

전자정보 시스템의 효율적 메뉴검색

신 동욱*, 남세진*, 배정일*, 박상규**, 장명욱**, 최기선***
충남대학교 컴퓨터공학과*, 한국전자통신 연구소**, 한국과학기술원***

Effective menu retrieval for electronic information system

Dongwook Shin, Sejin Nam, Jeongil Bae, Sangkyu Park, Myeongwook Jang,
and Keysun Choi

요약

Hitel 과 같은 전자정보 시스템은 사용자가 원하는 정보를 체계적으로 얻을 수 있도록 하기 위하여 메뉴들을 적당히 계층적으로 구성하여 제공하고 있다. 그러나, 보통 이 메뉴들의 계층이 정확한 분류법에 기초하여 만들어 지지 않았을 뿐 아니라 그 양도 엄청나게 방대하여, 이 메뉴 계층을 이용하여 사용자가 원하는 정보를 얻기가 쉽지 않다. 실험적으로 보통 Hitel을 자주 이용하는 사람들도 자신이 주로 이용하는 메뉴들의 구성만 이해하고 있을뿐, 사용하지 않는 부분의 메뉴들의 구성은 잘 알지 못하는 것이 일반적이었다. 따라서 Hitel을 자주 이용하는 사용자도 자신이 이용해 보지 않은 정보를 얻기 쉽지 않으며, 더더욱 초보자에게는 이 메뉴계층을 이용하여 원하는 정보를 얻기가 쉽지않은 실정이다.

본 연구에서는 정보검색 기술을 이용하여 Hitel과 같은 전자정보 시스템에서 사용자가 쉽게 자신이 원하는 정보를 얻을 수 있는 보조 시스템을 개발하고자 한다. 본 시스템은 사용자가 메뉴계층을 이용하기 전에 간략한 자연어로 입력을 주면, 여기에 적합한 메뉴나 실제 정보를 검색해 낸다. 따라서 사용자는 이 메뉴정보를 이용하여 메뉴계층을 쉽게 따라갈 수 있을뿐아니라, 경우에 따라서는 원하는 실제 정보를 검색하기 때문에 메뉴계층을 탐색할 필요가 없다.

본 연구에서는 자연어 입력을 최장 일치 방법으로 의미있는 명사들을 추출하여 불리안 질의어로 만든후, 명사들 사이의 관계가 표현된 시소러스를 이용하여 이 질의어를 확장시킨다. 다음에 이 질의어들을 메뉴들과 부분/정확부합을 통하여 관련된 메뉴들을 찾아낸후, 이들의 계층과계를 고려하여 최종 메뉴들을 검색한다. 본 시스템은 현재 C언어로 만들어져 구동중이며, 정확한 실험은 아직 하지 않았지만 높은 검색율을 보이고 있다.

1. 서론

최근에 컴퓨터가 점점 더 많이 가정에 보급되고 있고, 통신기술이 발달함에 따라 점점 더많은 정보서비스 시스템들이 통신망을 통하여 서비스를 하고 있다. 국내에서도 Hitel 및 천리안등 여러 시스템들이 정보를 제공해 오고 있다. 이들 전자 정보 시스템들은

제공하는 서비스가 매우 많고 다양하여 서비스를 제공하는 측에서 메뉴들을 일정한 계층구조로 제공하여 사용자가 이들 메뉴계층을 적당히 선택하여 원하는 정보를 검색하도록 하고 있다. 그러나, 보통 이 메뉴들의 계층이 정확한 분류법에 기초하여 만들어 지지 않았을 뿐 아니라 그 양도 엄청나게 방대하여, 이 메뉴 계층을 이용하여 사

* 본 연구는 한국전자통신 연구소로 부터 용역과제로 수행되었습.

용자가 원하는 정보를 얻기가 쉽지 않다. 실험적으로 보통 Hitel을 자주 이용하는 사람들도 자신이 주로 이용하는 메뉴들의 구성만 이해하고 있을뿐, 자신이 사용하지 않는 부분의 메뉴들의 구성은 잘 알지 못하는 것이 일반적이었다. 따라서 Hitel을 자주 이용하는 사용자도 자신이 이용해 보지 않은 정보를 얻기 쉽지 않으며, 더더욱 초보자에게는 이 메뉴계층을 이용하여 자신이 원하는 정보를 얻기가 쉽지않은 실정이다.

