한국어 시소러스를 이용한 웹 문서 추천 에이전트

서민혜* · 이성욱** · 서정연***

A Web-document Recommending System using the Korean Thesaurus

Min-Hye Seo* · Songwook Lee** · Jungyun Seo***

이 연구(논문)는 지식경제부 지원으로 수행하는 21세기 프론티어 연구개발사업
(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었습니다.

요 약

우리는 사용자의 행동을 관찰하고 학습하여 사용자 대신에 문서를 수집·분석함으로써 사용자에게 필요한 정보만을 추출하여 제공하는 웹 문서 추천 에이전트 시스템을 개발한다. 또한 우리는 이 시스템에 한국어 시소러스를 이용한 질의어 확장 방법의 적용을 제안한다. 한국어 시소러스를 이용한 질의어 확장을 위해, 새로운 웹 문서를 검색하기 위해 생성된 질의어를 한국어 시소러스를 통하여 그 하의어들을 찾아 후보 집합을 생성해 주고, TF-IDF와 상호 정보량을 이용하여 후보 집합 안에 있는 단어들 중에서 질의어와 가장 많은 관련 정보를 가지고 있는 단어를 추출함으로써 질의어를 확장해 주었다. 확장되지 않은 질의어만으로 웹 문서를 추천하게 되면 추천된 웹 문서의 수는 극히 제한적이지만, 질의어를 확장함으로써 보다 더 많은 유용한 웹 문서를 사용자에게 추천 및 제공 할 수 있다.

ABSTRACT

We build the web document recommending agent system which offers a certain amount of web documents to each user by monitoring and learning the user's action of web browsing. We also propose a method of query expansion using the Korean thesaurus. The queries to search for new web documents generate a candidate set using the Korean thesaurus. We extract the words which are mostly correlated with the queries, among the words in the candidate set, by using TF-IDF and mutual information. Then, we expand the query. If we adopt the system of query expansion, we can recommend a lot of web documents which have potential interests to users. We thus conclude that the system of query expansion is more effective than a base system of recommending web-documents to users.

키워드

웹 문서 추천 에이전트 시스템, 시소러스, 질의어 확장, TF-IDF, 상호정보량

* 한국신용정보(주)
** 충주대학교 (교신저자)
*** 서강대학교

접수일자 2008. 12. 10

Ⅰ. 서 론

최근 수년 동안 인터넷의 사용자와 정보량은 기하급수적으로 증가하고 있는 추세이다. 이러한 방대한 정보 중에서 사용자는 보통 자신에게 필요한 정보를 찾기 위해 직접 알고 있는 URL을 이용하거나, 검색 엔진을 통해 정보를 얻는다. 그러나, 이런 방법은 각 개인의 특성을 고려하지 않을 뿐 아니라, 그 정보에 대한 사용자의 의사를 정확하게 표현하기 어렵고, 적은 비용으로 원하는 정보를 찾기가 어렵다. 따라서, 사용자 개개인의 행동을 관찰하고 학습하여 사용자 관심분야에 대한 정보를 수집한 후, 이를 바탕으로 보다 적절한 웹 페이지를 추천하는 웹 문서 추천 에이전트 시스템이 필요하게 되었다.

웹 문서 추천 에이전트 시스템은 먼저 사용자로부터 관심 분야의 웹 문서에 대한 피드백 정보를 받아들인 후, 이를 형태소 분석과 구문 분석을 통하여 문서의 특징을 추출해 내고, 그 특징을 이용하여 사용자 프로파일(profile)을 생성한다. 생성된 사용자 프로파일은 질의어를 작성하기 위해 사용되고, 작성된 질의어는 다시 검색 엔진으로 보내져 사용자 관심 분야에 관한 정보를 갖고 있는 문서들을 검색하기 위해 사용된다. 시스템은 이렇게 검색된 문서들과 사용자 프로파일 사이의 유사도를 측정하고 이를 통해 높은 순위를 갖게 된 문서들만을 선별하여 이를 사용자에게 추천한다.

