

인용 지표를 이용한 재순위화 및 질의 확장의 성능 평가^{*}

- 인용색인 데이터베이스를 기반으로 -

Performance Evaluation of Re-ranking and Query Expansion for Citation Metrics: Based on Citation Index Databases

이 혜경 (HyeKyung Lee)^{**}
이용구 (Yong-Gu lee)^{***}

목 차

- | | |
|-----------|--------|
| 1. 서 론 | 4. 분 석 |
| 2. 이론적 배경 | 5. 결 론 |
| 3. 연구방법 | |

초 록

본 연구의 목적은 인용 지표가 인용 색인 데이터베이스의 검색성능 향상에 기여할 가능성을 파악하는 데에 있다. 이를 위하여 본 연구는 문헌정보학 분야 10개의 질의를 Web of Science에서 검색하여 수집한 3,467건의 문헌과 2000년부터 2021년까지 SSCI 문헌정보학 분야 저널 85종에 수록된 60,734건의 문헌을 기반으로 적합성 판단을 거쳐. 검색 결과의 상위 100순위에 대한 성능 및 검색 방식과 인용 지표를 활용한 재순위화, 그리고 벡터 공간모형 검색시스템 구축 등에 따른 질의 확장 실험을 수행하였다. 그 결과 첫째, 인용 지표를 단독으로 사용한 재순위화의 성능은 Web of Science의 검색성능과 상이하였으며. 인용 지표는 Web of Science 기준 시스템에 적용되지 않는 독립적인 지표로 사용하고 있었다. 둘째, 고유 질의어 수에 질의어의 총 출현 빈도를 조합하고 인용수를 보조적으로 사용했을 때, 성능에 긍정적인 영향을 미칠 것으로 확인하였다. 셋째, 질의 확장에서는 전반적으로 벡터 공간모형 기반 검색시스템의 기본 성능 대비 성능이 향상되었다. 넷째, 이용자 적합성을 통해 질의 확장을 적용한 경우가 시스템 적합성을 적용한 경우보다 성능이 향상 되었다. 다섯째, 피인용 수를 적합 문헌과 더불어 사용하면 최상위권 내 적합 문헌에서의 순위 변동 가능성을 보여주었다.

ABSTRACT

The purpose of this study is to explore the potential contribution of citation metrics to improving the search performance of citation index databases. To this end, the study generated ten queries in the field of library and information science and conducted experiments based on the relevance assessment using 3,467 documents retrieved from the Web of Science and 60,734 documents published in 85 SSCI journals in the field of library and information science from 2000 to 2021. The experiments included re-ranking of the top 100 search results using citation metrics and search methods, query expansion experiments using vector space model retrieval systems, and the construction of a citation-based re-ranking system. The results are as follows: 1) Re-ranking using citation metrics differed from Web of Science's performance, acting as independent metrics. 2) Combining query term frequencies and citation counts positively affected performance. 3) Query expansion generally improved performance compared to the vector space model baseline. 4) User-based query expansion outperformed system-based. 5) Combining citation counts with suitability documents affected ranking within top suitability documents.

키워드: 정보검색, 인용 지표, 재순위화, 질의 확장, 성능 평가

Information Retrieval, Citation Metrics, Re-Ranking, Query Expansion, Performance Evaluation

* 이 논문은 경북대학교 문헌정보학과 박사학위논문의 일부를 수정 및 요약한 것임.

** 경북대학교 문헌정보학과 강사(fantarias3@gmail.com / ISNI 0000 0004 9360 1074) (제1저자)

*** 경북대학교 문헌정보학과 부교수(yglee@knu.ac.kr / ISNI 0000 0004 6437 6752) (교신저자)

논문접수일자: 2023년 7월 22일 최초심사일자: 2023년 8월 2일 게재확정일자: 2023년 8월 12일

한국문헌정보학회지, 57(3): 249-277, 2023. <http://dx.doi.org/10.4275/KSLIS.2023.57.3.249>

※ Copyright © 2023 Korean Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited. the use is non-commercial and no modifications or adaptations are made.

1. 서 론

21세기 정보사회로의 도약은 세계의 다양한 학문 분야의 발굴과 연구 주제의 지평을 넓혀주는 계기가 되었다. 연구의 기원은 인류가 문명을 이룩하고부터 현재까지 끊이지 않고 있고, 과거의 연구는 현재의 생활과 미래를 발전시키는 매개가 되기도 한다. 이러한 연구의 비약적인 발전은 앞서 정보사회로의 도약이라고 언급한 바와 같이, 정보와 웹, 컴퓨터라는 매체의 발달과 그 흐름을 함께 한다.

Garfield(1964)는 인용과 이를 활용한 영향력 지수(Impact Factor)를 개발하여 SCI와 SSCI 등의 인용색인을 구축하였고, 인용색인은 추후 정보기술과의 접목을 통해 인용색인 데이터베이스로 발전하게 되었다. 그리고 최근의 연구자들이 주로 활용하는 인용색인 데이터베이스는 Web of Science, Scopus, Google Scholar로 나타났다(Bar-Ilan, 2008; Rovira et al., 2019; Anker et al., 2019; Martin-Martin et al., 2021).

인용색인 데이터베이스는 기본적으로 ‘인용’에 기반한 문헌과의 관계, 선별된 학술지들의 영향력들에 따른 기능이나 특징을 반영하기 때문에, 학술 데이터로 활용할 수 있다는 이점을 있다(Birkle et al., 2020). 따라서 인용색인 데이터베이스는 자신들이 보유한 학술지 논문 데이터를 제공하는 색인의 역할을 하지만, 데이터베이스마다 독자적인 논문의 영향력과 인용관계 산출을 통하여 논문들의 가치를 나타내주기도 한다.

그리므로 위와 같은 인용색인 데이터베이스의 이용자는 자신들의 정보요구에 따라 질의를 결정하고 각 데이터베이스에서 검색을 수행한 후 도

출된 검색 결과에 대하여 인용관 관련한 여러 요소를 적용해 자신에게 적절한 문헌을 선택한다.

이용자는 검색을 통해 가장 만족할만한 결과를 얻길 기대한다. 그러므로 좋은 검색시스템은 질의에 가장 적합한 결과로 응답하는 것이며(Rivas, Iglesias, & Borrajo, 2014), 정보검색에서 적합성은 검색시스템의 성능과 필연적인 관계로, 기준보다 적합한 결과의 도출을 위한 검색시스템 성능 향상과 개발은 지금도 지속해서 수행되고 있고, 연구자에게 영구적인 과제로 남아 있다(Ioannakis et al., 2017).

그리고 정보 검색에서 말하는 적합성은 일차 원적인 시스템 적합성 즉, 질의어와의 일치 정도를 확인하는 것이 아니라, 문헌의 내용과 인용(citation), 저자성(authorship) 등을 복합적으로 작용하였을 때 가장 적합한 적합성이라 판단한다(cooper, 1971; Lancaster, 1979; Mizzaro, 1998).

이러한 의미에서 인용색인 데이터베이스는 이들이 특징적으로 보유한 인용 요소와 지표가 있음에도 불구하고, 검색 수행에 인용 지표를 사용하지 않고 적합성 순위를 도출하는 것은 다소 의구심을 가질 만한 상황으로 여겨진다.

이러한 배경에서 본 연구는 인용색인 데이터베이스가 제공하는 검색 결과와 인용 지표 데이터를 기반으로 검색 결과의 재순위화와 질의 확장을 수행하여 인용색인 데이터베이스의 검색 시스템 성능 향상의 가능성을 확인해보고자 한다. 그리고 실제 인용색인 데이터베이스가 제공하는 서지데이터를 활용하여 자체 검색시스템을 구축하여 인용색인 데이터베이스의 결과와 비교해보고 시스템 자체의 성능도 더불어 파악하고자 한다.

2. 이론적 배경

2.1 인용색인 검색에서의 적합성 의미

검색시스템의 성능은 이용자의 질의에 대해서 얼마나 적합한 문헌을 보여주느냐에 달려 있다. 그러므로 적합성은 검색시스템에서 매우 중요한 개념으로 여겨진다.

적합성은 relevance, pertinence, aboutness 등 다양한 영문 단어로 표기되는데, 각 용어가 내포하는 적합의 범주는 학자마다 다소 상이하므로 국내에서도 학문의 범위나 연구 분야마다 연관성, 적합성 등의 용어들로 번역되기도 한다. 그리고 국내 정보검색 분야에서는 relevance를 주로 적합성으로 표기하고 있다.

인용색인 데이터베이스 역시 이용자 검색에 따라 정보를 제공하는 검색시스템에 기반하고 있으므로, 그에 따른 적합한 검색 결과를 보여주고 있다. <표 1>에 의하면, 주로 인용색인 데이터베이스는 질의어와 적합 문헌에서 등장하는 용어와의 유사도와 중복에 따라 적합성을

도출하고 있다. 이는 Mizzaro(1998)의 연구에서 제시한 정보검색 적합성의 4개의 차원 중 가장 낮은 첫번째 차원에 해당한다.

그리고 Spink et al.(1998)와 Maglaughlin과 Sonnenwald(2002)는 저자의 권위나 지위 상태 등이 문헌의 적합성에 작용할 가능성이 있음을 시사하였으며, 박정아, 손영우(2009)는 이용자가 판단하는 적합성에는 문헌의 내용과 더불어 정보검색의 상황과 정보의 신뢰성을 같이 확인해야 하는데, 이중 신뢰성은 저자의 권위를 포함하고 있다고 하였다.

결국 정보검색에서의 적합성은 주로 질의에 따른 검색의 결과로 나타난 문헌에 대한 적합성을 의미하나, 여기에서 적합성은 개인의 내재적인 요구에 합당한 적합성이 아닌 외부적인 시스템적 적합성을 기반으로 부차적인 콘텐츠를 포함한다. 더불어 적합성을 판단하는 요소로 정보요구(질의)와의 문헌의 용어 중복뿐만 아니라, 문헌이 지닌 내용과 주제, 인용이나 저자의 영향 등 다양하게 고려해야 한다.

<표 1> 인용색인 데이터베이스의 적합성의 범위와 고려도

	Web of Science ¹⁾	Scopus ²⁾
적합성의 범위	제목 초록 저자 키워드 키워드 플러스	서지사항에 등장하는 질의와 근접한 용어
적합성의 순서 영향력	제목, 초록에 등장하는 질의 유사 용어 기준치에 따라 순위화	조회수, 용어 위치에 따른 가중치, 용어 위치, 질의와의 근접성, 질의와의 완전성 등에 따른 순위화

1) Web of Science Core collection Help

https://images.webofknowledge.com/images/help/WoS/hs_sort_options.html [cited by 15/11/22]

2) https://service.elsevier.com/app/answers/detail/a_id/14182/suporthub/Scopus/ [cited by 15/11/22]

2.2 인용 지표

2.2.1 논문 수준

논문의 영향력을 판단하는 대표적인 지표로는 논문의 인용 수(citation count)를 고려할 수 있다. 일반적으로 인용 수는 논문의 과급력 혹은 영향력으로 해석하지만 과급력이 반드시 논문의 고품질을 의미하는 것은 아닌데, 이는 인용수가 높다고 무조건 논문의 내용이나 질이 좋다고는 판단할 수 없기 때문이므로 단순히 높은 인용수와 논문의 질을 연관하기에는 다소 무리가 있다.

