

# 자연어를 이용한 대화형 질의 처리 에이전트의 설계

주 형식, 김 남수, 강 현수, 이 용석

전북대학교 컴퓨터과학과

## Designing Interactive Query Processing Agent using Natural Language

Hyung-Sik Ju, Nam-Su Kim, Hyun-Su Kang, Yong-Seok Lee

Dept. of Computer Science, Chonbuk National University

### 요 약

인터넷은 지난 몇 년 동안 급성장을 이룩하였고, 근래에는 전문가 뿐만 아니라 일반 사용자들도 인터넷을 이용하여 정보를 검색, 활용하고 있다. 인터넷에서 검색 엔진의 도움 없이 정보를 검색한다는 것은 불가능한 일이 되었다. 그러나 일반 사용자가 인터넷에서 검색을 할 때 겪는 가장 큰 어려움은 검색 엔진의 특성과 연산자의 다양성, 불리언 질의의 형식화이다. 따라서 본 논문에서는 검색 엔진의 특성에 관계없이 검색을 할 수 있도록 하기 위하여 자연어를 사용하여 정보를 검색한다. 자연어 질의는 일상적인 생활에서 사용하는 언어를 기반으로 하기 때문에 인터넷에 친숙하지 못한 초보자들도 손쉽게 사용할 수 있다는 장점이 있다. 그리고 질의 처리 에이전트와의 대화를 통하여 사용자는 자신의 요구를 더욱 명확하게 할 수 있고, 그 결과 사용자는 자신이 필요로 하는 문서를 더욱 신속하게 찾을 수 있다.

### 1. 서론

지난 몇 년 동안 인터넷은 급성장을 이룩하였고, 근래에는 전문가 뿐만 아니라 일반 사용자들도 인터넷을 이용하여 자신이 필요로 하는 정보를 검색, 활용하고 있다.

정보의 바다라고 불리는 인터넷에서 검색 엔진의 도움 없이 정보를 찾는다는 것은 상상할 수 없는 일이 되어버렸다. 그렇지만 현존하는 검색 엔진의 수는 약 300여 개나 되고, 각 검색 엔진은 고유한 특성을 가지고 있다. 이러한 이유로 일반 사용자가 인터넷에서 정보를 검색할 때 겪는 가장 큰 어려움으로 각각의 검색 엔진의 특성과 연산자의 다양성[2]이 손꼽히고 있다.

이러한 불편을 해소하기 위하여 메타 검색엔진에서는 자동으로 사용자의 질의를 각각의 검색에 맞게 질의의 형식을 변형하여 검색을 수행한다. 다른 검색 엔진과 마찬가지로 메타 검색 엔진 또한 사용자가 형식화한 질의에 기반을 두고 검색을 실행하기 때문에 사용자의 질의 형식화 능력이 검색의 성능을 좌우하는 중요한 요소가 된다.

현재 모든 정보 검색 엔진에서는 불리언 질의를 사용하여 검색을 수행한다. 그렇지만 일반적으로 사용자가 자신의 의도를 불리언 질의로 표현하기란 쉽지 않다.

따라서 검색 엔진의 특성과 불리언 질의의 형식화에 관계없이 사용자의 질의를 검색 엔진에게 전달해 줄 수 있는 인터페이스를 필요로 한다. 이에 가장 적합한 인터페이스가 자연어 인터페이스[3]이다. 자연어 인터페이스는 사용자가 일상적인 생활 속에서 사용하는 언어를 기반으로 하기 때문에 질의 형식화의 어려움이 없으며 검색 시스템에 대한 사전 지식이 전혀 없더라도 손쉽게 사용할 수 있다는 장점이 있다. 이미 일부 검색 엔진[5]에서는 자연어 질의를 제공하고 있는 것 뿐만 아니라

동어의 확장, 발음 확장 등의 옵션을 통하여 자동으로 질의를 확장하여 검색 성능의 향상을 꾀하고 있다. 그렇지만 이 자연어 인터페이스를 타 검색 엔진에 적용할 수 없다.