본 연구에서는 정보검색 기술을 이용하여 Hitel과 같은 전자정보 시스템에서 사용자가 쉽게 자신이 원하는 메뉴들을 얻을 수 있는 보조 시스템을 개발하고자 한다. 일반적인 정보검색기술 [1,3]은 대상 문서들을 적당한 색인기술을 이용하여 색인하고, 사용자 질의어가 입력되면 이미 색인된 정보를 이용하여 적절한 문서를 검색해 낸다. 어떤 경우에는 문서들을 색인하지 않고, 사용자 질의어가 들어오면 대상문서들의 내용을 직접 보고 적합한 문서들을 찾아 내기도 한다. 그러나 본 연구에서는 일반 문서대신에 메뉴들을 대상으로 하기때문에 일반 정보검색 시스템과는 다음 두가지 점에서 다르다. 우선 메뉴들의 내용이 일반 문서들과는 달리 매우 짧다는 것과 둘째 많은 정보들이 메뉴내용 이외에 메뉴계층내에서 이 메뉴의 위치로부터 얻어진다는 것이다. 따라서 메뉴내용만을 가지고 색인할 경우에는 많은 정보들을 잃어버리게 되며, 메뉴가 위치한 계층정보를 모두이용하여 색인할 경우에는 색인화일이 실제 메뉴보다 엄청나게 커진다는 단점이 있다.

본 연구에서는 메뉴 시스템의 특징을 충분히 활용하기 위하여, 위에서 언급한 두가지 색인기술을 적당히 혼합하여 메뉴들을 색인하고, 자연어로 입력된 질의어를 명사들로만 구성된 불리언 질의어로 바꾸어 검색한다. 색인 방법은 모든 메뉴에 대하여 각 메뉴를 대표할 수 있는 색인어를 하나 선택하여 색인하며, 고유명사로 구성된 메뉴에 대

해서는 메뉴내용과 이 메뉴가 있는 위치정보 (계층구조에서의 길(path) 정보)쌍을 Dbase에 저장한다. 검색 방법은 우선 자연어로 입력된 질의어를 토큰들로 나눈 후 Dbase를 참조하여 검색한다. 그다음 자연어 질의어를 변환하여 불리언 질의어를 만든 후 이 질의어로부터 미리 구성된 시소러스를 이용하여 질의어 확장을 하고, 이를 메뉴들의 색인어와 비교하여 부합되는 메뉴들을 검색해 낸다. 이렇게 두가지 방법을 통하여 검색된 메뉴들을 대상으로 적당한 우선순위및 메뉴계층관계를 이용하여 사용자에게 주어져야 하는 최종 메뉴들을 선택한다. 본 시스템은 현재 C언어로 만들어져 구동중이며, 정확한 실험은 아직 하지 않았지만 매우 높은 검색율을 보이고 있다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 본 시스템의 기본적인 구조를 설명한 후, 3절에서는 메뉴들을 보고 색인어를 어떻게 결정하는지에 대하여 설명한다. 4절에서는 시소러스의 구성과 검색할때 시소러스를 어떻게 이용하는지를 설명하고, 5절에서는 메뉴데이터베이스에서 메뉴를 추출하고 부분/정확부합을 통하여 부분및 정확부합의 판별과정을 보인다. 6장에서는 Dbase의 구성방식에 대하여 설명하고 7장에서는 검색된 메뉴들이 메뉴계층구조에서 산재되어 있는 경우 이들을 어떻게 가지치기하여 최종메뉴를 선택할지를 보인다. 마지막으로 8절에서는 결론과 앞으로의 연구방향에 대하여 검토한다.

2. 시스템 구성

본 연구는 계층메뉴구조를 대상으로 먼저 용어사전및 시소러스를 개발하고, 메뉴들을 효율적으로 색인한후 여기에 기초하여 질의어에 적합한 메뉴들을 검색하는 과정을 포함하고 있다. 우선 용어사전은 본 연구가 명사들을 대상으로 하고 있기 때문에 명사들만 포함하고 있으며 시소러스는 메뉴시스

템에서 나오는 명사들을 대상으로 동의어, 광의어, 협의어등을 표현하고 있다. 둘째로 색인은 통제어 (controlled vocabulary)들을 이용하여 시소러스에서 용어들 사이의 일련의 관계를 보고 진행하며 색인화일을 효율적으로 만들기 위하여 B-tree를 이용하였다. 검색과정은 위에서 언급한대로 우선 Dbase에서 일차 검색하고, 그 다음에는 질의어를 시소러스를 이용하여 확장한후 이 확장된 질의어를 색인어와 비교하여 메뉴들을 검색한다. 이 검색과정을 도식적으로 표시하면 아래와 같다.

