

<http://dx.doi.org/10.7236/JIWIT.2012.12.2.69>

JIWIT 2012-2-9

웹 검색 개인화를 위한 개념네트워크 프로파일 기반 순위 재조정 기법

A New Re-ranking Technique based on Concept-Network Profiles for Personalized Web Search

김한준*, 노준호**, 장재영***

Han-joon Kim, Joonho Noh, Jaeyoung Chang

요약 본 논문은 웹 검색 개인화를 실현하기 위해 개념네트워크 구조의 사용자 프로파일에 기반한 새로운 형태의 순위 재조정 기법을 제안한다. 기본적으로 개인화 검색은 개인 사용자의 검색 성향을 담고 있는 사용자 프로파일을 기반으로 이루어지며, 이를 활용하여 초기에 주어진 검색 질의어를 확장하거나 검색결과의 순위를 재조정하게 된다. 제안 기법은 순위 재조정 기법을 주축으로 질의어 확장 기법을 융합한 형태를 취한다. 기본 아이디어는 사용자 프로파일에 의해 추천된 확장 질의어로부터 도출된 문서집합들에 공통적으로 출현하는 문서들의 중첩도를 평가하여, 그 결과값을 순위 재조정에 활용하는 것이다. 성향이 다른 다수의 실험자들이 검색 질의어 유형을 구분하여 실험을 수행함으로써 제안기법이 기존 기법에 비해 우수함을 보인다.

Abstract This paper proposes a novel way of personalized web search through re-ranking the search results with user profiles of concept-network structure. Basically, personalized search systems need to be based on user profiles that contain users' search patterns, and they actively use the user profiles in order to expand initial queries or to re-rank the search results. The proposed method is a sort of a re-ranking personalized search method integrated with query expansion facility. The method identifies some documents which occur commonly among a set of different search results from the expanded queries, and re-ranks the search results by the degree of co-occurring. We show that the proposed method outperforms the conventional ones by performing the empirical web search with a number of actual users who have diverse information needs and query intents.

Key Words : Information Retrieval, Web Search, Personalization, Ranking, User Profile, Concept Network

1. 서 론

오늘날 웹 검색(Web search)은 정보 획득의 필수적인 수단일 뿐만 아니라 기업의 이익 창출을 위한 중요한 마케팅 수단으로서 자리매김하였다^[1]. 향후 웹 검색 서비스

는 기능과 품질 측면에서 지속적으로 발전할 것이며, 그 가운데 핵심적인 기술로서 언급되는 것이 '개인화(personalization)' 기술이다. 웹 환경에서의 개인화 기술은 웹이 제공하는 서비스나 콘텐츠 등을 모든 사용자에게 공통적으로 제공하는 것이 아니라, 개인의 성향에 맞

*정회원, 서울시립대학교 전자전기컴퓨터공학부 (교신저자)

**준회원, 서울시립대학교 전자전기컴퓨터공학부

***정회원, 한성대학교 컴퓨터공학과

접수일자 2012.3.4, 수정완료 2012.3.30

게재확정일자 2012.4.13.

Received: 4 March, 2012, Revised: 30 March, 2012

Accepted: 13 April, 2012

*Corresponding Author: khj@uos.ac.kr

Dept. of Electrical and Computer Engineering, University of Seoul, Korea

게 선별하여 제공하는 것이다. 같은 맥락에서 웹 검색의 개인화는 동일한 검색 질의어에 대해서 사용자의 검색 의도 및 특성에 맞는 검색 결과를 전달하는 것이다.

웹 검색은 정보검색의 특수한 형태이며, 그것의 기본적인 목적은 사용자가 제시한 정보요구(information needs)에 적합한 문서집합을 최적화하여 제공하는 것이다. 이러한 정보검색의 목적은 사실 개인화 검색 개념을 내포하고 있다. 최근 검색엔진의 발전이 가속화되면서 개인화 검색은 검색 기술이 궁극적으로 나아가갈 방향으로 인식되고 있으며, 이에 따라 개인화 검색에 관한 다각적인 연구가 진행되고 있다.^{[2][3][4][5][6]}

