



IBM Text Analytics Group

Análisis conceptual de textos y relaciones en la WWW

Angel Hernandez Bravo
Executive I/T Architect
IBM Software

Requerimientos (Recolección de Datos)

- Ejecución de Búsquedas.
 - RF1: Capacidad de Multibuscador (tipo Copernic).
 - RF2: Posibilidad de búsquedas “avanzadas”.
 - RF3: Búsquedas en Bases de Datos documentales:
 - Sociología.
 - Educación.
 - Información sobre Estados.
 - Información sobre economía.
 - Información sobre Patentes / Tecnología / Ciencia.
 - Información sobre empresas.
 - RF4: Búsquedas en Web 2.0 (Foros de discusión, e-mails, news, blogs, redes sociales, comunidades virtuales)
 - RF5: Búsquedas en Imágenes, sonido y multimedia.
 - RF6: Búsquedas en revistas especializadas.

Requerimientos (Recolección de Datos)

- Generación de Alertas.
 - RF7: Creación y mantenimiento de listas de “favoritos”.
 - RF8: Creación y mantenimiento de “alertas”.

Requerimientos (Recolección de Datos)

- Generación de Informes.
 - RF9: Generación y mantenimiento de un Tesauro propio.
 - RF10: Creación de Bibliografías especializadas.
 - RF11: Generación de estadísticas de flujo informativo.
 - RF12: Webmetría.
 - RF13: Link análisis.
 - RF14: Generación automática de informes estándar de INTELIGENCIA

IBM Threat and Intelligence Solution

Analysis & Dispatch

Information from multiple sources



Databases/
Legacy Systems



World
Wide Web



E-mail



Broadcast /
Phone
Intercept



Watch Lists



Documents



Web Text
Media



Scanned
Documents



- Multiple Information Sources
- Extract key data from:
 - structured files
 - unstructured text
 - audio output
- Recognize, Resolve, Relate
- Generate Real Time alerts



Multiple Agencies



Analysts



Law Enforcement



Investigators



Customs, Ports &
Borders

The right information, to the right person, at the right time

Counter Terrorism Center

Welcome: Intelligence Analyst - LVMP-OID 001378
 Time and date: 10:22 AM Wednesday, 8/21/07



Alerts

National Threat Advisory:
ELEVATED

Significant Risk Of Terrorist Attacks

The threat level in the airline sector is HIGH or Orange. Read more

EAS Alert Status:
Conflict Alert:

●

No New Alerts

Last notification: 04: 15: 44
04/09/2007

Persistent Search Alert:

New alerts

Last notification: 09: 03: 33
04/09/2007

[View Alerts](#)

911 Tactical Alert Status

●

Currently there are no active tactical alerts

- Operations** [edit] [help] [close]
- [Records Management System](#)
 - [Registered Offenders](#)
 - [Gang Tracking System](#)
 - [Security Threat Group Hits](#)
 - [COMSTAT](#)
 - [OmniFind Yahoo Crawler](#)
 - [Rental Car Inquiries](#)
 - [Hotel Reservation Inquiries](#)
 - [McCarran Airport Inquiries](#)
 - [Jail Inquiries](#)
 - [Case Management](#)
 - [Computer Aided Dispatch](#)
 - [ScopeJLink](#)
 - [Crossroads](#)
 - [CrimeWeb](#)
 - [Family Tracs](#)

- Agency Links** [edit] [help] [close]
- [CAL-Gang](#)
 - [LEMS](#)
 - [NLETS](#)
 - [INS Registrants](#)
 - [Most Wanted Terrorists](#)

PERF [edit] [help] [close]

HOME

- [Home](#)
- [About PERF](#) ▶
- [PERF Membership](#) ▶
- [Join PERF](#)
- [Calendar](#)
- [Book Store](#)
- [News/Announcements](#)
- [Meeting Registration](#)
- [Fellowship Program](#)
- [Active Projects](#)

- Reference** [edit] [help] [close]
- [Global Name Reference Encyclopedia](#)
 - [Medical Information](#)
 - [Local News and Conditions](#)
 - [Statutes and Ordinances](#)
 - [General Info Lookup AT&T](#)
 - [Explosives Information](#)



