

지식 검색 서비스 개선을 위한 문서의 적합도 및 신뢰도 분석

Evaluation of the Documents from the Web-based Question and Answer Service

박 소 연(Soyeon Park)*
이 준 호(Joon Ho Lee)**
전 지 운(Jiwoon Jeon)***

목 차

- | | |
|----------------------|---------------------|
| 1. 서 론 | 4.4 답변의 신뢰도 평가 기준 |
| 2. 선행 연구 | 5. 평가 결과 |
| 3. 연구 방법 | 5.1 질문 제목 적합도 평가 결과 |
| 4. 평가 기준 | 5.2 질문 전체 적합도 평가 결과 |
| 4.1 질문 제목의 적합도 평가 기준 | 5.3 답변의 적합도 평가 결과 |
| 4.2 질문 전체의 적합도 평가 기준 | 5.4 답변의 신뢰도 평가 결과 |
| 4.3 답변의 적합도 평가 기준 | 6. 결 론 |

초 록

지식 검색 서비스가 국내 포털들의 대표적인 서비스로 정착되었음에도 불구하고 지금까지 지식 검색 서비스의 질적 향상을 위한 연구나 지식 검색 데이터베이스의 문서 평가에 대한 연구는 미흡한 실정이다. 이에 본 연구에서는 지식 검색 데이터베이스를 구성하는 지식 문서의 평가 기준을 제시하였다. 구체적으로 본 연구에서는 지식 문서를 구성하는 질문 제목, 질문 전체, 답변의 적합도 평가 기준을 제안하고, 답변의 신뢰도 평가 기준도 제시하였다. 이러한 평가 기준에 근거하여 본 연구에서는 지식 문서의 실제 평가를 수행하였다. 본 연구의 결과는 지식 검색 테스트 컬렉션 구축과 신뢰도 컬렉션 구축에 활용되어 궁극적으로 지식 검색 서비스 개선에 기여할 것으로 기대된다.

ABSTRACT

This study suggests evaluation criteria for the web-based question-answer databases provided by major Korean search portals. In particular, this study suggests evaluation criteria for the relevance of question titles, entire questions, and answers. The evaluation criteria for the quality of answers are also developed. Based on these criteria, evaluation of documents from Naver Knowledge-in are performed. The results of this study can be implemented to the development of test collection of question-answer databases. The implications for system designers and web content providers are discussed.

키워드: 지식 검색, 평가 기준, 검색 포털

Web-based Question and Answer Services, Evaluation Criteria, Search Portals

* 덕성여자대학교 문헌정보학과 조교수(sypark@duksung.ac.kr)

** 숭실대학교 정보과학대학 컴퓨터학부 부교수(joonho@computing.ssu.ac.kr)

*** Center for Intelligent Information Retrieval, University of Massachusetts-Amherst
(jeon@cs.umass.edu)

논문접수일자 2006년 5월 15일

게재확정일자 2006년 6월 13일

1. 서론

2002년 10월 네이버에서 지식인 서비스를 시작한 이후, 엠파스 야후 등과 같은 포털들도 유사한 서비스를 제공하면서 지식 검색 서비스는 국내 검색 포털들의 대표적 서비스로 성장하였다. 이용자들이 올리는 질문과 답변으로 구성되는 네이버의 “지식인” 데이터베이스는 현재 약 800만 건 이상의 문서를 축적하였다. 비즈니스 위크지가 구글이 NHN에게서 배워야 할 점으로 지식인 서비스를 주목했을 정도로(Moon, & Woyke, 2006), 지식인 서비스는 시작 직후부터 지금까지 국내 이용자들에게 폭발적인 인기를 끌고 있다. 이러한 인기는 누구나 어떠한 주제에 대해서도 질문과 답변을 할 수 있다는 개방적 구조와, 이렇게 축적된 자료를 공유할 수 있다는 커뮤니티적인 특성에 기인한다. 그러나 검증되지 않은 답변이나 추측성 답변들로 인하여 답변의 신뢰성, 정확성, 전문성이 저하되고 지식의 질적 하락이 초래될 수 있다는 비판도 공존한다.

지식 검색 서비스가 국내 포털들의 대표적인 서비스로 정착되었음에도 불구하고, 지금까지 지식 검색 서비스의 질적 향상을 위한 체계적인 연구는 매우 미흡한 실정이며, 이에 대한 원인들 중의 하나로서 개발 중인 검색 기법의 성능을 평가할 수 있는 테스트 컬렉션의 부재를 들 수 있다. 테스트 컬렉션은 질의 집합과 문서 집합 그리고 각각의 질의에 대한 적합 문서 리스트로 구성된다.

지식 검색 서비스에서 질의는 검색 포털들의 통합 검색 창에 입력되며, 지식 검색 시스템은 지식 검색 데이터베이스로부터 사용자 질의에

적합할 가능성이 높은 문서들을 검색한다. 따라서 지식 검색 테스트 컬렉션의 질의 집합은 통합 검색 창에 입력되는 질의들로부터 선정되고, 문서 집합은 지식 검색 데이터베이스로부터 선정되는 것이 바람직하다. 적합 문서 리스트는 테스트 컬렉션의 가장 중요한 요소이며, 본 연구에서는 지식 검색 데이터베이스로부터 통합 검색 질의에 대한 적합 문서를 선정하기 위한 평가 기준을 개발하고자 한다.

지식 검색 데이터베이스를 구성하는 하나의 문서는 질문 제목, 질문 본문, 답변으로 구성되어 있으며, 통합 검색 질의에 적합한 문서는 질문 제목과 질문 전체, 답변 모두가 적합한 문서로 정의되어야 한다. 일반적으로 통합 검색의 검색 결과에는 질문 제목만이 노출되고, 사용자는 통합 검색 질의에 적합하지 않은 질문 제목을 조회하지 않기 때문에, 통합 검색 질의에 적합한 지식 문서는 그 지식 문서의 질문 제목이 적합한 문서로 정의되는 것이 바람직하다. 따라서 본 연구에서는(i)통합 검색 질의에 대한 질문 제목의 적합도 평가 기준을 개발하고자 한다. 또한 질문 제목이 적합하더라도 전체 질문이 부적합하거나 질문 제목이 부적합하더라도 전체 질문이 적합한 경우도 존재하므로, 본 연구는(ii)통합 검색 질의에 대한 전체 질문의 적합도 평가 기준을 별도로 제안한다. 한편 답변이 부적합한 지식 문서는 질문의 적합성과 무관하게 부적합 문서로 정의되는 것이 바람직하므로, 본 연구에서는(iii)답변의 적합도 평가 기준을 분리하여 개발한다.