따라서 본 논문에서는 불리언 질의의 단점을 해결하고, 검색 엔진의 특성에 관계없이 질의를 가능하게 하기 위하여 “자연어를 이용한 대화형 질의 처리 에이전트”를 제안한다.

자연어 질의를 처리하기 위하여 본 시스템에서는 형태소 분석기와 구문 분석기를 이용하고 사용자의 정확한 검색 의도를 파악하기 위하여 와미 지표를 이용한다. 이 질의 처리 에이전트를 이용하여, 사용자는 검색 엔진의 특성을 고려하지 않고 사용자에게 친숙한 언어를 사용하여 질의를 할 수 있다. 또한 질의 처리 에이전트와의 대화를 통하여 사용자는 자신의 모호한 검색 의도를 더욱 명확하게 할 수 있고, 그 결과 사용자는 자신이 필요로 하는 문서를 더욱 신속하게 찾을 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 현재 이용되고 있는 검색 엔진의 일반적인 특징을 살펴보고, 본 논문에서 제안하는 대화형 질의 처리 에이전트에 대하여 설명한다. 3장에서는 본 시스템의 구조를 간략하게 설명하고, 4장에서는 대화형 질의 처리 에이전트의 활용 예를 보인다. 마지막으로 5장에서 결론을 맺는다.

### 2. 대화형 질의 처리 에이전트

본 장에는 기존 정보 검색 엔진의 검색 방법과 질의의 특징을 살펴보고 본 논문에서 제안하는 대화형 질의 처리 에이전트의 기능의 정의와 처리 절차를 설명한다.

2.1 기존 검색 엔진의 검색 방법 및 질의의 특징

현재 이용되고 있는 정보 검색 엔진의 검색 방법은 크게 2 가지이다. 첫째는 키워드 나열형으로 사용자가 찾고자 하는 문서에 대한 키워드를 입력하여 정보를 탐색하는 방법으로 올바른 키워드의 입력이 검색 결과의 성능을 좌우한다. 둘째는 분류 정보를 이용하는 방법으로 특정 정보에 대하여 포괄적인 정보를 얻고자 할 때 이용한다. 이 방법에서는 찾고자 하는 정보의 대분류 정보를 시작점으로 하여 보다 구체적인 소분류 정보를 이용하여 하향식 탐색을 반복함으로써 정보를 탐색한다.

그리고 검색에서 일반적으로 사용되는 질의를 살펴보면 2-3 개의 키워드들로 이루어져 있다. 그러나 이 키워드들을 이용하여 사용자의 정보 검색 욕구를 충분히 표현하기에는 질의가 너무 짧다. 따라서 사용자가 만족할 만한 결과들을 얻는다는 것은 힘든 일이다. 그러나 같은 2-3 개의 키워드들을 사용하는 자연어 질의의 경우에는 키워드 나열형 질의에 비하여 보다 풍족한 정보가 숨어 있음을 발견할 수 있다. 질의의 용언(형용사, 동사)은 키워드의 상태를 제한하여 키워드의 보다 구체적인 정보를 제공한다. 또한 자연어 질의의 경우, 부가적으로 사용자가 검색하고자 하는 분류 정보를 구할 수 있다.

자연어 질의를 이용함으로써 키워드 뿐만 아니라 용언 정보를 활용할 수 있기 때문에 질의를 확장할 수 있을 뿐만 아니라, 질의에서 구해진 분류 정보를 이용하여 사용자에게 보다 정확한 문서를 찾을 수 있도록 해준다. 즉 사용자는 검색 엔진이 제공하고 있는 키워드 검색 방법과 분류 검색 방법을 모두 사용할 수 있다.

2.2 대화형 질의 처리 에이전트의 기능

본 연구에서 제안하는 자연어 질의를 처리하는 대화형 질의 처리 에이전트는 다음과 같은 기능을 수행 한다.

