KiSTi 지식리포트

2013. 7. (No.35)

「정보서비스 공공기관 검색엔진 활용 만족도 조사」

이태석 · 신수미 · 유수현 · 정용일 · 이은정

KiSTi www.kisti.re.kr 한국과학기술정보연구원 Korea Institute of Science and Technology Information

목 차

《 요 약 》

□ 정보서비스의 필수요소인 검색엔진은 검색서비스 품질에 직접적인 영향을 주고 있으나, 검색엔진 도입 및 활용 경험에 대한 공유부족으로 검색서비스 개발과 운영 단계에서 많은 시행착오를 겪게 됨. 따라서 이번 공공 및 학술 분야의 검색서비스에 대한 담당자 면접조사를 통해 현장에서 느끼는 어려움과 만족도를 조사하여 정리함.

□ 검색서비스를 구축하기 위한 검색엔진의 필요성, 특성, 기능, 랭킹 기준 예시 등을 통해 검색엔진의 활용법을 이해하고, 최신 검색기술 발전 단계를 바탕으로 하여 새로운 검색엔진을 평가하고 선정 기준이 되는 주요 지표(재현율, 정확률)를 제시함.

□ 설문 조사를 통해 검색엔진을 활용하고 있는 기관의 종합 만족도와 세부적인 만족도와 함께 담당자 인터뷰에서 나타난 사실을 정리하고 사용하고 있는 검색엔진과 관련한 최근 이슈와 고려사항을 정리함.
- 검색엔진에 대한 기관별 종합 만족도를 살펴보면 KERIS(S3)가 80.0점으로 가장 높은 반면, KISTI(S1)는 46.7점으로 가장 낮은 수준임.
- 단계별 만족도를 살펴보면 개발과 운영 단계 모두 KERIS(S3)가 가장 높은 반면, 개발에서는 KISTI(S1), 운영에서는 KISTI(S1), KIPI(K2)가 가장 낮은 수준임.

□ 검색엔진 활용 고려사항으로 검색엔진 처리용량(성능, QPS)을 측정하고 적정한 용량을 산정하는 방법을 보임으로써 향후, 검색엔진을 교체하거나 기존 운영 성능을 향상시키기 위한 기준을 제시함.

□ 검색엔진 기능 점검사항으로 “[별첨-2] 검색엔진별 기능 비교” 와 같이 검색엔진의 기능을 “서버 구성과 색인처리”, “검색처리 기능”, “관리기능 및 기타” 부문으로 크게 나누어 검색엔진별 특성을 비교함.

※ 검색엔진 제품의 실제 명칭은 제조사 보호를 위해 S1~S4로 표시함.

1. 조사 개요

□ 조사 목적

- 공공 및 학술 분야의 정보서비스에서 활용되고 있는 검색엔진에 대한 개발과 운영단계의 만족도를 알아 봄.
- 검색엔진을 교체할 경우, 검색엔진을 활용하고 있는 기관들의 경험을 통해 적합한 제품을 결정할 수 있는 정보를 제공하는 것을 주요 목적으로 함.

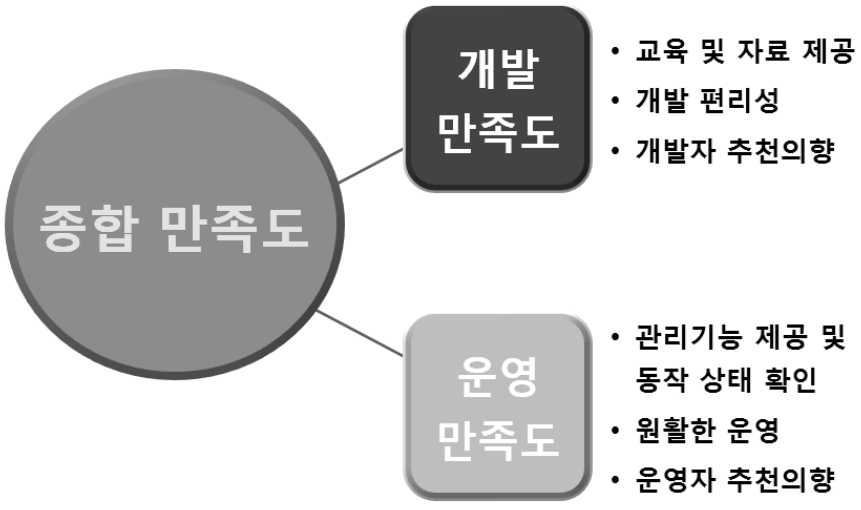
□ 조사 설계

- 조사 대상 : 국립중앙도서관, 국회도서관, 한국과학기술정보연구원(KISTI), 한국교육학술정보원(KERIS), 한국특허정보원(KIPI), LG상남도서관 등 총 6개 기관 (가나다 순서)
- 선정 방식 : 학술 및 연구 정보를 제공하는 기관 선정
- 조사 방법 : 기관 방문에 의한 개발 및 운영 담당자 일대일 면접조사
- 조사 기간 : 2013년 3월 27일 ~ 2013년 4월 26일

□ 만족도 산출

- 종합 만족도 : 개발단계와 운영단계를 종합한 만족 수준
- 개발 만족도 : 개발단계 세부 항목을 종합한 만족 수준
- 운영 만족도 : 운영단계 세부 항목을 종합한 만족 수준

〈그림 1〉 만족도 산출 체계



2. 검색엔진의 이해

1) 필요성 및 특성

□ 검색서비스에서 검색엔진의 필요성 [1][2]

- 대용량 자료에 대해 초고속 응답이 필요하거나, 전문 검색이 필요한 경우에 검색 엔진(역색인 처리)을 사용함.

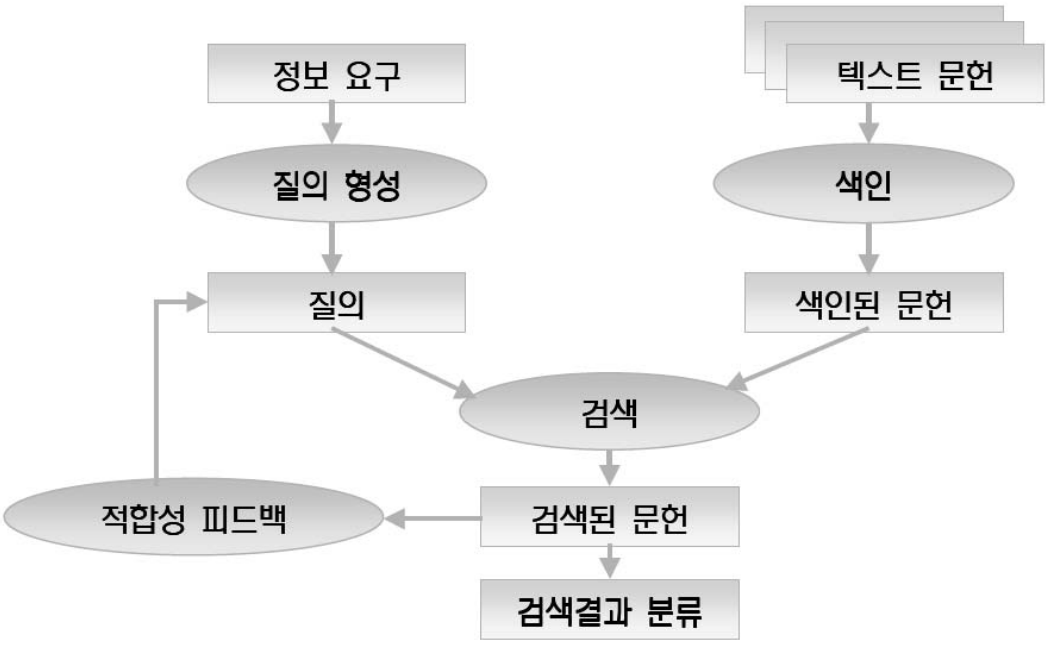
※ 작은 규모의 검색은 색인기능만 있으면 빠른 검색이 가능함.

※ 역색인 : 문서에서 검색이 될 만한 단어들의 존재여부와 위치를 미리 추출하여 빠른 탐색 자료구조를 구축(색인)한 후, 키워드 중심으로 빠르게 문서를 찾아가도록 재조정한 자료구조

□ 검색엔진 기능과 활용

- 질의식 처리, 검색, 적합성 피드백, 색인 기능으로 구분됨.
- 색인 처리기(한글 형태소분석기, Stemming, Tokenizer 등)와 질의 처리기가 일치되어야 높은 검색효율을 보여줌.

