.90	1.00
Factor Oops	0.97	0.97	0.85	0.90	0.97	0.97	0.85	0.90
Nombre comprimido	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Máximo comprimido	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Factor de puntuación comprimida	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sesgo a la izquierda	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Regularizado máx	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Longitud mínima de nombre abreviado	2	2	2	2	2	2	2	2
Longitud máxima de nombre abreviado	6	6	6	6	6	6	6	6
Puntuación de nombre abreviado	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Factor de validez de nombre abreviado	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
Proporción de validez de nombre abreviado	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Valor máximo de validez de nombre abreviado	200	200	200	200	200	200	200	200

### Parámetros de comparación de NameHunter (chino-germánico):

La tabla siguiente proporciona los parámetros de comparación predeterminados de los campos correspondientes al nombre de pila (GN) y apellido (SN) para los códigos de entorno cultural hispánico, hindú, indonesio y japonés.

	Chino - 3		Farsi - 12		Francés - 7		Alemán - 8	
	GN	SN	GN	SN	GN	SN	GN	SN
Umbral de nombre	0.70	0.70	0.65	0.65	0.65	0.65	0.65	0.65
Umbral de campo	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
Valor de ponderación del campo	0.80	1.00	1.00	1.00	0.80	1.00	0.80	1.00
Raíz faltante	0.70	0.70	0.99	0.98	0.99	0.98	0.99	0.98
Comparar variantes	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Comparar iniciales	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
Inicial en inicial	0.85	0.00	0.85	0.00	0.85	0.00	0.85	0.00
Inicial en cadena de caracteres	0.85	0.00	0.75	0.00	0.75	0.00	0.75	0.00
Nombre desconocido	0.65	0.65	0.60	0.75	0.65	0.65	0.60	0.60
Sin nombre	0.65	0.65	0.60	0.75	0.65	0.65	0.60	0.60
Modalidad de puntuación	LOW	AVG	AVG	AVG	AVG	AVG	AVG	AVG
Tipo de ancla	NONE	NONE	FIRST	NONE	NONE	LAST	NONE	LAST
Factor de ancla	1.00	1.00	0.90	1.00	1.00	0.85	1.00	0.85
Factor Oops	0.85	0.75	0.85	0.90	0.97	0.97	0.97	0.97
Nombre comprimido	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Máximo comprimido	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Factor de puntuación comprimida	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sesgo a la izquierda	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Regularizado máx	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Longitud mínima de nombre abreviado	2	2	2	2	2	2	2	2
Longitud máxima de nombre abreviado	6	6	6	6	6	6	6	6

	Chino - 3		Farsi - 12		Francés - 7		Alemán - 8	
Puntuación de nombre abreviado	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Factor de validez de nombre abreviado	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
Proporción de validez de nombre abreviado	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Valor máximo de validez de nombre abreviado	200	200	200	200	200	200	200	200

#### Parámetros de comparación de NameHunter (hispanico-japonés):

La tabla siguiente proporciona los parámetros de comparación predeterminados de los campos correspondientes al nombre de pila (GN) y apellido (SN) para los códigos de entorno cultural hispanico, hindú, indonesio y japonés.

	Hispanico - 4		Hindú - 14		Indonesio - 10		Japonés - 15	
	GN	SN	GN	SN	GN	SN	GN	SN
Umbral de nombre	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
Umbral de campo	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
Valor de ponderación del campo	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Raíz faltante	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.98
Comparar variantes	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Comparar iniciales	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
Inicial en inicial	0.85	1.00	1.00	0.00	0.85	0.00	0.80	0.00
Inicial en cadena de caracteres	0.85	0.85	0.85	0.00	0.75	0.00	0.80	0.00
Nombre desconocido	0.60	0.60	0.55	0.50	0.70	0.70	0.55	0.55
Sin nombre	0.60	0.60	0.55	0.50	0.70	0.70	0.55	0.55
Modalidad de puntuación	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG

	<b>Hispánico - 4</b>		<b>Hindú - 14</b>		<b>Indonesio - 10</b>		<b>Japonés - 15</b>	
Tipo de ancla	NONE	FIRST	NONE	NONE	NONE	NONE	NONE	NONE
Factor de ancla	1.00	0.70	1.00	1.00	1.00	1.00	1.00	1.00
Factor Oops	0.90	0.80	0.90	0.90	0.97	0.97	0.90	0.75
Nombre comprimido	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Máximo comprimido	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Factor de puntuación comprimida	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sesgo a la izquierda	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Regularizado máx	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Longitud mínima de nombre abreviado	2	2	2	2	2	2	2	2
Longitud máxima de nombre abreviado	6	6	6	6	6	6	6	6
Puntuación de nombre abreviado	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Factor de validez de nombre abreviado	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
Proporción de validez de nombre abreviado	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Valor máximo de validez de nombre abreviado	200	200	200	200	200	200	200	200

#### **Parámetros de comparación de NameHunter (coreano-ruso):**

En la tabla siguiente se proporcionan los parámetros de comparación predeterminado de los campos correspondientes al nombre de pila (GN) y apellido (SN) para los códigos de entorno cultural coreano, paquistaní, polaco, portugués y ruso

<b>Coreano - 5</b>		<b>Pakistaní - 13</b>		<b>Polaco - 18</b>		<b>Portugués - 19</b>		<b>Ruso - 6</b>	
GN	SN	GN	SN	GN	SN	GN	SN	GN	SN

	Coreano - 5		Pakistaní - 13		Polaco - 18		Portugués - 19		Ruso - 6	
Umbral de nombre	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
Umbral de campo	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
Valor de ponderación del campo	0.80	1.00	1.00	1.00	0.80	1.00	0.80	1.00	0.80	1.00
Raíz faltante	0.85	0.75	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.98
Comparar variantes	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Comparar iniciales	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
Inicial en inicial	0.85	0.00	0.85	0.00	1.00	0.00	1.00	0.00	0.80	0.00
Inicial en cadena de caracteres	0.85	0.00	0.75	0.00	0.85	0.00	0.85	0.00	0.85	0.00
Nombre desconocido	0.60	0.60	0.60	0.75	0.65	0.65	0.65	0.65	0.65	0.65
Sin nombre	0.60	0.65	0.60	0.75	0.65	0.65	0.65	0.65	0.65	0.65
Modalidad de puntuación	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG
Tipo de ancla	NONE	NONE	FIRST	LAST	NONE	NONE	NONE	NONE	FIRST	NONE
Factor de ancla	1.00	1.00	0.85	0.85	1.00	1.00	1.00	1.00	0.80	1.00
Factor Oops	0.85	0.75	0.85	0.90	0.90	0.80	0.90	0.80	0.90	0.85
Nombre comprimido	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Máximo comprimido	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Factor de puntuación comprimida	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sesgo a la izquierda	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Regularizado máx	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98

	Coreano - 5		Pakistaní - 13		Polaco - 18		Portugués - 19		Ruso - 6	
Longitud mínima de nombre abreviado	2	2	2	2	2	2	2	2	2	2
Longitud máxima de nombre abreviado	6	6	6	6	6	6	6	6	6	6
Puntuación de nombre abreviado	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97
Factor de validez de nombre abreviado	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75
Proporción de validez de nombre abreviado	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Valor máximo de validez de nombre abreviado	200	200	200	200	200	200	200	200	200	200

#### Parámetros de comparación de NameHunter (tailandés-yoruba):

La tabla siguiente proporciona los parámetros de comparación predeterminados de los campos correspondientes al nombre de pila (GN) y apellido (SN) para los códigos de entorno cultural tailandés, turco, vietnamita y yoruba.

	Tailandés - 9		Turco - 20		Vietnamita - 17		Yoruba - 11	
	GN	SN	GN	SN	GN	SN	GN	SN
Umbral de nombre	0.60	0.60	0.65	0.65	0.70	0.70	0.60	0.60
Umbral de campo	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
Valor de ponderación del campo	1.00	1.00	0.80	1.00	0.80	1.00	1.00	1.00
Raíz faltante	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.98
Comparar variantes	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

	Tailandés - 9		Turco - 20		Vietnamita - 17		Yoruba - 11	
Comparar iniciales	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
Inicial en inicial	0.90	0.00	1.00	0.00	0.85	0.00	0.85	0.00
Inicial en cadena de caracteres	0.85	0.00	0.85	0.00	0.85	0.00	0.85	0.00
Nombre desconocido	0.50	0.50	0.60	0.60	0.65	0.65	0.55	0.55
Sin nombre	0.50	0.50	0.60	0.60	0.65	0.65	0.55	0.55
Modalidad de puntuación	AVG	AVG	AVG	AVG	LOW	AVG	AVG	AVG
Tipo de ancla	NONE	NONE	NONE	LAST	NONE	NONE	NONE	NONE
Factor de ancla	1.00	1.00	1.00	0.70	1.00	1.00	1.00	1.00
Factor Oops	0.90	0.75	0.97	0.97	0.85	0.80	0.97	0.97
Nombre comprimido	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Máximo comprimido	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Factor de puntuación comprimida	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sesgo a la izquierda	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Regularizado máx	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Longitud mínima de nombre abreviado	2	2	2	2	2	2	2	2
Longitud máxima de nombre abreviado	6	6	6	6	6	6	6	6
Puntuación de nombre abreviado	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Factor de validez de nombre abreviado	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
Proporción de validez de nombre abreviado	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01

	Tailandés - 9		Turco - 20		Vietnamita - 17		Yoruba - 11	
Valor máximo de validez de nombre abreviado	200	200	200	200	200	200	200	200

### Parámetros de comparación de NameHunter (entornos culturales grupales):

En la tabla siguiente se proporcionan los parámetros de comparación predeterminados de los campos correspondientes al nombre de pila (GN) y apellido (SN) para todos los códigos de entornos culturales grupales:

#### Sudoeste asiático

Entorno cultural grupal que comprende las culturas afgana, árabe, persa (farsi), pakistaní y turca.

#### Europeo

Entorno cultural grupal que comprende las culturas anglosajona, francesa, germánica, hispánica y portuguesa.

**Han** Entorno cultural grupal que comprende las culturas china, coreana y vietnamita.

	Sudoeste asiático - 38		Europeo - 39		Han - 40	
	GN	SN	GN	SN	GN	SN
Umbral de nombre	0.65	0.65	0.65	0.65	0.70	0.70
Umbral de campo	0.49	0.49	0.49	0.49	0.49	0.49
Valor de ponderación del campo	1.00	1.00	0.80	1.00	0.80	1.00
Raíz faltante	0.99	0.98	0.99	0.98	0.70	0.70
Comparar variantes	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Comparar iniciales	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
Inicial en inicial	0.85	0.00	0.85	0.00	0.85	0.00
Inicial en cadena de caracteres	0.75	0.00	0.75	0.00	0.85	0.00
Nombre desconocido	0.60	0.75	0.60	0.60	0.60	0.60
Sin nombre	0.60	0.75	0.60	0.60	0.60	0.60
Modalidad de puntuación	AVG	AVG	AVG	AVG	LOW	AVG
Tipo de ancla	FIRST	NONE	NONE	NONE	NONE	NONE
Factor de ancla	0.90	1.00	1.00	1.00	1.00	1.00
Factor Oops	0.85	0.90	0.90	0.75	0.85	0.75
Nombre comprimido	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

	Sudoeste asiático - 38		Europeo - 39		Han - 40	
Máximo comprimido	0.95	0.95	0.95	0.95	0.95	0.95
Factor de puntuación comprimida	1.00	1.00	1.00	1.00	1.00	1.00
Sesgo a la izquierda	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Regularizado máx	0.98	0.98	0.98	0.98	0.98	0.98
Longitud mínima de nombre abreviado	2	2	2	2	2	2
Longitud máxima de nombre abreviado	6	6	6	6	6	6
Puntuación de nombre abreviado	0.97	0.97	0.97	0.97	0.97	0.97
Factor de validez de nombre abreviado	0.75	0.75	0.75	0.75	0.75	0.75
Proporción de validez de nombre abreviado	0,01	0,01	0,01	0,01	0,01	0,01
Valor máximo de validez de nombre abreviado	200	200	200	200	200	200

## Configurar conjuntos de reglas de transliteración para NameHunter

Para utilizar los conjuntos de reglas de transliteración adicionales en NameHunter, debe configurar NameHunter, NHServer y el proceso de Distributed Search.

### Configurar NameHunter para que utilice los archivos de reglas de transliteración

IBM InfoSphere Global Name Management contiene archivos de reglas que funcionan con la función de transliteración de NameHunter. Para poder utilizar los archivos de reglas, debe activar la transliteración y luego llamar a la función adecuada para cargar un archivo de reglas específico.

#### Antes de empezar

- Active la transliteración utilizando la función transliterate.
- Asegúrese de que el archivo de reglas que desea utilizar se encuentra en la vía de acceso.

## Acerca de esta tarea

Después de crear una instancia de NameHunter, llame a la función siguiente:

```
void NameHunter::loadTransRules(const std::string& nombreArchivoReglas);
```

sustituyendo *nombreArchivoReglas* por la ubicación del archivo de reglas correspondiente.

**Nota:** Para obtener los mejores resultados, el archivo de reglas debe encontrarse en la vía de acceso. Si el archivo de reglas no se encuentra en la vía de acceso, NameHunter lanza una excepción.

## Ejemplo de llamada al archivo de reglas de transliteración

Por ejemplo, para llamar al archivo de reglas de árabe, llame a la función

```
void NameHunter::loadTransRules(const std::string& arabicTransRule.ibm);
```

## Configurar el proceso de Distributed Search para que utilice archivos de reglas de transliteración

IBM InfoSphere Global Name Management contiene archivos de reglas que modifican la función de transliteración. Una vez modificado el archivo de configuración del proceso de Distributed Search para que incluya los nombres y ubicaciones de los archivos de reglas, el proceso de Distributed Search puede comparar los nombres escritos en los scripts manejados por dichos archivos de reglas.

### Procedimiento

1. Abra el archivo de configuración del proceso de Distributed Search. El archivo predeterminado es `dsconfig.ini`.
2. En la sección `[searcherCommon]`, especifique los archivos de transliteración que desea incluir. El ejemplo siguiente incluye los archivos de transliteración para coreano y japonés.

```
[searcherCommon]
```

```
koreanTransFile=koreanTransRule.ibm  
japaneseTransFile=japaneseTransRule.ibm
```

---

## Buscar nombres mediante Distributed Search

NameHunter Distributed Search permite escalar el número de búsquedas simultáneas permitidas y el número de nombres que deben buscarse (hasta 200 millones) añadiendo más recursos de procesador para dar soporte a servidores de búsqueda adicionales.

El diagrama siguiente muestra el diseño de alto nivel de Distributed Search. Varios clientes (que no forman parte de Distributed Search) se comunican con un proceso de middleware de comunicaciones central mediante XML a través de TCP/IP. El proceso de middleware coloca en cola las solicitudes de cliente y las envía a un conjunto de servidores de búsqueda. El proceso de middleware gestiona las respuestas de los servidores de búsqueda y devuelve una respuesta total al cliente solicitante. Los clientes no tienen conocimiento de que varios servidores han contribuido a la respuesta generada para sus solicitudes.

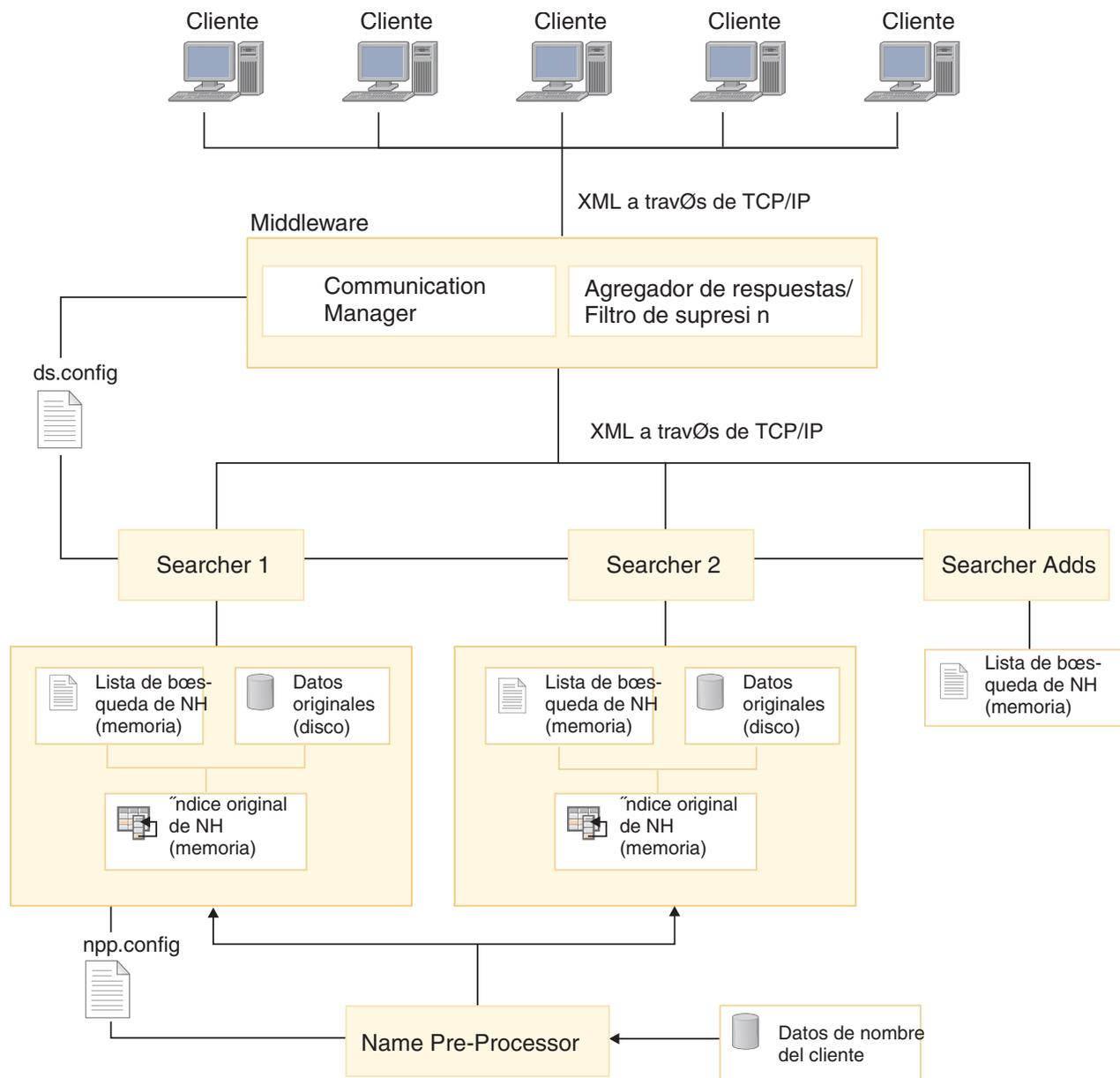


Figura 4. Diseño de alto nivel de Distributed Search

Cada servidor de Distributed Search gestiona una parte de una lista mayor de datos de nombre. La lista de datos gestionada por cada servidor de Distributed Search consta de una lista de búsqueda residente en memoria, junto con un índice de un repositorio basado en disco de los datos de nombre originales suministrados por el usuario. Generalmente, se asigna un servidor para manejar las actualizaciones que deban efectuarse en el repositorio combinado de nombres durante una sesión de Distributed Search, por ejemplo cuando se crean solicitudes de adición.

Los datos gestionados por cada servidor se crean mediante un programa de utilidad de Distributed Search denominado Name Preprocessor. Cuando se le suministra un archivo de datos de nombre delimitados por comas, Name Preprocessor analiza, clasifica, regulariza, elimina duplicaciones y particiona un archivo de nombres de gran tamaño en partes más pequeñas que todos los

servidores participantes en Distributed Search pueden compartir. La salida de Name Preprocessor es binaria y específica del sistema.

**Atención:** El programa de utilidad Name Preprocessor debe ejecutarse en un sistema que tenga el mismo sistema operativo (SO) y arquitectura que aquellos en los que se ejecutan los procesos de Distributed Search. Distributed Search fallará si Name Preprocessor se ejecuta en otro sistema operativo.

## Introducción a Name Preprocessor

Name Preprocessor es un programa de utilidad que convierte un archivo delimitado por comas de registros de nombre de cliente en los archivos necesarios para Distributed Search.

Name Preprocessor combina funciones de análisis de nombres (transliteración, análisis y clasificación) con los pasos internos de proceso de NameHunter (regularización y limpieza) para generar un conjunto de nombres adecuados como entrada de un proceso de Distributed Search.

**Atención:** El programa de utilidad Name Preprocessor debe ejecutarse en un sistema que tenga el mismo sistema operativo (SO) y arquitectura que aquellos en los que se ejecutan los procesos de Distributed Search. Distributed Search fallará si Name Preprocessor se ejecuta en otro sistema operativo.

### Visión general de los archivos de datos de Name Preprocessor

Name Preprocessor utiliza y genera varios archivos de texto delimitado por comas. Name Preprocessor divide la información en partes uniformes y elimina duplicaciones para que el usuario disponga de entradas de datos exclusivas con el formato esperado por los procesos de Distributed Search y Embedded Search.

El archivo Datos de cliente es la información de entrada de Name Preprocessor, donde se analiza clasifica, regulariza, translitera y se comprueba si existen duplicaciones. Este paso inicial genera un archivo temporal delimitado por comas conocido como archivo Datos provisionales. El resultado del proceso de eliminación de duplicaciones, en el que el archivo Datos provisionales se divide en fragmentos más pequeños, genera tres archivos de salida que se dividen en las estructuras de datos suministradas a los procesos de Distributed Search y Embedded Search:

- Lista de búsqueda de NH
- Datos originales
- Índice original de NH

La entrada de datos en Name Preprocessor no es el valor predeterminado al utilizar Distributed Search. Puede pasar por alto Name Preprocessor y cargar un proceso buscador directamente con datos de nombre, si ya ha procesado la información con otra aplicación. Si desea utilizar Name Preprocessor para procesar la información, especifique el archivo que desea procesar en el archivo de configuración de Name Preprocessor, npp.config.

### Flujo de Name Preprocessor

Considere el registro de nombre de entrada siguiente:

Jirí, Válek, 1234

Este registro pasará por las siguientes fases de proceso en Name Preprocessor:

1. El componente de transliteración de nombres elimina los caracteres ASCII ampliados para generar este formulario:  
JIRI, VALEK, 1234
2. El componente NameParser desglosa el nombre completo en los componentes de nombre de pila (GN) y apellido (SN) y comprueba que el orden del nombre suministrado originariamente en los datos de cliente es probable que sea incorrecto. Se genera un segundo orden de análisis alternativo para el nombre:  
Valek, Jiri, 1234
3. NameClassifier decide que SN y GN están asociados al entorno cultural europeo (código de entorno cultural 39).
4. Opcionalmente, pueden aplicarse al nombre de entrada reglas lingüísticas que suavizan los patrones principales de variación de deletreo en función del entorno cultural (regularización de nombres).

Así, el nombre de entrada original genera los siguientes registros de salida provisionales en formato CSV en Name Preprocessor:

```
1234,P,JIRI,VALEK,39,39,Jirí,Válek,0,0  
1234,P,VALEK,JIRI,39,39,Jirí,Válek,1,0
```

Los datos de nombre originales permanecen sin cambios. Durante el segundo paso del proceso realizado por Name Preprocessor, el archivo provisional se clasifica, se eliminan las duplicaciones y se divide en las estructuras de datos necesarias para los servidores de Distributed Search y Embedded Search.

#### **Preproceso de listas de nombres con Embedded Search:**

Embedded Search expone las prestaciones de búsqueda de NameHunter para IBM NameWorks en un único proceso en el que los datos de nombre se preprocesan cuando se inicializa IBM NameWorks, en lugar de hacerlo en un paso separado.

Embedded Search maneja el preproceso de nombres para reducir la sobrecarga de administración y aumentar el rendimiento global. A continuación, la aplicación podrá iniciar la búsqueda sin necesidad de preprocesar los datos.

Los archivos de datos de nombre que deben utilizarse con Embedded Search deben suministrarse en formato de valores separados por comas (CSV), en el que cada registro incluya entre tres y diez campos. Se presupone que la codificación de estos archivos de datos será ASCII o UTF-8. El texto con otras codificaciones provoca condiciones de error, que dan como resultado el lanzamiento de una excepción BADRECORD (GODW033E) que contiene el número y el contenido del registro de nombre.

**Importante:** Tanto Embedded Search como Distributed Search leen el archivo de datos provisionales de Name Preprocessor. Al preprocesar datos de nombre con Name Preprocessor, deben suministrarse todos los campos de datos, mientras que éstos son opcionales al preprocesar nombres con Embedded Search.

#### **Archivo Datos de cliente de Name Preprocessor:**

El archivo Datos de cliente constituye la información de entrada del programa de utilidad Name Preprocessor. Este archivo se procesa para generar el archivo Datos provisionales, que se divide en partes más pequeñas para poder pasarlo a Distributed Search.

El archivo Datos de cliente es un archivo delimitado por comas que contiene los campos siguientes:

**ID de registro**

Datos adicionales que deben asociarse con este registro (como por ejemplo un ID que señale al origen original de los datos). La longitud máxima es de 256 bytes.

**Categoría**

Categoría de nombre que se proporciona para el nombre. Si se deja en blanco, Name Preprocessor intentará determinar la categoría. Los valores válidos son:

- B = De persona y de organización
- O = Nombre de organización
- P = Nombre de persona

Los valores válidos provocarán un error del que se informará y se descartará el registro de nombres incorrectos.

**Tipo de script**

Proporciona el tipo de script del texto de nombre, en forma de número entre el 1 y el 8 para indicar un tipo de script específico. El 0 indica que no hay ningún tipo de script específico y el espacio en blanco indica que Name Preprocessor debería intentar determinar el tipo de script. Los valores válidos son:

- 0 = Ningún script específico.
- 1 = Script hanzi (chino)
- 2 = Script kanji/kana (japonés)
- 3 = Script devanagari (hindú)
- 4 = Script cirílico (ruso)
- 5 = Script latino (europeo, anglosajón)
- 6 = Script hangul (coreano)
- 7 = Script árabe (árabe, sudoeste asiático)
- 8 = Script griego (griego)

Los valores válidos provocarán un error del que se informará y se descartará el registro de nombres incorrectos.

**Apellido o nombre completo**

Si un registro de nombre contiene sólo tres campos, este campo se interpreta como el nombre completo. Si el registro de nombre se categoriza como nombre de organización (ya sea directamente en el archivo de datos o indirectamente mediante IBM NameWorks), se ignorarán los campos de nombre de pila y de entorno cultural del nombre de pila. En cualquier otro caso, el nombre se translitera, analiza, clasifica y añade a la lista de datos.

**Nombre de pila**

Elemento de nombre de pila, en el que el campo anterior se interpreta como elemento de apellido. Si no hay campos adicionales, los elementos de nombre de pila y apellido se transliteran, analizan, clasifican y añaden a la lista de datos.

**Código de entorno cultural de apellido**

Código de entorno cultural del apellido. Si este campo está en blanco o contiene el valor -1, IBM NameWorks clasifica el elemento de apellido y

utiliza el resultado de la clasificación como código de entorno cultural del apellido. Si este campo contiene un valor no válido, se notifica una condición de error.

#### **Código de entorno cultural de nombre de pila**

Código de entorno cultural del nombre de pila. Si este campo está en blanco o contiene el valor -1, IBM NameWorks clasifica el elemento de nombre de pila y utiliza el resultado de la clasificación como código de entorno cultural del nombre de pila. Si este campo contiene un valor no válido, se notifica una condición de error.

#### **Apellido original**

Este campo proporciona el elemento Apellido original. Este campo contiene el nombre completo original para un nombre de organización.

#### **Nombre de pila original**

Elemento de nombre de pila original.

#### **Distintivo de elemento alternativo**

Indica si el registro actual es un elemento alternativo del nombre original, independientemente de lo que pueda haberse determinado en campo anteriores (0 = no, 1 = sí).

#### **Archivo Datos provisionales de Name Preprocessor:**

El archivo Datos provisionales es un archivo temporal delimitado por comas que constituye la salida del primer paso del programa de utilidad Name Preprocessor. Este archivo es el resultado de aplicar Name Preprocessor (análisis, clasificación, etc.) al archivo Datos de cliente para que el archivo contenga toda la información necesaria para generar la entrada del servidor distribuido.

**Importante:** Tanto Embedded Search como Distributed Search leen el archivo de datos provisionales de Name Preprocessor. Al preprocesar datos de nombre con Name Preprocessor, deben suministrarse todos los campos de datos, mientras que éstos son opcionales al preprocesar nombres con Embedded Search.

En función de la categoría de nombres (de persona o de organización) del nombre, se incluyen campos diferentes. Si se determina que la categoría del nombre es Organization, sólo se incluyen los campos siguientes durante el preproceso de nombres:

#### **Formato de datos provisionales de nombre de organización**

- Datos complementarios
- Categoría
- Tipo de script
- Nombre
- Código de entorno cultural
- Nombre original

Las diferencias principales en el formato de datos provisionales consisten en que los nombres de organización se almacenan en un único campo.

El archivo de datos puede contener cualquiera de los campos siguientes. El número de campos suministrados determina los pasos de preproceso que se realizan.

#### **Datos complementarios**

Generalmente, un identificador de registro que puede utilizarse para

recuperar otra información relativa al registro de nombre de una o varias bases de datos, aunque este campo puede contener información adicional relativa al registro de nombre.

### **Categoría**

Una única letra que indica que el registro de nombre debe categorizarse como nombre de persona (P), nombre de organización (O) o ambos (B). Si este campo está vacío, IBM NameWorks categoriza el nombre antes de aplicar ningún otro proceso. Una letra no válida provocará una condición de error.

### **Tipo de script**

Proporciona el tipo de script del texto de nombre, en forma de número entre el 1 y el 8 para indicar un tipo de script específico. El 0 indica que no hay ningún tipo de script específico y el espacio en blanco indica que Name Preprocessor debería intentar determinar el tipo de script.

- 0 = Ningún script específico.
- 1 = Script hanzi (chino)
- 2 = Script kanji/kana (japonés)
- 3 = Script devanagari (hindú)
- 4 = Script cirílico (ruso)
- 5 = Script latino (europeo, anglosajón)
- 6 = Script hangul (coreano)
- 7 = Script árabe (árabe, sudoeste asiático)
- 8 = Script griego (griego)

### **Apellido o nombre completo**

Si un registro de nombre contiene sólo tres campos, este campo se interpreta como el nombre completo. Si el registro de nombre se categoriza como nombre de organización (ya sea directamente en el archivo de datos o indirectamente mediante IBM NameWorks), los campos adicionales se ignoran. En cualquier otro caso, el nombre se translitera, analiza, clasifica y añade a la lista de datos.

### **Nombre de pila**

Elemento de nombre de pila, en el que el campo anterior se interpreta como elemento de apellido. Si este campo está en blanco y el registro de nombre se categoriza como nombre de organización (ya sea directamente en el archivo de datos o indirectamente mediante IBM NameWorks), los campos adicionales se ignoran. Si no hay campos adicionales, los elementos de nombre de pila y apellido se transliteran, analizan, clasifican y añaden a la lista de datos.

### **Código de entorno cultural de apellido**

Código de entorno cultural del apellido. Si este campo está en blanco o contiene el valor -1, IBM NameWorks clasifica el elemento de apellido y utiliza el resultado de la clasificación como código de entorno cultural del apellido. Si este campo contiene un valor no válido, se notifica una condición de error.

### **Código de entorno cultural de nombre de pila**

Código de entorno cultural del nombre de pila. Si este campo está en blanco o contiene el valor -1, IBM NameWorks clasifica el elemento de nombre de pila y utiliza el resultado de la clasificación como código de

entorno cultural del nombre de pila. Si este campo contiene un valor no válido, se notifica una condición de error.

**Apellido original**

Elemento de apellido original. Este campo contiene el nombre completo original para un nombre de organización.

**Nombre de pila original**

Elemento de nombre de pila original.

**Distintivo de elemento alternativo**

Indica si el registro actual es un elemento alternativo del nombre original, independientemente de lo que pueda haberse determinado en campo anteriores (0 = no, 1 = sí).

*Archivo Lista de búsqueda de Name Preprocessor:*

El archivo Lista de búsqueda de NH es la entrada de Distributed Search que contiene la información de nombres de la que se han eliminado las duplicaciones.

Los archivos Lista de búsqueda, Datos originales e Índice original NH son el resultado del proceso de eliminación de duplicaciones, en el que el archivo Datos provisionales se divide en partes más pequeñas. Cada división del archivo provisional genera estos tres archivos.

El archivo Lista de búsqueda es un archivo delimitado por comas que es idéntico al archivo "Datos provisionales" con dos excepciones cuando está habilitada la modalidad de nombre exclusivo (createOrig=true en el archivo de configuración npp).

- El campo "Datos complementarios" contiene un valor de índice que se utiliza para buscar todos los registros coincidentes de un nombre exclusivo.
- Un valor adicional (numOrig) aparece como el último campo que indica el número de registros originales asociados a un nombre exclusivo.

**Datos complementarios**

Generalmente, un identificador de registro que puede utilizarse para recuperar otra información relativa al registro de nombre de una o varias bases de datos, aunque este campo puede contener información adicional relativa al registro de nombre.

**Categoría**

Una única letra que indica que el registro de nombre debe categorizarse como nombre de persona (P), nombre de organización (O) o ambos (B). Si este campo está vacío, IBM NameWorks categoriza el nombre antes de aplicar ningún otro proceso. Una letra no válida provocará una condición de error.

**Tipo de script**

Indica el tipo de script del texto del nombre. Un entero entre 1 y 8 para indicar un tipo de script específico o 0 para indicar ningún tipo de script específico.

**Apellido o nombre completo**

Si un registro de nombre contiene sólo tres campos, este campo se interpreta como el nombre completo. Si el registro de nombre se categoriza como nombre de organización (ya sea directamente en el archivo de datos o indirectamente mediante IBM NameWorks), los campos adicionales se ignoran. En cualquier otro caso, el nombre se translitera, analiza, clasifica y añade a la lista de datos.

**Nombre de pila**

Elemento de nombre de pila, en el que el campo anterior se interpreta como elemento de apellido. Si este campo está en blanco y el registro de nombre se categoriza como nombre de organización (ya sea directamente en el archivo de datos o indirectamente mediante IBM NameWorks), los campos adicionales se ignoran. Si no hay campos adicionales, los elementos de nombre de pila y apellido se transliteran, analizan, clasifican y añaden a la lista de datos.

**Código de entorno cultural de apellido**

Código de entorno cultural del apellido. Si este campo está en blanco o contiene el valor -1, IBM NameWorks clasifica el elemento de apellido y utiliza el resultado de la clasificación como código de entorno cultural del apellido. Si este campo contiene un valor no válido, se notifica una condición de error.

**Código de entorno cultural de nombre de pila**

Código de entorno cultural del nombre de pila. Si este campo está en blanco o contiene el valor -1, IBM NameWorks clasifica el elemento de nombre de pila y utiliza el resultado de la clasificación como código de entorno cultural del nombre de pila. Si este campo contiene un valor no válido, se notifica una condición de error.

**Apellido original**

Elemento de apellido original. En el caso de un nombre de organización, este campo se determina como el nombre completo original.

**Nombre de pila original**

Elemento de nombre de pila original.

**Distintivo de elemento alternativo**

Indica si el registro actual es un elemento alternativo del nombre original, independientemente de lo que pueda haberse determinado en campos anteriores (0 = no, 1 = sí).

**numOrig**

Indica el número de registros originales asociados con un nombre exclusivo.

*Archivo Datos originales de Name Preprocessor:*

El archivo Datos originales contiene datos de cliente con dos campos adicionales que muestran los resultados del preproceso.

Los archivos Lista de búsqueda, Datos originales e Índice original NH son el resultado del proceso de eliminación de duplicaciones, en el que el archivo de datos provisionales se divide en partes más pequeñas. Cada división del archivo provisional genera estos tres archivos.

El archivo Datos originales contiene los campos siguientes:

**SN** Apellido (preprocesado).

**GN** Nombre de pila (preprocesado).

**custID**

Datos adicionales que deben asociarse con este registro (como por ejemplo un ID que señale al origen original de los datos). La longitud máxima es de 256 bytes.

**altParseFlag**

¿Ha generado este nombre un elemento de nombre alternativo (0 = no, 1 = sí)?

**regFlag**

¿Ha necesitado regularizarse este nombre (0 = no, 1 = sí)?

*Archivo Índice original de Name Preprocessor:*

El archivo Índice original asocia los archivos Lista de búsqueda de NH y Datos originales.

Los archivos Lista de búsqueda, Datos originales e Índice original NH son el resultado del proceso de eliminación de duplicaciones, en el que el archivo de datos provisionales se divide en partes más pequeñas. Cada división del archivo provisional genera estos tres archivos.

Distributed Search utiliza el archivo Índice original para recuperar los datos de nombre originales de un resultado de NameHunter. El archivo Índice original contiene dos campos:

**nhID** ID numérico exclusivo asignado a cada nombre preprocesado exclusivo almacenado en la lista de búsqueda de NH.

**origOffset**

Desplazamiento al archivo Datos originales que señala al primer registro de cliente.

**Gestión de listas de nombres con el programa de utilidad Name Loader:**

Puede gestionar listas de nombres de Enterprise Name Search con el programa de utilidad Name Loader. El programa de utilidad Name Loader es un programa Java autónomo que toma los nombres de origen sin formato de listas de nombres, analiza los nombres mediante IBM NameWorks, y almacena los nombres y los resultados de los análisis de nombre en la base de datos ENS.

El programa de utilidad Name Loader para ENS es similar en función al programa de utilidad Name Preprocessor utilizado en Distributed Search, pero no lo sustituye. El programa de utilidad Name Loader carga los nombres a partir de un archivo de lista de nombres o del origen de datos externo como una lista "Passengers" o "Watch List". Los "nombres de origen" son los nombres que se han recibido originalmente, con su análisis y script originales. Éste es el formato en el que se visualizan los resultados de una búsqueda.

IBM NameWorks analiza, desglosa y transcribe (o translitera) los nombres de origen como corresponde. Los nombres resultantes se conocen como "nombres de búsqueda". Los nombres de búsqueda se almacenan en la base de datos ENS en formato interno (caracteres latinos en mayúsculas) y se cargan en IBM NameWorks para su uso en las búsquedas. ENS mantiene los dos formatos en las tablas de la base de datos y gestiona la correlación entre ellos.

Además, puede utilizar el programa de utilidad Name Loader para volver a cargar o eliminar nombres. La célula de búsqueda pueden permanecer activo al utilizar Name Loader para cargar, volver a cargar y eliminar nombres de una lista.