그럼에도 Smith(1981)는 이용자의 협조 없이 수치 그 자체로서 한 논문에 대한 이용자의 판단을 확인하고 편견 없이 논문을 평가할 수 있기 때문에 이러한 인용 수가 인용 지표로써 활용되어야 함을 피력하였다. 그러나 Cronin(1982)은 인용 수에 대한 인용 규범이나 규칙이 공식화되어 있지 않기 때문에 측정 자체가 불확실하므로, 이 결과로 도출된 인용 수는 신뢰하기에 무리가 있다고 하였다.

최근 Van Raan(2005)은 인용 과정만으로도 매우 복잡하고, 인용 행태와 인용 간의 명확한 연관성이 있으므로 인용수가 지니는 가치를 부인해서는 안 된다고 말하였다.

이처럼 인용 수는 인용색인 등장 이후 가장 전통적인 인용 지표로 여겨져 왔기에 오래전부터 인용 지표에 대한 논의가 끊이지 않고 있다.

최근 인용 수를 대체할 새로운 인용지표인 Altmetrics 등이 개발되기도 하였으나, 여전히 인용 수는 연구자들로부터 가장 기본적인 인용 지표로 여겨지고 있다.

2.2.2 저자 수준

Hirsch(2005)는 연구자의 연구산출물을 활용하여 연구자 자체의 영향력을 파악할 수 있는 h-index를 제안하였다. 보편적으로 연구자의 영향력을 판단할 때 고려할 수 있는 계량서지학 지표로는 연구자가 출판한 논문의 수와 앞 절에서 확인한 인용 수 등이 있다. 총 논문 출판 수는 연구자의 연구 생산성을 정량적으로 측정할 수 있지만 논문 자체의 중요성이나 영향력을 측정할 수 없다. 반면 논문의 인용 수는 연구자의 논문이 획득한 인용 수를 통해 영향력을 파악할 수는 있으나 논문의 참여 저자의 순위를 고려하지 않는다. 그리고 연구자가 출판한 논문 중 일부만이 상대적으로 높은 인용 수를 지녔음에도 불구하고 해당 연구자에 대한 영향력이 과도화 혹은 편향화될 가능성이 존재한다. 이러한 단점을 보완하기 위하여 연구자가 출판한 논문 중 주요 논문만을 활용하거나 최다수 인용 논문만을 성과 측정에 사용하는 등의 방법을 시도할 수 있지만, 이 또한 자의적이고 임의적인 부분이 존재하기 때문에 합리적으로 저자의 영향력을 평가하기에는 다소 무리가 있다(Hirsch, 2005).

이러한 배경에서 제안된 h-index는 연구자가 보여줄 수 있는 생산성과 영향력을 동시에 활용해 연구자의 영향력을 나타내고자 하였다.

그러나 h-index를 도출할 때 최고빈도의 피인용 수를 지닌 논문에 대해서는 상대적으로 고려도가 낮은 점등은 결국 h-index가 출판 논문의 최대 수치를 넘어서 수 없음을 의미한다. 그러므로 다른 연구자들보다 고빈도 인용 논문이 많지만, 출판 논문의 수 자체가 적은 연구자들은 상대적으로 h-index가 불리하게 작용할 수도 있다.

이에 따라 최근까지도 h-index를 보완하거나 발전시켜 파생된 다양한 저자 영향력 지표(ar-index, r-index, e-index 등)가 등장하고 있지만, 여전히 h-index는 인용 색인 데이터베이스에서 대표적인 저자 수준의 인용 지표로 제공되고 있는데, 저자 수준에서 가장 영향력 있는 지표로서 위치를 가늠할 수 있다.

2.3 검색시스템의 평가와 성능향상기법

2.3.1 검색시스템 성능의 측정

1992년 아래로 해마다 개최되는 TREC(Text REtrieval Conference)³⁾은 미국 국립표준기술연구소(National Institute of Standards and Technology)에서 주관하는 정보검색 및 정보 접근 관련 컨퍼런스이다. TREC에서 제공하는 테스트컬렉션은 실제의 검색과 유사한 환경을 구현하기 위하여 그 크기가 매우 방대하다. 그러므로 TREC의 참여 연구자들은 검색 순위에 기반한 실험과 다양한 성능 평가 척도를 사용한다.

TREC에서 공식적으로 사용하는 순위 기반의 대표적인 척도는 mAP(mean Average Precision), 11지점 평균 정확률(11-point average Precision, Precision at k, R-precision, bpref, DCG(Discounted Cumulative Gain), nDCG(normalized Discounted Cumulative Gain) 등이 있다(Soboroff, 2021).

이들 평가척도의 공통점은 언급한 대로 순위를 적용해서 성능을 판단하는 것인데, 검색 시스템의 전체 데이터양을 알 수 없으므로, 정해진 순위 혹은 상위 k 순위에 등장한 적합 문

헌의 비율에 따라 정확률을 도출한다.

이중 DCG는 최근 순위 기반의 추천 검색 시스템에서 일반적으로 사용되는 평가척도이다. DCG는 부분 적합 요소와 순위를 함께 고려하기 때문에 적합 문헌의 등급을 나누어 보다 세밀하게 성능을 평가할 수 있으며, 동시에 검색에서 도출되는 모든 적합 문헌의 순서나 위치마다 특정한 가중치를 부여하여 성능 평가 척도의 결과가 과대 측정되거나 편향될 가능성을 감소시킨다. 따라서 DCG를 활용하여 검색시스템을 안정화할 경우, 이용자가 적합 문헌을 더욱 상위의 문헌 집합에서 찾을 가능성이 커진다(Wang et al., 2013).

DCG 척도(〈수식 1〉 참조)를 산출하기 위해서는 우선 CG를 산출해야 한다. CG는 평가할 적합 문헌의 등급, 즉 문헌이 얼마나 적합한지에 따른 점수를 합한 값으로, DCG는 이 CG에 순위를 적용해 순위가 낮은 적합 문헌의 영향력을 낮추고 반대로 순위가 높은 적합 문헌의 영향력을 높여 검색시스템의 성능을 측정한다.

한편 nDCG는 이러한 DCG를 정규화한 척도로 〈수식 2〉와 같으며, DCG 값을 IDCG(Ideal Discounted Cumulative Gain)로 나누어 정규화한다. IDCG의 경우 순위가 높은 적합 문헌에 한하여 DCG 값을 도출한 것으로 이렇게 정규화할 경우, DCG 보다 안정적인 결과 성능 값을 확인할 수 있다.

$$DCG_p = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)}$$

〈수식 1〉 DCG 공식

3) Text Retrieval Conference(<https://TREC.nist.gov/>)

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$

〈수식 2〉 nDCG 공식

2.3.2 성능향상기법

(1) 재순위화

재순위화는 초기 검색 결과의 문헌을 특정한 알고리즘이나 규칙에 따라 이차 검색함으로써 문헌의 등장 순서를 조정하는 방법을 일컫는다.

재순위화를 위한 방법으로는 주로 검색시스템 이용자의 과거 이력을 활용하거나, 질의와 문헌 간의 유사 연관어(Natsev et al., 2007; 이승우, 송영인, 임해창, 2008; 장영진 외, 2020; Jiang et al., 2021) 등을 고려할 수 있다.

김한준, 노준호, 장재영(2012)은 개인화 검색에 이은 개인의 관심도나 이용자가 실제 적합 문헌으로 선택한 이력을 통하여 이용자별 질의어 프로파일을 제작해 검색 결과의 상위 문헌에 대하여 질의 확장을 수행하고 초기 질의어와 유사한 질의어에 따른 결과와의 중첩에 따라 재순위화를 수행하였다. 이승우, 송영인, 임해창(2008)은 사전 기반 방식이나 기계 학습 기반 방식을 사용하더라도 질의어와의 유사어를 확인하고 어휘 출현 횟수에 따라 적합 문헌으로 판단하고 이차 검색 이후 순위를 조정하였다.

최근에는 기계 학습이나 자연어 처리 모델이 직접적 검색시스템에 적용되면서 적합 문헌을 자동으로 판단하고 순위를 조정하기도 하였다. Gupta, Chinnakotla, Shrivastava(2018)는 합성곱 신경망(Convolutional Neural Network)을 사용하여 초기질의 결과의 특성을 추출하고 특성 간의 관계를 파악해 지식 그래프를 생성하고 이차 검색을 수행하였다.

그러나 어떠한 방법을 사용하여 순위 조정을 하더라도 재순위화는 초기 검색 결과 문헌에서 재순위화 요소를 추출하고 조정하거나, 질의 확장 기반의 이차 검색을 수행하게 된다.

(2) 질의 확장

질의 확장은 검색시스템의 성능 향상을 위해 사용되는 일반적인 방법(Rivas, Iglesias, & Borrajo, 2014; Jain, Seeja, & Jindal, 2021)으로, 정보요구에 부합할 특정 부가 정보를 제공하면 검색성능이 향상될 것이라는 가정하에 수행된다. 그리고 이미 질의 확장은 검색시스템의 성능향상에 기여하고 있음이 다수의 선행연구에서 밝혀졌다.

질의 확장은 크게 확장 문헌 범위에 따라 전역적 확장과 지역적 확장으로 구분할 수 있다. 이중 지역적 확장은 초기질의에 따른 일차 검색 후 적합 문헌에 기반하여 질의를 확장한다. 그러므로 지역적 확장 수행에서는 결과 문헌에 대한 적합 여부가 필수적이다.

최근의 검색시스템에서의 질의 확장은 대부분 시스템 적합성을 사용한다. 이용자가 실제로 검색 결과 문헌 전체에 대하여 적합성 여부를 판단하기에는 무리가 따르기 때문이다.

또한 지역적 확장은 질의 확장에 적용할 적합 문헌의 수도 고려한다. 최초 검색 결과 중 상위권에 등장한 소수의 문헌을 적용하는 것이 확실하게 성능 향상을 도모하기 때문에, 최상위로 등장한 첫 번째의 문헌을 적용하는 것이 성능 향상에 가장 효과적임이 최근의 선행연구(Xu & Croft, 2017)에서도 밝혀진 바가 있다. 그러나 Harman(1992)은 검색시스템에 마다 적합 문헌의 길이에 따른 변동이 존재할 수 있으므로, 상위에 등장하는 순서대로 적합 문헌을 적

용하기보다는 문헌의 중요도에 의하여 n개의 문헌을 적용하는 것도 방법임을 제시하였다.

즉, 최초 검색 결과에서 최상위 문헌이 적합 문헌일 가능성성이 높기 때문에, 최상위 문헌을 질의 확장에 사용하는 것이 일반적 방법으로 통용되지만, 검색시스템에 따라서는 그렇지 않을 가능성이 존재하므로 질의 확장에 적용할 문헌의 선택도 실험에서는 매우 중요한 요소로 간주 된다.

