

# 정보검색을 위한 의문 대명사를 가지는 한국어 질의어 분석

정회원 김기철\*

## Korean Language Query Analysis With an Interrogative Pronoun For Information Retrieval

Gi-cheol Kim\* *Regular Member*

요 약

초보자나 초등학생과 같은 정보검색 시스템 사용자는 정보를 검색하기 위해 주로 질의어를 사용한다. 일반적으로 대명사는 질의어에 자주 사용된다. 질의어로부터 사용자의 일반적인 요구사항들을 에러없이 뽑아내기 위해서는 의문대명사를 고려해야만 한다. 본 논문에서는 한국어로 질의어를 작성할 때 제한된 의문대명사의 사용을 허용하기 위해 4개의 대명사 패턴을 정의한다. 질의어에 있는 의문대명사가 같은 의미와 정보를 가지는 키워드로 대체되면 사용자의 질의 의도가 보다 명확해지게 된다. 이렇게 함으로써 사용자가 원하는 정답 문서만이 상위에 검색되도록 하고자 한다.

**Key Words** : 정보검색, 질의어, Information Retrieval, Query

ABSTRACT

Information retrieval system user such as novices or elementary school students are usually used in query language for IR. Generally, pronoun is scarcely used in query language. So, Interrogative pronoun has to be considered for elicitation of user's formal requirements without any error from Query Language. We defined the four pronoun patterns to allow using of restricted Interrogative pronoun when writing a Query Language in Korean Language. An interrogative pronoun in query language be replaced by keyword with same semantic and information. Then, user's intention for query language become more clearly. For this, only desired documents are retrieved by high level rank.

### I. 서론

정보화 사회에서 정보 검색 시스템의 사용은 필수적이다. 그러나 정보 검색 시스템의 사용자는 숙련자를 비롯하여 아직 한 번도 사용해 보지 않은 초보자에 이르기까지 다양하게 구성되며 검색을 위해 사용되는 질의어 역시 다양한 유형으로 나타난다. 따라서 정보 검색 시스템의 실용화를 위해서는 문

장의 형태나 단어의 쓰임에 관계없이 같은 의도를 가진 질의를 같은 유형으로 분류해 낼 수 있어야 한다. 예를 들어 “작년 인라인대회 우승자는?”과 “작년에 누가 인라인대회에서 우승했죠?”, “작년에 인라인대회에서 우승한 사람은 누구죠”, “작년 인라인대회 우승자” 등은 같은 의도를 가진 자연어 질의어이다. 이와 같은 질의어에서 만족할 만한 응답을 얻기 위해서는 의문형 질의어는 다양한 형태의 텍스트

※ 본 연구는 2006년도 원광보건대학 교내연구비 지원에 의하여 수행되었습니다.

\* 원광보건대학 유아교육과(kckim@wkhc.ac.kr)

논문번호 : 07107-1220, 접수일자 : 2007년 12월 20일

트에서 원하는 응답으로 처리할 수 있고 응용 영역의 변화에 유연하게 대처할 수 있어야 한다. 즉 질의응답 시스템이 하나의 소규모 대화형 시스템이라는 점을 감안하여, 사용자와 컴퓨터간에 지식을 서로 주고받을 수 있는 의미 있는 시스템이 개발되어야 한다. 본 논문에서는 정보검색을 위한 의문 대명사를 가지는 자연어 질의어 분석 방법을 통해 이를 구현하고자 한다. 본 시스템은 크게 두 단계로 구성된다. 첫 번째 단계는 정보 검색을 위한 의문대명사형 자연어 질의에 대하여 구문 정보 및 술어 정보에 기반한 질의 분석 및 확장을 통하여 의문대명사를 구체적인 의미의 키워드로 대체한다. 두 번째 단계는 사용자의 질의 의도에 맞도록 대체된 키워드를 이용하여 사용자가 원하는 정답 문서가 상위에 랭크되도록 한다. 이 과정을 통해 초보자나 정보 검색 시스템에 익숙하지 않은 사용자가 의문형 자연어 질의어로 정보 검색을 하더라도 사용자가 요구하는 문서가 검색됨을 알 수 있었다.