〈그림 2〉 검색엔진 동작

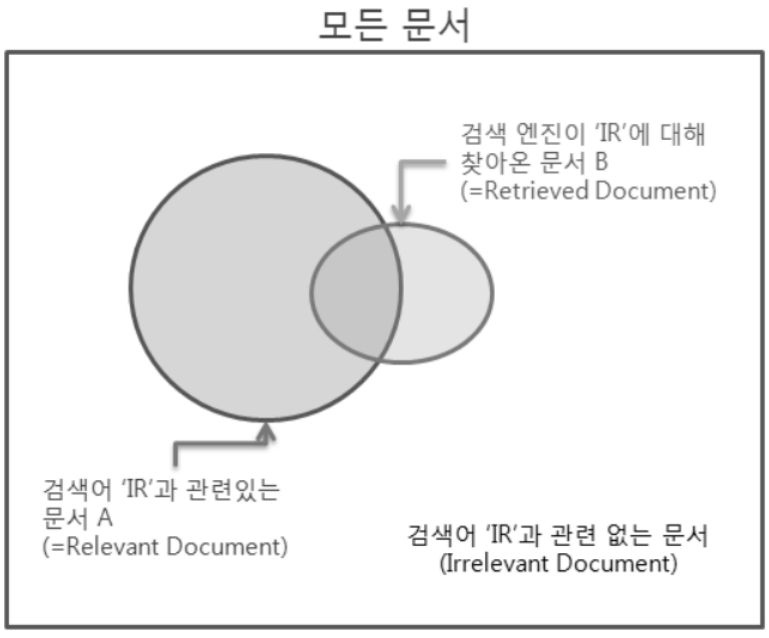


- 색인기 입력자료 = 문서의 텍스트, 제목, 키워드, Tag 등
- 색인기 : 색인 정보를 역색인하여 DB로 구축함.
  (색인되지 않은 정보는 검색이 되지 않음)
- 검색 결과를 분류하고 처리하여 검색서비스 개발에 활용함.

2) 성능 평가

□ 검색을 위한 키워드, 그 키워드에 관련된 문서(집합 A), 검색 엔진이 찾아온 문서 집합 B, 여기에는 관련 있는 문서와 관련 없는 문서가 섞여 있을 것임. [1][2]

〈그림 3〉 검색결과 평가 다이어그램



□ 재현율(Recall) = |A∩B|/|A|

- 재현율은 존재하는 전체 정답(A) 중 찾은 것(A∩B)의 비율
- 재현율 중심 색인 방법은 '엔그램(N-GRAM)'임. 검색할 문서수가 많지 않을 때 보통 두 글자를 패턴으로 하는 바이그램(Bi-GRAM)을 주로 사용함. 예) 검색어로 '인천국제공항'을 입력하면 '인천', '천국', '국제', '제공', '공항' 등 두 글자로 된 모든 문장을 찾는 방식임.
- 엔그램은 조사나 띄어쓰기 문제를 해결해 주는 강력한 기능을 발휘하지

만, 문서의 양이 많으면 불필요하게 많은 검색 결과를 보여준다는 단점이 발생함. 원시적인 방식이지만, 검색 재현율(recall)이 좋음.

□ 정확률(Precision) = |A∩B|/|B|
- 정확률은 시스템이 찾아온 것(B) 중 정답인 것(A∩B)의 비율
- 정확률 중심으로 처리하는 색인방식은 자연어 처리에 기반을 둔 형태소 분석기를 사용함. 과도한 색인/검색 결과에 대한 문제를 해결할 수 있음. 예) '인천국제공항'을 입력하면, '인천', '국제', '공항'이란 단어만 있으면 관련 정보를 찾아주게 됨.
- 형태소 분석기의 성능과 사전에 의해 정확률이 달라짐.

□ 재현율과 정확률은 적합한 문서집합인 A를 결정하기 어렵기 때문에 제한된 문서 자료에 대해서만 평가가 가능함. 대부분의 검색엔진은 색인기의 성능향상과 사전 구축 등을 통해 운영하면서 점진적으로 재현율과 정확률을 향상시킬 수 있는 유연한 구조로 되어 있음.

□ 그 외 색인속도(DPS)와 검색속도(QPS)가 있어 일반적으로 검색엔진의 BMT(Benchmarking Test)에 주로 사용함.
- 색인속도(DPS) : 초당 색인 문서수(Document Per Second)
- 검색속도(QPS) : 초당 쿼리 처리수(Query Per Second)

3) 검색결과의 적합성(랭킹)

□ 검색결과의 적합성은 내용 아래와 같은 정보검색 핵심 이론에 따라 다양하고 복합적으로 처리되고 있음. [1][2]
- TF/IDF : 내용상 중요성에 대해 단어의 출현빈도(Term Frequency)가 높고 소수의 문서에서 발견(Inverse Document Frequency)된다면, 해당 단어에 대한 정보성이 높음.
- Vector space model : 내용상 중요성에 대해 두 정보(문서)간의 유사성은 함께 나타나는 단어가 많을수록 높음.
- 관심(Attention)의 정도 반영 : 집단지성(Collective Intelligence)을 통해 관심이 높을수록 높음.

〈표 1〉 관심 정도에 따른 중요성 고려사항 [2]

고려사항	설 명
시간축의 반영	· 최근에 주목받은 정보의 가치가 높음.
여론의 반영	· 여러 사람이 지목(Link, Click)한 정보의 가치가 높음.
중복성 반영	· 여러 사람이 갈무리한 정보의 가치가 높음.

□ TF/IDF 중요성 판별에 의한 유사도 랭킹은 현실적으로 일부 사용
- TF/IDF 만으로 중요성을 판별하는 데 한계가 있음. 추가적인 인접도, 가중치 중심으로 보완하여 사용함.
- 랭킹 외에 날짜별, 가격별, 인기도별 Sorting을 함께 씀.

□ 검색 단어 또는 문장을 포함하고 있는 문서 랭킹 기준 (예시)
- 1순위 : 제목에서 입력형태 단어가 인접해서 나타나는 것
- 2순위 : 제목에서 언어처리 결과 원형 단어가 인접해서 나타나는 것
- 3순위 : 본문에서 입력형태 단어가 인접해서 나타나는 것
- 4순위 : 본문에서 언어처리 결과 원형 단어가 인접해서 나타나는 것
- 5순위 : 제목 또는 본문에서 단어의 거리가 가까운 것에서 먼 것까지 가까운 순서대로 표시함, 인접도가 같으면 단어 출현 빈도가 높은 것을 우선 표시, 인접도와 출현빈도가 같으면 필드 가중치에 따라 제목우선으로 표시함.

  ※ 인접도, 빈도 모두 일치할 경우 최신 자료, 제목 정렬순서로 표시함.

4) 검색의 발전단계

□ (1세대) 동일한 구조/형태 문서에 대한 검색 [2]
- 일종의 도서 검색 프로그램과 유사함.
- TF/IDF : 어떠한 단어에 대해서 그 단어가 여러번 나온 문서를 1순위로 결정하는 방식임.
- 문서의 구조를 알고 있으므로 제목에서 찾기, 두 단어가 가깝게 있는 문서 찾기 등이 가능함.
- 콘텐츠 분석이 중요기술임.

□ (2세대) 웹이라는 인터넷 정보공간에 카탈로그 같은 페이지 검색 [2]
- TF/IDF + PageRank : Web의 Link 구조를 분석해서 해당 문서의 중요도를 계산함.
- 다양한 문서 형식에 대해 문서 구조를 단순화시킴.
- 많은 문서를 수집하는 능력과 수집된 문서의 링크를 분석하는 것이 주요기술임.
- 콘텐츠 분석과 링크 분석이 중요기술임.

□ (3세대) Opinion 콘텐츠, 그리고 Communication 콘텐츠 [2]
- TF/IDF + PageRank + Human Factor
- 문서의 구조를 어느 정도 알고 있고, 문서에 링크가 존재하고, 문서에 이용 지표가 있음.
- 사용자의 행태 정보를 이용하여 추천시스템의 성능을 높임.
- 콘텐츠 분석, 링크 분석, 행태 분석이 중요기술임.

□ 향후 검색은 Search 2.0의 시대로 발전하면서 사용자 행태 정보를 이용하고 언어 분석을 심화하는 방향으로 발전하고 있음.

3. 면접조사 결과

1) 조사 기관 특성

- 조사한 기관은 아래의 표와 같은 정보 검색서비스를 운영하고 있음.