답변의 신뢰성, 정확성, 전문성을 향상시키기 위한 방법으로 답변의 신뢰도를 측정하여 이를 검색 기법에 반영한다면, 신뢰성이 낮은

지식 문서가 통합 검색 사용자에게 노출될 가능성을 줄일 수 있다. 이러한 기법의 개발을 지원하기 위하여 본 연구에서는(iv) 답변의 신뢰도 평가 기준을 개발하고자 한다.

이처럼 본 연구에서는 지식 문서를 구성하는 질문 제목, 전체 질문, 답변의 적합도 평가 기준을 제안하며, 답변의 신뢰도 평가 기준을 별도로 제안한다. 또한 본 연구에서는 이러한 평가 기준에 근거하여 이용자가 실제 입력한 질의에 대한 지식 문서의 실제 평가를 수행하고자 한다. 지식 문서에 대한 평가 기준을 제안하고 실제 평가를 수행한 본 연구의 결과는 지식 검색 테스트 컬렉션 구축과 답변의 신뢰도 컬렉션 구축에 직접적으로 활용될 수 있을 것으로 기대된다. 테스트 컬렉션은 시스템 혹은 관련 기술의 성능 평가를 통하여 시스템의 발전과 정보 유통에 있어서 경쟁력 강화에 필수적인 역할을 한다(맹성현 외 1999). 또한 본 연구에서 제시한 지식 문서 답변의 신뢰도 측정을 검색 기법 개발에 반영한다면 보다 질 높은 검색 결과를 이용자에게 제공할 수 있을 것이다. 이러한 테스트 컬렉션 구축과 검색 기법 개발은 향후 지식 검색 서비스 개선에 기여할 것으로 보인다.

2. 선행 연구

지식 검색 서비스가 본격적으로 상용화되기 시작한 것이 2000년 이후의 현상이므로 국내외 선행 연구들 중 지식 검색 데이터베이스의 문서 평가와 직접적으로 관련된 연구는 드문 실정이다. 국외 선행 연구의 경우 일반적인 웹 문서의 질 평가에 대해서는 상당한 연구가

수행되었으나(Strong, Lee, & Wang 1997; Zhu & Gauch, 2000; Zhou, & Croft 2005; Kleinberg 1999), 지식 검색 서비스와 같은 질의응답형 서비스의 문서 평가에 대한 연구는 찾아보기 어렵다. 또한 웹 문서의 질 평가에 있어서도 구체적이고 체계적인 평가 기준이 개발된 적은 드물다. 국외 연구 중 지식 검색 서비스와 관련된 일반적인 연구로는 지식 검색 서비스의 역사와 현황을 소개하거나(Bivens-Tatum, 2001; Pack, 2000) 간단한 성능 실험을 수행한 연구들을 들 수 있다(Janes, Hill, & Rolfe, 2001).

국내 관련 연구로는 박주범, 정동열(2004), 이두영, 강순희(2003), 최상희(2005)의 연구를 들 수 있다. 박주범, 정동열(2004)은 네이버, 야후, 엠파스의 지식 검색 서비스를 대상으로 72개의 질문을 이용한 실험을 통하여 응답률, 응답소요시간과 같은 간단한 성능을 비교하였다. 이들은 또한 253명의 지식 검색 서비스 이용자들에 대한 온라인 설문조사를 통하여 이용자들의 전반적인 특성, 정보이용행태 등을 조사하고, 지식검색 서비스를 개선하기 위한 방안을 지식의 효과성과 서비스의 효율성이라는 측면에서 제시하였다. 이두영, 강순희(2003)는 네이버, 야후, 엠파스가 제공하는 지식검색커뮤니티의 현황과 지식의 순환단계를 기술하였다. 최상희(2005)는 네이버 지식인의 질의응답문서들을 대상으로 클러스터링 기법과 수작업 주제 분류의 주제 표현 성능을 비교, 평가하였다.

3. 연구 방법

지식 검색 데이터베이스의 질문과 답변 평가

기준 개발을 위해서는 통합 검색 질의와 지식 문서의 전반적인 특성을 파악하는 작업이 수행되어야 한다. 이를 위하여 본 연구에서는 주중의 평일을 대표할 수 있는 2004년 6월 29일 하루 동안 네이버 통합 검색 창에 입력된 질의들과 “지식인” 데이터베이스에 축적된 지식 문서들 중의 일부를 다음과 같이 선정하였다. 본 연구에서 네이버를 선택한 이유는 네이버가 국내 검색 포털들 중 최초로 지식인 서비스를 시작하였으며, 네이버 지식인 서비스의 대중성이나 인지도가 타 검색 포털들보다 높기 때문이다.

- 통합 검색 질의 선정: 통합 검색 창에 입력된 질의들 중에서 무작위로 700개를 선정하고, 이들을 박소연, 이준호(2005)의 질의 형태 분류에 따라 “사이트 검색”과 “내용 검색”으로 분류하였다. 사이트 검색은 이용자의 검색 대상이 웹 사이트인 경우이며, 내용 검색은 이용자의 검색 대상이 특정한 주제에 관한 신문 기사, 게시판 글 “지식인”에 등록된 글들인 경우이다. 이러한 질의 형태 범주 도출 시 TREC내 Web Track의 검색 과제와 네이버가 제공하는 데이터베이스의 구성을 참고하였다. 질의의 형태 분류 후 질의와 결과가 일치하는 정도에 따라 사이트 검색 질의들을 “만족”과 “불만족”으로 분류하고, 내용 검색 질의들을 “만족”, “보통”, “불만족”으로 분류하였으며, 범주별로 25개씩 총 125개의 질의를 선정하였다. 이처럼 형태나 만족도가 상이한 질의를 선정하는 이유는 평가 기준 도출 시 보다 다양한 지식 문서를 참고하기 위해서였다.
- 지식 문서 선정: 선정된 통합 검색 질의

125개에 대해 지식 문서의 3개 필드(질문 제목, 질문 본문, 답변)를 검색하였다. 이때 2개의 상이한 검색 알고리즘을 사용함으로써 각각의 질의에 대하여 6개의 검색 결과 리스트들을 생성하였다. 이렇게 생성된 6개의 검색 결과 리스트들에서 상위 20개씩의 문서를 추출하여 총 120개의 문서들을 추출한 후, 중복된 문서를 제거함으로써 총 9,104개의 지식 문서들을 선정하였다.

