

멀티플랫폼용 IBM Content Manager



# 컨텐츠용 Information Integrator 관리

버전 8 릴리스 2



멀티플랫폼용 IBM Content Manager



# 컨텐츠용 Information Integrator 관리

버전 8 릴리스 2

주!

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 153 페이지의 『주의사항』에 있는 정보를 읽으십시오.

## 제 2 판 (2003년 3월)

이 개정판은 새 개정판에 별도로 명시하지 않으면 멀티플랫폼용 IBM Enterprise Information Portal(제품 번호 5724-B43)의 버전 8 릴리스 2 및 모든 후속 릴리스와 수정판에 적용됩니다.

Portions of this product are: Copyright © 1990-2000 ActionPoint, Inc. and/or its licensors, 1299 Parkmoor Drive, San Jose, CA 95126 U.S.A. All rights reserved.

Outside In<sup>®</sup> Viewer Technology, ©1992-2000 Inso Corporation. All Rights Reserved.

© Copyright International Business Machines Corporation 1999, 2003. All rights reserved.

# **목차**

이 책에 대하여 . . . . .	vii	Extended Search 커넥터에 대한 작업 . . . . .	29
이 책의 사용자 . . . . .	vii	연합 엔티티 작성 . . . . .	29
시스템 관리자에게 필요한 기술 . . . . .	viii	연합 엔티티 이해 . . . . .	29
비즈니스 분석자 및 프로세스 모델러에게 필요한 기술 . . . . .	viii	연합 엔티티 작성 마법사 사용 . . . . .	30
자세한 정보 찾기 . . . . .	viii	연합 텍스트 색인 작성 . . . . .	31
제품 패키지에 들어 있는 정보 . . . . .	viii	검색 템플릿 작성 . . . . .	31
웹에 제공된 지원 . . . . .	ix	검색 템플릿 정의 . . . . .	32
의견을 보내는 방법 . . . . .	x	검색 기준 정의 . . . . .	32
I EIP 버전 8.2의 새로운 기능 . . . . .	x	검색 설정 판별 . . . . .	32
		사용 권한 할당 . . . . .	33
<b>제 1 장 Enterprise Information Portal 소개 . . . . .</b>	<b>1</b>	<b>제 4 장 사용자 액세스 관리 . . . . .</b>	<b>35</b>
고객 정보 검색 . . . . .	1	사용자 ID 및 암호 작성 . . . . .	35
요구사항 . . . . .	2	I DB2 관리 권한의 이해 . . . . .	36
솔루션 . . . . .	2	I INI 파일을 사용해 DB2에 연결 . . . . .	36
개요 . . . . .	2	I 라이브러리 서버 및 시스템 관리자의 암호를 자원 관리자	
Enterprise Information Portal 구성요소 소개 . . . . .	3	I 로 변경 . . . . .	37
<b>제 2 장 관리 클라이언트 소개 . . . . .</b>	<b>9</b>	I 데이터베이스 액세스 암호 변경 . . . . .	37
EIP First Steps를 사용하여 시스템 관리 클라이언트 이해 . . . . .	9	I LDAP에서 사용자 가져오기 . . . . .	38
EIP 관리 . . . . .	9	I 사용 권한 소개 . . . . .	39
사용자 및 그룹 관리 . . . . .	9	사용 권한 세트 작성 . . . . .	40
관리 클라이언트 도구 사용 . . . . .	10	사용 권한 그룹 작성 . . . . .	40
사용 권한 소개 . . . . .	11	사용자에게 사용 권한 세트 할당 . . . . .	40
사용 권한 . . . . .	12	사용자 ID에 Grant Privilege Set 할당 . . . . .	41
제품 보기와 데이터베이스 전환 . . . . .	13	자원 관리자에 사용자 할당 . . . . .	41
관리 클라이언트 개선사항 및 확장사항 . . . . .	13	컬렉션에 사용자 할당 . . . . .	41
관리 클라이언트를 로컬 관리 데이터베이스에 연결 . . . . .	14	사용자 그룹 작성 . . . . .	41
관리 클라이언트를 원격 관리 데이터베이스에 연결 . . . . .	14	액세스 제어 목록 작성 . . . . .	42
I 1단계 - DB2 Configuration Assistant를 사용하여		액세스 제어 목록에 사용 권한 세트 할당 . . . . .	42
I 는 카탈로그 원격 데이터베이스 . . . . .	15	도메인 작성 . . . . .	42
2단계 - 서버 구성 유틸리티 사용 . . . . .	16	도메인 관리 . . . . .	43
3단계 - 원격 데이터베이스 연결을 테스트 . . . . .	17	도메인 액세스 . . . . .	44
문서 유형 정의 . . . . .	18	도메인에 사용자 할당 . . . . .	44
서버 MIME 유형 파일(cmbcc2mime.ini) 변경 . . . . .	18	도메인에 사용자 그룹 할당 . . . . .	44
<b>제 3 장 EIP 관리 클라이언트 기능 사용 . . . . .</b>	<b>21</b>	도메인에 사용 권한 세트 할당 . . . . .	44
연합 검색 작성 . . . . .	21	도메인에 자원 관리자 할당 . . . . .	45
서버 정의 . . . . .	22	도메인에 컬렉션 할당 . . . . .	45
서버 정의에 대한 지침 . . . . .	23	하나의 도메인에서 다른 도메인으로 사용자 이동 . . . . .	45
OnDemand 커넥터에 대한 작업: TCP/IP 조정 및 소켓 . . . . .	28	하나의 도메인에서 다른 도메인으로 사용자 그룹 이동 . . . . .	46
		하나의 도메인에서 다른 도메인으로 자원 관리자 이동 . . . . .	46

하나의 도메인에서 다른 도메인으로 컬렉션 이동	47	워크플로우 계획	113
하나의 도메인에서 다른 도메인으로 사용 권한 세		처리할 정보	113
트 이동	47	정보 처리 방법	114
하나의 도메인에서 다른 도메인으로 액세스 제어		수행할 조치	114
목록 이동	47	처리에서의 정보 이동 방법	114
		모든 작업이 서로 조화되는 방법	115
<b>제 5 장 Information Mining 관리</b>	49	Enterprise Information Portal 워크플로우 구성요	
Information Mining 개념	49	소 사용	115
Enterprise Information Portal Information		워크플로우 빌더 사용	115
Mining 서비스	49	워크플로우 서비스 사용	116
Information Mining 서비스의 구성요소	50	작업 목록 정의	116
비즈니스 환경에서 Information Mining 사용	52	조치 목록 정의	117
Information Mining 사용 예제	54	워크플로우 작성	117
지원되는 언어 및 형식	57	워크플로우 빌더 사용	117
개념	58	MQSeries 워크플로우 서버 시작	118
시스템 아키텍처	58		
Information Mining 개념	59	<b>제 7 장 IBM Web Crawler 샘플 파일</b>	119
Information Mining 도구	61	config-sample2.xml 샘플	119
인터페이스 프로그래밍	69	IBM Web Crawler 로그 분석 파일 예	121
First Steps	71		
분류법 구축	71	<b>제 8 장 텍스트 검색 및 QBIC® 사용</b>	127
Information Structuring Tool	71	텍스트 검색 엔진을 사용한 문서 검색	127
시작하기	72	텍스트 검색 서버 사용	127
액세스 권한	72	QBIC(Query By Image Content)를 사용하여 이미	
분류법 정의	73	지 검색	128
연계 문서 선택	75	이미지 검색 소개	128
연계 문서 업로드	75	이미지 검색 설정	128
카테고리 모델 평가	78	샘플 데이터 로드 및 색인화	131
카탈로그 연계	83	샘플 데이터를 로드하기 전에	131
성능 조정	83	텍스트 검색 색인 작성	132
IBM Web Crawler 사용	84	이미지 검색 데이터베이스, 카탈로그 및 기능 작	
IBM Web Crawler 성능	85	성	133
웹에 대한 IBM Web Crawler 구성 및 실행	85	로더 프로그램 실행	134
IBM Web Crawler 구성 파일	89	샘플 텍스트 데이터 색인화	136
IBM Web Crawler의 로그 기록	99		
문제점 해결	100	<b>제 9 장 문서 형식</b>	137
요약기 선택	101	Information Mining 문서 형식	137
Notes용 IBM Web Crawler	102	Word 처리: 일반	137
서버에서 IBM Web Crawler 제외	107	Word 처리: DOS	137
		Word 처리: International	138
<b>제 6 장 워크플로우 소개</b>	109	Word 처리: Windows	138
워크플로우 이해	109	Word 처리: Macintosh	139
워크플로우 사용 방법	109	스프레드시트 형식	139
사용자 ID 및 그룹 동기화	110	데이터베이스 형식	140
작동 워크플로우로 EIP 서버 재설치	111	표준 그래픽 형식	141
MQSeries 워크플로우와 EIP 데이터베이스 간		High-end 그래픽 형식	143
사용자 ID 및 그룹 갱신	112	프리젠테이션 형식	144

압축 및 인코드 형식 . . . . .	144	액세스 가능한 표시장치의 기능 . . . . .	151
기타 . . . . .	144	보조 기술과의 호환성. . . . .	152
<b>제 10 장 권한 관리 . . . . .</b>	<b>147</b>	내게 필요한 옵션 문서 . . . . .	152
지식 등록 정보 보호 . . . . .	147	<b>주의사항 . . . . .</b>	<b>153</b>
표시 기법 사용. . . . .	148	상표 . . . . .	155
비주얼 표시. . . . .	149	<b>용어집. . . . .</b>	<b>157</b>
감추기 표시. . . . .	149	<b>색인 . . . . .</b>	<b>165</b>
<b>제 11 장 내게 필요한 옵션 . . . . .</b>	<b>151</b>		
키보드 입력 및 탐색 . . . . .	151		





---

## 이 책에 대하여

이 책에서는 Enterprise Information Portal (EIP) 시스템 관리를 위해 알고 있어야 할 모든 기본 개념에 대해 설명합니다. EIP는 관리 클라이언트에서 관리할 수 있는 여러 구성요소를 제공하고 EIP를 사용하여 다른 제품 기능에 액세스할 수 있기 때문에, 이 안내서는 일반적인 시스템 관리 안내서가 아닙니다. 이 문서의 주제는 다음과 같습니다.

- 비즈니스 수요를 충족시키기 위한 EIP 사용
- 관리 클라이언트 액세스 및 사용
- 사용자 액세스 관리
- 관계형 데이터베이스에 저장되는 구조화 데이터, 구조화되지 않은 멀티미디어 콘텐츠 또는 텍스트 문서를 포함하여 복수 콘텐츠 서버에 있는 콘텐츠를 검색하기 위한 EIP 사용
- 워크플로우 설계, 구현 및 관리

---

## 이 책의 사용자

이 책에서는 EIP 시스템 관리자가 다음 작업을 수행하는 것을 도와줍니다.

### 시스템 관리

데이터베이스, 서버 및 네트워크 관리 포함

### 사용자 관리

개인 및 그룹에 대한 액세스 정의 및 부여, 액세스 제어 목록 유지보수

### 연합 검색

Content Management 시스템에서 콘텐츠를 검색하기 위한 연합 검색 템플릿 정의 및 사용

### Information Mining

문서에서 정보 추출, 문서 및 검색 결과 구분

### 웹 검색

IBM® Web Crawler를 사용하여 웹에서 콘텐츠 검색 및 가져오기

### 텍스트 검색

IBM DB2® TIE 또는 IBM 텍스트 검색 엔진(Content Manager 버전 7.1 및 이전 버전만)을 사용한 문서 검색 및 색인화

### 이미지 검색

Content Manager 버전 7.1(이전 버전)을 사용한 이미지 검색 수행

---

## 시스템 관리자에게 필요한 기술

수행하는 작업에 따라 다음을 이해해야 합니다.

- 사용자 액세스에 대한 보안 프로토콜
- Windows NT®, Windows® XP, Windows 2000, AIX® 또는 Solaris 운영 체제
- 네트워크 관리
- Content Management 시스템에서 콘텐츠 서버의 데이터 모델
- 데이터베이스 관리
- 검색 템플릿 작성 시 콘텐츠 및 검색 기준에 대한 실무 지식 적용 방법
- Information Mining 기술 및 도구
- 워크플로우 설계의 원리
- EIP 워크플로우로 지원하려는 비즈니스 처리

---

## 비즈니스 분석자 및 프로세스 모델러에게 필요한 기술

비즈니스 분석자와 프로세스 모델러도 이 안내서에서 자신의 기업을 위해 EIP 워크플로우를 정의하고 모델링하는 방법에 대한 개념적 정보를 찾을 수 있습니다.

Enterprise Information Portal 워크플로우 빌더를 사용하려면 다음을 수행해야 합니다.

- 기업의 비즈니스 처리에 사용되는 스탭프 요구사항, 프로그램 및 데이터 구조 이해
- 기업의 비즈니스 또는 워크플로우 처리에 관한 의사 결정

---

## 자세한 정보 찾기

제품 패키지에는 시스템 계획, 설치, 관리 및 사용에 필요한 완벽한 정보 세트가 포함되어 있습니다. 웹에서도 제품 문서 및 지원을 받을 수 있습니다.

### 제품 패키지에 들어 있는 정보

제품 패키지에는 Information Center 및 각각의 PDF(Portable Document Format) 포맷으로 된 서적이 들어 있습니다.

#### Information Center

제품 패키지에는 제품 설치 시 설치할 수 있는 Information Center가 들어 있습니다. Information Center 설치에 대한 자세한 정보는 *Content Management 시스템 계획 및 설치*를 참조하십시오.

Information Center는 Content Manager, Enterprise Information Portal 및 멀티플랫폼용 IBM Content Manager VideoCharger에 대한 문서를 포함합니다. 주제 기반 정보는 제품별 및 작업별로 구성됩니다(예: 관리). 제공된 탐색 메커니즘과 색인 이외에, 검색 기능은 검색 기능을 돕습니다.

## PDF 서적

운영 체제의 경우 Adobe Acrobat Reader를 사용하여 온라인으로 PDF 파일을 볼 수 있습니다. Acrobat Reader를 설치하지 않은 경우 <http://www.adobe.com>의 Adobe 웹사이트에서 다운로드할 수 있습니다.

표 1에서는 멀티플랫폼용 IBM Content Manager에 포함된 Content Manager 서적을 표시합니다.

표 1. Content Manager 서적

파일 이름	제목	서적 번호
설치	Content Management 시스템 계획 및 설치 <sup>1</sup>	GA30-1545-01
이주	Content Manager 버전 8로 이주	SA30-1598-01
sysadmin	시스템 관리 안내서	SA30-1546-01

멀티플랫폼용 IBM Content Manager를 주문하면 멀티플랫폼용 IBM Enterprise Information Portal도 받게 됩니다. 또는 별도로 멀티플랫폼용 IBM Enterprise Information Portal을 주문할 수 있습니다. 표 2는 제품에 포함된 Enterprise Information Portal 서적을 표시합니다.

표 2. Enterprise Information Portal 서적

파일 이름	제목	서적 번호
apgwork	Windows용 응용프로그램 프로그래밍 안내서 <sup>1</sup>	SA30-1551-01
ecliinst	eClient 설치, 구성 및 관리	SA30-1548-02
eipinst	컨텐츠용 IBM Information Integrator 계획 및 설치	GA30-1549-01
eipmanag	컨텐츠용 IBM Information Integrator 관리	SA30-1550-01
messcode	메시지 및 코드 <sup>2</sup>	SA30-1552-01

주:

1. Windows용 응용프로그램 프로그래밍 안내서에는 Content Manager 및 Enterprise Information Portal의 프로그래밍 응용프로그램에 대한 정보가 들어 있습니다.
2. 메세지 및 코드에는 Content Manager 및 Enterprise Information Portal의 메시지 및 코드가 들어 있습니다.

## 웹에 제공된 자원

제품 지원이 웹에 제공됩니다. 다음 제품 웹 사이트에서 **Support**를 누르십시오.

[www.ibm.com/software/data/cm/](http://www.ibm.com/software/data/cm/)

[www.ibm.com/software/data/eip/](http://www.ibm.com/software/data/eip/)

이 문서는 제품과 함께 소프트웨어로 제공됩니다. 웹의 제품 문서에 액세스하려면 제품 웹 사이트에서 **Library**를 누르십시오.

EDO(Enterprise Documentation Online)라는 HTML 기반 문서 인터페이스도 웹에서 구할 수 있습니다. 현재 API 참조 정보가 포함됩니다. EDO 액세스에 대한 정보를 보려면 Enterprise Information Portal 라이브러리 웹 페이지로 이동하십시오.

## 의견을 보내는 방법

피드백은 IBM에서 양질의 정보를 제공하는 데 도움이 됩니다. 이 책이나 다른 Content Manager 또는 Enterprise Information Portal 문서에 대한 어떤 의견이든지 보내주십시오. 다음과 같은 방법으로 의견을 보내주십시오.

- 웹을 통해 의견을 보내주십시오. 다음 IBM Data Management 온라인 독자 의견 양식(RCF) 페이지를 방문하십시오.

[www.ibm.com/software/data/rcf](http://www.ibm.com/software/data/rcf)

페이지를 사용하여 의견을 입력하고 보낼 수 있습니다.

- 전자 우편을 사용하여 [comments@vnet.ibm.com](mailto:comments@vnet.ibm.com)으로 의견을 보내주십시오. 제품 이름, 버전 번호, 책의 제목과 부품 번호(필요한 경우에는)를 기입하십시오. 특정 텍스트에 대한 의견을 보내는 경우, 텍스트 위치(예: 장과 절 제목, 테이블 번호, 페이지 번호 또는 도움말 주제 제목)를 기입하십시오.

---

## I EIP 버전 8.2의 새로운 기능

다음 변경사항이 제품에 작성되었습니다.

### Sun Solaris에 대한 지원

Solaris 시스템에서 커넥터, 기능 및 데이터베이스를 설치할 수 있습니다.

### 공용 시스템 관리

단일 클라이언트 응용프로그램은 Content Manager 및 Enterprise Information Portal 관리에 대한 별도의 액세스를 제공합니다.

### 새 커넥터

- Content Manager 버전 8 릴리스 1에 대한 ICM 커넥터는 Content Manager 버전 8의 강력한 문서 기억영역 기능을 이용하도록 허용합니다.
- 새 C++ Extended Search 버전 3.7 커넥터는 AIX에서 실행합니다.

### 개선된 커넥터

- 매개변수식 텍스트 검색이 연합 계층으로부터, 그리고 직접 Extended Search 연결을 통해 지원됩니다.
- 다음을 포함하여, OnDemand에 대한 기능 향상 및 성능 개선이 있습니다.

- OnDemand DDO의 구조에 대한 수정.
- 이제 비동기 검색이 지원됩니다.

### **새 Information Mining 서비스**

- 기능 추출
- 클러스터링
- 언어 ID

### **IBM Web Crawler**

IBM Web Crawler는 웹과 Lotus Notes® 데이터베이스에서 정보를 검색하고 요약하게 하는 기능입니다.

### **워크플로우 향상**

워크플로우는 이제 AIX 및 Solaris에서 전체 지원됩니다. 워크플로우 빌더, API 및 JavaBeans™는 개선된 워크플로우 기능 및 사용 가능성을 제공합니다.

### **Information Center**

브라우저 기반 Information Center는 Content Manager, Enterprise Information Portal 및 멀티플랫폼용 IBM Content Manager VideoCharger™에 대한 문서를 포함합니다. 주제 기반 정보는 제품별 및 작업별로 구성됩니다(예: 관리). 제공된 탐색 메커니즘과 색인 이외에, 검색 기능은 검색 기능을 돕습니다.

### **내게 필요한 옵션**

내게 필요한 옵션 기능은 거동이나 시각에 제한을 받는 신체적 장애가 있는 사람들이 소프트웨어 제품을 용이하게 사용할 수 있도록 해줍니다. 이 제품의 주요 내게 필요한 옵션 기능은 다음과 같습니다.

- 마우스 대신에 키보드를 사용하여 모든 기능을 작동할 수 있는 기능
- 향상된 표시장치 등록 정보 지원
- 비디오 및 오디오 경보 신호에 대한 옵션
- 보조 기술과의 호환성
- 운영 체제 내게 필요한 옵션 기능과의 호환성
- 내게 필요한 옵션 문서 형식



---

## 제 1 장 Enterprise Information Portal 소개

보험 회사 및 재정 기관과 같이 종이 사용이 많은 수많은 기업에서는 대량의 비즈니스 관련 콘텐츠를 관리합니다. 많은 산업은 비즈니스 정보 관리 및 액세스를 위한 기업 솔루션을 필요로 합니다.

콘텐츠 서버는 직원이 콘텐츠를 처리하고 작업할 수 있는 메타데이터와 함께 멀티미디어, 비즈니스 양식, 문서 및 관련 데이터를 저장하는 소프트웨어 시스템입니다. 서로 다른 콘텐츠 서버를 효과적으로 연결할 방법이 없을 때, 정보를 복사하거나 직원들이 다중 검색을 수행하도록 연계시키기 위해 시간과 돈을 낭비할 수 있습니다.

Enterprise Information Portal(EIP)에서는 워크스테이션 데스크탑에서 모든 기업 자원을 가져올 수 있는 최신 기술을 제공합니다. EIP를 사용하면 하나의 클라이언트로 서로 다른 콘텐츠 서버를 연결하여 정보 및 멀티미디어 자산의 가치를 극대화할 수 있습니다. EIP 클라이언트를 사용하면 사용자가 연결된 모든 콘텐츠 서버에 동시에 신속하게 액세스할 수 있습니다. 또한 사용자는 웹 또는 인트라넷을 포함한 콘텐츠 서버에서 Information Mining 또는 고급 검색을 수행할 수 있습니다. 사용자가 정의한 비즈니스 프로세스 내에서 워크플로우 작업을 수행할 수 있습니다.

EIP를 사용하면 기업의 응용프로그램을 사용자 조정할 수 있습니다. 응용프로그램 프로그래머는 EIP 샘플을 사용하여 데스크탑 및 웹 기반 응용프로그램을 모두 작성할 수 있습니다.

이 절에서는 EIP 개요에 대해 설명합니다. 가상의 보험 회사인 XYZ 보험에 대한 시나리오로 EIP의 특성 및 기능에 대해 설명합니다.

---

### 고객 정보 검색

대형 화재 및 생명 보험 회사인 XYZ 보험(XYZ)은 방대한 양의 사진, 청구, 보험 증권, 조정자의 메모, 감정서 및 기타 비즈니스 문서를 보유합니다.

XYZ는 Lotus® Domino™.Doc 파일 캐비닛의 건강 진단 및 평가 전자 양식과 함께 보험 계약자에게 보낸 모든 메모를 보관합니다. XYZ는 장기 저장 및 빠른 액세스를 위해 모든 보험 증서, 주의사항 및 송장을 Content Manager OnDemand 서버에 아카이브합니다. XYZ는 보험 계약자로부터 받은 모든 보험 지급 청구 양식, 사진 및 서신을 iSeries용 Content Manager 시스템 폴더에 저장합니다. 감정서는 DB2 DB2 Universal Database™(UDB) Data Warehouse Center Information Catalog Manager에 보관합니다. 고해상도 그래픽과 같은 광고, 홍보 및 새로운 비즈니스 분야에서 공유할 수 있는 Content Manager 시스템에 기업 미디어도 저장합니다. 또한 회사 프로시저와 같은 정보는 회사 인트라넷에 보관합니다.

## 요구사항

직원이 모든 고객 정보에 액세스해야 하므로 한 서버의 콘텐츠로 청구, 고객의 호출 및 일반 보험 계약자인 서비스를 처리할 수 없습니다. 고객 서비스를 제공하려면 직원은 동시에 다양한 콘텐츠 서버에 액세스해야 합니다. XYZ 보험에는 정보 검색을 위해 콘텐츠 서버와 회사 인트라넷을 연결하는 솔루션이 필요합니다. 또한 워크플로우 처리 사용을 확장하고자 합니다.

청구 담당자부터 조정자 및 조직원에 이르기까지 많은 직원이 문서에 액세스해야 합니다. XYZ에서는 특정 항목에 대한 액세스는 제한하는 반면, 다른 항목에 대한 액세스는 제한없이 제공해야 합니다. 또한 연계의 필요성을 줄일 수 있도록 사용이 쉬운 인터페이스도 필요합니다.

## 솔루션

XYZ 보험은 포괄적인 검색 기술을 사용하여 모든 콘텐츠 서버에 연결하여 데이터를 검색할 수 있으므로 EIP를 전개합니다. 이제 XYZ 고객 센터 담당자가 전화를 받으면 한 번의 연합 검색으로 모든 필수 보험 계약자 정보를 검색하게 됩니다.

XYZ 보험에서는 또한 EIP Information Mining 기능을 사용하여 회사의 인트라넷에서 정보를 검색합니다. 워크플로우 처리의 사용도 확장하고자 합니다.

---

## 개요

EIP는 제품의 구성요소들이 적절히 조화되어 기업에 적합한 고유 솔루션을 제공하는 포괄적인 제품입니다. 복수 층 아키텍처에 중점을 두고 있는 EIP에서는 검색 관리를 위한 관리 클라이언트, 검색을 실행하는 클라이언트 및 서로 다른 콘텐츠 서버(예: IBM Content Manager, OS/390®용 Content Manager ImagePlus®, Content Manager OnDemand, Lotus Domino.Doc, DB2 Universal Database, DB2 DataJoiner® 및 DB2 Data Warehouse Center Information Catalog Manager)에 연결하기 위한 커넥터를 제공합니다. EIP 커넥터 툴킷 및 샘플을 사용하여 추가 콘텐츠 서버에 대해 추가 커넥터를 작성할 수 있습니다.

EIP 아키텍처에서는 클라이언트 응용프로그램이 하나 이상의 콘텐츠 서버에 단일 검색을 실행할 수 있습니다. 검색을 수행하기 위해, 클라이언트는 EIP 시스템 관리자가 정의하는 검색 템플릿을 사용합니다.

검색 템플릿을 사용하여 클라이언트는 검색 템플릿에서 사용되는 연합 속성과 맵핑된 원래의 속성이 있는 콘텐츠 서버에 대해 동시에 실행되는 검색인 연합 검색을 실행합니다. EIP 검색 템플릿에는 검색 기준이 들어 있으며, 검색 기준은 각 콘텐츠 서버에 있는 원래 속성에 맵핑되는 연합 속성을 참조합니다. EIP 시스템 관리자는 검색 템



플리트를 작성합니다. EIP는 여러 콘텐츠 서버에 저장된 데이터를 액세스하고 검색하도록 커넥터를 제공합니다. 그러면 콘텐츠 서버는 데이터 오브젝트를 클라이언트에 리턴합니다.

EIP 아키텍처에는 다음과 같은 장점이 있습니다.

- e-business 트랜잭션 및 고객 서비스 응용프로그램을 지원하는 다양한 복수 콘텐츠 서버에 단일 조회를 사용하는 액세스.
- 웹을 포함하여 여러 콘텐츠 서버를 통한 Information Mining 기능.
- 복수의 이기종 콘텐츠 서버를 통한 데이터에 대한 워크플로우 처리 액세스.
- 클라이언트 응용프로그램, 색인 및 데이터의 분리로 인해 임의의 콘텐츠 서버의 데이터 위치에 관계없는 클라이언트 응용프로그램의 개발 지원.

## Enterprise Information Portal 구성요소 소개

이 절에서는 각 EIP 구성요소와 설치 옵션에 대해 설명합니다.

3 페이지의 표 3에서는 구성요소와 호환 가능한 운영 체제를 나열합니다.

표 3. EIP 구성요소 운영 체제 호환성

구성요소	Windows	AIX	Solaris	주
관리 데이터베이스	예	예	예	데이터베이스는 워크플로우 빌더 기능을 포함합니다.
관리 클라이언트	예	아니오	아니오	클라이언트는 Windows, AIX 또는 Solaris 운영 체제에 설치된 데이터베이스에 연결할 수 있습니다.
커넥터	예	예	예	
Information Mining	예	예	예	
IBM Web Crawler	예	예	예	
텍스트 검색 클라이언트	예	예	예	
이미지 검색 클라이언트	예	예	예	

표 3. EIP 구성요소 운영 체제 호환성 (계속)

구성요소	Windows	AIX	Solaris	주
커넥터 툴킷 및 샘플	예	예	예	<ul style="list-style-type: none"> <li>Windows 버전은 샘플 클라이언트를 컴파일할 소스 코드를 포함합니다. 어떤 샘플 클라이언트 코드도 AIX에 설치되어 있지 않습니다.</li> <li>워크플로우 샘플 및 API는 연합 커넥터 샘플로 설치됩니다.</li> </ul>
보기 프로그램	예	아니오	아니오	OnDemand 클라이언트 및 보기 프로그램을 설치합니다.
Information Center	예	예	예	

## 관리

관리 구성요소는 관리 데이터베이스 및 관리 클라이언트 하위 구성요소를 제공합니다. 관리 데이터베이스를 설치할 때, 워크플로우 기능도 설치합니다.

**관리 데이터베이스:** 관리 데이터베이스는 EIP 사용자와 그룹, 사용 권한 레벨, 암호, 사용자 ID 및 기타 정보에 대한 정보를 관리하는 DB2 데이터베이스입니다. 또한 데이터베이스는 워크플로우와 선택적으로 Information Mining 기능을 제공합니다. 여러 개의 데이터베이스를 설치할 수 있습니다. 각 데이터베이스는 EIP 워크플로우 기능을 제공합니다. Content Manager 버전 8 시스템이 있는 경우, EIP 관리 데이터베이스와 Content Manager 버전 8 라이브러리 서버 데이터베이스를 공유할 수 있습니다. 라이브러리 서버 데이터베이스에 EIP가 필요로 하는 정보가 모두 포함되어 있으므로 데이터베이스를 공유할 수 있습니다.

**관리 클라이언트:** 관리 클라이언트는 Windows 워크스테이션에서만 설치될 수 있습니다. 여러 개의 클라이언트를 설치할 수 있습니다. Content Manager 버전 8 시스템이 있는 경우, EIP 및 Content Manager 버전 8을 동일한 클라이언트에서 관리할 수 있습니다.

클라이언트가 제공하는 인터페이스에서 시스템 관리자는 다음을 수행할 수 있습니다.

- 연합 검색에 대해 각 콘텐츠 서버 정의.
- 콘텐츠 서버에서 원래의 엔티티와 속성을 식별하여 연합 엔티티로 맵핑.
- 모든 사용자 정의 콘텐츠 서버에 대한 원래 엔티티 및 속성 목록을 유지보수.
- 검색 템플릿 작성.

- 사용자 및 그룹 식별 및 관리.
- 사용자에게 권한 및 권한 세트 할당.
- 검색으로 얻은 정보로 사용자가 수행할 수 있는 조치에 대한 조건을 설정하고 검색 템플릿에 대한 액세스 정의.
- 비즈니스 워크플로우 처리 디자인 및 관리.

## 커넥터

커넥터는 EIP 클라이언트, 콘텐츠 서버 및 관리 데이터베이스간의 통신 인터페이스를 제공합니다. Content Manager 버전 7.1 커넥터와 같은 콘텐츠 서버 커넥터는 EIP가 서버에 로그인하고, 정보를 검색하고 관리나 일반 사용자 클라이언트로 정보를 리턴하게 하는 기능을 제공합니다. 연합 커넥터는 관리 클라이언트를 관리 데이터베이스로 연결합니다.

EIP는 다음과 같은 커넥터를 제공합니다.

- 연합 커넥터는 EIP 클라이언트를 관리 데이터베이스로 연결합니다.
- DB2 Universal Database 7.1, JDBC driver 1.3(Java™ 전용), ODBC 3.0(C++ 전용), DataJoiner 2.1.1용 관계형 데이터베이스 커넥터.
- Content Manager 버전 7.1 서버용 Content Manager 커넥터
- Content Manager 버전 8.2 서버용 Content Manager 커넥터
- Content Manager OnDemand 버전 7.1용 Content Manager OnDemand 커넥터
- 400® 버전 4.3 및 버전 5.1용 VisualInfo™용 Content Manager
- Image Plus/390 ODM Version 3.1, ImagePlus/390 Folder Application Facility Version 3.1용 OS/390용 Content Manager ImagePlus 커넥터.
- Domino.Doc Version 3.0a, Desktop Enabler Version 3.0a용 Lotus Domino.Doc 커넥터.
- 버전 3.7용 Extended Search 커넥터.
- DB2용 Information Catalog Manager 커넥터. Universal Database Visual Warehouse™ 버전 5.2, DB2 Universal Database Version 7.2.

## 기능

EIP는 네 개의 선택적 기능을 갖습니다.

### Information Mining

Information Mining은 콘텐츠 서버에 있는 텍스트 문서에서 감춰진 정보를 찾기 위한 언어 서비스를 제공합니다. 텍스트 문서 처리 시 요약, 카테고리 및 검색할 수 있는 메타데이터가 작성됩니다. WebSphere® Application Server 4.0(표준 또는 고급판)은 Information Mining 전제조건입니다. 또한 유사한 문서를 cluster할 수 있으며, 사람이나 회사 이름과 같은 기능들을 추출하고 문서의 언어를 결정할 수 있습니다.

## 이미지 검색 클라이언트

Content Manager 버전 7 콘텐츠 서버에서 이미지 검색 기능을 액세스하고 관리하는 데 필요한 인터페이스를 제공합니다.

## 텍스트 검색 클라이언트

텍스트 검색 서버에서 텍스트 검색 기능을 액세스하고 관리하는 데 필요한 인터페이스를 제공합니다.

## IBM Web Crawler

웹 크롤러는 Java 기반 콘텐츠 크롤러 및 마이너입니다. 웹 크롤러는 인트라넷, 외부넷 또는 인터넷 웹 Lotus Notes 데이터베이스 또는 Domino, 로컬 파일 시스템 및 FTP 컬렉션을 통해 콘텐츠를 크롤할 수 있습니다.

웹 크롤러는 여러 유형의 콘텐츠에서 메타데이터 및 텍스트를 마이닝할 수 있습니다. 예를 들어, HTML 콘텐츠는 저자, 키워드, 설명 등과 같은 메타태그, 마지막 수정 시간, 본문, 제목 및 URL로 마이닝될 수 있습니다. 사용자는 주어진 유형의 콘텐츠에 대해 미리 정의된 마이너 세트에서 선택합니다. 콘텐츠 및/또는 마이닝 메타데이터는 로컬 디스크로 저장됩니다.

## 내용 보기 프로그램

OnDemand 보기 프로그램을 설치하면 OnDemand 서버에서 검색된 문서를 보기에 필요한 OnDemand 클라이언트 및 기타 파일이 설치됩니다.

## 커넥터 툴킷 및 샘플

EIP는 다음과 같은 다양한 EIP 기능을 실험하고 테스트하는 데 사용할 수 있는 샘플 프로그램을 포함하는 커넥터 툴킷을 제공합니다.

- 콘텐츠 서버로 연결 또는 콘텐츠 서버에서 연결 해제
- SQL 및 기타 샘플 조회를 콘텐츠 서버에서 수행
- 콘텐츠 서버 MIME 유형 등을 판별

**Windows 커넥터 툴킷:** Windows 서버에서 커넥터 툴킷 및 샘플을 설치하려면, 개별 워크스테이션 기계 유형을 선택하십시오. 그리고 나서 커넥터 툴킷 및 샘플 구성요소를 선택하십시오. 모든 커넥터에 대해 샘플 프로그램을 설치하거나 개별 샘플을 선택하여 설치한 커넥터에 일치시킬 수 있습니다.

Windows 서버에서 커넥터 툴킷 샘플 프로그램은 다음과 같은 방식으로 조직됩니다.

```
c:\CMBROOT\SAMPLES\activex\xx
c:\CMBROOT\SAMPLES\cpp\xx
c:\CMBROOT\SAMPLES\java\xx
c:\CMBROOT\SAMPLES\jsp\xx
c:\CMBROOT\SAMPLES\server\xx
```

xx는 각 응용프로그램 커넥터(예: db2, od, d1 등)에 대해 샘플 프로그램을 포함하는 디렉토리 이름입니다.

AIX 서버에서 샘플 프로그램은 다음과 같은 방식으로 조직됩니다.

```
/usr/lpp/cmb/samples/cpp/icm
/usr/lpp/cmb/samples/java/xx
/usr/lpp/cmb/samples/jsp/xx
/usr/lpp/cmb/samples/server/exit
```

xx는 서브디렉토리 이름입니다(예: Beans, servlets 등).

Solaris 서버에서 샘플 프로그램은 다음과 같은 방식으로 조직됩니다.

```
/opt/IBMcmb/samples/java/xx
/opt/IBMcmb/samples/jsp/xx
/opt/IBMcmb/samples/server/exit
```

xx는 서브디렉토리 이름입니다(예: Beans, servlets 등).

샘플 프로그램은 프로그램을 설명하는 문서를 포함하고 샘플 코드를 사용하여 작업하는 데 필요한 서버 설정(환경 설정, 메모리 등)을 제공합니다.

## Information Center

Information Center 구성요소는 Enterprise Information Portal Information Center를 포함합니다. Information Center는 Enterprise Information Portal 라이브러리의 웹 기반, 검색 가능한 버전입니다.



---

## 제 2 장 관리 클라이언트 소개

관리 클라이언트는 EIP 관리 데이터베이스와 EIP 시스템 관리자 사이의 인터페이스를 제공합니다. 이 절에서는 클라이언트가 EIP 시스템 관리를 돕기 위해 제공하는 여러 기능에 대해 설명합니다.

클라이언트의 왼쪽 분할 영역에 있는 아이콘을 사용하여 서버 정의 및 사용자 관리와 같은 일부 기능에 액세스할 수 있습니다. 다른 기능들은 도구 메뉴 막대를 통해 액세스할 수 있습니다.

---

### EIP First Steps를 사용하여 시스템 관리 클라이언트 이해

*First Steps*는 EIP 설치 시 제공되는 모듈입니다. *First Steps*는 샘플 데이터를 제공하며 실제 데이터를 사용하지 않아도 되도록 오브젝트를 상주시킵니다. *First Steps*를 사용하여 관리 클라이언트의 기본 구조, 외관 및 느낌을 이해하도록 서버 정의, 사용자 및 그룹, 그리고 기타 기능을 검색하십시오.

---

### EIP 관리

시스템 관리자로서 관리 클라이언트를 통해 다음 작업 중 하나 이상을 완료할 수 있습니다.

- 콘텐츠 서버 정의
- 사용자 및 그룹 관리
- 사용 권한 및 액세스 레벨 관리
- 연합 검색 템플릿 작성
- 연합 엔티티 작성
- 관리 도메인이 작동된 경우, 서브도메인 작성.
- 워크플로우가 작동된 경우 워크플로우로 작업.
- Content Manager 버전 7에서 연합 텍스트 엔티티 작성.

---

### 사용자 및 그룹 관리

사용자 ID 및 사용 권한을 작성하여 다중 콘텐츠 서버에서 문서를 검색하고 작업하도록 사용자에게 액세스를 허용합니다. 적절한 사용 권한을 정의하여 사용자에게 할당함으로써 시스템에 저장된 데이터에 대한 액세스를 제한할 수 있습니다.

---

## 관리 클라이언트 도구 사용

이 절에서는 관리 클라이언트가 제공하는 도구에 대해 설명합니다.

### LDAP 구성

이 옵션을 누르면 EIP가 네 개의 탭이 있는 창을 시작합니다.

- LDAP 탭 - LDAP 서버에서 데이터 소스 가져오기를 사용 가능하게 하거나, LDAP 사용자 가져오기 및 인증을 사용 가능하게 하거나, 또는 이 둘을 모두 선택할 수 있습니다.
- 서버 탭 - 호스트 이름, 사용자 이름, 참조 유형 등을 포함하여 LDAP 서버 스펙을 정의하는 필드가 있습니다.
- 인증 탭 - 보안 소켓 계층 정보를 정의하는 필드가 있습니다.
- 고급 탭 - 최대 레코드 수 및 서버 시간종료에 대한 설정을 정의합니다.

### 사용자 맵핑 옵션

이 옵션을 사용하여 기본 설정인 사용자 맵핑 사용 기능을 사용 불가능 하게 할 수 있습니다.

### 연합 사용자 맵핑 편집기

연합 사용자 편집기는 사용자 목록을 표시하며 사용자를 특정 콘텐츠 서버로 맵핑하는 옵션을 제공합니다.

### 검색 템플릿 보기 프로그램

검색 템플릿 보기 프로그램은 모든 검색 템플릿에 대한 자세한 정보를 제공합니다. 이 보기 프로그램은 검색 템플릿 세부사항을 볼 수 있는 세 개의 옵션을 제공합니다.

- 연관된 맵핑(기본값) - 연합 엔티티에 대한 세부사항 및 검색 템플릿에 대한 기타 세부사항을 제공합니다.
- 검색 템플릿 - 기본 연산자, 기본값 등에 대한 세부사항을 제공합니다.
- 결과 표시 - 이름 표시, 너비 표시, 기준 순서 등에 대한 세부사항을 제공합니다.

### 서버 목록 보기 프로그램

선택된 서버의 목록을 표시합니다.

### 로그 표시기

로그 표시기를 사용하여 서버 목록을 화면 갱신한 후에 생성된 로그를 보십시오. 새 목록과 이전 목록간에 차이점이 있으면 로그에서 메시지 목록을 표시합니다.

서비스 서비스를 선택하여 워크플로우 및/또는 Information Mining을 사용 가능하게 하십시오.



## 관리 도메인

관리 도메인을 선택하여 사용 가능하게 하십시오. 관리 도메인은 사용 가능하게 된 후에 사용 불가능하게 될 수 없습니다.

## MIME 유형 편집기

MIME 유형 편집기는 각 콘텐츠 서버에 대한 다음 정보를 나열합니다.

- 콘텐츠 클래스
- 파일 확장자
- 관계형 데이터베이스(RDB) 열
- MIME 유형

MIME 유형 편집기에 나열되는 콘텐츠 서버 이름은 축약형이며 새 콘텐츠 서버를 정의할 때 나타나는 콘텐츠 서버 이름 목록과 일치합니다. 팁: DL은 Content Manager 버전 7.1 콘텐츠 서버에 대한 약어입니다. V4는 AS/400® 콘텐츠 서버에 대한 Content Manager의 약어입니다.

MIME 유형 편집기에서 기본 정보를 제거 및 편집할 수 있으며 정보를 추가할 수도 있습니다.

## MIME 대 응용프로그램 편집기

MIME 대 응용프로그램 편집기를 사용하여 다섯 개의 기본 MIME 대 응용프로그램 연관을 삭제 또는 편집하거나 다른 연관을 추가하십시오. MIME 대 응용프로그램 편집기에 정의된 값 및 설정은 일반 사용자 클라이언트에서 사용하는 보기 프로그램에 영향을 미칩니다.

## 서버 유형 정의

시스템 프로그래머가 개발한 사용자 조정 서버를 정의하려면 이 도구를 사용하십시오.

## DB2 ID/암호 변경

연결 전용 DB2 사용자 ID 및 암호를 수정하려면 이 옵션을 선택하십시오. 이것은 시스템 관리자 사용자 ID와는 완전히 별도의 사용자 ID입니다.

---

## 사용 권한 소개

이 절에서는 Enterprise Information Portal 사용 권한을 설명합니다. 권한 아이콘을 펼쳐서 네 개의 사용 권한에 액세스하십시오. 팁: 동일한 클라이언트에서 Content Manager 버전 8 및 EIP 버전 8을 관리할 수 있으므로, 클라이언트는 사용 권한 둘 다에 대한 모든 사용 권한을 클라이언트의 각 사이트에 표시합니다.

## 사용 권한

관리 클라이언트는 미리 정의된 사용 권한, 사용 권한 그룹 및 사용 권한 세트를 제공합니다. 사용 권한은 시스템 사용자(예: 시스템 관리자 및 클라이언트 일반 사용자, 권리 또는 권리들)에게 특정 방식으로 특정 오브젝트를 수행할 수 있도록 합니다.

### 사용 권한

EIP는 다중 사용 권한을 제공합니다. 사용 권한은 특정한 방법으로 특정 오브젝트를 수행하는 권한입니다. 예를 들어, 클라이언트 일반 사용자가 있는 경우 ItemAdd 및 ItemDelete 사용 권한을 할당할 수 있고 콘텐츠 서버에서 항목을 추가하거나 항목을 삭제할 수 있는 권한을 제공합니다. 사용 권한을 보려면, 인증 아이콘을 펼친 후 사용 권한을 두 번 누르십시오. 사용 권한을 작성하려면, 사용 권한을 마우스 오른쪽 버튼을 누른 후 신규 작성을 누르십시오.

### 사용 권한 그룹

EIP는 기본 사용 권한 그룹을 제공합니다. 사용 권한 그룹은 관련 사용 권한의 컬렉션입니다. 예를 들어, 시스템 관리자 EIP라는 사용 권한 그룹에는 일반적으로 EIP 시스템 관리와 연관된 다섯 개의 사용 권한이 포함되어 있습니다.

- EIPAdminServer
- EIPAdminEntity
- EIPAdminTextEntity
- EIPAdminTemplate
- EIPAdminInfoMining

사용 권한 그룹에 미리 할당된 사용 권한을 보거나 수정하려면, 인증 아이콘을 펼친 후 사용 권한 그룹 이름을 두 번 누르십시오. 사용 권한 그룹을 작성하려면 사용 권한 그룹을 두 번 누른 후 신규 작성을 누르십시오.

### 사용 권한 세트

EIP는 몇 개의 기본 사용 권한 그룹을 제공합니다. 사용 권한 세트는 사용자 역할을 정의하는 사용 권한의 컬렉션입니다. 예를 들어,

ClientUserCreateAndDelete라는 사용 권한 세트에는 클라이언트 일반 사용자 역할과 연관된 17개의 사용 권한이 포함됩니다. 예를 들어, Delete(항목 삭제), ItemAdd(항목 추가) 등. 클라이언트 사용자에게 대해 사용자 ID를 작성하고 사용 권한 세트 ClientUserCreateAndDelete를 할당하는 경우, 사용자는 콘텐츠 서버에 로그인한 후 사용 권한 세트에 포함된 17개의 사용자 역할 중 무엇이든 수행할 수 있습니다. 사용 권한 세트를 보거나 수정하려는 경우 인증 아이콘을 펼치고 사용 권한 세트를 누른 후 사용 권한 세트 이름을 두 번 누르십시오. 사용 권한 세트를 작성하려는 경우, 사용 권한 세트를 마우스 오른쪽 버튼을 누른 후 신규 작성을 누르십시오.

사용자 및 그룹을 관리할 때는 사용 권한 세트를 사용자 및/또는 사용자 그룹과 연관시킵니다. 사용 권한 세트를 사용자 그룹에 할당하는 경우, 그룹의 전체 사용자는 연관된 사용 권한 세트에 포함된 역할을 모두 수행할 수 있습니다.

---

## 제품 보기와 데이터베이스 전환

엔터프라이즈 솔루션의 부분으로 Content Manager 및 Enterprise Information Portal을 가지고 있는 경우, 하나의 사용자 인터페이스에서 시스템 관리 클라이언트 둘 다에 액세스할 수 있습니다. 이전에는 두 제품이 모두 설치되어 있는 경우 독립된 두 개의 클라이언트를 열어야 했습니다. 하나의 클라이언트 보기에서 다른 클라이언트 보기로의 전환은 양쪽 클라이언트에 적용되는 정보를 편리하게 수정할 수 있도록 하며 두 제품에 신속하게 액세스할 수 있도록 합니다.

로그오프하지 않고 EIP 관리에서 Content Manager 관리로 전환하려면, 주 시스템 관리 창으로 이동하여 왼쪽 분할 영역 위의 풀다운 메뉴를 사용하여 Content Manager를 선택하십시오.

연합 데이터베이스 간에 전환하려면, 클라이언트 창의 오른쪽 분할 영역으로 이동한 후 연합 데이터베이스 아이콘을 두 번 누르십시오.

또한 클라이언트를 종료하고 새 데이터베이스로 로그인하지 않고 다른 데이터베이스를 관리할 수 있습니다. 관리 클라이언트는 cmbds.ini 파일에 나열된 모든 관리 데이터베이스에 대한 아이콘을 표시합니다. 다른 데이터베이스로 전환하려면, 아이콘을 누르십시오. 사용자가 클라이언트에 로그인했을 때 입력한 사용자 ID와 다른 사용자 ID를 새 데이터베이스가 가지고 있는 경우, 다른 사용자 ID를 입력하도록 프롬프트됩니다.

---

## 관리 클라이언트 개선사항 및 확장사항

EIP 버전 8.2는 EIP 관리 클라이언트와 비교하여 그 기능이 크게 향상되었습니다. 향상된 기능에는 다음 사항이 포함됩니다.

### 개선된 마법사와 대화 상자

새 대화 상자가 사용자 관리를 더 용이하게 합니다. 새 마법사는 연합 엔티티와 검색 템플릿의 정의 및 수정을 보다 용이하게 합니다. 사용자는 EIP 버전 7.1에 의해 지원되는 대화 상자를 사용하도록 선택할 수 있습니다.

### 공유 관리 클라이언트

EIP 버전 8.2 및 Content Manager 버전 8.2를 동일한 시스템에 설치하면 이 두 제품이 하나의 관리 클라이언트를 공유합니다. 사용자가 양 제품에 대한 시스템 관리자인 경우, 한 번만 클라이언트에 로그인하면 되며 클라이언트 안에서 두 응용프로그램 사이를 교환할 수 있습니다. 또한 로그아웃하고 다시 로그인하지 않고 관리 데이터베이스간에 교환할 수도 있습니다.

## 도메인 관리자

정의된 도메인에 대해서만 관리 사용 권한을 갖는 도메인 관리자를 작성할 수 있습니다.

## 단일 사인온 및 LDAP 지원

EIP는 이제 Windows Active Directory 및 LDAP를 사용하여 사용자가 복수 콘텐츠 서버에 대한 단일 사인온 액세스를 가질 수 있게 합니다.

---

## 관리 클라이언트를 로컬 관리 데이터베이스에 연결

관리 클라이언트가 설치된 동일한 서버에 관리 데이터베이스를 설치하는 경우, 로컬 클라이언트 및 서버 연결에 필요한 정보는 이미 데이터베이스 연결 정보를 저장한 파일인 cmbds.ini에 저장되어 있습니다. 사후 설치 구성을 수행해선 안 되며 이 절의 단계를 사용하여 즉시 연결할 수 있습니다. **요구사항:** EIP 데이터베이스 설치 유틸리티를 사용하는 추가 로컬 데이터베이스를 작성하려면, 새 데이터베이스에 연결하기 전 필수 정보를 사용하여 cmbds.ini를 수동으로 수정해야 합니다.

1. 시작 --> 프로그램 --> 멀티플랫폼용 **Enterprise Information Portal** --> 관리를 누르십시오.
2. 서버 필드의 드롭다운 목록에서 로컬 데이터베이스를 선택하십시오.
3. 시스템 관리자 사용자 ID 및 암호를 입력한 후 확인을 누르십시오.
4. 시스템 관리 클라이언트가 열립니다. **팁:** EIP First Steps를 사용하는 경우, 샘플 데이터베이스는 클라이언트의 왼쪽 분할 영역에 표시됩니다.

---

## 관리 클라이언트를 원격 관리 데이터베이스에 연결

EIP 관리 클라이언트를 원격 AIX, Windows 또는 Solaris 데이터베이스에 연결하는 데에는 두 가지 방식이 있습니다.

- RMI 서버(자세한 정보는 콘텐츠용 *Information Integrator* 설치 및 설치 참조)를 통해 연결하십시오.
- DB2 Configuration Assistant를 사용하는 데이터베이스를 카탈로그한 후 EIP 서버 구성 유틸리티를 사용하는 서버 연결 매개변수를 정의하여 연결을 정의하십시오. 유틸리티는 정보(예: 데이터베이스 스키마 이름, 별명 이름, 운영 체제 등)를 cmbds.ini라는 파일로 복사합니다. 시스템 관리 클라이언트를 시작하는 경우, 로그를 기록할 수 있는 서버 목록은 cmbds.ini에서 정의된 서버에서 실행됩니다.

**요구사항:** 각 원격 데이터베이스를 별도로 카탈로그해야 합니다. 모든 원격 데이터베이스는 관리 클라이언트에서 연결할 수 있기 전에 cmbds.ini 파일에 나열되어야 합니다.

팁: 노련한 사용자인 경우, 서버 구성 유틸리티 단계를 건너뛰고 텍스트 편집기에서 cmbds.ini를 수정할 수 있습니다. cmbds.ini의 기본 경로는 C:\Program Files\IBM\CMgmt입니다.

중요: 제품을 설치한 사람이 연결하려는 원격 데이터베이스에 대해 데이터베이스 카탈로그값을 이미 구성한 경우, 그 데이터베이스에 대해 DB2 CCA 단계를 수행하지 않아야 합니다. 그러나 설치자가 데이터베이스 카탈로그값을 입력하지 않았거나 추가 원격 데이터베이스를 연결하려는 경우, DB2CCA를 사용하여 추가 데이터베이스에 대한 연결 매개변수로 cmbds.ini 파일을 수정해야 합니다.

