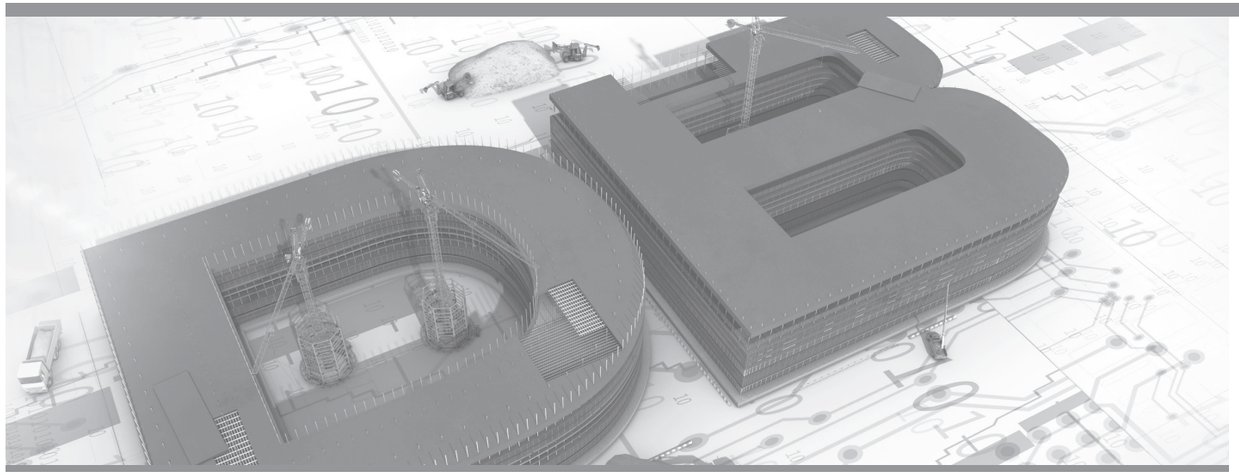
제7장



A Guide for DataBase Project Management

검색엔진 프로젝트 관리 가이드

제1절 | 개요

제2절 | RFP 작성 준비

제3절 | 표준 RFP 요구사항

제4절 | 프로젝트 점검사항

제5절 | 프로젝트 평가

제1절 | 개요

1.1. 검색엔진의 개요

일반적으로 검색엔진이라 하면 인터넷 상에 흩어져 있는 자료를 쉽게 찾을 수 있도록 도와주는 소프트웨어를 말한다. 반면, 전문 검색엔진은(Full-Text Retrival Engine)은 기관의 정보시스템 내 분산되어 있는 정보를 수집 및 색인하여 사용자의 질의에 적합한 결과를 제공하는 시스템을 말한다.

이러한 전문검색엔진은 기사 검색 시스템을 비롯해 각종 기술 정보, 문서 정보, 법률, 특허 정보 등의 시스템에서 원하는 결과를 찾을 때 활용할 수 있다. 최근 정보량이 늘어나면서 필요한 정보를 효과적으로 검색, 수집, 관리하는 방법이 중요한 과제가 되고 있어 이를 도와주는 검색 엔진의 효용성도 커지고 있다.