## II. 관련연구

### 2.1 자연어 질의 처리 기법

지금까지 상용되는 대부분의 정보 검색 시스템은 입력 질의 형식으로 불리언 질의 형태와 자연어 질의 형태를 사용한다. 각 정보 검색 시스템마다 약간씩 다른 불리언 연산자와 사용의 불편함 때문에 일반 사용자들은 사용이 용이한 자연어 질의를 이용하는 추세이다. 자연어 질의에 의한 검색 방식에 있어서도 첫째, 초기 자연어 검색, 둘째, 형태소 분석을 이용한 자연어 검색, 셋째, 구문 분석을 이용한 자연어 검색, 넷째, 의미 분석 및 화용 분석을 이용한 자연어 검색 방식으로 나눌 수 있다. 그러나 초등학교생이 다음과 같은 형태의 질의 즉, “삼국사기는 누가 썼나요?” 라는 형태의 질의로 정보 검색을 하고자 할 경우 인간은 후천적 학습과 경험정보에 의해 삼국사기의 저자는 ‘김부식’이라는 것을 알고 있다. 그러나 학습과 의미적 관계에 대한 경험적 정보를 갖추지 못한 컴퓨터에 있어서는 단어들 간의 의미적 관계를 표현해 놓은 의미망(WordNet)이나 온토로지(Ontology)와 같은 지식베이스의 도움 없이 문장 내 단어들 간의 의미적 내포 관계를 표현하는 것은 어려운 일이다.

본 논문에서는 사용자의 질의 의도를 좀 더 명확히 반영하기 위하여 질의에 대한 형태소 분석과 구문 분석과정을 통한 결과를 바탕으로 생성된 질의

에 대한 의미적 관계를 나타내는 자연어 질의 구조에 의한 질의는 WordNet 탐색 비용을 줄일 수 있을 뿐만 아니라 자동화된 구조를 생성하여 사용자의 편이성 관점에서의 제약을 극복할 수 있을 것이다 또한, 이렇게 생성된 자연어 질의 구조에는 각 노드마다 가중치를 부여하여 검색된 결과를 재순위화 함으로서 사용자의 질의 의도가 좀 더 명확히 반영된 검색된 결과를 얻을 수 있을 것이다.

### 2.2 국내 연구 동향

질의응답시스템에서는 사용자 질의의 의미를 분석하는 것과 문서에 나타난 개체들 중에서 답이 될 수 있는 것을 인식하는 것이 결정적인 요소라 할 수 있다. 이는 기존의 웹상에 존재하는 수많은 정보들 중에서 사용자가 원하는 정보를 정확히 검색하기 위한 많은 연구가 진행되어 왔다. 시간이 흐를수록 웹상의 정보량은 기하급수적으로 증가하고 있어서, 이들 정보들 중에서 사용자의 요구에 적합한 정보만을 검색하는 일은 더욱 어려워지고 있다.

한국어 질의 문장을 분석하기 위한 연구는 그다지 많이 이루어지지는 않았다. 그러므로 구문분석이나 의미분석 등이 독립되어 단계적으로 처리되는 문어체 문장 분석과는 다르게 형태소 분석, 구문 분석, 의미 분석, 때로는 그 이상의 화용 분석이 서로 밀접하게 연결되어 처리되어야 함을 제시하고 있다.

연구[1]에서는 질의응답시스템의 성능을 높이기 위해서는 정답 후보(Answer Candidate)를 랭킹하는 방법이 매우 중요하며, 앞으로 QA시스템의 정답 후보 랭킹 방법은 구문관계 보존성 혹은 의미 관계 보존성 등을 적용하는 방향으로 나아가야 하며, 그러기 위해 정교한 자연어 처리 기법의 활용과 충분한 지식 체계의 구축이 연구되어야 함을 제시해 준다.