아직 개인화 검색 기술은 효용성 측면에서 큰 성과를 거두지 못하고 있는 실정이지만, 최근 개인화를 표방한 개인화 웹포탈 및 개인화 검색엔진이 선을 보이고 있다. 개인화 웹포탈 서비스의 예로는 iGoogle^[7], MyYahoo^[8] 등을 들 수 있으며, 개인화 검색엔진의 예로는 Rollyo^[9], Ness^[10] 등이 소개되었다. 이러한 개인화 검색 시스템을 구현하기 위해서는 기본적으로 사용자 프로파일(user profile)을 구축하는 것이 필수적이다. 사용자 프로파일은 사용자의 관심 사항, 검색 패턴 등을 체계적으로 정리해 놓은 데이터 구조이며, 이는 사용자의 인구학적 정보 및 최근 검색 기록 등을 수동 또는 자동으로 수집, 분석하여 작성하게 된다. 이러한 사용자 프로파일의 품질은 개인화 검색의 성능을 결정짓는다.

본 논문에서 제안하는 개인화 기법은 이전 연구에서 제안한 개념네트워크(concept networks) 구조의 사용자 프로파일^[11]을 기반으로 하며, 사용자 프로파일을 통해 추천된 다수의 확장 질의어에 따라 도출된 문서집합들에 공통적으로 속하는 정도를 정량화하여 이를 순위 재조정에 활용한다. 본 연구에서 개인화 검색은 개념네트워크 프로파일에 기반한 질의 확장의 효과뿐만 아니라 순위 재조정 결과를 통해 성취된다.

본 논문의 구성은 다음과 같다. 2장에서는 개인화 검색 및 순위 재조정과 관련한 최신 연구내용을 살펴보고, 3장에서는 본 논문의 제안기법의 기반이 되는 개념네트워크 기반 질의확장 기법과 순위 재조정을 위한 검색결과 중첩도 척도를 제안한다. 4장에서는 제안기법의 효과를 검증하기 위한 실험 환경 및 방법, 성능 평가 내용을 담고 있다. 마지막으로 5장에서 결론을 맺는다.

II. 배경연구

1. 개인화 검색

앞서 기술한 바와 같이 개인화 검색을 실현하기 위해서는 각 개인의 검색 패턴을 형상화한 사용자 프로파일을 구성하는 것이 핵심적 요소가 된다. 최근 개인화 기술의 중요성이 인식되면서 개인화 웹 검색 목적의 사용자 프로파일의 생성을 위한 다양한 연구들이 진행되고 있다. [3]에서는 계층구조 형태의 사용자 프로파일을 자동 생성하여 이를 검색 결과에 적용해서 개인화하는 방법을 제안하였다. 그리고 [5]에서는 사용자의 과거 검색기록을 수집하여 사용자의 관심사항을 자동으로 찾아내는 기술을 제안하였다. [6]은 개인화 검색을 위해 사용자가 방문했던 기록을 바탕으로 질의 확장을 수행하는 방식을 취하였다. [12]에서 제안한 연구는 사용자 프로파일을 활용하지 않는 개인화 방안으로서 폭소노미(folksonomy) 데이터를 이용하여 사용자의 관심도에 따른 검색 결과를 보여주어 개인화를 하였으며, [13]은 검색 결과에 클러스터링 기법을 적용하여 개인화를 도모하였다. 본 연구와 유사한 연구로서 [14]는 주어진 사용자 프로파일을 확장하기 위해 관련 온톨로지의 개념집합과의 퍼지관계를 이용하였다.

사용자 프로파일을 이용한 개인화의 기법은 그림 1에서 보는 바와 같이 크게 두 가지로 분류할 수 있다. 한 가지 방법은 사용자 프로파일을 기반으로 하여 사용자의 질의를 확장(수정)시켜 그 확장된 질의를 검색엔진에 전달함으로써 개인화 검색을 실현하는 것이다 (그림 1 참조). 다른 하나는 사용자 질의를 검색엔진에 전달하여 얻은 결과를 사용자 프로파일에 의거하여 검색 순위를 재산정하여 개인화된 결과를 반환하는 것이다 (그림 2 참조). 본 연구에서 제안한 개인화 기법은 두 가지 방안을 융합한 것이라 할 수 있다.