본 연구에서는 선정된 125개 질의들과 9,104개 지식 문서들에 대한 내용 분석에 기초하여, 연구자들과 평가자들이 2005년 7월과 8월에 걸쳐 통합 검색 질의에 대한 질문 제목의 적합도 평가 기준과 질문 전체의 적합도 평가 기준을 귀납적으로 도출하였다. 또한 지식 문서의 질문에 대한 답변의 적합도 평가 기준과 답변의 신뢰도 평가 기준을 도출하였다. 적합도 평가 방식은 ‘적합,’ ‘부적합’의 이분법적인 방식이 아닌 ‘적합,’ ‘보통,’ ‘부적합’의 3점 척도를 사용하기로 결정하였다. 이처럼 3점 척도를 사용한 이유는 1950년대 말 이후부터 정보 검색 분야에서 적합성에 등급 또는 정도가 존재하므로 이분법적 평가 방법보다는 복수의 등급(degree)을 허용하는 것이 필요하다는 논의가 활발히 전개되어 왔기 때문이다(National Academy of Sciences 1959; Goffman 1964; Hillman 1964; Saracevic 1975). 적합도 평가 시 3점 척도나 5점 척도를 사용할 경우 적합 정도를 보다 섬세하게 표현할 수 있으며, 평가자 간에 발생할 수 있는 관점의 차이나 생각의 차이가 양극화되는 위험을 방지할 수 있다는 장점이 있다(맹성현 외 1999). 이에 정보 검색 분야의 많은 선행 연구들

이 3점 척도나 5 점 척도를 통한 적합도 평가를 수행하여 왔다(노정순 2004; 맹성현 외 1999).

본 연구에서는 이처럼 도출된 평가 기준에 대한 상세한 가이드라인을 작성하였으며, 이 가이드라인에 따라 9,104개의 지식 문서의 질문 제목의 적합도와 전체 질문의 적합도를 2005년 8월에 평가하였다. 문헌정보학 전공자인 평가자들은 평가 전에 각각의 질의들에 대해 이용자가 조회한 문서를 기록한 클릭 로그를 참고하여 이용자의 정보 요구를 파악하고, 질의의 주제 범위와 유의어 등을 확인하였다.

9,104개의 지식인 질문들 중 전체 질문이 적합하다고 판정된 2,177개의 질문과 전체 지식인 컬렉션 중 임의로 선정된 1,111개의 지식인 질문을 합하여 총 3,288개의 지식인 질문에 대한 답변 쌍을 선정하였다. 3,288개의 지식인 질문들에 대한 답변 중 질문자가 “Best Answer”로 선정한 답변들을 대상으로 적합도와 신뢰도를 2005년 10월과 11월에 평가하였다. “Best Answer”만을 분석 대상으로 한 이유는 하나의 질문에 대해 최대 수십 개의 답변들이 게시되지만 대부분의 이용자들이 “Best Answer”를 주로 조회하며, “Best Answer”가 답변의 리스트 중 최상단에 배열되므로 “Best Answer”의 적합도와 신뢰도를 평가하는 것이 중요하다고 판단되었기 때문이다.

4. 평가 기준

4.1 질문 제목의 적합도 평가 기준

일반적으로 통합 검색 결과 화면에는 질문

제목만 노출되고, 이용자는 통합 검색 질의에 적합하지 않은 질문 제목을 조회하지 않기 때문에, 질문 제목이 적합한 문서를 검색하여 상위에 배치하는 것이 중요하다. 이에 본 연구는 지식인 문서의 내용 분석에 근거하여 질문 제목의 적합도 평가 기준을 개발하였다. 대부분의 지식인 문서의 질문 제목은 한 단어에서 한 문장 정도의 길이로 짧은 경우가 많으므로, 질의가 질문 제목에 등장했는지 여부가 적합도 평가의 가장 중요한 기준으로 작용하였다. 또한 질의가 질문 제목에 등장하지 않더라도 질의에 대한 동의어나 유의어가 등장하는 경우 또는 질의와 밀접한 관련성이 있는 내용이 포함되는 경우 “적합”으로 평가하였다. 예를 들어 질의가 “연고전”이고 질문 제목이 “고연전은 언제 해요?”인 경우, 질의가 “sky”이고 질문 제목이 “서울대, 연대, 고대 어떻게 가요?”인 경우가 이에 해당한다. 질의에 대한 질문 제목의 평가 기준으로 “적합,” “보통,” “부적합”의 3점 척도를 사용하기로 하였으며, 구체적인 평가 항목 및 그 예는 아래와 같다.

4.1.1 질문 제목의 “적합” 평가 기준

1. 질의가 질문 제목에 모두 등장하는 경우

질의 : 파란

질문 제목 : 파란닷컴

2. 질의는 등장하지 않지만, 제목이 질의와 밀접한 연관성이 있고 클릭로그에서 큰 비중을 차지하는 경우

질의 : 황우석

질문 제목 : 줄기세포는 정말 없어요?

4.1.2 질문 전체의 “보통” 평가 기준

1. 질의의 일부분만 질문제목에 등장하는 경우
2. 질의의 일부분에 대한 동의어나 유의어가 등장하는 경우

4.1.3 질문 제목의 “부적합” 평가 기준

1. 질의가 질문 제목에 전혀 등장하지 않을 때
2. 질의의 일부분이 등장하긴 하지만, 전체 질의가 등장해야 의미가 있는 경우

질의 : 레인 포레스트

질문 제목 : 포레스트 검프

3. 질의의 일부분이 등장하지만, 초점이 벗어난 단어에 대해 물어볼 때

질의 : 통영 경매 정보

질문 제목 : 통영 국제 영화제(클릭로그에는 경매정보에 대한 내용뿐)

4.2 질문 전체의 적합도 평가 기준

지식인 질문에 대한 내용 분석 결과 이용자들이 질문하는 행태가 매우 다양한 것으로 나타났다. 예를 들어 질문 제목에 질문의 일부를 입력하고 질문 본문에 나머지 질문을 입력하는 경우를 들 수 있다. 즉 질문 제목이 “가요대상 대상 받은 가수” 질문 본문이 “1990년도 2002년도 대상 누가 차지하였는지 알려 주세요”인

경우이다. 또는 실제 질문 내용이 제목이 아닌 질문 본문에만 등장하는 경우도 있었다. 질문 제목이 “이것 좀 알려 주세요” 질문 본문이 “야 후 계정 어떻게 받나요?”와 같은 경우이다. 질의에 대한 질문 제목이 적합하더라도 전체 질문이 부적합한 경우 적합한 문서로 보기에 어려움이 있다. 따라서 질문 제목과 질문 본문을 통합하여 하나의 정보 요구로 보고 질의에 대한 전체 질문의 적합도를 평가하는 것이 필요하다.