## 1단계 - DB2 Configuration Assistant를 사용하는 카탈로그 원격 데이터베이스

DB2 Configuration Assistant(CCA)는 DB2의 원격 EIP 데이터베이스를 카탈로그합니다. DB2CCA를 사용하여 원격 데이터베이스를 카탈로그하려면, 원격 서버 호스트 이름, 데이터베이스 이름 및 데이터베이스 인스턴스 포트 번호를 알아야 하고 원격 데이터베이스에 대한 별명을 정의해야 합니다.

1a - 1f 단계는 데이터베이스 이름, 스키마 이름 및 연결 포트 번호를 찾는 방법에 대해 설명합니다. 이름 및 이름을 구성하는 연결 포트 번호와 관리 클라이언트 및 원격 데이터베이스 간 연결을 구성하는 포트 번호를 알아야 합니다.

1. 원격 데이터베이스 연결 정보를 찾으십시오.

- a. DB2 관리 권한이 있는 사용자 ID로 원격 AIX, Windows 또는 Solaris 서버에 로그인하십시오.
- b. db2 list db directory를 입력하십시오.
- c. 연결하려는 관리 데이터베이스의 이름을 선택하십시오. 여러 인스턴스는 여러 연결 포트 번호를 가질 수 있으므로 데이터베이스가 설치된 db2 인스턴스에 대해 참고하십시오.
- d. db2 connect to <database> user <userID> using <password>를 입력하십시오.
- e. db2 list tables를 입력하고 데이터베이스 스키마 이름(서버 구성 유틸리티가 필요로 하는)에 대한 메모를 작성하십시오.
- f. 원격 관리 데이터베이스와 연관된 연결 포트 번호를 찾으십시오.

Windows의 경우

- 1) 원격 Windows 서버에서 DB2 Control Center를 여십시오.
- 2) 로컬 시스템에 대해 사용 가능한 인스턴스 중 하나를 마우스 오른쪽 버튼으로 누르십시오.
- 3) "통신 설정"을 선택하십시오.

4) TCP/IP 선택사항의 오른쪽에서 "등록 정보" 버튼을 선택하십시오. 포트 번호가 창에 나열됩니다.

AIX 또는 Solaris의 경우

1) cd /usr/etc를 입력하십시오.

2) cat services 입력하십시오.

3) 원격 데이터베이스의 데이터베이스 인스턴스에 대해 연결 포트 번호를 찾을 때까지 서비스 목록을 살펴보십시오. 예를 들어, 데이터베이스가 db2inst1에 설치된 경우, 연결 포트는 50000일 수 있습니다.

2. DB2 Configuration Assistant를 사용하여 원격 데이터베이스를 카탈로그하십시오. 더 자세한 정보는 DB2CCA 도움말 파일을 참조하십시오.

a. 관리 클라이언트가 설치된 Windows 서버에 로그인하십시오. 모든 DB2ADM 사용 권한이 있는 사용자 ID로 로그인해야 합니다.

b. 시작 --> 프로그램 메뉴에서 DB2 Configuration Assistant를 탐색하십시오.

c. DB2 Configuration Assistant 프롬프트를 따라 원격 데이터베이스의 연결을 카탈로그한 후 테스트하십시오.

d. DB2 CCA 연결 테스트가 성공적인 경우, 『2단계 - 서버 구성 유틸리티 사용』의 단계를 따르거나 cmbds.ini 파일을 직접 수정하여 cmbds.ini에 저장된 원격 데이터베이스 연결 매개변수를 정의하십시오.

## 2단계 - 서버 구성 유틸리티 사용

서버 구성 유틸리티는 원격 데이터베이스에 대한 연결 정보(포트 번호, 호스트 이름 등)에 대해 프롬프트 표시한 후 cmbds.ini에 데이터를 저장합니다.

1. 시작 --> 프로그램 --> 멀티플랫폼용 IBM Enterprise Information Portal --> 서버 구성 유틸리티를 누르십시오.

2. 필드에 정보를 입력하십시오(표 4 참조).

표 4. 서버 구성 유틸리티

필드	정보	주
서버	데이터베이스 유형을 Content Manager 또는 EIP로 선택하십시오.	서버는 데이터베이스가 설치된 서버의 이름이 아니라 데이터베이스 유형을 의미합니다. <b>팁:</b> 시스템이 동일한 기계에 Content Manager 및 EIP 관리 클라이언트를 포함하는 경우에만 관리 클라이언트를 사용하여 두 데이터베이스 유형을 관리할 수 있습니다.
서버 이름	연결 중인 데이터베이스의 별명 이름을 입력하십시오. 요구사항: DB2CCA에 정의된 것과 동일한 별명 이름을 사용해야 합니다.	별명은 워크스테이션의 원격 데이터베이스를 식별하는 고유 이름을 제공합니다. 별명 이름은 최대 8자까지입니다. 예를 들어, 원격 데이터베이스 이름이 ICMNLSDB인 경우, 별명은 REMOTE1일 수 있습니다.

표 4. 서버 구성 유틸리티 (계속)

필드	정보	주
스키마 이름	원격 데이터베이스가 작성되었을 때 할당된 스키마 이름을 입력하십시오.	ICMADMIN은 EIP 및 Content Manager 데이터베이스에 대해 기본 스키마 이름입니다.
호스트 이름	원격 데이터베이스가 설치된 컴퓨터의 이름을 입력하십시오.	완전한 호스트 이름을 입력하거나 원격 데이터베이스를 설치한 컴퓨터의 IP 주소를 입력하십시오.
운영 체제	드롭다운 상자에서 운영 체제를 선택하십시오.	AIX, Sun Solaris 또는 Windows를 선택하십시오. OS/390 옵션은 EIP 8.2의 기능이 아닙니다.
포트 번호	원격 데이터베이스에 할당된 포트 번호를 입력하십시오.	50000은 Windows, AIX 및 Solaris에 설치된 EIP 및 Content Manager 데이터베이스에 대한 기본 연결 포트 번호입니다.
원격 데이터베이스 이름	원격 데이터베이스의 이름을 입력하십시오. 대문자를 사용하십시오.	ICMNLSDDB은 EIP 및 Content Manager 데이터베이스에 대해 기본 이름입니다.
노드 이름	원격 EIP 또는 Content Manager 데이터베이스의 노드 이름을 입력하십시오.	노드 이름은 원격 데이터베이스에 할당된 고유 이름으로 원격 데이터베이스에 대해 작성한 별명 이름과 유사합니다. Windows, AIX 또는 Solaris 서버에 설치된 데이터베이스의 노드 이름을 찾으려면, a. db2 명령행 세션을 여십시오. b. db2=> 프롬프트에서 LIST NODE DIRECTORY를 입력하십시오. c. DB2는 원격 서버에 설치 또는 정의된 모든 데이터베이스에 대해 노드 이름 및 기타 데이터를 표시합니다.
단일 사인온을 사용합니다.	데이터베이스 설치 시 단일 사인온이 작동한 경우 클릭하십시오.	기본 설정이 점검되지 않았습니다(사용할 수 없습니다).
보안 옵션	데이터베이스 작성 시 그 옵션이 선택된 경우 클라이언트 인증을 누르십시오.	기본 설정은 서버입니다.

3. 확인을 누르십시오.

### 3단계 - 원격 데이터베이스 연결을 테스트

1. 관리 클라이언트가 설치된 Windows 서버에 로그인하십시오.
2. 시작 --> 프로그램 --> 멀티플랫폼 8.2용 Enterprise Information Portal --> 관리를 누르십시오.
3. 서버 필드의 드롭다운 목록에서 원격 데이터베이스 별명 이름을 선택하십시오. 그 이름은 서버 구성 유틸리티 및 DB2 Configuration Assistant에 정의한 별명과 일치합니다.

4. 원격 데이터베이스와 연관된 사용자 ID 및 암호를 입력하십시오.
5. 확인을 누르십시오. 관리 클라이언트가 열립니다.

---

## 문서 유형 정의

EIP에서는 일부 문서 유형의 보기 프로그램 지원을 제공합니다. 서버에 문서 유형을 정의하는 경우, 원래의 해당 응용프로그램 내에서 문서를 시작할 수 있습니다. 예를 들어, Lotus Word Pro® 문서를 Content Manager OnDemand 서버에 저장하는 경우, 클라이언트 문서 보기 프로그램 대신 Lotus Word Pro에서 확장자가 .lwp인 문서를 시작하도록 EIP를 설정할 수 있습니다.

문서 유형을 정의하려면, 일반적으로 `x:\Program Files\IBM\CMgmt\CMBROOT` 디렉토리의 `cmbcc2mime.ini` 파일을 수정하십시오. 이 파일에는 사용자 조정 MIME 정의를 개발하는 방법에 대한 지시사항이 포함됩니다. 파일은 콘텐츠 클래스를 MIME 유형 스트림으로 변환하므로 클라이언트는 콘텐츠 서버에서 콘텐츠를 읽을 수 있습니다.

**중요사항:** MIME 유형을 기준으로 응용프로그램을 시작할 때는 기본 오브젝트만 표시됩니다. 문서에 있는 마크업은 표시되지 않습니다. 문서에 여러 부분이 있는 경우, 첫 번째 부분만 표시됩니다. 두 파일의 MIME 유형은 일치해야 합니다.

### 서버 MIME 유형 파일(`cmbcc2mime.ini`) 변경

서버 MIME 유형을 추가할 때, 추가 중인 문서 유형이 해당 파일에 작성된 MIME 유형인지 확인하십시오. 자세한 정보는 웹 사이트 `ftp://ftp.isi.edu/in-notes/iana/assignments/media-types`를 참조하십시오.

`cmbcc2mime.ini` 파일에 값을 추가하려면 다음 단계를 완료하십시오.

1. 텍스트 편집기에서 `cmbcc2mime.ini`를 여십시오.
2. 사용자 조정 값에 다음 형식을 사용하십시오.
  - 콘텐츠 클래스는 4096에서 시작됩니다.
  - 등호(=)가 콘텐츠 클래스 값 다음에 옵니다.
  - MIME 유형이 등호 다음에 옵니다. 이 값이 해당 콘텐츠 클래스의 표준 MIME 유형이 아닌 경우, 다음 단계를 따르십시오.
    - a. MIME 유형은 유형 및 하위 유형으로 구성됩니다. 올바른 유형은 응용프로그램, 텍스트, 이미지, 모델, 메시지, 오디오 및 비디오입니다.
    - b. 슬래시(/)가 유형 뒤에 옵니다.
    - c. 하위 유형을 작성하려면 토큰(x-)이 해당 문서에 사용된 토큰 앞에 와야 합니다. 예를 들어, 다음과 같습니다.  
`x-mydocumentclass (4096=application/x-mydocumentclass)`
  - 새 MIME 유형마다 2b 및 2c를 필요한 만큼 반복하십시오.



팁: OnDemand 콘텐츠 서버는 콘텐츠 클래스 숫자 값이 아닌 파일 확장자를 MIME 유형 스트림에 맵핑합니다.



---

## 제 3 장 EIP 관리 클라이언트 기능 사용

이 절에서는 EIP 시스템 관리자가 수행하는 몇 가지 공통 작업을 설명합니다.

---

### 연합 검색 작성

연합 검색은 하나 이상의 콘텐츠 서버를 동시에 검색하는 클라이언트 응용프로그램에서 발행된 조회입니다. EIP는 연합 검색을 위한 검색 템플리트를 작성하는 도구를 제공합니다. 각 콘텐츠 서버가 서로 다르게 정보를 저장하고 구성하기 때문에, 검색 템플리트는 각 서버에 대한 이러한 차이점을 고려해야 합니다. 검색 템플리트는 콘텐츠 서버를 검색하기 위해 연합 엔티티 및 해당 연합 속성을 원래의 속성에 매핑합니다.

연합 검색 작성에는 다음이 포함됩니다.

- EIP 커넥터를 사용하여 콘텐츠 서버에 대한 연결 정의
- 연합 엔티티 작성
  - 연합 엔티티 정의
  - 연합 속성 작성
  - 연합 속성을 원래의 속성에 매핑
  - 매개변수 할당
- 검색 템플리트 작성
  - 검색 템플리트 정의
  - 검색 기준 정의
  - 템플리트 설정 정의
  - 클라이언트 사용자에게 액세스 할당

EIP 버전 8.2에는 연합 엔티티 및 검색 템플리트를 더 쉽게 작성할 수 있도록 두 개의 마법사가 제공됩니다. 연합 엔티티 마법사는 원래의 속성 찾기를 용이하게 하기 위해 필터링할 수 있는 서버 목록을 포함합니다. 또한 연합 속성에 대한 올바른 기본 매개변수를 생성하여 구성의 오류 가능성을 줄입니다. 검색 템플리트 마법사는 검색 기준을 작성하도록 돕습니다. 또한 검색 기준 및 결과 표시 및 활동 방식을 디자인하도록 돕습니다. 심지어 검색 템플리트가 클라이언트 응용프로그램에서 어떻게 표시되는지에 대한 미리보기를 제공합니다. 그리고 EIP 버전 7.1에 대한 검색 템플리트와 연합 엔티티 작성을 위한 대화 상자가 이를 선호하는 사람에게 사용 가능합니다.

모든 마법사, 대화 상자 및 필드가 EIP 온라인 도움말에 문서화됩니다.

## 서버 정의

서버에 연결하여 서버 목록을 수행하려면 먼저 서버를 정의해야 합니다. 서버 아이콘을 오른쪽 마우스 단추로 누르고 신규를 누르면 클라이언트가 EIP에서 지원하는 모든 커넥터를 표시합니다. 서버를 정의하기 전에 커넥터에 대한 다음과 같은 기본 정보를 알고 있어야 합니다.

- 설치 프로그램이 어느 커넥터를 선택했습니까? 설치된 커넥터는 cmbcs.ini 구성 파일에 나열됩니다. Windows 서버에서, 기본 경로는 `x:\Program Files\IBM\CMgmt` 입니다. cmbcs.ini 파일의 위치에 대해서는 AIX 또는 Solaris 시스템 관리자에게 요청하십시오.
- 설치 프로그램이 로컬 및 원격 커넥터 옵션 중 어느 것을 선택했습니까? cmbcs.ini 파일은 로컬 또는 원격 커넥터 유형을 포함합니다.
- 시스템이 RMI용으로 구성되어 있습니까? RMI 서버가 시작되었습니까? 로컬 RMI 서버에서 RMI를 시작하려면 시작 --> 프로그램 --> 멀티플랫폼용 **IBM Enterprise Information Portal 8.2** --> RMI 서버 시작을 선택하십시오. 시스템이 원격 RMI를 사용하는 경우, cmbsvclnt.ini에서 RMI 커넥터가 설치된 원격 서버를 찾으십시오. 자세한 정보는 RMI 서버 시스템 관리자에게 요청하십시오.
- EIP를 설치한 사람이 AS/400용 CM 커넥터를 포함한 경우, frnolint.tbl 네트워크 테이블에는 어떤 정보가 포함되었습니까? AS/400 frnolint.tbl은 %CMBROOT%에 있습니다.
- 관계형 데이터베이스(예: Content Manager 버전 8 및 DB2, DataJoiner 및 Information Catalog)를 포함하는 원격 콘텐츠 서버를 정의하는 경우, 클라이언트를 사용 중인 워크스테이션에서 데이터베이스를 카탈로그하거나 추가해야 합니다.

다음 목록은 서버를 정의할 때 일반적으로 수행하는 단계입니다.

1. 서버를 오른쪽 마우스 단추로 누르고 신규를 선택하십시오.
2. 목록에서 서버를 선택하십시오. 새 서버 창이 나타납니다.
3. 일반 탭의 서버 이름 필드에 서버 이름 및 설명을 입력하십시오. 일부 서버는 데이터베이스 이름만을 입력합니다. 다른 서버는 데이터베이스가 설치된 서버의 완전한 이름을 입력합니다.
4. 필요한 경우, 초기화 매개변수를 지정하십시오. 일부 서버는 연결 문자열 및 구성 문자열과 같은 초기화 매개변수를 필요로 합니다. 다른 서버는 데이터베이스 이름만을 필요로 합니다.
5. 서버 테스트 연결을 누르십시오. 관리 클라이언트를 시작하기 위해 입력한 사용자 ID 및 암호를 사용하여 EIP가 몇몇 서버에 로그인합니다. 서버가 다른 사용자 ID 및 암호를 요구하면, EIP는 정의하는 콘텐츠 서버에 고유한 올바른 사용자 ID 및 암호의 입력을 요구하는 프롬프트를 표시합니다.

**팁:** 미리 정의된 서버 유형 중 하나가 아닌 콘텐츠 서버 유형을 정의할 수도 있지만, 새 서버 유형에 대한 Java 또는 C++ 커넥터 클래스 및 서버 정의 클래스를 제공해야 합니다. 또한 서버 목록을 실행하려면 Java 커넥터도 필요합니다. 콘텐츠 서버 추가에 대한 자세한 내용은 *워크스테이션 응용프로그램 프로그래밍 안내서* 및 온라인 API 참조서를 참조하십시오.

콘텐츠 서버에 대한 구성에 실패하는 경우, 문제점을 해결하는 방법 또는 수신된 오류 메시지에 대한 자세한 내용은 *메시지 및 코드*를 참조하십시오.

연결하려는 서버의 시스템 관리자에게 문의하여 추가 지원을 받을 수도 있습니다.

## 서버 정의에 대한 지침

이 절에서는 초기 서버 정의를 수행하는데 도움이 되는 지침을 제공합니다.

### DB2 (관계형) 데이터베이스에 연결

이 절은 DB2, DataJoiner, JDBC, ODBC, Information Catalog 및 Content Manager 버전 7과 버전 8 서버에 적용됩니다.

- **중요사항:** 서버를 정의하기 전에 각 DB2 데이터베이스를 카탈로그화해야 합니다. DB2 CCA를 사용하여 데이터베이스를 카탈로그화하거나 DB2 명령 프롬프트를 사용할 수 있습니다. 자세한 정보는 DB2 시스템 관리자에게 문의하십시오.
- 일반 탭의 서버 이름 필드에 연결할 데이터베이스의 이름을 입력해야 합니다. 서버 이름을 입력할 때는 대문자를 사용하십시오.
- DB2, DataJoiner, JDBC, ODBC 및 Information Catalog를 정의할 때, 초기화 매개변수 탭을 누르고 연결할 데이터베이스 테이블과 연관된 스키마 이름을 입력하십시오(예: SCHEMA=ICMADMIN).
- Content Manager 버전 7.1 또는 버전 8.2 서버를 정의하는 경우에는 데이터베이스 이름만 입력하면 됩니다. 초기화 매개변수 탭의 기본 설정을 변경하지 마십시오.
- Content Manager 버전 7.1 서버를 정의할 때는 로컬 드라이브의 x:\CMBROOT에 frnolint.tbl이라는 이름의 네트워크 테이블이 있어야 합니다. 이 네트워크 테이블에는 EIP가 Content Manager 버전 7.1 라이브러리 서버를 찾아 로그인하는데 필요한 호스트 이름, 포트 번호 및 서버 유형 정보가 들어 있습니다. 복수의 Content Manager 버전 7.1 서버를 정의할 경우에는, 서버를 정의하기 전에 각 서버가 frnolint.tbl 파일에서 독립된 항목을 가져야 합니다.
- DB2 DataJoiner에 연결하려면, Enterprise Information Portal에 대한 인증 방법이 DB2 Universal Database에 정의된 데이터베이스 인스턴스에 대한 서버로 정의됩니다.
- DataJoiner 2.1에 연결할 경우에는 DataJoiner 웹 사이트에서 바인드 프로그램을 다운로드하여 DataJoiner 데이터베이스를 바인드해야만 DataJoiner 서버를 정의할 수 있습니다.

## 텍스트 검색 서버에 연결

텍스트 검색 서버를 정의하려면, 텍스트 검색 서버와 연관된 Content Manager 버전 7.1 서버를 우선 정의해야 합니다.

드롭 다운 상자의 "연관된 Content Manager 버전 7.1 서버의 이름 선택"에 텍스트 검색 서버 이름을 입력합니다. 이 상자는 연관된 서버 탭에 있습니다.

Content Manager 버전 7.1 서버 및 텍스트 검색 서버가 시작되어 실행 중에 있어야만 EIP를 이들 서버에 연결할 수 있습니다.

## 복수의 AS/400용 Content Manager 서버에 연결

둘 이상의 AS/400 서버를 사용할 경우에는 네트워크 테이블에 추가 서버를 정의해야 합니다. 네트워크 테이블(frnlint.tbl)은 x:\<cmbrroot>에 있습니다. 새 서버의 경우, 서버 이름, 연결 유형(예: TCP/IP), 호스트 이름, 포트 및 서버 유형을 입력하십시오. 첫 번째 서버의 경우, 설치 프로그램은 frnlint.tbl을 작성하기 위해 설치 중 서버, 호스트 이름 및 포트값을 입력합니다.

다음은 frnlint.tbl에 저장되는 정보의 일반적인 예입니다.

```
/* VI/400 Network Table */
SERVER: VI400 REMOTE TCPIP
      HOSTNAME = vi400
      PORT     = 29000
      SERVER_TYPE = FRNLS400
```

## Extended Search 커넥터 구성

Extended Search 서버를 정의하기 위해 입력하는 정보는 다음 두 요소에 따라 달라집니다.

- Extended Server가 설치된 웹 서버 유형 - Domino Web Server, WebSphere, IIS.
- Extended Search가 설치된 웹 서버에 대해 정의된 포트 번호.

Extended Search 커넥터를 정의하려면 다음 단계를 수행하십시오.

1. 일반 탭의 서버 이름 필드에 Extended Search가 설치된 웹 서버의 완전한 호스트 이름을 입력하십시오.
2. 설치 프로그램이 Extended Search 설치 시 웹 서버 포트 번호에 대해 기본 설정을 선택한 경우, 초기화 매개변수 탭의 포트 번호 필드에 80을 입력하십시오.
3. 응용프로그램 ID 필드에 Demo를 입력하십시오. 표시된 대로 이름을 입력하십시오.
4. 암호 필드에 Demo를 입력하십시오.
5. 추가 매개변수 필드에서 다음을 수행하십시오.
  - a. Extended Search가 Domino 웹 서버에 설치되었고 설치 프로그램이 설치 시 웹 서버 및 Extended Search 포트 번호에 대해 기본 포트 번호 설정을 사용했을 을 알고 있는 경우에는 두 세미콜론을 변경하지 마십시오.

- b. 사용자 조정 설정을 사용하여 설치된 Extended Search 서버에 대한 추가 매개변수 필드를 수정하는 방법에 대해서는 아래 절을 참조하십시오.

WebSphere Application Server를 사용해 Extended Search 커넥터를 구성했거나 Extended Search 포트가 6001이 아니거나 Extended Search 서버 이름이 웹 서버 이름과 동일하지 않은 경우, ES servlet, ES 포트 번호 및 ES 서버 이름의 올바른 관련 경로를 찾으려면 커넥터를 구성해야 합니다.

WebSphere를 사용해 Extended Search 커넥터를 구성했거나 Extended Search 포트가 6001이 아니거나 확장 검색 서버 이름이 웹 서버 이름과 동일하지 않은 경우, 다음 단계를 완료하여 구성 파일(예: `desclient.cfg`)을 작성할 수 있습니다.

응용프로그램 또는 샘플이 있는 디렉토리로 그 디렉토리를 설정하십시오. 구성 파일(예: `desclient.cfg`)을 작성하십시오. 이 파일은 Enterprise Information Portal과 함께 제공되지 않습니다.

`desclient.cfg`에 `DESHOSTNAME=(ES host name)`를 추가하십시오.

Domino 응용프로그램 서버를 사용하여 ES를 구성한 경우, `DESREQURI=/servlet/ESAdmin`을 설정하십시오. WebSphere 응용프로그램 서버를 사용하여 ES를 구성한 경우, `servlet` 이 확장 검색 커넥터를 지원하는 HTTP 서버의 디렉토리 경로인 `DESREQURI=/lotuskms/ESAdmin`을 설정하십시오.

응용프로그램 서버가 WebSphere면, `DESREQURI`는 `/servlet/ESAdmin` 대신 `/lotuskms/ESAdmin`이 됩니다.

Thin/Fat 클라이언트에서 ES 소스를 검색하려면 관리 클라이언트의 DES 서버 정의 대화 상자에서

`"DESCFGPATH=<absolute path of desclient.cfg>"`

라는 추가 매개변수를 정의하십시오.

ES 샘플을 실행하려면 명령행 인수에서 `desclient.cfg`의 절대 경로를 패스하십시오.

예 1:

```
TConnectDES es.stl.ibm.com user password
PORT=80;DESAPPID=Demo;DESAPPPW=password;DESCFGPATH
=<absolute path of desclient.cfg>;
```

예 2:

```
java TConnectDES es.stl.ibm.com user password
PORT=80;DESAPPID=Demo;DESAPPPW=password;DESCFGPATH=<desclient.cfg의
desclient.cfg>;
```

## Information Catalog 서버 정의

서버를 정의하기 전에 Information Catalog 서버를 카탈로그해야 합니다. 서버 이름(예 : SAMPLE1)을 서버 이름 필드에 입력하십시오. 초기화 매개변수 탭에서, SCHEMA=<Schema name associated with SAMPLE1>을 입력하십시오.

## OnDemand 서버 정의

OnDemand 서버를 정의하려면 OnDemand 서버 및 라이브러리 서버 디먼이 실행 중이어야 합니다. EIP에서 OnDemand 서버를 정의하기 전에 OnDemand 서버를 Ping 하여 서버 및 디먼이 실행 중인지 확인할 수 있습니다.

일반 탭에서 OnDemand가 설치된 서버의 완전한 호스트 이름을 입력하십시오.

초기화 매개변수 탭에서 OnDemand 서버가 설치되었을 때 할당된 포트 번호를 입력하십시오. OnDemand를 설치한 사람이 OnDemand 설치 시 기본 포트값 0을 선택한 경우에는 포트 번호 필드에 0을 입력하십시오. 설치 프로그램이 다른 포트 번호를 선택했으면 # 부호와 함께 해당 포트 번호를 입력하십시오. 예를 들어, # 5000은 Windows 서버에서 OnDemand용으로 선택된 대체 포트 번호일 수 있습니다.

버전 4 소프트웨어를 실행하는 AS/400 서버에 설치된 OnDemand를 정의하는 경우, 추가 매개변수 필드에 STATECONNECT=#1을 입력해야 합니다.

버전 2.1 소프트웨어를 실행하는 OS/390 서버에 설치된 OnDemand 서버를 정의하는 경우, OnDemand가 OS/390 버전 2.1 서버에 설치될 때 지정된 사용자 조정 포트 번호를 입력하십시오.

OnDemand는 연결 동안 소켓이 활성 상태로 유지되어야 합니다.

## 텍스트 검색 서버 정의

텍스트 검색 서버를 정의하려면 먼저 텍스트 검색 서버와 연관된 Content Manager 버전 7 서버를 정의해야 합니다.

일반 탭의 서버 이름 필드에 텍스트 검색 서버의 이름을 입력하십시오. 연관된 서버 탭에 있는 드롭 다운 상자에서 연관된 Content Manager 버전 7.1 서버를 선택하십시오.

Content Manager 버전 7.1 서버 및 텍스트 검색 서버가 시작되어 실행 중에 있어야만 EIP를 이들 서버에 연결할 수 있습니다.

## Domino.Doc 서버 정의

서버 이름 필드에 Domino.Doc 서버의 서버 이름 및 라이브러리 이름까지의 경로를 입력하십시오. 예: oakley/DominoDoc1/Lib.nsf.



로컬 커넥터를 사용할 경우에는 EIP 클라이언트가 있는 워크스테이션에 Domino Doc Desktop Enabler를 설치해야 합니다. RMI를 사용할 경우에는 RMI 서버에 Domino Doc Desktop Enabler를 설치해야 합니다. Domino Doc Desktop Enabler는 Domino Doc 서버와 동일한 버전에 있어야 합니다.

초기화 매개변수 탭에 있는 두 개의 세미콜론을 수정하지 마십시오.

### OS/390용 ImagePlus 서버 정의

OS/390용 ImagePlus 서버를 정의할 때는 이 서버에 연결하는데 필요한 다음 매개변수를 확보해야 합니다. 아래 목록에는 샘플값이 들어 있습니다.

- FAF 포트 번호: 3061
- FAF 응용프로그램 ID: 01
- FAF 프로토콜: 4000
- FAF IP 주소: 9.67.43.83
- Object Distribution Manager CICS: 4000
- Object Distribution Manager IP 주소: 9.67.43.83
- Object Distribution Manager 포트 번호: 3082
- Object Distribution Manager 터미널 ID: 이 필드는 비워 두십시오
- 추가 매개변수: FAFSITE=CS61

### OS/390용 Content Manager ImagePlus에 추적 사용

OS/390용 Content Manager ImagePlus 서버에 연결할 수 없는 경우 추적으로 문제점을 해결하도록 도울 수 있습니다. OS/390용 Content Manager ImagePlus용 커넥터를 설치한 경우, cmbroot에 있는 eyepapi.ini 파일을 수정하여 OS/390용 ImagePlus에 대해 추적을 사용 가능화할 수 있습니다.

eyepapi.ini 파일에는 다음 행이 들어 있습니다.

```
; Path where the IPFAF files are stored
;   (MUST NOT have a trailing '\')
; -- default is the <ROOT Directory>\
;
IPFAFPath=d:\cmbroot
; Flag for Logging (EYPLmdd.LOG files)
;   -- default is Logging OFF (0)
; -- 0 All Logging OFF
; -- 1 Log files created only error conditions logged
; -- 2 Log files created all conditions logged
;
Logging = 0

;-----
;
; Flag for Logging the FAF Parameters Types created by APIs
;   -- default is Logging OFF (0)
```

```
;      -- 0 Parameter types Not logged
;      -- 1 Log Faf Parameter Types
;
FafTypeLogs = 0
```

### IPFAFPath

로그가 기록되는 디렉토리를 지정합니다. 로그 파일의 이름은 다음과 같습니다.

EYPmdd.LOG

여기서 *mdd*는 로그가 작성된 날짜(월 및 일)입니다.

### Logging

로그 파일이 작성되는 시점을 지정합니다.

- 0      로그하지 않습니다. 기본 설정은 0입니다.
- 1      작성된 로그 파일이 오류 조건만을 포함합니다.
- 2      작성된 로그 파일이 모든 조건을 포함합니다.

### FafTypeLogs

API에 의해 작성된 FAF 매개변수 유형에 대한 로그 기록을 지정합니다.

- 0      매개변수 유형을 로그하지 않습니다. 기본 설정은 0/입니다.
- 1      FAF 매개변수 유형을 로그합니다.

## OnDemand 커넥터에 대한 작업: TCP/IP 조정 및 소켓

알려진 Windows 문제점이 OnDemand 서버에 연결할 때 성능에 영향을 줄 수 있습니다. OnDemand 서버에 대한 반복되는 검색 중, 많은 Windows 소켓이 열리고 닫힙니다. 두 가지 기본 Windows 설정이 EIP와 OnDemand 서버 사이의 대량의 트래픽에 영향을 줄 수 있습니다.

- 응용프로그램이 Windows 소켓을 닫으면 Windows가 240초 동안 소켓 포트를 TIME\_WAIT 상태에 둡니다. 이 시간 동안 해당 포트를 재사용할 수 없습니다.
- Windows는 응용프로그램이 사용할 수 있는 포트 수를 5000으로 제한합니다.

문제점 발생을 방지하려면 Windows 레지스트리 편집기를 사용하여 포트 시간종료 대기 시간 및 번호에 대한 값을 변경하십시오.

- 시간종료 대기 시간의 값을 240초에서 더 낮은 숫자로 변경하십시오. (올바른 범위는 30 - 300초입니다.) 키의 이름은 HKEY\_Local\_Machine\System\CurrentControlSet\services\Tcpip\Parameters\TcpTimedWaitDelay입니다.
- 최대 포트 수를 기본값인 5000에서 더 높은 숫자로 증가시키십시오. (올바른 범위는 5000 - 65534입니다.) 키의 이름은 HKEY\_Local\_Machine\System\CurrentControlSet\services\Tcpip\Parameters\MaxUserPort입니다.

TcpTimedWaitDelay 및 MaxUserPort에 대한 자세한 정보는 Windows 문서를 참조하십시오.

## Extended Search 커넥터에 대한 작업

이 절에서는 EIP 버전 8.2의 Extended Search 커넥터 변경사항을 설명합니다.

DESLOCALE 키에 있는 로케일 값 전달로 사용자가 지정한 로케일이 지원됩니다. 사용자가 직접 ES 커넥터를 호출하는 경우, 명령행에서 이 쌍 값을 전달할 수 있습니다. ES 등록 정보에 대한 **Additional** 인수에서 이 값을 설정할 수 있습니다.

팁: EIP 버전 8.2 제품 상자에는 Extended Search 서버 소프트웨어가 포함되어 있습니다.

---

## 연합 엔티티 작성

컨텐츠 서버에 대한 연결을 정의한 후, 연합 검색 작성의 다음 단계에서는 검색 템플릿에 대한 기본 블록이 되는 연합 엔티티를 작성합니다. 이 절에서는 연합 엔티티 및 연합 엔티티 작성 마법사를 사용하는 방법에 대해 설명합니다.

### 연합 엔티티 이해

대부분의 경우, 클라이언트 응용프로그램 사용자는 서버별로 정보를 검색하려고 하지 않습니다. 대신 한 번의 연합 검색을 수행하려고 합니다. 검색 템플릿은 클라이언트 응용프로그램 사용자가 검색을 한 번의 조회로 묶을 수 있게 합니다. EIP 시스템 관리자는 클라이언트 응용프로그램에서 사용하기 위해 이들 검색 템플릿을 작성할 수 있습니다. 검색 템플릿을 작성하기 전에, 먼저 연합 속성을 컨텐츠 서버의 원래의 속성에 매핑하는 연합 엔티티를 작성해야 합니다.

예를 들어, DB2는 열이 테이블에 저장되는 정보의 속성을 표시하는 테이블에 정보를 저장합니다. Customer\_Demographics라는 테이블은 Name, Pol\_Number, Address, Phone 및 Occupation 같은 열을 포함할 수 있습니다.

한편, Content Manager는 테이블과 열 대신 항목, 항목 유형 및 속성을 사용합니다. DB2에 저장되는 것과 동일한 정보를 엔티티 이름 CustInfo에 저장할 수 있습니다. 엔티티의 속성은 CustName, Acct, HomeAddress, HomePhone 및 Job일 수 있습니다. 두 경우 모두, 동일한 정보가 저장되고 다르게 식별됩니다.

EIP는 컨텐츠 서버가 동일한 정보를 저장하는 여러 가지 방법을 모두 설명하기 위해 가지는 문제점을 해결합니다. 연합 엔티티는 사용자를 위해 이 정보를 추적합니다. 연합 엔티티는 실제로 데이터를 저장하지는 않습니다. 즉, 각 컨텐츠 서버가 데이터를 저장하는 방법에 대한 메타데이터를 저장합니다. 연합 엔티티를 작성하면 모든 해당 속성을, 조회하려는 컨텐츠 서버에 대해 해당하는 원래의 속성에 매핑합니다.

위에 제공된 예의 경우 연합 속성이 Policy\_Name, Policy\_Number, Home\_Address 인 Policy\_Info 및 Job\_Title이라는 연합 엔티티를 작성할 수 있습니다. 그런 다음, 연합 속성을 각각의 해당하는 원래의 속성에 매핑할 수 있습니다.

EIP는 이러한 정보가 들어 있는 서버 목록을 생성할 수 있습니다. 서버 목록에는 이 정보가 들어 있으며, 연합 엔티티 작성 마법사는 콘텐츠 서버에서 필터될 수 있는 서버 목록을 얻게 합니다. 필터링은 사용자가 연합 엔티티 작성 마법사를 사용하는 경우에만 사용 가능합니다. 필터링은 사용자가 수동 (마법사가 아닌) 메소드를 사용하여 연합 엔티티를 작성하는 경우에는 옵션이 아닙니다. 일단 서버 목록을 생성하면 연합 속성을 원래의 속성에 매핑하기 시작할 수 있습니다.

연합 속성을 원래의 속성에 매핑하는 것으로 충분하지 않습니다. 각각의 원래 속성은 여러 등록 정보 또한 가질 수 있습니다. 속성에는 (1) 널 값 가능, (2) 조회 가능, (3) 갱신할 수 없음 및 (4) 텍스트 검색 기능이 있습니다. 사용자가 선택하는 데이터 유형에 따라 데이터 길이, precision, 스케일 및 최소값 및 최대값에 관한 옵션을 가질 수도 있습니다.

이들 등록 정보를 정의하면 연합 속성에 매핑되는 원래의 속성에 의해 이미 정의되는 등록 정보보다 더 제한적으로 만들 수 없습니다. 마법사가 이 기준에 부합되는 기본 등록 정보를 제공합니다. 기본 연합 속성에 대한 기본 등록 정보를 사용자 조정한 후에 마법사가 제안하는 기본 등록 정보로 복귀하려는 경우, 여전히 기본 설정을 선택할 수 있습니다.

요약하면 연합 속성은 복수 콘텐츠 서버에 대해 해당하는 원래의 속성에 매핑합니다. 각 연합 속성의 등록 정보는 원래의 속성에 대한 모든 등록 정보를 포함합니다. 일단 연합 엔티티를 작성하면 이제 여러 콘텐츠 서버에 저장되는 정보에 대한 경로를 가지게 됩니다. 그런 다음, 연합 엔티티를 사용하여 특정 조회에 대한 검색 템플릿을 작성할 수 있습니다.

## 연합 엔티티 작성 마법사 사용

연합 엔티티 작성 마법사는 EIP 버전 8.2의 새 기능입니다. EIP 버전 7.1 이전에서와 같이 동일한 대화 상자를 사용하여 연합 엔티티를 작성할 수 있지만, 마법사로 연합 엔티티를 훨씬 쉽게 작성할 수 있습니다.

마법사를 사용하여 연합 엔티티를 작성하려면 아래에 나열된 단계를 수행하십시오.

1. **연합 엔티티** 이름을 정의하고 연합 엔티티를 설명합니다. 또한 연합 엔티티를 텍스트 검색 가능으로할 것인지 여부를 판별할 수 있습니다.
2. **연합 속성** 이름을 정의하고 연합 속성을 변경합니다.
3. **연합 속성 맵**은 연합 속성을 원래의 속성으로 매핑합니다. 서버 목록을 얻고 매핑하려는 원래의 속성을 선택하며, 나중에 매핑을 수정할 수 있는 도구가 제공됩니다.

4. 등록 정보 정의는 각 연합 속성에 대한 등록 정보를 정의합니다. 등록 정보를 사용자 조정하거나 기본 설정을 승인할 수 있습니다.
5. 연합 엔티티 확인은 연합 엔티티에 대해 선택한 설정을 검토할 수 있습니다. 이전 패널로 리턴하여 설정을 수정할 수 있습니다. 완료되면 완료를 누르십시오.

이들 단계는 마법사 사용에 해당합니다. 마법사 사용 방법에 대한 자세한 정보는 EIP 온라인 도움말을 참조하십시오.

---

## 연합 텍스트 색인 작성

텍스트 검색 엔진은 Content Manager 버전 7.1 콘텐츠 서버와 통합될 수 있으므로 Content Manager에 저장된 텍스트 정보를 자동으로 색인화 및 검색할 수 있습니다. 사용자는 단어 또는 구문을 검색하여 문서를 찾을 수 있습니다. 텍스트 검색 서버는 1바이트 문자와 2바이트 문자 세트를 모두 지원합니다.

텍스트 검색 엔진으로 Content Manager 버전 7.1 서버를 사용하는 경우, 연합 텍스트 색인을 작성할 수 있습니다. 그런 다음, 연합 텍스트 색인을 Content Manager 텍스트 검색 서버에 있는 Content Manager 텍스트 검색 색인에 맵핑합니다.

연합 텍스트 검색 색인을 작성할 경우, 결합된 검색에, 즉 원래의 텍스트 색인과 원래의 속성을 모두 검색할 때 이를 사용할 수 있습니다. 결합된 검색에 연합 텍스트 색인을 사용할 때, 연합 엔티티에 해당 색인을 맵핑하기도 합니다. 그런 다음, 연합 엔티티로 맵핑된 원래의 속성과 해당 연합 속성을 텍스트 검색 서버에 있는 원래의 텍스트 검색 색인에 맵핑합니다.

---

## 검색 템플릿 작성

연합 엔티티를 작성한 후, 검색 템플릿을 작성합니다. 검색 템플릿은 연합 엔티티를 콘텐츠가 저장되는 장소에 대한 맵핑으로 사용함을 기억하십시오. 검색 템플릿을 작성하면 사용자가 검색하려는 것, 검색 결과로 수행하려는 작업 및 템플릿을 사용하는 권한을 갖는 사람을 정의해야 합니다. 각 템플릿에 대해 하나의 연합 엔티티만을 사용할 수 있는 반면, 복수 템플릿에 대해 하나의 연합 엔티티를 사용할 수 있습니다. 또한 검색 기준으로서 연합 엔티티의 속성의 모든 조합에 대해 검색할 수 있습니다. 검색 템플릿을 작성하려면 검색 템플릿 마법사를 사용하여 다음 단계를 완료하십시오.

1. 검색 템플릿 정의
2. 검색 기준 정의
3. 검색 설정 판별
4. 액세스 사용 권한 할당

이들 단계는 검색 템플릿 마법사에 대한 단계에 해당합니다. 검색 템플릿 작성 처리를 완료하는 것에 대한 정보는 EIP 온라인 도움말을 참조하십시오.

## 검색 템플릿 정의

마법사를 시작한 후, 마법사는 사용자에게 검색 템플릿을 정의하도록 프롬프트합니다. 다음을 수행하십시오.

- 검색 템플릿에 대한 이름과 설명을 제공해야 합니다.
- 검색 템플릿에 대한 연합 엔티티를 선택해야 합니다. **제한사항:** 검색 템플릿당 하나의 연합 엔티티만을 사용할 수 있습니다.
- 필요한 경우에는 연합 텍스트 색인을 선택해야 합니다.

**팁:** 연합 텍스트 색인에 대한 체크 박스는 Content Manager 버전 7.1용 텍스트 검색 엔진을 사용 중인 경우에만 적용됩니다. 텍스트 검색을 위해 DB2 TIE를 사용 중인 경우, 이것은 매개변수식 검색이며 매개변수식 검색으로 검색 템플릿에 구성될 수 있습니다.

## 검색 기준 정의

검색 템플릿을 정의한 후, 마법사는 사용자에게 다음을 수행하도록 프롬프트합니다.

1. 속성이나 문서 중 하나인 검색 유형을 선택하십시오. 문서는 연합 텍스트 색인이 이전 단계에서 선택된 경우에만 사용 가능합니다.
2. 검색 기준 이름 지정
3. 연합 속성 선택
4. 사용 가능한 연산자 선택
5. 기본 검색 문자열 제공(문서 검색의 경우에만)

마법사는 선택된 연합 엔티티와 연관된 모든 연합 속성 목록이 있는 드롭 다운 메뉴를 제공합니다. 이들 속성이 검색 템플릿에 대한 검색 기준이 됩니다. 마법사는 또한 사용 가능한 연산자 목록을 제공합니다.

**팁:** 템플릿당 둘 이상의 검색 기준을 작성하거나 템플릿에서 기존 기준을 삭제할 수 있습니다.

## 검색 설정 판별

이 패널에서 기본 검색 설정, 기준 설정 및 표시값 설정을 정의할 수 있습니다. 이들 각 설정에는 사용자가 수정할 수 있는 기본값이 있습니다. 이들 설정을 수정하려면 각 설정에 해당하는 단추를 누르십시오.

기본 설정 창에서 다음을 수행할 수 있습니다.

- 클라이언트 응용프로그램 사용자가 검색 템플릿을 사용하려 할 때 서버가 사용 불가능한 경우, 발생하는 사항을 제어합니다.

- 매개변수식 검색에 대한 와일드카드 문자를 정의합니다.
- 검색 결과를 저장할 폴더의 이름을 지정할 수 있습니다.
- 모든 기준(AND) 또는 임의 기준(OR)을 사용해야 하는지 여부를 선택합니다.

기준 설정 창에서 검색 기준의 순서, 결과 표시열 순서, 열 헤더 및 열 폭을 제어할 수 있습니다.

표시값 설정 창은 검색 결과의 표시값을 정의하는 방법을 제공합니다. 실례로, 주중에 대한 값이 한 서버에서는 월요일이지만, 다른 서버에서는 월인 경우, 서버 둘 다의 검색 결과 표시값으로 월요일을 사용하도록 지정할 수 있습니다.

## 사용 권한 할당

찾을 위치(연합 엔티티를 사용하여), 찾을 내용(검색 기준) 및 결과 표시 방법(설정)을 정의하는 것 이외에, 검색 템플릿에 대한 액세스를 갖는 사람을 정의해야 합니다.

검색 템플릿 마법사의 사용 권한 할당 창은 기존 사용자 또는 사용자 그룹에 템플릿에 대한 액세스를 할당하는 도구를 제공합니다.

사용자에게 검색 템플릿에 대한 액세스 사용 권한을 할당하는 것이, 해당 사용자에게 템플릿에 매핑된 콘텐츠 서버에 대한 액세스를 부여하지는 않습니다. 사용자는 각 개별 콘텐츠 서버에 대한 보안 요구사항을 충족시켜야 합니다. 액세스 제어 목록과 사용자 관리를 사용하여 사용자에게 검색 템플릿에 대한 액세스를 할당하기 전에 해당 사용자가 적합한 사용 권한을 갖는지 확인해야 합니다.

마법사를 사용하여 사용자 또는 사용자 그룹을 검색하면 EIP는 필수 콘텐츠 서버에 대한 적절한 액세스를 갖는 사용자만을 리턴합니다.





---

## 제 4 장 사용자 액세스 관리

사용자는 사용자 ID, 암호 또는 사용 권한 세트 없이 EIP 시스템에 액세스할 수 없습니다. 그러나 사용자를 작성하여 사용자에게 사용 권한을 할당하기 전에, 시스템에 액세스할 수 있는 사용자와 그들의 작업에 필요한 사용 권한을 결정해야 합니다. 오브젝트 삭제 효과를 알지 못하는 사용자에게는 오브젝트 삭제 권한을 부여하지 않고자 할 것입니다. 반면 사용자에게 올바른 사용 권한을 부여하지 않음으로써 사용자가 그들의 작업을 수행하지 못하도록 하고 싶지도 않을 것입니다. 따라서 사용자에게 사용 권한을 할당하기 전에 각 작업에 필요한 작업 유형을 결정해야 합니다.

사용자는 EIP 시스템에서 오브젝트를 작성할 때 다른 사용자가 그 오브젝트에 대해 가지는 액세스를 정의해야 합니다. 오브젝트를 작성하는 사용자는 그 오브젝트에 액세스할 수 있는 사용자와 그 오브젝트에 대해 수행 가능한 조작을 정의해야 합니다. 이러한 정의는 EIP 시스템에 액세스 제어 목록 또는 ACL로서 알려집니다.

---

### 사용자 ID 및 암호 작성

시스템 관리 클라이언트에서 정의한 사용자 ID를 DB2 인증에도 사용하려면 사용자 ID는 DB2 이름 지정 규칙을 따라야 합니다. DB2 이름 지정 규칙은 최상위 관리자 또는 사용자 ID 연결에 사용하려는 사용자 ID에 적용됩니다. 다음 용어를 사용할 수 없습니다.

- USERS
- ADMINS
- GUESTS
- PUBLIC
- LOCAL
- SQL 예약어는 SQL 참조에 나열.

사용자 ID를 다음 문자로 시작할 수 없습니다.

- SQL
- SYS
- IBM

다음 문자를 사용할 수 있습니다.

- A에서 Z 제한사항: 몇몇 운영 체제는 대소문자 구분 사용자 ID 및 암호를 허용합니다. 대소문자 구분을 허용하는지 알고 싶으면 운영 체제 문서를 확인하십시오.
- 0에서 9

- #
- \$

제한사항: 사용자 ID는 30자를 초과할 수 없습니다.

## DB2 관리 권한의 이해

시스템 관리 클라이언트에 로그인할 경우, 데이터베이스 레벨 및 제품 레벨의 두 가지 인증 레벨이 있습니다. 시스템 관리자에는 관리 도메인 기능을 사용하는 경우 최상위 시스템 관리자와 하위 시스템 관리자의 두 가지 분류가 있습니다. 일반적으로 최상위 관리자만이 시스템 관리 클라이언트에 액세스할 수 있습니다.

최상위 관리자는 db2admin에 대한 DB2 사용 권한이 있어야 합니다. 즉 DB2에 대한 모든 관리 사용 권한을 필요로 합니다. 이 사용자 ID는 db2admin 사용 권한으로 운영 체제에서 정의되어야 합니다. 이 운영 체제 ID에 대한 암호는 DB2 연결 및 라이브러리 서버로의 로그인에 사용됩니다. 라이브러리 서버에 정의된 암호는 사용하지 않습니다. Content Manager 사용 권한: 모든 관리 활동을 수행하려면 사용자 ID는 모든 Content Manager 관리 사용 권한("AllPrivs")을 사용하여 라이브러리 서버에 정의됩니다.

하위 시스템 관리자는 DB2 사용 권한을 필요로 하지 않습니다. 하위 시스템 관리자는 라이브러리 서버의 특정 절만을 관리하므로 다음 두 가지 방법 중 하나로 시스템 관리 클라이언트에 로그인합니다.

- 사용자 ID가 운영 체제 사용자 ID인 경우, 운영 체제의 암호는 DB2에 연결하고 라이브러리 서버에 로그인하는 데 사용됩니다.
- 사용자 ID가 운영 체제 사용자 ID가 아닌 경우, cmbfedenv.ini (Enterprise Information Portal의 경우) 또는 cmbicmenv.ini (Content Manager의 경우)에 암호화된 사용자 ID 및 암호는 DB2에 연결하는 데 사용되며 로그인 창에 제공된 사용자 ID 및 암호는 라이브러리 서버에 로그인하는 데 사용됩니다.

라이브러리 서버에 로그인하는 방법에 대한 자세한 정보는 다음 절을 참조하십시오.

하위 시스템 관리자에게도 EIP 사용 권한이 필요합니다. 하위 시스템 관리자는 모든 하위 도메인 관리 활동에 대한 Domain 관리 사용 권한을 필요로 합니다.

## INI 파일을 사용해 DB2에 연결

INI 파일의 각 항목에는 라이브러리 서버와 DB2 연결을 위한 암호화된 사용자 ID 및 암호 한 쌍의 이름이 포함됩니다. 이 암호화된 사용자 ID(연결 사용자 ID로 알려짐) 및 암호는 제품을 설치한 시기에 정의됩니다. 연결 사용자 ID는 시스템 관리자의 사용자 ID와 달라야 합니다. Enterprise Information Portal은 DB2 연결에 cmbfedenv.ini를 사용하고 Content Manager는 cmbicmenv.ini를 사용합니다. 기본 연결 사용자 ID는 ICMCONCT입니다. 설치 시, 라이브러리 서버 및 자원 관리자에 대한 암호는 다음 세

군데에 포함됩니다. cmbicmenv.ini 파일은 사용자 ID 및 암호를 포함하여 라이브러리 서버에 액세스할 수 있습니다. 운영 체제는 라이브러리 서버 및 자원 관리자가 상주하는 데이터베이스에 대한 액세스를 정의합니다. ICMRM.properties 파일은 자원 관리자 사용자 ID 및 암호를 포함합니다.

INI 파일을 사용하는 경우, 즉 사용자 ID가 운영 체제 사용자 ID가 아닌 경우, INI 파일의 사용자 ID 및 연결 사용자 ID는 라이브러리 서버에 존재해야 합니다.

연결 사용자 ID는 라이브러리 서버 및 연결 체제에 정의되어야 합니다. UserDB2Connect 사용 권한을 필요로 합니다. INI 파일의 연결 사용자 ID 및 암호를 변경하려면, 관리 클라이언트 창에서 도구 --> 데이터베이스 ID/암호 변경을 선택하십시오.

## 라이브러리 서버 및 시스템 관리자의 암호를 자원 관리자로 변경

자원 관리자로 암호를 변경해야 하는 경우, 라이브러리 서버의 로그인 암호를 자원 관리자로 변경해야 하고 시스템 관리자의 암호를 자원 관리자로 변경해야 합니다. **중요:** 라이브러리 서버 및 시스템 관리자의 로그인 암호를 자원 관리자로 변경하는 경우 다음 단계를 순서대로 완료하십시오.

1. 시스템 관리 클라이언트에 로그인하십시오.
2. 자원 관리자 트리를 펼치십시오.
3. 이 트리를 수정 및 펼치고자 하는 자원 관리자를 누르십시오.
4. 서버 정의를 누른 후 등록 정보를 선택하십시오. 서버 패널 창이 열립니다.
5. 암호 필드에서 암호를 변경하십시오.
6. 확인을 누르십시오.
7. 펼친(2단계에서) 자원 관리자를 마우스 오른쪽 버튼을 누른 후 등록 정보를 선택하십시오. 자원 관리자 등록 정보 창이 열립니다.
8. 암호 필드에서 암호를 변경한 후 확인을 클릭하십시오.

