

의문의 초점을 고려한 자연어 기반의 정보검색 시스템

창신컴퓨터(주) 박홍원

Natural language based Information Retrieval System
considering the focus of the question.

ChangShin Computer Inc. Hong-won, Park

요약 본 논문에서는 기존의 키워드 검색 시스템의 불편함과 비효율성을 지적하고 이를 극복하기 위해 한국어 의문문 자체를 질의어로 채택하여 정보를 검색하는 자연어 기반의 정보검색 시스템을 제안하였다. 본 시스템은 주격 주제어와 서술격 주제어는 물론 의문의 초점과 초점 관련 어구에 대해서도 질의어 분석 단계에서 분석하여 검색자의 요구에 부응하는 응답문 검색이 가능하도록 설계하였다.

본 논문에서는 의문문 질의 시스템에 적합하도록 의문사를 5형태로 분류하고 실제 한국어 문장에서 이들 각각에 대한 처리를 규칙화시켜 질의어의 체계적인 분석을 시도하였다. 한편, 후보 문장 검색을 위한 색인어로 사용되는 주격 주제어와 서술격 주제어를 정해진 규칙을 통해 추출함으로써 체계적이고 정확도 높은 질의어 분석이 이루어지도록 했다. 뿐만 아니라 의문의 초점과 초점 관련 어구또한 정해진 규칙을 통해 분석·추출함으로써 응답문 검색의 정확성을 높였다.

1. 서론

1-1. 연구 배경

최근들어 단순한 키워드 검색의 차원을 벗어나 자연어 기반의 정보검색을 시도하려는 노력이 활발해지고 있다. 80년대부터 있어왔던 “지능형 정보검색 시스템”에 관한 연구는 자연어 형태의 질의문을 이해할 수 있는 데 초점을 맞추어 발전해왔다.[1]

자연어 질의 인터페이스를 비롯한 자연어 기반의 정보검색 시스템에 대해서는 자연어 질의어의 다양한 형태 연구 [2], 생략된 질의문 생성 [3], 자연어 질의 분해 기법 [4] 등에 관한 연구가 진행되어 왔지만 자연어 질의 시스템에서 정작 중요한 의문의 초점에 대한 연구는 부족한 것이 사실이었다.

한편, 실제 일반에서 널리 사용되고 있는 키워드 정보검색 시스템에서도 몇몇 업체들로부터 간단한 수준의 자연어 처리가 도입되고 있다. 국내에서 서비스하고 있는 9개의

대표적인 검색엔진(심마니, 정보탐정, 까치네, 넷서치, 미스다찾니, iBOT, 와카노, 애니서치, 웹글라이더, 코시크) 중에서 부분적이거나 자연어 처리 기능을 지원하는 엔진은 현재 2종류가 있다.[5] 그러나, 이러한 자연어 처리가 대부분 특정 품사에 대한 불용어 처리를 통해 색인어를 추출하는 정도의 수준이어서 사실상 자연어 기반의 정보검색 시스템이라고 보기에는 다소 무리가 있다.

이에 본 논문에서는 창신컴퓨터(주)가 지난 95년부터 최근까지 개발해 온 자연어 기반의 주문형 정보검색 시스템인 <ReQuest>를 토대로 하여 의문의 초점을 고려한 질의어 분석과 정보검색 시스템 개발에 대해 자세히 살펴보고자 한다.

1-2. 연구 방법과 범위

우선 본 논문에서는 의문의 초점을 고려한 자연어 기반의 의문문 질의 정보검색 시스템(이하, 의문문 질의 시스템)을 설명하기에 앞서 기존의 키워드 검색 시스템이 갖는 문제점과 한계성을 지적하였다. 그리고, 이러한 문제점에

대비시켜 본 의문문 질의 시스템의 필요성을 언급하였다.

그 다음으로 한국어 의문문 자체를 질의어로 채택하는 본 의문문 질의 시스템을 구현하는데 있어 고려해야 할 몇 가지 문제점에 대해 언급하였다. 본 논문에서는 이러한 몇 가지 문제점 각각을 해결하는 과정과 그 과정상에 나타난 자연어 처리 기법을 본 의문문 질의 시스템 구현의 핵심내용으로 서술하였다.

본 논문은 기존의 키워드 검색 시스템이나 자연어 질의 시스템에서 다루지 못했던 의문문 질의 인터페이스에 대해 다루고 있다. 따라서, 기타의 논고에서 다루어지지 않았던 의문의 초점과 의문사 분석에 대해서만 다루려 한다. 자연어 기반의 정보검색 시스템에서 흔히 다루어져야 하는 형태소 분석이나 구문구조 분석등에 대해서는 다루지 않았다.