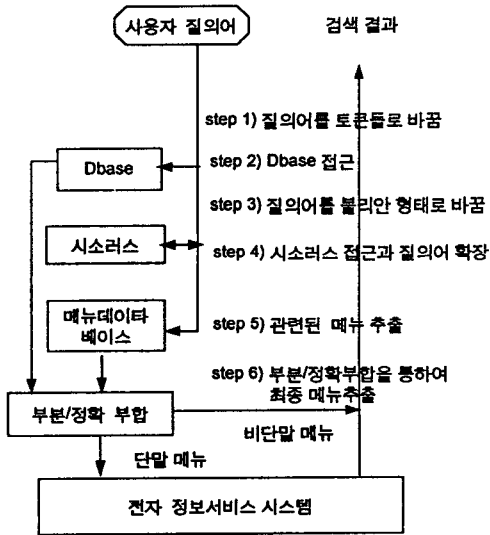


그림 1. 검색 시스템 개요

우선 자연어로 구성된 사용자 질의어가 입력되면 이를 토큰 (빈칸없이 연속된 글자모임)들로 잘라내고 이들중 일부가 메뉴와 일치하는지를 알아내기 위하여 Dbase를 검색한다. 이때 일치되는 메뉴가 있을 경우에는 이들을 추출하여 저장하여 둔다. 두번째로 토큰화된 질의어를 불용어처리를 한후, 명사들을 추출하여 명사들이 AND 연결어로만 연결된 불리안 질의어로 바꾼다. 세번째로 질의어의 각 명사들을 시소러스 용어들과

비교하여 일치되는 용어가 있으면 이 용어의 동의어와 협의어들을 전부 찾고 난후, 원래 명사를 이들 동의어와 협의어들을 OR로 연결한 절로 바꾼다. 네번째로는 이렇게 시소러스를 참조하여 확장된 질의어를 가지고 메뉴데이터베이스를 검색하여 색인어가 일치하는 메뉴들을 찾아낸다. 이때 찾아낸 메뉴들을 본연구에서 고안한 부분/정확부합을 이용하여 다시 필터링하고 이들을 Dbase에서 검색한 메뉴들과 비교 분석하여 최종 메뉴들을 찾아낸다.

3. 메뉴 색인

앞에서도 서술하였듯이, 메뉴항목은 일반 문서와는 달리 내용이 매우 짧을 뿐아니라 정보의 많은 부분이 메뉴항목에서 얻어지지 않고 메뉴의 위치에 의하여 얻어지므로 보통의 정보검색방법에서의 같은 색인방식을 사용하기는 곤란하다. 보통의 색인 방식은 문서들로부터 색인가능한 용어들을 추출하여 이들을 색인어로 사용하거나 이들을 가공하여 색인어로 사용하는데 [5], 메뉴를 대상으로 이러한 방법을 사용하면 색인어가 거의 없을뿐 아니라 사용자가 메뉴 제충을 염두에 두고 질의하였을 경우 적합한 메뉴들을 찾아내기가 매우 힘들기 때문이다. 따라서 본연구에서는 각 메뉴마다 가장 대표되는 색인어를 하나만 선택하여 색인하고, 메뉴제충에서 내포적으로 얻을 수 있는 정보는 시소러스에서 동의어와 협의어 관계를 이용하여 유추 가능하도록 하였다. 메뉴항목에서 색인어를 택하는 방법은 메뉴를 대표할 수 있는 용어 하나를 선택하는데, 선택가능한 용어가 여러개 있을 경우에는 시소러스를 참조하여 이들중 가장 협의어를 선택한다. 물론 이과정은 자동적으로 이루어지는 것이 아니라 사람이 해야되는 과정이며 어떤 용어를 색인어로 선택할지를 결정하기 힘든 경우도 종종 발생한다. 다음은 '컴퓨터'에 대하여 시소러스가 아래와 같이 구성되었다고 하자.