Rocchio(1971)는 적합성 퍼드백에서의 질의 확장을 <수식 3>같이 수행하였다.

$$Q = \alpha \cdot Q + \beta \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} R_i - \gamma \cdot \frac{1}{n_2} \sum_{i=1}^{n_2} S_i$$

<수식 3> 표준 로치오 공식

<수식 3>은 표준 로치오 공식으로 기존 질의 (Q)에 적합 문헌(R)의 벡터값을 가하고 반면 부적합 문헌(S)의 벡터값을 감하는 방식으로 기존질의와 각 적합 문헌의 벡터의 합, 부적합 문헌의 벡터의 합에는 각 α , β , γ 등의 가중치를 부여 하여 질의를 수정해 확장한다.

더불어 <수식 3>의 값은 모두 벡터로 Rocchio는 Salton(1965)의 벡터 공간검색에서 실험을 수행하였는데, 질의 용어와 문헌의 용어 간의 연결과 인덱싱 및 개념적 공간에서의 유사성이 필요한 이유임을 밝혔다.

문헌 내 용어의 유사도를 코사인 유사도나 TD-IDF로 표현해 자동으로 색인을 형성하여 구현하기 위해 등장한 벡터 공간 모형(Salton, Wong, & Yang, 1975)은 Salton과 McGill이 1983년 검색시스템으로 개발하면서, 기존 불리언 모형이 단순히 질의어와 문헌 용어의 일치로 검색 결과를 보이는 것 반해 질의어와 문헌

내 용어의 가중치에 따른 문헌의 유사도를 수치화하여 순위로 확인할 수 있도록 한 것이 가장 큰 특징이라 할 수 있다.

Rochhio의 질의 확장 이후 지역적 확장은 전통적으로 Salton과 McGill(1983)의 벡터 공간 모형 검색시스템에서 수행되었으나, 최근 들어 재순위화에서 언급한 것처럼 질의 확장은 대규모 문헌 집단에 기반하므로 BERT나 GPT를 사용(Zheng et al., 2020; Claveau, 2022)하는 등 자연어 처리 모델을 활용해 검색시스템에 반영하기도 하며, 기계 학습을 사용하여 질의 확장을 수행하기도 한다(Xu & Croft, 2017; Pereira & Etemad, 2020; Wang et al., 2021).

3. 연구방법

본 연구는 인용색인 데이터베이스의 데이터를 기반하여 검색시스템에서 인용 지표를 활용해 검색의 성능 향상을 모색하는 방법을 탐구하는 목적으로, 우선 Web of Science에서 실제 검색 실험을 진행하여 검색 결과를 획득하고, 이후 결과를 기반으로 적합성 판단과 각 서지데이터에 해당하는 인용 지표를 수집하였다. 그리고 벡터 공간모형 기반의 검색시스템을 구축하여 한정적인 공간에서의 검색 실험을 수행하고, 인용 지표를 사용한 재순위화와 질의 확장을 통해 검색성능에 미치는 영향을 파악하였다.

3.1 질의 생성

검색 결과를 획득하기 위해서는 무엇보다도 검색어가 선행되어야 한다. 본 연구는 Web of

Science에서 제공하고 있는 24개의 학문 대 분류 중 문헌정보학 분야를 선정하고, 검색을 위한 질의를 구축하였다. 이를 위해 Web of Science에서 2000년부터 2019년까지 총 66,830건의 문헌정보학 분야의 학술논문을 수집하여 Dynamic Topic Model(DTM)을 사용해 문헌정보학 분야의 주요 토픽을 분석하고(Gao et al., 2022), Web of Science에서 2001년부터 2021년 10월

까지의 문헌정보학 분야의 학술논문 55,442건을 기반으로 Latent Dirichlet Allocation(LDA)과 BERTopic을 사용하여 최근 문헌정보학계의 주제 영역을 탐색한 선행연구(김선욱, 양기덕, 2022)를 고찰해 질의 분야와 키워드를 〈표 2〉와 같이 선정하였다.

이어서 본 연구는 〈표 3〉과 같이 질의문의 키워드와 관련 질의개요문을 작성하였다.

〈표 2〉 질의 선정 영역과 키워드

질의	주제 영역		질의어 사용 키워드
	대주제	특정 영역	
1 문헌정보학 일반	정보 지식 조직과 개념, 이론 연구		information, knowledge, concept, practice, theory
2 정보검색	웹에서의 문서 기반 검색시스템 개발 연구		web, retrieval, search, information, document
3 정보서비스	대학도서관에서의 전자 접근 서비스 이용 연구		academic, library, service, digital, access
4 데이터 사이언스	시간적 요소를 포함한 네트워크 분석 연구		data, network, analysis, time, measure
5 의료 정보	환자 의료 정보를 활용한 학습데이터 관련 연구		health, patient, medical, learn, data
6 이용자 정보행태	웹상에서의 이용자 정보행태 연구		web, internet, social, user, behavior
7 계량 서지	논문 수준에서의 저널 인용색인의 영향력 지표에 관한 연구		journal, citation, index, article, impact
8 도서관 경영	도서관 마케팅 및 조직 경영 개발을 위한 연구		develop, market, business, management, organizations
9 정보 리터러시: 고등교육	대학도서관에서의 학생 대상의 정보 리터러시 교육 연구		university, library, student, literacy, education
10 정보 리터러시: 학교 도서관	학교 도서관(초중등) 학생 대상의 정보 리터러시 교육 연구		school, library, student, literacy, education

〈표 3〉 질의개요문 예시(질의 10)

- 질의 영역: 정보리터러시 - 학교도서관
- 키워드: school library student literacy education
- 개요: 초, 중, 고등학교의 학생을 대상으로 한 정보리터러시 교육에 해당하는 연구를 담은 문헌에 대해 적합성을 표기
- 비고: 유치원생을 포함한 정보리터러시 교육도 적합한 것으로 간주한다.
대학도서관 정보리터러시 교육이 내용을 포함한 연구일 경우, 연구의 중심이 학교도서관일 경우에만 적합한 것으로 간주한다.
학교 전 구성원을 대상으로 한 정보리터러시 교육 연구에서 학생을 포함한 내용이 있을 경우, 부적합한 것으로 간주한다.

본 연구에서의 검색은 연구자가 앞서 선정한 질의를 Web of Science에 입력하여 검색 결과를 획득하는 과정이며, 적합성 판단은 실제 평가자에게 검색 결과를 배포하여 질의개요문과의 적합성을 확인하여 판단하는 과정이다. 이에 질의개요문은 검색 결과의 적합성 판단에서 평가자들이 더욱 수월한 판단을 도모하기 위한 목적을 내포하고 있다.

〈표 3〉은 〈표 2〉에서 설정한 질의 영역과 질의 키워드에 해당하는 질의개요문의 예시로 질의10 '정보 리터러시-학교 도서관'에 대한 내용이다.

3.2 데이터 수집

3.2.1 서지 데이터

(1) 검색 결과 데이터

Web of Science가 제공하는 주요 검색항목은 모든 필드, 주제, 제목, 저자, 저널명, 출판연도, 기관명, 연구비 지원 기관, 출판사 등이 있으나, 본 연구는 질의 하나로 모든 필드 검색이 가능한 '모든 필드' 항목을 사용하였다.

이렇게 수집한 서지데이터는 10개 질의 결과 총 3,467건이었으며, 서지정보는 저자명, 저자전체명(Full name), 논문명, 학술지명, 초록, 출판연도, 문서 유형, 저자 키워드, 키워드 플러스(Keyword Plus)⁴⁾, 논문 수(Citation count), Web of Science 제어번호 등으로 총 10항목이다.

(2) 저자 데이터

본 연구는 인용 지표 중 저자 수준의 지표인

h-index를 활용하기 위하여 Web of Science가 제공하는 h-index를 수집하였다. 단, h-index의 경우 제1 저자의 h-index만을 사용하였는데, 단독연구의 경우는 저자가 1인이므로 h-index 수집에 있어 1인을 고려하나, 공동연구의 경우 저자의 모든 h-index를 수집하여 적용하는 것은 다소 무리가 있다고 판단하였으며, 또한 일반적으로 공동연구의 경우에서도 제1 저자가 이바지한 바를 높게 측정하는 경우가 존재하므로, 본 연구에서도 공동연구의 제1 저자만의 h-index를 고려하였다.

h-index를 수집하기 위하여 3,467건의 서지데이터에 해당하는 저자명 추출하였는데, 동명이인 식별을 위하여 저자의 전체명(full name)과 논문명을 부차적으로 활용하였다. 그 결과 서지데이터 3,467건에 대한 중복을 제외한 제1 저자는 총 2,667인이었다.

(3) 학술지 논문 데이터

본 연구는 질의 확장을 위하여 벡터공간모델 기반의 검색시스템을 구축하였다. 그리고 이를 위한 데이터로 SSCI에 수록된 문헌정보학 분야 학술지 85종에서 2000년 1월부터 2021년 12월까지 게재된 학술논문 90,948건의 서지데이터를 수집하였다.

수집 방식은 앞서 검색 결과 데이터와 동일하나 검색 질의로 학술지명을 입력한 후 도출된 검색 결과를 '내보내기' 기능으로 Excel 파일 형식으로 변환하고, 서지레코드를 알맞게 지정하였다. 내려받은 데이터 중 논문명과 초록이 누락 된 서지데이터 30,214건은 데이터로

4) 키워드 플러스는 저자 키워드와는 별도로 Web of Science에서 논문 콘텐츠를 고려해 부여하는 키워드이다.

서의 가치가 없음으로 판단하고 연구데이터에서 제외하였으며, 최종적으로 실험을 위해 수집한 데이터는 총 60,734건이다.

3.2.2 데이터 전처리

실험을 위해 수집한 서지데이터의 최대 인용수는 14,459회이며, 최소 인용수는 0회, 평균 23.38 회, 편차는 251.76이었다. 그리고 최대 h-index는 168, 최소 h-index는 0, 평균 10이었으며, 편차는 13.15인 것으로 나타났다.

따라서 본 연구는 원 데이터 수치의 사용으로 인한 실험에서의 오류 생성을 방지하고자 각 인용 지표의 최대 범주를 0에서 1 사이로 정규화하여 실험에 적용하였다.

또한 본 연구는 실험모델에서 문헌 용어로 사용할 논문명과 초록에 대해 첫째, 불용어(stopword)를 제거하였다. 둘째, 초록에 작성된 단어에 대하여 토크나이징(tokenizing) 처리를 진행하였다.셋째, 단어들의 어간을 추출하는 스테밍(stemming) 처리를 진행하였다. 더불어 Elsevier 社의 저작권 표기 문구가 작성된 초록이 다량 존재하여 다음과 동일하거나 유사한 저작권 문구를 삭제하였다.

(c) 2005 Elsevier B.V. All rights reserved.

(c) 2014 The Authors. Published by Elsevier Inc.

(c) 2008 Elsevier Inc. All rights reserved.