NEWS Thursday, August 30, 2007

"Good to Great" Policing: Application of Business Management Principles in the Public Sector

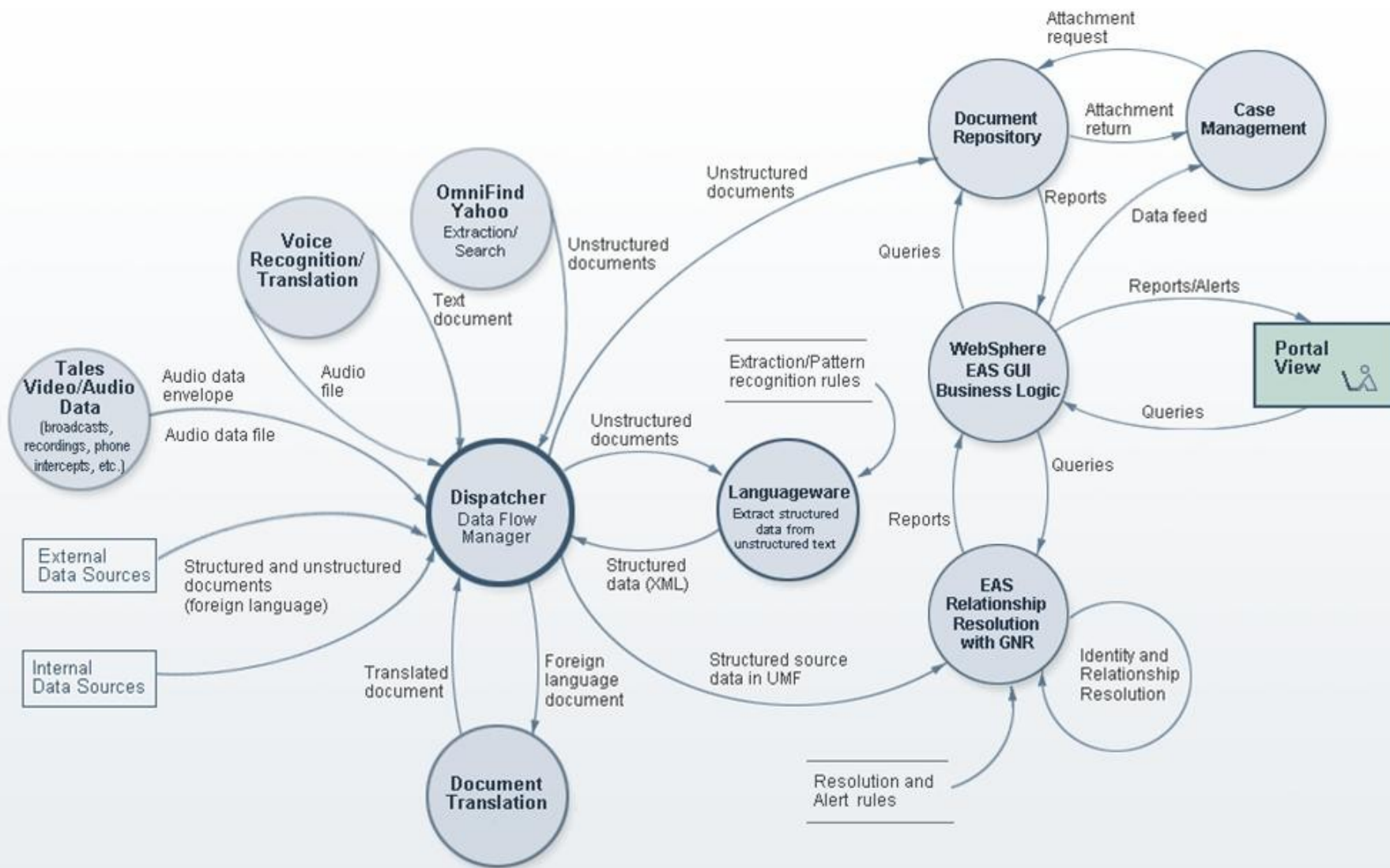
This publication, ["Good to Great" Policing: Application of Business Management Principles in the Public Sector](#), takes the core principles articulated by Jim Collins in his best-selling book Good to Great and applies it to law enforcement..