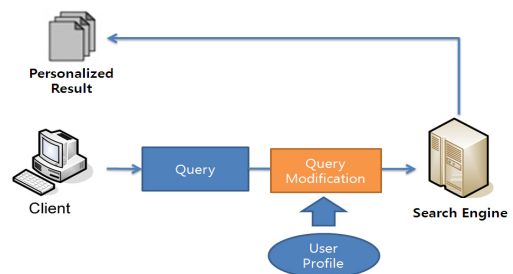


그림 1. 질의 확장을 통한 개인화 검색
Fig. 1. Personalized search by query expansion

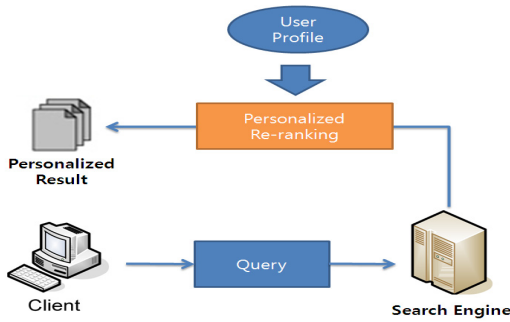


그림 2. 순위 재조정을 통한 개인화 검색
Fig. 2. Personalized search by re-ranking

2. 개념네트워크 기반 사용자 프로파일

본 연구의 순위 재조정 기법이 기반이 되는 개념네트워크 사용자 프로파일은 사용자 관심사항을 개념의 집합체로서 구성한 것이다. 여기서 개념은 Formal Concept Analysis (FCA)^[15]에서의 개념의 정의에 따른다. FCA 이론체계에서 하나의 개념은 외연(extent)과 내연(intent)로 구성되는데, 외연은 해당 개념에 포함되는 인스턴트(문서)들의 집합이며, 내연은 외연에 포함된 모든 인스턴트들의 공통된 속성들의 집합을 의미한다.

개념네트워크 형태의 기본 요소가 되는 개념은 매번 질의 수행될 때마다 생성됨을 가정한다. 그리고 위에서 설명한 FCA 이론의 개념 정의를 도입하여, 하나의 개념을 구성하는 외연은 각 질의로부터 얻어진 검색 결과에서 사용자가 방문한 문서집합으로 구성되며, 내연은 그 방문 문서집합으로부터 추출한 주요 키워드들의 집합으로 구성된다. 그런데 사실상 외연 문서집합으로부터 양질의 키워드집합을 얻어낼 수 있다면, 실제적으로 내연 키워드 집합만으로 하나의 개념을 정의할 수 있다. 정리하면, 개념네트워크 프로파일을 구성하는 기본 단위는 개념이며, 이는 사용자가 방문한 문서집합으로부터 추출된 키워드들로서 구성된다. 그리고 사용자 프로파일은 그 개념들의 네트워크 구조인 개념네트워크구조로서 표현된다. 질의가 지속적으로 진행되면서 다수의 개념들의 생성될 것이며, 점진적으로 개념네트워크는 지속적으로 진화할 수 있어야 한다. 그림 3은 개념네트워크의 예를 보여준다. 여기서 개념을 정의하기 위해 필요한 주요 키워드는 [11]에서 제안한 키워드 추출 기법을 통해 얻어진다.

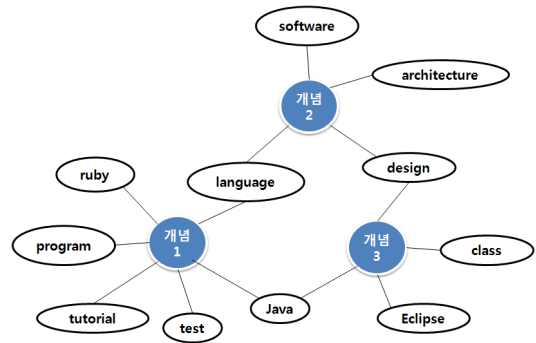


그림 3. 개념네트워크 사용자 프로파일의 예
Fig. 3. An example of concept network user profile

III. 중첩도 기반 개인화 검색

1. 중첩성 평가를 통한 순위 재조정 기법

순위 재조정(re-ranking)은 1차적으로 전달된 검색 결과에 대하여, 사전에 구성된 사용자 프로파일을 활용하여 사용자 관심도에 따라 웹문서의 순위를 변경하는 작업이다. 본 연구에서 순위 재조정을 위한 기본 전략은 각 질의어로부터 도출된 검색 결과의 중첩성(overlap) 정도를 정량적으로 평가하는 것이다.