2. 의문문 질의 시스템의 필요성

2-1. 키워드 검색의 문제점과 한계점

현재 일반인이 가장 쉽게 접할 수 있고 보편적으로 사용하고 있는 것이 바로 키워드 검색 시스템이다.[5] 이 키워드 검색은 다음과 같은 몇가지 문제점을 내포하고 있으므로 검색 시스템으로서의 한계성을 드러낸다.

- (1)문자열 자체에 대한 검색이 주로 이루어지므로 동의어나 유사어에 대한 검색이 고려되지 않는다.
- (2)주로 명사같은 체언 위주의 검색만이 가능하므로 동사나 형용사같은 용언의 검색이 불가능하다.
- (3)입력된 키워드가 어휘변별력을 크게 갖지 못할 경우, 불필요한 정보들까지도 검색하게 된다.
- (4)키워드를 통해 찾은 정보들로부터 실제로 원하는 정보를 검색하는데까지 엄청난 시간이 걸린다.

키워드 검색의 이러한 문제점은 본문 검색시 더욱 두드러지게 나타나 인터넷상의 웹문서 검색이나 신문기사 검색에서 그 한계성을 여실히 드러내고 있다.

2-2. 의문문 질의 시스템의 필요성

자연어 질의 인터페이스는 일상 생활에서 사용되는 자연어를 사용하여 질의할 수 있으므로 일반 사용자들이 쉽게

데이터베이스에 접근할 수 있도록 도와준다.[4] 의문문 질의 정보검색 시스템은 이같은 자연어 정보검색의 기능을 한층 강화시킨 것으로 자연어 중에서도 의문문을 질의어로 채택하는 형태이다.

예를 들어, 임진왜란이 몇 년도에 일어났는지에 대한 정보를 검색하고자 할 경우, 기존의 키워드 검색 시스템을 이용할 경우에는 “임진왜란”을 키워드로 하여 검색한 후 검색된 수십 수백개의 문서중에서 일일이 몇 년도에 일어났는지에 관한 정보를 찾아보아야 한다. 이러한 문제점은 “KEDO는 무엇의 약자인가?”, “여의도의 면적은 얼마인가?”라는 물음에 대해서도 공히 존재한다.

이와 같이 키워드 검색은 본문(full-text) 검색에 있어 상당히 비효율적일뿐 아니라 오히려 검색자에게 혼돈을 가져다 줄 수도 있다. 그러나, 의문의 초점을 고려하여 정보를 검색하도록 설계된 본 의문문 질의 시스템은 검색자가 요구하는 의문의 초점에 대해서 주목하여 정보를 검색하기 때문에 불필요한 정보를 검색할 가능성을 최소화시킬 수 있어 본문 검색에서 매우 효과적이다.

이러한 의문문 질의 시스템은 다음과 같은 상황에서 매우 효율적으로 활용될 수 있다.

- (1)신문기사, 백과사전, 법원판례등과 같이 검색하고자 하는 정보가 자연어 상태로 존재할 경우
- (2)동일하거나 유사한 내용의 정보집합에 대해 세부적이고 구체적인 정보 검색을 원할 경우
- (3)검색자가 키워드 검색에서 사용하는 검색식에 대해 전혀 모르거나 익숙치 않을 경우
- (4)동사와 형용사 같은 용언을 색인어로 하여 정보를 검색해야 할 경우

2-3. 의문문을 질의어로 채택할 경우의 고려 사항

의문의 초점을 고려하여 정보를 검색하도록 설계된 본 의문문 질의 시스템은 의문문 그 자체를 질의어로 채택하여 정보를 검색한다. 의문문을 질의어로 채택할 경우, 사용자의 검색 편의성이나 정보검색의 효율성은 극대화시킬 수 있겠지만 일반 키워드 검색에 비해 고려해야 할 사항들이 적지 않다. 고려되어야 할 사항들을 다음과 같이 정리해 보았다.

- (1)의문의 초점에 대한 정확한 분석을 통해 시스템이 반드시 해당 의문의 초점에 맞추어 응답문을 검색할 수 있도록 설계해야 한다.

