

필드검색시스템과 XML기반 구조검색시스템의 구축과 이용자 기반의 정확률 및 만족도 평가에 관한 연구

A Study on the Construction and User Base Evaluation of the Accuracy Rate and Satisfaction about Field Search System and XML Base Structure Search System

김덕성, 성균관대학교 문헌정보학과

Kim Duck-Sung

Sung Kyun Kwan University Department of Library & Information Science

많은 정보가 인터넷 상에 산재함에 따라 검색엔진의 한계를 느끼게 되었으며, 정보의 질적인 문제를 해결하기 하여 메타데이터가 등장하게 되었으며, 이러한 메타데이터를 대상으로 검색하는 시스템이 등장하게 되었다. 이에 본 연구는 2가지 메타데이터 검색시스템(필드검색시스템과 XML기반 구조검색시스템)을 구축하고 시스템의 차이와 이용자의 메타데이터 구조에 대한 이해의 차이가 이용자의 검색 만족도와 검색 정확율에 어떠한 영향을 미치는가를 밝히는 것이다.

I 서론

정보 기술의 발전으로 많은 사람들이 인터넷을 이용하고 있으며, 인터넷 초창기에 비하여 정보의 양이 급격히 증가함에 따라 많은 이용자들이 인터넷 검색엔진들을 사용하고 있다. 그러나 검색엔진의 발전에도 불구하고 검색엔진의 한계로 인하여 정보검색에 어려움을 겪고 있다. 위와 같은 정보 검색 환경의 변화에 따라 정보의 질적인 문제가 제기되었고 해결방안으로 메타데이터가 등장하게 되었다.

현재 메타데이터를 이용한 검색 시스템은 2가지 방식으로 구축되고 있다. 하나는 RDB 또는 BRS 검색엔진 등에 데이터를 저장하고 저장된 필드를 검색하는 시스템이며, 다른 하나는 XML 문서의 형태로 DB에 저장하고 XML문서의 요소

(Element)와 속성(Attribute)단위로 검색하는 시스템이다. 이 두 시스템은 데이터를 구조화하여 저장한다는 면에서는 유사하지만 검색하는 이용자 인터페이스 구조는 차이가 있다.

전자의 경우 메타데이터의 요소를 개별적인 필드로 나열하고 해당 필드에 대해 검색하는 필드 단위의 접근점을 가지는데 반하여, 후자는 트리(Tree)형식의 계층적인 검색 인터페이스를 제공하여 XML 문서의 요소와 속성에 대하여 구조적인 검색을 가능하게 한다.

현재까지의 연구는 각각의 구축된 필드검색 시스템에 대한 평가 또는 XML문서 저장방식이 주를 이루고 있어, 이용자의 검색인터페이스와 시스템의 차이에 따른 이용자 중심의 평가 또한 이루어지지 않아 이 분야에 대한 연구가 요구되고 있다.

본 연구의 목적은 동일한 Dublin Core 메타데이터 요소를 기반으로 필드검색 시스템과 XML 기반 구조검색 시스템을 구축하고, 검색 결과의 정확률과 검색 시스템에 대한 이용자의 만족도 측정을 통해 두 시스템에 대한 이용자 기반 평가를 하는 것이다. 또한 XML구조에 대한 선 이해가 정보 검색에 어떠한 영향을 미치는지를 연구하기 위하여, DC 메타데이터 구조에 대한 지식이 있는 이용자와 그렇지 않은 이용자 집단을 선정하여 검색 시스템의 차이에 따른 검색결과와 정확성과 검색 시스템에 대한 이용자의 만족도의 차이가 있는지를 밝히고자 한다.

