

한국어 대화체 문장 분석을 이용한 메타 정보검색

Meta Information Retrieval

using Sentence Analysis of Korean Dialogue Style

박 인 철(Park In Cheol)*

요약

오늘날 통신의 발전에 따라 인터넷상에 존재하는 정보의 양이 많아지고, 필요한 정보를 효율적으로 찾아내는 정보 검색 시스템의 중요성이 크게 대두되고 있다. 대부분의 정보 검색 시스템에서는 단순한 키워드나 키워드를 이용한 불리언 질의어를 바탕으로 필요한 문서를 검색해 내고 있다. 그러나, 키워드를 이용한 정보 검색은 사용자의 편의성 및 주어진 질의어에 대한 이해의 정확성 측면에서 우리가 일상생활에서 사용하는 대화체 문장을 이용한 질의어에 비해 많은 어려움을 가지고 있다. 본 논문에서는 이러한 문제점을 해결하기 위해 한국어 대화체 문장의 정보 검색을 위한 메타질의어처리시스템을 설계하고 구현한다. 본 논문에서 제안한 한국어 대화체 문장 분석을 이용한 정보 검색은 주어진 질의어에 대해 형태소 분석과 구문 분석 및 시소러스를 이용한 질의어의 확장을 통해 사용자가 원하는 질의어를 포함하는 새로운 질의어를 형성해 내며, 질의어에 포함된 중의성도 부분적으로 해결할 수 있었다.

ABSTRACT

Today, documents existing on internet by the development of communication network increase in number. And it is required the information retrieval system that can efficiently acquire the necessary information. Most information retrieval systems retrieve documents using a simple keyword or a boolean query of keywords. But, the method is not fit for novice users to use and has many difficulties than user's dialogue query from the viewpoint of convenience and precise understanding for query. So, this paper has an aim to suggest the method that will cope with above problems and to design and implement a meta query processing system for information retrieval using Korean dialogue sentences. The system implemented in this paper can generates a new boolean query for a given Korean dialogue sentence and resolve lexical ambiguities through morphological analysis, syntactic analysis and extension of query using thesaurus.

*정회원 호원대학교 컴퓨터학부 부교수

논문접수 : 2003. 10. 2.

심사완료 : 2003. 10.10.

*본 논문은 교내 학술연구조성비에 의해 이루어 졌음.

1. 서론

통신망의 발달로 사용자는 검색 엔진을 통해 인터넷상에서 필요한 정보를 빠르게 얻을 수 있게 되었으며, 대량의 정보를 다양한 방법에 의해 습득하는 현대인들은 정보를 효율적으로 찾아내는 방법에 대해 많은 관심을 가지고 있다. 현재 인터넷 환경에서 이러한 효율적인 정보의 탐색을 도와주는 다양한 형태의 많은 정보 검색 시스템이 존재하고 있다. 정보 검색 시스템이란 시스템 이용자가 필요로 하는 정보를 수집하여 내용을 분석한 후, 찾기 쉬운 형태로 조직하여 두었다가 정보에 대한 요구가 발생할 때 적절한 정보를 검색하여 제공하는 시스템이다[11]. 대부분의 정보 검색 시스템은 인터넷상에 존재하는 많은 문서를 수집하여 적절한 방법에 의해 문서들을 구분하기 위한 색인어를 추출하고, 저장한다. 그리고, 정보 검색 시스템의 사용자는 필요한 문서를 찾기 위해 적절한 단어를 이용하여 정보 검색 시스템에 질의를 하고, 정보 검색 시스템은 색인어를 기반으로 사용자의 질의어를 처리하여 해당하는 문서를 보여주게 된다. 이러한 정보 검색은 일반 사용자가 사용하기 쉽도록 자연어 질의어를 수용하고, 주제어를 파악하여 불필요한 문서를 가능한 한 배제하여 원하는 정보를 많은 시간과 노력을 요구하지 않으면서 획득할 수 있도록 해야한다.

