전문검색엔진을 위한 개념망의 개발

주 정 은* · 구 상 회**

Development of a Concept Network Useful for Specialized Search Engines

Joung Eun Ju* · Sang Hoe Koo**

Abstract

It is not easy to find desired information in the world wide web. In this research, we introduce a notion of concept network that is useful in finding information if it is used in search engines that are specialized in domains such as medicine, law or engineering. The concept network that we propose is a network in which nodes represent significant concepts in the domain, and links represent relationships between the concepts.

We may use the concept network constructor as a preprocessor to specialized search engines. When user enters a target word to find information, our system generates and displays a concept network in which nodes are concepts that are closely related with the target word. By reviewing the network, user may confirm that the target word is properly selected for his intention, otherwise he may replace the target word with better ones discovered in the network.

In this research, we propose a detailed method to construct concept network, implemented a prototypical system that constructs concept networks, and illustrate its usefulness by demonstrating a practical case.

Keyword : information retrieval, concept network, specialized search engine

* 고려대학교 일반대학원 디지털경영학과 박사과정(kquilt@korea.ac.kr)
** 고려대학교 경영정보학과 교수(skoo@korea.ac.kr)

1. 서 론

월드와이드 웹에서 제공되는 정보는 한곳에서 집중되어 관리되지 않고 웹에 연결된 수많은 컴퓨터에 분산되어 저장 관리된다. 또한 이렇게 분산되어 관리되는 정보는 성격, 내용, 표현 방식이 각 사이트별로 상이하게 다를 뿐 아니라, 시간의 흐름에 따라 역동적으로 변화한다. 따라서 웹에서 일반 사용자가 자신이 원하는 정보를 찾기란 쉬운 일이 아니다. 정보검색 서비스는 사용자들이 원하는 정보를 웹에서 효율적으로 찾아주는 기능으로서 현재 대다수의 포털 사이트에서 기본적으로 제공하고 있다[김태수, 이재윤, 1999].

현재 대부분의 정보검색 서비스는 질의어의 불리언 검색을 사용하여 웹페이지를 검색한 후, 검색된 페이지를 각자의 고유한 순위 알고리즘에 의해 나열하여 사용자에게 제공한다. 그러나 이렇게 찾아진 검색결과에 대한 사용자의 만족도는 낮은 편이다.

그 요인으로써 우선 사용자는 시스템이 보유하고 있는 색인에 대한 정보를 거의 알 수 없기 때문에 적절히 질의를 작성하지 않으면 원하는 정보를 검색 결과로써 확인하기가 쉽지 않다. 둘째로 검색을 시행하는 분야가 모든 분야에 걸쳐서 이루어지기 때문에 결과로써 나타나는 문서의 수가 엄청나게 많고 이 들 정보 중에서 사용자가 다시 원하는 정보를 찾으려면 시간과 노력이 더 소요된다.

이러한 문제점을 해결하기 위해서 본 연구에서는 개념망을 제시한다. 개념망은 전문 검색엔진에서 사용하기 위해 개발되었다. 전문검색엔진이란 의료, 법률, 전자상거래 등과 같이 특정 분야의 정보를 검색하여 주는 검색엔진을 의미한다. 개념망은 특정 도메인에 포함되는 주요 개념과 이들 개념 사이의 관계성을 네트워크 형태로 제시하여 전문검색엔진의 프리프로세서로 활용될 수 있다.

본 연구에서는 개념망 구축에 대한 알고리즘을 제시하고, 제시된 알고리즘을 사용하여 구체적인 개념망을 작성해 본다. 그리고 작성된 개념망이 실제 검색엔진에서 어떻게 사용될 수 있는지에 대한 방안도 제시한다.