〈표 2〉 면접 조사 기관의 서비스 및 주요 콘텐츠

기관 구분	설 명
국립중앙도서관	· 기관명 : 국립중앙도서관 · 서비스명 : 디브러리포털(dbrary) · 주요콘텐츠 : 단행자료, 연속간행물, 비도서자료, 온라인자료 등
국회도서관	· 기관명 : 국회도서관 · 서비스명 : 국회전자도서관 · 주요콘텐츠 : 도서자료, 학위논문, 학술정보 등
KISTI	· 기관명 : 한국과학기술정보연구원 · 서비스명 : 국가과학기술정보센터(NDSL) · 주요콘텐츠 : 학술정보(논문), 특허, 연구보고서, 동향, 표준 등
KERIS	· 기관명 : 한국교육학술정보원 · 서비스명 : 학술연구정보서비스(RISS) · 주요콘텐츠 : 해외 학술DB, 학위논문, 단행본, 공개강의 등
KIPI	· 기관명 : 한국특허정보원 · 서비스명 : 특허정보검색서비스(KIPRIS) · 주요콘텐츠 : 특허실용신안, 디자인, 상표 등
LG상남도서관	· 기관명 : LG상남도서관 · 서비스명 : LG ELIT · 주요콘텐츠 : 학술정보, 교육강의정보, 기술동향정보 등

※ 나열순서는 기관명 가나다 순서임.

- 콘텐츠 보유건수 1억건 이상 정보서비스를 운영하는 곳은 KISTI와 KIPI였으며, 일일 이용자가 가장 많은 사이트는 KIPI가 운영하는 KIPRIS임.
- 가장 많은 신규자료를 연간 구축하고 있는 기관은 KISTI임(1,000만 건).

〈표 3〉 기관별 검색서버 구성의 특징과 개발/운영 현황

		기관별					
		국립중앙 도서관	국회도서관	KISTI	KERIS	KIPI	LG상남 도서관
기본사항	검색엔진	S2	S4	S1	S3	K2	S4
	서버수량	4대	5대	10대	3대	6대	2대
	운영체제 (OS)	SUN	IBM	리눅스	IBM	IBM	SUN
	기본사양	8CPU 32GB	2CPU 48GB	2CPU 128GB	8CPU 32GB	16CPU 16GB	2CPU 32GB
	데이터 건수	400만건	750만건	1억건	5,000만건	1억건	1,000만건
	일일 이용자수	1만명 이내	5만~10만 명	1만~5만 명	1만명 이내	10만명 이상	1만명 이내
	연간 신규자료	100만건	100만건	1,000만건	100만건	100만건	100만건
개발단계	구축 경험	3년~5년	10년 이상	3년~5년	10년 이상	10년 이상	10년 이상
	구축 수량	3개	3개	4개	4개	2개	2개
운영단계	교체계획	5년 이상	4~5년	2~3년	5년 이상	2~3년	5년 이상
	운영인력 (OP)	3명	6명	3명	5명	10명	3명

※ 검색엔진 제품의 실제 명칭은 제조사 보호를 위해 S1~S4로 표시함.

- KISTI와 KIPI의 서버사양을 살펴보면 KISTI는 PC서버로 운영하고 KIPI는 IBM 대형서버로 운영하고 있으며, 검색 성능을 높이기 위해 메모리와 디스크에 자원을 적극적으로 투입하고 있음.
- 검색엔진은 국산(S2, S3)과 외산(S1, S4, K2)을 쓰고 있음.
- 개발단계에서 개발자의 검색엔진에 대한 경험은 10년 이상의 숙련자로서 검색엔진 제조사 또는 기술지원 전문 업체의 직원이 직접 투입되는 경우가 많음. 서비스의 안정성을 위해 경험이 많은 우수 개발자 투입이 매우 중요함.
- 담당자 인터뷰 결과 운영 인력측면에서 검색서비스 전담조직을 가지고 있고 KIPI가 가장 우수한 검색서비스 대응 역량을 가졌음. 또한, 특허정보 오픈서비스인 KIPRIS 플러스를 병행 운영하고 있었으며, 많은 시행

착오를 겪으면서 인력, 조직, 장비를 모두 분리함. (KIPRIS는 대전 정부통합전산센터, KIPRIS 플러스는 서울청사에서 운영함.)

- 조사 대상 기관에서는 대부분 장비와 운영체제의 안정성을 위해 IBM과 SUN장비를 사용하고 있음.

〈표 4〉 검색엔진 제품과 제조사 현황

설 명	비 고
· 검색엔진 : 코난 독크루저(DOCRUZER) · 제조사 : 코난테크놀로지 · 기술지원 : 제조사	국산
· 검색엔진 : FAST ESP · 제조사 : 마이크로소프트 · 기술지원 : 예쓰월드 등	외산
· 검색엔진 : IDOL(Intelligent Data Operating Layer) · 제조사 : HP · 기술지원 : 쓰리소프트플러스, 쓰리웨어	외산
· 검색엔진 : K2 엔터프라이즈 · 제조사 : VERITY · 기술지원 : 구 쓰리소프트	외산
· 검색엔진 : Search Formula-1 · 제조사 : 와이즈넛 · 기술지원 : 제조사	국산

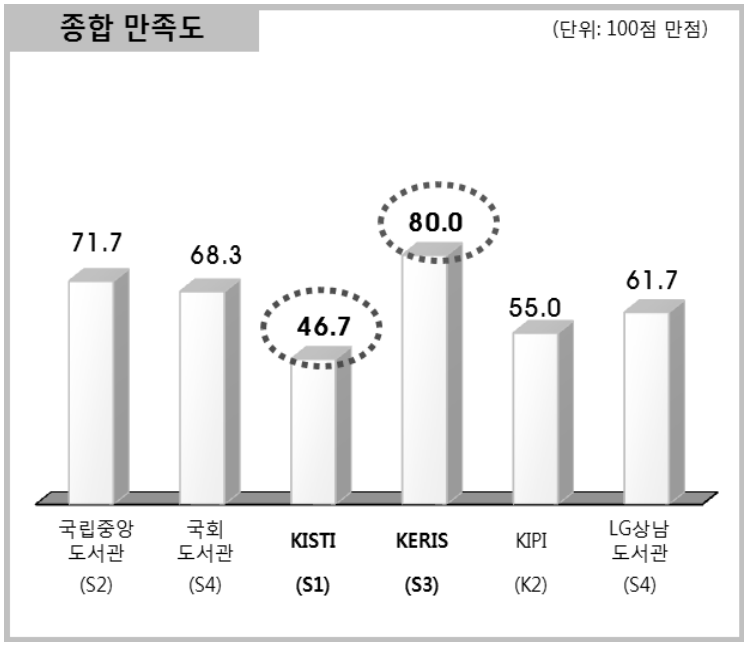
※ 나열순서는 제품명 알파벳 순서임.

2) 검색엔진에 대한 종합 만족도

- 검색엔진에 대한 기관별 종합 만족도를 100점 만점으로 조사한 결과 KERIS(S3)가 80.0점으로 가장 높은 반면, KISTI(S1)는 46.7점으로 가장 낮은 수준임.
- 단계별 만족도를 살펴보면 개발과 운영 단계 모두 KERIS(S3)가 가장 높은 반면, 개발에서는 KISTI(S1), 운영에서는 KISTI(S1), KIPI(K2)가 가장 낮은 수준임.
- 검색엔진에 대한 만족도가 낮은 KISTI와 KIPI에서 2~3년 내에 검색엔진을 교체할 계획을 가지고 있음.

- 만족도에 영향을 주는 주요 요인은 기술지원임. 검색 성능과 품질개선을 위해 최신 장비와 기술 변화에 대한 요구가 지속적으로 발생하기 때문임.
- KIPI에서 오래된 K2 검색엔진을 쓰는 이유는 최신 국산 및 외산 검색엔진이 특허정보 데이터 처리에 적합한 모듈이 없고, 관련 기술지원이 되지 않기 때문임. 일반 검색엔진은 특허자료 처리 전문 분야에 필요한 다양한 유틸리티가 없어 전문적인 기술지원이 필요함.