질문 전체의 적합도 평가에 있어서 첫째, 질의에 대한 등장 여부나 질의의 유의어나 동의어의 등장 여부가 가장 중요한 평가 기준으로 작용하였다. 이 때 질의의 주제 범위나 유의어, 동의어를 파악하기 위하여 클릭 로그를 참고하였다. 둘째, 질문 대상이 복수 개인 경우, 이 중에서 주어진 질의에 해당하는 질문 대상이 차지하는 비중을 고려하였다. 예를 들어 질의가 “김홍도”이고 질문이 “김홍도, 신윤복, 김두량, 김득신, 장승업(제목) 이 사람들이 만든 작품 내용에 대해 알고 싶습니다(본문)”인 경우 질문 대상 5개 중에서 질의에 해당하는 대상은 1개이므로 이러한 비중을 고려하였다. 본 연구에서는 질의에 대한 질문 전체의 적합도를 평가하기 위하여 “적합,” “보통,” “부적합”의 3점 척도를 사용하였으며, 구체적인 평가 기준과 그 예는 아래와 같다.

4.2.1 질문 전체의 “적합” 평가 기준

1. 질의가 질문 제목과 질문 본문에 모두 등장하며, 전체적인 질문이 질의와 관련이 있을 때

질의 : 야후
 질문 제목 : 우리나라 야후 아이디 가지고
 미국 야후 사용할 수 있나요
 질문 본문 : 만약 안 된다면 미국 야후 아이디
 무료 인가요?

- 2. 질의가 질문 제목에만 등장하고 질문 본문에는 등장하지 않지만, 질문 제목과 본문이 문장으로 연결되면서 전체 질문이 질의와 관련이 있을 때

질의 : 야후
 질문 제목 : 야후 탈퇴하려면
 질문 본문 : 어떻게 해야 하나요?

- 3. 질의가 질문 제목에는 등장하지 않고 질문 본문에만 등장하지만, 전체적인 질문이 질의와 관련이 있을 때

질의 : 야후
 질문 제목 : 이것 좀 알려주세요,
 질문 본문 : 야후 계정 어떻게 받나요?

- 4. 위의 세 가지 경우 중 질문 대상이 복수 개인 경우, 질문 대상 중에서 주어진 질의에 해당하는 질문 대상이 차지하는 비중이 50% 이상일 때

질의 : 야후
 질문 제목 : 네이버 야후 중 무엇이 더 좋은가요?(질문 대상이 2개이며 야후에 대한 부분이 1개)
 질문 본문 : 네이버와 야후 중 어느것이 더

역사가 깊은가요? 그리고 장단점은?

4.2.2 질문 전체의 “보통” 평가 기준

- 1. 질의의 일부가 질문 제목이나 질문 본문 중에서 한 번 이상 등장하지만, 질의의 성격상 전체 질의로 질문 하는 것이 더 적절한 경우

질의 : 마수철학원
 질문 제목 : 경기도 동두천시 근처 철학원 전화번호
 질문 본문 : 경기도 동두천시 근처 철학원 전화번호 알고 싶습니다.

- 2. 질의가 질문 제목이나 질문 본문 중에 한 번 이상 등장하지만 질문 내용에 질의 외의 다른 부수적인 내용이 추가되는 경우

질의 : 다음
 질문 제목 : 인터넷 다음 홈페이지 안 열려요
 질문 본문 : 네이버에서 다음 검색해서 그거 누르면 다음 홈페이지 연결되어야 하는데 잘 안 되요... 어떻게 해야 해요?

- 3. 위의 두 경우 중 질문 대상이 복수 개인 경우, 질문 대상 중에서 주어진 질의에 해당하는 질문 대상이 차지하는 비중이 20% 이상 50% 미만일 때

4.2.3 질문 전체의 “부적합” 평가 기준

- 1. 질의가 질문 제목이나 질문 본문 어디에

도 등장하지 않고, 전체적인 질문이 질의와 전혀 관련 없는 경우

- 2. 전체 질의의 일부 검색어가 질문 제목이나 질문 본문 중에 한 번 이상 등장하지만, 질의의 성격상 전체 질의로 질문했을 때만 의미 있는 경우

질의 : 레인 포레스트

질문 제목 : 포레스트 검프 영화평 부탁드립니다.

질문 본문 : 감상평이 숙제인데 부탁드립니다.

- 3. 질의가 질문 제목이나 질문 본문 중에서 별 의미 없이 한 두 번 등장하는 경우, 또는 질의가 아닌 다른 단어에 초점이 맞춰져 있는 경우

질의 : 다음

질문 제목 : 네이버 카페에서...

질문 본문 : 저는 다음(질문 본문에서 한 번 등장) 안하고 네이버 카페 쓰는 데...

- 4. 질의의 일부가 질문제목이나 질문 본문 중에서 한 번 이상 등장하지만, 질문 전체의 내용이 질의와 관련 없는 경우
- 5. 질의가 질문제목이나 질문본문 중에 한 번 이상 등장하고, 질문 대상이 복수 개인 경우, 질문 대상 중에서 주어진 질의에 해당하는 질문 대상이 차지하는 비중이 20% 미만일 때

4.3 답변의 적합도 평가 기준

통합 검색 질의에 대한 질문 제목과 전체 질문이 적합하더라도 질문에 대한 답변이 부적합한 경우 적합한 문서로 보기에는 어렵다. 따라서 본 연구에서는 질문에 대한 답변의 적합도 평가 기준을 별도로 개발하였으며, 이를 위하여 “적합,” “보통,” “부적합”의 3점 척도를 사용하였다. 답변의 적합도 평가 시 첫째 기준은 질문 제목과 질문 본문을 통합한 전체 질문 내용이 답변에 등장하였는지의 여부이다. 전체 질문 내용이 답변에 등장하고 답변이 질문 내용에 대해 적합한 정보를 제공할 때 답변이 적합한 것으로 평가하였다. 또한 질문 내용이 답변에 등장하지 않더라도 답변이 질문 내용에 대해 적합한 정보를 제공할 경우 적합한 것으로 평가하였다. 둘째, 질문자가 복수의 항목에 대해 질문하는 경우, 답변에서 얼마나 많은 항목에 대해 적합한 답변을 제공하였는지를 고려하였다. 셋째, 질문자가 질문하지 않은 항목에 대해서 답변하는 경우도 있었는데, 이 경우, 전체 답변 항목 중에서 질문 내용에 대해 적합한 답변이 차지하는 비중을 고려하였다. 답변의 구체적인 평가 기준과 예는 아래와 같다.