## 데이터베이스 액세스 암호 변경

데이터베이스 액세스 암호를 변경해야 하는 경우, 데이터베이스 연결 및 ICMRM.properties 파일에 대한 운영 체제 암호를 변경해야 하므로 자원 관리자는 새 암호를 식별할 수 있습니다.

데이터베이스 연결에 대해 운영 체제 암호를 변경하려면, 다음 단계를 수행하십시오.

1. 운영 체제에 따라 사용자 및 암호 유틸리티를 탐색하십시오.
2. ICMRM을 누르십시오.
3. 암호 설정을 선택하십시오.
4. 새 암호를 입력하십시오.

ICMRM.properties 파일을 변경하려면 다음 단계를 완료하십시오.

1. ICMRM.properties파일을 여십시오. 기본 위치는 <--

X가 Content Manager에 설치된 드라이브의 위치인-->  
X:\WebSphere\AppServer\installedApps\icrmr.ear\icrmr.war\WEB-INF\classes\com\ibm\mm\icrmr\ICMRM.properties(여기서 X는 Content Manager를 설치한 위치)입니다.

2. DBPassword를 변경하여 운영 체제 암호와 일치시키십시오.

3. ICMRM.properties 파일을 저장하십시오.

데이터베이스 암호를 변경한 다음 데이터베이스는 재시작하십시오. 또는 가만히 두면 두 개 또는 세 개의 오류를 발행한 뒤, 스스로 재설정됩니다.

시스템 관리 클라이언트의 자원 관리자에 대한 암호 및 기타 필드를 변경하는 방법에 대한 자세한 지시사항은, 시스템 관리 온라인 도움말을 참조하십시오.

---

## LDAP에서 사용자 가져오기

LDAP는 시스템별로 사용자 ID 및 암호 관리를 지원하는 것이 아니라 엔터프라이즈 레벨에서 사용자 ID 및 암호 관리를 지원합니다. EIP는 세 개의 LDAP 기술을 사용합니다. IBM Directory(이전 버전에서 IBM SecureWay Directory로 알려짐), Windows 2000 Active Directory, 그리고 Lotus Domino Directory Notes Address Book(NAB). 사용자 암호는 LDAP 서버에 상주합니다. 사용자가 또는 Enterprise Information Portal에 로그인 하는 경우, 사용자 ID 및 암호가 인증되어 사용자 ID의 고유 사용 권한이 EIP 데이터베이스에 있는 사용자 프로파일에 의해 점검됩니다. LDAP는 EIP 설치 시 작동했을 수도 있습니다. LDAP가 설치 시 작동하지 않은 경우에는 LDAP를 항상 활성화할 수 있습니다.

LDAP를 사용하려면, 시작 --> 프로그램 --> 멀티플랫폼용 EIP --> LDAP 사용자 ID 가져오기 스케줄러를 선택한 후 시스템 관리 클라이언트를 시작하십시오. LDAP 구성 창을 시작하십시오(도구 --> LDAP 구성). LDAP 사용자 가져오기 사용 및 인증 체크박스를 선택한 후 서버 페이지에서 LDAP 서버 정보를 제공하십시오.

LDAP가 사용 가능해지면, 새 사용자 창에서 LDAP 단추를 눌러 사용자를 가져올 수 있습니다. 이렇게 하면 LDAP 서버에서 사용자를 선택적으로 EIP로 가져올 수 있습니다. 대신 LDAP 사용자 ID 가져오기 스케줄러 유틸리티를 사용하는 그룹에서 사용자를 가져올 수 있습니다. 로그인 시, 라이브러리 서버는 자동으로 LDAP 서버에 연결하여 사용자를 인증합니다. LDAP 서버가 어떤 이유로든 사용자 암호를 확인할 수 없는 경우, 인증에 실패합니다.

기본 시스템 관리 클라이언트 창으로 가서 도구 -> LDAP 구성을 눌러 LDAP 서버 구성을 수정할 수 있습니다. EIP에서 메뉴 시작에서 LDAP 사용자 레지스트리 가져오기 유틸리티로 가서 현재 LDAP 서버도 변경할 수 있습니다. LDAP 계획에 대한 정보

는 *Content Management* 시스템 계획 및 설치를 참조하십시오. 시스템 관리 창에 LDAP 서버 정보를 구성하는 방법에 대한 정보는 시스템 관리 클라이언트 온라인 도움말을 참조하십시오.

LDAP의 계획에 대한 정보는 *Content Management* 시스템 계획 및 설치를 참조하십시오. LDAP를 구현하는 방법에 대해서는 시스템 관리 클라이언트 온라인 도움말을 참조하십시오.

---

## 사용 권한 소개

관리 클라이언트는 사용 권한 그룹, 사용 권한 세트 및 개별 사용 권한을 제공합니다. 결합된 Content Manager/EIP 시스템을 관리하는 경우, 사용 권한은 클라이언트의 두 부분에 공통입니다. 클라이언트에 빌드된 사용 권한은 시스템을 우선화할 수 있도록 합니다.

### 사용 권한 그룹

사용 권한 그룹은 시스템 관리자가 새 사용 권한 세트를 작성하거나 사용 권한 세트 대화 상자에서 역할을 사용하도록 돕는 것을 목적으로 하는 사용자 작업의 컬렉션입니다.

### 사용 권한 세트

사용 권한 세트는 사용자 역할의 컬렉션입니다.

### 사용 권한

사용 권한은 사용자 조치를 표시합니다. 예를 들어, 다음과 같습니다.

**예제 1 - 사용 권한:** 사용 권한 ClientScan 및 ClientImport를 일반적으로 Content Manager로 문서를 스캔 및 가져오는 데 클라이언트를 사용하는 사용자 그룹으로 할당하려 할 수 있습니다. 그 작업을 일반적으로 수행하는 사용자가 여럿 있는 경우, 하나의 사용자 ID(예: user1)를 작성합니다. 그러면 사용자 ID User1과 사용 권한 ClientScan 및 ClientImport를 연관시킵니다. 그리고 User1을 Group1이라는 그룹으로 할당합니다. 일반 사용자가 클라이언트에 로그인하고 Content Manager에 액세스하기 위해 user1을 입력한 경우, 그 사용자는 문서를 스캔하고 가져올 수만 있습니다.

**예제 2 - 사용 권한 그룹:** 모든 일반 클라이언트 작업에 액세스하기 위해 사용 권한을 필요로 하는 경험있는 일반 사용자 그룹이 있습니다. 사용자 ID(예: user2)를 작성합니다. 그리고 user2를 그룹(예: group2)로 할당합니다. 다음 ClientTaskAll이라는 사용 권한 그룹을 user1과 연관합니다. 일반 사용자가 클라이언트에 로그인하고 Content Manager에 액세스하기 위해 user2를 입력한 경우, 그 사용자는 ClientTaskAll이라는 사용 권한 그룹에 포함된 모든 작업을 수행할 수 있습니다.

**예제 3 - 사용 권한 세트:** 읽기 전용 액세스를 필요로 하는 사용자 그룹이 있습니다. 사용자 ID(dp: user3)를 작성합니다. 그리고 나서 user3을 그룹(예: group3)으로 할당합니다. ClientUserReadOnly라는 사용 권한 세트를 user3과 연관시킵니다. 일반 사용

자가 클라이언트에 로그인하고 Content Manager에 액세스하기 위해 user3을 입력한 경우, 그 사용자는 ClientUserReadOnly라는 사용 권한 세트에 포함된 작업만을 수행할 수 있습니다.

## 사용 권한 세트 작성

EIP 시스템 구성을 계획할 때는 시스템을 액세스하는 사용자와 그들이 시스템에 있는 오브젝트에 대해 가지는 액세스 종류를 결정해야 합니다. EIP 시스템은 사용 권한을 통해 액세스를 정의합니다.

사용 권한은 특정한 방법으로 특정 오브젝트에 액세스할 수 있는 권한을 부여합니다. 사용 권한에는 시스템에 오브젝트를 작성하고 저장된 오브젝트를 삭제 및 선택하는 권한 등이 포함됩니다. 사용자에게 할당되는 사용 권한 그룹이 사용 권한 세트입니다.

액세스 관리에 있어 첫 번째 작업은 사용자의 사용 권한 세트를 작성하는 것입니다. 사용 권한 세트는 사용자가 수행할 수 있는 작업 또는 조치를 식별합니다. 사용 권한 세트는 사용 권한을 결합하고 특정 유형의 사용자를 위해 조정됩니다. 예를 들어, 하나의 시스템 관리자 그룹이 문서 라우팅 서버를 관리하고 다른 시스템 관리자 그룹이 도메인을 관리하게 하고자 할 수도 있습니다. 시스템 관리자가 로그인하면, EIP는 시스템 관리자의 사용 권한 세트를 점검합니다.

시스템 관리 클라이언트에는 그룹으로 묶어 하나의 사용 권한 세트로 만들 수 있는 미리 정의된 여러 사용 권한이 있습니다. 작성한 사용 권한 세트는 개별 사용자에게 할당할 수 있습니다. 사용자 그룹에는 사용 권한 세트를 할당할 수 없습니다.

## 사용 권한 그룹 작성

사용 권한 그룹은 사용자 그룹과 유사합니다. 사용 권한 세트에 포함시킬 사용 권한을 쉽게 찾기 위해, 유사한 사용 권한을 모아 사용 권한 그룹을 작성합니다. 예를 들어, 시스템에 있는 거의 모든 사용자에게 할당할 두 개의 사용 권한이 있는 경우, 사용 권한 세트를 작성할 때마다 가지고 있는 여러 사용 권한을 검색하는 대신 이 두 기본 사용 권한을 BasicPrivs라고 하는 사용 권한 그룹으로 묶을 수 있습니다.

## 사용자에게 사용 권한 세트 할당

시스템 관리 클라이언트에는 그룹으로 묶어 하나의 사용 권한 세트로 만들 수 있는 미리 정의된 여러 사용 권한이 있습니다. 작성한 사용 권한 세트는 개별 사용자에게 할당할 수 있습니다. 사용자 그룹에는 사용 권한 세트를 할당할 수 없습니다.

사용 권한 이름을 작성할 수는 있으나 사용 권한 자체를 작성할 수는 없습니다. 시스템 관리 클라이언트에 정의되어 있지 않는 사용 권한을 작성하려면 시스템 프로그래머와 함께 작업하여 사용 권한을 작성해야 합니다.

EIP에서 제공하는 사용 권한 세트를 사용하거나 고유의 사용 권한 세트를 작성할 수 있습니다.

## 사용자 ID에 Grant Privilege Set 할당

사용자가 자신이 갖고 있는 것보다 더 많은 사용 권한을 가지는 사용자 ID를 작성하지 못하게 하기 위해, EIP가 사용 권한 세트 부여의 사용을 구현했습니다. 사용자 ID에 부여 사용 권한 세트를 할당하면 해당 사용자는 부여된 사용 권한 한계 내에서 사용자 ID를 작성할 수 있습니다. 예를 들어, 사용자 ID에 도메인을 관리하는 시스템 관리 사용 권한 세트를 부여할 수 있습니다. 그러나 이 사용자 ID가 사용자를 작성하는 사용 권한은 가지지 않도록 하고자 할 수도 있습니다. 그러므로 사용 권한 세트 부여 필드에서 이 사용자 ID를 작성할 때 "Noprivs"를 선택합니다. 결과적으로 이 사용자 ID는 도메인을 관리할 수는 있으나 그 도메인의 사용자를 작성할 수는 없습니다.

## 자원 관리자에 사용자 할당

사용자가 특정 자원 관리자에 액세스할 수 있도록 하기 위해, 사용자가 액세스 권한을 가지는 도메인에 자원 관리자를 할당합니다. 도메인에 자원 관리자를 할당하는 것에 대한 자세한 정보는 45 페이지의 『도메인에 자원 관리자 할당』을 참조하십시오.

## 컬렉션에 사용자 할당

사용자가 컬렉션에 액세스할 수 있도록 하기 위해, 사용자가 액세스 권한을 가지는 도메인에 자원 관리자 상의 컬렉션을 할당합니다. 도메인에 컬렉션을 할당하는 것에 대한 자세한 정보는 45 페이지의 『도메인에 컬렉션 할당』을 참조하십시오.

---

## 사용자 그룹 작성

중중 작업 설명이 동일한 사용자는 동일하거나 유사한 작업을 가지므로 시스템에 있는 오브젝트에 대해 동일한 액세스를 가집니다. 공통된 액세스를 필요로 하는 사용자들을 사용자 그룹으로 묶을 수 있습니다. 사용자 그룹을 중첩시킬 수 없습니다.

사용자 그룹은 작업이 유사한 개별 사용자들을 단지 편의를 위해 그룹화한 것입니다. 사용자 그룹에는 사용 권한 세트를 할당하지 않습니다. 사용자 그룹의 각 사용자는 고유의 사용 권한 세트를 가집니다. 사용자 그룹을 사용하면 시스템에 있는 오브젝트에 대한 액세스 제어 목록을 보다 쉽게 작성할 수 있습니다.

도메인을 사용 가능하게한 경우, 그룹에 사용자 ID를 할당하기 전에 해당 사용자 그룹이 특정 도메인 또는 PUBLIC 도메인에 있는지 점검하십시오(도메인에 대한 자세한 정보는 43 페이지의 『도메인 관리』를 참조하십시오). 사용자 ID를 포함시킬 도메인에 해당 사용자 그룹이 있는지 확인하십시오. 임의 도메인에 대한 사용자 ID를 작성하려면 사용자 그룹 창에서 새 사용자를 누르십시오. 작성한 사용자를 사용자 그룹에 추가한 다음, 그 사용자가 해당 도메인에 있는지 확인하십시오.

---

## 액세스 제어 목록 작성

사용자에게 그들의 작업을 완료하는 데 필요한 사용 권한을 제공하십시오. 오브젝트는 개별적으로 특정 액세스 제어 이슈를 가집니다.

액세스 제어 목록(ACL)은 하나 이상의 개별 사용자 ID 또는 사용자 그룹과 그와 연관된 사용 권한으로 구성된 목록입니다. ACL을 사용하여 EIP 시스템에서 오브젝트에 대한 사용자 액세스를 제어합니다. 액세스 제어 목록과 연관시킬 수 있는 오브젝트로는 사용자가 저장하는 데이터 오브젝트, 항목 유형, 항목 유형 서브세트, 작업 목록 및 처리가 있습니다.

사용 권한 세트는 시스템을 사용하는 개별 사용자의 최대 능력을 정의하는 반면, ACL은 오브젝트에 대한 개별 사용자의 액세스를 제한합니다. 한 사용자의 사용 권한 세트에 정의되지 않은 사용 권한을 가진 ACL은 해당 사용자에게 그 사용 권한을 부여하지 않습니다. 그 사용 권한을 가지는 사용자만이 오브젝트에 대해 그 사용 권한을 사용할 수 있습니다. ACL은 사용자 액세스를 제한하며 추가 액세스를 부여하지 않습니다. 액세스 제어 목록은 시스템을 관리할 때 또하나의 보안 레벨을 제공합니다.

### 액세스 제어 목록에 사용 권한 세트 할당

액세스 제어 목록(ACL)에 추가하는 각 사용자 ID는 그와 연관된 사용 권한 세트를 필요로 합니다. 사용자 ID 및 사용 권한 세트는 오브젝트에 대해 액세스를 가지는 사용자와 그들이 해당 오브젝트에 대해 가지는 액세스 종류를 정의합니다.

ACL에 없는 사용자는 어떠한 오브젝트에도 액세스할 수 없습니다. ACL에 사용자나 사용자 그룹을 추가하려면 해당 ACL에 대한 사용자 ID 및 사용 권한 세트를 선택하여 추가를 눌러야 합니다. 정의된 각 ACL의 사용자 ID 및 그룹이 액세스 제어 목록 창에 나열됩니다. 사용자 ID 및 그룹을 추가하거나 제거하여 이 테이블을 수정할 수 있습니다. ACL 작성 또는 수정에 대한 자세한 정보는 시스템 관리 클라이언트 온라인 도움말을 참조하십시오.

---

## 도메인 작성

도메인은 한 사람 이상의 시스템 관리자가 관리하는 관리 데이터베이스의 섹션입니다. 도메인은 사용자 ID, 사용자 그룹, 액세스 제어 목록, 사용 권한 세트, 자원 관리자 및 SMS 컬렉션으로 구성됩니다. 사용자들은 도메인을 볼 수 없으므로, 도메인의 이름은 도메인을 관리하는 시스템 관리자들에게만 의미가 있습니다. 사용자는 자신이 관리 데이터베이스의 부분으로 제한되었음을 알지 못하며, 이는 해당 도메인에 있는 항목에 대해서만 알고 있음을 의미합니다.

도메인은 관리 데이터베이스의 하위 절에 대한 관리와 사용자 액세스를 제한합니다. 관리 데이터베이스에 대해 전체 사용 권한을 가지는 시스템 관리자는 다른 시스템 관리자에게 제한된 관리 사용 권한을 위임할 수 있습니다. 전체 사용 권한을 가지는 시스템



관리자인 최상위 관리자는 관리 데이터베이스에 액세스할 수 있는 반면, 제한된 사용 권한을 가지는 시스템 관리자인 하위 관리자는 관리 데이터베이스의 절에만 액세스할 수 있습니다.

도메인은 액세스 제어 목록(ACL)에 대한 하위 관리자의 액세스를 제한합니다. 하위 관리자가 사용자 ID 및 사용자 그룹을 추가 또는 삭제하는 데 사용할 수 있는 ACL은 최상위 관리자만이 작성할 수 있습니다. 하위 관리자는 ACL을 작성, 갱신 또는 삭제할 수 없습니다.

하위 관리자는 그들의 도메인에 해당하는 최상위 관리자의 담당 업무를 분담할 수도 있습니다. 도메인을 작성하고 그러한 도메인을 관리하는 시스템 관리자를 할당함으로써, 최상위 관리자는 전체 시스템에 집중하여 하위 관리자가 그들 도메인 고유의 사용자 및 작업을 관리하도록 시스템을 효율적으로 관리하는 동안 하위 작업을 위임할 수 있습니다.

도메인을 사용 가능하게 하기 전에 다음 조건을 고려하십시오.

- 도메인을 사용 불가능하게 하거나 해제할 수 없습니다.
- 자원 관리자, 콜렉션, 사용자 ID 및 사용자 그룹은 한 번에 오직 하나의 도메인에만 있을 수 있습니다.
- 사용 권한 세트 및 액세스 제어 목록은 한 번에 둘 이상의 도메인에 존재할 수 있습니다.
- PUBLIC(공유) 도메인을 제외하고는, 도메인이 겹쳐지지 않습니다.
- 수퍼 관리 도메인에 작성된 오브젝트는 시스템이 생성했건 아니면 사용자가 작성했건 이동될 수 없습니다.

도메인을 사용하려면, 파일 메뉴로 이동하고, 도구 → 관리 도메인을 선택한 후 관리 도메인 사용을 선택하십시오. 도메인에 대해 시스템 관리 클라이언트를 재시작해서 적용해야 합니다. 도메인에 대한 관리 데이터베이스 구성 방법에 대한 특정 지시사항에 대해서는 시스템 관리 클라이언트 온라인 도움말을 참조하십시오.

## 도메인 관리

사용 권한 세트에 따라, 전체 관리 데이터베이스나 특정 도메인 중 하나를 관리합니다. 관리 데이터베이스에 대한 전체 액세스를 갖는 관리자는 최상위 관리자입니다. 하위 관리자는 특정 도메인의 오브젝트에 대한 전체 액세스를 갖습니다.

각 유형의 시스템 관리자는 그들의 도메인에서 사용자 및 콜렉션을 포함하여 오브젝트를 작성, 검색, 갱신 및 삭제할 수 있는 능력을 가집니다. 하위 관리자는 자신의 도메인 및 목록에서만 오브젝트를 보고 검색하거나 PUBLIC(공유) 도메인에서 검색할 수 있습니다.

## 도메인 액세스

하위 관리자는 임의 오브젝트의 도메인을 변경할 수 없습니다. 그러나 하위 관리자 자신의 도메인 및 목록의 콘텐츠를 액세스하거나 PUBLIC(공유) 도메인의 오브젝트를 검색할 수 있습니다.

최상위 관리자는 관리 데이터베이스에서 모든 도메인에 액세스할 수 있습니다. 최상위 관리자는 오브젝트를 작성하여 도메인에 할당할 수 있습니다. 사용 권한 세트 및 ACL과 같은 일부 오브젝트는 최상위 관리자만이 작성할 수 있으므로, 최상위 관리자가 이러한 오브젝트를 작성하여 하위 관리자가 사용하게 할 수 있습니다.

하위 관리자는 단지 자신의 도메인에 있는 임의 오브젝트를 작성, 검색, 갱신 및 삭제(CRUD)할 수 있습니다.

## 도메인에 사용자 할당

사용자 ID를 작성할 때는 사용자 ID를 도메인에 할당할 것인지 또는 기본 도메인에 둘 것인지 선택하게 됩니다. 나중에 사용자 등록 정보를 통해 사용자 ID의 도메인을 변경할 수 있습니다.

사용자 ID는 한 번에 하나의 도메인에 대해서만 액세스를 가집니다. PUBLIC(공유) 도메인에 사용자를 추가할 수 없습니다.

최상위 관리자만이 도메인을 작성하여 해당 도메인에 사용자를 할당할 수 있는 권한을 가집니다. 도메인은 둘 이상의 하위 관리자를 가질 수 있으나, 최상위 관리자만이 사용 권한 세트 안에서 시스템 관리 사용 권한을 부여하여 하위 관리자를 정의할 수 있습니다. 새 사용자 또는 사용자 등록 정보 창의 **사용 권한 세트 부여** 필드에 하위 관리자가 도메인에서 가지는 관리 사용 권한이 표시됩니다.

## 도메인에 사용자 그룹 할당

사용자 그룹을 임의 도메인에 할당하면 해당 사용자 그룹의 각 사용자 ID에 대해 지정된 도메인이 변경됩니다. 사용자 ID는 한 번에 하나의 도메인에 대해서만 액세스를 가집니다. 따라서 할당하는 그룹에 포함된 사용자 ID도 새 도메인으로 이동합니다.

사용자 그룹 이름은 한 번에 하나의 도메인에만 있을 수 없습니다. PUBLIC(공유) 도메인에 사용자 그룹을 할당할 수 있습니다.

## 도메인에 사용 권한 세트 할당

도메인에 사용자 ID를 추가할 때는 연관된 사용 권한 세트도 포함시켜야 합니다. 연관된 사용 권한 세트를 포함시키지 않으면 사용자가 작업을 수행할 수 없습니다. 임의 사용자에게 사용 가능하도록 사용 권한 세트를 저장하는 최적 위치는 PUBLIC(공유) 도메인입니다.

## 도메인에 자원 관리자 할당

특정 도메인에 특정 자원 관리자를 할당함으로써 해당 자원 관리자에 대한 사용자 액세스를 제한할 수 있습니다. 관리 데이터베이스가 액세스하는 새 자원 관리자를 정의할 때 도메인을 선택하는 옵션을 갖습니다.

모든 자원 관리자에 대한 기본값은 PUBLIC입니다. 자원 관리자로서의 액세스를 제한하려면 자원 관리자를 도메인으로 할당해야 합니다. 자원 관리자를 할당할 수 있는 도메인이 없는 경우에는 자원 관리자를 정의한 다음 필요한 도메인을 작성할 수 있습니다. 적절한 도메인을 정의했으면 자원 관리자 등록 정보를 열어 해당 도메인을 선택하십시오.

## 도메인에 컬렉션 할당

자원 관리자에 있는 특정 컬렉션을 특정 도메인에 할당함으로써 해당 컬렉션에 대한 사용자 액세스를 제한할 수 있습니다. 자원 관리자가 PUBLIC 도메인에 있으면 정의된 다른 도메인에 컬렉션을 할당할 수 있습니다. 그러나 자원 관리자가 이미 특정 도메인에 정의되어 있으면 해당 컬렉션을 다른 도메인에 할당할 수 없으며 PUBLIC 도메인으로도 할당이 불가능합니다.

사용자는 자원 관리자에 액세스할 수 있어야만 거기에 있는 컬렉션에 액세스할 수 있으므로, 자원 관리자에 대한 액세스를 제한하려면 거기에 있는 컬렉션에 대해 동일한 제한사항을 부과해야 합니다.

## 하나의 도메인에서 다른 도메인으로 사용자 이동

하나의 도메인에서 특정 사용자를 제거하고 다른 도메인에 추가하는 이유를 찾을 수 있습니다. 사용자가 속한 사용자 그룹을 기억하는 방법으로 사용자 조정 창에서 설명 필드 사용을 고려하십시오. 이것은 작업을 약간 더 용이하게 합니다.

**중요사항:** 이 작업은 매우 시간이 많이 걸리며 제대로 수행하지 않으면 시스템을 액세스할 때 문제점을 야기할 수 있습니다. 사용자의 도메인을 변경하려면 최상위 관리자가어야 합니다.

다음 단계를 주의하여 따르십시오.

1. 사용자가 속하는 모든 그룹을 찾으십시오.
2. 사용자가 속하는 모든 그룹에 대해, 이 그룹을 PUBLIC 도메인으로 이동시키거나 모든 그룹에서 사용자를 제거하십시오.
3. 목표 도메인으로 이동시킨 각 자원 관리자에 대한 모든 컬렉션이 뒤에 오도록 사용자와 연관된 자원 관리자를 PUBLIC 도메인으로 이동시키십시오.
4. 목표 도메인의 사용자와 연관된 모든 사용 권한 세트가 아직 목표 도메인에 없는 경우 작성하고 이동시키지 마십시오.

5. 이 사용자와 연관된 모든 액세스 제어 목록이 아직 목표 도메인에 없는 경우 작성하고 이동시키지 마십시오.
6. 사용자의 등록 정보를 열고 사용자의 도메인을 변경하여 목표 도메인으로 사용자를 이동시키십시오.
7. 선택적: PUBLIC 도메인에서 목표 도메인으로 단계 45 페이지의 1, 2 및 45 페이지의 3에서 이동시킨 그룹과 자원 관리자를 이동시킬 수 있습니다. 그렇지 않으면, 그룹 및 자원 관리자는 PUBLIC 도메인에 위치하여 다른 도메인에 있는 사용자의 공유를 허용할 필요가 있습니다.

**메모:** 사용자는 PUBLIC 도메인에 있을 수 없습니다. 사용자는 공유될 수 없습니다.

## 하나의 도메인에서 다른 도메인으로 사용자 그룹 이동

**중요사항:** 이 작업을 제대로 수행하지 않으면 시스템을 액세스할 때 문제점을 야기할 수 있습니다. 사용자 그룹의 도메인을 변경하려면 최상위 관리자이어야 합니다.

사용자 그룹을 다른 도메인으로 이동시키려면 다음 단계를 따르십시오.

- 사용자 그룹이 비어 있으면, 현재 도메인에서 그룹을 삭제하고 그룹을 다시 작성하여 목표 도메인으로 할당하십시오.
- 사용자 그룹이 비어 있지 않으면, 다음 단계를 따르십시오.
  1. 이 그룹에 속하는 모든 사용자를 찾으십시오.
  2. 모든 사용자를 삭제할 현재 도메인에서 그룹을 삭제하십시오.
  3. 그룹을 다시 작성하고 목표 도메인에 할당하십시오.
  4. 모든 사용자를 새로 작성된 이 그룹에 추가하십시오.

## 하나의 도메인에서 다른 도메인으로 자원 관리자 이동

자원 관리자의 도메인을 변경하려면 최상위 관리자여야 합니다.

자원 관리자를 다른 도메인으로 이동시키려면, 다음 단계를 따르십시오.

- 자원 관리자에 컬렉션이 없는 경우, 등록 정보를 열고 도메인을 목표 도메인으로 변경하여 목표 도메인으로 자원 관리자를 이동시키십시오.
- 자원 관리자에 컬렉션이 있는 경우, 다음 단계를 따르십시오.
  1. 자원 관리자를 PUBLIC 도메인으로 이동시키십시오.
  2. 등록 정보를 열고 목표 도메인을 선택하여 컬렉션을 목표 도메인으로 이동시키십시오.
  3. 등록 정보를 열고 목표 도메인을 선택하여 자원 관리자를 목표 도메인으로 이동시키십시오.

## 하나의 도메인에서 다른 도메인으로 컬렉션 이동

컬렉션의 도메인을 변경하려면 최상위 관리자이어야 합니다.

한 도메인에서 다른 도메인으로 컬렉션을 이동시키려면 다음 단계를 따르십시오.

1. 컬렉션이 속하는 자원 관리자를 찾으십시오.
2. 연관된 자원 관리자를 PUBLIC 도메인으로 이동시키십시오.
3. 등록 정보를 열고 목표 도메인을 선택하여 컬렉션을 목표 도메인으로 이동시키십시오.
4. 등록 정보를 열고 목표 도메인을 선택하여 자원 관리자를 목표 도메인으로 이동시키십시오.

## 하나의 도메인에서 다른 도메인으로 사용 권한 세트 이동

사용 권한 세트는 여러 도메인에 있을 수 있으므로, 사용 권한 세트를 이동시키지 않고 목표 도메인에 추가할 수 있습니다.

## 하나의 도메인에서 다른 도메인으로 액세스 제어 목록 이동

액세스 제어 목록은 여러 도메인에 있을 수 있으므로, 사용 권한 세트를 이동시키지 않고 목표 도메인에 추가할 수 있습니다.



---

## 제 5 장 Information Mining 관리

이 절은 Information Mining의 정의와 비즈니스 환경에서 Information Mining을 사용하는 방법으로 시작합니다. 이는 Information Structuring Tool의 Information Mining First Steps인 개념 Information Mining에 대한 절에 따르며 성능 조정에 대한 주석을 사용하여 완료합니다.

---

### Information Mining 개념

Enterprise Information Portal Information Mining은 회사가 정보 추출 및 분석의 많은 측면을 자동화하여 낮은 가격으로 사용자에게 관련 정보에 쉽게 액세스할 수 있도록 돕는 핵심 기술입니다.

Information Mining의 첫 번째 시도는 컴퓨터가 사용할 구조화되지 않은 텍스트의 정보에 쉽게 액세스하는 것입니다. 제한되지 않은 자연어로 표현된 실제 지식만의 완전한 해석은 여전히 현재 기술을 사용하여 접근하기 어렵습니다. 그러나 패턴 인식 기술과 발견적 학습을 적용하는 도구가 임의의 텍스트에서 유용한 정보를 추출할 수 있습니다. 추출된 정보의 범위는 문서에 언급되는 이름, 단체 또는 장소와 같이 소위 중요 단어를 식별하는 것부터 문서의 전체 요약까지 해당합니다.

또한 단일 문서로부터 단순히 단편적인 정보를 추출하는 것 이상이 Information Mining에 있습니다. 커다란 문서 컬렉션을 다룰 경우, Information Mining이 재생됩니다. "마이닝" 비유는 단일 문서에서 집대성된 정보의 식별과 추출, 그리고 메타데이터로 이 정보를 저장하는 것과 같은 지식 발견 프로세스뿐 아니라 관심있는 현상, 패턴 또는 경향을 발견하는 문서 컬렉션 사이의 이들 기능의 분배 분석 프로세스 모두에 사용됩니다.

### Enterprise Information Portal Information Mining 서비스

Enterprise Information Portal Information Mining 서비스는 개별 문서나 문서 컬렉션과 관련된 정보의 작성 및 유지보수를 위한 하부 구조를 제공합니다. 이 정보를 메타데이터라고 합니다. 문서 콘텐츠를 특징짓고 메타데이터로 저장되는 정보의 예에는 다음이 있습니다.

- 제목
- 추상 또는 요약
- 이름, 용어 또는 표현식
- 문서가 속하는 카테고리

문서를 메타데이터와 연관시키는 전통적인 메타데이터 저장과 Information Mining을 차별화하는 것은, 메타데이터가 명시적으로 사용 가능하지 않은 경우에도 Information Mining이 메타데이터를 자동으로 작성하는 기능을 제공한다는 점입니다. 마이닝과 검색 알고리즘이 메타데이터를 사용하여 검색 처리를 안내하거나 메타데이터에 대해 통계 모델을 실행하여 컬렉션의 개별 문서를 볼 때 명백하지 않을 수 있는 문서 사이의 관심있는 관계를 찾아서 거대한 문서 컬렉션의 관련 정보에 액세스할 수 있습니다.

마이닝과 검색 조작이 원래의 문서 콘텐츠를 고려하는 대신 잘 정의된 메타데이터의 세트에 대해 작업하기 때문에, 이들 처리의 속도는 메타데이터를 *Information Mining 데이터스토어*라는 전용 스토어에 보관하여 크게 증가될 수 있습니다. 응용프로그램이 단일 저장소로부터 메타데이터를 폐치하므로, 이것은 메타데이터에 대한 액세스를 크게 가속시키며 임의로 원격 콘텐츠 서버로 되돌아갈 필요가 없습니다. 메타데이터가 검색 및 탐색에 결정적이고 종종 검색 결과를 좁히는데 사용되므로, 콘텐츠 서버로 되돌아가지 않고도 많은 단계를 수행할 수 있습니다.

메타데이터 스토어 사용의 또다른 장점은 문서와 연관된 데이터를 실제 문서와는 별도로 보관하는 것입니다. 명백하게, 웹에 있는 외부 문서 같이 읽기 액세스만이 사용 가능한 문서의 경우, 콘텐츠와 메타데이터를 동일한 저장소에 보관하는 것이 전혀 선택사항이 아닙니다.

Enterprise Information Portal Information Mining 서비스는 메타데이터의 자동 작성을 위해 다음 메커니즘을 제공합니다.

- **카테고리** - 이것은 사용자 조정 분류법을 바탕으로 문서에 하나 이상의 카테고리를 할당합니다. 카테고리 구성요소에는 *Information Structuring Tool*이라는 분류법의 작성 및 유지보수를 위한 그래픽 사용자 인터페이스를 제공하는 응용프로그램이 포함됩니다.
- **요약** - 문서의 가장 중요한 문장을 추출하여 사용자가 전체 문서를 읽을지 여부를 결정하는 데 도움을 줄 수 있습니다. 사용자는 원하는 요약의 길이를 지정할 수 있으며 이 방법으로 추출된 메타데이터의 복잡도와 문서에 있는 정보량 사이의 균형에 직접적으로 영향을 줄 수 있습니다.
- **문서가 작성된 언어를 판별하는 언어 식별**. 이는 기타 Information Mining 서비스를 적용하기 전의 유용한 사전 처리 단계입니다.
- **텍스트 문서에서 자동으로 이름, 용어 및 표현식과 같은 중요한 어휘 항목을 인식하는 정보 추출**.
- **유사한 그룹이나 클러스터로 문서 세트를 나누는 clustering**. Cluster는 문서 컬렉션에서 자동으로 파생됩니다.

## Information Mining 서비스의 구성요소

Information Mining 응용프로그램은 일반적으로 다음 태스크를 포함합니다.

1. 찾아보고 탐색될 수 있는 그러한 방식으로 데이터 구성



2. 다른 데이터 원본 액세스

3. 예상이나 경향에 대한 필수 데이터를 거르기 위해 고급 검색 조작 사용

그림 1에서는 이들 Information Mining 작업을 설명합니다.

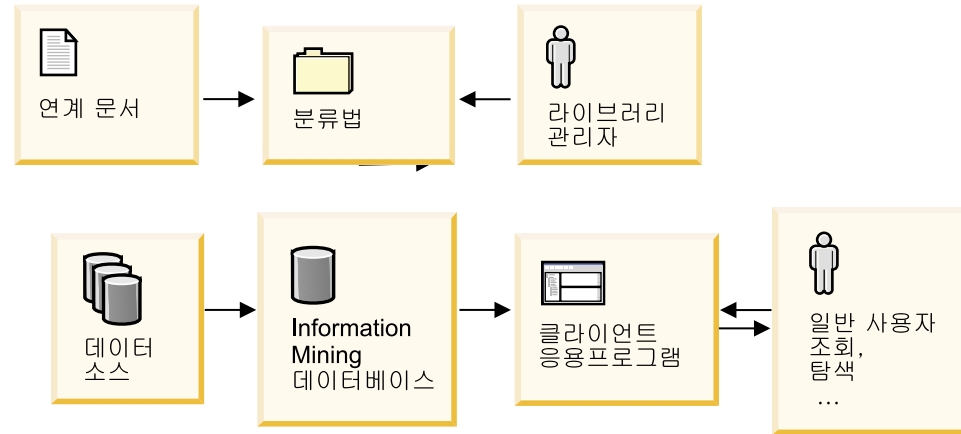


그림 1. Information Mining 태스크

Information Mining 기능을 활용하려면, 문서를 찾아보고 탐색할 수 있도록 구성해야 합니다. 작업은 보통 라이브러리 관리자나 지식 엔지니어에 의해 수행됩니다. 라이브러리 관리자는 문서에 있는 데이터의 계층 구조 주제 특성이 그 이상으로 트리 표시되도록 분류법을 정의하기 위해 Information Structuring Tool를 사용합니다. IST는 분류법 작성 및 유지보수를 위한 그래픽 사용자 인터페이스를 갖는 응용프로그램입니다. 카테고리는 연계되며, 분류법이 안정되면, 다른 데이터 원본을 액세스할 수 있습니다.

문서는 콘텐츠 서버 또는 웹 크롤러 서비스(JavaBeans에서만 사용 가능)를 사용하는 웹에서 가져오고 카테고리로 할당될 수 있습니다.

텍스트 분석 및 메타데이터 작성 기능은 비주얼이 아닌 JavaBeans 및 Java Service API의 레벨에서 프로그래밍 인터페이스로 사용될 수 있습니다.

Information Mining JavaBeans는 빠른 응용프로그램 개발을 위한 소프트웨어 구성요소이며 JavaBeans를 따릅니다. Java Service API에는 응용프로그램 작성을 위한 개별적인 빌딩 블록으로 전체 Information Mining 기능이 들어 있습니다. JavaBeans 기반 샘플 및 샘플 JSP가 응용프로그램 작성을 지원하기 위해 제공됩니다.

문서의 콘텐츠 식별은 문서를 처리하는 모든 Information Mining 조작에 대한 전제조건입니다. 작업의 하위 단계에는 다음이 포함됩니다.

1. 문서 코드 페이지 식별

2. 처리되어야 할 텍스트 절 식별, 즉 마크업 정보 또는 2진 데이터(예: 이미지) 무시

콘텐츠 서버의 문서가 제멋대로 구조화될 수 있기 때문에, Information Mining 서비스는 문서 형식에서 관련 텍스트 부분을 식별하고 추출하는 특정 모듈을 작성하는 방법을

제공합니다. 자주 사용되는 문서 형식을 다루는 기본 모듈도 Information Mining 서비스에서 사용할 수 있습니다. 지원되는 형식 목록에 대해서는 137 페이지의 제 9 장 『문서 형식』을 참조하고 샘플 기본 모듈의 사용 방법에 대한 세부사항에 대해서는 응용프로그램 프로그래밍 안내서를 참조하십시오.

선택된 각 문서에 대해 메타데이터를 작성하는 것은 문서의 콘텐츠를 처리하고 지식 자원(예: 사전 또는 주기 프로파일)을 기반으로 통계 방법이나 발견적 학습을 적용하는 것을 포함합니다.

Information Mining API는 다음 조작을 지원합니다.

- 요약
- 카테고리
- 언어 ID
- 정보 추출
- 클러스터링

모든 문서에 대해 작성된 메타데이터는 Information Mining 데이터 스토어에 저장됩니다.

일단 데이터 스토어가 채워지면, 문서 선택을 수행하기 위한 하나 이상의 옵션, 즉 이 데이터 스토어에 있는 정보를 기반으로 한 문서 선택이 가능하게 됩니다. 고급 검색 조작은 카테고리화 텍스트 조회를 결합하므로, 특정 카테고리에 속하는 문서에 대한 검색을 제한합니다.

## 비즈니스 환경에서 Information Mining 사용

모든 Information Mining 기술의 구현을 지원하는 조직의 하부 구조는 대개 최소한 다음 규칙으로 구성됩니다.

- 일반적으로 IT에 대한 시스템 관리자, 반드시 Information Mining으로 제한되지 않음
- 응용프로그램 프로그래머
- 라이브러리 관리자 또는 지식 엔지니어
- Information Mining 응용프로그램으로 작업하는 사람(일반 사용자)

응용프로그램의 본질에 따라, 위에서 나열된 역할에 추가하여 다음과 같은 좀 더 특정한 역할을 찾을 수도 있습니다.

- 웹 디자이너
- 제작자 또는 자문역

53 페이지의 그림 2에서는 Information Mining 역할 및 조치를 설명합니다.

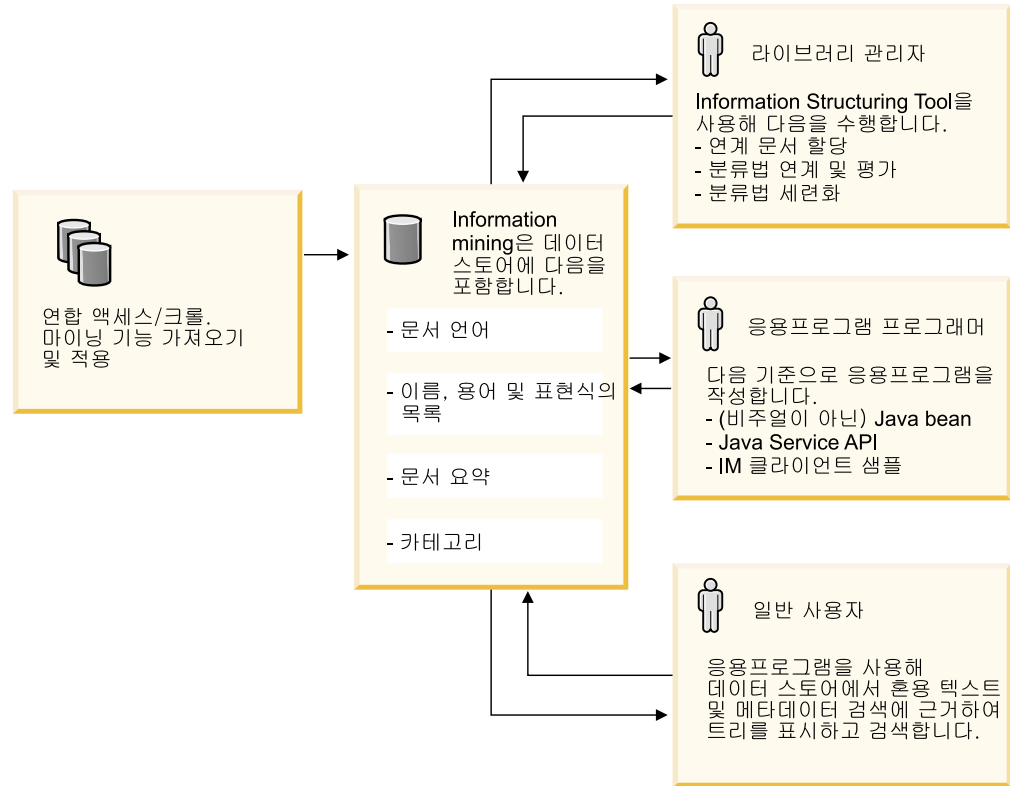


그림 2. Information Mining 역할 및 조치

시스템 관리자는 하드웨어 및 소프트웨어 환경을 설정하고 필수 자원(예: 파일 시스템 공간 및 액세스 권한)을 유지보수합니다. 시스템 관리자는 필수 Enterprise Information Portal 구성요소를 설치하고 콘텐츠 원본 및 Enterprise Information Portal 관리 응용 프로그램을 구성하므로 다양한 저장소에 대한 액세스가 사용 가능합니다. 또한 시스템 관리자는 데이터베이스 레벨에서 Information Mining 메타데이터 저장을 관리합니다.

응용프로그램 프로그래머는 JavaBeans 또는 Service API를 사용하는 응용프로그램을 작성합니다. JavaBeans 샘플에 대해서는 Windows용 응용프로그램 프로그래밍 안내서를 참조하고 Service API의 설명에 대해서는 응용프로그램 프로그래밍 안내서를 참조하십시오. 응용프로그램 프로그래머는 웹 디자이너의 도움으로 이를 수행할 수 있습니다.

라이브러리 관리자 또는 지식 엔지니어는 마이닝 및 검색에 사용될 문서 컬렉션 및 자원의 설정 및 유지보수를 담당합니다. 라이브러리 관리자는 Enterprise Information Portal 관리 응용프로그램을 사용하여 메타데이터 맵핑 및 검색 템플리트를 작성하고 Information Structuring Tool (71 페이지의 『분류법 구축』 참조)을 사용하여 카탈로그 및 분류법을 정의합니다. 데이터 원본이나 웹의 문서로 Information Mining 메타데이터 스토어를 채우고, 응용프로그램 프로그래머가 작성한 응용프로그램을 사용하는 것도 보통 라이브러리 관리자의 책임입니다.

일반 사용자는 응용프로그램 프로그래머가 작성한 응용프로그램으로 작업하여 라이브러리 관리자나 지식 엔지니어가 작성하고 유지보수하는 자원을 기반으로 Information Mining과 검색 조작을 수행합니다. 일반 사용자와 라이브러리 관리자 간의 작업 분배에 따라, 일반 사용자가 콘텐츠 서버에서 문서 선택 및 Information Mining 메타데이터 스토어 채우기에 참여할 수 있습니다.

## Information Mining 사용 예제

Electro Corp.은 대량 시장을 위한 전자 장치를 생산하는 회사입니다. 이 회사의 포트폴리오에는 다양한 개별 구성을 갖는 최소한 다섯 가지 제품이 포함됩니다.

영업 부서는 특정 장치 적용 분야의 고객 선호도에 대한 정보를 갖고 있습니다. 이들 사용 프로파일에서는 고객이 제품을 사용하는 방법과 다양한 구성 옵션에 대해 설명합니다. 각 프로파일은 특정 마케팅, 전시회 및 관계 관리 전략에 관련됩니다.

서비스 부서는 장치를 구성하는 부품, 부품 조립 방법, 해당 부품의 공급자 및 개별 부품의 유지보수성 및 신뢰성에 대한 정보를 갖고 있습니다.

계약 관리 부서는 대리점 및 하청업체에 대한 정보를 보관합니다. 또한 특정 유형의 계약에 대해 올바른 계약 조건과 관련된 법률 문서에 액세스합니다.

최근에 특정 고객의 매출이 현저하게 하락되었음을 알게 되었습니다. 다양한 전자 장치 적용 방법 사이의 경쟁이 변화되었고 고객 기대가 이러한 새로운 기술적 진보에 부합되도록 바뀌었습니다.

이러한 변화에 대응하고 자사의 이익에 맞게 이용하기 위해, Electro Corp.은 회사를 다시 업무를 진행시킬 전략을 개발하는 대책팀을 설정합니다.

First Steps는 입안자가 신속하게 정보에 정통한 결론을 내릴 수 있도록 이러한 관련 정보에 액세스하는 IT 하부 구조를 설계하는 것입니다.

이런 종류의 정보의 예는 다음과 같습니다.

- 경쟁사 제품, 기능, 가격 및 고객 수용성에 대한 데이터
- 고객의 관점에서 Electro Corps. 생산 라인의 강점과 약점에 대한 지식
- 이러한 제품이 일반적으로 사용되는 분야의 경향과 전망

이 정보는 서로 다른 하드웨어 및 소프트웨어 플랫폼, 구성 레벨(예: 계층구조, 색인화 또는 텍스트 파일) 및 문서 유형(예: 데이터베이스 레코드 또는 HTML)을 갖는 이기종의 데이터 소스에 상주합니다.

55 페이지의 그림 3에서는 Information Mining 사용을 설명합니다.

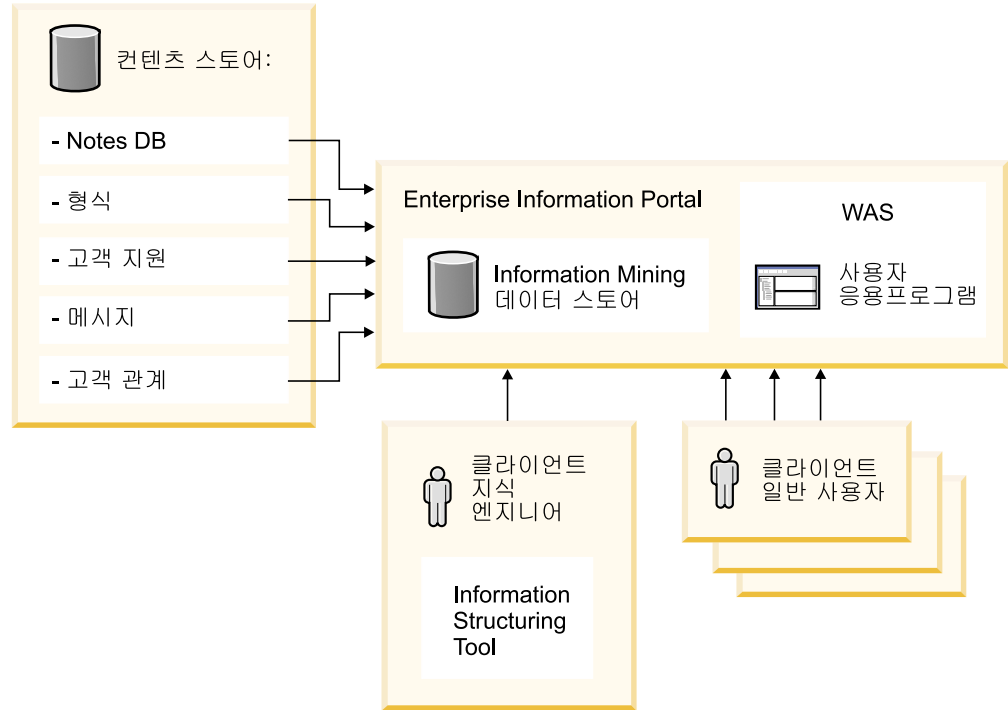


그림 3. Information Mining 예

대책팀은 입안자의 클라이언트 기계에 상주하는 thin 클라이언트(브라우저)에서 이러한 모든 다른 소스에 액세스하는 포털을 작성하기로 결정합니다. 대책팀은 Enterprise Information Portal을 선택하는데 그 이유는 비즈니스 응용프로그램 영역에서 표준이 희귀할 때 작성된 Legacy 데이터 원본에 사용자 지정 커넥터의 작성을 허용하고 그러한 포털을 작성하도록 모든 빌딩 블록을 제공하기 때문입니다.

다음 단계가 포함됩니다.

1. 하드웨어 및 소프트웨어 하부 구조 설정
2. 데이터 소스에 대한 액세스 방법 정의, 필요한 연결 설정 및 관련 데이터에 대한 맵핑 작성
3. 이 데이터를 입안자가 찾아보고 탐색할 수 있는 방법으로 데이터 구성
4. 일반 사용자 응용프로그램 작성

3단계는 Information Mining이 활약하는 단계입니다. 하부 구조가 사용 가능하고 실행 중이며, 데이터 소스가 식별되고, 대응하는 연결이 설정되고 관련 맵핑이 작성되면, 하나의 위치에서 필수 데이터에 액세스할 수 있고 연합 검색을 사용하여 이 데이터의 서브세트를 정의할 수 있습니다. 다음 질문은 특정 예측이나 경향에 대한 필수 데이터를 필터링하는 방법과 이 데이터를 전략적 계획 처리에 맞게 구성하는 방법에 대한 것입니다.

대책팀은 전략적 계획 정보의 유지보수, 구성 및 갱신을 담당하는 지식 엔지니어를 정의합니다. 다양한 데이터 소스에 상주하는 거대한 문서 세트로부터 관련 정보를 추출하

기 위해, 지식 엔지니어는 과거의 전략 계획에 참여한 스태프를 인터뷰하여 처리가 상호작용하는 방법과 좋은 실례인 것으로 증명된 것을 배우고 고객 관계 및 지원 데이터베이스를 검색합니다.

Enterprise Information Portal의 검색 성능을 사용하여 이 데이터베이스의 문서에서 고객 이름, 주소 또는 장치 등록 정보별로 쉽게 액세스할 수 있습니다. 그러나 사용법 프로파일을 판별하는 데 필요한 정보는 텍스트 내에 감춰져 있어서 이 정보를 얻는 유일한 방법은 Enterprise Information Portal Information Mining 서비스를 사용하여 문서 콘텐츠를 분석하는 것입니다.

이들 서비스가 제공하는 정보의 유용한 유형은 카테고리라는 문서 콘텐츠의 주제 특성입니다(예: 이 문서는 *PDA*에 관한 것임). Information Mining 카테고리 서비스는 문서 콘텐츠를 분석하여 문서를 카테고리에 할당합니다. 카테고리는 분류법이라는 주제에 관한 계층 구조로 구조화됩니다. 명시적으로 사용 가능하고 자동으로 작성된 메타데이터가 모두 가속된 액세스 및 검색을 도와주는 카탈로그라는 Information Mining 서비스가 유지보수하는 저장소에 상주합니다.

지식 엔지니어는 Information Structuring Tool을 사용하여 고객이 장치를 사용하는 방법을 예시하는 카탈로그를 정의합니다.