일반적으로, 검색을 하기 위한 사용자의 질의어는 질의어 방법에 따라 키워드를 이용한 질의어와 대화체 문장을 이용한 질의어로 분류할 수 있다[3][7]. 키워드를 이용한 질의어에서 질의어가 복잡한 내용을 포함하는 경우, 사용자는 질의어의 구조에 대한 이해를 바탕으로 질의어해야 한다. 그러나, 일반 사용자는 자신이 원하는 정보를 불리언(boolean) 형태의 질의어로 표현하는데 익숙하지 않다. 불리언 질의어(boolean query)란 검색어와 검색어 사이를 논리 연산자인 AND(&), OR(|), NOT(~)과 같은 논리적인 관계로 표현하는 것을 말한다

[3]. 반면에, 일상 생활에서 사용하는 대화체 문장을 이용한 질의어의 경우, 질의어의 복잡한 구조에 상관없이 일상 대화처럼 질의할 수 있기 때문에 사용자의 편의성을 높일 수 있으며, 대화체 문장 분석 처리의 과정에서 다양한 배경 지식을 이용하여 좀 더 정확한 질의어로 변환할 수 있어 사용자의 질의어 요구를 충실히 반영할 수 있다.

예를 들어, 다음은 “컴퓨터과학”과 관련된 질의어 형태이다[13]. 여기서, Q1과 Q2는 사용자가 불리언 질의어 형태로 질의하는 경우이며, NQ1과 NQ2는 대화체 문장 형태로 질의하는 것으로 질의어의 자연스러운 측면에서 보면 이들의 차이점을 명확하게 알 수 있다.

NQ1: 휴대용 컴퓨터나 PDA에서 사용되는 반도체나 또는 통신용 소프트웨어에 대해 알려주세요.

Q1: 반도체&휴대용 컴퓨터통신용 소프트웨어|반도체&PDA

NQ2: UNIX와 LINUX에 대해 다루어진 문서 중 MS를 제외한 문서는

Q2: UNIX&LINUX~MS

이와 같이, 대부분의 정보 검색 시스템은 색인기를 통해 구축된 색인어 데이터베이스를 기반으로 불리언 질의어에 기반한 키워드 검색만을 허용하거나, 대화체 문장 질의어를 허용하더라도 질의어에 나타나는 어휘를 단순히 패턴 매칭하거나, 형태소 분석을 통해 명사 또는 동사를 추출한 후, 불리언 질의어로 변환하고 있다. 이는 사용자의 자연어 질의어를 충분히 반영하여 검색하는 방법이 아니기 때문에 대부분의 정보 검색 시스템에서의 검색 결과의 질은 사용자의 요구에 만족할 만한 수준은 아니다.

이에 본 논문에서는 사용자가 한국어 대화체 문장의 질의어를 통해 원하는 최적의 정보를 검색할 수 있는 메타질의어처리시스템을

설계하고 구현한다. 이를 위해서 한국어 언어 처리 기술을 기반으로 한국어 대화체 문장 형태의 질의어에 대해 형태소 분석과 구문 분석 과정을 거친 후, 시소러스를 이용한 질의어 확장과 제약을 통해 사용자의 요구를 충분히 질의어에 반영하여 사용자의 질의어에 대한 최적의 불리언 질의어를 생성하는 방법을 보인다.

2. 한국어 대화체 문장 질의어의 특성

한국어 대화체 문장 형태의 질의어는 요구되는 대상을 구체화할 수 있는 관형구나 관형절과 같은 부가어의 수식을 받는 명사들과 이들 사이의 관계를 나타내는 용언으로 주로 구성되며, 의문문의 형태를 가진다. 예를 들어, 이러한 질의어는 대부분 “A는 무엇인가?”, “A는 B한가?” 등의 의문문이거나 “A를 C하라” 등과 같은 명령문으로 구성된다[7]. 또한, 많은 경우 용언이 생략되며, 이때 용언은 대부분 “알려줘”, “보여줘”, “말해줘”, “무엇인가?” 등이 생략된 것으로 생각할 수 있다. 이를 자연어의 문장 유형에 따라 분류하면 아래와 같다 [9].