본 논문에서는 이와 같은 웹 문서 추천 에이전트 시스템을 구축함과 동시에, 보다 효율적으로 사용자 관심분야에 관련된 문서들을 추천하기 위하여 한국어 시소러스(Thesaurus)를 이용한 질의어 확장(Query-Expansion) 방법을 제안한다. 즉, 한국어 시소러스를 이용하여 사용자 프로파일에 관련된 유의어들을 검색하고, 검색된 유의어들을 가지고 질의어를 확장함으로서, 사용자 관심분야에 대한 정보를 가지고 있는 문서들을 보다 더 많이 사용자에게 추천하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구들을 소개하고, 3장에서는 본 논문에서 제안하는 시스템의 구성에 대하여 설명한다. 4장에서는 제안 시스템의 성능에 대한 실험 및 실험 결과를 살펴보며, 마지막으로 5장에서는 결론과 향후 과제를 제시한다.

Ⅱ. 관련 연구

질의어 확장은 질의어로 사용된 단어를 가지고, 그와 관련된 다른 용어를 더하여 자동으로 질의어를 확장하는 방법이다. 질의어를 확장하기 위해 단어를 추출하는 방법으로는 질의어의 특정 단어 추출, 말뭉치의 특정 단어 추출, 언어 관련 특정 단어 추출 등의 방법이 대표적인 방법이다.

질의어의 특정 단어 추출은 특정한 질의어에 관련된 문서의 부분집합 안에서 새로운 단어가 획득된 것으로, 특정 질의어와 관련된 문서의 부분집합을 알아내기 위해 사용자의 확인절차가 필요하다. WebMate[9]에서는 상호 정보량(Mutual Information)을 이용하여 특정 영역에 제한된 트리거 쌍(trigger pair)을 생성한 후, 트리거 쌍 안에 있는 단어 중에서 몇 가지를 선별하여 질의어를 확장하였다. 말뭉치 특정 단어 추출 방법은 전체의 문서 내용을 분석함으로써 질의어와 가장 유사하게 쓰인 단어를 찾아 질의어를 확장하는 방법이다[4][7].

언어 관련 특정 단어 추출 방법은 일반적인 시소러스를 이용하여 단어를 획득하는 방법으로 워드넷(WordNet)[2]을 이용하는 방법이 대표적이다. 그러나, 단어간의 모호성(ambiguity) 때문에 시소러스의 사용이 어려우며, 게다가 워드넷과 같은 일반적인 시소러스의 경우, 특정 분야에 관한 수집에는 적합하지 않다는 단점을 가지고 있다. [5]는 워드넷을 이용하여 어휘-의미 관계에 의한 질의어 확장을 시도하였다.

질의어 확장 방법은 직접 사람이 질의어를 확장하는 방법, 시스템이 어느 정도 자동으로 골라낸 용어 중에서 사람이 적합하다고 판단되는 용어를 선택하여 질의어를 확장하는 방법, 시스템이 스스로 학습하여 질의어를 자동으로 확장하는 방법 등의 3가지로 분류할 수 있다.

Ⅲ. 웹 문서 추천 에이전트

본 논문의 웹 문서 추천 에이전트 시스템의 구조는 [그림 1]과 같다.