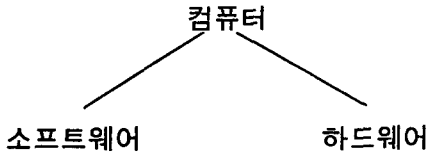


그림 2. 컴퓨터에 관한 시소러스의 구성예

이때 '컴퓨터'란 메뉴에 대해서는 대표할 수 있는 용어가 '컴퓨터' 하나이므로 당연히 '컴퓨터'로 색인하고, 'PC랑 나랑...'이란 메뉴에 대해서도 '소프트웨어' 및 '하드웨어' 어느것도 대표하지 못하므로 역시 '컴퓨터'로 색인한다. 그런데 '소프트웨어'라는 메뉴에 대해서는 대표할 수 있는 용어가 '컴퓨터' 및 '소프트웨어' 두개가 가능하므로 이들 중 협의어인 '소프트웨어'를 색인어로 선택한다. 여기서 대표한다는 의미는 메뉴의 의미를 포괄한다는 것을 나타내는 것으로 메뉴의 의미중 일부를 나타내는 용어는 메뉴를 대표한다고 할 수 없다.

4. 시소러스구성 및 질의어 확장

시소러스는 용어들과 이들 사이에 동의어, 광의어/협의어, 관계어등의 관계를 표현해 놓은 것으로서 [1], 정보검색 분야에서는 색인과정 및 검색과정에서 사용된다. 한글에 관한 시소러스는 JICST 시소러스 [2]의 번역판 및 한국과학기술원에서 한국통신 장기기초 과제수행중 만든 시소러스 [4]등이 있으나 이들은 너무 일반적이고 도서관의 분류표와 유사하여 메뉴계층에서 사용되는 특별한 용어들이나 특정한 분야에서 사용하기가 힘들다. 본 연구에서는 이들을 사용하는 대신에 Hitel에서 자주 사용되는 용어를 대상으로 동의어, 광의어/협의어들의 관계를 조사하여 시소러스를 만들었다. 이 시소러스는 와벽한 것은 아니며 계속 확장할 필요가 있는데, 현재는 1000여개의 용어에 대하여

이들 사이에 관계들을 표현하고 있다. 그림 3은 이 시소러스에 표현된 관계를 표시하고 있다. 여기서 '='로 표시된 관계는 동의어 관계를 나타내며, '-'로 표시된 관계는 광의어/협의어 관계를 나타낸다.

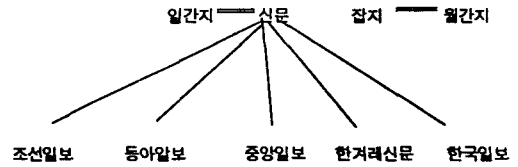


그림 3. 시소러스의 구조

검색과정에서 시소러스를 이용하는 방법은 다음과 같다. 2절에서 설명하였듯이 검색과정에서 시소러스를 참조하는 때는 사용자 질의어를 명사들로 바꾸고나서 인데, 예를 들어 사용자 질의어가 '신문 사설에 관한 메뉴'라면 이 질의어는 불용어처리와 명사추출 단계를 거쳐, '신문 AND 사설'이란 형태로 바뀌고 나서 시소러스를 참조하게 된다. 이 불리언 질의어를 가지고 시소러스를 참조하면, '신문'이 시소러스에 존재하므로 이 용어의 동의어 및 협의어를 전부 찾아 ORing을 하면 '(신문 OR 일간지 OR 조선일보 OR 동아일보 OR 중앙일보 OR 한겨레신문 OR 한국일보) AND 사설'로 바뀌게 된다. 만약 이 질의어에서 '사설'과 같이 명사가 시소러스에 없으면 이 용어는 더이상 확장되지 않고 그대로 질의어에 남게된다. 이렇게 확장된 질의어는 Conjunctive normal form으로 바뀌어 검색과정에 사용되는데, 위의 질의어는 '(신문 AND 사설) OR (일간지 AND 사설) OR (조선일보 AND 사설) OR (동아일보 AND 사설) OR (중앙일보 AND 사설) OR (한겨레신문 AND 사설) OR (한국일보 AND 사설)'로 바뀐다.