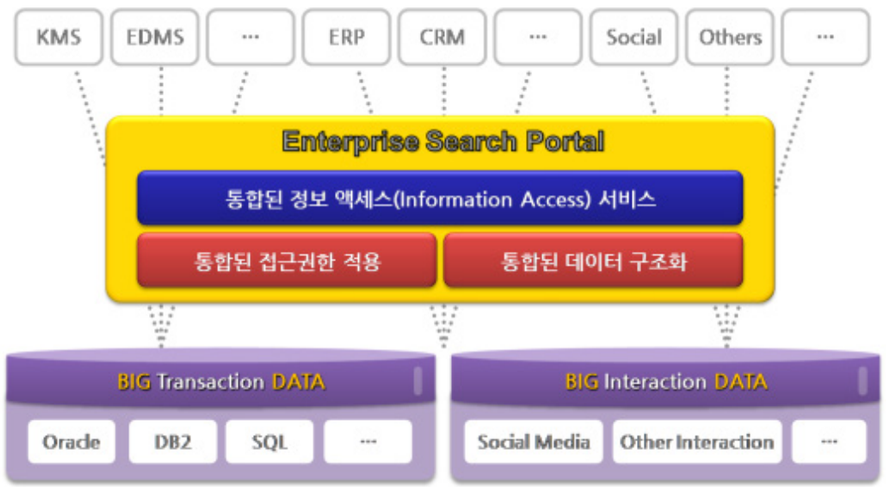
■ **검색엔진에 대한 이해**

[표 7-1] 검색엔진에 대한 이해

정보검색	가공되지 않은 데이터들에서 가치있는 정보를 추출(Extraction)한다는 뜻으로 검색이란 용어를 “찾다(Search)”라는 의미에서 한층 강화된 “선택해 찾는다(Retrieval)”라고 해석 정보의 집합에서 필요한 정보를 탐색해내는 인간의 행동 일정한 곳에 축적되어 있는 정보들 중에서 어떠한 목적에 부합되는 정보를 입수하는 행동
검색엔진	임의로 들어오는 사용자 질의에 대비하여 네트워크 상에 분산되어 있는 각종 정보를 수집/색인하고 질의에 적합한 결과를 정리하여 제공하는 시스템
웹 검색엔진	웹 사이트에 들어있는 정보를 색인해 두었다가 검색자가 필요한 정보와 관련된 단어(낱말)을 입력하면 검색어와 관련된 사이트 혹은 웹페이지들을 보여주는 프로그램
통합검색엔진	웹사이트 뿐만 아니라 DBMS, 비정형문서 등 거의 모든 유형의 문서에 대한 검색기능을 제공하는 솔루션

통합검색시스템은 정보 수요자가 필요하다고 예측되는 정보나 데이터를 미리 수집, 가공, 처리하여 찾기 쉬운 형태로 축적해 놓은 데이터베이스로부터 요구에 적합한 정보를 신속하게 찾아내어 정보요구자에게 제공하는 시스템이다.

공공기관은 홈페이지를 비롯하여 문서관리시스템, 지식관리시스템, 그룹웨어 등 다양한 시스템을 운영하고 있으며 각 시스템에 축적된 대량의 정보에서 원하는 것을 검색하여 사용자에게 제공해야 한다.



[그림 7-1]

이러한 기능을 전문검색엔진(Full-Text Search Engine)이 수행하며 전문적인 검색솔루션을 도입하여야 한다.

■ **검색엔진의 필요성**

- 정보화시스템의 보급증가에 따라 조직내 정보들이 기하급수적으로 증가
- 조직에 유용한 정보들이 각 시스템에 산재되어 있어 정보 검색의 어려움 존재
- 데이터량 증가에 따라 가치있는 정보를 찾기 위한 비용 증가
- 공공기관의 대국민 정보제공 서비스를 위한 필수 기능
- 기업의 업무향상성을 높이기 위한 주요 기능

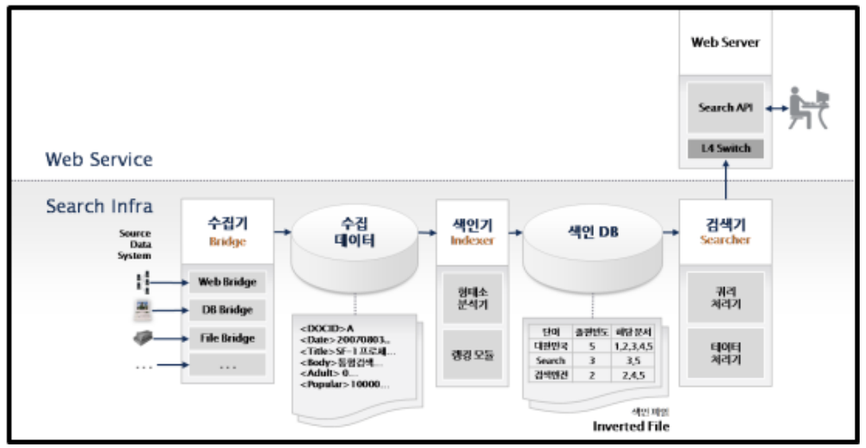
■ **검색엔진 기능**

검색엔진은 일반적으로 수집기, 색인기, 검색기로 구성된다.

수집기(Bridge 혹은 Gateway)는 웹사이트의 데이터나 정보시스템 내 저장된 데이터와 파일 등을 수집하고, 색인기(Indexer)는 수집된 원문을 기반으로 역파일

(Inverted File)로 변환하여 검색하기 쉬운 구조로 만든다. 검색기(Searcher)는 사용자의 검색질의에 대하여 검색결과값을 보여주는 기능을 한다.