그림 4에서 사용자의 질의어 q 에 의해 문서집합 D_q 가 얻어질 때, 개념네트워크 프로파일에 입각하여 자동 추출된 확장 질의어 q_1, q_2, \dots, q_k 에 의해 각기 문서집합 $D_{q_1}, D_{q_2}, \dots, D_{q_k}$ 가 검색결과를 얻을 수 있다고 하자. 여기서, D_{q_i} 는 초기 질의어 q 와 확장 질의어 q_i 의 배합으로 작성된 질의어 " $q \& q_i$ "에 의한 검색 문서집합이다. 예를 들어, 사용자가 'ruby' 라는 프로그래밍 언어에 관하여 질의를 하면, 그림 3의 개념네트워크 프로파일에서 'ruby'와 연관된 키워드인 'program', 'language', 'tutorial', 'linux', 'java', 'test'를 추출할 수 있다. 이 중에서 사용자가 질의 확장을 위해 'program', 'language', 'tutorial', 'test'를 선택하였다면, 이를 이용하여 "ruby & program", "ruby & test", "ruby & language", "ruby & tutorial" 등 4 가지 유형의 질의 확장을 수행할 수 있다. 이 확장 질의를 검색엔진에 전달하여 4가지 검색결과 집합을 얻게 된다. 그 검색결과인 $D_{q_1}, D_{q_2}, \dots, D_{q_k}$ 문서집합들에 공통적으로 나오는 웹 문서들은 사용자의 질의 의도와 관련성이 높다는 예측을 할 수 있다.

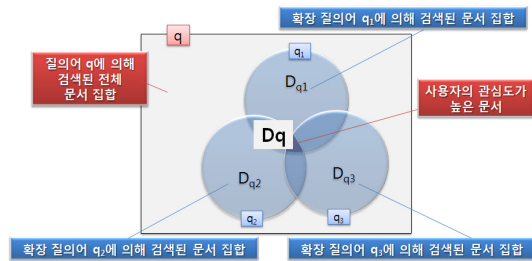


그림 4. 확장 질의어에 의해 검색되는 문서집합의 표현
 Fig. 4. Representation of document sets retrieved by expanded queries

그림 5는 위에서 예를 든 확장질의어에 대한 검색 결과를 보여준다. "ruby & program"의 검색결과 1위 문서와 "ruby & language"의 검색결과 3위 문서는 공통된 문서이다. 이렇게 검색 집합간에 공통으로 출현하는 웹문서는 개념네트워크 프로파일에 속한 다수의 키워드가 포함되어 있기 때문에 검색의도와 연관성이 높은 웹문서가 될 수 있다. 본 논문은 이렇게 웹문서들이 중복 출현하는 것에 초점을 맞추어 개인화를 위한 순위 재조정 문제를 풀고자 한다.

본래 검색엔진의 검색결과는 사용자 질의와 문서간의 관련도값을 계산하여 그 순서대로 정렬한 형태가 된다. 특히 상위 n개의 문서에 대한 순위는 매우 중요하다. 상위 n개의 검색 문서에 대하여 개인별로 상이한 순위를 제공하기 위해서는 추가적인 순위 척도가 필요하다. 문제는 동일한 질의로부터 얻어진 검색결과 집합에 특정 문서가 공통적으로 출현하는 정도를 어떻게 정량화하

는 것이다.

우선 확장 질의어로 검색 된 문서집합인 Dq_i 에 포함된 웹문서들에 대한 순위 점수는 식 (1)에 의해 계산된다. 개인화 검색의 효과를 높이기 위해 순위값 계산에 적용되는 순위의 범위는 상위 20개로 제한한다.

$$S(d_{q_i}^{(r)}) = \frac{1}{\log(r+1)} \quad (1)$$

$d_{q_i}^{(r)}$ 는 확장 질의어로 생성된 Dq_i 집합에서 순위가 r인 문서를 의미한다. 기본적으로 검색 문서집합에서 순위가 높은 웹문서가 그 중요도가 높기 때문에 순위를 기준으로 점수를 계산한다. 중복된 웹문서가 각 집합에서 순위가 낮은 웹문서일 경우에도 실제 재산정된 결과에서 상위에 위치시키기 위해서는 순위 간 점수 차이를 작게 하는 것이 바람직하다. 그래서 식 (1)에서 보는 바와 같이 순위 계산을 위해 분모 부분에 로그(log)를 적용한다. 로그항에 1을 더한 것은 순위가 1위인 경우에 분모가 0이 되는 경우를 방지하기 위함이다. 예를 들어, 특정 문서가 3개의 확장 질의어에 대한 순위가 1, 3, 6 등인 문서의 순위값은 각각 $1/\log(1+1) = 3.32$, $1/\log(3+1)=1.66$, $1/\log(6+1)=1.18$ 값을 가진다.

