유사문헌집단에서 적합/부적합정보의 유용성에 관한 연구*

A Study on the Utility of Relevance/Non-relevance Information in Homogeneous Documents

문성빈 (Sung-Been Moon)**

초 록

본 논문에서는 문헌의 적합성수준을 적합성정도에 따라 4그룹(부적합한, 조금 적합한, 적합한, 매우 적합한)으로 나눈 후 서로 다른 심사자가 적합성 판정을 내린 4개의 적합성 판정세트(A, B, C, D)에서 "조금 적합한" 문헌을 부적합문헌으로 분류했을 때와 적합문헌으로 분류하였을 때에, 초록/표제 시스템과 전문검색시스템에서 적합성피드백으로 인한 검색효율성의 증진은 어느 쪽이 더 혜택을 받게 되는 지를 연구하였다. "조금 적합한" 문헌을 적합문헌으로 포함시켰을 때 초록/표제시스템이 전문검색시스템보다 모든 적합성판정세트에서 검색효율성의 증가율이 높았고, 반면에 전문검색시스템에서는 "조금 적합한" 문헌을 적합문헌그룹에서 제외시켰을 때 검색효율성의 증가율이 일관성 있게 높아지는 것을 발견하였다. 이는 전문검색시스템에서는 적합문헌으로 포함된 "조금 적합한" 문헌으로부터 얻어지는 적합성피드백 정보는 잡음의 역할을 하게 되어 검색효율성의 증진에 도움이 안 되고 있음을 암시하고 있다. 특히, 매우 동질적인 문헌을 색인 및 검색대상으로 하고 있는 전문검색시스템에서는 잡음에 의해 초래되는 낮은 정확률을 개선하는 정교한 검색기법에 대한 연구가 지속되어야만 한다.

ABSTRACT

This study examined the relative retrieval effectiveness after relevance feedback between two systems (Title/Abstract and Full-text) using four different sets of relevance judgment. Four relevance levels (not relevant, marginally relevant, relevant, highly relevant) are also used, each of which is determined by referees giving a relevance score to documents. This study also investigated how much the average precision was improved after relevance feedback when "marginally relevant" documents are included in the relevant class with the Title/Abstract system, and with the Full-text retrieval system as well. It is found that the Title/Abstract system benefited from relevance feedback with the marginally relevant documents. In case of the Title/Abstract system, the higher percentage of improvement was consistently obtained when including the marginally relevant documents in the relevance class, however the result was vice versa in case of the Full-text retrieval system. It implied that the marginally relevant documents in the relevant class had caused noises in the Full-text retrieval system.

키워드: 적합성, 적합성판단, 적합성피드백, 전문검색시스템, 검색효율성
relevance, relevance judgment, relevance feedback, full-text retrieval system, retrieval effectiveness

---

* 이 논문은 연세대학교 학술연구비의 지원에 의해 연구되었음.

** 연세대학교 문헌정보학과 교수(sbmoon@yonsei.ac.kr)

■ 논문접수일자 : 2015년 8월 19일 ■ 최초심사일자 : 2015년 8월 29일 ■ 게재확정일자 : 2015년 9월 7일
■ 정보관리학회지, 32(3), 277-293, 2015. 〔http://dx.doi.org/10.3743/KOSIM.2015.32.3.277〕

1. 서 론

지난 수십 년 동안 적합성 개념에 대한 연구는 끊임없이 지속되어 오고 있다. 특히, 정보검색시스템의 검색효율성은 정확률과 재현율로 측정되는데, 적합성 개념은 이러한 평가척도를 위한 기반으로서 중요한 개념으로 자리매김하고 있다. 일반적으로 실험실 환경에서 행해지고 있는 정보검색실험은 검색시스템, 문헌, 질의어, 문헌과 질의 간의 적합성 판정 세트로 구성되어 있다. 문헌과 질의 간의 적합성 판정은 일반적으로 주제전문가 또는 실질적인 정보시스템 이용자에 의해서 이루어지기도 한다. 그러나, 적합성 판정에 영향을 미치는 요소들은 너무 다양하기 때문에 이들 요소들에 대한 연구 또한 끊임없이 지속되고 있다. 즉, 누구에 의해서 판정이 이루어졌는가 뿐만 아니라, 문헌의 어떤 구성요소를 판정의 기준으로 삼았는가에 따라 적합성 판정결과는 달라지기도 한다. 질의어와 문헌간의 직접적인 적합성 관계이외에도 다른 외부요인들이 적합성 판정에 영향을 주기도 한다는 연구결과도 보고되고 있다(Eisenberg, 1988; Eisenberg & Barry, 1988, Xu & Wang, 2008). 또한 적합성 판정을 위한 기준의 다양성은 검색효율성의 평가에 중요한 영향을 미치는 것으로 보고되고 있다(Vakkari & Sormunen, 2004; Xu & Chen, 2006).

본 연구에서 사용되고 있는 문헌과 질의 간의 적합성 판정은 전문(full-text)을 대상으로 이루어졌으며 색인 및 검색도 전문을 대상으로 행해졌다. 정보기술의 급격한 발전은 적은 비용으로 대용량 전문 데이터베이스의 구축을 가능하게 하였으며 전문검색시스템은 정보이용자가 자연언어로 전문을 대상으로 탐색하게 함으로써 표제, 초록, 통제언어를 이용해 탐색하는 다른 검색시스템의 대체시스템으로 인기를 모으고 있다. 전문검색시스템은 문헌에 할당된 색인어의 수에 의해 측정될 수 있는 색인의 망라성을 최대화시킴으로써 재현율을 증진시킨다. 즉, 문헌에 할당된 색인어를 증가시킨다는 것은 질의어가 검색대상이 될 수 있는 문헌에 출현할 수 있는 확률을 높여 준다는 것이다. 그러나, 전문을 이용한 문헌의 표현은 검색에 유용한 색인어 뿐 만 아니라 검색과정의 잡음(noise)이 될 수 있는 비주제색인어도 추가하게 되는 것이다(Blair & Maron, 1990). 이러한 잡음은 특히 주제가 매우 동질적인 문헌으로 구성되어 있는 실험문헌집단에서는 정확률을 떨어뜨리는 결정적인 요인이 되고 있다. 본 연구에서는 이러한 잡음의 영향을 최소화시키기 위해 적합문헌에서 자주 출현하는 질의어의 가중치를 높여 줄 수 있는 적합성 피드백 기법을 사용하였다.