5. 메뉴추출 및 부분/정확부합

4절에서 설명한 바와 같이 질의어가 확장되

면 질의어를 가지고 메뉴데이터베이스를 검색하게 되는데, 질의어에 나타나는 명사들과 메뉴의 색인어가 일치하면 이 메뉴는 일차 검색된다. 예를 들어 4절에서 확장된 질의어의 최종형태를 가지고 설명하면, 우선 '신문'이 색인어로 되어 있는 메뉴들을 검색한다. 본 연구에서는 이렇게 검색된 메뉴들을 그대로 선택하는 것이 아니라 이들에 대하여 부분/정확부합과정을 거치는데, 부분/정확부합이란 '신문'에 의하여 검색된 메뉴들이 '사설'이라는 용어를 가지고 있는지 검색하는 것이다. 이때 메뉴가 '사설'이라는 용어를 가지고 있는지는 색인어를 보고 결정하는 것이 아니라 이 메뉴의 내용을 직접 보고 여기에 '사설'이 나타나 있으면, 이 메뉴를 부분/정확부합되었다고 판단한다. 즉 부분/정확부합이라는 과정은 우선 질의어의 명사를 가지고 추출한 메뉴를 상대로 이 명사와 AND로 연결된 나머지 명사들이 메뉴의 내용에 직접 나오는지 확인하는 과정이다. 여기서 한가지 중요한 것은 AND로 연결된 용어들을 모두 포함한 메뉴는 정확부합 되었다고 하며, 이중 일부만을 포함한 용어는 부분부합되었다고 한다.

메뉴데이터베이스에서 검색한 메뉴들을 대상으로 AND로 연결된 나머지 명사들을 메뉴의 내용과 직접 비교하는 이유는 메뉴가 하나의 용어로만 색인되어 있어 나머지 용어들과 관련이 있는지는 메뉴의 내용을 직접보아야만 알 수 있기 때문이다. 메뉴데이터베이스에서 검색된 메뉴들이 질의어와 정확부합되었는지 혹은 부분부합되었는지의 정확한 정의는 다음과 같다.

정의 1) 질의어의 한 부분이 '명사1 AND 명사2 AND ... AND 명사N'이라고 할때 만약 명사M ($1 \leq M \leq N$)을 이용하여 메뉴들을 추출하였다고 가정하자. 이때 명사M을 제외하고 남은 나머지 명사들중 P개가 어떤 메뉴에 나타나 있다고 하면 다음 조건에 따라 이 메뉴는 정확부합 또는 부분부합되었다고 한다.

- (1) $N >= 3$ 일때, $P = N-1$ 이면 정확부합 되었다고 하고, $1 \leq P < N-1$ 이면 부분부합되었다고 한다.
- (2) $N \leq 2$ 일때, 부분부합되는 경우는 없고, $P = N-1$ 일 경우에만 정확부합 되었다고 한다.

이 정의에서와 같이 확장된 질의어에 나오는 모든 명사들을 가지고 메뉴데이터베이스의 메뉴들을 일차 검색한후, 검색된 메뉴들을 대상으로 부분부합되었는지, 정확부합되었는지를 가려내어 부분/정확부합된 메뉴들만을 선택하고 이들을 다음에 설명되는 가지치기 과정으로 넘기게 된다.

6. Dbase 구성

본 연구에서는 메뉴데이터베이스 이외에 Dbase라는 데이터베이스를 이용하는데 이 데이터베이스는 주로 고유명사를 포함하는 메뉴들에 대하여 그 위치정보를 포함하고 있는 데이터베이스이다. 앞절들에서 설명한 바와 같이 검색과정에서 고유명사들을 포함하는 메뉴들은 검색되지않는 경우가 많은데, 이는 고유명사들을 포함하는 메뉴는 다른 용어로 색인되는 것이 보통이며 또 고유명사들은 보통 시소러스에 포함되지 않기 때문에 사용자가 어떤 질의어를 주더라도 검색하기 힘들기 때문이다. 예를들어 '날씨'의 아래 항목중 '충청'이란 메뉴는 '날씨'로 색인되어 있어 '일기예보'라는 질의어에는 검색되지만 '충청도에 관한 정보'라고하면 검색되지 않는다. 따라서 이러한 부분을 검색하기 위하여 '날씨'아래의 '충청'이란 메뉴를 Dbase에 저장하고 사용자 질의어를 토큰화한후 이 토큰들을 가지고 Dbase를 검색하면 고유명사들이 있는 메뉴들을 검색할 수 있다. 본 연구에서는 Hitel메뉴를 대상으로 Dbase를 구성하였는데 전체메뉴의 30-40%정도가 Dbase에 저장되어 있다.