2. 관련 연구

본 연구에서 개발한 개념망은 특정 분야의 문서 집합으로부터 이를 대표하는 주요 개념을 추출하고, 추출된 개념들 사이에 관련성이 높은 경우, 링크를 설정함으로써, 개념을 노드로 그리고 이들 사이의 관련성을 링크로 구축한 네트워크를 의미한다. 본 연구는 먼저 명사를 추출한다는 측면에서 [이상준, 류근호, 2001 ; 이종인 외 3인, 1999 ; 정영미, 1993]등의 TF-IDF(termfrequency -inverted document frequency) 알고리즘과 유사하다. 그리고 이렇게 추출된 개념사이의 연관성을 분석하여 링크를 설정한다는 측면에서 [우선미외 2인, 2001 ; 임형근]과 같은 용어연관성 분석 방식과 유사하다. 연관성이 높은 개념사이에 링크를 설정하여 개념으로 구성된 네트워크를 구축한다는 의미에서 시소러스[김태수, 이재윤, 1999 ; 이종인외 3인, 1999 ; 정영미, 1993]나 워드넷[김민수외 2인, 1995 ; 김태수, 이재윤, 1999]과 유사하다. 본 장에서는 이들 연구와 본 연구를 비교 분석한다.

TF (term frequency)는 한 단어가 한 문서 내에 등장하는 횟수를 나타내고 DF (document frequency)는 한 단어가 검색된 N개의 문서의 집합 중에서 몇 개 문서에 등장하는 가를 나타낸다[이상준, 김태수, 2001 ; 이종인 외 3인, 1999 ; 정영미, 1993]. 특정 검색어가 한 문서에 많이 나타난다면 그 문서는 해당 검색어에 대해 중요한 문서라고 판단할 수 있지만 여러 문서에 걸쳐 모두 나타난다면 그 단어에 대한 중요도는 떨어진다고 볼 수 있다[Baeza-yates, 1999]. TF와 IDF가 동시에 높아지면, 특정 문서를 기타 문서와 구분하는 중요한 척도로 활용될 수 있다.

따라서 TF-IDF는 문서를 대표하는 개념의 추출을 목적으로 하는 본 연구와는 달리 문서의 분류에 유용하게 사용된다.

현재 연구된 대부분의 용어연관성 분석방식은 공기(共起 ; co-occurrence)와 공기빈도(共起頻度 ; co-occurrence frequency)를 활용한 통계적 방법에 기이한다[우선미외 2인, 2001 ; 임형근, 2001]. 본 연구의 링크 설정 방식 역시 이들 연구결과를 활용하였기에 용어연관성에 있어서는 이들 연구와 차이가 없다.

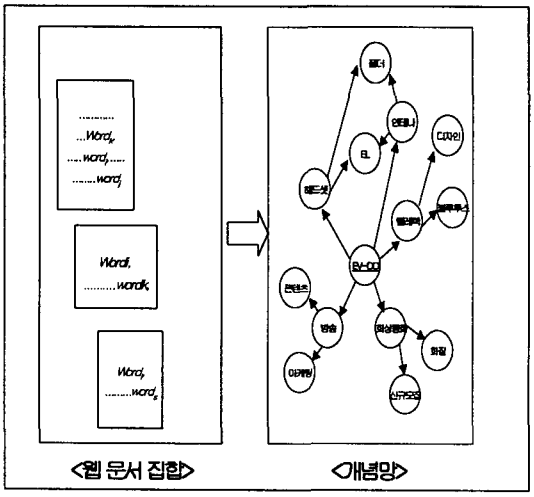
시소러스는 특정 분야에 존재하는 개념간의 관계를 명시하기 위하여 조직화한 색인어의 집합이다[김태수외 2인, 1995 ; 이종인외 3인, 1999 ; 최명목, 김민구, 2000 ; Baeza-yates, 1999]. 시소러스는 대규모 정보를 효율적으로 검색하는데 사용된다는 측면에서 본 연구의 목적과 크게 차이가 없으나, 구축방식과 관계성 등에는 차이가 있다. 시소러스는 상하관계, 동의관계, 대체관계 등 정교한 관계성이 제공되나 시소러스의 구축이 수작업에 의한다는 측면에서 웹정보와 같이 역동적으로 변화하는 소스의 분석에는 활용하기가 어렵다.