[Full Story](#)

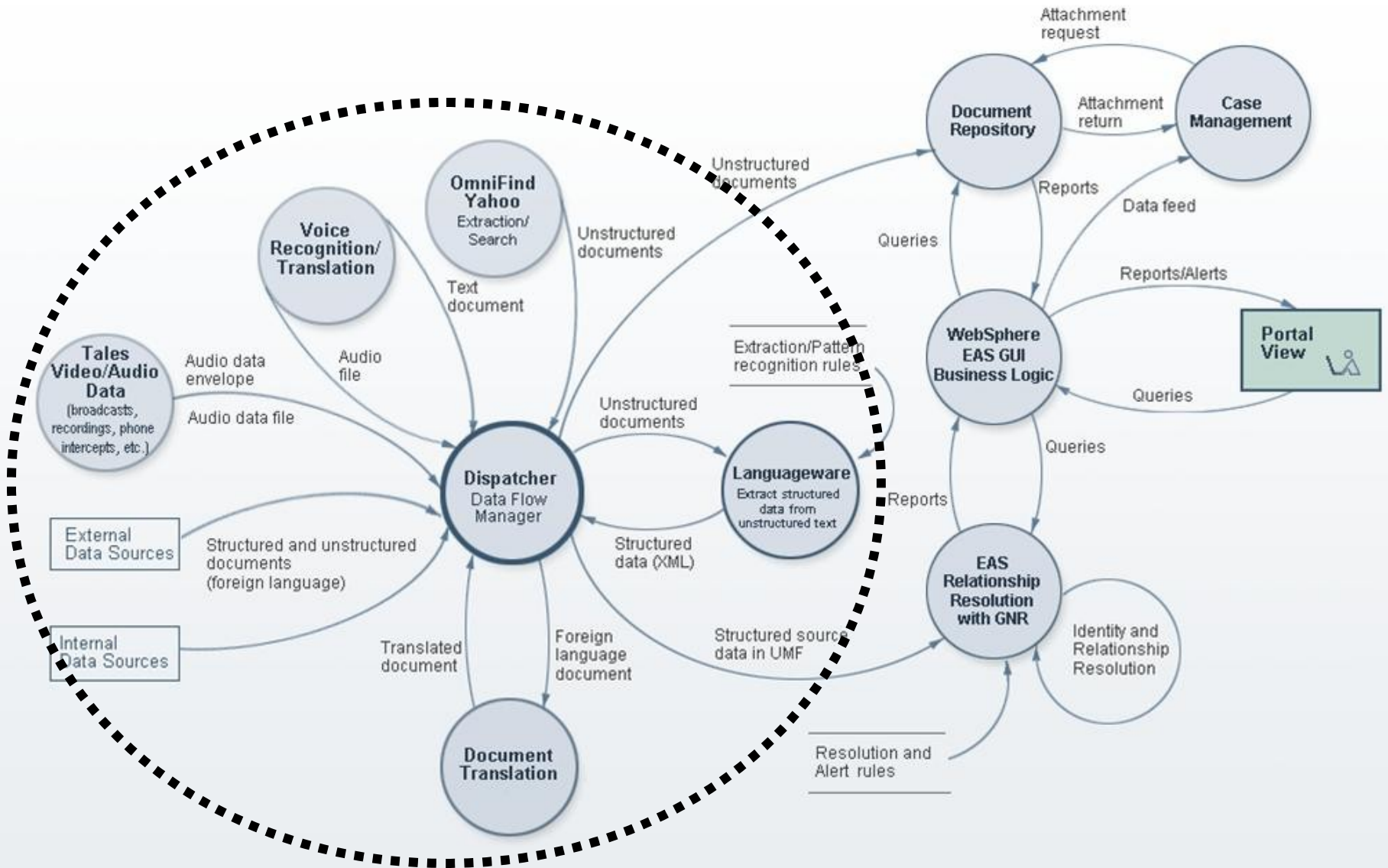
Violent Crime in America: 24 Months of Alarming Trends