일반적인 검색엔진의 구조는 아래와 같다.



[그림 7-2]

■ **수집기**

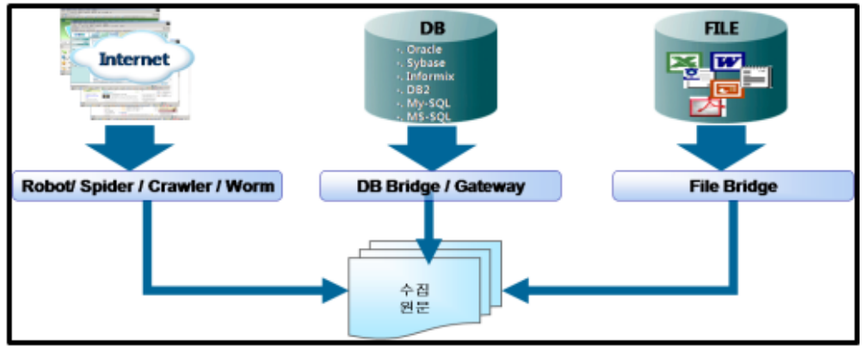
검색엔진의 수집기는 웹 수집기, DB 수집기, 파일 수집기로 구분할 수 있다.

웹 수집기(Web Crawler)는 인터넷상의 HTML문서를 수집하여 색인가능한 형태로 만드는 기능을 수행한다. 보통 홈페이지의 유관기관 사이트를 수집할 때 사용하며, 웹 사이트 수집기 중에서도 게시판을 전문적으로 수집하는 수집기가 있는 이를 활용하여 유관 게시판을 집중적으로 수집/저장하여 모니터링 할 수도 있다.

DB 수집기는 조직 내 저장된 여러 종류의 데이터베이스를 읽어들여 색인가능한 파일로 만드는 기능을 수행한다. 검색업체별로 Bridge, Gateway, Gatherer 등의 명칭을 사용하기도 한다.

파일(File)수집기는 시스템에 첨부된 문서나 특정 디렉토리 내 전자문서(MS 오피스 파일, 아래한글, PDF, 훈민정음, TXT 파일 등)을 수집하여 색인 가능한 형태로 만

드는 기능을 수행한다. 이 기능은 첨부파일에서 텍스트만 추출하는 파일필터가 반드시 있어야 가능하므로, 파일수집기 기능에 필터기능이 있는지를 검토하는 것이 매우 중요하다.



[그림 7-3]

■ **색인기(Indexer)**

색인기는 대량의 문서를 빠르게 검색하기 위하여 수집된 데이터를 다양한 방식으로 색인하는 모듈이다.

[표 7-2]

종류	용도	적용 예
형태소분석기를 사용한 색인	언어 분석기와 언어사전을 이용하여 단어 단위의 색인을 필요로 할 때 사용	데이터베이스진흥원 → 데이터베이스진흥원, 데이터, 베이스, 진흥원, 데이터베이스
TOKEN	아무런 언어분석도 사용하지 않을 경우에 사용. 일치 검색과 같은 효과를 내기 위해 사용.	데이터베이스진흥원 → 데이터베이스진흥원
bi-gram	바이그램으로 색인어를 추출할 때 사용	데이터베이스진흥원 → 데이, 이터, 터베, 베이, 이스, 스진, 진흥, 흥원
LIKE	데이터베이스의 like검색과 비슷한 효과를 내기 위해 사용 (색인 용량이 매우 커짐)	데이터베이스진흥원 → 데이, 이터, 터베, 베이, 이스, 스진, 진흥, 흥원, 데이터, 이터베, 터베이, 베이스, 이스진, 스진흥, 진흥원, ….

색인하는 방법은 아래와 같다.