## Lectura de nombres de origen desde los archivos de listas de nombres

Los archivos de listas de nombres son archivos de texto plano con formato CSV (valores separados por comas). Cada fila de la lista de nombres representa un nombre de origen. El archivo CSV contiene la información siguiente para cada fila:

Tabla 10. Campos de los archivos CVS de nombres entrantes

Campos	Significado	Ejemplo	Valor predeterminado	Notas
0	Apellido	"SMITH"	blanco	O el apellido o el nombre de pila puede ser un espacio en blanco, pero no los dos.
1	Nombre de pila	"JOHN"	blanco	O el apellido o el nombre de pila puede ser un espacio en blanco, pero no los dos.
2	idData	300011	sin valor predeterminado	Obligatorio. Identificador que debe ser exclusivo dentro de un origen de datos.
3	Cultura del apellido	3	blanco	Opcional
4	Cultura del nombre de pila	3	blanco	Opcional
5	Categoría	0	blanco	Opcional. Indica el tipo de nombre, ya sea de persona, de organización o no especificado.

### El componente Analyzer de Name Loader

El componente Analyzer de Name Loader realiza el análisis siguiente en cada nombre de origen de entrada:

- Procesa cada nombre de origen mediante la transliteración del nombre.
- Categoriza el nombre en un tipo de nombre si el elemento "Category" está en blanco en el archivo CSV plano.
- Busca el mejor análisis de la forma original del nombre si la categoría es "personal".
- Busca uno o varios nombres de búsqueda transliterados en el nombre de origen.

### Nombre de origen y nombres de búsqueda

Los *nombres de origen* son los nombres que se han cargado originalmente desde una lista de nombres. En cambio, los *nombres de búsqueda* son los nombres que resultan del análisis que realiza IBM NameWorks. El esquema almacena los dos tipos, nombre de origen y nombre de búsqueda.

Los nombres de origen se encuentran en la tabla ENS\_SOURCE\_NAME. Los nombres de búsqueda se encuentran en la tabla ENS\_SEARCH\_NAME. Enterprise Name Search gestiona la correlación entre los dos tipos. Por razones de rendimiento, cuando el nombre se graba en las tablas de esquema, la operación de escritura se realiza mediante varias hebras. Cada hebra utiliza una actualización de base de datos por lotes para añadir o actualizar varios registros. El número de hebras y el tamaño del lote se pueden configurar en el archivo de configuración del programa de utilidad Name Loader.

Si dos nombres de búsqueda son idénticas excepto por el entorno cultural, ambos nombres de búsqueda se almacenan por separado. Por ejemplo, "June Park " con una designación de entorno cultural coreano se almacena por separado de "June Park" con una designación de entorno cultural anglosajón porque las reglas de coincidencia de estos dos entornos culturales son distintas.

Si dos nombres de búsqueda son idénticos excepto por el distintivo de elemento alternativo, ambos nombres también se almacenan por separado. Por ejemplo, el nombre de origen "Elton John" es diferente del nombre de origen "John Elton". Una solicitud de búsqueda de nombres de "Elton John" busca e informa los dos nombres de origen, pero el nombre de origen de John Elton se notifica con una puntuación de coincidencia inferior porque está basado en un elemento alternativo.

### **Archivo loader.config**

El programa de utilidad Name Loader tiene su propio archivo de configuración, denominado loader.config, que especifica el comportamiento de la instancia de IBM NameWorks utilizada por Enterprise Name Search. Toda la información sobre las opciones y configuración del programa de utilidad Name Loader se especifica en este archivo.

### **Archivo de configuración de Name Preprocessor**

La entrada, los pasos y la salida de Name Preprocessor están controlados por un archivo de configuración, npp.config, cuyo formato es el de un archivo .ini de Microsoft Windows.

La categoría de nombre (NameCategory) del nombre de entrada determina la lista de datos (de organización o de persona) a la que se añade el nombre. Si no se especifica ningún valor para NameCategory, Name Preprocessor utiliza NameSifter para determinarla. Si el valor de NameCategory es unknown (desconocida) y la opción de examen está establecida en false (doCategorize=false), Name Preprocessor trata el nombre asociado que no es de entidad como un nombre de persona.

Para especificar los archivos de cadenas de caracteres externas, añada cadenas de caracteres de análisis personalizadas a la sección Custom Tokens del archivo npp.config.

### **Archivo Sample npp.config**

```
[npp]
inFile=names.txt
interimFile=names.npp.txt
nppOutFile=names.npp.out
origDataFile=names.orig.dat
origIndexFile=names.orig.idx

sifterRulesFile=SifterRules.ibm

arabicTransFile=arabicTransRule.ibm
cyrillicTransFile=cyrillicTransRule.ibm
greekTransFile=greekTransRule.ibm
koreanTransFile=koreanTransRule.ibm
latinTransFile=latinTransRule.ibm
japaneseTransFile=japaneseTransRule.ibm
chineseTransFile=chineseTransRule.ibm

reportIncrement=100000
doFullName=true
doCategorize=false
```

```

doTransliterate=true
doParse=true
doClassify=true
doNhClean=true
parseThreshold=0.5

maxRecsPerSplit=2000000
createNpp=true
createNh=true
numNppOutFiles=2
deleteTempFiles=true
deleteNppFile=false
createOrig=true

logDebug=cout
logError=cout
logEvent=cout

maxGnCacheSize=4000000
maxOnCacheSize=0
maxSnCacheSize=4000000

```

### Proceso de Distributed Search cargado directamente

Un proceso de Distributed Search puede cargarse directamente con datos de nombre sin preprocesar para pasar por alto el preproceso de nombres. En este caso, sólo la lista de búsqueda de NH es necesaria para el proceso. Los datos originales y el índice original de NH creados por Name Preprocessor no se utilizan.

En esta modalidad, el formato del archivo Lista de búsqueda de NH es el siguiente: Todos los campos de estos archivos de formato CSV admiten campos entrecomillados, que permiten la aparición de comas en los archivos de entrada.

#### Nombre de persona

```
ID,category,SN,GN,cultureSN,cultureGN,origSN,origGN,altParse,0,count
```

#### Nombre de organización

```
ID,category,ON,origON,altParse,0,count
```

**ID** ID numérico exclusivo asignado a cada nombre no preprocesado exclusivo.

#### **category**

Categoría del nombre; P (nombre de persona) o 0 (nombre de organización)

**SN** Apellido

**GN** Nombre de pila

**ON** Nombre de organización

#### **cultureSN**

Código de entorno cultural del apellido

#### **cultureGN**

Código de entorno cultural del nombre de pila

#### **origSN**

Apellido original

#### **origGN**

Nombre de pila original

**altParse**

Indica si el nombre se ha generado a partir de un elemento de nombre alternativo (0 = no, 1 = sí). Especifique el valor 1 si sabe que el nombre es un elemento de nombre alternativo.

**count** Indica el número de registros que hace referencia a este ID. Si está ejecutando en modalidad de búsqueda completa, cada ID está vinculado a un registro, por lo que el recuento será cero. Si está en modalidad de nombre exclusivo, cada registro puede hacer referencia a varios ID, por lo que el recuento será mayor que cero.

**Ejecutar Name Preprocessor**

Puede ejecutar Name Preprocessor desde un indicador de mandatos.

**Antes de empezar**

Si no se especifica un nombre de archivo en la línea de mandatos, Name Preprocessor abre su archivo de configuración predeterminado, `npp.config`, en el directorio actual durante la invocación. Generalmente, este es el directorio desde el que se invoca Name Preprocessor. Por ejemplo:

```
C:> npp
```

Puede especificar otro directorio para el archivo de configuración en la línea de mandatos al ejecutar Name Preprocessor. Por ejemplo:

```
C:> npp /usr/GNR/data/npp.config
```

**Consideraciones relativas a la configuración de Name Preprocessor**

La decisión de cómo configurar Name Preprocessor requiere algunas consideraciones y experimentación para comprobar lo que mejor se ajusta a los datos y al entorno de sistema.

Deben tenerse en cuenta varias consideraciones en función de las tareas que desee ejecutar en los datos:

**Transliteración**

Convierte los nombres para utilizar sólo las 26 letras del alfabeto latino básico. La transliteración elimina los acentos de las letras latinas acentuadas y convierte los nombres escritos en los alfabetos árabe, chino, cirílico, griego, kana (hiragana y katakana japonés) y hangul (coreano) a las formas equivalentes con caracteres del alfabeto latino. Si los datos que deben procesarse ya utilizan el alfabeto básico latino de 26 letras, la transliteración puede inhabilitarse para mejorar el tiempo de preproceso.

**Análisis**

Divide un nombre en sus componentes de nombre de pila (GN) y apellido (SN). El análisis genera elementos de nombre adicionales si la confianza de NameParser es inferior al umbral. Generalmente, el análisis de datos de nombre aumenta el número de nombres en un 10%. Puede inhabilitar el análisis y ahorrar tiempo de preproceso suministrando nombres que ya estén divididos en los componentes de GN y SN.

**Clasificación**

Ejecuta NameClassifier para asignar códigos de entorno cultural a los apellidos y nombres de pila. El hecho de no clasificar los nombres ralentiza el tiempo de búsqueda e inhabilita la regularización, por lo que es muy aconsejable realizar la clasificación.

### Regularización

Genera formas normalizadas de los nombres de entrada si existen reglas de regularización de IBM para el entorno cultural. La regularización puede mejorar significativamente la calidad de la búsqueda, pero también ralentiza el proceso de búsqueda debido al mayor número de nombres de entrada.

### NhClean

La función de limpieza de NameHunter elimina todos los caracteres excepto los espacios y las letras mayúsculas de la A a la Z. La operación de limpieza es rápida y debe aumentar el número de entradas duplicadas, lo cual conduce a una reducción del número de registros que Distributed Search debe cargar.

### Dimensionamiento

El número de buscadores que deben ejecutarse simultáneamente es una función que afecta al número de nombres, las opciones habilitadas y el número de procesadores disponibles. Las opciones seleccionadas determinan el rendimiento básico y afectan al número de buscadores utilizados, en función de los requisitos. El ejemplo siguiente ayuda a explicar las implicaciones del dimensionamiento:

- Se empieza con 50.000.000 nombres de entrada.
- El análisis añade 5.000.000 nombres, lo que da un total de 55.000.000.
- La regularización añade 20.000.000 nombres, lo que arroja un nuevo total de 75.000.000.
- La eliminación de duplicaciones de Name Preprocessor reduce el número de nombres exclusivos a 50.000.000. Ésta es una reducción típica, pero puede ser muy superior o inferior.
- Hay ocho procesadores disponibles para búsquedas distribuidas. Uno de ellos está reservado para adiciones, por lo que los 50.000.000 nombres exclusivos pueden dividirse por siete, suministrando aproximadamente 7.000.000 nombres por procesador. Por tanto, establecerá la entrada de configuración **numNppOutFiles** en 7.

Puede utilizar Name Preprocessor para dividir la lista de datos en tantas sublistas como sea necesario mediante la opción de configuración **numNppOutFiles**.

## Visión general del rendimiento y la configuración de Distributed Search

Esta sección ofrece información relativa a cómo configurar, ajustar y ejecutar Distributed Search.

Distributed Search consume intensivamente recursos de memoria y procesador. Un proceso de Distributed Search consume el 100% de la capacidad de proceso de un sistema cuando se procesa una solicitud. Por tanto, configure Distributed Search de modo que tenga como mínimo tantos procesadores como procesos de búsqueda. El gestor de comunicaciones no requiere un procesador dedicado siempre y cuando el número de búsquedas que soporte sea razonable.

La velocidad con la que Distributed Search puede manejar consultas es directamente proporcional al número de procesos de búsqueda que operan en paralelo. A medida que aumenta el número de búsquedas y disminuye el tamaño de los datos soportados por cada búsqueda, el rendimiento decrece, ya que el coste de la gestión de mensajes pasa a ser más y más significativo.

## Memoria de Distributed Search

Cada proceso de Distributed Search debe cargar todos sus datos de nombre en la memoria. Si un proceso de búsqueda debe paginar en disco, el rendimiento será inaceptable.

La estimación de la cantidad de memoria consumida por cada proceso de búsqueda resulta afectado por muchos factores, que incluyen el promedio de longitud de los nombres de entrada, si se utiliza o no la regularización y cómo Name Preprocessor ha preprocesado los nombres. En la mayoría de los casos, si sabe cuántos nombres se cargarán en un buscador, puede utilizar la regla empírica siguiente para realizar la estimación de la memoria necesaria:

$$100 \text{ MB} + \text{NúmeroDeNombres} * 300 = \text{memoria necesaria}$$

100 MB es la cantidad de memoria necesaria si carga todos los archivos de soporte de NameHunter. 300 bytes por nombre es un cálculo prudente del almacenamiento necesario. Por tanto, si carga 10 millones de nombres, la memoria necesaria será:

$$100 \text{ MB} + 10.000.000 * 300 = 3,1 \text{ GB}$$

Además de la memoria para nombres de datos, NameHunter también requiere memoria para almacenar posibles coincidencias durante la búsqueda. Cada búsqueda requiere hasta 1.000 bytes para cada posible coincidencia hallada con una puntuación superior al umbral de coincidencias. Puede que sea necesario realizar un seguimiento de 40.000 o más posibles coincidencias incluso cuando solamente se devuelven las 50 primeras. También deberán examinarse todos los nombres para localizar lo que tengan puntuaciones de similitud más elevadas. Los servidores de Distributed Search suelen ser de una sola hebra, por lo que solamente es necesaria una lista de posibles coincidencias en cualquier momento en concreto. Es prudente permitir unos 100 MB adicionales, o más, para recopilar los resultados de la búsqueda.

## Velocidad de transacción de Distributed Search

En tamaños superiores a un millón de nombres, el tiempo necesario para realizar una búsqueda es muy superior al necesario para gestionar los mensajes XML. Suponiendo que la asignación de procesador y de memoria sean suficientes, un proceso de Distributed Search puede realizar alrededor de 10 consultas por segundo con respecto a un millón de nombres. Esta ecuación aumenta linealmente. Con 10 millones de nombres, un proceso de búsqueda puede realizar una consulta por segundo. El rendimiento del sistema depende de la composición de los datos y de la potencia del procesador.

Por ejemplo, una lista de 100 millones de nombres se reducirá a 64 millones de nombres exclusivos después del preproceso de nombres. En un servidor con 8 procesadores, esta lista puede dividirse en subconjuntos de 8 millones de nombres. Con esta configuración, Distributed Search puede dar soporte a más de una consulta por segundo. Si hay 16 procesadores, el subconjunto se reduce a 4 millones de nombres y la velocidad de transacción será ligeramente superior a dos consultas por segundo.

## Archivo de configuración y valores de Distributed Search

Distributed Search consta de un proceso que gestiona las comunicaciones (commgr) y de uno o varios procesos que realizan búsquedas y actualizaciones (searcher). Cada uno de los buscadores (searcher) necesita acceso a un conjunto de archivos de soporte lingüístico de tiempo de ejecución y necesita saber dónde buscar su parte de datos de nombre correspondiente. El archivo de configuración predeterminado que se suministra con Distributed Search se denomina ds.config. Este archivo contiene los valores de configuración que cada proceso de búsqueda

necesita cuando se inicia una sesión. Algunos de estos valores se comparten entre todos los buscadores, y algunos de ellos son utilizados únicamente por una instancia específica del proceso de búsqueda.

### Archivo de ejemplo ds.config

A continuación se muestra el contenido del archivo de configuración predeterminado. Estos valores configuran un sistema de Distributed Search con tres procesos de búsqueda. Los dos primeros procesos comparten partes del archivo de datos original, y el tercero se ejecuta sin datos, pero se configura para procesar adiciones.

Los elementos de nombre de archivo esperan que los archivos se encuentren en el directorio actual, a menos que se especifiquen nombres de archivo totalmente calificados con una vía de acceso (por ejemplo, \user\GNR\data\names.txt). Todos los elementos booleanos aceptan la entrada siguiente como true:

- true o t
- yes o y

Los elementos no son sensibles a las mayúsculas y minúsculas y cualquier otro valor se considerará falso (false).

```
[commgr]
listenPort=2345
ipv6=
sleepMsec=10
waitConnectSec=5
heartbeatSec=60
msgBuffSize=1000000
logDebug=
logError=cout
logEvent=cout
logMessage=
numSearchers=3

[searcherCommon]
compparmsDefaults=compparms.config

ibmTaqFile=taq.ibm
ibmGnvFile=gnv.ibm
ibmSnvFile=snv.ibm
ibmOnvFile=onv.ibm
ibmOnTermFile=terms.ibm

custTaqFile=
custGnvFile=
custSnvFile=
custOnvFile=
custOnTermFile=

arabicTransFile=arabicTransRule.ibm
cyrillicTransFile=cyrillicTransRule.ibm
greekTransFile=greekTransRule.ibm
koreanTransFile=koreanTransRule.ibm
latinTransFile=latinTransRule.ibm
japaneseTransFile=japaneseTransRule.ibm

afghaniRegFile=swasianRegRule.ibm
angloRegFile=angloRegRule.ibm
arabicRegFile=swasianRegRule.ibm
chineseRegFile=chineseRegRule.ibm
europeanRegFile=
farsiRegFile=farsiRegRule.ibm
```

```
frenchRegFile=frenchRegRule.ibm
genericRegFile=angloRegRule.ibm
germanRegFile=germanRegRule.ibm
hanRegFile=
hispanicRegFile=hispanicRegRule.ibm
indianRegFile=indianRegRule.ibm
indonesianRegFile=indoRegRule.ibm
japaneseRegFile=japaneseRegRule.ibm
koreanRegFile=koreanRegRule.ibm
pakistaniRegFile=swasianRegRule.ibm
russianRegFile=russianRegRule.ibm
southwestAsianRegFile=swasianRegRule.ibm
thaiRegFile=thaiRegRule.ibm
vietnameseRegFile=
yorubanRegFile=
```

```
genericOnRegFile=genericOnRegRule.ibm
```

```
isUnique=true
```

```
doTransliterate=true
doRegularize=true
doOnToPNListSearch=false
doCompressedBitsig=true
```

```
defaultMaxResults=100
defaultAltScoreFactor=0.97
```

```
numRecords=100000
reportIncrement=100000
```

```
allowFnuLnu=false
allowFnuInit=false
allowInitLnu=false
allowInitInit=false
```

```
[searcher1]
hostname=localhost
port=2346
ipv6=
doAdds=false
nameFile=names.nh.txt.1
origDataFile=names.orig.dat.1
origIndexFile=names.orig.idx.1
logDebug=
logError=cout
logEvent=cout
logMessage=
```

```
[searcher2]
hostname=localhost
port=2347
ipv6=
doAdds=false
nameFile=names.nh.txt.2
origDataFile=names.orig.dat.2
origIndexFile=names.orig.idx.2
logDebug=
logError=cout
logEvent=cout
logMessage=
```

```
[searcher3]
hostname=localhost
port=2348
ipv6=
doAdds=true
```

```
nameFile=  
origDataFile=  
origIndexFile=  
logDebug=  
logError=cout  
logEvent=cout  
logMessage=
```

### **Opciones de configuración de Distributed Search para el gestor de comunicaciones:**

Existen varias opciones de configuración para el proceso del gestor de comunicaciones de Distributed Search.

Las siguientes opciones de configuración controlan el gestor de comunicaciones (commgr).

#### **listenPort**

Puerto TCP/IP utilizado por los clientes para conectarse al Servidor distribuido.

**ipv6** Una entrada opcional que indica si se debe utilizar el protocolo IPv6 para la comunicación con los procesos de llamada. Este valor debe coincidir con el que utilizan los procesos de llamada como, por ejemplo, en los datos de configuración NameWorks. Si no está presente o si se ha inhabilitado, se utilizará el protocolo IPv4.

#### **sleepMsec**

Número de milisegundos que el gestor de comunicaciones permanecerá inactivo entre bucles de proceso de mensajes. El valor predeterminado debería ser suficiente, pero se acepta cualquier valor de 10 o superior.

#### **waitConnectSec**

Intervalo entre intentos de conexión de commgr para procesos de búsqueda.

#### **heartBeatSec**

Número de segundos entre mensajes de latido entre commgr y los procesos de búsqueda. Los mensajes de latido conservan activa una conexión durante períodos de inactividad.

#### **msgBuffSize**

Tamaño en bytes del almacenamiento intermedio TCP/IP utilizado para enviar y recibir mensajes. El valor predeterminado, 1.000.000, debe ser más que suficiente.

#### **logDebug, logError, logEvent, logMessage**

Distributed Search tiene cuatro niveles de anotaciones que puede controlar con el archivo de configuración. Cada entrada reconoce lo siguiente:

- **filename** – nombre del archivo donde deben colocarse las entradas de anotaciones.
- **cout, cerr, clog** – las corrientes de salida estándar.
- **blanco** – no hay nada que indique a Distributed Search que esta anotación no es necesaria.

#### **logDebug**

Muestra mensajes de depuración durante el desarrollo. En producción, habrá muy poca salida en estas anotaciones.

**logError**

Muestra condiciones de error tales como entrada de usuario no válida, errores del sistema y problemas de comunicaciones.

**logEvent**

Anota eventos significativos tales como condiciones de inicio, transmisión y recepción de mensajes.

**logMessage**

Anota todos los mensajes entrantes y salientes. Genera gran cantidad de salida y debe inhabilitarse durante la operación habitual.

**numSearchers**

Número de búsquedas en ejecución.

**Opciones de configuración comunes de Distributed Search:**

Varias opciones son comunes a todos los procesos de Distributed Search. Estas opciones comunes se encuentran en la sección [searcherCommon].

Las entradas siguientes son comunes a todos los procesos de Distributed Search [searcherCommon]:

*Tabla 11. Opciones comunes de Distributed Search*

<b>Entrada</b>	<b>Descripción</b>
ibmTaqFile	Ubicación del archivo TAQ (fórmulas de tratamiento, afijos, calificadores) suministrado por IBM (taq.ibm).
ibmGnvFile	Ubicación del archivo de variantes de nombre de pila suministrado por IBM (gnv.ibm).
ibmSnvFile	Ubicación del archivo de variantes de apellido suministrado por IBM (snv.ibm).
ibmOnvFile	Ubicación del archivo de variantes de organización suministrado por IBM (onv.ibm).
ibmOnTermFile	Ubicación del archivo de términos proporcionado por IBM (terms.ibm).
custOnTermFile	Ubicación de un archivo términos personalizado.
custTaqFile	Ubicación de un archivo TAQ (fórmulas de tratamiento, afijos y calificadores) personalizado.
custGnvFile	Ubicación de un archivo de variantes de nombre de pila.
custSnvFile	Ubicación de un archivo de variantes de apellido personalizado.
arabicTransFile	Ubicación del archivo de reglas de transliteración del árabe de IBM (arabicTransRule.ibm).
cyrillicTransFile	Ubicación del archivo de reglas de transliteración del cirílico de IBM (cyrillicTransFile.ibm).
greekTransFile	Ubicación del archivo de reglas de transliteración del griego de IBM (greekTransFile.ibm).
latinTransFile	Ubicación del archivo de reglas de transliteración del latino de IBM (latinTransFile.ibm).
angloRegFile	Ubicación del archivo de reglas de regularización del inglés de IBM (angloRegRule.ibm).
arabicRegFile	Ubicación del archivo de reglas de regularización del árabe de IBM (arabicRegRule.ibm).

Tabla 11. Opciones comunes de Distributed Search (continuación)

Entrada	Descripción
chineseRegFile	Ubicación del archivo de reglas de regularización del chino de IBM (chineseRegRule.ibm).
farsiRegFile	Ubicación del archivo de reglas de regularización del farsi de IBM (farsiRegRule.ibm).
frenchRegFile	Ubicación del archivo de reglas de regularización del francés de IBM (frenchRegRule.ibm).
germanRegFile	Ubicación del archivo de reglas de regularización del alemán de IBM (germanRegRule.ibm).
hispanicRegFile	Ubicación del archivo de reglas de regularización del entorno cultural hispano de IBM (hispanicRegRule.ibm).
indianRegFile	Ubicación del archivo de reglas de regularización del hindú de IBM (indianRegRule.ibm).
indonesianRegFile	Ubicación del archivo de reglas de regularización del indonesio de IBM (indonesianRegRule.ibm).
japaneseRegFile	Ubicación del archivo de reglas de regularización del japonés de IBM (japaneseRegRule.ibm).
koreanRegFile	Ubicación del archivo de reglas de regularización del coreano de IBM (koreanRegRule.ibm).
russianRegFile	Ubicación del archivo de reglas de regularización del ruso de IBM (russianRegRule.ibm).
southwestasianRegFile	Ubicación del archivo de reglas de regularización del sudoeste asiático de IBM (southwestasianRegRule.ibm).
thaiRegFile	Ubicación del archivo de reglas de regularización del tailandés de IBM (thaiRegRule.ibm).
genericRegFile	Ubicación del archivo de reglas de regularización genéricas de IBM (genericRegRule.ibm).
isUnique	Indica a Distributed Search si Name Preprocessor ha eliminado las duplicaciones de los datos de entrada y que existen archivos de datos originales.
doTransliterate	Indica si se habilita la transliteración para los nombres de entrada.
doRegularize	Indica si deben regularizarse los nombres de entrada.
defaultAltScoreFactor	Valor predeterminado de un factor de penalización que se aplica si se utiliza un análisis alternativo para la comparación de nombres. Este valor puede alterarse temporalmente mediante mensajes de parámetro y mensajes de búsqueda.
defaultMaxResults	Número máximo de resultados predeterminado que una solicitud de búsqueda puede devolver. Puede alterarse temporalmente mediante mensajes de parámetro y búsqueda.
allowFnuLnu	Controla si están permitidas las consultas sin apellido y nombre de pila (Nombre de pila desconocido, Apellido desconocido). Normalmente, este valor está establecido en false, ya que esta consulta podría colapsar fácilmente el sistema con demasiadas respuestas.
allowFnuInit	Controla si se permiten consultas con un único apellido inicial y un nombre de pila en blanco. Normalmente, este valor está establecido en false, ya que esta consulta podría colapsar fácilmente el sistema con demasiadas respuestas.

Tabla 11. Opciones comunes de Distributed Search (continuación)

Entrada	Descripción
allowInitLnu	Controla si se permiten consultas con un único nombre de pila inicial y un apellido en blanco. Normalmente, este valor está establecido en false, ya que esta consulta podría colapsar fácilmente el sistema con demasiadas respuestas.
allowInitInit	Controla si se permiten consultas con un único nombre de pila inicial y un único apellido inicial. Normalmente, este valor está establecido en false, ya que esta consulta podría colapsar fácilmente el sistema con demasiadas respuestas.

### Opciones de Distributed Search exclusivas de búsquedas individuales:

Varias opciones de configuración son exclusivas de búsquedas individuales. Cada cabecera del archivo de configuración tiene un número exclusivo al final (por ejemplo, search1). Este número indica el proceso de búsqueda al que se aplican las entradas de configuración. Debe tener una de estas secciones para cada búsqueda.

Las entradas siguientes son exclusivas de búsquedas individuales.

#### hostname

El nombre de host de la máquina que ejecuta el buscador.

**port** El número de puerto utilizado para comunicarse con el gestor de comunicaciones.

**ipv6** Una entrada opcional que indica si se debe utilizar el protocolo IPv6 para la comunicación con el gestor de comunicaciones. Si no está presente o si se ha inhabilitado, se utilizará el protocolo IPv4.

#### doAdds

¿Responderá este proceso de búsqueda a solicitudes de adición? Sólo una búsqueda debe tener esta entrada establecida en true.

#### nameFile

La ubicación de los datos de nombre preprocesados. En un plano ideal, debe producirla Name Preprocessor.

#### origDataFile

Salida del archivo de datos originales de Name Preprocessor asociada con nameFile.

#### origIndexFile

Salida del índice de datos originales de Name Preprocessor asociada con origDataFile.

#### logDebug, logError, logEvent, logMessage

Distributed Search tiene cuatro niveles de anotaciones que puede controlar con el archivo de configuración. Cada entrada reconoce lo siguiente:

- **filename** – nombre del archivo donde deben colocarse las entradas de anotaciones.
- **cout, cerr, clog** – las corrientes de salida estándar.
- **blanco** – no hay nada que indique a Distributed Search que esta anotación no es necesaria.

#### logDebug

Muestra mensajes de depuración durante el desarrollo. En producción, habrá muy poca salida en estas anotaciones.

**logError**

Muestra condiciones de error tales como entrada de usuario no válida, errores del sistema y problemas de comunicaciones.

**logEvent**

Anota eventos significativos tales como condiciones de inicio, transmisión y recepción de mensajes.

**logMessage**

Anota todos los mensajes entrantes y salientes. Genera gran cantidad de salida y debe inhabilitarse durante la operación habitual.

**Ejecutar Distributed Search**

Para poder ejecutar Distributed Search, los datos de nombre en los que opera se pueden procesar con el programa de utilidad Name Preprocessor. Como parte de ese proceso, el archivo original de datos de nombre puede dividirse en el número de partes deseado. Como alternativa, el archivo original de datos de nombre puede dividirse manualmente sin utilizar Name Preprocessor. Cada parte requiere una instancia dedicada del proceso de Distributed Search y un procesador dedicado para obtener un rendimiento y unos tiempos de respuesta óptimos.

**Antes de empezar**

Después de dividir el archivo de datos originales mediante Name o mediante la división manual, debe determinar si es o no necesario dar soporte a la adición de nombres a la lista de datos de nombre después de iniciar una sesión del servidor de Distributed Search. Si se ha realizado una división manual del archivo original de datos de nombre, puede añadir el nombre o nombres nuevos a una de las partes y luego reiniciar el servidor de Distributed Search. Generalmente, los nombres deben añadirse con un mensaje de adición (add) sin finalizar una sesión de Distributed Search. En tal caso, debe configurarse una instancia de la búsqueda para transacciones de adición y asignarse un procesador independiente para esta instancia de adición. A continuación, estas decisiones relativas al soporte de recursos y entorno de ejecución se registran como valores en el archivo de configuración de Distributed Search.

**Acerca de esta tarea**

Para ejecutar Distributed Search, siga estos pasos:

**Procedimiento**

1. Inicia cada instancia configurada de un proceso de Distributed Search especificando su ID de búsqueda y, opcionalmente, un archivo de configuración. El ID de búsqueda corresponde al número que figura en la cabecera del archivo de configuración de Distributed Search. Por ejemplo, 1 utiliza la configuración de [searcher1], por lo que la sección de la cabecera del archivo de configuración será searcher 1 ds.config.
2. Inicie el gestor de comunicaciones desde la línea de mandatos, especificando (opcionalmente) un archivo de configuración. Por ejemplo: commgr ds.config.

**Qué hacer a continuación**

Todos los procesos de búsqueda deben terminar la carga y la inicialización para que el gestor de comunicaciones (commgr) pueda iniciarse. Un proceso de búsqueda estará preparado cuando observe el mensaje siguiente en la consola:

en  
espera de conexión en el puerto <número de puerto>

Como en los procesos de búsqueda, el argumento de archivo de configuración es opcional y, si no se especifica, toma por omisión el valor ds.config del directorio actual. Si el gestor de comunicaciones no puede conectarse a uno o varios de los buscadores, envía continuamente un mensaje que se direcciona al dispositivo de consola en el formato siguiente:

```
06:25:53 no ha podido establecerse conexión con el servidor - localhost:2348
```

Commgr rechaza las solicitudes de conexión de usuario hasta que todos sus procesos de búsqueda asociados se han conectado satisfactoriamente. Una vez conectado a todos los procesos de búsqueda configurados, commgr emite un mensaje al dispositivo de consola con el formato siguiente:

```
06:25:54 conectado a todos los buscadores, en espera de solicitudes
```

Si se configura con opciones de anotaciones predeterminadas (por ejemplo anotar eventos y errores), un proceso de Distributed Search genera la salida siguiente en la consola durante la inicialización de la sesión:

```
>searcher 1
11:13:26
searcher1.hostname ----- localhost
searcher1.port ----- 2346
searcher1.doAdds ----- 0
searcher1.isUnique----- 1
searcher1.doRegularize----- 1
searcher1.doTransliterate---- 0
searcher1.nameFile ----- names1m.nh.txt.1
searcher1.numRecords----- 400000
searcher1.reportIncrement---- 100000
searcher1.defaultMaxResults-- 100
searcher1.ibmTaqFile----- taq.ibm
searcher1.ibmGnvFile----- gnv.ibm
searcher1.ibmSnvFile----- snv.ibm
searcher1.ibmBnvFile----- bnv.ibm
searcher1.custTaqFile-----
searcher1.custGnvFile-----
searcher1.custSnvFile-----
searcher1.arabicTransFile --- arabicTransRule.ibm
searcher1.cyrillicTransFile - cyrillicTransRule.ibm
searcher1.greekTransFile ---- greekTransRule.ibm
searcher1.latinTransFile ---- latinTransRule.ibm
searcher1.angloRegFile ----- angloRegRule.ibm
searcher1.arabicRegFile ----- swAsianRegRule.ibm
searcher1.germanRegFile -----
searcher1.indianRegFile -----
searcher1.russianRegFile ----
searcher1.thaiRegFile -----
searcher1.origDataFile----- names1m.orig.dat.1
searcher1.origIndexFile----- names1m.orig.idx.1
searcher1.logDebug -----
searcher1.logError ----- cout
searcher1.logEvent ----- cout
searcher1.logMessage -----
11:13:26 cargando taqs desde taq.ibm
11:13:26 cargando gnv desde gnv.ibm
11:13:30 cargando snv desde snv.ibm
11:13:33 cargando archivo trans árabe desde arabicTransRule.ibm
11:13:35 cargando archivo trans cirílico desde cyrillicTransRule.ibm
11:13:35 cargando archivo trans griego desde greekTransRule.ibm
11:13:35 cargando archivo trans latín desde latinTransRule.ibm
11:13:35 cargando archivo reg inglés desde angloRegRule.ibm
11:13:35 cargando archivo reg árabe desde swAsianRegRule.ibm
11:13:35 cargando nombres desde names1m.nh.txt.1
11:13:35 reservando espacio para 400000 registros
11:13:38 -- núm cargados = 100000
```

```

11:13:42 -- núm cargados = 200000
11:13:46 -- núm cargados = 300000
11:13:51 -- núm cargados = 400000
11:13:56 -- núm cargados = 500000
11:13:57 numLoaded      = 533025
11:13:57 cargando datos orig
núm leídos = 100000
núm leídos = 200000
núm leídos = 300000
núm leídos = 400000
núm leídos = 500000
11:13:58 esperando conexión en el puerto 2346
11:13:58 conectado, esperando datos...

```

El ejemplo siguiente muestra cómo podría ser la salida de la consola del inicio satisfactorio de una sesión de commgr:

```

>commgr
11:13:58
commgr.listenPort ----- 2345
commgr.sleepMsec ----- 10
commgr.logDebug -----
commgr.logError ----- cout
commgr.logEvent ----- cout
commgr.logMessage -----
commgr.numSearchers ----- 3
11:13:58 a la escucha en el puerto 2345
11:13:58 conectado a buscador localhost:2346
11:14:14 no se ha podido establecer conexión con el servidor - localhost:2347
11:14:15 no se ha podido establecer conexión con el servidor - localhost:2347
11:14:15 conectado al buscador localhost:2347
11:14:15 conectado al buscador localhost:2348
11:14:15 esperando solicitudes...

```

## Opciones de configuración y consideraciones acerca de Distributed Search

Distributed Search está diseñada para funcionar en una amplia variedad de entornos operativos.

Distributed Search puede operar con la misma facilidad tanto en una máquina de un solo procesador como en un servidor dedicado con 32 ó 64 procesadores, debido a su sistema de mensajería simplificado (conexión TCP/IP) y a la capacidad de distribuir una búsqueda por todos los procesadores disponibles. La decisión de eliminar nombres duplicados o dejarlos en los datos de nombre originales tiene importantes implicaciones para el rendimiento y la calidad de los resultados de búsqueda de Distributed Search. No existe ningún conjunto de procedimientos recomendados para tales decisiones, ya que el mejor rendimiento depende de la naturaleza y características de los propios registros de datos de nombre originales.

Cuando los registros de datos de nombre originales se procesan manualmente o con el programa de utilidad NamePreprocessor para eliminar duplicaciones, cada solicitud de búsqueda opera sobre un número reducido de nombres candidatos con las correspondientes mejoras en los tiempos de respuesta de búsqueda y en niveles de rendimiento de transacción. NamePreprocessor puede reducir una lista de nombres que contenga 100 millones de nombres a aproximadamente 33 millones de nombres limpiando los datos y eliminando duplicaciones.

El grado de reducción obtenido por el proceso de limpieza y eliminación de duplicaciones varía según el conjunto de datos de nombre. Algunas listas de datos tienen un mayor volumen de nombres duplicados, mientras que otras tienden a contener nombres exclusivos. Las palabras superfluas, las incoherencias de formato

y otros factores aleatorios o basados en el entorno cultural también pueden influir en el tamaño final de la lista de nombres limpia y sin duplicaciones.

La reducción del tiempo de búsqueda puede verse modificada por la necesidad de que cada proceso de búsqueda vaya seguido de algún número de transacciones de solicitud de datos originales, que intercambian el ID de nombre exclusivo generado internamente que se asigna durante el proceso de eliminación de duplicaciones por un conjunto de uno o varios ID de nombre real de los datos de nombre originales. Sin embargo, el tiempo ahorrado en la búsqueda en un número muy reducido de nombres contrarresta la sobrecarga impuesta por las transacciones de solicitud de datos originales. Si esta sobrecarga o el tiempo de preproceso necesario para identificar y eliminar nombres duplicados sobrepasa los límites aceptables, Distributed Search puede seguir operando con los datos de nombre originales, que se dividen manualmente o mediante NamePreprocessor en segmentos más pequeños que permiten aumentar el proceso paralelo y utilizar más recursos de procesador para cada solicitud de búsqueda.

Los registros de nombres preprocesados también pueden mejorar en gran medida los resultados de búsqueda de nombres comunes. Cuando una solicitud de búsqueda se refiere a un nombre común, es posible que coincidencias de nombre válidas y potencialmente útiles con ligeras variaciones de deletreo sean eliminadas o "bloqueadas" por las coincidencias de nombre exactas. Debe tener en cuenta este resultado cuando Distributed Search realice búsquedas en una lista de datos de nombre de gran tamaño, para asegurarse de que no se eliminen de los resultados de búsqueda coincidencias de nombre significativas.

## Interfaz XML de NameHunter Distributed Search

Distributed Search utiliza una interfaz XML para manipular y buscar una lista de datos de nombre. Todos los mensajes constan de una solicitud de un cliente y una respuesta de Distributed Search.