56 페이지의 그림 4에서는 카탈로그를 표시합니다.

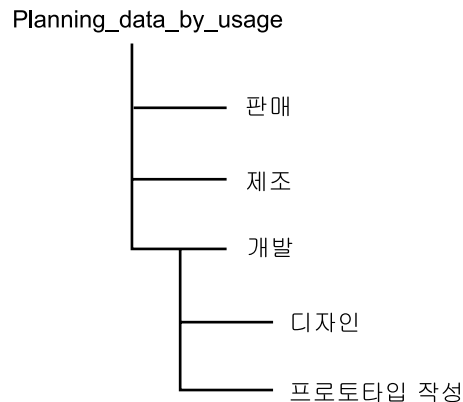


그림 4. 카탈로그 예

지식 엔지니어는 매출 및 지원 데이터베이스에서 모든 고객 및 고객이 장치를 사용하는 방법에 대해 연합 검색을 수행하여 분류법의 각 카테고리를 대표하는 연계 문서 세트를 식별할 수 있습니다. 이것은 관련 데이터를 보다 상세하게 조사할 때 새로운 구분이 나타날 수 있기 때문에 분류법의 재구성을 가져올 수 있습니다.

일단 분류법이 안정되고 각 카테고리에 충분한 수의 연계 문서가 있는 경우, 지식 엔지니어는 Information Structuring Tool을 사용하여 분류법을 연계합니다. 연계는 카테고리 서비스를 사용하여 문서를 카테고리에 할당하는 데 사용될 수 있는 카테고리 모델을 작성합니다.

반면 IT 부서의 프로그래머는 Enterprise Information Portal thin 클라이언트와 JavaBeans 또는 Java Service API를 사용하여 전략 계획자를 위한 일반 사용자 응용 프로그램을 작성합니다.

이 응용프로그램은 전략 계획에 사용하기 위해 사용자 조정된 많은 검색 템플리트로 구성됩니다. 이 템플리트를 사용하여 계획자는 혼합된 다양한 백엔드에서 검색된 문서를 사용하여 카탈로그를 채울 수 있습니다. 카탈로그를 채울 때, Information Mining 서비스가 자동으로 문서에 카테고리를 할당합니다. 새 사용 프로파일이 식별되는 경우, 지식 엔지니어는 입안자가 연계 자료로 식별한 새 문서를 사용하여 적절하게 분류법을 재구성합니다. 카탈로그를 재연계하고 새 결과가 입안자에게 전달됩니다.

위의 예에서는 고객 기대 및 수요를 반영하도록 Information Mining 기능을 결합하여 Electro Corp. 같은 회사가 시장의 변화에 대응하고 경쟁력을 유지할 수 있음을 보여줍니다.

## 지원되는 언어 및 형식

Enterprise Information Portal Information Mining 서비스는 다음 언어를 지원합니다 (표 5참조).

표 5. 지원되는 언어

언어	언어 ID	정보 추출	요약	카테고리	클러스터링
영어	x	x	x	x	x
독어	x		x	x	x
불어	x		x	x	x
덴마크어	x				
핀란드어	x				
이탈리아어	x		x	x	x
노르웨이어	x				
포르투갈어	x		x	x	x
스페인어	x		x	x	x
스웨덴어	x				
한국어	x		x	x	x
일본어	x	x	x	x	x
중국어(대만 및 중국)	x		x	x	x

지원되는 문서 형식 목록에 대해서는 137 페이지의 제 9 장 『문서 형식』을 참조하십시오.

## 개념

소비되는 정보량은 계속 증가합니다. 대부분의 조직은 잠재적으로 큰 값의 정보가 들어 있는 더 커지고 증가하는 수의 온라인 문서의 수를 갖습니다. 예를 들어, 고객 피드백 데이터, 아주 경쟁적인 시장에서 필수적인 전략 정보 또는 새로 변화하는 비즈니스 기회에 대한 안목을 제공하는 정보 등이 있습니다. Information Mining 서비스는 많은 온라인 문서를 처리하는 응용프로그램으로 사용되도록 디자인되었습니다.

## 시스템 아키텍처

그림 5에서는 Information Mining 시스템 아키텍처를 설명합니다.

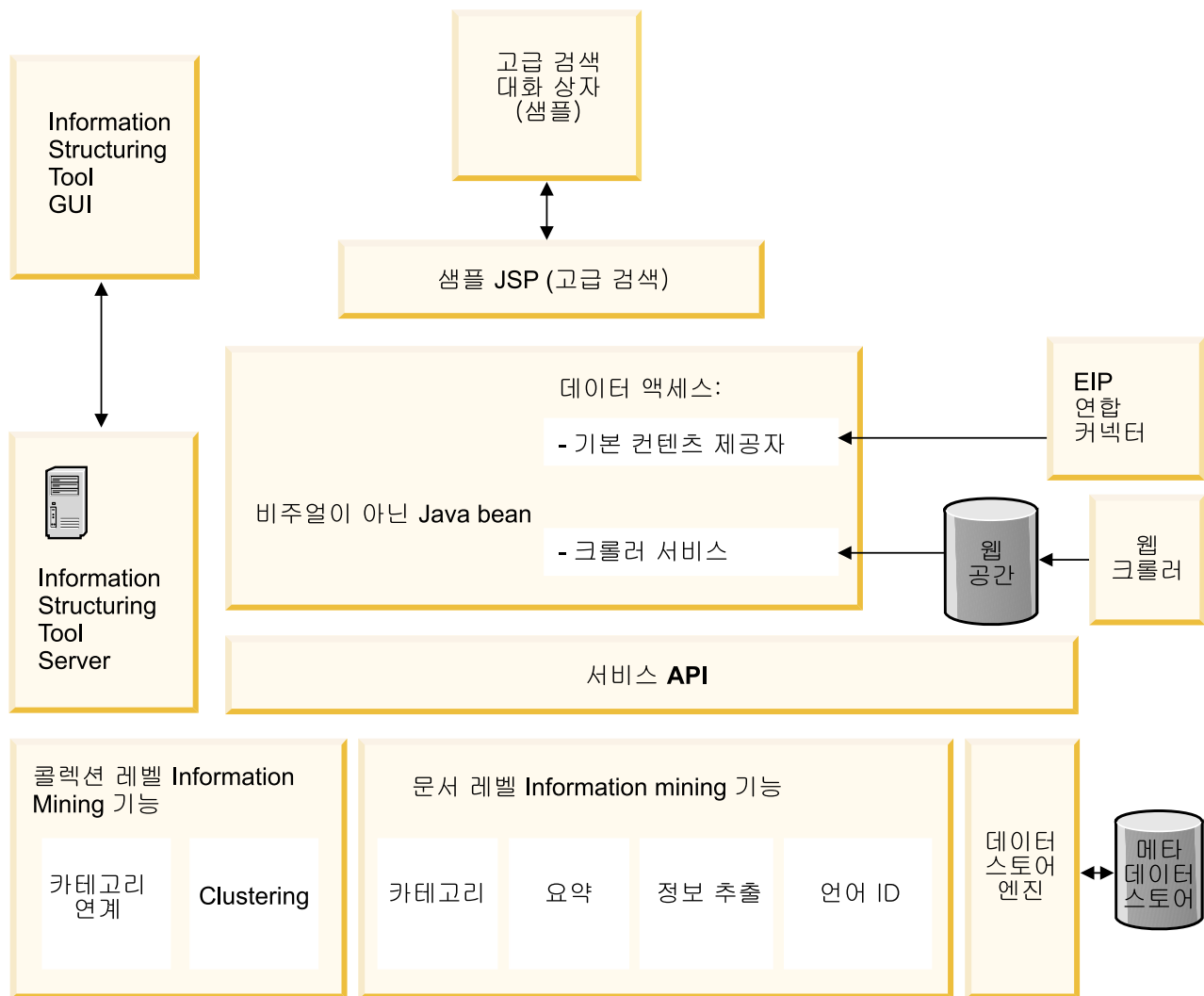


그림 5. Information Mining 시스템 아키텍처

도표의 오른쪽에 있는 상자는 Information Mining 서비스가 사용하는 구성요소를 나타내지만 다음과 같은 서비스 부분으로 간주되지 않습니다.

- 연합 커넥터(OO API의 부분)



- 웹 크롤러

Information Mining 기능의 여러 계층이 있습니다.

#### 1. Java Service API

이 계층은 Information Mining 기능 및 메타데이터 지속성을 일관된 Java API로 노출합니다.

#### 2. 비시각적 JavaBeans

이 계층은 표준 Bean에서 이벤트 유형 및 규약을 적용하는 JavaBeans 스펙에 근거하는 사용 준비 구성요소를 사용합니다.

#### 3. 샘플 Java 서버 페이지

이 레벨은 고급 검색(즉, 카테고리 제한사항이 있는 텍스트 검색)에 대한 응용프로그램을 설명하는 비시각적 JavaBeans를 사용하여 샘플 코드를 구성합니다.

#### 4. Information Structuring Tool

분류법 작성 및 유지보수를 위한 그래픽 사용자 인터페이스를 갖는 응용프로그램입니다.

## Information Mining 개념

Information Mining 기능을 효과적으로 사용할 수 있게 하고 완전히 이해하기 위해 이 절에서는 기본 개념을 다룹니다.

Information Mining 서비스는 개별 문서나 문서 컬렉션과 관련된 정보의 작성 및 유지보수를 위한 하부 구조를 제공합니다. 문서에 대한 이 정보는 **메타데이터**로 나타냅니다.

라이브러리는 Information Mining 데이터베이스의 콘텐츠의 개념적인 보기입니다. 라이브러리는 카탈로그 세트를 포함합니다.

카탈로그는 텍스트 문서에 대한 메타데이터 스토어이며 다음을 포함합니다.

- 각 문서에 대해 어떤 속성이 저장되는가를 정의하는 **카탈로그 스키마**.
- 카테고리의 계층 구조 트리 구조인 **분류법**.
- 자동으로 문서에 카테고리를 할당하기 위해 사용될 수 있는 문서 연계 결과에 근거한 카테고리 모델. 이 모델은 분류법이 작성되고 연계될 수 있는 **Information Structuring Tool**을 사용하여 생성됩니다. 모델은 카테고리 서비스에 대한 입력으로 사용됩니다.

스키마는 카탈로그에서 문서에 대해 생성되고 저장될 수 있는 속성 이름과 유형을 지정합니다. 스키마가 미리 정의되고 다음 속성을 포함합니다.

- 문자열 유형 IKF\_CONTENT
- 문자열 유형 IKF\_TITLE

- 문자열 유형 IKF\_AUTHOR
- 문자열 유형 IKF\_CATEGORIES
- 문자열 유형 IKF\_SUMMARY
- 문자열 유형 IKF\_LANGUAGE
- 문자열 유형 IKF\_FEATURES
- 문자열 유형 IKF\_COMMENTS
- 시간소인 유형 IKF\_DATE
- 정수 유형 IKF\_IDNUMBER

카탈로그는 카탈로그 스키마에 따라 레코드를 작성하여 가져온 문서에서 추출되거나 작성된 정보를 저장합니다. 레코드는 고유한 식별자와 이름 값 쌍 세트를 갖습니다. 지속적인 오브젝트 ID나 PID로 알려진 고유한 식별자(ID)는 작성된 레코드를 다시 원래의 문서 소스에 링크합니다.

60 페이지의 그림 6에서는 샘플 레코드를 표시합니다.

Record	
IKF_TITLE	"새"
IKF_AUTHOR	"J. Smith"
IKF_SUMMARY	"이 책은 새에 대한 요약입니다."
IKF_CATEGORIES	새/곤충을 먹는 동물
IKF_DATE	07/01/2001

그림 6. 샘플 레코드

JavaBeans 또는 Service API를 사용하여 Information Mining 레코드값을 설정하려면, 사용 값이 정의된 크기 제한 내에 있는지를 확인하십시오. 그렇지 않으면 `DKIKFSizeOutOfBoundsException`을 얻습니다. 정의된 제한은 다음과 같습니다.

표 6.

키	바이트의 최대 크기
IKF_CONTENT	209715
IKF_TITLE	2048
IKF_AUTHOR	2048
IKF_CATEGORIES	8192
IKF_SUMMARY	8192
IKF_LANGUAGE	8
IKF_FEATURES	524288
IKF_COMMENTS	8192

레코드가 작성되고 나면, 카탈로그에 할당하여 카탈로그에 저장됩니다. 카테고리는 레코드에 저장된 다른 값에 의해 선택될 수 있지만 보통, 카테고리 결과는 적절한 카테고리를 선택하기 위해 사용될 수 있습니다. 이와 같은 카테고리로 할당되어야 하는 레코드는 텍스트 검색을 사용하기 위해 문서 콘텐츠 색인화를 포함합니다. 각 카탈로그는 모든 검색 결과가 항상 자동으로 카탈로그 검색 범위 내에 있음을 의미하는 단일 텍스트 색인을 갖습니다.

데이터 스토어 엔진은 지속적 데이터 스토어에 대한 액세스를 유지보수하는 구성요소입니다.

## Information Mining 도구

Information Mining 서비스는 온라인 문서를 취급하기 위한 기능을 제공합니다. 이것은 다음을 포함합니다.

- Information Structuring Tool은 카탈로그를 작성하고 유지보수합니다.
- 언어 식별 서비스는 문서가 작성된 언어를 자동으로 발견합니다.
- 카테고리 서비스는 Information Structuring Tool을 사용하여 이전에 정의된 카테고리에 문서를 자동 할당합니다.
- 요약 서비스는 문서의 요약을 생성하기 위해 문서에서 단어와 구문을 분석합니다.
- 정보 추출 서비스는 도메인 종속 어휘 정의를 요구하지 않고, 자동으로 텍스트에서 중요한 항목을 인식합니다.
- 클러스터링 서비스는 그룹이나 클러스터로 문서 세트를 나눕니다. 각 클러스터의 문서는 공통 기능을 공유합니다. Cluster는 미리 정의되지 않으며, 자동으로 파생됩니다.
- 고급 검색은 특정 카탈로그로 제한된 카탈로그에 저장된 문서에서 텍스트를 검색합니다.

### Information Structuring Tool

Information Structuring Tool은 라이브러리라는 카탈로그 세트를 작성하고 유지보수하는 수단을 제공하는 웹 기반 응용프로그램입니다. 카탈로그는 문서에서 추출된 메타데이터를 저장하는 데 사용되며 미리 정의된 조직에 대한 문서를 구성하는 데 사용되는 분류법과 연관됩니다. 분류법은 문서를 주제에 관한 콘텐츠에 따라서 분류하는 카테고리의 계층적 구조입니다.

예를 들어, Information Structuring Tool을 사용하여 라이브러리 관리자가 새의 식습관을 설명하는 카탈로그를 정의할 수 있습니다.

62 페이지의 그림 7에서는 샘플 카탈로그를 표시합니다.

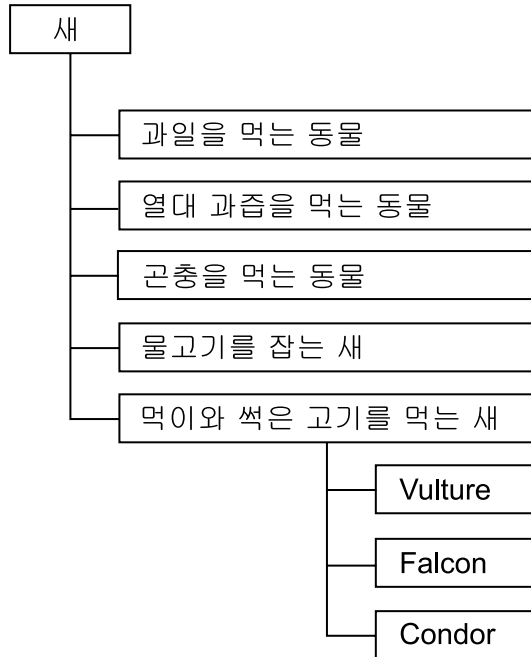


그림 7. 샘플 카탈로그

카테고리의 잘 구조화된 시스템은 대량 데이터에서 관련 정보를 찾는 데 상당히 도움을 줍니다. 카테고리는 문서 컬렉션의 의도했던 사용을 일치시키기 위해 선택되며 미리 샘플 문서를 사용하여 연계되어야 합니다. Information Structuring Tool을 사용하여 작성된 카테고리 모델은 차후에 카테고리 서비스에서 사용되어 자동으로 문서에 카테고리를 할당할 수 있습니다.

Information Structuring Tool 기능은 다음을 포함합니다.

- 카탈로그를 작성, 이름 바꾸기 및 삭제합니다.
- 카탈로그의 설명을 편집합니다.
- 카테고리를 작성, 이름 바꾸기 및 삭제합니다.
- 연계 문서를 카테고리로 그리고 카테고리로부터 추가하고 제거합니다.
- 연계 문서의 콘텐츠를 봅니다.
- 카탈로그 연계 처리를 시작하고 중지합니다.
- 카탈로그에서 연계 데이터의 품질에 대한 피드백을 얻습니다.

Information Structuring Tool 설치 방법과 사용에 대한 자세한 정보는 71 페이지의 『분류법 구축』을 참조하십시오.

## 언어 ID

언어 식별 서비스는 주어진 언어 세트에 대해 텍스트 문서가 쓰여진 가장 유사한 언어를 선택합니다.

언어 식별 서비스는 각 문서에 대한 언어와 신뢰값의 등급 목록을 리턴합니다. 언어 키는 ISO 표준 639에 따라 두 문자 코드로 지정됩니다. 신뢰값은 문서가 언어에 맞는 정도를 나타내는 측정이며 0(불량) 및 1(양호) 간의 범위를 갖는 부동 소수점 숫자로 표시됩니다. 언어 식별 알고리즘은 1개 국어를 사용하는 문서를 판별하도록 디자인되었습니다. 그러므로 다국어 문서의 경우, 신뢰값에 의한 등급은 문서의 올바른 언어를 반영하도록 보증될 수 없습니다.

다음과 같은 언어가 발견될 수 있습니다.

- 영어 EN
- 독어 DE
- 불어 FR
- 덴마크어 DA
- 핀란드어 FI
- 이탈리아어(IT)
- 노르웨이/북말(NB)
- 노르웨이/니노르스크(NO)
- 포르투갈어(PT)
- 스페인어(ES)
- 스웨덴어(SV)
- 한국어(KO)
- 일본어(JA)
- 중국어 및 대만어 ZH

다음 등록 정보가 설정될 수 있습니다.

- **maxResults**(Java Service API 사용 시에만):

각 문서에 대해 결정되고 리턴되는 언어의 최대 수. 0보다 크거나 같은 정수값입니다. 기본값은 1이며, 최적 등급 결과가 리턴되었음을 의미합니다. 값이 0으로 설정된 경우, 언어에 대한 신뢰값이 0.01 아래인 경우를 제외하고 인식된 모든 언어는 등급 순서로 리턴됩니다.

다른 Information Mining 서비스에 대한 사전 처리 단계로 언어 식별을 사용할 수 있습니다. 예를 들어, 정보 추출을 수행하기 전에 영어나 일본어 문서 전체를 찾기 위해 언어 식별을 사용할 수 있습니다.

## 카테고리

카테고리는 Information Structuring Tool을 사용하여 작성된 미리 정의된 조직 모델에 대한 문서를 조직하며 카테고리를 문서에 할당하는 수단입니다.

그러므로 카테고리 서비스를 사용하기 전에, Information Structuring Tool을 사용하여 그러한 모델을 작성하기 위한 분류법을 정의하고 연계해야 합니다.

카테고리 결과에는 문서가 카테고리에 맞는 정도를 나타내는 카테고리 신뢰값이 들어 있습니다. 그러한 결과의 세트는 각 문서에 대해 리턴됩니다. 목록은 리턴된 신뢰값에 따라 등급화됩니다.

다음 등록 정보가 설정될 수 있습니다.

- **maxResults:**

각 문서에 대해 리턴할 카테고리의 최대 수. 기본값은 1이며, 모든 카테고리가 리턴됨을 의미합니다. 결과 목록이 등급화됩니다.

- **minConfidence:**

문서에 할당된 신뢰값은 문서가 카테고리에 맞는 정도를 나타냅니다. 매개변수 minConfidence는 0(다소 덜 적합) 및 1(보다 적합) 간의 범위를 갖는 최소값을 지정합니다. 기본값은 0으로 설정되며, 문서에 할당된 모든 카테고리가 리턴됨을 의미합니다. 결과 목록은 다소 낮게 적합되는 것부터 다소 덜 적합한 등급까지 나뉩니다.

- **catalogName:**

이것은 카테고리에 사용되는 카탈로그를 지정합니다. 카탈로그는 Information Structuring Tool를 사용하여 카테고리용으로 작성되고 연계될 수 있습니다.

지원되는 언어 목록에 대해서는 57 페이지의 『지원되는 언어 및 형식』을 참조하십시오.

## 요약

문서 요약은 문서 콘텐츠의 특성인 문서에서 추출된 구문의 컬렉션으로 구성됩니다. 예를 들어, 요약 도구는 문서가 관련되는지 여부를 결정하도록 도우며 완전히 읽을 수 있어야 하거나, 조회 결과의 부분으로 리턴될 때 문서 링크를 따를 지 여부를 결정하도록 돕습니다.

요약 결과에는 각 구문을 선택하고 서로 인접하여 발생했는지 여부를 결정하기 위해 응용프로그램에 의해 사용될 수 있는 데이터 구조(행렬)로 된 요약과 단일 문자열로 된 요약이 들어 있습니다.

요약 서비스는 다른 모드에서 사용될 수 있습니다. 모드는 값 maxLength와 ratio가 요약 길이를 판별하는 데 사용되는 방식을 결정합니다.

- **최대 길이:**

요약에서 최대 구문 수. 작성된 요약은 maxLength 보다 길지 않습니다. 기본값은 3입니다.

- **비율:**

문서의 총 길이에 대한 구문 수. 작성된 요약 길이는 문서의 전체 길이로 판별됩니다. 기본값은 0.1입니다.

- **모드:**

요약 길이 설정에 필요한 maxLength와 비율 간의 관계를 결정합니다. 서로 다른 모드는 다음과 같습니다.

– MODE\_LESS\_THAN\_MAXLENGTH:

요약은 기껏해야 maxLength 수의 구문을 갖습니다. 이것이 기본 모드입니다.

– MODE\_EQUALS\_RATIO:

요약은 정확히 비율 수의 구문을 갖습니다. 이것은 문서의 총 구문 수와 비율을 곱해 결정됩니다.

– MODE\_EQUALS\_RATIO\_BUT\_AT\_MOST\_MAXLENGTH:

요약은 최소 비율 수의 구문을 갖지만(문서의 총 수와 비율 곱), 구문의 maxLength 수보다 크지 않습니다.

지원되는 언어 목록에 대해서는 57 페이지의 『지원되는 언어 및 형식』을 참조하십시오.

## 정보 추출

문서 분석 시 중요한 작업은 문서 콘텐츠에 대한 정보를 제공하는 항목의 추출입니다. 이 주요 요소는 다음과 같이 사용될 수 있습니다.

- 문서가 중요한 지 여부를 평가하도록 돕는 중요한 정보를 나타내기 위해
- 조회 세분을 위해 사용할 주요 개념을 찾고 저장하기 위해
- 관련된 문서 수집을 위한 기준으로

주요 요소의 예는 단어, 이름 또는 다중단어 용어와 같은 어휘 항목입니다.

영어의 경우, 정보 추출 서비스는 동일한 엔티티를 나타내거나 동일한 개념을 표시하는 경우 텍스트에서 그러한 키 요소의 발생을 함께 그룹화하고 찾는 키 요소를 정규화합니다. 예를 들어, James J. Smith, Mr. Smith, James and Smith가 문서에 있는 경우, 동일한 정규화 양식으로 매핑하여 동일한 사람을 나타내는 것으로 모두 표시됩니다. 단어의 변화된 양식도 정규화 양식으로 매핑됩니다(예: children을 child로 매핑).

그러나 일본어의 경우, 모든 키 요소는 문서에서 발생할 때 추출됩니다. 정규화가 ISO8601 및 ISO4217을 준수하는 곳에서, 날짜, 시간 및 통화 표현식을 제외하고 정규화는 수행되지 않습니다.

정보 추출 서비스를 사용하여 다음에 대한 문서를 분석할 수 있습니다.

- 단일 및 다중단어 어휘 항목(예: 공고, 값, 제품 주기)
- 장소, 사람 및 조직의 이름(예: Washington, Bush, Data Management Academy)
- 약어(예: MB(megabytes))
- 날짜, 돈 및 숫자에 대한 용어(예: 1958년 1월 11일, 01/11/58, \$30, 30 펜스, 4.5, 5000)

다음은 추출하기 위해 세 가지 유형의 정보를 지정할 수 있습니다.

- 이름
- 용어
- 표현식

Java Service API를 사용하여, 아래 나열된 하위 유형을 식별하며 하위 유형이 추출된 어휘 항목에 어느 정도 맞는지 언급하는 신뢰값을 식별할 수도 있습니다. 신뢰값의 범위는 0(불량) 및 1(양호)입니다. 유형 이름, 용어 및 표현식의 하위 유형은 다음을 포함합니다.

- 이름
  - 위치(예: Montreal 또는 London)
  - 사람(예: Tim Brown)
  - 조직(예: Smith and Son)
  - 알 수 없음(예: Smashing Pumpkins, Silicon Valley, CCTV(전체 양식을 갖지 않은 약어))
  - 기타(예: AIS Plan, ISO Conference, Internet, Privacy Act Officer, JCAHO Performance Report)
- 용어
  - 지정되지 않은 용어(예: 복잡한 오락, 예술 세계, 클래스 변수, 소스 코드, 데이터 정의, 처리 개선 발의)
- 표현식
  - 기수(예: 4, 50, 70)
  - 서수(예: 40번째, 50번째)
  - 퍼센트(예: 12%, 60 퍼센트)
  - 날짜(예: 07/28/98)
  - 시간(예: 18시간, 4시)
  - 돈(예: DM90, 30 파운드)
  - 약어(예: NY)

정보 추출은 영어와 일본어 문서에서만 작업합니다. 영어나 일본어가 아닌 문서 컬렉션에서 문서를 식별하기 위한 사전 처리 단계로 언어 식별 서비스를 사용할 수 있습니다. 정보 추출 서비스를 다른 마이닝 기능과 결합할 수 있습니다. 예를 들어, 요약 서비스에 대한 사전 처리 단계로 사용하여 Texas의 통치자가 아닌 대통령으로서의 Bush에 대한 정보만을 요약할 수 있습니다.

## 클러스터링

클러스터링 도구는 문서의 컬렉션을 배열하여 유사한 문서를 함께 그룹화하고, 다른 그룹(클러스터)의 문서가 콘텐츠에 대해 서로 구분되도록 합니다. 이러한 방식으로, 클러스터링은 많은 문서 컬렉션의 개요를 제공하며 관련된 문서를 식별하는 수단으로 사용



될 수 있습니다. 유사하게 응용프로그램 영역 내에서 연계 문서를 클러스터링하여 Information Structuring Tool를 사용하여 분류법 빌드를 지원하기 위해 사용될 수 있습니다. 또한 클러스터링은 새로운 경향이나 새로운 기술을 지시할 수 있는 컬렉션 내에서 유사한 문서와 복제본이나, 경쟁적 분석에 대한 흥미를 일으킬 수 있는 매우 유사한 문서를 찾는 데 유용합니다.

클러스터링은 각 클러스터의 문서가 콘텐츠에 대해 서로 가능한 유사한 클러스터로 문서를 구성하는 반복적인 처리이며, cluster는 가능한 한 서로 다릅니다. 클러스터링은 전체로서 문서 컬렉션에서 작업하며 문서 레벨에서 작업하는 카테고리 또는 요약 등 위의 Information Mining 서비스와 대조됩니다. 클러스터링은 각 문서의 대표적인 기능을 서로 비교하고 기능 유사성에 따라 문서를 그룹화하여 작업합니다.

클러스터링 단계 중, 새로운 문서는 문서 세트에 추가될 수 없습니다.

다음 등록 정보가 설정될 수 있습니다.

- **maxClusterCount**  
리턴되는 클러스터의 최대 수.
- **minClusterCount**  
리턴되는 클러스터의 최소 수.
- **clusterFeatureCount**  
Cluster당 리턴되는 레이블(키워드) 수.

그러나 이 값은 clusterer로 바인딩되지 않지만, 단지 지침 경계로 사용됩니다. 클러스터링 서비스의 출력은 결과 목록입니다.

클러스터링은 영어 문서에서만 작업합니다. 영어가 아닌 문서 컬렉션에서 문서를 식별하기 위한 사전 처리 단계로 언어 식별 서비스를 사용할 수 있습니다.

## 고급 검색

전체 Enterprise Information Portal 콘텐츠 서버에서 수행되는 표준 Enterprise Information Portal 검색과 대조적으로, 소위 고급 검색은 Information Structuring Tool으로 작성된 카탈로그에 저장된 ID가 있는 문서에서만 검색합니다. 심지어 검색을 더 좁히기 위해, 고급 검색 조회는 텍스트만을 검색할 뿐 아니라 특정 카테고리의 문서로 검색을 제한할 수도 있습니다.

다음 매개변수가 설정될 수 있습니다.

- **catalogName:**  
이것은 검색에 사용되는 카탈로그를 지정합니다. 카탈로그는 Information Structuring Tool를 사용하여 작성되고 연계될 수 있습니다.
- **maxResults:**

각 조회에 대해 리턴된 검색 결과의 최대 수. 기본값은 0이며 모든 결과가 리턴되었음을 의미합니다.

다음 조회 유형이 제출될 수 있습니다.

1. 순수한 텍스트 조회. 이 검색은 텍스트 조회와 일치하는 모든 문서를 리턴합니다. 결과 목록이 관련에 의해 순서화됩니다.
2. 순수한 카테고리 검색. 이 검색은 카테고리에 할당된 모든 문서를 리턴합니다. 결과는 임의의 순서입니다.
3. 결합된 텍스트 및 카테고리 검색. 이 검색은 텍스트 조회와 일치하며 카테고리에 할당된 모든 문서를 리턴합니다. 결과 목록이 관련에 의해 순서화됩니다.

시스템에 대해 제출된 고급 검색 조회는 항상 특정 카탈로그에 바인드됩니다. 이것을 카탈로그 검색 범위라고 합니다. 카탈로그는 참조되어야 하는 가져온 문서에서의 보기를 나타내므로 상호 카탈로그 검색이 가능하지 않습니다.

조회 문자열의 BNF(조회 구문규칙)는 다음과 같습니다.

```

query_string      ::= term
term              ::= ( term )
                  ::= single_term
                  ::= compound_term
single_term       ::= category_term
                  ::= text_search_term
                  ::= string_term
                  ::= number_term
compound_term     ::= term binary_bool_operator term
                  ::= unary_bool_operator single_term
category_term     ::= ( DKIKFCategory category_operator category_path_value )
text_search_term  ::= ( "attribute name" CONTAINS text_search_value )
string_term       ::= ( "attribute name" string_operator string_value )
number_term       ::= ( "attribute name" basic_operator number_value )
binary_bool_operator ::= AND | OR
unary_bool_operator  ::= NOT
category_operator  ::= >= | =
string_operator     ::= LIKE | basic_operator
basic_operator      ::= > | < | <= | >= | != | =
category_path_value ::= "category path"
text_search_value   ::= "'string'"
string_value        ::= "string"
number_value        ::= "integer" | "decimal number"

```

- 단말기, 문자열 및 숫자는 일반 용어를 나타냅니다.
- 카테고리 연산자 '='는 오직 하나의 카테고리로 검색 범위를 제한합니다.
- 카테고리 연산자 '>='는 이 카테고리와 카테고리 트리에서의 모든 관련 서브카테고리에 대한 검색을 펼칩니다.
- CONTAINS 절에서의 문자열 검색은 단일 문자에 대한 총칭 문자('\_')와 임의의 문자량에 대한('%') 총칭 문자를 포함할 수 있습니다. 예를 들어, \_LOB는 BLOB 및 CLOB와 일치할 수 있으며, 반면에 %name은 filename과 일치할 수 있습니다. 검색 가능한 것으로 플래그된 스키마 속성만(예: IKF\_CONTENT) 문자열 연산자 CONTAINS를 사용하여 조회될 수 있습니다.

- LIKE 절의 검색 문자열은 SQL에서 사용된 충칭 문자를 포함할 수 있습니다.
- 현재 지원되는 속성 이름의 전체 목록에 대해서는 59 페이지의 『Information Mining 개념』의 내용을 참조하십시오

조회 예:

- 순수한 텍스트 조회:

```
("IKF_CONTENT" CONTAINS "'southern Africa'") AND NOT
("IKF_CONTENT" CONTAINS "'Cape'")
```

- 순수한 카테고리 조회:

```
("DKIKFCATEGORY" >= "birds/Fruit eaters")
```

- 결합된 텍스트 및 카테고리 조회:

```
("IKF_CONTENT" CONTAINS "'South Africa'") AND
("DKIKFCATEGORY" >= "birds/Birds of prey and scavengers/Falcon")
```

- 속성 조회:

```
("IKF_SUMMARY" LIKE "humming birds in the tropics")
```

또는

```
("IKF_FEATURES" LIKE "Goethe") AND ("IKF_TITLE" = "Faust")
```

## 인터페이스 프로그래밍

Information Mining 기능은 다음과 같은 응용프로그램을 빌드하는 데 사용 가능합니다.

- Java Service API
- Information MiningJavaBeans

**Java Service API**는 Enterprise Information Portal 서비스로 Information Structuring Tool의 부분인 카탈로그 유지보수를 제외한 전체 Information Mining 기능을 통합합니다. Java RMI에 근거한 클라이언트/서버 통신을 제공합니다.

Java Service API를 사용하는 응용프로그램은 다음을 수행할 수 있습니다.

- 문서가 작성된 언어 판별
- 텍스트 문서의 요약 작성
- 카테고리를 문서에 할당
- 예를 들어, 텍스트 문서에서 이름, 용어 또는 표현식과 같은 정보 추출
- 유사한 문서 함께 그룹핑
- 카탈로그에서 문서에 대한 메타데이터를 저장하고 찾아 보기
- 예를 들어, 요약과 같이 특정 카테고리로 제한된 문서와 속성에서 텍스트 검색 수행

Java Service API는 직접 메소드 호출을 사용하는 로컬 모드 또는 Java RMI(Remote Method Invocation)를 사용하는 원격 모드에서 실행할 수 있습니다. 원격 실행은 웹 응용프로그램을 실행하는 응용프로그램 서버로 하나의 서버를 구성하고 텍스트 분석, 색인화 및 검색을 수행하는 Information Mining 서버로 또다른 서버를 구성할 수 있습니다. 서버 작업 메커니즘은 전체 작업이 Information Mining 서버(원격 시스템)로 송신될 수 있으며 전체 처리가 해당 시스템에서 수행되는 수단입니다.

웹 크롤러 사용 시, 해당 액세스 메커니즘은 JavaBeans를 사용하여 구현되어야 합니다. 웹 크롤러 액세스는 Java Service API 레벨에서 사용할 수 없습니다.

Information Mining Java Service API의 자세한 설명의 경우, Windows용 응용프로그램 프로그래밍 안내서를 참조하십시오.

70 페이지의 그림 8에서는 Information Mining 원격 구성을 설명합니다.

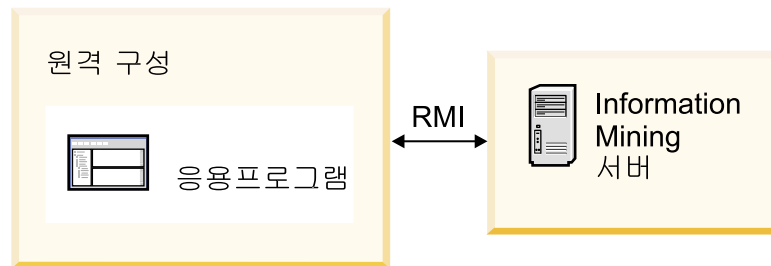


그림 8. Information Mining 원격 구성

**Information Mining Beans**는 빠른 응용프로그램 개발을 위한 상위 레벨 Java API이며 JavaBeans 스펙에 따라 빌드됩니다. Bean이 서버 작업 처리를 지원하지 않으므로, 성능의 이유로 Bean을 사용하는 모든 응용프로그램 개발은 동일한 시스템에서 수행되어야 합니다.

각 Bean은 이벤트 단위로 수행될 수 있으며 일부는 직접 호출될 수 있는 메소드를 제공합니다. 기존 Enterprise Information Portal Beans을 사용하는 통합은 Enterprise Information Portal Beans으로 사용된 결과 이벤트의 지원을 통해 수행됩니다. 이는 연합 검색 및 웹 크롤러 결과 이벤트가 Information Mining 서비스와 호환 가능하며 Information Mining 결과는 EIP로 처리될 수 있다는 것을 의미합니다. Information Mining Beans의 자세한 설명에 대해서는 Windows용 응용프로그램 프로그래밍 안내서를 참조하십시오.

---

## First Steps

Enterprise Information Portal Information Mining First Steps는 실제 시나리오에 기초하여 비즈니스 환경에 IBM Information Mining 기술이 적용될 수 있는 방법에 대한 지식 엔지니어, 시스템 관리자 또는 응용프로그램 프로그래머를 표시하는 스토리보드 자습서입니다. 자습서는 다음과 같이 구성됩니다.

- 간략한 소개
- First Steps가 시작되기 전에 수행해야 하는 사항
- 샘플 데이터 구성
- 데이터 액세스
- 샘플 클라이언트 사용
- 샘플 데이터 제거
- 자세한 읽기 참조

Information Mining First Steps를 액세스하려면,  
<CMBROOT>\ikf\firststeps\first\_steps.html을 실행하십시오.

Windows에서 Information Mining First Steps를 실행하려면, 시작 버튼을 누르고 멀티플랫폼용 **Enterprise Information Portal 8.1** --> **<Information Mining>** --> **First Steps**를 선택하십시오.

Enterprise Information Portal Information Mining First Steps는 설치 확인으로도 사용할 수 있습니다.

---

## 분류법 구축

Information Structuring Tool은 라이브러리라는 카탈로그 세트를 작성하고 유지보수하는 수단을 제공하는 웹 기반 응용프로그램입니다. 카탈로그는 메타데이터를 저장하는 데 사용되고 정보를 카탈로그로 구성하는 분류법과 연관됩니다. 분류법은 문서를 주제에 관한 콘텐츠에 따라서 분류하는 카테고리의 계층적 구조입니다. 예를 들어, 분류법의 맨 위 레벨은 **비즈니스**, **문화**, **스포츠** 같은 카테고리를 포함하는 반면, **스포츠**는 다음 하위 레벨에서 **팀 스포츠**와 **경기로** 분할됩니다. 한 레벨 더 아래에서 **팀 스포츠**는 **축구**, **야구** 및 **테니스**로 분할됩니다.

카테고리에 연계 문서를 할당하고 카탈로그를 연계하면 Information Structuring Tool에서는 그 후에 문서에 카테고리를 할당하기 위해 카테고리 서비스 Bean이 사용할 수 있는 카테고리 모델을 작성합니다.

## Information Structuring Tool

Information Structuring Tool은 servlet 컨테이너(예: IBM WAS servlet 엔진)에서 웹 응용프로그램으로 전개되어야 합니다.

그러나 동일한 Information Mining인스턴스에 대해 두 개의 Information Structuring Tool 웹 응용프로그램(예: 하나는 IST1, 다른 하나는 IST2) 작업은 허용되지 않는다는 점에 유의하십시오.

Information Structuring Tool을 전개하기 전에, 올바른 WAS Version이 설치 및 구성되었는지를 확인하십시오. 세부사항에 대해서는 *Enterprise Information Portal* 계획 및 설치를 참조하십시오.

Information Structuring Tool의 개발에 필요한 사용자 액세스 권한은 다음과 같습니다.

- Windows의 경우: 시스템 관리자 권한
- AIX의 경우: 루트 사용자 권한
- Sun의 경우: 루트 사용자 권한

## 시작하기

이 응용프로그램에서는 분류법을 정의, 유지보수 및 연계할 수 있는 웹 기반 인터페이스를 제공합니다. 두 개의 프레임이 있습니다. 왼쪽 프레임을 카탈로그 보기라고 하며, 이 프레임은 분류법을 작성 및 유지보수하는 데 사용됩니다. 오른쪽 프레임을 노트북이라고 하며, 이 프레임에서는 응용프로그램 사용에 대한 정보를 제공합니다. 노트북은 카탈로그에 대한 문서를 연계하고 평가하는 데 사용되는 일련의 탭을 표시합니다. 응용프로그램을 사용하려면 먼저 왼쪽 카탈로그 보기 프레임에 카탈로그와 카테고리가 작성되어야 합니다.

## 액세스 권한

Information Structuring Tool에서, 사용자 이름 및 암호 정보는 Enterprise Information Portal로 유지보수됩니다. Enterprise Information Portal에 알려진 사용자 이름 및 암호만을 입력할 수 있습니다.

Information Structuring Tool 시작 시, 연계 문서 업로딩에 사용되는 Java Applet이 파일 시스템에 대해 읽기 액세스를 필요로 하므로 보안 경고 메시지가 표시됩니다. 감퇴 시, 연계 문서를 업로드할 수 없으므로 분류법을 연계할 수 없습니다.

Information Structuring Tool는 다중 사용자 환경에서 실행할 수 있습니다. 복수 사용자가 동일한 분류법으로 볼 수 있도록 허용하기 위해, Information Structuring Tool에서는 카탈로그 및 해당 카테고리에 대한 액세스를 제어하는 잠금 메커니즘을 제공합니다.

사용자는 카탈로그를 사용하여 작업하기 전에 카탈로그를 선택하고 잠가서 카탈로그를 명시적으로 잠그거나 카탈로그를 사용하는 작업(예: 연계 문서 추가)을 시작할 수 있는

데, 이 경우 카탈로그는 액세스 충돌을 예방하기 위해 자동으로 잠깁니다. 다른 사용자는 이 카탈로그를 볼 수 있지만 카탈로그를 잠근 사용자가 다시 잠금 해제할 때까지는 카탈로그의 어떤 것도 변경할 수 없습니다.

Information Structuring Tool이 전개되는 응용프로그램 서버가 종료되는 경우, 모든 잠금이 삭제됨에 유의하십시오.

## 분류법 정의

카탈로그는 카테고리로 구성되는 트리형의 구조인 분류법의 고정 지점입니다.

새 카탈로그를 정의하고 해당 카테고리를 선택하는 단계는 다음과 같습니다.

### 1. 새 분류법을 정의하고 작성할 카테고리를 결정합니다.

카탈로그 보기에서 라이브러리를 선택하고 오른쪽 마우스 단추를 누르십시오. 메뉴가 표시됩니다. 새 카탈로그를 선택하면 카탈로그 아이콘이 작성됩니다. 카탈로그 이름을 입력하여 이 아이콘의 이름을 바꾸고 **Enter** 단추를 누르십시오. 이름이 동일한 폴더가 작성됩니다. 이것이 루트 카테고리입니다. 노트북의 콘텐츠가 변경됩니다.

새 카탈로그를 라이브러리에 추가하는 다른 대체 메소드는 Information Structuring Tool 외부에서(예: 파일 시스템에서), 작성된 기존 분류법을 가져오는 것입니다. 자세한 정보는 75 페이지의 『연계 문서 업로드』를 참조하십시오.

활동적으로 분류법에서 작업하려면 카탈로그를 잠궈야 합니다. 카탈로그 작성 시, 이것은 자동으로 수행됩니다. 기존 카탈로그를 잠그려면 카탈로그를 선택하고 오른쪽 마우스 단추를 누르십시오. 메뉴가 표시됩니다. 카탈로그 잠금을 선택하면 카탈로그 상태 아이콘이 변경됩니다. 카탈로그 상태의 여러 유형을 표시하기 위해 아이콘이 사용됩니다.



분류법 트리가 접히고 어떤 사용자도 분류법 트리를 잠글 수 없습니다.



분류법 트리가 확장되고 어떤 사용자도 분류법 트리를 잠글 수 없습니다.



분류법 트리가 접히고 현재 사용자에게 의해 잠겨집니다.



분류법 트리가 확장되고 현재 사용자에게 의해 잠겨집니다.



분류법 트리가 접히고 어떤 사용자에 의해 잠겨집니다.



분류법 트리가 확장되고 어떤 사용자에 의해 잠겨집니다.

그림 9에서는 두 가지 카탈로그의 예를 나타냅니다.

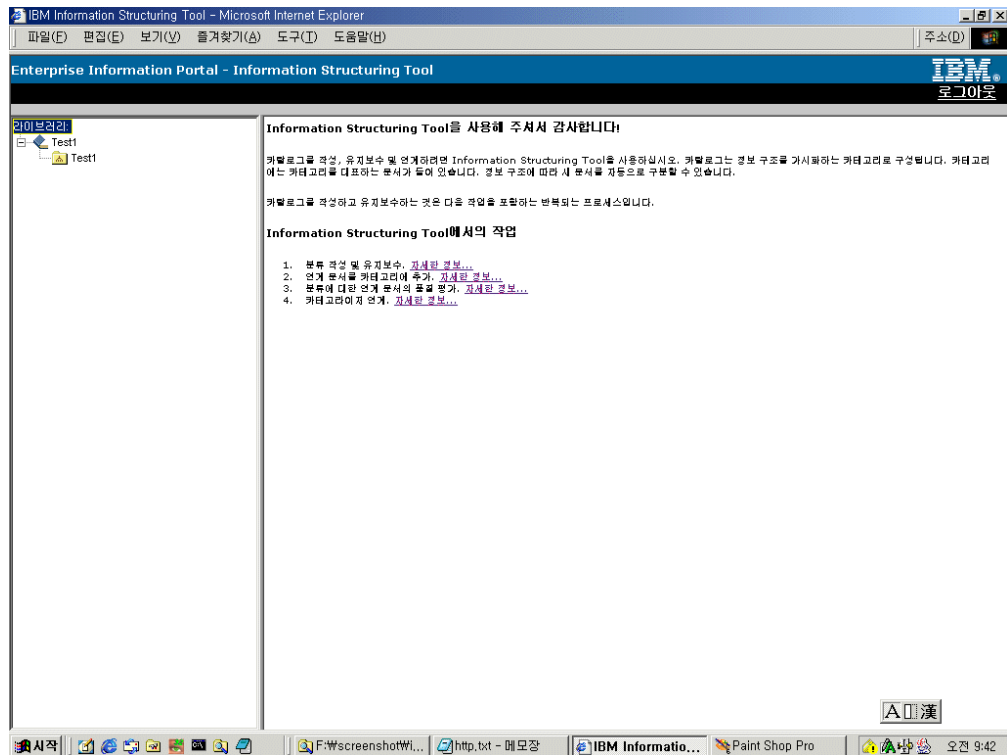


그림 9. 카탈로그 예

2. 일단 카탈로그가 작성되고 잠기면 카테고리를 추가, 이름 바꾸기 또는 삭제할 수 있습니다. 처음으로 새 카탈로그를 작성하면 카탈로그와 이름이 동일한 루트 카테고리도 작성됩니다. 이 카테고리의 이름을 바꿀 수 있습니다. 이름을 바꾸려면 카테고리를 선택하고 오른쪽 마우스 단추를 누르십시오. 이름 바꾸기를 선택하고 새 이름을 입력한 후, **Enter**를 누르십시오.

새 하위 카테고리를 추가하려면 카테고리를 강조표시하고 오른쪽 마우스 단추를 누르십시오. 새 카테고리를 선택하고 새 이름을 입력한 후, **Enter**를 누르십시오. 트리 구조에서 동일한 레벨에 있는 카테고리 이름은 고유해야 합니다.

루트 카테고리는 삭제할 수 없습니다. 카탈로그가 삭제될 때만 루트 카테고리가 삭제됩니다. 카테고리를 삭제하면 카테고리에 있는 모든 연계 문서, 기록 및 하위 카테고리(트리 구조의 하위 카테고리)도 제거됩니다.



3. 선택한 카탈로그에 관련된 더욱 설명적인 정보가 등록 정보 탭에 표시됩니다. 처음에, 설명 필드는 비어 있습니다. 카탈로그 설명을 추가 또는 편집하려면 설명 편집 단추를 누르십시오. 새 창이 표시되어, 설명을 입력할 수 있습니다.

## 연계 문서 선택

카테고리 모델의 품질은 각 카테고리에 할당된 연계 문서의 품질에 의해 크게 좌우됩니다.

연계 문서는 Enterprise Information Portal이 지원하는 문서 형식 중 하나여야 합니다. 지원되는 형식 목록에 대해서는 137 페이지의 제 9 장 『문서 형식』을 참조하십시오.

연계 문서의 적절한 세트를 선택하는 것은 필수적입니다. 문서는 다음과 같아야 합니다.

- 카테고리를 대표해야 합니다
- 너무 많은 마크업이나 단어 목록 없이 상당한 양의 설명적 텍스트를 포함해야 합니다
- 모두 동일한 쓰기 양식으로 작성되어야 합니다. 예를 들어, 다른 문서가 보고서 양식으로 되어 있는 경우 산문 쓰기는 피하십시오
- 모두가 거의 비슷한 길이이고 가능한 너무 길지 않아야 합니다. 연계 문서의 길이는 또한 카테고리 서비스를 사용하여 구분하려는 문서와 거의 동일해야 합니다.

카테고리당 약 40개의 연계 문서의 컬렉션이 바람직합니다. 그러나 선택한 카테고리가 더 일반적인 경우, 더 많은 문서가 필요합니다. 카테고리는 주의하여 선택되고 의미가 있어야 하며, 인간 색인자에게 모호하고 불분명하게 표시되는 연계 문서와 카테고리는 분명히 자동 처리 중 문제점을 제기합니다.

## 연계 문서 업로드

연계 문서를 추가하려면 각 카테고리를 선택하십시오. 연계 문서 목록 창이 표시됩니다. 이 목록에 문서를 추가하려면 문서 추가... 단추를 누르십시오.

연계 문서 추가 창이 나타납니다. 파일 업로드 후에 이 창을 닫지 말아야 하며, 다른 카탈로그에 업로드하는 차후 문서나 다른 카탈로그에 사용될 수 있습니다.

연계 문서를 추가하려면 찾아보기... 단추를 누르고 열기 창에서 관련 파일 또는 디렉토리를 선택하십시오.

동일한 디렉토리나 전체 디렉토리에서 업로드할 하나 이상의 파일을 선택할 수 있습니다. 선택한 디렉토리가 비어 있는 경우, 사용자에게 알립니다.

이것은 Information Structuring Tool 외부(예: 파일 시스템)에서 작성된 기존 분류법으로 가져오고 작업할 수 있게 합니다.

파일 시스템에서 디렉토리를 선택한 경우(예: Development), 서브디렉토리 Design이 들어 있으며, 분류법 트리에서 카테고리 Development에는 서브카테고리 Design이 들어 있으며, 서브디렉토리 내의 파일이 서브카테고리에 추가됩니다. 서브카테고리가 없는 경우, 작성되며 해당 파일이 이 새로 작성된 서브카테고리에 추가됩니다.

문서 언어 및 선택된 전체 파일의 형식을 선택하십시오. 일반 텍스트 파일을 제외하고는 항상 "자동 탐지" 형식을 사용하십시오.

연계 문서 목록에 파일을 추가하려면 제출 단추를 누르십시오.

그림 10에서는 연계 문서 추가를 표시합니다.

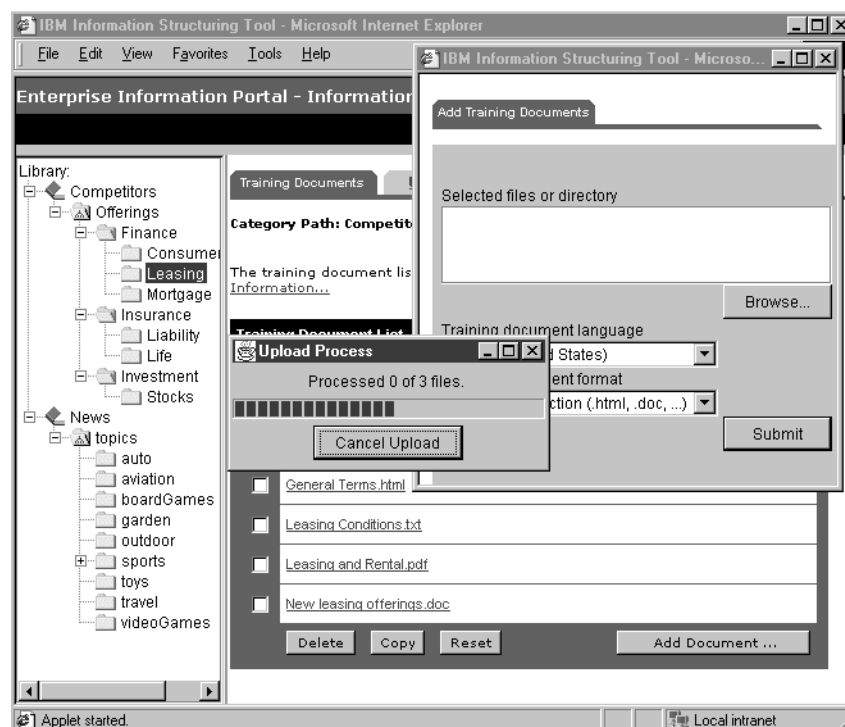


그림 10. 연계 문서 추가

업로드 처리 창이 나타납니다. 이 창에서 문서 업로드 처리를 취소할 수 있습니다. 업로드 처리가 취소되면, 더 이상 파일이 추가되지 않습니다. 이미 연계 문서로 추가된 파일은 제거됩니다. 그러나 업로드 중 서브카탈로그가 추가된 경우, 카탈로그에서 삭제되지 않습니다.