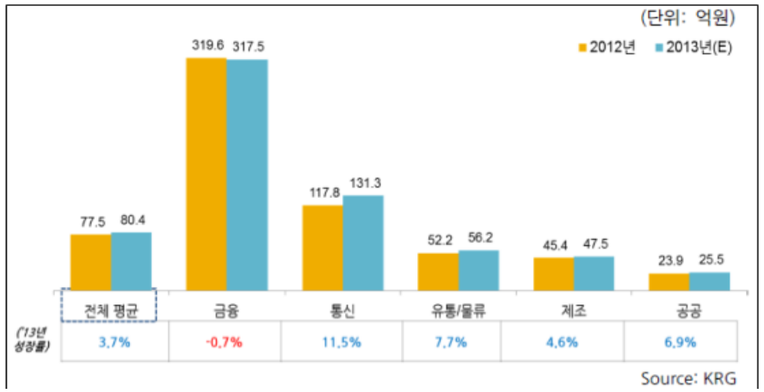
일반적으로는 형태소분석기를 사용한 색인을 사용하여 검색시스템을 구축한다. 필요에 따라 BI-Gram, LIKE 등의 옵션으로 색인을 하기도 하지만 이런 경우 색인용량이 크고 속도가 느릴 수 있으므로 확인을 해야 한다.

1.2. 검색엔진 동향

2013년 검색부문은 이전과 같이 통합검색, 검색고도화 이슈를 유지하긴 하나 빅데이터 / 소셜분석 등과 같이 비즈니스와 연계된 IT사업을 중점적으로 진행한 것으로 보인다.

2013년 검색시장의 Hot Keyword로는 빅데이터, BI, 문서중앙, 시스템 고도화 등으로 요약되며, 공공부문에서는 SW진흥법 개정안 발효로 인하여 중견SI의 공공시장 본격 진출과 공공정보화사업 선진화, 그리고 정부의 중점 육성 분야인 정보교육화 사업, 헬스케어 IT 분야들이 주목되고 있다. 또한 일부 시범 프로젝트에서는 오픈소스 기반의 검색엔진 도입도 추진되고 있으며, 검색결과의 활용도를 높이는 측면에서 N-Screen 기반의 검색UI가 많이 제안되고 있다.

기업부문에서는 ERP와 GW, KMS 고도화와 CRM 신규 도입 등 기간 시스템 투자 증가 및 빅데이터, Cloud Computing, Mobile 부문에 대한 관심이 높아지고 있으며 빅데이터의 경우 마케팅 영역을 중심으로 활용도가 높아지는 추세이다.



[그림 7-4] Fig. 2013 년 업종별 평균 IT 투자

공공/기업부문 모두 내부 데이터량이 증가함에 따라 이미 도입한 검색엔진의 업그레이드, 시스템 고도화를 병행하여 진행한 사례가 많았으며, 특히 빅데이터의 이슈가 증가함에 따라 외부 Web & Social 데이터를 수집하여 내부 데이터와 교차분석 후 그 결과를 활용하는 프로젝트가 진행되었다.

1.3. 검색엔진 도입 효과

각 정보시스템(그룹웨어, 지식관리시스템, 문서관리시스템, SAP, ERP 등)에는 기본적으로 DB검색기능이 내장되어 있다. 따라서, 일반적인 제목, 제목+본문 검색은 수행되나 상세한 조건식을 가지고 수행하는 상세조건검색은 수행할 수 없다.

또한, 각 시스템에서 보유하는 데이터량이 증가되고 사용자가 늘어남에 따라 검색효율이 서서히 둔화되어 사용하지 어렵게 되거나 원하는 검색결과를 얻을 수 없게 된다.

반면, 전문적인 검색솔루션을 도입하면 정보시스템 내 데이터를 빠르고 효과적으로 검색할 수 있고 다양한 검색조건을 활용한 상세조건 검색결과를 얻을 수 있다. 또한, 제목, 본문, 작성자, 날짜와 같은 일반적인 검색과 AND, OR, NOT 연산자를 활용한 Boolean 검색 및 정보시스템 내에 함께 저장된 첨부파일(MS Office, 아래한글, PDF 등)도 검색되므로 사용자에게 최적의 검색결과를 제공할 수 있다.

■ 기본 연산자 설명

기본연산자는 효율적이고 정확한 검색을 위해 검색어 사이의 논리적인 관계를 지정해 주는 검색 연산자 입니다. AND, OR, NOT 연산자가 이에 포함된다.

[표 7-3] 기본연산자(Boolean Operation)의 표현

검색연산자	사용사례
ADD연산자[&],[]	검색&기업 또는 검색 기업
OR 연산자[|]	검색|기업
NOT 연산자[!]	검색!기업

검색 & 기업: '검색'과 '기업'두개의 키워드가 모두 포함된 결과를 찾음

검색 | 기업: '검색'이나 '기업' 둘 중의 하나의 검색어라도 포함되면 검색 결과로 찾음. 보다 포괄적인 검색결과를 얻을 수 있음.