다수의 확장 질의어에 따른 검색 문서집합들에 중복되는 웹문서는 각 확장 질의어에 대한 검색 결과가 되는 것이므로, 각 확장 질의어에 대한 순위를 합산하여 식 (2)와 같이 해당 웹문서 d_q 의 순위 점수 $S(d_q)$ 를 정의할 수 있다.



그림 5. 확장 질의어에 대한 검색 결과 및 중복된 웹문서의 예
 Fig. 5. Examples of search results for expanded queries and overlapped web documents

$$S(d_q) = \sum_{i=1}^k S(d_{q_i}^{(r)}) \quad (2)$$

여기서, k 는 확장 질의어의 개수를 의미한다. 특정 문서 d_q 가 확장 질의에 따른 검색된 문서집합 Dq_i 에 모두 중복되어 나타났다면, k 개의 순위 점수값 $S(d_{q_i}^{(r)})$ 을 모두 합산하게 된다. 여기서 문서집합 간에 동일 문서가 20 순위 이내에 존재하지 않는 경우에는 순위 점수값을 0으로 설정한다.

여기서 식 (2)의 중첩도를 측정하는데 있어, 중첩된 검색결과를 구분하는 기준은 웹페이지 식별자인 URL (uniform resource locator)이다. 그런데 많은 경우 URL이 다르다할지라도 유사한 내용의 웹문서들이 동일한 도메인상에서 분화된 것들이 많다. 그러므로 URL이 아닌 도메인명(domain name)을 기준으로 중첩도를 평가하는 방법이 의미가 있다. 실제적으로 도메인명을 기준으로 하는 방안이 URL 기준 방안보다 개인화 검색에 기여하는 정도가 다소 향상됨을 보일 것이다.

2. 프로토타입 시스템 구조

그림 6은 개인화 웹검색 프로토타입 시스템의 구조도를 보여주며, 제안 기법은 이 시스템의 주요 모듈로서 구현되었다. 개인화 검색을 위한 핵심 모듈은 '확장 질의어 추출모듈', '개념네트워크 관리 모듈', '순위 재조정' 모듈 등이다. 그리고 검색엔진 구축을 위해 오픈 소스로 널리 알려진 lucene^[6]을 활용하였다. 본 시스템은 사용자가 과거에 질의했던 정보를 수집하여 개념네트워크 사용자 프로파일을 자동 생성한다. 시스템은 이를 통해 사전에 사용자 프로파일 데이터베이스를 구축해놓고 개인화를 위한 두 가지 작업, 즉 질의 확장과 순위 재조정 작업을 실시간으로 수행한 준비를 한다. 사용자가 질의를 수행하면 확장질의어 추출 모듈에 의해 최적의 질의어를 사용자에게 추천하게 되고, 사용자가 자동 또는 수동으로 적절한 질의어를 선택한다. 이를 활용하여 확장된 질의어는 검색엔진에 전달되고 해당 검색결과를 생성한다. 이후 확장 질의어 별로 도출된 검색집합에 대하여 중첩도(식 (2) 참조)를 계산하여 개인별로 새롭게 조정된 순위의 검색결과를 사용자에게 최종적으로 전달하게 된다. 본 논문에서 제안한 중첩도 평가는 순위 재조정 모듈에서 이루어진다.

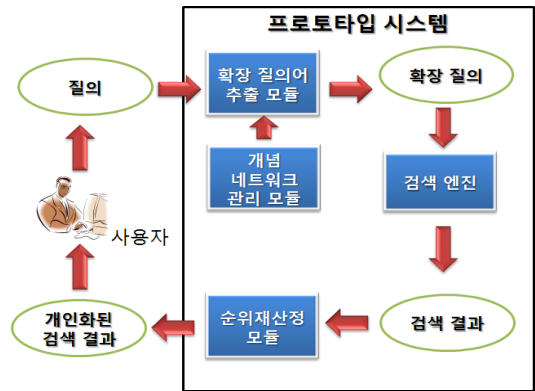


그림 6. 개인화 검색을 위한 프로토타입 시스템 구조
Fig. 6. A prototype system architecture for personalized search

