



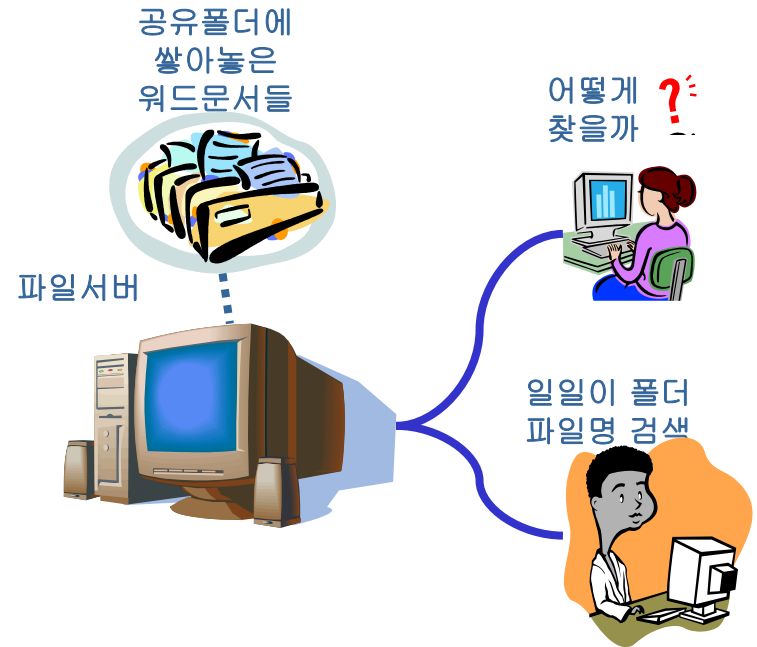
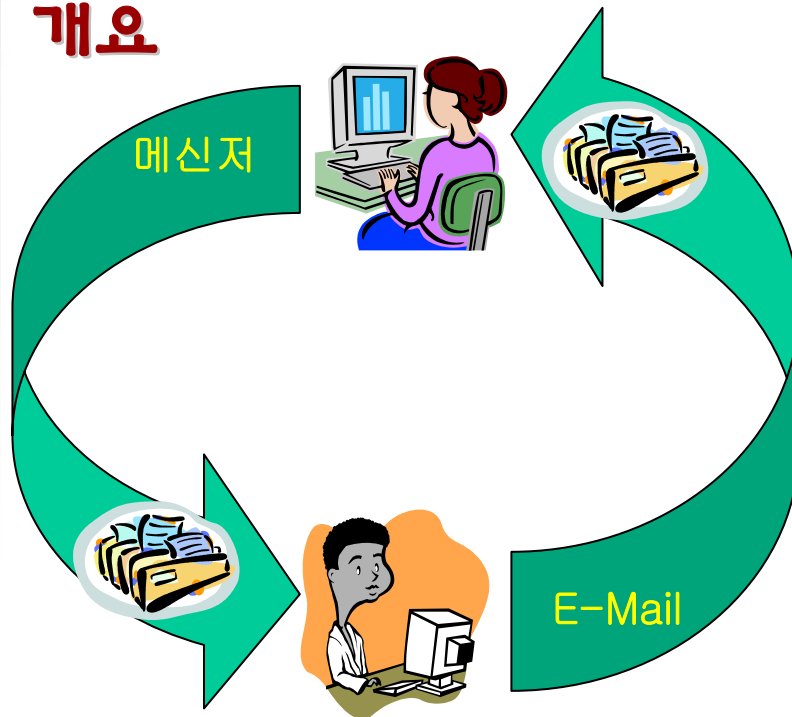
공유 워드문서 전문검색엔진 - JSearch-Document 1.0



[주] 코 아 뱅 크

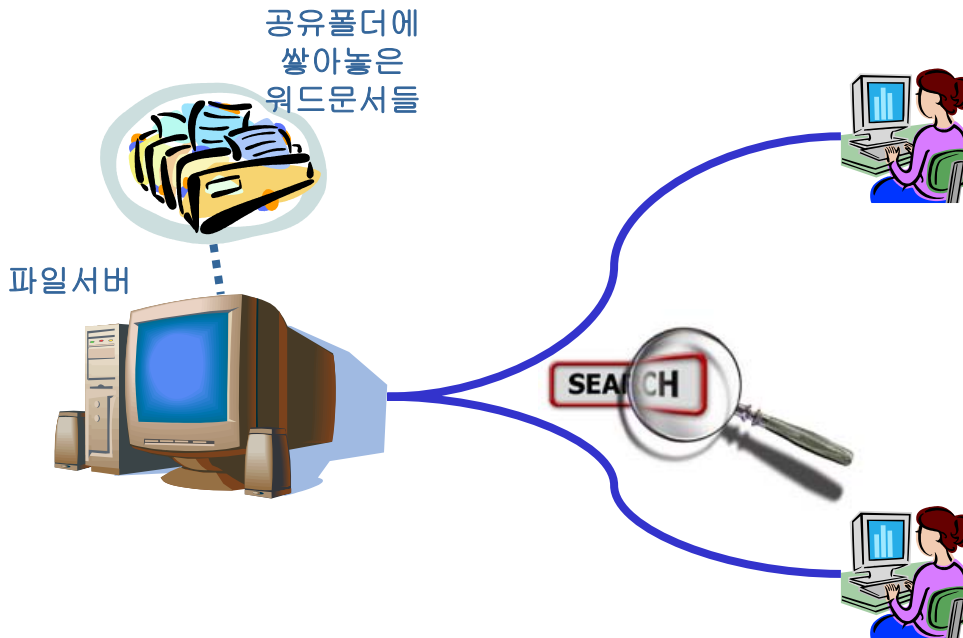


개요



❖ 정보습득능력은 고도로 정보화 되어가는 현대사회의 기업이나 개인의 경쟁력이다. 일반적인 포탈 검색 엔진 업체를 통해 정보를 얻는 데는 수월해 졌으나, 기업 내부의 회사 문서들을 찾는 데는 많은 시간과 방법이 요구되고 있다.

❖ 기업내부의 문서들은 기업구성원간 이메일 또는 메신저로 공유되며 더 나아가 파일서버를 두어 구성원 간의 문서 공유를 하게 된다. 그러나, 이러한 공유방식은 데이터의 백업이 되지 않거나 정보검색이 되지 않아 문서의 재사용이 용의하지 않다.