검색 ! 기업: '검색'이 포함된 결과 중에서 '기업'이 포함되지 않은 결과를 검색함. 이 NOT 연산은 검색결과에서 원하지 않은 문서를 검색어를 통해서 제외하는데 있어서 매우 유용한 연산자임.

■ **상세조건 검색 설명**

상세조건 검색을 통해 사용자는 보다 정확한 검색결과를 얻을 수 있다. 찾고자 하는 검색대상을 명확히 하고 검색조건을 세분화할 수 있으며 검색기간을 지정하여 입력하므로 다량의 검색결과에서 원하는 검색결과를 추출할 수 있다.

일반적으로 홈페이지 영역은 아래와 같은 방식으로 상세조건을 구성한다.

[그림 7-5] 홈페이지 상세검색 #1

[그림 7-6] 홈페이지 상세검색 #2

검색대상이 많은 도서관, 자료관 등은 아래와 같이 검색대상(자료유형)을 뿐 아니라 검색키워드도 자유롭게 입력하는 상세조건을 많이 사용한다.

[그림 7-7] 도서관 상세검색 #3

기관내 통합기록관시스템의 상세검색은 아래와 같은 상세검색조건으로 구성한다.

[그림 7-8] 자료관 상세검색 #4

최근에는 높은 검색 품질 이슈 이외에도 최신 정보의 실시간 검색 요구가 많아지고 있는 추세이므로 통합검색솔루션을 도입하여 사용자가 요구하는 실시간 검색이슈에 대응하고 있다.

제2절 | RFP 작성 준비

2.1. 범위 설정

검색솔루션 도입을 위한 사업유형을 나열하면

① 검색솔루션 신규 도입 및 구축

② 기존 검색엔진의 버전 업그레이드 및 검색대상 확대

③ 기존 검색엔진을 활용한 검색대상(컬렉션) 추가에 따른 용역

④ 검색엔진의 유지보수 용역으로 구분된다.

검색과 관련하여 사업의 범위를 정의할 때는 정해진 규칙이 있는 것이 아니라 각 기관별로 사안에 맞게 설정하면 된다.

일반적으로 신규 시스템을 구축할 때 저장될 데이터를 원할하게 찾기 위하여 검색솔루션을 새로 도입하며, 기존에 도입된 검색엔진이 도입연한에 의해서 제 성능을 발휘하지 못할 때 검색엔진의 업그레이드 사업을 진행하거나 다른 검색엔진의 신규 도입을 추진한다.

기존 검색엔진을 활용한 검색대상(컬렉션) 추가 용역 사업의 경우는 신규 검색엔진을 도입하여 1차 사업을 완료한 후에 시간이 지남에 따라 검색대상시스템(포탈, GW, 게시판 등)이 다양한 형태로 진화하여 초기에 계획된 검색대상에 비해 다수의 검색대상이 추가되었을 경우 진행한다.

한편, 검색대상은 그대로 유지하더라도 홈페이지 개편, 시스템 개편 등을 통해 User Interface가 변경되는 경우에도 검색용역 사업을 진행하여 변경된 UI가 검색결과 화면에도 반영되도록 한다.

2.2. 투입 인력

검색엔진은 패키지 형태로 공급된다. 검색엔진을 우선 서버에 탑재하여 기본 테스트를 진행한 후에 검색대상 시스템과 요구사항을 확인하고 이 수준에 따라 검색엔진의 구축 및 UI 작업을 진행할 개발자를 투입하게 된다. 따라서, 검색을 적용할 시스템과 요구사항의 난이도를 먼저 살펴보고 투입인력에 대하여 구축업체와 논의하는 것이 좋다.

사업규모에 따른 기간과 적정 투입인력을 살펴보면 아래와 같다.