질의응답시스템의 질의 형태는 자연어 문장이기 때문에 질의의 의미를 파악하기 위해서는 면밀한 분석을 해야 한다. 이러한 질의 분석을 위해 단순히 의문사 정보만을 이용한 기존 연구도 있었고, 규칙 기반 부분 분석을 적용한 연구도 있었다. 의문사 정보만을 이용한 경우는 “누가”, “언제”, “어디에서”, “무엇을”, “어떻게” 등의 의문사로 시작하는 의문사에 대해 사람, 시간, 장소, 방법, 개체 등으로 정답이 될 단어의 범주를 매우 포괄적으로 예측하였다.

연구[2]에서는 한국어 질의응답시스템에서 정확한 정답추출을 위한 구문정보에 기반한 질의분석을 제안했는데, 질의분석은 세부 정답유형 결정, 세분화된 키워드 추출을 통해 정확한 정답추출을 목적으로

로 하며, 술어 유형정보를 이용하여 질의분석을 수행하고, 구문구조 정보를 이용하여 중요키워드와 일반 키워드를 추출한 후 정답 유형 자질 명사를 이용하여 세부 정답 유형을 결정하였다.

이에 본 논문에서는 세부 정답 유형을 결정하고 질의의 초점을 나타내는 중요 키워드를 추출하는 질의 분석을 통해 정답 추출단계에서 보다 정확한 정답을 추출할 수 있을 것으로 보이며, 국내 검색엔진에서는 적용되고 있지 않은 자연어 질의를 기반으로 하여 검색엔진을 한번도 접해보지 않은 초보자라도 쉽게 검색이 가능하도록 편리한 사용자 인터페이스를 제공하고자 한다.

### 2.3 국외 연구동향

연구[3]에서는 질의처리 과정에서 질의 유형이 질의 응답시스템이 문서에서 정답이 될 수 있는 정답후보들을 추출하는데 중요한 정보를 제공해 주며, 최근 TREC에서 소개된 질의응답 시스템들은 대부분 질의유형 분류를 위한 모듈을 포함하고 있다는 사실을 제시해준다

연구[4],[5]에서 질의유형은 사용자의 질의 의도를 특정한 범주에 할당하는 것으로 질의 응답 시스템 연구의 한 분야로 진행되어 왔다. 일반적으로 규칙에 기반한 접근 방법을 채택한 질의응답 시스템들은 다음과 같은 장점을 가지고 있음을 알려준다.

- ▶ 질의 유형 분류 과정이 유한 상태 오토마타로 구현되므로 사용자의 질의에 대해서 즉각적으로 질의 유형을 분류해 낼 수 있다.
- ▶ 수동으로 기술된 규칙에 따라 질의 유형을 분류하므로 질의 유형을 잘못 분류해서 전혀 관계없는 엉뚱한 대답을 하는 경우를 방지할 수 있다.

그러나 규칙에 기반한 질의 유형 분류 방법은 규칙을 수정하기 위해서 전문적인 지식을 가진 사람들의 노력이 필요하고 규칙과 일치되지 않는 질의가 들어왔을 때는 질의 유형을 분류할 수 없는 문제점을 가지고 있다.

본 논문에서는 질의의 유형을 몇 가지로 나누어 보고, 각각을 규칙으로 표현하여 질의 작성이 용이한 자연어 질의 방식을 도입하여 형태소 분석과 구문 분석, 그리고 유의어 사전 등을 통하여 사용자의 질의를 분석하고 이 과정에 의해 생성된 질의를 통하여 수없이 많은 정보들 중에서 사용자가 원하는

정보만을 검색하기 위한 자연어 처리 접근 방식으로 도입하여 정답후보에 가장 핵심적인 키워드를 추출하고 색인을 관리하는 검색방법과 실험 배경을 중심으로 소개하는 방식을 사용하는 독창성을 가진다.