- (2)의문사 각각에 대해 통사론적, 어휘론적 분석을 통해 의문문 내에서 사용되는 의문사의 역할을 시스템이 정확히 파악할 수 있도록 한다.
- (3)일반적인 자연어 질의 시스템에서와 마찬가지로 형태소 분석을 비롯한 문장 분석 등의 과정도 필수적으로 요구된다.
- (4)다양한 의문형 어미와 검색자의 다양한 질의어 사용에 대처할 수 있도록 충분한 문형정보와 어미 정보를 시스템이 갖고 있어야 한다.

3. 의문문 유형과 의문의 초점

3-1. 의문문의 유형과 질의어

의문문의 유형은 발생한 의문에 대해 가지는 화자의 불확실성의 정도에 따라 대체로 설명의문문(내용의문문), 선택의문문, 판정의문문(찬부의문문), 확인의문문, 수사의문문 등 5개로 나누어 볼 수 있다.[6][7] 일상 대화 속에서는 이들 5개의 의문 유형이 다양하게 사용될 수 있지만 정형화된 텍스트에서 특정 정보를 검색하는데 필요한 의문문의 유형은 대부분 설명의문문에 국한되므로 본 논의도 설명의문문에 초점을 맞추어 전개하고자 한다.

<p>ㄱ. 설명의문문 : 예)여의도의 면적은 얼마입니까?</p> <p>ㄴ. 선택의문문 : 예)기차가 빠릅니까, 비행기가 빠릅니까?</p> <p>ㄷ. 판정의문문 : 예)일본은 섬나라입니까?</p> <p>ㄹ. 확인의문문 : 예)고래는 포유류지, 그렇지?</p> <p>ㅁ. 수사의문문 : 예)철수가 가겠어?</p>
--

그림 1. 의문문의 유형별 예문

그림 1에서 보는 바와 같이 설명의문문을 제외한 나머지 4가지 의문문의 유형은 객관적인 자료나 정보에 대한 물음으로 사용되기 힘든 면이 있다는 것을 예문의 통해서도 쉽게 확인할 수 있다. 또한, 선택의문문, 확인의문문, 수사의문문은 기존의 대화 맥락을 통해서만 짐작해서 알 수 있는 화용적 전제(pragmatic presupposition)가 반드시 고려되어야 하므로 질의어로서 사용되기 힘든 측면이 있다. 한편, 판정의문문은 기능적인 측면에서 설명의문문을 질의어로 채택하였을 경우보다 검색자 입장에서 효과적이지 못하다고 판단되기 때문에 질의어로 사용되기 곤란한 것이 사실이어서 의문문 질의어로는 설명의문문이 가장 일반적으로 사용될 수 있다.

설명문의문에는 대체로 의문사가 하나의 의문의 초점으로 나타나는 것이 통례이다. 다수의 의문사가 나타나는 점의문문의 유형도 있으나 언어 수행상의 난점으로 인하여 두 개 이상의 의문사가 동시에 출현하기는 실제로 드문 것이 사실이다.[6] 또한, 의문의 초점도 일반적으로 미정사에 놓이게 되어 해당 문장은 이 미정사에 대하여 설명하는 새로운 정보(신정보)를 포함한 문장을 응답문으로 요구하게 된다.

3-2. 의문의 초점

의문의 초점은 ‘발생한 의문에 대해 발화된 화자의 불확실성’으로, 의문문 내의 어떤 성분도 의문의 초점이 될 수 있다.[6] 동일한 어휘 항목(lexical item)이라 하더라도 의문의 초점의 부여 여부에 따라 의문사와 부정사로 결정되므로 설명의문문은 의문의 초점과 불가분의 관계에 있다.[8][9][10]

국어학에서의 ‘초점’은 「주제-설명(Theme-Rheme)」구조를 정보구조면에서 살펴볼 때, 설명 부분에서 신정보를 전달하는 요소를 초점(focus)라고 보고 있다.[9] 즉, 의문문에서의 ‘초점’은 발화자가 전혀 모르고 있는 정보에 대한 요구로 파악할 수 있다. 즉, 의문의 초점은 설명의문문에서 특정 단어에 부여되어 신정보를 요구하는 역할을 한다.

한편, 질의-응답의 관계를 「주제-설명」구조에서 한단계 발전시켜 「전제-대답」의 관계로 파악하기도 한다. 즉, 서술문은 구체적 ‘물음’에 대한 구체적 ‘대답’이라고 보고 그 대답을 요구한 물음으로서의 의문문을 대답의로서의 서술문에 대한 하나의 「전제」라고 가정한다.[11] 이러한 전제는 본 의문문 질의 시스템에서 응답문을 검색하는 키워드의 역할을 한다.