[표 7-4]

사업 범위	사업기간	적정 투입 M/M	권장 기술 등급
홈페이지, 게시판 등	1~2개월	2~4M/M	초상급
내부 포탈, 대외 포탈 등	3~4개월	5~7M/M	초상급
GW, 통합자료관 등 내부 시스템	3~5개월	6~8M/M	중급

※위 내용은 일반적인 내용을 기준으로 산정되었으므로 모든 기관에 동일하게 적용되지 않음. 검색대상 시스템의 복잡도와 데이터 량에 따라 사업기간 및 투입 M/M는 증가될 수 있음

새로 구축되는 서비스(또는 시스템)이나 기존 서비스에 검색을 적용하려면 DB 스키마를 파악하여 검색대상을 정확히 정의하는 것이 꼭 필요하다. DB에 저장된 데이터를 추출하는 SQL과 첨부파일 저장 경로, 권한 등의 정보를 습득하여 검색엔진에 이 정보를 전달해야 검색이 이루어진다.

이러한 작업은 검색엔진 쪽에서 임의로 파악할 수 있는 부분이 아니므로 기존 서비스(또는 시스템)을 구축한 업체나 운영/유지보수의 업체와의 협업이 반드시 필요하다.

신규 서비스인 경우에는 현재 사업을 진행하고 있는 시스템통합(SI) 사업자와의 협업을 통해 정보를 취득해야 하므로 기관-주관사-검색업체의 협조체계가 잘 갖춰져야 한다.

또 한가지 짚어야 할 내용은 검색엔진을 적용 및 구축하는 사업기간이다. 예를 들어, 전체 사업기간이 10개월이라 할지라도 검색엔진의 도입 및 구축은 사업의 후반기에 집중적으로 이뤄지는 경우가 많다.

사업 초기에는 검색엔진의 라이선스를 도입하고 요구사항 분석 등을 수행하게 되지만 실제 검색할 데이터가 존재하지 않은 경우가 많기 때문에 데이터가 생성된 후에 사업에 참여하게 된다.

따라서, 이러한 내용을 고려하여 전체 사업기간에서 검색엔진의 구축 기간을 분할하여 시스템별로 적용할 것인지 아니면 후반부에 집중 투입할 것인지를 결정하면 사업진행이 수월할 것이다.

2.3. 예산 규모 추정 방법(투입 MM 산정)

■ **고려항목**

검색솔루션 도입 사업의 예산 규모를 일괄적으로 정의하긴 어렵다. 각 공급사마다 라이선스 정책이 상이하고 가격도 다르기 때문이다.

예산을 산정할 때 가장 중요한 요소는 도입될 검색솔루션의 라이선스 수(Copy)이다. 각 기관의 검색대상 시스템 수와 향후 저장될 데이터 용량에 따라 1개의 검색엔진 라이선스부터 N개의 라이선스가 도입될 수 있으므로 사전 업무분석을 통해 검색대상에 따른 라이선스 수를 파악해야 한다.

그리고, 안정성을 위하여 시스템 이중화를 고려하고 있다면 산정된 라이선스 수에 2배를 고려해야 하고 도입비용도 역시 그만큼 증가한다.

■ **산정기준**

위에서 기술한 바와 같이 검색솔루션을 신규로 도입하는 경우에는 각 공급사마다 가격에 차이가 있으므로 기준을 정하기 어렵다. 따라서, RFP 단계 이전에 각사로부터 검색엔진 공급에 대한 사전 견적을 수령하여 사업예산 규모를 산정하는 것이 바람직하다.

라이선스 도입이 사업이 아닌 용역 사업의 경우는 기존에 도입된 검색엔진을 활용하여 사업을 진행하므로 투입인력 등급에 따른 M/M와 기술료를 고려하여 예산을 산정한다.

2.4. 선행과제

통합검색솔루션을 도입하기 위하여 사전에 수행되어야 하는 선행과제는 필요치 않다. 일반적으로 홈페이지 관련 사업, 인트라넷 관련 여러가지 사업(H/W 도입, DB 구축 등)에 묶여서 진행될 수도 있고 기 구축된 시스템에 대해서 단독으로 검색사업이 진행될 수도 있다

제3절 | 표준 RFP 요구사항

각 기관/기업에서 검색엔진을 도입할 때 요구되는 요구사항은 검색엔진 필수기능, 관리자 기능, 시스템 성능 필수사항으로 나눌 수 있는데 그 내용은 아래와 같다.