워드넷은 단어 간의 관계를 단어 중심으로 표현된 사전과는 달리 의미중심으로 표현한 어휘 데이터베이스이다[김민수 외 2인, 1995 ; 김태수, 이재윤, 1999]. 워드넷의 노드는 synset을 이루고 링크는 이들간의 의미적인 관계를 표현한 네트워크형태로 되어 있다. 여기서 synset은 비슷한 의미를 가지는 단어들의 집합을 뜻한다. 워드넷 역시 정보검색의 효과를 높일 수 있다는 측면에서는 본 연구와 유사하나, 수작업에 의존하여 구축하는 워드넷과는 달리 개념망은 완전히 자동화되어 생성된다.

3. 개념망

3.1 개념망의 정의

개념망은 특정 도메인(domain)에 포함되는 주요 개념(concept)과 이들 개념사이의 관계성(relation)을 망(network)로 표현한 것이다. 즉, 특정 도메인에서 문서 집합을 선정한 뒤 이들 문서 집합을 대표할 수 있는 단어를 추출하고 이들 단어들 사이의 관계성을 분석하여 관계성이 높은 단어들 사이에 링크를 연결함으로써 단어들의 네트워크를 형성한 것이다(<그림 1>).



<그림 1> 개념망의 정의

개념망은 검색엔진의 프리프로세서로 활용될 수 있다. 이 경우, 포털 사이트와 같은 일반적인 검색엔진보다는 전문 검색 엔진을 대상으로 할 때, 효과적으로 활용될 수 있다. 따라서 특정 도메인을 선정하여 이 도메인을 대상으로 하여 개념망을 구성하게 된다. 특정 도메인은 기존의 산업 분류에서 말하는 특정 산업 분야가 될 수도 있고 포탈 사이트에서 제공되는 카테고리가 될 수도 있다. 웹상에는 분야별로 전문화된 정보와 서비스를 제공하는 사이트들이 많이 존재한다. 이들 사이트들이 특정 도메인의 문서 집합이 될 수 있다.

3.2 개념망 구축 절차

개념이란 문서집합 전체를 대표하는 단어를 의미한다. 주요 개념으로 채택된 단어를 대상으

로 각각의 단어 쌍을 형성하여 동일 문서 내에 공동 존재 빈도를 조사한다. 공동 존재의 빈도가 높으면 이 들 단어사이의 관계성이 높은 것으로 가정하고 이들 사이에 링크를 연결하면 개념망이 구축된다. 개념망을 구축하는 순서를 나타내면 다음과 같다. <그림 2>는 개념망 구축 절차를 도식화한 것이다. 이를 요약하면 다음과 같다.

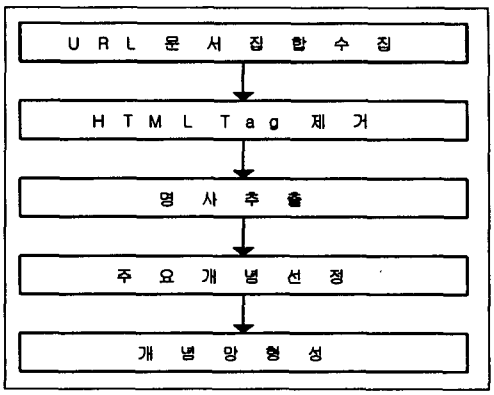
① URL 문서 집합 수집 : 개념망 구축 대상이 되는 영역에 대하여 정보의 질이 우수하고 잘 알려진 사이트 또는 해당 영역 전문가가 추천하는 사이트를 선정하여 이들의 URL를 수집하여 놓은 문서들의 집합이다.

② HTML Tag 제거 : HTML 파서를 이용하여 수집된 문서들의 HTML 태그를 제거하여 컨텐츠(contents)만을 추출하기 위한 과정이다. 본 연구에서 제시한 개념망은 태그와 같은 구조적인 정보를 고려하지 않으므로 모든 HTML 태그를 제거한 순수 컨텐츠만을 분석 대상으로 한다.