II 연구 방법론

2.1 연구 디자인

필드검색 시스템과 XML기반 구조검색 시스템의 인터페이스의 차이와 DC 구조에 대한 지식여부에 따른 검색 시스템의 평가를 위하여 시스템을 구축한다. 시스템 구축을 위한 데이터 요소로는 Gateway to Faculty Syllabus (GFS) 메타데이터 요소를 추출하여 구축하며, GFS 메타데이터 시스템을 구축할 때 사용한 문헌정보학 대학 강의안 데이터를 이용하여 구축한다. 필드검색 시스템은 Oracle RDBMS의 테이블에 메타데이터 요소를 저장하고 필드 검색 인터페이스를 통하여 검색할 수 있도록 구축하며, XML기반 구조검색 시스템은 Oracle8i에서 XML구조검색을 지원하기 위하여 고안한 interMedia Text 옵션을 이용하여 CLOB(Character Large Object) 데이터 타입의 필드에 XML문서를 저장하고 트리 형식의 구조적인 접근이 가능한 인터페이스를 통하여 검색할 수 있도록 구축한다.

구축된 검색 시스템의 비교 평가를 위하여 다

음과 같은 실험을 실시한다. 필드검색 시스템과 XML기반 구조검색 시스템의 비교 평가와 DC 메타데이터에 대한 구조가 시스템 평가에 영향을 미치는지 평가하기 위한 실험에서 선정된 변인은 다음과 같다. 독립변인으로는 두 시스템의 차이와 DC 메타데이터 구조에 대한 지식 여부이며, 종속 변인으로는 이용자의 평가에 의한 검색결과와 정확률과 검색 시스템에 대한 이용자의 만족도이다. 본 연구를 수행하기 위하여 선정된 독립변인과 종속 변인은 <표 1>과 <표 2>로 나타내었다.

<표 1> 독립 개념과 종속 개념

독립 개념	종속 개념
1. 필드검색 시스템과 XML기반 구조검색 시스템	a. 검색결과와 정확률
2. DC 메타데이터 구조에 대한 지식 여부	b. 검색 시스템의 이용자 만족도

<표 2> 변인 및 측정요소

독립 변인	종속 변인
1.1 필드검색 시스템	a' - 요약 검색결과에 의거한 정확률 - 상세 검색결과에 의거한 정확률 - 실제 사이트에 의거한 정확률
1.2 XML기반 구조검색 시스템	b' # 검색 시스템에 나타난 화면에 대한 만족도 - 화면 구성의 복잡 정도 - 검색 요소 제시방법의 편리 정도 - 화면 연결의 자연스러움 정도
2.1 DC 메타데이터 구조에 대한 지식이 있는 이용자	b' # 검색 시스템 기능에 대한 만족도 - 검색결과 수의 적절성 정도 - 이용자의 숙련도에 따른 고려 정도 - 이용자의 정보요구 해결 정도 - 검색을 위해 접근해야하는 클릭 수
2.2 DC 메타데이터 구조에 대한 지식이 없는 이용자	# 검색 시스템에서 사용되는 용어와 시스템 정보 - 검색 작업과 시스템 용어와의 관련성 정도 - 검색 시스템 용어 선정의 적절성 정도 # 검색 시스템을 사용하기 위한 학습에 관한 만족도 - 직관적인 검색 시스템 사용 방법의 이해 정도 - 검색 방법 학습의 어려움 정도 - 시행착오를 통한 새로운 기능 습득 어려움 정도

실험을 위한 피실험자의 선정은 DC메타데이터 구조에 대한 지식이 시스템 평가에 영향을 미치는지 측정하기 위하여 DC메타데이터 구조에 대한 지식이 있는 이용자와 그렇지 않은 이용자 두 그룹으로 나누어 선정한다. 피실험자는 두개의 검색 시스템을 이용하여 피실험자 본인이 선택한 임의의 검색 질의에 대하여 검색을 수행한다. 한 시스템에서 얻은 경험이 다른 시스템의 이용에 영향을 미치는 전이효과를 상쇄시키기 위하여 피

실험자 집단을 다시 두 개의 그룹으로 나누어 한 그룹은 필드검색 시스템을 다른 그룹은 XML기반 구조검색 시스템을 먼저 수행하게 한다.

실험 중 종속변인인 정확율을 평가하기 위하여 실험자는 피실험자가 평가한 검색 결과의 적합자료 수를 기록하고, 실험을 끝낸 후 피실험자는 설문지를 통하여 제시된 검색 시스템에 대한 만족도에 대한 답을 함으로써 실험을 모두 마치게 된다.