❖ JSearch-Document

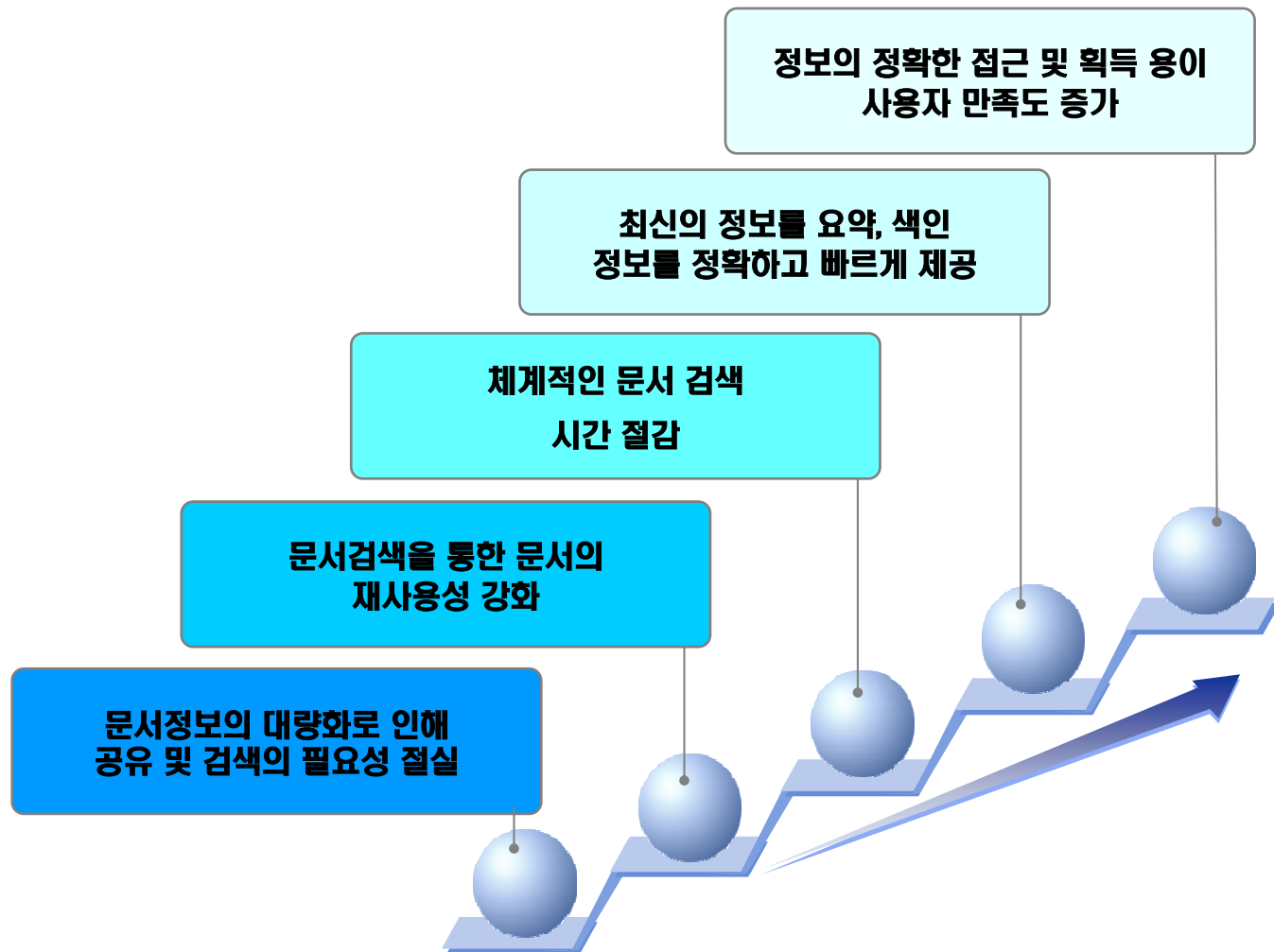


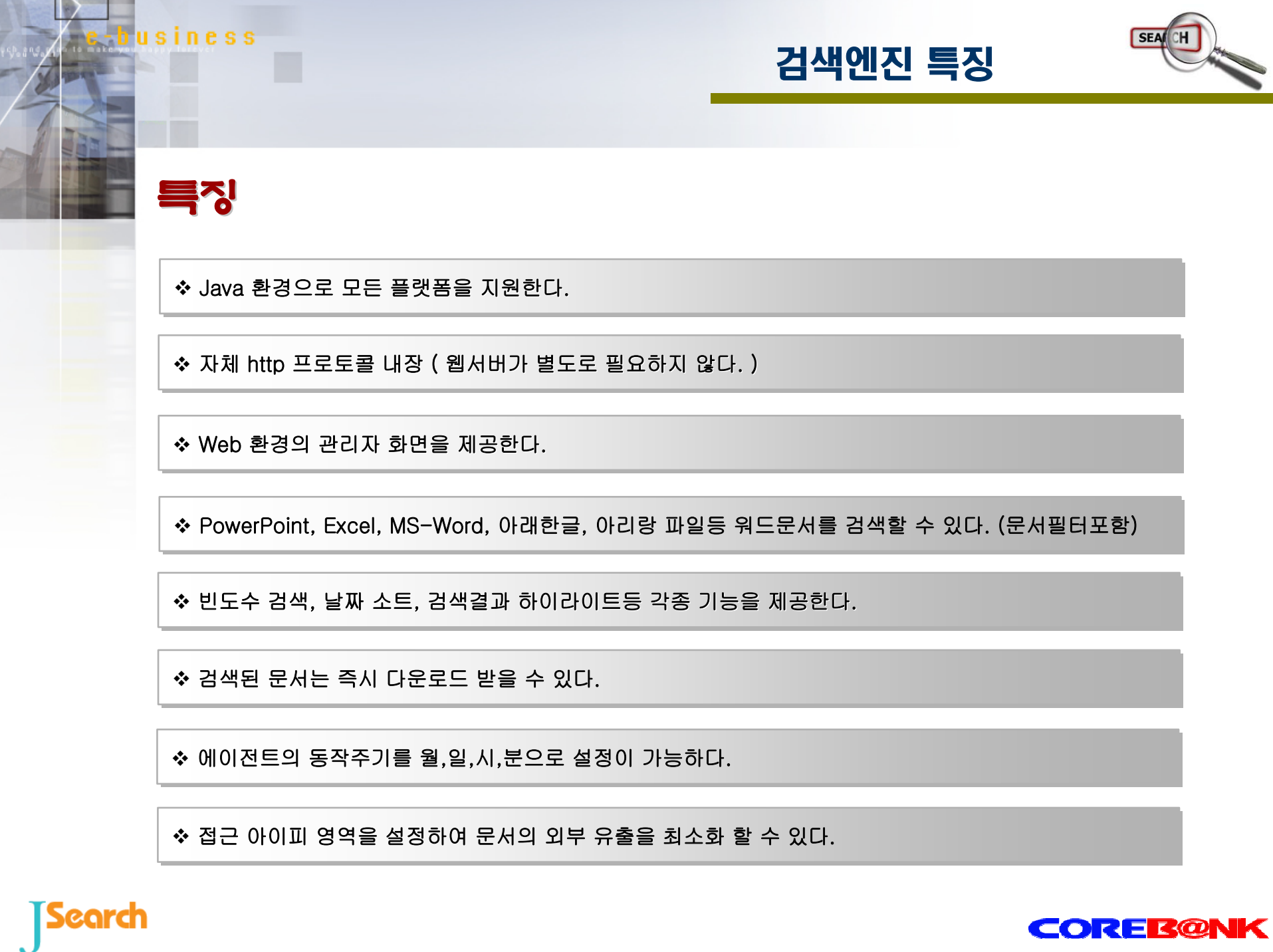
❖ 파일서버 내부의 공유 폴더를 주기적으로 감시하여 파일명 및 내용 검색이 가능하게 함으로써 정보의 검색 및 재사용이 가능하도록 하는 검색엔진이 필요하다.