## Ⅲ. 질의처리

### 3.1 의문대명사형 질의의 특성

많은 수의 초등학생들이 인터넷을 통하여 질문에 대한 정확한 답을 얻고자 하는데 “누구인가?” “무엇인가?” “왜 그런 것인가?” 등의 의문형 대명사를 이용하여 질의를 구성한다. 이러한 현상은 많지 않은 어휘만을 습득한 경우 자신의 의사 표현에 적합한 어휘 선정의 어려움에서 기인한다고 할 수 있다. 비단 적은 어휘만을 습득한 경우 뿐 아니라 학습과 경험에 의해 많은 어휘를 습득한 경우라고 하더라도 순간적인 어휘 선택 장애로 인해 이러한 현상을 경험 할 수 있다. 예를 들어, 동의보감의 저자를 알고자하여 질의를 구성 할 때 “저자”라는 어휘의 선택에 장애가 있을 경우, “동의보감은 누가 썼나요?”와 같이 의문대명사를 포함한 형태로 질의를 구성하게 된다. 또한, 적합한 어휘를 선택했다고 하더라도 한국어 의문형 문장의 특성상 완성된 문장을 구성하기위해 중복된 의미로 의문대명사를 사용하는 경우가 있다. 위의 예에서 “저자”라는 어휘를 선택하였다고 하더라도 질의 의문형 문장의 완성을 위해 “동의보감의 저자는 누구인가요?”와 같이 의문대명사를 포함하게 된다. 이러한 경우, 특별한 후처리 없이 일반적인 자연어 처리 과정만 거치게 된다면 “누구”도 키워드로 추출되게 된다. 이는 검색에 있어서 올바른지 못한 문서를 정답 문서로 검색해 주게 되는 부정적 확률을 높여주게 되는 하나의 원인이 될 수 있다.

일반적으로 문장은 일정한 문법적 규칙에 따라 이루어지므로 구조적 형식에 있어서 공통되는 일정한 틀을 갖고, 제한된 수의 유형으로 나뉠 수 있다. 따라서 먼저 질의에 사용될 동사의 유형을 분류하기 위해 국립국어원 21세기 세종계획에서 제공하는 ‘현대 국어 사용 빈도 조사 결과’의 동사를 빈도수별로 고빈도, 중빈도, 저빈도, 중의성 동사를 통해 행위주체, 행위대상을 정리하고, 네이버 지식인을 통하여 본 연구에 적합한 질의 문장들을 수집하였다.

본 연구에서는 분류된 동사들을 중심으로 동사정보를 통하여 질의에 포함된 의문대명사를 구체적인 키워드로 변환이 가능함을 보이고자 한다. 따라서 많

은 동사들 중에서 사용 빈도수가 확연히 차이가 나는 경계지점을 중심으로 분류하였으며, 다의성이 있는 일부 동사 중에서 빈도수가 높은 동사를 선택하였다.

의문대명사형 질의 문장의 분석을 통한 의문대명사의 구체화를 위해 정보검색시스템에서 사용되는 질의어의 문장 구조를 파악하여 문장의 의미를 제약하면 보다 정확한 분석 결과를 얻을 수 있다.

의문대명사형 질의 문장은 서술어(동사형)에 따라 의문대명사를 주어형, 목적어형, 부사어형, 술어형으로 분류하였다.

표 1. 주어형 의문질의어 패턴

	질의	동사	행위주체
1	삼국사기는 누가 썼나요? (*모자, 약, 도구, 먼지)	쓰다	사람→저자, 지은이, 작가, 글쓴이
2	누가 전화기를 발명했나요?	발명하다	사람→발명가
3	한글은 누가 만들었나요? (*노래, 교과서)	만들다	사람→발명가
4	지하철역 이름은 누가 어떻게 짓나요? 지하철역 이름은 누가 어떤식으로 짓나요?	짓다	사람→작명가 어떻게→방식, 방법

표 2. 목적어형 의문질의어 패턴

	질의	동사	행위주체	행위대상
1	코끼리는 무엇을 먹나요?	먹다	코끼리	먹이
2	천주교에서는 누구를 믿나요?	믿다	천주교	교주
3	단군이야기에서 태백산은 어디를 의미합니까?	의미하다	태백산	지역
4	예날 미인들은 미를 위해 무엇을 했나요?	하다	예날미인	방법