Los ejemplos siguientes de solicitud y respuesta XML incluyen ejemplos de códigos y respuestas asociadas con descripciones de elementos clave. Algunos de los elementos son meramente informativos y no se describen. Por ejemplo, la información de versión se coloca en todas las cabeceras de mensaje:

```
<NHServerMessage protocol_version="4.1">
```

Con la excepción de la respuesta de resultados de búsqueda, las respuestas de Distributed Search son simples acuses de recibo de que la solicitud se ha procesado satisfactoriamente.

Todas las solicitudes entre Distributed Search y las aplicaciones cliente deben terminar en nulo (NULL).

### Solicitudes XML de Distributed Search

Las solicitudes XML de Distributed Search permiten manipular y realizar búsquedas en una lista de datos de nombre. Todas las solicitudes reciben una respuesta de Distributed Search que describe el resultado de la solicitud.

#### Solicitud de adición:

Con la solicitud de adición, puede solicitar a Distributed Search que añada un nombre al conjunto de nombres residentes en memoria en los que pueden realizarse búsquedas. Si es satisfactoria, se devuelve un mensaje de respuesta con

el valor de ID de solicitud suministrado en el mensaje de la solicitud de adición asociado, para que un proceso de cliente pueda emparejar las solicitudes y las respuestas.

El campo NEC (Named Entity Category) determina el tipo de nombre y los campos que deben utilizarse al actualizar una lista de búsqueda para Distributed Search:

- O – los nombres de organización utilizan el campo NAME
- P – los nombres de persona utilizan los campos SN y GN
- A – se utilizan los tres campos y se añaden dos nombres

Ejemplo de solicitud de adición

```
< NHServerMessage protocol_version="4.1">
  <BASIC_REQUEST_INFO request_type="A" request_id="-1"/>
  <NAME_TO_ADD
    SN="Acme Light Industries, Inc."
    GN=""
    NAME="Acme Light Industries, Inc."
    NEC="A"
    NAME_ID="123456"
    CULTURE_SN="4"
    CULTURE_GN="4"
    ALT_PARSE="N" />
  <DATA_LIST_NAME value="" />
</ NHServerMessage>
```

#### BASIC\_REQUEST\_INFO

Atributo	Obligatorio	Límites
request_type	Sí	Debe ser 'A'
request_id	No	Un número que Distributed Search devuelve en la respuesta. Utilice -1 para indicar a Distributed Search que debe generar un número.

#### NAME\_TO\_ADD

Atributo	Obligatorio	Límites
SN	Sí <sup>1</sup>	Apellido, 128 caracteres como máximo.
GN	Sí <sup>1</sup>	Nombre de pila, 128 caracteres como máximo.
NAME	Sí <sup>2</sup>	Nombre, 128 caracteres como máximo. Se utiliza para contener nombres sin campos, como por ejemplo nombres de organización.
NEC	Sí	NEC (Named Entity Category). Los valores válidos son: O = Organización P = Personal A = Todos

Atributo	Obligatorio	Límites
NAME_ID	Sí	ID del nombre que debe añadirse, de 256 caracteres alfanuméricos como máximo. No es necesario que el ID sea exclusivo.
CULTURE_SN	No	Código de entorno cultural del apellido.
CULTURE_GN	No	Código de entorno cultural del nombre de pila.
ALT_PARSE	No	Elemento alternativo de un nombre (Y o N)

<sup>1</sup>Sólo es obligatorio si NEC es P o A.

<sup>2</sup>Sólo es obligatorio si NEC es O o A.

#### DATA\_LIST\_NAME

Atributo	Obligatorio	Límites
value	Sí	No implementado actualmente en el Servidor distribuido.

#### Solicitud de supresión:

Con la solicitud de supresión, puede suprimir uno o varios nombres de Distributed Server. Esta solicitud suprime todos los registros de Distributed Server que tengan el ID de nombre (Name\_ID) suministrado por el cliente especificado. Si es satisfactoria, se devolverá un mensaje de respuesta satisfactoria con el ID de solicitud original del remitente.

#### Ejemplo de solicitud de supresión

```
<NHServerMessage protocol_version="4.2">
  <BASIC_REQUEST_INFO request_type="D" request_id="-1"/>
  <RECORD_ID_TO_DELETE value="123456"/>
  <DATA_LIST_NAME value=""/>
</NHServerMessage>
```

#### BASIC\_REQUEST\_INFO

Atributo	Obligatorio	Límites
request_type	Sí	Debe ser D.
request_id	No	Un número que Distributed Search devuelve en la respuesta. Utilice -1 para indicar a Distributed Search que debe generar un número.

#### RECORD\_ID\_TO\_DELETE

Atributo	Obligatorio	Límites
value	Sí	ID de nombre (Name_ID) suministrado por cliente que debe suprimirse. Se suprimirán todos los nombres con este ID de nombre.

#### DATA\_LIST\_NAME

Atributo	Obligatorio	Límites
value	No	No implementado actualmente en Distributed Search.

#### Solicitud de datos originales:

Con la solicitud de datos originales, puede solicitar a Distributed Search que envíe todos los registros de datos originales correspondientes a los nombres asociados con esta coincidencia. Esta solicitud sólo es necesaria al utilizar Name Preprocessor para eliminar las duplicaciones de los datos de nombre. Si es satisfactoria, se devolverá un mensaje de resultados de búsqueda con el ID de solicitud original del remitente.

#### Ejemplo de solicitud de datos originales

```
<NHServerMessage protocol_version="4.2">
  <BASIC_REQUEST_INFO request_type="G" request_id="-1"/>
  <DATA_LIST_NAME value=""/>
  <NAME_ID value="123456"/>
</NHServerMessage>
```

#### BASIC\_REQUEST\_INFO

Atributo	Obligatorio	Límites
request_type	Sí	Debe ser G.
request_id	No	Un número que Distributed Search devuelve en la respuesta. Utilice -1 para indicar a Distributed Search que debe generar un número.

#### DATA\_LIST\_NAME

Atributo	Obligatorio	Límites
value	Sí	No implementado actualmente en el Servidor distribuido.

#### NAME\_ID

Atributo	Obligatorio	Límites
value	Sí	ID de NameHunter del nombre para el que deben recuperarse datos originales. Este valor es el ID de nombre (NAME_ID) devuelto en el mensaje de resultados de búsqueda si Name Preprocessor ha eliminado las duplicaciones de la lista de nombres original.

### Solicitud de búsqueda:

Con la solicitud de búsqueda, puede solicitar a Distributed Search que busque un nombre. Si es satisfactoria, se devolverá una respuesta de resultados de búsqueda con el ID de solicitud original del remitente.

Opcionalmente, la solicitud de búsqueda también puede especificar uno o varios valores de parámetro que deben especificarse para la solicitud realizada. Puede especificar cualquiera de los valores, o todos ellos, en cada uno de los siguientes grupos de parámetros (GENERAL,\_PARMS, COMP\_PARMS\_GN, COMP\_PARMS\_SN y COMP\_PARMS\_ON) en la solicitud de búsqueda. Las alteraciones temporales y ajustes del parámetro de comparación (CompParm) se aplican a los campos GnCulture, SnCulture u OnCulture del nombre de la búsqueda. Si no se especifica ningún entorno cultural, las alteraciones temporales y ajustes se aplican al entorno cultural predeterminado (CultureUnknown).

El campo NEC (Named Entity Category) determina el tipo de nombre de consulta y los campos que deben utilizarse al ejecutar una búsqueda:

- O – los nombres de organización utilizan el campo NAME
- P – los nombres de persona utilizan los campos SN y GN

El NEC predeterminado para una solicitud de búsqueda es *P* (nombre de persona). El NEC *A* (todos) no está permitido para solicitudes de búsqueda e IBM NameWorks lo correlaciona con *P*.

El ejemplo siguiente muestra una búsqueda típica que utiliza un nombre de organización sin parámetros.

Ejemplo de solicitud de búsqueda (sin especificar parámetros)

```
<NHServerMessage protocol_version="4.1">
  <BASIC_REQUEST_INFO request_type="S" request_id="-1"/>
  <DATA_LIST_NAME value="data_list_1"/>
  <SEARCH_NAME NAME="Acme Light Industries, Inc."
    NEC="0"
    Srch Opt=3/>
</NHServerMessage>
```

La solicitud de búsqueda que sigue contiene factores de ajuste para un nombre de organización. Los factores de ajuste se especifican añadiendo *\_ADJ* a un parámetro y especificando un valor válido para el ajuste.

Ejemplo de solicitud de búsqueda (especificando parámetros)

```

<NH_SERVER_MESSAGE protocol_version="false">
  <BASIC_REQUEST_INFO request_type='S' request_id='-1' />
  <DATA_LIST_NAME value='' />
  <SEARCH_NAME NAME="Kidder Byron Licensed Land Surveyor"
    NEC="0" CULTURE_SN="" CULTURE_GN=""
    SRCH_OPT="2" />
  <GENERAL_PARMS SHOULD_USE_INDEX="Y" NAME_THRESH="0.75"
    MAX_RETURN_NAMES="50" RETURN_ORIG_NAMES="N" />
  <COMP_PARMS_ON INITIAL_TOKEN_SCORE_ADJ="0.98"
    INITIAL_INITIAL_SCORE_ADJ="1.0"
    NAME_UNKNOWN_SCORE_ADJ="0.98"
    NO_NAME_SCORE_ADJ="0.97"
    ANCHOR_FACTOR_ADJ="0.98"
    OOPS_FACTOR_ADJ="1.05"
    COMPRESSED_SCORE_MAX_ADJ="1.0"
    FIELD_THRESH_ADJ="0.97"
    FIELD_WEIGHT_ADJ="0.98"
    MISSING_STEM_FACTOR_ADJ="0.98"
    FIELD_VARIANT_SCORE_ADJ="1.05" />

</NH_SERVER_MESSAGE>

```

La solicitud de búsqueda que sigue contiene valores específicos y factores de ajuste para un nombre de persona.

Ejemplo de solicitud de búsqueda (especificando parámetros)

```

<NHServerMessage protocol_version="4.1">
  <BASIC_REQUEST_INFO request_type="S" request_id="-1" />
  <DATA_LIST_NAME value="data_list_1" />
  <SEARCH_NAME SN="Freeman"
    GN="Harlow J"
    CULTURE_SN="1"
    CULTURE_GN="1" />
  <GENERAL_PARMS
    NAME_THRESH="0.600000"
    MAX_RETURN_NAMES="15"
    RETURN_ORIG_NAMES="N" />
  <COMP_PARMS_GN
    SCORE_MODE="1"
    OOPS_FACTOR="0.600000"
    NO_NAME_SCORE="0.750000"
    NAME_UNKNOWN_SCORE="0.750000"
    MISSING_TAQ_FACTOR="0.970000"
    MISSING_STEM_FACTOR="0.950000"
    MATCH_VARIANTS="Y"
    MATCH_INITIALS="Y"
    LEFT_BIAS="N"
    INITIAL_TOKEN_SCORE="0.800000"
    INITIAL_INITIAL_SCORE="0.900000"
    FIELD_WEIGHT="0.800000"
    FIELD_THRESH="0.500000"
    DO_COMPRESSED_SCORE="Y"
    COMPRESSED_SCORE_MAX="0.950000"
    DIF_TAQ_FACTOR="0.990000"
    CULTURE="4"
    ANCHOR_TYPE="1"
    ANCHOR_FACTOR="0.600000"
    MATCH_FIELD_VARIANTS="Y"
    FIELD_VARIANT_SCORE="0.95" />
  <COMP_PARMS_SN
    SCORE_MODE="0"
    OOPS_FACTOR="0.750000"
    NO_NAME_SCORE="0.750000"
    NAME_UNKNOWN_SCORE="0.700000"
    MISSING_TAQ_FACTOR="0.980000"
    MISSING_STEM_FACTOR="0.960000"

```

```

MATCH_VARIANTS="Y"
MATCH_INITIALS="Y"
LEFT_BIAS="N"
INITIAL_TOKEN_SCORE="0.850000"
INITIAL_INITIAL_SCORE="0.900000"
FIELD_WEIGHT="1.000000"
FIELD_THRESH="0.550000"
DO_COMPRESSED_SCORE="Y"
COMPRESSED_SCORE_MAX="0.950000"
DIF_TAQ_FACTOR="0.990000"
CULTURE="4"
ANCHOR_TYPE="1"
ANCHOR_FACTOR="0.800000"
MATCH_FIELD_VARIANTS="Y"
FIELD_VARIANT_SCORE="0.95"/>

```

</NHServerMessage>

#### BASIC\_REQUEST\_INFO

Atributo	Obligatorio	Límites
request_type	Sí	Debe ser S.
request_id	No	Un número que Distributed Search devuelve en la respuesta. Utilice -1 para indicar a Distributed Search que debe generar un número.

#### DATA\_LIST\_NAME

Atributo	Obligatorio	Límites
value	Sí	No implementado actualmente en Distributed Search.

#### SEARCH\_NAME

Atributo	Obligatorio	Límites
SN	Sí <sup>1</sup>	Apellido, 128 caracteres como máximo.
GN	Sí <sup>1</sup>	Nombre de pila, 128 caracteres como máximo.
NAME	Sí <sup>2</sup>	Nombre, 128 caracteres como máximo. Se utiliza para contener nombres sin campos, como por ejemplo nombres de organización.
NEC	Sí	NEC (Named Entity Category). Los valores válidos son: O = Organización P = Personal
CULTURE_GN	No	Código de entorno cultural del nombre de pila. Si no se suministra el valor de este campo, se utilizan los parámetros actuales.

Atributo	Obligatorio	Límites
CULTURE_SN	No	Código de entorno cultural del apellido. Si no se suministra el valor de este campo, se utilizan los parámetros actuales.
SRCH_OPT	No	1 = buscar sólo en la lista de nombres de persona 2 = buscar sólo en la lista de organizaciones 3 = buscar en las listas de nombres de persona y de organización

<sup>1</sup>Sólo es necesario si NEC es P.

<sup>2</sup>Sólo es necesario si NEC es O.

### Solicitud de conclusión:

Con la solicitud de conclusión, puede iniciar una conclusión controlada de todos los buscadores y del gestor de comunicaciones (commgr) de Distributed Search. Las transacciones en cola finalizan antes de que se realice la conclusión.

Una solicitud válida contiene una contraseña sensible a mayúsculas y minúsculas que coincide con la contraseña especificada en el archivo de configuración de Distributed Search. Justo antes de la conclusión de commgr, se devolverá al proceso cliente solicitante una respuesta satisfactoria que contendrá el ID de solicitud original del remitente. En un despliegue operativo típico de Distributed Search, la utilización de este mensaje está limitada al personal administrativo y de soporte.

Ejemplo de solicitud de conclusión

```
<NHServerMessage protocol_version="4.2">
  <BASIC_REQUEST_INFO request_type="X" request_id="-1"/>
  <SHUTDOWN_PASSWORD value="NHSERVER"/>
</NHServerMessage>
```

### BASIC\_REQUEST\_INFO

Atributo	Obligatorio	Límites
request_type	Sí	Debe ser X.
request_id	No	Un número que Distributed Search devuelve en la respuesta. Utilice -1 para indicar a Distributed Search que debe generar un número.

### SHUTDOWN\_PASSWORD

Atributo	Obligatorio	Límites
value	No	No se utiliza.

### Solicitud de estado:

Con la solicitud de estado, puede solicitar a Distributed Search que notifique el estado actual de su proceso y cola. Distributed Search responde con una respuesta de estado que contiene el ID de solicitud original del remitente.

Ejemplo de solicitud de estado

```
<NHServerMessage protocol_version="4.2">
  <BASIC_REQUEST_INFO request_type="T" request_id="-1"/>
</NHServerMessage>
```

### BASIC\_REQUEST\_INFO

Atributo	Obligatorio	Límites
request_type	Sí	Debe ser T.
request_id	No	Un número que Distributed Search devuelve en la respuesta. Utilice -1 para indicar a Distributed Search que debe generar un número.

### Solicitud de actualización:

Con la solicitud de actualización, puede actualizar un nombre que ya se encuentre en una lista de nombres de Distributed Search residente en la memoria (lista de datos) durante una sesión de Distributed Search. Si es satisfactoria, se devolverá un mensaje de respuesta satisfactoria que contendrá el ID de solicitud original del remitente.

Esta solicitud modifica todos los registros del sistema cuyo campo NAME\_ID coincide con el valor especificado en el atributo RECORD\_ID\_TO\_UPDATE. La solicitud de actualización puede ser lenta, ya que realiza una solicitud de supresión y luego una solicitud de adición. Pueden resultar afectados varios registros si comparten el mismo NAME\_ID.

El campo NEC (Named Entity Category) determina el tipo de nombre y los campos que deben utilizarse al actualizar una lista de búsqueda para Distributed Search:

- O – los nombres de organización utilizan el campo NAME
- P – los nombres de persona utilizan los campos SN y GN
- A – todos, indicando que se utilizan los tres campos y se añaden dos nombres

Ejemplo de solicitud de actualización

```
< NHServerMessage protocol_version="4.1">
  <BASIC_REQUEST_INFO request_type="U" request_id="-1"/>
  <UPDATED_NAME
    SN="Acme Light Industries Inc."
    GN=""
    NAME="Acme Light Industries Inc."
  >
  NEC="A"
  NAME_ID="123456"
  CULTURE_SN="1"
  CULTURE_GN="1"
```

```

    ALT_PARSE="N" />
    <RECORD_ID_TO_UPDATE value="123456"/>
    <DATA_LIST_NAME value="" />
  </ NHServerMessage>

```

#### BASIC\_REQUEST\_INFO

Atributo	Obligatorio	Límites
request_type	Sí	Debe ser U.
request_id	No	Un número que Distributed Search devuelve en la respuesta. Utilice -1 para indicar a Distributed Search que debe generar un número.

#### UPDATED\_NAME

Atributo	Obligatorio	Límites
SN	Sí <sup>1</sup>	Apellido, 128 caracteres como máximo.
GN	Sí <sup>1</sup>	Nombre de pila, 128 caracteres como máximo.
NAME	Sí <sup>2</sup>	Nombre, 128 caracteres como máximo. Se utiliza para contener nombres sin campos, como por ejemplo nombres de organización.
NEC	Sí	NEC (Named Entity Category). Los valores válidos son: O = Organización P = Personal A = Todos
NAME_ID	Sí	ID del nombre que debe actualizarse, de 256 caracteres alfanuméricos como máximo. No es necesario que el ID sea exclusivo.
CULTURE_SN	No	Código de entorno cultural del apellido.
CULTURE_GN	No	Código de entorno cultural del nombre de pila.
ALT_PARSE	No	¿Es este un elemento alternativo de un nombre? (Y o N)

<sup>1</sup>Sólo es necesario si NEC es P.

<sup>2</sup>Sólo es necesario si NEC es O.

#### RECORD\_ID\_TO\_UPDATE

Atributo	Obligatorio	Límites
value	Sí	Nombre que debe actualizarse, de 256 caracteres como máximo. NAME_ID debe coincidir con este campo.

#### DATA\_LIST\_NAME

Atributo	Obligatorio	Límites
value	No	No implementado actualmente en Distributed Search.

### Respuestas XML de Distributed Search

Después de emitir una solicitud XML de Distributed Search y de que ésta se haya procesado, recibirá un mensaje de respuesta que indica si la solicitud ha sido satisfactoria o ha fallado. También puede recibir una respuesta Status después de emitir la solicitud Status.

#### Respuesta de error:

Se obtiene una respuesta de error si Distributed Search encuentra un error al procesar una solicitud. La respuesta indica el ID de solicitud original del remitente para el que se ha producido el error. En otras secciones se describen los códigos de error y mensajes específicos.

#### Ejemplo de respuesta de error

```
<NHServerMessage server_version="4.2" >
  <REQUEST_ID value="4070517"/>
  <ERROR supplied_data_list_name=""
    severity=""
    error_msg="El buscador 3 no responde"
    error_code="GNRDS-123"/>
</NHServerMessage>
```

#### REQUEST\_ID

Atributo	Obligatorio	Límites
value	N/A	ID de la solicitud que ha provocado el error.

#### ERROR

Atributo	Obligatorio	Límites
severity	N/A	No implementado actualmente en Distributed Search.
error_msg	N/A	Descripción del error.
error_code	N/A	Código exclusivo del mensaje de error.

## Respuesta de resultados de búsqueda:

Una respuesta de resultados de búsqueda se devuelve una vez que el usuario ha emitido una solicitud de búsqueda o una solicitud de datos originales. La respuesta indica la solicitud suministrando el ID de solicitud original del remitente. Ambas solicitudes utilizan la mayoría de los campos de esta respuesta, pero no todos.

El ejemplo siguiente muestra una respuesta de resultados de búsqueda que contiene dos nombres coincidentes.

### Ejemplo de respuesta de resultados de búsqueda

```
<NHServerMessage protocol_version="4.1">
  <SEARCH_RESULTS>
    <NAME
      SN="FREEMAN"
      GN="HARLOW J"
      NAME=""
      NEC="P"
      NAME_ID="123456"
      SN_SCORE="1.000000"
      GN_SCORE="1.000000"
      FULL_NAME_SCORE="1.000000"
      NAME_CNT="1"
      IS_REG="N"
      ALT_PARSE="N"/>
    <NAME
      SN=""
      GN=""
      NAME="Acme Light Industries Inc."
      NEC="0"
      NAME_ID="73"
      SN_SCORE="0.7700"
      GN_SCORE="0.85000"
      FULL_NAME_SCORE="0.82000"
      NAME_CNT="5"
      IS_REG="N"
      ALT_PARSE="N"/>
  </SEARCH_RESULTS>
  <REQUEST_ID value="9070517"/>
</NHServerMessage>
```

### REQUEST\_ID

Atributo	Obligatorio	Límites
value	N/A	ID de la solicitud de búsqueda.

### SEARCH\_RESULTS\_NAME

Atributo	Obligatorio	Límites
SN	N/A	Apellido, 60 caracteres como máximo.
GN	N/A	Nombre de pila, 60 caracteres como máximo.

Atributo	Obligatorio	Límites
NAME	No	Nombre, 128 caracteres como máximo. Se utiliza para contener nombres sin campos, como por ejemplo nombres de organización.
NEC	No	NEC (Named Entity Category). Los valores válidos son: O = Organización P = Personal
NAME_ID	N/A	Para nombres exclusivos, este valor es el NH_ID; para nombres originales, este valor es el ID de cliente. NAME_ID puede tener 256 como máximo.
SN_SCORE	N/A	Puntuación del apellido, 0 – 1.00
GN_SCORE	N/A	Puntuación del nombre de pila, 0 – 1.00
FULL_NAME_SCORE	N/A	Puntuación del nombre completo, 0 – 1.00
NAME_CNT	N/A	El número de nombres originales a los que se aplica este nombre exclusivo. Establézcalo en 1 cuando se devuelva el nombre original.
ALT_PARSE	N/A	¿Es este el resultado de un elemento de nombre alternativo? Sólo se utiliza para nombres originales.
IS_REG	N/A	¿Es este el resultado de una regularización de nombre? Sólo se utiliza para nombres originales.

### Respuesta de estado:

La respuesta de estado devuelve el estado actual del proceso y la cola como respuesta a una solicitud de estado y contiene el ID de la solicitud de estado asociada.

### Ejemplo de respuesta de estado

```
<NHServerMessage server_version="4.2" >
  <REQUEST_ID value="4070517"/>
  <STATUS_MESSAGE
    value="Servidor en ejecución con 0 búsquedas en ejecución
    y 0 en cola\"/>
</NHServerMessage>
```

REQUEST\_ID

Atributo	Obligatorio	Límites
value	N/A	ID de la solicitud de estado asociada.

#### STATUS\_MESSAGE

Atributo	Obligatorio	Límites
value	N/A	Texto que describe el estado del Servidor distribuido.

#### Respuesta satisfactoria:

La respuesta satisfactoria es la confirmación utilizada por Distributed Search para indicar que un mensaje se ha recibido y procesado correctamente. La respuesta indica la transacción que ha sido satisfactoria suministrando el ID de solicitud original del remitente.

#### Ejemplo de respuesta satisfactoria

```
<NHServerMessage protocol_version="4.2">
  <REQUEST_ID value="1070555"/>
</NHServerMessage>
```

#### REQUEST\_ID

Atributo	Obligatorio	Límites
value	N/A	ID del mensaje de solicitud originador.

## Búsqueda de nombres con Enterprise Name Search

Enterprise Name Search (ENS) ofrece una infraestructura para distribuir búsquedas de nombres de empresas de gran volumen y a gran escala en listas de nombres con muchos datos. ENS saca el máximo rendimiento de IBM NameWorks al realizar funciones de búsqueda de nombres de manera eficaz, gestionar listas de nombres y configurar, gestionar y supervisar el proceso de búsquedas de nombres. ENS proporciona opciones adicionales para el escalado horizontal en función de las necesidades de rendimiento, despliegue con alta disponibilidad y migración tras error, e interfaces de SOAP y REST para la integración de clientes con software de terceros y sistemas heredados existentes.

Puede utilizar las prestaciones de búsqueda de ENS de su propio programa cliente mediante la API de servicio web o puede utilizar la interfaz gráfica de usuario que se proporciona en ENS. Incluso si intenta utilizar los servicios web, la GUI puede resultarle útil al realizar búsquedas y a la hora de entender las prestaciones y el comportamiento de ENS, puesto que utiliza los servicios web para las búsquedas.

## Gestión de la seguridad del usuario de Enterprise Name Search

ENS incorpora el servidor de aplicaciones WebSphere Liberty para alojar sus componentes web y se basa en la implementación de seguridad de WebSphere Liberty para las funciones de autenticación y autorización.

De forma predeterminada, WebSphere Liberty utiliza un simple almacén basado en archivos para obtener información sobre los usuarios, los grupos y las contraseñas. En ENS, el archivo que contiene esta información se halla en *ensroot/ibm-home/wlp/users.xml*. Puede añadir, eliminar y gestionar usuarios editando este archivo. Estas acciones se pueden llevar a cabo con el sistema en ejecución.

Los clientes con requisitos de seguridad más complejos pueden cambiar la configuración de WebSphere Liberty después de la instalación de ENS para utilizar un registro de usuarios de LDAP en lugar del simple almacén basado en archivos. Este tema está fuera del ámbito de la documentación de GNM pero se describe en la documentación sobre seguridad de WebSphere Liberty.

Adicionalmente, los usuarios pueden cambiar sus propias contraseñas de ENS en una pantalla de la GUI de cambio de contraseña.

## Sintaxis de la línea de mandatos del programa de utilidad **wpswd**

El programa de utilidad *wpswd* es un programa de utilidad de línea de mandatos que puede utilizar para gestionar el archivo de contraseñas *ens.passwd*. El archivo de contraseñas contiene la lista de nombres de usuario, contraseñas y grupos del rol de seguridad asignados a cada usuario de Enterprise Name Search.

Para utilizar el programa de utilidad *wpswd*, vaya al directorio *<install>/bin* y especifique los mandatos en la línea de mandatos. La sintaxis del programa de utilidad *wpswd* es la siguiente:

- Para Linux:  
`./wpswd ubicación_vía_acceso_opcional_del_archivo_contraseñas opciones`
- Para Windows:  
`wpswd.bat  
ubicación_vía_acceso_opcional_del_archivo_contraseñas  
opciones`

Si no se especifica una vía de acceso al archivo de contraseñas, se tomará el valor predeterminado del archivo *ens.passwd* ubicado en el directorio *<install>/ibm-home/ens*.

## Opciones de mandato

- c *nombre\_usuario*  
Cree un usuario nuevo con el nombre de usuario especificado y genere una contraseña para el usuario.
- c *nombre\_usuario contraseña*  
Cree un usuario nuevo con el nombre de usuario y contraseña especificados.
- d *nombre\_usuario*  
Suprima un usuario nuevo con el nombre de usuario y contraseña especificados. Al suprimir un usuario, se eliminan el nombre de usuario y la contraseña, y se elimina el usuario de todos los grupos del rol de seguridad al que pertenece.
- l  
Lista todos los usuarios actuales de Enterprise Name Search por nombre de usuario.
- p *nombre\_usuario*  
Genere una nueva contraseña para el usuario especificado.

- p *nombre\_usuario contraseña*  
Cambie la contraseña al valor especificado para el usuario especificado.
- a *nombre\_usuario grp1 grp2 ...*  
Asigne el usuario a grupos de rol de seguridad. Los grupos del rol de seguridad determinan a qué componentes ENS pueden acceder los usuarios y los tipos de funciones que pueden realizar.
- r *nombre\_usuario grp1 grp2 ...*  
Vuelva a establecer los grupos del rol de seguridad asignados al usuario especificado.
- u *nombre\_usuario grp1 grp2 ...*  
Elimine el usuario de uno o más grupos de roles de seguridad.

**Nota:** Los usuarios, contraseñas y roles debe utilizar caracteres a-Z (mayúsculas y minúsculas) y 0-9.

## Ejemplo

Para crear un usuario, especifique una contraseña y asigne un grupo de rol de seguridad al nuevo usuario en un entorno Windows:

1. Cree el usuario con un nombre de usuario de "ConsoleUser1" y asigne una contraseña de "ens1234":

```
wpswd.bat -c ConsoleUser1 ens1234
```

2. Asigne el rol de seguridad del grupo de "admin" al usuario:

```
wpswd.bat -a ConsoleUser1 admin
```

## Creación de usuarios y asignación de grupos de seguridad

Los usuarios y los permisos se definen y gestionan añadiendo y modificando los elementos <usuario> y <miembro> en el archivo *raíz\_ens/ibm-home/wlp/users.xml*.

### Acerca de esta tarea

El archivo *users.xml* tiene una sección cerca de la parte superior del comentario "Lista de usuarios aquí":

```
<!-- Lista de usuarios aquí -->
<user name="ensadmin" password="{aes}AICwWs/XAgx70bztj0s09zks31vubq0rERfNM/o81K"/>
```

Los usuarios se crean añadiendo una línea en esta sección. No quite ni modifique las dos entradas que se han creado aquí durante la instalación. Cuando añada un usuario nuevo, otorgue una contraseña inicial con texto sin formato similar al siguiente:

```
<user name="smith" password="changeMe"/>
```

El nombre de usuario y la contraseña sólo deben contener letras (a-z y A-Z) y dígitos (0-9). Tras añadir una entrada de usuario, el usuario debería cambiar su contraseña igual que en "Cambio o restablecimiento de las contraseñas de usuario".  
[Enlace a este tema]

Para definir la pertenencia de un usuario a un grupo, además de añadir el elemento <usuario>, tendrá que obtener una lista del usuario en un grupo o en más grupos para especificar qué operaciones puede llevar a cabo el usuario nuevo. ENS permite que los usuarios estén en los grupos siguientes:

Tabla 12.

Grupo	Derechos
admins	<ul style="list-style-type: none"> <li>• Utilizar todas las operaciones de la consola para ver y cambiar el estado del sistema.</li> <li>• Buscar y gestionar cualquier lista de nombres.</li> <li>• Utilizar el panel de instrumentos.</li> </ul>
console_users	<ul style="list-style-type: none"> <li>• Utilizar todas las operaciones de la consola para ver y cambiar el estado del sistema.</li> <li>• Utilizar el panel de instrumentos.</li> </ul>
dashboard_users	<ul style="list-style-type: none"> <li>• Utilizar el panel de instrumentos para ver el estado del sistema.</li> </ul>
managers_all_lists	<ul style="list-style-type: none"> <li>• Buscar y gestionar (añadir y suprimir desde) cualquier lista de nombres.</li> </ul>
managersNNN* (por ejemplo, managers101)	<ul style="list-style-type: none"> <li>• Buscar y gestionar listas de nombres para las que se ha especificado este rol de gestor a la hora de crear la lista en NameLoader.</li> </ul>
searchers_all_lists	<ul style="list-style-type: none"> <li>• Buscar cualquier lista de nombres.</li> </ul>
searchersNNN * (por ejemplo, searchers101)	<ul style="list-style-type: none"> <li>• Buscar listas de nombres para las que se ha especificado este rol de buscador a la hora de crear la lista en NameLoader.</li> <li>• Se utiliza para un control preciso.</li> </ul>

La pertenencia al grupo "admins" proporciona acceso a todas las operaciones de ENS. No es necesario añadir un usuario de este grupo en ningún otro grupo.

La pertenencia al grupo "managers\_all\_lists" permite que el usuario busque y gestione las listas de nombres. No es necesario añadir un usuario de este grupo en ningún otro grupo de buscadores o gestores.

La pertenencia al grupo "searchers\_all\_lists" permite que un usuario realice búsquedas en todas las listas de nombres. No es necesario añadir un usuario de este grupo en ningún otro grupo de buscadores.

Los grupos "searchersNNN" y "managersNNN" proporcionan control preciso de acceso a listas concretas. La mayor parte de instalaciones de ENS no utilizan esta función. Para poner un usuario en un grupo, busque el elemento del <grupo> correspondiente en users.xml y añada un elemento <miembro> para el usuario. Por ejemplo, podemos hacer que un usuario existente, "smith", sea miembro del grupo console\_users de una forma similar a la siguiente:

```
<group name="console_users">
<member name="smith"/>
</group>
```

Puede eliminar "smith" del grupo console\_users eliminando el elemento <member name="smith">.

### **Cambio o restablecimiento de las contraseñas de usuario**

Las contraseñas de usuario para acceder a Enterprise Name Search se gestionan a través del archivo users.xml.

## Acerca de esta tarea

Cuando un administrador crea un usuario nuevo añadiendo una entrada de usuario en la entrada users.xml, pueden conferirle una contraseña inicial de texto sin formato como ésta:

```
<user name="jdoe" password="changeMe"/>
```

Luego, el usuario puede cambiar su propia contraseña yendo a la página de cambio de contraseña (en un servidor ENS en ejecución) en `http://host:puerto/changePassword.jsp`. Se le solicitará que inicie sesión utilizando su contraseña antigua (la inicial), a continuación, se le presentará un formulario donde se podrá indicar la contraseña nueva dos veces. Tras hacerlo y pulsar "Cambiar contraseña", se actualizará el archivo users.xml para que contenga la nueva contraseña y ahora estará codificada en AES:

```
<user name="jdoe" password="{aes}ACdUKKZ0yQi09b4jY83d19UNGUmBbdT9znIT8kmUPmu3"/>
```

Las contraseñas deben tener una longitud entre 5 y 30 caracteres. Una contraseña de usuario puede no ser la misma que el nombre de usuario.

Si un usuario olvida la contraseña, un administrador puede restablecerla cambiando la entrada

```
<usuario>
```

para que vuelva a ser una conocida de texto sin formato, como anteriormente:

```
<user name="jdoe" password="changeMe"/>
```

Una vez hecho esto, el usuario deberá cambiarla de nuevo, como anteriormente.

## Supresión del acceso de usuarios a las funciones de Enterprise Name Search

Puede eliminar un usuario de un grupo determinado (eliminando su acceso a una parte concreta de la funcionalidad de ENS) editando el archivo users.xml y suprimiendo el nodo `<miembro...>` para ese usuario en el grupo relevante.

## Acerca de esta tarea

Puede eliminar un usuario de todo el ENS eliminándolo de todos los grupos y extrayendo su nodo de `<usuario>`.

## Control de acceso preciso para listas específicas

El archivo users.xml de ENS tiene un grupo denominado "searchers\_all\_lists" cuyos miembros pueden realizar búsquedas en todas las listas de nombres. Además, tiene un grupo denominado "managers\_all\_lists" cuyos miembros pueden realizar búsquedas en todas las listas y también pueden gestionarlas añadiendo, suprimiendo y obteniendo nombres.

## Acerca de esta tarea

Para muchas instalaciones, estos grupos genéricos de buscadores y gestores, o incluso el grupo más genérico "admins", proporcionan un modelo de control de acceso conveniente.

Es posible que algunos clientes quieran utilizar un modelo preciso en que determinados usuarios puedan realizar búsquedas en algunas listas de nombres y no en otras. ENS da soporte a esta función utilizando grupos "searchersNNN" y "managersNNN" enumerados.

Cuando se añade una lista de nombres a ESN utilizando NameLoader, está proporcionando el nombre de un rol de buscador (por ejemplo, "searcher101") para la lista. Cualquier usuario del grupo asociado ("searchers101") también puede realizar búsquedas de nombres en esa lista.

De igual modo, especifique el nombre de un rol de gestor (por ejemplo, "manager101") para la lista. Cualquier usuario del grupo asociado ("managers101") puede añadir y suprimir nombres de esa lista a través de los servicios web.

Fíjese que los nombres de grupos tal como se muestran en users.xml están en plural ("searchers101") pero los correspondientes nombres de rol que aparecen en NameLoader están en singular ("searcher101").

Estos roles precisos son opcionales. No es necesario que los utilice para descubrir que los grupos genéricos "searchers\_all\_lists" y "managers\_all\_lists" son suficientes, aunque siga necesitando especificar los roles cuando añada nombres con NameLoader.

De forma predeterminada, los grupos disponibles en users.xml son searchers101 a través de searchers120, y managers101 a través de managers120. Si necesita un rango más amplio de grupos, hasta searcher1-searcher500 y manager1-manager500, puede definirlos añadiendo más elementos de <grupo> en el archivo users.xml y más elementos <seguridad-rol> en el archivo applications.xml. Ambos archivos se hallan en la carpeta *ensroot/ibm-home/wlp*. Consulte los comentarios en esos archivos para obtener más información.

Cuando realice búsquedas de nombres en ENS, utilizando la GUI de búsqueda o los servicios web, solamente verá los resultados de las listas de nombres para las que disponga de permisos de búsqueda. Por ejemplo, supongamos que: la lista de nombres CLIENTES está asociada al rol searcher101, la lista de nombres EMPLEADOS está asociada al rol searcher102 y el usuario "joselm" es solamente un miembro del grupo searchers101. Entonces:

- Cuando el usuario joselm busca "John Smith", ya sea en la GUI de búsqueda de ENS o utilizando los servicios web, solamente se hallarán resultados de la lista CLIENTES y no de EMPLEADOS.
- Cuando el usuario joselm utiliza la GUI de búsqueda de ENS, la lista de nombres en las que pueden realizarse búsquedas que se muestra incluye CLIENTES pero no EMPLEADOS.
- Cuando el usuario utiliza el servicio web "get namelists", la lista de resultados incluye CLIENTES pero no EMPLEADOS.

## **Gestión de listas de nombres con el programa de utilidad NameLoader**

Puede gestionar listas de nombres para Enterprise Name Search con el programa de utilidad NameLoader. El programa de utilidad NameLoader es un programa Java autónomo que toma los nombres de origen sin formato de listas de nombres, analiza los nombres mediante IBM NameWorks y almacena los nombres y los resultados de los análisis de nombre en la base de datos de ENS.