- Tools** [edit] [help] [close]
- ✉ [E-mail](#)
 - [Instant Messaging](#)
 - [Photo Imaging System](#)
 - [Web CAM Viewing](#)
 - [Language Translation Processing](#)
 - [License Plate Reader System](#)

IBM Solución de Inteligencia



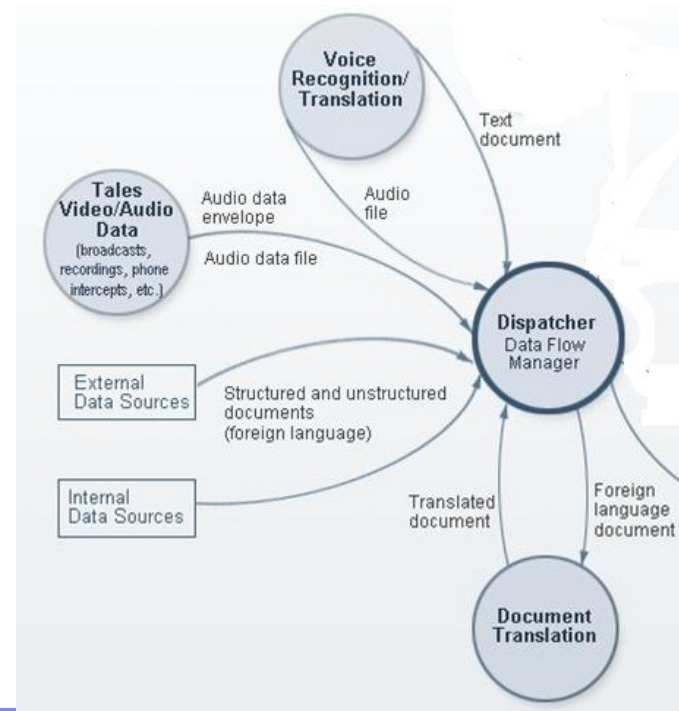
IBM Solución de Inteligencia



Tres fuentes de datos

1) Broadcast Video and Phone intercepts

Name ▲	Type	Packed Size	Has a password	Size	R...	Date
ar-Al_Arabiya-Ne...	Text Document	1 KB	No	2 KB	48%	11/5/2007 12:47 PM
ar-Al_Arabiya-Ne...	MPEG-4 Movie	10,506 KB	No	10,836 KB	4%	10/24/2007 2:38 PM
ar-Al_Arabiya-Ne...	QTT File	1 KB	No	2 KB	47%	10/24/2007 2:39 PM
ar-Al_Arabiya-Ne...	SMIL File	1 KB	No	1 KB	48%	11/5/2007 12:47 PM
ar-Al_Arabiya-Ne...	Text Document	1 KB	No	2 KB	59%	11/5/2007 12:47 PM



Tres fuentes de datos

:

2) WWW search, Crawling files & DBases with OMNIFIND

- Websites, Social Networks
 - Sweeping websites, and sending data for annotation and translation to LanguageWare for further analysis.

- Files
 - User defines which Files and databases to crawl

- Manual Searches
 - User defined manual searches, extract and

Web Site Crawler

"Perhaps the most shocking aspect of this product is how easy it is to get up and running – especially coming from a company like IBM." - Forrester

IBM OmniFind Yahoo! Edition

Open Search Page About Help Log Out

YAHOO!S

Dashboard

Crawl Web Sites

Crawl Files

Manage Search Experience

Manage System

Web sites

Crawler status: **Not running**

Start Crawling Web sites [Add, Remove or Edit Web Sites](#)

Files

You have not entered any directories to crawl.

Enter a directory to crawl:

Example: T:\shared\human_resources

Search

168 documents in the index; index size is 18 MB.

Average query response time is 0 milliseconds.

Average queries per minute is 0.

