순위 휴리스틱을 이용한 단음절 질의어의 검색 성능 향상에 관한 실험적 연구

An Experimental study on enhancing retrieval performance for monosyllabic queries using ranking heuristics

박양하, 한국교육학술정보원, adelante@keris.or.kr
문성빈, 연세대학교 문헌정보학과, sbmoon@yonsei.ac.kr

Yang-Ha Park, Korea Education & Research Information Service
Sung-Been Moon, Dept. of Library and Information Science, Yonsei University

온라인 대용량 데이터베이스 검색에서 널리 사용되는 불리언 검색의 가장 큰 단점은 검색된 문헌을 적합한 정도에 따라 순위화하기 어렵다는 것이었다. 본 연구에서는 부적합문헌이 다른 질의어에 비해 많이 검색되는 단음절 명사 질의어를 사용하여 이용자의 노력은 최소화 하면서 검색 성능은 향상 시킬 수 있는 순위 휴리스틱 적용에 대해 연구하였다.

1. 서론

오늘날에 있어서 인터넷의 발달은 정보를 저장하는 데이터베이스 기술의 발달과 함께 대용량화 되어가고 있다. 정보의 집중은 원하는 자료를 한곳에서 모두 찾을 수 있는 장점이 있는 반면, 여러 정보 중에서 정확히 요구와 일치되는 정보를 찾아내는 능력도 요구하게 되었다. 대용량 데이터베이스 환경에서 탐색의 문제는 적합문헌이 한 건도 검색이 되지 않는 문제와 적합문헌을 찾기 어려울 정도로 많은 문헌이 검색되는 문제로 귀결된다. 적합문헌이 검색되지 않는 문제에 대한 연구는 탐색실패(no posting)의 원인을 찾고 해결책을 제시하는 방법으로 이루어지고 있다. 그러나 부적합 문헌이 많이 검색되는 문제를 해결하려는 연구들에서는 시스템으로 해결하고자 하는 노력보다 탐색자의 탐색능력을 높이는 것에 의존하려는 경향이 많았다(노정순 2001).

인터넷의 등장으로 인해서 큰 변화를 겪고 있는 분야 중 하나가 교육으로, 인터넷을 단지 정보의 창고가 아닌 보다 효율적이고 효과적인 교육을 시킬 수 있는 도구로 활용하게 되었다. 웹기반 학습 환경의 특징은 학습자들이 능동적으로 정보를 탐색하고 공유할 수 있는 상호작용이 가능하다는 것으로, 교실에서 일방적으로 이루어지고 있는 주입식 학습을 보완하는 대안으로 여겨지고 있다. 초・중・고등학교 학생들의 경우 검색식을 사용해야하는 복잡한 검색보다는 단순히 질의어 하나를 입력하여 결과를 원하는 경우가 많으며, 결과에 대한 적합 문헌의 선택 기술인 이용자 피드백도 일반인이나 연구자에 비해 정확도를 기대하기 어렵다. 최근 학교도서관의 발전과 더불어 점점 더 많은 교수-학습자료들이 집중되어지고

있다. 그 속에서 적합문헌을 빠르게 찾아내도록 하기 위해서는 검색결과 제시 시 적합문헌을 상위에 위치하도록 시스템 내에서 처리하여 이용자에게 지원하여야 한다.

따라서 본 연구에서는 부적합한 자료가 많이 검색될 확률이 높은 단음절 명사 질의어에 대해 정확률이 높은 검색 결과를 제공하기 위한 몇 가지 순위 휴리스틱 기법을 적용하고, 실험결과를 통해 검색 성능을 비교해 보고자 한다.