③ 명사 추출 : 태그가 제거된 컨텐츠들을 대상으로 주요 개념으로 추출하기 위해 명사만을 선별하는 과정이다. 본 연구에서 제시한 개념망은 명사를 대상으로 하므로 그 외의 품사는 필요하지 않아 형태소 분석기 [13]를 이용하여 명사만을 뽑는다.

④ 주요개념 선정 : 앞의 세 단계의 전처리 과정을 거쳐 얻어진 명사들을 분석하여 주요 개념을 추출한다. 방법은 본 연구에서 제시하는 주요 개념 추출 알고리즘을 적용하여 일정 임계치 이상의 단어들을 주요 개념으로 선정한다.

⑤ 개념망 구축 : 앞 단계에서 선정된 주요 개념은 관계성을 조사하여 관계성이 높다고 여겨지면 링크를 설정한다. 여기서 관계성이란 한 쌍의 단어가 같은 문서에 출현하는 빈도가 높을수록 관계가 높다고 간주한다.



<그림 2> 개념망 구축 순서

3.3 개념 추출 방법

주요 개념이란 선택된 문서들을 가장 잘 대표하는 명사의 집합이다. 이들 주요 개념은 전체문서를 대상으로 고려한 절대적 중요도와 개별 문서간의 크기를 고려한 상대적 중요도의 곱으로 계산한다.

절대적 중요도는 특정 단어가 전체 문서에서 출현하는 총 출현 빈도를 의미하고, 상대적 중요도는 특정 단어가 각각의 개별 문서 내에서의 단어 출현 빈도를 전체 단어 출현 빈도로 나누어서 계산한 것을 합한 것이다. 크기가 작은 문서에서 100번 출현한 단어가, 크기가 큰 문서에서 100번 출현한 단어보다 중요하다. 상대적 중요도는 이러한 문서간의 상대적 크기를 고려한 중요도를 나타낸다. 주요 개념 추출에 사용할 식은 다음과 같다. 식에서 임계치는 실험을 통해 추출된 값을 조사함으로서 추정하였다.

$$w_i = \text{절대가중치}_i \times \text{상대가중치}_i > \theta_1$$

$$= \text{워드}\,i\text{의 총빈도} \times \sum_j \frac{\text{워드}\,i\text{의 문서}\,j\text{에서의 빈도}}{\text{문서}\,j\text{의 모든 단어 빈도}} > \theta_1$$

$$= \sum_{j=1}^{n} wf_{ij} \times \sum_{j=1}^{n} \frac{wf_{ij}}{\sum_{k=1}^{m} wf_{kj}} > \theta_1$$

wf_{ij} : $word_i$의 doc_j에서의 *frequency*

m : *word*개수

n : *doc*(문서)개수

θ_1 : 임계치

3.4 망 구축 방법

개념망 구축의 목적은 사용자가 입력한 단어와 유관한 개념들을 사용자에게 제시함으로써 사용자가 입력한 키워드가 그의 검색 의도를 적절히 반영한 단어인지 확인하고, 사용자의 검색 의도에 더욱 적절한 다른 검색어가 존재하는지를 조사하고, 또한 다른 검색어를 추가함으로써 보다 효과적인 검색을 수행할 수 있도록 하는데 있다.

개념사이의 관계성은 단어쌍의 동일 문서 내 공동존재 정도에 의해서 결정된다. 어떤 두 단어가 같은 문서에서 매우 자주 출현한다면 두 단어는 관계성이 높은 것으로 간주하며 이 관계성이 특정 임계치 이상이면 두 개념 사이에 관계가 존재하는 것으로 간주하여 링크를 설정한다.

문서의 크기가 커지면, 별로 관계없는 두 단어의 공동 출현 빈도도 비례하여 커진다. 이러한 오류를 배제하기 위하여, 공동존재 빈도를 표준화할 필요가 있다. 본 연구에서는 두 단어의 공동출현 빈도를 두 단어의 총 출현빈도로 나누어 표준화한다.