일반적으로 적합성판단의 수준은 이진수준(적합 또는 부적합)에 의해 이루어지고 있으나, 몇몇 연구자들은 적합성의 수준을 "부적합한", "조금 적합한", "적합한", 그리고 "매우 적합한" 등으로 구분하여 이러한 적합성수준 및 이를 위한 기준에 대한 이용자 관점을 분석하였으며, 특히 "조금 적합한"이라는 적합성수준이 무엇을 의미하는지 그리고, 이를 위해 어떤 기준이 적용되었는지에 대해 보고하였다(Shamber, 1994; Spink et al., 1998; Voorhees, 2001; Maglaughlin & Sonnenwald, 2002). Belkin(1984)은 적합성 판정에 영향을 줄 수 있는 상황적인 요소들은: 1) 이용자가 찾고 있는 주제, 즉 주제성

(topicality), 2) 검색된 정보를 이미 가지고 있는 경우, 3) 이용자의 지식의 상태 또는 문제점에 대한 신뢰, 그리고 4) 이용자의 목적 또는 의도 등이라고 언급하고 있다. Shaw et al.(1991)은 전문가그룹에 의해 신중하게 이루어진 적합성판정이야 말로 검색결과에 대한 의미있는 평가를 가능하게 할 것이라고 주장하고 있다.

본 연구에서는 "조금 적합한" 문헌으로 분류된 문헌들을 처음에는 부적합문헌으로 분류하고 이들의 적합성수준에 따라 단계적으로 적합문헌그룹에 포함시킴으로써 적합문헌의 수를 증가시켰을 때 초록/표제를 이용한 시스템과 전문(full-text)을 사용한 검색시스템에서 평균정확률의 증감을 측정하였다. 이를 통해 부적합문헌으로 분류되었던 "조금 적합한" 문헌이 적합문헌그룹으로 포함되었을 때 그 결과가 두 시스템의 검색효율성에 미치는 영향을 연구하였다.

2. 이론적 배경

2.1 적합성 개념

적합성의 개념에 대한 구체적인 논의는 Saracevic(1975)과 Swanson(1986)에 의해 이루어져 왔고, 그 이후 여러 학자에 의해 다양한 시각에서 재조명되어 왔다(Schamber et al., 1990; Harter, 1992; Schamber, 1994; Mizzaro, 1997; Borlund, 2003; Saracevic, 2007a; Saracevic, 2007b; Huang & Soergel, 2013). 적합성 개념은 일반적으로 주관적 또는 객관적 개념으로 나누고 있으며, 이를 각각 이용자 지향적 또는 시스템 지향적이라고 부르곤 한다(Swasnson, 1986). 시스템 지향적 관점은 이용자의 정보요구가 표현된 정적인 질의어와 문헌간의 주제적 관련성(topicality)에 의해 결정되는데 이는 질의에 출현하는 용어와 문헌을 대표하는 색인어와의 일치정도에 의해 결정된다. 반면에 이용자 지향적인 관점은 적합성 개념을 정적인 개념으로 보지 않고 동적인 개념으로 보고 있다. 즉 적합성 판단은 이용자의 인지적 또는 상황적 요인에 의해 지배받을 수 있다는 것이다(Shamber et al., 1990; Harter, 1996).

또한 검색효율성에 영향을 미치는 적합성 판단을 위한 기준(relevance criteria), 적합성수준(relevance level)에 대해서도 끊임없이 논의되어 오고 있다(Harter, 1996; Spink et al., 1998; Tang et al., 1999; Spink & Greisdorf, 2001; Kekäläinen & Järvelin, 2002; Maglaughlin & Sonnenwald, 2002; Sormunen, 2002; Greisdorf, 2003; Vakkari & Sormunen, 2004; Xu & Chen, 2006; Xu & Wang, 2008). 그리고, 상이한 적합성 판정세트가 검색효율성에 미치는 영향에 대한 연구도 행해져 왔지만 그다지 영향이 미치지 않는 것으로 알려져 왔다(Burgin, 1992; 문성빈, 1997).

Vakkari & Sormunen(2004)은 이용자와의 상호작용, 즉 적합성피드백을 통해서 적합성 수준(relevance levels)이 검색효율성에 미치는 영향을 연구하였고, 적합성피드백을 위한 질의확장의 효과는 "매우 적합한" 문헌을 얼마만큼 성공적으로 식별하고 검색해 낼 수 있는가에 달려 있다고 보고하고 있다. 일반적으로 적합성의 개념은 단순히 주제적 관련성만을 보고

판단하는 “적합 또는 부적합” 수준으로 측정되고 있다. TREC(Text REtrieval Conference)에서의 적합성 판단을 위한 지침에서도 “적합” 또는 “부적합”으로만 측정하도록 되어 있으며 일반적으로 문헌의 극히 일부분만이라도 주제와 관련이 있다고 판단되면 그 문헌은 적합한 문헌으로 간주되고 있다.

Sormunen(2002)은 적합성 판정기준을 확장시켜 문헌이 주제를 다루고 있는 적합성 정도(topical relevance degree)에 따라 셋으로 구분하여 조작적 정의를 내리고 있다. “조금 적합한(marginally relevant)” 문헌은 단지 주제를 가리키고 있을 뿐인 것이고 “보통 적합한(fairly relevant)” 문헌은 더 많은 정보를 포함하곤 있지만 망라적이지는 못한 것인 반면에 “매우 적합한(highly relevant)문헌은 주제의 핵심을 망라적으로 논의하고 있는 것으로 보고 있다. 그는 TREC7/TREC8에서 사용된 적합문헌들의 적합성 평가을 다시 시도해보았는데, 50%의 적합문헌은 “조금 적합한”, 34%가 “보통 적합한”, 16%가 “매우 적합한”으로 판단되었다. TREC에 의해 적합하다고 판단된 문헌의 25%가 위에서 언급한 적합성 기준에 의해서는 부적합한 문헌으로 판정되었다고 보고하고 있다.

“매우 적합한” 문헌들은 주제와 연관된 단어들은 더 많이 포함하고 있으며, 일반적으로 저자들은 논의하고자 하는 개념을 언급하기 위해 여러 가지 표현을 쓰고 있다. 이는 “조금 적합한” 문헌들은 “매우 적합한” 문헌들보다 탐색주제와는 관련 없는 텍스트(문단, 문장, 단어)를 더 많이 포함하고 있음을 암시하고 있는 것이다(Sormunen et al., 2001). “조금 적합한” 문헌은 이용자로 하여금 정보요구와의 주제적 유사성을 잘 인지하지 못하게 하며 검색실험대상 문헌집단내에서 적합문헌의 분포는 대부분 “조금 적합한” 문헌들 쪽으로 치우쳐 있는 것을 볼 수 있다. 적합성 판단이 관대한 평가기준 또는 이진수준에서 의해서 이루어진다면, “조금 적합한” 문헌들도 대다수의 적합문헌들로 분류되어 질 것이다. 이는 “조금 적합한” 문헌들이 전반적인 검색효율성 평가의 결과에 영향을 미치며 “매우 적합한” 문헌들의 영향력은 간과될 수 있음을 의미한다(Sormunen, 2002; Voorhees, 2001). 시스템의 순위는 시스템평가에 사용된 적합성정도에 의존하여 바뀐다는 Voorhee(2001)의 연구결과가 위의 주장을 지지해주고 있다.