4.3.1 답변의 “적합” 평가 기준

- 1. 질문 내용이 답변에 모두 등장하고, 답변이 질문 내용에 대해 적합한 정보를 제공할 때

질문 제목 : 스카이 라이프 해지

질문 본문 : 스카이 라이프 해지하려고 하는데 어떻게 하나요?

답 변 : 스카이라이프를 시청하시다 해지를 하시려면 위약금과 철거비 2만원 내셔야 합니다. 스카이라이프 해지 문의는 스카이라이프고객센터 (1588 - ****)로 문의 하시면 됩니다.

2. 질문 내용이 답변에 등장하지는 않지만 답변이 질문내용에 대해 적합한 정보를 제공할 때

질문 제목 : 다음카페 다음 넘어가지 않아요.
질문 본문 : 다음카페 메인화면에서 게시판으로 넘어가지 않아요.

답 변 : 저는 도구로 인터넷 옵션에 들어가서 보안 상태를 모두 최소로 했거든요. 그랬더니 잘 되더군요.

3. 질문자가 복수의 항목에 대해 질문하는 경우, 답변에서 질문 항목의 60% 이상을 충족하는 적합한 답변을 제공하는 경우

질문제목 : 덕성여대가 궁금해요.
질문본문 : 덕성여대 어떻게 가요? 덕성여대 학생들이 예쁘다던데 정말인가요? 덕성여대의 맛있는 음식점은 뭐가 있어요?(질문 3개)

답 변 : 덕성여대는 수유역에서 내려서 마을버스02번 타고 10분가면 금방가요. 그리고 덕성여대학생들은 대학설문조사 “미”기준에서 1위를 차지하였습니다.(답변 2개 : 66.6%)

4. 질문자가 질문하지 않은 항목에 대해서도

답변하는 경우, 전체 답변 항목 중에서 질문 내용에 대한 적합한 답변이 차지하는 비중이 60%이상일 때

질문 제목 : 덕성여대와 동덕여대 중 어디가 더 좋아요

질문 본문 : 덕성여대 동덕여대 한 번도 안가봐서 몰라서 그런데 캠퍼스랑 성적 뭐 이런 거 어때요?

답 변 : 덕성여대는 이렇고, 동덕여대는 이렇고 서울여대는 저렇다.(비중 : 66.6%)

4.3.2 답변의 “보통” 평가 기준

1. 질문내용의 일부에만 해당하는 답변을 할 경우

질문 제목 : 통영의 경매정보
질문 본문 : 통영에 시는데, 통영에서 이뤄지는 경매정보에 대해서 궁금해요. 관련 사이트 같은 곳 알려주세요.

답 변 : 통영은 잘 모르겠고 경매정보와 관련된 좋은 사이트는 무엇 무엇이 있다. (클릭로그를 보면 경매 사이트가 주를 이룸)

2. 질문자가 복수의 항목에 대해 질문하는 경우, 답변에서 질문 항목의 30%이상 60% 미만을 충족하는 적합한 답변을 제공하는 경우

3. 답변자가 질문자가 질문하지 않은 항목에 대해서도 답변하는 경우, 전체 답변 항목 중에서 질문 내용에 대한 적합한 답변이 차지하는 비중이 30-60%에 해당할 때

4.3.3 답변의 “부적합” 평가 기준

1. 질문 내용이 답변에 등장하지 않으며, 답변이 질문과 전혀 관련 없는 내용을 다룰 때

질문 제목 : 꿈 해석 부탁

질문 본문 : 제가 덕성여대 다니는데 꿈에 저희 학교가 나오는 거예요... 근데 그곳에서 어떤 귀인을 만나...꿈 해석좀 해주세요.

답변 : 보통 여대라고 하면 많은...(여대에 대한 설명)

2. 질문 내용의 일부에 대한 정보를 제공하지만 답변이 질문 내용의 초점에서 벗어났을 때

질문 제목 : 꾸러기 많이 사용하나요

질문 본문 : 야후에 꾸러기 있잖아요. 그거 초등학교 사이에 인기 있잖아요

답변 : 제 동생이 초등학교인데요, 어찌나 장난꾸러기인지...동생을 장난꾸러기로 소개(클릭 로그에는 야후의 “꾸러기” 사이트, TV의 “꾸러기” 만화 프로그램이 대부분임)

3. 질문자가 복수의 항목에 대해 질문하는 경우, 답변에서 질문 항목의 30% 미만을 충족하는 답변을 제공하는 경우
4. 질문자가 질문하지 않은 항목에 대해서도 답변자가 답변하는 경우, 전체 답변 항목 중에서 질문 내용에 대한 적합한 답변이 차지하는 비중이 30% 미만인 해당할 때

4.4 답변의 신뢰도 평가 기준

지식 검색 서비스의 대중적 인기에도 불구하고, 검증되지 않은 답변이나 주관적인 답변, 출처가 불분명한 답변으로 인해 답변의 질, 신뢰성, 전문성이 떨어진다는 비판이 제기되어 왔다. 또한 비방, 욕설, 유해 정보, 음란 정보, 광고성 글들이 질문이나 답변으로 게시되는 점도 지식인 서비스가 개선해야 할 사항 중의 하나이다. 따라서 답변의 신뢰도 평가 기준을 구축하여 신뢰도가 낮은 답변을 판별하는 것이 시급하다.

한편 지식 검색 데이터베이스에 등록된 질문들은 생활 상식이나 신변잡기적인 내용을 묻는 범주와 전문적인 지식을 필요로 하는 범주로 구분될 수 있다. 박주범, 정동열의 연구(2004)에서도 253명의 지식 검색 이용자들에게 자주 이용하는 지식의 내용을 문의한 결과 생활 상식 관련 지식(38.3%)과 전문지식(36.4%)이 가장 많은 것으로 나타났다. 생활 상식 관련 질문과 전문 지식 관련 질문에서 요구되는 답변의 수준이 다르므로, 본 연구에서는 지식 검색 데이터베이스의 질문들을 지식형 질문과 생활형 질문으로 구분하여, 질문 유형에 따라 신뢰도 평가 항목을 차별화하여 적용하였다. 좀 더 구체적으로 지식형 질문이란 질문자가 전문적 지식, 정확한 근거가 필요한 객관적 지식을 요구하는 경우이며, 그 예로는 “배아줄기 세포 이식법으로 가능한 유전병은?” “상장 폐지된 주식은 어떻게 됩니까?” 등을 들 수 있다. 생활형 질문이란 질문자가 전문 지식이 필요 없는 생활 상식이나, 개인적인 의견을 묻는 경우이며, 그 예로는 “본 프레레 어떻게 생각해요?”