■ 검색엔진 필수기능

- 분산검색구조 지원(로드밸런싱 및 멀티쓰레드 지원)
- HWP, MS Office, PDF 등 다양한 파일 포맷에 대한 필터 제공
- 오라클, MS-SQL을 포함한 다양한 DB에 대한 수집/색인기능 제공
- 자체 형태소분석기를 통한 형태소 분석 기능
- 검색어 추천, 인기 검색어, 급등 검색어, 바로가기 기능 제공
- 검색어 자동완성 기능 제공
- 유사어, 불용어, 금지어 기능 제공
- 중요도, 관련도, 등록일 등을 포함한 다양한 조건에 의한 검색결과 정렬기능
- 결과내 검색, 카테고리별 결과내 검색 기능
- 검색필드에 대한 AND, OR, NOT 검색기능 제공
- 복수의 정렬 항목에 대하여 다중 정렬 기능 제공
- 키워드 하이라이팅 기능 제공
- 중복문서 그룹핑 기능 제공
- 문서권한 분리기능 제공
- 24시간 무정지 시스템을 위한 색인 이중화 및 복구 기능 제공
- 디스크 공간효율성 확보 및 색인 보안을 위한 색인 압축기능 제공
- 대소분서 구별 색인, 한자/한글 변환 색인, 유니코드 지원

■ 관리자기능

- 웹 기반의 관리자 환경
- 사용자사전, 동의어사전, 금칙어 등을 관리할 수 있는 사전관리기능
- 중요도, 관련도, 등록일 등을 포함한 다양한 조건에 의한 검색결과 정렬기능
- 서비스 이용 통계, 각종 사전관리, 색인 데이터 관리 기능
- 인기검색어 모니터링
- 검색서비스 기동 및 검색엔진 상태 모니터링
- 검색어 로그 관리기능

■ 시스템 성능 필수 사항

- SW 신뢰성, 안정성을 보장하는 최신 GS 인증 제품
- 행정업무용 소프트웨어 등록제품
- 다양한 OS 지원(.Windows, Linux, Solaris, AIX, HP 등)
- 다양한 개발언어 지원(JAVA, JSP, ASP, PHP, .Net 등)
- 색인 데이터 메모리 로딩 처리 기능(In Memory)
- 안정적인 유지보수를 위한 자체 개발 엔진

■ 기타

- 다국어 형태소 분석기 제공(한/중/영/일 등)
  - — 한글, 영문 외 중국어, 일본어 등으로 웹페이지를 구성하고 검색해야 하는 경우에는 다국어 형태소 분석기를 갖추는 것이 좋다. 그렇지 않으면, 다국어의 경우에는 형태소분석에 의한 색인이 아닌 TOKEN 방식 또는 띄어쓰기로 구분된 색인으로 검색 서비스를 해야한다.
- 각종 인증 제품
  - — 각 기관의 필요에 따라 GS 인증 제품, 행정업무용 소프트웨어 등록제품, 범용성 테스트 인증 제품, 조달 등록 제품 등 다양한 인증 제품을 명시한다.

제4절 | 프로젝트 점검사항

■ 검색대상 정의

통합검색의 대상이 되는 시스템은 매우 다양하다. 외부 데이터를 보면 홈페이지의 모든 데이터, 게시판, 첨부파일, 유관기관의 웹페이지를 검색대상으로 볼 수 있다. 내부 데이터를 보면 전자결재, 지식관리시스템, 이메일, 각종 커뮤니티, 게시판, 문서관리시스템, 통합자료관, 첨부파일, 기관 내부포탈 등 데이터가 저장되는 모든 시스템이 검색대상으로 정의된다.

통합검색 사업을 진행할 때 가장 먼저 해야할 것이 검색대상의 정의이다.위에서 언급된 다양한 데이터에서 어떤 시스템을 검색할 것인지에 따라 검색솔루션의 엔진 개수, 개발기간 및 비용이 산정된다.

■ 하드웨어 구성

통합검색 프로젝트는 홈페이지 개편, 내부 정보시스템 도입사업과 같은 SI사업의 일부 영역으로 진행되거나 기존에 도입된 검색엔진의 고도화(업그레이드)로 진행되는 경우가 일반적이다. SI사업에 포함되어 진행되는 경우, 검색엔진을 어느 서버에 설치/운영해야 하는지가 중요한 결정사항으로 대두된다. SI사업의 경우 WEB/WAS 서버, DB서버, 각종 분석 및 Visual Tool 서버 등 다양한 서버가 도입되는데 가장 좋은 것은 검색엔진도 별도의 서버에 탑재하여 운영하는 좋다.

