

IBM DB2 Information Integrator  
OmniFind Edition



# 릴리스 정보

버전 8.2 *Fix Pack 1*



IBM DB2 Information Integrator  
OmniFind Edition



# 릴리스 정보

버전 8.2 *Fix Pack 1*

주:

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 반드시 “주의사항”에 나오는 일반 정보를 읽으십시오.

이 문서에는 IBM의 소유권 정보가 들어 있습니다. 이 문서는 라이선스 계약에 따라 제공되고 저작권법에 의해 보호됩니다. 본 서적에 포함된 정보는 어떠한 제품 보증을 포함하지 않으며 본 매뉴얼에서 제공되는 어떠한 사항도 이와 같이 해석되지 않습니다.

온라인으로 또는 해당 지역의 IBM 담당자를 통해 IBM 책을 주문할 수 있습니다.

- 온라인으로 책을 주문하려면 [www.ibm.com/shop/publications/order](http://www.ibm.com/shop/publications/order)의 IBM Publications Center로 찾아가십시오.
- 해당 지역의 IBM 담당자를 찾으려면 [www.ibm.com/planetwide](http://www.ibm.com/planetwide)의 전세계 IBM 사이트로 찾아가십시오.

IBM에 정보를 보내면, IBM은 귀하의 권리를 침해하지 않는 범위 내에서 IBM이 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

# 목차

1 릴리스 정보에 대한 정보 . . . . .	v	1 일본어 및 중국어 문서의 공백에서 줄 바꾸기 문자	
1 웹의 기술 지원 정보. . . . .	v	1 제거. . . . .	26
1 DB2 II OmniFind Edition, 버전 8.2.1 설치 . . . . .	1	1 컬렉션 구문 분석기 및 Stellent 세션에 대한 문서 유	
1 DB2 II OmniFind Edition 설치 제거 . . . . .	3	1 형 구성. . . . .	27
1 이 릴리스의 새로운 기능 . . . . .	5	1 컬렉션 구문 분석기에 대한 문서 유형 구성. . . . .	28
1 새 지원 운영 체제 및 데이터 소스 유형 . . . . .	5	1 Stellent 세션에 대한 문서 유형 구성. . . . .	30
1 사용자 정의 컬렉션 ID 지정 . . . . .	5	1 쿼리 처리의 향상 . . . . .	35
1 암호 보호 웹 사이트 크롤링 . . . . .	6	1 동일한 웹 사이트의 결과 축소 . . . . .	35
1 웹 페이지의 프레임 세트 크롤링. . . . .	9	1 검색 결과의 강조표시 확장 . . . . .	35
1 웹 문서의 META 태그로 보안 토큰 구성 . . . . .	9	1 검색 응용프로그램의 시간종료 지원 . . . . .	36
1 금지된 문서 링크의 앵커 텍스트 인덱스화 . . . . .	10		
1 웹 페이지에서 no-index 및 no-follow 지시문 겹쳐		알려진 제한사항, 문제점 및 해결 방법 . . . . .	37
1 쓰기. . . . .	10	AIX용 DB2 UDB tar 파일 압축 풀기 . . . . .	37
1 웹 크롤러에 대한 글로벌 크롤 스페이스 구성 . . . . .	11	AIX의 공유 메모리 세그먼트 . . . . .	38
1 다중 Lotus Notes 서버 구성 . . . . .	13	Windows에서 설치 제거한 후 시스템 재시작 . . . . .	38
1 페이지를 크롤할 Notes 서버 지정 . . . . .	13	Linux 및 AIX용 상위 항목이 삭제된 URLFetcher	
1 VeniceBridge 크롤러의 성능 향상. . . . .	16	프로세스 . . . . .	38
1 Windows 파일 시스템의 로그인 정보 지정. . . . .	16	서버에서의 인덱스 위치 . . . . .	39
1 외부 이름이 표시 이름으로 변경됨. . . . .	17	힙(heap) 크기 한계. . . . .	39
1 데이터 리스너 구성 및 모니터링 . . . . .	17	대형 오브젝트(LOB)가 있는 DB2 UDB 테이블 . . . . .	39
1 시스템 보기: 데이터 리스너 페이지 . . . . .	18	인덱스 상태가 표시되지 않음 . . . . .	39
1 시스템 보기 모니터: 데이터 리스너 페이지 . . . . .	19	Linux 및 AIX용 사용자 ID 권한. . . . .	40
1 데이터 리스너 세부사항 페이지. . . . .	19	1 FileNet 콘텐츠 서비스에 대한 직접 문서 액세스 사	
1 데이터 리스너 구성요소 중지 및 시작. . . . .	20	1 용 . . . . .	40
1 컬렉션의 문서 수 계산. . . . .	21	1 별칭 테이블 크롤링 제한사항 . . . . .	40
1 컬렉션 마법사에서 시스템 자원 자동 점검 . . . . .	22		
1 구문 분석기 활동 모니터링 . . . . .	22	주의사항 . . . . .	43
1 구문 분석기 세부사항 페이지 . . . . .	22	상표. . . . .	45
1 HTML 메타데이터를 검색 필드에 맵핑 . . . . .	25		
1 HTML 메타데이터 필드 맵핑 페이지. . . . .	25	IBM에 문의 . . . . .	47
		제품 정보 . . . . .	47
		문서에 대한 의견 제공 . . . . .	47



---

## 1 릴리스 정보에 대한 정보

1 릴리스 정보 문서에서는 이 릴리스의 새로운 기능, 알려진 제한사항, 문제점, 문제해결  
1 방법 및 IBM® DB2® Information Integrator OmniFind™ Edition, 버전 8.2.1(버전  
1 8.2와 Fix Pack 1)의 문서 갱신에 대해 설명합니다.

1 IBM DB2 Information Integrator OmniFind Edition, 버전 8.2.1은 번역되지 않았습  
1 니다. 제품 인터페이스 및 메시지의 변경사항은 영어로만 제공됩니다.

---

## 1 웹의 기술 지원 정보

1 DB2 II OmniFind Edition 제품은 최근에 WebSphere® Information Integrator  
1 OmniFind Edition으로 이름이 바뀌었습니다. 제품 관련 웹 페이지에서는 WebSphere  
1 Information Integrator OmniFind Edition에 대한 언급을 볼 수 있지만, 제품 인터페  
1 이스, 문서 및 문제점 해결 정보(기술 정보)에서는 여전히 DB2 상표를 반영합니다.

1 다음 웹 사이트에서 기술 지원 정보를 찾을 수 있습니다.

### 1 **IBM DB2 Information Integrator OmniFind Edition**

1 [www.ibm.com/software/data/integration/db2ii/supportomni.html](http://www.ibm.com/software/data/integration/db2ii/supportomni.html)

### 1 **IBM WebSphere Information Integrator**

1 <http://www.ibm.com/software/data/integration/db2ii/support.html>

### 1 **IBM WebSphere Application Server**

1 <http://www.ibm.com/software/webservers/appserv/was/support/>

### 1 **Linux™, UNIX® 및 Windows®용 IBM DB2 Universal Database™ 버전 8 제품**

1 [www.ibm.com/software/data/db2/udb/winos2unix/support](http://www.ibm.com/software/data/db2/udb/winos2unix/support)





## 1 DB2 II OmniFind Edition, 버전 8.2.1 설치

1 DB2 II OmniFind Edition 설치 런치패드를 사용하여 전체 제품(버전 8.2.1)을 설치하  
1거나 기존 설치에 버전 8.2.1 업그레이드만 설치할 수 있습니다.

1 버전 8.2.1 업그레이드를 설치하기 전에 기존 설치를 백업해야 합니다. 시스템 백업 및  
1복구에 대한 정보는 *Administering Enterprise Search*를 참조하십시오. 설치 중에 기  
1존 데이터 디렉토리의 경로를 지정하십시오. 설치 런치패드가 기존 설치 정보를 사용하  
1여 업그레이드를 설치합니다.

1 전체 버전 8.2.1 제품을 설치하려고 하고 이전에 DB2 II OmniFind Edition 버전 8.2  
1를 설치한 경우, 먼저 DB2 II OmniFind Edition을 설치 제거해야 합니다. 기존 설치  
1위에 전체 제품을 설치하거나, 기존 시스템을 백업하고 버전 8.2.1을 설치한 후 버전 8.2  
1구성 설정을 버전 8.2.1 시스템으로 리스토어할 수 없습니다.

1 버전 8.2.1을 설치하려면 다음을 수행하십시오.

- 1 1. 설치 런치패드를 시작하십시오.

옵션	설명
Linux	root 사용자로 로그인하고 ./install-linux.bin을 입력하십시오.
AIX	root 사용자로 로그인하고 ./install-aix.bin을 입력하십시오.
Windows	Administrator 권한이 있어야 합니다. install-win32.exe 파일을 더블 클릭하거나 install-win32.exe를 입력하십시오.

- 1 2. 설치 런치패드가 열리면 제품 설치를 누르십시오.

- 1 3. 버전 8.2.1 업그레이드를 설치하려면 다음을 수행하십시오.

- 1 a. 업그레이드 설치 창에서 기존 **DB2 II OmniFind Edition** 설치 업그레이드를  
1선택하십시오. 이 옵션을 선택하면, 기존 버전 8.2 시스템 구성이 버전 8.2.1 시  
1스템으로 이주됩니다.

- 1 b. 기존 데이터 디렉토리의 경로를 입력하십시오. 다음을 누르십시오.

- 1 c. 요약 창의 설정값을 검토하십시오. 다음을 눌러 설치를 시작하십시오.

- 1 4. 전체 버전 8.2.1 제품을 설치하려면 다음을 수행하십시오.

- 1 a. 새 버전의 **DB2 II OmniFind Edition** 설치를 선택하십시오. 이전에 DB2 II  
1OmniFind Edition 버전 8.2를 설치한 경우, 이 옵션을 선택하기 전에 이를 설  
1치 제거하십시오.

- 1                                    b. *Installation Guide for Enterprise Search*의 프로시저에 따라 제품을 설치하십시오.
- 1

## 1 DB2 II OmniFind Edition 설치 제거

1 설치 제거 프로그램은 엔터프라이즈 검색 구성요소를 제거합니다. 모든 데이터 시스템  
1 구성을 제거할 수도 있습니다.

1 DB2 II OmniFind Edition을 설치 제거하기 전에 시스템을 백업해야 합니다.

1 다중 서버 구성의 경우, 각 서버에서 설치 제거 프로그램을 실행하십시오. DB2 II  
1 OmniFind Edition을 설치 제거하려면 다음을 수행하십시오.

- 1 1. Linux 및 AIX의 경우, root 사용자로 로그인하십시오.
- 1 2. 설치 디렉토리(ES\_INSTALL\_ROOT)에서 \_uninst 디렉토리를 찾아보거나 변경한  
1 후 설치 제거 프로그램을 시작하십시오.

1

옵션	설명
Linux 또는 AIX	./uninstall.bin을 입력하십시오.
Windows	uninstall.exe 파일을 더블 클릭하거나 입력하십시오.

1

1

1

1

- 1 3. 선택적: 모든 엔터프라이즈 검색 관리자 구성 정보를 제거하려면 Remove all data  
1 and configuration files 선택란을 선택하십시오.

1 **경고:** 이 선택란을 선택하면, 시스템의 모든 콜렉션에 대한 모든 데이터가 제거되  
1 1니다.

1 Windows의 경우, DB2 II OmniFind Edition을 다시 설치하기 전에 컴퓨터를 재  
1 시작해야 합니다.



## 1 이 릴리스의 새로운 기능

1 DB2 II OmniFind Edition의 새 기능에는 검색 관리 향상, 검색 품질과 성능 향상 및  
1 Microsoft® Windows 2003 지원이 포함됩니다.

## 1 새 지원 운영 체제 및 데이터 소스 유형

1 DB2 II OmniFind Edition은 32비트 Windows Server 2003 운영 체제(Standard Server,  
1 Advanced Server 및 Datacenter Server)를 지원하며 Microsoft SQL Server 데이터  
1 베이스를 크롤할 수 있습니다.

1 Windows Server 2003을 설치하려면, *Installation Guide for Enterprise Search*에서  
1 Windows Server 2000과 동일한 지시사항을 따르십시오.

1 SQL Server 데이터베이스를 크롤하려면 엔터프라이즈 검색 관리 콘솔을 사용하여 DB2  
1 크롤러를 작성하고 크롤할 데이터 소스를 구성하십시오.

## 1 사용자 정의 컬렉션 ID 지정

1 컬렉션을 작성할 때, 시스템이 작성하는 내부 ID를 사용하기로 선택하거나 사용자 정  
1 의 컬렉션 ID를 지정할 수 있습니다.

1 사용자 정의 컬렉션 ID를 지정하는 경우, 검색 응용프로그램이 엔터프라이즈 검색  
1 API(Application Programming Interface)를 사용하여 시스템이 작성하는 잠재적 암호  
1 화 ID 대신 이 ID를 갖는 컬렉션을 호출할 수 있습니다.

1 사용자 정의 컬렉션 ID를 지정하는 경우, 컬렉션을 작성할 때 지정해야 합니다. 나중에  
1 컬렉션을 편집하여 이 ID를 변경할 수 없습니다.

1 사용할 컬렉션 ID 유형을 지정하려면 다음을 수행하십시오.

- 1 1. 엔터프라이즈 검색 관리 콘솔을 시작하십시오. 보안이 사용 가능한 경우, 엔터프라이  
1 즈 검색 관리자 권한이 있는 사용자로 로그인하십시오.
- 1 2. 마법사를 사용하여 컬렉션을 작성할 것인지 여부에 따라 컬렉션 작성 또는 컬렉션  
1 마법사를 누르십시오.
- 1 3. 컬렉션 ID 필드에서 다음 옵션 중 하나를 선택하십시오.

1 **디폴트 컬렉션 ID**

1 시스템이 컬렉션의 내부 ID를 작성하도록 하려면 이 옵션을 선택하십시오.

## 사용자 정의 컬렉션 ID

컬렉션에 대해 사용자 고유 ID를 지정하려면 이 옵션을 선택하십시오. 지정하는 값에는 영숫자 문자 및 밑줄 문자만 포함될 수 있습니다.

---

## 암호 보호 웹 사이트 크롤링

DB2 II OmniFind Edition 웹 크롤러는 HTTP 기본 인증을 사용하고 쿠키에 정보를 보유함으로써 암호로 보호된 웹 사이트를 크롤할 수 있습니다.

### HTTP 기본 인증 지원

엔터프라이즈 검색 관리자는 크롤을 시작하기 전에 크롤할 암호 보호 웹 사이트를 식별하고 `httpauth.ini` 구성 파일에 필요한 증명서를 지정할 수 있습니다.