El programa de utilidad NameLoader para ENS está relacionado con el programa de utilidad Name Preprocessor utilizado en Distributed Search, pero no lo sustituye. El programa de utilidad NameLoader carga nombres de un archivo de lista de nombres y los almacena en la base de datos de dos formas distintas: como "nombres de origen" y "nombres de búsqueda".

Los nombres de origen son los nombres tal como se han recibido originalmente, con su análisis y script original. NameLoader utiliza GNM para analizar estos nombres de origen con el fin de generar formatos analizados o nombres de búsqueda. Este análisis incluye la determinación de una categoría (de persona o de organización), un entorno cultural y (para nombres de persona) un análisis separando el nombre de pila del apellido. NameLoader busca o añade los registros de base de datos para almacenar nombres de origen, nombres de búsqueda y una correlación entre ellos.

También puede utilizar el programa de utilidad NameLoader para volver a cargar o eliminar nombres. La célula de ENS puede permanecer activo cuando se utiliza NameLoader para cargar, volver a cargar y eliminar nombres de una lista, aunque para conseguir un funcionamiento de eliminación lo más eficiente posible ("borrar todo"), la célula debe estar inactiva.

## Lectura de nombres de origen desde los archivos de listas de nombres

Los archivos de listas de nombres son archivos de texto plano con formato CSV (valores separados por comas). Si se utilizan caracteres no latinos, el archivo debería utilizar la codificación UTF-8. Cada fila de la lista de nombres representa un nombre de origen y suele contener la información siguiente:

Tabla 13. Campos en archivos CVS para los nombres de entrada: formato normal

Campo	Significado	Ejemplo	Valor predeterminado	Nota
0	Apellido	"SMITH"	blanco	O el apellido o el nombre de pila puede ser un espacio en blanco, pero no los dos.
1	Nombre de pila	"JOHN"	blanco	O el apellido o el nombre de pila puede ser un espacio en blanco, pero no los dos.
2	idData	300011	sin valor predeterminado	Obligatorio. Identificador que debe ser exclusivo dentro de un origen de datos.
3	Cultura del apellido	3	blanco	Opcional
4	Cultura del nombre de pila	3	blanco	Opcional
5	Categoría	0	blanco	Opcional. Indica el tipo de nombre, ya sea de persona, de organización o no especificado. Utilice "p" para de persona u "o" para de organización. Si lo prefiere, puede utilizar el dígito "0". para de persona y el "1" para de organización pero las letras pueden prestarse más fácilmente a confusión.

Algunos ejemplos de líneas de entrada válidas utilizando el formato normal:

Smith,John,x101,1,1,p - John Smith, ID externo x101, entorno cultural anglosajón para sn/gn. Nombre de persona.

John Smith,,x102,1,1,p - nombre de campo único en lugar de sn/gn. Fíjese en las comillas dobles.

Smith,John,x103,, - se omite el entorno cultural y la categoría. GNM determina ambas cosas mediante el análisis

Smith,John,x104 - ditto. Cuando se omiten los campos finales, no es necesario utilizar comas.

△,△,x105,,,p - que muestra que esos nombres pueden utilizar un script no latino

Smith Corp,,x106,,,o - nombre de organización en el campo sn (tenga en cuenta las comas dobles), categoría "o" explícita.

Smith Corp,,x107 - nombre de organización, sin marcar la categoría. GNM realiza la determinación mediante análisis

### **Lectura de nombres de origen desde los archivos de listas de nombres: formato de entrada ampliado**

En la mayoría de casos, el formato de entrada normal descrito anteriormente resulta adecuado pero existe un formato alternativo para una situación en que se utiliza un script no latino para la entrada. La mayoría de tipos de script puede reconocerse de forma automática pero ENS y GNM no distinguen automáticamente hanzi de kanji en nombres de entrada.

Si su lista de nombres contiene un nombre de organización japonés escrito utilizando solamente caracteres chinos, no kana, GNM no podrá reconocerlo automáticamente como nombre japonés. Para obtener una lista de nombres que incluya este tipo de nombres, debería utilizar el nuevo formato de archivo ampliado que le permite especificar un tipo de script explícito cuando sea necesario. Este formato coloca los campos en un orden diferente y añade un campo en el que puede especificar el tipo de script de forma explícita. Mediante este formato, el archivo CSV contiene la información siguiente para cada fila.

Cuando ejecute NameLoader con este tipo de archivo, también tendrá que especificar "-extendedInputFormat" en la línea de mandatos de NameLoader o "extendedInputFormat=true" en el archivo de configuración del cargador.

*Tabla 14. Campos en archivos CVS para nombres de entrada: formato ampliado, una línea de datos para un nombre de persona*

<b>Campo</b>	<b>Significado</b>	<b>Ejemplo</b>	<b>Valor predeterminado</b>	<b>Notas</b>
0	ID externo	300011	Ningún valor predeterminado	Obligatorio. Identificador que debe ser exclusivo dentro de un origen de datos.

Tabla 14. Campos en archivos CVS para nombres de entrada: formato ampliado, una línea de datos para un nombre de persona (continuación)

Campos	Significado	Ejemplo	Valor predeterminado	Notas
1	Categoría	0	blanco	Opcional. Indica el tipo de nombre, ya sea de persona, de organización o no especificado. Utilice "p" para de persona u "o" para de organización. Si lo prefiere, puede utilizar el dígito "0" para de persona y "1" para de organización pero las letras pueden prestarse más fácilmente a confusión.
2	Tipo de script		blanco	Para nombres de persona, no se utiliza este campo y se debería dejar en blanco.
3	Apellido o nombre de persona completo	"Smith" o "John Smith"	blanco	O el apellido o el nombre de pila puede ser un espacio en blanco, pero no los dos.
4	Nombre de pila	"JOHN"	blanco	O el apellido o el nombre de pila puede ser un espacio en blanco, pero no los dos.
5	Cultura del apellido	1	blanco	Opcional
6	Cultura del nombre de pila	1	blanco	Opcional

Tabla 15. Campos en archivos CVS para nombres de entrada: formato ampliado, una línea de datos para un nombre de organización

Campos	Significado	Ejemplo	Valor predeterminado	Notas
0	ID externo	300011	Ningún valor predeterminado	Obligatorio. Identificador que debe ser exclusivo dentro de un origen de datos.
1	Categoría	0	blanco	Opcional. Indica el tipo de nombre, ya sea de persona, de organización o no especificado. Utilice "p" para de persona u "o" para de organización. Si lo prefiere, puede utilizar el dígito "0" para de persona y "1" para de organización pero las letras pueden prestarse más fácilmente a confusión.
2	Tipo de script			Sugerencia opcional explícita de scriptType, casi siempre se deja en blanco. Actualmente el único uso para este campo es distinguir entradas de nombre de organización escritas en hanzi de las que están escritas en kanji especificando CHINESESCRIPT o JAPANESESCRIPT. Otros tipos de script se determinan automáticamente a partir del texto del nombre.
3	Nombre de organización	"Smith Corp"	Ningún valor predeterminado	No se puede dejar en blanco

Tabla 15. Campos en archivos CVS para nombres de entrada: formato ampliado, una línea de datos para un nombre de organización (continuación)

Campos	Significado	Ejemplo	Valor predeterminado	Notas
4	Entorno cultural del nombre de organización	1	blanco	Opcional

## Componente analizador de NameLoader

El componente Analyzer de Name Loader realiza el análisis siguiente en cada nombre de origen de entrada:

- Procesa cada nombre de origen mediante la transliteración del nombre.
- Categoriza el nombre en un tipo de nombre si el elemento "Category" está en blanco en el archivo CSV plano.
- Busca el mejor análisis de la forma original del nombre si la categoría es "personal".
- Busca uno o varios nombres de búsqueda transliterados en el nombre de origen.

## Nombre de origen y nombres de búsqueda

Los *nombres de origen* son los nombres que se han cargado originalmente desde una lista de nombres. En cambio, los *nombres de búsqueda* son los nombres que resultan del análisis que realiza IBM NameWorks. El esquema almacena los dos tipos, nombre de origen y nombre de búsqueda.

Los nombres de origen se encuentran en la tabla ENS\_SOURCE\_NAME. Los nombres de búsqueda se encuentran en la tabla ENS\_SEARCH\_NAME. Enterprise Name Search gestiona la correlación entre los dos tipos. Por razones de rendimiento, cuando el nombre se graba en las tablas de esquema, la operación de escritura se realiza mediante varias hebras. Cada hebra utiliza una actualización de base de datos por lotes para añadir o actualizar varios registros. El número de hebras y el tamaño del lote se pueden configurar en el archivo de configuración del programa de utilidad NameLoader.

Si dos nombres de búsqueda son idénticas excepto por el entorno cultural, ambos nombres de búsqueda se almacenan por separado. Por ejemplo, "June Park " con una designación de entorno cultural coreano se almacena por separado de "June Park" con una designación de entorno cultural anglosajón porque las reglas de coincidencia de estos dos entornos culturales son distintas.

Si dos nombres de búsqueda son idénticos excepto por el distintivo de elemento alternativo, ambos nombres también se almacenan por separado. Por ejemplo, el nombre de origen "Elton John" es diferente del nombre de origen "John Elton". Una solicitud de búsqueda de nombres de "Elton John" busca e informa los dos nombres de origen, pero el nombre de origen de John Elton se notifica con una puntuación de coincidencia inferior porque está basado en un elemento alternativo.

## Archivo loader.config

El programa de utilidad NameLoader tiene su propio archivo de configuración, denominado loader.config, que especifica el comportamiento de la instancia de IBM NameWorks utilizada por Enterprise Name Search. Toda la información sobre

las opciones y configuración del programa de utilidad NameLoader se especifica en este archivo.

## Mandatos de NameLoader

El programa de utilidad NameLoader se utiliza para gestionar listas de nombres para su uso con Enterprise Name Search. Ejecute NameLoader escribiendo un sencilla mandato en el indicador de mandatos del sistema operativo para realizar una operación determinada. Puede controlar las operaciones de NameLoader mediante el uso de opciones y valores en la línea de mandatos que se especifican en el archivo de configuración de NameLoader.

Para ejecutar el programa de utilidad NameLoader, vaya al directorio <vía acceso instalación>/bin/ y escriba **nameLoader** con uno o varios parámetros o acciones siguientes:

Tabla 16. Mandatos de NameLoader de ENS

Acción	Efecto	Ejemplos
-load	(Valor predeterminado) Carga nombres de origen desde un archivo CSV, los procesa en nombres de búsqueda y añade o actualiza entradas en las tablas siguientes: <ul style="list-style-type: none"> <li>• <b>ens_source_name</b></li> <li>• <b>ens_search_name</b></li> <li>• <b>ens_search_source_name</b></li> <li>• <b>ens_name_list</b></li> </ul> De forma opcional también añade registros a las tablas <b>ens_search_name_revs</b> y <b>ens_search_name_adds</b> para realizar un seguimiento de las revisiones.	<p>Cargar una única lista de nombres utilizando los valores predeterminados:</p> <pre>&lt;vía acceso instalación&gt;/bin/nameLoader -load -nlc EMPLOYEES -nld "Current employees" -in employees.csv</pre> <p>Cargar una única lista de nombres mediante un archivo de configuración de Loader propio:</p> <pre>&lt;vía acceso instalación&gt;/bin/nameLoader -load -nlc EMPLOYEES -nld "Current employees" -in employees.csv myLoaderConfigFile.config</pre> <p><b>Nota:</b> Puede especificar uno de los archivos de configuración como último argumento de línea de mandatos en cualquiera de estos ejemplos.</p>
-reload	Combina operaciones de borrar y cargar.	<p>Borrar y volver a cargar una única lista de nombres mediante los valores predeterminados:</p> <pre>&lt;vía acceso instalación&gt;/bin/nameLoader -reload -nlc EMPLOYEES -nld "Current employees" -in employees.csv</pre>
-clear	Elimina las entradas <b>ens_name_list</b> y <b>ens_source_name</b> de una lista de nombres determinada en la base de datos. No elimina las entradas de <b>ens_search_name</b> .	<p>Borrar una única lista de nombres:</p> <pre>&lt;vía acceso instalación&gt;/bin/nameLoader -clear -nlc EMPLOYEES</pre>
-clearAll	Elimina todas las entradas de <b>ens_name_list</b> , <b>ens_source_name</b> y <b>ens_search_name</b> .	<p>Borrar todas las listas de nombres:</p> <pre>&lt;vía acceso instalación&gt;/bin/nameLoader -clearAll</pre>
-usage	Muestra todas las opciones del programa de utilidad NameLoader.	<p>Consulte las opciones adicionales para el programa de utilidad NameLoader:</p> <pre>&lt;vía acceso instalación&gt;/bin/nameLoader -usage</pre>

El archivo predeterminado loader.config que ha creado el instalador contiene comentarios que describen cada entrada. Si se manipula una gran cantidad de

nombres, posiblemente desee realizar ajustes en los parámetros relacionados con el rendimiento en este archivo. Esto puede implicar una diferencia significativa en el rendimiento de NameLoader. Consulte los comentarios del archivo si desea una explicación de cada parámetro de rendimiento.

Puede especificar un archivo de configuración de NameLoader como el último argumento de línea de mandatos a la hora de ejecutar NameLoader. Si lo omite, se utilizará el archivo `loader.config` predeterminado.

La mayoría de opciones y detalles de configuración, incluida la vía de acceso del archivo de configuración de Global Name Management, pueden especificarse en el archivo de configuración de NameLoader. De forma alternativa, puede especificar los detalles que normalmente varían de una lista de nombres a otra como argumentos de la línea de mandatos. Esto permite que un archivo de configuración se pueda utilizar para diversas listas de nombres.

Cuando ejecute NameLoader con argumentos no válidos o con el argumento `"-usage"`, se mostrará un resumen de los argumentos de línea de mandatos válidos.

### **Archivo de configuración de NameLoader**

La función NameLoader está controlada mediante una combinación del archivo de configuración y los argumentos de línea de mandatos. El archivo de configuración de NameLoader es independiente del archivo de configuración de Global Name Management, que especifica el comportamiento de NameWorks. Esta separación de los archivos de configuración permite que los archivos de configuración de NameLoader de varias listas de nombres compartan un archivo de configuración común de Global Name Management.

Al instalar ENS, se genera un archivo de configuración de NameLoader predeterminado con los valores adecuados de la instalación. Este archivo reside en *vía de acceso de la instalación*/`data/loader.config`. Si ejecuta el mandato `nameLoader` sin especificar otro archivo de configuración, se utilizará el predeterminado. Puede editar este archivo para ajustar sus valores, o bien puede crear archivos de configuración personalizados para diferentes propósitos. El archivo de configuración predeterminado tiene comentarios que describen cada valor.

El archivo de configuración de NameLoader se especifica en un argumento de línea de mandatos al ejecutar NameLoader. Todas las demás opciones y detalles de configuración, incluida la vía de acceso del archivo de configuración de Global Name Management, pueden especificarse en el archivo de configuración de NameLoader. De forma alternativa, puede especificar los detalles que normalmente varían de una lista de nombres a otra como argumentos de la línea de mandatos. Esto permite que un archivo de configuración se pueda utilizar para diversas listas de nombres.

El archivo `loader.config` de ejemplo que se incluye con ENS tiene comentarios que describen cada entrada. NameLoader muestra información sobre los argumentos de línea de mandatos cuando se ejecuta con el argumento `"-usage"` o con argumentos no válidos.

Tabla 17. Entradas del archivo de configuración de NameLoader de ENS

Clave del archivo de configuración de NameLoader	Tipo	Valor típico	Notas	Opción de línea de mandatos correspondiente
nameListCode	Serie	CLIENTES	Nombre de la lista de nombres externa. Normalmente se proporciona en los argumentos de línea de mandatos a menos que confeccione los archivos de configuración de cargador según la lista de nombres.	-nameListCode o -nlc
nameListDescription	Serie	"Nuestro clientes"	Descripción del origen de datos. Normalmente se proporciona en los argumentos de línea de mandatos a menos que confeccione los archivos de configuración de cargador según la lista de nombres.	nameListDescription o -nld
nameListManagerRole	Serie	manager107	Rol de usuario de ENS obligatorio para realizar cambios en esta lista. Los valores aceptados son "manager0" a "manager400".	-managerRole o -mr
nameListSearcherRole	Serie	usuario de ENS obligatorio para buscar en esta lista. El rango disponible predeterminado es "searcher101" en "searcher120".*	Rol de usuario de ENS obligatorio para buscar en esta lista o para ver si existe. Los valores válidos son "searcher0" a "searcher400".	-searcherRole o -sr
inFileName	Serie	project1/badGuys.in	Vía de acceso del archivo CSV que se va a utilizar como entrada.	-inFileName o -in
jdbcDriver	Serie	com.ibm.db2.jcc.DB2Driver	Nombre de la clase de controlador JDBC.	
dbUrl	Serie	jdbc:db2:ENS	URL de la base de datos.	
dbSchema	Serie	ens	Nombre de esquema de base de datos.	

Tabla 17. Entradas del archivo de configuración de NameLoader de ENS (continuación)

Clave del archivo de configuración de NameLoader	Tipo	Valor típico	Notas	Opción de línea de mandatos correspondiente
dbUserId	Serie	yourDBUser	ID de usuario para el acceso a la base de datos. Puede ser en forma de argumentos de línea de mandatos para evitar que se almacene en el archivo de configuración. Si no se especifica en ningún sitio, el sistema se lo solicita al usuario en la consola.	-databaseUserid o $\Delta$ dbu
dbPassword	Serie	yourDBPass	Contraseña para acceder a la base de datos. Puede ser en forma de argumentos de línea de mandatos para evitar que se almacene en el archivo de configuración. Si no se especifica en ningún sitio, el sistema se lo solicita al usuario en la consola.	-databasePassword o -dbp
doFullName	Booleano	true	Controla si los nombres de origen de un solo campo (con apellidos, pero sin nombre de pila) se analizan en nombre de pila y apellido durante la operación de carga.	
parseThreshold	Entero	65 de 0 a 100	Controla el uso de análisis alternativos en los nombres de origen, si doFullName es true y se analiza un nombre de un solo campo.	
doCategorize	Booleano	false	Controla si NameWorks clasifica los nombres de origen de un solo campo sin categoría (person/org) durante la operación de carga.	
gnrConfigFile	Serie	/abc/ gnr.config	Vía de acceso del archivo de configuración de Global Name Management que controla el desglose y análisis.	
numThreadsAnalyzer	Entero	** del 1 al 64	Número de hebras utilizadas para el análisis de nombres. Se ajusta para mejorar el rendimiento.	
numThreadsWriter	Entero	** del 1 al 64	Número de hebras utilizado para la escritura de los nombres en la base de datos o en el archivo de salida. Se ajusta para mejorar el rendimiento.	

Tabla 17. Entradas del archivo de configuración de NameLoader de ENS (continuación)

Clave del archivo de configuración de NameLoader	Tipo	Valor típico	Notas	Opción de línea de mandatos correspondiente
dbBatchSize	**Entero32	del 1 al 256	Número de nombres procesado en el momento de escribir en la base de datos. Se ajusta para mejorar el rendimiento.	
threadQueueSize	**Entero	del 1 al 9999	Número de nombres retenidos en las colas para el análisis o para la salida. Se ajusta para mejorar el rendimiento.	

\* - Para el control de acceso preciso según se ha descrito en Control de acceso preciso para listas específicas. Si el rango entre 101 y 120 no es suficiente, puede ampliarlo tal como se describe en esa sección.

\*\* - Para obtener sugerencias sobre el ajuste de cada uno de estos parámetros de rendimiento, consulte los comentarios correspondientes en el archivo loader.config de su instalación.

### Listas de nombres

El archivo de configuración de NameLoader y los argumentos de línea de mandatos especifican lo siguiente para una lista de nombres:

- Código de la lista de nombres
- Descripción
- Rol del buscador
- Roles del gestor

Si la tabla de lista de nombres de ENS (ens\_name\_list) contiene una entrada con el código de lista de nombres especificado, indica que se ha actualizado. Si no, se añade uno.

### Cargar nombres de listas de nombres mediante el programa de utilidad Name Loader

Siempre que tenga una nueva lista de nombres o actualizaciones para una lista de nombres, puede utilizar el programa de utilidad Loader Nombre para procesar y cargar los nombres en el esquema de base de datos. Para obtener las actualizaciones para una lista de nombres existente, también puede utilizar los servicios web "addName" y "deleteName".

### Acerca de esta tarea

Los argumentos especificados en la línea de mandatos pueden alterar temporalmente los valores predeterminados y los valores del archivo de configuración. En la línea de mandatos sólo es necesario utilizar el nombre del archivo de configuración. Todo lo demás puede especificarse en el archivo de configuración. Una vez finalizada la instalación, puede encontrarse una plantilla del archivo de configuración de Name Loader en <vía acceso instalación>/data/loader.config. Se completa de forma automática con la

información de conexión de base de datos ENS, excepto el nombre y la contraseña del usuario de la base de datos.

## Procedimiento

En una línea de mandatos, inicie el programa de utilidad Name Loader mediante la especificación del mandato siguiente:

### UNIX/Linux

```
<vía acceso instalación>/bin/nameLoader <opciones> <archivoConfig>
```

### Windows

```
<vía acceso instalación>\bin\nameLoader.bat <opciones> <archivoConfig>
```

## Ejemplo

Para obtener un mensaje de uso de Name Loader que contenga todas las opciones, escriba lo siguiente:

```
vía acceso instalación/bin/nameLoader -usage
```

que genera la salida siguiente:

```
Uso: nameLoader <opciones> <archivoConfig>
```

Opciones:

```
-in customers.csv      o -inFileName customers.csv  
-nlc CUSTOMERS        o -nameListCode CUSTOMERS  
-nld "Our customers"  o -nameListDescription "Our customers"
```

```
-sr searcher101       o -searcherRole searcher101  
-mr manager101        o -managerRole manager101  
-dbu myDatabaseUserid o -dbUserid myDatabaseUserid  
-dbp myDatabasePassword o -dbPassword myDatabasePassword  
-revtrack o -norevtrack o -revisionTracking o -noRevisionTracking o -autorevtrack o -autoRevisionT
```

```
-extendedInputFormat (si procede)
```

```
-verbose o -trace o -quiet (como máximo uno de estos)
```

```
-load, -reload, -clear o -clearAll (como máximo uno de estos; el valor predeterminado es -load)
```

Sólo se necesita **configurationFileName** en la línea de mandatos. Todo lo demás puede especificarse en el archivo de configuración. O, si desea utilizar el archivo de configuración predeterminado, puede omitir el archivo de configuración de la línea de mandatos y especificar los argumentos **-nlc**, **-nld** e **-in**, tal como se muestra a continuación:

```
nameLoader -nlc "CUSTOMERS" -nld "Nuestros clientes" -in customers.csv -dbu mydbuserid -dbp mydbpa
```

Si omite el ID de usuario y la contraseña de la base de datos en ambos lugares, el programa se los solicitará en la consola. Acciones:

- **-load** añade/actualiza la base de datos sin eliminar nada.
- **-reload** elimina los nombres de origen de esta lista de nombres y, a continuación, vuelve a añadir los nombres de origen y de búsqueda
- **-clear** elimina los nombres de origen de esta lista de nombres sin añadir nada
- **-clearAll** elimina TODOS los nombres de origen, nombres de búsqueda y listas de nombres de la base de datos

Suponiendo que el archivo de configuración sólo contiene los valores de su conexión de base de datos ENS y los parámetros de rendimiento, a continuación figura un ejemplo de un mandato que carga la lista de nombres CUSTOMERS en un sistema UNIX/Linux mientras que la célula de ENS está apagada/inactiva.

```
vía acceso instalación/bin/nameLoader -load -in /CUSTOMERS.txt -nlc CUSTOMERS -nld "Listar con los n
```

Donde

- **-load** especifica la acción que debe llevarse a cabo. Cargar (load) es el valor predeterminado.
- **-in <vía acceso instalación>** especifica la vía de acceso del archivo de texto de lista de nombres con formato de tipo CSV.
- **-nlc CUSTOMERS** especifica que el "código de lista de nombres" (identificador corto de lista de nombres) es "CUSTOMERS"
- **-nld <descripción>** especifica una descripción de la lista de nombres que el usuario puede entender.
- **-sr <rol de seguridad>** otorga acceso de lectura para esta lista de nombres a todos los usuarios que tengan el rol de seguridad "searcher120". Estos roles deben tener un formato de "searcher1" a "searcher500".
- **-mr <rol de seguridad>** otorga acceso de grabación (gestión) para esta lista de nombres a todos los usuarios que tengan el rol de seguridad "manager120". Estos roles deben tener un formato de "manager1" a "manager500".

A continuación figura un ejemplo de cómo borrar todas las listas de nombres en UNIX/Linux, cuando no se esté ejecutando la célula de ENS:

```
<vía acceso instalación>/bin/nameLoader -clearAll <vía acceso instalación>/data/loader.config
```

A continuación figura un ejemplo de cómo cargar la lista de nombres VENDORS en UNIX/Linux, mientras se está ejecutando la célula de ENS:

```
<vía acceso instalación>/bin/nameLoader -load -revtrack -in /VENDORS.txt -nlc VENDORS -nld "Lista con los nombres de todos los proveedores" -sr searcher150 -mr manager150 <vía acceso instalación>/data/loader.config
```

**-revtrack** activa el seguimiento de las revisiones para las revisiones de nombres. Esta característica permite a los buscadores de una célula de ENS que esté en ejecución saber qué nombres se han insertado recientemente, y añadirlos gradualmente a sus particiones de lista de nombres que están en la memoria. Esto le permite añadir nombres procedentes de una lista de nombres mientras se está ejecutando el sistema. Si esto no se especifica cuando se ejecuta Name Loader mientras la célula de ENS está en ejecución, ENS no sabrá que existen nombres nuevos hasta que se reinicie la célula.

## Actualización de las estadísticas de base de datos

Tras utilizar NameLoader para cargar algunos nombres en las tablas de ENS de la base de datos, es importante optimizar el rendimiento utilizando mandatos de la base de datos con el fin de actualizar las estadísticas acerca de tablas que retienen nombres. De esta forma se permite que la base de datos optimice la forma de acceder a estas tablas y es lo que realmente diferencia el rendimiento de ENS.

### Acerca de esta tarea

Se llevará a cabo esta tarea tras cargar ciertos nombres. Puede que le parezca importante ejecutar estadísticas tras añadir 200.000 nombres y hacerlo de nuevo con tamaños mayores (digamos, por ejemplo, tras añadir un millón de nombres y de nuevo tras añadir 5 millones, etc.). Los mandatos de DB2 que aparecen a continuación se pueden ejecutar mientras NameLoader está en ejecución.

Esta tarea suele llevarla a cabo un DBA. Si usted está solo (supongamos un sistema de demo), los mandatos de DB2 para llevar a cabo esta tarea, suponiendo un nombre de esquema de "ENS", serían:

```
RUNSTATS on table ENS.ENS_SEARCH_NAME ON KEY COLUMNS;  
RUNSTATS on table ENS.ENS_SOURCE_NAME ON KEY COLUMNS;  
RUNSTATS on table ENS.ENS_SEARCH_SOURCE_NAME ON KEY COLUMNS;  
RUNSTATS on table ENS.ENS_SEARCH_NAME_ADDS ON KEY COLUMNS;  
RUNSTATS on table ENS.ENS_SEARCH_NAME_REVS ON KEY COLUMNS;  
COMMIT;
```

La equivalencia en Oracle es DBMS\_STATS. Encontrará más información disponible en: [http://docs.oracle.com/cd/B19306\\_01/appdev.102/b14258/d\\_stats.htm#i103646](http://docs.oracle.com/cd/B19306_01/appdev.102/b14258/d_stats.htm#i103646) .

Fíjese que algunos sistemas ejecutan estadísticas de forma automática por las noches con el fin de optimizar su funcionamiento. Si su sistema lleva a cabo esta tarea en una hora en que las tablas de ENS están vacías de nombres, puede deshacerse el efecto generado anteriormente por la ejecución manual de estadísticas. En tal caso, puede que tenga que volver a ejecutar las estadísticas con nombres cargados, como se mostraba más arriba.

### **Eliminar nombres del esquema de Enterprise Name Search**

Mediante el programa de utilidad NameLoader, puede suprimir de forma eficiente todos los nombres de las listas de nombres del esquema de Enterprise Name Search con **-clearAll**, o suprimir solamente una lista de nombres específica con **-clear**.

#### **Acerca de esta tarea**

La opción **-clearAll** suprimirá todos los nombres y listas de nombres. Esta operación requiere que la célula esté inactiva.

```
nameLoader -clearAll
```

La opción **-clear** suprimirá un solo nombre. El hecho que los nombres de origen de diversas listas de nombres se puedan correlacionar con los mismos nombres de búsqueda significa que esta operación es más compleja e inherentemente menos eficiente que **-clearAll**.

```
nameLoader -clear -nlc CUSTOMERS
```



---

## Capítulo 6. Configuración de IBM NameWorks

La especificación de valores predeterminados, alteraciones temporales, información de estrategias de búsqueda y otros elementos afectan al funcionamiento de IBM NameWorks. Para utilizar la API de IBM NameWorks de forma eficaz, configure IBM NameWorks especificando valores predeterminados e información de configuración en el archivo de configuración de IBM NameWorks o mediante la clase `Configuration()`.

El uso de un archivo externo, como `nw.config`, o de la clase `Configuration()` depende de su configuración y del número de CPU disponibles para el proceso. El uso de la clase `Configuration()` podría aumentar el rendimiento durante la inicialización porque un archivo de configuración externo no necesita añadirse a la memoria, leerse ni procesarse. Sin embargo, esta mejora es normalmente pequeña y varía en función de las distintas configuraciones.

### Archivo de configuración de IBM NameWorks

Puede especificar información de configuración de IBM NameWorks utilizando valores de configuración en un único archivo de texto que se lee durante la inicialización del sistema. La información de este archivo de texto tiene el formato siguiente, en el que nombres de sección entre corchetes delimitan listas de pares clave/valor:

```
[Nombre sección] Clave=Valor
```

Los nombres de sección y de clave no son sensibles a mayúsculas y minúsculas y los valores pueden contener cualquier carácter.

Puede controlar el límite del número de hebras necesarias en tiempo de ejecución para las búsquedas en un proceso de IBM NameWorks. Un mecanismo de agrupación de hebras agrupa las actividades de generación de hebras para garantizar que el número de hebras generadas por Embedded Search no sobrepase el límite especificado. Puede establecer este límite mediante la entrada `MaxThreads=` de la sección `[General]` del archivo de configuración de IBM NameWorks.

El enfoque recomendado para especificar alteraciones temporales de parámetros de comparación es crear un archivo alteraciones temporales de parámetros de comparación predeterminado que se ajuste a cada lista de datos, aplicar ese archivo a la lista de datos, y luego utilizar alteraciones temporales relativas en la sección `[Search Strategy]` del archivo de configuración de IBM NameWorks. Una vez que las alteraciones temporales predeterminadas se han asociado a una lista de datos, no deberá especificar alteraciones temporales en cada consulta.

**Nota:** La utilización de una hebra para cada núcleo de CPU disponible en la máquina destino ayuda a alcanzar un rendimiento óptimo de IBM NameWorks.

### La clase `Configuration()`

Mediante la clase `Configuration()` de IBM NameWorks, puede especificar valores predeterminados y mantenerlos en la memoria como parte de su programa en lugar de utilizar un archivo de configuración externo. Puede utilizar esta clase para especificar los valores de configuración predeterminados y, a continuación, especificar alteraciones temporales utilizando una de las opciones siguientes:

- Cree un conjunto de parámetros de comparación predeterminados que se ajustan para cada lista de datos mediante el uso de la clase Strategy() e implemente las alteraciones temporales utilizando los métodos de alteración temporal, como addGivenNameOverride(), addSurnameOverride() y addOrganizationNameOverride().
- Utilice la clase CompParmsOverrides() para proporcionar alteraciones temporales de parámetros de comparación de nombres de pila, apellidos y nombres de organización a los métodos Scoring.search() y Scoring.compare().

El uso de la clase Configuration() le permite almacenar su configuración de IBM NameWorks en una base de datos y especificar dinámicamente configuraciones nuevas o actualizadas. El uso de esta opción es conveniente si desea guardar la configuración de IBM NameWorks en una base de datos y utilizar otra aplicación para crear instancias de IBM NameWorks.

---

## Especificación de valores de configuración mediante el archivo de configuración de IBM NameWorks

Puede establecer y actualizar la información de configuración de IBM NameWorks mediante la modificación de valores en el archivo de configuración que se lee durante la inicialización del sistema. La codificación de este archivo debe ser UTF-8.

### Acerca de esta tarea

Esta tarea se completa normalmente por un administrador del sistema, que realiza cambios en el archivo de configuración y reinicia el servidor de aplicaciones para que los cambios entren en vigor. Cada sección del archivo de configuración de IBM NameWorks contiene diferentes parámetros. Consulte los temas de esta sección para obtener más información sobre los parámetros que puede modificar en cada una de las secciones del archivo de configuración.

### Procedimiento

1. Vaya al directorio donde reside el archivo de configuración de IBM NameWorks.

Opción	Descripción
Si utiliza IBM NameWorks en la aplicación cliente	El archivo de configuración existe en cualquier directorio que se utiliza para la aplicación cliente personalizada.
Si utiliza IBM NameWorks como servicio web	La ubicación predeterminada del archivo de configuración es <i>dir_instalación/data</i> , donde <i>dir_instalación</i> es el directorio donde se ha instalado IBM InfoSphere Global Name Management .

2. Abra el archivo de configuración en un editor de texto y realice los cambios oportunos. Las secciones y parámetros que incluya variarán en función de su configuración.
3. Guarde el archivo de configuración.
4. Detenga y reinicie la aplicación cliente o servicio web para que los cambios entren en vigor.

Opción	Descripción
Si utiliza IBM NameWorks en la aplicación cliente	<ol style="list-style-type: none"> <li>1. Detenga la aplicación de cliente.</li> <li>2. Reinicie la aplicación cliente.</li> </ol>
Si utiliza IBM NameWorks como servicio web	<ol style="list-style-type: none"> <li>1. Vaya al directorio donde ha instalado IBM InfoSphere Global Name Management</li> <li>2. Ejecute el mandato <b>stopGNR</b> para detener el servicio web.</li> <li>3. Ejecute el mandato <b>startGNR</b> para reiniciar el servicio web.</li> </ol>

## Archivo de configuración de ejemplo

```
[General]
MaxThreads=4
CompParamsDefaults=/gnr/data/compparms.config

[Custom Tokens]
OR=GivenName
BARONESS=Title

[DataList:Distributed]
Type=1
Server=localhost|4250|1|

[DataList:Embedded]
Type=0
List=part1.csv
List=part2.csv
CompressedBitSig=true
TAQ=/gnr/data/taq.ibm
GNV=/gnr/data/gnv.ibm
SNV=/gnr/data/snv.ibm
PNREG=/gnr/data/angloRegRule.ibm,Anglo
PNREG=/gnr/data/swasianRegRule.ibm,Arabic
PNREG=/gnr/data/swasianRegRule.ibm,Pakistani
PNREG=/gnr/data/russianRegRule.ibm,Russian

[DateCompare]
EII=93

[Strategy:Broad]
[GNParams:Broad]
ANCHOR_FACTOR=0.95
COMPRESSED_SCORE_MAX=1.00
DO_COMPRESSED_SCORE=Y
FIELD_THRESH=0.50
FIELD_WEIGHT=0.40
INITIAL_INITIAL_SCORE=0.70
INITIAL_TOKEN_SCORE=0.75
MATCH_INITIALS=Y
OOPS_FACTOR=0.95

[SNParams:Broad]
ANCHOR_FACTOR=0.95
COMPRESSED_SCORE_MAX=1.00
DO_COMPRESSED_SCORE=Y
FIELD_WEIGHT=0.60
NAME_UNKNOWN_SCORE=0.50
OOPS_FACTOR=0.95
```

```

[Strategy:Narrow]
[GNParms:Narrow]
ANCHOR_FACTOR=0.85
COMPRESSED_SCORE_MAX=0.00
DO_COMPRESSED_SCORE=Y
FIELD_THRESH=0.60
FIELD_WEIGHT=0.40
INITIAL_INITIAL_SCORE=0.65
INITIAL_TOKEN_SCORE=0.70
NAME_UNKNOWN_SCORE=0.50
OOPS_FACTOR=0.85
[SNParms:Narrow]
ANCHOR_FACTOR=0.85
COMPRESSED_SCORE_MAX=1.00
DO_COMPRESSED_SCORE=Y
FIELD_WEIGHT=0.60
NAME_UNKNOWN_SCORE=0.40
OOPS_FACTOR=0.85

[Strategy:Standard]
[GNParms:Standard]
[SNParms:Standard]
[ONParms:Standard]

[Transliteration Modules]
Module=/gnr/data/arabicTransRule.ibm
Module=/gnr/data/cyrillicTransRule.ibm

[Reference Files]
NameSifter=/gnr/data/SifterRules.ibm

[Comparison Files]
TAQ=/gnr/data/taq.ibm
GNV=/gnr/data/gnv.ibm
SNV=/gnr/data/snv.ibm
PNREG=/gnr/data/angloRegRule.ibm,Anglo
PNREG=/gnr/data/swasianRegRule.ibm,Arabic
PNREG=/gnr/data/swasianRegRule.ibm,Pakistani
PNREG=/gnr/data/russianRegRule.ibm,Russian

```

## Sección General del archivo de configuración

La sección [General] contiene información que está disponible en todas las características. El soporte para agrupaciones de hebras y alteraciones temporales para los parámetros de comparación (CompParms) predeterminados se listan bajo esta sección del archivo de configuración.

Los campos siguientes pueden configurarse en la sección [General] del archivo de configuración:

```

[General]
MaxThreads=
CompParmsDefaults=DefaultAltScoreFactor=

```

### **MaxThreads=*n***

Indica el número máximo de hebras utilizadas para dar soporte a operaciones de búsqueda simultáneas. Este valor se aplica tanto a motores de búsqueda externos como a búsquedas incorporadas, y puede ser cualquier entero positivo igual ( $\geq$ ) o superior a cero. Sin embargo, los valores superiores al número de núcleos de procesador pueden degradar el rendimiento. El valor predeterminado es `MaxThreads=0`, que indica que no deben utilizarse hebras de búsqueda independientes. Si el valor especificado es inferior a cero se genera una excepción de parámetro no válido (GODW031E), y los valores no válidos provocan una condición de error.

**CompParmsDefaults=nombre de archivo que contiene alteraciones temporales**

Especifica el nombre de un archivo que contiene las alteraciones temporales de los parámetros de comparación predeterminados. El formato del archivo es el mismo que el utilizado por Distributed Search. El valor predeterminado es en blanco, que indica que no debe utilizarse ningún archivo de alteraciones temporales. Si se especifica un nombre de archivo no válido o datos no válidos, se generará un error de parámetros de comparación incorrectos (GODW035E).