홈페이지 사업의 경우 WEB/WAS 서버에 검색엔진을 함께 설치 운영하고, 내부 시스템 사업인 경우 DB 서버에 함께 탑재되어 검색엔진을 운영하는 경우가 있으나 안정성 측면에서 권장할만한 것은 아니다.

검색엔진은 수집된 대량의 데이터를 색인하여야 한다. 따라서, 색인된 데이터가 저장될 스토리지가 요구된다. 홈페이지나 조직 내 일반적인 게시판의 경우라면 검색엔진이 탑재된 서버의 로컬 디스크로도 충분하겠지만 문서관리시스템, 통합자료관시스템, 다년치의 전자결재, 다수의 유관기관 데이터를 연계하여 저장하는 등 대량의 문서를 수집/색인하고자 하면 별도의 스토리지를 구성한 후 색인데이터를 보관하여야 한다.

일반적인 경우, 한 대의 검색서버에서 최대 1TB 정도의 색인데이터를 처리할 수 있다.

■ 이중화 구성

검색시스템을 구성할 때 검색엔진의 이중화는 필수요소는 아니다. 적은 규모의 시스템 예를 들어, 일반적인 홈페이지, 게시판, 단일시스템의 검색 등은 검색엔진을 Stand-Alone으로 구성한다.

그러나, 홈페이지 중에서도 포탈형식의 홈페이지처럼 외부사용자가 많고 민원업무와 직접적으로 연결된 것은 가능한 이중화를 하여 검색서비스의 안정성을 이루는 것이 좋다.

또한, 통합자료관, 문서관리시스템처럼 대용량 데이터를 색인하여 검색하는 경우에도 검색시스템을 이중화하여 원활한 검색서비스를 제공하는 것이 바람직하다.

■ 필수 산출물

사업의 종류에 따라서 다양한 산출물이 작성되지만 원활한 운영과 향후 유지보수를 위해서 꼭 필요한 산출물은 아래와 같다.

① 검색화면 설계서

② 검색색인(컬렉션) 정의서

③ 검색시스템 구성도

④ 사용자 매뉴얼

⑤ 운영자 매뉴얼

제5절 | 프로젝트 평가

■ 성능평가

검색엔진의 성능은 주로 검색결과 처리시간으로 측정된다. 데이터가 크지 않은 경우는 일반적으로 처리시간이 1초 이내로 측정되어야만 정상적으로 구축한 프로젝트라 평가할 수 있다.

통합자료관이나 수 년치의 전자결재 혹은 문서관리시스템을 검색하여 색인데이터가 1TB 이상 되는 프로젝트는 현실적으로 약 3초이내의 처리시간으로 측정하는 것이 좋다. 검색대상 컬렉션을 나누거나 검색엔진을 여러 대의 서버로 분산사여 검색처리 시간을 단축시키는 기법을 사용할 수 있겠지만 테라바이트 이상의 검색시스템은 3초 이내의 처리시간으로 평가하는 현실적이다.

■ 기능평가

검색기능은 한두가지가 아닌 다양한 기능이 존재한다. 각 기관의 검색대상 특성에 맞는 검색기능을 요구하게 되므로 요구사항 정의시 기관만의 검색기능을 명확히 전달하여야 한다.

기능평가는 화면상에 구현되는 기능을 갖고 평가하게 되므로 RFP에 검색결과의 화면, 검색결과내 각종 정렬 및 상세검색 조건 등을 가능한 상세히 언급하는 것이 좋다.

참고할만한 검색대상 기능으로는

① 검색대상의 통합검색

② 검색어 자동완성

③ 각종 정렬(날짜, 작성자, 제목, 필드별 이름)의 오름차순/내림차순

④ 인기검색어

⑤ 검색어 하이라이팅 기능

⑥ 관리자기능

등이 있다.

■ 안정성 평가

검색엔진은 다수의 사용자가 보낸 동시에 보낸 질의에 대해서 정상적으로 검색결과를 내보아야 한다. 검색시스템을 구축한 후에 부하 테스트 솔루션을 통해 검색엔진의 부하 테스트를 진행하여 검색엔진의 안정성을 평가토록 한다.

내부시스템의 경우에는 조직내 내부 사용자를 대상으로 하므로 부하테스트가 필수요소는 아니겠지만 홈페이지나 외부 포탈의 경우에는 다수의 외부 사용자가 동시에 검색시스템을 사용할 수 있으므로 부하테스트를 거쳐서 안정성을 검증하는 것이 좋다.