표 3. 부사형 의문질의어 패턴

	질의	동사	행위대상
1	수정과는 무엇으로 만드나요?	만들다	재료/도구
2	선인장은 무엇으로 세나요?	세다	단위/도구
3	불국사는 어디에 있나요?	있다	위치
4	벽화는 무엇으로 그렸을까요?	그리다	재료/도구

표 4. 술어형 의문질의어 패턴

	질의	subj
1	코끼리의 먹이는 무엇인가요?	subj+무엇 → subj=먹이
2	까마귀의 학명은 무엇인가요?	subj+무엇 → subj=학명
3	동의보감의 저자는 누구인가요?	subj+누구 → subj=저자
4	링컨 대통령의 저격법은 누구인가요?	subj+누구 → subj=저격법

의문질의어 문장 분석을 자연어 처리에 적합한 구문 및 의미분석 단계에서 이용 가능하도록 의미적인 결합관계가 적용된 새로운 의문질의어 문장을 고찰하였다. 이러한 질의 유형은 자연어 인터페이스가 다양한 도메인으로 확장되어도 유사한 형태를 가지며, 단지 “누가”, “언제”, “어디에” “무엇으로” 등에 해당되는 객체만이 주어진 도메인에 영향을 미치게 될 것이다. 이외에도 술어가 가지는 중의적 성질을 고려하여 질의 유형을 분류하여야 한다.

예를 들어, 의문대명사형 질의 문장에 “쓰다”와 같이 다의어가 포함된 경우에는 사용자 의도에 적합하지 않은 많은 문서들이 검색되게 된다. 이런 경우 질의 문장의 정보만으로는 다의어의 의미를 정확하게 파악하기가 어렵다. 본 논문에서는 의문형 질의 처리 에이전트를 이용하여 시스템과의 대화를 통해 다의어의 경우 사용자로 하여금 그 의미를 결정하도록 하여 이러한 문제를 해결한다. 예를 들어, “동의보감은 누가 썼나요?”와 같은 의문대명사형 자연어 질의 문장에서 “쓰다”라는 동사는 “모자를 쓰다”, “약을 쓰다”, “돈을 쓰다”, “글을 쓰다”, “맛이 쓰다” 등 많은 의미를 가지고 있다. 이러한 경우 일반적인 자연어 처리 방법으로 처리하였을 경우 “누구”라는 키워드가 구체적인 키워드인 “저자”로 변형되기 위해서는 동사 “쓰다”의 여러 의미 중 “글을 쓰다”의 의미를 사용자로 하여금 선택하도록 하고, 행위 주체 및 행위 대상 정보를 통하여 그 의미를 가지는 다른 키워드로 대체한다. “저자”라는 키워드로 변형 후 검색을 하게 되면 사용자의 검색 의도에 보다 적합한 문서만을 검색하게 될 것이다.

(Q1) 동의보감은 누가 썼나요? -> s [ 썼나요 ] (쓰) -> obj [ 동의보감은 ] (동의보감) -> subj [ 누가 ] (누구)
------------------------------------------------------------------------------------------

그림 1. 주어형 질의에 대한 구문분석 결과

표 5. 중의성 동사 “쓰다”의 정보

질의문장	동의보감은 누가 썼나요?		
동사	행위주체	행위대상	일반 키워드
쓰다	사람 └ 저자, 작가, 지은이...	글, 책, 원고, ...	누구
	사람 └ 어른, 아기, ...	모자, ...	
	사람 └ 연필, 붓, ...	도구	

구문 분석 결과에서 동사 “쓰다”의 의미를 사용자로 하여금 결정할 수 있도록 제공을 하고, 결정된 의미의 동사에 대해 동사 정보 테이블로부터 의문대명사 “누구”에 대한 구체적인 키워드로 변형을 하게 된다.

### 3.2 의문대명사형 질의 문장 분석

일반적으로 정보검색시스템에서는 사용자의 질의 의도를 정확히 분석하는 것과 검색된 문서들 중에서 정답 문서들의 순위를 상위에 랭크되도록 하는 것이 결정적인 요소라 할 수 있다. 기존의 검색 연구들은 주로 명사의 확장 및 명사 개념망을 이용하거나 가중치 부여 방법에 관한 연구 쪽에 초점이 맞추어져 있었다.