**DefaultAltScoreFactor=factor**

Indica el factor predeterminado que se debe aplicar a la puntuación de similitud final para las coincidencias halladas en los análisis alternativos. Este valor se puede alterar temporalmente para listas de datos específicas. El valor debe ser un valor real positivo (coma flotante) inferior o igual a 1,0; de lo contrario, se generará una excepción de parámetro no válido (GOD031E) cuando se detecte un valor no válido.

## Sección Custom tokens (cadenas de caracteres personalizadas) del archivo de configuración

La sección [Custom Tokens] del archivo de configuración lista los tipos de cadenas de caracteres de análisis personalizadas.

Las cadenas de caracteres de análisis personalizadas se indican en la sección [Custom Tokens] con este formato:

```
[Custom Tokens]  
cadena de caracteres=tipo [,comentario ]
```

**cadena de caracteres**

Texto de la cadena de caracteres personalizada.

**tipo**

Tipo de la cadena de caracteres, que puede ser cualquiera de los siguientes:

**Nombre de pila**

Una cadena de caracteres de nombre normal, como John o Michael, utilizada generalmente como nombre de pila. Este tipo de cadena de caracteres también puede aparecer como apellido, pero es cuatro veces más probable que se trate como un nombre de pila.

**Apellido**

Cadena de caracteres de nombre normal, como McGillicuddy o Wiltshire, utilizada generalmente como apellido. Este tipo de cadena de caracteres también puede aparecer como nombre de pila, pero es cuatro veces más probable que se trate como un apellido.

**Initial** Una única letra que debe tratarse como inicial en lugar de como número romano u otro tipo de cadena de caracteres.

**Title** Serie que refleja generalmente una posición social y aparece habitualmente antes de otras cadenas de caracteres de nombre.

**Prefix** Partícula que se incluye en la misma frase de nombre que la cadena de caracteres raíz de nombre subsiguiente. Las cadenas de caracteres como *de* y *la* son prefijos.

**Suffix** Partícula que se incluye en la misma frase de nombre que la cadena de caracteres raíz de nombre precedente. Una cadena de caracteres como *aldeen* es un sufijo.

**Qualifier**

Calificador que generalmente indica relaciones generacionales o una

posición social. Cadenas de caracteres como *Jr.* (Junior) y *Esq.* (Esquire) son calificadores. Los calificadores no se incluyen en los campos de nombre de pila ni de apellido.

### **Conjunction**

Palabras como *and*, que unen varios nombres.

### **comentario**

Descripción opcional de la cadena de caracteres.

## **Sección Datalist del archivo de configuración**

Las descripciones de lista de datos se almacenan en secciones [`Datalist:nombre`] individuales, donde el valor *nombre* representa el nombre de la lista de datos que se pasa en el argumento *listas\_datos* de la llamada al método `search()`.

Las alteraciones temporales por lista de datos se aplican sólo a las listas de datos incorporadas. Puede especificar alteraciones temporales de parámetros de comparación para listas de datos individuales si añade una o más entradas `CompParamsDefaults=` a la sección [`Datalist:`] del archivo de configuración. Si utiliza esta opción, puede establecer los valores predeterminados para listas de datos específicas, mientras que las estrategias de búsqueda y las alteraciones temporales que se indican mediante la clase `CompParamOverrides` sólo alteran temporalmente los valores predeterminados de una consulta individual.

### **Información de Distributed Search en el archivo de configuración**

La sección [`Datalist:nombre`] del archivo de configuración admite entradas de Distributed Search, que se utilizan cuando Distributed Search está habilitado.

Los parámetros siguientes se aplican a las secciones [`Datalist:name`] del archivo de configuración de IBM NameWorks para búsquedas de Distributed Search.

Puede facilitarse más de una entrada `Server=`, lo que indicará que se han asociado varios servidores a la lista de datos indicada. Por tanto, las consultas deben federarse y se acumularán varios resultados. Si se asocia el distintivo `add` con una entrada `Server=`, ese servidor se utiliza para las operaciones `addName()`. Sólo puede configurarse con el distintivo `add` un servidor por cada lista de datos.

```
[Datalist:Distribuida]
Type=1
Server=host|puerto|nombre_lista[|add]IncludeTAQs=
```

#### **Type=*n***

Indica si esta lista de datos es de tipo búsqueda completa (`Type=1`) o de nombres exclusivos (`Type=2`). Si no se encuentra ninguna entrada `Type=n`, se presupone que la lista de datos es de tipo Búsqueda completa.

#### **host**

Dirección IP (nombre simbólico o expresión numérica) de la máquina de host. Si la dirección IP se escribe entre corchetes (igual que en `Server=[host]|puerto`) se utilizará el protocolo IPv6 para las comunicaciones con el motor de Distributed Search (que también se ha de configurar para que utilice el protocolo IPv6).

#### **puerto**

Dirección de puerto IP decimal (en el rango 0:65535)

#### **nombre\_lista**

Nombre de lista de datos interna que se utiliza en la instancia de Distributed Search. Este parámetro no se utiliza actualmente.

### ***add***

Indica que se debe utilizar este servidor cuando se añaden registros de nombres nuevos a una lista de datos. Cualquier servidor sin la opción *add* se tratará como de sólo lectura.

### **IncludeTAQs=**

Entrada única que indica si los valores de fórmulas de tratamiento y calificadores deben incluirse con las entradas de lista de datos. Puede especificar este valor en el archivo de configuración de IBM NameWorks para aplicarlo a las entradas de lista de datos durante las operaciones de adición, actualización y búsqueda especificando IncludeTAQs=true.

**Nota:** Si se incluye la información TAQ, las fórmulas de tratamiento se añaden al campo de nombre especificado y los calificadores se añaden al campo de apellido antes de que se utilicen los datos de nombre en las operaciones de comparación de comparación por pares o búsqueda.

## **Información de Embedded Search del archivo de configuración**

Las entradas de Embedded Search están soportadas en la sección [Datalist:*nombre*] del archivo de configuración y se utilizan si se habilita Embedded Search. Los valores no válidos provocan una condición de error para todos los parámetros.

Los parámetros siguientes son aplicables al realizar búsquedas en una lista de datos incorporada con el método search(). Todos los nombres de entrada pueden especificarse en mayúsculas, minúsculas o ambas mezcladas.

[Datalist:*Incorporada*]

Type=0

List=

CompressedBitSig=

TAQ=

GNV=

SNV=

ONV=

PNREG=

ONREG=

ONTERM=

ONTOPN=NativeTaq=

NativePnVar=

NativeOnVar=

NativeOnReg=

NativeOnReg=

### **Type=*n***

Las listas de datos de Embedded Search se identifican mediante una entrada Type=0 en la sección Datalist del archivo de configuración. El valor predeterminado de esta entrada es Type=0 (listas de datos de Embedded Search). Las entradas de tipo Server= asociadas con una lista de datos Type=0 se ignorarán.

### **List=*nombre del archivo que contiene registros de nombre***

Las listas de nombres individuales asociadas con una lista de datos se identifican mediante una entrada List=0 en la sección Datalist del archivo de configuración. Se aceptan varias entradas List=, y cada una de ellas se trata como una lista de nombres independiente asociada con la lista de datos. Los archivos de lista de nombres deben especificarse en formato .csv. Puede especificarse una lista de adiciones determinada añadiendo la serie |add al nombre de archivo. La lista de adiciones recibe nombres que se han añadido después de cargar la lista de nombres. Si una descripción de lista de datos de

Embedded Search no incluye entradas List=, se crea una única lista de nombres vacía que permite añadir nombres. La lista de nombres vacía se marca como lista de adiciones.

Si se especifica un nombre de archivo no válido para este parámetro, se generará un error de archivo de datos incorrecto (GODW032E).

**CompressedBitSig=*n***

Determina si deben incluirse firmas de bits para formatos comprimidos de los nombres. Este valor puede ser 0 ó 1, siendo 1 el valor predeterminado.

Si se especifica un valor de parámetro no válido, se generará un error de valor de parámetro no válido (GODW031E).

**TAQ=*nombre de vía de acceso de lista de TAQ***

Pueden asociarse uno o varios archivos de alteraciones temporales de TAQ con una lista de datos. Los archivos de alteraciones temporales de TAQ se aplican a cada lista de nombres asociada con una lista de datos.

Si se especifica un nombre de archivo incorrecto o si el contenido del archivo no puede cargarse, se generará un error de lista de datos de NameHunter (GODW037E).

**GNV= | SNV= | ONV=*nombre de vía de acceso de lista de variantes***

Pueden asociarse uno o varios archivos de variantes con una lista de datos. Los archivos de variantes se aplican a cada lista de nombres asociada con una lista de datos.

Si se especifica un nombre de archivo incorrecto o si el contenido del archivo no puede cargarse, se generará un error de lista de datos de NameHunter (GODW037E).

**PNREG= | ONREG=*nombre de vía de acceso de archivo de reglas de regularización, nombre de entorno cultural***

Pueden asociarse uno o varios archivos de regularización con una lista de datos. Los archivos de regularización se aplican a cada lista de nombres asociada con una lista de datos. El formato de esta entrada es el mismo que el de entradas similares de la sección [Comparison Files] del archivo de configuración de IBM NameWorks.

Si se especifica un nombre de archivo incorrecto o si el contenido del archivo no puede cargarse, se generará un error de lista de datos de NameHunter (GODW037E).

**ONTERM=*nombre de vía de acceso del archivo de términos***

Pueden asociarse uno o varios archivos de términos con una lista de datos.

Si un archivo de términos tiene un término en blanco, se genera un error de lista de datos de texto en blanco (GODH065E). Es posible que el término en el archivo conste de caracteres no permitidos

**ONTOPN=*n***

Determina si debe incluirse soporte adicional para buscar nombres de organización en comparación a nombres de persona. Este valor puede ser 0 ó 1, siendo 0 el valor predeterminado.

Si se especifica un valor de parámetro no válido, se generará un error de valor de parámetro no válido (GODW031E).

**NativeTaq=*nombre de vía de acceso de lista de TAQ de script nativo***

Pueden asociarse uno o varios archivos de alteraciones temporales de TAQ de

script nativo con una lista de datos. Los archivos de alteraciones temporales de TAQ de script nativo se aplican a cada lista de nombres asociada con una lista de datos.

Si se especifica un nombre de archivo incorrecto o si el contenido del archivo no puede cargarse, se generará un error de lista de datos de NameHunter (GODW037E).

***NativePnVar=nombre de vía de acceso de lista de variantes de nombre de persona de script nativo***

Pueden asociarse uno o varios archivos de variantes de nombre de persona de script nativo con una lista de datos. Los archivos de variantes se aplican a cada lista de nombres asociada con una lista de datos.

Si se especifica un nombre de archivo incorrecto o si el contenido del archivo no puede cargarse, se generará un error de lista de datos de NameHunter (GODW037E).

***NativeOnVar=nombre de vía de acceso de lista de variantes de nombres de organización de script nativo***

Pueden asociarse uno o varios archivos de variantes de nombre de organización de script nativo con una lista de datos. Los archivos de variantes se aplican a cada lista de nombres asociada con una lista de datos.

Si se especifica un nombre de archivo incorrecto o si el contenido del archivo no puede cargarse, se generará un error de lista de datos de NameHunter (GODW037E).

***NativePnReg=nombre de vía de acceso del nombre de script y del archivo de regularización de nombres de persona de script nativo***

Pueden asociarse uno o varios archivos de regularización de nombre de persona de script nativo con una lista de datos. Los archivos de regularización se aplican a cada lista de nombres asociada con una lista de datos. El formato de esta entrada es el mismo que el de entradas similares de la sección [Comparison Files] del archivo de configuración de IBM NameWorks.

Si se especifica un nombre de archivo incorrecto o si el contenido del archivo no puede cargarse, se generará un error de lista de datos de NameHunter (GODW037E).

***NativeOnReg=nombre de vía de acceso del nombre de script y del archivo de regularización de nombres de organización de script nativo***

Pueden asociarse uno o varios archivos de regularización de nombre de organización de script nativo con una lista de datos. Los archivos de regularización se aplican a cada lista de nombres asociada con una lista de datos. El formato de esta entrada es el mismo que el de entradas similares de la sección [Comparison Files] del archivo de configuración de IBM NameWorks.

Si se especifica un nombre de archivo incorrecto o si el contenido del archivo no puede cargarse, se generará un error de lista de datos de NameHunter (GODW037E).

## **Sección Search Strategy del archivo de configuración**

La información de estrategias de búsqueda se almacena en varias secciones del archivo de configuración. Puede establecer parámetros individuales para nombres de persona en las secciones GNParms y SNParms y parámetros de nombres de organización en la sección ONParms.

[Strategy:nombre]  
GNCulture=  
SNCulture=

```
ONCulture=  
MinScore=  
MaxReplies=  
SearchOpt=  
IncludeTAQs=  
[GNParams:name]  
...  
[SNParams:name]  
...  
[ONParams:name]  
...
```

#### **name**

El valor de *nombre* hace referencia al nombre de la estrategia de búsqueda. Estas secciones contienen los parámetros de comparación del nombre de pila, apellido y nombre de organización en el formato *nombre=valor* previsto por NameHunter Distributed Search. Estas secciones se ignorarán si no hay ninguna sección [*Strategy:nombre*] asociada.

#### **GNCulture | SNCulture**

Indica el código de entorno cultural que debe utilizarse para el nombre de pila y el apellido. Los valores válidos están en el rango -1:20. Si ninguna de estas entradas está presente, sus valores respectivos tomarán por omisión el valor -1.

#### **ONCulture**

Indica el código de entorno cultural que debe utilizarse para nombres de organización. Este valor está soportado, pero no se utiliza actualmente.

#### **MinScore**

Valor mínimo de puntuación de nombres para las coincidencias devueltas, en el rango 0:100. Este número es un filtro de las mejores coincidencias, clasificadas por puntuación de nombre completo. Si este valor es -1, IBM NameWorks busca una entrada de alteración temporal *minScore=* en la estrategia de búsqueda especificada y utiliza ese valor si se ha suministrado. Si no se especifica ningún valor de sustitución, se utiliza el valor predeterminado de NameHunter para las culturas dadas.

#### **MaxReplies**

Número máximo de coincidencias que deben devolverse. Este número es un filtro de las mejores coincidencias, clasificadas por puntuación de nombre completo. Si este valor es -1, IBM NameWorks busca una entrada de alteración temporal *MaxReplies=* en la estrategia de búsqueda especificada y utiliza ese valor si se ha suministrado. Si no se ha especificado un valor de alteración temporal, el número de coincidencias no está limitado.

#### **SearchOpt**

Especifica el tipo de lista de nombres en la que debe buscarse. Si este valor es 0, IBM NameWorks busca en todas las listas de nombres (*SearchOpt=3*).

- 1 = Buscar sólo en la lista de nombres de persona
- 2 = Buscar sólo en la lista de nombres de organización
- 3 = Buscar en las listas de nombres de persona y de organización

#### **IncludeTAQs=**

Entrada única que indica si los valores de fórmulas de tratamiento y calificadores deben incluirse con las entradas de lista de datos. Puede especificar este valor en el archivo de configuración de IBM NameWorks para aplicarlo a las entradas de lista de datos durante las operaciones de adición, actualización y búsqueda especificando *IncludeTAQs=true*.

**Nota:** Si se incluye la información TAQ, las fórmulas de tratamiento se añaden al campo de nombre especificado y los calificadores se añaden al campo de apellido antes de que se utilicen los datos de nombre en las operaciones de comparación de comparación por pares o búsqueda.

### Index

Indica si debe utilizarse el índice de NameHunter al realizar búsquedas. Los valores válidos son Y o N. Si esta entrada no está presente, el valor predeterminado es Y.

[GNParms:*nombre*] | [SNParms:*nombre*] | [ONParms:*nombre*]

## Sección Date compare del archivo de configuración

Puede alterar temporalmente los valores de puntuación de diversas pruebas de comparación de fechas en una sección [DateCompare].

Las alteraciones temporales pueden aparecer en el formato siguiente:

[DateCompare]  
clave=valor

### key

Valor de nombre que se toma de la tabla siguiente. Los valores de clave no válidos se ignoran.

### value

Los valores válidos deben ser un número en el rango 1:100. Los valores externos a este rango se ignoran.

Tabla 18. Alteraciones temporales de Date Compare y descripciones

Clave	Descripción	Puntuación predeterminada
ESS	A=A, M&D transpuestos	99
EET	A=A, M=M, dígitos de D transpuestos	98
ETE	A=A, dígitos de M transpuestos, D=D	97
ETT	A=A, dígitos de M transpuestos, dígitos de D transpuestos	96
E EI	A=A, M=M, D se ignora	95
EIE	A=A, M se ignora, D=D	94
ETI	A=A, dígitos de M transpuestos, D se ignora	93
EIT	A=A, M se ignora, dígitos de D transpuestos	92
VEE	A +/-5, M=M, D=D	91
VET	A +/-5, M=M, dígitos de D transpuestos	90
VTT	A +/-5, dígitos de M transpuestos, dígitos de D transpuestos	89
EII	A=A, M se ignora, D se ignora	88
TEE	Dígitos de A transpuestos, M=M, D=D	87
TET	Dígitos de A transpuestos, M=M, dígitos de D transpuestos	86
TTE	Dígitos de A transpuestos, dígitos de M transpuestos, D=D	85
TTT	Dígitos de A transpuestos, dígitos de M transpuestos, dígitos de D transpuestos	84

Tabla 18. Alteraciones temporales de Date Compare y descripciones (continuación)

Clave	Descripción	Puntuación predeterminada
TII	Dígitos de A transpuestos, M se ignora, D se ignora x	83
VII	A +/-5, M se ignora, D se ignora	82
XEE	A +/-10, M=M, D=D	81
XET	A +/-10, M=M, dígitos de D transpuestos x	80
XTE	A +/-10, dígitos de M transpuestos, D=D	79
XTT	A +/-10, dígitos de M transpuestos, dígitos de D transpuestos	78
XII	A +/-10, M se ignora, D se ignora	77
OB1	fecha +/- 1 día	76
OB2	fecha +/- 2 días	75
OB3	fecha +/- 3 días	74
OB4	fecha +/- 4 días	73
OB5	fecha +/- 5 días	72

## Sección Transliteration modules del archivo de configuración

La sección *Transliteration Modules* (módulos de transliteración) indica los módulos de transliteración que están instalados.

Los módulos de transliteración instalados se indican en la sección [Transliteration Modules], con este formato:

[Transliteration Modules]

Module=nombre de vía de acceso del módulo de transliteración

Los siguientes módulos de transliteración pueden utilizarse con IBM InfoSphere Global Name Management :

- arabicTransRule.ibm
- chineseTransRule.ibm
- cyrillicTransRule.ibm
- greekTransRule.ibm
- japaneseTransRule.ibm
- koreanTransRule.ibm
- latinTransRule.ibm
- chineseOnTransRule.ibm
- cyrillicOnTransRule.ibm
- hindiOnTransRule.ibm
- japaneseOnTransRule.ibm
- koreanOnTransRule.ibm
- anyTransRule.ibm\*

**Nota:** anyTransRule.ibm es un módulo de transliteración que se ha añadido como último recurso para casos excepcionales. La utilización de anyTransRule.ibm evita excepciones cuando se utilizan scripts no admitidos y no se recomienda para instalaciones típicas.

Los módulos de transliteración se deben especificar en el archivo de configuración en el orden siguiente para garantizar que las reglas de un módulo no interfieran con las de otro:

1. Archivos de transliteración de nombres de personas
2. Archivos de transliteración de nombres de organizaciones
3. anyTransRule.ibm

## Sección Reference files del archivo de configuración

Las ubicaciones de los archivos de datos de referencia se indican en la sección [Reference Files].

En la sección Reference Files del archivo de configuración se admiten las entradas siguientes:

```
[Reference Files]
NameSifter=<nombre vía acceso>
CustomCultures=<nombre vía acceso>
```

### nombre vía acceso

El nombre completo de la vía de acceso a la lista de archivos de reglas de NameSifter, delimitados por signos de punto y coma (;). También se admiten los signos de dos puntos (:) en máquinas Unix.

Nombre completo de la vía de acceso a cualquier archivo de reglas .cc (Custom Cultures o culturas personalizadas). Si utiliza un directorio que tiene otros tipos de archivos, incluya el nombre de archivo en la vía de acceso. Por ejemplo: CustomerCultures=<nombre de vía de acceso>/italian.cc

## Sección Comparison Files (archivos de comparación) del archivo de configuración

Las ubicaciones de los archivos de soporte de *NameHunter* se indican en la sección [Comparison Files].

Los parámetros siguientes son aplicables al realizar una comparación por pares con el método compare(). Esta sección es opcional, pero puede utilizar ONREG y PNREG como atributos si desea suministrar nombres de archivo específicos. Cada uno de estos formatos está soportado en la sección [Comparison Files] del archivo de configuración de *NameWorks*.

```
[Comparison Files]
TAQ=taq.ibm
GNV=gnv.ibm
SNV=snv.ibm
ONV=onv.ibm
ONTERM=terms.ibm
PNREG=angloRegRule.ibm,Anglo
PNREG=swasianRegRule.ibm,Arabic
PNREG=chineseRegRule.ibm,Chinese
PNREG=frenchRegRule.ibm,French
PNREG=germanRegRule.ibm,German
PNREG=hispanicRegRule.ibm,Hispanic
PNREG=indianRegRule.ibm,Indian
PNREG=koreanRegRule.ibm,Korean
PNREG=swasianRegRule.ibm,Pakistani
PNREG=polishRegRule.ibm,Polish
PNREG=portugueseRegRule.ibm,Portuguese
PNREG=russianRegRule.ibm,Russian
PNREG=thaiRegRule.ibm,Thai
PNREG=turkishRegRule.ibm,Turkish
PNREG=swasianRegRule.ibm,SouthwestAsian
ONREG=genericOnRegRule.ibm,Ambiguous
ONREG=angloOnRegRule.ibm,Anglo
```

ONREG=chineseOnRegRule.ibm,Chinese  
ONREG=hispanicOnRegRule.ibm,Hispanic  
ONREG=koreanOnRegRule.ibm,Korean  
ONREG=polishOnRegRule.ibm,Polish  
ONREG=portugueseOnRegRule.ibm,Portuguese  
ONREG=russianOnRegRule.ibm,Russian  
NativeTaq=ctaq.ibm  
NativePnVar=cnv.ibm  
NativeOnVar=conv.ibm  
NativeOnReg=chineseOnRegRule.ibm,Hanzi  
NativeOnReg=japaneseOnRegRule.ibm,Kanji

**TAQ=nombre\_vía\_acceso\_archivo**

Nombre de vía de acceso de la lista de TAQ.

**GNV=nombre\_vía\_acceso\_archivo**

Nombre de vía de acceso de la lista de variantes de nombre de pila.

**SNV=nombre\_vía\_acceso\_archivo**

Nombre de vía de acceso de la lista de variantes de apellidos.

**ONV=nombre\_vía\_acceso\_archivo**

El nombre de la vía de acceso de la lista de variantes de nombres de organizaciones.

**ONTERM=nombre\_vía\_acceso\_archivo**

Nombre de vía de acceso de la lista de términos.

**ONREG=nombre\_vía\_acceso\_archivo,nombre\_entorno\_cultural**

Nombre de vía de acceso del archivo de reglas de regularización y nombre del entorno cultural. Los nombres de los entornos culturales válidos figuran a continuación.

**PNREG=nombre\_vía\_acceso\_archivo,nombre\_entorno\_cultural**

Nombre de vía de acceso del archivo de reglas de regularización y nombre del entorno cultural. Los nombres de los entornos culturales válidos figuran a continuación.

**NativeTaq=nombre\_vía\_acceso\_archivo**

Nombre de la vía de acceso de la lista de TAQ de script nativo.

**NativePnVar=nombre\_vía\_acceso\_archivo**

El nombre de la vía de acceso de la lista de variantes de nombres de personas de script nativo.

**NativeOnVar=nombre\_vía\_acceso\_archivo**

El nombre de la vía de acceso de la lista de variantes de nombres de organizaciones de script nativo.

**NativePnReg=nombre\_vía\_acceso\_archivo,nombre\_script**

Nombre de la vía de acceso de un archivo de reglas de regularización de nombre de persona de script nativo y el nombre asociado del script. Los nombres válidos de script aparecen a continuación.

**NativeOnReg=nombre\_vía\_acceso\_archivo,nombre\_script**

Nombre de la vía de acceso de un archivo de reglas de regularización de nombre de persona de script nativo y el nombre asociado del script. Los nombres válidos de script aparecen a continuación.

Los nombres de entornos culturales pueden ser uno de los valores siguientes:

- Afgano
- Anglosajón
- Árabe

- Chino
- Farsi
- Francés
- Genérico (o ambiguo)
- Alemán
- Hispano
- Hindú
- Indonesio
- Japonés
- Coreano
- Paquistaní
- Polaco
- Portugués
- Ruso
- Tailandés
- Turco
- Vietnamita
- Yoruba
- -----
- Europeo
- Han
- Sudoeste asiático
- -----
- Un identificador de entorno cultural personalizado (Custom1..Custom20)

Los nombres de script pueden ser uno de los valores siguientes:

- Hanzi
- Kanji
- Devanagari
- Cirílico
- Latín
- Hangul
- Árabe
- Griego

## **Especificación de los valores de configuración mediante la clase Configuration**

Puede establecer y actualizar la información de configuración de IBM NameWorks escribiendo un programa que utilice la clase Configuration. El uso de esta clase le permite especificar valores de configuración predeterminados y especificar dinámicamente alteraciones temporales mediante la clase CompParmsOverrides.

### **Acerca de esta tarea**

Esta tarea la realiza normalmente un desarrollador de programa que trabaja directamente con la API de IBM NameWorks. Consulte la documentación de

referencia de la API para obtener más información acerca de cómo utilizar la clase Configuration, sus constructores y otros métodos relacionados.

**Nota:** El programa puede variar en función de las necesidades de la aplicación cliente y cómo se utilicen los métodos de la clase Configuration. La información aquí contenido se ha diseñado para proporcionar una descripción de alto nivel de los pasos que son necesarios para utilizar la clase Configuration en lugar del archivo de configuración de IBM NameWorks. El siguiente procedimiento incluye ejemplos de código para Java, pero los pasos para crear un programa en C++ son los mismos.

## Procedimiento

1. Con la aplicación de desarrollo, cree un objeto Configuration. La siguiente línea de código crea un objeto Configuration vacío. Los pasos siguientes ilustran la información adicional que forma parte del objeto Configuration.

```
Configuration configuration = new Configuration();
```

2. Cree un objeto Datalist para especificar la lista de datos que desea utilizar para realizar la búsqueda. El ejemplo siguiente crea un objeto Configuration que contiene una lista de datos denominada *Customers* y especifica que no se deben incluir TAQ en la búsqueda. Puede especificar parámetros adicionales, como por ejemplo añadir una entrada de lista a la lista de datos, según el tipo de la búsqueda.

```
Configuration configuration = new Configuration();  
Datalist customers = configuration.addDatalist("Customers");  
customers.setIncludeTaq(false);
```

3. Cree un objeto Strategy mediante la clase Strategy. Puede llamar a esta estrategia de búsqueda en el programa para que se utilice en comparaciones por pares y búsquedas. El ejemplo siguiente crea una estrategia de búsqueda de tipo *Broad* con varias variables especificadas.

```
Strategy broad = configuration.addStrategy("Broad");  
broad.setMinScore(75)  
    .setMaxReply(1000)  
    .setSearchOptions(EnumSet.of(NameCategory.PERSONAL));
```

4. Llame a diversos métodos para especificar los datos de configuración, como el número máximo de hebras que se va a utilizar en la búsqueda y el nombre del archivo que contiene las alteraciones temporales de parámetros de comparación predeterminadas.

```
configuration.setMaxThreads(8)  
    .setDefaultCompParmsOverridesFile("compparms.config");
```

5. Pase el objeto Configuration al constructor Analytics(), al constructor Scoring(), o a ambos para crear objetos Analytics y Scoring. Una vez creados estos objetos, el objeto Configuration se puede rechazar y los cambios subsiguientes no tendrán efecto a no ser que desee utilizar el mismo objeto Configuration para crear más objetos Analytics o Scoring.

```
Analytics analytics = new Analytics(configuration);  
Scoring scoring = new Scoring(configuration);
```

## Actualizar la configuración de IBM NameWorks para utilizar archivos adicionales de reglas de transliteración

IBM NameWorks utiliza archivos de reglas para determinar cómo se transliteran los nombres. Un administrador del sistema debe modificar el archivo de configuración de IBM NameWorks para que IBM NameWorks pueda hacer uso de los archivos de reglas adicionales. La versión 6.0 añade soporte para las reglas de

transliteración específicas de los nombres de organizaciones además de los nombres de personas. Las reglas de transliteración para nombres de organizaciones incluyen "On" en el nombre de archivo.

## Procedimiento

1. Modifique el archivo de configuración de IBM NameWorks para que incluya los archivos de reglas en la cabecera [Transliteration Modules] mediante la sintaxis siguiente:

`Module= vía_acceso_completa/nombre_archivo_reglas`

donde *vía\_acceso\_completa* es la vía de acceso completa y el nombre del directorio, y */nombre\_archivo\_reglas* es el nombre específico del archivo de reglas que se debe utilizar.

Puede utilizar con IBM NameWorks los siguientes archivos de reglas de transliteración válidos:

### **arabicTransRule.ibm**

Reglas de transliteración para nombres de personas escritos en script árabe.

### **chineseTransRule.ibm**

Reglas de transliteración para nombres de personas escritos en script chino (hanzi).

### **cyrillicTransRule.ibm**

Reglas de transliteración para nombres de personas escritos en script árabe.

### **greekTransRule.ibm**

Reglas de transliteración para nombres de personas escritos en script griego.

### **hindiTransRule.ibm**

Reglas de transliteración para nombres de personas escritos en script devanagari.

### **japaneseTransRule.ibm**

Reglas de transliteración para nombres de personas escritos en script japonés (kana).

### **koreanTransRule.ibm**

Reglas de transliteración para nombres de personas escritos en script coreano (hangul).

### **chineseOnTransRule.ibm**

Reglas de transliteración para nombres de organizaciones escritos en script chino (hanzi).

### **cyrillicOnTransRule.ibm**

Reglas de transliteración para nombres de organizaciones escritos en script cirílico.

### **hindiOnTransRule.ibm**

Reglas de transliteración para nombres de organizaciones escritos en script devanagari.

### **japaneseOnTransRule.ibm**

Reglas de transliteración para nombres de organizaciones escritos en script japonés (kanji/kana).

**koreanOnTransRule.ibm**

Reglas de transliteración para nombres de organizaciones escritos en script coreano (hangul).

**anyTransRule.ibm**

Un conjunto especial de reglas de transliteración que puede procesar cualquier script pero que solamente representa un formato rudimentario, con frecuencia poco adecuado, para el análisis y la puntuación de nombres. Este conjunto de reglas de transliteración solamente se debería utilizar en último recurso, para evitar scripts no admitidos o combinaciones de scripts que provoquen errores de transliteración. Este archivo siempre debe consultar el último lugar para garantizar que solamente se utiliza como último recurso, cuando no hay ningún otro módulo de transliteración que pueda gestionar un nombre.

2. Detenga y reinicie los servidores apropiados para reinicializar IBM NameWorks, a fin de que los servidores utilicen la información del archivo de configuración actualizado. El proceso de instalación del servicio web de IBM NameWorks incluye mandatos de detención y inicio para estas operaciones.

**Ejemplo**

Por ejemplo, para incluir el archivo de reglas de nombres de personas del árabe situado en el directorio \NW de la unidad C:\, actualizaría el archivo de configuración de esta manera:

```
[Transliteration Modules]  
Module=C:\NW\arabicTransRule.ibm
```

---

## Capítulo 7. Resolución de problemas y soporte

Para detectar y solucionar problemas relacionados con la instalación de IBM InfoSphere Global Name Management , utilice la información de soporte y resolución de problemas para identificar la causa de un problema, cómo recopilar información de diagnóstico, dónde obtener arreglos y en qué bases de conocimientos hay que realizar búsquedas.

Si tiene que ponerse en contacto con el centro de soporte de IBM, utilice esta información para reunir la información de diagnóstico que necesita el servicio técnico para poderle a ayudar a resolver un problema.

---

### Lista de comprobación para la resolución de problemas de IBM InfoSphere Global Name Management

Al responder a una serie de preguntas que están estructuradas en una lista de comprobación, se puede identificar la causa de un problema y averiguar por sí mismo una solución al problema.

Responder a las siguientes preguntas puede ayudar a identificar el origen de un problema que se está produciendo en IBM InfoSphere Global Name Management :

1. ¿Está soportada la configuración? Consulte el tema sobre los requisitos del sistema para garantizar que el sistema cumple todos los requisitos de hardware, sistema operativo y software.
2. ¿Ha aplicado los últimos arreglos en IBM InfoSphere Global Name Management ? Visite IBM InfoSphere Global Name Management Support Portal.
3. ¿Qué hace cuándo ocurre el problema?

#### **¿El programa de instalación del producto le solicita el CD o el DVD incorrectos?**

En caso afirmativo, compruebe para verificar que la unidad de disco local contiene la estructura de directorio completa. Para que el programa de instalación funcione, la unidad de disco duro local debe contener la estructura de directorio completa hasta el directorio Disk1. Si el único programa de instalación está en la unidad de disco duro local, copie la estructura de directorio completa desde el CD o el DVD.

#### **¿El programa de instalación le informó de que uno o más componentes no se instalaron satisfactoriamente?**

En caso afirmativo, examine los archivos de registro de la instalación para corregir el problema. Luego, utilice el programa de instalación para volver a instalar dichos componentes.

#### **¿Está iniciando IBM NameWorks y ha recibido el mensaje de error GODW033E?**

En caso afirmativo, este mensaje de error indica que uno o varios registros de nombres del archivo de datos de nombre no son válidos porque se han codificado incorrectamente. El mensaje de error contiene el archivo de datos problemáticos. Intente guardar los archivos de datos en el formato UTF8 y, a continuación, reinicie IBM NameWorks.

### ¿Faltan datos de nombres nuevos cuando cierra el producto o el componente?

#### ¿Está utilizando Distributed Search?

En caso afirmativo, ¿estaba añadiendo nombres a la lista de búsqueda mientras ejecutaba el producto o el componente? Si utiliza varios servidores para procesar nombres Distributed Search, compruebe el archivo `ds.config` para asegurarse de que los siguientes valores están configurados:

- Debajo de la sección **[commgr]**, marque el número de buscadores (`numSearchers=n`). Asegúrese de que *n* es igual al número de listas de búsqueda más un buscador Add dedicado.
- Asegúrese de que al menos uno de los buscadores configurados está dedicado a las incorporaciones (al menos un buscador está establecido en `doAdds=true`).

Si cambia estos valores, reinicie los servidores de Distributed Search. Consulte *Archivo de configuración y valores de Distributed Search* en el Information Center para obtener más información.

#### ¿Ha recibido el mensaje de error GODS054E?

En caso afirmativo, la causa común de que reciba este mensaje de error es que falta el nombre de archivo de datos NameSifter o la vía de acceso de datos especificada no es válida en el archivo de configuración de IBM NameWorks. Examine el archivo de configuración debajo de la sección **[Reference Files]** y compruebe el valor configurado para el valor `NameSifter=`. Si el nombre de archivo y la vía de acceso no son correctos, puede que el problema sea que el tiempo de ejecución Java no deja suficiente memoria para cargar los datos NameSifter. Si cree que este error es que no hay suficiente memoria, utilice `-Xmx128m` o un parámetro similar para que haya suficiente memoria.

Si estaba intentando desactivar NameSifter, elimine el nombre de archivo y la sentencia de vía de acceso en el valor `NameSifter=`. Este valor se encuentra en el archivo de configuración IBM NameWorks debajo de la sección **[Reference Files]**. Un ejemplo de por qué sería conveniente desactivar NameSifter es que los datos del nombre no contienen nombres de organización, por lo que no es necesario que IBM NameWorks categorice los datos de nombres por personas u organizaciones. Además, debe especificar esta entrada si ya sabe que la categoría del nombre es un nombre de persona o un nombre de organización.

4. ¿Ha examinado los archivos de registro del componente para ver si contienen algún mensaje sobre el problema?
5. ¿Qué sucedería si se emitieran alguno de los mensajes de error o códigos de error? Consulte los *mensajes y la información de códigos de error* en el centro de información para obtener más información.
6. ¿Ha examinado las bases de conocimientos del producto para obtener información que pudiera resolver el problema?
7. Si ha probado cada una de estas opciones aplicables y el problema persiste, póngase en contacto con el soporte de software de IBM.

---

## Códigos de error de componente de la API para C++

Se devuelven códigos de error numéricos cuando IBM InfoSphere Global Name Management encuentra un error. Si se produce un error, consulte la documentación de IBM InfoSphere Global Name Management para conocer el número de código de error y así obtener información sobre el tipo de error, el lugar donde se produjo y cómo corregirlo.

### Sintaxis del código de error

La clase `exception`, `ibmgnr::Exception`, se utiliza para notificar información sobre errores. Los errores se dividen en tres categorías, errores de datos, errores de entrada y errores internos, que pueden utilizarse para diferenciar entre la causa y la gravedad del error. Esta clase incluye varios métodos que pueden ser utilizados por aplicaciones clientes. Se puede recuperar información de texto invocando el método `ibmgnr::Exception::text()`, y se pueden recuperar valores enteros invocando el método `ibmgnr::Exception::value()`. El ejemplo siguiente muestra el aspecto que podría tener una cláusula `catch` básica para este tipo de excepción.

```
catch (ibmgnr::Exception & e)
{
    ibmgnr::Exception Type type = e.type(),
    char component = e.component(),
    int code = e.code(),
    std::string text = e.text(),
    reportComplexError(type, component, code, text),
};
```

#### `ExceptionType type()`

Este enumerador describe el tipo de error que se ha encontrado. Se pueden devolver tres valores diferentes para `type()`:

##### **Interno**

Error interno, no se puede continuar.

##### **Datos de referencia**

Corrupción de datos de referencia, no se puede continuar.

##### **Entrada**

Datos de entrada no válidos.