IV. 실험 및 결과

1. 실험 데이터 및 성능평가 척도

제안 기법의 성능평가는 개인화 검색의 효능을 평가하는 실험을 수반하기 때문에 실험 데이터 및 실험 수행자 선정에 주의를 기울였다. 우선 실험 데이터의 구축을 위해 영문 뉴스 사이트인 www.foxnews.com에서 10만 개의 문서를 수집하였고, 뉴스 주제별로 질의어를 선정하였다. foxnews는 entertainment, health, scitech, travel, leisure, world, sports 등의 주제로 분류되어 있는데, 대표적인 주제를 추려내어 각 주제별로 실험자 4명을 할당하였다. 실험자는 이미 주어진 질의어들에 대한 질의의도를 결정하고 시스템에 질의를 수행하여 웹문서를 탐색하는 과정을 거치는 동안 시스템은 자동으로 사용자 프로파일을 생성하게 된다. 결과적으로 시스템은 20명의 사용자에게 대한 개념네트워크 프로파일을 생성하게 된다.

그리고 검색 결과를 비교하기 위한 측정치로서 DCG(Discounted Cumulative Gain)^[17]를 활용하였다. DCG는 검색엔진의 효율성을 측정하는 도구로 가장 많이 쓰이는 측정값들 중의 하나이다. DCG는 검색결과로 도출된 각 문서에 대해서 검색어와의 실제 관련도에 따라 점수를 부여하는 방식을 사용한다. 그런데 DCG에는 검색 결과의 순위(랭킹)보다 결과들의 관련도만을 정량화하여 계산하는 CG(Cumulative Gain) 방식이 있고, 검색 결과의 랭킹이 중요한 요소로 활용하면서 정규화된 측정 방식이 nDCG(Normalized Discounted Cumulative Gain)

가 있다. 본 실험에서는 nDCG 방식을 채택한다.

DCG는 관련도가 높은 문서가 검색 결과에 우선적으로 랭킹된다는 사실에 입각해서 그렇지 않을 경우 감점을 부과하는 방식을 사용한다. 따라서 DCG는 식 (3)과 같이 정의한다.

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log 2(1 + i)} \quad (3)$$

일반적으로 DCG는 위의 수식 그대로 사용하지 않고 0부터 1까지의 정규화된 수치인 nDCG로 표현하며 그 수식은 식 (4)와 같다.

$$nDCG_q = M_q \sum_{i=1}^K (2^{rel(i)} - 1) / \log(1 + i) \quad (4)$$

여기서, i 는 검색결과안에서의 순위값이며, M_q 는 정규화 상수를 의미한다. 그리고 $rel(i)$ 는 순위 i 에서의 결과 관련성을 의미하며, 상위 20개의 검색 문서에 대하여 관련성 정도에 따라 0에서 3까지의 정수값을 갖도록 하였다. 그 관련성은 해당 사용자가 직접 개입하여 결정해야 한다. 결국 nDCG값은 1에 가까울수록 우수한 검색 결과를 나타낸다.

개인화 검색의 효과를 극명하기 비교하기 위해 질의어 선정에 신중을 기하였으며 주로 다중 의미를 가지거나 핫이슈가 되는 중요 단어를 채택하였다. 대표적인 질의어의 예는 'apple', 'ruby', 'yankees', 'pirates', 'north korea', 'champions', 'trade', 'linux', 'jordan', 'Obama', 'Bible', 'rockets', 'LPGA' 등이 포함된다.

2. 성능평가 결과 및 분석

제안기법과의 비교 평가를 위해 비교 기준에 해당하는 개인화 기법을 설정하는 것이 필요하다. 기존 연구에서는 주로 질의 확장을 통해 개인화를 시도하였기 때문에, 본 실험에서는 비교 기준에 해당하는 개인화 기법으로서 모든 확장질의어를 초기 질의어에 병합시키는 기법을 설정하였다. 예를 들어, 그림 3의 개념네트워크 프로파일의 사용자가 질의확장을 위해 'ruby'와 연관된 키워드인 'program', 'language', 'tutorial', 'test'를 선택하였다면, 초기 질의어와 연관 키워드를 모두 포함한 "ruby & program & test & language & tutorial"라는 질의로 확장하는 것이다. 이 기법은 개념네트워크 사용자 프로

파일에 기반하여 확장질의어를 채택하는 것인데다가 전통적 질의확장 기법과 거의 동일하므로, 제안 기법과 비교하기에 공정한 기법이라 할 수 있다. 비교 기법은 초기 질의에 대한 확장 질의가 1개만 생성되므로 중첩도 개념이 적용되지 않지만, 제안 기법은 중첩도를 평가하는 기준에 따라 URL 기준 및 도메인명 기준 기법으로 구분된다.