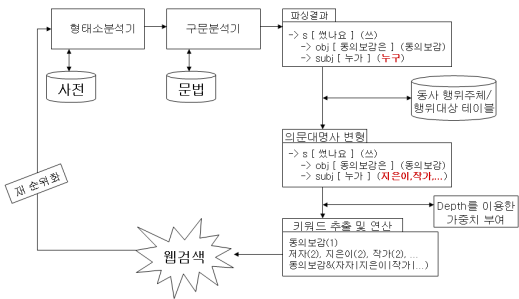


그림 2. 전체 처리 과정 구성도

본 논문에서는 그림 2와 같은 절차로 의문대명사가 포함된 질의문을 분석하여 의문대명사가 의미하는 구체적인 개념어로 질의를 재구성 하게 된다. 이렇게 질의를 분석함으로써 사용자가 원하는 정확한 개념어를 제시하게 되고, 사용자의 질의 의도를 반영할 수가 있다. 또한 이러한 질의분석을 수행함으로써 질의의 유형과 질의의 제약을 설정할 수 있다.

질의분석을 통해 구체적 개념을 파악하기 위해서는 질의의 의미를 어떻게 알아낼 것인가에 대한 고찰이 선행되어야 한다. 본 연구에서 제안하는 시스템의 질의 범주는 활용빈도가 높은 동사를 고빈도, 중빈도, 저빈도로 나누어 주어로서 의문대명사 20개, 목적어로서 의문대명사 40개, 부사어로서 의문대명사 40개, 술어로서 의문대명사 50개를 초등학교 아이들이 잘 표현할 수 있는 것들을 중심으로 선택하였다.

질의의 의미를 알아내기 위해서는 질의에서 범주를 결정짓는 요소가 무엇인가를 파악해야 한다. 가장 쉬운 방법으로는 질의의 의문사를 보고 질의 범

주를 결정하는 것이 있다. 그러나 어떤 질의의 경우에는 범주를 결정하는 데 의문사만을 고려하는 것이 충분하지 않을 수 있다. 따라서 질의의 유형에 따라 범주를 결정할 때 고려해야 하는 요소를 구분하였다. 첫째, “누구”, “어디”, “언제” 등의 의문문은 의문사만으로 정답의 범주를 쉽게 결정할 수 있고, “어떻게” 경우는 “어떻게” 다음에 나타난 형용사나 부사에 따라 범주를 결정함을 알 수 있다. 의문사가 “무엇” “어느것”인 의문문과 기타 의문문의 경우는 정답의 범주를 결정하기 어렵기 때문에 동사 정보를 이용한다.

### 3.3 의문대명사형 질의 처리

이 절에서는 의문대명사형 질의 처리 에이전트의 구성을 위한 알고리즘을 의사 코드(pseudo-code)를 통하여 살펴보는데 주어로서의 의문대명사에 대한 처리 과정만을 보인다.

표 6에서는 주어로서의 의문대명사에 대한 구문 분석 결과에 나타나는 문법 자질 정보를 이용하여 구조를 생성하는 과정을 나타내고 있다.

표 6에서 predi=“쓰다”의 술어 정보를 이용하여 주어로서의 의문대명사인 subj=“누가”는 “저자”를 의미적으로 내포하고 있음을 알 수 있다. 구문 분석 결과를 바탕으로 한 이러한 단서들을 중심으로 어절간의 의미적 포함 관계가 형성된다.