#### `component()`

Devuelve un código, formado por una sola letra, que identifica el componente de IBM InfoSphere Global Name Management donde se produjo el error. Los valores válidos para la función `component()` son los siguientes:

<b>Identificador</b>	<b>Componente</b>
A	Country of Association
C	NameClassifier
D	Distributed Search
H	NameHunter
I	Global error <sup>1</sup>
N	NameClassifier Country of Association
P	NameParser
S	NameSifter
T	NameTransliterator
V	NameVariantGenerator
W	NameWorks

**Identificador**

1

**Componente**

Los errores globales (001–006) se muestran como GODInnnE, donde *nnn* es el código numérico devuelto. La letra *I* denota un error global, que puede ser notificado por un componente cualquiera. Por ejemplo, el error GODH002E significa que se ha producido un error de apertura de archivo (002) en NameHunter (H). Este mismo error se puede producir en otro componente, tal como NameParser, donde el error se mostraría como GODP002E. Cuando consulte la documentación para los errores 001–006, examine el código, formado por una sola letra, que precede al código de error numérico para identificar el componente donde se produjo el error.

**code()**

Devuelve el código de error que está asociado a una condición de error determinada.

**const throw()**

Devuelve la información de texto asociada que puede acompañar a un error.

**value() const**

Devuelve el valor entero que puede estar asociado a un error.

**id() const throw() | wid() const throw()**

Devuelve una serie de caracteres, con el formato GODcnnnE, que identifica la condición de error.

**GOD**

Prefijo de identificación de errores, formado por tres letras, que se asigna a productos de IBM InfoSphere Global Name Management .

**c** Identificador de componentes, formado por un solo carácter, que es devuelto por la función `char component()`.

**nnn**

Código de error numérico que es devuelto por la función `code()`.

**E** Indicador estándar de IBM para mensajes de error.

**const throw()**

Devuelve una serie de caracteres que incluye el identificador de la condición de error así como el valor entero y la información de texto que puedan estar asociados, separados por un espacio en blanco.

## Códigos de error de datos de referencia

Los errores de datos de referencia indican que existe corrupción en un archivo de datos de referencia. Si este error se repite, significa que la aplicación cliente está sobrescribiendo estructuras de datos internas de IBM InfoSphere Global Name Recognition .

---

### GODD901E Configuración no válida o datos no válidos

**Explicación:** Existen datos no válidos en el archivo de configuración de NamePreprocessor, `npp.config`. Este error se debe normalmente a un intento de regularizar una lista de datos no exclusiva. La regularización no puede producirse sin la creación de una lista de datos exclusiva.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Lea el mensaje de error para determinar el problema y, a continuación, abra el archivo `npp.config` y solucione la entrada de datos no válida.

---

### GODS051E Error de datos de NameSifter.

**Explicación:** Falta uno de los archivos de reglas especificados o se ha producido un error de E/S al intentar abrir el archivo de reglas.

**Acción del Sistema:** Los módulos GNR lanzan la

excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Asegúrese de que el archivo de reglas existe y de que el nombre de archivo se ha escrito correctamente.

#### GODS052E Error de datos de NameSifter.

**Explicación:** Uno de los archivos de reglas especificados contiene un error de sintaxis.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Corrija el error de sintaxis del archivo de reglas y pase de nuevo el nombre a `NameSifter`.

#### GODS053E Error de datos de NameSifter.

**Explicación:** El interlocutor del constructor de `NameSifter` ha especificado un nombre de lista de reglas principales que no existe realmente entre los archivos de reglas especificados.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Asegúrese de que el nombre de lista de reglas principales existe en el archivo de reglas.

#### GODS054E Error de datos de NameSifter.

**Explicación:** Falta el archivo `NameAnalyzer.dat` o se ha producido un error de E/S al intentar leer el archivo.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Asegúrese de que el archivo `NameAnalyzer.dat` existe y que el nombre de vía de acceso es correcto. Si el problema persiste, obtenga una copia nueva del archivo afectado.

#### GODS055E Error de datos de NameSifter.

**Explicación:** El archivo `NameAnalyzer.dat` está dañado de algún modo.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Obtenga una copia del archivo `NameAnalyzer.dat`.

#### GODT031E No se puede leer el archivo de reglas.

**Explicación:** Un archivo de reglas de transliteración (`xxxTransRule.ibm`) o de regularización (`xxxRegRules.ibm`) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Asegúrese de que el archivo se encuentra en la ubicación correcta y que el nombre de vía de acceso es correcto. Coloque el archivo en la ubicación correcta o corrija el nombre de vía de acceso incorrecto.

#### GODT032E La cabecera no incluye la propiedad module-name.

**Explicación:** La propiedad `module-name` no se ha incluido en el archivo de cabecera. Este error indica generalmente que un archivo de reglas de transliteración (`xxxTransRule.ibm`) o de regularización (`xxxRegRules.ibm`) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

#### GODT033E Error de lectura de archivo después de reglas de transliteración.

**Explicación:** Un archivo de reglas de transliteración (`xxxTransRule.ibm`) o de regularización (`xxxRegRules.ibm`) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

#### GODT034E Faltan reglas de Transliterator.

**Explicación:** Faltan las reglas de `Transliterator`. Este error indica generalmente que un archivo de reglas de transliteración (`xxxTransRule.ibm`) o de regularización (`xxxRegRules.ibm`) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT035E La última regla tiene un nombre diferente del especificado en el atributo Transliterator-ID.**

**Explicación:** Un archivo de reglas de transliteración (xxxTransRule.ibm) o de regularización (xxxRegRules.ibm) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgmr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT036E No hay ningún signo de dos puntos en la línea de cabecera.**

**Explicación:** Este error indica generalmente que un archivo de reglas de transliteración (xxxTransRule.ibm) o de regularización (xxxRegRules.ibm) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgmr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT037E Nombre de propiedad vacío.**

**Explicación:** El nombre de la propiedad está vacío. Este error indica generalmente que un archivo de reglas de transliteración (xxxTransRule.ibm) o de regularización (xxxRegRules.ibm) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgmr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT038E Valor de propiedad vacío.**

**Explicación:** El valor de la propiedad está vacío. Este error indica generalmente que un archivo de reglas de transliteración (xxxTransRule.ibm) o de regularización

(xxxRegRules.ibm) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgmr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo.

---

**GODT039E Valor de propiedad no válido.**

**Explicación:** Un archivo de reglas de transliteración (xxxTransRule.ibm) o de regularización (xxxRegRules.ibm) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgmr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT040E Nombre de propiedad desconocido.**

**Explicación:** Un archivo de reglas de transliteración (xxxTransRule.ibm) o de regularización (xxxRegRules.ibm) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgmr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT041E Tipo de conjunto de reglas desconocido.**

**Explicación:** El tipo de conjunto de reglas es desconocido. Este error indica generalmente que un archivo de reglas de transliteración (xxxTransRule.ibm) o de regularización (xxxRegRules.ibm) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgmr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT042E No se ha especificado el ID de transcriptor.**

**Explicación:** Un archivo de reglas de transliteración (xxxTransRule.ibm) o de regularización (xxxRegRules.ibm) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT043E Faltan campos en la regla de la lista de búsqueda.**

**Explicación:** Este error indica generalmente que un archivo de reglas de transliteración (xxxTransRule.ibm) o de regularización (xxxRegRules.ibm) está dañado. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT045E Se han encontrado caracteres UTF-8 no válidos.**

**Explicación:** La serie de caracteres de entrada de NameTransliterator contiene caracteres UTF-8 no válidos. Este error indica generalmente que la serie de caracteres de entrada tiene una codificación que no es UTF-8. Esta excepción sólo se notifica cuando algún componente llama a NT para leer un archivo habiendo pasado un nombre de archivo para cargar.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Elimine los caracteres no válidos e intente de nuevo pasar la serie de caracteres de entrada a NameTransliterator. La serie de caracteres de entrada y la posición de los caracteres aparecen en los campos de texto y valor de `ibmgnr::Exception`, y también en la serie `what()`.

---

**GODT046E ID de codificación desconocido.**

**Explicación:** Indica que un interlocutor ha pasado un ID de codificación desconocido a NameTransliterator.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT047E ID de transcriptor desconocido.**

**Explicación:** Se ha pasado un ID de transcriptor no válido a NameTransliterator.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo afectado.

---

**GODT048E No se puede decodificar la entrada.**

**Explicación:** Indica que una serie de entrada no está en la codificación de caracteres especificada.

**Acción del Sistema:** En la excepción se indica el número de línea del archivo de transliteración, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Asegúrese de pasar una serie de caracteres en el formato de codificación esperado por NameTransliterator. Si no ha especificado la codificación, ésta debe coincidir con la del valor predeterminado de la plataforma.

---

**GODT049E No es posible transliterar la entrada.**

**Explicación:** Uno o varios caracteres de un nombre de entrada no pueden transliterarse porque están fuera del rango de caracteres soportados para el módulo de transliteración seleccionado.

**Acción del Sistema:** En la excepción se indica la posición del primer carácter no soportado, que puede recuperarse mediante el método `ibmgnr::Exception::value()`.

**Respuesta del Usuario:** Cargue el módulo de transliteración si aún no lo ha hecho. Elimine también los caracteres no soportados que puedan existir en el nombre de entrada.

## Códigos de error global

Los errores globales se pueden producir en diversos componentes y no son necesariamente específicos de un aspecto cualquiera del análisis de nombres.

Los errores globales (001–006) se muestran como GODI $nnn$ E, donde  $nnn$  es el código numérico devuelto. La letra  $I$  denota un error global, que puede ser notificado por un componente cualquiera. Por ejemplo, el error GODH002E significa que se ha producido un error de apertura de archivo (002) en NameHunter (H). Este mismo error se puede producir en otro componente, tal como NameParser, donde el error se mostraría como GODP002E. Cuando consulte la documentación para los errores 001–006, examine el código, formado por una sola letra, que precede al código de error numérico para identificar el componente donde se produjo el error.

---

### GODI001E Error de aserción

**Explicación:** Se ha producido un error interno al intentar clasificar un nombre.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Reinicie la aplicación. Póngase en contacto con IBM Product Support si el error persiste después de varios reinicios de la aplicación. Asegúrese de capturar cuándo y dónde se ha producido el error, así como la acción que estaba intentando realizar.

---

### GODI002E No se puede abrir el archivo.

**Explicación:** El archivo especificado no ha podido abrirse desde una de las funciones siguientes, posiblemente porque no existe:

- `ConfigHandler::load()`
- `NameHunter::loadFieldVariants()`
- `NameHunter::loadTags()`
- `NameHunter::loadRegRules()`
- `NameHunter::loadTransRules()`
- `NameHunter::loadVariants()`

**Acción del Sistema:** Los mensajes de error se devuelven mediante una de las funciones de NameHunter. Por ejemplo, la función `NameHunter::fieldVariantError()` devuelve una descripción si la función `addFieldVariant` devuelve `false`.

**Respuesta del Usuario:** Compruebe que el archivo existe y que tiene los permisos adecuados para acceder al mismo.

---

### GODI003E Error de análisis interno.

**Explicación:** Indica que se ha producido un error interno. En las excepciones que notifican este error, el valor `text()` contiene una palabra clave que puede utilizarse para facilitar el diagnóstico de la causa del problema.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Reinicie la aplicación. Póngase en contacto con IBM Product Support si el error persiste después de varios reinicios de la aplicación. Asegúrese de capturar cuándo y dónde se ha producido el error, la palabra clave asociada con el error y la acción que estaba intentando realizar.

---

### GODI004E Error de método interno.

**Explicación:** Se ha producido un error interno al intentar clasificar un nombre.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Reinicie la aplicación. Póngase en contacto con IBM Product Support si el error persiste después de varios reinicios de la aplicación. Asegúrese de capturar cuándo y dónde se ha producido el error, así como la acción que estaba intentando realizar.

---

### GODI005E Falta el archivo NameAnalyzer.dat.

**Explicación:** El archivo `NameAnalyzer.dat` no se ha encontrado en la ubicación especificada.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Asegúrese de que el archivo se encuentra en la ubicación correcta y que el nombre de vía de acceso es correcto. Coloque el archivo en la ubicación correcta o corrija el nombre de vía de acceso incorrecto.

---

### GODI006E Se ha especificado una vía de acceso del archivo NameAnalyzer.dat diferente.

**Explicación:** Un componente ha utilizado un nombre de archivo para crear un objeto y se ha utilizado un nombre diferente para crear un segundo objeto, mientras que el primero seguía activo. Los componentes siguientes pueden generar posibles objetos:

- NameParser
- Country of Association (COA)
- COA de NameClassifier
- NameVariantGenerator

**Acción del Sistema:** Los módulos GNR lanzan la

## Códigos de error de entrada

Los errores de entrada indican que la serie de caracteres de entrada está en un formato que no se puede leer. Normalmente este error se produce como consecuencia de una secuencia de caracteres UTF-8 incorrecta. Debe corregir la serie de caracteres de entrada para que el proceso pueda continuar.

---

### GODH008E Después del análisis, GN y SN están en blanco.

**Explicación:** Se ha pasado un nombre vacío a la función SearchList::add().

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgmr::Exception.

**Respuesta del Usuario:** Especifique un nombre válido para la entrada que debe añadirse a la lista de datos y vuelva a intentar la función de adición (add).

---

### GODH009E El primer nombre está en blanco.

**Explicación:** La función NameHunter::addVariant() ha detectado que la primera cadena de caracteres del par de variantes está en blanco.

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgmr::Exception.

**Respuesta del Usuario:** Compruebe los siguientes archivos de variantes para localizar la cadena de caracteres en blanco.

- ibmGnvFile
- ibmSnvFile
- ibmBnvFile
- ibmFieldVarFile

A continuación, debe especificar un nombre para la primera cadena de caracteres del par de variantes.

---

### GODH010E El segundo nombre está en blanco.

**Explicación:** La función NameHunter::addVariant() ha detectado que la segunda cadena de caracteres del par de variantes está en blanco.

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgmr::Exception.

**Respuesta del Usuario:** Compruebe los siguientes archivos de variantes para localizar la cadena de caracteres en blanco.

- ibmGnvFile
- ibmSnvFile
- ibmBnvFile

excepción ibmgmr::Exception.

**Respuesta del Usuario:** Asegúrese de que el nombre del archivo NameAnalyzer.dat sea correcto y de no haber pasado otro nombre de archivo a un componente diferente.

- ibmFieldVarFile

A continuación, debe especificar un nombre para la segunda cadena de caracteres del par de variantes.

---

### GODH011E Nombre de grupo no válido.

**Explicación:** Una entrada de variante de campo contiene un campo de texto vacío.

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgmr::Exception.

**Respuesta del Usuario:** Obtenga una copia nueva del archivo en el que se ha producido el error.

---

### GODH012E Puntuación no válida (debe estar entre 0 y 1).

**Explicación:** La función NameHunter::addVariant() ha encontrado una puntuación de variante no válida.

**Acción del Sistema:** Este código de error registra el número de línea y la información de error, que puede extraerse llamando al método NameHunter::ConfigHandler::errorList().

**Respuesta del Usuario:** Compruebe los siguientes archivos de variantes para localizar la cadena de caracteres en blanco.

- ibmGnvFile
- ibmSnvFile
- ibmBnvFile
- ibmFieldVarFile

Obtenga una copia nueva del archivo en el que se ha producido el error.

---

### GODH013E Entrada duplicada.

**Explicación:** Una de las funciones siguientes ha encontrado una entrada duplicada:

- ConfigHandler::load()
- NameHunter::addTaq()
- NameHunter::loadTaq()
- NameHunter::addVariants()

- NameHunter::loadVariants()

**Acción del Sistema:** Este código de error registra el número de línea y la información de error, que puede extraerse llamando al método NameHunter::ConfigHandler::errorList().

**Respuesta del Usuario:** Obtenga una copia nueva del archivo en el que se ha producido el error.

#### **GODH014E Tipo de campo desconocido.**

**Explicación:** Se ha pasado un tipo de campo desconocido (no un nombre de pila ni un apellido) a las estructuras de datos de CompParms (setDefault() o setParmsDefault()).

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgnr::Exception.

**Respuesta del Usuario:** Especifique uno de los tipos de campos válidos en la enumeración NameFieldType (NameConstants.h). Los siguientes tipos de campos son válidos:

- GivenName
- SurName
- OrgName

#### **GODH015E Código de entorno cultural desconocido.**

**Explicación:** Se ha pasado un código de entorno cultural desconocido a una de las funciones siguientes:

- CompParms::setDefault()
- NameHunter::loadFieldVariants()
- NameHunter::addTaq()
- NameHunter::loadTaq()
- NameHunter::addVariant()
- NameHunter::loadVariants()

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgnr::Exception.

**Respuesta del Usuario:** Especifique uno de los códigos de entorno cultural válidos en la enumeración Culture (NameConstants.h).

#### **GODH016E Tipo de TAQ desconocido.**

**Explicación:** Se ha pasado un tipo de TAQ desconocido a NameHunter::addTaq(). Los tipos de TAQ válidos provienen del campo Tokentypes de la enumeración NameConstants.h.

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgnr::Exception.

**Respuesta del Usuario:** Especifique uno de los tipos de campos válidos.

#### **GODH017E Cabecera de configuración vacía o ausente.**

**Explicación:** La función ConfigHandler::load() no ha encontrado una cabecera válida (por ejemplo, [ParmsGnAnglo]). Corrija el archivo de configuración.

**Acción del Sistema:** Este código de error registra el número de línea y la información de error, que puede extraerse llamando al método NameHunter::ConfigHandler::errorList().

**Respuesta del Usuario:** Corrija el archivo de configuración especificando una cabecera válida.

#### **GODH018E No se ha encontrado el delimitador de valor de código (=).**

**Explicación:** La función ConfigHandler::load() ha encontrado un par de valores sin el delimitador, que generalmente es un signo igual (=).

**Acción del Sistema:** Este código de error registra el número de línea y la información de error, que puede extraerse llamando al método NameHunter::ConfigHandler::errorList().

**Respuesta del Usuario:** Corrija el archivo de configuración especificando un delimitador válido.

#### **GODH019E No se ha podido encontrar la cabecera de versión en el archivo de referencia.**

**Explicación:** Se ha especificado un archivo TAQ obsoleto o no válido para que lo cargue NameHunter. Este error indica generalmente que no se ha encontrado la cabecera de versión en el archivo de referencia (por ejemplo taq.ibm o var.ibm).

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgnr::Exception.

**Respuesta del Usuario:** Obtenga la versión más reciente del archivo TAQ (GNR versión 4.1 o posterior).

#### **GODH020E Tipo de factor TAQ no válido.**

**Explicación:** Se ha pasado un tipo de factor TAQ no válido a NameHunter. El tipo de factor TAQ debe ser 1 (diferente) o 2 (ausente).

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgnr::Exception.

**Respuesta del Usuario:** Corrija el tipo de factor TAQ no válido y especifique un valor válido.

#### **GODH021E El texto de TAQ no puede estar en blanco.**

**Explicación:** La entrada de texto de TAQ está vacía. Debe suministrarse un valor para el texto de TAQ.

**Acción del Sistema:** Los módulos GNR lanzan la excepción ibmgnr::Exception.

**Respuesta del Usuario:** Especifique un texto de TAQ válido y pase el nombre a NameHunter.

---

**GODH022E** Se ha especificado un nombre de categoría no válido.

**Explicación:** Se ha pasado una categoría de nombre no válida a NameHunter.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Especifique una categoría de nombre válida a NameHunter. Las categorías de nombre válidas son Personal y Organización.

---

**GODH065E** Texto vacío en archivo de términos

**Explicación:** Existe una entrada vacía en el archivo de términos predeterminado, `terms.ibm`. Este error puede deberse a la especificación del tipo de campo como un nombre de pila o un apellido, lo que genera la eliminación de determinados términos de organización como parte de proceso de limpieza de nombres. Las entradas en el archivo `terms.ibm` son sólo para nombres de organización.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Si desea utilizar el archivo de términos predeterminado, `terms.ibm`, debe especificar el tipo de campo de los nombres de organización en la

## Códigos de error interno

Los errores internos indican que la aplicación se ha corrompido de alguna manera. Después de producirse un error interno, debe reiniciar la aplicación para poder continuar. Normalmente este tipo de error es debido a que la aplicación cliente está sobrescribiendo estructuras de datos internas de IBM InfoSphere Global Name Recognition .

---

**GODH007E** La memoria se ha agotado.

**Explicación:** Una clase `SearchList` notifica este error si, al añadir nombres, no puede obtener memoria suficiente.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Aumente la cantidad de memoria disponible o reduzca el número de nombres cargados en la misma.

---

**GODP024E** No se ha podido crear la instancia de `NameAnalyzer`.

**Explicación:** Se ha producido un problema al intentar inicializar la biblioteca `NameAnalyzer`. Este error indica generalmente que el archivo `NameAnalyzer.dat` no se encuentra en la ubicación especificada.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

función `loadTerms()`. Por ejemplo, `loadTerms("terms.ibm",OrgName)`.

---

**GODP026E** Error al convertir desde Unicode.

**Explicación:** Se ha producido un error cuando `NameParser` ha intentado convertir una serie de caracteres desde Unicode.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Asegúrese de pasar una serie de caracteres en el formato de codificación esperado por `NameParser`. Si no ha especificado la codificación, ésta debe coincidir con la del valor predeterminado de la plataforma.

---

**GODP028E** Error al crear la serie de caracteres Unicode.

**Explicación:** `NameParser` ha recibido una serie de caracteres en un formato de codificación inesperado.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Asegúrese de pasar una serie de caracteres en el formato de codificación esperado por `NameParser`. Si no ha especificado la codificación, ésta debe coincidir con la del valor predeterminado de la plataforma.

**Respuesta del Usuario:** Asegúrese de que el nombre de vía de acceso del archivo `NameAnalyzer.dat` sea correcta. Si el nombre de vía de acceso es incorrecto, especifique el nombre correcto de la vía de acceso donde se encuentra el archivo `NameAnalyzer.dat`.

---

**GODP025E** Error al abrir el convertidor.

**Explicación:** El interlocutor ha pasado un nombre de codificación no válido a `NameParser`.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgnr::Exception`.

**Respuesta del Usuario:** Compruebe el código de la aplicación y asegúrese de que el nombre que ha pasado a `NameParser` es un nombre de conjunto de caracteres IANA (Internet Assigned Numbers Authority) correcto.

---

**GODP027E** Error al convertir a UTF-8.

**Explicación:** Se ha producido un error interno cuando

NameParser ha intentado convertir una serie de caracteres al formato UTF-8.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgmr::Exception`.

**Respuesta del Usuario:** Asegúrese de pasar una serie de caracteres en el formato de codificación esperado por NameParser. Si no ha especificado la codificación, ésta debe coincidir con la del valor predeterminado de la plataforma.

---

**GODP029E Error al crear Transliterator.**

**Explicación:** Se ha producido un error de sintaxis o de desbordamiento en la lista de filtros de palabras superfluas de NameParser.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgmr::Exception`.

**Respuesta del Usuario:** Reduzca el número de filtros de palabras superfluas incluidos en el archivo de configuración de NameWorks.

---

**GODP030E Error al crear Transliterator a partir de las reglas.**

**Explicación:** Generalmente, este mensaje indica que se ha producido un error de sintaxis en uno de los filtros de la lista de filtros de palabras superfluas de NameParser o que hay demasiados filtros en la lista de filtros.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgmr::Exception`.

**Respuesta del Usuario:** Corrija el error de sintaxis del filtro de palabras superfluas afectado. Además, puede reducir el número de filtros de palabras superfluas incluidos en el archivo de configuración de NameWorks, suponiendo que llame a NameParser mediante NameWorks.

---

**GODT044E Error de ICU.**

**Explicación:** La biblioteca ICU ha devuelto un error.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgmr::Exception`.

**Respuesta del Usuario:** Reinicie la aplicación. Póngase en contacto con IBM Product Support si el error persiste después de varios reinicios de la aplicación. Asegúrese de capturar cuándo y dónde se ha producido el error, así como la acción que estaba intentando realizar.

---

**GODV050E El análisis ha fallado.**

**Explicación:** Indica que se ha producido un error interno en NameVariantGenerator al buscar en el archivo NameAnalyzer.dat.

**Acción del Sistema:** Los módulos GNR lanzan la

excepción `ibmgmr::Exception`.

**Respuesta del Usuario:** Reinicie la aplicación. Póngase en contacto con IBM Product Support si el error persiste después de varios reinicios de la aplicación. Asegúrese de capturar cuándo y dónde se ha producido el error, así como la acción que estaba intentando realizar.

---

**GODW101E Código de entorno cultural no válido.**

**Explicación:** Se ha encontrado un código de entorno cultural no válido en IBM NameWorks. El número de código de entorno cultural sigue al mensaje de error.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgmr::Exception`.

**Respuesta del Usuario:** Reinicie la aplicación. Póngase en contacto con IBM Product Support si el error persiste después de varios reinicios de la aplicación. Asegúrese de capturar cuándo y dónde se ha producido el error, así como la acción que estaba intentando realizar.

---

**GODW102E Mapa de bits de conjunto de entorno cultural no válido.**

**Explicación:** Se ha encontrado un mapa de bits de conjunto de entorno cultural no válido en IBM NameWorks. El entero del mapa de bits sigue al mensaje de error.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgmr::Exception`.

**Respuesta del Usuario:** Reinicie la aplicación. Póngase en contacto con IBM Product Support si el error persiste después de varios reinicios de la aplicación. Asegúrese de capturar cuándo y dónde se ha producido el error, así como la acción que estaba intentando realizar.

---

**GODW103E Código de categoría de nombres no válido.**

**Explicación:** Se ha encontrado un código de categoría de nombres no válido en IBM NameWorks. El código de categoría de nombres sigue al mensaje de error.

**Acción del Sistema:** Los módulos GNR lanzan la excepción `ibmgmr::Exception`.

**Respuesta del Usuario:** Reinicie la aplicación. Póngase en contacto con IBM Product Support si el error persiste después de varios reinicios de la aplicación. Asegúrese de capturar cuándo y dónde se ha producido el error, así como la acción que estaba intentando realizar.

**GODW104E Mapa de bits de conjunto de categoría de nombres no válido.**

**Explicación:** Se ha encontrado un mapa de bits de conjunto de categoría de nombres no válido en IBM NameWorks. El entero del mapa de bits sigue al mensaje de error.

**Acción del Sistema:** Los módulos GNR lanzan la

excepción `ibmgmr::Exception`.

**Respuesta del Usuario:** Reinicie la aplicación. Póngase en contacto con IBM Product Support si el error persiste después de varios reinicios de la aplicación. Asegúrese de capturar cuándo y dónde se ha producido el error, así como la acción que estaba intentando realizar.

---

## Códigos de error de IBM NameWorks

Se devuelven códigos de error numéricos cuando IBM NameWorks encuentra un error. Si se produce un error, consulte la documentación de IBM InfoSphere Global Name Management para conocer el número de código de error y así obtener información sobre el tipo de error, el lugar donde se produjo y cómo corregirlo.

Los códigos de error se utilizan para diferenciar entre la causa y la gravedad del error. Las clases de excepción siguientes se utilizan para presentar información de error procedente de las clases Scoring y Analytics de IBM NameWorks para C++ y Java. Estas clases incluyen varios métodos que pueden ser utilizados por las aplicaciones.

*Tabla 19. Clases de excepción para C++ y Java*

Excepción de C++	Excepción de Java
<code>ibmgmr::NwException</code>	<code>java.lang.RuntimeException</code>

### Ejemplo de cláusula catch de C++

```
catch (ibmgmr::NwException & e)
{
  int code = e.code(),
  std::string text = e.text(),
  reportComplexError(type, code, text),
};
```

#### **int code() const**

Devuelve el código de error que está asociado a una condición de error determinada.

#### **const char \*text() const throw()**

Devuelve la información de texto asociada que puede acompañar a un error.

#### **const char \*what() const throw()**

Devuelve una serie de caracteres que incluye el identificador de la condición de error así como el valor entero y la información de texto que puedan estar asociados, separados por un espacio en blanco.

### Ejemplo de cláusula catch de Java

```
catch ( Throwable exception )
{
  StringWriter stackTrace = new StringWriter();
  PrintWriter printer = new PrintWriter(stackTrace);
  exception.printStackTrace(printer);
  printer.close();
  reportError(exception.getMessage(), stackTrace);
}
```

#### **reportError()**

Permite notificar diversas excepciones diferentes.

## Códigos de error de IBM NameWorks para C++

Estos códigos de error son muy específicos de errores individuales de IBM NameWorks. Una llamada a un método (como, por ejemplo Analytics::analyze()) puede devolver varios errores diferentes, dependiendo del componente de IBM InfoSphere Global Name Management donde se produjo el error.

---

**GODW001E** Se ha producido un error al cargar el archivo de configuración.

**Explicación:** Este error indica generalmente que el nombre del archivo de configuración es incorrecto.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Compruebe que el nombre de archivo sea correcto y que el archivo de configuración esté en la ubicación correcta.

---

**GODW002E** No se ha podido crear el objeto de transliteración.

**Explicación:** El objeto de transliteración no ha podido crearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

**GODW003E** Bloqueo de NameTransliterator.

**Explicación:** El objeto transcriptor no ha podido bloquearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Intente de nuevo la acción que estaba intentando realizar. Si el problema persiste, reinicie la aplicación.

---

**GODW004E** Error de NameTransliterator.

**Explicación:** Se ha producido un error de transliteración. El mensaje de error de NameTransliterator se devuelve mediante la serie `what()` (por ejemplo, `GODTnnnE`, donde `nnn` es el número de error) del objeto `ibmgnr::NwException`.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

**GODW005E** Error de construcción de NameParser.

**Explicación:** El objeto analizador no ha podido crearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

**GODW006E** Error de bloqueo de NameParser.

**Explicación:** El objeto analizador no ha podido bloquearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Intente de nuevo la acción que estaba intentando realizar. Si el problema persiste, reinicie la aplicación.

---

**GODW007E** Error de NameParser.

**Explicación:** Se ha producido un error de análisis. El mensaje de error de NameParser se devuelve mediante la serie `what()` (por ejemplo, `GODPnnnE`, donde `nnn` es el número de error) del objeto `ibmgnr::NwException`.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

**GODW008E** Cadena de caracteres personalizada no válida.

**Explicación:** La cadena o cadenas de caracteres personalizadas no son válidas.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Corrija la información de cadena de caracteres personalizada del archivo de configuración y reinicie la aplicación.

---

**GODW009E** Error de construcción de NameVariantGenerator.

**Explicación:** El objeto NameVariantGenerator no ha podido crearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

**GODW010E Error de bloqueo de NameVariantGenerator.**

**Explicación:** El objeto NameVariantGenerator no ha podido bloquearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Intente de nuevo la acción que estaba intentando realizar. Si el problema persiste, reinicie la aplicación.

---

**GODW011E Error de NameVariantGenerator.**

**Explicación:** El objeto NameVariantGenerator no ha podido crearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

**GODW012E Error de construcción de COA (Country of Association).**

**Explicación:** El objeto COA no ha podido crearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

**GODW013E Error de bloqueo de COA (Country of Association).**

**Explicación:** El objeto COA no ha podido bloquearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Intente de nuevo la acción que estaba intentando realizar. Si el problema persiste, reinicie la aplicación.

---

**GODW014E Error de COA (Country of Association).**

**Explicación:** Se ha producido un error de análisis. El mensaje de error de COA se devuelve mediante la serie `what()` (por ejemplo, `GODAnnnE`, donde `nnn` es el número de error) del objeto `ibmgnr::NwException`.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

---

**GODW015E Error de construcción de la implementación analítica.**

**Explicación:** El objeto Analytics no ha podido crearse debido a falta de memoria.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Asigne más memoria al proceso e intente de nuevo la acción que estaba intentando realizar.

---

**GODW016E Error de construcción de la implementación de puntuación.**

**Explicación:** El objeto Scoring no ha podido crearse debido a falta de memoria.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Asigne más memoria al proceso e intente de nuevo la acción que estaba intentando realizar.

---

**GODW017E Error de construcción de NameSifter.**

**Explicación:** El objeto NameSifter no ha podido crearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

**GODW018E Error de construcción de DateCompare.**

**Explicación:** El objeto DateCompare no ha podido crearse debido a falta de memoria.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Asigne más memoria al proceso e intente de nuevo la acción que estaba intentando realizar.

---

**GODW019E Error de construcción de NameHunter.**

**Explicación:** El objeto NameHunter no ha podido crearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la acción que estaba intentando realizar.

---

---

**GODW020E Error de comparación de nombres en NameHunter.**

**Explicación:** Se ha producido un error de análisis. El mensaje de error de NameHunter se devuelve mediante la serie **what()** (por ejemplo, GODHnnnE, donde *nnn* es el número de error) del objeto `ibmgnr::NwException`.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Examine el mensaje de error, corrija el error e intente de nuevo la operación de comparación.

---

**GODW021E Falta la lista de datos.**

**Explicación:** El nombre de lista de datos especificado no aparece en el archivo de configuración.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Compruebe que el nombre de lista de datos sea correcto. Corrija el archivo de configuración o el nombre de la lista de datos e intente de nuevo la acción que estaba intentando realizar.

---

**GODW022E Lista de datos con datos de servidor incorrectos.**

**Explicación:** Entrada "Server=" incorrecta para una lista de datos.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Corrija los datos de servidor en el archivo de configuración y reinicie la aplicación.

---

**GODW023E Lista de datos sin datos de servidor.**

**Explicación:** La lista de datos no contiene entradas "Server=".

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Corrija el archivo de configuración y reinicie la aplicación.

---

**GODW024E Falta estrategia de búsqueda.**

**Explicación:** El nombre de estrategia de búsqueda especificado no aparece en el archivo de configuración.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Compruebe que el nombre de la estrategia de búsqueda sea correcto. Corrija el archivo de configuración o el nombre de la estrategia de búsqueda e intente de nuevo la acción que estaba intentando realizar.

---

**GODW025E Error de construcción de buscador.**

**Explicación:** No se ha podido realizar una búsqueda debido a falta de memoria.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Asigne más memoria al proceso e intente de nuevo la acción que estaba intentando realizar.

---

**GODW026E Error de búsqueda.**

**Explicación:** El motor de búsqueda ha devuelto un error.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Un error en el mensaje indica que el proceso de Distributed Search ha devuelto información de error. De lo contrario, se ha producido un error de comunicación y el mensaje contiene el código de error. Corrija el error y reintente la operación de búsqueda.

---

**GODW027E Error de adición a lista de datos.**

**Explicación:** La operación de adición ha fallado porque no se ha especificado ningún servidor de adición (mensaje en blanco) o porque el mensaje contiene información de error desde un motor de búsqueda.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Un error en el mensaje indica que el proceso de Distributed Search ha devuelto información de error. De lo contrario, se ha producido un error de comunicación y el mensaje contiene el código de error. Corrija el error y reintente la operación de adición.

---

**GODW028E Error de actualización de lista de datos.**

**Explicación:** La operación de actualización ha fallado debido a un error de comunicaciones.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Corrija el error de comunicaciones y reintente la operación de actualización.

---

**GODW029E Error de supresión de lista de datos.**

**Explicación:** La operación de supresión ha fallado debido a un error de comunicaciones.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Corrija el error de comunicaciones y reintente la operación de supresión.

---

**GODW030E Error de extracción de lista de datos.**

**Explicación:** La operación de extracción ha fallado debido a un error de comunicaciones.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Corrija el error de comunicaciones y reintente la operación de extracción.

---

**GODW031E Valor de parámetro no válido.**

**Explicación:** El objeto `NameVariantGenerator` no ha podido crearse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`.

**Respuesta del Usuario:** Corrija el parámetro no válido e intente de nuevo la acción que estaba intentando realizar.

---

**GODW032E Archivo de datos incorrecto.**

**Explicación:** El archivo de datos de nombre no ha podido abrirse.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`. Esta excepción se notifica como `RunTimeException Java` para los servicios Web y las API Java.

**Respuesta del Usuario:** Asegúrese de que el archivo de datos de nombre existe y que el nombre de vía de acceso es correcto. Si el problema persiste, obtenga una copia nueva del archivo afectado.

---

**GODW033E Registro incorrecto.**

**Explicación:** Hay un registro de nombre no válido en el archivo de datos de nombre.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`. Esta excepción se notifica como `RunTimeException Java` para los servicios Web y las API Java.

**Respuesta del Usuario:** Corrija el registro no válido e intente de nuevo la acción que estaba intentando realizar.

---

**GODW034E Bloqueo de identificador de transacción.**

**Explicación:** Se ha producido un error interno en esta transacción.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`. Esta excepción se notifica como `RunTimeException Java` para los servicios Web y las API Java.

**Respuesta del Usuario:** Reinicie la aplicación para resolver el error.

---

**GODW035E Parámetros de comparación predeterminados incorrectos.**

**Explicación:** Se ha producido un error en el archivo de alteración temporal de parámetros de comparación (`CompParms`) predeterminado.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`. Esta excepción se notifica como `RunTimeException Java` para los servicios Web y las API Java.

**Respuesta del Usuario:**

---

**GODW036E Anomalía de bloqueo de regularización.**

**Explicación:** Se ha producido un error interno de regularización.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`. Esta excepción se notifica como `RunTimeException Java` para los servicios Web y las API Java.

**Respuesta del Usuario:** Reinicie la aplicación para resolver el error.

---

**GODW037E Nombre de parámetro de comparación no válido.**

**Explicación:** Existe un nombre de alteración temporal de parámetro de comparación (`CompParm`) no válido.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`. Esta excepción se notifica como `RunTimeException Java` para los servicios Web y las API Java.

**Respuesta del Usuario:** Corrija el nombre de parámetro de comparación no válido. La serie de texto que acompaña al mensaje de error indica el nombre del archivo donde se ha producido el error.

---

**GODW038E Valor de parámetro de comparación no válido.**

**Explicación:** Existe un valor de alteración temporal de parámetro de comparación (`CompParm`) no válido.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`. Esta excepción se notifica como `RunTimeException Java` para los servicios Web y las API Java.

**Respuesta del Usuario:** Corrija el valor de parámetro de comparación no válido. La serie de texto que acompaña al mensaje de error indica el nombre del archivo donde se ha producido el error.

**GODW039E Tipo de lista de datos no válido.**

**Explicación:** Existe un tipo de lista de datos no válido en el archivo de configuración de IBM NameWorks.

**Acción del Sistema:** Los módulos IBM NameWorks lanzan la excepción `ibmgnr::NwException`. Esta excepción se notifica como `RunTimeException` Java para los servicios Web y las API Java.

**Respuesta del Usuario:** Corrija el parámetro de tipo de lista de datos no válido. Los valores siguientes son entradas válidas para el parámetro `Type=`:

- 0 = búsqueda incorporada
- 1 = búsqueda completa
- 2 = búsqueda de nombres exclusivos

---

## Códigos de error de Distributed Search

Los errores de Distributed Search se devuelven al cliente en el código <ERROR> de la respuesta XML. Cada error notifica una clasificación de gravedad, un código de error y un mensaje.

---

**gnrds-001 No se ha podido encontrar el principio del mensaje, se descarta.**

**Explicación:** Distributed Search no ha podido encontrar el código que identifica el principio de un mensaje. El código es <NHServerMessage> o <NH\_SERVER\_MESSAGE>. Este error es una indicación de un error de programación o de comunicaciones.