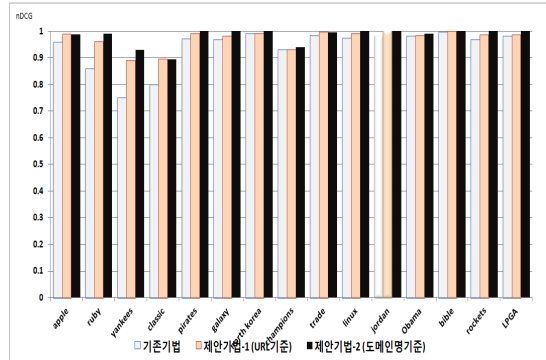


그림 7. 주제별 질의어에 대한 성능 비교
Fig. 7. Performance comparison for queries of various topics

그림 7은 3가지 개인화 기법에 대한 주제별 질의어에 대한 nDCG값을 보여준다. 그림에서 보는 바와 같이 제안 기법의 nDCG값이 기존 기법보다 평균적으로 3-5% 가량 우세하게 나타났다. nDCG값의 특성상 2%이상의 상승은 쉽지 않는 것이어서 이 향상 비율은 제안 기법의 개인화 효능이 큰 의미가 있음을 보여준다. 특히 제안 기법의 성능이 높은 질의어는 Scitech, Sports 분야의 다중 의미를 가지는 단어들이므로 파악되었다.

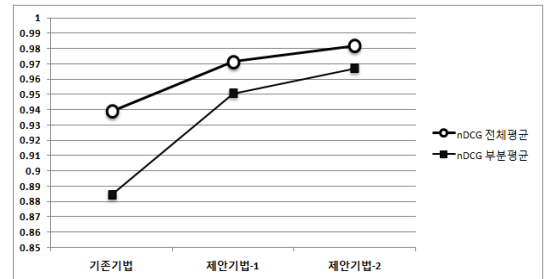


그림 8. 질의어 그룹별 평균 성능 비교
Fig. 8. Comparison of average performance measures by query groups

그림 8은 개인화 검색의 차별화 효과를 세부적으로 관

찾하기 위해 다중 의미가 두드러진 단어들과 그렇지 않은 단어들을 구분하여 nDCG 평균값을 비교한 것이다. 이 그림에서 nDCG전체평균과 nDCG부분평균은 각각 전체 질의어에 대한 nDCG 평균, 다중 의미도가 높은 질의어들만의 nDCG 평균을 의미한다. nDCG부분평균 측면에서 제안기법은 기준기법에 비해서 7~9%의 성능향상을 보이고 있다. 다중 의미도가 높은 질의어 집합의 경우, 전체적인 성능 향상에 비해 2배 이상의 효과를 보이고 있는 것은 제안 기법의 개인화 효능이 매우 효과적임을 나타낸다. 그리고 중첩도 계산의 기준으로서 URL에 비해 도메인명을 기준으로 설정한 것이 개인화 검색의 효과가 평균 1.1%, 최고 2.2% 우세하게 나타났다. 결과적으로 도메인명을 기준으로 한 중첩도 평가값이 순위 재조정애 활용됨으로써 웹검색 개인화에 크게 기여할 수 있음을 보였다.

V. 결론

사용자를 구별하지 않고 일률적인 검색 결과를 제공하는 기존 검색 방식에 비해 개인화 검색은 사용자의 특성에 맞춰 검색 결과를 제공하는 것이어서 이는 향후 웹 검색엔진이 성취해야 할 중요한 목표가 된다. 본 논문에서는 개인화 검색을 실현하기 위해 개념네트워크 기반의 사용자 프로파일을 적극 활용하여 새로운 형태의 순위 재조정 기법을 제안하였다. 이는 사용자 프로파일 기반 질의 확장에 대한 검색결과 집합으로부터 URL 및 도메인명에 대한 중첩도를 정량화함으로써 개인 성향에 따른 순위 재조정을 실현한 것이다. 본 연구 수행을 위해 개인화 검색을 위한 프로토타입 시스템을 구현하였으며 다른 성향의 실험자 그룹에 의한 다양한 검색 질의 실험을 통해 제안 기법이 전통적 개인화 기법에 비해 성능이 우수함을 보였다. 향후 개인 성향 및 집단지성(collective intelligence)을 동시에 고려한 순위 재조정 기법을 고안할 예정이다.