3.2.3 적합성 판단

본 연구는 Web of Science에서 각 질의에 따라 관련성 기준으로 전체의 검색 결과를 수집하였다. 그리고 실제의 적합성 판단데이터를 획득하기 위해 전공자 3인⁵⁾이 질의 생성 단계에서 작성된 질의개요문(〈표 3〉 참조)을 토대로 검색 결과에 관한 적합성 판단을 수행하도록 하였다.

적합성을 판단하는 평가자들에게는 질의의 검색 결과로 나타난 문헌의 논문명과 초록을 제공하였으며, 때에 따라 원문을 파악하여 질의와 검색 결과 문헌의 적합 여부를 적합성 판단 평가자에 기표하도록 하였다. 그리고 문헌의 적합 판단이 모호하지만, 조금이라도 적합하다고 여겨진다면 적합하다고 표시할 수 있도록 안내하였다. 이는 TREC에서 최소한이라도 적합하면 적합 문헌으로 간주한다는 적합성 판단 기준에 기인한 것이다(Voorhees & Harman, 2005).

평가자들은 Web of Science에서의 최초 질의 검색 결과 문헌 3,467건, 벡터 공간모형 검색시스템에서의 최초 검색 결과에서 추가로 나타난 문헌 703건, 질의 확장 이후의 검색 결과에서 추가로 나타난 문헌 445건 등으로 1인당 총 4,615건의 문헌에 대한 적합성 판단을 수행하였으며, 최종적으로 수집한 적합성 판단 건수는 13,845건이다.

본 연구는 〈표 4〉처럼, 이렇게 수집한 평가자의 적합성 판단데이터를 이진(binary) 적합과 부분(particinal) 적합으로 구분하여, 문헌에 적용해 최종 적합성을 판단하였다.

5) 전공자 3인은 최소 문헌정보학 석사학위 이상 및 관련기관 경력 3년 이상 혹은 박사학위자로 구성하였다. 각 질의에 대한 자세한 설명과 적합성 판단 여부에 대한 안내를 미리 전달하였으며, 적합성 판단을 위한 소요 기간으로 약 2주간 진행하였다.

〈표 4〉 문헌의 적합 산정 기준

점수	구분		산정 기준
	이진	부분	
3	적합	완전 적합	3인 중 3인 적합
2		부분 적합	3인 중 2인 적합
1	부적합	부분 적합	3인 중 1인 적합
0		완전 부적합	3인 중 0인 적합

〈표 5〉에 따르면 이진 형식에서 질의 문헌당 평균 약 42% 정도가 적합 문헌인 것으로 나타났으며, 가장 낮은 적합 문헌을 보여준 질의는 8번 질의로 문헌의 약 21%가 적합한 것으로 확인되었고, 가장 높은 적합 문헌을 보여준 질의는 7번 질의로 57% 정도가 적합한 것으로 나타났다.

질의마다의 적합 문헌과 부적합 문헌의 비율은 다소 상이하나, 적합성 판단에 따른 검색 결과의 부적합 문헌 비율이 적합 문헌 비율 보다 높게 나타난 것은 선행연구(김홍렬, 이두영, 1999; Tayler, 2012)에서도 언급되었던바, 본 실험에 사용할 적합성 판단은 적절히 수행된 것으로 판단하였다.

한편, 부분 적합의 경우 〈표 6〉에서와 같이

이진 형식에서 적합 문헌으로 판단된 1,601건의 문헌 중 완전 적합한 문헌은 847건으로, 약 52% 정도에서 평가자 모두 적합 문헌으로 여겼다.

이는 검색시스템 이용자가 검색 결과에 대하여 적합 문헌을 판단할 때 시스템 적합성과 이용자 적합성이 상이할 수 있으며, 또한 이론적 배경에서 언급한 것처럼 이용자마다 스스로 생각하는 적합성 기준이 다르기 때문이라는 선행 연구의 결과를 본 적합성 판단 수행에서도 여실히 보여준다.

그리고 질의 8은 질의 10건 중에서 이진 적합성 21%로 질의 중에서 가장 낮은 적합 문헌 비율을 보였으며, 이중 완전 적합은 14건으로 전체문헌의 약 5%를 차지하였다. 이에 질의8은 타 질의군에 비하여 평가자들이 매우 보수적으

〈표 5〉 WoS의 검색 결과에 대한 적합성 판단 결과(이진 형식)

질의	검색 결과 문헌수	적합 문헌수		부적합 문헌수	
		적합 문헌수	비율	부적합 문헌수	비율
1	193	48	25%	145	75%
2	329	185	56%	144	44%
3	277	111	40%	166	60%
4	128	64	50%	64	50%
5	290	113	39%	177	61%
6	101	55	54%	46	46%
7	1,000	572	57%	428	43%
8	240	51	21%	189	79%
9	659	341	52%	318	48%
10	250	61	24%	189	76%

〈표 6〉 WoS의 검색 결과에 대한 적합성 판단 결과(부분 적합 형식)

질의	완전 적합(3)	부분 적합(2)	부분 적합(1)	부적합(0)	총 문헌수
1	17	31	42	103	193
2	112	73	84	60	329
3	57	54	61	105	277
4	18	46	42	22	128
5	65	48	85	92	290
6	24	31	37	9	101
7	380	192	241	187	1000
8	14	37	49	140	240
9	137	204	38	280	659
10	23	38	49	140	250
문헌수	847	754	728	1,138	3,467
표준편차	106.47	62.47	58.50	75.71	-

로 적합성을 판단한 것으로 확인할 수 있었다.

이는 〈표 2〉에서의 질의 8번의 주제 영역인 ‘도서관 경영’과 선택한 질의어 ‘develop, market, business, management, organizations’의 연관성이 부족하여 나타난 결과로 판단하였는데. 본 연구에서는 선행연구를 토대로 주제 영역을 선정하고 질의어를 조합하였으나, 두 선행연구 모두 문헌정보학 분야의 학술논문을 활용했다는 점에 있어 ‘library’라는 용어가 일반적인 용어로 분류되어 ‘도서관 경영’이라는 주제 영역에서 등장하지 않았다. 그러므로 본 연구는 질의 8의 적합성 판단 결과를 이상치로 판단하고, 검색 결과 성능 평가에서 제외하기로 하였다.

3.3 검색시스템 구축

3.3.1 실험환경

본 연구는 실험을 위하여 OS Windows 11의 컴퓨팅 연산 장비 1대를 사용하였다. 또한 성능

측정 도구로 사용한 pyTREC_eval⁶⁾ (Gysel & Rijke, 2018) 패키지의 경우, 특정 프로그래밍 언어 버전의 최적 사양을 명시하였기에 Python 3.5버전과 Visual C++ 14.0버전에 한정한 가상 Python 환경을 별도로 구현하여 성능 측정을 수행하였다. 한편 실험 데이터 수집과 정리, 실험 결과의 분석 등에는 Windows 기반 스프레드시트 소프트웨어인 MS Excel과 Python 패키지를 사용하였다.

3.3.2 데이터셋

본 연구는 질의 검색을 통해 수집한 문헌의 서지데이터 3,467건과 문헌정보학분야 SSCI 학술지 85종에 게재된 문헌의 서지데이터 60,734 건에 기반하여 각각의 데이터셋을 구축하였다. 기본적인 데이터셋의 문헌 레코드 주요 속성은 총 11요소로 질의번호, 문헌 번호, 검색 결과 순위, 논문명, 저자명, 초록, 제 1저자의 h-index, 논문 수, 출판 연도, 이진 형식의 적합성 판단

6) https://github.com/cvangysel/pyTREC_eval

결과, 부분 적합 점수로 구성되었으나, 질의 확장용 벡터공간모형 검색시스템 구축에 사용한 60,734건의 데이터셋의 경우, h-index 수집을 진행하지 않았으므로 h-index 관련 요소는 속성에서 제외하였다.

3.3.3 벡터 공간모형 검색시스템 구축

벡터 공간모형에서의 문헌과 질의 간의 유사도 측정은 코사인 유사계수를 사용하였으며, 이때 적용한 용어 가중치는 〈표 7〉과 같다.

〈표 7〉 유사도 도출을 위한 산출식

질의	$TF = tf$
문헌	$TF = \log(tf) + 1$
	$IDF = \log\left[\frac{1+N}{1+df(t)}\right] + 1$

질의의 경우 각 질의에 출현하는 단어마다의 가치가 동등하므로 단순 TF를 사용하였다. 다만, 질의의 모든 단어는 한번 출현하였기에 실제로는 이진 TF와 같다. 그리고 문헌에 대한 단어 빈도의 경우, Buckley와 Salton과 Allan이 제안한 SMART시스템(1993)의 logTF를 사용함으로써, 문헌에 각 최소 출현 및 최대 출현의 단어에 대한 영향력을 보완하고자 하였다.

한편, 질의에 따른 문헌의 유사도 산출은 기본적으로 질의 확장을 위한 검색시스템을 위하여 수행된 것이므로 학술지 논문 서지데이터 60,734 건을 우선 산출하였다.

그리고 검색시스템 구축에는 활용되지 않거나 재순위화의 속성으로 사용하기 위해 Web of Science 초기 검색 결과의 문헌 서지데이터 3,467건에 대해서도 유사도 산출을 수행하면서 결과적으로 총 64,201건의 서지데이터의 유사

도를 산출하였다.

3.3.4 검색성능의 측정

본 연구는 이진 형식과 부분 적합 형식으로 구분한 적합성 판단 데이터와 각 문헌 데이터의 지표별 순위 데이터를 조합한 결과치를 사용하여 검색 결과의 상위 100위까지의 성능을 측정하였다.

이에 따라 본 연구는 이진형식의 평가척도인 Precision at k와 부분 적합 형식의 성능 측정 척도인 nDCG를 사용하여 성능을 확인하였다. 또한 임의 순위 k는 상위 10순위(p@10), 20순위(p@20), 30순위(p@30), 50순위(p@50), 100순위(p@100)로 지정하며, nDCG도 마찬가지로 위와 같은 k 값을 적용하여 두 가지 측정 척도가 동일 순위에서의 성능을 나타낼 수 있도록 하였다.

3.4 재순위화

본 연구는 순위에 대하여 실제 적합성 판단을 수행한 데이터와 인용 지표를 사용하여 Web of Science 검색 결과 데이터를 재순위화하고 성능을 측정하였으며, 세부 설명은 다음과 같다.

첫째, 결과 데이터에 인용지표순으로 재순위화를 진행하였다. 둘째, 질의어에 해당하는 단어 출현 빈도에 대하여 고유로 등장하는 질의어 횟수와 질의어의 총 출현 횟수를 토대로 인용지표를 조합하여 재순위화를 진행하였다. 셋째, 질의와 문헌 간의 코사인 유사도를 토대로 인용지표를 조합하여 재순위화를 진행하였다.

이러한 방식으로 본 연구에서 수행한 재순위화의 종류는 총 11건이다.

3.5 질의확장

본 연구는 벡터 공간모형에 기반한 검색시스템이므로, 표준 로치오 공식(〈수식 3〉 참조)을 사용해 질의 확장을 진행하였다.