**Respuesta del Usuario:** Puede intentar reenviar el mensaje anómalo, pero puede que la única solución sea reiniciar el sistema.

una indicación de un error de programación.

---

**gnrds-002 No se ha podido encontrar el final del mensaje, se descarta.**

**Explicación:** Distributed Search no ha podido encontrar el código que identifica el final de un mensaje. El código es </NHServerMessage> o </NH\_SERVER\_MESSAGE>. Este error es una indicación de un error de programación o de comunicaciones.

**Respuesta del Usuario:** Puede intentar reenviar el mensaje anómalo, pero puede que la única solución sea reiniciar el sistema.

---

**gnrds-005 No se ha encontrado el código – CÓDIGO.**

**Explicación:** Distributed Search no ha encontrado un código obligatorio (por ejemplo, `request_type`). El código que falta se mostrará en lugar de "CÓDIGO". Este error es una indicación de un error de programación.

---

**gnrds-006 Tipo de mensaje no válido – X**

**Explicación:** El valor del campo "request\_type" no es válido. El valor erróneo se mostrará en lugar de "X".

---

**gnrds-007 El valor debe ser un número – CÓDIGO=VALOR.**

**Explicación:** Se ha especificado un valor no numérico para un campo numérico. En el mensaje se mostrarán el CÓDIGO y el VALOR incorrectos.

---

**gnrds-008 El valor debe estar entre 0.0 y 1.0 – CÓDIGO=VALOR.**

**Explicación:** Un valor de escala (por ejemplo, umbral) está fuera del rango obligatorio. En el mensaje se mostrarán el CÓDIGO y el VALOR incorrectos.

---

**gnrds-003 No se ha podido determinar el tipo de mensaje.**

**Explicación:** El campo `request_type` contiene un valor no válido o Distributed Search no reconoce un mensaje de respuesta. Este error es una indicación de un error de programación o de comunicaciones.

**Respuesta del Usuario:** Puede intentar reenviar el mensaje anómalo, pero puede que la única solución sea reiniciar el sistema.

---

**gnrds-009 Código de entorno cultural no válido – CÓDIGO=VALOR.**

**Explicación:** Un código de entorno cultural está fuera del rango soportado. Consulte la tabla de códigos de entorno cultural para conocer los valores válidos. En el mensaje se mostrarán el CÓDIGO y el VALOR incorrectos.

---

**gnrds-004 No se ha encontrado la cabecera de registro – CABECERA\_REGISTRO**

**Explicación:** Distributed Search no ha encontrado una cabecera de registro obligatoria (por ejemplo, `SEARCH_NAME`). Esta cabecera que falta se mostrará en lugar de "CABECERA\_REGISTRO". Este error es

---

**gnrds-010 Valor booleano no válido – CÓDIGO=VALOR.**

**Explicación:** Un valor booleano no contiene un valor válido. Distributed Search acepta "T, TRUE, Y, YES, ON, 1" como true y "F, FALSE, N, NO, OFF, 0" como

false. Los valores son sensibles a mayúsculas y minúsculas. En el mensaje se mostrarán el CÓDIGO y el VALOR incorrectos.

---

**gnrds-011** El valor debe ser un número mayor que 0 – CÓDIGO=VALOR.

**Explicación:** Se ha especificado un número menor que 1 en un campo que requiere un número positivo. En el mensaje se mostrarán el CÓDIGO y el VALOR incorrectos.

---

**gnrds-012** El tipo de ancla debe ser 0, 1 ó 2 – CÓDIGO=VALOR.

**Explicación:** El valor de ANCHOR\_TYPE no es válido.

---

**gnrds-013** La modalidad de puntuación debe ser 0, 1 ó 2 – CÓDIGO=VALOR.

**Explicación:** El valor de SCORE\_MODE no es válido.

---

**gnrds-014** No hay buscadores configurados para dar soporte a adiciones o actualizaciones.

**Explicación:** Distributed Search ha recibido una solicitud de adición o actualización, pero no hay buscadores que den soporte a adiciones.

**Respuesta del Usuario:** Si deben admitirse las adiciones, un buscador debe tener el valor de configuración "doAdds=true".

---

**gnrds-015** Las consultas en las que falta el nombre de pila (GN) y el apellido (SN) no están permitidas.

**Explicación:** Se ha recibido una solicitud de búsqueda con un nombre de pila (GN) y un apellido (SN) en blanco, y Distributed Search se ha configurado para rechazar este tipo de consultas. El archivo de configuración contiene la entrada "allowFnuLnu=false".

---

**gnrds-016** Las consultas en las que falta el nombre de pila (GN) y con una sola inicial para el apellido (SN) no están permitidas.

**Explicación:** Se ha recibido una solicitud de búsqueda con un nombre de pila (GN) en blanco y una sola inicial para el apellido (SN), y Distributed Search se ha configurado para rechazar este tipo de consultas. El archivo de configuración contiene la entrada "allowFnuInit=false".

---

**gnrds-017** Las consultas con una sola inicial para el nombre de pila (GN) y en las que falta el apellido (SN) no están permitidas.

**Explicación:** Se ha recibido una solicitud de búsqueda

con una sola inicial para el nombre de pila (GN) y un apellido (SN) en blanco, y Distributed Search se ha configurado para rechazar este tipo de consultas. El archivo de configuración contiene la entrada "allowInitLnu=false".

---

**gnrds-018** Las consultas con una sola inicial para el nombre de pila (GN) y una sola inicial para el apellido (SN) no están permitidas.

**Explicación:** Se ha recibido una solicitud de búsqueda con una sola inicial para el nombre de pila (GN) y una sola inicial para el apellido (SN), y Distributed Search se ha configurado para rechazar este tipo de consultas. El archivo de configuración contiene la entrada "allowInitInit=false".

---

**gnrds-020** Uno o varios buscadores no responden.

**Explicación:** Uno o varios buscadores no responden y no puede crearse una respuesta completa.

**Respuesta del Usuario:** Probablemente, deberá reiniciarse Distributed Search.

---

**gnrds-021** El ID de nombre no puede estar en blanco.

**Explicación:** Las solicitudes de adición y actualización deben suministrar un ID. Cuando se devuelve este mensaje, significa que el ID está en blanco.

---

**gnrds-022** El ID de nombre para la actualización no puede estar en blanco.

**Explicación:** Las solicitudes de adición y actualización deben suministrar un ID para la actualización. Cuando se devuelve este mensaje, significa que el valor ID\_TO\_UPDATE está en blanco.

---

**gnrds-023** Una respuesta del buscador era demasiado grande para el almacenamiento intermedio de mensajes.

**Explicación:** Casi con seguridad, este error se debe a un mensaje de resultados de búsqueda que contiene demasiadas respuestas que van de un buscador a commgr.

**Respuesta del Usuario:** Las consultas que generan demasiados resultados deben evitarse; sin embargo, puede aumentar el tamaño del almacenamiento intermedio de mensajes por medio del valor msgBuffSize de ds.conf. Su valor predeterminado es de 1Mb.

---

**gnrds-024** No se ha podido analizar un mensaje.

**Explicación:** NameParser no ha podido analizar el mensaje que ha recibido.

---

**gnrds-025** No se ha podido realizar la transliteración, UTF8 no válido.

**Explicación:** NameTransliterator ha detectado un UTF8 no válido en los campos SN o GN de una consulta o mensaje de adición. Este mensaje también puede grabarse en las anotaciones de error durante el inicio y el preproceso si se detecta un UTF8 no válido en el archivo de entrada.

---

**gnrds-026** El tipo de puntuación debe ser 0 ó 1.

**Explicación:** Se ha especificado un valor que no 0 o 1 para el tipo de puntuación en el archivo compparms.config. Este error es extremadamente raro y no debería producirse en condiciones de operación normales.

**PRECAUCIÓN:**

La modificación del valor de score\_type cambia el algoritmo de puntuación utilizado por Distributed Search. No debe modificar este valor.

---

**gnrds-027** La categoría de nombre debe ser P, O o A.

**Explicación:** Se ha especificado una categoría de nombre que no es de persona (P), de organización (O) o Todos (A) para el nombre de consulta.

**Respuesta del Usuario:** Elimine el valor no válido e inserte una categoría de nombre válida.

---

**gnrds-028** El nombre no puede estar vacío.

**Explicación:** Se ha pasado un nombre de consulta vacío a Distributed Search.

---

**Respuesta del Usuario:** Especifique un nombre válido como nombre de consulta e intente de nuevo la operación de búsqueda.

---

**gnrds-029** La opción de búsqueda debe ser 1, 2 ó 3.

**Explicación:** Se ha especificado un valor que no es 1, 2 ó 3 para el parámetro SearchOpt= de la estrategia de búsqueda.

**Respuesta del Usuario:** Especifique un valor válido para el parámetro SearchOpt= de la estrategia de búsqueda.

- 1 = Buscar sólo en la lista de nombres de persona
  - 2 = Buscar sólo en la lista de nombres de organización
  - 3 = Buscar en las listas de nombres de persona y de organización
- 

**gnrds-030** El número no debe ser negativo.

**Explicación:** Este error puede producirse con cualquier número, ya que Distributed Search no admite valores negativos.

**Respuesta del Usuario:** Corrija el valor no válido especificando un valor igual o superior a cero.

---

**gnrnh-001** Después del análisis, GN y SN están en blanco.

**Explicación:** Este mensaje indica que un intento de añadir un registro ha fallado debido a que el nombre de pila y el apellido están en blanco. Puede que el usuario los haya especificado como espacios en blanco o que se hayan convertido a espacios en blanco mediante la transliteración.

---

## Códigos de error de Enterprise Name Search

Los errores de Enterprise Name Search se devuelven al cliente en el código <ERROR> de la respuesta XML, en una lista de errores en la salida JSON o en un diálogo visualizado en una GUI y se suelen registrar también en los registros del servidor de ENS. Cuando en la salida de servicio web se muestra una indicación de error, ésta incluye un código de error y posiblemente un mensaje.

### Códigos de error de la consola ENS

Estos mensajes y los códigos relacionados pueden aparecer en los diálogos de error en la GUI de la consola de ENS y donde proceda en el archivo de registro del servidor.

**CDHNC1001E**

No se ha especificado ninguna acción.

**CDHNC1002E**

Se ha producido una excepción al iniciar o detener la célula. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1301E**

Se ha producido un error interno, que ha provocado un estado de datos incoherentes. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1302E**

Se ha producido un error interno: se han definido varias células. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1303E**

Se ha producido un error interno: no se ha definido ninguna célula. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1304E**

Se ha producido un error de servidor interno: el producto no puede localizar la célula. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1305E**

Valor no permitido para redundancyType. El valor debe ser uno de los siguientes: NONE, MIRRORED u OVERLAPPING.

**CDHNC1306E**

El número de duplicación sólo puede especificarse cuando el tipo de redundancia es DUPLICADO; el número debe ser mayor que 1.

**CDHNC1307E**

Se ha producido un error de base de datos interno: una actualización no ha sido satisfactoria. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1308E**

Se ha producido un error interno: no se ha creado la partición. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1309E**

Se ha producido un error interno, que ha provocado un estado de datos incoherentes. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1310E**

No se puede configurar una célula en el estado Activo. Cambie el estado de la célula a Inactiva y, a continuación, configure la célula.

**CDHNC1311E**

Se ha producido un error interno, que ha provocado un estado de datos incoherentes. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1312E**

Se ha producido un error interno, que ha provocado un estado de datos incoherentes. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1313E**

Se ha producido un error interno, que ha provocado un estado de datos incoherentes. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1314E**

Utilice el formato siguiente para el ID de servidor: ipAddress\_ProfileName.

**CDHNC1315E**

El campo (*fieldname*)( debe ser una matriz no vacía.

**CDHNC1316E**

El campo (*fieldname*) debe ser un valor no nulo de tipo (*type*).

**CDHNC1317E**

Se ha producido un error interno: se ha producido una excepción desconocida al recopilar el estado de la célula. Posiblemente, la base de datos no esté en ejecución. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1318E**

La célula incluye ahora el servidor (*serverID*), que no está listado aquí. Renueve esta pantalla y vuelva a intentar la operación de configuración de célula.

**CDHNC1319E**

Se ha producido un error interno: el servidor (*serverID*) que se lista aquí ya no forma parte de la célula. Renueve esta pantalla y vuelva a intentar la operación de configuración de célula. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1320E**

Se ha producido un error interno: falta (*autoStartFlag*) para el servidor (*serverID*). Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1321E**

Se ha producido un error interno: el servidor (*serverID*) se ha listado varias veces. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (*errorTimestamp*).

**CDHNC1322E**

Se ha producido un error interno: al menos debe seleccionarse un buscador. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (*errorTimestamp*).

**CDHNC1323E**

Se ha producido un error interno: entrada de *mirrorCount* no válida. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (*errorTimestamp*).

**CDHNC1324E**

Se ha producido un error interno: entrada de *redundancyType* no válida. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (*errorTimestamp*).

**CDHNC1401E**

No se ha especificado ninguna acción.

**CDHNC1402E**

Se ha producido un error interno mientras se gestionaba el servidor. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (*errorTimestamp*).

**CDHNC1501E**

No se ha especificado ninguna acción.

**CDHNC1502E**

Se ha producido un error interno al iniciar o detener un asignador. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (*errorTimestamp*).

**CDHNC1601E**

No se ha especificado ninguna acción.

**CDHNC1602E**

Se ha producido un error interno mientras se gestionaba un buscador. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (*errorTimestamp*).

**CDHNC1604E**

No se puede suprimir un buscador activo.

**CDHNC1801E**

Se ha producido un error interno mientras se gestionaba el panel de instrumentos. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (*errorTimestamp*).

**CDHNC1802E**

No se ha podido analizar serverID como un entero

**CDHNC1803E**

No se ha podido recuperar serverID.

**CDHNC1804E**

No se han definido parámetros.

**CDHNC1805E**

Se ha producido un error interno: se ha asignado más de un servidor al mismo ID. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1806E**

Se ha producido un error interno: el producto no puede localizar un servidor con este ID en la base de datos. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1807E**

Se ha producido un error interno: la base de datos no se ha podido actualizar satisfactoriamente. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1809E**

No se ha especificado ninguna acción.

**CDHNC1901E**

Se ha producido un error interno, que ha provocado un estado de datos incoherentes. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1902E**

No hay ninguna acción a realizar.

**CDHNC1903E**

Se ha producido un error interno: el producto ha encontrado más de una célula activas. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1904E**

Se ha producido un error interno: el producto no puede localizar una célula configurada. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1906E**

Se ha producido un error interno: el producto no puede encontrar la célula a modificar. Póngase en contacto con el administrador del sistema.El

producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1907E**

Se ha producido un error interno: el producto no ha podido actualizar satisfactoriamente la base de datos. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC1908E**

Se ha producido un error interno durante la transferencia de la consola. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2101E**

No se ha podido recuperar serverID del URL.

**CDHNC2102E**

No se ha podido analizar serverID.

**CDHNC2103E**

Se ha producido un error interno: el producto ha encontrado más de un servidor en la base de datos con el ID especificado. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2104E**

El producto no puede encontrar el ID de servidor en la base de datos. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2105E**

El valor del parámetro (*paramName*) debe ser true o false.

**CDHNC2106E**

El componente ya está en el proceso de ser iniciado o detenido.

**CDHNC2107E**

El componente no se puede iniciar o detener, debido a que el servidor no está activo.

**CDHNC2108E**

Se ha producido un error interno: la actualización de la base de datos no ha sido satisfactoria. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2109E**

Se ha producido un error interno, que ha provocado un estado de datos incoherentes. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2110E**

Excepción al intentar bloquear la tabla de estado del servidor. Póngase en

contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2111E**

Se ha producido un error interno mientras se accedía a la base de datos. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2112E**

Se ha producido un error interno mientras se grababa en la base de datos. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2113E**

Las particiones deben definirse antes de que el buscador se pueda establecer para el inicio automático.

**CDHNC2114E**

Las particiones deben definirse antes de que el buscador se pueda iniciar.

**CDHNC2115E**

No hay servidores del producto actualmente activos.

**CDHNC2116E**

Se ha producido un error interno: el producto no puede obtener el estado de la célula. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2117E**

No se puede determinar el URL.

**CDHNC2118E**

Se ha pasado un nombre de parámetro inesperado al servicio. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2119E**

El valor de mirroredFrom es idéntico al de serverID. No se puede duplicar el servidor. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2120E**

No se puede realizar la operación iniciar o detener mientras la célula esté inactiva. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNC2121E**

El valor de newServer es idéntico al de serverID. No se puede transferir la consola. Póngase en contacto con el administrador del sistema.El producto

ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

#### **CDHNC2122E**

Se ha pasado un nombre de parámetro vacío inesperado al servicio. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

#### **CDHNC2123E**

Se ha producido un error interno. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

#### **CDHNC2124E**

Este servidor no se puede duplicar, porque los otros servidores en la célula ya tienen los componentes correspondientes configurados. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

## **Códigos de error de búsqueda de ENS**

Estos mensajes y los códigos relacionados pueden aparecer en los diálogos de error en la GUI de la búsqueda de ENS, en las respuestas de error del servicio web y donde proceda en el archivo de registro del servidor.

#### **CDHNS3000E**

No se ha recibido ninguna respuesta del servidor en el intervalo de tiempo esperado.

#### **CDHNS3001E**

No se han encontrado listas de nombres para las que tenga permiso de búsqueda.

#### **CDHNS3002E**

No se han encontrado estrategias.

#### **CDHNS3003E**

No se han encontrado culturas.

#### **CDHNS3010E**

Se ha producido un error interno: No se puede reinicializar el asignador. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

#### **CDHNS3011E**

El parámetro nameType no puede ser nulo ni estar vacío.

#### **CDHNS3012E**

El parámetro nameText no puede ser nulo ni estar vacío.

#### **CDHNS3013E**

El valor de nameType debe ser ORGANIZATION o PERSON.

#### **CDHNS3014E**

Se ha producido un error interno mientras se comunicaba con los buscadores. Póngase en contacto con el administrador del sistema.El

producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3015E**

Se ha producido un error interno: algunos componentes necesarios del buscador no están disponibles. Póngase en contacto con el administrador del sistema para comprobar el estado de los servidores.

**CDHNS3016E**

Se ha producido un error interno: El producto no puede leer la respuesta del buscador. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3017E**

Se ha producido un error interno: El producto no puede construir la solicitud del buscador. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3018E**

Se ha producido un error interno: El producto no puede obtener la lista de buscadores activos. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3019E**

Se ha producido un error interno: El producto no puede localizar la instancia de asignador en este servidor. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3020E**

Se ha producido un error interno: El asignador no está activo. Póngase en contacto con el administrador del sistema para comprobar el estado de los servidores.

**CDHNS3021E**

El parámetro nameListCode no puede ser nulo ni estar vacío.

**CDHNS3022E**

El parámetro externalID no puede ser nulo ni estar vacío.

**CDHNS3023E**

Su ID de usuario no tiene permisos de buscador establecidos para esta lista de nombres. Póngase en contacto con el administrador del sistema.

**CDHNS3024E**

No se puede obtener detalles de nombre debido a que falta nameListCodes o éste no es exclusivo.

**CDHNS3025E**

Se ha especificado un nombre de origen no válido. El nombre no existe.

**CDHNS3026E**

Se ha producido un error interno: El nombre de origen debe ser exclusivo, pero el producto ha encontrado varios nombres de origen. Póngase en

contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3027E**

No se ha podido añadir el nombre debido a que faltan criterios de nombre.

**CDHNS3028E**

El externalID especificado en los parámetros de la solicitud no coincide con el externalID especificado en el cuerpo del mensaje.

**CDHNS3029E**

El nameListCode especificado en los parámetros de la solicitud no coincide con el nameListCode especificado en el cuerpo del mensaje.

**CDHNS3030E**

Su ID de usuario no tiene permisos de gestor establecidos para esta lista de nombres. Póngase en contacto con el administrador del sistema.

**CDHNS3031E**

El parámetro nameCategory no puede ser nulo ni estar vacío.

**CDHNS3032E**

El parámetro nameCategory no es una categoría de nombres válida.

**CDHNS3033E**

El nombre no se puede suprimir debido a que falta nameListCodes o éste no es exclusivo. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3034E**

Los parámetros especificados no están permitidos cuando se añade un nombre de persona: *paramNames*.

**CDHNS3035E**

El valor del parámetro includeAlternateParses debe ser true o false.

**CDHNS3036E**

El parámetro *paramNames* no es válido cuando se añade un tipo de nombre FULL.

**CDHNS3037E**

Los parámetros givenName y surname no pueden ser nulo o estar vacíos cuando se añade un tipo de nombre PARSED.

**CDHNS3038E**

El nombre no se puede añadir debido a que falta o hay varios nameListCodes. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3039E**

Los parámetros especificados no están permitidos cuando se añade un nombre de organización: *paramNames*.

**CDHNS3040E**

El parámetro maneText no puede ser nulo o estar vacío cuando se añade un nombre de organización.

**CDHNS3041E**

No se ha podido recuperar referencias externas debido a que faltan criterios.

**CDHNS3042E**

Se ha producido un error interno al intentar obtener el número de puerto del buscador. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3043E**

Se ha producido un error interno al intentar establecer una conexión SSL con un buscador. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3044E**

Se ha producido un error interno: La sesión SSL del buscador no es válida. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3045E**

Se ha producido un error interno al intentar configurar la conexión SSL con un buscador. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3046E**

Se ha producido un error interno: Todos los componentes de buscador disponibles han sobrepasado el intervalo de tiempo de espera especificado. Póngase en contacto con un administrador para comprobar el estado de los servidores. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3047E**

Se ha producido un error interno: componentes necesarios del buscador no están disponibles. Póngase en contacto con un administrador para comprobar el estado de los servidores. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3048E**

Se ha producido un error interno durante la transliteración del nombre. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3049E**

Ha especificado una o más listas de nombres que no existen o en las que no puede realizar búsquedas con sus permisos actuales.

**CDHNS3050E**

No hay listas de nombres para las que tenga permiso de búsqueda.

**CDHNS3051E**

Se ha producido un error interno: un buscador devuelve un mensaje de error. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3052E**

maxResults debe ser un entero mayor que 0 (cero). Como alternativa, especifique -1 para utilizar el valor de maxResults predeterminado o 0 (cero) para el límite final tal y como está configurado.

**CDHNS3053E**

minScore debe ser un entero mayor o igual a 0 (cero), o especifique -1 para utilizar el valor de minScore predeterminado tal y como está configurado.

**CDHNS3061E**

No se puede analizar el nombre, posiblemente debido a caracteres no válidos.

**CDHNS3062E**

El nombre de búsqueda no se puede analizar, posiblemente debido a caracteres no válidos.

**CDHNS3063E**

Se ha producido un error interno: el producto no puede comprobar la existencia del nombre de búsqueda. El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3072E**

La lista de las listas de nombres que se han pasado a searchName no pueden tener más entradas del número máximo configurado.

**CDHNS3092E**

Se ha producido un error desconocido al intentar recuperar los códigos de estrategia. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3093E**

Se ha producido un error desconocido al intentar recuperar las culturas. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3094E**

Se ha producido un error desconocido al intentar realizar la búsqueda del nombre. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3095E**

Se ha producido un error desconocido al intentar recuperar las listas de nombres. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3096E**

Se ha producido un error desconocido al intentar recuperar los detalles del nombre. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3097E**

Se ha producido un error desconocido al intentar añadir un nombre. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3098E**

Se ha producido un error desconocido al intentar eliminar un nombre. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS3099E**

Se ha producido un error desconocido al intentar recuperar referencias externas. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS4001E**

Se ha producido un error en el servidor. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS4002E**

Debe proporcionarse un valor para *paramName*.

**CDHNS4003E**

Se ha especificado un valor no válido para el parámetro *paramName*.

**CDHNS4004E**

*minScore* debe ser un entero entre 0 y 100 inclusivo. Como alternativa, especifique -1 para utilizar el valor predeterminado de la estrategia de búsqueda.

**CDHNS4005E**

*maxResults* debe ser un entero mayor que cero. Como alternativa, especifique -1 para utilizar el valor predeterminado de la estrategia de búsqueda.

**CDHNS4006E**

Se ha producido un error de base de datos en el servidor. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

**CDHNS4007E**

Se ha producido un error del producto en el servidor. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

#### CDHNS4008E

se han pasado parámetros inesperados al servicio de búsqueda. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

#### CDHNS4009E

El nombre especificado incluye caracteres que no se pueden transliterar. Esto puede ser debido a la utilización de un conjunto de caracteres no soportado o una combinación de juegos de caracteres diferentes en el mismo nombre.

#### CDHNS4010E

Se ha producido un error interno. Póngase en contacto con el administrador del sistema.El producto ha registrado un mensaje en el archivo de registro del servidor de aplicaciones con una indicación de fecha y hora de aproximadamente (errorTimestamp).

---

## Buscar en las bases de conocimientos

A menudo puede encontrar soluciones a los problemas buscando en las bases de conocimientos de IBM. Puede optimizar los resultados utilizando los recursos, las herramientas de soporte y los métodos de búsqueda disponibles.

### Acerca de esta tarea

Puede encontrar información útil buscando en el Information Center para IBM InfoSphere Global Name Management , pero algunas veces necesita mirar más allá del Information Center para responder a sus preguntas o resolver problemas.

### Procedimiento

Para buscar en las bases de conocimientos la información que necesita, utilice uno de los métodos siguientes:

- Busque el contenido que necesita utilizando IBM Support Portal for IBM InfoSphere Global Name Management .  
IBM Support Portal es una vista centralizada y unificada de todas las herramientas de soporte técnico e información para todos los sistemas, software y servicios de IBM. IBM Support Portal permite acceder a la cartera de soporte electrónico de IBM de un lugar a otro. Puede personalizar las páginas para centrarse en la información y los recursos que necesita para prevenir los problemas y resolverlos con mayor rapidez. Familiarícese con IBM Support Portal gracias a los vídeos de demostración ([https://www.ibm.com/blogs/SPNA/entry/the\\_ibm\\_support\\_portal\\_videos](https://www.ibm.com/blogs/SPNA/entry/the_ibm_support_portal_videos)) sobre esta herramienta. Estos vídeos ofrecen una introducción a la herramienta IBM Support Portal, exploran la resolución de problemas y otros recursos, y muestran cómo se puede personalizar la página, moviendo, añadiendo y suprimiendo portlets.
- Busque contenido acerca de IBM InfoSphere Global Name Management utilizando uno de los recursos técnicos adicionales siguientes:
  - Notas técnicas de IBM InfoSphere
  - APAR (informes de problemas) de IBM InfoSphere
- Busque contenido utilizando la búsqueda de cabecera de IBM. Puede utilizar la búsqueda de cabecera de IBM escribiendo la serie de búsqueda en el campo Search (Buscar)= situado en la parte superior de cualquier página ibm.com.

- Busque contenido utilizando cualquier motor de búsqueda externo como Google, Yahoo o Bing. Si utiliza un motor de búsqueda externo, los resultados incluirán probablemente información que se encuentra fuera del dominio ibm.com. Aunque, algunas veces, puede encontrar información útil de resolución de problemas sobre productos de IBM en los grupos de noticias, foros y blogs que no están en ibm.com.

**Consejo:** Incluya "IBM" y el nombre del producto en la búsqueda si busca información sobre un producto de IBM.

---

## Archivos de registro

La información se graba en los archivos de registro cuando se produce una condición de habilitación en un componente de sistema específico, como por ejemplo cuando el componente se instala, se inicia o se produce un error durante el proceso.

Puede definir los nombres de archivo de registro y las ubicaciones utilizando los archivos de configuración, como, por ejemplo, el archivo de configuración IBM NameWorks. También puede establecer el nivel de registro, incluido el activado de rastreo. Toda la información de rastreo se incluye en los registros.

IBM NameWorks utiliza los mecanismos estándar encontrados en el paquete de registro `java.util.logging`, para que se puedan habilitar todos los aspectos del registro y el rastreo realizado llevado a cabo las funciones de IBM NameWorks utilizando las clases encontradas en el paquete Java. El valor predeterminado es que el registro esté inhabilitado.

El nombre del registrador que IBM NameWorks utiliza es `com.ibm.gnr.NameWorks`. Para obtener acceso al control de registro, utilice un código similar al siguiente:

```
java.util.logging.Logger gnrLogger =  
java.util.logging.Logger.getLogger( "com.ibm.gnr.NameWorks" );
```

A continuación, utilice los métodos del registrador estándar para establecer el nivel de registro y el manejador.

---

## Rastreo

Los rastreos son registros de proceso de componentes o transacciones. La información recopilada de un rastreo se puede utilizar para evaluar los problemas y el rendimiento.

En IBM NameWorks, el archivo de registro cronológico de rastreo (`trace.log`) se encuentra en el directorio `/ewas/profiles/NameWorksProfile/logs/NameWorksServer` del directorio de instalación del producto predeterminado.

IBM NameWorks soporta la mayor parte de los valores de nivel de registro proporcionados por WebSphere Application Server (por ejemplo, Severe, Fine, Finer y Finest). Para cambiar los valores de rastreo de IBM NameWorks, edite el archivo `server.xml` en el directorio `/ewas/profiles/NameWorksProfile/config/cells/DefaultNode/nodes/DefaultNode/servers/NameWorksServer` del directorio de instalación predeterminado y modifique el parámetro **`startupTraceSpecification`** para incluir el valor de nivel de registro apropiado. Por ejemplo, para establecer el valor de nivel de registro en `fine`, especifique:  
`startupTraceSpecification="*=info:com.ibm.gnr.NameWorks=fine".`

---

## Cómo ponerse en contacto con el soporte de IBM

El soporte de IBM proporciona asistencia para defectos del producto, respondiendo a las preguntas más frecuentes y realizando el redescubrimiento.

### Antes de empezar

Después de intentar hallar la respuesta o la solución utilizando otras opciones de autoayuda como las notas técnicas, puede ponerse en contacto con el soporte de IBM. Antes de ponerse en contacto con el soporte de IBM, su empresa debe tener un contrato vigente de mantenimiento de IBM y debe estar autorizado a notificar problemas a IBM. Para obtener información sobre los tipos de soporte disponibles, consulte el tema Support portfolio en la publicación *Software Support Handbook*.

Si prefiere tener una copia en PDF del manual, puede descargarlo en inglés. Si prefiere tener una copia del manual en otro idioma (por ejemplo, japonés o francés), consulte la página de bienvenida en el manual *Software Support Handbook* debajo de **Other languages**.

### Procedimiento

Complete los pasos siguientes para ponerse en contacto con el soporte de IBM con un problema:

1. Defina el problema. Describa el problema y los síntomas del modo más específico posible. Para obtener más información consulte el tema Getting IBM support en el manual *Software Support Handbook*.
2. Determine el nivel de gravedad del problema. Si es necesario, consulte el tema Getting IBM support en el manual *Software Support Handbook* para obtener definiciones útiles de niveles de gravedad.
3. Recopile información general, que incluya los datos siguientes: data:
  - ¿Qué versión y nivel de release del producto está utilizando?
  - ¿Qué niveles de software estaba ejecutando cuando se produjo el problema? Incluya los números de versión y números de paquete de servicio de la base de datos y el sistema operativo, así como todos los productos relacionados.
  - ¿Este problema ya se había producido antes o se trata de un hecho aislado?
  - ¿Puede reproducir el problema? En caso afirmativo, ¿cuáles son los pasos necesarios para reproducirlo?
  - ¿Es usted el principal contacto para solucionar este problema y cómo prefiere que le contacten (por correo electrónico o por teléfono)? ¿Existe más de un número de teléfono, página o dirección de correo electrónico donde se le pueda contactar? ¿Cuál es su disponibilidad? ¿Cuándo está en disposición de trabajar sobre este problema con el soporte de software de IBM?
  - ¿Existe algún contacto alternativo bien informado con quién el soporte de software de IBM pueda hablar?
4. Recopile los mensajes o datos de diagnóstico disponibles (como por ejemplo, los archivos de registro) que se han producido. Es de gran utilidad tener los números de mensaje de cualquiera de los mensajes que haya recibido cuando llamó al servicio de soporte.
5. Notifique el problema al soporte de IBM de una de las maneras siguientes:
  - En línea a través de IBM Support Portal: puede abrir, actualizar y ver todas las solicitudes de servicio desde el portlet Service Request en la página de solicitudes de servicio.

- Por teléfono: para conocer el número de teléfono al que debe llamar en su país, consulte la página web Directory of worldwide contacts.

## Resultados

Si el problema que notifica es debido a un defecto de software o a documentación omitida o inexacta, el soporte de IBM crea un Informe autorizado de análisis de programa (APAR). El APAR describe el problema con detalle. Siempre que es posible, el soporte de IBM proporciona una solución alternativa que el usuario puede aplicar hasta que se resuelva el APAR y se proporcione un arreglo. IBM publica diariamente los APAR resueltos en el sitio web de soporte de IBM a fin de que otros usuarios que tengan el mismo problema puedan beneficiarse de la misma solución.

---

## Suscripción a actualizaciones de soporte

Para mantenerse informado de la información relevante sobre los productos de IBM que utilice, puede suscribirse a actualizaciones.

### Acerca de esta tarea

Suscribiéndose para recibir actualizaciones, puede recibir información técnica y actualizaciones importantes para herramientas y recursos de soporte específicos. Puede suscribirse a actualizaciones utilizando uno de los métodos siguientes:

#### Canales de información de RSS y suscripciones a soporte social

Para obtener información general sobre RSS, incluidos los pasos sobre cómo empezar y una lista de las páginas web de IBM habilitadas para RSS, visite el sitio Canales de información de soporte de software de IBM..

#### Mis notificaciones

Con Mis notificaciones, puede suscribirse a actualizaciones de soporte para cualquiera de los productos de IBM. Puede especificar si desea recibir anuncios por correo electrónico a diario o semanalmente. Puede especificar qué tipo de información desea recibir (por ejemplo, publicaciones, sugerencias y consejos, flashes del producto (conocidos también como alertas), descargas y controladores). Mis notificaciones permite personalizar y clasificar los productos sobre los que desea estar informado así como los métodos de entrega que satisfacen mejor sus necesidades.

## Procedimiento

Para suscribirse a las actualizaciones de soporte:

1. Para suscribirse a las notificaciones de IBM InfoSphere Global Name Management , vaya a IBM InfoSphere Global Name Management Support Portal y pulse **Crear o actualizar su suscripción para este producto** en el portlet **Notificaciones**.
  - a. Si ya está registrado en Mi registro, inicie la sesión y salte al próximo paso. Si aún no se ha registrado, pulse **Regístrese ahora**. Complete el formulario de registro utilizando la dirección de correo electrónico como ID de IBM y pulse **Enviar**.
  - b. En el campo **Avisarme por**, seleccione uno o más de los métodos que desea utilizar para recibir actualizaciones, entre ellos el correo electrónico y los canales de información de RSS.
  - c. Si desea enviar notificaciones a una carpeta específica, seleccione la carpeta o escriba el nombre de una carpeta en **Opciones**.

- d. En la lista **Tipos de documento**, seleccione todos los tipos de documentos sobre los que desee recibir notificaciones en la lista.
    - e. Pulse **Enviar**.
  2. Para suscribirse a Mis notificaciones para otros productos de IBM, acceda primero a IBM Support Portal y pulse **Mis notificaciones** en el portlet **Notificaciones**.
    - a. Si ya está registrado en Mi registro, inicie la sesión y salte al próximo paso. Si aún no se ha registrado, pulse **Regístrese ahora**. Complete el formulario de registro utilizando la dirección de correo electrónico como ID de IBM y pulse **Enviar**.
    - b. Pulse **Edit profile**.
    - c. Pulse **Add products** y elija una categoría de productos; por ejemplo, **Software**. Se visualizará una segunda lista.
    - d. En la segunda lista, seleccione un segmento de productos; por ejemplo, **Data & Information Management**. Se visualizará una tercera lista.
    - e. En la tercera lista, seleccione un subsegmento de productos; por ejemplo, **Databases**. Se visualizará una lista de productos aplicables.
    - f. Seleccione los productos para los que desea recibir actualizaciones.
    - g. Pulse **Add products**.
    - h. Después de seleccionar todos los productos de interés, pulse **Subscribe to email** en la pestaña **Edit profile**.
    - i. Seleccione **Please send these documents by weekly email**.
    - j. Actualice su dirección de correo electrónico según sea necesario.
    - k. En la **Documents list**, seleccione la categoría de productos; por ejemplo, **Software**.
    - l. En la **Documents list**, seleccione la categoría de productos; por ejemplo, **Software**.
    - m. Pulse **Update**.

## Resultados

Hasta que modifique las fuentes y las preferencias de Mis notificaciones, recibirá notificaciones de las actualizaciones que haya solicitado. Puede modificar las preferencias cuando sea necesario (por ejemplo, si deja de utilizar un producto y empieza a utilizar otro).

### Información relacionada

-  [Canales de información de RSS de soporte de software de IBM](#)
-  [Suscribirse a las actualizaciones de contenido de soporte para Mis notificaciones](#)
-  [Mis notificaciones para el soporte técnico de IBM](#)
-  [Visión general de Mis notificaciones para el soporte técnico de IBM](#)



---

## Apéndice. Glosario

Este glosario incluye términos y definiciones de IBM InfoSphere Global Name Management.

En este glosario se utilizan las siguientes referencias cruzadas:

- Véase le remite de un término a un sinónimo preferido, o de un acrónimo o abreviatura a la definición completa de éstos.
- Véase también le remite a un término relacionado u opuesto.

Para consultar los glosarios de otros productos de IBM, diríjase a la dirección [www.ibm.com/software/globalization/terminology](http://www.ibm.com/software/globalization/terminology) (se abre en una ventana aparte).

---

### A

**afijo** Elemento dependiente de un nombre que se añade al principio (prefijo), en el medio (infijo) o al final (sufijo) de un nombre y que modifica su significado. Un afijo puede asociarse directamente al nombre (como "Mac" en "Macintosh"), separarse de la raíz del nombre mediante signos de puntuación (como "O" en "O'Connell") o separarse de la raíz del nombre mediante un espacio en blanco (como "Abd" en "Abd Allah"). Los afijos son más comunes en apellidos y a veces puede identificar el origen étnico. Véase también raíz de nombre.

**analizar**

Proceso que analiza el texto para determinar su estructura y divide el texto en cadenas de caracteres individuales.

**apellido**

Nombre que se añade a un nombre de pila e identifica a un individuo como parte de un grupo de personas, como por ejemplo una familia, tribu o clan. El apellido es un elemento clave de un nombre de persona, junto con el nombre de pila. Los apellidos no se utilizan en algunas áreas del mundo. Véase también nombre de pila, nombre de persona.