❖ JSearch-Document 는 검색의 3대 요소인 'Easy Search', 'Fast Search', 'Accurate Search'를 실현시킬 수 있도록 만들어져 최고의 정확성과 초고속의 Performance를 보장하는 지능형 검색엔진 패키지이다.



기대효과





특징

❖ Java 환경으로 모든 플랫폼을 지원한다.

❖ 자체 http 프로토콜 내장 (웹서버가 별도로 필요하지 않다.)

❖ Web 환경의 관리자 화면을 제공한다.

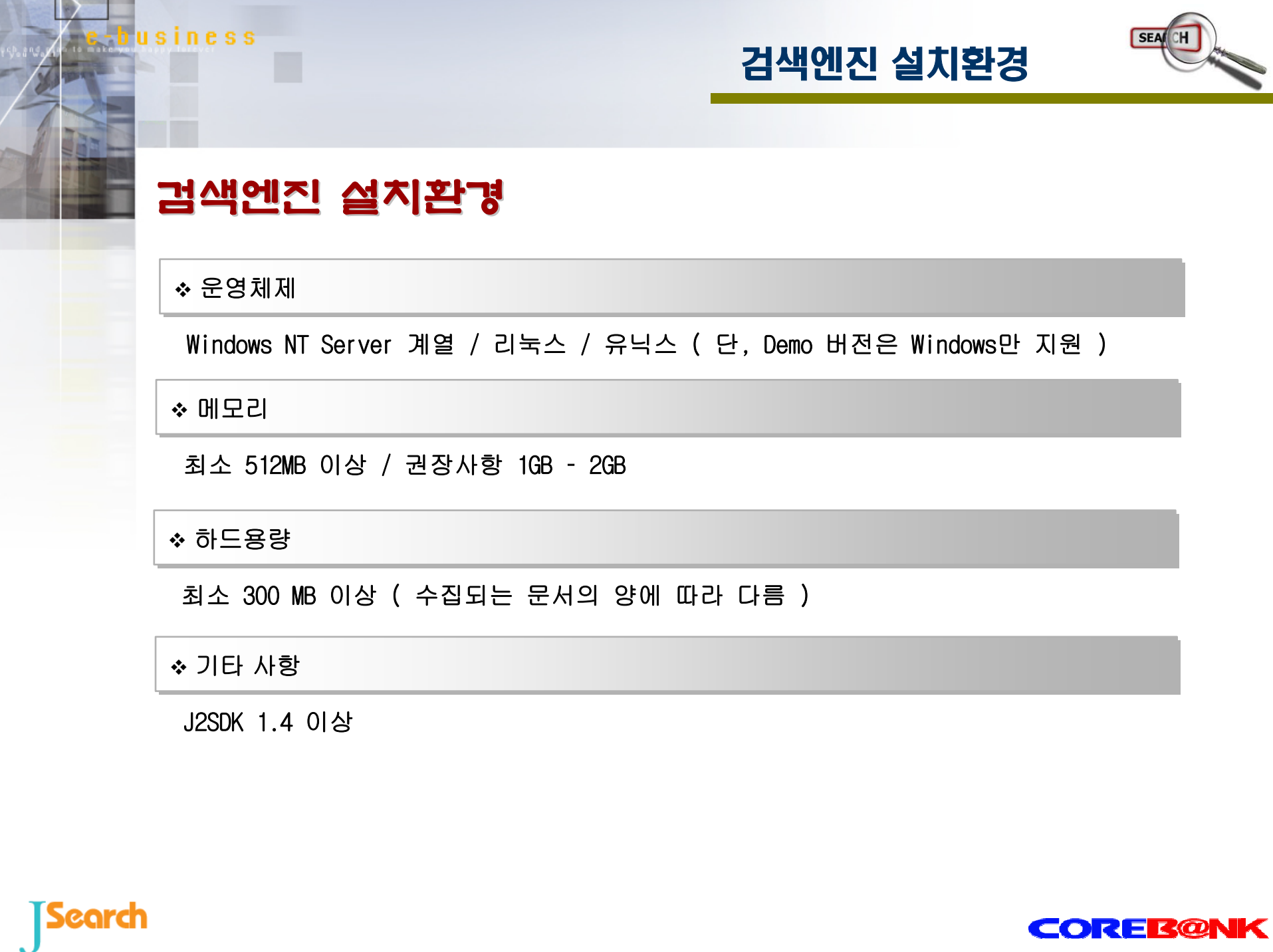
❖ PowerPoint, Excel, MS-Word, 아래한글, 아리랑 파일등 워드문서를 검색할 수 있다. (문서필터포함)

❖ 빈도수 검색, 날짜 소트, 검색결과 하이라이트등 각종 기능을 제공한다.

❖ 검색된 문서는 즉시 다운로드 받을 수 있다.

❖ 에이전트의 동작주기를 월,일,시,분으로 설정이 가능하다.

❖ 접근 아이피 영역을 설정하여 문서의 외부 유출을 최소화 할 수 있다.



검색엔진 설치환경

❖ 운영체제

Windows NT Server 계열 / 리눅스 / 유닉스 (단, Demo 버전은 Windows만 지원)