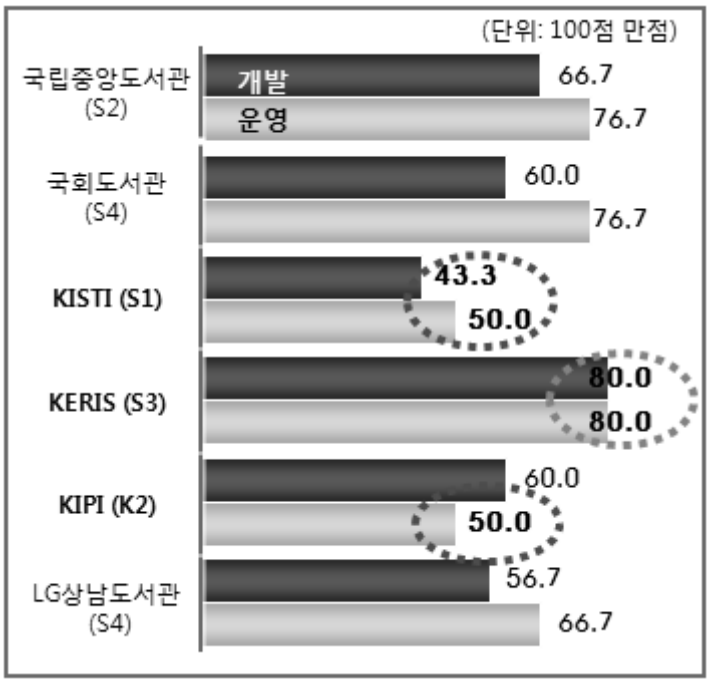
〈그림 4〉 검색엔진에 대한 개발/운영 종합만족도



- 운영단계에서 부족한 부분이 지속적으로 개선되어 개발단계보다 만족도가 높으나, KIPI의 경우는 단종된 K2를 2000년부터 지금까지 쓰면서 운영 만족도가 많이 떨어진 상태임.
- 2010년 S3로 교체한 KERIS는 다른 검색엔진은 쓰는 기관보다 80점으로 만족도가 높으며, 개발과 운영 만족도가 모두 높음.
- KERIS의 경우 검색엔진 제조사의 직원이 투입되어 전략적 성공사례로 집중 지원하여 RISS 서비스를 구축하였으며, 구축 후에 발생하는 문제에 대해서도 밀착지원을 하면서 운영담당자의 만족도가 높은 것으로 면접결과 확인됨.
- 국립중앙도서관에서는 운영 업체가 검색엔진과 관련하여 기술적으로 큰 어려움을 겪었으나, 이를 극복하기 위해 S2 검색엔진을 개발한 전문 엔

지니어를 참여시키면서 개선사항을 효과적으로 처리하게 되었고, 최근 안정적인 운영을 하여 운영 만족도가 높아지고 있음.

〈그림 5〉 개발 및 운영 단계 만족도



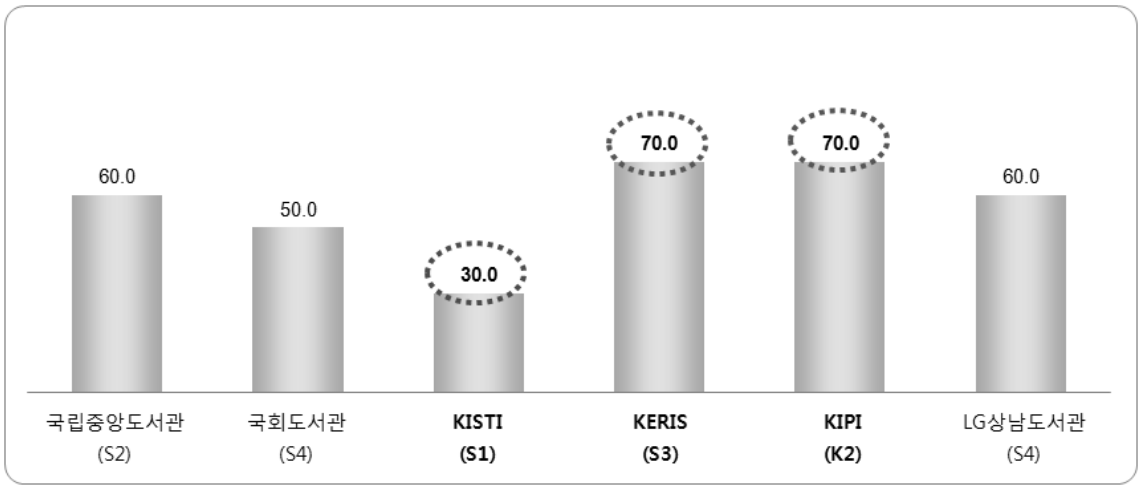
- 국회도서관의 경우 검색엔진 도입에서 충분한 BMT 과정을 거쳐 검색엔진에 대한 신뢰가 높은 편이며, 운영에서도 전문 업체에서 밀착 기술지원을 통해 지속적인 개선이 가능함에 따라 운영 만족도가 개발 만족도보다 높아짐.
- LG상남도서관은 국내에서 S4를 초기에 도입한 경우로 초기에 한글처리 등 문제로 개발과 운영 어려움이 있었으나, 검색엔진 오류(장애, 중단)가 없을 뿐만 아니라 영문 자료에서 만족한 결과를 보여주어 현재는 상당히 만족하고 있음. (외산 검색엔진의 체계적인 버전관리와 안정성 부분은 국산 제품이 참고할 만함.)

3) 개발단계 만족도

□ 교육 및 자료 제공

- 검색엔진의 충분한 교육 및 자료 제공에 대한 만족도를 살펴보면 KIPI(K2)과 KERIS(S3)가 70.0점으로 가장 높고, 다음으로 국립중앙도서관(S2)과 LG상남도서관(S4), 국회도서관(S4) 순이며 KISTI(S1)는 30.0점으로 가장 낮은 수준임.

〈그림 6〉 교육 및 자료제공

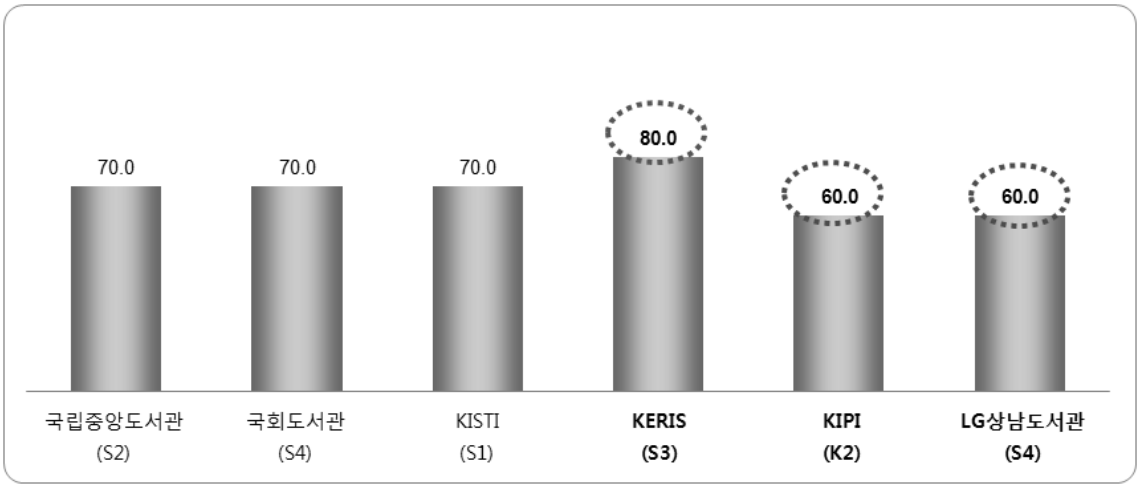


- 공식적인 검색교육과 매뉴얼뿐만 아니라, 검색엔진에 대해 개발자 스스로 작성하고 공유한 자료와 온라인 커뮤니티 활성화 여부가 가장 큰 영향을 주는 것으로 인터뷰에서 나타남.
- K2의 경우 2000년 초에 좋은 검색엔진으로 인정받아 조사 기관 모두 K2 검색엔진을 사용한 경험을 가지고 있었음.
- 검색엔진 S3 제조사는 국내 및 해외에까지 검색엔진을 알리기 위한 행사와 마케팅을 통해 많이 알려지고 있음.
- 상대적으로 외산 검색엔진 S1과 S4는 성능과 관계없이 만족도가 낮음. 그 원인은 우리나라에서 쉽게 볼 수 있는 기술자료의 보급 수준이 낮고 전문 기술지원업체를 통해야 하는 번거로움이 있음.
- 현재는 하나의 검색엔진이 널리 쓰이는 것이 아니라, 검색서비스 특성에 잘 맞추어 주는 검색엔진이 많은 관심을 받고 있는 추세임.

□ 개발 편리성

- 검색엔진의 개발 편리성에 대한 만족도를 살펴보면 KERIS(S3)가 80.0점으로 가장 높고, KISTI(S1)와 국립중앙도서관(S2), 국회도서관(S4)이 70.0점으로 다음 순이며 KIPI(K2)와 LG상남도서관(S4)은 60.0점으로 가장 낮은 수준임.