2. 문헌순위화

정보검색 시스템에서 검색대상 문헌이 질의에 적합한 정도를 기준으로 최적문헌부터 순서를 매기는 방법을 순위화(ranking)라 한다. 정보검색에서 검색된 문헌의 양이 너무 많을 경우 이용자가 그 모두를 확인해 선택한다는 것은 어려운 일이기 때문에 적합성이 높은 것을 먼저 출력해 이용자의 편의를 도모하는 기능을 한다. 특히, 대용량의 정보를 축적하고 있는 인터넷에서의 정보검색은 검색되는 문헌이 지나치게 많으므로 이용자가 검색결과 중에서 최적합문헌을 찾도록 순위화하여 정보의 선택을 용이하게 해주는 것이 매우 중요하다. 순위화는 검색의 정확성을 확보하는 방법으로 검색효율의 측면에서 매우 중요한 개념이다. 그러나 상용화 된 데이터베이스들이 많이 사용하는 전통적인 불리언 기법으로는 적합수준을 구분하여 순위화 하는 것이 불가능하고 또한 이용자가 자신의 정보요구를 불논리로 정확히 표현하기 어렵다. 이런 경우 검색결과에 대해 문서의 우선 순위를 매겨 상위 문서들을 사용자에게 제공하는 것이 가장 효율적이다. 우선 순위에 따른 검색결과는 사용자에게 필요한 정보를 얻는 시간을 최소화 시켜준다는 장점이 있다(김진숙 1999).

이용자 피드백은 초기 검색 결과 검색된 문헌 중에서 이용자가 직접 선택한 적합문헌을 기반으로 적합한 문헌에 출현한 용어의 가중치는 높여주고 부적합한 문헌에 출현한 용어의 가중치는 낮추는 방식으로 질의를 확장하는 것으로서 적합성 피드백이라고도 한다(Salton 1990). 이용자 피드백을 통한 질의확장은 성능이 우수하다. 그러나 이용자가 정확한 적합성 정보를 줄 수 있다면 검색 성능은 향상될 것이나 대용량 서지데이터베이스의 경우 이용자가 많은 초기 검색결과에 대한 적합성 판정을 직접 해야 하는 번거로움이 있으며, 적합성 판정 기준이 이용자 의존적이므로 반드시 검색성능이 향상된다고 단정 짓기는 어렵다(Harman 1988; Ekmekcioglu, Robertson, and Willett 1992; Efthimiadis 1993).

시스템적합성 피드백은 초기 탐색결과에서 높은 순위가 할당된 문헌에 포함된 색인어를 활용하여 질의를 확장한다. 시스템피드백은 이용자의 개입 없이 초기 검색 결과 상위 순위에 검색된 문헌을 분석하여 질의확장을 하는 것이다. 신은자(1998)의 연구에서는 사전 이용자 모형을 구축하여 시스템 내에서 자동적으로 적합성을 평가하는 이용자 모형 기반 피드백 검색 기법을 제안하였다. 이를 통해 이용자가 직접 적합성 평가를 해야 하는 부담을 덜어주는 방법을 제안하였다.

이용자 모형을 통한 자동적합성 피드백과는 달리 시소러스나 의미망과 같은 지식베이스를 이용해서 질의확장 대신 탐색어와 질의어 사이의 거리를 계산하고 질의어와의 유사한 단어를 활용하여 질의를 확장하는 방식을 순위화에 이용하기도 한다. 이때 사용되는 지식베이스는 수작업으로 작성할 수도 있고, 동시출현어분석을 통해 통계적 연관성을 계산하여 자동으로 생성할 수도 있다. 박지연(2001)의 논문에서는 이용자 피드백 없이 동시출현 기반 질의-용어간 유사도를 이용한 질의확장을 통해 단락검색의 성능을 향상시키는 방안

을 연구했다.

문헌 순위화를 위해 문헌의 구조를 이용하여 제목 등에 나타난 색인어에 가중치를 더 주는 방법을 사용할 수도 있는데, 초창기 인터넷 검색시스템에서 주로 사용한 방법이다. 노정순(2001)의 연구에서는 불논리 질의 기반의 OPAC에서 정렬과 적합성 순위화 알고리즘의 효과를 분석하기 위한 실험, 순위화 성능을 평가하는 척도를 평가하는 실험이 함께 실시되었다.

최근에는 인터넷 검색 시스템이 발전해 가면서 다른 사람들의 정보이용행태나 해당 탐색자의 신상정보를 통해 순위결정에 필요한 정보를 추출하는 데이터마이닝을 통해 피드백 정보를 획득하는 방식도 나타났다. 일반적으로 온라인환경에서의 정보검색시스템은 반복적인 탐색을 통해 가능한 적합 문헌은 많이 검색하되 부적합문헌은 검색하지 않는 것이 목적이 되고 있다.