파일을 정상적으로 업로드할 수 없는 경우, 업로드 상태 창은 파일이 갱신될 수 없는 이유에 대한 정보와 함께 자동으로 표시됩니다. 예를 들어, 다음과 같습니다.

- 같은 이름을 가진 파일이 이미 존재합니다.
- 파일이 비어 있습니다.
- 파일이 서버로 업로드될 수 없습니다.

업로드 상태 창에서, **연계 문서** 탭을 눌러 연계 문서 목록으로 리턴하십시오. 카테고리  
로 정상 업로드된 전체 연계 문서가 표시됩니다. **이전**과 **다음** 단추를 사용하여 목록의  
전체 문서를 볼 수 있습니다.

연계 문서를 둘 이상의 카테고리에 추가하려면 파일을 첫 번째 카테고리로 업로드한 다  
음 이를 다른 카테고리로 복사하십시오. 이 파일은 다시 업로드하지 마십시오. 복사하  
려면, 연계 문서 목록 창에서 하나 이상의 문서를 선택하고 복사를 누르십시오. 표시되  
는 창에서 **찾아보기...**를 눌러 문서를 복사하려는 하나 이상의 카테고리를 선택한 후 제  
출을 누르십시오.

다음 조치는 파일 업로드 중 허용되지 않습니다.

- 카탈로그 잠금 해제
- Information Structuring Tool로부터 로그아웃
- 카탈로그 연계 또는 평가 시작
- 카탈로그 이름 바꾸기
- 카탈로그의 업로드 상태 정보 삭제

그러나 다음 조치가 허용됩니다.

- 동일한 카테고리나 다른 카테고리로 다른 파일 업로드 처리 시작
- 다른 카탈로그에서 작업

78 페이지의 그림 11에서는 연계 문서 목록을 표시합니다.

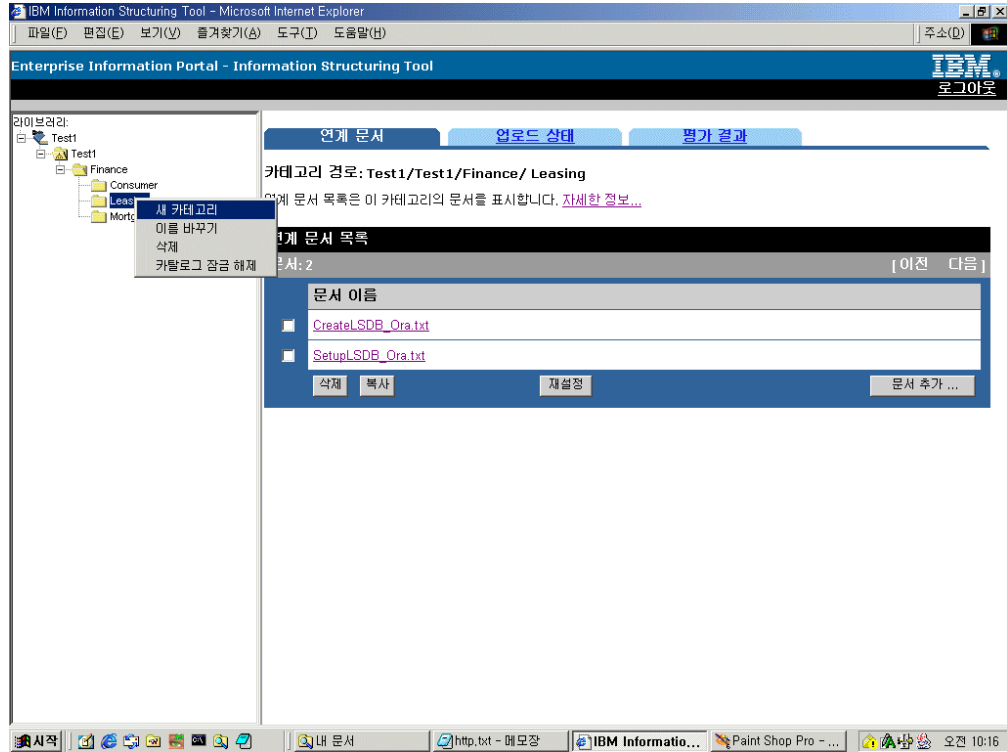


그림 11. 연계 문서 목록

파일 업로드 중 브라우저가 닫히면, 파일이 이미 서버로 전송된 경우 해당 파일이 연계 문서로 추가됩니다. 그렇지 않은 경우, 파일이 추가되지 않습니다.

전체 연계 문서를 업로드한 후에, 평가를 시작하려면 카탈로그를 선택하십시오.

## 카테고리 모델 평가

분류법이 정의되고 연계 문서가 각 카테고리에 할당된 경우, 분류법이 평가되어야 합니다. 분류법 평가는 연계 문서가 미리 정의된 분류법에 대해 어느 정도 좋은지 판단하도록 돕습니다. 이것은 다음 단계로 구성된 반복 처리입니다.

1. 평가를 시작하십시오.
2. 평가 결과를 액세스하십시오.
3. 분류법이나 연계 문서에 변경사항을 작성하십시오.
4. 평가를 재실행하십시오.

각 평가 반복 중, 평가 처리는 다음과 같습니다.

- 연계 문서를 연계 세트(문서의 약 80%)와 테스트 세트(문서의 약 20%)로 나눕니다. 연계 세트를 사용하여 카탈로그를 연계하고 테스트 세트에 대해서는 카테고리 서비스를 사용합니다.

- 문서가 충분히 높은 신뢰값을 갖고 올바른 카테고리에 할당되는지 점검하십시오. 범위는 0 - 1이며, 1은 문서가 이상적임을 의미합니다. 신뢰값을 설정할 수 있습니다. 기본값은 0.5입니다.

3 - 5 회 반복을 선택할 수 있습니다. 3회 반복이 기본값이며 분류법의 허와 실에 대한 의미있는 조망을 제공합니다. 5회를 선택하면, 전체 문서가 연계 및 테스트 세트 둘다에 속합니다.

평가 처리를 시작하려면 평가 시작을 누르십시오.

각 카테고리의 경우, 각 평가 반복 중 다음이 계산됩니다.

- *Correct* 문서. 평가 중 c로 할당된 카테고리 c의 연계 문서 수.
- *Outbound* 문서. 평가 중 다른 카테고리로 할당된 카테고리 c의 연계 문서 수.
- *Inbound* 문서. 평가 중 카테고리 c로 할당되었지만 다른 카테고리에서 기원하는 연계 문서의 수.
- *Unassigned* 문서. 평가 중 임의의 카테고리로 할당되지 않은 카테고리 c의 연계 문서 수. 이것은 카테고리에 할당되었지만 주어진 신뢰값 아래에 있는 문서를 포함할 수 있습니다.

평가 처리가 카탈로그 레벨에서 처리되는 방식에 대한 개요는 다음을 표시하며 표시됩니다.

- 평가가 실행 중인지 정지하였는지를 표시하는 평가 상태
- 마지막 평가 날짜를 나열하는 마지막 평가
- 완료된 평가 반복의 횟수
- 총 반복 수
- 인바운드 문서뿐 아니라 카테고리에 원래 할당된 올바른 문서의 백분율을 표시하는 평균 전체 정확도
- 올바른 카테고리에서 문서 백분율을 표시하는 평균 전체 재호출
- 제대로 할당된 문서 수를 나타내는 correct 문서 수
- Inbound 또는 outbound 문서 수인 잘못 위치한 문서 수
- Unassigned 문서 수

Precision과 recall은 지정된 신뢰값과 밀접하게 관련됩니다. 신뢰값이 낮으면, precision이 떨어지며 recall은 증가하며, 그 반대로 성립합니다. 높은 precision 값은 여러 연계 문서가 제대로 할당되었음을 의미하며, 반면 높은 recall 값은 대부분의 연계 문서가 카테고리로 할당되어, unassigned 문서가 없거나 극히 적음을 의미합니다.

## 평가 결과

상세 평가 결과는 평가 결과 탭을 눌러 얻습니다. 카탈로그를 선택하고 탭을 누른 경우, 전체 카테고리에 대한 결과가 표시되며, 카테고리를 선택한 경우 카테고리에 대한 결과가 표시됩니다.

그림 12에서는 카탈로그 레벨에서 평가 결과를 표시합니다.

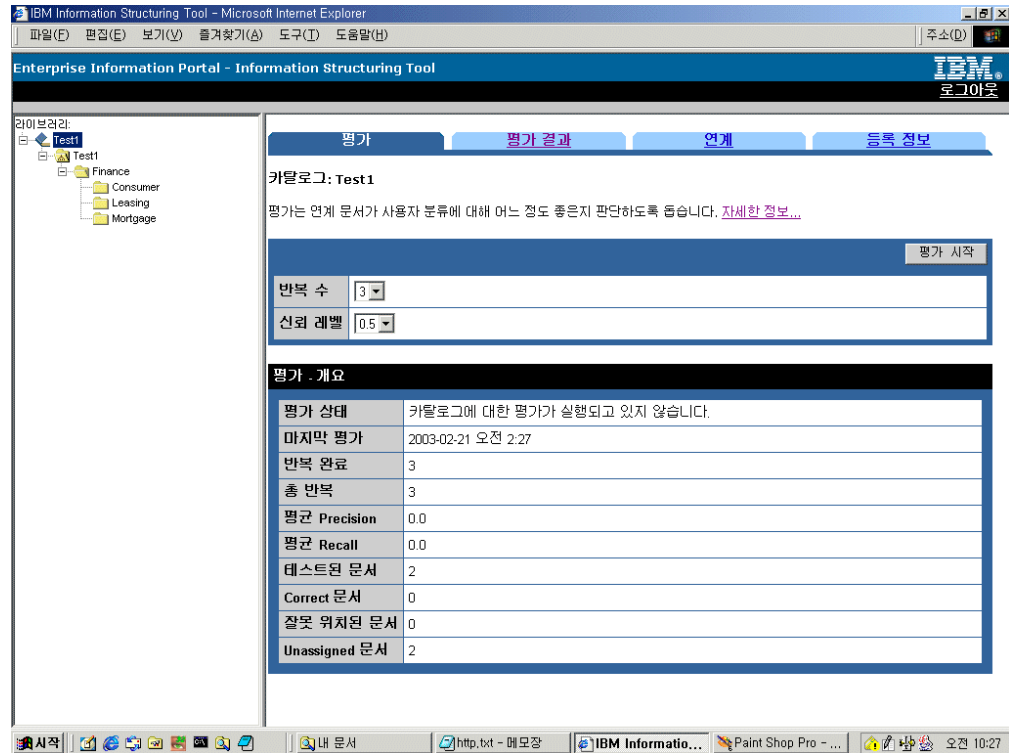


그림 12. 카탈로그 레벨에서 평가 결과

전체 카탈로그 결과로 시작하십시오. 적색(중요) 및 청색(덜 중요) 값은 카테고리나 연계 문서 중 하나에서 어떤 것이 적합하지 않음을 나타냅니다. 카테고리의 품질은 이의 연계 문서에 의해 정의되므로, 카테고리에서 (inbound)와 (outbound) 사이의 문서 이동을 연구하는 좋습니다.

카테고리와 연계 문서에 대한 변경은 precision과 recall이 균등하게 강조되는지 또는 precision만 강조되는지에 크게 의존합니다. Precision 값이 클수록 카테고리가 분류상의 다른 카테고리와 관련한 변별력은 더 좋아집니다. 반면 recall 값이 클수록 unassigned 연계 문서는 더 적어집니다.

평가 결과는 두 가지 레벨로 표시됩니다.

### 1. 카탈로그 레벨:

카탈로그의 각 카테고리:

- Precision과 recall 백분율

- Inbound 문서의 백분율
- Outbound 문서의 백분율
- Correct 문서의 백분율
- Unassigned 문서의 백분율

## 2. 카테고리 레벨:

Inbound와 outbound 문서 유형:

- 연계 문서와 원래 또는 대상 카테고리

Correct 및 Unassigned 문서 유형:

- 연계 문서

## 평가 결과 해석

다음 절에서는 평가 결과를 해석하는 방식을 제안하지만, 분류법은 전체로서 동작함을 항상 기억하십시오. 즉, 분류법의 한 절에 작성된 변경사항은 분류법의 어딘가에서 생성 되는 결과에 역효과를 가져올 수 있음을 의미합니다.

그림 13에서는 카테고리 레벨에서 평가 결과를 표시합니다

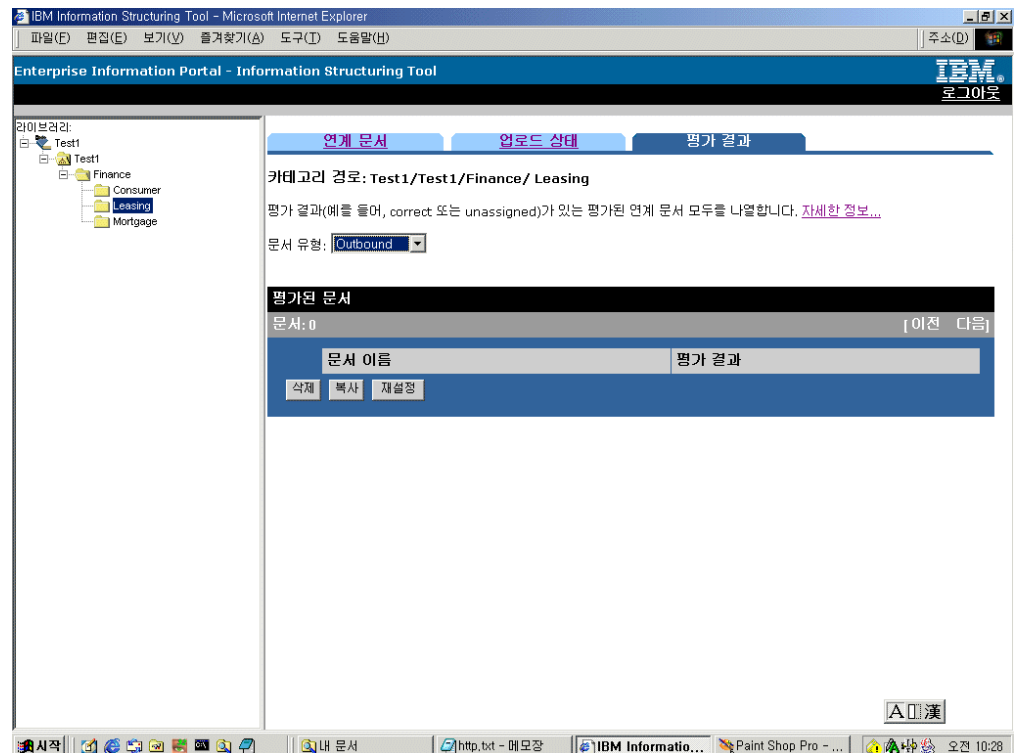


그림 13. 카테고리 레벨에서 평가 결과

카탈로그 레벨의 평가 결과로 시작하고, 잠정적으로 낮은 precision과 recall 값을 갖는 카테고리화 적색(더 중요) 또는 청색(덜 중요)으로 표시된 연계 문서를 선택하십시오. 차례로 각 카테고리에 대해(왼쪽 프레임에서 선택 가능), 다음을 점검하십시오.

- 다수의 **inbound** 문서:
  - 카테고리가 많은 다른 카테고리에서 문서를 얻는 경우:
    - 카테고리에 충분한 변별력이 없습니다. 이들 문서를 이 카테고리에 복사하십시오.
    - 더 적절한 연계 문서를 선택하여 이 카테고리의 변별력을 높이거나 카테고리를 서브카테고리로 나누십시오.
  - 카테고리가 하나 또는 두 개의 카테고리에서 문서를 얻는 경우:
    - 이들 문서를 이 카테고리에 복사하십시오.
    - 원래 카테고리를 점검하십시오. 이 카테고리를 보존하는 것이 적절합니까? 아니면 카테고리를 병합해야 합니까? 병합하려면, 전체 연계 문서를 보존하려는 카테고리로 복사한 후 다른 카테고리는 삭제하십시오.
- 다수의 **outbound** 문서:
  - 카테고리가 많은 다른 카테고리로 문서를 잃는 경우:
    - 카테고리에 충분한 변별력이 없습니다. 더 적합한 연계 문서를 선택하거나 카테고리 삭제를 고려하십시오.
  - 카테고리가 하나 또는 두 개의 카테고리로 문서를 잃는 경우:
    - 다른 카테고리로 문서를 복사하십시오.
    - 이 카테고리를 점검하십시오. 이 카테고리를 보존하는 것이 적절합니까? 아니면 카테고리를 병합해야 합니까?
- 다수의 **unassigned** 문서:
  - Unassigned 문서와 correct 문서와의 비교:
    - 크기
 

Unassigned 문서가 더 짧으면, 예를 들어 두 개의 문서를 결합하여 더 긴 새 문서로 만들어 correct 문서의 크기와 일치시키도록 고려하십시오.
    - 양식
 

양식이 다른 경우, 문서를 삭제하십시오.
    - 주제
 

주제가 약간 다르고, correct 문서의 주제와 여전히 관련되어 있는 경우, 이 주제를 다루는 추가적인 연계 문서를 찾아 동일한 카테고리로 업로드하거나 새로 작성된 카테고리로 업로드하십시오.

Recall과 precision 값이 양호한데도 동일한 문서가 반복적인 평가 중에 unassigned 상태로 남아 있으면 이들 문서를 삭제하십시오.

분류법을 변경한 경우, 예를 들어 카테고리를 삭제하거나 두 개의 카테고리를 병합한 경우 또는 여러 문서를 이동시킨 경우에만 평가 프로세스를 다시 시작해야 합니다. 여러 카테고리에 대해 새로운 연계 문서를 추가하거나 문서를 복사하는 등의 카테고리 레벨



에서의 미미한 변경에 대해서는 평가 프로세스를 재시작할 필요가 없습니다. 변경사항을 거꾸로 하려는 경우 실행 취소 기능이 없음을 유의하십시오.

관심있는 적색 또는 청색 값이 더 이상 없을 때 또는 90% 이상의 precision과 recall 수준을 획득하면 평가를 정지하십시오.

## 카탈로그 연계

분류법을 평가하고 결과에 만족한 후에는 메타데이터의 특수 유형, 즉 카테고리 모델을 생성하도록 연계 문서를 모두 사용하여 연계되어야 하며, 이는 후에 Information Mining 카테고리 서비스를 사용하여 새 문서를 구분하는 데 사용할 수 있습니다.

연계 단계를 시작하려면 연계가 필요한 카탈로그를 선택하고 연계 탭을 누른 다음 연계 시작 단추를 누르십시오.

선택한 카탈로그의 이름이 오른쪽 프레임의 맨 위에 표시됩니다. 또한 카탈로그의 연계 상태도 볼 수 있습니다. 연계 상태의 유형은 다음과 같습니다.

- 문서를 구분할 수 없습니다. 카탈로그를 연계해보고 카테고리라이저를 갱신하십시오. (이것은 또한 새로 작성되는 카탈로그에 대한 기본값입니다. 이 경우, 최종 연계 날짜가 비어 있습니다.) 카탈로그가 마지막으로 연계된 이후에 카테고리가 삭제되거나 이름이 바뀌었기 때문에 최종 연계 결과가 올바르지 않거나 카탈로그가 신규라서 전혀 연계되지 않았습니다. 새 문서를 구분하기 위해 카테고리 서비스 사용을 시도하면 오류가 발생합니다.
- 카탈로그의 연계 문서 세트가 변경되었습니다. 문서는 여전히 마지막 연계 결과에 따라 구분됩니다. 카탈로그를 연계해보고 카테고리라이저를 갱신하십시오.
- 연계가 카탈로그에서 실행 중입니다.
- 카탈로그가 최신이므로 연계가 필요하지 않습니다.

연계 처리를 중단하려면, 연계 정지 단추를 누르십시오. 연계가 정지되는 경우, 카탈로그 연계 처리가 재실행되어야 함에 유의하십시오.

연계 중인 동안 카탈로그로 새 연계 문서를 업로드할 수 없습니다.

---

## 성능 조정

카탈로그에서 레코드 작성 시, 텍스트 검색 가능한 속성에 대한 텍스트 색인(예를 들어, IKF\_CONTENT)은 매우 커지며 검색 성능은 상당히 저하됩니다. 기억영역을 최적화하고 성능을 증가시키려면, 주기적으로 텍스트 색인을 재구성하고, 색인을 크게 갱신한 이후에는 특히 재구성할 필요가 있습니다.

텍스트 색인의 재구성은 예를 들어, 밤과 같은 컴퓨터 오프피크 시간에 최적으로 시작됩니다. 색인 조직화를 실행하려면 다음과 같이 전환하십시오.

- Windows용: ... \ikf\IkfReorg.cmd

- AIX용: .../ikf/bin/IkfReorg.sh
- Solaris의 경우: .../ikf/bin/IkfReorg

매개변수는 IkfReorg <UserID><Password><DBName>을 포함합니다.

## IBM Web Crawler 사용

이 절에서는 IBM Web Crawler 기능을 구성하는 방법을 설명합니다. 기능 체크 박스를 선택한 경우 EIP 설치 프로그램으로 기능이 설치됩니다.

EIP 버전 8.2에는 검색된 파일, HTML 기반 문서, 구성 예제 및 지원 유틸리티에서 데이터를 추출하는 웹 크롤러, Lotus Notes 크롤러 및 요약기가 들어 있습니다. IBM Web Crawler(또한 GCS라고도 함)는 Java 버전 1.3 이상을 요구합니다.

IBM Web Crawler는 Java 기반 콘텐츠 크롤러이며 마이너입니다. 콘텐츠를 가리킬 때, 해당 콘텐츠를 획득하여 마이닝합니다.

IBM Web Crawler는 인트라넷, 엑스트라넷 또는 인터넷 웹에서, Lotus Notes 데이터 베이스에서 본래대로 또는 Domino를 통해 로컬 파일 시스템에서 콘텐츠를 검색할 수 있습니다. IBM Web Crawler는 새로운 프로토콜을 손쉽게 추가하기 위해 빌드됩니다. 콘텐츠는 임의 유형(예: HTML, Notes 첨부 및 멀티미디어)일 수 있습니다.

IBM Web Crawler는 여러 유형의 콘텐츠에서 메타데이터 및 텍스트를 마이닝할 수 있습니다. 예를 들어, HTML 콘텐츠는 다음에 대해 마이닝될 수 있습니다.

- URL
- 제목
- 본문
- 마지막 수정 시간
- 저자, 키워드, 설명 등과 같은 메타 태그

사용자는 주어진 유형의 콘텐츠에 대해 미리 정의된 마이너 세트에서 선택합니다. 콘텐츠 및/또는 마이닝 메타데이터는 로컬 디스크로 저장됩니다. IBM Web Crawler는 Network Solutions Outside In 기술을 사용하여 200가지보다 많은 콘텐츠 유형에서 텍스트를 추출하는 것으로, 검색 응용프로그램에서 사용하는 이상적인 협력입니다. IBM Web Crawler는 새로운 프로토콜을 손쉽게 추가하기 위해서도 빌드됩니다.

IBM Web Crawler는 Windows NT 4.0 및 Windows 2000 운영 체제에서 사용 가능합니다. 약 반 시간 정도 IBM Web Crawler를 설치, 구성 및 사용할 수 있습니다. 500 MHz PC에서 초당 약 10개의 파일로 콘텐츠를 얻고 마이닝합니다. 백만 오브젝트로 스케일하도록 테스트되었습니다(200000 Notes). 여러 사용자와 사용자당 여러 크롤/마이닝 구성을 지원하며, 사용자가 선호하는 자국어어를 지원할 수 있습니다.

## IBM Web Crawler 성능

설치 프로그램은 두 가지 파일을 설치합니다.

`x:/<install directory>/run`

일괄처리 파일과 샘플 구성을 위한 IBM Web Crawler.

`x:/<install directory>/notes-run`

주 일괄처리 파일 및 샘플 구성을 위한 웹 크롤러.

`x:/<install directory>/lib`

IBM Web Crawler .jar와 .zip 및 필터링 파일.

## 웹에 대한 IBM Web Crawler 구성 및 실행

이 절에서는 웹에 대한 IBM Web Crawler 구성 및 실행 방법에 대해 설명합니다. 웹에 대한 IBM Web Crawler는 HTTP, FTP, 뉴스 또는 파일 서버에 액세스하며 HTML 문서와 기타 오브젝트에 대한 요약을 작성합니다. 요약은 파일이며, 오브젝트 및 전체 텍스트를 포함하며 문서 또는 오브젝트당 한 개의 요약이 있습니다.

### 기본 구성

이 절에는 XML 형식으로 IBM Web Crawler 구성 파일을 편집할 수 있는 방법을 설명하는 지시사항이 들어 있습니다. 두 개의 샘플 구성이 사용자의 시작을 돕기 위해 제공됩니다.

- config-db2.xml 파일은 DB2 UDB와 함께 IBM Web Crawler를 사용하도록 제공됩니다.
- config-sample.xml 파일은 DB2 UDB 없이 IBM Web Crawler를 사용하도록 제공됩니다.

1. 명령 프롬프트를 여십시오.
2. IBM Web Crawler를 설치한 실행 서브디렉토리로 디렉토리를 변경하십시오. 예를 들어, IBM Web Crawler를 Windows 서버에 설치한 경우, `cd x:<cmdbroot>\gcs\run`을 입력하십시오. AIX에 IBM Web Crawler를 설치한 경우, `cd /usr/lpp/cmb/gcs`를 입력하십시오.

팁: 원래 파일의 사본을 보존하는 것은 매우 중요합니다. 파일 오류로 IBM Web Crawler가 중단될 수 있습니다. 편집 중인 동안 주의하십시오.

3. DB2 UDB 데이터베이스를 사용하여 IBM Web Crawler를 실행하려면(더 확장 가능하며, 더 느림), config-db2.xml 파일을 편집하십시오. 예를 들어, 명령 프롬프트에 `edit config-db2.xml`을 입력하십시오.
4. DB2 UDB 데이터베이스를 사용하지 않고 IBM Web Crawler를 실행하려면(덜 확장 가능하며, 더 빠름), config-sample.xml 파일을 편집하십시오. 예를 들어, 명령 프롬프트에 `config-sample.xml`을 입력하십시오.

데이터베이스 없이  $n$  URL의 크롤을 실행하려면, 검색된 URL 메타데이터를 포함하기 위해 대략  $n/1000$  MB의 RAM이 시스템에 필요합니다. 예를 들어, 512 MB의 RAM이 필요한 500,000 URL을 검색합니다. 이 메모리를 이용하려면, crawlweb.bat 파일을 편집하고 JVMXmx 값을 증가시키십시오.

## Information Mining에 대한 웹 크롤러 구성

Information Mining을 사용하여 웹 크롤러를 사용하려면, 즉, 크롤된 문서에 정보 마이닝을 적용할 수 있으려면, 위에 언급된 샘플 구성 파일 (config-sample.xml 및 config-db2.xml)에 주어진 설정과 다른 다음 특정 구성 파일을 필요로 합니다.

```
<globals ...
    max-urls="5000"
    temp-filepool-class="FullPathFilePool"
    summaries-dir="webspaces/ikf/disks/1/"
    summaries-filepool-class="DistributedEIPFilePool"
    ... >
...
</globals>
<group-list>
  <group ...>
    ...
    <summarizer-config>
      ...
      <resource-handler content-type="*"
        summarizable="EipHtmlSummarizable"
        summary-maker="EipHtmlRawSummaryMaker" />
      ...
    </summarizer-config>
    ...
  </group>
</group-list>
```

이 설정은 Information Mining 구성요소인 im-crawler-config-sample.xml에 대한 샘플 구성 파일의 부분으로 다음 디렉토리에 위치합니다.

Windows의 경우:

```
<CMBROOT>\samples\java\Beans\infomining\webcrawler\
```

UNIX(AIX 및 Solaris)의 경우:

```
<CMBROOT>/samples/java/Beans/infomining/webcrawler/
```

## IBM Web Crawler DB2 옵션 구성

DB2 옵션을 구성하려면 데이터베이스를 작성해야 합니다. 이것은 DB2 시스템 관리자 권한을 필요로 합니다. DB2 시스템 관리자 계정으로 교환할 필요가 있을 수 있습니다. DB2가 허용하는 대로 데이터베이스의 이름을 지정할 수 있지만, 데이터베이스 이름이 gcs가 아닌 경우, 웹 크롤러 구성 파일에서 dbname을 갱신해야 합니다.

데이터베이스 관리자 권한을 가진 경우, DB2 명령 프롬프트에서 이 명령을 실행하여 데이터베이스를 작성하십시오.

```
db -createdb <user><password>[database_name]
```

데이터베이스 이름을 지정하지 않은 경우, gcs가 사용됩니다. 데이터베이스가 작성되고 나면, 다음 명령을 발행하여 IBM Web Crawler 표를 추가하십시오.

```
db -createtables<user><password>[database_name]
```

IBM Web Crawler 데이터베이스 및 표 작성은 IBM Web Crawler와 함께 DB2를 사용하기 위해 완료되어야 합니다.

다음 구성 파일 설정(urlpool-config 절에서)은 새 데이터베이스(dbname:)를 사용하기 위해 필요합니다.

- 데이터베이스의 이름(위에서 작성한 대로), 예: gcs.
- 사용자 이름: 사용자 이름, 예: db2admin
- 암호: 사용자 암호, 예: db2admin.

데이터베이스, 사용자 이름 및 암호 등록 정보를 해당 값으로 설정하십시오. 캐시 크기나 드라이버를 변경하지 마십시오. 사용자 시스템의 크롤 범위를 설정하려면 파일 편집을 계속하십시오.

### 크롤 범위 설정

이 구성 파일 설정은 DB2를 사용할 지 여부에 관계 없이 크롤 범위를 설정하기 위해 필요합니다.

크롤러 config 절에서 다음 설정을 점검하고 사용자 필요에 맞게 이 항목을 설정하십시오.

#### seed list

하나 이상의 시작하는 절대 URL. URL은 사용 가능해야 합니다. 브라우저 사용을 검증하십시오. 예: http://www.<mysite>.com/

#### content-type-pattern-list

파일 확장자가 이 패턴과 일치하는 경우에만 페이지에 있는 크롤 URL. 예: htm\*

#### include-pattern-list

이 패턴과 일치하는 경우에만 페이지에 있는 크롤 URL. 예: <mysite>.com

또한 이들 항목을 설정할 수도 있습니다.

#### recursion-depth

임의의 시작점에서 검색할 링크의 최대 거리. 무제한 깊이의 경우 -1을 사용합니다.

### exclude-pattern-list

이 패턴과 일치하지 않는 경우에만 페이지에 있는 크롤 URL. 예: \*cgi-bin\*

### 시스템 등록 정보

un-socksified 시스템에서 방화벽을 통해 검색하려면, 이 파일에 socksProxy 값을 설정할 필요가 있습니다.

### IBM Web Crawler 시작

.xml 구성 파일을 편집한 경우, 저장하십시오.

IBM Web Crawler를 시작하려면, crawlweb 일괄처리 파일 및 구성 파일을 사용하십시오. 명령 프롬프트를 열고 다음을 입력하십시오.

- Windows의 경우: crawlweb.bat<CONFIGFILE>
- AIX의 경우: crawlweb.sh<CONFIGFILE>

DB2 UDB를 사용하여 실행하려면, crawlweb config-db2.xml을 입력하고 Enter를 누르십시오. DB2 UDB를 사용하지 않고 실행하려면, crawlweb config-sample.xml을 입력하고 Enter를 누르십시오.

팁: 정규 기준으로 크롤/요약 진행을 보고하도록 계획하십시오. 목표가 검색되면, summaries-dir에서 구성된 위치에 요약이 쓰여집니다. 기본값은 트리에서 .html 파일로 메타데이터 프로로그에다가 원래 오브젝트를 더하여 씁니다. 검색 중이나 검색 후에 자세한 정보를 위해 로그 파일을 조사할 수 있습니다.

### 고급 구성

이 시점에서 구성 옵션을 공부할 수 있습니다. 구성 예에 대해서는 119 페이지의 제 7 장 『IBM Web Crawler 샘플 파일』의 config-sample2.xml 파일을 참조하십시오. 샘플 구성은 다음과 같습니다.

- 크롤러 및 요약기 스템드
- 그래프 작성 모니터
- 로그 기록 옵션
- SOCKS
- Lotus Domino 크롤링
- 복수 콘텐츠 유형
- 추가 제외
- .pdf 파일 같은 오브젝트 요약 확보를 위한 InsoSummarizable 사용

구성 파일에서 사용 가능한 매개변수의 형식적 정의에 대해 config.dtd 파일을 참조하십시오. **권장사항:** 이 파일을 편집하지 마십시오. 파일 사본을 작성한 후, 사본 이름을 바꾸십시오.

## IBM Web Crawler 구성 파일

구성 파일은 수집할 웹 기반 자원 및 해당 자원 요약 방법에 대해 IBM Web Crawler에 알려주는 XML 파일입니다. 이 절에서는 config.xml에서 설정할 수 있는 각 요소 및 속성에 대해 설명합니다. Notes용 IBM Web Crawler 사용 방법에 대한 정보는 102 페이지의 『Notes용 IBM Web Crawler』를 참조하십시오.

IBM Web Crawler는 구성 파일의 콘텐츠가 gcs-config.dtd로 컴파일되었는지 점검합니다. 중요한 오류가 있는 경우, 예를 들어 검색할 URL이 없는 경우, IBM Web Crawler가 종료하고 오류 메시지를 인쇄합니다. 사소한 문제점(알 수 없는 속성 또는 값)이 발생하면 프로그램이 로그 파일에 경고를 기록하고 계속 진행됩니다. **권장사항:** 편집 전에 구성 파일을 백업하십시오. 파일 오류로 IBM Web Crawler가 중단될 수 있습니다.

샘플 구성 파일이 IBM Web Crawler와 함께 선적됩니다.

### <gcs-config>

gcs-config 파일은 **globals** 및 **group-list**의 두 절이 있습니다. gcs-config 파일의 예에 대해서는 119 페이지의 제 7 장 『IBM Web Crawler 샘플 파일』을 참조하십시오.

#### **globals**

globals 요소는 파일 시스템, 성능 및 네트워크 정보 같은 IBM Web Crawler 설정을 캡처합니다.

#### **group-list**

group-list 요소는 그룹의 크롤 및 요약을 구성합니다. 여기서 group은 비즈니스 또는 네트워크 도메인 같은 자원 세트입니다.

### <globals>

globals 요소는 IBM Web Crawler 글로벌 설정을 표시합니다. 설정은 global 속성 및 하위 요소로 인코딩됩니다.

다음 목록에서는 global 속성을 정의합니다. 글로벌 하위 요소의 정의에 대해서는 92 페이지의 『<logger-config>』의 내용을 참조하십시오.

#### **max-urls**

검색할 최대 URL 수. 이 수는 양수여야 하며 기본값은 100,000입니다.

#### **summaries-dir**

자원 요약을 작성하는 디렉토리. 기본적으로, summaries/ 디렉토리가 사용됩니다.

#### **summaries-filepool-class**

자원 요약에 사용되는 파일 풀 유형. 요약 파일의 이름 지정 방법 및 사용할 서브디렉토리 구조(있는 경우)를 판별합니다. 기본적으로, FullPathFilePool

이 사용되며, 이 값은 호스트의 디렉토리를 작성한 후, URL과 동일한 파일 이름 및 서브디렉토리 구조를 사용합니다.

### **num-crawlers**

사용할 크롤러 스레드 수. 이 수는 양수여야 하며 기본값은 20입니다.

### **num-summarizers**

사용할 요약기 스레드 수. 이 수는 양수여야 하며 기본값은 5입니다. 이 단계를 사용하여 num-crawler 및 num-summarizer를 구성하십시오.

1. 크롤러를 기계 속도 MHz/20으로 설정하십시오. 예를 들어, 600 MHz 시스템에서는 30을 사용하십시오.
2. 요약기 숫자를 1단계의 첫 번째 숫자로 설정하십시오. 예를 들어, 8로 설정하십시오.
3. Windows 작업 관리자 성능 패널을 주시하면서 시험 작업을 해보십시오. CPU가 100%의 2초 이상으로 진행되는 경우, 1단계로 돌아가 더 작은 숫자로 설정하십시오. 예를 들어, CPU 사용이 100%가 되지 않을 때까지는 나머지 시간의 세 번째 숫자로 설정하십시오.

시험 작업 동안 텍스트 모니터에 Summarizer: ToDo의 수가 구성된 요약기의 수보다 적은 것으로 계속 보고되는 경우, 3단계를 위반하지 않는 한 요약기의 수를 줄이고(적을수록 좋음) 크롤러의 수를 늘리십시오(많을수록 좋음). 성능을 최대화하려면 가능한 경우 최고속 네트워크를 사용하여 개별 디스크에 요약, DB, 임시 공간 및 로그를 배치하십시오.

### **text-monitor**

on으로 설정하면 text-monitor가 IBM Web Crawler 상태를 5초 간격으로 standard out으로 인쇄합니다. 10진수 값을 입력하면 텍스트 모니터가 텍스트 출력의 화면 갱신 간격(초 단위)을 설정합니다. 기본 설정은 off입니다.

### **graph-monitor**

on으로 설정하면 그래프 GUI를 사용하여 IBM Web Crawler 상태를 표시하도록 graph-monitor가 IBM Web Crawler에 알립니다. 10진수 값을 입력하면 그래프 모니터가 모니터 GUI의 화면 갱신 간격(초 단위)을 설정합니다. 기본 설정은 off입니다.

### **log-file**

사용할 주요 로그 파일을 지정합니다. 기본값은 log/log.txt입니다.

팁: logger-config 요소에 추가 로거 정보를 지정할 수 있습니다.

### **log-priority**

기본 로그 우선순위를 설정합니다. info, warn 또는 error 값을 입력하십시오. 기본값은 warn입니다.

팁: logger-config 요소에 추가 로거 정보를 지정할 수 있습니다.



**temp-dir**

임시 파일을 작성할 디렉토리. **팁:** IBM Web Crawler를 사용하여 이 디렉토리의 모든 파일을 삭제할 수 있습니다. `x:/temp/gcs` 기본 설정에서 이 값을 변경해서는 안됩니다.

**temp-filepool-class**

임시 파일에 사용할 파일 풀 유형. **권장사항:** TempFilePool 기본 설정에서 이 값을 변경하지 마십시오.

**content-dir**

IBM Web Crawler가 콘텐츠 파일을 작성하는 디렉토리. 보통 `content-dir`은 `temp-dir`과 동일합니다.

**content-filepool-class**

콘텐츠 파일에 사용할 파일 풀 유형. 일반적으로 이것은 `temp-filepool-class`와 동일합니다.

**how-often-to-gc**

가비지 콜렉션 요청 간에 검색할 URL 수. **권장사항:** 정수  $\geq 50$ 을 정의하십시오. 기본 설정은 100입니다.

**max-resource-pool-size**

요약을 기다리는 최대 자원 대기열 크기. **권장사항:** 정수  $\geq 10$ 을 정의하십시오. 기본 설정은 요약기당 10개의 대기 자원을 허용합니다.

**connect-timeout**

네트워크 연결 시간종료 전의 대기 시간(밀리초 단위)을 정의합니다. 기본값은 4,000입니다. 올바른 범위는 1,000 - 60,000입니다.

**read-timeout**

네트워크 읽기 시간종료 전의 대기 시간(밀리초 단위)을 정의합니다. 기본값은 6000입니다. 올바른 범위는 1,000 - 60,000입니다.

**cookies**

HTTP 헤더의 쿠키 점검 및 해당 쿠키의 데이터베이스 저장 여부를 정의합니다. 기본 설정은 `off`입니다. 이 값을 `on`으로 설정하면 쿠키를 사용할 수 있습니다.

**locale** 요약 및 로그 기록에 사용할 언어를 정의합니다. 기본값은 `en_US`입니다.

글로벌 하위 요소는 `logger-configs`, `urlpool-config` 및 `system-properties`를 포함합니다.

## <logger-config>

logger-config 파일에는 기록 대상에 대한 고급 제어, 형식화 방법 및 로그 파일 작성 장소가 포함됩니다. default log-file 및 log-priority는 globals 속성으로 지정됩니다. 로그 기록에 대한 자세한 정보는 99 페이지의 『IBM Web Crawler의 로그 기록』을 참조하십시오.

### category

예를 들어, gcs.crawler와 같은 구성되는 로거 카테고리. 지정하지 않은 경우, 기본 로거가 구성됩니다. 특정 카테고리에 대한 설정은 모든 하위 카테고리에 영향을 줍니다.

### priority

메시지를 기록해야 하는 최소 우선순위. 지정하지 않은 경우, 이 로거가 상위 카테고리에서 우선순위를 가져옵니다. (궁극적으로는 globals default log-priority에서 가져옵니다.)

### log-file

로그 파일 작성 장소를 정의합니다. '+'로 시작하는 경우, 다른(상위) 로그 파일과 이 로그 파일을 사용합니다. 지정하지 않은 경우, 상위 로그 파일을 사용합니다. (궁극적으로 globals default log-file을 사용합니다.)

팁: 서로 다른 로거를 겹쳐쓸 수 있으므로 여러 로거에 동일한 log-file을 지정하지 마십시오.

### log-layout

로그 파일에 인쇄될 각 메시지에 사용되는 배치를 정의합니다.

## <urlpool-config>

urlpool-config 파일은 URL이 저장된 IBM Web Crawler의 구성요소를 구성합니다. URL 풀에는 여러 옵션이 있습니다. 메모리에 풀을 저장하거나 DB2를 사용하거나 또는 각 URL 정보만큼 저장하지 않는 특정 소량 메모리 버전을 사용할 수 있습니다. urlpool-config 요소를 지정하지 않는 경우, URL 풀이 메모리에 저장됩니다. urlpool-config에는 데이터베이스 정보 같은 특정 내용을 지정하는 하위 urlpool-param 요소가 있을 수 있습니다.

### urlcontainer-class

사용할 URL 컨테이너 유형. 다음과 같이 지정하십시오.

- DB2 UDB로 검색하려면 DB2URLContainer
- DB2 UDB(기본값) 없이 검색하려면 MemoryURLContainer
- DB2 UDB 없이 검색하고 여분의 메모리를 사용하려면(일부 참조 URL 및 기타 정보 저장) BigMemoryURLContainer

### urlcollection-class

사용할 URL 컬렉션 유형. 다음과 같이 지정하십시오.

- DB2 UDB로 검색하려면 DB2URLCollection
- DB2 UDB(기본값) 없이 검색하려면 MemoryURLCollection
- DB2 UDB 없이 검색하고 여분의 메모리를 사용하려면(일부 참조 URL 및 기타 정보 저장) BigMemoryURLCollection

### <urlpool-param>

urlcollection-class로 매개변수를 전달하기 위해 사용됩니다. 예에 대해서는 119 페이지의 제 7 장 『IBM Web Crawler 샘플 파일』에서 DB2 UDB를 사용하여 샘플 구성의 데이터베이스 연결 정보를 참조하십시오.

**name** 매개변수 이름을 정의합니다.

**value** 매개변수 값을 정의합니다.

팁: 이 매개변수의 오류 점검이 없으므로 매개변수 사용에 유의하십시오.

### <system-properties>

시스템 등록 정보는 시스템 등록 정보 설정 목록을 표시합니다.

#### <property>

예를 들어, 고급 구성 샘플의 SOCKS 게이트웨이 사용에 대해서는 구성을 참조하십시오.

**name** 매개변수 이름

**value** 매개변수 값

또다른 방법으로, 다음을 사용하여 PROXY 게이트웨이를 통해 외부 서버에 대한 IBM Web Crawler 액세스를 구성할 수도 있습니다.

```
<system-properties>
  <property name="proxySet" value="true"/>
  <property name="proxyHost" value="proxy.hostname"/>
  <property name="proxyPort" value="80"/>
</system-properties>
```

팁: 이 매개변수에 대한 오류 점검이 없으므로 사용에 유의하십시오.

### <group-list>

group-list는 하나 이상의 그룹 요소의 목록입니다.

#### <group>

그룹 요소는 유사한 방법으로 검색 및 요약될 단일 자원 그룹을 표시합니다. 각 그룹에는 고유 이름 속성 및 검색할 대상을 알려줄 최소한 하나의 crawler-config 하위 요소가 있어야 합니다. 기본 요약기를 사용하지 않으려는 경우, 그룹에는 하위 summarizer-config 요소가 있을 수 있습니다. **팁:** 그룹이 겹쳐지면(두 개 이상의 그룹

에 동일한 URL이 있는 경우) 예상치 못한 결과가 발생할 수도 있습니다. 복수의 그룹화된 URL은 해당 URL이 발견된 첫 번째 그룹과만 연관됩니다.

**name** 이 그룹의 고유 이름(필수).

### <crawler-config>

이 규칙을 사용하여 검색 범위를 설정하십시오. 크롤러는 seed-list의 각 URL을 검색하고 해당 콘텐츠의 URL을 구문 분석한 후, 다음과 일치하는 해당 URL을 to-be-crawled 목록에 추가합니다.

- content-type-pattern-list의 최소 한 가지 규칙
- include-pattern-list의 최소 한 가지 규칙
- exclude-pattern-list에 해당하는 규칙이 없음

또한 crawler-config에는 단일 속성 **recursion-depth**가 필요합니다. recursion-depth는 각 seed에서 몇 개의 링크 반경을 크롤러가 통과해야 하는지 정의합니다. 기본값은 -1이며 무한 깊이를 표시합니다.

### <seed-list>

인증 정보를 포함할 수 있는 URL seed 목록입니다.

### <seed>

seed는 크롤러를 시작하기 위한 URL seed를 표시하며 URL 속성 및 인증 정보가 포함됩니다. 각 seed는 절대 URL(예: http://<your.server>.com/)이어야 합니다. 재지정했거나 사용할 수 없거나 텍스트가 아닌 페이지로 지정된 seed는 사용하지 마십시오. seed를 포함하도록 편집한 페이지로 지정하는 것이 유용합니다. 이러한 페이지는 갱신, 검토 및 브라우저로 테스트가 용이합니다.

URL 검색을 시작하려는 seed URL.

### <authentication>

rfc2617에 정의된 대로 기본 인증이 보호하는 seed URL에 전송된 선택적 인증

#### username

인증에 사용되는 사용자 이름

#### password

인증에 사용되는 암호

예를 들어, 다음과 같습니다.

```
<seed url="http://your.server.com/"><authentication username="me"
password="mine"/></seed>
```

### <content-type-pattern list>

검색될 콘텐츠 유형을 포함하기 위한 패턴 목록이며 파일 확장자로 식별합니다. 이 목록의 url-name-pattern과 일치하는 URL 파일 확장자(.html, .gif, .doc 등)는 이 테스트를 패스합니다. 확장자가 없는 URL은 기본적으로 테스트를 패스합니다. content-type-pattern-list를 지정하지 않았거나 비어 있는 경우, 파일 확장자가 없는 URL만을 승인합니다.

### <include-pattern list>

이것은 예를 들어 서버나 도메인 이름에 의해 검색되는 URL을 포함하는 패턴 목록입니다. 이 목록의 url-obj-pattern, url-regex-pattern, url-name-pattern 또는 url-predicate-pattern과 일치하는 URL은 테스트를 패스합니다. include-pattern-list를 지정하지 않았거나 비어 있는 경우, 모든 URL이 승인됩니다.

### <exclude-pattern list>

검색될 URL을 제외하기 위한 패턴 목록입니다. 이 목록의 url-obj-pattern, url-regex-pattern, url-name-pattern 또는 url-predicate-pattern과 일치하는 URL은 검색되지 않습니다. exclude-pattern-list를 지정하지 않았거나 비어 있는 경우, 어떤 URL도 거부되지 않습니다.

### <url-obj-pattern>

URL(프로토콜, 호스트 등)의 다른 부분과 와일드 카드를 일치시키는 패턴입니다. exclude-pattern-list 및 include-pattern-list에서 모두 사용될 수 있습니다. 각 부분의 패턴에는 시작과 끝에 '\*' 와일드 카드가 있으며 이는 모든 표기와 일치합니다. 그러나 패턴 중간에는 와일드 카드가 올 수 없습니다. 대소문자를 구분합니다. 생략된 URL 부분 패턴은 자동으로 일치합니다.

다음 목록에는 Java 및 IBM Web Crawler가 <http://www.ibm.com/products/index.html?query#ref> URL을 구분하는 예가 들어 있습니다.

- 프로토콜은 http임
- 호스트는 www.ibm.com임
- 포트는 -1임(지정하지 않음)
- 파일은 /products/index.html?query임
- 경로는 /products/index.html임
- 디렉토리는 /products/임
- 파일 이름은 index.html임
- 확장자는 .html임
- 조회는 query임
- 참조는 ref임

다음 목록에서는 url-obj-pattern의 각 요소의 세부사항에 대해 설명합니다.

#### 프로토콜

URL 프로토콜이 일치해야 하는 와일드 카드 패턴(예: http).

**호스트** URL 호스트가 일치해야 하는 와일드 카드 패턴(예: \*.ibm.com).

**포트** URL 포트가 일치해야 하는 와일드 카드 패턴(예: 80).

**파일** URL 파일이 일치해야 하는 와일드 카드 패턴(예: \*.htm\*). URL의 파일 부분은 호스트 다음의 첫 번째 슬래시로 시작하고 참조가 아닌 조회를 포함할 수 있습니다. http://www.ibm.com/products/index.html?query#ref의 파일 부분은 /products/index.html?query입니다.

**경로** URL 경로가 일치해야 하는 와일드 카드 패턴(예: \*.htm\*). URL의 경로 부분은 호스트 다음의 첫 번째 슬래시로 시작하고 조회 또는 참조를 포함하지 않습니다. 예에서 http://www.ibm.com/products/index.html?query#ref의 경로 부분은 /products/index.html입니다.

#### 디렉토리

URL의 디렉토리가 일치해야 하는 와일드 카드 패턴(예: /products/). 디렉토리는 첫 번째 슬래시로 시작하여 마지막 슬래시로 끝나는 경로 부분입니다. 예에서 http://www.ibm.com/products/index.html?query#ref의 디렉토리 부분은 /products/입니다. 조회 또는 참조는 포함하지 않습니다. 그러나 마지막 슬래시가 없는 잘못된 URL(예: http://www.ibm.com/products)의 경우에는 디렉토리가 정확히 일치하지 않습니다. 잘못된 URL 예에서 URL 디렉토리는 /입니다.

#### 파일 이름

URL의 파일 이름이 일치해야 하는 와일드 카드 패턴. 예에서 index.html입니다. 파일 이름은 마지막 슬래시 다음의 경로 부분입니다. 예에서 http://www.ibm.com/products/index.html?query#ref의 파일 이름은 index.html입니다. 조회 또는 참조는 포함하지 않습니다.

**확장자** URL 파일 확장자가 일치해야 하는 와일드 카드 패턴(예: htm\*). 가능하면 content-type-pattern-list를 사용하는 것이 바람직합니다.

**조회** URL 조회가 일치해야 하는 와일드카드 패턴.

**참조** URL 참조가 일치해야 하는 와일드 카드 패턴(HTTP에는 사용되지 않음). 예를 들어, <url-obj-pattern host="\*.ibm.com"/>은 모든 IBM 사이트의 HTML 페이지와 일치합니다.

#### <url-regex-pattern>

url-regex-pattern은 일반 표현식을 사용하는 URL과 일치하는 패턴입니다. exclude-pattern-list 또는 include-pattern-list에서 모두 사용할 수 있습니다. com.ibm.regex 패키지(regex4j)를 사용하며 대부분의 Perl 5 일반 표현식 기능을 갖

고 있습니다. 두 가지 일반 표현식을 가질 수 있습니다. 한 가지는 URL이 일치해야 하는 표현식이고 다른 한 가지는 URL이 일치해서는 안되는 표현식입니다. 대소문자 비구분을 위한 `i`와 같은 다른 옵션을 지정할 수 있습니다. 자세한 내용은 `Regex4j` 일반 표현식을 참조하십시오.

**match** URL이 일치해야 하는 Perl 5 양식의 일반 표현식.

**no-match**

URL이 일치해서는 안되는 Perl 5 양식의 일반 표현식.

**options**

대소문자 비구분을 위한 `i`와 같은 선택적 수정자.

예를 들어, `<url-regex-pattern match="^http://www\.ibm\.com/.*\.`  
`html?$/>`는 기본 IBM 웹 사이트의 HTML 페이지와 일치합니다.

### **<url-name-pattern>**

전체 URL 또는 URL 파일 확장자와 일치하는 와일드 카드가 있는 단일 패턴입니다. `content-type-pattern-list`, `include-pattern-list` 및 `exclude-pattern-list`에서 사용할 수 있습니다. 패턴 문자열의 시작과 끝에 모든 표기와 일치하는 '\*'가 있을 수 있습니다. 그러나 이름 중간에는 와일드 카드가 올 수 없습니다. 대소문자를 구분합니다.