`httpauth.ini` 파일은 다음 포매팅 규칙을 따라야 합니다.

- `httpauth.ini` 파일의 내용은 텍스트이며 공백을 제외한 비인쇄 문자를 포함할 수 없습니다. 데이터 항목은 줄 바꾸기로 구분됩니다.
- 파일에 권한 부여 항목, 주석 및 공백 라인이 임의의 순서대로 포함될 수 있습니다.
- 주석은 공백이 아닌 첫 번째 문자가 #인 행입니다.
- 주석은 주석이 아닌 텍스트가 있는 행에 표시될 수 없습니다.
- 각 권한 부여 항목에는 다음 정보가 단일 행에 다음 순서대로 포함되어야 합니다.
  - 목표 URL(권한 부여가 필요한 서버 파일 시스템의 루트)
  - 서버에 의미 있는 범주 이름
  - 로그인에 사용되는 사용자 이름
  - 로그인에 사용되는 암호
  - 재인증 URL
  - 세미콜론으로 분리된 보안 토큰 목록. 이 목록은 비어 있을 수 있지만 필드가 다섯 번째 필드 뒤에 쉼표 구분자와 함께 여전히 존재해야 합니다.
- 행의 필드는 쉼표로 분리됩니다. 쉼표 주위에 공백이 허용됩니다.
- 각 항목의 암호는 DB2 II OmniFind Edition 나머지 부분에서 사용된 것과 동일한 암호화 메소드로 암호화됩니다.
- URL에 대한 권한 부여는 서버 파일 시스템에서 지정된 URL 아래의 모든 URL에 대해 유효한 것으로 가정됩니다.

`httpauthedit`라는 명령행 유틸리티를 사용하여 `httpauth.ini` 파일을 수정할 수 있습니다. 이 유틸리티를 사용하면, 항목을 작성하거나 편집할 때 명백한 텍스트(암호화되지 않은)로 암호를 입력할 수 있지만 유틸리티가 암호를 디스크에 기록하기 전에 암호화합니다. 암호를 변경하려면 항목을 삭제한 후 다시 추가해야 합니다.

`httpauthedit` 유틸리티를 시작하려면 다음 명령을 한 행에 입력하십시오.

```
java -classpath ES_INSTALL_ROOT/lib/URLFetcher.jar:esinstall/lib/es.oss.jar
com.ibm.es.wc.uf.HTTPAuthEdit configdir
```

`ES_INSTALL_ROOT`는 DB2 II OmniFind Edition 설치 디렉토리이며, `configdir`은 `httpauth.ini` 파일을 작성할(또는 수정) 크롤러 인스턴스의 마스터 구성 디렉토리입니다.

마스터 구성 디렉토리는 `ES_NODE_ROOT/master_config/CR_INST_DIR`에 있으며, 여기서 `ES_NODE_ROOT`는 `ES_CFG` 파일에서 `NodeDirectory`의 값이고 `CR_INST_DIR`은 새 웹 크롤러 인스턴스에 대해 작성한 디렉토리입니다(관리 소프트웨어에 의해 임의의 이름이 제공됨). 올바른 인스턴스 구성으로 작업하고 있는지 확인하십시오. 변경사항은 다음에 크롤러를 시작할 때까지 적용되지 않습니다.

`httpauthedit` 유틸리티는 다음 명령을 허용합니다.

표 1. `HTTPAUTHEDIT` 유틸리티에서 허용되는 명령

명령	설명
?	도움말: 명령 세트에 대해 설명합니다.
b	찾아보기: 파일 내용을 인쇄합니다.
a	<p>추가: 새 권한 부여 항목을 파일에 추가합니다. 이 명령은 단일 행에서 쉼표로 분리된 5개 또는 6개의 인수를 필요로 합니다. 예: <code>http://www.ibm.com/solutions/</code>, <code>solutions</code>, <code>wwwsearch</code>, <code>foobar</code>, <code>xyz123</code>, <code>http://www.ibm.com/solutions/</code>, <code>group1;group2;admin3</code></p> <ul style="list-style-type: none"> <li>• 목표 URL: <code>http://www.ibm.com/solutions/</code></li> <li>• 범주 이름: <code>solutions</code></li> <li>• 사용자 이름: <code>wwwsearch</code>, <code>foobar</code></li> <li>• 암호: <code>xyz123</code></li> <li>• 재인증 URL: <code>http://www.ibm.com/solutions/</code></li> <li>• 선택적: 보안 토큰: <code>group1;group2;admin3</code></li> </ul>
d	삭제: 파일에서 권한 부여 항목을 제거합니다. 이 명령에는 하나의 인수가 필요합니다: <code>http://www.ibm.com/solutions/</code> 와 같은 파일에서 제거할 인증 항목의 목표 URL.
q	나감: 프로그램을 종료합니다.

재인증 URL은 목표 URL과 동일할 수 있습니다.

- 재인증 URL은 권한 부여 없이 요청된 경우, 클라이언트가 권한 부여를 제출할 범주에 대해 HTTP 기본 요청으로 서버의 401 응답을 트리거하는 URL이어야 합니다.
- 재인증 URL은 경로 재지정 없이 이 응답을 트리거해야 합니다(재인증 URL은 301, 302 등이 아니며, 클라이언트를 다른 페이지로 경로 재지정하는 순수한 목적지 페이지이어야 합니다). 재인증 URL은 이 URL이 권한 부여하는 범주의 루트 URL과 동일하지 않은 경우도 있습니다. 재인증 URL이 루트 URL과 동일한지 여부는 웹 사이트의 내부 논리에 따라 결정됩니다. 웹 사이트 관리자가 세부사항을 제공할 수 있습니다.

1 앞뒤 공백은 필드 값에서 제거됩니다. 유효하지 않은 값을 입력하면, httpauthedit가 오  
1 류 메시지를 표시합니다.

1 변경사항은 quit 명령을 입력할 때 저장됩니다. 명시적인 저장 명령은 없습니다. 변경사  
1 항을 취소하기 위해 quit을 사용하는 대신 종료함으로써 프로그램을 종료할 수 있습니  
1 다.

## 1 쿠키

1 웹 크롤러는 서버에서 수신된 쿠키를 보유하여 크롤러 인스턴스 주기 동안 이 쿠키를  
1 사용합니다. 크롤러가 종료되면, 만기되지 않은 모든 쿠키를 디스크에 저장하고 다음에  
1 크롤링 세션을 시작할 때 다시 로드합니다. 쿠키 파일은 cookies.ini이며 크롤러 데이터  
1 디렉토리에 있습니다.

1 크롤러에서 쿠키 조절은 HTTP 인증(httpauth.ini 파일)과는 독립적으로 수행되지만, 웹  
1 서버에서 콘텐츠를 보호하기 위해 쿠키를 사용하는 경우에는 수동으로 제공한 쿠키라도  
1 인증이 필요할 수 있습니다. 웹 사이트 관리자는 사이트 크롤에 필요한 정보를 제공할  
1 수 있어야 합니다.

1 엔터프라이즈 검색 관리자는 크롤러가 발견한 쿠키 이외에 텍스트 편집기를 사용하여  
1 cookies.ini 파일에 쿠키를 추가할 수 있습니다. 이 파일은 크롤러가 종료될 때마다 크  
1 롤러에 의해 재작성되므로 수동으로 입력한 쿠키를 다른 파일에 보관하고 크롤을 시작  
1 하기 전에 필요에 따라 이 쿠키를 추가하십시오. (크롤러는 만기되지 않은 쿠키를 버리  
1 지 않지만, 문제가 발생하여 전체 쿠키 컬렉션을 작성할 수 없는 경우에는 수동으로 지  
1 정한 쿠키가 손실되지 않기를 원합니다.)

1 웹 크롤러에 쿠키를 지정하려면 cookies.ini 파일을 편집해야 합니다. cookies.ini 파일  
1 에서 각 행은 항목입니다. 공백 행 및 주석이 허용되지만 보존되지는 않습니다. 각 항  
1 목의 형식은 다음과 같습니다. 각 항목은 한 행이어야 합니다.

```
1 Cookie(53,40)ASPSESSIONIDQSTACSD=SLNSIDFNLI SDNFLSIDNFLSNL;path=/  
1 https://www.ibm.com:443/help/solutions/
```

1 Cookie는 항목의 시작을 나타내는 필수 키워드입니다. 숫자 53은 쿠키 이름과 값 쌍  
1 및 쿠키와 함께 전송할 기타 내용(경로, 보안 설정값 등)의 문자 길이입니다. 숫자 40  
1 은 유효성 확인을 위해 연관된 URL의 길이입니다.

1 ASPSESSION. . . 의 경우, 이 53개의 문자는 시작 서버로 전송할 쿠키의 내용입니  
1 다. 이 문자열 뒤에는 쉼표(,) 구분자가 나옵니다.

1 URL https://www.ibm.com:443/help/solutions/는 도메인 이름과 같은 요소를  
1 제공하여 쿠키의 유효성을 확인할 수 있습니다. 수동으로 추가한 쿠키의 경우, 쿠키에  
1 대한 보안 및 개인 정보 보호 제한사항을 만족시키는 유효성 확인 URL을 포함시켜야  
1 합니다.



---

## 1 웹 페이지의 프레임 세트 크롤링

1 이제, 웹 크롤러는 HTML 프레임으로 작성된 웹 페이지를 크롤할 수 있습니다.

1 DB2 II OmniFind Edition 버전 8.2에서는 웹 크롤러가 프레임 세트를 구문 분석하  
1 나 프레임에서 다른 HTML 페이지로 링크를 추출할 수 없었습니다.

---

## 1 웹 문서의 META 태그로 보안 토큰 구성

1 웹 크롤러의 새 보안 옵션을 통해 구성 파일에서 META 태그의 이름을 지정할 수 있  
1 습니다. 이러한 태그의 내용은 태그가 있는 문서에 대한 액세스를 제어하기 위한 보안  
1 토큰으로 사용됩니다.

1 DB2 II OmniFind Edition, 버전 8.2에서, 보안 토큰은 글로벌로 정의되었으며, 웹 크  
1 롤러에 대한 Document-level security의 사용 가능 여부에 따라 모든 문서에 적용되지  
1 나 아무 문서에도 적용되지 않았습니다.

1 이제, META 태그를 보안 토큰으로 지정하여 문서에 대한 액세스를 제어할 수 있습니  
1 다. META 태그 기반 보안을 사용하려면 웹 크롤러에 대해 Document-level security  
1 를 사용 가능하게 하고 구성 파일을 편집하여 META 태그 이름을 지정해야 합니다(관  
1 리 콘솔에서 META 태그 이름을 지정할 수 없음).

1 보안 토큰으로 사용할 META 태그를 지정하려면 다음을 수행하십시오.

- 1 1. 엔터프라이즈 검색 관리자로 크롤러 서버에 로그인하십시오(이 사용자 ID 및 암호  
1 는 DB2 II OmniFind Edition 설치 중에 구성되었습니다).
- 1 2. 텍스트 편집기를 사용하여 master\_config 디렉토리 트리에 있는 crawl.properties  
1 파일을 편집하십시오.
- 1 3. 필요할 경우, security\_meta\_tags 매개변수를 삽입하십시오. (security\_meta\_tags  
1 매개변수가 이미 있는 경우, 디폴트값은 nil입니다.)
- 1 4. META 태그 이름을 쉼표 분리 목록으로 지정하십시오.

1 예:

1 security\_meta\_tags=security,acl,access

1 태그 이름의 규칙:

- 1 • 구성 파일에서 읽은 태그 이름은 소문자로 변환됩니다.
  - 1 • 크롤된 문서에 표시된 태그 이름은 구성 파일의 이름과 비교하기 전에 소문자로  
1 변환됩니다.
  - 1 • 이름 목록의 유일한 구분자 문자는 쉼표입니다(공백으로 이름이 분리되지 않음).
  - 1 • 이름에 임베드된 공백은 보유되며, 앞뒤 공백은 제거됩니다.
- 1 5. 파일을 저장하고 종료하십시오.

1 security\_meta\_tags 매개변수에 대해 태그 이름이 구성된 경우, 다운로드한 모든  
1 HTML 페이지를 처리하는 중에 문서의 META 태그를 security\_meta\_tags 목록과  
1 비교하여 점검합니다.

1 보안 토큰을 사용하여 문서 액세스를 제어하는 방법에 대한 자세한 정보는 *Administering*  
1 *Enterprise Search*의 *Document-level security* 주제를 참조하십시오.

---

## 1 금지된 문서 링크의 앵커 텍스트 인덱스화

1 문서에 웹 크롤러가 크롤할 수 없는 문서에 대한 링크가 있는 경우, 이러한 링크의 앵  
1 커 텍스트를 인덱스에 보유할 것인지 여부를 지정할 수 있습니다.

1 robots.txt 파일 또는 metatag 스펙의 설정값으로 웹 크롤러가 웹 사이트의 특정 문  
1 서에 액세스하지 못하도록 할 수 있습니다. 웹 크롤러가 크롤할 수 있는 문서에 금지된  
1 문서에 대한 링크가 있는 경우, 고급 웹 크롤러 등록 정보 페이지에서 이러한 링크의  
1 앵커 텍스트 처리 방법을 지정할 수 있습니다.

- 1 1. 금지된 문서 링크의 앵커 텍스트를 인덱스에 포함시키려면 **금지된 문서 링크의 앵  
1 커 텍스트 인덱스화** 선택란을 선택하십시오. (앵커 텍스트는 웹 문서에서 <A>와 </A>  
1 태그 사이에 표시되는 설명적 텍스트입니다.) 이 선택란을 선택하면, 금지된 페이지  
1 를 가리키는 링크의 앵커 텍스트에 있는 텍스트를 검색하여 해당 페이지를 찾을 수  
1 있습니다.
- 1 2. 인덱스에서 금지된 링크의 앵커 텍스트를 제외시키려면(금지된 문서도 제외), **금지된  
1 문서 링크의 앵커 텍스트 인덱스화** 선택란을 지우십시오.

1 웹 크롤러를 작성할 때 고급 웹 크롤러 등록 정보를 지정하거나 기존 웹 크롤러의 크롤  
1 스페이스를 편집하여 등록 정보를 지정할 수 있습니다.

---

## 1 웹 페이지에서 no-index 및 no-follow 지시문 겹쳐쓰기

1 구성 파일에서 웹 크롤러가 no-follow 또는 no-index 지시문을 포함하는 페이지 또는  
1 인덱스 페이지에 대한 링크를 따르는지 여부를 제어할 수 있습니다. 엔터프라이즈 검색  
1 관리 콘솔에서는 이 기능을 지원하지 않습니다.