3. 순위 휴리스틱을 사용한 단음절 명사 질의어 검색 실험

3.1 실험 설계

본 연구에서는 학교도서관을 위해 개발된 DLS(Digital Library System)를 사용하는 이용자들이 사용한 단음절 명사 질의어에 대하여 별다른 논리식을 작성하지 않은 상태에서 적합문헌의 고순위화를 통한 검색 성능 향상을 측정하고자 하였다. 즉 상대적으로 부적합문헌이 검색되기 쉬운 단음절 질의어가 입력되는 경우, 초기 결과를 바탕으로 시스템 내에서 자동으로 적합문헌을 재검색 하도록 검색식을 조정하고 재탐색하여 적합문헌 순위를 높이고자 하였다. 이를 위해서 초기 검색결과를 보고 경험적으로 검색 결과의 정확률을 높이기 위한 순위 휴리스틱 기법을 적용하여 각 순위 성능 평가를 실시하였다.

실험에 질의어로 사용할 단음절 명사는 「한국어 형태소 및 어휘사용 빈도의 분석」에 나온 일반명사를 참조하여 5개의 단음절 명사를 선택하였다. 선택된 5개의 단음절 명사는 의존성이 없고 동음이의어가 없는 명사인 강(river), 개(dog), 새(bird), 별(star), 집(house)를 선택하였다.

현재 DLS의 서명 검색 방법은 질의어를 포함하는 문자열을 지닌 서명의 서지번호순으로 결과가 순위화 되어 이용자들에게 제공되고 있다. 질의어를 포함하는 문자열이 존재하는 자료가, 목록이 구축될 때 순서대로 부여되는 서지번호에 의해 순위화 되어 화면에 제시되어지고 있기 때문에, 질의어에 대한 검색 목적에 정확히 부합되는 자료가 고순위로 검색될 확률은 높지 않은 단점을 가지고 있다. 그 예로 실험대상 시도의 DLS에서 2003년 9월 19일자로 반출 받은 종합목록의 서지 데이터 119,744건 중 '개'라는 질의어를 주고 서명 검색을 한 경우 단행본 자료만 1,524개의 자료가 검색이 되었으며, 첫 페이지에 제시되는 10권의 자료 중 개(dog)라는 정확한 질의어에 부합되는 자료는 7위에 「북극의 개」 한 권만 검색되었다. 1위 검색된 자료는 서지번호 228번의 「철학개론」이라는 자료였다.

본 실험을 위해 별도의 휴리스틱 적용 없이 기본적으로 시스템 내에서 검색된 200위까지의 검색 결과를 바탕으로 별도로 이용자가 새로운 식을 작성하지 않고 간략히 시스템에서 자동으로 부적한 문헌을 재검색 할 수 있는 순위 휴리스틱을 사용하였다. 순위 휴리스틱의 적용은 부적합문헌을 탈락시켜서 성능을 향상시키고자 했던 기존의 실험에서 사용되었던 유형들에서 아이디어를 얻어 작성하였다.

본 연구에서 이용자에게 적합한 문헌을 상위로 제시할 것으로 예상된 순위 휴리스틱을 유형별로 살펴보면 다음과 같다.

첫 번째, 문자열 내에서 질의어의 출현 위

치를 고려하여 재검색하는 방법이다. 기본 검색 상위 결과를 살펴보았을 때도 부적합한 단어에 질의어가 포함되어 의미가 전혀 다른 문헌이 적합한 것으로 검색되어진 것을 확인 할 수 있었다.

두 번째, 문헌의 표제에서 질의어와 동일한 길이의 문자열을 재검색하는 방법을 사용하였다. 기본 검색 결과를 분석해 보았을 때, 문자열의 길이가 짧은 문헌일수록 적합한 문헌일 확률이 높았다.

세 번째, 문헌의 표제에서 질의어를 포함하는 문자열 내에 질의어 다음 형태소가 미리 작성해 둔 조사파일과 비교하여 일치하는 문헌을 재검색하는 방법을 사용하였다. 인접형태소 분석을 통해 문자열의 의미나 패턴을 파악하여 적합문헌으로 인정하는 실험들을 참고하였다. 이 방법은 질어어가 문자열의 가운데 위치하더라도 적합한 문헌을 상위 순위로 검색하기 위해 실시하였다.

네 번째, 가장 간단한 방법이면서 성능을 향상 시킬 수 있는 불용어처리를 위한 휴리스틱 적용이다. 미리 작성한 명사어 파일을 불용어리스트처럼 활용하여 자동으로 재검색하여 적합문헌을 상위에 검색되도록 조절하는 순위 휴리스틱도 적용해 보았다.