- ① 보편적 객체에 관한 것
 - 파충류의 생활에 대하여
 - 8월의 꽃은
- ② 구체적 객체에 대한 것
 - 세계에서 가장 높은 빌딩은
 - 가장 빠른 비행기에 대해
 - 긴 잎을 가진 식물을 말해 주시오
- ③ 서술 형태의 제약
 - 지구가 도는 이유는
 - 다람쥐와 비슷한 동물은
- ④ 사용자가 해당 형태 요구
 - 원자의 종류는 몇가지인가
 - 영어로 “바둑”이란
 - 정보과학논문에서 컴퓨터를 포함하는

장을 찾아라

⑤ 예/아니오 질문

- 미국에 북한대사관이 있는가
- 인류의 조상은 원숭이인가

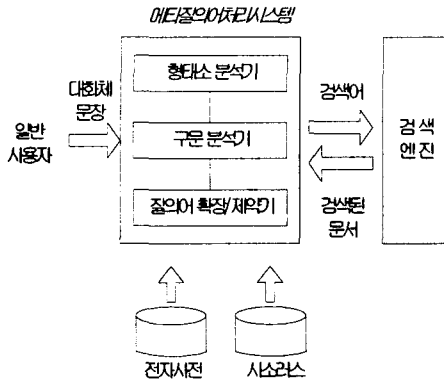
특히, 이들 질의어의 유형에서 구체적인 객체에 대한 질문(②)과 서술 형태의 제약(③)에 관한 질문은 형용사나 부사 등이 포함되어 있으며, 이들은 사용자가 원하는 문서를 검색하는데 있어 고려하지 않아도 되는 어휘이다.

일반적으로, 이와 같이 다양한 형태의 한국어 대화형 질의어를 처리하는 방법에는 2가지로 분류할 수 있다. 하나는 질의어를 단순히 형태소 분석하고, 이를 바탕으로 명사를 추출하여 키워드 중심으로 검색하는 방법이다. 이때, 추출된 명사를 모두 ‘&’ 연산을 통해 불리언 질의어로 변환하는 것이다. 예를 들어, “휴대용 컴퓨터 중 PDA를 제외한 모든 문서를 찾아라”와 같은 질의어에 대해 “휴대용&컴퓨터&PDA&제외&문서”와 같은 불리언 질의어가 생성된다. 이러한 방법은 명사들만을 추출하여 검색하기 때문에 사용자가 명사들을 나열하는 방법과 큰 차이가 없다. 또 다른 방법으로 일정한 패턴을 수집하고 수집된 패턴 내에서 불리언 질의어 유형을 추출하여 부정확한 불리언 질의어를 보완하는 방법이다. 이는 구문 분석의 초기 단계인 부분 파싱 수준으로 해석할 수 있으며, 형태소 분석만을 이용하는 불리언 질의어 생성에 비해 좀 더 나은 결과를 얻을 수 있다.

3. 한국어 대화체 문장 분석을 위한 메타질의어 처리 시스템

한국어 대화체 문장 형태의 질의어를 처리하기 위해 본 논문에서 제안한 메타질의어처리시스템은 전자 사전과 형태소 분석기, 구문 분석기로 구성되어 있다. 또한, 질의어를 제약하거나 유사한 의미로 확장하여 사용자의 정확한 질의어를 반영하기 위해 시소러스를 구축하였다. 다음 [그림 1]은 메타질의어처리시스템의

구조를 나타내고 있다.



[그림 1] 메타질의어처리시스템의 구조

3.1 전자사전의 구조

본 논문에서는 형태소 분석 및 CFG를 기반으로 하는 구문 분석 기법을 이용하여 한국어 문장을 분석하는데 필요한 약 10만개의 한국어 어휘를 추출하여 트라이 구조[1]로 전자사전을 구축하였다. 특히, 단어, 품사 정보, ‘하다’ 정보, 불규칙, 접두사, 접미사, 부가적 용법은 [2]를 기반으로, 구문적 모호성을 해결하기 위해 문형과 공기 정보는 [6]을 기반으로 구축하였다. 본 논문에서 구축한 전자사전의 구조는 다음과 같다.

- ① 사전 구조 : 단어@@(품사/문형)구문 분석에 필요한 정보
- ② 구문 분석에 필요한 정보 : #{품사}*{하다정보}*%불규칙정보[!접두사][!\$접미사](&문형+1의 의미지표^2의 의미지표)*

또한, 형태소 분석 과정에서는 필요할 경우, 미지어 추정 루틴을 통해 사전에 포함되지 않은 새로운 단어에 대해서도 쉽게 불리언 질의어를 생성해 낼 수 있도록 하였다.