1 일부 웹 페이지는 no-index 또는 no-follow 지시문으로 지정되며, 이 지시문은 로봇(예  
1 : 웹 크롤러)이 해당 페이지의 콘텐츠를 인덱스에 포함시키지 않고 해당 페이지에 있는  
1 링크를 따르지 않거나 둘다 수행하지 않도록 지시합니다.

1 DB2 II OmniFind Edition 버전 8.2에서, 이러한 제한사항은 웹 사이트 관리자 또는  
1 콘텐츠 작성자에 의해 시행됩니다. 그러나, 이러한 설정을 제어하면 크롤의 질이 향상됩  
1 니다. 예를 들어, 수천 개의 링크가 포함되지만 기타 유용한 콘텐츠가 없는 디렉토리 페  
1 이지 세트가 있습니다. 이러한 페이지를 크롤해야 하며 링크를 따르지만 페이지 자체를  
1 인덱스화해서는 안됩니다.

크롤러 관리자가 계층 구조 아래쪽으로 크롤하지 않도록 하려고 하지만, 원하는 리프 페이지에 링크가 포함되어 있고 no-follow 지시문이 포함되어 있지 않은 경우도 있습니다. 이러한 페이지 중 일부는 자동 생성되고 소유자가 없으므로 이러한 페이지의 콘텐츠 소유자가 필요한 지시문을 삽입하도록 할 수 있거나 못할 수도 있습니다.

버전 8.2.1 크롤러는 구성 디렉토리에서 followindex.rules라는 새 구성 파일을 찾습니다. 이 파일에는 임의의 수의 follow 및 index 규칙이 포함될 수 있습니다.

```
forbid follow URLprefix
allow follow URLprefix
forbid index URLprefix
allow index URLprefix
```

follow 및 index 규칙에서 IP 주소 및 DNS 호스트 이름을 지정할 수 없으며, URL 접두부(와일드 카드 포함)만 지정할 수 있습니다. 순서는 크롤링 규칙을 지정할 때와 같이 중요합니다. Following 또는 indexing을 명시적으로 허용하거나 금지하는 규칙이 정의된 경우, 목표 문서에 있는 설정을 포함한 기타 설정을 겹쳐씁니다.

---

## 웹 크롤러에 대한 글로벌 크롤 스페이스 구성

구성 파일에서 웹 크롤러에 대한 글로벌 크롤 스페이스를 작성하는 규칙을 작성하여 인덱스에서의 URL 제거를 더 효과적으로 제어할 수 있습니다. 엔터프라이즈 검색 관리 콘솔에서는 이 기능을 지원하지 않습니다.

각 웹 크롤러는 크롤하거나 크롤하지 않을 URL을 정의하는 크롤 스페이스와 함께 구성됩니다. 크롤 스페이스에서 발견된 URL은 나중에 크롤할 수 있도록 데이터베이스에 보유되며, 크롤 스페이스에 없는 URL은 제거됩니다. 크롤러가 빈 데이터베이스로 시작하는 경우, 크롤 스페이스 정의와 데이터베이스는 크롤러가 실행되는 동안 일관성을 유지합니다.

크롤러가 중지되어 크롤 스페이스가 감소하는 경우도 있습니다(페이지를 금지하는 새 규칙을 추가하여). 크롤러가 재시작되면 크롤 스페이스 정의와 데이터베이스가 불일치하게 됩니다. 데이터베이스에는 더 작아진 새 크롤 스페이스에 없는 URL(일부는 크롤되고 일부는 크롤되지 않은)이 포함됩니다. DB2 II OmniFind Edition 버전 8.2에서, 웹 크롤러는 이러한 URL의 HTTP 리턴 코드를 760으로 변경하고(크롤 스페이스 제외) 인덱스에서 현재 제외된 페이지를 제거함으로써 일관성을 리스토어합니다. 이 웹 크롤러가 컬렉션의 유일한 웹 크롤러이고 크롤 스페이스가 인덱스에서 허용되는 모든 웹 문서를 식별하는 것으로 가정하기 때문에 이를 수행합니다.

크롤러가 대형 크롤 스페이스에서 잠시 동안 작동한 후, 관리자가 둘 이상의 크롤러에 있는 크롤 스페이스를 분할하기로 결정하는 경우도 있습니다(일부가 나머지 부분보다 더 자주 크롤되도록 하기 위해). 새 크롤러가 작성되고, 각 크롤러는 독립적인 고유 데이터

1 베이스 테이블(초기에 비어 있음) 및 원래 크롤 스페이스의 다른 부분을 갖게 됩니다.  
1 그런 다음, 원래 크롤러의 크롤 스페이스는 새 크롤러가 크롤할 부분을 제거한 후에 남  
1 은 부분으로 감소합니다.

1 원래 크롤러가 버전 8.2 크롤러인 경우, 크롤러는 위에 설명된 방식으로 일관성을 데이  
1 터베이스로 리스토어합니다. 이는 관리자가 원하는 작동이 아닙니다. 이동된 URL이 인  
1 텍스에서 제거되지 않으며, 다른 크롤러가 이 URL을 크롤합니다.

1 DB2 II OmniFind Edition 버전 8.2.1에서는 현재 크롤러가 크롤하지 않을 URL을 나  
1 타내지만 삭제되지 않는 상위 레벨의 보조 크롤 스페이스를 제공합니다. (두 스페이스  
1 에 없는 URL은 발견 중에 계속 거부되고 이전처럼 다시 스캔할 때 인덱스에서 제거됩  
1 니다.)

1 두 번째 글로벌 크롤 스페이스는 크롤러의 구성 디렉토리에 있는 global.rules 파일의  
1 내용에 의해 정의됩니다. 파일이 있는 경우, 크롤러 초기화 중에 이 파일을 읽습니다.  
1 global.rules 파일이 없는 경우, 크롤러는 단일 레벨의 크롤 스페이스로 작동합니다(버  
1 전 8.2에서처럼). 글로벌 스페이스가 정의된 경우, 크롤러는 이전처럼 URL을 관리하지  
1 만 URL이 두 스페이스에 없는 경우에만 인덱스에서 URL을 제거하도록 지시합니다.

1 global.rules 파일의 구문은 로컬 crawl.rules 파일의 구문과 동일하지만 도메인 이름 규  
1 칩만 의미를 갖습니다. (이에 따라, IP 주소 또는 URL 접두부 패턴이 아닌 DNS 호스  
1 트 이름을 사용해서만 크롤러 간에 크롤 스페이스를 파티션할 수 있습니다.)

1 global.rules 파일이 있으면 글로벌 크롤 스페이스 기능이 사용 가능하고 이 파일이 없  
1 으면 사용 불가능합니다. 로컬 크롤 스페이스의 URL 접두부 또는 IP 주소 규칙에 의  
1 해 제외된(crawl.rules 파일에 정의된 대로) URL은 글로벌 크롤 스페이스의 영향을 받  
1 지 않으므로 계속 제외된 상태로 유지됩니다. 글로벌 크롤 스페이스는 로컬 도메인 규  
1 칩에 의해 한 크롤러의 크롤 스페이스에서 제외된 URL의 인덱스에서 제거하지 못하도  
1 록 하기 위해서만 사용할 수 있습니다.

1 규칙의 작동 방법은 다음과 같습니다.

- 1 1. 로컬 접두부 규칙 또는 주소 규칙에 의해 크롤러 데이터베이스에서 URL이 제외된  
1 경우, URL에 리턴 코드 760이 지정되고 인덱스에서 제거됩니다. 이 단계가 크롤  
1 스페이스 처리의 마지막 단계입니다. 이 URL은 다시 크롤되지 않습니다.
- 1 2. 1단계 이후: 로컬 도메인 규칙에 의해 크롤러 데이터베이스에서 URL이 제외되고 글  
1 로벌 크롤 스페이스가 없는 경우, URL에 리턴 코드 760이 지정되고 인덱스에서 제  
1 거됩니다. 이 단계가 크롤 스페이스 처리의 마지막 단계입니다. 이 URL은 다시 크  
1 롤되지 않습니다.
- 1 3. 1단계 이후: 로컬 도메인 규칙에 의해 크롤러 데이터베이스에서 URL이 제외되었지  
1 만 글로벌 크롤 스페이스의 규칙에 의해 명시적으로 허용되는 경우, URL에 리턴  
1 코드 761이 지정됩니다. 크롤러가 이 URL을 다시 크롤하지 않지만 인덱스에서 제  
1 거되지는 않습니다(일부 다른 크롤러의 로컬 크롤 스페이스에 있는 것으로 가정됨).

1 4. 1단계 이후: 로컬 도메인 규칙에 의해 크롤러 데이터베이스에서 URL이 제외되었지  
1 만 글로벌 크롤 스페이스의 규칙에 의해 명시적으로 허용되지 않는 경우, URL에  
1 리턴 코드 760이 지정되고 인덱스에서 제거됩니다. 이 단계가 크롤 스페이스 처리  
1 의 마지막 단계입니다.

1 이것은 중요한 사항입니다. 글로벌 크롤 스페이스는 로컬 크롤 스페이스에 의해 이  
1 미 제외된 URL이 삭제되지 않도록 하기 위해서만 사용되므로 후보 URL에 적용되  
1 는 규칙이 없는 경우, 글로벌 크롤 스페이스의 디폴트 결과는 금지하는 것입니다. 이  
1 는 로컬 크롤 스페이스의 디폴트값과 반대됩니다.

1 global.rules 파일은 글로벌 크롤 스페이스를 공유하는 모든 크롤러의 master\_config 디  
1 렉토리에 있어야 하며, 모든 글로벌 파일 사본과 개별 로컬 crawl.rules 파일이 서로 일  
1 관되도록 주의하여 편집해야 합니다.

---

## 1 다중 Lotus Notes 서버 구성

1 추가 크롤러가 서버를 크롤하도록 구성하지 않고도 다중 Lotus Notes® 서버에서 동일  
1 한 크롤 스페이스로 데이터베이스를 추가할 수 있습니다. DB2 II OmniFind Edition  
1 버전 8.2에서는 각 Notes 크롤러가 단일 Lotus Notes 서버를 크롤하도록 제한되었습  
1 니다.

1 Notes 크롤러가 한 Lotus Notes 서버의 데이터베이스를 크롤하도록 구성한 후, 다음  
1 조치를 수행하여 다른 Lotus Notes 서버에서 동일한 크롤 스페이스로 데이터베이스를  
1 추가하십시오.

- 1 1. 크롤 페이지에서 구성할 Notes 크롤러를 선택한 후 크롤 스페이스 아이콘을 누르  
1 십시오.
- 1 2. Notes 크롤 스페이스 페이지에서 데이터베이스 추가를 누르십시오.
- 1 3. 다음 창을 사용하여 크롤할 Lotus Notes 서버에 대한 정보를 지정하십시오: 『페이  
1 지를 크롤할 Notes 서버 지정』

1 크롤러 마법사가 표시하는 후속 페이지에 대한 도움이 필요하면, 도움말을 누르십시오.

1 **팁:** 크롤러가 Lotus Notes 서버에 액세스하기 위해 사용하는 암호를 변경하려면 Notes  
1 크롤 스페이스 페이지에서 서버를 선택한 후, 서버 편집을 누르십시오. (Notes 크  
1 롤러가 단일 Lotus Notes 서버를 크롤하도록 제한되었을 때, 이 조치는 Notes 크  
1 롤 스페이스 페이지 위쪽에 제공되었습니다.)

### 1 페이지를 크롤할 Notes 서버 지정

1 기존 크롤 스페이스에 대해 구성할 Lotus Notes 서버를 지정하려면 크롤할 Notes 서  
1 버 지정 페이지를 사용하십시오. 크롤 스페이스에 서버를 추가하는 경우, 크롤러가 해  
1 당 서버의 데이터베이스에 액세스할 수 있도록 하는 정보를 지정하십시오.

Notes 크롤러가 크롤할 수 있도록 Notes® 환경을 설정하는 방법에 대해서는 *Administering Enterprise Search*를 참조하십시오.

**중요사항:** 다음을 누르면 시스템이 지정한 Notes 서버에 액세스할 수 있는지 확인합니다. 서버에 액세스할 수 없으면 오류 메시지가 표시됩니다. 서버 이름 및 프로토콜 정보를 확인하여 정정하고 다음을 다시 누르십시오.

## 필드 및 제어사항

### 기존 서버

크롤 스페이스에 이미 존재하는 Lotus Notes 서버에서 데이터베이스를 추가하도록 지정합니다.

이 옵션을 선택하는 경우, 추가할 데이터베이스가 속하는 Lotus Notes 서버를 선택하십시오.

### 새 서버

새 Lotus Notes 서버에 대한 정보를 크롤 스페이스에 추가하고 해당 서버에서 크롤할 데이터베이스를 선택하도록 지정합니다.

이 옵션을 선택하는 경우, Notes 크롤러가 서버의 데이터베이스에 액세스할 수 있도록 하는 옵션을 지정하십시오.

### Lotus Notes 서버 이름

이 크롤러가 크롤할 Notes 데이터베이스(.nsf 파일)를 호스트하는 Domino® 서버의 완전한 호스트 이름을 입력하십시오(예: server1.ibm.com).

NRPC 프로토콜을 사용하는 Notes 서버의 디폴트 포트는 1352입니다. DIIOP 프로토콜을 사용하는 Notes 서버의 디폴트 포트는 80입니다. 서버가 디폴트 포트에서 대기(listen)하지 않는 경우, 포트 번호(예: server1.ibm.com:1353 또는 server1.ibm.com:81)를 지정하십시오.

### 프로토콜

서버가 사용하는 통신 프로토콜에 대한 정보를 지정하십시오.

### NRPC(Notes Remote Procedure Call)

Domino 서버가 NRPC(Notes Remote Procedure Call) 프로토콜을 사용하여 서버와 웹 브라우저 간의 정보를 교환하는 경우, 이 옵션을 선택하십시오.

**중요사항:** 이전에 이 컬렉션에 대해 NRPC 프로토콜을 사용하는 Notes 크롤러를 구성한 경우, 이 Notes 크롤러에 대한 정보를 구성하기 전에 해당 크롤러를 중지하십시오. NRPC를 사용하는 엔터프라이즈 검색 발견 프로세스와 NRPC를 사용하는 크롤러 프로세스



1

## 1 VeniceBridge 크롤러의 성능 향상

1  
1  
1  
1

Ventica VeniceBridge는 이제 DB2 Information Integrator Content Edition(DB2 II Content Edition)이라는 IBM 제품입니다. DB2 II OmniFind Edition 제품 인터페이스 및 문서에서 여전히 VeniceBridge를 언급하므로 이러한 릴리스 정보에서도 VeniceBridge 용어를 사용하여 사용자 인터페이스의 개선 기능을 설명합니다.