- 1) 사용자에게 검색 엔진의 특성과 관계없이 질의할 수 있어야 한다.
- 2) 사용자는 자신이 표현하기 쉬운 일상 생활 용어로 질의할 수 있어야 한다.
- 3) 사용자의 질의 모호성을 해소하기 위하여 자동 혹은 대화를 통하여 질의 모호성을 해소 할 수 있어야 한다.
- 4) 사용자는 자신의 기호에 맞는 검색 엔진을 결정할 수 있어야 한다.
- 5) 검색 엔진의 검색 방법(키워드 검색, 분류검색)을 활용할 수 있도록 하기 위하여 대화형 질의 처리 에이전트는 자동 혹은 대화를 통하여 질의를 생성할 수 있어야 한다.
- 6) 사용자는 확장된 질의에 대하여 수정할 수 있어야 한다.
- 7) 인터넷 상의 주요 검색 엔진에 대한 도움달 기능을 제공해야 한다.

대화형 질의 처리 에이전트는 사용자와의 대화를 통하여 현재 검색 엔진이 제공해 주고 있는 키워드 나열형 검색과 분류 검색을 모두 제공하여 보다 효율적인 탐색이 가능하도록 한다.

2.3 대화형 질의 처리 에이전트의 처리 절차

대화형 질의 처리 에이전트가 질의를 처리하는 절차는 크게 3 단계로 나누어진다. 첫째 단계는 사용자의 질의의 형식화 단계로 사용자 질의로부터 자동으로 불리언 질의를 생성하는 단계이다. 둘째 단계는 생성된 불리언 질의에 대한 사용자의 피드백 단계로 사용자는 확장된 질의를 더욱 구체화하거나 시스템의 오류를 수정한다. 셋째 단계는 사용자가 선택한 검색 엔

진의 특성에 적합하게 질의를 형식화하는 단계이다.

2.3.1 질의 형식화 단계

질의의 모호성을 해소하기 위하여 한국어 처리 시스템을 활용한다. 구문 분석과 의미 지표를 이용하여 질의를 보다 구체화한다. 동의어의 경우 자동으로 질의를 확장하고, 다의어를 포함하고 있는 경우(“말과 당나귀에 관하여”, “배가 아플 때 어떻게 하나”)는 단어의 관계를 이용하여 질의를 확장한다. 이에 대하여 4장에서 그 예를 보인다. 그러나 그 밖의 경우에는 사용자와의 대화를 통하여 의미를 한정하고 적절한 용어를 추가하여 질의를 형식화한다.

2.3.2 질의 수정 단계

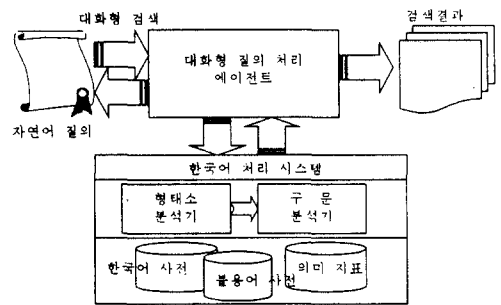
사용자는 질의 처리 에이전트에 의하여 구체화된 질의를 보고 질의를 수정한다. 사용자는 질의 처리 에이전트에 의해 확장된 질의를 더욱 구체화하거나, 또는 질의 처리 에이전트의 오류를 수정하여 자신의 의도를 충분히 반영할 수 있도록 한다.

2.3.3 검색 엔진의 특성에 적합한 질의 생성 단계

질의의 형식화 절차가 끝난 후 사용자는 자신의 특성에 맞는 검색 엔진을 선택한다. 사용자가 검색 엔진을 선택하면 검색 엔진에서 제공하고 있는 다양한 기능에 대한 환경을 설정하는 절차가 이루어진다. 검색 방법(분류검색, 키워드 검색), 범위(웹, 유즈넷,...), 문서의 수 등을 사용자가 선택한 검색 엔진에 맞게 질의를 생성한다.

최종 질의 생성이 끝난 후 질의 처리 에이전트는 검색을 시작한다. 질의 처리 에이전트는 질의에 대한 검색 결과가 없을 경우, 확장된 질의에서 첨가된 단어를 제거하고 재검색 한다.