성능비교를 위한 평가척도로는 10위내 정확률(P@10)과 10위내 순위정확률(RP@10)을 함께 측정하여 높은 값을 가지는 실험결과를 비교하였다(이재윤 2003).

10위내 정확률(P@10)

· P@10= 10위내 적합문서의 수/10

10위내 순위 정확률(RP@10)

· RP@10= (P@1+P@2+...+P@10)/10

검색된 웹 문서의 순위결과를 평가하는 성능 평가 척도인 위 두 가지 척도를 본 실험에 적용하여 질의어의 원래 의미와 부합되는 서명을 지닌 문헌의 경우 적합문헌으로 그렇지 않은 경우를 부적합문헌으로 판단하여 각 정확률값이 높은 휴리스틱 처리방법이 높은 검색 성능을 지닌 것으로 평가한다.

산술적인 평가 방법 외에도 각 순위 휴리스틱에 따라 검색에서 누락된 적합 문헌도 고려하여 적용된 휴리스틱이 시스템 적용하는 것이 적합한지 판단하였다.

3.2 실험내용

실험에 선정된 단음절 명사 질의어의 기본 검색 결과를 토대로 상위 200위까지 문헌의 분석을 통해 적합 문헌의 고순위화라는 최적해를 얻기 위한 휴리스틱을 구상하고, 이를 적용하여 결과를 도출하도록 수행하였다. 별도의 휴리스틱을 적용하지 않고 검색된 문서들 중 상위 10위까지의 검색 결과를 베이스라인으로 평가하고 각각의 휴리스틱을 적용한 후 상위 10위까지의 검색결과를 바탕으로 검색의 성능을 비교 분석하였다.

적용된 휴리스틱은 앞서 기술한 네 가지 유형에서 구체적으로 다음의 여섯 개의 순위 휴리스틱을 통해 실험하였다.

첫째, 문헌의 표제에 포함된 문자열의 첫 글자에 질의어가 위치하는 문헌을 재검색 하였다. 이 순위 휴리스틱의 적용은 기존의 비통계적 가중치 기법으로 서지 데이터에서 용어가 출현하는 필드에 따라 가중치가 부여되는 필드 가중치 기법에서 착안하여, 위치에 따른 우선 순위 적용을 고려하게 되었다.

둘째, 문헌의 표제에 포함된 문자열의 끝 글자에 질의어가 위치하는 문헌을 재검색 하였다. 이 순위 휴리스틱의 적용은 기본 검색 상위 200위에 검색된 문헌들의 제목을 살펴본 결과 "진돗개", "삽살개", "들개" 혹은 "한강", "두만강" 등의 고유명사로 표현된 적합문헌이 존재할 수 있음을 경험적으로 판단하여 순위 휴리스틱으로 반영하였다.

셋째, 위의 두 휴리스틱 결과를 혼합하여 문헌의 표제에 포함된 문자열의 첫 글자나 끝

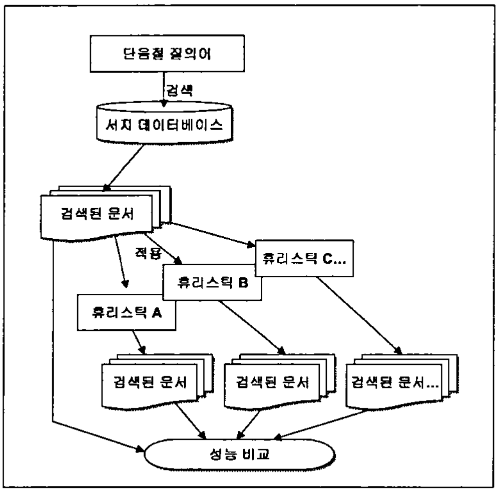
글자에 질의어가 위치하는 문헌을 재검색 하였다. 이는 단순히 첫 글자로 출현하는 결과만이 상위 순위에 있거나 끝 글자만이 상위 순위에 배치되는 경우 발생할 수 있는 적합문헌의 누락에 대한 대안으로 혼합의 방법을 사용하였다.

넷째, 문헌의 표제에 포함된 문자열의 길이가 질의어와 동일하게 단음절인 문헌을 재검색 하였다.