“연애 잘하는 법 좀 알려주세요.” “어떻게 해야 공부 좀 잘할 수 있죠?” “공짜로 음악 다운받을 수 있는 곳은?” 등을 들 수 있다.

본 연구에서는 “높음,” “보통,” “낮음”의 3 점 척도를 사용하여 답변의 신뢰도 평가를 위한 기준을 개발하였다. 구체적인 답변의 신뢰도 평가 기준은 <표 1>과 같으며, 이 기준의 평가 항목 중 하나 이상을 만족시키면 기준에 부합하는 것으로 평가하였다. 지식형 질문의 경우 객관적으로 확실한 근거나 출처를 밝혀 주는 답변을, 생활형 질문의 경우에는 상식적이면서 논리적인 전개가 뛰어난 답변을 신뢰도가 높은 것으로 평가하였다. 지식형 질문과 생활형 질문의 경우 모두 답변의 내용이 비방, 욕설, 음란성 글이거나 질문과 전혀 관계가 없고 의미가 없는 답변의 경우에는 신뢰도를 낮게 평가하였다. 또한 추측성 답변이나, 출처가 제시되지 않고 근거 없는 주관적 의견의 경우에도 신뢰도를 낮게 평가되었다.

5. 평가 결과

5.1 질문 제목 적합도 평가 결과

본 연구에 포함된 9,104개의 지식인 질문 중 제목이 적합하다고 평가된 질문은 전체의 25.3%인 2303개였으며, 보통으로 평가된 질문은 1134개(12.5%)였으며, 부적합하다고 평가된 질문은 5667(62.2%) 개였다. 제목이 적합한 것으로 판정된 질문의 수가 비교적 적은 것으로 나타났다. 한편 질의의 주제 범위가 “다음”이나 “박스”와 같이 광범위하고 일반적인 경우 질문 제목이 “적합”으로 평가받는 비중이 높았고, 질의의 주제 범위가 구체적일수록 “부적합”으로 평가받는 비중이 높았다.

5.2 질문 전체의 적합도 평가 결과

9,104개의 지식인 질문 중 제목과 본문을 통합하여 질문 전체가 적합하다고 평가된 질문은 2177개(23.9%)였으며, 보통으로 평가된 질문

<표 1> 답변의 신뢰도 평가 기준

신뢰도 정도	지식형	생활형
높음	<ul style="list-style-type: none"> • 공신력 있는 정확한 출처 • 객관적으로 확실한 근거(이론적 학문적인 예시 등) • 논리적 설명 • 자료 첨가(표, 그림, 사진) • 질문에 대한 핵심 답변 	<ul style="list-style-type: none"> • 논리적인 개인의견 • 속담, 격언, 생활지식 등 학문적 근거는 없지만 상식적 • 질문에 대한 핵심 답변
보통	<ul style="list-style-type: none"> • 답변을 하였으나 근거부족 	<ul style="list-style-type: none"> • 정확한 출처가 나오지는 않고 답변자의 의견에 의존하나 어느 정도 논리적임
낮음	<ul style="list-style-type: none"> • 비방, 욕설 음란한 글 • 명예훼손성 글 • 추측성 답변 • 근거가 없는 개인의견 • 질문과는 전혀 관련 없는 답변 	

은 651개(7.2%)였으며, 부적합하다고 평가된 질문은 6276개(68.9%)였다. 또한 전체 9,104개의 질문의 분포를 질문 제목의 적합도와 전체 질문의 적합도에 따라 살펴보면 <표 2>와 같다. 곧 질문 제목과 전체 질문이 모두 적합한 문서는 총 질문의 18.68%로, 질문 제목과 전체 질문이 모두 부적합한 문서의(56.35%) 약 3분의 1에 해당하는 것으로 나타났다.

5.3 답변의 적합도 평가 결과

본 연구에서는 9,104개의 지식인 질문들 중 전체 질문이 적합하다고 판정된 2,177개의 질문과 전체 지식인 컬렉션 중 무작위로 선정된 1,111개의 지식인 질문을 합하여 총 3,288개의 지식인 질문에 대한 답변 쌍을 선정하였다. 2,177개의 질문들에 대한 답변의 적합도 분포가 무작위로 선정된 질문 그룹에도 적용되는지를 조사하기 위하여 전체 지식인 컬렉션에서 무작위로 선정된 질문을 분석에 추가하였다. 3,288개의 지식인 답변들 중 695개에 대해서는 분석이 불가능하였다.

분석이 불가능한 이유로는 첫째, 일부 질문들에 대해서는 답변이 없거나 또는 틀린 문서가 나타났다기 때문이다. 둘째, 지식인 취지에 맞지 않는 광고성 글, 오락성 글, 난센스 퀴즈를 물어보는 질문들, 동정성 답변 일반 게시판에 올리는 유형의 글들이 게시되는 경우가 있기 때문이다. 3,288개의 질문들 중 이처럼 분석이 불가능한 문서들을 제외한 후 남은 2,593개의 질문들에 대한 답변 쌍 중 “Best Answer”의 적합도를 평가하였다. 2,593개의 답변들 중 적합하다고 평가된 답변은 2,409개(92.9%), 보통으로 평가된 답변은 105개(4.05%), 부적합으로 평가된 답변은 79개(3.16%)로 나타났다. 전반적으로 질문에 대한 지식인 답변의 적합도는 높게 나타났다.