2.2 귀무 가설

1. DC 메타데이터 구조에 대한 지식이 있는 이용자에게 있어, 필드검색 시스템과 XML기반 구조검색 시스템의 차이에 따라 이용자가 평가한 검색 결과에 대한 정확율에 차이가 없을 것이다.
2. DC 메타데이터 구조에 대한 지식이 있는 이용자에게 있어, 필드검색 시스템과 XML기반 구조검색 시스템의 차이에 따라 이용자의 검색 시스템에 대한 만족도에 차이가 없을 것이다.
3. DC 메타데이터 구조에 대한 지식이 없는 이용자에게 있어, 필드검색 시스템과 XML기반 구조검색 시스템의 차이에 따라 이용자가 평가한 검색 결과에 대한 정확율에 차이가 없을 것이다.
4. DC 메타데이터 구조에 대한 지식이 없는 이용자에게 있어, 필드검색 시스템과 XML기반 구조검색 시스템의 차이에 따라 이용자의 검색 시스템에 대한 만족도에 차이가 없을 것이다.
5. XML기반 구조검색시스템에 있어 DC 메타데이터 구조에 대한 지식이 없는 이용자와 있는 이용자 그룹의 차이에 따라 이용자가 평가한 검색 결과에 대한 정확율에 차이가 없을 것이다.

6. XML기반 구조검색시스템에 있어 DC 메타데이터 구조에 대한 지식이 없는 이용자와 있는 이용자 그룹의 차이에 따라 이용자의 검색 시스템에 대한 만족도에 차이가 없을 것이다.

2.3 평가 기준 및 평가 방법

2.3.1 검색결과의 정확율

정확율이란 검색된 결과자료 가운데 적합한 결과자료의 비율로서, 본 연구에는 이용자 기반의 평가를 주목적이며 실험주제를 문헌정보학에 한정하고 문헌정보학 전공자를 피실험자로 선정하여 평가하기 때문에 검색결과의 정확율은 피실험자의 적합성 판정 여부만을 사용하였다.

피실험자를 대상으로 검색 결과에 대한 적합성 판정 척도는 3-5개 수준¹⁾으로 선정하였다. 본 연구에서는 '적합하다', '부분적으로 적합하다', '적합하지 않다'의 3점 기준으로 적합성을 판정하게 하였으며, 정확률의 측정 시 부분적으로 적합한 검색결과에 대한 판정을 위하여 정확율 이외에 넓은 의미의 정확율인 확장 정확율을 이용하여 측정하였다.

검색결과의 순위화(Ordering)는 피실험자가 선정한 검색어와 가장 많이 일치하는 자료를 우선 순위로 정렬하며, 총 검색결과가 20개 이상의 자료가 검색되었을 경우 20개 자료에 대해서만 적합성 판정을 하게 하였다.

(1) 요약정보 및 상세정보에 의한 정확률

요약정보는 처음 출력되는 검색 결과로서 본 시스템에서는 교과목명, 표준학과명, 대학명, 교수명이 출력된다. 상세정보 하이퍼링크를 클릭하면 메타데이터 요소로 선정된 모든 요소에 대한 정

1) Robert R. Korfhange, 류근호, 이세한 옮김. "정보저장 및 검색". 서울: 시그마프레스. 2000. p.218

보를 확인할 수 있다.

요약 및 상세 정보에 의한 정확율은 피실험자에게 시스템에서 제공해주는 요약정보와 상세정보를 대상으로 3점 기준에 따라 적합성 판정을 하게 하는 것이다.

(2) 실제 사이트에 의거한 정확률

실제 사이트에 의한 정확율은 상세정보에서 해당 사이트의 URL(Uniform Resource Locator) 클릭을 통하여 실제 대학 강의안을 제공해주고 있는 사이트로 이동한 후 피실험자로 하여금 적합성 판정을 하게 하는 것이다.