1  
1  
1  
1  
1

이제, 재크롤링 성능을 향상시키기 위해 개별 VeniceBridge 항목 클래스에 대한 규칙을 구성할 때 데이터 맵을 선택할 수 있습니다. 데이터 맵은 작성 날짜, 개정 날짜, 정렬 키 및 내용 크기와 같은 항목 클래스에 대한 VeniceBridge 크롤러 정보를 제공하는 DB2 II Content Edition 구조입니다. 데이터 맵을 크롤러 구성의 항목 클래스와 연관시키면 크롤러가 해당 항목 클래스의 내용을 신속하게 재크롤하는 데 도움이 됩니다.

1  
1  
1

Documentum 및 FileNet 콘텐츠 서비스의 경우, VeniceBridge 크롤러에 디폴트 데이터 맵이 포함되어 있으므로 데이터 맵을 지정하지 않고도 이 기능을 사용할 수 있습니다.

1

### 중요사항:

1  
1

이러한 성능 개선 기능을 사용하려면 DB2 II Content Edition Fix Pack 1을 실행해야 합니다.

1  
1

데이터 맵을 기존 VeniceBridge 크롤 스페이스의 항목 클래스와 연관시키려면 다음을 수행하십시오.

1  
1  
1  
1  
1  
1  
1  
1  
1

1. VeniceBridge 크롤러가 속하는 콜렉션을 편집하십시오.
2. 크롤 페이지를 누르십시오.
3. VeniceBridge 크롤러 이름을 찾은 후 크롤 스페이스 단추를 누르십시오.
4. VeniceBridge 크롤 스페이스 페이지에서 구성할 항목 클래스를 찾은 후 편집을 누르십시오.
5. VeniceBridge 항목 클래스에 대한 규칙 구성 페이지의 데이터 맵 필드에서 이 항목 클래스에 사용할 데이터 맵의 이름을 선택하십시오.
6. 확인을 누르십시오.

1  
1  
1

VeniceBridge 크롤러를 작성하거나 기존 크롤 스페이스에 항목 클래스를 추가하는 경우, VeniceBridge 항목 클래스에 대한 규칙 구성 페이지에서 기타 옵션을 지정할 때 데이터 맵 이름을 선택하십시오.

1

---

## 1 Windows 파일 시스템의 로그인 정보 지정

1  
1

Windows 서브디렉토리에 대한 크롤링 규칙을 구성할 때, 크롤러가 해당 서브디렉토리의 데이터에 액세스할 수 있도록 하는 사용자 ID 및 암호를 지정할 수 있습니다.



1 기존 Windows 파일 시스템 크롤 스페이스의 서브디렉토리에 대한 로그인 증명서를 구  
1 성하려면 다음을 수행하십시오.

- 1 1. Windows 파일 시스템 크롤러가 속하는 콜렉션을 편집하십시오.
- 1 2. 크롤 페이지를 누르십시오.
- 1 3. Windows 파일 시스템 크롤러 이름을 찾은 후 크롤 스페이스 단추를 누르십시오.
- 1 4. Windows 크롤 스페이스 페이지에서 구성할 서브디렉토리를 찾은 후 편집을 누르  
1 십시오.
- 1 5. Windows 서브디렉토리에 대한 규칙 구성 페이지에서 사용자 ID 및 암호 필드에  
1 이 서브디렉토리의 유효한 사용자 ID 및 암호를 입력하십시오.
- 1 6. 확인을 누르십시오.

1 Windows 파일 시스템 크롤러를 작성하거나 기존 크롤 스페이스에 서브디렉토리를 추  
1 가하는 경우, Windows 서브디렉토리에 대한 규칙 구성 페이지에서 기타 옵션을 지정  
1 할 때 로그인 증명서를 지정하십시오.

---

## 1 외부 이름이 표시 이름으로 변경됨

1 엔터프라이즈 검색 관리 콘솔에서 이 필드에 지정된 값이 표시 목적으로만 사용된다는  
1 점을 더 명확히 전달하기 위해 크롤러 구성 마법사에서 외부 이름이 표시 이름으로 변  
1 경되었습니다.

1 표시 이름을 지정할 때, 오브젝트의 내부 이름을 변경하지 마십시오. 예를 들어,  
1 lastUpdateDateTime 데이터베이스 필드에 더 이해하기 쉬운 표시 이름 Last updated  
1 를 지정할 수 있습니다.

1 의미 있는 표시 이름을 지정하면 사용자가 검색할 필드를 더 쉽게 인식할 수 있습니다.  
1 또한, 사용자가 Author와 같은 한 필드에서 검색하여 각 데이터 소스에서 서로 다른 내  
1 부 필드 이름(예: Author, Creator, Producer 등)을 갖는 문서를 찾을 수 있습니다.

---

## 1 데이터 리스너 구성 및 모니터링

1 엔터프라이즈 검색 관리 콘솔을 사용하여 사용자 정의 데이터 리스너 응용프로그램에  
1 대한 지원을 구성할 수 있습니다.

1 DB2 II OmniFind Edition, 버전 8.2에서, 응용프로그램이 엔터프라이즈 검색 콜렉션  
1 을 갱신하려면 먼저 여러 파일을 수동으로 편집해야 했습니다. 다음 정보는 *Administering  
1 Enterprise Search*의 *Configuring support for external crawlers*에 있는 지시사항을  
1 대체합니다.

1 데이터 리스너 응용프로그램은 콜렉션에 데이터를 추가하고 콜렉션에서 URI를 제거하  
1 며, 웹 크롤러가 URL을 방문하거나 재방문하도록 지시할 수 있습니다. 사용자(클라이

1 언트)가 데이터 리스너 구성요소에 연결할 때, 클라이언트 ID, 암호 및 갱신할 콜렉션  
1 의 ID를 전달해야 합니다. 데이터 리스너 구성요소는 클라이언트 ID 및 암호를 인증하  
1 고, 해당되는 경우 응용프로그램이 지정된 콜렉션을 갱신할 수 있는 권한을 부여합니다.

1 데이터 리스너 응용프로그램에 대한 지원을 구성하려면 관리 콘솔 도구 모음에서 시스  
1 템 옵션을 선택하십시오. 다음 창을 사용하여 데이터 리스너 응용프로그램에 대한 정보  
1 를 지정하고 데이터 리스너 활동을 모니터하십시오.

- 1 • 『시스템 보기: 데이터 리스너 페이지』
- 1 • 19 페이지의 『시스템 보기 모니터: 데이터 리스너 페이지』
- 1 • 19 페이지의 『데이터 리스너 세부사항 페이지』

## 1 시스템 보기: 데이터 리스너 페이지

1 사용자 정의 데이터 리스너 응용프로그램이 콜렉션을 갱신할 수 있도록 하는 정보를 지  
1 정하려면 시스템 보기의 데이터 리스너 페이지를 사용하십시오.

1 데이터 리스너 응용프로그램을 사용하면 DB2 Information Integrator OmniFind Edition  
1 에 제공된 디폴트 크롤러로는 크롤할 수 없는 데이터 소스 유형을 크롤할 수 있습니다.  
1 응용프로그램은 콜렉션에 데이터를 추가하고 콜렉션에서 데이터를 제거하며, 웹 크롤러  
1 가 URL을 방문하거나 재방문하도록 지시할 수 있습니다.

### 1 필드 및 제어사항

#### 1 스레드 수

1 데이터 리스너 구성요소가 데이터 리스너 응용프로그램의 요청을 처리하기 위  
1 해 작성할 수 있는 최대 작업 스레드 수를 입력하십시오. 디폴트값은 6입니다.

#### 1 데이터 리스너 포트

1 데이터 리스너 구성요소의 포트 번호를 입력하십시오. 디폴트값은 6668입니다.

1 **중요사항:** 포트 번호를 변경하는 경우, 변경사항을 적용하려면 데이터 리스너 구  
1 성요소를 재시작해야 합니다.

#### 1 데이터 리스너 클라이언트 ID 추가

1 새 데이터 리스너 응용프로그램에 대한 정보를 지정할 수 있도록 데이터 리스  
1 너 클라이언트 ID 목록에 공백 행을 추가합니다.

#### 1 데이터 리스너 클라이언트 ID

1 엔터프라이즈 검색 콜렉션에 대한 액세스 권한을 부여할 데이터 리스너 응용프  
1 로그램의 클라이언트 ID를 입력하십시오. 각 ID는 엔터프라이즈 검색 시스템  
1 내에서 고유해야 합니다.

1 **암호** 데이터 리스너 응용프로그램의 암호를 입력하십시오. 엔터프라이즈 검색 콜렉션  
1 에 액세스하려면 데이터 리스너 응용프로그램이 지정된 증명서(ID 및 암호)를  
1 사용해야 합니다.

## 컬렉션 이름

데이터 리스너 응용프로그램이 갱신할 컬렉션을 선택하십시오. 각 데이터 리스너 응용프로그램은 한 컬렉션에 있는 콘텐츠만 갱신할 수 있습니다.



## 제거

엔터프라이즈 검색 시스템에서 선택한 데이터 리스너 응용프로그램을 제거합니다. 이 응용프로그램은 엔터프라이즈 검색 컬렉션에 데이터를 추가하거나 제거하는 데 사용되지 않습니다.

## 시스템 보기 모니터: 데이터 리스너 페이지

데이터 리스너 구성요소 및 데이터 리스너 응용프로그램 활동에 대한 세부사항을 보기 위해 시스템을 모니터링하는 경우, 데이터 리스너 페이지를 사용하십시오.

### 필드 및 제어사항

#### 새로 고침

이 페이지의 정보를 새로 고칩니다. 상태 메시지는 이 정보가 최종 갱신된 시기를 표시합니다.

**이름** 구성요소 이름을 표시합니다(데이터 리스너).



#### 세부사항

데이터 리스너 활동에 대한 자세한 정보를 표시합니다. (데이터 리스너가 중지되면 세부사항을 볼 수 없습니다.)

**상태** 다음 아이콘은 데이터 리스너의 현재 상태를 표시합니다.



#### 실행 중

데이터 리스너 구성요소가 활성 상태임을 표시합니다. 엔터프라이즈 검색 시스템이 시작되면 데이터 리스너 구성요소가 자동으로 시작됩니다. 일반적으로, 데이터 리스너에 대해 구성된 포트 번호를 변경하지 않으면 이를 중지하거나 재시작할 필요가 없습니다.



#### 중지됨


데이터 리스너 구성요소가 중지되었음을 표시합니다.


## 데이터 리스너 세부사항 페이지

스레드 상태 및 데이터 리스너 응용프로그램 요청에 대한 정보와 같은 현재 데이터 리스너 활동에 대한 세부사항을 모니터링하려면 데이터 리스너 세부사항 페이지를 사용하십시오.

1 필드 및 제어사항  
1 새로 고침  
1 이 페이지의 정보를 새로 고칩니다. 상태 메시지는 이 정보가 최종 갱신된 시  
1 기를 표시합니다.

1 상태 다음 아이콘은 데이터 리스너의 현재 상태를 표시합니다.

1  실행 중  
1 데이터 리스너 구성요소가 활성 상태임을 표시합니다. 엔터프라이즈 검색  
1 색 시스템이 시작되면 데이터 리스너 구성요소가 자동으로 시작됩니다.  
1 일반적으로, 데이터 리스너를 중지하거나 재시작할 필요가 없습니다.

1  중지됨  
1 데이터 리스너 구성요소가 중지되었음을 표시합니다.

1 처리할 요청 수  
1 처리를 위해 큐에 대기 중인 데이터 리스너 응용프로그램의 요청 수를 표시합  
1 니다.

1 총 스레드 수  
1 데이터 리스너 구성요소가 데이터 리스너 응용프로그램의 요청을 처리하기 위  
1 해 작성할 수 있는 총 스레드 수를 표시합니다.

1 스레드 상태  
1 가능한 스레드 상태를 보여줍니다.  
1 • 스레드가 초기화됨  
1 • 스레드가 클라이언트 요청을 읽는 중  
1 • 스레드가 클라이언트 요청을 처리 중  
1 • 스레드가 클라이언트로 응답을 보내는 중  
1 • 스레드가 클라이언트로 응답 전송을 완료함  
1 • 스레드가 다음 클라이언트 요청을 대기 중  
1 • 스레드가 일시정지 상태임  
1 • 스레드를 중지 중

1 스레드 수  
1 주어진 스레드 상태에 대해 사용 중인 스레드 수를 표시합니다(예: 5개의 스레  
1 드가 현재 데이터를 페치 중).

## 1 데이터 리스너 구성요소 중지 및 시작

1 데이터 리스너 구성요소에 대해 구성된 포트 번호를 변경하는 경우, 구성요소를 중지한  
1 후 재시작해야 합니다.

## 프로시저

데이터 리스너 구성요소를 중지 및 재시작하려면 다음을 수행하십시오.

1. 인덱스 서버에서 엔터프라이즈 검색 관리자로 로그인하십시오. 이 사용자 ID는 DB2 II OmniFind Edition 설치 중에 지정됩니다.
2. ES\_NODE\_ROOT/master\_config/datalistener 디렉토리로 변경하십시오.
3. 다음 명령을 입력하여 데이터 리스너를 중지하고 재시작하십시오.

```
esadmin datalistener stop  
esadmin datalistener start
```

---

## 컬렉션의 문서 수 계산

컬렉션을 작성할 때, 컬렉션이 보유할 것으로 예상되는 문서 수에 대한 추정치를 제공할 수 있습니다. DB2 II OmniFind Edition의 자원 관리자는 이 수를 사용하여 컬렉션에 필요한 메모리 및 디스크 자원을 계산합니다.

메모리 및 디스크 자원 요구사항 추정치가 현재 시스템에 사용 가능한 크기를 초과하면 자원 관리자가 경고를 발행합니다. 경고에 따라 이후의 자원 부족 문제점을 방지할 수 있습니다.

모니터는 계산된 문서 수를 경고 수신에 대해 구성하는 임계값 백분율과 함께 사용하여 컬렉션이 접근하는 최대 문서 수 관련 통지를 보내는 시기를 결정합니다.

컬렉션의 잠재적 크기 추정치를 제공하려면 컬렉션을 작성할 때 추정 문서 수 필드에 숫자를 입력하십시오. 디폴트값은 1백만 개의 문서입니다(1000 000).

컬렉션을 작성한 후, 컬렉션을 편집하고 일반 페이지에서 새 값을 지정하여 이 추정치를 변경할 수 있습니다.

컬렉션이 추정 크기로 커지면, 시스템이 인덱스에 문서 추가를 중지하지 않습니다. 그러나, 컬렉션에 대해 경보를 구성하고 인덱스의 문서 수가 한계를 초과할 때 옵션을 통해 이를 알리도록 지정하는 경우, 컬렉션 크기가 지정한 통지 백분율에 도달하면 사용자가 통지를 수신하게 됩니다.