2.2 적합성 피드백

적합성피드백은 질의를 수정하는 과정에서 적합문헌에서 자주 출현하고 있는 용어들의 가중치를 증가시킴으로써 질의를 부적합문헌으로부터는 멀어지게 하고 적합문헌의 방향으로 이동시키는 것이다. 이렇게 함으로써 다음 단계의 검색에서 원치 않는 문헌들은 검색되지 않게 하며, 원하는 문헌들을 더 많이 검색해줄 수 있는 이점을 가지고 있다(Salton & Buckley, 1990). 지난 수십년간 적합성피드백 기법 및 이들이 검색효율성의 증진에 미치는 긍정적인 영향은 이미 많은 학자들에 의해 연구되어 그 결과들이 보고되었다(Dang et al., 2010; López-Pujalte et al., 2003; López-Pujalte et al., 2002; Quiroga & Mostafa, 2002; Belkin et al., 2001; Amati & Crestani, 1999; Spink &

Losee, 1996).

Amati & Crestani(1999)는 적합성피드백을 통한 정보필터링을 위해 학습모델을 제안하였는데 이는 보편화된 확률검색모형을 적용한 것으로 적합성피드백을 위한 학습표본을 위해 적합문헌과 부적합문헌을 사용하였으며 특히 학습샘플규모가 작을 때 적합성피드백에서 부적합문헌의 정보가 미치는 영향을 평가하였다. 그리고 부적합문헌의 정보를 효과적으로 이용하면 정보필터링에서 필요한 패러매터값을 신속하게 측정할 수 있게 된다는 것을 증명해보이려 했다.

Belkin et al.(2001)은 TREC의 상호작용 트랙(Interactive Track)에서 이용자와의 상호작용을 통해 질의 재구성을 지원해 줄 수 있는 인터페이스 설계 및 용어추천 방법을 제안하였다. 그들은 적합성피드백을 자동질의확장이라기 보다는 용어추천도구로써 구현하였으며 적합문헌 및 부적합문헌 모두를 이용하여 적합성판단을 허용함으로써 좋은 용어는 양수의 가중치, 나쁜 용어는 음수의 가중치가 함께 추천될 수 있도록 하였다. 이들은 또한 새로운 버전의 적합성피드백 메카니즘을 제안하였다. 동일한 용어를 포함하고 있는 적합한 텍스트과 부적합한 텍스트의 차별화, 즉 서로 다른 문맥상에서 사용된 용어, 주제에 대한 언급이 극히 부분적으로 다루어진 경우 또는 부적절한 관점에서 또는 다의적인 용어로 사용된 경우를 구별할 수 있어야 하고, 질의어에 포함되거나 또는 적합하다고 판단된 텍스트에서 출현된 용어는 부적합한 텍스트에서 출현여부에 관계없이 좋은 용어로 추천되어야 한다는 것이다.

Quiroga & Mostafa(2002)는 정보필터링 뿐만 아니라 정보검색에서 이용자 적합성피드백의 역할과 적합성의 개념에 대해서 재조명하였다. 이들은 적합성피드백 과정에서 상호작용에 대한 보다 나은 이해가 필요하며, 부분적인 적합성 또는 부적합성을 표현하기 위한 효과적인 방법 및 특정한 과제를 수행하기 위한 최상의 정보원, 즉, "매우 적합한" 문헌을 위한 평가기준이 필요함을 인지하였으며 부적합 문헌을 이용한 피드백도 중요할 수 있다는 것을 제안하고 있다. 또한 적합성 피드백은 검색된 문헌의 적합성 여부뿐 만아니라 적합성수준도 표현할 수 있도록 허용해 주어야 만 한다고 제안하고 있다.

López-Pujalte et al.(2003)과 López-Pujalte et al.(2002)은 정보검색에서 적합성피드백에 유전자 알고리즘을 적용시킴으로써 다른 기법과의 상대적 효율성을 측정하고 비교하였다. Dang et al.(2010)은 문헌에 출현하는 용어의 빈도수에만 의존하고 있는 전통적인 TF-IDF 가중치 기법보다는 질의용어를 중심으로 문맥(context)내에서 출현하는 용어들을 질의 용어의 가중치의 증감에 반영하는 문헌의 문맥-의존적(context-dependent)인 가중치 기법의 필요성에 대해 언급하고 있다. 즉, 질의 용어(q_i)가 적합문헌내의 q_i중심의 문맥상에서 출현한 용어들에 의해 둘러싸여 있다면 q_i의 가중치를 높혀 주고(Boosted), 반대의 경우에는, 즉 질의 용어(q_i)가 부적합문헌내의 q_i중심의 문맥상에서 출현한 용어들에 의해 둘러싸여 있다면 그 용어들의 가중치를 낮춰 주는(Discounted) 것이다. 이들은 이러한 기법을 B&D(Boost & Discount)과정이라고 부르고 있다.

3. 연구방법 및 실험환경

일반적으로, 정보검색실험을 위한 구성요소로는 정보검색시스템, 문헌집단, 질의, 적합성 판정세트 등이 있다. 본 연구의 실험을 위해 사용된 SMART시스템은 1960년도에 코넬대학에 개발되어 지난 50년 동안 사용되고 있는 검색시스템이다. 그리고 문헌집단 CF392는 전문으로 이루어진 데이터베이스로서 1974년부터 1979년까지 6년간의 MEDLINE 데이터베이스로부터 "Cystic Fibrosis"라는 키워드에 의해 1차로 검색된 1,239개의 문헌가운데 요약문(extract)이 아닌 초록(abstract)을 가지고 있는 784개의 문헌 중에서 임의로 하나씩 건너서 392개의 문헌을 선정하고 이들의 전문을 스캔하여 전문데이터베이스를 구축한 것이다.