두 단어의 연관성 정도를 나타내는 식은 아래와 같다. 식에서 임계치는 실험을 통해 추출된 값을 조사함으로서 추정하였다.

$$R(w_i, w_j) = \frac{\text{단어}i\text{와 단어}j\text{의 문서내 공동출현빈도}}{\text{단어}i\text{와 단어}j\text{의 전체출현빈도}} > \theta_2$$

$$= \frac{\sum_{k=1}^{n} MIN(wf_{ik}, wf_{jk})}{\sum_{k=1}^{n} wf_{ik} + \sum_{k=1}^{n} wf_{jk}} > \theta_2$$

$R(w_i, w_j)$: $word_i$와 $word_j$사이의관계성

wf_{ij} : $word_i$의 doc_j에서의 *frequency*

θ_2 : 임계치

4. 개념망의 구현과 사례

4.1 시스템 구현

본 연구에서는 앞서 설명한 개념추출 방법과 개념망 구축방법을 구현해 보았다. 구현 언어는 비주얼베이직을 이용하였고, 문서 집합에서 추출된 결과들을 단계별로 저장하고 각 단계별 결과를 저장하기 위해 MS-ACCESS를 이용하였다.

문서집합에서 주요 개념을 추출하기 위한 전처리 과정으로 명사를 추출하기 위해 국민대학교 자연언어처리 연구소에서 개발한 HAM 5.0 [13]을 이용하였다.

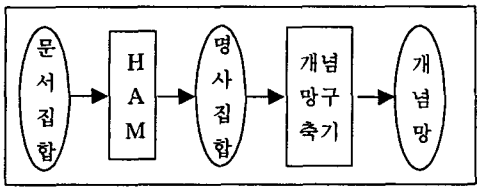
본 연구에서 분석한 문서는 KTSET[14]문서와 유명 검색엔진의 카테고리를 활용하여 수작업으로 수집한 문서 150개를 대상으로 실시하였다. KTSET을 활용한 이유는 본 시스템의 분석 대상이 웹 문서인데, KTSET 문서가 일반적인 웹 문서와 구조적으로 매우 유사하며 정보 검색의 성능을 판단하기 위해 가장 널리 사용되는 표준화된 문서 집합이기 때문이다.

수작업으로 수집한 문서 집합은 기존의 포탈 사이트의 분류에서 정보 통신 분야 중 휴대폰에 관련된 사이트로 한정하여 무작위로 선정하였다. 실험 상의 편의를 위해서 쇼핑몰과 같이 상업적인 사이트를 제외하고 정보를 주로 제공하는 사이트들을 선별하여서 문서 집합으로 채택하였다.

4.2 시스템 구조

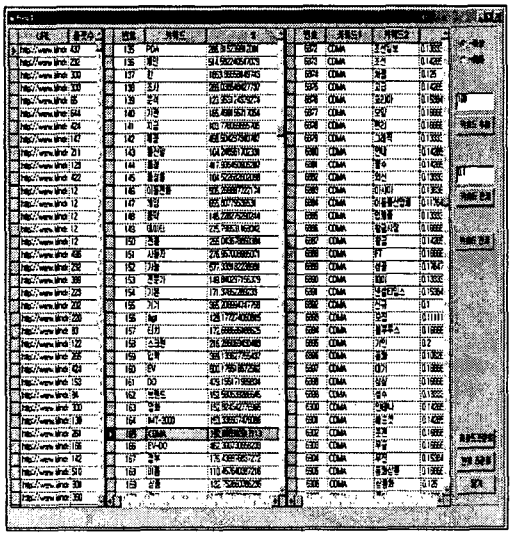
본 프로그램의 구조는 <그림 3>과 같다. 모든 작업은 단계별로 이루어지고 결과를 확인할 수 있도록 설계되었으며 앞 절에서 설명한 바와 같이 명사 추출 과정에는 HAM5.0이 실행되고 명사 추출과정과 개념망 형성과정의 결과는 DB

로 저장되어 프로그램의 실행이 종료된 후 결과를 분석할 수 있다.