3 시스템 구조



[그림 1] 시스템 구성도

본 연구에서 사용되는 전체 시스템 구조는 [그림 1]과 같다. 본 시스템은 크게 2 부분으로 나눌 수 있다. 하나는 한국어 처리 시스템으로 형태소 분석기[6], 구문 분석기[11]로 이루어져 있으며, 자연어를 해석하는데 이용된다. 다른 하나는 대화형 질의 처리 에이전트로 한국어 시스템과 대화형 모델에 기반하여 사용자와의 대화를 주도한다. 사용자는 찾고자 하는 정보를 얻기 위하여 대화형 질의 처리 에이전트에게 자연어 질의를 입력

하면, 대화형 질의 처리 에이전트는 한국어 처리 시스템을 이용하여 사용자와의 대화를 통하여 사용자의 요구를 더욱 명확하게 하여 질의를 확장하고, 보다 신속하게 문서를 찾도록 한다.

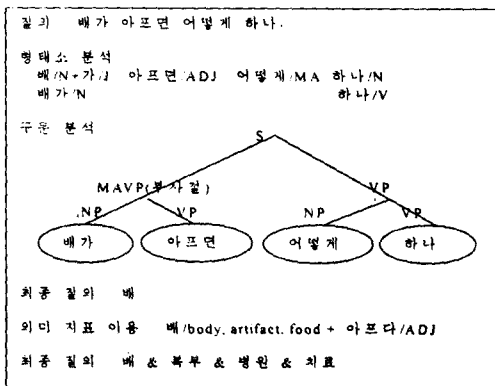
### 3.1 한국어 처리 시스템

본 연구에서는 형태소 분석기, 구문 분석기, 의미 지표 등을 이용하여 사용자의 질의를 분석한다. 사용자가 시스템에 입력한 질의는 형태소 분석과 구문 분석 과정을 거쳐 나온 트리 구조를 이용하여 불리언 질의를 자동으로 생성한다. 다의어의 경우, 의미 지표를 이용하여 단어의 모호성을 해소한다.

의미 지표에 포함된 단어 정보는 다음과 같다. 명사들의 경우 단어들의 계층 정보와 분류 정보가 들어 있다. 계층 정보는 한국어 명사 중 키워드로서의 의미가 있는 명사들로 한정하여 워드넷의 계층 구조를 참조하여 구축 중이다. 분류 정보는 검색 엔진 심마니[12]의 HTML 문서를 이용하여 구축하였다. 검색 엔진 심마니는 14개의 대분류 정보와 이에 해당하는 각각의 소분류 정보로 이루어져 있다. 본 논문에서는 실험의 편의를 위하여 14개의 대분류 정보만을 이용하였다. 또한 동사들의 경우에는 문형 정보를 가지고 있다.

### 4 대화형 질의 처리의 예

[그림 2]는 질의 “배가 아프면 어떻게 하나”에 대한 한국어 처리 시스템을 이용한 질의 처리 과정이다. 질의 “배가 아프면 어떻게 하나”는 형태소 모호성을 포함하고 있지만, 구문 분석 과정을 통하여 모호성이 해소됨을 볼 수 있다. 그러나 구문 분석 과정을 통하여 최종 생성된 질의는 ‘배’라는 단어로 의미 모호성을 지닌다. 본 연구에서는 다의어의 의미를 한정하기 위하여 의미지표를 이용하였다. 본 질의의 경우 ‘아프다’라는 단어에 의하여 ‘배’의 의미가 ‘body’로 한정됨을 볼 수 있다. 단어와 단어의 상호 관계에 의미가 한정되기도 하지만 적용이 어려운 경우에는 사용자와의 대화를 통하여 의미를 한정한다. ‘body’로 한정된 ‘배’의 의미를 더욱 구체화하기 위하여, 관련어를 추가하여 질의를 확장한다. 최종 확장된 질의는 “배 & 복부 & 병원 & 치료”이다. 본 질의의 경우 분류 정보는 사용자와의 대화를 통하여, ‘건강/의학’ 분야로 한정된다.