1992년부터 시작된 TREC(Text REtrieval Conference)에서는 이전에 소규모 문헌집단을 이용하여 수행되어 왔던 검색실험결과의 신뢰성에 대한 문제점을 지적하며 매우 규모가 큰 문헌집단을 이용한 검색실험의 결과를 보고하고 있다(Voorhees & Harman, 2005). 그러나, 정보검색의 주요 목적은 우선 대규모 문헌집단에서 이용자의 질의에 적합한 문헌집단을 찾아내고 이들 매우 유사한 적합문헌집단 가운데서 이용자의 정보요구를 충족시킬 수 있는 더욱 적합한 문헌을 검색해 내는 것이다. 이것은 검색시스템의 정교한 검색기법에 의해서만 가능하게 될 것이다. 본 연구에서 사용되고 있는 CF392 전문데이터베이스는 이러한 연구 목적을 충족시킬 수 있는 매우 유사한 문헌들의 집합으로서 392개의 모든 문헌들이 위에서 이미 언급한 "Cystic Fibrosis"라는 소아폐질환관련 세부주제를 다루고 있다(Shaw et al., 1991).

문헌과 질의의 적합성평가 과정에서 심사자들은 문헌의 내용을 평가하고 적합성 정도를 점수로 부여하기 위해 전문을 살펴보았으며 적합성 평가결과는 전문내의 테이블이나 그림에 의해 영향을 받기도 하였다. 〈표 1〉과 같이 적합성 판정결과는 적합성 정도에 따라 점수가 부여되었고, 〈표 2〉와 같이 4그룹의 심사자들이 각각 부여한 점수에 따라 적합문헌과 부적합문헌을 구분하였다.

본 연구에서는 〈표 2〉에서와 같이 "조금 적합한" 문헌이 적합문헌(A1, B1, C1, D1, X1)으로 간주되었을 때와 그렇지 않았을 때(A2, B2, C2, D2, X2)의 검색결과를 비교해보았다. 그리고 서지전문가를 제외한 주제전문가 세 그룹의 점수의 합에 의해 적합문헌으로 선정되었을 때의 서로 다른 적합문헌세트, 즉 합이 1 이상일 때(Q1), 3 이상일 때(Q3), 5 이상일 때(Q5)의 검색결과도 살펴보았다. 즉 Q1은 적합성 판정 기준이 가장 관대한 경우에 얻어진 적합성 판정세트이고, Q5는 가장 엄격하게 적용

〈표 1〉 적합성 정도와 판정의 정의

적합성 정도(점수)	판정	정의
2	매우 적합한	문헌은 질의에 대한 즉각적인 응답으로 간주
1	조금 적합한	문헌은 주제적으로 관련 있으나 질의에 대한 즉각적인 응답으로 볼 수 없다.
0	부적합한	문헌은 질의와 관련이 없다.

〈표 2〉 적합성판정을 내린 심사자그룹 및 결과로 얻어진 상이한 적합문헌세트

적합문헌 판정세트	심사자
A2, A1	Cystic Fibrosis에 연구관심을 가지고 있는 질의서를 작성한 소아과 교수가 부여한 점수(1, 2)에 따라 적합문헌을 선정, 즉 A1은 점수가 1점 이상인 문헌집합. A2는 점수가 2점 이상인 문헌집합.
B2, B1	질의서를 작성한 교수의 동료교수
C2, C1	소화호흡기분야의 박사 후 연구 과정에 있는 연구원
D2, D1	의학분야 온라인탐색가로서 서지전문가
Q1, Q3, Q5	A, B, C 그룹의 적합성점수의 합에 따라 적합문헌을 선정, 예를 들면 Q3는 점수 합이 3보다 큰 문헌을 적합문헌으로 간주
X1, X2	X1 = A1, B1, C1, D1의 합집합 X2 = A2, B2, C2, D2의 합집합

하였을 때 얻어진 판정세트인 것이다. 그리고, 본 연구에서는 최초 부적합문헌으로 분류되었던 "조금 적합한" 문헌들로부터 얻을 수 있는 정보를 "부적합정보"라고 하였으며, 이들이 단계적으로 적합한 문헌으로 포함되었을 때 이들로부터 얻어진 정보가 초록/표제 시스템과 전문검색시스템의 검색효율성에 어떠한 영향을 미치는 지 살펴보고자 하는 것이다.

〈표 2〉의 결과에 의해 얻어진 각 적합문헌 판정세트 내에서의 적합문헌의 수는 아래의 〈표 3〉에 나타나 있다. A, B, C, D, X의 각 판정세트에서 보여 지듯이 적합성 점수가 1보다 큰 경우 적합문헌으로 포함시킨 판정세트 A1의 적합문헌의 수(717)가 적합성 점수가 2보다 큰 경우 포함시킨 A2(408)보다 더 많은 것을 발견할 수 있다. 이는 당연한 것으로써 다른 판정세트에서도 동일한 결과(B1>B2, C1>C2, D1>D2, X1>X2, Q1>Q2>Q3)를 보여 주고 있다. 또한 CF1239를 위해 만들어진 100개의 질의수도 전체 적합문헌의 수에 의존하여 그 수가 〈표 3〉과 같이 다소 달라지고 있다. 이는 적합문헌이 존재하지 않는 질의는 포함시키지 않았기 때문이다.

또한 〈표 4〉와 〈표 5〉에서는 각 질의번호(1-100)에 해당하는 적합성 판정세트(A1, A2, B1, B2, C1, C2, D1, D2, X1, X2, Q1, Q3, Q5)의 적합문헌 수를 보여주고 있다.

본 연구에서는 문헌의 가중치를 측정하기 위해 문헌 내에 나타나는 용어의 실제 출현빈도(TF)를 반영하였고 질의 용어는 IDF의 값을 적용하여 문헌과 각 질의와의 초기검색을 위한 유사도를 계산하였다. 또한 적합성피드백을 위해서는 검색된 적합문헌에 나타나는 질의용어의 실제 출현빈도를 반영한 투포아송(Two-Poisson) 독립모형을 이용하여 초록/표제 시스템과 전문검색시스템의 검색효율성에 어떻게 영향을 미치는지를 살펴보았다. 전문검색시스템에서 투

〈표 3〉 각 적합문헌 판정세트 내에서의 적합문헌과 질의의 수

판정세트	A1	A2	B1	B2	C1	C2	D1	D2	X1	X2	Q1	Q3	Q5
적합문헌수	717	408	682	340	660	355	933	449	1372	619	979	496	387
질의수	96	84	96	80	96	84	91	87	98	91	98	91	83