Tres fuentes de datos

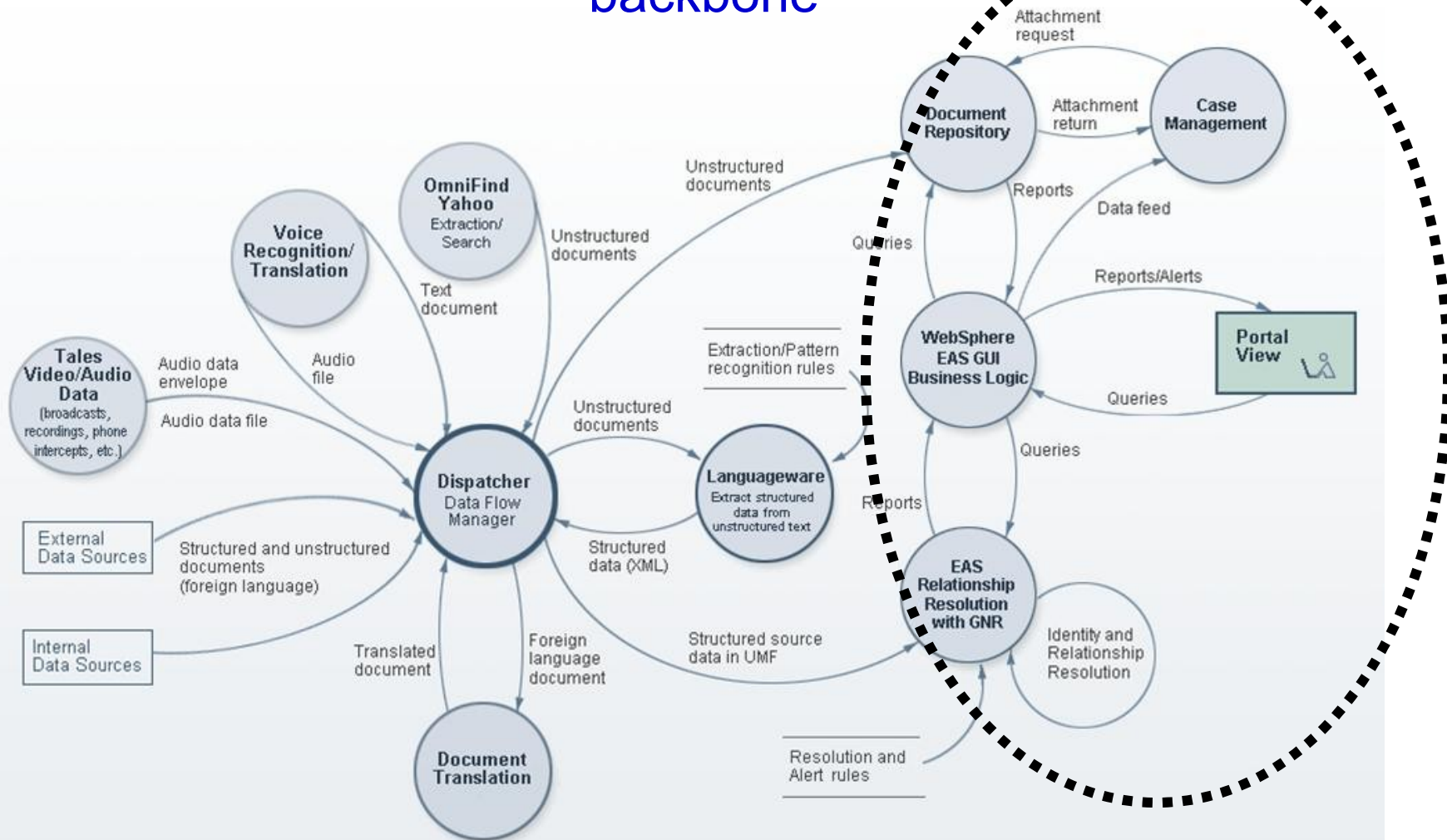
:

3) Unstructured Text / Text Analytics

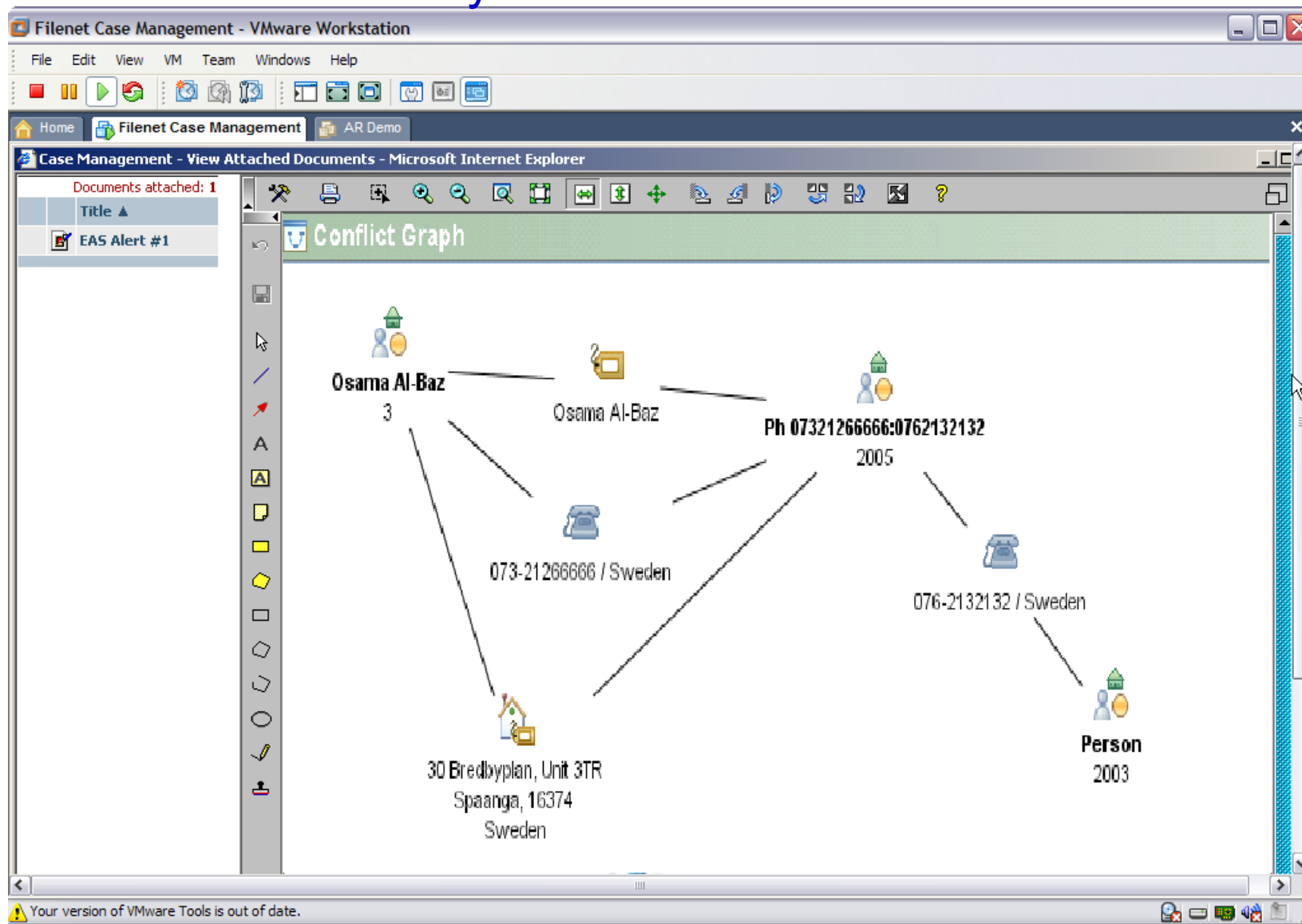
- Turn unstructured text into structured, analyse it, and translate it into 22 languages.
 - Search, Indexing & provide Text Analysis using Triples building Data Segments