예를 들어, `<url-name-pattern name="*.ibm.com/*"/>`는 IBM 사이트의 모든 파일과 일치하지만 `<url-name-pattern name="*.ibm.com/*.html"/>`은 중간에 와일드 카드가 있어 올바르지 않습니다.

이름 URL 문자열이 일치해야 하는 와일드 카드 패턴. 시작과 끝에 선택적인 '\*' 와일드 카드가 있습니다.

### **<url-predicate-pattern>**

이 패턴은 URL과 일치하는 Java `UnaryPredicate` 클래스를 로드합니다. `exclude-pattern-list` 또는 `include-pattern-list`에서 모두 사용할 수 있습니다. 클래스에는 URL과 술어가 일치하는 경우에 `true`를 리턴하는 `public boolean execute(URL url)` 메소드가 있어야 합니다.

클래스 완전한 `UnaryPredicate` 클래스 이름.

### **<summarizer-config>**

하위 `resource-handlers` 목록이 있는 요약기의 구성입니다. 대개 그룹당 `summarizer-config` 하나만이 있을 수 있습니다.

### **<resource-handler>**

해당 콘텐츠 유형(`text/html`) 또는 파일 이름 확장자(`htm`)에 기초하여 자원(예: 웹 페이지 또는 뉴스 그룹 기사)에 작성되는 요약 유형을 결정합니다. 자원을 요약할 준비가

되면 IBM Web Crawler가 resource-handler를 순서대로 점검하여 콘텐츠 유형 또는 파일 확장자와 일치하는 첫 번째 resource-handler를 사용합니다. 일치하는 resource-handler가 없는 경우에는 기본적으로 Copy2RdfSummarizable 및 Copy2RdfSummaryMaker를 사용합니다. resource-handler를 추가하여 목록 마지막의 content-type 또는 file-extension없이 가져올 수 있습니다.

또한 resource-handler에는 특정 매개변수를 해당 SummaryMaker 클래스로 전달하는 summarizer-param 하위가 있을 수 있습니다.

#### **content-type**

자원의 content-type이 일치해야 하는 와일드 카드 패턴(예: \*htm\*).

#### **file-extension**

자원 파일 확장자가 일치해야 하는 와일드 카드 패턴(예: htm\*).

#### **summarizable**

자원 Summarizable 클래스 이름(예: HtmlRawSummarizable).

#### **summary-maker**

자원 SummaryMaker 클래스 이름(예: HtmlRawSummaryMaker).

content-type 및 file-extension 패턴은 와일드 카드를 허용합니다. 패턴에는 패턴 문자열 시작과 끝에 모든 표기와 일치하는 \*가 있을 수 있습니다. 그러나 중간에는 와일드 카드가 올 수 없습니다. 대소문자를 구분합니다.

content-type 및 file-extension 패턴이 일치하면 resource-handler도 일치하고 지정하지 않은 패턴은 언제나 일치합니다. 따라서

```
<resource-handler content-type="*htm*"
summarizable="*HtmlRawSummarizable" summary-maker=
"HtmlRawSummaryMaker"/>
```

는 파일 확장자에 관계없이 콘텐츠 유형이 text/html인 모든 파일과 일치합니다.

summarizable 및 summary-maker에 대해 com.ibm.IBM Web Crawler.summarizer.resource 패키지에 클래스가 있는 경우, 클래스 전체 경로를 지정하지 않아도 됩니다.

#### **<summarizer-param>**

SummaryMaker 클래스로 전달되는 특수 매개변수입니다. 사용법은 해당 클래스에 고유합니다.

**name** 매개변수 이름

**value** 매개변수 값.

팁: 이 매개변수에 대한 오류 점검이 없을 수 있으므로, 사용에 유의하십시오.



## IBM Web Crawler의 로그 기록

여기서는 IBM Web Crawler의 로그 기록 기능에 대해 설명합니다.

IBM Web Crawler에서는 기록 대상, 기록 장소 및 형식화 방법을 제공합니다. 예를 들어, 한 파일에 검색된 각 페이지의 응답 코드, 다른 파일에 대한 IBM Web Crawler 상태(검색된 URL 수, 작동 중인 스레드 수 등), 세 번째 파일에 요약된 URL, 네 번째 파일에 대한 모든 IBM Web Crawler 경고 및 디버깅을 위한 다른 파일로의 넷 유틸리티 패키지의 모든 로그 메시지 작성을 선택할 수 있습니다.

로그 분석 파일의 샘플에 대해서는 121 페이지의 『IBM Web Crawler 로그 분석 파일 예』를 참조하십시오.

### 로그 사용

로그 기록은 network/web/crawl/summarization 계정, 다른 응용프로그램 구성요소와의 통신 및 IBM Web Crawler 디버깅에 유용합니다.

크롤 및 마이닝 계정은 예를 들어, 잘못 구성된 서버, 누락 페이지 및 콘텐츠 유형당 오브젝트 수 등의 여러 흥미로운 기능을 표시합니다. loganalysis.pl Perl 스크립트는 로그 요약 계정의 샘플을 제공합니다. 응용프로그램에는 콘텐츠 제거 시기 같은 IBM Web Crawler 정보가 필요합니다.

### 로거 구성

IBM Web Crawler config.xml 파일에 하나 이상의 로거 구성을 지정할 수 있습니다. globals 요소의 log-priority 및 log-file 속성은 기본 로그 기록 방침을 설정합니다.

로그 기록 방침을 확장하려면 globals 하위 요소로서 logger-config 명령문을 작성하십시오. 각 명령문은 IBM Web Crawler 로그 메시지의 서브세트를 선택하여 해당 메시지를 특정 파일에 경로지정한 후, 특정 형식을 사용하여 작성합니다. 우선순위 및 카테고리 속성을 사용하여 기록된 메시지의 서브세트를 선택합니다. 올바른 우선순위 값은 trace, debug, info 및 warn입니다(대소문자 비구분).

- 우선순위 값을 설정하면 가장 상세한 추적을 통해 로거의 상세도를 판별할 수 있습니다.
- trace 및 debug는 유지보수 레벨입니다. 메시지는 영어로 하드 코드화됩니다.
- info 및 warn은 자국어로 지원되는 사용자 레벨입니다.
- info는 많은 메시지를 작성합니다. warn 우선순위를 지정하여 메시지 출력을 줄이십시오.

### 구성 예 로그 기록

date/time/thread 정보 없이 from/to 하이퍼링크를 log/fromto.txt 파일에 기록

```
<logger-config category="gcs.url.fromto" priority="info"
log-layout="%m\n" log-file="log/fromto.txt"/>
```

요약된 오브젝트를 **log/resources.txt** 파일에 기록

```
logger-config category="gcs.summaries.list.resource" priority="info"
log-file="log/resources.txt"/
```

건너뛴 **URL** 및 해당 이유 기록

```
<logger-config category="gcs.url.skipped"
priority="info" log-file="log/urls_skipped.txt"/>
```

특별히 처리한 **HTTP** 응답 코드 기록

```
<logger-config category="gcs.http.302"
priority="info" log-file="log/urls_redirected.txt"/>

<logger-config category="gcs.http.404"
priority="info" log-file="log/urls_not_found.txt"/>
```

해당 우선순위를 포함한 요약기 카테고리의 모든 메시지 기록

```
<logger-config category="gcs.summarizer"
priority="TRACE" log-file="summarizer_trace.txt"
log-layout="%d: %t: %c: %p: %m\n"/>
```

## 문제점 해결

문제점이 발생한 경우, 우선 점검해야 할 사항은 다음과 같습니다.

**seed** 목록 페이지에 접근할 수 있습니까?

페이지가 존재해야 하며(재지정한 seed는 포함) 시스템에서(SOCKS가 사용 중인 경우, SOCKS를 통해) 접근할 수 있어야 합니다.

**seed** 목록 페이지는 일반 **HTML**입니까?

프레임, 플래시, 자바스크립트 및 기타 유사 항목이 seed로 바람직하지 않습니다. 일반 **HTML** 페이지를 선택하십시오.

**DB2 UDB**를 사용하고 있는 경우, 이미 검색했습니까?

DB2 UDB는 검색한 사항을 추적합니다. 모든 페이지를 처리한 경우, DB2 UDB는 응답이 없습니다. db -emptytables 명령을 사용하여 새 크롤을 시작하십시오.

**DB2 UDB**를 사용하고 있는 경우, 구성 파일 데이터베이스 액세스 항목이 올바른지?

데이터베이스 연결에 실패하면 크롤도 실패합니다.

구성 파일 편집을 주의깊게 점검했습니까?

오류로 인해 IBM Web Crawler가 중단될 수 있습니다. 제한적인 max-urls 또는 recursion-depth 값을 점검했습니까?

문제점이 계속 발생합니까?

구성 파일을 편집하여 로그 우선순위를 "debug"로 변경하십시오. 웹 크롤러를 다시 시작하여 중단한 후, 로그 파일을 점검하십시오.

## 요약기 선택

요약기의 목적은 자원(예: 웹 페이지) 또는 호스트(예: 웹 서버)를 사용하여, 원하는 정보를 포함하고 있는 파일을 사용이 용이한 형식으로 작성하는 것입니다.

IBM Web Crawler에는 다른 콘텐츠 유형을 처리하고 다른 종류의 데이터를 자원에서 추출하여 다른 파일 형식으로 출력하는 다양한 요약기가 포함됩니다. 이 섹션에서는 사용 가능한 요약기의 기능 및 요구사항에 대해 설명합니다. 원하는 기능을 갖고 있는 요약기가 없는 경우에는 사용자 고유의 요약기를 작성할 수도 있습니다.

IBM Web Crawler에는 두 가지 종류의 요약기가 있습니다. 자원 요약기는 웹 페이지 같은 단일 자원 요약을 작성하고 호스트 요약기는 웹 서버 같은 호스트 요약을 작성합니다. 현재는 자원 요약기만을 구성할 수 있습니다.

### 자원 요약기 선택

자원 요약기 선택 또는 작성 시 고려해야 할 사항은 다음과 같습니다.

- 입력 형식은? HTML 웹 페이지, PDF, WordPro 문서 및 XML 파일
- 추출하려는 메타데이터는? HTTP 헤더, 제목, 주석 링크 및 본문 텍스트
- 원하는 출력 형식은? XML, HTML 및 RDF

IBM Web Crawler config 파일의 resource-handler 요소를 사용하여 특정 자원 유형에 사용할 요약기를 지정합니다. 먼저 요약기에 사용할 콘텐츠 유형과 파일 이름 확장자를 지정합니다. 그런 다음, 해당 작업을 수행할 Java Summarizable 및 SummaryMaker 클래스를 지정합니다. summarizable 클래스는 요약할 자원을 표시하고 summary-maker 클래스는 작성할 요약 유형을 표시합니다.

### 기본 요약기(Copy + RDF 요약기)

copy + RDF 요약기는 다른 요약기에 의한 처리에 content-type이 명시적으로 구성되지 않은 임의의 오브젝트에 적용되는 요약기입니다. 이 요약기는 모든 자원 유형에 사용되며 두 가지 파일을 작성합니다. 첫 번째 파일은 정확한 원래 자원 사본이며, 두 번째 파일은 원래 URL, 저장 파일의 파일 이름 및 HTTP 헤더 정보를 포함하고 있는 RDF 요약입니다. 또한 DefaultSummarizable 및 Copy2RdfSummaryMaker를 사용하여 이 파일을 명시적으로 구성할 수 있습니다.

### HTML 페이지용 요약기(raw HTML 요약기)

HTML 자원의 경우, 원시 HTML 요약기는 URL 및 HTTP 헤더 정보가 맨 위의 주석에 포함된 파일 사본을 간단히 작성합니다. HtmlSummarizable 및 HtmlRawSummaryMaker를 사용하여 원시 HTML 요약기를 구성합니다.

```
<resource-handler content-type="*htm*"
    summarizable="HtmlSummarizable"
    summary-maker="HtmlRawSummaryMaker" />
```

## HTML 페이지용 요약기(EIP HTML 요약기)

HTML 자원의 경우, 원시 HTML 요약기는 URL 및 HTTP 헤더 정보가 맨 위의 주석에 포함된 파일 사본을 간단히 작성합니다. EIPHtmlSummarizable 및 EIPHtmlRawSummaryMaker를 사용하여 이 요약기를 구성합니다.

```
<resource-handler content-type="*htm*"
                  summarizable="EIPHtmlSummarizable"
                  summary-maker="EIPHtmlRawSummaryMaker" />
```

## HTML 페이지용 요약기(HTML 요약기 쓰기 없음)

이 요약기는 HTML을 검색하고 링크를 따르지만 파일의 요약을 작성하지는 않습니다. 예를 들어, 이 요약기는 사이트의 모든 PDF 파일을 검색하지만(INSO 대 XML 요약기 사용) HTML 파일은 저장하지 않으려 할 때 유용합니다. HtmlSummarizable 및 NoWriteSummaryMaker를 사용하여 이 요약기를 구성할 수 있습니다.

```
<resource-handler content-type="*htm*"
                  summarizable="InsoSummarizable"
                  summary-maker="InsoSummaryMaker" />
```

## 다른 콘텐츠 유형의 요약기(INSO 대 XML 요약기)

이 요약기는 Microsoft Word 문서, PDF 파일 및 파워 포인트 프리젠테이션 같은 200개가 넘는 자원 유형의 XML 요약을 작성합니다. 이 요약기에는 Network Solutions INSO 필터(INSO 사용권 필요)가 추출하는 본문 텍스트 및 일부 메타 정보가 들어 있습니다. InsoSummarizable 및 InsoSummaryMaker를 사용하여 이 요약기를 구성합니다.

```
<resource-handler content-type="pdf"
                  summarizable="InsoSummarizable"
                  summary-maker="InsoSummaryMaker" />
```

## 다른 요약기

다른 자원 유형을 요약하고 다른 데이터를 마이닝하거나 다른 형식으로 출력해야 하는 경우에는 IBM에 문의하거나 사용자 고유의 요약기를 작성하십시오.

## Notes용 IBM Web Crawler

이 절에서는 Notes용 IBM Web Crawler에 대한 구성 및 실행 방법에 대해 설명합니다. Notes용 IBM Web Crawler는 Notes 데이터베이스에 액세스하여 Notes 문서 및 첨부 요약을 작성합니다. 요약은 XML 형식 파일로서, 오브젝트 및 전체 텍스트를 포함하며 문서 또는 첨부당 한 개의 요약이 있습니다.

### 전제조건

다음 전제조건은 Notes에 대한 IBM Web Crawler를 실행하기 전에 필요합니다.

- Lotus Notes 버전 5.0.5 이상
- 자동 압축 풀기 zip 파일인 첨부 파일을 핸들하려면 PKZIP 버전 2.50이 필요합니다.

## 테스트 크롤 수행

시작 --> 프로그램 --> 명령 프롬프트를 선택하십시오. 새 창에서 Notes용 IBM Web Crawler를 설치한 곳으로 디렉토리를 이동한 후, notes-run 서브디렉토리로 이동하십시오. 예를 들어, 다음과 같습니다.

```
cd c:\<install directory>\gcs\notes-run
```

Notes 크롤은 다음 두 가지 파일에 의해 제어됩니다.

- 검색할 수 있는 Notes 데이터베이스를 식별하기 위해 편집하는 소스 목록. Notes 서버 이름, IP 주소, .nsf 파일 이름 등이 포함됩니다. 예를 들어, 소스 목록에는 34개의 Notes 데이터베이스 이름이 지정됩니다. 검색되는 이 Notes 데이터베이스는 아래에서 설명하는 구성으로 설정됩니다. 소스 목록은 an.xml 파일 또는 Notes 데이터베이스(.nsf 파일)일 수 있습니다.
- 소스 목록, 검색할 소스, 처리할 첨부 유형, 출력 형식 등을 지정하는 구성 파일. 구성 파일은 언제나 .xml 파일입니다.

IBM Web Crawler를 올바르게 설치했는지 확인하려면 테스트 데이터베이스를 검색하십시오. 편집기를 사용하여 testSources.xml 소스 목록의 test.nsf 데이터베이스에 대한 경로와 파일 이름이 올바른지 확인하십시오. test.nsf 데이터베이스는 IBM Web Crawler를 설치한 notes-run 서브디렉토리에 있습니다. 원래 파일의 백업 사본을 작성하십시오. **권장사항:** 주의깊게 편집하십시오. 파일 오류로 인해 IBM Web Crawler에 장애가 발생할 수 있습니다. 모든 변경사항을 저장하십시오.

포함된 test.nsf 데이터베이스를 검색하여 설치를 테스트하십시오. 유형은 crawlNotes crawlTestXml입니다.

crawlNotes.bat 파일은 crawlTestXml을 구성 파일로 사용하여 IBM Web Crawler를 시작합니다. .xml은 구성 파일 이름에 자동 추가됩니다. IBM Web Crawler는 각각 첨부이 있는 두 가지 문서의 요약 및 검색을 보고해야 합니다.

IBM Web Crawler가 완료되면, 요약 디렉토리에서 요약을 보고, 구성 파일에 지정된 로그 요약 디렉토리에서 크롤 로그 파일을 볼 수 있습니다.

## 사용자 조정 Notes 크롤 구성

정상적인 테스트 크롤 이후에 다른 데이터베이스를 검색하려고 합니다.

1. 데이터베이스 소스 목록을 작성하십시오. 검색할 Notes 데이터베이스를 소스 파일에 추가하십시오.

XML 파일에서 검색할 Notes 데이터베이스를 식별하려면 testSources.xml 파일을 편집하여 시작하십시오. Notes 데이터베이스에서 검색할 Notes 데이터베이스를 식별하려면 Notes를 사용하여 testSources.nsf Notes 데이터베이스를 열어 갱신하십시오. 소스 파일에 설정할 수 있는 매개변수는 소스 목록 편집에서 설명합니다.

2. 크롤러 구성을 설정하십시오. XML 형식으로 구성 파일을 편집해야 합니다.

- 소스가 XML 파일에 나열되어 있는 경우, crawlTestXml.xml을 편집하여 시작한 후, sourcesInXmlFile을 소스 파일 지점으로 설정하십시오.
- 소스가 Notes 데이터베이스에 있는 경우, crawlTestNsf.xml을 편집하여 시작한 후, sourcesInNotesDB를 소스 데이터베이스 지점으로 설정하십시오. 구성 파일에 설정할 수 있는 매개변수가 89 페이지의 『IBM Web Crawler 구성 파일』에서 설명됩니다.

일단 소스 목록 및 구성을 완료하면 IBM Web Crawler: crawlNotes your\_config를 호출하십시오.

또는 소스 목록이 Notes 데이터베이스에 있는 경우, IBM Web Crawler를 다음 작업 예와 같이 시작하십시오. crawlNotes crawlTestNsf

IBM Web Crawler Notes 크롤러가 완료되면, 요약 디렉토리에서 요약을 보고, 구성 파일에 지정된 로그 요약 디렉토리에서 크롤 로그 파일을 볼 수 있습니다.

## 소스 목록 매개변수

소스 목록에는 검색할 수 있는 Notes 데이터베이스 설명이 포함됩니다. file.xml 형식의 소스 목록에는 하나 이상의 oneDBInfo 요소를 포함하는 notesDataSources 요소가 들어 있습니다. 각 oneDBInfo 요소에 포함된 요소는 다음과 같습니다.

**id** 이 데이터베이스의 숫자 ID. 구성 파일의 범위 매개변수가 참조합니다.

### serverName

데이터베이스를 작동하는 서버 이름. 로컬 데이터베이스에는 널(null) 문자열 ""을 사용하십시오.

### pathAndFileName

서버 데이터베이스의 전체 경로 및 파일 이름. 경로 및 파일 이름을 .nsf로 종료하십시오.

### viewName

검색할 데이터베이스 Notes 보기의 이름.

### ipAddress

선택적. 서버 IP 주소를 입력하는 경우, DNS가 사용되지 않습니다. DNS가 목표 서버 이름을 해석하지 못하는 경우, 여기에서 해당 IP 주소를 지정하게 됩니다. Windows의 경우에는 nslookup server\_name 명령을 사용하여 IP 주소를 판별할 수 있습니다.

### dateLastCrawled

선택적. 데이터베이스가 마지막으로 검색된 데이터. 구성 파일에서 update Date Last Crawled를 no로 설정하지 않으면 자동으로 변경됩니다.

**tries** 선택적. 검색되지 않는 경우(예: 시간종료), 데이터베이스 다시 검색을 시도한 횟수.

## fieldSubstitutions

출력 XML 문서에 Notes 데이터베이스 필드 이름을 교체하는 방법 맵핑. 하나 이상의 대체 요소를 포함하고 있으며 각 요소의 두 가지 속성은 다음과 같습니다.

- Original: 출력 XML 문서(있는 경우)에 교체될 필드 이름.
- Replace: 출력 XML 문서에 원래 필드 이름을 교체할 새 필드 이름.

Notes 데이터베이스의 소스 목록은 Notes 클라이언트를 사용하여 점검 및 갱신할 수 있습니다. Notes를 시작한 후, 파일 --> 데이터베이스 --> 열기를 선택하십시오. x:\<install directory>/gcs/notes-run 디렉토리에서 testSources.nsf 데이터베이스를 찾아 열려면 찾아보기를 누르십시오. 테스트 데이터베이스는 Notes 클라이언트를 사용하여 점검 및 갱신할 수도 있습니다. Notes를 시작한 후, 파일 --> 데이터베이스 --> 열기를 선택하십시오. x:\<install directory>/gcs/notes-run 디렉토리에서 test.nsf 데이터베이스를 찾아 열려면 찾아보기를 누르십시오.

## 구성 파일 매개변수

구성 파일에서 설정할 수 있는 매개변수는 아래에서 설명합니다. 나열한 기본값이 있는 매개변수는 생략할 수 있습니다.

sourcesInXml 파일 또는 sourcesInNotesDB 요서는 소스 목록을 식별합니다. 소스는 XML 형식 파일 또는 Notes 데이터베이스로서, 각각 검색할 데이터베이스를 식별하는 정보를 포함합니다.

단일 실행을 제어하는 매개변수를 포함하는 runinfo 요소. 즉, 이 매개변수는 크롤러의 해당 용도에서 검색되는 모든 Notes 데이터베이스에 적용됩니다.

## rangeSpecify

검색할 데이터베이스 ID. ID는 소스 목록의 ID 필드에 입력되는 숫자입니다. 1-4, 15, 25-31과 같이 개별 ID 및 하이픈 표시 범위의 쉼표로 구분되는 목록으로 지정할 수 있습니다.

## SummaryDirectory

출력 요약의 루트 디렉토리를 지정합니다. 요약은 이 디렉토리의 서브디렉토리에 작성됩니다.

## MaxThreads

병렬 크롤 스레드 수를 지정합니다. 각 Notes 데이터베이스는 단일 스레드로 검색됩니다. 복수의 데이터베이스를 병렬로 검색합니다.

## doIncrementalCrawl

기본값은 no입니다. yes인 경우, 'summarizeThisDateAndLater' 이후 수정하거나 새로운 Notes 문서만을 처리합니다. 'summarizeThisDateAndLater'를 지정하지 않은 경우, 데이터베이스 소스 목록에 지정된 각 데이터베이스 고

유의 DateLastCrawled 필드를 크롤러가 사용합니다. doIncrementalCrawl=no 인 경우, 날짜에 관계없이 모든 문서를 처리합니다.

#### **summarizeThisDateAndLater**

이 필드의 형식은 MM/dd/yyyy hh:mm a tz입니다(예: 01/01/2000 01:11 PM PDT). 날짜 및 시간이 입력되지 않은 경우, 소스 목록에서 마지막으로 기록한 크롤 이후(doIncrementalCrawl을 yes로 설정한 경우) 또는 모든 경우(doIncrementalCrawl을 no로 설정한 경우)의 모든 문서를 요약하십시오.

#### **detachAttachments**

기본값은 yes입니다. yes인 경우, 첨부 파일을 분리 및 요약하십시오. 처리할 첨부 파일 유형을 구성 파일에서 나열합니다. no인 경우, 첨부 파일을 무시합니다.

#### **attachmentFilenameFormat**

기본값은 l(long)입니다. s(short)도 가능합니다. 긴 파일 이름은 유형, 서버, 데이터베이스 이름 및 Notes ID를 인코딩합니다. 짧은 파일 이름은 유형 및 Notes ID를 인코딩합니다.

#### **processAttachmentsAfterwards**

기본값은 no입니다. yes인 경우, Notes 데이터베이스 요약 동안 첨부 파일을 요약하지 않습니다. 대신 파일 요약 명령을 지정하는, 각 첨부 파일에 대한 notesCrawl-attachments.bat에 레코드가 작성됩니다. 첨부 파일을 요약한 후 삭제하는 일괄처리 파일을 작성하고 실행합니다. 이후 첨부 처리를 위해서는 일반적으로 디스크 기억영역이 많이 필요합니다.

#### **saveAttachmentFiles**

기본값은 no입니다. yes인 경우, 첨부 파일 원본은 처리 이후에 지워지지 않습니다. 이 옵션은 processAttachmentsAfterwards가 no인 경우에만 유효합니다. processAttachmentsAfterwards가 yes로 설정된 경우, 저장된 첨부 파일을 저장하는 데 디스크 공간이 필요합니다.

#### **MaximumNumberOfDetachingErrors**

기본값은 10입니다. 첨부 처리 시의 최대 오류(예: 첨부 파일 저장 시 디스크 용량 부족) 수로서 크롤 중단 전에 크롤러가 허용하는 오류 수입니다.

#### **saveURLsToFile**

기본값은 no입니다. 이 값을 지정한 경우, Notes 문서 항목의 URL이 다음과 같은 양식의 이름을 가진 파일에 작성됩니다.

databaseName(without path and .nsf) + ".html".

#### **updateDateLastCrawled**

기본값은 yes입니다. 이 값이 no인 경우, 소스 파일의 dateLastCrawled를 갱신하지 마십시오.

#### **tempDirectory**

기본값은 c:\temp입니다. 이 디렉토리는 모든 임시 파일 작성에 사용됩니다.



## logSummaryDirectory

기본값은 log입니다. 로그 파일이 저장될 디렉토리입니다.

## loggerPriority

기본값은 info입니다. 이 설정은 로거의 우선순위를 정의합니다. 이 설정은 최상위에서 최하위까지 error, warn 또는 info일 수 있습니다. 예를 들어, 로거 우선순위를 warn으로 설정한 경우, 우선순위가 warn 및 error인 로그 메시지만을 기록합니다.

처리할 첨부 파일 확장자를 식별하는 포함 요소를 포함하는 첨부 요소입니다(예: .prz).

## 서버에서 IBM Web Crawler 제외

보안 및 성능상의 이유로 EIP 시스템 관리자는 특정 서버 또는 페이지가 검색되지 않기를 원할 수 있습니다. 서버 및 페이지의 크롤러 활동을 제한할 수 있어야 합니다.

액세스 방침 파일을 사용하는 서버나 페이지를 피하도록 IBM Web Crawler에 지시할 수 있습니다. 이 파일은 로봇 제외를 위한 표준에 발표된 지침에 따라 작성됩니다 (<http://info.webcrawler.com/mak/projects/robots/norobots.html>을 참조).

- IBM Web Crawler는 서버를 검색하기 전 및 그 이후에 주기적으로 액세스 방침 파일 *yourserver/robots.txt*를 요청합니다.
- 이 파일은 다음 양식의 행으로 구성됩니다.

*field*:<optionalspace>*value*<optionalspace>

*field*가 User-Agent이고 *value*가 IBM-WebCrawler 또는 \*인 경우, 다음 Disallow 행(다음 User-Agent 행까지)이 피해야 할 부분 주소를 지정합니다. 이는 전체 경로 또는 부분 경로가 될 수 있으며, 이 값으로 시작하는 주소는 검색되지 않습니다.

예를 들어, 다음과 같습니다.

Disallow: /help

는 /help.html 및 /help/index.html을 모두 허용하지 않습니다.

Disallow: /help/

는 /help/index.html을 허용하지 않지만 /help.html은 허용합니다.

빈 값을 사용하면 모든 주소가 검색됩니다.

- 이들 행은 공백 행으로 구분될 수 있습니다.
- # 문자를 입력하여 주석을 포함시킬 수 있습니다. 나머지 행은 주석으로 간주됩니다.

다음은 그 예입니다.

- 이 /robots.txt는 모든 로봇이 이 서버를 피하도록 지정합니다.

```
# disallow everybody
```

```
User-agent: *  
Disallow: /
```

- 이 /robots.txt는 IBM Web Crawler만이 이 서버를 검색할 수 있고 크롤러는 제한을 갖지 않도록 지정합니다.

```
# allow only IBM
```

```
User-agent: *  
Disallow: /  
User-agent: IBM-WebCrawler  
Disallow: # disallow nothing
```

- 이 /robots.txt는 모든 로봇이 temp, development 및 testing htmldocs 트리 안에 있는 주소를 피해야 하고, IBM-WebCrawler를 호출한 사용자 에이전트가 개발 및 트리 테스트에 대해 예외를 가지며(여기서는 허용됨), xyz 및 wxyz 로봇이 완전히 떨어져 있도록 지정합니다.

```
# a more realistic example
```

```
User-agent: *  
Disallow: /htmldocs/temp  
Disallow: /htmldocs/development  
Disallow: /htmldocs/testing
```

```
User-agent: xyz  
User-agent: wxyz  
Disallow: /
```

```
User-agent: IBM-WebCrawler  
Disallow: /htmldocs/temp
```

---

## 제 6 장 워크플로우 소개

EIP 워크플로우를 사용하여 비즈니스 작업의 흐름 및 성능을 제어할 수 있습니다. 사용자가 연합 검색 결과로 작업하면 종종 수행할 조치를 결정해야 합니다. 사용자가 EIP 워크플로우를 사용하여 작업을 수행할 방법을 미리 결정할 수 있습니다.

워크플로우 구성요소가 함께 작동하는 방법을 제어하는 프로파일 및 규칙을 설정하여 워크플로우를 자동화할 수 있습니다. 또한 사용 권한 세트 및 액세스 제어 목록을 통해 사용자 액세스 및 권한을 제어하여 시스템을 어느 정도 제한하는지 선택할 수도 있습니다.

---

### 워크플로우 이해

대부분의 비즈니스 조작은 상호 관련된 일련의 처리로 규정될 수 있습니다. 작업은 한 직원에서 다른 직원으로, 한 부서에서 다른 부서로 이동합니다. 일부 단순한 처리에서는 몇 가지 단계만 요구될 수 있는 반면, 보다 복잡한 처리에는 여러 부서의 다수 직원들이 관계됩니다.

워크플로우는 처리를 통해 작업을 이동하고 처리 전반의 작업에 대해 의사 결정을 할 수 있습니다. 예를 들어, XYZ 보험은 대량의 청구 양식을 우편으로 받습니다. 확인 처리 시, 보험 회사 지급 청구 조정자는 사진, 평가 및 감정서 같은 문서를 수집해야 합니다. 직원이 최종 승인에 적절한 문서를 수집하고, 정보를 열고 정렬, 정리 및 모니터링하는 데 매일 몇 시간이 소요됩니다.

이러한 정보는 정보가 수신되어 점검되면 한 직원에서 다른 직원으로 이동됩니다. 청구가 완료되면 둘 이상의 부서의 직원이 처리할 수 있습니다.

---

### 워크플로우 사용 방법

XYZ 보험의 예에서처럼, 문서를 처리하는 대부분의 회사들은 다음 작업의 일부 또는 전부를 수행합니다.

- 차후 검색을 위한 문서 정리
- 다른 소스에서 문서, 양식, 보고서 및 정보를 수집한 후, 이러한 문서를 처리할 위치에 전달
- 수신 메일을 현재 처리되고 있는 문서와 비교

워크플로우는 작업의 흐름을 나타냅니다. 워크플로우에서는 하나 이상의 문서 또는 콘텐츠 그룹에서 수행할 수 있는 조치와, 워크플로우를 통해 이 문서 그룹이 가져온 경로에 대해 설명합니다. 워크플로우는 범위와 경계가 명백하게 정의된 작업이 수행되는 방

법을 반영합니다. 활동과 작업의 순서, 그리고 이들 활동과 작업 간의 연결 및 관계를 정의합니다. 워크플로우는 작업 흐름에 관해 의사 결정하는 데 사용되는 기준을 판별합니다. 워크플로우 작성 처리에 대한 자세한 정보는 워크스테이션 응용프로그램 프로그래밍 안내서의 내용을 참조하십시오. 워크플로우로 클라이언트 사용에 대한 자세한 정보는 *eClient* 설치, 구성 및 관리를 참조하십시오.

---

## 사용자 ID 및 그룹 동기화

이 절에서는 EIP, Content Manager 및 MQ Series 워크플로우 간 사용자 ID 및 그룹을 동기화하는 방법에 대해 설명합니다.

MQSeries 워크플로우가 시스템의 부분인 경우 Content Manager 또는 Enterprise Information Portal에서 사용자를 관리한다면, MQSeries 워크플로우에서도 사용자를 관리할 수 있습니다. 그러므로 Content Manager 또는 Enterprise Information Portal에서 사용자 ID 또는 그룹을 작성, 수정하거나 삭제하는 경우에는 MQSeries 워크플로우 서버에서도 동일하게 수행해야 합니다.

Content Manager 및 Enterprise Information Portal이 사용자 ID 및 그룹을 공유하므로, 사용 가능한 워크플로우 서비스를 사용하여 시스템 관리 클라이언트에서 사용자 ID 또는 그룹을 작성할 때 MQSeries 워크플로우 서버가 실행중이 아니면, 사용자 또는 그룹이 MQSeries 워크플로우 서버에서 작성되지 않았다는 오류를 수신하게 됩니다. 사용자 ID는 Content Manager 또는 Enterprise Information Portal에서 사용자로 여전히 존재하지만 MQSeries 워크플로우 서버에서는 존재하지 않습니다.

MQSeries 워크플로우의 Content Management 사용자 ID 및 그룹을 동기화하려면, 워크플로우 사용자 동기화 유틸리티를 실행해야 합니다. 수동으로 MQSeries 워크플로우 서버를 설치했다면, MQSeries 워크플로우 서버가 시작했는지 확인하십시오. 서비스를 열고 MQSeries 워크플로우 X.X에 대한 상태를 점검하십시오 - FMC(X.X가 설치한 워크플로우 버전임). MQSeries 워크플로우가 실행되지 않는 경우 서비스를 시작하거나, 조용히 설치된 경우, WFInstall 디렉토리로 가서 일괄처리 파일 CMBWFStart.bat을 실행하십시오. 다음 단계를 완료하고 동기 유틸리티를 실행하십시오.

1. Enterprise Information Portal을 설치한 디렉토리로 변경하십시오. 기본 디렉토리는 C:\CMBROOT입니다.
2. EIPUser2WF.bat를 입력하십시오.

**중요:** 원격 플로우 서버가 있는 경우, EIPUser2WF.bat을 입력하기 전에 RMI 서버 일괄처리 또는 셸 파일(cmbsvregist81.bat 또는 cmbsvregist81.sh)을 시작해야 합니다. 또한 EIPUser2WF.bat는 로컬 또는 원격 서버를 찾기 위해 두 개의 INI 파일 cmbsvcs.ini 및 cmbsvclient.ini에서 정보를 사용합니다. cmbsvcs.ini가 서버가 로컬인 경우 LOCAL 또는 서버가 원격인 경우 REMOTE인지를 확인하십시오. cmbsvclient.ini 파일에는 서버 위치가 포함되어 있습니다.

3. EIP 데이터베이스 이름, 사용자 ID, 암호 및 스키마에 대해 필수 정보를 입력하십시오.

필수 정보를 입력하면 동기화 유틸리티는 사용자를 Content Manager 또는 Enterprise Information Portal 서버에서 MQSeries 워크플로우 서버로 복사합니다. 유틸리티가 완료되면 사용자 또는 그룹은 MQSeries 워크플로우 서버에 알려지지 않은 오류는 수신하지 않아야 합니다.

Content Manager 또는 Enterprise Information Portal에서 사용자 ID 또는 그룹을 삭제하는 경우, MQSeries 워크플로우 서버에서도 사용자 ID 또는 그룹을 삭제해야 합니다. 사용자 ID 또는 그룹이 MQSeries 워크플로우 서버에 존재하지 않는 경우에는 Content Manager 또는 Enterprise Information Portal에서 ID 또는 그룹을 삭제할 수 없습니다. 예를 들어, 사용 불가능한 EIP 워크플로우 서버 옵션을 사용하여 Content Manager 또는 Enterprise Information Portal에 사용자 ID 또는 그룹을 작성할 수 있습니다. 그리고 나서 사용자 ID 또는 그룹을 삭제하려는 경우 워크플로우 서버 옵션을 사용할 수 있습니다. 사용자 ID 또는 그룹이 MQSeries 워크플로우에 존재하지 않으므로 사용자 ID가 MQSeries 워크플로우 서버에 존재하지 않음을 설명하는 오류를 수신합니다. 오류를 정정하려면 EIPUser2WF 유틸리티를 실행하여 사용자 ID 및 그룹을 동기화한 후 사용자 ID 및 그룹을 삭제해야 합니다.

## 작동 워크플로우로 EIP 서버 재설치

EIP 서버에 작동 EIP 워크플로우가 있고 새 EIP 서버를 재설치하려는 경우, MQSeries 워크플로우 서버에서 EIP 워크플로우 데이터를 모두 제거해야 합니다.

EIP 서버를 제거하기 전에, 다음 단계를 완료하여 MQSeries 워크플로우 서버에서 EIP 워크플로우 데이터를 모두 제거하십시오. **중요:** 아래 설명한 순서대로 단계를 수행하십시오.

1. eClient 또는 워크플로우 API를 사용하여 워크플로우 인스턴스를 모두 종료하십시오. 워크플로우 인스턴스를 종료하면 모든 작업 항목을 삭제합니다.
2. EIP 시스템 관리 클라이언트 또는 시스템 관리 API를 사용하여 작업 목록 및 워크플로우 템플릿을 삭제하십시오.
3. 워크플로우가 작동하는 경우 EIP에서 모든 사용자 및 사용자 그룹을 삭제하면, 모든 사용자 ID 및 사용자 그룹이 MQSeries 워크플로우 서버에서도 삭제됩니다.
4. EIP용 MQSeries 워크플로우 서버 재구성
  - a. EIP 설치 CD-ROM을 삽입하십시오. 명령 프롬프트를 사용하여 WfInstall 디렉토리로 변경하십시오.
  - b. 명령 프롬프트에 다음과 같이 입력하십시오.

```
fmcibie -iCMBWAdmin.fdl -uadmin -ppassword -o -f
```

**중요:** EIP 및 MQSeries 워크플로우에서 사용자 ID 및 그룹을 삭제하기 전에 EIP 데이터베이스를 제거하고 삭제한 EIP 데이터베이스에 있던 것과 동일한 사용자 ID 및 그룹을 작성하려는 경우, 사용자 및 그룹을 EIP에 추가할 수 없다고 표시하는 오류가 수신됩니다. 오류 메시지는 다음과 같습니다. DGL2616A: Fail to add the user: XXX -DGL2485A: 이 워크플로우 사용자가 이미 존재합니다. 다음 단계를 완료하여 문제점을 해결하십시오.

1. EIP 시스템 관리 클라이언트에 EIP 워크플로우 서비스를 사용하지 않습니다.
2. EIP 시스템 관리 클라이언트를 로그오프하고 다시 로그인하십시오. 워크플로우 서비스를 사용하지 않고 로그인하는 경우, EIP에 있고 MQSeries 워크플로우 서버도 존재하는 것과 동일한 사용자 ID 및 그룹을 작성할 수 있습니다.
3. MQSeries 워크플로우 서버에 이미 존재하는 사용자 ID 및 그룹을 작성하면 EIP 워크플로우 서비스가 사용 가능합니다.

## MQSeries 워크플로우와 EIP 데이터베이스 간 사용자 ID 및 그룹 갱신

MQSeries 워크플로우에 존재하지만 EIP 데이터베이스에는 존재하지 않는 각 사용자 ID 또는 그룹의 경우, 다음 단계를 완료하고 사용자 ID 및 그룹을 동기화할 수 있습니다.

1. 파일을 작성하고 다음 텍스트를 입력하십시오.

- CODEPAGE 1252
- FM\_RELEASE V3R3 2
- DELETE PERSON 'User1'
- DELETE PERSON 'User2'
- DELETE PERSON 'User3'

User1, User2 및 User3이 삭제하고자 하는 사용자입니다. 필요한 만큼 많은 사용자를 나열할 수 있습니다.

2. 파일을 저장한 후 명령 프롬프트에서 다음 명령을 실행하십시오.

```
fmcibie -u admin -ppassword -i DeletePersons.fdl -f -o
```

DeletePersons.fdl은 이전 단계에 작성한 파일 이름입니다.

EIP 데이터베이스에 더 이상 존재하지 않는 MQSeries 워크플로우 서버에 있는 각 그룹의 경우, 다음 단계를 완료할 수 있습니다.

1. 텍스트 파일을 작성하고 다음 정보를 입력하십시오.

- CODEPAGE 1252
- FM\_RELEASE V3R3 2
- DELETE ROLE 'Group1'
- DELETE ROLE 'Group2'

- DELETE ROLE ‘Group3’

Group1, Group2 및 Group3은 삭제하려고 하는 그룹입니다. 필요한 만큼 많은 그룹을 나열할 수 있습니다.

2. 파일을 저장한 후 명령 프롬프트에 다음 명령을 입력하십시오.

```
fmcibie -u admin -p password -i DeleteGroups.fdl -f -o
```

DeleteGroups.fdl은 이전 단계에서 작성한 파일 이름입니다.

EIP에서 사용자 ID를 삭제하여 워크플로우 지원이 작동할 경우, 다음 오류가 표시될 수도 있습니다.

```
Failed to delete a user [RC=12]
```

사용자 조작 삭제가 실패한 경우, 시스템은 temp.log를 실행하여 장애에 대한 세부사항을 제공합니다. temp.log 파일이 x:\CMBROOT에 쓰입니다. **팁:** MQSeries 워크플로우 클라이언트로 로그인하기 위해 사용한 사용자 ID를 삭제하려는 경우, 사용자 조작 삭제가 실패합니다.

## 워크플로우 계획

워크플로우 정의를 시작하기 전에 비즈니스가 수행되는 작업, 위치와 방법 및 사용자를 분석해야 합니다. 시스템 관리자 또는 비즈니스 분석가가 이 계획 단계를 수행합니다.

최종 제품이란? 최종 제품은 사용자의 비즈니스, 비즈니스의 한 부서 또는 여러 부서의 특정 직원에 의해 수행된 모든 작업의 결과일 수 있습니다. 예를 들어, XYZ 보험의 청구 보상 처리의 최종 제품은 청구를 승인 또는 거부하는 보험 계약자에게 보내는 서신입니다.

최종 제품을 생성하기 위해 처리할 정보를 분석하고, 수행해야 할 조치와 이 조치가 수행되는 위치를 결정하며, 워크플로우를 통해 정보가 이동하는 방법을 결정하십시오.

## 처리할 정보

기업에서 사용자가 처리해야 하는 정보를 고려하십시오. 최종 제품을 지원하는 입력 유형은 무엇입니까? 처리해야 하는 특정 문서는 무엇입니까?

작업 항목은 실제 문서가 아닙니다. 작업 항목에는 문서 참조와 문서, 문서 상태, 작성 날짜 등에 대한 기타 정보도 포함되어 있습니다. 작업 항목은 콘텐츠 서버의 임의의 콘텐츠(문서 또는 오브젝트)일 수 있습니다. 예를 들어, XYZ 보험은 우선 청구 양식을 받은 후 사진, 평가 및 감정서 같은 후속 문서를 받습니다.

## 정보 처리 방법

처리의 각 단계를 누가 가장 잘 처리할 수 있습니까? 예를 들어, 관리 보조는 청구 양식이 완전한지 검증한 후, 보험 계약자로부터 특정 문서를 받을 때까지 청구 양식을 정리합니다. 문서가 도착하면 청구 조정자는 문서를 청구 양식과 비교하고 해당 문서 승인을 담당할 수도 있습니다.

작업 목록은 하나 이상의 직원이 사용할 수 있도록 작성한 작업 대기열로 간주될 수 있습니다. 청구 양식은 다수의 청구 조정자가 액세스할 수 있는 작업 목록으로 그룹화될 수 있습니다. 작업 목록은 작업 항목의 필터링된 보기입니다. 직원은 볼 수 있도록 허용된 작업 목록의 항목만 봅니다.

작업 목록은 사진, 평가 및 감정서 수집 같은 청구 처리의 각 부분을 처리하는 방식으로 작업 항목을 필터하도록 정의될 수 있습니다. 작업 목록은 여러 워크플로우의 작업으로 구성될 수도 있습니다. 예를 들어, 한 청구 조정자의 작업 목록에 하나의 청구를 위한 평가, 두 번째 청구를 위한 사진 및 세 번째 청구를 위한 감정서를 포함할 수 있습니다. 이 작업 목록의 각 항목에 대해 조정자가 수행하는 조치는 다를 수 있습니다. 조정자는 평가를 검토하여 첫 번째 청구를 승인할 수 있습니다. 사진에 대한 조치를 취하기 전에 두 번째 청구에 대한 자세한 정보를 보기 위해 대기해야 할 수 있습니다. 세 번째 청구의 경우, 조치에 대한 감정서를 다른 직원에게 보낼 수 있습니다.

## 수행할 조치

워크플로우 중 작업 항목의 콘텐츠에 취해야 할 조치를 고려하십시오. 예를 들어, 청구 조정자는 청구 양식을 승인하거나 이를 미완성으로 거부할 수 있습니다. 조치 목록에서는 사용자가 작업에 대해 수행할 수 있는 조치를 정의합니다.

예를 들어, 청구가 초기 요구사항에 맞는지에 따라 조정자는 한 옵션을 선택하여 워크플로우를 통해 청구 양식을 계속하거나 다른 옵션을 선택하여 양식을 거부할 수 있습니다.

## 처리에서의 정보 이동 방법

정보 및 활동을 이동할 방법을 고려하십시오. 예를 들어, 초기 청구 양식이 검토된 시기는 언제입니까? 처리에서 다음 단계로 이동해야 하는 지원 문서는? 청구가 승인 또는 거부되는지를 판별하는 기준은? 이 정보 흐름은 워크플로우의 기본입니다.

워크플로우는 처리를 통해 작업을 안내하는 경로로 구성됩니다. 입력이 시작되는 위치는? 워크플로우는 특정 지점에서 시작되어야 합니다. XYZ 보험의 경우, 보험 계약자가 제출한 청구 양식이 워크플로우를 시작하는 문서입니다.

모든 문서를 받으면 작업 항목은 최종 조치(예: 청구 승인)에 대한 경로를 따라 계속할 수 있습니다.



## 모든 작업이 서로 조화되는 방법

처리할 정보를 분석하고, 수행할 조치를 결정하며, 정보 이동 방법을 결정한 후, 워크플로우의 그래픽 표시인 워크플로우 도표를 작성할 준비가 되어 있습니다. EIP 워크플로우 빌더 기능을 사용하여 도표를 작성할 수 있습니다.

워크플로우 도표에는 다양한 활동을 통해 작업이 이동하는 방법이 표시되며 활동이 관련된 작업을 나타냅니다. 워크플로우의 흐름, 기본 요소 및 핵심에 대해 설명합니다.

워크플로우 도표의 각 기호는 작업이 수행되는 지점을 나타냅니다. 보험 청구는 검토되어야 하며, 지원 문서가 수집되어야 하고, 특정 기준에 따라 청구가 승인되거나 거부되어야 합니다. 워크플로우 빌더에서 사용되는 처리 기호에 대한 자세한 내용은 117 페이지의 『워크플로우 작성』을 참조하십시오.

---

## Enterprise Information Portal 워크플로우 구성요소 사용

이 절에서는 워크플로우 구성요소에 대해 설명합니다. 관리 클라이언트를 통해 모든 구성요소를 액세스합니다. **팁:** EIP 버전 8 워크플로우는 새 Content Manager 버전 8 아키텍처를 수용하기 위해 버전 7.1 컨테이너에 대한 변경사항과 함께 여러 변경사항을 포함합니다.

### 워크플로우 빌더 사용




워크플로우 빌더를 사용하여 작업 그룹, 부서 또는 기업을 그래픽으로 정의하고 구축할 수 있습니다. **제한사항:** EIP 이주 처리는 버전 7.1 데이터베이스에서 사용자를 이주합니다. EIP 버전 8.2는 워크플로우 데이터의 자동화된 이주를 제공하지 않습니다. EIP 버전 8.2 워크플로우 빌더를 사용하여 버전 7.1 워크플로우 도표를 다시 그린 후 EIP 버전 7.1 워크플로우 처리를 다시 전개해야 합니다.

워크플로우 빌더를 사용하여 워크플로우 모델을 작성하기 전에, 사용 권한 세트, 액세스 제어 목록, 사용자, 사용자 그룹, 조치, 조치 목록 및 작업 목록을 정의해야 합니다. 관리 클라이언트에서 워크플로우를 정의하면 전체 워크플로우의 기본 조치 목록을 설정할 수 있습니다. 또한 워크플로우의 각 노드에 다른 조치 목록을 할당할 수도 있습니다. 이러한 작업에 대한 자세한 내용은 117 페이지의 『조치 목록 정의』, 116 페이지의 『작업 목록 정의』 및 온라인 도움말을 참조하십시오.

워크플로우 빌더를 사용하여 워크플로우를 구축한 경우에도, 워크플로우 빌더를 사용하여 워크플로우를 실행할 수 없습니다. 사용자는 클라이언트를 사용하여 작업 목록 및 작업 항목에 대한 조치를 보거나 수행합니다. Enterprise Information Portal 워크플로우로 작업하기 위해 클라이언트를 프로그래밍하는 방법에 대한 자세한 정보는, 워크스테이션 응용프로그램 프로그래밍 안내서 및 온라인 API 참조서를 참조하십시오. 116 페이지의 표 7은 모든 워크플로우에 공통인 세 개의 워크플로우 아이콘을 설명합니다. 기

타 워크플로우 아이콘(예: 하위 워크플로우, 이벤트, 콜렉션 및 사용자 종료)이 있습니다. EIP 온라인 도움말은 도구 모음 아이콘을 자세히 설명합니다.

표 7. 기본 워크플로우 아이콘

아이콘	설명
	시작 노드는 워크플로우 처리를 시작합니다. 워크플로우 처리 도표는 오직 하나의 시작 노드를 가져야 합니다.
	정지 노드는 워크플로우 처리를 종료합니다. 모든 새 워크플로우 처리 도표에는 정지 노드가 있습니다. 처리 작성 시, 정지 노드가 생성됩니다. 그리기 표면의 임의 위치로 정지 노드를 이동시킬 수 있습니다. 워크플로우 처리 도표는 오직 하나의 정지 노드를 가져야 합니다.
	작업 노드는 워크플로우 처리에서 특정 위치에 대한 작업 목록 및 조치 목록을 연관합니다. 작업 노드는 작업을 수행하는 워크플로우 처리에서 위치를 표시합니다. 시작 및 중지 노드를 포함한 각 노드는 조치 세트 및 조치 세트를 수행하도록 할당된 관리자를 필요로 합니다.

## 워크플로우 서비스 사용

Enterprise Information Portal에서는 워크플로우 정보를 유지보수하는 워크플로우 서비스를 제공합니다. 워크플로우 빌더를 사용하여 작성된 워크플로우 및 조치 목록 정의는 Enterprise Information Portal 관리 데이터베이스 및 IBM MQSeries® 워크플로우 데이터베이스에서 유지보수됩니다.

시스템 관리자가 작업 목록을 작성하면, 작업 목록과 연관된 정보는 관리 데이터베이스에 영구적으로 저장됩니다. 시스템 관리자는 EIP를 사용하여 작업 목록을 갱신, 삭제 및 추가할 수 있습니다. 시스템 관리자가 워크플로우를 체크아웃하면 워크플로우는 Enterprise Information Portal 데이터베이스에서 잠기고 사용자에게 체크 아웃된 것으로 데이터베이스 내에서 표시되며, 사용자가 완료할 때까지 다른 사람이 갱신하지 못하게 합니다.

## 작업 목록 정의

작업 목록은 사용 가능한 작업 필터로 생각될 수 있습니다. 작업 목록은 특정 사용자나 사용자 그룹에 할당된 항목의 필터링된 목록입니다. 사용자가 Enterprise Information Portal에 로그인하면 이들에 할당된 작업 항목의 필터링된 목록을 볼 수 있습니다. Enterprise Information Portal 관리 클라이언트를 사용하여 작업 목록을 정의합니다.

작업 목록 정의에는 작업 항목의 프리젠테이션, 상태 및 보안을 관리하는 규칙이 포함됩니다. 작업 목록을 작성하는 동시에 작업 목록마다 규칙을 지정합니다. 작업 목록에 대한 액세스를 관리하기 위해 작업 목록의 액세스 제어 목록을 작성합니다. 작업 목록 정의 방법에 대한 자세한 내용은 온라인 도움말을 참조하십시오. 작업 목록 정의에는 다음 사항이 포함됩니다.