〈표 4〉 질의번호(1-50)에 해당하는 각 적합성 판정세트의 적합문헌의 수

질문서	A1	A2	B1	B2	C1	C2	D1	D2	X1	X2	Q1	Q3	Q5
1	5	2	5	2	6	2	10	1	14	3	8	3	2
2	1	0	1	0	0	0	0	0	2	0	2	0	0
3	6	1	2	2	10	2	5	1	16	3	13	2	2
4	1	0	1	0	1	0	0	0	2	0	2	0	0
5	11	8	11	7	13	5	26	7	34	13	17	9	8
6	2	0	0	0	1	0	2	0	5	0	3	0	0
7	0	0	4	0	3	0	2	1	9	1	7	0	0
8	3	1	2	0	2	1	3	2	4	2	3	2	0
9	2	1	2	0	2	0	2	1	4	2	3	1	0
10	7	2	6	2	6	3	7	4	10	4	9	4	3
11	3	2	2	2	4	2	4	3	6	3	4	2	2
12	3	2	2	2	2	1	2	1	4	2	3	2	2
13	2	0	2	1	2	1	3	1	6	2	4	1	1
14	7	5	8	3	7	4	5	3	11	6	9	6	5
15	21	10	14	8	18	9	25	13	33	15	24	12	9
16	22	9	16	7	21	10	31	14	45	18	28	13	11
17	2	2	2	0	4	2	5	2	7	3	4	2	2
18	7	2	3	0	3	1	2	1	7	2	7	3	1
19	4	4	4	4	4	3	4	2	5	4	4	4	4
20	9	6	7	4	8	6	8	5	9	7	9	7	5
21	4	3	5	1	4	2	9	4	10	4	5	4	1
22	5	0	5	0	4	2	10	3	12	4	8	2	0
23	11	4	6	4	5	4	6	5	11	6	11	5	4
24	4	2	3	2	2	2	9	2	11	3	4	2	2
25	7	4	4	1	6	1	10	3	14	6	8	4	1
26	4	2	5	1	6	2	3	3	6	4	6	4	1
27	3	2	3	1	2	2	5	2	6	2	4	2	2
28	2	0	3	0	2	0	10	4	10	4	4	1	0
29	4	0	2	0	3	1	6	5	6	5	4	1	0
30	1	1	1	0	1	0	3	2	3	2	1	1	0
31	6	4	6	2	5	4	9	6	9	6	7	5	4
32	3	1	2	1	2	2	4	2	5	2	3	2	1
33	9	6	6	5	8	5	11	7	16	10	11	6	5
34	5	5	6	4	6	4	8	5	10	6	7	5	5
35	3	2	3	1	3	1	2	1	6	2	5	2	1
36	1	1	1	1	1	1	0	0	1	1	1	1	1
37	15	12	12	10	16	11	21	14	22	14	18	12	11
38	8	2	2	1	3	1	7	4	12	4	9	2	1
39	20	9	11	1	14	9	37	15	41	18	23	10	8
40	16	8	15	3	26	8	21	8	47	13	33	11	7
41	4	3	4	3	4	4	4	3	4	4	4	4	3
42	6	3	3	3	5	2	12	4	13	4	6	3	3
43	14	7	20	7	9	5	18	7	32	10	23	8	7
44	41	34	35	32	36	26	43	32	47	36	42	34	33
45	3	1	3	2	3	0	2	2	4	2	4	2	1
46	5	4	4	3	4	4	5	4	6	4	5	4	4
47	9	5	6	5	9	2	4	3	14	7	13	5	5
48	4	3	5	3	7	3	5	3	11	4	9	3	3
49	13	10	12	12	13	11	9	8	14	12	14	12	11
50	3	1	4	2	2	2	7	2	9	2	5	2	2

〈표 5〉 질의번호(50-100)에 해당하는 각 적합성 판정세트의 적합문헌의 수

질문서	A1	A2	B1	B2	C1	C2	D1	D2	X1	X2	Q1	Q3	Q5
51	44	31	51	34	41	22	56	28	74	43	57	39	30
52	0	0	0	0	0	0	0	0	0	0	0	0	0
53	6	0	3	0	2	0	3	1	9	1	7	1	0
54	16	11	17	6	16	12	17	11	22	14	21	13	10
55	6	1	5	1	4	1	3	2	11	2	10	1	1
56	2	0	8	0	1	0	0	0	10	0	10	0	0
57	13	5	14	7	13	5	6	1	21	10	20	10	5
58	28	24	27	17	29	18	24	17	37	27	32	26	23
59	19	13	26	8	21	12	44	23	51	28	32	15	11
60	7	2	5	1	6	3	3	2	10	4	8	4	3
61	2	2	6	1	5	1	9	3	16	4	9	2	2
62	26	5	14	3	11	2	18	5	40	10	33	7	3
63	8	6	12	6	8	5	6	2	17	7	14	7	6
64	5	2	3	2	5	2	9	2	15	2	9	2	2
65	17	9	13	5	15	7	30	15	36	17	19	11	8
66	9	3	8	3	5	3	17	5	22	7	11	4	3
67	1	0	2	0	0	0	0	0	3	0	3	0	0
68	5	1	2	1	2	0	3	2	8	2	7	1	1
69	1	1	1	1	1	1	1	1	1	1	1	1	1
70	4	4	5	4	4	4	5	4	6	4	5	4	4
71	0	0	0	0	1	0	0	0	1	0	1	0	0
72	2	2	2	1	2	2	2	2	3	3	2	2	2
73	2	2	2	2	3	2	2	2	3	2	3	2	2
74	0	0	0	0	0	0	0	0	0	0	0	0	0
75	5	1	3	1	4	1	3	1	8	1	8	1	1
76	2	1	1	1	2	1	3	2	3	2	2	1	1
77	9	6	10	8	9	7	15	8	16	11	10	9	7
78	7	3	11	5	9	3	14	4	21	6	15	6	4
79	5	3	7	3	3	3	4	1	11	3	9	3	3
80	3	0	5	0	1	0	3	0	7	0	6	1	
81	2	2	2	2	4	1	3	1	6	2	4	2	2
82	5	5	5	3	5	4	10	4	10	5	5	5	4
83	4	1	6	2	3	3	5	0	10	3	7	3	2
84	3	1	2	1	2	1	2	0	4	1	4	1	1
85	5	3	5	3	4	3	9	6	11	7	6	4	3
86	4	2	5	1	4	1	9	2	12	3	7	2	1
87	6	4	8	4	4	4	10	4	13	4	10	4	4
88	1	1	3	1	2	1	1	1	3	1	3	1	1
89	2	1	4	2	3	1	3	2	6	3	4	2	1
90	5	3	6	4	6	3	8	5	9	5	6	5	4
91	39	28	48	26	39	23	70	31	88	48	53	37	26
92	13	8	11	4	12	5	16	4	22	10	15	9	5
93	1	1	1	1	1	1	6	3	6	3	1	1	1
94	7	6	6	3	8	6	16	10	17	10	8	6	6
95	3	2	2	0	2	2	1	1	4	3	3	2	2
96	2	1	1	1	2	1	5	1	5	1	3	1	1
97	4	2	3	2	2	1	4	1	6	2	5	2	2
98	3	2	4	2	3	2	6	3	7	4	5	2	2
99	1	0	1	1	1	1	0	0	1	1	1	1	1
100	2	1	5	1	2	1	3	1	5	1	5	1	1