**주:** 버전 8.2에서는 이 경보를 구성할 때, 인덱스에 포함될 수 있는 문서 수를 입력하여 인덱스 문서 수의 한계를 지정했습니다. 버전 8.2.1에서는 추정 문서 수 필드에 지정하는 수가 한계 필드에 자동으로 표시되며, 인덱스의 문서 수가 이 한계 백분율에 도달할 때 통지를 수신하도록 지정함으로써 통지 임계값을 구성합니다.



## 구문 분석기 상태 요약

이 영역의 옵션은 구문 분석기의 상태를 표시하며, 구문 분석되었지만 아직 인덱스화되지 않은 문서에 대한 통계 정보를 제공합니다.

**상태** 상태 행의 첫 번째 아이콘은 크롤러의 현재 상태를 표시합니다.



### 실행 중

구문 분석기가 실행 중임을 표시합니다.

구문 분석기의 실제 상태에 따라 현재 구문 분석 활동에 대한 더 자세한 정보를 제공하는 메시지가 표시될 수 있습니다. 예를 들어, 구문 분석기가 현재 유틸 문서(구문 분석할 문서가 더 많아질 때까지 정지, 오류 조건으로 인해 재시작 대기 또는 구문 분석기 서비스의 재시작 대기) 또는 일시정지된 문서(예: 인덱스의 재구성 완료 대기)를 구문 분석 중일 수 있습니다.



### 중지됨

구문 분석기가 중지되었음을 표시합니다.

구문 분석기의 현재 상태 및 관리 콘솔에 따라 다음 아이콘을 사용하여 구문 분석기 활동을 관리할 수 있습니다.



### 시작

구문 분석기를 시작합니다.

구문 분석기가 시작된 후, 구문 분석기는 대개 더 이상 구문 분석할 문서가 없을 때까지 크롤된 문서를 구문 분석합니다. 그런 다음, 구문 분석기는 정지 모드가 되어 다음에 새로 크롤된 구문 분석할 문서가 있는지 점검할 때까지 실행되지 않습니다.



### 중지

구문 분석기를 중지합니다.

예를 들어, 범주 유형을 변경하거나 범주 규칙을 추가하는 경우, 구문 분석기를 중지한 후 재시작해야 합니다.

### 데이터 저장소에서 인덱스화해야 할 문서 수

구문 분석기가 구문 분석하여 데이터 저장소에 기록한 문서 수를 표시합니다. 이러한 문서는 아직 엔터프라이즈 검색 인덱스에 추가되지 않았습니다.

컬렉션의 각 크롤러에 대해 다음 통계가 사용 가능합니다.

### 스냅샷 통계

이 영역의 옵션은 구문 분석기 활동의 현재 상태를 표시하며, 개별 크롤러에 의

1 해 수집된 문서에 대한 통계 정보를 제공합니다. 최상의 검색 결과를 얻으려면  
1 구문 분석기가 컬렉션의 여러 크롤러에 의해 크롤된 문서를 모두 구문 분석할  
1 때까지 인덱스를 빌드하지 마십시오.

### 1 스냅샷 시작

1 구문 분석기 상태의 현재 보기에 있는 통계가 수집된 날짜 및 시간을  
1 표시합니다.

### 1 구문 분석 비율

1 구문 분석기가 초당 구문 분석할 페이지 수를 표시합니다.

1 컬렉션의 각 크롤러에 대해 다음 통계가 사용 가능합니다.

### 1 크롤러 이름

1 컬렉션에 있는 각 크롤러의 이름을 표시합니다.

### 1 크롤러 유형

1 각 크롤러에 대해 크롤러가 크롤하는 데이터 유형을 표시합니다.

### 1 구문 분석된 문서

1 각 크롤러에 대해 구문 분석할 수 있는 문서 수 및 지금까지 구문 분  
1 석된 문서 수를 표시합니다.

1 상태 각 크롤러에 대해 문서 구문 분석 활동의 상태를 표시합니다.

### 1 실행 중

1 구문 분석기가 현재 이 크롤러에 의해 크롤된 문서를 구문 분  
1 석 중임을 나타냅니다.

### 1 중지됨

1 구문 분석기가 이 크롤러에 대한 문서를 구문 분석하지 않음  
1 을 나타냅니다(구문 분석할 문서가 없기 때문일 수 있음).

### 1 일시정지됨

1 이 크롤러가 크롤한 문서의 구문 분석이 일시정지되었음을 나  
1 타냅니다.

### 1 오류

1 이 크롤러가 크롤한 문서를 구문 분석하는 중에 오류가 발생  
1 했음을 나타냅니다. 오류를 검토하려면 컬렉션을 모니터링하는  
1 중에 로그 페이지를 열어 구문 분석기의 로그 파일을 보십시  
1 오.





완료

구문 분석기가 이 크롤러에 의해 크롤된 문서의 구문 분석을 완료했음을 나타냅니다.

## HTML 메타데이터를 검색 필드에 맵핑

구문 분석기가 HTML 메타데이터 요소를 필드 이름에 맵핑할 수 있습니다. 컬렉션에 대한 구문 분석기를 구성할 때, 검색 필드로 인덱스화할 HTML 메타데이터 요소를 지정할 수 있습니다. 그런 다음, 사용자가 쿼리에서 이 필드 이름을 지정하여 HTML 문서의 특정 부분을 검색할 수 있습니다.

HTML 메타데이터를 필드에 맵핑하면, 사용자가 검색 필드에 액세스할 수 있도록 할 것인지 여부를 지정하고 매개변수식 쿼리로 필드를 검색하거나 검색 결과에서 필드를 볼 수 있습니다.

HTML 메타데이터 요소를 검색 필드에 맵핑하려면 다음 조치를 수행하십시오.

1. 관리 콘솔에서 변경할 컬렉션의 편집 아이콘을 누르십시오.
2. 구문 분석 페이지를 선택하십시오.
3. **HTML 메타데이터를 필드에 맵핑**을 누르십시오.
4. 다음 창을 사용하여 검색할 메타데이터 필드에 대한 정보를 지정하십시오.

- 『HTML 메타데이터 필드 맵핑 페이지』

## HTML 메타데이터 필드 맵핑 페이지

HTML 메타데이터 요소를 필드 이름에 맵핑하려면 HTML 메타데이터 필드 맵핑 페이지를 사용하십시오. 쿼리에서 맵핑된 필드 이름을 지정함으로써 HTML 문서의 특정 부분을 검색하여 더 정확한 검색 결과를 얻을 수 있습니다.

HTML 필드 맵핑을 작성하는 경우, 변경사항은 다음에 구문 분석기를 중지하고 재시작할 때 적용됩니다. 새 맵핑은 이미 구문 분석되고 인덱스화된 데이터에 영향을 미치지 않습니다.

### 필드 및 제어사항

#### 필드 추가

다른 HTML 메타데이터 요소를 검색 필드 이름에 맵핑할 수 있는 공백 라인을 추가합니다.

검색 필드 이름에 맵핑하려는 각 HTML 메타데이터 요소에 대해 다음 정보를 제공하십시오.

#### 필드 이름

HTML 메타데이터 요소와 연관시킬 이름을 입력하십시오. 사용자가 이 컬렉션

에서 HTML 문서를 쿼리할 때 이 필드 이름을 지정할 수 있습니다. 예를 들어, publisher라는 HTML 메타데이터 태그를 맵핑하기 위해 검색 필드 이름으로 publisher를 입력할 수 있습니다.

이 값은 영숫자 문자만 포함하는 단일 단어이어야 합니다.

### HTML 메타데이터 필드

필드 이름 필드에 지정한 필드 이름으로 맵핑되는 HTML 메타데이터 요소의 이름을 입력하십시오. 동일한 예를 사용하려는 경우, HTML 메타데이터 요소 publisher를 검색 필드 publisher에 맵핑하므로 이 필드에 publisher를 입력하십시오.

이 값은 영숫자 문자만 포함하는 단일 단어이어야 합니다.

### 필드 이름별 검색

필드 이름 필드에 지정한 필드 이름을 지정하여 HTML 문서를 검색할 수 있도록 하려면 이 선택란을 선택하십시오. 이 선택란이 지워져 있으면, 사용자가 이 컬렉션에서 HTML 문서를 검색할 때 필드 이름을 지정할 수 없습니다.

### 매개변수식 검색

사용자가 이 필드를 검색할 때 매개변수식 쿼리를 공식화할 수 있도록 하려면 이 선택란을 선택하십시오. 매개변수식 쿼리를 사용하면 숫자 및 날짜 값을 평가할 수 있습니다. 예를 들어, 사용자가 이 필드의 값이 다른 값보다 작거나 큰 지 판별하려고 할 수 있습니다.

이 선택란을 선택하는 경우, HTML 메타데이터 요소의 데이터 유형이 DECIMAL, DOUBLE, INTEGER, SHORT, TIME 또는 TIMESTAMP인지 확인하십시오.

### 검색 결과에 표시

이 필드를 검색 결과에 표시하려면 이 선택란을 선택하십시오. 이 선택란이 지워져 있으면, 이 필드가 검색 결과에 포함되지 않습니다.



### 삭제

선택한 HTML 메타데이터 필드 맵핑을 삭제합니다.

---

## 일본어 및 중국어 문서의 공백에서 줄 바꾸기 문자 제거

일본어 및 중국어와 같이 공백이 단어 바운더리를 분리하는 데 사용되지 않는 경우, 구문 분석기가 줄 바꾸기를 유발하는 특정 공백 문자를 제거하도록 구성할 수 있습니다.

컬렉션에 대해 이 기능을 사용 가능하게 하려면 다음을 수행하십시오.

1. 엔터프라이즈 검색 관리자로 로그인하십시오(이 사용자 ID 및 암호는 DB2 II OmniFind Edition 설치 중에 설정되었습니다).

1 2. 텍스트 편집기를 사용하여 `collection_ID.parserdriver/collection.properties`  
1 파일을 편집하십시오. 여기서, `collection_ID`는 컬렉션을 작성할 때 시스템이 콜  
1 렉션에 지정한 ID입니다.

1 3. `removeCjNewlineChars` 등록 정보의 값을 `false`에서 `true`로 변경하십시오.

1 이 기능이 사용 가능하고 지정된 컬렉션의 문서가 일본어 또는 중국어인 경우, 구문 분  
1 석기가 ASCII가 아닌 문자 범위에서 두 문자를 분리하는 일련의 줄 바꾸기 또는 탭  
1 문자를 제거합니다. 다음 문자가 제거됩니다: Tab(0x09 ), LF 또는 줄 바꾸기(0x0A)  
1 및 CR 또는 캐리지 리턴 0x0D.

---

## 1 컬렉션 구문 분석기 및 Stellent 세션에 대한 문서 유형 구성

1 DB2 II OmniFind Edition 버전 8.2에서, 파일 유형 및 콘텐츠 유형은 특정 컬렉션  
1 구문 분석기 및 Stellent 문서 필터링 세션과 사전 연관되었습니다. 이제, 구성 파일을  
1 작성하여 특정 유형 문서의 구문 분석 방법을 지정할 수 있습니다.

1 버전 8.2에서는 특정 유형의 문서(예: PostScript 문서)가 구문 분석되지 않으며, 일부  
1 는 처리를 위해 Stellent 세션으로 전송되도록 하드 코딩되었습니다(Microsoft Word,  
1 Microsoft Excel, Microsoft PowerPoint, Lotus Freelance, Lotus 123, PDF, RT 및  
1 Ichitaro). 기타 모든 문서 형식은 내장 HTML 또는 XML 구문 분석기에 의해 처리됩  
1 니다. 메타데이터는 올바르지 않을 수 있으므로 일반 텍스트 및 HTML 문서가 오류  
1 상태의 Stellent 세션으로 전송된 후, 다시 내장 구문 분석기 중 하나로 전송될 수 있습  
1 니다(성능에 영향을 미칠 수 있는 상황).

1 크롤 스페이스의 문서가 정확하고 효율적으로 구문 분석되도록 하기 위해 구성 파일을  
1 작성하여 컬렉션 구문 분석기가 구문 분석할 문서 유형 및 Stellent 세션에서 구문 분석  
1 할 문서 유형을 지정할 수 있습니다.

1 문서 유형을 컬렉션 구문 분석기 및 Stellent 세션과 연관시키기 위한 태스크는 다음과  
1 같습니다.

- 1 1. 컬렉션 구문 분석기에 대한 문서 유형을 구성하십시오. 이 단계에는 문서 유형을 콜  
1 렉션에서 사용하는 구문 분석기로 맵핑하는 구성 파일의 작성이 포함됩니다. 각 콜  
1 렉션에 대해 이러한 구성 파일 중 하나를 작성할 수 있습니다.
- 1 2. Stellent 세션에 대해 문서 유형을 구성하십시오. 이 단계에는 문서 유형을 컬렉션에  
1 서 사용하는 Stellent 문서 필터로 맵핑하는 구성 파일의 작성이 포함됩니다. 각 콜  
1 렉션에 대해 이러한 구성 파일 중 하나를 작성할 수 있습니다.
- 1 3. 변경사항을 적용하려면 엔터프라이즈 검색 관리 콘솔을 사용하여 문서 유형을 구성  
1 한 컬렉션을 모니터링하고 구문 분석기를 중지한 후 재시작하십시오.

## 컬렉션 구문 분석기에 대한 문서 유형 구성

문서의 특정 유형을 컬렉션 구문 분석기와 연관시키기 위해 구성 파일을 작성할 수 있습니다. 엔터프라이즈 검색 관리 콘솔에서는 이 기능을 지원하지 않습니다.

구성 파일은 다음 정보를 지정합니다.

- 파일 확장자 또는 콘텐츠 유형에 따라 Stellent 세션으로 전송할 문서
- 미완료 메타데이터로 인해 유형을 알 수 없는 문서의 구문 분석 방법

구성 파일이 없는 경우, 컬렉션 구문 분석기는 DB2 II OmniFind Edition, 버전 8.2의 동작에 해당하는 디폴트 구성 설정값을 사용합니다. 이러한 디폴트 설정값에 대한 설명은 29 페이지의 『디폴트 구문 분석기 서비스 규칙』을 참조하십시오.

문서 유형을 컬렉션 구문 분석기와 연관시키려면 다음을 수행하십시오.

1. 인덱스 서버에서 엔터프라이즈 검색 관리자로 로그인하십시오. 이 사용자 ID는 DB2 II OmniFind Edition 설치 중에 지정됩니다.
2. 다음과 같이 구성 파일을 작성하십시오. 여기서, *collection\_ID*는 구성할 컬렉션을 나타냅니다.

```
ES_NODE_ROOT/master_config/collection_ID.parserdriver/parserTypes.cfg
```

이 파일은 Stellent 세션으로 보낼 파일 확장자 및 내용 유형을 나열합니다. 파일 형식은 일련의 행으로 되어 있으며, 각 행은 다음 규칙 중 하나입니다.