표 6. 주어로서의 의문대명사

동의보감은 누가 썼나요?
-> s [ 썼나요 ] (쓰) -> obj [ 동의보감 ] (동의보감) -> subj [ 누가 ] (누구)
while(!EOF){ if(read_line[0] == s){ if(read_line[2] != intpronoun){ predi = read_line[2]; // root of s } } else{ if(read_line[2] != intpronoun){ head = read_line[2]; } else{ if(read_line[0] == subj){ CL1 = SearchVerbTbl(predi).central; } else if(read_line[0] == obj){ CL1 = SearchVerbTbl(predi).target; } } } line++;

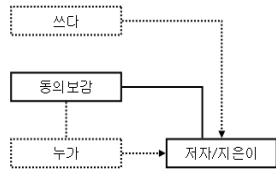


그림 3. 주어로서의 의문대명사 구조

따라서 구문 분석 결과로 나타나는 문법 자질 정보와 이를 통한 문맥 구조 생성 과정에 의해 그림 3과 같은 구조를 생성하게 된다.

#### IV. 실험 및 평가

##### 4.1 실험 방법

실험에서는 의문대명사가 포함된 질의를 각 유형별로 선택하여 일반적인 자연어 질의 형태로 검색한 결과와 본 논문에서 제안한 방법에 의한 검색 결과를 비교하였다.

##### 4.2 실험 결과 및 평가

실험에 사용된 각 유형별 의문대명사형 자연어 질의를 통하여 검색된 결과는 표 7과 같다.

표 7. 의문대명사형 자연어질의에 의한 검색 문서 수

	질의	상위 20위 안의 정답 수
NQ1	동의보감은 누가 썼나요?	0
NQ2	누가 전구를 발명했나요?	11 (최상위 4위)
NQ3	지렁이는 무엇을 먹나요?	0
NQ4	수정과는 무엇으로 만드나요?	0
NQ5	벽화는 무엇으로 그렸을까요?	0
NQ6	미국의 수도는 어디인가요?	7 (최상위 5위)
NQ7	구교와 신교의 차이점은 무엇인가요?	15 (최상위 1위)

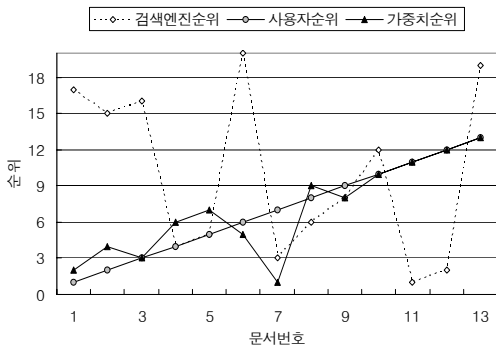


그림 4. NQ1에 대한 검색 결과 비교

그림 4~그림 7까지는 각 질의에 대해 피 실험자에 의해 결정된 검색 문서의 순위(사용자순위)와 상용 검색엔진에 의한 순위(검색엔진순위), 그리고 본 논문에서 제안한 방법에 의한 순위(가중치순위)에 대한 결과를 비교를 그래프로 보이고 있다.