❖ 메모리

최소 512MB 이상 / 권장사항 1GB - 2GB

❖ 하드용량

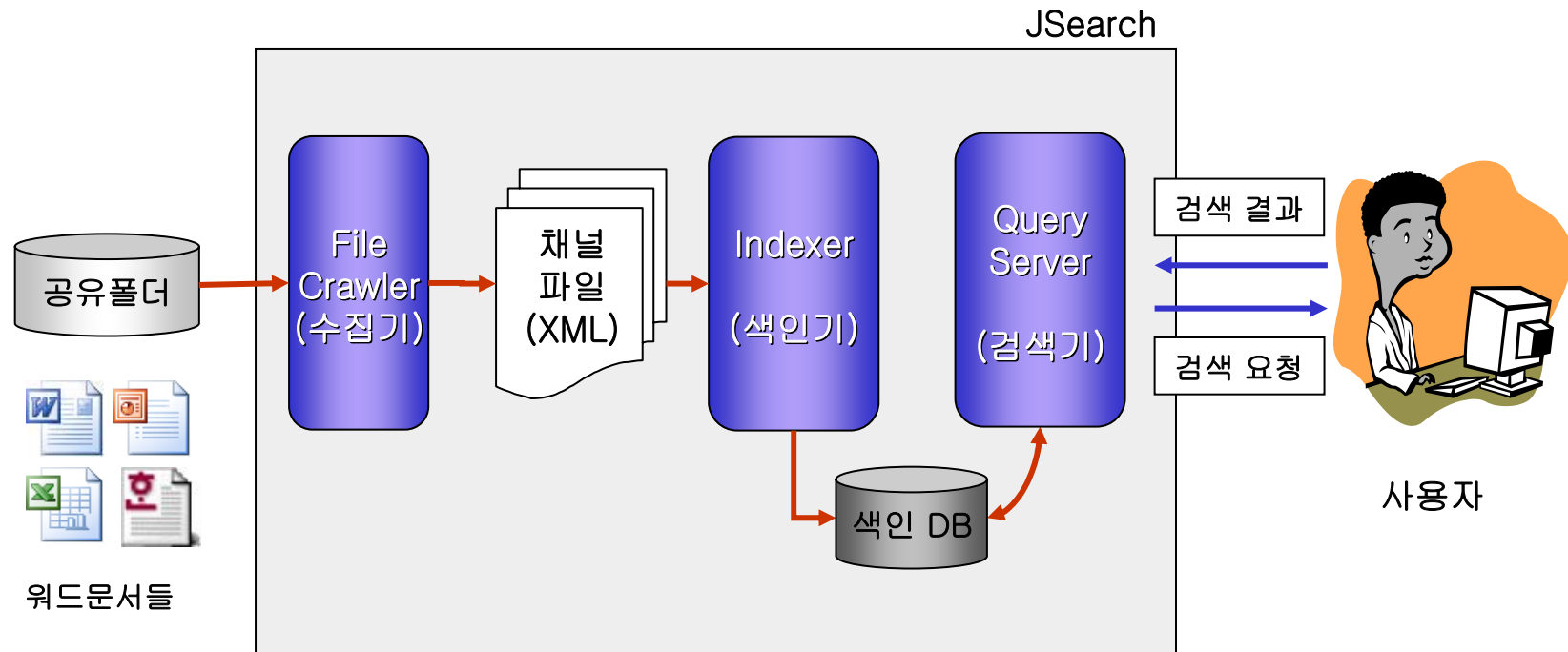
최소 300 MB 이상 (수집되는 문서의 양에 따라 다름)

❖ 기타 사항

J2SDK 1.4 이상



구성





주의사항

❖ 시스템 재부팅시에 필히 검색엔진을 정지시킨 후 하십시오.!

문서수집과 색인 에이전트가 동작중일때 재부팅 및 시스템 다운을 시킬 경우에는 색인데이터가 깨질 위험이 있습니다. 항상, 검색서버 정지 스크립트를 실행시킨 후 해당 프로세서가 정지된 것을 확인한 후 재부팅을 하여야 합니다.

❖ 수집되는 문서의 양이 많을 경우 시스템 사양이 충분해야 합니다.

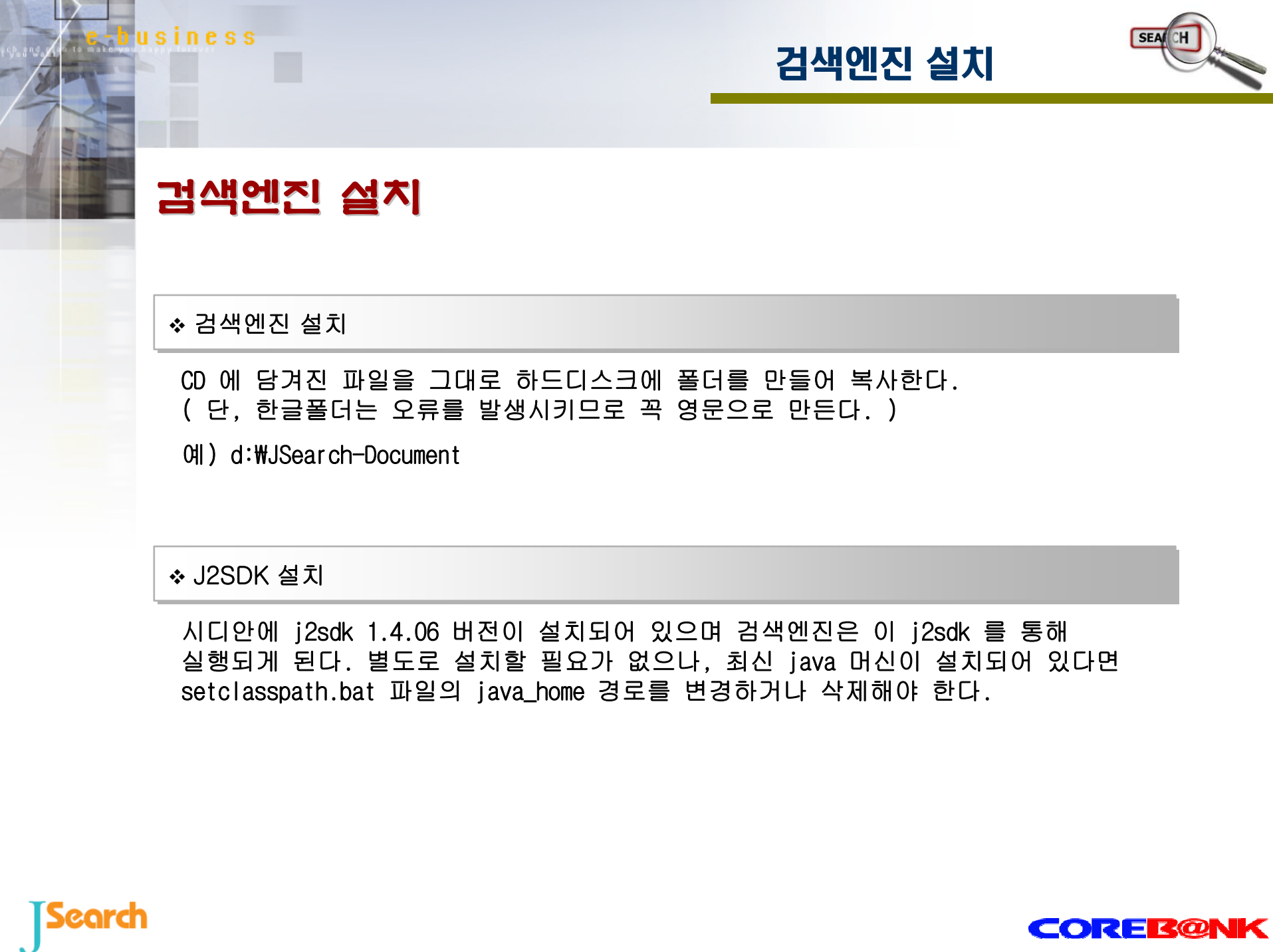
메모리가 2GB 이상이 되고 CPU 가 2개 이상이 되면 안정적인 운영이 가능합니다.