### **EXTENSION** *extension parser*

지정한 구문 분석기가 URL이 지정된 확장자로 끝나는 모든 문서를 처리합니다. 확장자에 마침표를 포함시키지 마십시오. 비교는 대소문자가 구분되지 않습니다.

### **CONTENTTYPE** *type/subtype parser*

지정한 구문 분석기가 지정된 콘텐츠 유형이 지정된 유형/부속 유형과 일치하는 모든 문서를 처리합니다. 문서의 콘텐츠 유형이 t/s인 경우, t가 유형과 같고 s가 부속 유형과 같거나 부속 유형이 와일드 카드(별표 \*)이면 일치가 발생합니다.

### **UNKNOWN** *parser*

지정한 구문 분석기가 확장자 및 콘텐츠 유형을 알 수 없는(즉, 크롤러에 의해 사용 가능해지지 않음) 모든 문서를 처리합니다.

### **DEFAULT** *parser*

기타 규칙에 의해 처리되지 않은 모든 문서가 지정된 구문 분석기에 의해 처리됩니다.

모든 경우, *parser*는 html, xml, stellent 또는 none이어야 하며, 여기서 none은 문서가 구문 분석되지 않음을 의미합니다.

1           **규칙 우선순위:** 둘 이상의 규칙이 문서와 일치하는 경우, 규칙의 표시 순서에 상관  
1 없이 더 구체적인 규칙이 적용됩니다. 규칙의 구체성은 위 목록에 표시된 순서대로  
1 줄어듭니다.

- 1           • EXTENSION 규칙은 CONTENTTYPE 규칙보다 더 구체적입니다.
- 1           • 부속 유형이 있는 CONTENTTYPE 규칙은 와일드 카드가 있는 규칙보다 더 구  
1 체적입니다. 예를 들어, 콘텐츠 유형 application/postscript의 규칙은  
1 application/\*의 규칙보다 우선합니다.
- 1           • 동일한 확장자 또는 콘텐츠 유형에 대해 두 개의 규칙이 있으면 안됩니다. 이 경  
1 우, 규칙의 구현 방법에 따라 우선순위가 제공됩니다.

## 1           예

1           이 예에서, 내장 HTML 구문 분석기는 확장자가 txt, htm 또는 html이고 문서 유형  
1 이 text/로 시작되거나, 알 수 없는 확장자 및 콘텐츠 유형을 갖는 모든 문서를 처리  
1 합니다. 내장 XML 구문 분석기는 확장자가 xml이거나 콘텐츠 유형이 text/xml인 모  
1 든 문서를 처리합니다. 콘텐츠 유형이 application/으로 시작되는 문서를 포함한 기  
1 타 모든 문서는 Stellent 세션으로 전송됩니다.

```
1           EXTENSION doc stellent  
1           EXTENSION txt html  
1           EXTENSION htm html  
1           EXTENSION html html  
1           EXTENSION xml xml  
1           EXTENSION ps none  
1           CONTENTTYPE text/xml xml  
1           CONTENTTYPE text/* html  
1           CONTENTTYPE application/* stellent  
1           UNKNOWN html  
1           DEFAULT stellent
```

## 1           디폴트 구문 분석기 서비스 규칙

1           파일 유형 및 콘텐츠 유형을 컬렉션 구문 분석기에 맵핑하는 구성 파일을 작성하지 않  
1 으면, 디폴트값을 사용하여 문서를 구문 분석합니다.

1           컬렉션 구문 분석기가 사용하는 디폴트값은 다음과 같습니다.

```
1           EXTENSION pdf stellent  
1           EXTENSION ppt stellent  
1           EXTENSION prz stellent  
1           EXTENSION lwp stellent  
1           EXTENSION doc stellent  
1           EXTENSION rtf stellent  
1           EXTENSION xls stellent  
1           EXTENSION 123 stellent  
1           EXTENSION vsd stellent  
1           EXTENSION vdx stellent  
1           EXTENSION jxw stellent  
1           EXTENSION jsw stellent  
1           EXTENSION jtw stellent  
1           EXTENSION jaw stellent
```

```

1      EXTENSION juw stellent
1      EXTENSION jbw stellent
1      EXTENSION jvw stellent
1      EXTENSION jfw stellent
1      EXTENSION jtt stellent
1      EXTENSION jtd stellent
1      EXTENSION jttc stellent
1      EXTENSION jtdc stellent
1      EXTENSION jtdx stellent
1      EXTENSION ps none
1      EXTENSION xml xml
1      EXTENSION txt text
1      EXTENSION htm html
1      EXTENSION html html
1      EXTENSION shtml html
1      EXTENSION xhtml html
1      EXTENSION asp html
1
1      CONTENTTYPE application/postscript none
1      CONTENTTYPE application/* stellent
1      CONTENTTYPE text/rtf stellent
1      CONTENTTYPE text/richtext stellent
1      CONTENTTYPE text/xml xml
1      CONTENTTYPE text/html html
1      CONTENTTYPE text/plain text
1
1      UNKNOWN html
1      DEFAULT html

```

## Stellent 세션에 대한 문서 유형 구성

Stellent 문서 필터가 구문 분석할 문서 유형을 지정하려면 구성 파일을 작성하십시오. 엔터프라이즈 검색 관리 콘솔에서는 이 기능을 지원하지 않습니다.

구성 파일은 다음 정보를 지정합니다.

- Stellant 세션이 구문 분석할 파일 유형. 파일 유형은 Stellant 라이브러리에서 인식되는 파일 유형 중 하나에 해당합니다.
- 내장 구문 분석기 중 하나로 처리하기 위해 컬렉션 구문 분석기로 다시 전송할 파일 유형. (이 조치는 잘못된 메타데이터로 인해 컬렉션 구문 분석기가 문서를 오류가 있는 Stellant 세션으로 보내기 때문에 필요합니다.)
- DB2 II OmniFind Edition에서 지원되지 않기 때문에 거부할 파일 유형.

구성 파일이 지정되었지만 존재하지 않으면 구문 분석기가 시작되지 않습니다. stellent.properties 파일의 OutsideInSupportedTypes 등록 정보에 대한 구성 파일을 지정하지 않으면, 디폴트 구성 설정값이 사용됩니다. DB2 II OmniFind Edition 버전 8.2 동작에 해당하는 이 디폴트 설정값에 대한 설명은 32 페이지의 『Stellent 세션의 디폴트 구문 분석 규칙』을 참조하십시오.

Stellent 세션에 대해 문서 유형을 구성하려면 다음을 수행하십시오.

- 1 1. 인덱스 서버에서 엔터프라이즈 검색 관리자로 로그인하십시오. 이 사용자 ID는 DB2  
1 II OmniFind Edition 설치 중에 지정됩니다.
- 1 2. ES\_NODE\_ROOT/master\_config/collection\_ID.stellent/stellent.  
1 properties 파일을 편집하십시오. 여기서, *collection\_ID*는 구성할 컬렉션을 나  
1 타냅니다.
- 1 3. OutsideInSupportedTypes 등록 정보로 작성할 구성 파일의 절대 경로를 지정하  
1 십시오.

1 예를 들어, 단일 컬렉션에 대해 다음 구성 파일을 작성하여 다른 컬렉션 특정 파일  
1 로 이를 저장할 수 있습니다.

1 ES\_NODE\_ROOT/master\_config/collection\_ID.stellent/stellenttypes.cfg

1 모든 컬렉션에 동일한 설정값을 사용하기 위해 다음 구성 파일을 작성하여 다른 시  
1 스템 레벨 파일로 이를 저장할 수 있습니다. (이 방법을 사용하는 경우, 2단계에 지  
1 정된 대로 *stellent.properties* 파일에 이 경로를 지정해야 합니다.)

1 ES\_INSTALL\_ROOT/default\_config/stellent/stellenttypes.cfg

- 1 4. 구성 파일을 작성하십시오. 이 파일은 문서 유형 및 문서 처리 방법을 나열합니다.  
1 파일 형식은 일련의 행으로 되어 있으며, 각 행은 다음 규칙 중 하나입니다.

```
1 accept DEFAULT
1 accept ALL doctype
1 accept type doctype
1 native DEFAULT
1 native type doctype
1 reject type
```

1 여기서:

1 *doctype*

1 doctype 쿼리 토큰에 사용할 값입니다. doctype 토큰은 쿼리에서 지정하여  
1 (예: doctype:pdf) 검색할 수 있습니다.

1 *type* Stellant 라이브러리에 있는 FI\_ 값 중 하나이며, doctype은 규칙이 적용  
1 되는 경우 doctype 토큰에 사용되는 값입니다.

#### 1 **DEFAULT**

1 규칙 유형에 따라 승인된 유형 또는 원시(native) 유형 목록에 디폴트 목록  
1 이 포함됨을 의미합니다. 이 옵션을 사용하면 디폴트 구성을 대체하지 않고  
1 확장할 수 있습니다.

1 **A11** 명시적으로 나열되지 않은 모든 유형이 지정된 doctype 토큰에서 승인됨을  
1 의미합니다.

1 **규칙 처리:**

- 1 • accept DEFAULT가 지정된 경우의 디폴트 목록을 포함하여 *type*에 대해 accept  
1 규칙이 있는 경우, 유형이 승인됩니다.

- 1 • 그렇지 않은 경우, *type*에 대해 reject 규칙이 있으면 유형이 승인되지 않습니다.
- 1 • 그렇지 않은 경우, accept ALL을 지정하면 유형이 승인됩니다.
- 1 • 그렇지 않으면, 유형이 승인되지 않습니다.

1 문서 유형이 승인된 경우, 적용된 규칙에 지정된 *doctype* 값이 사용됩니다. 이 값  
1 은 구문 분석된 내용과 함께 컬렉션 구문 분석기로 다시 전송됩니다. 문서 유형이  
1 승인되지 않은 경우에는 다음 동작이 발생합니다.

- 1 • native DEFAULT가 지정된 경우의 디폴트 목록을 포함하여 *type*에 대해 원시  
1 (native) 규칙이 있는 경우, 문서는 다시 내장 구문 분석기로 전송되어 이 규칙  
1 에 따라 지정된 *doctype* 토큰의 값을 나타냅니다. *doctype*의 값은 txt, htm  
1 또는 xml이어야 하며, 이는 각각 일반 텍스트, HTML 또는 XML을 나타냅니  
1 다.
- 1 • 그렇지 않으면, 문서가 거부되어 구문 분석되지 않습니다.

## 1 예

1 다음 구성 파일에서 Stellent 세션은 디폴트 목록 이외에 Microsoft Visio 형식을 승인  
1 합니다.

```
1 accept DEFAULT
1 accept FI_VISI03 visio
1 accept FI_VISI04 visio
1 accept FI_VISI05 visio
1 accept FI_VISI06 visio
```

1 다음 구성 파일의 경우, Postscript 문서가 지원되며 ps 문서 유형으로 검색할 수 있습  
1 니다. X pixmap 형식(XPM)은 다시 내장 텍스트 구문 분석기로 전송되고 PNG 이미  
1 지 형식은 거부되며, 기타 모든 파일 유형이 승인되고 other 문서 유형으로 검색할 수  
1 있습니다.

```
1 accept DEFAULT
1 accept FI_POSTSCRIPT ps
1 native FI_XPIXMAP txt
1 accept ALL other
1 reject FI_PNG
```

## 1 Stellent 세션의 디폴트 구문 분석 규칙

1 파일 유형을 Stellent 세션 문서 필터로 맵핑하는 구성 파일을 작성하지 않으면, 디폴트  
1 규칙을 사용하여 문서를 구문 분석합니다.

1 Stellent 세션에서 사용하는 디폴트 규칙은 다음과 같습니다.

```
1 ACCEPT FI_WORD4 doc
1 ACCEPT FI_WORD5 doc
1 ACCEPT FI_RTF rtf
1 ACCEPT FI_WINWORD1 doc
1 ACCEPT FI_WINWORD1COMPLEX doc
```



1	ACCEPT FI_WINWORD2	doc
1	ACCEPT FI_WORD6	doc
1	ACCEPT FI_WINWORD6	doc
1	ACCEPT FI_ICHITAR03	jxw
1	ACCEPT FI_ICHITAR04	jsw
1	ACCEPT FI_WINWORD1J	doc
1	ACCEPT FI_WINWORD5J	doc
1	ACCEPT FI_RTFJ	rtf
1	ACCEPT FI_WINWORD7	doc
1	ACCEPT FI_WORDPRO	lwp
1	ACCEPT FI_WINWORD97	doc
1	ACCEPT FI_ICHITAR08	jtd
1	ACCEPT FI_WORDPRO97	lwp
1	ACCEPT FI_WINWORD2000	doc
1	ACCEPT FI_WINWORD2002	doc
1	ACCEPT FI_WINWORD2003	doc
1	ACCEPT FI_123R1	123
1	ACCEPT FI_123R2	123
1	ACCEPT FI_123R3	123
1	ACCEPT FI_EXCEL	xls
1	ACCEPT FI_EXCEL3	xls
1	ACCEPT FI_EXCEL4	xls
1	ACCEPT FI_123R4	123
1	ACCEPT FI_EXCEL5	xls
1	ACCEPT FI_123R6	123
1	ACCEPT FI_EXCEL97	xls
1	ACCEPT FI_123R9	123
1	ACCEPT FI_EXCEL2000	xls
1	ACCEPT FI_EXCEL2002	xls
1	ACCEPT FI_EXCEL2003	xls
1	ACCEPT FI_FREELANCE	prz
1	ACCEPT FI_POWERPOINT4	ppt
1	ACCEPT FI_POWERPOINT3	ppt
1	ACCEPT FI_POWERPOINT7	ppt
1	ACCEPT FI_FREELANCE3	prz
1	ACCEPT FI_POWERPOINTMAC3	ppt
1	ACCEPT FI_POWERPOINTMAC4	ppt
1	ACCEPT FI_PDF	pdf
1	ACCEPT FI_EXTPOWERPOINT4	ppt
1	ACCEPT FI_EXTPOWERPOINTMAC4	ppt
1	ACCEPT FI_POWERPOINTMACB3	ppt
1	ACCEPT FI_POWERPOINTMACB4	ppt
1	ACCEPT FI_POWERPOINT97	ppt
1	ACCEPT FI_PDFMACBIN	pdf
1	ACCEPT FI_POWERPOINT9597	ppt
1	ACCEPT FI_POWERPOINT2000	ppt
1	ACCEPT FI_POWERPOINT2	ppt
1		
1	NATIVE FI_HTML	htm
1	NATIVE FI_HTML_LATIN2	htm
1	NATIVE FI_HTML_JAPANESESJIS	htm
1	NATIVE FI_HTML_JAPANESEEUC	htm
1	NATIVE FI_HTML_CHINESEBIG5	htm
1	NATIVE FI_HTML_CHINESEEUC	htm
1	NATIVE FI_HTML_CHINESEGB	htm
1	NATIVE FI_HTML_KOREANHANGUL	htm
1	NATIVE FI_HTML_CYRILLIC1251	htm
1	NATIVE FI_HTML_CYRILLICKO18	htm