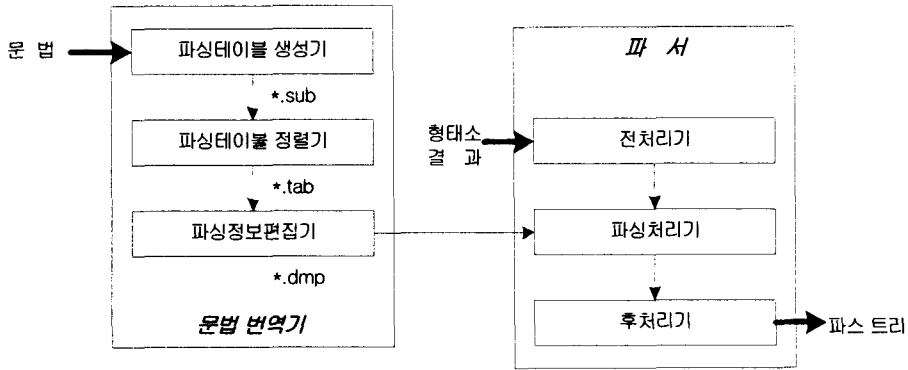
3.2 형태소 분석기와 구문 분석기

형태소 분석기는 자연어 질의로부터 명사 상당어구를 추출하는 역할과 구문 분석기의 입력문을 생성하여 제공하는 역할을 수행한다[5]. 본 논문에서는 형태소 분석 과정에서 숫자뿐만 아니라 영문자 특수 기호 등과 같은 한국어를 제외한 부분을 미리 분리하였으며, 명사 상당어구를 추출할 때 사전에 포함된 표제어를 추출하는 것 이외에 복합 명사를 분리하여 검색의 효율성을 높히도록 하였다. 특히, 표제어의 원형을 찾아내는 원형 복원은 사전 탐색의 횟수를 증가시켜 형태소 분석의 성능을 떨어뜨릴 수 있기 때문에 사전과 원형 복원을 결합하여 사전 탐색의 횟수를 줄이도록 하였다. 또한, 구문 분석은 한국어의 언어적 특성상 분석 중간에 문장 성분간의 관계를 검사하는 동적인 단일화를 선택적으로 지원하는 조건 단일화에 기반한 GLR 파서를 사용하여 한국어가 가지는 문법적 특징을 수용하면서도 빠른 구문 분석 결과를 가지도록 하였다[4][8].

본 논문에서는 한국어 대화체 문장 형태로 입력된 분장의 구조가 문법에 맞는지를 검사하기 위한 파서를 두 부분으로 분리하였다. [그림 2]는 두 부분으로 나뉘어진 파서의 전체적인 시스템의 구조를 나타내고 있다.

하나는 한국어의 구문 규칙을 문법으로 기술하고 문법 번역기를 통해 파싱 과정에 사용할 수 있는 정보로 변환해 주는 부분이다. 다른 하나는 형태소 분석 결과와 파싱 정보를 이용해 입력된 문장을 분석하여 파스 트리로 만들어 주는 부분이다.

문법 번역기는 파싱 테이블 생성기, 파싱 테이블 정렬기, 파싱 정보 편집기로 구성되어 있다. 각각의 기능을 살펴보면 다음과 같다.



[그림 2] 구문 분석기의 구조

① 파싱 테이블 생성기는 한국어 구문 규칙을 기록한 문법을 입력받아 파싱 테이블을 만들어낸다.

② 파싱 테이블 정렬기는 신속한 처리를 위해 생성된 파싱 테이블의 내용을 상대순으로 정렬한다.

③ 파싱 정보 편집기는 파서가 이용하기 위한 편한 형태로 정보를 편집해 주는 역할을 한다.

파싱 단계에서 전처리는 형태소 분석기로부터 출력된 결과를 입력받아 파서가 이용할 수 있는 형태로 재편집하고, 후처리는 파싱된 결과를 파스 트리 형태로 만들어 준다.