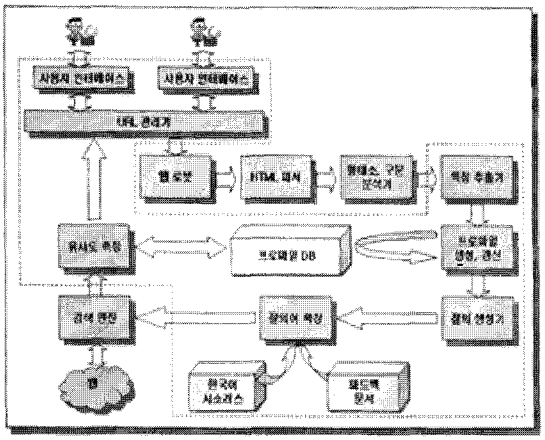


그림 1. 웹 문서 추천 에이전트의 구조
Fig. 1 Architecture of the system

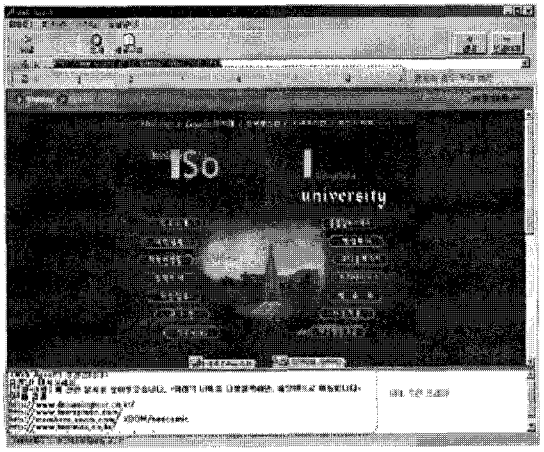


그림 2. 사용자 인터페이스
Fig. 2 User interface

사용자가 사용자 인터페이스(UI)를 통하여 인터넷을 검색하거나, 에이전트가 추천한 URL을 방문하여 평가하면, UI는 이를 URL관리기에게 넘겨주고, 이는 분석되어 사용자 프로파일을 생성한다. 시스템은 생성된 사용자 프로파일을 이용하여 사용자가 관심을 갖고 있는 웹 문서를 추천한다. 기존의 웹 문서 추천 에이전트는 단순하게 사용자 프로파일을 이용하여 질의어를 생성하고 그 질의어에 해당하는 검색엔진의 결과만을 추천해 주었지만, 본 연구에서는 정확률을 유지하면서, 한국어 시소러스(Thesaurus)를 이용하여 질의어를 확장하고, 확장된 질의어를 통한 검색엔진의 결과를 지속적으로 사용자 프로파일과 비교함으로서 더 많은 웹 문서를 추천하고자 한다.

3.1 사용자 인터페이스

[그림 2]는 본 시스템의 사용자 인터페이스를 나타낸 그림이다. 사용자는 연결버튼을 통하여 서버에 연결하게 되고, 서버에 연결된 이후에 사용자 인터페이스는 사용자의 웹 문서 검색 결과나 해당 웹 문서의 평가 결과를 URL관리자에게 넘겨준다. 맨 아래의 프레임(frame)에는 시스템이 추천한 웹 문서의 URL을 사용자에게 보여주고, 바로 그 위의 프레임에서는 사용자가 웹을 검색한 결과를 보여준다.

사용자는 자유롭게 웹 문서를 볼 수도 있고, 또는 찾아본 웹 문서에 대하여 점수를 매겨 그 문서에 대한 관심도를 표현할 수도 있다. 이렇게 사용자가 직접적으로 관심도를 표현한 문서는 명시적 피드백으로 반영되고, 사용자가 관심도를 표현하지 않고 탐색한 웹 문서는 암시적 피드백으로 반영된다.

3.2 전처리기

전처리기는 URL관리기, 웹 로봇, HTML 파서(Parser), 형태소 분석기, 구문 분석기, 자질(Feature) 추출기로 구성되어 있다. 먼저 사용자 인터페이스를 통하여 받아들여진 평가 URL과 피드백이 URL관리기로 넘겨져 통합 관리된다. 즉, URL관리기는 중복된 URL을 병합하고 검색 엔진 관련 URL을 삭제하는 역할을 한다. 병합된 URL은 웹 로봇[1]에게 넘겨져 그 URL에 해당하는 웹 문서의 HTML문서를 그대로 가져오게 되고, HTML 파서는 이러한 HTML 문서 내에 있는 태그(tag)등을 삭제하여 텍스트(text) 문서의 형태로 바꿔주게 된다.

3.3 문서 자질 추출

문서의 자질 추출에서 uni-gram을 사용할 경우 한 단어만을 이용하기 때문에 단어와 단어 사이의 문맥(context) 정보를 반영할 수 없고, 단순 n-gram을 사용할 경우 여러 단어를 사용하므로 단어 사이의 문맥 정보를 어느 정도 반영할 수는 있으나, 많은 문맥 정보를 지니고

1) 본 논문에서 사용된 웹 로봇은 wget[11]이다.

있는 지배소와 의존소 사이의 관계를 사용할 수 없다. 그러므로, 슬라이딩 윈도우(sliding window)와 의존 구조에 의한 단어쌍을 n-gram과 함께 이용함으로서, 비록 인접하지는 않았지만 상관관계가 있는 단어들을 문서의 특징으로 사용하여 사용자의 관심도를 더욱 잘 반영할 수 있게 한다[1].