지식인 답변 적합도의 분포를 보다 세밀하게 분석하기 위하여 전체 질문이 적합한 질문 그룹, 질문 제목과 전체 질문이 모두 적합한 질문 그룹, 무작위로 선택된 질문 그룹을 분리하여 답변의 적합도를 별도로 분석하였다. <표 3>에 따르면 그룹별로 답변의 적합도 분포가 유사한 것으로 나타났다. 이는 답변 적합도의 분포가

<표 2> 지식인 질문의 분포

질문 제목	전체 질문		
	적합	보통	부적합
적합	1,701(18.68%)	185(2.03%)	417(4.58%)
보통	160(1.76%)	246(2.71%)	728(8.0%)
부적합	316(3.5%)	220(2.40%)	5,131(56.35%)

<표 3> 지식인 답변 적합도의 분포

	적합	보통	부적합	총계
질문 전체가 적합한 데이터	1,569(92.23)	73(4.35)	58(3.42)	1,700
무작위로 선택된 데이터	840(93.96)	32(3.58)	21(2.35)	894
질문 제목과 전체 질문이 적합한 데이터	1,213(91.75)	55(4.16)	54(4.08)	1,322

질문 표본 추출 방법에 크게 영향 받지 않음을 시사한다.

5.4 답변의 신뢰도 평가

본 연구에서 분석된 2,593개 답변들 중 신뢰도가 높은 것으로 평가된 답변은 1,701개(65.6%)였으며, 보통으로 평가된 답변은 606개(23.37%), 낮은 것으로 평가된 답변은 286개(11.03%)로 나타났다. 지식인 답변들 중 “Best Answer” 만을 분석 대상으로 하였다. 이를 고려하면, 답변의 신뢰도가 높지는 않은 것으로 보인다.

한편 지식인 답변에 대한 내용 분석을 통하여 신뢰도가 높은 답변들과 낮은 답변들의 특징이 아래와 같이 조사되었다. 전반적으로 신뢰도가 높은 답변들은 아래와 같은 특징을 지닌 것으로 나타났다.

- 서술형 보다는 소항목으로 나누어 체계적으로 설명한 경우
- 전문용어를 사용하는 경우(의학용어, 컴퓨터용어, 경제학 용어 등)
- 길이가 긴 답변(예외도 있음)
- 추가적으로 참고 사이트까지 제시한 경우
- 적절한 정보와 함께 자신의 경험담을 첨가하는 경우

신뢰도가 낮은 답변들의 경우 아래와 같은

특징을 지닌 것으로 나타났다.

- 이모티콘, 욕설, 은어를 과도하게 사용하고, 맞춤법이 틀린 경우
- 답변의 길이가 너무 짧은 경우
- 답변 끝에 ‘허접한 답변 죄송합니다.’ ‘그냥 썼는데 괜찮을지’와 같이 자신의 답변에 대해 지나치게 겸손한 말 사용한 경우
- 답변의 내용이 지나치게 긴 경우: 답변의 내용이 논리정연하게 배열되지 못하고 길이만 긴 경우는 신뢰도가 떨어진다. 이런 경우 대체로 질문 본문이 지나치게 길어서 무엇을 질문하는지 파악하기 힘든 경우가 많았다.
- 추측형 문장이 들어간 경우: 답변자가 “~ 하지 않을 까요”라고 하거나 “~일 것 같아요” 등의 추측성 내용을 쓴 경우는 신뢰도 평가 시 낮게 평가되었다.

5.5 답변의 적합도와 신뢰도의 관계

답변의 적합도와 신뢰도 간의 관계를 조사하기 위하여 연구에 포함된 2,593의 지식인 답변을 대상으로 적합도 범주별로 신뢰도의 분포를 조사한 결과는 <표 5>와 같다. 적합한 답변의 경우 신뢰도가 높은 경우가 많았고, 부적합한 답변의 경우 신뢰도도 낮은 경우가 많았다.

<표 4> 지식인 답변 신뢰도 분포

	높음	보통	낮음	총계
질문 전체가 적합한 데이터	1,099(64.65)	393(23.12)	208(12.24)	1,700
무작위로 선택된 데이터	602(67.45)	213(23.91)	78(8.75)	891
질문 제목과 전체 질문이 적합한 데이터	844(63.84)	299(22.62)	179(13.54)	1,322

〈표 5〉 답변의 적합도와 신뢰도에 대한 분할표

	답변 신뢰도 높음	답변 신뢰도 보통	답변 신뢰도 낮음	계
답변 적합	1,655	546	208	2,409
답변 적합 보통	35	51	19	105
답변 부적합	11	9	59	79
계	1,701	606	286	2,593

귀무가설: 지식 문서 답변의 적합도와 신뢰도는 독립적이다.

연구가설: 지식 문서 답변의 적합도와 신뢰도는 관계가 있다.

이 데이터를 이용하여 Two-way 카이제곱법을 적용한 결과, 답변의 적합도 범주별로 신뢰도의 분포에 통계적으로 유의한 차이가 있는 것으로 나타났다. $\chi^2(4, N=2593)=393.746$, $p < 0.001$. 곧 답변의 적합도는 신뢰도에 영향을 미치며, 신뢰도는 적합도에 영향을 미친다고 볼 수 있다. 이는 지식 문서 테스트 컬렉션 구축 시 적합도 테스트 컬렉션과 신뢰도 테스트 컬렉션이 배타적이지 않으며, 밀접한 관련성을 가짐을 시사한다.

6. 결론

본 연구에서는 지식 검색 서비스 개선을 위하여 지식 검색 데이터베이스를 구성하는 문서 평가 기준을 개발하였다. 또한 이러한 평가 기준에 근거하여 네이버 지식인을 대상으로 지식 문서 질문 제목의 적합도, 전체 질문의 적합도, 지식 문서 답변의 적합도 및 신뢰도를 분석, 평가하였다.

본 연구의 조사 결과 첫째, 질의에 대비하여

제목이 적합하다고 평가된 질문은 전체의 25.3%로 비교적 적은 편이었다. 제목이 보통으로 평가된 질문은 12.5%, 부적합하다고 평가된 질문은 62.2%개로 나타났다. 또한 분석된 지식인 질문 중 제목과 본문을 통합하여 질문 전체가 적합하다고 평가된 질문은 23.9%, 보통으로 평가된 질문은 7.2% 부적합하다고 평가된 질문은 68.9%로 조사되었다. 질문 제목과 전체 질문이 모두 적합한 문서는 총 질문의 18.68%로, 질문 제목과 전체 질문이 모두 부적합한 문서의(56.35%) 약 3분의 1에 불과한 것으로 나타났다. 따라서 지식인 데이터베이스 검색 시 제목과 전체 질문이 모두 적합한 질문을 검색하기 위한 알고리즘의 개선이 필요할 것으로 보인다.