**apodo** Nombre alternativo de un nombre de persona, a menudo derivado de otros elementos del nombre.

---

### C

**cadena de caracteres de nombre**

Elemento indivisible más pequeño de un nombre, delimitado por un espacio en blanco o un signo de puntuación. Las cadenas de caracteres de nombre se combinan para formar frases de nombre y campos de nombre. Una cadena de caracteres de nombre puede contener varios componentes de nombre. Las cadenas de caracteres de nombre son afijos o raíces. La función exacta de una cadena de caracteres de nombre depende de su posición en el nombre de persona. Véase también campo de nombre, frase de nombre, raíz de nombre.

**calificador**

Término o frase añadida al final de un nombre de persona para distinguirlo mediante la especificación de una posición generacional (como

por ejemplo Junior o Senior, o "fils" en francés en lugar de Junior), una distinción académica o un rango religioso que una persona ha obtenido (por ejemplo, Ph.D.) o una cualificación profesional de cualquier tipo (por ejemplo, D.D.S.). A efectos de comparación de nombres, un calificador se considera un componente periférico o menor de un nombre de persona. Véase también honorífico/a.

**campo de nombre**

Construcción de datos que consta de una o varias frases de nombre o cadenas de caracteres de nombre. Véase también frase de nombre, cadena de caracteres de nombre.

**célula** Grupo de procesos gestionados que están federados en el mismo gestor de despliegue y pueden incluir grupos principales de alta disponibilidad.

---

## D

**diacrítico/a**

Marca que indica un cambio del valor fonético de un carácter o combinación de caracteres.

**distribuido**

Relativo a programas y fuentes computerizadas de información de un entorno de informática que están físicamente ubicados en distintos sistemas, mientras siguen trabajando conjuntamente como una única unidad lógica.

---

## E

**estrategia de búsqueda**

Colección con nombre de valores de parámetros de comparación que alteran temporalmente los valores existentes o predeterminados y que se utilizan para realizar una búsqueda. Por ejemplo, las estrategias de búsqueda existentes incluyen las estrategias estándar (valores predeterminados), amplia (valores que amplían la búsqueda) y limitada (valores que restringen la búsqueda). Los administradores pueden definir su propio conjunto de valores de parámetros de comparación y guardarlos como estrategia de búsqueda.

---

## F

**fórmula de tratamiento**

Componente de un nombre de persona que representa una posición social, religiosa o académica, como por ejemplo "Dr.", "Ms." o "Colonel". La fórmula de tratamiento es un componente opcional de un nombre de persona, y generalmente precede al nombre de pila. A efectos de comparación de nombres, una fórmula de tratamiento se considera un componente periférico o menor de un nombre de persona. Véase también honorífico/a.

**fórmulas de tratamiento, afijos y calificadores (TAQ)**

Cadena de caracteres de nombre que ayuda a identificar que una serie de texto representa un nombre. Estas cadenas de caracteres normalmente dicen algo sobre el nombre, pero sólo los afijos forman parte del nombre real. Véase también fórmula de tratamiento, afijo y calificador.

**frase de nombre**

Unidad indivisible que consta de una raíz de nombre y los afijos asociados

con ella. Algunas frases de nombre pueden estar formadas por varias raíces, como en el nombre chino Mei-Hui o el nombre inglés Mary-Anne. Pueden combinarse una o varias frases para crear un campo de nombre. Véase también campo de nombre, raíz de nombre, cadena de caracteres de nombre.

---

## H

### **honorífico/a**

Prefijo o sufijo que indica la posición social alcanzada por una persona o conferida a ella. Véase también calificador, fórmula de tratamiento.

---

## I

**inicial** Una cadena de caracteres de nombre que consta de un solo carácter que representa el primer carácter de un nombre. Las cadenas de caracteres de nombre de un solo carácter ("iniciales") se manejan de forma diferente que las cadenas de caracteres de nombre de varios caracteres ("nombres").

---

## J

### **JavaScript Object Notation (JSON)**

Formato de intercambio de datos ligero que se basa en la notación literal de objetos de JavaScript. JSON es independiente del lenguaje de programación, sin embargo utiliza convenciones de lenguajes entre los que se incluyen C, C++, C#, Java, JavaScript, Perl, Python.

**JSON** Véase JavaScript Object Notation.

---

## L

### **lista de datos**

Lista de nombres residente en memoria que se construye a partir de un origen de datos externo. Las solicitudes de búsqueda se realizan en una o varias listas de datos.

---

## M

### **marca de acento**

Diacrítico que se utiliza para marcar el tono de una sílaba. Véase también diacrítico.

### **marcador de parentesco**

Término que se incluye en un nombre de persona y que indica una relación de parentesco entre individuos. Por ejemplo, en el nombre "Karim bin Hassan," el marcador de parentesco "bin" significa "hijo de".

### **matronímico**

Elemento derivado del nombre de la madre o de otro antepasado femenino de una persona.

---

## N

### **nombre**

Nombre de pila de los nombres anglosajones. Véase nombre de pila y segundo nombre.

---

**nombre analizado**

Nombre cuya estructura sintáctica (es decir, frases de nombre, campos de nombre, fórmulas de tratamiento y calificadores) se ha definido y representado como salida del proceso de análisis.

**nombre de análisis alternativo**

Posible variación de un nombre, que se utiliza para mejorar el análisis y puntuación del nombre. Véase también variante de nombre.

**nombre de familia**

Véase apellido.

**nombre de organización**

Nombre no personal que hace referencia a un cuerpo estructurado por una o más personas que realizan funciones comunes. Los nombres de organización incluyen generalmente algún tipo de indicador, patrón o palabras que los identifican como nombres no personales.

**nombre de persona**

Nombre que hace referencia a un ser humano individual y que consta de uno o varios nombres de pila, apellidos, fórmulas de tratamiento o calificadores. Un nombre de persona completo hace referencia a un individuo y puede codificar información que indique la clase social, los antecedentes religiosos y políticos, el nivel de formación, el origen étnico o cultural y la procedencia regional. Un nombre de persona está formado por uno o dos campos de nombre. Véase también nombre de pila, apellido, fórmula de tratamiento, calificador, campo de nombre.

**nombre de pila**

Nombre que se utiliza para identificar una persona dentro de un grupo, como por ejemplo una familia. Una persona puede tener varios nombres de pila. El nombre de pila es el elemento clave de un nombre de persona. Un nombre de pila puede ser el único elemento de nombre que tiene carácter universal entre todos los nombres del mundo. Véase también apellido, nombre de persona.

**nombres conjuntos**

Agrupación de segmentos de nombres que contienen dos o más nombres de pila, dos o más fórmulas de tratamiento, pares formados por una fórmula de tratamiento y un nombre de pila, dos o más nombres completos o cualquier combinación de estos elementos de nombre. Por ejemplo, "Mr. and Mrs. John Smith" contiene fórmulas de tratamiento conjuntas y "Mr. and Mrs. John and Mary Smith" contiene fórmulas de tratamiento conjuntas y nombres de pila conjuntos en la construcción de nombres.

---

**P****palabra ignorada**

Palabra que no añade ningún significado a un nombre de organización y no se incluye en ninguna comparación ni puntuación de nombres.

**patronímico**

Elemento derivado del nombre del padre o de otro antepasado masculino de una persona. Tanto los apellidos como los nombres de pila actúan como nombres patronímicos en diversas áreas del mundo.

**precisión**

Medición de recuperación de información que especifica la proporción de datos relevantes en relación a todos los datos recuperados. La precisión es

un valor predictivo positivo. El mejor procedimiento para medir la recuperación de información es utilizar tanto la precisión como la recuperación. Véase también recuperación.

**prefijo**

Afijo que aparece al principio de un nombre. Por ejemplo, en el apellido "de Rosa" el afijo "de" es un prefijo.

**puntuación**

Proceso que calcula el grado de coincidencia de los atributos de una identidad de entrada con los atributos de una entidad existente.

---

**R**

**raíz de nombre**

Elemento de nombre que puede aparecer de forma autónoma o combinado con afijos u otras raíces para formar un nombre completo o una frase de nombre. Véase también frase de nombre, cadena de caracteres de nombre.

**recuperación**

Medición de recuperación de información que especifica el porcentaje de datos relevantes recuperados en relación a todos los datos relevantes disponibles. La recuperación es una medida de confiabilidad. El mejor procedimiento para medir la recuperación de información es utilizar tanto la precisión como la recuperación. Véase también precisión.

**redundancia**

Uso de varias unidades funcionales idénticas, como por ejemplo varias unidades de disco o sistemas de fuente de alimentación, dentro de un sistema informático para proporcionar seguridad de datos y cierto grado de tolerancia de errores en caso de anomalías de hardware.

**regularización**

Proceso de normalización de las cadenas de caracteres de nombres que, una vez normalizados, se añaden a una lista de datos. Véase también cadena de caracteres de nombre y lista de datos.

**Representational State Transfer (REST)**

Estilo arquitectónico de software para sistemas de hipermedia distribuida, como por ejemplo World Wide Web. El término se suele emplear para describir cualquier interfaz simple que utiliza XML (o YAMLS, JSON y texto sin formato) a través de HTTP sin una capa de mensajería adicional, como por ejemplo SOAP.

**REST** Véase Representational State Transfer.

**romanización**

Proceso de transcribir cualquier texto no latino con el alfabeto latino. Véase también transliterar.

---

**S**

**score** Resultado de un análisis computacional. Véase también puntuación.

**segundo nombre**

Segundo nombre de los nombres anglosajones. Véase nombre de pila y nombre.

**sintaxis**

Organización y relación entre los elementos de un nombre (u otra

expresión o frase). Por ejemplo, la sintaxis de los nombres ingleses distingue los nombres de pila y los apellidos en: Todd Lane y Lane Todd.

**SOAP** Protocolo ligero basado en XML para intercambiar información en un entorno distribuido descentralizado. SOAP se puede utilizar para consultar y devolver información e invocar servicios en Internet.

**sufijo** Afijo que aparece al final de un nombre. Por ejemplo, el afijo "eddin" del nombre "Nur-eddin" es un sufijo.

---

## T

**TAQ** Véase fórmulas de tratamiento, afijos y calificadores.

**transliterar**

Conversión de texto de alfabetos no latinos a sus caracteres latinos equivalentes. Consulte también "romanización" en la página 249.

---

## V

**variante de nombre**

Alternativa de un nombre especificado que se considera equivalente a dicho nombre, pero que difiere de él en su forma externa particular. Las variantes de nombre surgen de variaciones en la escritura (por ejemplo, Geoff y "Jeff"), apodos (por ejemplo, "Bill" por "William"), abreviaturas (por ejemplo, "GPE" por "Guadalupe"), traducciones (por ejemplo, "Peter" por "Pierre") u otros procesos.

---

## Avisos

Esta información se ha escrito para productos y servicios ofrecidos en los Estados Unidos de América.

Es posible que en otros países IBM no ofrezca los productos, los servicios o las características que se describen en este documento. El representante local de IBM le puede informar acerca de los productos y servicios que actualmente están disponibles en su localidad. Las referencias hechas a productos, programas o servicios IBM no pretenden afirmar ni dar a entender que únicamente puedan utilizarse dichos productos, programas o servicios IBM. Puede utilizarse en su lugar cualquier otro producto, programa o servicio funcionalmente equivalente que no vulnere ninguno de los derechos de propiedad intelectual de IBM. No obstante, es responsabilidad del usuario evaluar y verificar el funcionamiento de cualquier producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patente pendientes de aprobación que cubran los temas descritos en este documento. La posesión de este documento no le otorga ninguna licencia sobre dichas patentes. Puede enviar consultas sobre las licencias, por escrito, a:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
EE.UU.

Para consultas sobre licencias relativas a la información de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM en su país o envíe las consultas, por escrito, a:

IBM World Trade Asia Corporation  
Licensing 2-31 Roppongi 3-chome, Minato-ku  
Tokio 106-0032, Japón

**El párrafo siguiente no es aplicable en el Reino Unido ni en ningún otro país en el que tales disposiciones sean incompatibles con la legislación local:**

INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍA DE NINGUNA CLASE, EXPLÍCITA O IMPLÍCITA, INCLUIDAS, PERO SIN LIMITARSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN Y DE COMERCIALIZACIÓN O IDONEIDAD PARA UN PROPÓSITO DETERMINADO. Algunas legislaciones no contemplan la declaración de limitación de garantías, ni implícitas ni explícitas, en determinadas transacciones, por lo que cabe la posibilidad de que esta declaración no se aplique en su caso.

Esta información puede contener imprecisiones técnicas o errores tipográficos. La información incluida en este documento está sujeta a cambios periódicos, que se incorporarán en nuevas ediciones de la publicación. IBM puede efectuar mejoras y/o cambios en los productos y/o programas descritos en esta publicación en cualquier momento y sin previo aviso.

Las referencias hechas en esta publicación a sitios Web que no son de IBM se proporcionan únicamente por cortesía y de ningún modo deben interpretarse como promoción de dichos sitios Web. Los materiales de estos sitios Web no forman parte de los materiales de IBM para este producto, y el usuario será responsable del uso que se haga de estos sitios Web.

IBM puede utilizar o distribuir la información que usted le suministre del modo que IBM considere conveniente sin incurrir por ello en ninguna obligación para con usted.

Los licenciatarios de este programa que deseen obtener información acerca del mismo con el fin de: (i) intercambiar la información entre programas creados independientemente y otros programas (incluido este) y (ii) utilizar mutuamente la información que se ha intercambiado, deben ponerse en contacto con:

IBM Corporation  
J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003  
EE.UU.

Esta información puede estar disponible, sujeta a los términos y condiciones pertinentes, e incluir en algunos casos el pago de una cantidad.

El programa bajo licencia que se describe en este documento, así como todo el material bajo licencia disponible que contiene, lo proporciona IBM bajo los términos del Acuerdo de Cliente de IBM, el Acuerdo Internacional de Programas bajo Licencia de IBM o cualquier otro acuerdo equivalente entre ambas partes.

Los datos de rendimiento incluidos aquí se determinaron en un entorno controlado. Por lo tanto, los resultados que se obtengan en otros entornos operativos pueden variar significativamente. Algunas mediciones pueden haberse realizado en sistemas que estén en fase de desarrollo y no existe ninguna garantía de que esas mediciones vayan a ser iguales en los sistemas disponibles en el mercado. Además, es posible que algunas mediciones se hayan estimado mediante extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben verificar los datos aplicables a su entorno específico.

La información concerniente a productos que no son de IBM se ha obtenido de los suministradores de dichos productos, de sus anuncios publicados o de otras fuentes de información pública disponibles. IBM no ha comprobado dichos productos y no puede afirmar la exactitud en cuanto a rendimiento, compatibilidad u otras características relativas a productos no IBM. Las consultas acerca de las prestaciones de los productos que no son de IBM deben dirigirse a los suministradores de tales productos.

Todas las declaraciones relacionadas con las direcciones o intenciones futuras de IBM están sujetas a cambios o renunciadas sin previo aviso y representan únicamente metas y objetivos. Esta información sólo está pensada a la planificación. La información aquí contenida está sujeta a cambios antes de que los productos descritos estén disponibles.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlas de la forma más completa posible, los ejemplos incluyen nombres de personas, empresas, marcas y productos. Todos estos

nombres nombres son ficticios y cualquier parecido con nombres y direcciones utilizados por una empresa real es mera coincidencia.

#### LICENCIA DE COPYRIGHT:

Esta información contiene programas de aplicación de ejemplo en lenguaje fuente, que muestran técnicas de programación en diversas plataformas operativas. Puede copiar, modificar y distribuir los programas de ejemplo de cualquier forma, sin tener que pagar a IBM, con intención de desarrollar, utilizar, comercializar o distribuir programas de aplicación que estén en conformidad con la interfaz de programación de aplicaciones (API) de la plataforma operativa para la que están escritos los programas de ejemplo. Los ejemplos no se han probado minuciosamente bajo todas las condiciones. Por lo tanto, IBM no puede garantizar ni dar por sentada la fiabilidad, la facilidad de mantenimiento ni el funcionamiento de los programas.

© Copyright IBM Corp. (2003, 2016). Reservados todos los derechos.

Si visualiza esta información en soporte software, es posible que las fotografías y las ilustraciones en color no aparezcan.



---

## Marcas registradas

Las marcas registradas de IBM y determinadas marcas registradas que no son de IBM se han marcado cuando aparecen por primera vez en esta información con el símbolo correspondiente.

IBM, el logotipo de IBM e [ibm.com](http://ibm.com) son marcas registradas de International Business Machines Corp., registradas en muchas jurisdicciones en el mundo entero. Los nombres de otros productos y servicios pueden ser marcas registradas de IBM o de otras compañías. En la página [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) existe una lista actualizada de las marcas registradas de IBM bajo el epígrafe "Avisos de copyright y marcas registradas".

Los términos siguientes son marcas registradas de otras empresas:

Adobe, el logotipo de Adobe, PostScript y el logotipo de PostScript son marcas registradas de Adobe Systems Incorporated en los Estados Unidos y/o en otros países.

IT Infrastructure Library es una marca registrada de Central Computer and Telecommunications Agency, que ahora forma parte de la oficina de comercio gubernamental.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas registradas de Intel Corporation o de sus filiales en los Estados Unidos y/o en otros países.

Linux es una marca registrada de Linus Torvalds en Estados Unidos y/o en otros países.

Microsoft, Windows, Windows NT y el logotipo de Windows son marcas registradas de Microsoft Corporation en los Estados Unidos y/o en otros países.

ITIL es una marca registrada y una marca registrada y comunitaria de la oficina de comercio gubernamental y está registrada en Estados Unidos, en la Oficina de patentes y marcas.

UNIX es una marca registrada de The Open Group en los Estados Unidos y en otros países.

Cell Broadband Engine es una marca registrada de Sony Computer Entertainment, Inc. en los Estados Unidos y/o en otros países y es utilizada de acuerdo con la licencia correspondiente.

Java y todas las marcas basadas en Java son marcas registradas de Sun Microsystems, Inc., en los Estados Unidos y/o en otros países.

El Servicio postal de los Estados Unidos posee las marcas registradas siguientes: CASS, CASS Certified, DPV, LACSLink, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS y United States Postal Service. IBM Corporation es un licenciataria de DPV y LACSLink no exclusivo del Servicio postal de los Estados Unidos.

Otros nombres de empresas, productos o servicios pueden ser marcas registradas o de servicio de terceros.

---

## Términos y condiciones

Los permisos para utilizar estas publicaciones están sujetos a los siguientes términos y condiciones.

**Uso personal:** puede reproducir estas publicaciones para uso personal (no comercial) siempre y cuando incluya una copia de todos los avisos de derechos de autor. No puede distribuir ni visualizar estas publicaciones ni ninguna de sus partes, como tampoco elaborar trabajos que se deriven de ellas, sin el consentimiento explícito de IBM.

**Uso comercial:** puede reproducir, distribuir y visualizar estas publicaciones únicamente dentro de su empresa, siempre y cuando incluya una copia de todos los avisos de derechos de autor. No puede elaborar trabajos que se deriven de estas publicaciones, ni tampoco reproducir, distribuir ni visualizar estas publicaciones ni ninguna de sus partes fuera de su empresa, sin el consentimiento explícito de IBM.

Aparte de la autorización que se concede explícitamente en este permiso, no se otorga ningún otro permiso, licencia ni derecho, ya sea explícito o implícito, sobre las publicaciones, la información, los datos, el software o cualquier otra propiedad intelectual contenida en ellas.

IBM se reserva el derecho de retirar los permisos aquí concedidos siempre que, según el parecer de IBM, las publicaciones se utilicen en detrimento de sus intereses o cuando, también según el parecer de IBM, no se sigan debidamente las instrucciones anteriores.

No puede bajar, exportar ni reexportar esta información si no lo hace en plena conformidad con la legislación y normativa vigente, incluidas todas las leyes y normas de exportación de Estados Unidos.

IBM NO PROPORCIONA NINGUNA GARANTÍA SOBRE EL CONTENIDO DE ESTAS PUBLICACIONES. LAS PUBLICACIONES SE PROPORCIONAN "TAL CUAL" Y SIN GARANTÍAS DE NINGUNA CLASE, BIEN SEA EXPLÍCITAS O IMPLÍCITAS, INCLUIDAS, PERO SIN LIMITARSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE COMERCIALIZACIÓN, NO VULNERACIÓN E IDONEIDAD PARA UN FIN DETERMINADO.



---

# Índice

## Caracteres Especiales

[Datalist:], sección  
CompParmsDefaults= 68

### A

actualizar  
  nombres en listas de datos 55  
addName()  
  métodos 54  
addNameToDatalist(), método 54  
Afijo de organización  
  OA 28  
afijos  
  description 24  
alteraciones temporales  
  parámetros de comparación 68  
analizar  
  nombres analizados 20  
  utilizando el método analyze() 41  
analizar nombres 41  
  realizar análisis completo de un nombre 41  
analizar nombres (desglosar) 35  
  descripción del árbol de análisis 19  
  en componentes individuales 35  
  método analyzeForSearch() 60, 72  
  para la búsqueda 60, 72  
  utilización de NameWorks 35  
analyze()  
  métodos 41  
analyze(), método 41  
analyzeForSearch()  
  métodos 60, 72  
añadir  
  nombres a listas de datos 54  
apellidos  
  description 21  
  generar variantes de nombres 51  
  identificar entorno cultural 45  
  identificar país de asociación 48, 49  
API  
  códigos de error 209  
Aplicación de ejemplo search de NameHunter 99  
Aplicación de ejemplo why de NameHunter 99  
aplicaciones cliente 11  
Aplicaciones de ejemplo de NameHunter 99  
arabicTransRule.ibm  
  configurar para utilizar con el proceso de Distributed Search 132  
  configurar para utilizar con NameHunter 131  
árbol de análisis  
  description 19

árboles de análisis  
  crear 35  
Archivo Datos provisionales 137  
Archivo de configuración de IBM NameWorks 197  
  Distributed Search 194  
  Embedded Search 195  
archivo de configuración de NameLoader 181  
archivo de configuración de NameWorks 192, 193, 194, 199, 200, 201  
archivos  
  archivo de contraseñas (ens.passwd), Enterprise Name Search 171  
archivos de registro  
  establecer nivel de registro para IBM NameWorks 240  
  resolución de problemas 240  
archivos de reglas  
  transliteración 30  
archivos de reglas de transliteración  
  configurar para utilizar con el proceso de Distributed Search 132  
  configurar para utilizar con IBM NameWorks 205  
  configurar para utilizar con NameHunter 131

### B

bases de conocimientos  
  buscar soluciones a los problemas 239  
buscar nombres 53, 80  
  analyzeForSearch(), método 60, 72  
  mediante IBM NameWorks 64  
  mediante NameHunter 82  
  recuperar datos complementarios 81  
  utilizar Enterprise Name Search 170  
búsquedas completas  
  description 64  
búsquedas de nombres completos 80  
búsquedas de nombres exclusivos 80  
  description 64  
  recuperar datos complementarios 81

### C

cadena de caracteres  
  cadena de caracteres raíz 28  
cadena de caracteres de nombre  
  description 23, 91  
cadena de caracteres raíz  
  description 28  
calificadores  
  description 21, 24  
campos de nombre  
  apellidos 21  
  calificadores 21

campos de nombre (*continuación*)  
  conjunciones precedentes 21  
  description 21  
  fórmulas de tratamiento 21  
  generación de variantes para nombres de pila y apellidos 51  
  generar variantes de nombres completos 51  
  identificar el entorno cultural de campos de nombre 45  
  nombre de pila 21  
canales de información de RSS  
  resolución de problemas 242  
características  
  novedad en la versión 6.0 2  
casos de ejemplo de utilización  
  buscar nombres 61  
categorías de nombres 14, 73  
  códigos de razón 77  
categorize(), método 76  
clase de configuración  
  alteraciones temporales 68  
  alteraciones temporales de parámetros de comparación 71  
  especificar valores de configuración 203  
clasificar nombres  
  entorno cultural 60, 72  
códigos  
  códigos de entorno cultural 43  
  códigos de razón 77  
códigos de error 219, 224, 226  
Códigos de error 209  
códigos de error C++ 220  
códigos de error de componentes  
  errores generales  
    GODC001E 214  
    GODC002E 214  
    GODC003E 214  
    GODC004E 214  
    GODC005E 214  
    GODC006E 214  
    GODD001E 214  
    GODD002E 214  
    GODD003E 214  
    GODD004E 214  
    GODD005E 214  
    GODD006E 214  
    GODH001E 214  
    GODH002E 214  
    GODH003E 214  
    GODH004E 214  
    GODH005E 214  
    GODH006E 214  
    GODL001E 214  
    GODL002E 214  
    GODL003E 214  
    GODL004E 214  
    GODL005E 214  
    GODL006E 214  
    GODP001E 214

- códigos de error de componentes (continuación)
    - errores generales (continuación)
      - GODP002E 214
      - GODP003E 214
      - GODP004E 214
      - GODP005E 214
      - GODP006E 214
      - GODS001E 214
      - GODS002E 214
      - GODS003E 214
      - GODS004E 214
      - GODS005E 214
      - GODS006E 214
      - GODT001E 214
      - GODT002E 214
      - GODT003E 214
      - GODT004E 214
      - GODT005E 214
      - GODT006E 214
      - GODV001E 214
      - GODV002E 214
      - GODV003E 214
      - GODV004E 214
      - GODV005E 214
      - GODV006E 214
      - GODW001E 214
      - GODW002E 214
      - GODW003E 214
      - GODW004E 214
      - GODW005E 214
      - GODW006E 214
  - códigos de razón
    - categorías de nombres 77
  - comentarios
    - enviar v
  - comparación de nombres
    - procedimientos 13
    - visión general 13
  - comparar fechas
    - determinar diferencias entre fechas utilizando IBM NameWorks 79
    - mediante IBM NameWorks 79
  - comparar nombres
    - mediante IBM NameWorks 78
  - componente
    - API 9
  - componentes del nombre 19
  - CompParms
    - Distributed Search 121
    - NameHunter 121
  - CompParms (parámetros de comparación) de NameHunter
    - entorno cultural coreano-ruso 126
  - entornos culturales
    - ambiguo-árabe 122
  - entornos culturales chino-germánico 124
  - entornos culturales grupales 130
  - entornos culturales hispanico-japonés 125
  - entornos culturales tailandés-yoruba 128
  - Nombres de empresa 128
  - CompParmsOverrides, clase
    - alteraciones temporales 68
  - CompParmsOverrides, clase (continuación)
    - alteraciones temporales de parámetros de comparación 71
  - conexiones IPv6
    - archivo de configuración de IBM NameWorks 194
  - conjunciones precedentes
    - description 21
  - contactar
    - soporte de IBM 241
  - cyrillicTransRule.ibm
    - configurar para utilizar con el proceso de Distributed Search 132
    - configurar para utilizar con NameHunter 131
- D**
- dataFetch()
    - métodos 81
  - dataFetch(), método 81
  - datalists
    - añadir nombres 54
  - dateCompare(), método 79
  - dateDifference(), método 79
  - Datos TAQ de NameHunter 83
  - deleteNameFromDatalist()
    - métodos 57
  - deleteNameFromDatalist(), método 57
  - description 10
  - Designador de organización
    - OD 25
  - directrices legales
    - publicaciones
      - reutilizar 257
    - términos y condiciones 257
  - Distributed Search 157, 224
    - aplicaciones de servidor 11
    - archivo de configuración de IBM NameWorks 194
    - códigos de error 224
- E**
- Embedded Search
    - archivo de configuración de IBM NameWorks 195
    - preproceso de listas de nombres 135
  - Enterprise Name Search 226
    - actualización de las estadísticas de base de datos 186
    - archivo de contraseñas (ens.passwd) 171
    - asignar usuarios a grupos de seguridad 172
    - cambiar las contraseñas de usuario 174
    - cargar nombres de listas de nombres en un esquema 184
    - códigos de error 226
    - control de acceso 174
    - crear nombres de usuario y contraseñas 172
- Enterprise Name Search (continuación)**
- gestión de la seguridad de usuario, restablecimiento de contraseñas de usuario 174
  - gestionar la seguridad de usuario 171, 174
  - gestionar listas de nombres 141, 176
  - restablecimiento de contraseñas de usuario 174
  - seguridad, sintaxis del programa de utilidad wpspwd 171
  - sintaxis del programa de utilidad wpspwd 171
  - suprimir nombres de listas de nombres 187
  - suprimir todos los nombres del esquema de Enterprise Name Search 187
  - suprimir usuarios 174
- entorno cultural**
- identificación utilizando NameWorks 42
  - método analyzeForSearch() 60, 72
- entornos culturales**
- códigos de entorno cultural 43
  - identificación de un nombre de organización 45
  - identificar para campos de nombre 45
  - identificar para nombre completo 44
- enviar comentarios v**
- esquemas**
- cargar nombres de listas de nombres a un esquema de Enterprise Name Search 184
  - suprimir todos los nombres del esquema de Enterprise Name Search 187
- estrategia de búsqueda**
- alteraciones temporales 68
  - amplia 65
  - de ejemplo
    - amplia 65
    - limitada 66
  - limitada 66
- estrategias de búsqueda**
- alteraciones temporales de parámetros de comparación 70
- crear**
- mediante el archivo de configuración 67
  - mediante la clase Strategy 67
- gestionar**
- IBM NameWorks 65
  - IBM NameWorks 65
- modificar**
- mediante el archivo de configuración 67
- suprimir 68**
- F**
- fechas**
- comparar mediante IBM NameWorks 79
  - determinar diferencia entre fechas utilizando IBM NameWorks 79

- fórmulas de tratamiento
  - description 21, 24
- frases de nombre
  - description 23
  - identificar país de asociación 48, 49
- funciones nuevas
  - versión 6.0 8

## G

- genderize(), método 46
- genderizeField(), método 47
- generar
  - variantes de nombres completos 51
  - variantes para campos de nombre 51
- género
  - identificación utilizando NameWorks 46
  - identificar para nombre 47
  - identificar para nombre completo 46
  - identificar utilizando el método analyze() 41
- getVariants(), método 51
- Global Name Management
  - Descripción 1
- glosario 245
- greekTransRule.ibm
  - configurar para utilizar con el proceso de Distributed Search 132
  - configurar para utilizar con NameHunter 131

## H

- herramientas
  - herramientas de soporte 239

## I

- IBM NameWorks 10, 220
  - alteraciones temporales de parámetros de comparación 68
  - API 10
  - archivo de configuración 190
  - cargar datos actualizados 190
  - clase de configuración 203
  - configurar 189
  - configurar para utilizar archivos de reglas de transliteración 205
  - Embedded Searching 135
  - estrategias de búsqueda 65
  - gestionar
    - estrategias de búsqueda 65
  - gestionar listas de datos 53
  - información de migración 57
  - suprimir estrategias de búsqueda 68
- ibmgnr::Exception 209
- ibmgnr::NwException 219
- identificación del entorno cultural
  - description 42
- identificar
  - categorías de nombres 76
  - entorno cultural de campos de nombre 45
  - entorno cultural de nombre completo 44

- identificar (*continuación*)
  - entorno cultural de un nombre de organización 45
  - género de nombre completo 46
  - género de nombre de pila 47
  - género utilizando NameWorks 46
  - país de asociación para nombres completos 48
  - país de asociación para nombres de pila y apellidos 49
- identificar el entorno cultural
  - description 42
  - utilización de NameWorks 42
- información de requisitos previos v
- información relacionada v
- Interfaz de mensajes XML de Distributed Search 157

## L

- LDAP 171
- lista de cadenas de caracteres externa
  - description 29
- lista de cadenas de caracteres personalizada
  - description 29
- listas de comprobaciones
  - lista de comprobación para la resolución de problemas de productos generales 207
- listas de datos
  - actualizar nombres e información asociada 55
  - description 54
  - gestionar 53
  - suprimir nombres 57
- listas de nombres
  - description 28
  - gestionar listas de nombres de Enterprise Name Search 141, 176
  - lista de cadenas de caracteres externa 29
  - lista de cadenas de caracteres personalizada 29

## M

- mandatos de NameLoader 180
- marcas registradas 255
- mejoras
  - versión 6.0 8
- mejoras del producto
  - versión 6.0 2
- método analyzeForSearch() 60, 72
  - crear objetos QueryName 62
- Método compare() 78
- Método createName()
  - crear objetos Name 62
- Método search() 64, 80
- métodos
  - categorize() 76
  - compare() 78
  - dateCompare() 79
  - dateDifference() 79
  - genderize() 46
  - genderizeField() 47

- métodos (*continuación*)

- getVariants() 51
- parse() 35
- search() 64
- modificar 100

## N

- Name data archive
  - description 28
- Name Preprocessor 137
- NameHunter
  - configurar para utilizar archivos de reglas de transliteración 131
  - parámetros de comparación 100
  - Parámetros de comparación 100
- NameHunter Developer's Toolkit 95
- NameWorks 219
  - códigos de error 219
  - identificar el entorno cultural 42
- NDA
  - description 28
- nombres
  - afijos 24
  - analizados 20
  - buscar 80
  - cadenas de caracteres raíz 28
  - calificadores 24
  - categorías 14, 73
  - categorizar 76
  - componentes y estructura 14, 73
  - conjuntos 17
  - fórmulas de tratamiento 24
  - generación de variantes de nombre utilizando NameWorks 49
  - identificación de un entorno cultural utilizando NameWorks 42
  - identificación del entorno cultural 42
  - identificación del entorno cultural de nombres de organización 45
  - identificar entorno cultural de nombres completos 44
  - identificar género 47
  - identificar género de nombre completo 46
  - modelo de nombres 14, 73
  - nombres de organización 18, 75
  - nombres de persona 14, 73
  - recuperar datos complementarios 81
  - TAQ 24
  - transliteración 29
  - visión general 13
- nombres conjuntos 17
- nombres de organización
  - afijo de organización 25
  - description 18, 75
  - designador de organización 25
  - palabra ignorada 25
  - TAQ 25
- nombres de persona
  - afijos 24
  - calificadores 24
  - comparar mediante IBM NameWorks 78
  - componentes y estructura 14, 73
  - fórmulas de tratamiento 24
  - modelo de nombres 14, 73

- nombres de persona (*continuación*)
  - TAQ 24
- nombres de pila
  - description 21
  - generar variantes de nombres 51
  - identificar entorno cultural 45
  - identificar país de asociación 48, 49
- nombres de variante
  - programa de utilidad de conversión 91
- notas
  - release 2
- novedad en la versión 6.0 2

## O

- Objeto Name
  - crear 62
- Objeto QueryName
  - crear 62

## P

- país de asociación
  - identificación utilizando NameWorks 47
  - identificar para nombres completos 48
  - identificar para nombres de pila y apellidos 49
  - identificar utilizando el método analyze() 41
- país de origen
  - description 47
  - frecuencia 47
  - importancia 47
  - seguridad 47
- Palabra ignorada
  - SW 27
- parámetros de comparación 100
  - alteraciones temporales 68
    - CompParmOverrides, clase 71
    - Configuration, clase 71
    - estrategias de búsqueda 70
  - Distributed Search 121
  - NameHunter 121
- Parámetros de comparación 100
  - NameHunter 100
- parse(), método 35
- preproceso de nombres 137
- problemas y soluciones
  - buscar en las bases de conocimientos 239
- Proceso de Distributed Search
  - configurar para utilizar archivos de reglas de transliteración 132
- producto, arquitectura
  - Global Name Management 9
- programa de utilidad Name Loader
  - cargar nombres de listas de nombres 184
- programa de utilidad NameLoader
  - suprimir nombres de listas de nombres 187

- programa de utilidad NameLoader (*continuación*)
  - suprimir todos los nombres del esquema de Enterprise Name Search 187
- programas de utilidad
  - sintaxis del programa de utilidad wpspwd, Enterprise Name Search 171

## R

- rastreo
  - establecer nivel de registro para IBM NameWorks 240
- recursos técnicos
  - buscar 239
  - suscribirse a actualizaciones de soporte 242
- regularización de nombres 93
- resolución de problemas
  - buscar en las bases de conocimientos 239
  - comprobar archivos de registro 240
  - contactar con el soporte de IBM 241
  - lista de comprobación para la resolución de problemas de productos generales 207
  - rastreo 240
  - suscribirse a actualizaciones de soporte 242

## S

- search()
  - métodos 80
- Sección Comparison Files (archivos de comparación) 201
- Sección Custom (cadenas de caracteres personalizadas) 193
- sección Datalist 194
- Sección DateCompare 199
- sección General 192
- Sección Reference Files (archivos de referencia) 201
- sección Search Strategy 197
- Sección Transliteration Modules (módulos de transliteración) 200
- seguridad
  - archivo de contraseñas (ens.passwd), Enterprise Name Search 171
  - asignar grupos de seguridad a usuarios, Enterprise Name Search 172
  - asignar grupos de seguridad basados en roles a usuarios, Enterprise Name Search 172
  - cambiar contraseñas de usuario, Enterprise Name Search 174
  - control de acceso, Enterprise Name Search 174
  - crear contraseñas, Enterprise Name Search 172
  - crear nombres de usuario, Enterprise Name Search 172

- seguridad (*continuación*)
  - restablecimiento de contraseñas de usuario, Enterprise Name Search 174
  - sintaxis del programa de utilidad wpspwd, Enterprise Name Search 171
  - suprimir usuarios, Enterprise Name Search 174
- Solicitudes XML de Distributed Search 157
- solicitudes XML de DS 157
- soporte
  - buscar soluciones a los problemas 239
  - contactar 241
  - suscribirse a actualizaciones de soporte 242
- soporte de IBM
  - contactar 241
- Strategy, clase
  - alteraciones temporales 68
- suprimir
  - nombres en listas de datos 57
- suscripción
  - suscribirse a actualizaciones de soporte 242

## T

- TAQ
  - afijo de organización 25
  - description 24
  - designador de organización 25
  - formato de archivo 83, 88
  - palabra ignorada 25
- TAQ de nombres de organización
  - Afijo de organización 28
  - Designador de organización 25
  - Palabra ignorada 27
- tareas de administrador del sistema
  - gestionar listas de datos 53
- términos
  - formato de archivo 91
- transliteración 29
  - archivos de reglas 30

## U

- updateName()
  - métodos 55
- updateNameInDatalist(), método 55

## V

- valores de configuración
  - modificar y actualizar 190
- variante
  - formato de archivo 89
- variantes
  - description 49
  - generación utilizando NameWorks 49
  - generar para nombres completos 51
  - generar para nombres de pila y apellidos 51

- Variantes de cadena de caracteres de nombre de NameHunter 88
- variantes de nombre
  - description 49
  - generación utilizando NameWorks 49
  - generar para nombres completos 51
  - generar para nombres de pila y apellidos 51
- variantes de nombres
  - generar lista utilizando el método analyze() 41
- visión general
  - Global Name Management 1
- Visión general de NameHunter 82







Impreso en España