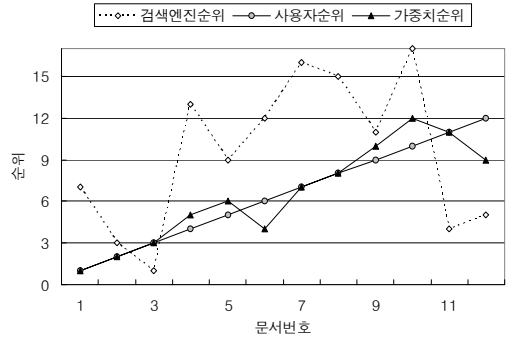


그림 5. NQ2에 대한 검색 결과 비교

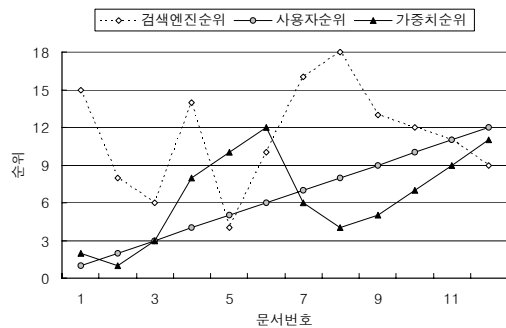


그림 6. NQ3에 대한 검색 결과 비교

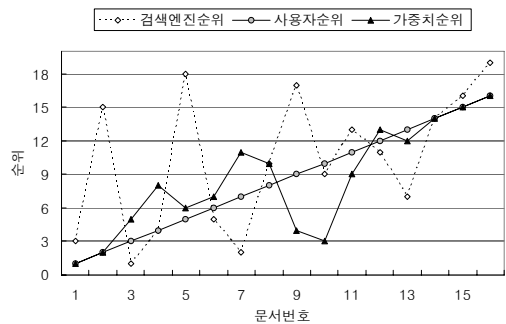


그림 7. NQ7에 대한 검색 결과 비교

그림 4~그림 7에 보인 그래프에서 본 논문에서 제안한 방법에 의한 결과가 보다 사용자 순위에 근접한 결과를 보임을 알 수 있고, 이는 질의에 대한 사용자 의도가 보다 명확히 반영된 결과임을 알 수 있다.

### V. 결론 및 향후 연구

본 논문에서는 의문대명사가 포함된 자연어 질의에 대하여 동사 정보를 이용한 의문대명사의 구체화 및 문장 유형의 분석을 통해 키워드에 가중치를 부여함으로써, 사용자의 질의 의도에 보다 적합한 검색 결과를 얻을 수 있음을 보였다. 또한 이를 통하여 일반적인 질의응답시스템에서 사용되는 질의를 웹상의 검색 엔진에 적용하여 별도의 데이터베이스를 구축하지 않고 웹을 질의응답시스템으로 활용 가능함을 보여주었다.

본 논문에서 제안한 방법에 의한 검색 결과는 사용자 순위에서 좀 더 근접한 결과를 보였다. 불리언 질의가 아닌 자연어 질의를 기본 입력 형태로 하였기 때문에 초보자라도 쉽게 사용할 수 있으며 서로 다른 검색 엔진마다 적합한 불질의를 생성해 줄 수 있으므로 사용자는 연산자를 알 필요가 없을 뿐만 아니라 검색엔진에 제약을 받지 않아도 된다. 또한 대화를 통해 사용자의 의도를 정확히 파악하기 때문에 사용자가 원하는 문서만 검색 될 뿐만 아니라 불필요한 문서의 검색을 방지하여 검색 효율을 향상시킬 수 있다.

향후 연구로는 본 연구를 토대로 보다 많은 동사에 대한 정보의 구축이 진행되어야 하고, 수동형 문장을 포함하여 폭넓은 의문대명사형 질의에 대한 분석이 있어야 할 것이다.

### 참 고 문 헌

- [1] 나승훈, 강인수, 권오욱, 이종혁, “질의 응답 시스템에서 구문 근접성에 기반한 정답후보 랭킹 방법”, 한국정보과학회 2002년 춘계학술대회, VOL.29 NO.01 pp.478-480, 2002.04.
- [2] 신승은, 서영훈, “한국어 질의응답시스템에서 구문정보에 기반한 질의분석”, 한국정보과학회 2004년 춘계학술대회, VOL.31 NO.01 pp.931-933, 2004.04.
- [3] G.S.Mann, “A Statistical Method for Short Answer Extraction”, In Proceedngs of the ACL Workshop Open-Domain Question Answering, pp.13-30, 2001.
- [4] Moldovan D., Harabagiu S., Pasca M., “LASSO: A Tool for Surfing the Answer Net”, In Proceedings of The Eighth Text REtrieval Conference(TREC-8), pp.175, 1999.

- [5] Parikh, J., Narasimha Murty, M., “Adapting Question Answering Techniques to the Web”, In Proceedings of the Language Engineering Conference 2002, pp.163-171, 2002.12.

김 기 철 (Gi-cheol Kim)

정회원



1984년 2월 전북대학교 전산통계학과 졸업

1986년 2월전북대학교 전산통계학과 석사

1995년 8월전북대학교 전산통계학과 박사

1988년~현재 원광보건대학 유아

교육과 교수

<관심분야> 정보검색, 자연어 처리, 음성인식, 프로그래밍 언어론