<그림 3> 시스템 구조

<그림 4> 실행화면



<그림 5> 개념망 구축 결과화면

<그림 4>는 개념망을 구축하는 프로그램의 실행화면이며, <그림 5>는 주요개념을 추출하고 관계성이 높은 단어들을 결과로써 확인하는 화면이다.

4.3 개념망 구축 사례

우선 주요 개념을 추출하기 위한 임계치를 선정하기 위해서 주요개념 추출 알고리즘을 계산하여 임계치를 25, 50, 100 등으로 몇 가지 단계로 나누어서 비교해 보았다. 그 결과 임계치가 100인 경우에 가장 적절한 명사들이 주요 개념으로 선택되는 것을 확인할 수 있었다.

주요 개념 알고리즘을 적용하여 명사들을 추출한 결과를 분석해 보면 임계치 100 이상 일 때 가장 적절한 단어들이 추출되었다. 임계치가 너무 낮은 경우에는 주요 개념의 범위가 너무 커지고 임계치가 너무 높을 경우에는 문서 집합을 대표하는 단어도 있으나 일부분은 일반적인 단어도 포함되어 있어서 범위를 적절히 하는 것이 더 효과적이었다.

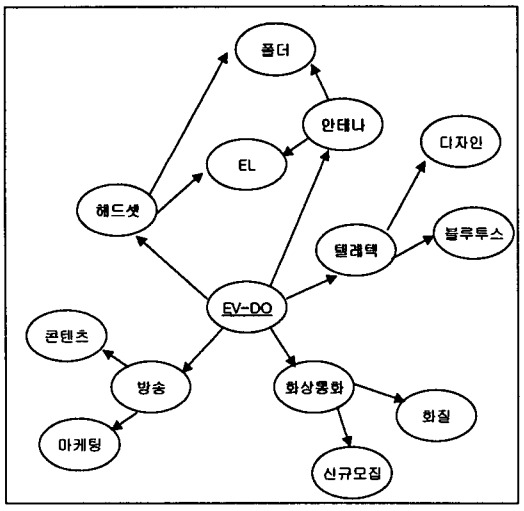
<표 1>"EV-DO"의 개념망 구축 결과

단어1	단어2	결과치	단어1	단어2	결과치
EV-DO	브랜드	0.5	EV-DO	데이터	0.150
	방송	0.368		영화	0.143
	텔레택	0.308		경쟁	0.138
	비디오	0.294		명의	0.133
	주문	0.267		LG전자	0.125
	EV-DO	0.250		멀티미디어	0.125
	DO	0.241		화상통	0.118
	EV	0.250		CDMA	0.111
	화상	0.200		4	0.108
	상태	0.182		콘텐츠	0.105
	전용	0.158		마케팅	0.105
	실시	0.154		제품	0.103

개념망 구축 시 개념망 구축 알고리즘을 적용한 결과 임계치가 0.5 이하의 값들이 나타나는 것을 볼 수 있었다. 임계치를 0.1로 설정하여 개

념망을 구축하였다. 결과 중에서 단어 "EV-DO"를 중심으로 정리한 결과가 <표 1>과 같으며, 이를 다시 네트워크 형태로표시하면 <그림 6>과 같다. 단어를 중복 허용하고 한 단어가 여러 단어와 관계성이 높은 경우도 있기 때문에 실제는 아주 복잡하게 얽혀 있는 망으로 구성되지만, <그림 6>은 "EV-DO"를 중심으로 관련 있는 단어를 먼저 작성하고 또 이 단어가 관계성이 높은 단어들을 뽑아서 간략하게 도식화한 것이다.

검색망을 실제 검색엔진에서 활용할 때는 트리 형태로 표현하여 사용자들에게 알아보기 쉽도록 하였다.