❖ 색인데이터가 깨졌을 경우 ?

관리자로 로그인한 후 [검색세트 관리] 메뉴를 클릭하여 [색인 문서 관리] 를 선택하시고, [전체삭제] 를 클릭하시면 전체가 삭제됩니다. 삭제후에 documentCrawler 를 다시 실행하시면 됩니다.

❖ 정보수집기와 색인기를 실행했는데, 색인데이터가 없습니다.

Jsearch-Document 가 설치된 폴더가 한글이거나, 상위 디렉토리가 한글로 되어 있을경우 에러를 발생합니다. 루트 디렉토리에 영문으로 폴더를 만드십시오.



검색엔진 설치

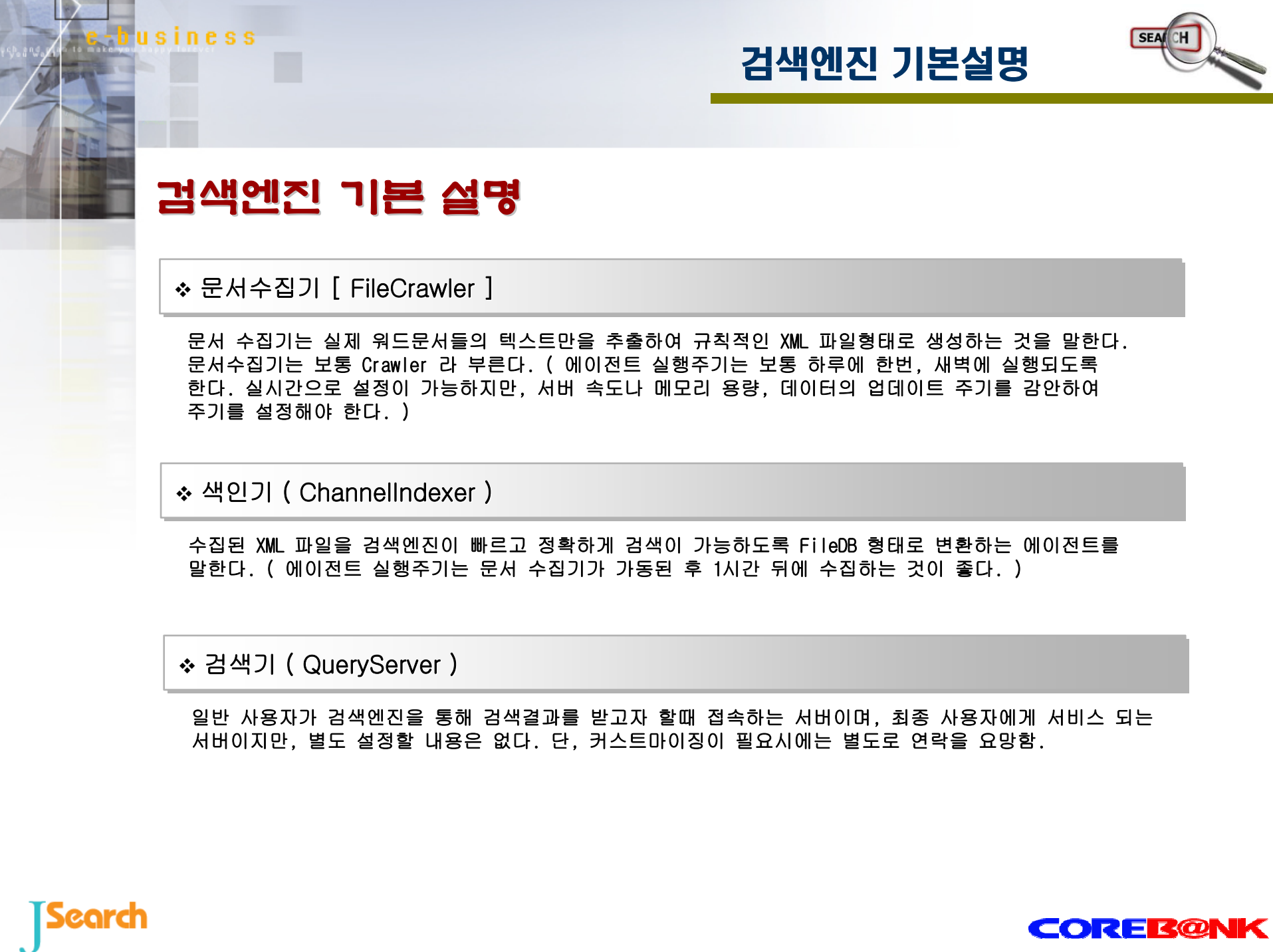
❖ 검색엔진 설치

CD 에 담겨진 파일을 그대로 하드디스크에 폴더를 만들어 복사한다.
(단, 한글폴더는 오류를 발생시키므로 꼭 영문으로 만든다.)

예) d:\WJSearch-Document

❖ J2SDK 설치

시디안에 j2sdk 1.4.06 버전이 설치되어 있으며 검색엔진은 이 j2sdk 를 통해 실행되게 된다. 별도로 설치할 필요가 없으나, 최신 java 머신이 설치되어 있다면 setclasspath.bat 파일의 java_home 경로를 변경하거나 삭제해야 한다.



검색엔진 기본 설명

❖ 문서수집기 [FileCrawler]