〈그림 7〉 개발 편리성

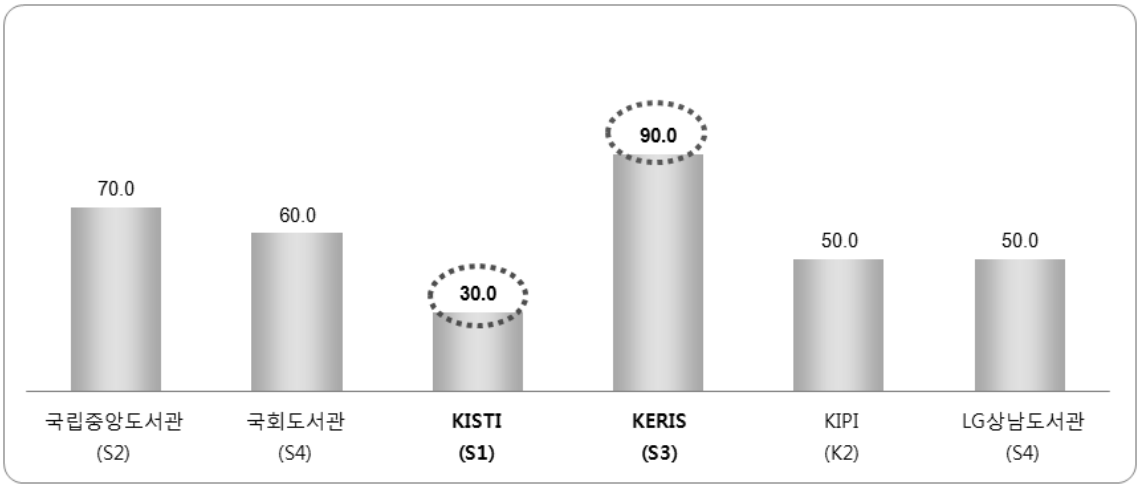


- 교육 및 자료제공 만족도와 다르게 개발 편리성은 모두 비슷한 수준을 보여줌.
- 실제 개발에 필요한 몇 가지 기능을 중심(샘플 소스 등)으로 개발을 수행하기 때문에 개발편리성에서 큰 차이가 나지 않는 것으로 판단됨. 특히 개발자들은 복잡하고 어려운 기능은 검증되지 않았기 때문에 결과물의 신뢰성이나 일정을 고려하여 사용하지 않음.
- 예외적으로 검색엔진의 고급 기능을 사용하는 경우는 서비스의 핵심요구사항으로 응용프로그램으로 처리할 수 없는 검색엔진의 고유한 유사도 처리, 요약, 랭킹 등의 기능에 대한 것임.

□ 개발자 추천의향

- 검색엔진에 대한 개발 추천의향을 살펴보면 KERIS(S3)가 90.0점으로 가장 높고, 다음으로 국립중앙도서관(S2), 국회도서관(S4), KIPI(K2)와 LG상남도서관(S4) 순이며 KISTI(S1)는 30.0점으로 가장 낮은 수준임.

〈그림 8〉 개발자 추천 의향



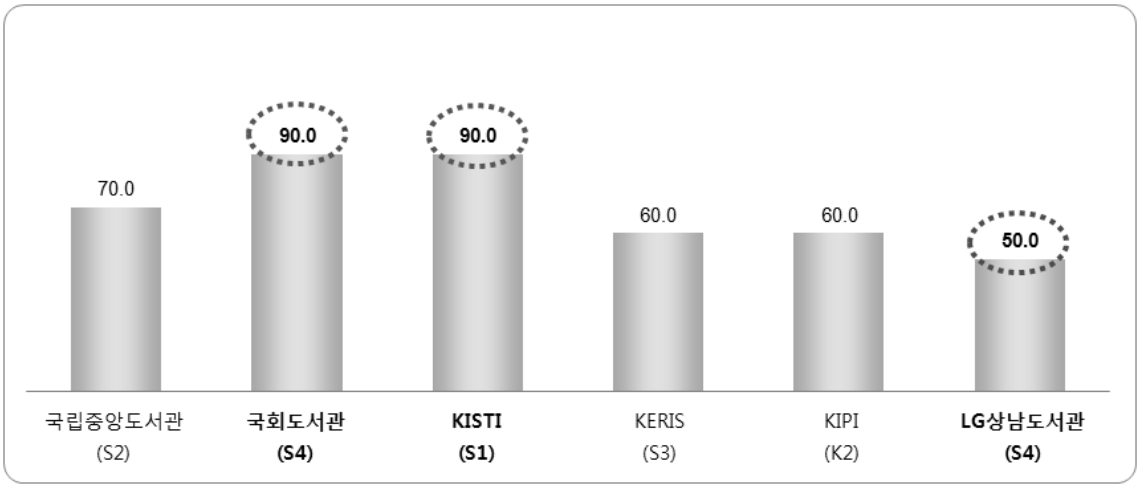
- 국산 검색엔진의 추천의향이 외산 검색엔진보다 월등히 높은 것은 국산 검색엔진의 성능과 기술 수준이 올라가 좋은 인식이 많이 알려졌기 때문임.
- 또한, 외산 검색엔진의 경우 한글 처리는 국내의 기술력에 의존하고 있어서 개선 사항 처리가 복잡하고 오래 걸림.
- 면접 조사에서 외산 검색엔진을 추천하는 경우는 검색엔진의 완성도가 높으며, 체계적인 관리가 계속 유지되기 때문임. 이러한 요소는 장기적으로 국산 검색엔진이 갖추어야 할 요소임.
- 국산 검색엔진의 실패 사례로 고객의 다양한 요구사항에 맞추어 검색엔진의 핵심 기능을 변경함에 따라 관련 기술자의 의존성이 높아지고 체계적인 버전관리를 못하여 어려움을 겪는 경우가 있었음.

4) 운영단계 만족도

□ 관리기능 제공 및 동작 상태 확인

- 검색엔진의 관리기능 제공 및 동작 상태 확인에 대한 만족도를 살펴보면 KISTI(S1)와 국회도서관(S4)이 90.0점으로 가장 높고, 다음으로 국립중앙도서관(S2), KIPI(K2)와 KERIS(S3) 순이며 LG상남도서관(S4)이 50.0점으로 가장 낮은 수준임.

〈그림 9〉 관리기능 제공 및 동작 상태 확인

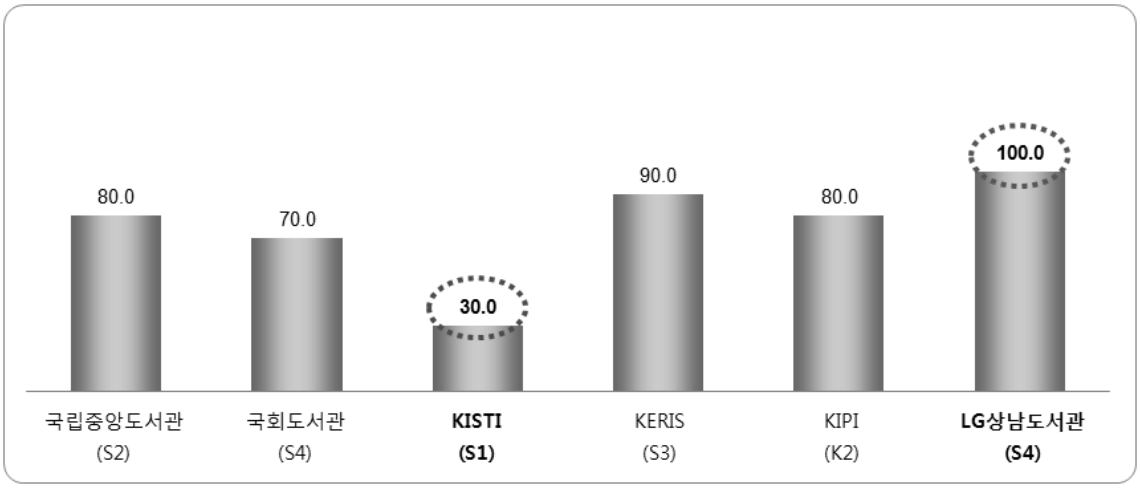


- 운영단계에서 관리기능 부분에 대한 만족도는 외산 검색엔진 KISTI(S1)와 국회도서관(S4)이 높은 만족도를 보임. (LG상남도서관 S4는 국회도서관보다 낮은 초기 버전에 대한 첫 도입사례로 시행착오가 있었음)
- 국산 검색엔진의 기술수준은 높으나, 운영단계에 필요한 기능에 대해 충분히 검증되지 않았다는 것을 알 수 있음.

□ 원활한 운영

- 검색엔진의 원활한 운영에 대한 만족도를 살펴보면 LG상남도서관(S4)이 100점으로 가장 높고, 다음으로 KERIS(S3), 국립중앙도서관(S2)과 KIPI(K2), 국회도서관(S4) 순이며 KISTI(S1)는 30.0점으로 가장 낮은 수준임.