일반적으로 로치오를 사용하여 질의 확장을 수행한 선행연구에서는 α 파라미터를 1로 고정해 초기질의의 변동을 두지 않고, 초기질의로서의 단일 역할만 주로 수행하도록 하는 경우가 있으므로, 본 연구 역시 α 파라미터에 대하여 1로 고정하였다.

Rocchio(1971)는 적합 문헌의 가중치 β 에 대해 부적합 문헌의 가중치 γ 보다 커야 하며 그 합은 1이어야 한다고 하였다. 그러나 본 연구에서 도출한 문헌의 코사인 유사도는 질의에 비하여 상대적으로 매우 긴 문헌으로서 그 유사도의 최대치가 0.6 이상의 결과 값은 가지지 못하였다. 마찬가지로 최저치는 0으로 최대와 최소의 차이가 일반적인 코사인 유사도의 형태와 상이하였다. 따라서 유사도가 최저인 문헌에 γ 의 가중치를 적용할 경우, 상대적으로 높은 유사도를 지닌 문헌이라 할지라도, 절대적인 유사도가 낮아서 이러한 경향이 질의 확장에 영향을 미칠 수 있을 것이라 추측하였으므로, 본 연구에서는 γ 를 적용하지 않고, 적합 문헌의 가중치인 β 만을 적용하는 방식으로 질의 확장을 수행하였다. 더불어 선행연구에서 언급한 대로 γ 가 0인 경우, β 는 1이 되어야 하나, 실험 대상의 문헌의 전반적인 유사도 값이 낮고, 질의 확장 이후 이차 검색 결과에 인용 지표와 적합성 판단데이터를 가중하여 적용할 것이므로, β 를 0.5로 고정하여 실험을 수행하였다.

한편, Web of Science의 검색 결과에 대한

재순위화 실험에서는 기본 인용 지표로 저자 수준 지표인 h-index와 논문 수준 지표인 인용 수를 사용하였지만, 질의 확장에서는 SSCI 85 종의 학술논문 데이터 60,734건으로 질의에 대한 벡터 공간모형 검색시스템을 구축하고, 인용수만을 실험에 사용하였다. 인용수의 경우 Web of Science에서 현재 서지데이터를 내려 받을 시점의 수치를 함께 제공받을 수 있으나, h-index의 경우 그렇지 못하기 때문에 연구자가 문헌의 저자와 원문 및 저자 검색을 통하여 h-index를 수집해야 한다. 이러한 과정은 실험을 준비하기까지의 다량의 시간이 소요되므로 불가피하게 h-index를 활용한 질의 확장은 수행하지 못하였다.

그리고 본 연구의 질의 확장은 데이터셋이 다른 별도의 검색시스템을 구성하여 인용 지표의 영향을 파악하는 것으로 Web of Science와는 별개의 결과로 확인하였다. 그러나 질의 이후 검색 결과는 Web of Science의 검색 결과와 동일한 양을 출력하기 위하여, Web of Science에서 질의마다 검색 결과로 도출된 최대 순위와 동일하게 질의 확장의 결과로 결정하고 질의 확장에 대한 재순위화 및 성능을 파악하였다.

또한 본 연구는 질의 확장을 위해 문헌 단위로 범주를 설정하였다. 선행연구에 따르면 상위 최소 문헌일 때 성공적인 질의 확장을 도모한다고 하였으나(Harman, 1988: 1992; Xu & Croft, 2017), 해당 문헌이 부적합 문헌이라면 그 결과 역시 부정적인 결과를 초래할 가능성 이 높다. 이러한 상황을 방지하기 위하여 본 연구는 상위 3 문헌에 해당하는 질의 확장을 수행하였다. 이에 벡터 공간모형 검색시스템에서 유사도가 가장 높은 순위의 상위 3 문헌을 적용한

질의 확장 방식, 인용수에서의 상위 3 문헌을 적용한 질의 확장 방식, 유사도 상위 문헌 중 인용수 상위 3 문헌을 적용한 질의 확장 방식, 유사도 상위 문헌 중 실제로 적합 문헌으로 판정된 3 문헌을 적용한 질의 확장 방식, 그리고 유사도 상위 문헌 중 실제로 적합 문헌으로 판정된 인용수 상위 3 문헌을 적용한 질의 확장 방식으로 총 5가지의 질의 확장을 수행하고 성능을 측정하였다.

4. 분석

4.1 검색시스템 기본 성능 평가

4.1.1 Web of Science 성능

Web of Science의 성능으로 10순위 기준 평균은 0.80, 상위 100순위 기준 평균 0.67로 나타났으며, 질의 2, 3, 7, 9는 10순위 기준 성능이 1인 것으로 10건의 문헌 모두 적합한 것을 <표

8>과 같이 확인할 수 있다. 이 중 질의 2와 9는 모두 완전 적합한 문헌인 것으로 나타났으며, 부분 적합점수에 따른 성능의 100순위 기준 성능은 평균은 0.88인 것으로 <표 9>를 통해 알 수 있다.

그리고 부분 적합 점수의 순위별 성능 편차는 0.07에서 0.09인 반면, 이진 적합 순위별 성능 편차는 100순위 기준 최대 0.20으로 나타나는데, 이는 연구 방법의 <표 4>와 <표 6>처럼 비록 이진에서는 부적합 문헌이더라도 부분 적합에서는 1점인 문헌이 존재하여 전반적인 성능치가 이진 적합보다도 높은 것에 기인하였다.

한편, 본 연구는 Web of Science의 인용색인 데이터를 사용하였기에 그 검색 결과에 해당하는 성능을 측정하였으나, Web of Science의 검색시스템 알고리즘에 대하여 공식적으로 알려진 바가 없으므로, 본 연구에서 진행된 재순위화나 질의 확장 실험의 성능과 Web of Science의 검색시스템과의 직접적인 비교는 불가능하다고 판단하였다.

<표 8> WoS의 질의별 기본 성능(p@k)

질의	10	20	30	50	100
1	0.50	0.60	0.63	0.50	0.37
2	1.00	1.00	1.00	0.98	0.97
3	1.00	0.85	0.87	0.88	0.84
4	0.80	0.70	0.70	0.62	0.55
5	0.70	0.75	0.73	0.66	0.63
6	0.60	0.70	0.70	0.70	0.55
7	1.00	1.00	0.93	0.92	0.94
9	1.00	1.00	0.97	0.90	0.77
10	0.60	0.55	0.50	0.42	0.38
평균	0.80	0.79	0.78	0.73	0.67
편차	0.18	0.16	0.15	0.18	0.20
최대	1.00	1.00	1.00	0.98	0.97
최소	0.50	0.55	0.50	0.42	0.37

〈표 9〉 WoS의 질의별 기본 성능(nDCG@k)

질의	10	20	30	50	100
1	0.92	0.89	0.88	0.84	0.81
2	1.00	1.00	0.99	0.98	0.96
3	0.94	0.91	0.91	0.92	0.94
4	0.94	0.93	0.92	0.89	0.87
5	0.95	0.95	0.93	0.90	0.86
6	0.89	0.89	0.88	0.89	0.88
7	0.98	0.99	0.96	0.97	0.98
9	1.00	0.99	0.98	0.96	0.90
10	0.69	0.69	0.72	0.70	0.69
평균	0.92	0.92	0.91	0.89	0.88
편차	0.09	0.08	0.07	0.08	0.08
최대	1.00	1.00	0.99	0.98	0.98
최소	0.69	0.69	0.72	0.70	0.69

4.2 재순위화에 따른 성능 평가

본 연구는 인용수(c), h-index(h)를 적용한 재순위화와 결과 문헌에 출현한 고유 질의어의 수(Unique Query Number: UQN)와 질의어의 총 출현빈도(Term Frequency: TF), 결과 문헌의 코사인 유사도(Cosine Similarity: CoS) 등을 활용하면서 인용 지표를 조합한 재순위화

를 수행하고 성능을 측정하였다.

4.2.1 인용 지표

Web of Science 검색 결과 데이터에 대하여 h-index와 인용 수를 이용하여 재순위화한 결과에 이진 적합을 적용한 성능은 〈표 10〉, 〈표 11〉과 같다.

h-index에 따른 100순위 기준에서 이진 적

〈표 10〉 h-index 순위 성능(p@k)

질의	10	20	30	50	100
1	0.30	0.40	0.27	0.22	0.23
2	0.50	0.70	0.70	0.70	0.59
3	0.30	0.30	0.30	0.34	0.38
4	0.40	0.40	0.47	0.50	0.50
5	0.40	0.25	0.23	0.26	0.31
6	0.60	0.60	0.57	0.58	0.54
7	0.70	0.65	0.60	0.62	0.56
9	0.50	0.40	0.53	0.58	0.48
10	0.40	0.25	0.27	0.20	0.31
평균	0.46	0.44	0.44	0.44	0.43
편차	0.12	0.15	0.16	0.17	0.12
최대	0.70	0.70	0.70	0.70	0.59
최소	0.30	0.25	0.23	0.20	0.23

〈표 11〉 인용 수 순위 성능(p@k)

질의	10	20	30	50	100
1	0.20	0.20	0.27	0.28	0.26
2	0.60	0.65	0.67	0.64	0.61
3	0.10	0.30	0.47	0.40	0.44
4	0.50	0.70	0.67	0.56	0.54
5	0.10	0.35	0.40	0.32	0.33
6	0.40	0.50	0.50	0.54	0.54
7	0.30	0.55	0.50	0.52	0.55
9	0.50	0.50	0.47	0.46	0.49
10	0.20	0.20	0.20	0.28	0.27
평균	0.32	0.44	0.46	0.44	0.45
편차	0.17	0.17	0.14	0.12	0.12
최대	0.60	0.70	0.67	0.64	0.61
최소	0.10	0.20	0.20	0.28	0.26

합 성능은 0.43으로 확인할 수 있었으며, 편차는 50순위 성능에서 최대 0.17, 최소는 10순위와 100순위 기준에서 0.12인 것으로 나타났다.

그리고 인용수의 경우 100순위 기준 평균 0.45의 성능을 보였으며, 10순위와 20순위 기준에서 편차가 0.17, 50순위와 100순위 기준에서

편차가 0.12인 것으로 나타났다.

한편, h-index와 인용 수에 부분 적합 점수를 적용하여 순위화 한 결과의 성능은 〈표 12〉와 〈표 13〉과 같다. h-index의 100순위 기준 평균 성능은 0.80으로, 단위 순위 기준 중에서 가장 높은 성능이었으며, 인용 수에서의 100순위 기준 평균 성능은 0.78로 나타났다.