3.3 시소러스 사전

시소러스란 해당 주제 분야에서 필요한 모든 개념을 수집하여 이들에 대한 개념의 대소관계나, 동의어, 동형 이의어, 관련어 등을 적절히 조절하여 정보시스템과 문헌생산자, 색인 작성자, 이용자간에 통일적으로 사용할 수 있도록 통제하여둔 용어 통제어표를 말한다 [10][12]. 일반적으로, 질의어에 포함된 하나의 검색어만으로 해당 주제를 전부 검색할 수 없

으므로 그 검색어에 관련된 개념의 대소, 관련어 등을 모두 검색하여야 하는 불편이 뒤따르게 된다. 이를 위해서는 시소러스를 이용한 질의어의 확장이 필요하다.

본 논문에서는 시소러스를 이용하여 질의어의 제약과 확장에 이용될 수 있도록 하였다. [표 1]은 본 논문에서 구현한 시스템에서 구축한 약 8만여 어휘 중 시소러스의 일부분을 나타낸 것이다. 하나의 어휘에 대해 의미적으로 명확하게 구분되는 다른 뜻이 있을 경우, 이를 분리하고 각 의미에 대해 의미적으로 관계가 있는 단어를 구축하였다. 또한, 영어 단어를 포함하는 문서 검색에 대비하여 필요한 경우, 해당 어휘에 대한 영어 표현을 포함하였다.

4. 한국어 대화체의 질의어 처리

한국어 대화체 문장 형태의 질의어를 불리언 질의어 형태로 변환하기 위해서는 먼저 질의어 분석이 이루어져야 한다. 질의어 분석이 이루어진 후, 먼저 분석된 결과로부터 주제어를 추출하고 주제어 사이의 연산 관계, 즉 검색 연산자를 결정하게 된다.

주제어를 추출하기 위해서는 질의어의 분석 결과의 트리를 순회하며, 단말 노드의 범주(category) 정보(명사를 포함한 체언구)를 검사하여 유용한 키워드인지 아닌지를 결정한다.

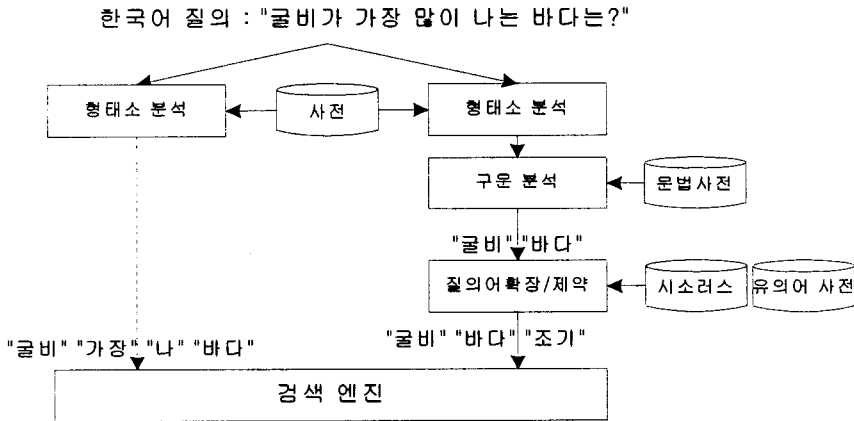
[표 1] 질의어 처리에서 사용한 시소러스의 일부

[가곡]=무언가 가사|song#노래 가락|song#곡 시조|song
 [가나]=허라가나 가다가나|Ghana#가나공화국|Ghana
 [가래]=가래 삼|spade#침 타액|sputum
 [가마]=상여 전기로 자갈밭 건식 제련|palanquin#장차 가열 장치 가열실 연소실
 |kiln#짚곡식 그릇|sack#탈것 교역|palanquin#바스코다가마인도
 캘리컷 희망봉

예를 들어, [그림 3]은 "굴비가 가장 많이 나는 바다는?"이란 사용자의 검색 질의어에 대해 적절한 불리언 질의어를 생성하는 과정을 나타내고 있다.

사용자의 검색 질의어에 대해 형태소 분석이나 부분적 패턴 정보에 의하여 불리언 질의어를 생성할 경우, 아래와 같은 불리언 질의어를 생성한다.