포아송모형의 효율성은 선행연구에서 이미 보고하고 있다(문성빈, 1993). 검색효율성은 평균정확률(Average Precision), 즉 0.25, 0.5, 0.75의 재현율 지점에서의 정확률들의 값을 합하여 셋으로 나눈 평균으로 보여주고 있다. 적합성 피드백은 투포아송모형에 의해 하나의 문헌이 검색될 때마다 행해지고 있으며, 적합성피드백 후에 의해 얻어지는 문헌가중치 값은 아래와 같다.

$$\text{문헌가중치} = \sum_{i=1}^{n} d_i * \log \frac{\frac{1+(f_i/N)+s_i}{1+r}}{\frac{f_i - s_i}{N-r}}$$

위의 공식에서 d_i는 문헌 내 i번째 질의용어의 실제출현빈도를 나타내고, s_i는 검색된 적합문헌 내에서 i번째 질의용어의 출현빈도의 합, r은 질의어에 의해 검색된 적합문헌의 수, 1과 f_i/N는 적합문헌이 검색되지 않았을 때의 교정인수의 역할을 하는데 f_i는 i번째 질의용어의 장서빈도를 나타낸다. N은 전체문헌의 수를 의미한다. 즉, (f_i/N)은 i번째 질의용어의 전체문헌 내에서의 평균출현빈도, $(f_i-s_i)/(N-r)$에서 (f_i-s_i)는 전체출현빈도에서 검색된 적합문헌에서의 총 출현빈도를 뺀 것이고, $(N-r)$는 전체문헌수에서 검색된 적합문헌의 수를 뺀 것이므로 전체 부적합문헌 내에서 i번째 질의용어의 평균출현빈도를 나타낸다.

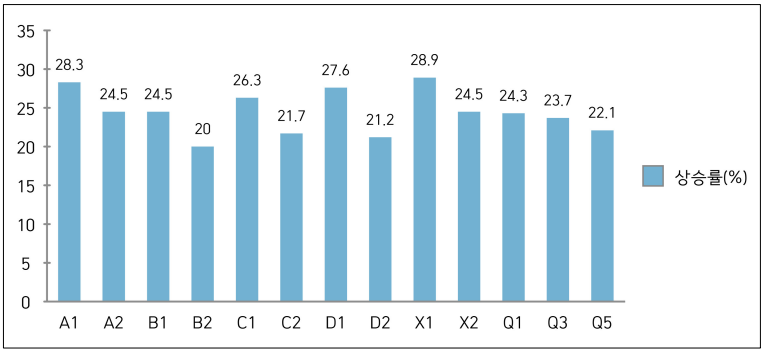
4. 실험결과

각 적합성 판정세트에서 초록/표제와 전문 시스템의 적합성피드백 전후에 검색효율성의 증감상황을 〈표 6〉과 〈표 7〉에서 보여 주고 있으며 이는 다시 〈그림 1〉과 〈그림 2〉에서 시각적으로 구별이 용이하게 나타나있다. 전반적으로 초록/표제에서는 A1에서 적합성피드백후의 상승률(28.3%)이 A2에서의 상승률(24.5%)보다 높이 나타났고 B, C, D, X, Q의 모든 적합성판정세트에서 동일한 경향을 〈그림 1〉에서 명확하게 보여주고 있다. 이는 초록/표제에서는 적합성점수가 1이상일 때 즉, 적합문헌의 기준이 엄격할 때보다 적합성판정 기준이 관대함으로써 "조금 적합한" 문헌들이 적합문헌으로 포함됨으로써 적합문헌의 수가 많아졌을 때, 적합성피드백으로 인한 혜택을 더 누릴 수 있다는 것을 보여 주고 있다. Sparck Jones(1974)는 검색효율성의 차이가 0.05-0.1에 있을 때는 "주목할 만한(noticeable)", 0.1 이상일 때는 "실질적인(material)" 차이라고 정의하였는데 이에 따르면 적합성피드백후의 검색효율성의 차이는 모든 적합성세트에서 "주목할 만한" 차이를 보여주고 있다.

그러나, 전문에서는 A2에서 적합성피드백후의 상승률(25.4%)이 A1에서의 상승률(15.0%)보다 높이 나타났으며, B만 제외하고 C, D, X, Q의 모든 적합성판정세트에서 초록/표제와는 정반대의 경향이 나타나고 있으며, 이는 또한 〈그림 2〉에서 시각적으로 명확하게 볼 수 있다. 전문에서는 적합성점수가 2 이상일 때 즉, 적합문헌의 기준이 관대할 때보다 적합성 판정기준이 엄격함으로써 "매우 적합한" 문헌들만이 적합문헌으로 포함되었을 때 적합성피드백으로 인한 혜택을 많이 보고 있다는 것을 보여 주고 있다. 전문에서는 확률검색모형의 기반이 되는 적합할 확률을 산출해내기 위한 실마리(clue)가

〈표 6〉 초록/표제에서 적합성피드백 후 검색효율성의 상승률

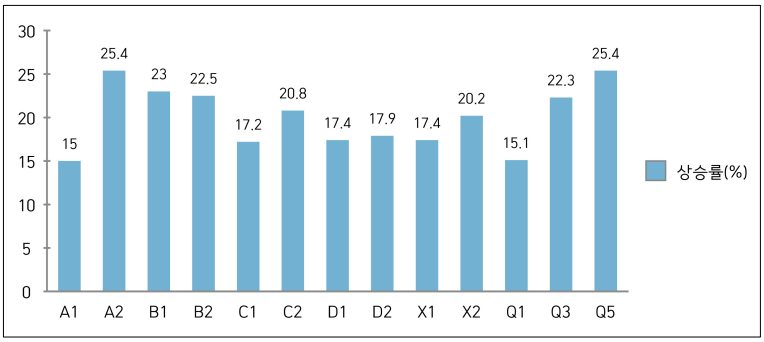
적합성 판정세트	No Feedback	Feedback	피드백 후 상승률(%)
A1	0.2449	0.3141	28.3
A2	0.3049	0.3797	24.5
B1	0.2461	0.3065	24.5
B2	0.2656	0.3189	20.0
C1	0.2579	0.3256	26.3
C2	0.2452	0.2984	21.7
D1	0.1931	0.2463	27.6
D2	0.2733	0.3312	21.2
X1	0.2150	0.2772	28.9
X2	0.2735	0.3404	24.5
Q1	0.2442	0.3035	24.3
Q3	0.2779	0.3438	23.7
Q5	0.2812	0.3434	22.1