〈그림 10〉 원활한 운영

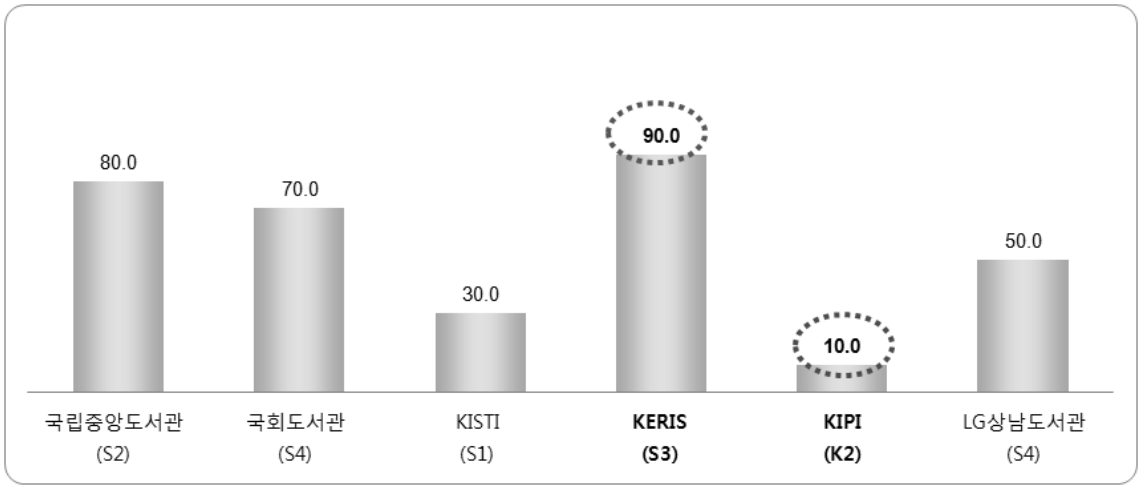


- LG상남도서관은 검색엔진 장애가 거의 없이 운영이 잘 되고 있음.
- KISTI는 높은 사양의 장비를 쓰고 있지만, S1 검색엔진이 리눅스 서버 환경에서 충분한 성능을 발휘하지 못함.
- 각 기관에서 검색서비스를 잘 운영하기 위한 장비가 필요하나, 적정한 사양 및 구성을 산정하는 데에 어려움이 있음.

□ 운영자 추천의향

- 검색엔진에 대한 추천의향을 살펴보면 KERIS(S3)가 90.0점으로 가장 높고, 다음으로 국립중앙도서관(S2), 국회도서관(S4), LG상남도서관(S4), KISTI(S1) 순이며 KIPI(K2)는 10.0점으로 가장 낮은 수준임.

〈그림 11〉 운영자 추천의향



- K2의 경우 단종된 제품으로 추천의 의미가 없음.
- KERIS에서 사용 중인 S3에 대한 추천이 높은 이유는 제조사에서 다각적이고 신속한 지원을 충분하게 해주기 때문임.
- 외산 검색엔진 S1과 S4가 추천의향이 낮은 이유는 전문 기술지원 업체를 두고 있지만 국내 제조사처럼 밀착지원이 되지 않기 때문임.

4. 검색엔진 활용 고려사항

1) 검색엔진 성능 요건

□ 검색서비스 처리용량(성능, QPS) 계산 [3]

- 검색엔진 성능을 측정할 때에는 단위 시간 동안 처리 가능한 질의수를 측정함.
- 10,000명의 사용자가 10분 동안 3회 검색을 하는 경우 50회/초 처리용량이 필요함.

〈표 5〉 처리용량(성능, 속도) 계산 예

처리시간 = 10분 = 600초 질의수 = 10,000명 * 3회/명 = 30,000회 처리용량 = 30,000회 / 600초 = 50회 / 초

- 최대 처리용량 계산 방법
  : 하루 중 이용자가 주로 이용하는 시간이 특정 시간대에 몰리는 것이 일반적이고, 피크타임의 처리용량과 전체 평균적인 처리용량의 비율이 대략 2~3배 높음. 경우에 따라서 하루 중 4~5시간 정도의 업무시간에 접속이 집중되면, 피크타임의 처리용량이 하루 평균의 4~5배 정도를 처리되어야 함.

〈표 6〉 최대 처리용량 계산

최대 처리용량 = 2~4 * 일평균처리용량(건/초)

- 일평균 10만 건의 페이지뷰가 발생하는 사이트의 평균적 처리용량은 초당 1.16건이고, 최대 처리용량은 4.64건임.

〈표 7〉 평균 처리용량

10만건 / 1일 = 10만건 / 86,400초 = 1.16건 / 초

〈표 8〉 최대 처리용량 계산

최대 처리용량	= 평균 처리용량 * 4 = (1.16건 / 초) * 4 = 4.64건 / 초

- 최대 처리용량으로부터 일평균 처리용량 계산
  : 검색시스템이 최대 초당 10건 가량의 검색질의를 처리하는 처리용량(성능)이면, 집중적인 업무시간에 검색질의가 몰리는 것을 감안하여 피크타임의 처리용량을 평균의 4배로 잡아 일평균 처리용량을 계산함.

〈표 9〉 최대 처리용량으로부터 일평균 처리용량 계산

일평균 처리용량	= 1 / 4 * 최대 처리용량 = 1 / 4 * (10건 / 초) = 2.5건 / 초 = 21만6천건 / 1일

: 일평균 약 21만건의 검색질의를 무난히 처리할 수 있음.

□ BMT 테스트 환경 및 성능 기준
- 검색엔진을 선정하기위한 테스트는 동일한 사양의 장비를 사용함.
- 테스트 대상 검색엔진이 모두 동작할 수 있는 사양과 테스트할 색인 사이즈를 고려하여 장비를 결정함.

〈표 10〉 검색엔진 BMT 평가 기준(예)

구 분	설 명	비 고
색인 Size	· 논문 6,000만 건	서비스 자료를 대상으로 함
서버 사양	· CPU : 2개 (4core * 2) 2.6Ghz · 메모리: 32GB · 디스크 : 1TB (DAS)	
평가 기준	· 최대검색 성능 : 100QPS 5분 이상, 3초 이내 · 최대색인 성능 : 100DPS 6,000만건 / 주	NDSL 서비스 기준

2) 검색엔진 기능 점검사항

□ “[별첨-2] 검색엔진별 기능 비교” 와 같이 검색엔진의 기능을 “서버 구성과 색인처리”, “검색처리 기능”, “관리기능 및 기타” 부문으로 크게 나누어 검색엔진별 특성을 비교함.

□ 검색 서비스의 목적과 필요에 따라 요구하는 기능을 제시하고 별도 기능 테스트를 실시하는 것을 권장함.

※ “[별첨-2] 검색엔진 기능 부문별 기능 비교” 자료는 제조사와 합의되지 않은 내용이며, NDSL 정보검색서비스를 기준으로 필요한 기능을 중심으로 작성된 것임.

5. 결론

□ 본 조사를 통해 공공 및 학술 분야의 공공기관이 사용하고 있는 검색엔진과 검색서비스 개발과 운영 상태를 살펴보았으며, 검색엔진의 활용과 도입을 위해 필요한 사항을 참고할 수 있는 자료를 제시하였음.

- 기술지원에 대한 만족도는 국산 검색엔진으로는 S3가 KERIS에서 호평을 받았고, 외산 검색엔진으로는 S4가 국회도서관에서 좋은 평가를 받았음.
- S4는 국내 기술지원업체로 2개의 전문업체가 경쟁하고 있으며, 이 업체를 중심으로 한글형태소분석 기술개발과 유지보수 지원이 이루어지고 있음.
- KERIS와 국회도서관이 쓰고 있는 검색엔진 S3과 S4를 차기 검색엔진으로 검토해 볼 만함.

□ 면접조사 결과 검색엔진 교체시기에 솔루션의 특성과 기술지원 체계를 잘 고려하여 준비함으로써 어려움을 최소화하고 안정적인 서비스로 발전시킬 수 있었음.

- 따라서 각 기관별 검색서비스 담당자 상호 교류를 통해 검색관련 지식을 지속적으로 공유하는 것이 필요함.
- 특히 검색엔진별로 다양한 경험을 모두 해 볼 수 없기 때문에 다른 사이트의 경험을 활용할 필요가 있음.
- 빈번한 검색엔진의 수정(설정)과 검색 핵심기능에 대한 응용프로그램 처리 비중이 높아짐에 따라 그 만큼 검색엔진 교체와 같은 변경 작업은 감당하기 힘들어지기 때문에 관련한 기술적 사항, 전문 인력, 프로그램 모듈(버전, 소스 등 명세서)을 잘 관리하는 체계가 필요함.

□ 운영단계에서 고려할 사항

- 원활하게 운영하기 위한 적정 규모의 장비는 서비스 모니터링을 통해 일평균 처리용량, 최대 처리용량을 산정하여야 함.
- 장애예방을 위한 여분의 성능과 용량을 확보할 필요가 있음.
- KIPI는 검색서비스운영 전담부서를 조직하고, 서비스별 부하분산과 재난

대응을 위해 지역적으로 시스템을 분리하는 등 다른 기관들 보다 효과적으로 관리하는 좋은 사례임.