다섯째, 문헌의 표제에 질의어가 출현하는 문자열의 조사가 미리 작성해 둔 조사 파일에 있는 경우 그 문헌을 재검색 하였다. 이 순위 휴리스틱의 적용은 단음절 명사 질의어이므로 명사 다음에는 조사가 많이 수반되는 특성을 고려하였다. 이 순위 휴리스틱은 형태소의 비교를 위해 「한국어 형태소 및 어휘사용 빈도의 분석」에서 분석한 124개의 고빈도 조사 중 121개의 조사를 별도의 파일로 처리하여 재검색 시 비교토록 하는 방법을 사용하였다. 제외된 3개의 조사는 'ㄴ, ㄹ, ㄹ랑' 이다. 예를 들어 「개와 고양이」라는 서명과 「개구장이」라는 서명이 존재하는 경우 '개'라는 음절 다음 음절인 '와'가 조사 파일에 존재하고 있으므로「개와 고양이」를 상위의 검색결과로 올리고「개구장이」를 검색에서 배제시키는 것이다.

여섯째, 표제에 질의어가 출현하는 문자열이 미리 작성해 둔 명사 파일에 있는 경우 그 문헌을 검색에서 배제하고 재검색하였다. 즉 단음절 질의어가 들어 온 경우에만 미리 준비한 명사사전파일과 비교하여 불용어 처리를 하는 것이다. 불용어 처리는 기존에 상용화된 시스템에서 정확률을 높일 수 있는 가장 간단하지만 정확한 수단으로 사용되어졌다. 기본 검색 상위 결과로 제공된 문헌을 분석한 결과, 질의어를 포함하는 다음절 명사어들이 검색결과로 많이 검색되어 부적합문헌의 수를 늘이는 경우가 많았다. 이는 일반적으로 사용되는 [NOT]검색과 동일하게 처리되는 방식이다. 주로 한국어에서 불용어로 처리되는 대상은 명사, 수사, 대명사, 동사, 형용사, 관형사, 감탄사 등이다. 그러나 본 실험에서는 반대로 미리 준비한 일반 명사들을 비교하여 문자열에 명사 파일과 일치하는 다음절어가 있는 경우 검색에서 배제시키고 나머지 결과의 순위를 높이는 방법을 사용하였다. 명사사전은 질의어 당 일반적으로 사용되는 용어를 국어사전을 참고하여 작성하였으며, 각 질의어 당 열 개의 명사를 포함하는 파일을 만들어 비교한다. 예를 들어, 질의어가 '개(dog)'인 경우 '개구장이', '개성', '개발' 등의 명사는 검색되지 않게 하는 방법이다.

본 실험의 전체적인 설계도는 <그림 1>과 같다.



<그림 1> 실험 설계도

4. 휴리스틱별 성능평가

6가지의 휴리스틱을 적용한 결과는 <표 1>과 <표 2>와 같이 길이 휴리스틱을 사용한 경우가 가장 성능이 높았다. 그리고 질의어에 따라 편차가 가장 큰 휴리스틱은 첫 글자 휴리스틱이었다. 조사의 경우 '새', '별', '집'과

같이 다른 다음절 단어에 포함되기 어려운 단음절 질의어는 검색 성능이 높게 나타난 결과를 확인할 수 있었다.

휴리스틱을 적용하여 성능을 평가한 결과와 실험 과정에서 발견 된 분석의 결과 길의 휴리스틱, 조사 파일 비교 휴리스틱, 명사 파일을 통한 불용어 처리 휴리스틱이 결과가 전반적으로 성능이 높은 것을 볼 수 있다.

<표 1> 10위내 정확률(P@10)

휴리스틱 / 질의어	베이스 라인	첫 글자	끝 글자	혼합	길이	조사	명사
[강]	0.4	0.3	0.6	0.4	1	0.4	0.5
[개]	0.2	0.1	0.5	0.2	1	0.2	0.3
[새]	0.6	0.4	0.9	0.5	0.9	0.7	0.7
[별]	0.3	0.4	0.9	0.4	1	1	0.5
[집]	0.2	0.8	0.4	0.3	1	0.7	1
평균	0.34	0.40	0.66	0.36	0.98	0.60	0.60

<표 2> 10위내 순위정확률(RP@10)