〈표 12〉 h-index 순위 성능(nDCG@k)

질의	10	20	30	50	100
1	0.80	0.80	0.76	0.74	0.73
2	0.95	0.93	0.92	0.90	0.88
3	0.92	0.82	0.78	0.79	0.80
4	0.90	0.83	0.81	0.80	0.88
5	0.65	0.69	0.69	0.71	0.78
6	0.67	0.74	0.75	0.79	0.87
7	0.97	0.93	0.93	0.93	0.88
9	0.60	0.59	0.64	0.70	0.66
10	0.67	0.68	0.70	0.63	0.69
평균	0.79	0.78	0.78	0.78	0.80
편차	0.13	0.10	0.09	0.09	0.08
최대	0.97	0.93	0.93	0.93	0.88
최소	0.60	0.59	0.64	0.63	0.66

〈표 13〉 인용 수 순위 성능(nDCG@k)

질의	10	20	30	50	100
1	0.70	0.67	0.69	0.68	0.67
2	0.85	0.86	0.86	0.86	0.87
3	0.52	0.54	0.65	0.64	0.74
4	0.92	0.91	0.90	0.88	0.84
5	1.00	0.84	0.83	0.77	0.80
6	0.95	0.85	0.84	0.85	0.90
7	0.68	0.72	0.71	0.77	0.84
9	0.81	0.80	0.78	0.76	0.69
10	0.69	0.66	0.68	0.71	0.66
평균	0.79	0.76	0.77	0.77	0.78
편차	0.14	0.11	0.08	0.07	0.08
최대	1.00	0.91	0.90	0.88	0.90
최소	0.52	0.54	0.65	0.64	0.66

이로써 〈표 8〉의 Web of Science 평균 성능이 이진 적합 100순위 기준 0.67로, h-index 성능보다 0.24 높으며, 인용수 성능보다 0.22 높은 것에 비하여, 〈표 9〉와 같이 Web of Science의 부분 적합 점수 100순위 성능 평균이 0.88이며, h-index 성능보다 0.08 높고, 인용수 성능보다 0.1 높은 것은, 상대적으로 부분 적합 점수에서의 성능이 이진 적합 성능보다 Web of Science 와의 성능에 근접한 것을 확인할 수 있었다. 그러나 단순하게 인용 지표를 재순위화한 성능은 Web of Science의 기본 성능보다 모든 순위

구간에서 성능이 하락하였다.

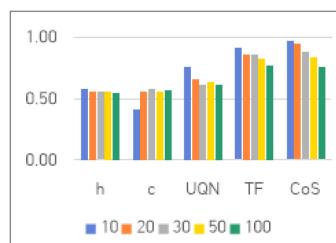
4.2.2 출현 빈도 및 유사도

인용 지표와 함께 질의어 출현 빈도, 고유 질의어 수, 코사인 유사도를 적용한 이진 적합의 재순위화 기본 성능은 〈표 14〉, 〈그림 1〉과 같다.

50순위 기준까지 코사인 유사도를 사용한 재순위화 방식이 가장 성능이 높았으며, 100순위 기준에서 출현 빈도가 성능이 높은 것으로 나타났다. 그리고 인용 지표만을 적용한 순위 성능의 경우, 모든 순위 기준에서 50% 이상의

〈표 14〉 인용 지표 및 검색 방식의 기본 성능(p@k)

기준	10	20	30	50	100
h	0.46	0.44	0.44	0.44	0.43
c	0.32	0.44	0.46	0.44	0.45
UQN	0.60	0.52	0.49	0.50	0.49
TF	0.72	0.68	0.68	0.65	0.61
CoS	0.77	0.75	0.70	0.66	0.60



〈그림 1〉 인용 지표 및 검색 방식의 기본 성능 그래프(p@k)

성능을 찾아보기 어려웠다. 또한 검색데이터에 h-index만을 적용한 성능은 모든 순위 구간에서 최하의 성능을 보여주었다.

〈표 15〉와 〈그림 2〉의 부분 적합 점수에서도 50순위 기준까지 코사인 유사도가 가장 성능이 높았으며, 100순위 기준에서는 출현 빈도가 성능이 높은 것으로 나타났다. 그리고 인용 지표만을 적용한 순위 성능의 경우, 전 순위 구간에서 0.7 후반의 성능을 확인하였고, 이진 적합의 결과와는 다르게 인용수가 가장 낮은 성능을 보여주었다.

〈표 16〉은 검색 방식에 인용 지표를 조합한 성능 결과를 종합한 것으로, 30위 기준이다.

이진 적합의 경우, 단순 적용에서 h-index가 인용수보다 0.02 낮은 성능을 보였으나, 출현 빈도와 조합하면 최대 37% 상승하여 0.69의

성능 확인할 수 있었으며, 부분 적합의 경우, 단순 적용에서 인용수가 h-index보다 0.01정도 낮은 성능을 보였으나, 출현 빈도나 고유 질의어 수에서 충 출현 빈도를 적용하면 10% 정도 성능이 상승한 0.85의 성능을 확인하였다.

〈그림 3〉과 〈그림 4〉는 〈표 16〉을 그래프화 한 것으로, 적합 구분에 따라 인용 지표의 단순 적용 성능이 상반되지만, 전체적으로 인용 지표를 직접 적용하는 것보다 검색 방식에 인용 지표를 조합해 부가적인 요소로 사용하는 것이 검색 결과의 성능에 영향을 미치고 있음을 확인할 수 있다.

특히, 고유 질의어 수에 충 출현 빈도를 적용한 검색 결과에 인용 지표를 보조적 도구로 사용할 때 성능 향상의 가능성을 확인하였고, 그 밖의 검색 방식에서는 기본 성능이 인용 지표

〈표 15〉 인용 지표 및 검색 방식의 기본 성능(nDCG@k)

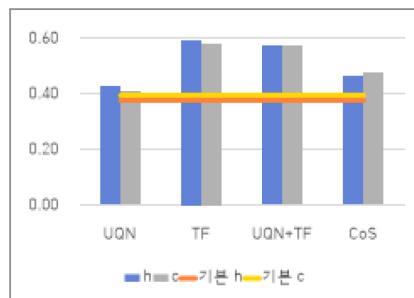
기준	10	20	30	50	100
h	0.79	0.78	0.78	0.78	0.80
c	0.79	0.76	0.77	0.77	0.78
UQN	0.82	0.80	0.79	0.80	0.81
TF	0.88	0.88	0.88	0.87	0.85
CoS	0.91	0.90	0.88	0.87	0.84



〈그림 2〉 인용 지표 및 검색 방식의 기본 성능 그래프(nDCG@k)

〈표 16〉 30위 기준 검색 방식과 인용 지표 조합 성능

적합 구분	이진 적합		부분 적합	
	h	c	h	c
인용 지표	0.44	0.46	0.78	0.77
단순 적용	0.50	0.47	0.80	0.79
UQN	0.69	0.67	0.84	0.85
TF	0.66	0.66	0.84	0.85
UQ+TF	0.54	0.55	0.80	0.83



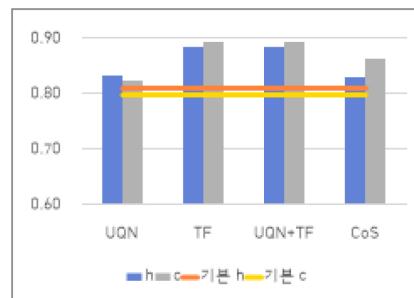
〈그림 3〉 인용 지표의 적용 성능(p@k)

를 조합한 성능보다 높아서, 여전히 인용 지표가 별개의 영역으로 작용하고 있음을 파악할 수 있었다.

4.3 질의 확장에 따른 성능 평가

4.3.1 벡터 공간모형 검색시스템 기본 성능

질의 확장의 성능을 확인하기 위하여 자체 구축한 검색시스템의 기본 성능을 측정한 결과는 〈표 17〉처럼 이진 적합에서 100순위 기준



〈그림 4〉 인용 지표의 적용 성능(nDCG@k)

평균 0.57의 성능을 보였으며, 10순위 기준은 0.66의 성능으로 순위 기준이 넓어질수록 그 성능이 낮아졌다. 그러나 각 순위 기준별 성능 편차가 0.22에서 0.23으로 거의 동일한 수준인 것은 순위 구간에서의 질의별 성능 차이가 구간마다 매우 상이한 수준은 아닌 것으로 확인되었다.

〈표 18〉의 부분 적합 적용 결과에서는 100순위 기준의 평균 성능은 0.67인 것으로 나타났으며, 앞서 재순위화의 여러 성능의 결과에서 나

〈표 17〉 벡터 공간모형 검색시스템 기본 성능(p@k)

질의	10	20	30	50	100
1	0.50	0.65	0.77	0.74	0.60
2	1.00	1.00	0.93	0.92	0.90
3	0.90	0.70	0.57	0.56	0.58
4	0.40	0.25	0.20	0.26	0.27
5	0.40	0.45	0.37	0.24	0.24
6	0.60	0.65	0.63	0.58	0.59
7	0.90	0.95	0.87	0.90	0.87
9	0.80	0.80	0.83	0.78	0.77
10	0.40	0.35	0.33	0.36	0.28
평균	0.66	0.64	0.61	0.59	0.57
편차	0.22	0.23	0.23	0.23	0.23
최대	1.00	1.00	0.93	0.92	0.90
최소	0.40	0.25	0.20	0.24	0.24

〈표 18〉 벡터 공간모형 검색시스템 기본 성능(nDCG@k)

질의	10	20	30	50	100
1	0.70	0.67	0.69	0.68	0.67
2	0.85	0.86	0.86	0.86	0.87
3	0.52	0.54	0.65	0.64	0.74
4	0.92	0.91	0.90	0.88	0.84
5	1.00	0.84	0.83	0.77	0.80
6	0.95	0.85	0.84	0.85	0.90
7	0.68	0.72	0.71	0.77	0.84
9	0.81	0.80	0.78	0.76	0.69
10	0.69	0.66	0.68	0.71	0.66
평균	0.79	0.76	0.77	0.77	0.78
편차	0.14	0.11	0.08	0.07	0.08
최대	1.00	0.91	0.90	0.88	0.90
최소	0.52	0.54	0.65	0.64	0.66

타난대로 이진 적합보다 부분 적합의 성능이 전체적으로 높으며, 편차도 다소 적은 것을 벡터 공간모형 검색시스템의 기본 성능에서도 확인할 수 있었다.