IBM Solución de Inteligencia

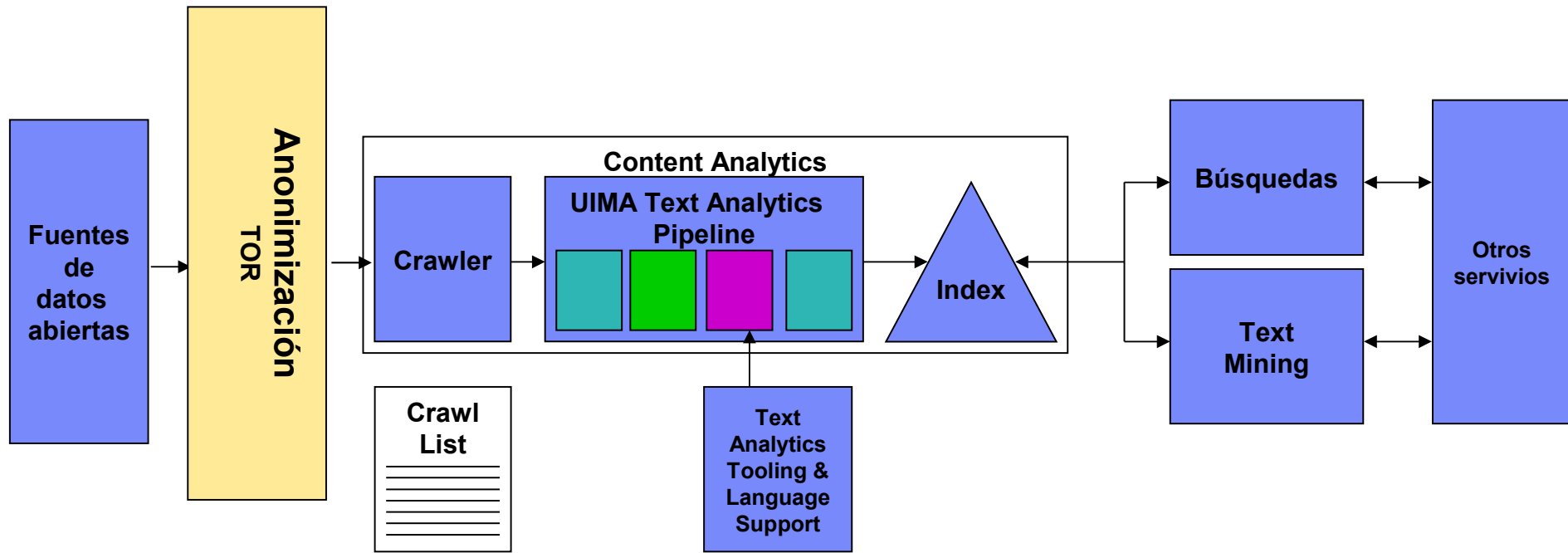
- Document/Case Management and Entity Analytics as backbone