3.4 사용자 프로파일 생성

사용자의 프로파일을 생성하기 위해, 자질 추출기가 먼저 불용어를 제거한 다음 각 문서에 나타나는 단어들에 대해 가중치를 부여해주고, 그 중에서 가중치가 높은 상위 N개의 단어를 추출하여 자질로 사용한다. 가중치를 부여하는 과정에서, 사용자의 관심도를 반영하기 위하여 각 단어의 가중치에 문서에 대한 사용자의 평가를 합산하였다. 특히, 자질 추출을 위한 각 단어의 가중치를 구하는 방법으로는 TF-IDF(Term Frequency-Inversed Document Frequency)를 이용하였다[6]. 여기에 사용자의 관심도를 반영하기 위하여 TF값에 사용자의 평가 점수를 합산함으로써, TF-IDF의 식을 식 (1)과 같이 변형하였다. 식 (1)에서 w는 단어 또는 단어쌍을, TF(w)는 w의 빈도수를, score(D)는 문서 D에 대한 사용자 평가 점수이다. C는 사용자의 관심분야에 대해 긍정적 피드백을 받은 문서의 집합, $\bar{c}$는 부정적 피드백을 받은 문서의 집합을 각각 나타낸다. DF(w,C)는 C에서 w가 나타난 문서의 수를 뜻한다.

$$word\_weight = (TF(w) + score(D)) \log \frac{P(w|C)}{P(w|\bar{C})} \quad (1)$$

$$P(w|C) = \frac{DF(w,C)/(\sum C_i)}{C/(\sum C_i)} \quad (2)$$

3.5 유사도 측정

웹 문서 추천 시스템에서 생성된 질의어로 검색 엔진을 통해 새로운 웹 문서들을 가져왔을 경우, 그 웹 문서를 그대로 모두 추천한다면 그 정확률은 상당히 낮아지게 될 것이다. 그러므로, 유사도 측정을 통하여 임계값 이상의 유사도 값을 갖는 문서들만을 사용자에게 추천함으로써 보다 사용자의 관심 분야에 가까운 문서를 추천해 주게 된다.

새로운 웹 문서와 사용자 프로파일 사이의 유사도를 측정하기 위해, 식(3)과 같은 단순 베이시안 모델(Naïve Bayesian classifier)을 사용하였다[8]. 식(3)은 새로운 웹 문서 Doc가 특정 프로파일 cj에 속할 확률로서, TF(Fj,doc)는 웹 문서 doc에서 특징으로 추출된 자질 Fj의 빈도수를 나타낸다. 또한, 식(4)는 특정 프로파일 c에서 문서에 포함하고 있는 자질 Fj를 가질 확률을 나타내며 |V|는 전체 자질의 개수이며, TF(Fi,c)는 프로파일 c안에 포함된 문서에 나타난 자질 Fi의 빈도수를 나타낸다.

$$P(c_j|Doc) = \frac{P(c_j)\prod_{F_j \in Doc} P(F_j|c_j)^{TF(F_j,doc)}}{\sum_{c' \in C} P(c')\prod_{F_j \in Doc} P(F_j|c')^{TF(F_j,doc)}} \quad (3)$$

$$P(F_j|c) = \frac{1 + TF(F_j,c)}{|V| + \sum_i TF(F_i,c)} \quad (4)$$

3.6 정보 여과 방법 및 피드백

본 논문에서는 정보 여과 방법으로 내용(Content) 기반 여과 방법을 사용하였고, 사용자 피드백 방법으로는 명시적 피드백과 암시적 피드백을 사용하였다.

내용 기반 여과 방법은 이전에 사용자가 관심도를 표현했던 문서들의 내용을 분석하여 각 단어에 가중치를 부여한 후, 가중치가 높은 단어들을 추출하여 그것을 기반으로 새로운 문서들 중에서 사용자가 관심을 가질 만한 문서들만을 여과하여 추천하는 방법이다.

명시적 피드백은 사용자에게 직접 문서에 대한 평가를 요구하는 방법으로 본 논문에서는 사용자가 -3점부터 3점까지 점수를 부여할 수 있도록 하였다. 암시적 피드백은 사용자의 웹 탐색 행위만을 관찰함으로서 웹 문서에 대한 가중치를 부여하는 방법으로 본 논문에서는 사용자가 방문하는 웹 문서의 URL을 관찰하여 피드백으로 이용하였다.