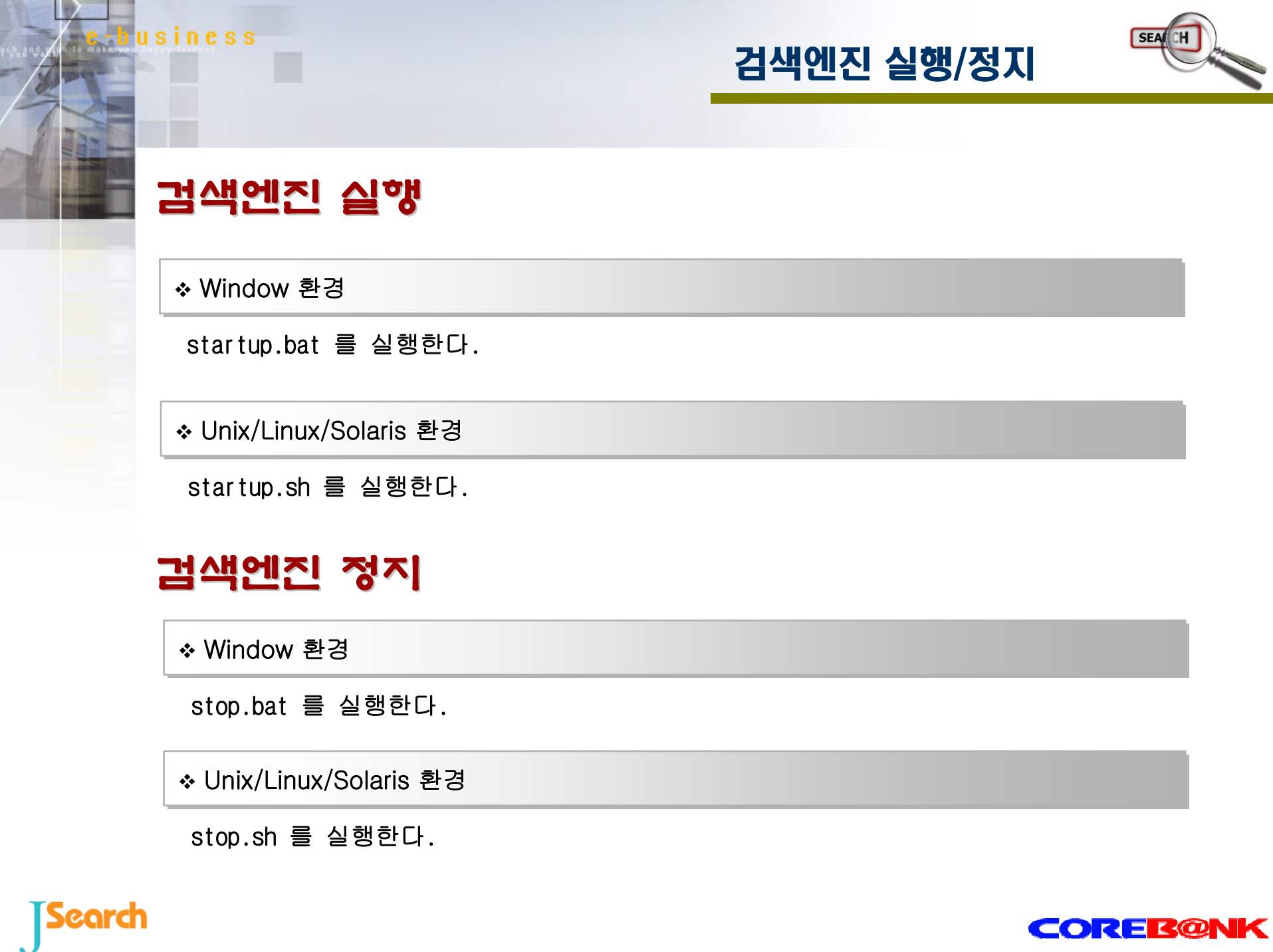
문서 수집기는 실제 워드문서들의 텍스트만을 추출하여 규칙적인 XML 파일형태로 생성하는 것을 말한다. 문서수집기는 보통 Crawler 라 부른다. (에이전트 실행주기는 보통 하루에 한번, 새벽에 실행되도록 한다. 실시간으로 설정이 가능하지만, 서버 속도나 메모리 용량, 데이터의 업데이트 주기를 감안하여 주기를 설정해야 한다.)

❖ 색인기 (ChannelIndexer)

수집된 XML 파일을 검색엔진이 빠르고 정확하게 검색이 가능하도록 FileDB 형태로 변환하는 에이전트를 말한다. (에이전트 실행주기는 문서 수집기가 가동된 후 1시간 뒤에 수집하는 것이 좋다.)

❖ 검색기 (QueryServer)

일반 사용자가 검색엔진을 통해 검색결과를 받고자 할때 접속하는 서버이며, 최종 사용자에게 서비스 되는 서버이지만, 별도 설정할 내용은 없다. 단, 커스트마이징이 필요시에는 별도로 연락을 요망함.



검색엔진 실행

❖ Window 환경

startup.bat 를 실행한다.

❖ Unix/Linux/Solaris 환경

startup.sh 를 실행한다.

검색엔진 정지

❖ Window 환경

stop.bat 를 실행한다.

❖ Unix/Linux/Solaris 환경

stop.sh 를 실행한다.



검색엔진 관리자 로그인

• JSearch-Enterprise 검색엔진을 실행한 후에 기본적인 설정을 하기 위해서는 관리자 화면에 접근한다.

• 인터넷 브라우저를 실행시켜 해당 컴퓨터의 아이피와 검색엔진이 점유하고 있는 포트를 이용하여 접속을 한다.

<http://서버아이피/admin/login.jlet>

예)
<http://127.0.0.1/admin/login.jlet>

• 사용자ID 와 비밀번호를 입력한다.
예) admin/manager

검색세트

JSearch-Document 3.2 - Microsoft Internet Explorer

파일(F) 편집(E) 보기(V) 즐겨찾기(A) 도구(T) 도움말(H)

뒤로

앞으로

홈

검색

즐거찾기

미디어

인tranet

이동

연결

주소 http://127.0.0.1:7000/admin/main.jlet

JSearch-Document

검색세트 관리

에이전트 동작

검색스킨 관리

추가기능 관리

JSearch-Document >> 검색세트 관리

검색세트	저장위치	색인방법	문자세트	관리기능	삭제
document	data/docFile	unicode	한국어 (euc-kr)	에이전트 정보관리 색인 문서 관리	삭제

Copyright (c) 2005 CoreBank corporation. All rights reserved.

- 검색세트란 ?

컴퓨터에서 하드디스크의 파티션과 같이 검색할 데이터를 하나의 데이터 폴더에 한 개 이상의 에이전트들이 수집한 데이터를 모으는 개념이다.
- 본 프로그램은 하나의 검색세트가 기본으로 설정되어 있으며, 삭제가 가능하지 않다.
- 최초 로그인을 하게 되면 한개의 검색세트가 보이게 된다.
- [검색세트 관리] 상위메뉴를 누르고 [에이전트 정보관리] 를 클릭한다.