### 액세스 제어 목록

액세스 제어 목록은 하나 이상의 개별 사용자 ID나 사용자 그룹 및 이들과 연관된 사용 권한 세트로 구성됩니다. 사용 권한 세트는 작업에 대해 특정 작업에 액세스하거나 특정 작업을 수행할 사용자 권한을 정의하는 데 사용됩니다. 액세스 제어 목록을 사용하여 작업 목록에 있는 항목에 대한 사용자 액세스를 제한합니다.

### 작업 목록 필터링 및 정렬

사용자가 필터링되어 정렬된 작업 목록을 볼 수 있는 기준

### 작업 목록에 있는 최대 항목 수

작업 목록에 포함시키려는 최대 항목 수

---

## 조치 목록 정의

조치 목록은 사용자가 워크플로우에서 작업을 수행할 수 있는 종합적인 모든 조치 목록입니다.

관리 클라이언트 온라인 도움말은 조치와 조치 목록을 정의하는 방법을 설명하는 단계별 지시사항을 제공합니다.

---

## 워크플로우 작성

조치, 조치 목록 및 작업 목록을 정의한 후, 워크플로우 빌더를 사용하여 워크플로우의 모델을 작성합니다. 관리 클라이언트 온라인 도움말은 조치와 조치 목록을 정의하는 방법을 설명하는 단계별 지시사항을 제공합니다. 워크플로우 빌더는 워크플로우 작성을 위한 비주얼 신호를 제공합니다.

---

## 워크플로우 빌더 사용

이 단계에서, 관리 데이터베이스에서 워크플로우를 시작합니다. **제한사항:** 워크플로우에 대해 사용자가 선택한 데이터베이스는 MQ Series를 설치하고 MQSeries 서비스가 시작되어야 하는 동일 서버에 있어야 합니다.

EIP 워크플로우를 사용하고 워크플로우 정의를 작성하려면 다음과 같이 하십시오.

1. 관리 클라이언트에 로그인하십시오.
2. 여러 관리 데이터베이스가 있는 경우, 워크플로우를 사용하려는 데이터베이스에 대한 아이콘을 누르십시오.
3. Tools->Services를 누르십시오. 워크플로우 사용을 누르십시오.
4. 클라이언트를 로그아웃하고 다시 로그인하십시오. 여러 데이터베이스가 있는 경우, EIP 워크플로우를 작동시킨 데이터베이스에 대한 아이콘을 선택하십시오. 워크플로우 폴더 아이콘이 표시됩니다.

5. Enterprise Information Portal 관리 기본 창의 왼쪽 분할 영역에서 워크플로우 폴더 워크플로우를 두 번 누르십시오.
6. 워크플로우 정의 아이콘을 두 번 누르고 신규를 선택하여 워크플로우 정의를 작성하십시오.

**요구사항:** 워크플로우를 정의하기 전에 최소한 하나의 액세스 제어 목록, 하나의 조치 및 하나의 조치 목록을 작성해야 합니다.

---

## MQSeries 워크플로우 서버 시작

명령 프롬프트에 `cmbwfstart`를 입력하여 MQSeries 워크플로우 서버를 시작하십시오. MQSeries 워크플로우 서버에 대해 두 개의 창이 열립니다. 이 명령 창을 열린 상태로 두고 서버 실행을 계속하십시오.

초기 Enterprise Information Portal 설치 후에 워크플로우를 설치하려는 경우, 워크플로우 기능에 대해 Enterprise Information Portal 시스템을 구성해야 합니다. 또한 관리 클라이언트가 설치된 워크스테이션과 다른 워크스테이션에 워크플로우 기능을 설치하는 경우, 구성을 변경해야 합니다.

1. 관리 창에서 파일 구성원 도구를 누르십시오.
2. 메뉴에서 서비스를 선택하십시오.
3. 워크플로우 체크 박스를 선택하십시오.
4. 구성이 완료된 후, Enterprise Information Portal 관리 클라이언트를 로그오프하고 다시 로그인하여 워크플로우 기능을 초기화하십시오. Enterprise Information Portal 관리 클라이언트에 로그인한 후, 워크플로우 정의 아이콘이 왼쪽 분할 영역에 나타납니다.

**팁:** 시스템 관리자가 워크플로우 기능을 관리하는 권한을 갖고 있지 않으면 워크플로우 정의 아이콘을 볼 수 없습니다. 워크플로우 기능에 대한 액세스를 제한하려는 경우, 각 콘텐츠 서버에 대한 해당 시스템 관리 서적을 참조하십시오. 시스템 관리자에게 워크플로우 기능 관리 권한 부여에 대한 자세한 내용은 온라인 도움말을 참조하십시오.

EIP Connector Toolkit 및 샘플을 사용하여 사용자 조정 응용프로그램에서 클라이언트를 작성하거나, EIP 샘플 클라이언트를 사용할 수 있습니다.

---

## 제 7 장 IBM Web Crawler 샘플 파일

이 절에서는 두 가지 코드 샘플을 제공합니다. config-sample2.xml 샘플 파일은 <gcs-config> 구성 매개변수의 예를 제공합니다. 로그 분석 샘플은 완료된 크롤로부터 정보를 포함하는 보고서 예를 제공합니다.

---

### config-sample2.xml 샘플

이 절의 코드 샘플은 gcs-config 파일의 예입니다.

```
<!DOCTYPE gcs-config SYSTEM "config.dtd">
<gcs-config>
  <!-- Global settings: -->
  <globals max-urls="1000000"
    num-crawlers="30"
    num-summarizers="8"
    summaries-dir="summaries"
    log-file="log/LOG.txt"
    temp-dir="temp"
    log-priority="warn"
    text-monitor="60"
    graph-monitor="2"
    connect-timeout="120"
    read-timeout="100">

    <!-- sample logger settings -->
    <logger-config category="gcs.summaries.list.resource"
      priority="info" log-file="log/resources.txt"/>
    <logger-config category="gcs.summaries.list.host"          priority="info"
      log-file="log/hosts.txt"/>
    <logger-config category="gcs.url.skipped"                  priority="info"
      log-file="log/skipped_urls.txt"/>
    <logger-config category="gcs.url.fromto"                   priority="info"
      log-layout="%m\n" log-file="log/fromto.txt"/>
    <logger-config category="gcs.http"                          priority="info"
      log-file="log/http.txt"/>
    <logger-config category="gcs.http.connect"                 priority="info"
      log-file="log/connecterrs.txt"/>

    <!--use this to specify a database
    <urlpool-config urlcontainer-class="DB2URLContainer"
      urlcollection- class="DB2URLCollection">
      <urlpool-param name="dbname"      value="gcs"/>
      <urlpool-param name="user"        value="xxxxxx"/>
      <urlpool-param name="password"    value="xxxxxx"/>
      <urlpool-param name="cachesize"   value="1000"/>
      <urlpool-param name="driver"
      value="COM.ibm.db2.jdbc.app.DB2Driver"/>
    </urlpool-config> -->

    <!--use this to specify a SOCKS proxy
    <system-properties>
      <property name="socksProxySet" value="true"/>
      <property name="socksProxyHost" value="socks2.server.ibm.com"/>
```

```

        <property name="socksProxyPort" value="1080"/>
    </system-properties> -->

</globals>

<group-list>
    <group name="ibm">

        <crawler-config recursion-depth="-1">
            <seed-list>
                <!-- URLs to start crawling at: -->
                <seed url="http://gcs.stl.ibm.com/gcs/testurl.html"/>
                <seed url="http://gcs.stl.ibm.com/gcs/stl.html"/>
                <seed url="http://gcs.stl.ibm.com/gcs/ibm.html"/>
            </seed-list>

            <content-type-pattern-list>
                <!-- URL file extensions that don't match these patterns
                won't be crawled: -->
                <url-name-pattern name="htm*" />
                <url-name-pattern name="pdf" />
                <url-name-pattern name="gif" />
                <url-name-pattern name="zip" />
                <url-name-pattern name="txt" />
            </content-type-pattern-list>

            <include-pattern-list>
                <!-- URLs that don't match these patterns won't be crawled: -->
                <url-obj-pattern host="*.ibm.com" />
                <!-- sbo - url-obj-pattern query="*OpenDocument*" / -->
                <!-- sbo - url-obj-pattern query="*OpenView*" / -->
                <!-- url-obj-pattern query="*OpenDocument =>
OpenDocument&amp;ExpandAll*" / -->
                <!-- url-obj-pattern query="*OpenView =>
OpenView&amp;ExpandAll&amp;Count=999999*" / -->
            </include-pattern-list>

            <exclude-pattern-list>
                <!-- URLs that match these patterns won't be crawled: -->
                <!-- skip these common patterns in our intranet -->
                <url-obj-pattern file="*news*" />
                <url-obj-pattern file="*search*" />
                <url-obj-pattern file="*/afs/" />
                <url-obj-pattern file="*/.../" />
                <url-obj-pattern file="*bluepages*" />
                <!-- skip personal home pages -->
                <url-obj-pattern file="*/~*" />
                <!-- skip SOCKS: no URL should specify this directly -->
                <url-regex-pattern match=".*:1080/.*" />
                <!-- skip gateways: recommended for mere mortals -->
                <url-regex-pattern match=".*[?\\=\\+\\;\\%&quot;&amp;].*" />
                <!-- else crawl gateways as configured here... -->
                <!-- skip Domino -->
                <url-obj-pattern file="*.nsf*" />
                <!-- else crawl Domino: allow only OpenDocument -->
                <url-obj-pattern query="*OpenServer*" />
                <url-obj-pattern query="*OpenDatabase*" />
                <url-obj-pattern query="*OpenElement*" />
                <url-obj-pattern query="*OpenView*" />
                <url-obj-pattern query="*OpenAbout*" />
                <url-obj-pattern query="*OpenHelp*" />
                <url-obj-pattern query="*OpenIcon*" />
            </exclude-pattern-list>
        </crawler-config>
    </group>
</group-list>

```

```

        <url-obj-pattern    query="*OpenForm*" />
        <url-obj-pattern    query="*OpenNavigator*" />
        <url-obj-pattern    query="*OpenAgent*" />
        <url-obj-pattern    query="*CreateDocument*" />
        <url-obj-pattern    query="*DeleteDocument*" />
        <url-obj-pattern    query="*EditDocument*" />
        <url-obj-pattern    query="*SaveDocument*" />
        <url-obj-pattern    query="*SearchSite*" />
        <url-obj-pattern    query="*SearchView*" />
        <url-obj-pattern    query="*&login*" />
        <url-obj-pattern    query="*Command*" />
        <!-- crawl Domino: avoid OpenDocument permutations -->
        <url-obj-pattern    query="*ExpandSection*" />
        <url-obj-pattern    query="*Navigate*" />
        <url-obj-pattern    query="*Start*" />
    <!-- -->

</exclude-pattern-list>
</crawler-config>

<summarizer-config>
<!-- Copy2Rdf is the default summarizer.  For these types, use: -->

    <resource-handler content-type="*htm*"
                        summarizable="EipHtmlSummarizable"
                        summary-maker="EipHtmlRawSummaryMaker" />
    <resource-handler content-type="*pdf"
                        summarizable="InsoSummarizable"
                        summary-maker="InsoSummaryMaker" />
</summarizer-config>
v    </group>
    </group-list>
</gcs-config>

```

---

## IBM Web Crawler 로그 분석 파일 예

```
D:\gcs\run\log>perl loganalysis.pl log.txt
```

```
Elapsed time in Log.txt for 7710 lines was 1.84 minutes.
```

```

GCS was configured for 20 crawlers
  999 total crawls attempted
  137 - total crawl failures:
      21 GCSHttpConnection.ABANDONING
      12 GCSHttpConnection.CONNECT_ERROR
      16 GCSHttpConnection.UNKNOWN_HOST
       4 HTTP 403
      29 HTTP 404
       2 HTTP 500
       8 HTTP 599
       1 Read timed out
      39 Robots not allowed
       4 over max redirects
       1 unknown protocol

-----
      862 = successfully crawled
       0 - unchanged since earlier crawl
-----
      862 = new or changed

```

468 crawled per minute

GCS was configured for 5 summarizers

855 total summaries attempted

0 - total summary failures:

-----

855 = successfully summarized

144 gcs.summaries.list.host

855 gcs.summaries.list.resource

465 summarized per minute

GCS successfully crawled 134 servers to obtain 862 URL:

afqa0854.mop.ibm.com: 15  
als1f1.yamato.ibm.com: 1  
apache.btv.ibm.com: 1  
apc.endicott.ibm.com: 2  
as400service.ibm.com: 1  
atlas.bocaraton.ibm.com: 1  
autoproxy.ibm.com: 1  
cer.si.ibm.com: 1  
commerce.www.ibm.com: 1  
crmweb.boulder.ibm.com: 3  
d02ntcl01.ibm.com: 1  
dacs.endicott.ibm.com: 1  
duke.toraix.can.ibm.com: 1  
ebcweb.austin.ibm.com: 1  
ecspubs.ibmus2.ibm.com: 5  
edaw3.fishkill.ibm.com: 1  
endwww.endicott.ibm.com: 1  
gcs.stl.ibm.com: 1  
gustwick.austin.ibm.com: 1  
ibmfnsys.somers.hqregion.ibm.com: 1  
ibmpny1.somers.hqregion.ibm.com: 2  
ifw-www.mul.ie.ibm.com: 1  
iplswww.nas.ibm.com: 2  
itirc.ibm.com: 1  
logosite.services.ibm.com: 1  
lt.lahulpe.ibm.com: 17  
messaging.ibm.com: 1  
mrsmrn04.leeds.uk.ibm.com: 1  
online.lahulpe.ibm.com: 1  
page.sg.ibm.com: 1  
procure.sbyl.ibm.com: 1  
reso.somers.hqregion.ibm.com: 1  
risctal.leipzig.de.ibm.com: 1  
rrhhar.argentina.ibm.com: 1  
seashore.stl.ibm.com: 1  
secureway.raleigh.ibm.com: 15  
service.software.ibm.com: 1  
software.ibmus2.ibm.com: 1  
techcenter.austin.ibm.com: 1  
tr2.fishkill.ibm.com: 8080: 1  
ucd.torolab.ibm.com: 1  
usmweb.boulder.ibm.com: 1  
w3-1.ibm.com: 32  
w3-2.ibm.com: 3



w3-3.ibm.com: 108  
w3-5.ibm.com: 4  
w3.a-nz.au.ibm.com: 1  
w3.academy.ibm.com: 1  
w3.almaden.ibm.com: 2  
w3.alphaworks.ibm.com: 1  
w3.ap.ibm.com: 1  
w3.asca.ibm.com: 7  
w3.austin.ibm.com: 3  
w3.boulder.ibm.com: 1  
w3.br.ibm.com: 1  
w3.btv.ibm.com: 1  
w3.can.ibm.com: 40  
w3.chq.ibm.com: 4  
w3.coc.ibm.com: 1  
w3.corporatetechnology.ibm.com: 1  
w3.cupertino.ibm.com: 1  
w3.dds.dfw.ibm.com: 17  
w3.demopkg.ibm.com: 4  
w3.design.ibm.com: 1  
w3.developer.ibm.com: 3  
w3.education.ibm.com: 1  
w3.emea.ibm.com: 14  
w3.enterlib.ibm.com: 7  
w3.finsys.ibm.com: 1  
w3.gcg.ibm.com: 1  
w3.globalfinancing.de.ibm.com: 1  
w3.hakozaki.ibm.com: 1  
w3.houston.ibm.com: 1  
w3.hursley.ibm.com: 5  
w3.iabc.ibm.com: 1  
w3.ibm.com: 180  
w3.ibmfax.ibm.com: 1  
w3.ibm1a.ibm.com: 14  
w3.isicc.de.ibm.com: 1  
w3.itso.ibm.com: 1  
w3.japan.ibm.com: 1  
w3.knowledge.raleigh.ibm.com: 1  
w3.linux.ibm.com: 1  
w3.marketiq.ibm.com: 1  
w3.micro.ibm.com: 2  
w3.mtlisc.can.ibm.com: 1  
w3.munich.ibm.com: 1  
w3.ode.raleigh.ibm.com: 1  
w3.paylink.au.ibm.com: 1  
w3.pisc.uk.ibm.com: 1  
w3.pl.ibm.com: 1  
w3.printers.ibm.com: 1  
w3.pssc.mop.ibm.com: 1  
w3.pssed.au.ibm.com: 1  
w3.raleigh.ibm.com: 3  
w3.rchland.ibm.com: 1  
w3.research.ibm.com: 3  
w3.reserve.ibm.com: 1  
w3.rs6000.ibm.com: 1  
w3.security.ibm.com: 1  
w3.software.ibm.com: 6  
w3.ssd.ibm.com: 1

w3.stl.ibm.com: 1  
w3.techline.ibm.com: 1  
w3.techsupp.yamato.ibm.com: 1  
w3.torolab.ibm.com: 2  
w3.usergroup.ibm.com: 1  
w3.vendor.pok.ibm.com: 1  
w3.viewblue.ibm.com: 1  
w3.watson.ibm.com: 2  
w3.wdg.uk.ibm.com: 1  
w3.ytal.yasu.ibm.com: 1  
w3.zurich.ibm.com: 1  
w3chq.disbursements.ibm.com: 1  
w3is.lagaude.ibm.com: 1  
w3md.btv.ibm.com: 1  
w3ssd.mainz.de.ibm.com: 1  
w3vm.demopkg.ibm.com: 1  
widweb.raleigh.ibm.com: 1  
wtscpok.itso.ibm.com: 1  
wwas.raleigh.ibm.com: 1  
www-1.ibm.com: 63  
www-3.ibm.com: 4  
www-4.ibm.com: 86  
www.almaden.ibm.com: 1  
www.as400.ibm.com: 1  
www.chips.ibm.com: 1  
www.ibm.com: 52  
www.ieg.ibm.com: 1  
www.patents.ibm.com: 1  
www.pc.ibm.com: 2  
www.rs6000.ibm.com: 23  
www.software.ibm.com: 9  
www.storage.ibm.com: 1  
www.watson.ibm.com: 1

GCS timed out 1 times:

w3-3.ibm.com: 1

GCS ignored 42 URL prohibited by robots.txt:

reso.somers.hqregion.ibm.com: 1  
w3.education.ibm.com: 1  
w3.rchland.ibm.com: 34  
w3.zurich.ibm.com: 1  
www.ibm.com: 5

GCS skipped 3846 URL (requires gcs.url logging)

59 specified an unsupported protocol:

protocol not supported gopher: 6

protocol not supported mailto: 53

1206 had content-types (lower or UPPER case, > 10) that were not included

.2: 12

.faq: 13

.1: 14

.asp: 16

.cgi: 21

.shtml: 90

.pl: 92

.gif: 157

.nsf: 160

```
.jpg: 214
.css: 240
516 URL were on servers and/or paths that were not included
2065 were excluded for these reasons:
  URL longer than 254: 1
  excluded by rule 1: 1210
  excluded by rule 2: 854
```



---

## 제 8 장 텍스트 검색 및 QBIC<sup>®</sup> 사용

이 부록의 첫 번째 절에서는 텍스트 검색과 QBIC(Query by Image Content)의 구성 및 사용 방법에 대해 설명합니다. EIP 설치 시 Content Manager 버전 7.1 커넥터를 선택하는 경우에만 두 가지 기능을 사용할 수 있습니다. 이 부록의 두 번째 절에서는 샘플 응용프로그램에 사용되는 이미지 데이터 및 샘플 텍스트 로드와 관련된 정보에 대해 설명합니다.

---

### 텍스트 검색 엔진을 사용한 문서 검색

텍스트 검색과 Content Manager 버전 7.1 서버를 통합하여, Content Manager에 저장된 문서를 자동으로 색인화, 검색 및 조회할 수 있습니다. 사용자는 단어나 구문을 검색하여 문서를 찾을 수 있습니다. 텍스트 검색 서버는 1바이트 문자 세트 및 2바이트 문자 세트를 모두 지원하며 AIX 및 Windows에서 모두 실행됩니다.

텍스트 검색에는 XML, HTML 및 태그가 있는 ASCII 문서에 대한 구조화 문서 지원이 포함되며, 지정된 문서 절 내에서 용어를 검색할 수 있습니다. 내포된 절에 있는 데이터를 검색할 수 있습니다. 전체 XML 구문으로 검색할 수 있습니다. 예를 들어, 제목 안에서 IBM을 검색하고 특정 절에 있는 제목 안에서 IBM을 검색할 수 있습니다. DTD 경로를 지정하면 DTD에 대한 참조가 해당 문서의 메타데이터로 저장된다고 가정하고 각 문서에 적합한 DTD를 동적으로 사용할 수 있습니다.

텍스트 검색을 사용한 EIP 시스템 계획 및 설치에 대한 자세한 내용은 *Enterprise Information Portal 계획 및 설치*를 참조하십시오.

### 텍스트 검색 서버 사용

텍스트 검색 서버를 사용하려면 멀티플랫폼용 IBM Content Manager 관리 클라이언트를 시작하기 전에 서버 관리를 사용해야 합니다. 관리를 사용하려면 다음을 수행하십시오.

1. 멀티플랫폼용 IBM Content Manager 라이브러리 서버를 시작하십시오.  
라이브러리 서버로 색인 클래스를 구축하십시오.
2. 다음을 입력하여 텍스트 검색 서버가 설치된 워크스테이션에서 이 서버를 시작하십시오.

```
imlss -start dlinst
```

여기서 *dlinst*는 설치 시 또는 *imlcfgsv* 명령 유틸리티를 사용할 때 선택된 텍스트 검색 서버 인스턴스의 이름입니다.

---

## QBIC(Query By Image Content)를 사용하여 이미지 검색

이 절에서는 QBIC(Query By Image Content)에 대해 소개하며 QBIC 구성 및 사용 방법에 대해 설명합니다. QBIC 기능은 Content Manager 버전 7.1 커넥터를 설치한 경우에만 사용 가능합니다. QBIC는 Windows 및 AIX 운영 체제와 호환 가능합니다.

### 이미지 검색 소개

이미지 검색 서버는 IBM의 QBIC(Query By Image Content) 기술을 사용하여 색상 및 텍스처 같은 특정 비주얼 등록 정보로 오브젝트를 검색할 수 있도록 도와 줍니다. 이미지 검색 서버는 이미지를 분석하여 이미지 정보를 데이터베이스에 저장합니다. 그런 다음, 사용자는 이미지의 비주얼 등록 정보를 사용하는 이미지 조회를 실행하여 단어로 설명하지 않고도 색상, 텍스처 및 그 위치를 일치시킬 수 있습니다. 이미지 및 멀티미디어 데이터의 보다 강력한 검색을 위해 콘텐츠 기반 조회와 텍스트 및 키워드 검색을 결합할 수 있습니다.

각 이미지 검색 서버에는 하나 이상의 이미지 검색 데이터베이스를 포함하는 데이터 디렉토리가 있으며, 이미지 검색 데이터베이스는 이미지 검색 카탈로그를 보유합니다. 이미지 검색 카탈로그는 이미지 컬렉션의 비주얼 기능에 대한 데이터를 저장합니다. 실제 이미지 오브젝트는 멀티플랫폼용 IBM Content Manager 시스템의 오브젝트 서버에 저장됩니다. 이미지 검색 서버는 AIX 및 Windows에서 실행됩니다.

이미지 검색 설치에 대한 자세한 내용은 *Enterprise Information Portal* 계획 및 설치를 참조하십시오.

### 이미지 검색 설정

이 지시사항은 Content Manager Version 7.1 커넥터를 선택한 경우에 자동으로 설치되는 이미지 검색을 설치한 후, 적용됩니다. 이미지 검색 설정은 다음으로 구성됩니다.

1. 환경 설정
2. 이미지 검색 서버 구성
3. 이미지 검색 클라이언트 구성
4. 샘플 이미지 데이터 로드

**AIX에서 설치 마법사를 사용한 경우:** 구성을 실행하거나 스크립트를 설정하거나 서버 구성 명령을 발생하지 않아도 됩니다. 마법사가 사용자를 위해 이 작업을 완료합니다.

**Windows에 설치하는 경우:** 이 작업을 완료해야 합니다.

#### 환경 설정

서버 및 클라이언트 기계에서 모두 이 절에 있는 환경 설정 작업을 완료하십시오. 이미지 검색 서버에는 다음 환경 변수가 필요합니다.

## QBICTOP

이미지 검색 구성 시 파일 이름을 해석합니다.

## QbicImagePath

서버 이미지 파일에서 파일 이름을 해석합니다.

## QbicMaskPath

서버 마스크 파일에서 파일 이름을 해석합니다.

## QbicSketchPath

서버 스케치 파일에서 파일 이름을 해석합니다.

## QbicTextPath

서버 텍스트 파일에서 파일 이름을 해석합니다.

이미지 검색 클라이언트에는 QBICTOP 환경 변수만 필요합니다.

**AIX 예:** AIX의 경우, 설정 스크립트를 생성하는 구성 스크립트를 실행한 후, 설정 스크립트를 실행하여 환경을 설정하십시오.

1. 다음 구성 스크립트를 실행하십시오.

```
/usr/lpp/cmb/bin/frnconfig.iss QBICTOP
```

여기서 QBICTOP는 제어 파일(\*.ini)의 디렉토리 경로입니다. QBICTOP를 /user1/cmb/qbic로 설정하십시오. 여기서 /user1은 이미지 검색 관리 사용자 ID의 홈 디렉토리입니다. 이미지 검색 사용자 ID는 이 디렉토리에 대한 읽기/쓰기 액세스를 갖고 있어야 합니다.

이 스크립트는 frnsetup.iss라는 설정 스크립트를 생성합니다.

2. 이미지 검색 사용자 ID의 홈 디렉토리에서 다음을 실행하십시오.

```
.. ./frnsetup.iss
```

이 스크립트는 이미지 검색 서버와 클라이언트의 환경 변수를 채웁니다.

**Windows 예:** 환경 변수를 설정하려면 다음을 수행하십시오.

1. 시작 --> 설정 --> 제어판을 선택하십시오.
2. 시스템을 두 번 누르십시오.
3. 환경 탭을 누르십시오.
4. 적당한 필드에 값을 입력하여 130 페이지의 표 8에 표시된 변수와 값을 설정하고 설정을 누르십시오.

**요구사항:** 이미지 검색 클라이언트에는 QBICTOP 변수만 필요합니다. 클라이언트 환경에는 QBICTOP 변수만 설정하십시오.

표 8. 이미지 검색 환경 변수

변수	값
QBICTOP	d:\cmbroot\iss
QbicImagePath	d:\cmbroot\iss
QbicMaskPath	d:\cmbroot\iss
QbicSketchPath	d:\cmbroot\iss
QbicTextPath	d:\cmbroot\iss

여기서 d:는 이미지 검색이 설치된 드라이브입니다.

## 이미지 검색 서버 구성

이미지 검색 서버를 시작하기 전에 이미지 검색 서버를 구성해야 합니다. 서버 구성은 초기 구성 완료와 연결 검증으로 구성됩니다.

이미지 검색 서버를 구성하려면 다음을 수행하십시오.

1. qbicadm을 입력하여 명령 해석기를 시작하십시오.
2. **config server** 명령을 입력하십시오. 예를 들어, 다음과 같습니다.

```
config server LIBSRVRN FRNADMIN PASSWORD 9999
```

여기서 LIBSRVRN은 라이브러리 서버 이름, FRNADMIN은 Content Manager 사용자 ID, PASSWORD는 Content Manager 암호, 9999는 이미지 검색 서버의 포트 번호입니다.

자세한 내용은 『연결 검증』을 참조하십시오.

## 이미지 검색 클라이언트 구성

이미지 검색 시스템 관리 프로그램을 포함하여 이미지 검색 클라이언트를 시작하기 전에 이미지 검색 클라이언트를 구성해야 합니다. Content Manager 시스템 관리에는 한 개의 별명을 할당해야 합니다. 연결을 검증하여 구성을 테스트하십시오.

**별명 할당:** 이미지 검색 클라이언트로 작용하는 이미지 검색 Content Manager 시스템 관리를 사용하기 전에 최소한 한 개의 서버 별명을 할당해야 합니다.

서버 별명을 할당하려면 다음을 수행하십시오.

1. qbicadm을 입력하여 명령 해석기를 시작하십시오.
2. **add alias** 명령을 입력하십시오. 예를 들어, 다음과 같습니다.

```
add alias QBICSRV HOSTNAME 9999
```

여기서 QBICSRV는 별명 이름, HOSTNAME은 이미지 검색 서버의 호스트 이름, 9999는 이미지 검색 서버의 포트 번호입니다.

**연결 검증:**



### 중요사항:

1. 이미지 검색 서버에 연결하려면 라이브러리 서버가 실행 중이어야 합니다.
2. 이미지 검색 시스템 관리에는 기존의 Content Manager 사용자 ID가 필요합니다. 라이브러리 서버에 성공적으로 연결하려면 이미지 검색 사용자 ID와 라이브러리 서버 사용자 ID가 같아야 합니다. 이 사용자 ID의 기본값은 frnadmin입니다. 이 값을 변경하는 경우, ID가 일치하는지 확인하십시오.

연결을 검증하려면 다음을 수행하십시오.

1. 이미지 검색 서버를 구성하고 별명을 추가한 후, 서버 명령행에서 commsrv를 입력하여 서버를 시작하십시오.
2. 명령 해석기를 시작하려면 qbicadm을 입력하십시오.
3. 명령 해석기에서 connect 명령을 입력하십시오.

```
connect QBICSRV FRNADMIN PASSWORD
```

여기서 QBICSRV는 별명 이름, FRNADMIN은 Content Manager 사용자 ID, PASSWORD는 Content Manager 암호입니다.

연결 완료 후, 라이브러리 서버는 LIBSRVRN입니다라는 메시지가 표시됩니다.

4. 서버에서 연결 해제하려면 disconnect를 입력하십시오.
5. 명령 해석기를 종료하려면 quit를 입력하십시오.

---

## 샘플 데이터 로드 및 색인화

이 절에서는 샘플 응용프로그램에 사용하는 샘플 텍스트와 이미지 데이터의 로드 및 색인 방법에 대해 설명합니다. 이 절의 내용은 Content Manager 버전 7.1 커넥터를 설치하고 텍스트 검색 옵션을 선택한 경우에만 적용됩니다.

Enterprise Information Portal CD에는 여러 샘플 로더가 제공됩니다. 이 절에서는 LoadSampleTSQBICDL 샘플 로더를 사용하여 이미지와 텍스트 데이터를 모두 로드하는 방법에 대해 설명합니다. 텍스트와 이미지 검색 데이터를 별도로 로드하여 두 기능이 모두 올바르게 작동하도록 할 수 있습니다.

### 샘플 데이터를 로드하기 전에

로더 프로그램을 실행하기 전에 다음을 수행해야 합니다.

1. EIP 관리 클라이언트에 로그인하십시오. 시작 --> 프로그램 --> 멀티플랫폼용 **Enterprise Information Portal 8.2** --> 관리를 누르십시오.
2. 데이터베이스를 선택한 후, 올바른 사용자 ID 및 암호를 사용하여 로그인하십시오. 기본 데이터베이스 icmn1sdb를 선택한 경우, **icmadmin**을 사용자 ID로 입력하고 암호 필드에 암호를 입력하십시오. 다른 데이터베이스를 사용하는 경우, 적용 가능한 사용자 ID를 입력하십시오.

3. Content Manager 시스템 관리 프로그램을 사용하여 라이브러리 서버 구성을 작성하십시오. 이러한 작업에 대한 지원은 시스템 관리 프로그램 온라인 도움말을 참조하십시오.
4. 다음 단계를 완료하여 라이브러리 서버 구성의 액세스 등록 정보를 변경하십시오.
  - a. 새 구성을 오른쪽 마우스 단추로 누르고 등록 정보를 눌러 등록 정보 노트북을 여십시오.
  - b. 액세스 탭을 누르십시오.
  - c. 워크스테이션의 무제한 세션 라디오 단추를 누르십시오.

## 텍스트 검색 색인 작성

데이터를 로드하기 전에 텍스트 샘플을 색인화하는 데 사용할 수 있는 빈 텍스트 검색 색인을 작성해야 합니다. **팁:** Content Manager 버전 6.1 또는 버전 7.1 서버에서만 텍스트 검색 색인을 작성할 수 있습니다.

텍스트 검색 색인을 작성하려면 다음을 수행하십시오.

1. 다음 명령을 사용하여 텍스트 검색 서버가 설치된 워크스테이션에서 텍스트 검색 서버를 시작하십시오.

```
imlss -start dlinst
```

여기서 dlinst는 설치 시에 또는 imlcfgsv 명령 유틸리티를 사용할 때 선택된 텍스트 검색 서버 인스턴스의 이름입니다.

2. Content Manager 시스템 관리 프로그램을 시작하고 이 프로그램에 로그인하십시오.
3. 왼쪽 맨 위 분할 영역에 있는 목록에서 텍스트 검색을 선택하십시오.
4. 왼쪽 분할 영역에서 검색 서버를 두 번 누르십시오.
5. **TM**을 두 번 누르십시오. TM은 텍스트 검색 서버의 검색 서버 별명입니다.
6. 왼쪽 분할 영역에서 색인 폴더를 두 번 누르십시오. RC\_EMPTY\_LIST 메시지가 표시되는 경우, 메뉴 막대에서 선택 --> 신규를 눌러 색인을 작성하십시오.
7. 새 색인 창에서 색인을 정의하십시오. 각 필드의 상세한 설명을 보려면 도움말을 누르십시오.

예를 들어, 다음과 같습니다.

### Windows의 경우:

이름    TMINDEX

유형    Precise

색인 파일

x:\cmbroot\ts\index\tmlindex. 여기서 x는 설치 드라이브입니다. 경로가 없는 경우, 해당 경로가 작성됩니다.

#### 색인 작업 파일

x:\cmbroot\ts\work\tmlindex. 여기서 x는 설치 드라이브입니다. 경로가 없는 경우, 해당 경로가 작성됩니다.

#### 정보 항목

Content Manager 라이브러리 서버의 이름.

클라이언트 및 서버 기본 DLL 이름을 변경하지 마십시오.

#### AIX의 경우:

이름 TMINDEX

유형 Precise

#### 색인 파일

/home/cltadmin/tsindex/index/tminindex. 경로가 없는 경우, 해당 경로가 작성됩니다.

#### 색인 작업 파일

/home/cltadmin/tsindex/work/tminindex. 경로가 없는 경우, 해당 경로가 작성됩니다. 사용자에게 디렉토리에 쓸 권한이 있는지 확인하십시오.

#### 정보 항목

Content Manager 라이브러리 서버의 이름.

8. 확인을 누르십시오.

9. TMINDEX를 두 번 눌러 TMINDEX 관리 노트북을 여십시오.

## 이미지 검색 데이터베이스, 카탈로그 및 기능 작성

샘플 텍스트 검색 데이터의 텍스트 검색 색인을 작성한 후, 샘플 이미지 데이터의 이미지 검색 데이터베이스와 카탈로그를 작성해야 합니다.

이미지 검색 데이터베이스, 카탈로그 및 기능을 작성하려면 다음을 수행하십시오.

1. 다음 명령을 입력하여 이미지 검색 서버가 설치된 워크스테이션에서 이미지 검색 서버를 시작하십시오.

```
commsrv
```

2. Content Manager 시스템 관리 프로그램을 시작하고 이 프로그램에 로그인하십시오.
3. 왼쪽 맨 위 분할 영역에서 이미지 검색을 선택하십시오.
4. 왼쪽 분할 영역에서 이미지 검색 서버를 누르십시오.
5. QBICSRV를 누르십시오.

여기서 QBICSRV는 설치 중 사용자가 지정한 이미지 검색 서버 이름입니다.

6. 왼쪽 분할 영역에서 데이터베이스를 오른쪽 마우스 단추로 누르고 새 데이터베이스를 선택하십시오.
7. 새 데이터베이스 창에서 이름 필드에 SAMPLEDB를 입력하고 확인을 누르십시오.
8. 왼쪽 분할 영역에서 데이터베이스를 눌러 왼쪽 분할 영역에 **SAMPLEDB** 아이콘을 표시하십시오.
9. **SAMPLEDB**를 누르십시오.
10. 왼쪽 분할 영역에서 카탈로그를 오른쪽 마우스 단추로 누르고 새 카탈로그를 누르십시오.
11. 새 카탈로그 창에서 이름 필드에 SAMPLECAT를 입력하고 확인을 누르십시오.
12. 왼쪽 분할 영역에서 카탈로그를 눌러 **SAMPLECAT** 아이콘을 표시하십시오.
13. **SAMPLECAT**를 누르십시오.
14. 왼쪽 분할 영역에서 기능을 오른쪽 마우스 단추로 누르고 새 기능을 누르십시오.
15. 새 기능 창의 이름 필드에서 각 기능을 선택하고 적용을 누르십시오. 네 가지 기능이 모두 선택되면 다음 메시지가 표시됩니다.  
가능한 모든 기능이 카탈로그에 추가되었습니다.
16. 확인을 누르십시오.
17. 취소를 누르십시오.

## 로더 프로그램 실행

샘플 데이터를 로드하여 텍스트 및 이미지 검색을 테스트할 수 있습니다.

샘플 이미지 데이터는 다음 파일에 있습니다.

**Windows의 경우:**

```
x:\cmbroot\samples\java\d1\samples.jar
```

**AIX의 경우:**

```
/usr/lpp/cmb/samples/java/d1/samples.jar
```

샘플 로더 프로그램은 데이터를 Content Manager로 로드하여 색인화합니다. 프로그램 실행 구문 규칙에 대한 자세한 내용은 소스 프로그램의 서문을 참조하십시오. 샘플 로더 프로그램은 다음과 같습니다.

**Windows의 경우:**

```
x:\cmbroot\samples\java\d1\LoadSampleTSQBICDL.jar
```

**AIX의 경우:**

```
/usr/lpp/cmb/samples/java/d1/LoadSampleTSQBICDL.jar
```

샘플 데이터 로더 프로그램을 실행하려면 다음을 수행하십시오.

1. 다음을 입력하여 .jar 파일을 압축 해제하십시오.

```
jar -xvf samples.jar
```

파일은 적절한 디렉토리로 압축 해제됩니다.

2. 다음 작업을 완료하여 워크스테이션 환경 변수가 샘플 로더 프로그램을 컴파일하도록 설정하십시오.

**Windows의 경우:**

- a. 텍스트 편집기에서 x:\cmbroot\cmbenv71.bat를 열고 다음과 같이 처음 세 행을 변경하여 워크스테이션 환경 변수를 설정하십시오.

```
set CMBROOT = e:\cmbroot
set DB2HOME = e:\sqllib
set JAVAHOME = d:\jdk117
```

- b. cmbenv71.bat를 저장하고 다음을 입력하여 환경 변수를 설정하십시오.

```
cmbenv71
```

**AIX의 경우:**

- a. /usr/lpp/cmb/bin/으로 가서 다음을 입력하여 설정을 실행하십시오.

```
. /cmbenv71.sh
```

- b. 모든 사용자가 /usr/lpp/cmb/samples/java/d1/ 아래의 서브디렉토리 및 샘플 파일을 기록할 수 있는지 확인하십시오.

3. 다음 대소문자 구분 명령을 입력하여 로더 프로그램을 컴파일하십시오.

```
javac LoadSampleTSQBICDL.java
```

4. **요구사항:** 로더 프로그램을 시작하기 전에 다음 서버가 실행 중이어야 합니다.

- 라이브러리 서버
- 오브젝트 서버
- 텍스트 검색 서버
- 이미지 검색 서버

Content Manager의 자국어 버전을 실행 중인 경우, 로더 프로그램을 실행하기 전에 FRNDEFLANG 변수를 ENU로 설정하십시오. 환경 변수 설정을 위한 AIX 명령은 다음과 같습니다. export FRNDEFLANG=ENU

5. 다음을 입력하여 로더 프로그램으로 샘플 데이터를 로드하십시오.

```
java LoadSampleTSQBICDL sampleQBIC.dat load.log frnadmin password LIBSRVRN
```

여기서 사용자 ID는 frnadmin, 암호는 password, 라이브러리 서버는 LIBSRVRN입니다.

6. load.log를 점검하여 샘플 데이터가 제대로 로드되었는지 확인하십시오.

샘플 데이터 로드를 완료한 후, Content Manager 시스템 관리 프로그램이나 텍스트 검색 명령 행 도구를 사용하여 샘플 텍스트 데이터를 색인화하십시오.

## 샘플 텍스트 데이터 색인화

샘플 데이터를 색인화하려면 다음을 수행하십시오.

1. Content Manager 시스템 관리 프로그램을 시작하고 이 프로그램에 로그인하십시오.
2. 왼쪽 맨 위 분할 영역에 있는 목록에서 텍스트 검색을 선택하십시오.
3. 검색 서버를 두 번 누르십시오.
4. **TM**을 두 번 누르십시오. TM은 텍스트 검색 서버의 검색 서버 별명입니다.
5. 새 텍스트 색인을 오른쪽 마우스 단추로 누르고 등록 정보를 누르십시오.
6. 등록 정보 노트북의 명시적 페이지에서 화면 갱신을 누르십시오.
7. 색인 수 필드는 로더 프로그램으로 로드된 문서의 수를 표시합니다.
8. 색인을 눌러 파일을 색인화하십시오.
9. 몇 분 후, 화면 갱신을 눌러 기본 문서 색인 필드에서 성공적으로 색인화된 문서 수를 보십시오.

데이터를 색인화한 후, 샘플 Java 응용프로그램을 사용하여 컬렉션을 조회하거나 `imlsrch` 명령행 도구를 사용하여 간단한 조회를 실행할 수 있습니다.

---

## 제 9 장 문서 형식

---

### Information Mining 문서 형식

이 부록에서는 Information Mining이 지원하는 문서 형식을 설명합니다.

#### Word 처리: 일반

ANSI 텍스트(7 & 8 비트)	전체 버전
ASCII 텍스트(7 & 8 비트 버전 사용 가능)	전체 버전
HTML	3.0까지의 버전(몇 가지 제한사항)
IBM FFT	전체 버전
IBM Revisable Form Text	전체 버전
Microsoft Rich Text Format (RTF)	전체 버전
Unicode Text	전체 버전

#### Word 처리: DOS

DEC WPS Plus (DX)	4.0까지의 버전
DEC WPS Plus (WPL)	4.1까지의 버전
DisplayWrite <sup>®</sup> 2 & 3 (TXT)	전체 버전
DisplayWrite 4 & 5	릴리스 2.0까지의 버전
Enable	버전 3.0, 4.0 및 4.5
First Choice	3.0까지의 버전
Framework	버전 3.0
IBM Writing Assistant	버전 1.01
Lotus Manuscript	2.0까지의 버전
MASS11	8.0까지의 버전
Microsoft Word	6.0까지의 버전
Microsoft 작업	2.0까지의 버전

<b>MultiMate</b>	4.0까지의 버전
<b>Navy DIF</b>	전체 버전
<b>Nota Bene</b>	버전 3.0
<b>Office Writer</b>	버전 4.0 - 6.0
<b>PC-File Letter</b>	5.0까지의 버전
<b>PC-File+ Letter</b>	3.0까지의 버전
<b>PFS:Write</b>	버전 A, B 및 C
<b>Professional Write</b>	2.1까지의 버전
<b>Q&amp;A</b>	버전 2.0
<b>Samna Word</b>	4.0까지의 버전
<b>SmartWare II</b>	버전 1.02
<b>Sprint</b>	버전 1.0
<b>Total Word</b>	버전 1.2
<b>Volkswriter 3 &amp; 4</b>	1.0까지의 버전
<b>Wang PC (IWP)</b>	2.6까지의 버전
<b>WordMARC</b>	Composer Plus까지의 버전
<b>WordPerfect</b>	6.1까지의 버전
<b>WordStar</b>	7.0까지의 버전
<b>WordStar 2000</b>	3.0까지의 버전
<b>XyWrite</b>	III Plus까지의 버전

## **Word 처리: International**

<b>JustSystems Ichitaro</b>	버전 5.0, 6.0, 8.0, 9.0 및 10.0
-----------------------------	------------------------------

## **Word 처리: Windows**

<b>AMI/AMI Professional</b>	3.1까지의 버전
<b>Windows용 Corel WordPerfect</b>	9.0까지의 버전
<b>JustWrite</b>	3.0까지의 버전
<b>Legacy</b>	1.1까지의 버전
<b>Lotus WordPro (Win32 / Intel 플랫폼)</b>	SmartSuite® 96, 97 및 Millennium



**Lotus WordPro(Unix 플랫폼 - 텍스트 경우만)**

SmartSuite 97 및 Millennium

**Microsoft Windows Works** 4.0까지의 버전

**Microsoft Windows Write** 3.0까지의 버전

**Microsoft Word 97** Word 97

**Microsoft Word 2000** Word 2000

**Windows용 Microsoft Word** 7.0까지의 버전

**Microsoft WordPad** 전체 버전

**Novell Perfect Works** 버전 2.0

**Windows용 Novell WordPerfect**

7.0까지의 버전

**Professional Write Plus** 버전 1.0

**Windows용 Q&A Write** 버전 3.0

**Windows용 WordStar** 버전 1.0

**Word 처리: Macintosh**

**Microsoft Word** 4.0 - 6.0 버전

**Microsoft Word 98** Word 98

**WordPerfect** 1.02 - 3.0 버전

**Microsoft 작업** 2.0까지의 버전

**MacWrite II** 버전 1.1

**스프레드시트 형식**

**VP Planner 3D** 버전 1.0

**Enable** 버전 3.0, 4.0 및 4.5

**First Choice** 3.0까지의 버전

**Framework** 버전 3.0

**Lotus 1-2-3<sup>®</sup> (DOS & Windows)**

5.0까지의 버전

**SmartSuite용 Lotus 1-2-3** SmartSuite 97 및 Millennium

**Lotus 1-2-3 Charts (DOS & Windows)**

5.0까지의 버전

**Lotus 1-2-3 (OS/2<sup>®</sup>)** 2.0까지의 버전

<b>Lotus 1-2-3 Charts (OS/2)</b>	2.0까지의 버전
<b>Lotus Symphony</b>	1.0, 1.1 및 2.0 버전
<b>Microsoft Excel 97</b>	Excel 97
<b>Microsoft Excel 2000</b>	Excel 2000
<b>Microsoft Excel Macintosh</b>	3.0 - 4.0, 98 버전
<b>Microsoft Excel Windows</b>	2.2 - 7.0 버전
<b>Microsoft Excel Charts</b>	2.x - 7.0 버전
<b>Microsoft Multiplan</b>	버전 4.0
<b>Microsoft Windows Works</b>	4.0까지의 버전
<b>Microsoft Works (DOS)</b>	2.0까지의 버전
<b>Microsoft Works (Mac)</b>	2.0까지의 버전
<b>Mosaic Twin</b>	버전 2.5
<b>Novell Perfect Works</b>	버전 2.0
<b>QuattroPro for DOS</b>	5.0까지의 버전
<b>Windows용 QuattroPro</b>	9.0까지의 버전
<b>PFS:Professional Plan</b>	버전 1.0
<b>SuperCalc 5</b>	버전 4.0
<b>SmartWare II</b>	버전 1.02

## 데이터베이스 형식

<b>SmartWare II</b>	버전 1.02
<b>Access</b>	2.0까지의 버전
<b>dBase</b>	5.0까지의 버전
<b>DataEase</b>	버전 4.x
<b>dBXL</b>	버전 1.3
<b>Enable</b>	버전 3.0, 4.0 및 4.5
<b>First Choice</b>	3.0까지의 버전
<b>FoxBase</b>	버전 2.1
<b>Framework</b>	버전 3.0
<b>Microsoft Windows Works</b>	4.0까지의 버전
<b>Microsoft Works (DOS)</b>	2.0까지의 버전

<b>Microsoft Works (Mac)</b>	2.0까지의 버전
<b>Paradox (DOS)</b>	4.0까지의 버전
<b>Paradox (Windows)</b>	1.0까지의 버전
<b>Personal R:BASE</b>	버전 1.0
<b>R:BASE 5000</b>	3.1까지의 버전
<b>R:BASE System V</b>	버전 1.0
<b>Q &amp; A</b>	2.0까지의 버전
<b>Reflex</b>	버전 2.0

## 표준 그래픽 형식

### **PNG - Portable Network Graphics Internet Format**

버전 1.0

### **Binary Group 3 Fax** 전체 버전

### **BMP(including RLE, ICO, CUR & os/2 DIB)**

Windows

### **CDR(TIFF 이미지가 병합된 경우)**

Coral Draw 버전 2.0 - 9.0

### **CGM - Computer Graphics Metafile**

ANSI, CALS, NIST, 버전 3.0

### **CMX - Corel Clip Art Format**

5 - 6 버전

### **DCX (multi-page PCX)** Microsoft Fax

### **DRW - Micrografx Designer**

버전 3.1

### **DRW - Micrografx Draw** 4.0까지의 버전

### **DXF (Binary and ASCII) AutoCAD Drawing Interchange Format**

14까지의 버전

### **EMF**

Windows Enhanced Metafile

### **EPS Encapsulated PostScript**

TIFF 이미지가 병합된 경우

### **FMV - FrameMaker graphics**

버전 5.0까지의 벡터 및 래스터 형식

### **FPX - Kodak Flash Pix**

특정 형식 없음

<b>GDF - IBM Graphics Data Format</b>	버전 1.0
<b>GEM - Graphics Environment Manager Metafile</b>	비트맵 및 벡터
<b>GIF - Graphics Interchange Format</b>	Compuserve
<b>GP4 - Group 4 CALS Format</b>	유형 I 및 유형 II
<b>HPGL - Hewlett Packard Graphics Language</b>	버전 2.0
<b>IMG - GEM Paint</b>	특정 버전 없음
<b>JFIF (JPEG가 TIFF 형식이 아님)</b>	전체 버전
<b>JPEG - Joint Photographic Experts Group format</b>	전체 버전
<b>MET - OS/2 PM Metafile</b>	버전 3.0
<b>PBM - Portable Bitmap</b>	특정 버전 없음
<b>Kodak Photo CD</b>	버전 1.0
<b>PCD - PCX Bitmap</b>	PC Paintbrush
<b>Perfect Works (Draw)</b>	Novell 버전 2.0
<b>PGM - Portable Graymap</b>	특정 버전 없음
<b>PIC - Lotus 1-2-3 Picture File Format</b>	특정 버전 없음
<b>PICT1 &amp; PICT2 (Raster)</b>	Macintosh Standard
<b>PIF - IBM Picture Interchange Format</b>	버전 1.0
<b>PNTG</b>	MacPaint
<b>PPM - Portable Pixmap</b>	특정 버전 없음
<b>Progressive JPEG</b>	특정 버전 없음
<b>PSP - Paintshop Pro (Win32의 경우에만)</b>	버전 5.0, 5.0.1
<b>RND - AutoShade Rendering File Format</b>	버전 2.0

**SDW Ami Draw Snapshot (Lotus)**

전체 버전

**SRS - Sun Raster File Format**

특정 버전 없음

**Targa**

Truevision

**TIFF**

6까지의 버전

**TIFF CCITT Group 3 & 4**

팩스 시스템

**VISO (버전 4의 경우만 페이지 미리보기 모드) Visio 4, 5, 2000**

Visio 4, 5, 2000

**WMF**

Windows Metafile

**WordPerfect Graphics [WPG and WPG2]**

2.0까지의 버전

**XBM - X-Windows Bitmap** x10 호환 가능

**XPM - X-Windows Pixmap** x10 호환 가능

**XWD - X-Windows Dump** x10 호환 가능

**High-end 그래픽 형식**

**PSD - Adobe Photoshop File Format**

버전 4.0

**AI - Adobe Illustrator File Format**

7.0까지의 버전

**CDR - Corel Draw**

8.0까지의 버전

**DSF - Micrografx Designer** Windows 95, 버전 6.0

**DWG - AutoCAD Native Drawing Format**

12 - 14 버전

**IGES - Initial Graphics Exchange Specification**

버전 5.1

**PDF - Portable Document Format**

일본어 PDF를 포함하는 Acrobat 버전 2.1, 3.0, 4.0

**PS - Postscript**

레벨 2

## 프리젠테이션 형식

### Macintosh용 Microsoft PowerPoint

버전 4.0, 98

### Corel Presentations

버전 8.0 및 9.0

### Novell Presentations

버전 3.0 및 7.0

### DOS용 Harvard Graphics

버전 2.x 및 3.x

### Harvard Graphics

Windows 버전

### Freelance 96

Freelance 96

### Windows 95용 Freelance

SmartSuite 97 및 Millennium

### Windows용 Freelance

버전 1.0 및 2.0

### OS/2용 Freelance

2.0까지의 버전

### Windows용 Microsoft PowerPoint

7.0까지의 버전

### Microsoft PowerPoint 97

PowerPoint 97

### Microsoft PowerPoint 2000

PowerPoint 2000

## 압축 및 인코드 형식

### ZIP PKWARE

2.0g까지의 버전

### GZIP

특정 버전 없음

### LZA Self Extracting Compress

특정 버전 없음

### LZH Compress

특정 버전 없음

### Microsoft Binder

버전 7.0, 바인더 97

### MIME (텍스트 메일)

특정 버전 없음

### UUEncode

특정 버전 없음

### UNIX<sup>®</sup> Compress

특정 버전 없음

### UNIX TAR

특정 버전 없음

## 기타

### vCard Electronic Business Card

버전 2.1

### Executable (EXE, DLL)

특정 버전 없음

### Windows NT용 실행 파일

특정 버전 없음

**MSG** (텍스트의 경우만)      Microsoft Outlook 메일 형식  
**Microsoft Project** (텍스트의 경우만)  
Project 98





---

## 제 10 장 권한 관리

이 장에서는 EIP 권한 관리 기능을 소개하며 권한 관리 개념을 설명합니다. 등록 정보를 보호하기 위해 사용할 수 있는 표시 기법을 설명합니다.