의어에서 "가장"은 부사와 명사로 분류될 수 있는 품사 모호성을 가지며, "나는"은 "나/대명사+는/조사"의 분석과 "날/동사+는/관형형어미", "나/동사+는/관형형어미"의 분석으로 인해 "나"가 불리언 질의어로 선택된다. 이러한 모호성은 질의어 문장에서 "가장"과 "나는"이 어떠한 문법적 역할을 수행하고 있는지를 분석해야만 해결될 수 있는 문제들이다. 즉, 한국어 단



[그림 3] 한국어 대화체 질의어의 처리 과정

- ① 대화체 문장 질의어 : 굴비가 가장 많이 나는 바다는?
- ② 형태소 분석 후 불리언 질의어 : 굴비 & 가장 & 나 & 바다
- ③ 올바른 불리언 질의어 : 굴비 & 바다

즉, [그림 3]에 나타난 바와 같이 위의 질

어가 가지는 품사 모호성과 하나의 단어가 다른 여러 형태소들의 결합으로 이루어질 때 나타나는 어휘 모호성으로 인해 사전과 형태소 분석만을 이용하는 경우는 ②와 같은 불리언 질의를 생성하게 됨으로써 사용자가 원하지 않는 문서들까지도 추출하게 되어 많은 시간과 노력을 요구하게 된다.

따라서, 이러한 문제를 해결하고 사용자

질의어에 대한 정확한 이해를 위해서는 부분 파싱 내지는 전체 문장의 파싱 과정이 요구되며, [그림 4]의 구분 분석 결과의 트리에서 나타난 바와 같이 체언구(NP)에 포함되는 “굴비”와 “바다”가 키워드로 채택된다. 그리고, 키워드의 범주에 속하지만 검색 결과의 신뢰성을 높이기 위해 키워드로서 생략을 해야만 하는 경우, 본 논문에서는 [표 2]와 같은 불용어 패턴을 이용하여 분석 과정에서 불용어(stopword)들을 제거하였다.

[표 2] 불용어 처리를 위한 패턴 모음

~에 대한 문서/논문/문헌/기사/연구
 ~에 관한 기사/문서/논문/문헌/연구
 ~을/를 다룬 것/논문/연구/기사/문헌
 ~을/를 응용한/이용한 연구

또한, 명사를 포함한 체언구뿐만 아니라, 의미 있는 동사들도 유사 키워드로 생성하여 키워드 확장 루틴에서 동사 유의어 사전을 기반하여 동사의 유사 명사로 확장되도록 하였다. 예를 들어, “엘비스프레슬리의 죽음”이라는 기사가 담긴 문서를 검색한다 가정했을 때, 사용자가 질의어로서 “엘비스프레슬리는 언제 죽었나?”라고 입력했다면 ‘엘비스프레슬리’만을 키워드로 사용하기보다는 동사를 명사화하여 ‘엘비스프레슬리’와 ‘죽임’을 이용하여 검색한 결과에서 보다 나은 검색 성능을 기대할 수 있다. 이러한 유사 키워드로 다루어지는 동사들로는 [표 3]과 같은 종류들이 있으며, 이들은 테이블로 구성되어 질의어를 확장할 때 이용된다.

[표 3] 유사 키워드로서의 동사의 종류

죽다	다치다	이기다
아프다	싸우다	
나타내다	좋아하다	꿈꾸다
		사라

지다 바라다
 잃다 원하다

분석 과정에서 추출된 키워드들은 연산자를 선택하는 방법에 따라 키워드들 사이에 연산자를 부여하게 된다. 본 논문에서는 [표 5]와 같이 명사에 붙는 조사나 관형형어미, 부사격어미 또는 접속사에 의해 연산자를 결정하게 되며, “통신, 네트워크”와 같이 명사들을 나열한 경우에는 기본적으로 AND 연산자를 부여하였다.