에이전트 수정

JSearch-Document 1.0 - Microsoft Internet Explorer

파일(F) 편집(E) 보기(V) 즐겨찾기(A) 도구(T) 도움말(H)

뒤로 - - - - - 검색 - - - - - 즐겨찾기 - - - - - 미디어 - - - - - 이동 - - - - - 연결 >>

주소(D) http://127.0.0.1:7000/admin/main.jlet

JSearch-Document

검색세트 관리 에이전트 동작 검색스킨 관리 추가기능 관리

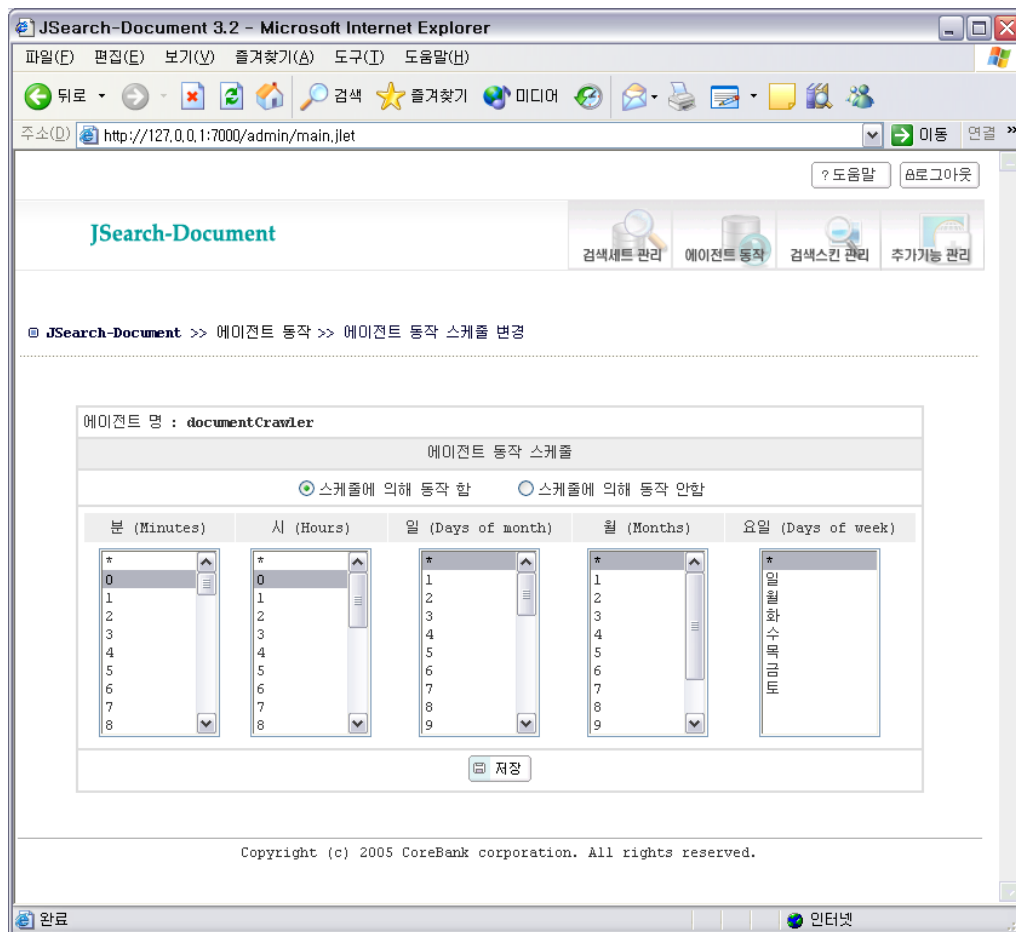
JSearch-Document >> 검색세트 관리 >> (document) 에이전트 관리 >> 파일크롤러 등록정보 변경

에이전트 기본 정보	
에이전트 명	documentCrawler
에이전트 클래스	com.janusystem.janus.agent.FileCrawler
로그기록 처리수준	정보, 경고, 오류 (3)
디렉토리 경로 정보	
디렉토리 전체 경로 1	c:\Documents and Settings\김문수\My Documents\문서작업\
디렉토리 전체 경로 2	c:\DOCS\
디렉토리 전체 경로 3	
디렉토리 전체 경로 4	
색인방법	☒ 자동색인 ☐ 스케줄 색인
추가 필드 정보	
extra.field.1	
extra.field.2	

- 에이전트란?
에이전트는 검색세트의 동작하는 프로그램들이다. 최소한 한 개의 색인기(Indexer)와 정보수집기(Crawler)가 존재해야 한다.
- [documentCrawler] 이름을 클릭한다.
- [디렉토리 전체 경로 1]에 문서를 수집할 절대 경로 입력
- 정보수집기가 동작한 후 색인기를 자동으로 동작시키려면 [색인방법] 항목에 자동색인을 선택한다.
- 다른 폴더도 추가하고 싶으면 [디렉토리 전체 경로 2] ... 4 까지 차례로 전체경로를 입력한다.
- [저장]버튼을 누르면 에이전트를 설정이 완료된다.



에이전트 스케줄 설정



• 에이전트 스케줄이란 ?

해당 에이전트가 주기적인 동작을 위하여 linux 의 cron 과 같이 주기를 설정하는 것을 말한다.

• 멀티셀렉트를 지원하므로 특정 시간의 동작이 가능하다.

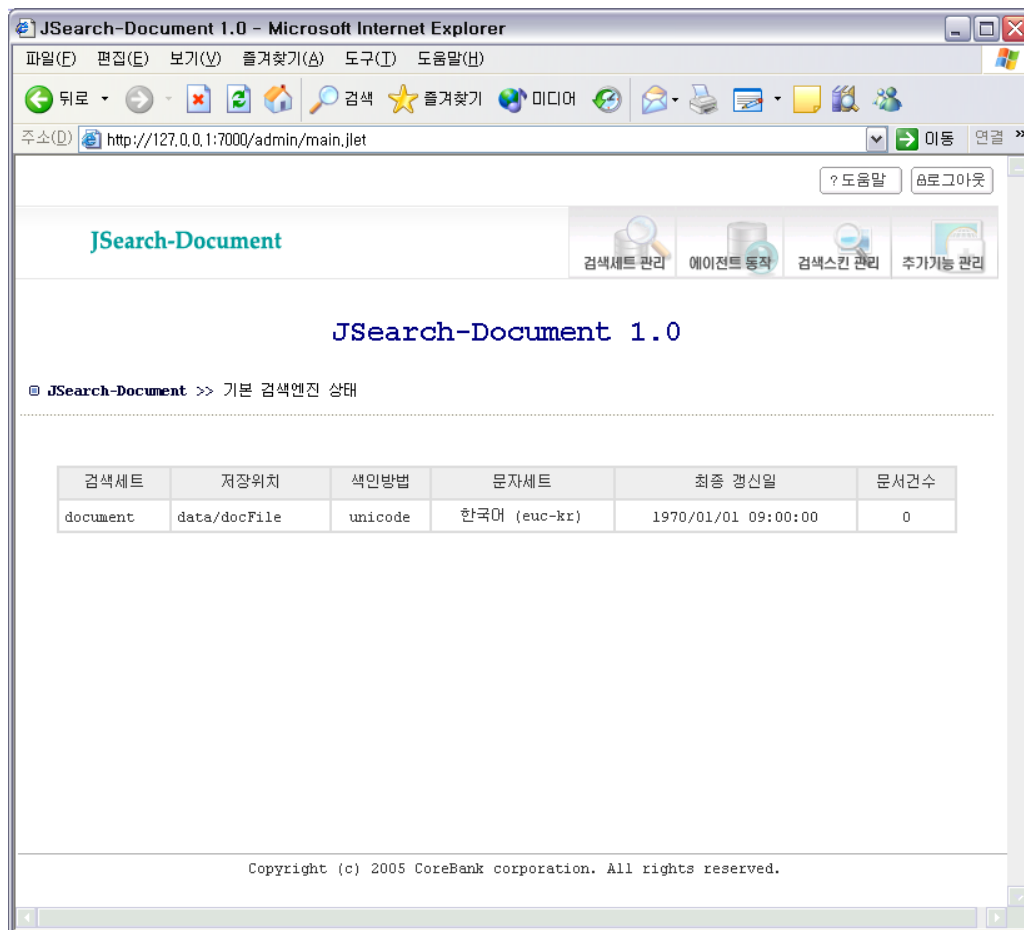
• 옆의 설정은 매요일,매월,매일, 매시,0분에 동작하는 설정이다.즉,한 시간에 한번씩 돌아간다.