---

### 지식 등록 정보 보호

디지털 양식으로 저장된 멀티미디어 오브젝트는 지식 등록 정보입니다. 이 오브젝트의 보호는 비즈니스 목표에 중요할 수 있으며, 특히 복사가 상대적으로 쉬운 WWW(World Wide Web)에 이 오브젝트가 있을 때 그러합니다. 보호를 위해 멀티미디어 디지털 오브젝트를 표시하거나 식별을 위해 오브젝트의 지문을 채취하여 지식 등록 정보의 권한이 없는 사용을 저지시킬 수 있습니다.

귀중한 오브젝트에 표시를 적용하여 다음을 수행할 수 있습니다.

- 권한이 없는 복사나 재사용을 저지하기 위해 소스를 식별합니다. 이것을 수위표라고 하며 보통 볼 수 있습니다.
- 권한이 없는 복사나 재사용을 저지하기 위해 콘텐츠를 받는 사람을 식별합니다. 이것을 지문이라고 하며 보통 볼 수 없습니다.
- 추가 정보를 얻을 수 있는 문의처를 제공합니다.
- 분배의 부가 가치 체인에서 사용할 시간과 날짜와 같은 정보를 제공합니다.

오브젝트가 고객에게 전달되기 전에 디지털 오브젝트를 표시할 수 있습니다. 전달하기 전에 수위표와 지문 둘 다 적용될 수 있습니다. 그러나 전달 전에 지문을 적용하면 받는 사람이 알려지며 전달 처리에서 동적으로 적용되도록 표시를 요구할 수 있습니다. 전달 전에 사용자 자신의 제어 환경에서 표시를 적용하면 변경의 위험이 덜하므로 보안을 증진시킵니다.

관리 및 전달 처리의 여러 단계에서 표시를 적용할 수 있습니다. 사용자 상황은 적절한 내용에 영향을 미칩니다. 사용자 처리의 다음 위치에서 표시가 적용될 수 있습니다.

- 오브젝트가 저장되기 전

공용 표시가 오브젝트(예: 소유자를 식별하기 위한 비주얼 수위표)에 대해 사용되는 경우, 오브젝트를 저장하기 전이나 저장하는 동안 표시를 적용할 수 있습니다. 원래 의, 표시되지 않은 오브젝트와 표시된 오브젝트를 둘 다 Content Manager 시스템에 저장할 수 있습니다. 또는 표시된 오브젝트만을 저장하고 표시되지 않은 오브젝트를 별도의 저장소에 보관할 수 있습니다.

- 오브젝트가 저장된 후

Content Manager 시스템에 저장된 오브젝트를 표시하려는 경우, 오브젝트를 검색하고, 표시하고, 표시되지 않은 오브젝트를 표시된 버전으로 교체하거나 표시된 오브젝트를 새 항목으로 저장할 수 있습니다.

- 오브젝트가 검색되었을 때

적용되는 표시가 받는 사람에 근거하여 변화하는 경우, 오브젝트를 검색한 후에 동적으로 표시를 적용할 수 있습니다. 표시된 오브젝트는 원래의 오브젝트 대신 전달될 수 있습니다.

표시되지 않은 시스템에 여러 legacy 오브젝트를 가지며 시간이 걸리게 하거나 자원을 사용하여 다시 돌아가 각 자원을 표시하게 하지 않으려는 경우, 검색되었을 때 오브젝트를 표시할 수 있습니다.

---

## 표시 기법 사용

선택할 수 있는 다양한 콘텐츠 표시 기법이 있습니다. 각 기법은 특정 문제점을 지정하며 제거와 수정에 대한 저항에서 다릅니다.

표시는 다음에 의해 특징지웁니다.

- 전달된 정보

**수위표** 콘텐츠의 소스를 식별합니다. 오브젝트의 소유자와 버전과 같은 정보를 포함할 수 있습니다.

**지문** 콘텐츠의 받는 사람을 식별합니다. 오브젝트가 전달된 장소와 사람과 같은 정보를 포함할 수 있습니다.

- 가시성

**비주얼** 표시가 보이며 주의시킬 수 있습니다.

**감추기** 표시가 이미지 내에서 감추어집니다.

- 무결성

**사라지기 쉬움**

수정으로 표시가 사라집니다.

**강함** 크기 재조정, 압축, 회전 및 자르기와 같이 표시된 오브젝트에 대한 수정에 견딥니다.

- 응용프로그램의 시간

- 오브젝트가 캡처되었을 때

- 오브젝트가 저장되었을 때

- 고객에게 분배하기 위해 오브젝트가 검색되었을 때

- 받는 사람의 워크스테이션에서 오브젝트를 받았을 때

- 위치

- 비주얼 표시가 불법 재사용을 저지하기 위한 것인 경우, 이미지의 더 큰 부분에 적용될 수 있습니다. 이미지의 좀 더 텍스처된 영역을 제거하기 더 어렵게 하기 위해 겹치도록 표시를 이동할 수 있습니다.
- 감추기 표시가 사용된 경우, 이미지의 텍스처된 영역은 이미지의 최소 영향으로 데이터가 병합되게 합니다.
- 비주얼 표시가 소유권을 표시하기 위한 것인 경우, 눈에 거슬리지 않게 이미지의 구석에 위치될 수 있습니다.
- 비주얼 및 감추기 표시가 둘 다 사용되는 경우, 비주얼 표시가 우선 적용되어야 합니다.

#### • 형식

**2진** 표시는 이미지를 통해 반복되는 임의 비트 순서일 수 있습니다. 이 임의 순서는 이미지를 표시하거나 표시해제하기 위해 사용될 수 있는 키입니다.

표시는 이미지일 수도 있습니다.

#### 구조화 데이터

표시는 병합된 텍스처 데이터일 수 있습니다.

## 비주얼 표시

비주얼 표시는 마스크와 이미지 둘다가 보이는 방식으로 이미지 위에 지정되는 투명성 마스크입니다. 제거하기 어려운 비주얼 표시는 오브젝트의 남용을 효과적으로 저지시킵니다.

다음 상황에서 비주얼 표시를 사용하십시오.

- 고객에게 검토할 이미지를 제공하지만 고객이 검토 사본을 재사용하지 못하게 하려고 할 때
- WWW(World Wide Web)의 광고에 이미지를 사용하려고 할 때

## 감추기 표시

감추기 표시는 이미지가 변경되지 않은 채로 표시되는 방식으로 이미지 내에 감춘 데이터입니다. 응용프로그램은 표시를 적용, 발견 및 해독하는 데 필요합니다.

다음 상황에서 감추기 표시를 사용하십시오.

- 표시된 오브젝트의 불법 복사를 저지하고 소유권을 식별하기 위해 정보를 병합하려고 할 때(수위표 지정)
- 분배 경로를 추적하기 위해 정보를 병합하려고 할 때(지문 지정)
- 이미지에서 캡션이나 주석을 병합하려고 할 때



---

## 제 11 장 내게 필요한 옵션

이 제품은 사용이 불가능한 사람들에게 액세스할 수 있게 하는 몇 가지 기능을 포함합니다. 이 기능은 다음을 포함합니다.

- 마우스 대신에 키보드를 사용하여 모든 기능을 작동할 수 있는 능력
- 향상된 표시장치 등록 정보 지원
- 보조 기술과의 호환성
- 운영 체제 내게 필요한 옵션 기능과의 호환성
- 내게 필요한 옵션 문서 형식

---

### 키보드 입력 및 탐색

다음 기능은 키보드 입력과 탐색에 사용 가능합니다.

#### 키보드 입력

제품을 조작하려면 마우스 대신 키보드를 사용할 수 있습니다.

메뉴 항목 및 제어는 제어를 활성화하거나 키보드에서 직접 메뉴 항목을 선택하도록 허용하는 액세스 키를 제공합니다. 이 키는 자체 문서화하며, 액세스 키는 표시된 메뉴나 제어에서 밑줄그어져 있습니다.

#### 키보드 포커스

Windows 기반 시스템에서 키보드 포커스의 위치는 강조표시되며, 어떤 창의 영역이 활성화되었는지 어디에서 키스트로크가 영향을 미치는지 나타냅니다.

#### 응답 시간 조정

Windows 기반 시스템에서, 제어판을 통해 응답 시간을 조정할 수 있습니다.

---

### 액세스 가능한 표시장치의 기능

클라이언트는 사용자 인터페이스를 확장시키며 시력이 낮은 사용자의 내게 필요한 옵션을 향상시키는 몇 가지 기능을 갖습니다. 이러한 향상된 기능은 높은 명암 설정과 사용자 조정 가능한 글꼴 등록 정보에 대한 지원을 포함합니다.

#### 높은 명암 모드

클라이언트는 운영 체제에서 제공한 높은 명암 모드 옵션을 지원합니다. 이 기능은 배경 및 전경색 간의 더 높은 명암을 지원합니다.

### 글꼴 설정

Windows 기반 시스템에서, 메뉴와 대화 상자 창에서 텍스트에 대한 색, 크기 및 글꼴을 결정하는 표시 설정을 지정할 수 있습니다. 클라이언트는 문서 목록에 대한 글꼴을 선택하게 합니다.

### 색에 대한 비 종속성

이 제품의 기능을 사용하려면 색 간을 구별할 필요가 없습니다.

---

## 보조 기술과의 호환성

클라이언트는 Narrator 및 Via Voice와 같은 화면 판독기 응용프로그램과 호환 가능합니다. 클라이언트는 시각적으로 손상된 사용자가 화면상의 정보를 사용할 수 있도록 하는 내게 필요한 옵션 응용프로그램에 필요한 등록 정보를 갖습니다.

---

## 내게 필요한 옵션 문서

이 제품의 문서는 PDF 형식으로 사용 가능합니다. [access.adobe.com](http://access.adobe.com)의 Adobe에서 사용 가능한 여유 도구를 사용하여 PDF 파일을 HTML이나 텍스트로 변환할 수 있습니다. 이것은 사용자 브라우저에 설정된 표시 환경 설정에 따라 사용자가 문서를 볼 수 있게 합니다. 또한 화면 판독기나 다른 보조 기술의 사용을 허용합니다.

---

## 주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다.

IBM은 다른 국가에서는 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 재산을 침해하지 않는 한, 기능상 동등한 제품, 프로그램 또는 서비스를 대신 사용할 수 있습니다. 그러나 비IBM 제품, 프로그램 또는 서비스의 운영에 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 국가에서는 적용되지 않습니다. IBM은 타인의 권리 침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여 (단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증없이 이 책을 『현상 그대로』 제공합니다. 일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책 사항을 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.

(1) 독립적으로 작성된 프로그램 및 기타 프로그램(본 프로그램 포함)간의 정보 교환 및  
(2) 교환된 정보의 상호 이용을 목적으로 정보를 원하는 프로그램 라이선스 사용자는 다음 주소로 문의하십시오.

135-270

서울특별시 강남구 도곡동 467-12, 군인공제회관빌딩

한국 아이.비.엠 주식회사

고객만족센터

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

이 정보에 기술된 라이선스가 있는 프로그램 및 이 프로그램에 대해 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 성능은 개발 레벨 상태의 시스템에서 측정되었을 수 있으므로 이러한 측정치가 일반적으로 사용되고 있는 시스템에서도 동일하게 나타날 것이라고 보증할 수 없습니다. 또한, 일부 성능은 추정치일 수도 있습니다. 실제 결과는 다를 수 있습니다. 이 문서의 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 배상 청구에 대해서는 확신할 수 없습니다. 비IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

IBM의 향후 방향 또는 의도에 관한 모든 언급은 별도의 통지없이 변경될 수 있습니다.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이 예제에는 개념을 완벽하게 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 라이선스:

이 정보에는 여러 가지 운영 플랫폼에서의 프로그래밍 기법을 보여주는 원시 언어로 된 샘플 응용프로그램이 포함되어 있습니다. 귀하는 샘플 프로그램의 작성 기준이 된 운영 플랫폼의 응용프로그램 프로그래밍 인터페이스에 부합하는 응용프로그램을 개발, 사용, 마케팅 및 배포하기 위한 목적으로 이러한 샘플 프로그램을 IBM에 추가 비용없이 어떤 형태로든 복사, 수정 및 배포할 수 있습니다. 이러한 예제는 모든 조건하에서 철저히 테스트된 것은 아닙니다. 따라서 IBM은 이러한 프로그램의 신뢰성, 서비스 가능성 또는 기능을 보증하거나 암시할 수 없습니다. 귀하는 IBM의 응용프로그램 프로그래밍 인터페이스에 부합하는 응용프로그램을 개발, 사용, 마케팅 및 배포하기 위한 목적으로 이러한 샘플 프로그램을 IBM에 추가 비용없이 어떤 형태로든 복사, 수정 및 배포할 수 있습니다.

---

## 상표

다음 용어는 미국 또는 기타 국가에서 사용되는 IBM Corporation의 상표입니다.

IBM	DisplayWrite	PowerPC
400	e-business	PTX
Advanced Peer-to-Peer Networking	HotMedia	QBIC
AIX	Hummingbird	RS/6000
AIXwindows	ImagePlus	SecureWay
APPN	IMS	SP
AS/400	Micro Channel	VideoCharger
C Set ++	MQSeries	Visual Warehouse
CICS	MVS/ESA	VisualAge
DATABASE 2	NetView	VisualInfo
DataJoiner	OS/2	WebSphere
DB2	OS/390	
DB2 Universal Database	PAL	

Approach, Domino, Lotus, Lotus 1-2-3, Lotus Notes 및 SmartSuite는 미국 또는 기타 국가에서 사용되는 Lotus Development Corporation의 상표 또는 등록상표입니다.

Intel 및 Pentium은 미국 또는 기타 국가에서 사용되는 Intel Corporation의 상표 또는 등록상표입니다.

Microsoft, Windows 및 Windows NT는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 등록상표입니다.

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

UNIX는 미국 또는 기타 국가에서 사용되는 Open Group의 등록상표입니다.

기타 회사, 제품 및 서비스 이름은 해당 회사의 상표 및 서비스표입니다.

## 용어집

이 용어집은 이 시스템에 고유한 용어 및 약어를 정의합니다. 기울임꼴로 표시된 용어는 이 용어집의 다른 곳에 정의되어 있습니다.

## 가

**검색 기준(search criteria).** Enterprise Information Portal에서 관리자가 사용자에게 사용 가능한 선택사항을 제한하거나 자세히 정의하는 검색 템플릿에 대해 정의하는 특정 필드.

**검색 템플릿(search template).** 관리자가 설계한 검색 기준으로 구성된 특정 유형의 연합 검색을 위한 양식. 시스템 관리자는 각 검색 템플릿에 액세스할 수 있는 사용자 및 사용자 그룹도 식별합니다.

**게이트웨이(gateway).** 다른 네트워크 아키텍처를 가진 두 개의 컴퓨터 네트워크를 상호 연결하는 기능적 단위. 게이트웨이는 다른 아키텍처를 가진 네트워크 또는 시스템을 연결합니다. 브릿지는 동일하거나 유사한 아키텍처를 가진 네트워크 또는 시스템을 상호 연결합니다.

**결합 검색(combined search).** 매개변수식, 텍스트 또는 이미지와 같은 하나 이상의 검색 유형이 결합된 조회.

**교환(interchange).** CIF(Common Interchange File)와 CIU(Common Interchange Unit)를 사용하여 하나의 OS/390용 Content Manager ImagePlus 시스템에서 다른 ImagePlus 시스템으로 색인과 함께 이미지를 가져오거나 내보내는 기능.

**구성요소(component).** 루트 구성요소 또는 하위 구성요소의 일반 용어.

**근거리 통신망(LAN: Local Area Network).** 장치 세트가 통신을 위해 다른 장치 세트에 연결되고 보다 큰 네트워크에 연결할 수 있는 네트워크.

**기가바이트(gigabyte: GB).** (1) 프로세서 기억영역, 실제 및 가상 기억영역, 채널 볼륨의 경우,  $2^{30}$  또는 1 073 741 824바이트입니다. (2) 디스크 기억영역 용량 및 통신 볼륨의 경우, 1 000 000 000바이트입니다.

**기능(feature).** 이미지 검색 서버에 저장된 비주얼 콘텐츠 정보. 또한 이미지 검색 응용프로그램이 일치 여부를 결정할 때 사용

하는 비주얼 특성. QBIC 기능에는 평균 색, 히스토그램 색, 위치 색 및 텍스처의 네 가지가 있습니다.

**기본 행수(cardinality).** 데이터베이스 테이블의 행 수.

## 나

**네트워크 테이블 파일(network table file).** Content Manager 시스템의 각 노드에 대한 시스템 고유의 구성 정보가 들어 있는 텍스트 파일. 시스템의 각 노드에는 노드를 식별하고 연결에 필요한 노드를 나열하는 네트워크 테이블 파일이 있어야 합니다.

네트워크 테이블의 이름은 FRNOLINT.TBL입니다.

## 다

**데이터 형식(data format).** MIME 유형을 참조하십시오.

**데이터스토어(datastore).** (1) 데이터가 저장되는 위치(예: 데이터베이스 시스템, 파일 또는 디렉토리)에 대한 일반 용어. (2) 응용프로그램에서 콘텐츠 서버의 가상 표현.

**등급(rank).** 조회 결과에 대한 주어진 부분의 관련도를 나타내는 정수값. 등급이 높을수록 일치도가 높습니다.

**등록 정보(property).** 오브젝트에 대해 설명하는 오브젝트의 특성. 등록 정보를 변경하거나 수정할 수 있습니다. 등록 정보의 예로는 유형 양식이 있습니다.

## 라

**라이브러리 서버(library server).** 항목에 대한 조회를 저장, 관리 및 처리하는 Content Manager 시스템의 구성요소.

**라이브러리 클라이언트(library client).** 라이브러리 시스템에 하위 레벨의 프로그래밍 인터페이스를 제공하는 Content Manager 시스템의 구성요소. 라이브러리 클라이언트에는 소프트웨어 개발자 킷의 일부인 API가 포함됩니다.

**렌더(render).** 일반적으로 이미지 지향이 아닌 데이터를 가져와 이미지로 표현하거나 표시하는 작업. Content Manager에서 워드프로세싱 문서는 표시 목적으로 이미지로 렌더될 수 있습니다.

**루트 구성요소(root component).** 시스템에서 정의하거나 사용자가 정의한 속성으로 구성된 계층 구조 항목 유형의 첫 번째 레벨 또는 유일한 레벨.

**링크(link).** 두 항목(원본 및 대상)간의 방향 관계. 링크 세트를 사용하여 일 대 다 연관을 모델링할 수 있습니다. 참조와 다릅니다.

## 마

**매개변수식 검색(parametric search).** 오브젝트의 등록 정보를 기본으로 한 오브젝트에 대한 조회.

**멀티미디어 파일 시스템(multimedia file system).** 비디오 및 오디오의 저장과 전달에 알맞게 최적화된 파일 시스템.

**멀티미디어(multimedia).** 컴퓨터에서 표시 및 제어를 위해 여러 가지 미디어 요소(텍스트, 그래픽, 오디오, 스틸 이미지, 비디오, 애니메이션)를 결합한 것.

**메소드(method).** Java 설계 또는 프로그래밍에서 조작에 지정된 작동을 구현하는 소프트웨어. C++의 구성원 함수와 동일한 의미입니다.

**문서(document).** Content Manager 시스템과 사용자 사이에서 별도의 단위로 저장, 검색 및 교환할 수 있는 항목. 문서 의미 유형이 있는 항목은 문서를 이루는 정보를 포함하도록 예상되나, Content Manager 문서 모델의 구현을 의미하지는 않습니다.

항목 유형이 분류된 문서에서 작성된 항목(Content Manager 문서 모델의 특정 구현)은 문서 부분을 포함해야 합니다. 분류된 항목 유형 문서를 사용하여 문서 또는 폴더 의미 유형을 갖는 항목을 작성할 수 있습니다.

문서 부분은 텍스트, 이미지 및 스프레드시트를 비롯하여 여러 가지 유형의 콘텐츠를 포함할 수 있습니다.

**미디어 서버(media server).** 비디오 파일의 저장 및 액세스에 사용되는 Content Manager 시스템의 AIX 기본 구성요소.

## 바

**반복자(iterator).** 한 번에 오브젝트 컬렉션 전체를 처리하는 데 사용하는 클래스 또는 구조.

**볼륨(volume).** 시스템의 오브젝트가 저장된 실제 물리적 기억 영역 장치 또는 장치를 나타냅니다.

**부분(part).** 오브젝트를 참조하십시오.

## 사

**사용 권한세트(privilege set).** 시스템 구성요소 및 기능에 대한 작업에 필요한 사용 권한 컬렉션. 시스템 관리자는 사용자(사용자 ID) 및 사용자 그룹에 사용 권한 세트를 할당합니다.

**사용 권한(privilege).** 특정 방법으로 특정 오브젝트에 액세스하는 권한. 사용 권한에는 시스템에 저장된 오브젝트의 작성, 삭제 및 선택과 같은 권한이 포함됩니다. 사용 권한은 시스템 관리자에 의해 할당됩니다.

**사용 기록(history log).** 워크플로우의 활동 레코드를 보관하는 파일.

**사용자 그룹(user group).** 한 명 이상의 정의된 개별 사용자로 구성된 그룹으로서, 단일 그룹 이름으로 식별됩니다.

**사용자 맵핑(user mapping).** Enterprise Information Portal 사용자 ID 및 암호를 하나 이상의 콘텐츠 서버에 있는 해당 사용자 ID 및 암호에 연관시키는 것. 사용자 맵핑은 Enterprise Information Portal 및 다중 콘텐츠 서버에 대한 단일 로그 온을 가능하게 합니다.

**사용자 종료 루틴(user exit routine).** 미리 정의된 사용자 종료에서 제어를 수신하는 사용자가 작성한 루틴.

**사용자 종료(user exit).** 사용자 종료 루틴에 제어를 제공할 수 있는 IBM 제공 프로그램의 지점.

**사용자(user).** Enterprise Information Portal에서는 Enterprise Information Portal 관리 프로그램에서 식별되는 모든 사람을 지칭합니다.

**색인 클래스 보기(index class view).** 이전 Content Manager에서 색인 클래스 서브세트에 대해 API에서 사용되는 용어.

**색인 클래스 서브세트(index class subset).** 이전의 Content Manager에서 응용프로그램이 폴더 및 오브젝트를 저장, 검색 및 표시하는 데 사용하는 색인 클래스의 보기.

**색인 클래스(index class).** 항목 유형을 참조하십시오.

**색인(index).** 나중에 검색할 수 있도록 특정 항목 또는 오브젝트를 식별하는 속성값을 추가하거나 편집하는 것.

**생성자(constructor).** 프로그래밍 언어에서 클래스와 동일한 이름을 가지며 해당 클래스에 대한 오브젝트를 작성 및 초기화하는 데 사용되는 메소드.

**서버 목록(server inventory).** 지정된 콘텐츠 서버에 있는 원래의 엔티티 및 원래의 속성의 포괄적인 목록.

**서버 유형 정의(server type definition).** Enterprise Information Portal에 대해 특정 유형의 사용자 조정 서버를 고유하게 식별하는 데 필요한 특성 목록으로 시스템 관리자에 의해 식별됩니다.

**서버 정의(server definition).** Enterprise Information Portal에 대해 서버를 고유하게 식별하는 콘텐츠 서버의 특성.

**서브클래스(subclass).** 다른 클래스에서 파생된 클래스 클래스와 서브클래스 사이에 하나 이상의 클래스가 있을 수 있습니다.

**속성(attribute).** 항목의 특정 특성 또는 등록 정보(예: 이름, 주소, 연령 등)를 기술하는 데이터 단위로, 해당 항목을 찾는 데도 사용될 수 있습니다. 속성은 해당 속성별로 저장되는 정보 범위를 나타내는 유형과, 해당 범위 내에 있는 값을 가집니다. 예를 들어, 제목, 실행 시간 또는 인코딩 유형(MPEG1, H.263 등)과 같은 멀티미디어 파일 시스템 내의 파일에 대한 정보가 있습니다. Enterprise Information Portal의 경우, 연합 속성 및 원래의 속성도 참조하십시오.

**수퍼 클래스(superclass).** 클래스가 파생된 클래스 클래스와 수퍼 클래스 사이에 하나 이상의 클래스가 있을 수 있습니다.

**스테이징 영역(staging area).** 자원 관리자의 작업 기억영역. 자원 관리자 캐시라고도 합니다.

**스테이징(staging).** 일반적으로 시스템 요구 또는 사용자 요청이 있을 때 저장된 오브젝트를 오프라인 또는 우선순위가 낮은 장치에서 다시 온라인 또는 우선순위가 높은 장치로 이동하는 프로세스. 사용자가 영구 기억영역에 저장된 오브젝트를 요청하면, 작업 사본이 스테이징 영역에 쓰여집니다.

**스트림 데이터(streamed data).** 지정된 전송률로 네트워크 연결을 통해 전송되는 모든 데이터. 스트림은 하나의 데이터 유형 또는 조합된 유형일 수 있습니다. 데이터 비율(초당 비트 수로 표시됨)은 스트림 및 네트워크의 여러 유형에 따라 다양합니다.

## 아

**액세스 제어 목록(access control list).** 하나 이상의 사용자 ID 또는 사용자 그룹 및 연관된 사용 권한으로 구성된 목록.

Enterprise Information Portal 시스템에서는 액세스 제어 목록을 사용하여 검색 템플릿에 대한 사용자 액세스를 제어합니다.

**액세스 제어(access control).** 특정 함수 및 저장된 오브젝트에 권한이 부여된 사용자만이 권한이 부여된 방법으로 액세스할 수 있게 하는 프로세스.

**연합 검색(federated search).** 이기종일 수 있는 하나 이상의 콘텐츠 서버에서 동시에 데이터를 검색하는 Enterprise Information Portal에서 발행된 조회.

**연합 데이터스토어(federated datastore).** Content Manager와 같은 여러 개의 특정 콘텐츠 서버를 가상적으로 표현한 것.

**연합 속성(federated attribute).** 하나 이상의 콘텐츠 서버에서 원래의 속성에 매핑된 Enterprise Information Portal 메타데이터 카테고리. 예를 들어, 연합 속성인 보험 증권 번호는 Content Manager에서는 보험 증권 번호 속성에 매핑될 수 있고 OS/390용 Content Manager ImagePlus에서는 보험 증권 ID 속성에 매핑될 수 있습니다.

**연합 엔티티(federated entity).** 연합 속성으로 구성되었으며 선택적으로 하나 이상의 연합 텍스트 색인에 연관된 Enterprise Information Portal 메타데이터 오브젝트.

**연합 컬렉션(federated collection).** 연합 검색에서 발생한 오브젝트 그룹.

**연합 텍스트 색인(federated text index).** 하나 이상의 콘텐츠 서버에서 하나 이상의 원래 텍스트 색인에 매핑된 Enterprise Information Portal 메타데이터 오브젝트.

**오버레이(overlay).** 선, 음영, 텍스트, 상자 또는 로고와 같이 인쇄 중에 페이지에서 변수 데이터에 병합될 수 있는 미리 정의된 컬렉션.

**오브젝트 서버 캐시(object server cache).** 자원 관리자 캐시를 참조하십시오.

**오브젝트 서버(object server).** 자원 관리자를 참조하십시오.

**오브젝트(object).** 사용자가 하나의 단위로 저장, 검색 및 조작할 수 있는 모든 디지털 콘텐츠로서 JPEG 이미지, MP3 오디오, AVI 비디오 및 책의 텍스트 블록을 예로 들 수 있습니다.

**와일드 카드 문자(wildcard character).** 하나 이상의 문자를 나타내는 데 사용될 수 있는 별표(\*) 또는 물음표(?)와 같은 특수 문자. 모든 문자 또는 문자 세트를 와일드 카드 문자로 대체할 수 있습니다.

**워크플로우 상태(Workflow State).** 전체 워크플로우 상태.

**워크플로우(workflow).** Enterprise Information Portal에서 작업 패키지, 문서 또는 폴더가 처리되는 동안 거치는 일련의 작업 단계 및 이러한 단계를 제어하는 규칙.

예를 들어, 지불 요구 승인은 개별 보험 지불 요구가 승인을 받기 위해 따라야 하는 프로세스에 대해 설명합니다.

**원래의 색인 텍스트(native text index).** 특정 콘텐츠 서버에서 관리되는 텍스트 항목의 색인. 예를 들어, Content Manager 콘텐츠 서버의 단일 텍스트 검색 색인이 있습니다.

**원래의 속성(native attribute).** 특정 콘텐츠 서버에서 관리되고 해당 콘텐츠 서버에 고유한 오브젝트의 특성. 예를 들어, 키 필드인 보험 증권 번호는 Content Manager 콘텐츠 서버의 원래의 속성인 반면, 보험 증권 ID 필드는 Content Manager OnDemand 콘텐츠 서버의 원래의 속성일 수 있습니다.

**원래의 엔티티(native entity).** 특정 콘텐츠 서버에서 관리되고 원래의 속성으로 구성된 오브젝트. 예를 들어, Content Manager 색인 클래스는 Content Manager 키 필드로 구성된 원래의 엔티티입니다.

**의미 유형(semantic type).** 항목에 대한 사용법 또는 규칙. 기본, 주식, 메모는 Content Manager에서 제공하는 의미 유형입니다. 사용자도 자신의 고유한 의미 유형을 정의할 수 있습니다.

**일시중단(suspend).** 워크플로우에서 오브젝트를 제거하고 이를 활성화하는 데 필요한 일시중단 기준을 정의하는 것. 나중에 오브젝트를 활성화하면 처리를 계속할 수 있습니다.

## 자

**자원 관리자.** 오브젝트를 관리하는 Content Manager 시스템의 구성요소. 이러한 오브젝트는 라이브러리 서버에 저장된 항목에 의해 참조됩니다.

**자원 관리자 캐시(resource manager cache).** 자원 관리자의 작업 기억영역. 스테이징 영역이라고도 합니다.

**작업 단계(work step).** 개별 작업 항목, 문서 또는 폴더가 통과해야 하는 워크플로우 또는 문서 경로지정 프로세스의 분리된 지점.

**작업 목록(worklist).** 사용자에게 할당된 작업 항목, 문서 또는 폴더의 컬렉션.

**작업 상태(work state).** 개별 작업 항목, 문서 또는 폴더 상태.

**작업 패킷(work packet).** Enterprise Information Portal 버전 7.1의 한 위치에서 다른 위치로 경로지정된 문서의 컬렉션. 사용자는 작업 목록을 통해 작업 패킷에 액세스하여 작업합니다.

**작업 항목(work item).** 이전의 Content Manager 워크플로우 및 Enterprise Information Portal 고급 워크플로우에서, 워크플로우 내에서 활성화된 모든 작업 활동.

**조치 목록(action list).** 사용자가 워크플로우 또는 문서 경로지정 프로세스에서 수행할 수 있는 승인된 조치 목록으로, 시스템 관리자 또는 다른 워크플로우 조정자에 의해 정의됩니다.

**조회 문자열(query string).** 조회의 등록 정보 및 등록 정보 값을 지정하는 문자열. 응용프로그램에 조회 문자열을 작성하여 조회로 전달할 수 있습니다.

## 차

**참조(reference).** 루트 또는 하위 구성요소와 다른 루트 구성요소 간의 단방향성을 가진 일 대 일 연관. 링크와 다릅니다.

**추상 클래스(abstract class).** 개념을 나타내는 객체 지향 프로그래밍 클래스. 여기에서 파생된 클래스는 개념의 구현을 나타냅니다. 추상 클래스 오브젝트는 구성할 수 없습니다. 즉, 인스턴스로 작성될 수 없습니다.

## 카

**캐시(cache).** 기본 기억영역보다 작고 빠른 특수한 목적의 버퍼로, 자주 액세스될 수 있는 데이터의 사본을 보유하는 데 사용됩니다. 캐시를 사용하면 액세스 시간은 단축되나 메모리 요구사항이 증가될 수도 있습니다.

**커넥터 클래스(connector class).** 특정 콘텐츠 서버에 고유한 API에 대한 표준 액세스를 제공하는 객체 지향 프로그래밍 클래스.

**커서(cursor).** 응용프로그램에서 일부 순서화된 행 세트 내의 특정 행을 가리키는 데 사용하는 명명된 제어 구조. 커서는 세트에서 행을 검색하는 데 사용됩니다.

**콘텐츠 서버(content server).** 멀티미디어 및 비즈니스 데이터와 사용자가 이 데이터에 대해 작업하는 데 필요한 관련 메타데이터를 저장하는 소프트웨어 시스템. 콘텐츠 서버의 예로는 OS/390용 Content Manager ImagePlus 및 Content Manager가 있습니다.

**컬렉션(collection).** 관리 규칙 세트가 유사한 오브젝트의 그룹.

**클라이언트 응용프로그램(client application).** Enterprise Information Portal에서 콘텐츠 서버에 액세스하기 위해 객체 지향 또는 인터넷 API를 사용하여 작성된 응용프로그램.

**클라이언트/서버(client/server).** 통신에서 한 사이트에 있는 프로그램이 다른 사이트에 있는 프로그램으로 요청을 보낸 다음, 응답을 기다리는 분산 데이터 처리에서의 상호작용 모델. 요청 프로그램을 클라이언트라고 하고 응답 프로그램을 서버라고 합니다.

**클래스(class).** 객체 지향 설계 또는 프로그래밍에서 공통 정의, 공통 등록 정보, 조작 및 작동을 사용하는 오브젝트를 작성하기 위해 인스턴스로 작성될 수 있는 모델 또는 템플릿. 오브젝트는 클래스의 인스턴스입니다.

**키 필드(key field).** 속성을 참조하십시오.

## 과

**파일 시스템(file system).** AIX에서 기억영역에 대해 하드 드라이브를 파티션하는 메소드.

**패키지(package).** 액세스 보호 및 이름 공간 관리를 제공하는 관련 클래스 및 인터페이스의 콜렉션.

**폴더 관리자(folder manager).** 데이터를 온라인 문서 및 폴더로 관리하기 위한 Content Manager 모델. 응용프로그램과 Content Manager 컨텐츠 서버 사이의 기본 인터페이스로 폴더 관리자 API를 사용할 수 있습니다.

**폴더(folder).** 분류에 관계없이 폴더 의미 유형이 있는 항목 유형의 항목. 폴더 의미 유형이 있는 항목은 Content Manager에서 제공하는 특정 폴더 기능을 포함합니다. 또한 문서 또는 자원 항목과 같은 항목 유형 분류에서 사용 가능한 추가 기능 및 모든 비자원 항목 성능을 포함합니다. 폴더는 문서 및 서브폴더를 포함하는 많은 항목 유형을 포함할 수 있습니다. 폴더는 속성에 의해 색인화될 수 있습니다.

## 하

**하위 구성요소(child component).** 계층 구조 항목 유형의 두 번째 이하 레벨로서, 이는 선택적입니다. 각 하위 구성요소는 직접 상위 레벨에 연관됩니다.

**항목 유형 분류(item type classification).** 해당 항목 유형의 항목을 보다 자세히 식별하는 항목 유형 내의 구분. 동일한 항목 유형을 가진 모든 항목은 동일한 항목 유형 분류를 갖습니다. Content Manager는 폴더, 문서, 오브젝트, 비디오, 이미지 및 텍스트와 같은 항목 유형 분류를 제공하며, 사용자가 고유한 항목 유형 분류를 정의할 수도 있습니다.

**항목 유형(item type).** 항목과 같이 정의한 후 나중에 찾을 수 있는 템플릿로서, 루트 구성요소, 0개 이상의 하위 구성요소 및 분류로 구성됩니다.

**항목(item).** Enterprise Information Portal에서 관리하는 최소 정보 단위에 대한 일반 용어. 각 항목에는 ID가 있습니다. 예를 들어, 항목은 폴더 또는 문서일 수 있습니다.

**해제(release).** 항목에서 일시중단 기준을 제거하는 것. 일시중단 항목은 기준에 해당되거나 적절한 권한을 가진 사용자가 기준을 대체하고 수동으로 이를 해제할 경우에 해제됩니다.

**핸들(handle).** 오브젝트를 나타내고 오브젝트를 검색하는 데 사용되는 문자열.

## A

**ADSM.** Tivoli® Storage Manager를 참조하십시오.

**API.** API(Application Programming Interface)를 참조하십시오.

**API(Application Programming Interface).** 응용프로그램이 다른 응용프로그램과 통신할 수 있도록 하는 소프트웨어 인터페이스. API는 기본 라이선스 프로그램에서 제공하는 특정 기능 및 서비스를 확보하기 위해 응용프로그램에서 코드화될 수 있는 프로그래밍 언어 구성 또는 명령문 세트입니다.

**AVI.** AVI(Audio/Video Interleaved)를 참조하십시오.

**AVI(오디오/비디오 인터리브: Audio/Video Interleaved).** 파일에 오디오 및 비디오 데이터를 삽입할 수 있도록 하는 RIFF(Resource Interchange File Format) 파일 스펙. 파일 장치에서의 순차 액세스를 유지보수하는 동안 재생 또는 레코딩을 할 때 대체 청크에서 개별 트랙에 액세스할 수 있습니다.

## B

**BLOB.** BLOB(Binary Large Object)를 참조하십시오.

**BLOB(Binary Large Object).** 크기의 범위가 0에서 2기가바이트인 일련의 바이트. 이 문자열에는 연관된 코드 페이지 및 문자 세트가 없습니다. 이미지, 오디오 및 비디오 오브젝트는 BLOB로 저장됩니다.

## C

**CGI.** CGI(Common Gateway Interface)를 참조하십시오.

**CGI 스크립트(CGI script).** 웹 서버에서 실행되며, CGI(Common Gateway Interface)를 사용하여 일반적으로 웹 서버에서 수행되지 않는 작업(예: 데이터베이스 액세스 및 양식 처리)을 수행하는 컴퓨터 프로그램. CGI 스크립트는 Perl과 같은 스크립트 언어로 작성된 CGI 프로그램입니다.

**CGI(Common Gateway Interface).** 웹 서버와, 웹 서버 외부 프로그램 간의 정보 교환을 위한 표준. 외부 프로그램은 웹 서버가 실행되는 운영 체제에서 지원하는 프로그램 언어로 작성될 수 있습니다. CGI 스크립트를 참조하십시오.

**CIF.** CIF(Common Interchange File)를 참조하십시오.

**CIF(Common Interchange File).** 하나의 IPIA(ImagePlus Interchange Architecture) 데이터 스트림이 들어 있는 파일.

**CIU.** CIU(Common Interchange Unit)를 참조하십시오.

**CIU(Common Interchange Unit).** CIF(Common Interchange File)의 독립적인 전송 단위. 이는 수신 데이터베이스와의 관계를 식별하는 CIF의 일부입니다. CIF는 여러 개의 CIU를 포함할 수 있습니다.

## D

**DDO.** DDO(Dynamic Data Object)를 참조하십시오.

**DDO(Dynamic Data Object).** 응용프로그램에서 해당 오브젝트를 기억영역 내부 또는 외부로 이동하는 데 사용되는 저장 오브젝트의 일반적인 표현 방법.

**DTD.** DTD(Document Type Definition)를 참조하십시오.

**DTD(Document Type Definition).** XML 문서의 특정 클래스에 대한 구조를 지정하는 규칙. DTD는 요소, 속성 또는 표기법을 사용하여 구조를 정의하며 문서의 특정 클래스에서 각 요소, 속성 또는 표기법을 사용할 수 있는 방법에 대한 제한조건을 설정합니다. DTD는 특정 마크업 언어의 구조에 대해 완전히 설명할 수 있다는 점에서 데이터베이스 스키마와 유사합니다.

## G

**GB.** 기가바이트를 참조하십시오.

## H

**HTML.** HTML(Hypertext Markup Language)을 참조하십시오.

**HTML(Hypertext Markup Language).** SGML 표준을 따르고 기본적으로 하이퍼텍스트 링크를 포함하는 온라인 텍스트 및 그래픽 정보를 지원하도록 설계된 마크업 언어.

## I

**Information Mining.** 텍스트에서 핵심 정보를 추출하는 자동화된 프로세스로서(요약) 문서 컬렉션에서 핵심 주제를 찾고(카테고리) 강력하고 유연한 조회를 사용하여 관련 문서를 검색합니다.

**IOCA.** IOCA(Image Object Content Architecture)를 참조하십시오.

**IOCA(Image Object Content Architecture).** 이미지를 상호 교환하고 표시하는 데 사용되는 구조의 컬렉션.

## J

**JavaBeans.** 『Bean』이라고 하는 재사용 가능한 Java 구성요소를 빌드하기 위한 플랫폼 독립 소프트웨어 구성요소 기술. 빌드한 후, 이 Bean을 다른 소프트웨어 엔지니어가 사용하거나 Java 응용프로그램에서 사용할 수 있습니다. 소프트웨어 엔지니어는 JavaBeans를 사용하여 그래픽 끌어서 놓기 개발 환경에서 Bean을 조작 및 어셈블할 수 있습니다

**JPEG.** JPEG(Joint Photographic Experts Group)를 참조하십시오.

**JPEG(Joint Photographic Experts Group).** (1) 디지털화된 연속 톤 이미지 압축의 표준을 설정하기 위해 작업하는 그룹. (2) 이 그룹에서 개발한 스틸 그림에 대한 표준.

## L

**LAN.** 근거리 통신망을 참조하십시오.

## M

**MediaArchiver.** 오디오 및 비디오 스트림 데이터를 저장하는 데 사용되는 실제 장치. VideoCharger는 MediaArchiver의 한 유형입니다.

**MIME 유형(MIME type).** 인터넷에서 전송되고 있는 오브젝트의 유형을 식별하기 위한 인터넷 표준. MIME 유형에는 여러 가지 변형된 오디오, 이미지 및 비디오가 포함되어 있습니다. 각 오브젝트에는 MIME 유형이 있습니다.



**MIME(Multipurpose Internet Mail Extensions)** . *MIME* 유형을 참조하십시오.

## O

**OLE.** *OLE(Object Linking and Embedding)*를 참조하십시오.

**OLE(Object Linking and Embedding).** 다른 응용프로그램 내에서 활성화될 수 있도록 응용프로그램 링크 및 포함 모두에 필요한 Microsoft® 스펙.

## P

**PID.** *PID(지속 식별자)*를 참조하십시오.

**PID(지속 식별자).** 저장된 위치에 관계없이 오브젝트를 고유하게 식별하는 ID. PID는 항목 ID와 위치로 구성됩니다.

## Q

**QBIC.** *QBIC(Query By Image Content)*를 참조하십시오.

**QBIC(Query By Image Content).** 일반 텍스트가 아닌 기능이라고 하는 비주얼 콘텐츠를 기본으로 검색할 수 있게 하는 조회 기법. QBIC를 사용하여 색 및 텍스처와 같은 비주얼 특성을 기본으로 오브젝트를 검색할 수 있습니다.

## R

**README 파일(README file).** 파일과 연관된 프로그램을 설치하거나 실행하기 전에 보아야 하는 파일. README 파일에는 일반적으로 최신 제품 정보, 설치 정보 또는 제품 사용에 대한 팁이 들어 있습니다.

**RIFF.** *RIFF(Resource Interchange File Format)*를 참조하십시오.

**RIFF(Resource Interchange File Format)** . 다른 유형의 컴퓨터 장비에서 재생할 소리 또는 그래픽을 저장하는 데 사용됩니다.

**RMI 서버(RMI server).** Java *RMI(Remote Method Invocation)* 분산 오브젝트 모델을 구현하는 서버

**RMI(Remote Method Invocation).** 분산 프로그래밍을 가능하게 하는 API 세트. 한 JVM(Java Virtual Machine)에 있는 오브젝트는 다른 JVM의 오브젝트에 있는 메소드를 호출할 수 있습니다

## T

**thin 클라이언트(thin client).** 설치된 소프트웨어는 거의 없거나 접속된 네트워크 서버에 의해 관리 및 전달되는 소프트웨어에 대한 액세스 권한은 갖고 있는 클라이언트. thin 클라이언트는 워크스테이션과 같은 전기능 클라이언트에 대한 대안입니다.

**Tivoli StorageManager(TSM).** 이기종 환경에서 기억영역 관리 및 데이터 액세스 서비스를 제공하는 *클라이언트/서버* 제품. TSM은 여러 가지 통신 메소드를 지원하고 파일의 백업 및 기억영역을 관리하기 위한 관리 기능을 제공하며 백업 조작을 스케줄하기 위한 기능을 제공합니다.

**TSM.** *Tivoli Storage Manager*를 참조하십시오.

**TSM 볼륨(TSM volume).** *Tivoli Storage Manager*에서 관리하는 논리 영역

## U

**URL(Uniform Resource Locator).** 컴퓨터 또는 인터넷과 같은 네트워크의 정보 자원을 나타내는 문자 순서. 이 문자 순서에는 정보 자원에 액세스하는 데 사용되는 프로토콜의 약어 이름과, 정보 자원을 찾기 위해 프로토콜에서 사용하는 정보가 포함됩니다. 예를 들어, 인터넷 구문에서 http, ftp, gopher, telnet 및 news는 여러 가지 정보 자원에 액세스하는 데 사용되는 프로토콜의 약어 이름 중 일부입니다.

## X

**XDO.** *XDO(Extended Data Object)*를 참조하십시오.

**XDO(Extended Data Object).** 응용프로그램에서 오브젝트를 기억영역 내부 또는 외부로 이동하는 데 사용되는 저장된 복합 멀티미디어 오브젝트의 일반적인 표현 방법. XDO는 주로 DDO에 포함되어 있습니다.

**XML.** *XML(Extensible Markup Language)*을 참조하십시오.

**XML(Extensible Markup Language).** 마크업 언어를 정의하는 표준 다중 언어로 SGML에서 파생되었으며 SGML의 서브세트입니다. XML은 좀더 복잡하고 많이 사용되지 않는 SGML의 부분을 없애며, 좀더 쉽게 문서 유형 및 작성자를 처리하고 구조화 정보를 관리하며 다양한 컴퓨터 시스템에서 구조화 정보를 전송 및 공유하는 응용프로그램을 작성할 수 있게 합니다. XML을 사용하는 데는 SGML에서는 필수였던 강력한 응용프로그램 및 처리는 필요하지 않습니다. XML은 W3C(World Wide Web Consortium)의 후원 아래 개발 중입니다.



# 색인

## [가]

검색 기준  
정의 및 맵핑 32  
검색 템플릿, 작성 31  
계획  
Enterprise Information Portal 2  
관리 도메인 41  
관리 클라이언트  
작성  
검색 기준 32  
검색 템플릿 31  
워크플로우 117  
정의  
작업 목록 116  
조치 117  
조치 목록 117

## [나]

내용 보기 프로그램 옵션 6

## [다]

도메인 46  
이해 42  
작성 42  
최상위 관리자 사용 권한 43, 44  
하위 관리자 사용 권한 43, 44

## [마]

멀티플랫폼용 IBM Enterprise Information  
Portal  
구성요소 3  
메타데이터 저장  
Information Mining 사용 49

## [바]

분류법  
정보 구조 도구 사용 71

## [사]

사용 권한 그룹 40  
사용 권한 세트 35, 40  
도메인 이동 47  
작성 40  
사용자 35  
도메인 이동 45  
사용 권한 세트 41  
사용자 그룹 41  
도메인 이동 46  
사용자 ID 35  
샘플 데이터 로드 131  
샘플 로더 프로그램, 실행 134  
서버 목록 21  
시작  
워크플로우 빌더 117

## [아]

액세스 가능성 151  
액세스 제어 목록  
도메인 이동 47  
워크플로우  
개념 109  
계획 113  
작성 117  
워크플로우 기능  
구성 109  
구성요소 115  
워크플로우 빌더  
설명 115  
시작 117  
워크플로우 작성 117  
웹 크롤러  
EIP 옵션 6  
이미지 검색  
구성 128  
별명 할당 130  
설정 128  
연결 검증 130  
이미지 검색 구성 128  
이미지 검색 옵션 5

## [자]

자원 관리자  
도메인에 할당 45  
사용자 할당 41  
자원 관리자, 도메인 이동 46  
작업 목록  
설명 114  
정의 116  
작업 패키지, 설명 113  
장애 151  
정보 구조 도구  
분류법 연계 83  
분류법 정의 73  
분류법 평가 78  
설명 71  
설치 71  
시작하기 72  
연계 문서 선택 75  
잠금 메커니즘 72  
WAS 사용 71  
조치 목록  
미리 정의된 조치 117  
정의 114, 117  
조치, 정의 117

## [카]

카탈로그  
삭제 73  
연계 83  
연계 문서 추가 75  
이름 바꾸기 73  
추가 73  
평가 78  
커넥터 5  
컨텐츠 서버  
정의 21  
컬렉션  
도메인 이동 47  
도메인에 할당 45  
키보드 151

## [ 타 ]

텍스트 검색

설정 127

XML 지원 127

텍스트 및 이미지 검색 문서 로드 134

## C

cmbcc2mime.ini 18

## E

EIP

관리 구성요소 4

내용 보기 프로그램 클라이언트 6

웹 크롤러 6

이미지 검색 클라이언트 5

커넥터 5

커넥터 툴킷 6

텍스트 검색 클라이언트 6

Information Center 구성요소 7

Information Mining 옵션 5

EIP 구성요소

관리 4

내용 보기 프로그램 6

운영 체제 호환성 3, 4

웹 크롤러 6

이미지 검색 5

커넥터 5

텍스트 검색 6

Information Center 7

Information Mining 5

Enterprise Information Portal

작성

검색 기준 32

검색 템플릿 31

워크플로우 117

정의

작업 목록 116

조치 117

조치 목록 117

## G

Grant Privilege Set 41

## I

Information Mining

구성요소 50

목표 그룹 52

분류법 빌드 71

비즈니스 환경에서 작업 52

서비스 49

설명 49

예 54

지원되는 문서 형식 57

지원되는 언어 57

## L

LDAP

가져오기 38

구성 38

## M

MIME 유형 사용자 조정 18

MIME 유형 파일

서버를 위해 변경 18

# IBM 한글 지원에 관한 설문



FAX : (02) 3787-0123

보내 주시는 의견은 더 나은 고객 지원 체제를 위한 귀중한 자료가 됩니다.  
독자 여러분의 좋은 의견을 기다립니다.

책 제목: 멀티플랫폼용 IBM Content Manager  
컨텐츠용 Information Integrator 관리  
버전 3 릴리스 2

책 번호: SA30-1550-01

성명			직위/담당업무	
회사명			부서명	
주소				
전화번호			팩스번호	
전자우편 주소				
사용중인 시스템	<input type="checkbox"/> 중대형 서버 <input type="checkbox"/> UNIX 서버 <input type="checkbox"/> PC 및 PC 서버			

1. IBM에서 제공하는 한글 책자와 영문 책자 중 어느 것을 더 좋아하십니까? 그 이유는 무엇입니까?  
☐ 한글 책자                      ☐ 영문 책자  
 (이유: \_\_\_\_\_)
2. 본 책자와 해당 소프트웨어에서 사용된 한글 용어에 대한 귀하의 평가 점수는?  
☐ 수                      ☐ 우                      ☐ 미                      ☐ 양                      ☐ 가
3. 본 책자와 해당 소프트웨어에서 번역 품질에 대한 귀하의 평가 점수는?  
☐ 수                      ☐ 우                      ☐ 미                      ☐ 양                      ☐ 가
4. 본 책자의 인쇄 상태에 대한 귀하의 평가 점수는?  
☐ 수                      ☐ 우                      ☐ 미                      ☐ 양                      ☐ 가
5. 한글 소프트웨어 및 책자가 지원되는 분야에 대해 귀하는 어떻게 생각하십니까?  
☐ 한글 책자를 늘려야 함                      ☐ 현재 수준으로 만족  
☐ 그다지 필요성을 느끼지 않음
6. IBM은 인쇄물 형식(hardcopy)과 화면 형식(softcopy)의 두 종류로 책자를 제공합니다. 어느 형식을 더 좋아하십니까?  
☐ 인쇄물 형식(hardcopy)                      ☐ 화면 형식(softcopy)                      ☐ 둘 다

☎ IBM 한글 지원 서비스에 대해 기타 제안사항이 있으시면 적어주십시오.

☺ 설문에 답해 주셔서 감사합니다.

귀하의 의견은 저희에게 매우 소중한 것이며, 고객 여러분들께 보다 좋은 제품을 제공해 드리기 위해 최선을 다하겠습니다.



프로그램 번호: 5724-B43

Printed in U.S.A

SA30-1550-01