[표 5] 불리언 질의어 생성을 위한 연산자 유형

OR	AND	NOT
A 나 B	A와 B	A를제외한
A 또는 B	A의 B	A가 아닌
A 및 B	A B	A를포함하
	A. B	지 않는
	A 방식의 B	
	A를 이용한 B	
	A를 위한 B	
	A에 사용되는 B	
	A에 대한 B	

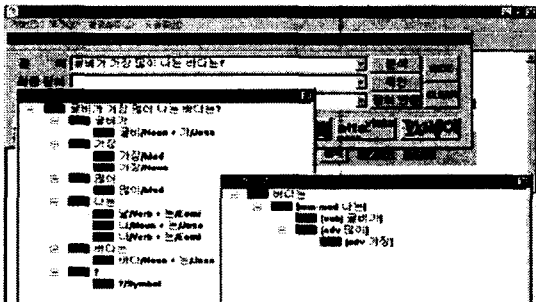
마지막으로, 본 논문에서 제안하는 메타 정보 검색은 불리언 질의어를 생성한 후, 생성된 키워드들에 의해 유의어 사전을 통한 질의어의 확장 및 시소러스를 이용하여 자동으로 다의어의 제약이 이루어지도록 하였다. 이러한 질의어의 확장 및 제약을 통해 사용자가 원하는 문서를 정확하게 추출할 수 있다.

5. 구현 및 실험 결과

본 논문에서 제안한 시스템은 윈도우 환경에서 C/C++ 언어를 이용하여 구현하였으며, 한국어 대화체 문장 분석을 이용한 메타 검색 시스템이다. 이를 이용한 검색은 먼저 한국어 메타질의어처리시스템을 통해 사용자의 질의어

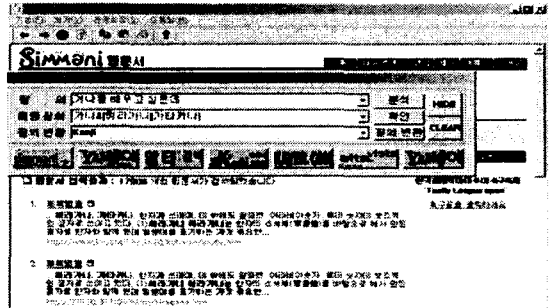
를 분석한 후, 사용자의 요구에 적합한 불리언 질의어를 생성하게 되며, 생성된 불리언 질의어를 기반으로 현재 인터넷상에서 이용 가능한 다양한 검색 엔진을 호출하여 검색 결과를 보여주게 된다.

먼저 사용자의 이해를 돕기 위해 질의어를 입력하면 앞에서 설명한 형태소 분석과 구문 분석 결과를 화면에 보여지도록 하였으며, 주어진 질의어에 의미적 중의성이 포함된 경우, 사용자가 적절한 의미로 제약할 수 있도록 구성되어 있다. 다음 [그림 4]는 사용자의 질의어 "굴비가 가장 많이 나는 바다는?"에 대한 형태소 분석 및 구문 분석 결과를 보여주고 있으며, [그림 5]는 사용자의 질의어를 바탕으로 불리언 질의어를 확장한 것이다.

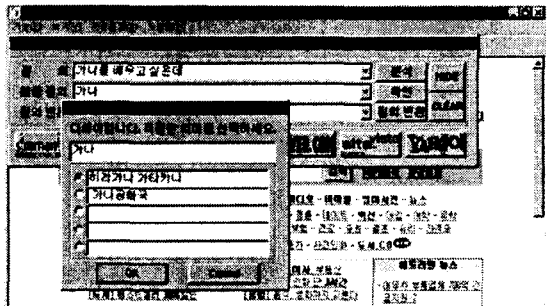


[그림 4] 형태소 및 구문 분석 결과

또한, 만약 질의어에 다의어가 포함된 경우 가능한 한 원하지 않는 문서의 검색을 배제하기 위하여 사용자로 하여금 구축된 시소러스를 이용하여 자동으로 다의어 제약을 가할 수 있도록 하였다. 예를 들어, [그림 6]은 질의어 "가나를 배우고 싶은데"에서 "가나"라는 검색어가 두 가지 의미("히라가나 가다카나" 또는 "가나공화국")를 가지고 있으므로 사용자가 적절한 제약을 통해 검색 결과를 통제할 수 있음을 보여주고 있다.