4.3.2 질의 확장

이진 적합 기준에서 시스템 피드백(Sys_feedback_CosSim) 적용, 시스템 피드백에 단순히

인용수를 적용(Sys_Feedback_Citation), 시스템 피드백의 최상위 문헌에 인용수를 적용(Sys_feedback_CosSim_Citation), 그리고 적합 문헌을 활용하여 이용자 피드백을 적용(User_Feedback_RelDoc.), 적합 문헌과 인용수를 모두 적용(User_Feedback_RelDoc_Citation) 하여 질의 확장을 수행한 결과는 〈표 19〉처럼 100순위 기준에서 벡터 공간모형의 기본 시스템이 0.57

〈표 19〉 질의 확장 성능(p@k)

기준	10	20	30	50	100
VSM_baseline	0.66	0.64	0.61	0.59	0.57
질의 확장	Sys_feedback_CosSim 1.7%	0.67 -1.7%	0.63 3.0%	0.63 -0.7%	0.56 -1.2%
	Sys_Feedback_Citation -8.5%	0.60 -5.2%	0.61 -3.0%	0.59 -1.1%	0.55 -2.7%
	Sys_feedback_CosSim_Citation -6.8%	0.61 -7.8%	0.59 -2.4%	0.60 -0.7%	0.55 -2.5%
	User_Feedback_RelDoc. 3.4%	0.68 -2.6%	0.63 3.6%	0.60 0.4%	0.56 -0.8%
	User_Feedback_RelDoc_Citation 6.8%	0.70 -5.2%	0.61 1.8%	0.60 0.4%	0.56 -0.4%

〈표 20〉 질의 확장 성능(nDCG@k)

기준	10	20	30	50	100
VSM_baseline	0.78	0.79	0.79	0.78	0.80
질의 확장	Sys_feedback_CosSim	0.83 7.0%	0.84 6.2%	0.84 6.3%	0.81 3.4%
	Sys_Feedback_Citation	0.76 -2.6%	0.77 -2.1%	0.77 -1.8%	0.78 -0.7%
	Sys_Feedback_CosSim_Citation	0.79 1.2%	0.80 1.2%	0.80 1.9%	0.79 0.6%
	User_feedback_RelDoc.	0.80 2.3%	0.79 0.6%	0.80 1.9%	0.80 0.8%
	User_Feedback_RelDoc_Citation	0.79 1.2%	0.79 0.0%	0.80 1.2%	0.81 -0.5%
					0.5%

로 가장 높은 성능을 보였다.

상위 30순위 기준에서는 적합 문헌을 사용한 이용자 피드백 상위 3 문헌으로 질의 확장한 성능이 3.6% 상승한 0.63으로 가장 높았으며, 상위 10순위에서는 이용자 피드백의 문헌 최상위 문헌 중 상위 인용 수의 적합 문헌으로 질의 확장한 결과가 기본 성능에 비해 6.8% 상승한 0.70로 가장 우수한 성능을 나타내었다.

반면 단순 인용 수 기준으로 시스템 피드백 한 질의 확장한 결과는 순위 기준마다 차이는 있으나, 전반적으로 성능이 저조하였으며, 10 순위 기준에서는 최대 8.5% 가량 하락하였다. 그리고 시스템 피드백으로 상위 5 문헌에 대하여 인용수 상위 3 문헌을 자동 적용한 질의 확장의 경우에도 전체 성능이 저하된 것을 확인 할 수 있었다.

〈표 20〉은 부분 적합에 따른 질의 확장 성능 결과로, 상위 10순위에서 유사도에 따른 시스템 피드백의 질의 확장이 7% 높은 0.83의 성능을 보였으나, 100순위에서 0.79로 가장 낮은 성능을 확인할 수 있었다. 한편 인용 수만 적용한 시스템 피드백 질의 확장의 성능은 이전 적합

의 결과와 마찬가지로 대부분의 순위에서 성능이 하락하였다.

〈표 21〉은 질의마다 상위 10순위까지 등장한 적합 문헌의 순위를 확인한 것으로, a는 유사도와 적합 문헌의 질의 확장한 순위이며, b는 a에 인용수를 적용한 질의 확장한 순위이다. ‘질의 1’의 경우 적합 문헌의 등장 수는 5건으로 같았으나, 인용수를 적용한 방식(b)의 검색 결과 6위에 적합 문헌이 등장하였다.

그리고 ‘질의 4’, ‘질의 5’, ‘질의 6’의 경우 b 방식이 a 방식보다 1건 가량 적합 문헌 수가 더 많았으며, ‘질의 2’, ‘질의 7’, ‘질의 10’은 적합 문헌 수와 순위가 동일한 것으로 확인되었다. 물론 ‘질의 9’와 같이 b 방식의 적합 문헌 수가 줄어들거나, ‘질의 3’처럼 상대적으로 하위 적합 문헌의 등수가 변동하는 사례가 있었지만, 본 현황을 통해 대체로 인용수가 순위 변동에 긍정적인 영향이 있는 것으로 확인할 수 있었다.

결과적으로 여타의 선행연구들과 마찬가지로 본 연구의 검색시스템에서의 질의 확장도 성능향상에 영향이 있음이 밝혀졌다. 그리고 이전

〈표 21〉 질의별 적합 문헌 상위 10순위

질의 1		질의 2		질의 3		질의 4		질의 5		질의 6		질의 7		질의 9		질의 10	
a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	4	3	3	5	3	3	3	3	3	5	5
6	4	4	4	4	10	7	8	5	6	4	4	4	4	4	8	8	8
10	6	5	5	5		10		10	7	7	5	5	5	5	9	9	9
	6	6	6	6					10	8	6	6	6	6	10	10	
	7	7	7	7						10	8	8	7	7			
	8	8	8	8							9	9	10				
	9	9	9	10							10	10					
	10	10															

의 재순위화에서와 같이 인용수를 단독으로 사용했을 경우에는 검색성능에 특별히 기여하지 못하는 것으로 나타났다.

한편, 질의 확장에서 적합 문헌을 적용하면, 유사도만으로 질의 확장 하였을 때보다 이진 형식, 상위구간일수록 좋은 성능을 가져왔으며, 특히 인용 수를 보조적으로 사용한 경우, 최상 위권에서의 순위 변동에 영향이 있음을 확인할 수 있었다.

5. 결 론

본 연구는 인용색인 데이터베이스의 검색 결과를 바탕으로, 인용 지표가 인용색인 검색시스템 성능에 미치는 영향과 성능향상의 가능성을 파악하고자 하였다. 이를 위하여 본 연구는 Web of Science에서 문헌정보학 분야에서 10개의 질의를 생성하고, 이에 따른 검색 결과 3,467건의 문헌과, 2000년부터 2021년까지 SSCI 문헌정보학 분야 저널 85종에 수록된 60,734건의 문

헌에서 검색을 수행하고 이에 따른 결과 상위 100순위에 대한 성능과 검색 방식과 인용 지표를 활용한 재순위화, 그리고 벡터 공간모형 검색시스템 구축 등에 따른 질의 확장 실험을 수행하였다.

우선 Web of Science의 검색 결과에 대하여 실제로 이용자가 적합성 판단을 수행하였을 경우, 주제 영역에 따라 다소 편차가 존재하나, 평균 질의당 42건 정도 적합한 문헌인 것으로 나타났다.

그리고 인용 지표를 이용한 재순위화 및 질의 확장 실험을 수행한 결과는 다음과 같다.

첫째, 인용 지표를 단독으로 사용한 재순위화의 성능은 Web of Science의 검색성능과 상이 했는데, 이는 이론적 배경에서 밝힌 것처럼 Web of Science가 시스템 적합성에 기반한 적합성 순위이므로, 인용 지표는 Web of Science가 가진 기존 시스템에는 적용되지 않는 독립적인 지표로 그 순위에 작용하고 있었다.

둘째, 따라서 고유 질의어 수에 질의어의 총 출현 빈도를 조합하여 인용 지표를 보조적으로

적용한 경우, 특히 h-index보다 인용수를 기반하는 것이 성능에 긍정적인 영향을 미칠 것으로 확인하였다.

셋째, 전반적으로 인용수가 전체 성능의 향상을 도모하지는 않았지만 유사한 성능에서 적합 문헌의 순위를 상승시키는 경향과 적합성이 높은 문헌을 확인할 수 있었는데, 이는 인용수를 보조적으로 적용한다면, 이용자가 적합 문헌이라 판단할 문헌을 상위로 노출 시킬 가능성이 존재함을 확인하였다.

이후 벡터 공간모형을 기반으로 검색시스템을 구축하고 표준 로치오 공식에 인용 지표를 이용하여 질의 확장을 수행한 결과는 다음과 같다.

첫째, 본 연구에서 구축한 벡터 공간모형 기반의 검색시스템의 기본 성능에 비하여 전반적으로 질의 확장 성능이 향상되어, 여타 선행연구에서 밝혀진 결과를 검증하였다.

둘째, 이용자 적합성을 통하여 질의 확장한 경우가 시스템 적합성을 적용한 경우보다 성능 향상을 도모하였다.

셋째, 인용 수를 단독으로 질의 확장에 적용하면 성능향상에 기여하지 못하였으나, 적합 문헌과 더불어 사용하면 최상위권 내 적합 문헌에서의 순위 변동 가능성을 보여주었다.

질의 확장에 따른 검색 결과의 성능은 질의에 해당하는 용어를 수정 혹은 추가하여 이차 검색을 수행하고, 그 결과에 다시 적합성 피드백을 진행해 검색 결과의 성능을 둘이켜 봐야 더욱 신뢰 높은 판별이 가능하다. 따라서 본 연구는 질의 확장에 따른 추가 적합성 판단을 지속해서 수행하고 결과를 획득하였으나, 상위 100위에 한정하여 그 성능을 확인하였다. 실제로 본

연구가 구축한 벡터 공간모형의 검색시스템에 사용한 데이터는 60,473건이며, 검색 결과 중 100위는 매우 상위권에 해당한다. 그럼에도 불구하고 질의 확장의 성능이 향상되었고, 특히 질의 확장 결과 중 최상위권 내 적합 문헌의 변동은 추후 성능 평정의 순위 구간을 넓혀 볼 가치가 있을 것으로 판단한다.

또한 본 연구는 인용색인 데이터베이스에서의 인용 지표를 사용하여 검색에 적용해 보았다는 탐색적 연구로 의의를 찾을 수 있다.

더불어 실제 이용자 적합성 판단에서, 적합 문헌의 이진 적합뿐만 아니라, 부분 적합 점수를 부여하여 적합 문헌 간의 등급을 나누고 순위와 조합한 후 더욱 세밀한 성능 측정을 수행하고자 했다는 점에서 본 연구의 가치를 고려 할 수 있다.

본 연구는 적은 수에 해당하는 질의와 제한적인 학문 분야의 데이터셋을 활용하였으므로 대규모로의 적용과 즉각적인 일반화는 다소 어려울 수 있다.

이에 본 연구를 기반한 후속 연구의 제안은 다음과 같다.

첫째, 현재 사용한 Web of Science뿐만 아니라, 타 인용색인 데이터베이스인 Scopus와의 비교 분석을 통하여 인용 지표의 적용 가능성 을 보다 자세히 고찰 볼 수 있을 것이다.

둘째, 초기질의의 숫자와 기본 데이터셋을 대량화하여 많은 사례 검증을 확보하고, 인용 지표에 적용하여 안정적인 검색시스템 성능향상의 구체적인 형태를 구현해 볼 수 있을 것이다.

셋째, 인용수와 h-index에 해당하는 직접적인 정보를 수집하여 검색시스템의 적용 방안을 강구하고, 성능 영향 패턴이나 기점을 확인 해

볼 수 있을 것이다.

본 연구의 결과는 추후 인용색인 데이터베이스뿐만 아니라 일반의 검색시스템에서의 인용지표 적용 및 도입 등의 기초적인 자료로 기여

할 것으로 생각하며, 본 연구가 인용과 검색의 연관 연구에서도 다양한 시각을 열어주는 계기가 될 것으로 기대하는 바이다.