휴리스틱 / 질의어	베이스 라인	첫 글자	끝 글자	혼합	길이	조사	명사
[강]	0.188	0.128	0.505	0.207	1	0.4	0.443
[개]	0.035	0.025	0.022	0.045	1	0.1	0.117
[새]	0.273	0.151	0.815	0.213	0.707	0.471	0.39
[별]	0.072	0.141	0.879	0.151	1	1	0.208
[집]	0.034	0.556	0.127	0.075	1	0.416	1
평균	0.120	0.200	0.470	0.138	0.941	0.477	0.432

그러나 이와 같은 부적합문헌의 배제를 통한 검색 성능 향상 실험에서 간과되는 것이 바로 적합 문헌의 누락이다. 산술적인 평가 결과와 함께 실제 시스템에 적용하기 위해서는 적합 문헌의 누락도 함께 고려되어야 한다.

첫 글자 휴리스틱의 경우, 가운데 위치하거나 끝 글자로 위치하는 경우 누락되게 된다. 따라서 끝 글자 휴리스틱의 적합 문헌이나, 조사 파일 비교 휴리스틱에서 적합하다고 검색된 문헌의 누락이 일부 발생하였다. 예로는 「두만강 해란강의 전설」, 「사슴과 사냥개」, 「초록별의 비밀」, 「도도새와 카바리아 나무」 등과 같은 문헌이 검색된 서지 번호 내에서 누락되었다. 끝 글자의 경우는 반대로 첫 글자로 위치하는 문헌이 누락되게 된다. 또한 뒤에 조사가 붙어 끝 글자로 인식되지 못한 문헌들도 누락되었다. 예로는 「금강을 노래한 시인 신동엽」, 「개가 무서워요」, 「별찾기」, 「새들의 세계」, 「집의 사회사」 등과 같은 문헌이 누락되었다. 혼합의 경우는 위의 두 휴리스틱보다 적합문헌의 누락 비율이 적었지만, 역시 조사에 의해 문헌의 문자열에서 가운데 위치한 문헌은 누락되는 단점이 있다. 예를 들어 「진돗개의 특성」, 「작은새가 들어올린 하마」 등의 문헌은 검색되지 못하였다. 길이의 경우는 모든 질의어에서 높은 검색 성능을 보여서, 시스템 적용 시 가장 우선되는 경우이나, 적합 문헌의 누락 정도도 심하여 신중히 고려되어야 한다. 다른 휴리스틱에서 검색된 적합 문헌 중 단음절이 아닌 적합 문헌의 모두가 누락된다. 불용어 처리 휴리스틱의 경우 검색 결과 문헌의 내용을 비교해 보았을 때, 부적합 문헌의 배제와 동시에 적합 문헌의 누락도 일어나지 않는 것을 발견하였다.

5. 결론 및 제언

실험을 통해 단음절 질의어 입력 시 재검색을 통해 학교도서관 지원 시스템인 DLS 내에서의 검색성능을 향상시킬 수 있는 순위 휴리스틱을 적용하여 그 결과를 비교해 보았다.

본 연구를 통한 휴리스틱의 유형별 결과는 다음과 같다.

(1) 문서의 표제에 질의어가 출현하는 문자열에서, 위치한 순서를 고려하여 순위를 조정한 경우, 첫 글자 위치를 재검색 하는 것보다 끝 글자만을 재검색하는 것이 성능이 더 뛰어났다. 또한 일반적으로 다음절 명사어에 포함될 확률이 높은 질의어의 경우 베이스라인보다 성능이 낮을 수도 있었다.

(2) 문헌의 표제에 질의어가 출현한 문자열의 길이가 짧은 문헌을 우선으로 재검색 하도록 조정한 경우 다른 휴리스틱에 비해 성능이 월등히 높은 것으로 밝혀졌다. 그러나 그만큼 적합 문헌이 검색에서 누락될 위험이 있다.

(3) 문헌의 표제에 질의어가 출현한 문자열에서 질의어 다음 글자가 미리 준비한 조사파일과의 비교를 통해 검색 결과를 재조정한 경우, 위치에 의한 휴리스틱보다는 성능이 월등히 높았다. 그러나 위치에 의한 휴리스틱과 유사하게 다음절 단어에 포함될 확률이 높은 경우 성능이 낮아질 수 있다.