2.3.2 검색 시스템의 사용자 만족도

친(Chin 1998)²⁾은 이용자의 주관적인 만족도를 측정할 수 있는 설문지에 관한 연구를 하였으며, 시스템에 대한 전체적인 이용자 반응(overall reaction to the software), 화면(screen), 용어와 시스템 정보(terminology and system information), 학습(learning) 그리고 시스템 기능(system capabilities)에 있어 신뢰성이 높은 것으로 나타났다.

본 연구에서는 친(1998)의 연구를 바탕으로 이용자 만족도의 변인으로 검색 시스템의 화면, 검색 시스템의 기능, 검색 시스템에서 사용한 용어 그리고 검색 시스템을 사용하기 위한 학습 정도를 선정하였다. 본 검색 시스템의 만족도는 각각의 시스템에 대한 실험을 마친 후, 피실험자에게 설문지를 제공함으로써 평가한다. 설문지를 통한

질문은 7점 척도(1:매우 불만족, 7:매우 만족)로 답하게 하며, 이용자 만족도에 관한 설문지는 부록 2에 수록하였다.

2.4 피실험자 선정

본 연구에서 선정된 실험대상자들은 성균관대학교에 재학 중인 문헌정보학 전공 대학원생, 학부생 40명이며, DC 메타데이터 구조에 대한 지식이 있는지 여부에 따라 DC 메타데이터 구조에 대한 지식이 있는 이용자 그룹과 DC 메타데이터 구조에 대한 지식이 없는 이용자 그룹으로 나뉜다. DC 메타데이터 구조에 대한 지식이 있는지 여부에 따른 그룹의 선정 기준은 다음과 같다.

DC 메타데이터 구조에 대한 지식이 있는 이용자 그룹 : DC 메타데이터 구조에 대한 지식이 있는 이용자 그룹은 메타데이터에 대한 과목을 이수한 학생 및 DC 15개 요소와, 한정어에 대하여 15개 요소 중 어느 요소의 한정어에 해당하는지를 말할 수 있는 학생 그룹으로 이루어진다. 본 연구에서는 총 40명 중 20명을 실험 대상으로 하여 신청서를 배포하고 내용을 설명한 후 흥미가 있고 참여의사가 있는 학생이 자발적으로 참여하도록 한다.

DC 메타데이터 구조에 대한 지식이 없는 이용자 그룹 : DC 메타데이터 구조에 대한 지식이 없는 이용자 그룹은 메타데이터에 대한 과목을 이수하지 않으며, DC 15개 요소에 대하여 말할 수 없는 학생 그룹으로 이루어진다. 본 연구에서는 총 40명 중 20명을 실험 대상으로 하여 신청서를 배포하고 내용을 설명한 후 흥미가 있고 참여의사가 있는 학생이 자발적으로 참여하도록 한다.

또한 필드검색과 XML기반 구조검색을 실시하기 위하여 각각의 20명의 이용자 그룹의 학생들에게 고유의 번호를 부여한 후, Visual Basic 6.0

2) John P. Chin, Virginia A. Diehl, Kent L. Norman. "Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface." Conference proceedings on Human factors in computing systems May 15-19, Washington United States. 1988. pp.213-218.

의 random함수를 이용하여 각 그룹을 다시 2그룹으로 나누었다. 2그룹은 필드검색을 먼저 수행하는 그룹과 XML기반 구조검색을 먼저 수행하는 그룹으로 총 4개의 그룹을 선정한다.

2.5 실험 절차

본 연구는 필드검색 시스템과 XML기반 구조검색 시스템의 검색 결과의 정확율과 이용자의 만족도를 평가하며, DC 메타데이터에 대한 지식 여부에 따라 검색 결과의 정확율과 이용자의 만족도에 차이가 있는지를 검증하기 위하여 다음 순서에 따라 실험을 실시한다.

가. 실험자는 실험내용과 실험절차에 대하여 피실험자에게 간단히 설명하며, 실험자가 선정한 간단한 검색 질의를 수행하게 함으로써 피실험자에 대한 튜토리얼을 마친다.

나. 피실험자는 문헌정보학 주제에 관하여 피실험자 자신이 임의로 선정한 검색질의를 가지고 두 검색 시스템 중 실험자가 선정한 검색 시스템 이용하여 검색을 수행한다. 이때 실험자는 피실험자의 현재 수행한 검색결과에 전이 효과를 상쇄하기 위하여 실험 집단을 이용자 그룹에 따라 통제한다.