참고 문헌

- [1] A. Ramos, and S. Cota, "Search Engine Marketing", McGraw-Hill, 2008
- [2] P-A. Chirita, C. S. Firan, and W. Nejdl, "Personalized Query Expansion for the Web", Proceedings of Annual ACM Conference on Research and Development in Information Retrieval, 2007, pp.7-14
- [3] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-based personalized search and browsing", Web Intelligence and Agent Systems, Vol. 1, No. 3-4, 2003, pp.219-234
- [4] J. Hu, and P.K. Chan, "Personalized Web Search by Using Learned User Profiles in Re-ranking", Proceedings of the Internal Workshop on Web Mining and Web Usage Analysis, 2008, pp.78-83
- [5] F. Qiu, and J. Cho, "Automatic identification of user interest for personalized search", Proceedings of International World Wide Web Conference, 2006, pp.727-736
- [6] X. Shen, B. Tan, and C. Zhai, "Implicit user modeling for personalized search", Proceedings of the International Conference on Information and Knowledge Management, 2005, pp. 825-831.
- [7] <http://www.google.com/ig>
- [8] <http://my.yahoo.com>
- [9] <http://www.rollyo.com>
- [10] <http://likeness.com>
- [11] S. Lee, H. Kim, B. Lee, and S. Kang, "Concept Network-based User Profile Construction for Personalized Web Search" Proceedings of Korea Computer Congress, 2009, pp. 203-208.
- [12] S. Xu, S. Bao, B. Fei, Z. Su, Y. Yu, "Exploring folksonomy for personalized search", Proceedings of Annual ACM Conference on Research and Development in Information Retrieval, 2008, pp. 155-162.
- [13] P. Ferragina, A. Gulli, "A Personalized Search Engine Based on Web-Snippet Hierarchical Clustering", Proceedings of International World Wide Web Conference, 2005, pp. 801-810.
- [14] F. Akhlaghian, B. Arzanian, P. Moradi, "A Personalized Search Engine Using Ontology-Based Fuzzy Concept Networks", Proceedings in

International Conference on Data Storage and Data Engineering, 2010, pp. 137-141.

[15] U. Priss, "Formal Concept Analysis in Information Science", Annual Review of Information Science and Technology, Vol. 40, No. 1, 2006, pp.521-543.

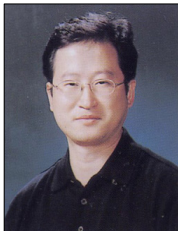
[16] <http://lucene.apache.org>

[17] H. Valizadegan, R. Jin, R. Zhang, and J. Mao, "Learning to Rank by Optimizing NDCG Measure", Proceeding of Neural Information Processing Systems, 2010, pp.41-48

※ 이 논문은 2011년도 서울시립대학교 연구년교수 연구비에 의하여 연구되었음.

저자 소개

김 한 준(정회원)



- 1994년: 서울대학교 계산통계학과 졸업 (이학사)
- 1996년: 서울대학교 전산과학과 대학원 졸업 (이학석사)
- 2002년: 서울대학교 컴퓨터공학부 대학원 졸업 (공학박사)
- 2002년~2002년: 서울대학교 공과대

학 Post-Doc

• 2002년~현재: 서울시립대학교 전자전기컴퓨터공학부 부교수
 <주관심분야> 데이터마이닝, 정보검색, 기계학습, 데이터베이스

장 재 영(정회원)

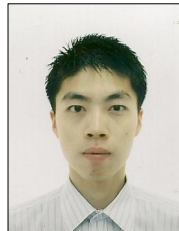


- 1992년: 서울대학교 계산통계학과 (이학사)
- 1994년: 서울대학교 계산통계학과 (이학석사)
- 1999년: 서울대학교 계산통계학과 (이학박사)
- 2000년~현재: 한성대학교 컴퓨터공

학과 교수

<주관심분야> 데이터베이스, 정보검색, 데이터마이닝

노 준 호(준회원)



- 2011년: 서울시립대학교 전자전기컴퓨터공학부(공학사)
- 2011년~현재: 서울시립대학교 전자전기컴퓨터공학부 대학원 재학
 <주관심분야> 정보검색, 데이터마이닝