1	NATIVE FI_CYRILLIC1251	txt
1	NATIVE FI_CYRILLICKO18	txt
1	NATIVE FI_W2KHTML	htm
1	NATIVE FI_XL2KHTML	htm
1	NATIVE FI_PP2KHTML	htm
1	NATIVE FI_XML	xml
1	NATIVE FI_WML	xml
1	NATIVE FI_HTML_JAPANESEJIS	htm
1	NATIVE FI_WML_CHINESEBIG5	xml
1	NATIVE FI_WML_CHINESEEEUC	xml
1	NATIVE FI_WML_CHINESEGB	xml
1	NATIVE FI_WML_CYRILLIC1251	xml
1	NATIVE FI_WML_CYRILLICKO18	xml
1	NATIVE FI_WML_JAPANESEJIS	xml
1	NATIVE FI_WML_JAPANESESJIS	xml
1	NATIVE FI_WML_JAPANESEEEUC	xml
1	NATIVE FI_WML_KOREANHANGUL	xml
1	NATIVE FI_WML_LATIN2	xml
1	NATIVE FI_HTMLUNICODE	htm
1	NATIVE FI_XML_DOCTYPE_HTML	htm
1	NATIVE FI_XHTML	htm
1	NATIVE FI_ASCII	txt
1	NATIVE FI_ANSI	txt
1	NATIVE FI_UNICODE	txt
1	NATIVE FI_ASCII8	txt
1	NATIVE FI_ANSI8	txt
1	NATIVE FI_MAC	txt
1	NATIVE FI_MAC8	txt
1	NATIVE FI_SHIFTJIS	txt
1	NATIVE FI_CHINESEGB	txt
1	NATIVE FI_HANGEUL	txt
1	NATIVE FI_CHINESEBIG5	txt
1	NATIVE FI_LATIN2	txt
1	NATIVE FI_JAPANESE_EUC	txt
1	NATIVE FI_HEBREW_OLDCODE	txt
1	NATIVE FI_HEBREW_PC8	txt
1	NATIVE FI_HEBREW_E0	txt
1	NATIVE FI_HEBREW_WINDOWS	txt
1	NATIVE FI_ARABIC_710	txt
1	NATIVE FI_ARABIC_720	txt
1	NATIVE FI_ARABIC_WINDOWS	txt
1	NATIVE FI_7BITTEXT	txt
1	NATIVE FI_JAPANESE_JIS	txt
1	NATIVE FI_CENTRALEU_1250	txt
1	NATIVE FI_UTF8	txt
1	NATIVE FI_EBCDIC_37	txt
1	NATIVE FI_EBCDIC_273	txt
1	NATIVE FI_EBCDIC_277	txt
1	NATIVE FI_EBCDIC_278	txt
1	NATIVE FI_EBCDIC_280	txt
1	NATIVE FI_EBCDIC_284	txt
1	NATIVE FI_EBCDIC_285	txt
1	NATIVE FI_EBCDIC_297	txt
1	NATIVE FI_EBCDIC_500	txt
1	NATIVE FI_EBCDIC_870	txt
1	NATIVE FI_EBCDIC_871	txt
1	NATIVE FI_EBCDIC_1026	txt

1

## 1 쿼리 처리의 향상

1

다중 서버 구성에서 장애 복구 보호는 서버 레벨뿐만 아니라 컬렉션 레벨에서도 사용 가능합니다. 특정 이유로 인해 검색 서버에서 컬렉션이 사용 불가능하게 되면, 해당 컬렉션에 대한 쿼리가 자동으로 다른 검색 서버로 라우트됩니다.

1

1

## 1 동일한 웹 사이트의 결과 축소

1

SiteDefs.txt 파일을 편집하여 동일한 웹 사이트의 결과를 축소하는 옵션을 지정할 때, 문자열을 URL 접두부와 함께 사이트 ID로 지정할 수 있습니다.

1

1

DB2 II OmniFind Edition, 버전 8.2에서 SiteDefs.txt 파일에는 URL 접두부 목록만 포함되었습니다. 이제, 각 사이트의 ID로 문자열을 지정할 수 있습니다. 다음 예와 같이, 행 시작 부분에 URL 접두부를 입력하고 선택적으로 오른쪽에 문자열 ID를 입력하십시오.

1

1

1

1

http://mycompany.server1.com/hr/	hr
http://mycompany.server2.com/hr/	hr
http://mycompany.server3.com/hr/	hr
http://mycompany.server1.com/finance/	finance
http://mycompany.server1.com/news	news
http://mycompany.server1.com/USA	

1

1

1

1

1

검색 응용프로그램은 URL 접두부 또는 ID를 사용하여 검색 결과의 문서를 그룹화하거나 축소할 수 있습니다. 예를 들어, hr ID를 사용하면 인적 자원 문서를 호스팅하는 세 서버의 검색 결과를 검색 결과에서 함께 축소할 수 있습니다.

1

1

1

엔터프라이즈 검색 쿼리 구문은 이 기능을 지원하도록 확장되었습니다. 예를 들어, 같은 그룹내의 문서를 검색하려면 다음 구문을 사용하십시오. 여기서 *group*은 그룹에 포함된 그룹ID 또는 완전한 URL입니다.

1

1

1

```
samegroupas:group
```

1

예:

1

```
samegroupas:hr
samegroupas:http://mycompany.server3.com/hr/
```

1

1

SiteDefs.txt 파일 갱신에 대한 자세한 정보는 *Administering Enterprise Search*의 *동일한 웹 사이트의 결과 축소* 주제를 참조하십시오.

1

1

## 1 검색 결과의 강조표시 확장

1

쿼리 인터페이스의 등록 정보를 사용하여 검색 결과 세부사항의 여러 영역에서 쿼리 조항을 강조표시할 수 있습니다. DB2 II OmniFind Edition에서는 강조표시가 요약 필드로 제한되었습니다.

1

1

1 검색 응용프로그램을 작성할 때, 쿼리 인터페이스의 getProperty 및 setProperty 메  
1 소드를 사용하여 쿼리 처리를 조작할 수 있습니다. 이러한 메소드에서 사용할 수 있는  
1 새 등록 정보 HighlightingMode를 통해 검색 결과에서 쿼리 조항의 강조표시 방법을  
1 제어할 수 있습니다. 이 등록 정보의 값은 다음과 같습니다.

### 1 **DefaultHighlighting**

1 요약 필드에서만 쿼리 조항을 강조표시합니다. 검색 응용프로그램이  
1 HighlightingMode 등록 정보를 생략하는 경우, 이 값이 디폴트 조치입니다.

### 1 **ExtendedHighlighting**

1 제목, URL 및 기타 필드와 같은 다른 검색 결과 영역으로 쿼리 조항의 강조표시  
1 를 확장합니다.

1 getProperty 또는 getProperties 메소드를 사용하여 결과 오브젝트의 등록 정보에  
1 서 강조표시된 URL 필드를 검색할 수 있습니다. 등록 정보 이름은  
1 HighlightedDocumentID이며, 이 등록 정보의 값은 강조표시된 URL입니다. 결과 메  
1 소드 getDocumentID는 계속 강조표시되지 않은 URL을 리턴합니다.

---

## 1 **검색 응용프로그램의 시간종료 지원**

1 리모트 API(Application Programming Interface)의 등록 정보를 사용하여 리모트 요  
1 청에 대한 시간종료 값을 지정할 수 있습니다.

1 검색 응용프로그램은 다음 메소드에서 Properties 오브젝트로 시간종료 등록 정보를  
1 전달할 수 있습니다.

1 SearchFactory.getSearchService(Properties)  
1 BrowseFactory.getBrowseService(Properties)

1 예:

```
1 // create a new Properties object.  
1 Properties config = new Properties();  
1 config.setProperty("hostname", hostname);  
1 config.setProperty("port", portNumber);  
1 config.setProperty("locale", "en_US");  
1 config.setProperty("timeout", "60");  
1 SearchService searchService = factory.getSearchService(config);
```

1 리모트 요청이 시간종료되기 전에 대기하는 초 수를 나타내는 시간종료 값은 정수(예:  
1 60.5 또는 sixty가 아닌 60)이어야 합니다. 시간종료 값을 지정하지 않으면, 리모트 API가  
1 디폴트값 30초를 사용합니다.

---

## 알려진 제한사항, 문제점 및 해결 방법

이 정보는 DB2 Information Integrator OmniFind Edition, 버전 8.2의 알려진 제한사항, 문제점 및 해결 방법에 대해 설명합니다.

한계 및 제한사항이 제품의 다른 릴리스에 적용될 수도 있고 적용되지 않을 수도 있습니다.

---

## AIX용 DB2 UDB tar 파일 압축 풀기

AIX용 DB2 Universal Database Enterprise Server Edition, 버전 8.2의 경우, DB2 UDB를 설치하려면 먼저 CD 이미지의 압축을 풀어야 합니다.

엔터프라이즈 검색 설치를 시작하기 전에 CD의 압축을 푸십시오.

CD 압축 풀기에 대한 자세한 정보는 DB2 Information Center, 버전 8.2: <http://publib.boulder.ibm.com/infocenter/db2help/index.jsp>로 이동하십시오. DB2 정보 센터에서 데이터베이스 시스템 → 설치 → **Linux, UNIX 및 Windows용 DB2 Universal Database** → **DB2 서버** → **DB2 UDB Enterprise Server Edition**(파티션되지 않음) → **AIX** → **DB2 설치 마법사** 시작을 누르십시오.

엔터프라이즈 검색 설치 프로그램이 DB2 UDB를 설치할 수 있게 하려면 먼저 이 파일의 압축을 풀어야 합니다.

CD 이미지의 압축을 풀려면 다음을 수행하십시오.

1. product.tar.Z를 복사하십시오. 여기서 product는 임시 파일 시스템에 설치하도록 사용이 허가된 제품입니다.
2. DB2 설치 마법사를 시작하려면 다음 명령을 입력하십시오.

```
zcat product.tar.Z | tar -xf - ; ./product/db2setup
```

예를 들어, DB2 UDB Enterprise Server Edition의 제품 이름이 ese이면 다음 명령을 입력하십시오.

```
zcat ese.tar.Z | tar -xf - ; ./ese/db2setup
```

엔터프라이즈 검색 설치 중에 압축을 푼 파일 경로에 db2setup 파일을 포함하는 디렉토리를 제공하십시오. 예를 들어, 압축을 푼 파일의 디렉토리는 .#ese.sbcsaix1일 수 있습니다.

---

## AIX의 공유 메모리 세그먼트

AIX에서 SQL1224N 메시지가 표시되는 경우, AIX에서 허용되는 공유 메모리 세그먼트 수가 초과되었을 수 있습니다.

이 문제점을 해결하려면 IBM 지원 웹 사이트에 있는 SQL1224N 공유 메모리 세그먼트 항목을 참조하십시오.

---

## Windows에서 설치 제거한 후 시스템 재시작

Windows 컴퓨터에서 DB2 II OmniFind Edition을 설치 제거하는 경우, 모든 프로그램 파일의 제거를 완료하려면 시스템을 재시작해야 합니다.

엔터프라이즈 검색 설치 디렉토리도 수동으로 삭제해야 합니다. 시스템을 재시작하지 않거나 프로그램을 삭제하지 않고 엔터프라이즈 검색을 다시 설치할 경우, 문제점이 발생할 수 있습니다.

---

## Linux 및 AIX용 상위 항목이 삭제된 URLFetcher 프로세스

문제점: 웹 크롤러는 두 개의 프로세스를 사용합니다. 한 프로세스가 종료되면 두 번째 프로세스에게 종료 신호를 보냅니다. 종종 한 프로세스가 예기치 않게 종료하면 다른 프로세스가 무한대로 실행될 수 있습니다.

UNIX ps 명령의 출력에 다음 프로세스와 같은 상위 항목이 삭제된 프로세스가 표시될 경우, 이런 일이 발생했는지 여부를 판별할 수 있습니다.

```
F S UID      PID  PPID  C PRI  NI ADDR  SZ WCHAN  STIME TTY      TIME CMD
0 S user1 26222    1  0  75   0   - 152641 schedu 10:34 pts/8 00:00:00
```

```
/opt/IBMJava2-141/jre/bin/java -classpath ../../URLFetcher.jar:...
```

다음 사항을 참조하십시오.

- PPID(상위 프로세스 ID)가 1이면 프로세스의 상위 항목이 삭제되었음을 의미합니다.
- 클래스 경로에 URLFetcher.jar가 있으면 이것이 웹 크롤러 프로세스임을 의미합니다.

### 솔루션

상위 항목이 삭제된 프로세스의 프로세스 ID(PID)를 찾으십시오. 앞의 예제에서 PID는 26222이며 이를 종료하십시오. Linux 및 AIX의 경우, 인수 없이 kill 명령을 사용하십시오. 프로세스 테이블에 『하위』 항목이 있을 수도 있습니다. 하위 항목은 상위 항목이 삭제된 프로세스에 의해 작성된 스레드를 나타냅니다. 상위 구성원 프로세스를 종료할 때 하위 항목을 삭제해야 합니다.

상위 항목이 삭제된 URLFetcher 프로세스가 있을 경우, 이 프로세스가 프로세스 테이블에 있으면 운영 체제가 이에 대한 자원을 예약하기 때문에 프로세스를 종료해야 합니다. 특히 웹 크롤러가 비정상적으로 종료되었다고 의심되는 경우, 상위 항목이 삭제된 프로세스가 있는지 여부를 정기적으로 점검해야 합니다. 그런 프로세스가 존재하면 종료하십시오.

주: DB2 II OmniFind Edition 버전 8.2.1에서, C++ 구성요소가 사라짐을 통보하면 URL 페처가 자체 종료되므로 상위 항목이 삭제된 프로세스가 거의 발생하지 않습니다. 문제점이 발생하는 경우, 실수로 다른 사용자의 httpauth.ini 편집 세션을 종료하지 않도록 주의하십시오(HTTPAuthEdit 유틸리티는 URLFetcher JAR 파일에서 실행되는 JVM이며, 프로세스 테이블에서 유사하게 보일 수 있습니다).

---

## 서버에서의 인덱스 위치

컬렉션을 작성할 때 디폴트 위치가 아닌 인덱스 위치를 지정할 경우, 서버 데이터 루트 디렉토리에 없는 인덱스 위치를 지정하면 디폴트가 아닌 인덱스 위치에 대한 디스크 스페이스 사용량은 모니터링되지 않습니다.