3.7 질의어 확장에 의한 검색 성능 개선

우리는 질의어를 확장하기 위하여 언어 특정 관련 단어의 확장을 이용하였으며, 이를 위해 먼저 한국어 시소러스를 이용하여 후보 집합을 생성해주게 되고, 이 후 두 가지 종류의 여과 방법을 이용하여 질의어를 확장할 후보를 선택하였다.

[그림 3]은 한국어 시소러스를 이용하여 질의어를 확장하기 위해, 해당하는 질의어에 대한 후보 집합 생성의 예를 보여주고 있다. 후보 집합을 생성하기 위해서는 사용자 프로파일에 있는 단어를 이용하여 질의어를 생성한 다음, 그 질의어에 해당하는 단어를 한국어 시소러스에서 찾아 그 단어의 하의어를 모두 후보 집합에 등록한다.

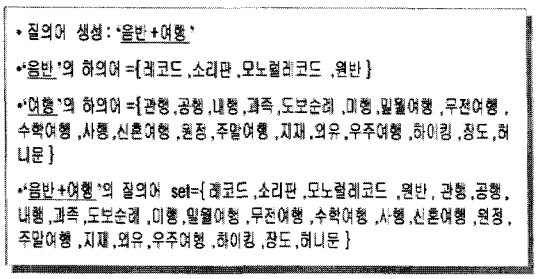


그림 3. 한국어 시소러스를 이용한 후보 집합 생성 예
Fig. 3 Example of candidate set by using Korean thesaurus

그러나, 후보 집합에 등록된 모든 단어를 질의어 확장에 사용하는 것은 아니다. 후보 집합 중에서도 사용자의 관심 분야와 관련이 있는 단어들을 추출하기 위하여 TF-IDF[6]와 상호정보량(Mutual Information)[3]을 각각 사용하였다. 사용한 TF-IDF의 식은 다음과 같다.

$$TF(w,d) \quad IDF(w) = \log \frac{|D|}{DF(w)} \tag{5}$$

$$d^{(i)} = TF(w_i,d) \times IDF(w_i) \tag{6}$$

상호정보량은 단어 c와 단어 w사이의 관련 정도를 나타내는 값으로, 식 (7)과 같이 계산할 수 있다. fc, fw 는 c, w 각각의 빈도수를 나타내며, fcw는 c와 w가 함께 나타난 빈도수이다.

$$MI(cw) = \log_2\left(\frac{Nf_{cw}}{f_c f_w} + 1\right) \tag{7}$$

$$MI_{tot} = MI_p - MI_N \tag{8}$$

이와 같은 식을 사용하여, 후보 집합 안에 있는 모든 단어들에 대하여 질의어 c와 후보 집합의 단어 w사이의 상호정보량을 긍정적 피드백 관련 문서와 부정적 피드백 관련 문서로 대상을 각각 달리하여 따로 계산을 한 후, 식 (8)과 같이 긍정적 피드백 문서의 상호정보량 MIP와 부정적 피드백 관련 문서 MIN의 차를 최종 결과값으로 사용하였다. TF-IDF와 MI는 둘 다 임계치 이상의 TF-IDF와 MI값을 가진 단어들만을 가지고 질의어를 확장하였다.

Ⅳ. 실험 및 평가

실험은 사용자 10명을 대상으로 하였고, 사용자들은 사용자 인터페이스를 이용하여 자유롭게 웹 문서를 검색하면서 명시적 피드백을 표현하고자 하는 문서에 대해서는 점수를 부여하였다. 시스템은 하루에 한번 사용자들의 정보를 모아 각 사용자들의 관심 분야와 관련이 있는 새로운 웹 문서를 추천해 주었다. 본 시스템에서는 사용자들이 검색한 총 1,623개의 HTML문서를 데이터로 사용하였고, 질의어 확장을 위해 ETRI 시소러스를 이용하였다.

제안된 시스템의 성능 개선 여부를 알아보기 위하여 질의어를 확장하지 않은 경우와 TF-IDF를 이용하여 질의어를 확장한 경우, 상호 정보량을 이용하여 질의어를 확장한 경우 각각에 대하여 웹 문서 추천하는 실험을 하였다. 이러한 시스템에 대한 객관적인 평가 방법이 아직 없으므로, 본 실험에서는 시스템을 질의어 확장에 따른 추천 문서의 수와 추천 웹 문서에 대한 정확률을 가지고 평가하였다.