둘째, 본 연구에서 분석한 2,593개 답변들 중 적합하다고 평가된 답변은 92.9%, 보통으로 평가된 답변은 4.05%, 부적합으로 평가된 답변은 3.16%로 나타났다. 신뢰도가 높은 것으로 평가된 답변은 65.6%, 보통으로 평가된 답변은 23.37%, 낮은 것으로 평가된 답변은 11.03%로 나타났다. 지식인 질문에 대한 답변의 적합도는 대체로 높은 것으로 나타났으나 답변의 신뢰도는 떨어지는 것으로 나타났다. 신뢰도가 낮은 답변들의 비중이 10%가 넘는 것은 유해정보, 음란정보, 추측성 답변들에 부분적으로 기인하는 것으로 보인다. 신뢰도가 높은 문서들을 판별하기 위하여 유해정보, 음란정보 등

에 대해서는 모니터링이 필요할 것으로 보인다. 본 연구에서는 답변의 신뢰도가 낮은 문서들과 높은 문서들의 특징을 내용 분석을 통하여 도출하였다. 이러한 특징들에 근거하여 문서의 신뢰도를 예측할 수 있다면 검색의 질과 이용자의 만족도를 개선할 수 있을 것으로 기대된다. 본 연구자들은 현재 지식 문서의 길이, 조회 수, 질문자의 평, 전문가의 평 등과 같은 문서의 자질(feature)들에 근거하여 문서의 신뢰도를 예측할 수 있는 검색 알고리즘을 개발, 시험 중이다.

셋째, 답변의 적합도와 신뢰도 간의 관계가 통계적으로 유의한 것으로 나타났다.

넷째, 지식인 취지에 맞지 않는 질문이나 답변들이 게시되는 경우가 드물게 존재하는 것으로 나타났다. 즉 광고성 글, 오락성 글, 난센스 퀴즈들, 동정성 답변, 일반 게시판에 올리는 유형의 글들이 게시되는 경우가 있었고, 향후 적극적으로 지속적인 모니터링을 통하여 이러한 문서들을 판별하는 것이 요청된다.

이러한 본 연구의 결과는 지식 검색 테스트 컬렉션 구축과 답변의 신뢰도 컬렉션 구축에 직접적으로 활용되어 향후 지식 검색 서비스 개선에 기여할 것으로 기대된다. 테스트 컬렉션은 질의 집합, 문헌 집합, 적합성 판단 집합으

로 구성되며, 테스트 컬렉션은 시스템 혹은 관련 기술의 성능 평가를 통하여 시스템의 발전과 정보 유통에 있어서 경쟁력 강화에 필수적인 역할을 한다. 위에서 언급되었듯이 지식 문서 답변의 신뢰도 측정을 검색 기법 개발에 반영한다면 보다 질 높은 검색 결과를 이용자에게 제공할 수 있을 것이다.

한편 본 연구에서는 하루 동안 입력된 질의에 대한 지식인 질문과 답변들의 평가를 수행하였다. 보다 장기간에 걸쳐 입력된 다양한 질의들을 대상으로 평가 작업을 수행한다면, 본 연구의 결과의 타당도가 향상될 수 있을 것으로 기대된다. 또한 향후 연구에서는 네이버 외에 타 포탈 업체들의 지식 검색 문서들에 대한 평가 및 비교, 분석이 필요할 것으로 보인다. 후발 지식 검색 서비스들의 경우 네이버의 지식인을 벤치마킹하여 매우 유사한 문서 구조로 구성되어 있으므로 본 연구에서 제시한 평가 기준이 적용될 수 있을 것으로 보인다. 마지막으로, 이용자들의 지식 검색 서비스에 대한 만족도나 개선이 필요한 사항 등을 심층적으로 파악하기 위해서는 실제 이용자들을 대상으로 한 인터뷰, 관찰 실험 등의 별도의 연구가 필요할 것으로 보인다.

참 고 문 헌

노정순. 2004. Invisible Web 탐색도구의 성능 비교 및 분석. 『정보관리학회지』 21(3): 203-225.

맹성현, 이석훈, 이준호, 이응봉, 송사광. 1999. 정보 검색 시스템 평가를 위한 균형 테스트 컬렉션 구축. 『정보관리학회지』 16(2):

- 135-148.
- 박소연, 이준호. 2005. 국내 웹 이용자의 검색 행태 추이 분석. 『한국문헌정보학회지』, 39(2): 147-160.
- 박주범, 정동열. 2004. 지식검색서비스 이용에 관한 실증적 연구. 『정보관리학회지』, 21(3): 83-98.
- 이두영, 강순희. 2003. 웹 검색엔진의 지식검색커뮤니티 기반 지식관리에 관한 연구. 『제 10회 한국정보관리학회 학술대회 논문집』: 143-152.
- 최상희. 2005. 문서구조를 이용한 질의응답문서 클러스터링에 관한 연구. 『한국문헌정보학회지』, 39(4): 105-118.
- Bivens-Tatum, R. W. 2001. "Expert services on the Web: the commercial competition for libraries." *College & Research Libraries News*, 62(7): 714-716.
- Goffman, W. 1964. "On relevance as a measure." *Information Storage and Retrieval*, 2(3): 201-203.
- Hillman, D. J. 1964. "The Notion of Relevance(1)." *American Documentation*, 15(1): 26-34.
- Janes, J., Hill, C., and Rolfe, A. 2001. "Ask-an-expert services analysis." *Journal of the American Society for Information Science and Technology*, 52(13): 1106-1121.
- Kleinberg, J. M. 1999. "Authoritative sources in hyperlinked environment." *Journal of the ACM*, 46(5): 604-632.
- Moon, I., and Woyke, E. 2006. "NHN: The little search engine that could." *Businessweek*, 30 January.
- National Academy of Sciences. 1959. *Proceedings of the International Conference on Science Information*, 2 Vols. National Academy of Sciences, Washington, D. C.
- Pack, T. 2000. "Human search engines: the next killer app." *EContent*, 23(6): 16-22.
- Saracevic, T. 1975. "Relevance: A review of and a framework for the thinking on the notion in information science." *Journal of the American Society for Information Science*, 26(6): 321-343.
- Strong, D. M., Lee, Y. W., & Wang, R. Y. 1997. "Data quality in context." *Communications of the ACM*, 40(5): 103-110.
- Zhou, Y., and Croft, W. B. 2005. "Document quality models for web ad hoc retrieval." in *Proceedings of the ACM Fourteenth Conference on Information and Knowledge Management*, 331-332.
- Zhu, X., & Gauch, S. 2000. "Incorporating quality metrics in centralized/ distributed information retrieval on the world wide web." In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 288-295.