[그림 7] 시소러스를 이용한 검색 결과



[그림 6] 시소러스를 이용한 불리언 질의어의 제약

본 논문에서는 지금까지 설명한 과정으로 메타질의어처리시스템을 통해 다양한 형태의 한국어 대화체 문장 형태의 질의어를 이용한 검색을 실험하였다. 일반적으로 상용되고 있는 5개의 검색 엔진을 이용하여 100개의 한국어 대화체 문장 형태의 질의어에 대해 실험한 결과를 분석하면 본 논문에서 제안한 메타 검색 방법을 이용함으로써 36% 정도의 불필요한 문서를 필터링할 수 있었다. 또한, 시소러스를 이용한 질의어 제약을 통해 검색된 상위 문서중에서 사용자의 의도가 정확히 반영된 문서들을 얻을 수 있었다.

6. 결론

정보검색 엔진은 사용자가 원하는 문서를 빠르고 정확하게 찾을 수 있어야 하며, 사용자와의 쉬운 인터페이스를 가지고 있어야 한다. 검색 엔진이 사용하기 용이한 인터페이스를 가지려면 익숙하지 않은 불리언 질의어 형태보다는 일상 생활에서 사용하는 자연어를 검색 엔진에서 사용할 수 있어야 한다.

본 논문에서는 한국어 대화체 문장 형태를 기본 입력 형태로 하여, 사용자의 편의성을 높이고 사용자의 의도를 정확히 반영하는 메타 질의어처리시스템을 설계 및 구현하였다. 이를 위해 본 논문에서는 한국어 질의어를 분석할 수 있는 전자 사전을 구축하였으며, 구문 분석 및 부분 구문 분석 기술을 적용하여 한국어 대화체 문장의 질의어 형태로부터 불리언 질의어를 생성하였다. 또한, 사용자의 선택에 따라 구축된 시소러스를 통한 유사어의 확장, 동사의 의미 확장, 대화형 질의어 확장을 가능하도록 하였다. 본 논문에서 제안한 방법은 사용자가 원하지 않는 문서를 가능한 한 배제할 수 있어 향상된 검색 결과를 기대할 수 있으며, 한국어 인터페이스 기술을 요구하는 응용 및 정보 검색 시스템의 사용자 질의어 처리 시스템에 적용할 수 있을 것이다.

한 한국어 구문분석”, 정보과학회논문지, 22권 4호, pp.653-662, 1995.

[5] 이근용, “효율적인”.

[6] 연세한국어사전, 언어정보연구원, 1999.

[7] 이석호, 김성기, “자연어 한글 질의어 처리를 위한 인터페이스의 설계 및 구현”, 한국정보과학회 논문지(C), Vol. 12, No. 1, pp. 31-44, 1985.

[8] 이용석, “구문분석 실용화를 위한 구문분석기의 개발”, 한국전자통신연구원 최종보고서, 1999.

[9] 이승률, 강현규, 박세영, 이상조, “자연어 질의어 정보 검색 시스템의 비주제어 탐색 방법을 통한 성능 개선”, 제6회 한글 및 한국어 정보처리 학술대회, pp.374-377, 1994.

[10] 이영자, 이경호, 정보학개론, 경북대학교출판부, 1994.

[11] 정영미, “정보검색론”, 구미무역출판부, 1993.

[12] Allen. J., Natural Language Understanding, The Benjamin/Cummings publishing Company Inc, 1994.

[13] KTSET95,

<http://nlp.korea.ac.kr/~cmj/kirs/cgi/ktset.html>

참고 문헌

[1] 김철수, “한국어 형태소 분석 환경을 효율적으로 지원하는 사전 구조”, 전북대학교 박사학위 논문, 1998.

[2] 동아 새국어 사전 제3판, 두산동아출판사, 2002.

[3] 박미화, 원형석, 이원일, 이근배, “구문분석에 기반한 자연어 질의로부터 불리언 질의 생성”, 1998.

[4] 양승원, “조건단일화기반 PATR II를 이용

박인철



- 1984 전북대학교 전산통계학졸업
(공학 학사)
- 1986 전북대학교 대학원 전산통계
학졸업 (공학 석사)
- 1998 전북대학교 대학원 전산통계
학졸업 (공학 박사)
- 1999. 10.- 현재 호원대학교 컴퓨
터학부 부교수