다중 서버 설치의 경우, 컬렉션을 작성할 때 디폴트 위치가 아닌 인덱스 위치를 지정하려면 인덱스 위치가 우선 네 개의 서버 모두에 있어야 합니다.

---

## 힙(heap) 크기 한계

문서 구문 분석기 Java™ 세션(parserservice라고 함)의 힙(heap) 크기는 400MB로 제한됩니다.

힙(heap) 크기 값은 다음과 같이 저장됩니다.

```
ES_NODE_ROOT/config/services.ini:session7.max_heap=400
```

parserservice 세션 중에 OutOfMemory 메시지가 시스템 오류 로그에 표시되면 services.ini 파일에서 parserservice의 max\_heap 값을 늘린 다음 parserservice 세션을 재시작하십시오.

---

## 대형 오브젝트(LOB)가 있는 DB2 UDB 테이블

128MB 이상의 LOB 데이터가 있는 DB2 Universal Database 테이블은 크롤할 수 없습니다. 크기가 8MB 이상인 문서는 인덱스화할 수 없습니다.

---

## 인덱스 상태가 표시되지 않음

이전 설치를 제거하지 않고 새 버전의 DB2 II OmniFind Edition을 재설치하는 경우, 인덱스 모니터링 상태가 올바르게 표시되지 않을 수 있습니다.

인덱스 빌드를 시작하면 관리 콘솔의 인덱스 페이지에 빌드 상태가 올바르게 표시되지 않습니다. 그러나 인덱스는 빌드되며, 빌드가 완료되면 인덱스 페이지에 인덱스화된 올바른 문서의 수와 올바른 인덱스 빌드 횟수가 표시됩니다.

새 버전을 설치하기 전에 이전 버전의 DB2 II OmniFind Edition을 제거하도록 권장됩니다.

---

## Linux 및 AIX용 사용자 ID 권한

새 엔터프라이즈 검색 사용자 ID에 대한 쓰기 권한을 추가해야 할 수 있습니다.

엔터프라이즈 검색 관리자와 다른 Linux 또는 AIX 사용자로 DB2 II OmniFind Edition을 다시 설치하는 경우, DB2 크롤러를 추가하고 카탈로그되지 않은 리모트 데이터베이스(DB2 Universal Database Type 4 JDBC(Java DataBase Connectivity) 드라이버를 사용하는 데이터베이스)를 크롤하도록 지정하면 오류가 발생할 수 있습니다. 이 경우, 새 엔터프라이즈 검색 관리자 사용자에게 대한 쓰기 권한을 `/home/db2inst1/sqllib/tmp/0.spf` 파일에 추가해야 합니다

---

### 1 FileNet 콘텐츠 서비스에 대한 직접 문서 액세스 사용

1 VeniceBridge 크롤러를 사용하여 DB2 Information Integrator Content Edition(이전  
1 의 VeniceBridge) 서버에 의해 관리되는 FileNet 콘텐츠 서비스 저장소를 크롤하는 경  
1 우, 목표 FileNet 콘텐츠 서비스 라이브러리 및 FileNet 웹 서비스는 FileNet 콘텐츠  
1 서비스 커넥터와 작업하며 직접 문서 액세스를 사용하도록 구성해야 합니다. 직접 문서  
1 액세스 지원을 설정하려면 FileNet 웹 서비스 관리자 안내서 및 FileNet 콘텐츠 서비스  
1 관리 문서를 참조하십시오.

---

### 1 별칭 테이블 크롤링 제한사항

1 z/OS®용 IBM DB2 Universal Database 및 Microsoft SQL Server에 대한 IBM DB2  
1 Information Integrator 별칭 테이블을 크롤하는 데 제한사항이 있습니다.

1 목표 테이블이 z/OS용 DB2 Universal Database 또는 Microsoft SQL Server에 대  
1 한 DB2 Information Integrator 별칭 테이블이고 문서 내용 필드가 지정된 경우, 고유  
1 ID에 다음 데이터 유형의 컬럼이 포함될 수 없습니다.

- 1 • LONG VARCHAR
- 1 • LONG VARCHAR FOR BIT DATA
- 1 • LONG VARGRAPHIC

1 목표 테이블이 Microsoft SQL Server용 DB2 Information Integrator 별칭 테이블이  
1 고 별칭 테이블이 UTF-8 코드 세트 데이터베이스에 작성된 경우, BLOB 데이터 유형  
1 을 갖는 컬럼의 문서 내용이 손상될 수 있습니다.



1                    목표 테이블이 z/OS용 DB2 Universal Database 또는 Microsoft SQL Server에 대  
1                    한 DB2 Information Integrator 별칭 테이블이고 문서 내용 필드가 지정된 경우, 다음  
1                    단계에 따라 DB2 크롤러의 성능을 향상시키십시오.  
1                    1. COLLATING\_SEQUENCE 옵션을 Y로 설정하여 서버를 작성하십시오.  
1                    2. 이 서버 아래에서 별칭을 작성하십시오.  
1                    옵션을 Y로 설정하기 전에 별칭 테이블이 이미 정의되어 있는 경우, 별칭을 삭제하고  
1                    다시 작성하십시오.



---

## 주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다. IBM은 모든 국가에서 이 자료에 기술된 제품, 서비스 또는 기능을 제공하지 않을 수도 있습니다. 현재 사용할 수 있는 제품 및 서비스에 대한 정보는 한국 IBM 담당자에게 문의하십시오. 이 책에서 IBM 제품, 프로그램 또는 서비스를 언급하는 것이 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산권을 침해하지 않는 한, 기능상으로 동등한 제품, 프로그램 또는 서비스를 대신 사용할 수 있습니다. 그러나 비IBM 제품, 프로그램 또는 서비스의 운용에 대한 평가 및 검증은 사용자의 책임입니다.

IBM은 이 책에서 다루고 있는 특정 내용에 대해 특허를 보유하고 있거나 현재 특허 출원 중일 수 있습니다. 이 책을 제공한다고 해서 특허에 대한 라이선스까지 부여하는 것은 아닙니다. 라이선스에 대한 의문사항은 다음으로 문의하십시오.

135-270

서울특별시 강남구 도곡동 467-12, 군인공제회관빌딩

한국 아이.비.엠 주식회사

고객만족센터

전화번호: 080-023-8080

2바이트(DBCS) 정보에 관한 라이선스 문의는 한국 IBM 고객만족센터에 문의하거나 다음 주소로 서면 문의하시기 바랍니다.

IBM World Trade Asia Corporation

Licensing

2-31 Roppongi 3-chome, Minato-ku

Tokyo 106, Japan

다음 단락은 현지법과 상충하는 영국이나 기타 국가/지역에서는 적용되지 않습니다. IBM은 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여(단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증없이 이 책을 『현재상태대로』 제공합니다. 일부 국가에서는 특정 거래에 있어서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로 이 사항이 적용되지 않을 수도 있습니다.

이 정보에는 기술적으로 부정확한 내용이나 인쇄상의 오류가 있을 수 있습니다. 이 정보는 주기적으로 변경되며, 변경된 사항은 최신판에 통합됩니다. IBM은 이 책에서 설명한 제품 및(또는) 프로그램을 사전 통지없이 언제든지 개선 및(또는) 변경할 수 있습니다.

이 정보에서 언급되는 비IBM의 웹 사이트는 단지 편의상 제공된 것으로, 어떤 방식으로든 이들 웹 사이트를 옹호하고자 하는 것은 아닙니다. 해당 웹 사이트의 자료는 본 IBM 제품 자료의 일부가 아니므로 해당 웹 사이트 사용으로 인한 위험은 사용자 본인이 감수해야 합니다.

IBM은 귀하의 권리를 침해하지 않는 범위 내에서 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

- (i) 독자적으로 작성된 프로그램과 다른 프로그램(본 프로그램 포함) 간의 정보 교환 및
- (ii) 교환된 정보의 상호 이용을 목적으로 본 프로그램에 관한 정보를 얻고자 하는 라이선스 사용자는 다음 주소로 문의하십시오.

135-270

서울특별시 강남구 도곡동 467-12, 군인공제회관빌딩

한국 아이.비.엠 주식회사

고객만족센터

이러한 정보는 해당 조건(예를 들어, 사용료 지불 등)에 따라 사용할 수 있습니다.

이 정보에 기술된 라이선스가 있는 프로그램 및 사용 가능한 모든 라이선스가 있는 자료는 IBM이 IBM 기본 계약, IBM 프로그램 라이선스 계약(IPLA) 또는 이와 동등한 계약에 따라 제공한 것입니다.

본 문서에 포함된 모든 성능 데이터는 제한된 환경에서 산출된 것입니다. 따라서 다른 운영 환경에서 얻어진 결과는 상당히 다를 수 있습니다. 일부 측정치는 개발 레벨 시스템에서 작성되었을 수 있으며, 따라서 이러한 측정치가 일반적으로 사용되고 있는 시스템에서도 동일하게 나타날 것이라고는 보장할 수 없습니다. 또한, 일부 성능은 추정을 통해 추측되었을 수도 있으므로 실제 결과는 다를 수 있습니다. 이 책의 사용자는 해당 데이터를 사용자의 특정 환경에서 검증해야 합니다.

비IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 비IBM 제품을 테스트하지 않았으므로, 이들 제품과 관련된 성능의 정확성, 호환성 또는 기타 주장에 대해서는 확신할 수 없습니다. 비IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

IBM이 제시하는 방향 또는 의도에 관한 모든 언급은 특별한 통지없이 변경될 수 있습니다.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이 예제에는 가능한 완벽하게 개념을 설명하기 위해 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며, 실제 기업의 이름 및 주소와 유사하더라도 이는 전적으로 우연입니다.

저작권:

이 정보에는 여러 운영 플랫폼에서의 프로그래밍 기법을 보여주는 원어로 된 샘플 응용프로그램이 들어 있습니다. 귀하는 샘플 응용프로그램의 작성 기준이 된 운영 플랫폼의 응용프로그램 프로그래밍 인터페이스에 부합하는 응용프로그램의 개발, 사용, 마케팅 또는 배포를 목적으로 이들 샘플 프로그램을 추가 비용없이 복사, 수정 및 배포할 수 있습니다. 이러한 예제는 모든 조건하에서 철저히 테스트된 것은 아닙니다. 따라서 IBM은 이러한 프로그램의 신뢰성, 서비스 가능성 또는 기능을 보증하거나 암시하지 않습니다. 귀하는 IBM의 응용프로그램 프로그래밍 인터페이스에 부합하는 응용프로그램을 개발, 사용, 마케팅 및 배포하기 위한 목적으로 이러한 샘플 프로그램을 추가 비용 없이 어떤 형태로든 복사, 수정 및 배포할 수 있습니다.

이러한 샘플 프로그램 또는 파생 제품의 각 사본이나 일부에는 반드시 다음과 같은 저작권 표시가 포함되어야 합니다.

Outside In (®) Viewer Technology, © 1992-2004 Stellent, Chicago, IL., Inc. All Rights Reserved.

IBM XSLT Processor Licensed Materials - Property of IBM © Copyright IBM Corp., 1999-2004. All Rights Reserved.

---

## 상표

이 주제는 IBM 상표 및 IBM 상표가 아닌 특정 상표를 나열합니다.

다음은 미국과 기타 국가에서 International Business Machines Corporation의 상표입니다.

IBM  
AIX  
AIX 5L  
DB2  
DB2 Universal Database  
Domino  
Informix  
Lotus  
Lotus Notes  
Notes  
OmniFind  
WebSphere  
xSeries  
z/OS

다음은 타사의 상표 또는 등록상표입니다.

Java 및 모든 Java 기반 상표와 로고는 미국 또는 기타 국가에서 Sun Microsystems, Inc.의 상표 또는 등록상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

Intel, Intel Inside(로고), MMX 및 Pentium은 미국 또는 기타 국가에서 사용되는 Intel Corporation의 상표입니다.

UNIX는 미국 또는 기타 국가에서 Open Group의 등록상표입니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 상표입니다.

기타 회사, 제품 및 서비스 이름은 타사의 상표 또는 서비스표입니다.

---

## IBM에 문의

미국 또는 캐나다에 있는 IBM 고객 지원 센터는 1-800-IBM-SERV(1-800-426-7378)로 문의하십시오.

사용 가능한 서비스 옵션에 대해 알아보려면 다음 번호로 문의하십시오.

- 고객만족센터: 080-023-8080

해당 국가 및 지역의 IBM 지사를 찾으려면 IBM Directory of Worldwide Contacts([www.ibm.com/planetwide](http://www.ibm.com/planetwide))를 참조하십시오.

---

## 제품 정보

DB2 Information Integrator 관련 정보는 전화나 웹에서 얻을 수 있습니다.

DB2 Information Integrator 관련 정보는 전화나 웹에서 얻을 수 있습니다. 여기에 제공된 전화번호는 미국 지역에서 유효합니다.

1. 제품을 주문하거나 일반 정보를 얻으려면 1-800-IBM-CALL(1-800-426-2255)로 문의하십시오.
2. 책을 주문하려면 1-800-879-2755로 문의하십시오.
3. 웹 사이트 [www.ibm.com/software/data/integration/db2ii/support.html](http://www.ibm.com/software/data/integration/db2ii/support.html)을 방문하십시오.

이 사이트에는 다음 최신 정보를 제공합니다.

- 기술 라이브러리
- 책 주문
- 클라이언트 다운로드
- 뉴스그룹
- Fix Paks
- 뉴스
- 웹 자원에 대한 링크

---

## 문서에 대한 의견 제공

이 책 또는 기타 DB2 Information Integrator 문서에 대한 의견을 보내주십시오.

귀하의 피드백은 IBM이 고품질의 정보를 제공하는데 도움이 됩니다. 이 책 또는 기타 DB2 Information Integrator 문서에 대한 의견을 보내주십시오. 다음 방법을 사용하여 의견을 제공하실 수 있습니다.

1. [www.ibm.com/software/data/rcf](http://www.ibm.com/software/data/rcf)에 있는 Online Reader's Comment Form을 사용하여 귀하의 의견을 보내 주십시오.
2. 전자 우편을 통해 [ibmkspoe@kr.ibm.com](mailto:ibmkspoe@kr.ibm.com)으로 귀하의 의견을 보내 주십시오. 가능하다면, 제품의 이름, 제품의 버전 번호 및 책 이름, 부품 번호(해당되는 경우)를 포함시키십시오. 특정 텍스트에 대한 주석을 다는 경우 텍스트의 위치(예: 제목, 테이블 번호 또는 페이지 번호)를 적어주십시오.









**IBM**



**Java**<sup>™</sup>  
**COMPATIBLE**