• [저장] 을 누른다. 그러면 검색세트에 방금 설정한 에이전트가 등록된다.

• 색인기의 스케줄 설정은 항상 documentCrawler 에이전트가 실행된 후 일정시간 후에 실행을 시켜야 한다. 단, 자동색인일 경우 스케줄 설정을 안해도 된다.



설정완료



- 해당 에이전트들은 설정한 주기에 의해 자동으로 동작을 하게 된다.
- 에이전트 동작을 즉시 실행하고자 할 경우에는 상위 메뉴의 [에이전트 동작] 을 선택하여 해당 에이전트를 실행시키면 즉시 동작을 하게 된다.
- 문서의 수집과 색인이 완료가 되면 이제 검색엔진을 이용한 검색을 시작할 수 있게 된다.



접근 영역 설정

JSearch-Document 3.2 - Microsoft Internet Explorer

파일(F) 편집(E) 보기(V) 즐겨찾기(A) 도구(T) 도움말(H)

뒤로, 앞으로, 새로고침, 검색, 즐겨찾기, 미디어, 이동, 연결

주소(D) http://127.0.0.1:7000/admin/main.jlet

JSearch-Document

검색세트 관리, 에이전트 동작, 검색스킨 관리, 추가기능 관리

◎ JSearch-Document >> 접근 아이피 관리

수정

**** 아이피 등록 방법 ****

1. 해당 아이피간의 구분자는 공백(space) 입니다. 예) 211.222.221.1 211.222.221.7
2. 아이피 영역을 표시할때에는 -(마이너스)로 구분을 합니다. 예) 211.222.221.1-211.222.221.30
3. 특정 아이피 영역을 전체를 제한할때에는 * 로 구분합니다. 예) 211.222.221.*

아이피 영역	
허용 아이피 영역	
거부 아이피 영역	192.168.1.77

Copyright (c) 2005 CoreBank corporation. All rights reserved.

완료 인터넷

• 접근 아이피 관리란?

검색엔진을 사용하여 검색하고 다운로드 받지 않아야 할 컴퓨터의 접근을 제한하는 것이다. 설정을 하면 관리자 화면 접근이나 검색이 허용되지 않는다.

• 접근 아이피 허용과 거부가 있는데 중복되는 경우 거부가 허용보다 우선순위에 들어간다.

• 해당 아이피간의 구분자는 공백(space) 입니다.
예) 211.222.221.1 211.222.221.7

• 아이피 영역을 표시할때에는 -(마이너스)로 구분을 합니다.
예) 211.222.221.1-211.222.221.30

• 특정 아이피 영역을 전체를 제한할 때에는 * 로 구분합니다.
예) 211.222.221.*



검색 페이지

JSearch-Document - Microsoft Internet Explorer

파일(F) 편집(E) 보기(V) 즐겨찾기(A) 도구(T) 도움말(H)

뒤로 - 앞으로 - 새로고침 - 검색 - 즐겨찾기 - 미디어 - 인쇄 - 연결

주소(D) http://127.0.0.1:7000/query/index.html 이동 연결 >>

JSearch-Document

검색어	파일명+본문	검색
정렬기준	☒ 정확도 ☐ 최근문서 ☐ 과거문서	

▶ 검색어를 입력하십시오.

▶ 검색 연산자

AND 검색 : 한국 && 은행 / 한국 AND 은행
 OR 검색 : 한국 은행 / 한국 || 은행 / 한국 OR 은행
 NOT 검색 : 한국 -은행 / 한국 NOT 은행

Copyright © 2005 CoreBank corporation. All Rights Reserved.

완료 인터넷

- 왼쪽의 그림과 같이 검색서버가 설치된 서버의 아이피와 포트를 입력하면 검색기본 페이지가 출력된다.

예) <http://192.168.1.43:7000/>

- 파일명 및 본문 검색이 기본으로 설정되어 있으며, 각각의 옵션을 설정하고 검색을 시작한다.

검색 결과 및 다운로드

JSearch-Document

검색어: 파일명+내용 은행 검색

정렬기준: ☒ 정확도 ☐ 최근문서 ☐ 과거문서

>> 검색결과(1 - 10) 총 50 문서가 검색되었습니다.

1. 등록및관리내용(결재원) TYPE : hwp 등록일 : 20050217

... L/C 등록 보관 시스템통지은행으로부터 전송 받은 L/C를 KNET으로 전송과 동시에 은행공통의 e-L/C 관리시스템에 등록 보관하는 시스템① 해외 L/C 개설은행은 개설된 L/C를 스위프트 네트워크를 통하여 국내 통지은행으로 전송② 통지은행은 전송받은 L/C의 진정성을 확인하고 해당 L/C를, ...

- PATH : c:\#DOCS\#등록및관리내용(결재원).hwp

2. e-LC관련은행현황 TYPE : doc 등록일 : 20050329

... 5월준비된 은행순으로 테스트 시행2) 2005.6월2개은행 시범 실시(외환, 하나)3) 2005년 하반기희망은행 순서대로 이행 실시

- 현재 업무 분석중- 현재의 시스템 상황에서는 하반기전은행 시행 시 적용 불확실함(주요 Issue 사항 참조) EDI 내국신용장 결제 시스템 - EDI 내국, ...

- PATH : c:\#DOCS\#e-LC관련은행현황.doc

3. (406)대내외.장단기구분 TYPE : xls 등록일 : 20010811

... 대외장기 5 대내소계 한국은행 계산용 6 대외소계 한국은행 계산용 7 단기소계 한국은행 계산용 8 장기소계 한국은행 계산용 9 합계 한국은행 계산용 +2000.02산설 부 문대내외/장단기구분외환정보망작 성 자료도함목 분석표시 스텝Code NO외화 B/S406디맨드명작성일자박노열2000.02.29, ...

- PATH : c:\#DOCS\#(406)대내외.장단기구분.xls

완료 인터넷

- 검색어를 입력하여 검색결과를 출력시키는데, 검색결과의 파일명을 클릭하면 웹페이지로 해당 파일을 다운로드 할 수 있다.