- 대용량 검색서비스 운영을 위한 전문인력 양성과 업무프로세스 정착을 위해 지속적인 관심과 지원이 필요함.

□ 최근 특성화된 검색엔진의 성격을 잘 파악하여 검색서비스를 기획하는 것이 중요하고, 서비스 개발/설계과정에서 적합한 검색엔진을 선택하는 것이 바람직함.

□ 서비스에서 검색은 필수요소로서 어떻게 개발하고 운영할 것인지 지속적인 관리가 필요하며, 작업에 있어서 아래의 격언이 주는 의미를 깊게 생각해 보는 것이 필요함.

〈표 11〉 검색 관련 격언

- 검색은 길을 만드는 작업이다.
- 양이 질을 담보한다.
- 검색은 신속하고 정확한 것이 최고다. 다른 기능은 방해가 된다.
- 검색의 품질은 콘텐츠의 양과 색인어 추출기에 달려있다.
- 검색결과 없음은 그 정보가 없을 때 나타나야 한다.
- 사용자도 중요한 정보 원천이다. (선순환구조)
- 억지로 Collective Intelligence를 모으면 역효과(abusing)가 커진다.

〈참고문헌〉

[1] Introduction to Modern Information Retrieval, Gerard Salton, Michael J. McGill, , New York: Mc Graw Hill, 1983.

〈웹사이트〉

[2] 검색기술의 흐름과 동향, <http://www.moransoft.com/materials.html>
[3] 검색엔진의 성능측정 기준, <http://www.powerbox.pe.kr/4>

〈감사의 글〉

이번 『KISTI 지식리포트』 - 정보서비스 공공기관 검색엔진 활용 만족도 조사 - 가 발간되기까지 작업에 기여해준 모든 분들에게 진심으로 감사드립니다. 특히 본 면접조사는 대외적으로 알려지는 것이 민감할 수 있는 사항임에도 불구하고 공공기관의 정보서비스 활용 증진을 위해 국립중앙도서관, 국회도서관, 한국과학기술정보연구원, 한국교육학술정보원, 한국특허정보원, LG상남도서관이 함께 만들어낸 결과물입니다. 조사에 협력해 주신 해당 기관의 검색서비스 전문가 여러분의 귀중한 기여에 다시 한 번 감사의 말씀을 드리고자 합니다.

[별첨-1] 설문지

▣ 대상 : 국립중앙도서관, 국회도서관, KISTI, KERIS, KIPI, LG상남도서관 검색 서비스 운영/관리자, 개발자

▣ 목적 : 공공 및 학술연구 분야의 대용량 검색서비스 운영에 대한 상호 지식을 공유하기 위한 조사입니다. 개발단계와 운영단계에서 서비스 개발자와 운영자가 느끼는 만족도를 조사하고자 합니다. 조사된 내용은 조사목적에 국한하여 활용될 것이며, 보고서로 발간하여 참여기관과 공유하겠습니다.

▣ 조사기간 : 2013. 3. ~ 2013. 4.

▣ 조사자 : KISTI NDSL서비스실 이태석 연구원, 02-3299-6074

(기본 사항)

1. 검색엔진의 종류는 무엇입니까?
   가. IDOL(HP Autonomy)　　나. Search Formula-1(㈜와이즈넛)
   다. 독크루저(코난테크놀로지)　　라. FAST ESP(Microsoft)
   마. 기타(　　)

2. 귀사가 운영하는 서비스를 위해 사용 중인 검색 서버는 몇 대입니까?
   가. 1대~5대　　나. 5대~10대
   다. 10대~20대　　라. 20대 이상

3. 검색 서버의 운영체제(OS)는 무엇입니까?
   가. 리눅스　　나. MS Windows Server
   다. 유닉스(HP-UX, 솔라리스 등)　　라. 기타 (　　)

4. 검색 서버의 기본 사양은 어떻게 됩니까?

가. CPU2개, 8GB RAM　　나. CPU4개 32GB RAM

다. CPU8개, 64GB RAM　　다. 기타 (CPU　　개, GB RAM)

5. 검색서비스 데이터 건수는 몇 건입니까?

가. 1,000만 건 이하　　나. 1,000만 건 ~ 5,000만 건

다. 5,000만 건 ~ 1억 건　　라. 1억 건 이상

6. 일일 검색 이용자수(IP 수)는 몇 명입니까?

가. 1만 명 이내　　나. 1만 명 ~ 5만 명

나. 5만 명 ~ 10만 명　　다. 10만 명 이상

7. 연간 신규(갱신)자료는 어느 정도 입니까?

가. 100만 건　　나. 500만 건

다. 1,000만 건　　라. 1,000만 건 이상

8. 검색 시스템 구성 및 특이사항? (관리화면, 검색엔진 자료, 구성도, 사진 등)

(　　)

(개발 단계 설문)

9. 검색서비스 구축 경험은 몇 년입니까?

가. 3년 이하　　나. 3년 ~ 5년

다. 5년 ~ 10년　　라. 10년 이상

10. 검색서비스 구축 경험이 있는 검색엔진은 몇 개입니까?

가. 1개　　나. 2개

다. 3개　　라. 4개 이상

11. 해당 검색서비스 구축에 사용된 검색엔진(1번 문항)에 대한 교육이나 자료는 충분하고 만족 합니까?

ⓞ	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩

12. 사용한 검색엔진(1번 문항)이 개발하기에 편리하다.

ⓞ	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩

13. 사용한 검색엔진(1번 문항)에 대해 추천 의향은?

ⓞ	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩

14. 검색서비스 개발에서 겪는 애로사항에 대해 자유롭게 적어주세요.

(운영 단계 설문)

15. 검색엔진 관리기능이 제공되고 동작 상태 확인이 편리하다.

ⓞ	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩

16. 검색서비스가 중단되는 일 없이 원활하게 운영 되고 있다.

ⓞ	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩

17. 관리중인 사전이 있으면 모두 체크해 주세요.

가. 불용어 사전　　나. 동의어 사전

다. 신조어(복합명사) 사전　　라. 기타 (　　)

18. 검색서비스와 관련하여 검색식, 키워드관리를 하고 있습니까?

가. Zero 쿼리　　나. 최빈 쿼리

라. 자동완성 검색어　　마. 연관 검색어

19. 기관에서 검색엔진을 어떤 방식으로 선정합니까?

가. 정한 사양의 최저가 입찰을 한다.

나. 경험 있는 전문 업체의 추천으로 선정한다.

다. 성능과 기능 중심으로 제품을 선택한다.

라. 유사 사이트 도입 경험을 토대로 선정한다.

20. 검색엔진에 대해 정해진 기능과 성능 요건은 무엇이라 생각합니까?

(　　)

21. 사용 중인 검색엔진(1번 문항)에 대한 추천 의향은 어느 정도 입니까?

ⓞ	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩

22. 검색엔진을 교체할 계획이 있습니까?

가. 1년 이내　　나. 2 ~ 3년

라. 4 ~ 5년　　마. 5년 이상

23. 검색서비스 운영과 관련하여 궁금한 사항, 어려운 점, 공유할 사항, 기타 관리업무 등 자유롭게 의견을 적어주세요.

조사 일시	2013. . .	기관명	
담 당 자		직 급	
기술 협력자		직급	
서비스 명			
URL			
운영 인력		_____명	

감사합니다.(끝)

[별첨-2] 검색엔진별 기능 비교

■ 검색 서버 구성 및 색인 처리 부문

No	기능 설명	검색엔진별 수준(상,중,하)			
		S1	S2	S3	S4
1	검색서버 복제, 서버 수직 확장 - 검색서버 복제를 통해 검색처리 성능(QPS) 증대	상	상	상	중
2	분산검색, 서버 수평 확장 - 검색데이터 서버 분산으로 처리 용량 증대	상	상	상	중
3	다국어 (한국어, 중어, 일어, 영어) 형태소 색인	상	중	중	상
4	같은 필드에 대해 한영 자동 색인 및 검색 지원 - 한국어 조사 처리 - 영어 단/복수, 형용사/부사 급 변화	불가	중	하	하
5	사전 관리 (불용어, 동의어, 한국어 복합명사 등)	중	상	하	중
6	묶음 필드 색인 및 검색	중	상	상	중
7	한자 음 색인 및 검색	중	중	상	중
8	한글 형태소 분서 및 조사제거 색인 및 검색	하	상	중	중
9	PDF, 워드 문서 색인	중	상	상	중
10	색인 문서 1건 최대 Size	2GB	무제한	무제한	무제한

※ 검색엔진 제품의 실제 명칭은 제조사 보호를 위해 S1~S4로 표시함.