참 고 문 헌

- 김선우, Yang, Kiduk (2022). LDA와 BERTopic을 이용한 토픽모델링의 증강과 확장 기법 연구. 정보 관리학회지, 39(3), 99-132. <https://doi.org/10.3743/KOSIM.2022.39.3.099>
- 김한준, 노준호, 장재영 (2012). 웹 검색 개인화를 위한 개념네트워크 프로파일 기반 순위 재조정기법. 한국인터넷방송통신학회논문지, 12(2), 69-76. <http://dx.doi.org/10.7236/JIWIT.2012.12.2.6>
- 김홍렬, 이두영 (1999). 연구진행정도에 따른 검색문헌의 적합평가기준 연구. 한국정보관리학회 학술대회 논문집, 5-8.
- 박정아, 손영우 (2009). 정보검색에서의 사용자 중심 적합성 판단 모형. 한국감성과학회, 12(4), 489-500.
- 이승우, 송영인, 임해창 (2008). 혼합 방식에 기반한 의견 문서 검색 시스템. 정보관리학회지, 25(4), 115-129. <https://doi.org/10.3743/KOSIM.2008.25.4.115>
- 장영진, 권오옥, 김학수 (2020). 정보 검색 기반 기계독해 시스템을 위한 단락 재순위화 모델. 한국정보 과학회 학술발표논문집, 47(1), 410-412.
- Anker, M. S., Hadzibegovic, S., Lena, A., & Haverkamp, W. (2019). The difference in referencing in Web of Science, Scopus, and Google Scholar. ESC Heart Failure, 6(6), 1291-1312. <https://doi.org/10.1002/ehf2.12583>
- Bar-Ilan, J. (2008). Which h-index? – a comparison of WoS, Scopus and Google Scholar. Scientometrics, 74(2), 257-271. <https://doi.org/10.1007/s11192-008-0216-y>
- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. Quantitative Science Studies, 1(1), 363-376. https://doi.org/10.1162/qss_a_00018
- Buckley, C., Salton, G., & Allan, J. (1993, March). The smart information retrieval project. In Proceedings of the workshop on Human Language Technology, 392-392.
- Claveau, V. (2021, December). Neural text generation for query expansion in information retrieval. In IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 202-209. <https://doi.org/10.1145/3486622.3493957>

- Cooper, W. S. (1971). A definition of relevance for information retrieval. *Information storage and retrieval*, 7(1), 19-37. [https://doi.org/10.1016/0020-0271\(71\)90024-6](https://doi.org/10.1016/0020-0271(71)90024-6)
- Cronin, B. (1982). Norms and functions in citation - the view of journal editors and referees in psychology. *Social Science Information Studies*, 2, 65-78. [https://doi.org/10.1016/0143-6236\(82\)90001-1](https://doi.org/10.1016/0143-6236(82)90001-1)
- Gao, Q., Huang, X., Dong, K., Liang, Z., & Wu, J. (2022). Semantic-enhanced topic evolution analysis: a combination of the dynamic topic model and word2vec. *Scientometrics*, 127(3), 1543-1563. <https://doi.org/10.1007/s11192-022-04275-z>
- Garfield, E. (1964). "Science citation index"-a new dimension in indexing. *Science*, 144(3619), 649-654. <https://doi.org/10.1126/science.144.3619.649>
- Gupta, V., Chinnakotla, M., & Shrivastava, M. (2018, November). Retrieve and re-rank: a simple and effective IR approach to simple question answering over knowledge graphs. In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER), 22-27. <https://doi.org/10.18653/v1/W18-5504>
- Harman, D. (1988, May). Towards interactive query expansion. In Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 321-331. <https://doi.org/10.1145/62437.62469>
- Harman, D. (1992, June). Relevance feedback revisited. In Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1-10. <https://doi.org/10.1145/133160.133167>
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569-16572. <https://doi.org/10.1073/pnas.0507655102>
- Ioannakis, G., Koutsoudis, A., Pratikakis, I., & Chamzas, C. (2017). RETRIEVAL—an online performance evaluation tool for information retrieval methods. *IEEE Transactions on Multimedia*, 20(1), 119-127. <https://doi.org/10.1109/TMM.2017.2716193>
- Jain, S., Seeja, K. R., & Jindal, R. (2021). A fuzzy ontology framework in information retrieval using semantic query expansion. *International Journal of Information Management Data Insights*, 1(1), 100009. <https://doi.org/10.1016/j.ijimei.2021.100009>
- Jiang, Z., Tang, R., Xin, J., & Lin, J. (2021, November). How does BERT rerank passages? an attribution analysis with information bottlenecks. In Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, 496-509. <https://doi.org/10.18653/v1/2021.blackboxnlp-1.39>

- Lancaster, F. W. (1979). Information Retrieval Systems: Characteristics, Testing, and Evaluation. New York: Wiley.
- Lv, Y. & Zhai, C. (2010, July). Positional relevance model for pseudo-relevance feedback. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 579-586. <https://doi.org/10.1145/1835449.1835546>
- Maglaughlin, K. L. & Sonnenwald, D. H. (2002). User perspectives on relevance criteria: a comparison among relevant, partially relevant, and not relevant judgments. Journal of the American Society for Information Science and Technology, 53(5), 327-342. <https://doi.org/10.1002/asi.10049>
- Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. Scientometrics, 126(1), 871-906. <https://doi.org/10.1007/s11192-020-03690-4>
- Mizzaro, S. (1998). How many relevances in information retrieval? Interacting with computers, 10(3), 303-320. [https://doi.org/10.1016/S0953-5438\(98\)00012-5](https://doi.org/10.1016/S0953-5438(98)00012-5)
- Natsev, A., Haubold, A., Tešić, J., Xie, L., & Yan, R. (2007, September). Semantic concept-based query expansion and re-ranking for multimedia retrieval. In Proceedings of the 15th ACM international conference on Multimedia, 991-1000. <https://doi.org/10.1145/1291233.1291448>
- Pereira, M., Etemad, E., & Paulovich, F. (2020, March). Iterative learning to rank from explicit relevance feedback. In Proceedings of the 35th Annual ACM Symposium on Applied Computing, 698-705. <https://doi.org/10.1145/3341105.3374002>
- Rivas, A. R., Iglesias, E. L., & Borrajo, L. (2014). Study of query expansion techniques and their application in the biomedical information retrieval. The Scientific World Journal, 2014. <https://doi.org/10.1155/2014/132158>
- Rocchio, J. (1971). Relevance feedback in information retrieval. The Smart Retrieval System-experiments in Automatic Document Processing, 313-323.
- Rovira, C., Codina, L., Guerrero-Solé, F., & Lopezosa, C. (2019). Ranking by relevance and citation counts, a comparative study: Google Scholar, Microsoft Academic, WoS and Scopus. Future Internet, 11(9), 202. <https://doi.org/10.3390/fi11090202>
- Salton, G. & Lesk, M. E. (1965). The SMART automatic document retrieval systems—an illustration. Communications of the ACM, 8(6), 391-398. <https://doi.org/10.1145/364955.364990>
- Salton, G. & McGill, Michael J. (1983) Introduction to Modern Information Retrieval. New York: McGraw-Hill.

- Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620. <https://doi.org/10.1145/361219.361220>
- Smith, L. C. (1981). Citation Analysis. *Library Trends*, 30(1), 83-106. <https://hdl.handle.net/2142/7190>
- Soboroff, I. (2021). Overview of TREC 2021. In 30th Text REtrieval Conference. Gaithersburg, Maryland. Available: <https://trec.nist.gov/pubs/trec30/papers/Overview-2021.pdf>
- Spink, A., Greisdorf, H., & Bateman, J. (1998). From highly relevant to not relevant: examining different regions of relevance. *Information processing & management*, 34(5), 599-621. [https://doi.org/10.1016/S0306-4573\(98\)00025-9](https://doi.org/10.1016/S0306-4573(98)00025-9)
- Taylor, A. (2012). User relevance criteria choices and the information search process. *Information Processing & Management*, 48(1), 136-153. <https://doi.org/10.1016/j.ipm.2011.04.005>
- Van Gysel, C. & de Rijke, M. (2018, June). Pytrec_eval: an extremely fast python interface to trec_eval. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 873-876. <https://doi.org/10.1145/3209978.3210065>
- Van Raan, A. F. (2005). For your citations only? Hot topics in bibliometric analysis. *Measurement: interdisciplinary research and perspectives*, 3(1), 50-62. https://doi.org/10.1207/s15366359mea0301_7
- Voorhees, E. M. & Harman, D. K. ed. (2005). TREC: Experiment and evaluation in information retrieval (Vol. 63). Cambridge: MIT Press. Available: <http://aclanthology.lst.uni-saarland.de/J06-4008.pdf>
- Wang, X., Yang, H., Zhao, L., Mo, Y., & Shen, J. (2021, July). Refbert: Compressing bert by referencing to pre-computed representations. In 2021 International Joint Conference on Neural Networks (IJCNN), 1-8. IEEE. <http://doi.org/10.1109/IJCNN52387.2021.9534402>
- Wang, Y., Wang, L., Li, Y., He, D., Chen, W., & Liu, T. Y. (2013, June). A theoretical analysis of NDCG ranking measures. In Proceedings of the 26th Annual Conference on Learning Theory (COLT 2013), 25-54. <https://doi.org/10.48550/arXiv.1304.6480>
- Xu, J. & Croft, W. B. (2017, August). Quary expansion using local and global document analysis. In Acm Sigir Forum, 51(2), 168-175. <https://doi.org/10.1145/3130348.3130364>
- Zheng, Z., Hui, K., He, B., Han, X., Sun, L., & Yates, A. (2020). BERT-QE: contextualized query expansion for document re-ranking. In Findings of the Association for Computational Linguistics: EMNLP 2020, 4718-4728. <https://doi.org/10.18653/v1/2020.findings-emnlp.424>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Jang, Youngjin, Kwon, Oh-Woog, & Kim, Harksoo (2020). Passage re-ranking model for information retrieval based machine reading comprehension. Conference of Computing Science and Engineering, 410-412.
- Kim, HanJoon, Noh, Joonho, & Chang, Jae-Young (2012). A new re-ranking technique based on concept-network profiles for personalized web search. The Journal of The Institute of Internet, Broadcasting and Communication, 12(2), 69-76.
<http://dx.doi.org/10.7236/JIWIT.2012.1226>.
- Kim, HongRyul & Lee, Too-Young (1999). A study on relevance criteria of retrieved documents according to the research stage. Conference of the Korean Society for Information Management, 5-8.
- Kim, SeonWook & Yang, Kiduk (2022). Topic model augmentation and extension method using LDA and BERTopic. Journal of the Korean Society for Information Management, 39(3), 99-132. <https://doi.org/10.3743/KOSIM.2022.39.3.099>
- Lee, Seung-Wook, Song, Young-In, & Rim, Hae-Chang (2008). An opinionated document retrieval system based on hybrid method. Journal of the Korean Society for Information Management, 25(4), 115-129. <https://doi.org/10.3743/KOSIM.2008.25.4.115>
- Park, JungAh & Sohn, YoungWoo (2009). User-centered relevance judgement model for information retrieval. The Korean Society For Emotion & Sensibility, 12(4), 489-500.