(4) 명사파일을 통한 불용어 처리 휴리스틱의 경우 부적합문헌을 가장 객관적이고 강력하게 제거함으로 적합 문헌의 순위를 높일 수 있는 방법이었다.

모든 휴리스틱을 시스템에 한꺼번에 적용할 경우 시스템 속도의 문제라든지, 부적합문헌을 제거하려는 목적에 적합 문헌이 누락되는 역효과가 발생할 우려가 있을 수 있다. 따라서 실제로 DLS 시스템에서 단음절 질의어가 입력되었을 때, 자동으로 시스템에서 미리 명사어 사전을 마련하여 불용어 처리를 하여 결과로 제공하고, 문자열의 길이가 짧은 것을 고순위화 하도록 적절한 가중치 부여하는 방법을 고려해 볼 수 있다. 나머지 휴리스틱은 이용자의 선택사항으로 남겨두되, 학생이란 특성을 고려하여 연산자를 추가하기보다는 '~과', '~를 제외하고' '마지막 글자에' 또는 '순서대로' 등 별도의 교육 없이도 사용 가능한 용어로 바꾸어 주는 것이 적당하다. 방대한 양의 양질의 데이터베이스가 단순히 이용자들의 특성을 파악하지 못하여, 적절한 검색 결과를 제공해 주지 못하거나, 부적합문헌 검색의 책임을 질의어를 입력하는 이용자의 잘못으로 돌리는 것은 무성의한 일이다.

본 연구는를 실제 시스템에 적용하기에는 단음절어만을 고려한 점, 초기 검색 결과의 상위 200위내의 문헌만을 분석에 포함한 점 등 DLS라는 시스템의 전반적인 성능향상을 위한 연구로는 부족한 점이 있었다. 또한 분석결과를 통해 사용된 휴리스틱의 적용 기준에 연구자의 주관이 개입되었을 한계점이 있다. 앞으로 이런 한계점을 보완하여 실제 시스템으로 반영하기 위해 실험용이 아닌 완벽한 불용어 처리용 명사사전과 주제영역에 구애받지 않는 휴리스틱의 개발이 필요하다. 기술의 발달로 빠른 시간 내에 복잡한 처리가 가능해지고 있으므로 시스템 내에서 이용자에게 재현률만을 강조하는 검색 시스템이 아닌 정확률을 강조하는 시스템으로 발전하려는 노력이 필요하다.

참고문헌

김진숙, 1999. 정보검색시스템 KRISTAL-II에서의 우선순위 부여모델 구현. 『KORDIC Newsletter』, 제 15호.

노정순, 2001. OPAC에서 서명단어탐색의 문헌순위화에 관한 연구. 『정보관리학회지』, 18(2).

문성빈, 1993. 적합성피이드백을 이용한 전문검색시스템의 검색효율성 증진을 위한 연구. 『정보관리학회지』, 10(2).

박지연, 2001. 『질의확장에 의한 단락검색의 성능향상에 관한 연구』. 석사학위논문, 연세대학교대학원, 문헌정보학과.

신은자, 정영미. 1998. 피드백 정보를 이용한 불논리 검색 시스템의 성능 증진에 관한 실험적 연구. 『정보관리학회지』, 15(1).

이재윤. 2003. 질문유형에 따른 인터넷 검색엔진의 성능비교. 『제10회 한국정보관리학회 학술대회』.

정영미. 1993. 『정보검색론』. 서울:구미무역.

Efthimiadis, Efthimis N. 1993. "A user-centered evaluation of ranking algorithms for inter active query expansion." *Proceedings of ACM SIGIR International Conference on Research and Development in Information Retrieval.* 재인용 : 박지연. 2001. 『질의확장에 의한 단락검색의 성능 향상에 관한 연구』. 석사학위논문, 연세대학교대학원, 문헌정보학과.

Ekmekcioglu , F. Cu n a, Alexan der M. Robertson, an d Peter Willett . 1992. "Effectiveness of query expansion in ranked-output document retrieval systems." *Journal of Information science*, 18(2). 재인용 : 상동

Harman, D. 1988. "Toward s interactive query expansion." *Proceedings of A CM SIGIR International Conference on Research and Development in Information Retrieval.* 재인용 : 상동.

Salton, G. and C. Buckley. 1990. "Improving retrieval performance by relevance feedback ." *Journal of the American Society for Information Science.* 41