다. 검색 수행을 통한 검색결과가 주어지면 피실험자는 출력된 요약정보와 상세정보를 읽고 자료의 적합성여부를 판단하며, 실험자는 피실험자가 판단한 적합한 자료의 건수와 총 열람한 자료의 건수를 기록한다.

라. 피실험자는 '다'의 결과 중 원하는 자료에 대하여 실제 사이트에 접근함으로써 적합성 여부를 판단하며, 실험자는 피실험자가 실제 사이트에 접속한 총 건수와 피실험자가 판단한 적합한 자료의 건수를 기록한다.

마. 실험을 마친 피실험자는 이용자의 만족도 평

가를 위해 설문지에 제시된 질문에 대답하도록 한다.

바. 한 시스템에 대한 실험이 끝난 피실험자는 '가'에서 '마'까지 수행한 절차에 따라 다른 검색 시스템을 이용하여 실험에 임하도록 한다.

사. 한 명의 피실험자의 실험이 끝난 후 웹사이트에 대하여 방문했던 흔적을 지우기 위하여 웹 브라우저에 남아있는 Cash와 History정보를 삭제한다.

2.6 자료 수집

피실험자 그룹의 선정은 2001년 8월 1일까지 완료하며, 실험 및 자료의 수집기간은 2001년 9월 11일부터 2001년 9월 30일까지 20일간 실시한다.

자료수집을 위해서 실험을 실시하고 정확율 산정을 위하여 실험 결과를 실험자가 기록하고, 실험 후 피실험자를 대상으로 만족도에 대한 설문지를 웹 상으로 받아 데이터베이스에 저장한다.

참고문헌

- 김현희, 장혜원. "디지털도서관 문서 양식으로서의 XML과 HTML의 특성 및 검색 기능 비교 연구". 정보관리학회지 16(2). 1999. pp.105-134
- 오삼균. "인터넷 대학 강의안의 검색을 위한 Metadata DB 구축". 정보관리학회지 16(2). 1999. pp.149-163
- 윤승구. "구조와 메타정보 검색을 지원하는 SGML웹 검색엔진의 개발". 연세대학교 대학원 석사학위논문. 1998.
- 이은경. "인터넷상의 전산학 분야 대학 강의안 메타데이터 시스템 구축 및 평가". 성균관대학교 대학원 석사학위논문. 2000.
- 채승기. "XML 저장 기법 및 활용방안". 정보관리강좌 2001. 정보관리학회. 2001. pp61-81.

한국 더블린 코어 메타데이터. "자주 질문되는 물음들 (FAQ)". 2001

<http://www.dublincore.or.kr/faq.htm>.

Bruce, Harry. "User Satisfaction with Information Seeking on the Internet" *Journal of the American Society for Information Science*. 49(6). 1998. p.542

Chin, John P, Diehl, Virginia A, Norman., Kent L. "Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface." *Conference proceedings on Human factors in computing systems May 15-19, Washington United States*. 1988. pp.213-218.

Chu, Heting, Rosenthal, Marilyn. "Search Engines for the World Wide Web :A Comparative Study and Evaluation Methodology," *ASIS 1996 Annual Conference Proceedings October 19-24*. 1996.

Gilliland-Swetland, Anne J, Kafai, Yasmin B, Landis, William E. "Application of Dublin Core Metadata in the Description of Digital primary Source in Elementary School Classrooms". *Journal of the American Society for Information Science* 51(2). 2000. pp.193-201

Su, Louise T, Chen, Hsin-liang. "Evaluation of Web Search Engines by Undergraduate Students". *ASIS'99 Proceedings of the 62nd ASIS Annual Meeting* 36. 1999. pp.98-114

Zeng, Marcia Lei. "Metadata Element for Object Description and Representation: A Case Report from a Digitized Historical Fashion Collection Project". *Journal of the American Society for Information Science* 50(13). 1999. pp.1193-1208