[그림 4]는 시스템이 5회 추천할 때, 질의어 확장을 하지 않은 기본 시스템(Base)의 추천 URL의 수와 기본 시스템에 질의어 확장을 추가한 시스템의 추천 URL 수를 나타내었다. 이처럼 질의어를 확장하지 않은 기본 시스템보다 TF-IDF나 상호 정보량을 이용하여 질의어를 확장한 것이 더 많은 문서를 사용자에게 추천하였으며, TF-IDF를 이용한 방법과 상호 정보량을 이용한 방법에서 총 누적 추천 수는 각각 70개, 68개로 별 차이를 보이지 않았다.

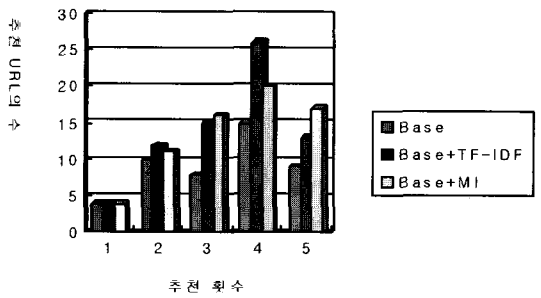


그림 4. 질의어 확장에 따른 추천 문서의 누적분포수
Fig. 4 The number of documents recommended by query expansion

[그림 5]는 질의어 확장에 따른 정확률 측정을 한 결과이다. 각 추천 문서에 대한 정확성 판단의 근거로는 개별적 사용자들의 피드백을 이용하였다. 질의어를 확장하지 않은 웹 문서 추천과 TF-IDF 또는 상호 정보량을 이용하여 질의어를 확장한 웹 문서 추천 사이의 정확률은 근소한 차이를 보였다. 이는 질의어를 확장하여 더 많은 문서를 추천하였음에도 정확률이 낮아지지 않았다는 것을 뜻한다.

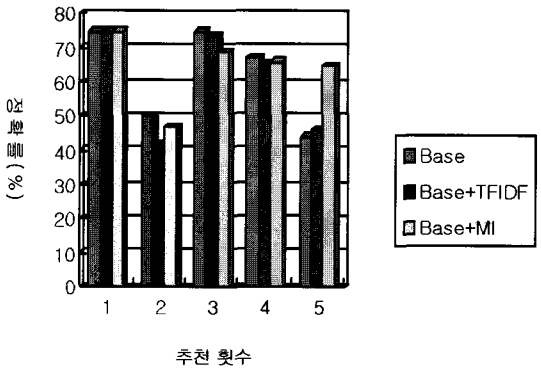


그림 5. 질의어 확장에 따른 정확률 측정
Fig. 5 Accuracy of query expansion

[표 1]은 전체 시스템의 추천 URL의 수와 그 중 사용자에게 적합한 연관 URL의 수를 나타내고 그 정확률을 나타낸 표이다. [표 1]에서 나타나듯이, 본 시스템은 정확률을 유지하면서 질의어 확장을 통해 사용자에게는 보다 많은 웹 문서를 추천할 수 있었다. 즉 질의어 확장에 의한 재현율의 향상 효과가 있다고 할 수 있다. 만약 전체 문서 컬렉션에서 사용자가 원하는 문서의 개수가 50개라고 가정한다면 기본시스템의 재현율 54%를 질의어를 확장함으로써 각각 82%, 80%로 향상시키는 결과를 가져온다고 할 수 있다.

표 1. 전체 시스템 평가 결과
Table. 1 Results of experiments

	Base	TF-IDF	MI
추천 URL수	46	70	68
연관 URL수	27	41	40
정확률(%)	58.7	58.6	58.8

V. 결론 및 향후 과제

우리는 웹 문서 추천 에이전트 시스템을 개발하였고, 보다 많은 웹 문서를 효과적으로 사용자에게 추천하고자 한국어 시소러스의 하의어를 이용하여 후보 집합을 설정하고, 후보 집합에 있는 단어들 중에서 TF-IDF와 상호 정보량을 각각 사용하여 사용자 관심 분야와 관련 있는 단어를 선택함으로서 질의어를 확장하였다. 또 이를 이용하여 웹 문서를 추천한 결과, 질의어를 확장하지 않은 시스템보다 질의어를 확장한 시스템이 기존의 정확률을 유지하면서 더 많은 웹 문서를 추천하는 효과를 가져왔다. 즉 정확률을 유지하며 재현율을 향상시키는 효과를 가져와 전체적인 시스템 성능을 향상시킬 수 있었다.