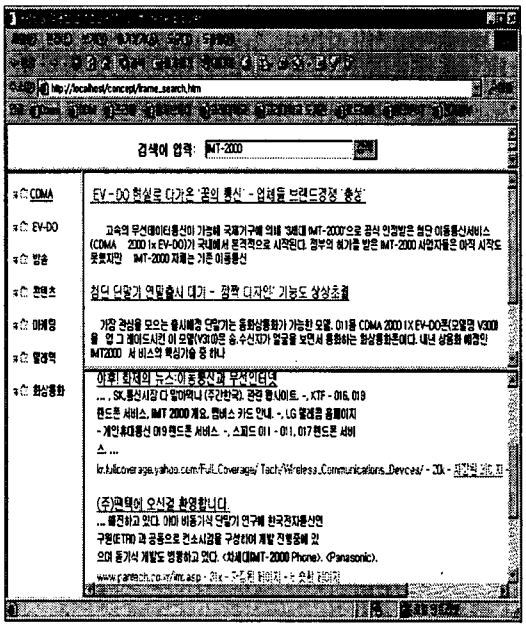
<그림 6> 개념망 구축 사례

5. 논의 및 결론

본 연구에서 제시한 개념망은 특정 도메인을 대표하는 문서 집합을 대상으로 이들을 대표할 수 있는 주요개념을 추출하며, 이들 주요 개념 사이의 관계성을 분석하여 중요한 관계성을 가지는 단어 사이에 링크를 설정하여 만든 단어들의 네트워크이다.

개념망은 개발 목적이 전문검색엔진의 프리프로세싱에 있다. 즉, 사용자가 검색을 원하는 단어를 입력하면, 본 시스템은 입력한 단어의 개념망을 보여준다. 사용자는 개념망을 조사함으로써 입력한 단어가 자신이 검색하고자 하는 목표 개념을 잘 반영하고 있는지를 알 수 있으며, 자신이 미처 생각지 못한 의미있는 개념을 발견하고 검색에 활용할 수 있다. 따라서 개념망을 조직이나 개인의 지식관리 시스템(Knowledge Management Systems)의 전문검색엔진에 추가하여 활용하면 조직외적 지식을 효과적으로 검색할 수 있을 것이다.

<그림 7>은 본 개념망을 전문검색엔진에 활용한 예이다. 그림은 사용자가 "IMT-2000"이라는 질의어를 입력하고 검색을 실행한 결과화면을 나타낸 것이다. 사용자가 검색어를 입력하면, 화면에서처럼, 좌측에 검색어의 개념망이 트리 형태로 나타나며, 우측에는 다양한 검색결과가 나타난다. 사용자는 개념망과 검색결과를 확인한 후, 질의어를 적절히 수정하여 검색을 다시 할 수 있다.



<그림 7> 개념망을 이용한 검색 실행화면

개념망은 시소러스나 워드넷과는 달리 매우 쉽고 빠르게 자동으로 구축된다. 따라서 개념망

은 다이나믹하게 변화하는 시사성 있는 웹 정보에 대한 검색에도 적합할 뿐 아니라, 쇼핑몰이나 EC 시스템에서도 효과적으로 활용될 수 있다.

본 연구의 개념망은 특정 도메인을 대표하는 사이트들이 가지고 있는 정보의 질이 우수하다는 가정 하에 이루어 졌다. 즉, 문서집합이 가지는 정보의 질이 좋으면 개념망의 결과들이 상당히 우수하고, 정보의 질이 낮으면 개념망에 제시되는 결과의 질이 떨어지는 것을 본 연구를 수행하면서 확인할 수 있었다. 하지만 특정 도메인에서 그들을 대표하는 수준 높은 사이트들은 어느 정도 구축되어 있기 때문에 이중 질 높은 사이트들을 선별해서 선택하게 되면 될 것이다.

본 연구는 웹 문서 중에서 컨텐츠만을 대상으로 하였다. 하지만, HTML 문서의 태그를 보면 이들 태그가 가지는 의미가 상당히 크다. <B>, <title>과 같은 태그들은 웹 문서들의 핵심 단어를 나타내는 경우가 많다. 따라서 태그 정보와 같은 웹 문서의 구조적인 정보를 활용하는 방안을 개발하여 적용한다면 단어의 출현 빈도에 의존하여 생기에 되는 한계점을 보완할 수 있을 것이다.