■ 검색 관련 기능(검색 연산자, 정렬, 랭킹) 부문

No	기능 설명	검색엔진별 지원 여부			
		S1	S2	S3	S4
1	Refine 검색, 검색결과에서 지정 필드 조합으로 조회 - 검색결과에서 발행일, 저자, 발행국, 저널별 조회	중	상	상	중
2	정렬 - 다중 정렬 옵션 (날짜 내림, 저자 올림 동시적용)	상	상	상	상
3	랭킹 & 정렬 - 동일한 랭크 문서에 대한 추가 정렬 옵션	불가	상	중	중
4	랭킹(필드가중치, 최신성, 인접성, 빈도) - 묶음 필드 랭킹, 서로 다른 랭킹프로파일 지원	중	중	중	중
5	랭킹 조정 요구사항 (인접도>빈도 조합) - 완전 일치, 키워드 인접상태, 출현 빈도순서로 랭킹을 조정할 수 있음.	불가	상	하	하
6	유사문서 검색 - 입력한 문서(키워드 집합)와 비슷한 문서 검색	중	중	상	중
7	논리 검색 연산자 (and, or, not)	상	중	상	상
8	인접 검색 연산자 (within, near), 범위 지정 - 예) 정보 /w3 처리, 정보* /n3 지원	상	하	중	중
9	구 검색, 입력된 키워드와 순서가 일치된 문장 검색	상	하	상	상
10	범위 연산자 (Range 검색) - 숫자, 날자 필드에 대해 검색 제한을 함.	상	상	상	상
11	Count 연산자, 특정 단어가 지정 횟수보다 많이 출현한 문서만 검색	상	불가	불가	상
12	Boundary Matching, 앞 또는 뒤에서 일치된 입력된 문장으로 시작하는 모든 문서 검색	상	상	불가	불가
13	절단 연산자(Wildcard), 일부가 같은 단어 검색 - 예) 정보*표현, 수?정보*처리	중	불가	중	하
14	Field Collapsing, 콜렉션(뷰)별 검색결과 출력 - 한번검색으로 논문, 특허, 동향에서 각 각 10개씩 출력	하	상	상	불가
15	문서별 중요 색인어 제시(단어별 가중치, DocVector) - 기사(문서) 내용에 나온 주요 키워드	하	상	중	불가
16	검색결과에서 히트된 문장 부분 동적인 출력 - 하이라이트 및 Teaser(요약) 표시	중	상	상	중
17	검색결과 조회(출력) 최대 한계 극복 방법(4,000건)	없음	상	하	없음

※ 검색엔진 제품의 실제 명칭은 제조사 보호를 위해 S1~S4로 표시함.

■ 관리 기능 및 기타 부문

No	기능 설명	검색엔진별 지원 여부			
		S1	S2	S3	S4
1	Admin 관리자 도구 (웹, GUI) 지원 - Zero Query, 최빈 Query, 느린 Query 모니터링 등	하	하	중	하
2	Admin 관리자 도구 (웹, GUI) 지원 - 색인 및 복사 진행 상태 모니터링	중	하	중	하
3	Admin 관리자 도구 (웹, GUI) 지원 - Query 부하 모니터링	중	하	중	하

※ 검색엔진 제품의 실제 명칭은 제조사 보호를 위해 S1~S4로 표시함.

◀ 저 자 ▶

이 태 석	· KISTI 정보서비스실 선임연구원 · tsyi@kisti.re.kr
신 수 미	· KISTI 정보서비스실 선임연구원 · sumi@kisti.re.kr
유 수 현	· KISTI 정보서비스실 선임연구원 · yoosu@kisti.re.kr
정 용 일	· KISTI 정보서비스실 선임연구원 · yijeong@kisti.re.kr
이 은 정	· KISTI 정보서비스실 연구원 · jeong2@kisti.re.kr

KISTI 지식리포트 제35호

정보서비스 공공기관 검색엔진 활용 만족도 조사

인 쇄 2013년 7월 29일
발 행 2013년 7월 31일
펴낸곳 KiSTi 한국과학기술정보연구원 Korea Institute of Science and Technology Information www.kisti.re.kr
펴낸이 박영서
편집인 김혜선
주 소 서울시 동대문구 회기로 66
전화 02-3299-6114
등 록 1991. 2. 12, 제5-258호
ISBN 978-89-294-0299-0 93560
인쇄처 승림디엔씨

□ KISTI 지식리포트 발행 목록

호	서명	저자	발간일
1	학술지 수집 정책 수립을 위한 국내 현황 분석	이재윤, 김혜선, 이혜진	2009.06.11
2	국내 과학기술지식의 글로벌 확산 전략	서태설, 최현규	2009.09.04
3	국가 과학기술 진흥을 위한 KISTI의 전략적 정보자원 개발방안	황혜경, 최호남, 윤희윤	2009.09.21
4	학술논문 오픈 액세스를 위한 공공접근정책 방향	서태설, 허 선, 노경란	2009.10.23
5	과학데이터의 공유와 활용	이상환, 심원식	2009.11.10
6	E.infrastructure기반 국가 R&D 정보서비스의 지능화 방안	송인석, 오세홍	2009.12.04
7	한국과학기술인용색인서비스(KSCI)의 현황 및 발전 전략	최선희, 이재윤	2010.01.29
8	학술정보센터의 새로운 서비스 모델 : 오픈 액세스 출판	노경란, 이혜진	2010.02.11
9	과학기술정보의 아카이빙 체제 구축	황혜경, 이선희, 최호남, 서혜란	2010.02.16
10	세계 주요 과학기술 정보기관의 최근 동향	이상환, 노경란, 김혜선, 황혜경, 정은경	2010.03.02
11	과학기술 정보자료 보존관리 : 현황분석 및 미래예측	이선희, 황혜경, 류범종, 윤희윤, 김석영	2010.06.30
12	리포지터리 사례분석 및 시사점 도출	이상기, 정영미	2010.10.07
13	디지털 콘텐츠 유통을 위한 저작권 쟁점 분석	유수현	2010.10.07
14	이공계 대학 교수의 과학기술정보 이용 현황	김환민, 김재훈	2010.12.02
15	계량서지적 분석용 공개 소프트웨어 활용 방안	최선희, 김희정, 이재윤	2011.01.24
16	2010년도 한국 과학자의 SCI 논문 계량분석	김완종, 노경란, 박민수, 최현규	2011.04.06
17	국내 과학기술정보 이용실태 조사 분석	박민수, 이상환, 최현규, 정정수	2011.04.08
18	국내 과학기술정보 이용자 니즈 및 형태 연구	박민수, 이상환, 최현규, 정정수	2011.04.08
19	학술지 유통환경 변화와 국내 학술지의 국제화	서태설, 김규환, 최현규	2011.06.07
20	우리나라 SCI급 논문의 영향력 분석 : NCR for Korea 1981-2010을 기준으로	김완종, 노경란, 최현규, 박민수	2011.08.17
21	연구자들의 소셜 미디어 이용	노경란, 최현규	2011.10.19
22	연구자를 위한 소셜 미디어 활용 가이드	노경란, 최현규	2011.10.19
23	서비스 사이언스 기반 과학기술 콘텐츠 서비스 방안	김지영, 신기정, 황혜경, 조부연	2011.10.26
24	효율적 연구를 위한 소셜 미디어 활용	노경란, 유수현, 최현규	2011.11.09
25	SEO(검색엔진최적화)를 통한 검색순위 올리기 전략	현미환, 이태석, 문영수, 권정혁	2011.11.11
26	연구자 협업지원형 정보서비스 사례연구	이혜진, 현미환, 김혜선, 박민수, 최현규	2011.12.01
27	과학기술 R&D 라이프사이클 연구 : 생명공학 및 나노분야를 중심으로	김혜선, 권나현, 정은경, 이정연, 최현규	2011.11.30
28	웹사이트의 사용성 개선을 위한 단계별 전략	현미환, 박민수, 이태석, 최현규	2011.12.05
29	국내 과학기술 연구자의 소셜 미디어 활용 현황	현미환, 이혜진, 김혜선, 박민수, 최현규	2011.12.08
30	이용현황 분석을 통한 학술정보 활용지표 개발	이혜진, 유수현, 김혜선 이재윤	2012.01.31
31	FTA와 개정 저작권법이 정보유통에 미치는 영향	유수현, 이대희	2012.02.21
32	고품질 정보서비스 지원을 위한 컴퓨팅자원 인프라 구축	정영임, 신용주, 한성근, 김재훈, 김정환, 최호남	2012.03.30
33	모바일 앱(App) 개발을 위한 특화 기술 분석	현미환, 신수미, 김혜선	2012.08.10
34	한국과학기술정보연구원(KISTI) 원문제공서비스 현황 분석	이선희, 김지영, 문영수, 신기정	2012.12.31