앞으로, 좀더 좋은 성능의 형태소・구문 분석기 등의 언어처리기가 필요하다. 또한 본 시스템에서는 단순 베이시안 분류를 유사도 측정에 사용하므로, 단어 발생의 독립성을 가정한 단점이 있을 수 있다. 따라서 이를 보완할 수 있는 알고리즘을 연구한다면 더 효과적인 프로파일 생성이 가능할 것이다. 시스템 성능 향상을 위한 적절한 혼합 여과 방법 등의 연구도 필요하다. 본 시스템에서는 질의어 확장을 위해 TF-IDF와 상호 정보 척도를 각각 사용하였으나, 질의어 확장을 위한 좀 더 다양한 접근 방법이 필요하며, 질의어 확장을 통해 웹 문서를 추천하는 데 있어서는 무엇보다도 추천에 대한 정확률을 향상시킬 수 있는 방법에 관한 연구가 요구된다.

참고문헌

[1] 윤윤경, 효과적인 웹 문서 추천을 위한 동적 사용자 프로파일 생성 기법, 서강대학교 석사 학위 논문, 1999

[2] George Miller. Special Issue, "WordNet : An on-line lexical database", International Journal of Lexicography, 3(4), 1990

[3] Church K. W. and Hanks P., "Word Association Norms, Mutual Information and Lexicography", Computational Linguistics, 16(1), pp. 22-29, 1990

[4] Chengfeng han, Hideo Fujii, W. Bruce Croft, "Automatic Query Expansion for Japanese Text Retrieval", UMass Technical Report, 1994

[5] Ellen M. Voorhees, "Query Expansion using Lexical-Semantic Relations", SIGIR '94, 1994

[6] lewis, D., D., Gale, W., A., "A Sequential Algorithm for Training Text Classifiers", Proceeding of the 7th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, Dubline, 1994

[7] Susan Gauch and Jianying Wang, "A Corpus Analysis Approach for Automatic Query Expansion", CIKM '97, 1997

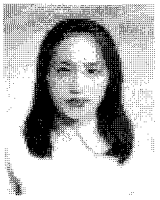
[8] Joachims, T., "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization", Proceedings of the 14th International Conference on Machine Learning ICML97, 1997

[9] Liren Chen and Katia Sycara, "WebMate : A Personal Agent for Browsing and Searching", 2nd International Conference on Autonomous Agents and Multi Agent System, 1998

[10] Davide Trucato, Fred Popowich, Janine Toole, Dan Fass, Devlan Nicholson, Gordon Tisher, "Adapting a synonym database to specific domains", ACL 2000, 2000

[11] www.gnu.org/software/wget/wget.html

저자소개



서민혜(Min-Hyeo Seo)

2001년 서강대학교 컴퓨터학과 석사
2001년-현재 (주) 한국신용정보 연구원
※관심분야 : 인터넷응용시스템, 한국어정보처리

이성욱(Songwook Lee)

1996년 서강대학교 컴퓨터학과 학사
1998년 서강대학교 컴퓨터학과 석사
2003년 서강대학교 컴퓨터학과 박사
2004-2005년 LG전자기술원 선임연구원
2005-2007년 동서대학교 컴퓨터정보공학부 전임강사
2007년-현재 국립충주대학교 컴퓨터과학과 전임강사
※관심분야 : 인터넷응용시스템, 한국어정보처리

서정연(Jungyun Seo)

1981년 서강대학교 수학과 학사
1985년 University of Texas at Austin, Computer Science, M.S.
1990년 University of Texas at Austin, Computer Science, Ph.D
1990-1991년 UniSQL Inc. 선임연구원
1991-1995년 KAIST 전산학과 조교수
1995년 - 현재 서강대학교 컴퓨터공학과 교수
※관심분야 : 자연어처리, 대화형 에이전트, 한국어 정보처리, 텍스트마이닝