7. 가지치기

Dbase로부터 검색된 메뉴들과 메뉴데이터 베이스에서 부분/정확부합을 통하여 검색된 메뉴들에서 가지치기를 통하여 최종적으로 사용자 출력으로 검색되는 메뉴들이 결정된다. 본 논문에서 가지치기라는 용어는 '메뉴계층구조에서 동일한 길(path)에 존재하는 메뉴들중에서 하나를 선택'하거나 '여러 나무 tree)중에서 일부만을 선택'하는 것을 의미한다. 예를 들어 사용자 질의어가 '증권'일 경우 부분/정확부합을 통하여 선택된 메뉴들은 '증권'을 포함하여 '상한가', '하한가' 등 '증권'의 아래 메뉴들이 상당수 검색되는데, 이들은 모두 '증권'을 뿌리(root)로 하는 나무에 속하며 '증권'이 사용자 질의어에 정확부합하므로 가지치기를 통하여 이들 중 '증권'만을 최종적으로 선택한다. 또한 사용자 질의어가 '충청도 날씨'일 경우 '날씨' 메뉴 아래 '충청도'가 정확부합되어 선택되고 '충청도'에 관한 다른정보들은 Dbase에서 선택되는데 이 경우에는 가지치기를 통하여 정확부합된 메뉴를 선택하고 Dbase에서 선택된 다른 가지들은 버린다. 그이유는 부분/정확부합된 메뉴가 사용자 질의어에 좀더 많이 유사하며, Dbase에서 선택된 메뉴들은 사용자 질의어중 고유명사 하나에만 부합되면 선택되기 때문에 사용자 질의어에 덜 적합하다고 보기 때문이다. 즉 가지치기란 6절까지 설명된 방법들을 통하여 추출된 메뉴들중 사용자가 정말 요구하는 메뉴들을 최종적으로 선택하는 과정이라 할 수 있다.

가지치기는 두가지 과정으로 구성되는데 메뉴계층구조에서 동일한 길에 존재하는 메뉴들중에서 하나를 선택'하는 과정과 '부분/정확부합을 통하여 선택된 메뉴와 Dbase에서 선택된 메뉴들중에서 일부만을 선택'하는 과정으로 구성된다. 첫번째 과정은 메뉴 두개이상인 동일한 길에 존재할 경우는 둘다 사용자에게 보여줄 필요없이 둘중에 하나만 보여주면 되는데 다음과 같은 원칙에

의하여 보여줄 메뉴를 선택한다.

- 1-1) 동일한 길에서 두개 이상의 메뉴가 전부 부분부합 되었을 경우, 이중 가장 아래메뉴를 선택한다.
- 1-2) 동일한 길에서 위의 메뉴는 정확부합되고 아래 메뉴는 부분부합되었을 경우, 위의 메뉴를 선택한다.
- 1-3) 동일한 길에서 위의 메뉴나 아래 메뉴가 전부 정확부합되었을 경우, 아래 메뉴를 선택한다.

이 원칙의 기본 아이디어는 동일한 길에 있는 메뉴들중에는 가능하면 단말 메뉴에 가까운 메뉴를 선택하며, 만약 어떤 메뉴가 정확부합되었을 경우에는 다른 메뉴들은 삭제하고 정확부합된 메뉴를 선택한다는 것이다. 이는 사용자는 가능하면 단말메뉴를 얻기를 원하며, 정확부합된 경우는 사용자가 그 메뉴가 메뉴계층에서 어디에 있는지를 알기 원할때가 많다는 데에 근거한다.

두번째 과정은 부분/정확부합을 통과한 메뉴들과 Dbase에서 얻어진 메뉴들 중에서 어떤 메뉴들을 최종적으로 선택할 것인가 하는 과정으로서, 기본적으로 부분/정확부합을 통하여 얻어진 메뉴들이 우선순위가 높으며 아래의 원칙에 따라 선택된다.