〈그림 1〉 초록/표제의 적합성피드백 후 각 적합문헌 판정세트(A, B, C, D, X, Q)의 상승률(%)

〈표 7〉 전문(Full-Text)에서 적합성피드백 후 검색효율성의 상승률

적합성 판정세트	No Feedback	Feedback	피드백 후 상승률(%)
A1	0.3098	0.3562	15.0
A2	0.2974	0.3728	25.4
B1	0.2678	0.3294	23.0
B2	0.2747	0.3366	22.5
C1	0.2968	0.3479	17.2
C2	0.2552	0.3084	20.8
D1	0.2528	0.2968	17.4
D2	0.3025	0.3565	17.9
X1	0.2735	0.3211	17.4
X2	0.3053	0.3671	20.2
Q1	0.3034	0.3493	15.1
Q3	0.3013	0.3686	22.3
Q5	0.2810	0.3525	25.4



〈그림 2〉 전문의 적합성피드백 후 각 적합문헌 판정세트(A, B, C, D, X, Q)의 상승률(%)

최대한 제공되고 있지만, 이는 "조금 적합한" 문헌을 포함시킴으로써 결과적으로 "매우 적합한" 문헌뿐만 아니라 "조금 적합한" 문헌에서 출현하는 질의용어들이 잡음(noise)의 역할을 함으로써 초기검색의 경우뿐만 아니라 적합성피드백 후에 "조금 적합한" 문헌들이 먼저 검색되게 하는 결과를 초래하고 있다. 이러한 잡음은 특히 주제가 매우 동질적인 전문(full-text)으로 구성되어 있는 실험문헌집단에서는 평균정확률을 떨어뜨리는 결정적인 요인이 되고 있다 것을 보여주고 있다. 그러므로, 전문에서는 "매우 적합한" 문헌들만 적합문헌으로 포함시킴으로써 이러한 잡음의 영향력을 최소화하고 적합성피드백의 효과를 최대화해야만 할 것이다. 또한 적합성피드백후의 검색효율성의 차이가 0.044인 적합성세트 D1만 제외하고 모든 적합성세트에서 "주목할 만한" 차이를 보여주고 있다.

5. 연구결과에 대한 기대효과 및 활용방안

지난 수십 년간 적합성 개념에 대한 논의는 계속되어 왔고 적합성은 정보검색 실험에서 검색효율성을 측정하는 척도인 정확률과 재현율의 기본 개념이 되고 있다. 또한 적합성판단의 기준(criteria)과 수준(level)도 다양해져서 이에 대한 논의도 끊임없이 지속되고 있다. 정보검색실험에서 적합한 정도를 이진수준이 아닌 두 개 이상의 수준으로 판정한 후 이들 중 "조금 적합한" 문헌이 어느 그룹에 배정되느냐에 따라 적합문헌의 수는 달라질 수 있고 적합/부적합정보를 이용한 적합성피드백 기법을 적용하였을 때 검색효율성 또한 다른 결과를 가져올 수 있다. 본 연구에서는 초록/표제 그리고 전문을 색인 및 검색대상으로 하였을 때 적합성피드백 후 검색효율성의 증감은 상대적으로 어느 쪽이 더 혜택을 받고 있는지 살펴보았고, 또한 각 적합성판정세트가 적합성피드백 후에 검색효율성이 "주목할 만한"차이를 보여주고 있는지를 면밀히 살펴보았다.

전반적으로 모든 적합성 판정세트의 적합성피드백 후의 검색효율성은 "주목할 만한" 차이를 보였고 초록/표제는 "조금 적합한" 문헌을 적합문헌으로 포함시켰을 경우가 그렇지 않았을 때보다 적합성피드백 후 높은 증가율을 보

여 주었으나, 전문에서는 "조금 적합한" 문헌을 적합문헌에서 제외시켰을 경우가 일관성 있게 높은 증가율을 보여 주고 있었다. 전문검색시스템에서 확률모형의 유용성에 대한 연구는 Maron(1988)에 의해 이미 보고 된 바 있다. 그러나 본 연구에서 적합문헌에 포함된 "조금 적합한" 문헌은 확률모형을 적용한 전문검색시스템의 적합성피드백에서 잡음을 초래하고 있음을 암시하고 있다. 즉, 전문검색시스템의 대상이 되는 문헌들에 대한의 적합성 판단은 엄격하게 이루어져야 한다는 것이다. 본 연구결과는 전문검색시스템이 제공하는 검색효율성의 문제점, 즉 낮은 정확률을 적합성 판정의 엄격한 기준을 적용한 적합성피드백 기법을 사용함으로써 부분적으로 해결할 수 있는 방안을 제시하였다.

본 연구의 제한점은 이공계열의 한 분야인 의학분야의 특정주제를 다루는 매우 유사문헌들을 색인 및 검색실험대상으로 하였기 때문에 연구결과의 적용이 인문사회분야의 문헌으로 구성된 실험문헌집단에서는 다른 결과가 나타날 수 있다. 본 연구결과가 일반화되기 위해서는 이공계열보다 어휘의 다양성이 존재하는 인문사회분야의 유사문헌집단을 대상으로 추가 검색실험이 행해져야 할 것이다.

지난 수십 년간 정보기술 및 네트워크의 기술은 끊임없이 발전되어 왔으며 전 세계를 연결하는 정보고속도로라는 개념의 초고속 통신망을 이용한 학술정보의 유통과 이에 대한 요구는 주목할 만큼 증가하고 있을 뿐만 아니라 온라인 상용시스템에서의 전문검색시스템에 이용되는 전문데이터베이스가 차지하는 비율은 날로 증가하고 있다. 또한 이러한 통신망을 이용하여 구축되고 있는 디지털도서관은 이용자의 정보요구에 적합한 문헌의 원문제공서비스를 가능하게 하고 있다. 미래의 디지털도서관은 문헌의 원문제공서비스 뿐만 아니라 전문데이터베이스를 대상으로 한 색인 및 검색이 가능하도록 전문검색시스템에 대한 지속적인 연구결과의 응용은 필수적이다. 이는 미래의 디지털도서관 시대에 전문데이터베이스를 이용한 색인 및 정교한 검색시스템의 가능성 및 활용성을 높이게 될 것이고, 그러므로 전문검색시스템이 기존의 초록과 표제를 기반으로 한 온라인 상용시스템을 대체하는데 기여할 것으로 확신한다.

참 고 문 헌

문성빈 (1993). 적합성피드백을 이용한 전문검색시스템의 효율성 증진을 위한 연구. 정보관리학회지, 10(2), 43-67.