본 연구에서는 개념망 구축에 중점을 두어 이루어 졌다. 따라서 개념망에 제시된 단어들이 속하는 사이트들의 적합성에 따른 순위화에 대한 방안은 고려되어 있지 않다. 앞으로는 검색된 페이지의 순위화에 대한연구를 수행할 것이다. 또한, 개념망의 구축 후, 개념망의 재구축주기, 개념망의 적절한 임계치 설정 등 여러 가지 실험적인 연구를 계획하고 있다.

참 고 문 헌

[1] 김민수, 김태연, 노봉남, "국어사전을 이용한 한국어 명사에 대한 상위어 자동 추출 및 WordNet의 프로토타입 개발", *한국정보처리학회 논문지*, 제2권 제6호, 1995, pp. 847-856.

[2] 김태수, 이재윤, "WordNet과 시소러스", *언어정보의 탐구*, 1999, pp.232-269.

[3] 신진섭, 이창훈, "단어의 연관성을 이용한 문서의 자동분류", *한국정보처리학회 논문지*, 제6권 제9호, 1999, pp.2422-2430.

[4] 우선미, 유춘식, 김용성, "용어 연관성 분석을 이용한 사용자 위주의 문서순위결정 기법", *정보과학회논문지 : 소프트웨어 및 응용*, 제28권 제2호, 2001, pp.149-156.

[5] 이상준, 류근호, "게임 정보검색을 위한 자동색인 및 신조어 처리 시스템 구현", *한국정보처리학회 춘계학술발표논문집*, 제8권 제1호, 2001, pp.51-54.

[6] 이종인, 한광록 양승현,김영섬, "한국어 명사의 시소러스 구축을 위한 시스템 설계 및 구현", *한국정보처리학회 논문지*, 제6권 제2호, 1999, pp.347-356.

[7] 임형근, "색인어 연관성을 이용한 의료정보문서 분류에 관한 연구", *정보처리학회 논문지*, B 제8-B권 제5호, 2001, pp.469-476.

[8] 임해창, 윤보현, 강승식, "한국학 서지 정보와 전자텍스트를 위한 자동색인 및 검색 시스템 개발 연구", *한국어전산학 제2집*, 1996, pp.279-292.

[9] 주정은, 구상회, "효과적인 정보검색을 위한 개념망의 구축", *한국정보과학회 추계 논문집II*, 2002, pp.295-297.

[10] 정영미, "정보검색론", 구미무역(주)출판부, 1993.

[11] 최명목, 김민구, "정보검색에서 시소러스를 이용한 효율적이고 효과적인 질의 평가 방법", *한국퍼지 및 지능시스템학회 논문지*, 2000, Vol. 10 No. 6, 2, pp.605-615.

[12] Baeza-yates 외, *최신 정보검색론*, 홍릉과학출판사, 1999.

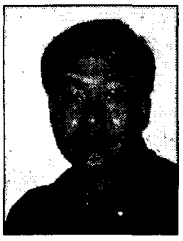
[13] http://nlp.kookmin.ac.kr/, 국민대학교자연언어 정보검색 연구실.

[14] http://green.skhu.ac.kr/~skhuir/, KTSET 문서.

■ 저 자 소 개

주 정 은

덕성여자대학교 경영학과 학사, 고려대학교 일반대학원 디지털경영학 석사. 현재 고려대학교 일반대학원 디지털경영학 박사과정 재학 중. 연구 분야는 인공지능, e비즈니스, 전자상거래 등을 포함.



구 상 회

고려대학교 경영학과 학사, 미국 남가주대학교(University of Southern California) 전산과학과 석사 및 박사. 현재 고려대학교 서창캠퍼스 경영정보학과 교수로 재직 중. 연구분야는 인공지능, e비즈니스, 전자상거래 등을 포함.

◈ 이 논문은 2003년 4월 6일 접수하여 2차 수정을 거쳐 2003년 6월 16일 게재확정되었습니다.