- 2-1) Dbase에서 얻어진 메뉴들중에서 부분/정확부합을 통하여 얻어진 항목과 공통되는 항목이 있는 경우는 공통되는 항목을 최종 메뉴로 선택한다.
- 2-2) Dbase에서 얻어진 메뉴들중에서 부분/정확부합을 통하여 얻어진 항목과 공통되는 항목이 없는 경우는 부분/정확부합을 통하여 얻어진 항목만을 최종메뉴로 선택한다.
- 2-3) Dbase에서 얻어진 메뉴들만 있고 부분/정확부합을 통하여 얻어진 항목이 없는 경우는 Dbase에서 얻어진 항목을 최종메뉴로 선택한다.

- 2-4) Dbase에서 얻어진 메뉴들은 없고 부분/정확부합을 통하여 얻어진 항목만 있을 경우는 부분/정확부합을 통하여 얻어진 항목을 최종 메뉴로 선택한다.

본 연구에서는 제안된 방법의 성능을 정확히 평가하지 않았지만 여러 사용자를 대상으로 평가한 결과 높은 검색률을 보이고 있음을 확인하였다.

8. 결론

본 연구에서는 정보검색 기술을 이용하여 Hitel과 같은 전자정보 시스템에서 사용자가 쉽게 자신이 원하는 정보를 얻을 수 있도록 하는 메뉴검색시스템을 개발하였다. 현재까지의 대부분의 전자정보시스템은 사용자에게 메뉴들을 계층적으로 제공하고, 사용자로 하여금 이 메뉴계층을 적절히 선택하여 자신이 원하는 정보를 얻도록 하고 있다. 그러나 이 메뉴계층은 정확한 분류법에 근거하여 만들어지지 않은 경우가 대부분이어서 사용자가 메뉴계층을 따라 자신이 원하는 정보를 얻기가 쉽지 않다. 본 연구에서는 이러한 점에 착안하여 사용자가 간략한 자연어로 입력을 주면 여기에 적합한 메뉴들과 이들의 위치정보를 찾아냄으로써 사용자가 쉽고 빠르게 자신이 원하는 정보를 얻을 수 있도록 하는 검색시스템을 개발하였다.

본 시스템은 사용자가 자연어로 입력을 주면, 이들중에서 명사만을 골라내어 불리언 질의어로 만든후 시소러스를 이용하여 질의어를 확장한다. 그다음 확장된 질의어를 가지고 메뉴데이터베이스를 접근하여 색인어가 일치하는 메뉴들을 추출하고 이들을 부분/정확부합 알고리즘을 이용하여 여기에 통과되는 메뉴들을 선택한다. 이러한 방식으로는 고유명사로 구성된 질의어에 적합한 메뉴들을 잘 검색하지 못하는데 이러한 경우를 보강하기 위하여 고유명사가 포함된

메뉴들로 구성된 Dbase를 첨가하였다. 검색 알고리즘에서는 최종적으로 부분/정확부합을 거쳐 얻어진 메뉴들과 Dbase에서 얻어진 메뉴들을 우선순위를 비교하여 사용자에게 보여주는 메뉴들을 선택한다. 본 시스템은 현재 C언어로 만들어져 구동중이며, 정확한 실험은 아직 하지 않았지만 높은 검색율을 보이고 있다. 향후 연구과제로는 자연어 질의어를 처리할 수 있는 방법과 시소러스를 보완하고 부분/정확부합과 가지치기 알고리즘을 좀더 정교하게 만드는 것들을 들수 있다.

참고문헌

1. W.B. Frakes, Information Retrieval, Data structure and algorithms, Prentice Hall, 1992.
2. JICST Thesaurus, The Japan Information Center of Science and Technology, 1987.
3. G. Salton, Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison Wesley Publishing Company, 1989.
4. 시소러스 종류, 문화, 한국과학 기술원, 지능형 정보검색 연구 최종연구보고서, 1993.
5. 임형국, 정상철, 신동욱, 김형근, 최기선, "시소러스를 기반으로 하는 자동색인 시스템에 관한 연구," 94 춘계 학술발표 논문집, 한국정보과학회, 1994, pp. 173-176.