문성빈 (1997). 상이한 적합성 판정과 전문검색시스템의 평가에 관한 연구. 정보관리학회지, 14(2), 123-141.

Amati, G., & Crestani, F. (1999). Probabilistic learning for selective dissemination of information. Information Processing and Management, 35(5), 633-654.

Belkin, N.J. (1984). Cognitive models and information transfer. Social Science Information Studies, 4, 111-129.

Belkin, N.J., Cool, C., Kelly, D., Lin, S.J., Park, S.Y., Perez-Carballo, J., & Sikora, C. (2001). Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. Information Processing and Management, 37(3), 403-434.

Blair, D.C., & Maron, M.E. (1990). Full-text information retrieval: further analysis and clarification. Information Processing and Management, 26(3), 437-447.

Borlund, P. (2003). The concept of relevance in IR. Journal of the American Society for Information Science and Technology, 54(10), 913-925.

Burgin, R. (1992). Variations in relevance judgements and evaluation of retrieval performance. Information Processing and Management, 28(5), 619-627.

Dang, E.K.F., Luk, R.W.P., Allan, J., Ho, K.S., Chan, S.C.F., Chung, K.F.L., & Lee, D.L. (2010). A new context-dependent term weight computed by boost and discount using relevance information. Journal of the American Society for Information Science and Technology, 61(12), 2514-2530.

Eisenberg, M.B. (1988). Measuring relevance judgments. Information Processing and Management, 24(4), 373-389.

Eisenberg, M.B., & Barry, C. (1988). Order effect: A study of the possible influence of presentation order on user judgments of document relevance. Journal of the American Society for Information Science, 39(1), 37-49.

Greisdorf, H. (2003). Relevance threshold: A multi-stage predictive model of how users evaluate information. Information Processing and Management, 39(3), 403-423.

Harter, S.P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43(9), 602-615.

Harter, S.P. (1996). Variation in relevance assessments and the measurement of retrieval effectiveness. Journal of the American Society for Information Science, 47(1), 37-49.

Hjørland, B. (2010). The Foundation of the concept of relevance. Journal of the American Society for Information Science and Technology, 61(2), 217-237.

Huang, X., & Soergel, D. (2013). Relevance: An improved framework for explicating the notion. Journal of the American Society for Information Science and Technology, 64(1), 18-35.

Kekäläinen, J., & Järvelin, K. (2002). Using graded relevance assessments in IR evaluation. Journal of the American Society for Information Science and Technology, 53(13), 1120-1129.

López-Pujalte, C., Guerrero Bote, V. P., & Moya Anegón, F. (2002). A test of genetic algorithms

in Relevance Feedback. Information Processing and Management, 38(6), 793-805.

López-Pujalte, C., Guerrero Bote, V.P., & Moya Anegón, F. (2003). Order-based fitness functions for genetic algorithms applied to relevance feedback. Journal of the American Society for Information Science and Technology, 54(2), 152-160.

Maglaughlin, K.L., & Sonnenwald, D.H. (2002). User perspectives on relevance criteria: A comparison among relevant, partially relevant, and not relevant. Journal of the American Society for Information Science and Technology, 53(5), 327-342.

Maron, M.E. (1988). Probabilistic design principles for conventional and full-text retrieval systems. Information Processing and Management, 24(3), 249-255.

Mizzaro, S. (1997). Relevance: The whole history. Journal of the American Society for Information Science, 48(9), 810-832.

Quiroga, L.M., & Mostafa, J. (2002). An experiment in building profiles in information filtering: The role of context of user relevance feedback. Information Processing and Management, 38(5), 671-694.

Salton. G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41(4), 288-297.

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.

Saracevic, T. (2007a). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. Journal of the American Society for Information Science and Technology, 58(13), 1915-1933.

Saracevic, T. (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. Journal of the American Society for Information Science and Technology, 58(13), 2126-2144.

Schamber, L. (1994). Relevance and information behavior. Annual Review of Information Science and Technology, 29(1), 3-48.

Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational, definition. Information Processing & Management, 26(6), 755-776.

Shaw, W.M. Jr., Wood, R.E., & Tibbo, H.R. (1991). The cystic fibrosis database: Content and research opportunities. LISR, 13, 347-366.

Sormunen, E. (2002). Liberal relevance criteria of TREC-counting on negligible documents? In M. Beaulieu, R. Baeza-Yates, S. Myaeng, and K. Järvelin (Eds.), Proceedings of the SIGIR

2002 (pp. 324-330). New York: ACM.

Sormunen, E., Kekäläinen, J., Koivisto, J., & Järvelin, K. (2001). Document text characteristics affect the ranking of the most relevant documents by expanded structured queries. Journal of Documentation, 57(3), 358-376.

Spink, A., & Greisdorf, H. (2001). Regions and levels: Measuring and mapping users' relevance judgments. Journal of the American Society for Information Science and Technology, 52(2), 161-173.

Spink, A., & Losee, R.M. (1996). Feedback in information retrieval. Annual Review of Information and Science and Technology, 31, 33-78.

Spink, A., Greisdorf, H., & Bateman, J. (1998). From highly relevant to nonrelevant: Examining different regions of relevance. Information Processing and Management, 34(5), 599-622.

Swanson, D.R. (1986). Subjective versus objective relevance in bibliographic retrieval system. The Library Quarterly, 56, 389-398.

Tang, R., Shaw, W.M., & Vevea, J.L. (1999). Towards the identification of the optimal number of relevance categories. Journal of the American Society for Information Science, 50(3), 254-264.

Vakkari, P., & Sormunen, E. (2004). The influence of relevance levels on the effectiveness of interactive information retrieval. Journal of the American Society for Information Science and Technology, 55(11), 963-969.

Voorhees, E. (2001). Evaluation by highly relevant documents. In W. Croft, D. Harper, D. Kraft, and J. Hobel (Eds.), Proceedings of the SIGIR 2001 (pp. 74-82). New York: ACM.

Voorhees, E.M., & Harman, D.K. (Eds.) (2005). TREC: Experiment and evaluation in information retrieval. Cambridge: MIT Press.

Xu, Y., & Chen, Z. (2006). Relevance judgment: What do information users consider beyond topicality. Journal of the American Society for Information Science and Technology, 57(7), 961-973.

Xu, Y., & Wang, D. (2008). Order effect in relevance judgment. Journal of the American Society for Information Science and Technology, 59(8), 1264-1275.

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

Moon, Sung-Been (1993). Enhancing performance of full-text retrieval systems using relevance

feedback. Journal of the Korean Society for Information Management, 10(2), 43-67.

Moon, Sung-Been (1997). Variations in relevance assessments and evaluation of the performance of full-text retrieval system. Journal of the Korean Society for Information Management, 14(2), 123-141.