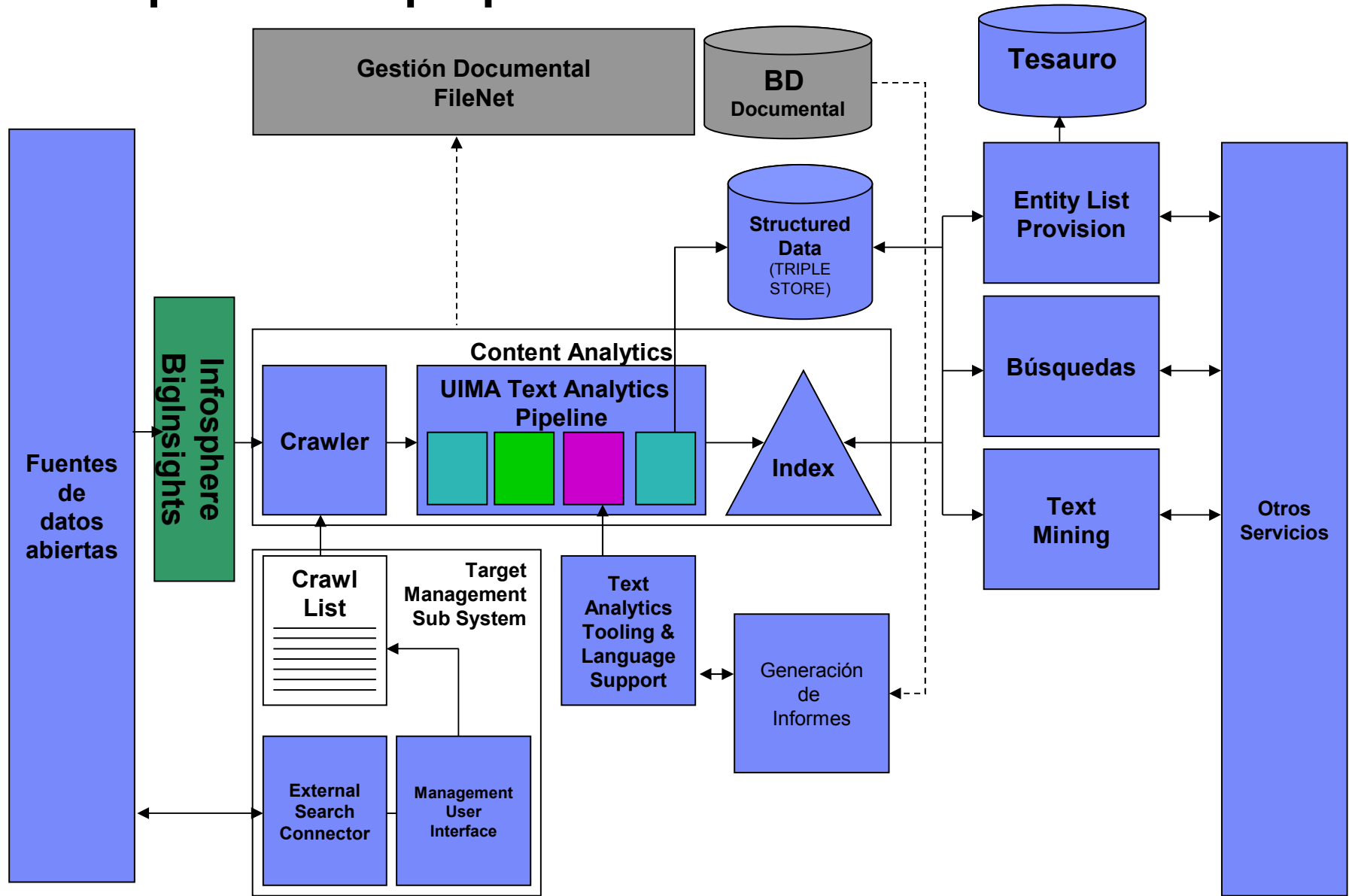
Análisis de identidades y relaciones



Arquitectura propuesta



Arquitectura propuesta



Dossier en FileNet

FileNet Case Management - VMware Workstation

FileNet Case Management - VMware Workstation

FileNet Case Management - Microsoft Internet Explorer

Home | FileNet Case Management | AR Demo

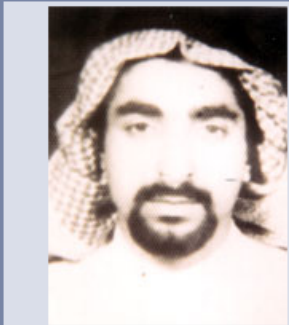
Case Management - Microsoft Internet Explorer

Add Document | Attach Document | Save | Close | Action | Mode: Browse | Create Case | Search | Preferences | Sign Out

FILENET

Dossier | Relationship | Case Notes | Attachments | Audit

General Information

Full Name:	Last name	First name	Middle name/initial	
	Rasheed	Hans		
DOB:	1967/06/26			
Status Indicator/ORI:	CONSOLIDATED WATCHLIST T Indicator			
Demographics:	Ethnicity	Gender	Age	
	ARABIC	M	30	
	Weight	Hair Color	Eye Color	
	145	BLACK	BROWN	
	SMT			
	T-RH-STAR			
Special Designation:	Known Terrorist			

Entity ID: 2 Case Number: 65875

Contact Type	Date	Agency/ORI
ARREST	Thu Nov 08 09:21:34 PST 2007	ABC067671-01

Address	Date	Agency/ORI
4041 E Sunset Rd, Henderson 89014 USA	Thu Nov 08 09:21:34 PST 2007	ABC067671-01

Vehicle Information

Make	Model	Year	License Plate Number	State	Color	Date	Agency/ORI
GMC	ENVOY	1999	ABC 123	NV	BLACK	2007-11-08 12:15:37.0	ABC067671-01

Case ID:

Received Date:

Inbaskets

- Cases (14)
- My Work (0)

Your version of VMware Tools is out of date.