[그림 2] 한국어 처리 시스템을 이용한 질의 처리 과정

질의 생성이 끝이 나면 사용자는 이용하고자 하는 검색 엔진을 선택하고 검색 엔진의 특성에 따라 각종 옵션을 지정한다.

### 5 결론

본 연구에서는 인터넷 정보 검색 엔진에 익숙하지 못한 일반 사용자들의 편의를 도모하기 위하여 자연어 질의를 이용한 대화형 질의 처리 에이전트를 제안하였다. 대화형 질의 처리 에이전트를 이용함으로써 사용자는 검색 엔진의 특성에 관계없이 질의를 할 수 있으며, 시스템과의 대화를 통하여 자연스럽게 검색 엔진의 특성을 반영할 수 있고, 사용자의 명확하지 못한 질의의 표현 능력을 더욱 구체화할 수 있다. 그러나 자연어 질의는 특성상 한국어 처리 시스템의 견고성에 기반했을 때 더 의미가 있다. 한국어 처리 시스템의 견고성 문제는 꾸준히 연구되어야 할 분야이다. 그리고 현재 구축 중인 의미 지표의 계층 구조의 범위와 정확성 문제도 고려의 대상이 되어야 할 것 같다. 의미 지표의 분류 정보는 실험의 편의상 대분류 정보만을 사용하고 있지만, 세분류 정보를 이용하여 사용자에게 보다 구체적인 정보를 제공해야 할 것으로 보인다. 또한 질의를 확장하기 위해 사전을 사용하고 있지만 용어의 희소성 문제가 예상된다. 이를 위해 코퍼스 분석을 통하여 관련어를 보충할 수 있는 방법에 대한 연구를 필요로 한다.

### [참고문헌]

- [1] 강현규, 양지현, 김영성, 서영훈, “정보 검색에서 질의 형식화를 도와주는 “개념 마법사”의 설계”, 한국 정보 과학회 가을 학술 발표 논문집(II), 1997.
- [2] 고형일, 김병만, 김주연, 김정호, 허남걸, “Web 상에서의 정보검색을 위한 지능형 에이전트의 설계 및 구현”, 한국정보과학회 학술발표논문집, 1997.
- [3] 김경선, 서정연, “지능형 서식 검색 시스템을 위한 자연어 대화 인터페이스”, 한국정보과학회 학술발표논문집(B), 1999
- [4] 박미화, 원형식, 이원일, 이근배, “구문분석에 기반한 자연어 질의로부터의 불리언 질의 생성”, 제 10 회 한글 및 한국어 정보처리 학술대회, 1998.
- [5] 윤보현, 백대호, 김상범, 한경수, 임해창, “대화적 질의 확장을 통해 의미적 용어불일치를 완화하는 정보검색 방안”, 한국정보과학회 학술발표논문집, 1999.
- [6] 이근용, 이용석, 이기오, “효율적인 한국어 분석을 위한 확장된 최장일치법”, 제 8 회 한글 및 한국어 정보처리 학술대회, 1996.
- [7] 이수현, 박동인, “한글 인터넷 정보검색 시스템의 기술방향”, 시스템 공학연구소
- [8] 정영미, “정보검색론”, 구미무역 출판부, 1993
- [9] J. Allen, Natural Language Understanding 2 nd ed. The Benjamin Cummings, 1995.
- [10] William B.Frakes, Richard Baeza-Yates, “Information Retrieval”, Prentice-Hall, 1992.
- [11] Yi-Gyu Hwang, Hyun-Young Lee, Yong-Seok Lee, Gi-o Lee, “Resolution Strategy of Morphological Ambiguity for Korean Parsing”, '99 ICCPOL, 1999.
- [12] 검색 엔진 심마니, <http://www.simmani.com/>
- [13] 검색 엔진 네이버, <http://www.naver.com/>