- ㄱ. 영희가 어제 신촌에서 안경을 샀습니다.
- ㄴ. 누가 어제 신촌에서 안경을 샀습니까?
- ㄷ. 영희가 언제 신촌에서 안경을 샀습니까?
- ㄹ. 영희가 어제 어디에서 안경을 샀습니까?
- ㅁ. 영희가 어제 신촌에서 무엇을 샀습니까?
- ㅂ. 영희가 어제 신촌에서 안경을 어찌했습니까?
- ㅅ. 영희가 뭘 했습니까?
- ㅇ. 누가 뭘 했습니까?

그림 2. 초점에 따른 다양한 물음

의문의 초점은 초점을 두는 위치에 따라 의문문의 형태를 크게 변화시킬 수 있다. 그림 2에서 보는 바와 같이 “영희가 어제 신촌에서 안경을 샀습니다.”(ㄱ)라는 문장에 대해 의문의 초점을 어디에 두느냐에 따라 여러개의 의문문을 구성할 수 있다.

시간 속성에 중점을 둘 경우 ‘초점’은 ‘언제’가 되고(ㄷ), 장소 속성에 중점을 둘 경우 ‘초점’은 ‘어디’가 되며(ㄹ), 행동 주체에 중점을 둘 경우 ‘초점’은 ‘누구’가 될 것이다(ㄴ). 신정보와 구정보의 관점에서 보면 시간, 장소와 관계없이 ‘언제’와 ‘어디’, ‘누구’ 등은 신정보 단위로 볼 수 있다. 즉, ‘언제’, ‘어디’, ‘누구’ 등이 신정보를 요구하는 의문의 초점이 되고 나머지는 구정보로 처리된다.

4. 의문문 질의어를 위한 의문사 분석

4-1. 의문사의 분류

의문사들은 응답의 논리적 범주에 대하여 행하는 제약에 근거하여 크게 ‘한정의문사’와 ‘비한정의문사’로 나누어진다.[12] 이러한 분류는 지극히 언어학적인 분류에 해당되지만 실제로는 수식가능 여부와 조사후접 여부 등 여러 문법상황을 고려하여 재분류되어야 할 필요가 있다.

그림 3은 본 시스템에서 사용하기 위해 편의상 재분류한 의문사의 분류표이다. 이것은 의문문 질의 시스템을 구현함에 있어 정보검색시 질의어와 검색대상문에 나타나는 한국어 정보들의 품사 표지와 어휘 결합성에 주목하여 의문사를 크게 5가지로 분류한 것이다. 이는 응답문 검색을 위한 검색대상 어휘의 범위도 고려하여 설정된 것이다.

- 1류 : ‘의문사(관형사)+명사’ 형태의 의문 유형에 사용되는 의문사
예)어느, 어떤, 무슨, 몇
- 2류 : 명사, 부사(조사의 결합 없이 단독으로) 등 2갈래로 처리할 수 있는 의문사
예)언제
- 3류 : 명사, 부사(조사와 결합해서) 등 2갈래로 처리할 수 있는 의문사
예)어디
- 4류 : 반드시 명사로만 사용되는 의문사(검색대상도 반드시 명사 혹은 명사에 대한 설명)
예)누구, 무엇(검색대상 어구의 마지막 어구는 반드시 해당 의문사와 동일한 속성)
- 5류 : 동사구 전반이나 어구 전체를 검색대상으로 하는 의문사
예) 왜, 어떻게

그림 3. 본 시스템을 위한 의문사 분류

4-2. 의문사와 의문의 초점

본 절(本節)에서는 전 절(前節)에서 제시된 ‘5분류’에 의해 분류된 각각의 의문사들을 의문의 초점과 연관시켜 문장 내에서 어떻게 기능하는지에 대해 그림 4와 같이 정리해 보았다. 그림 4는 각각의 의문사를 기본문형에 대해서 품사 표지만을 사용하여 그 구조를 정리한 것이다.

1류는 그 예에서 보는 바와 같이, 관형사로서 사용된 의문사들로 구성된다. 따라서, 후접하는 품사는 명사이며 후접하는 명사는 모두 의문사의 수식을 받는다. ‘어느 나라’, ‘어떤 사람’, ‘무슨 요일’, ‘몇 일’ 등에서 ‘나라’, ‘사람’, ‘요일’, ‘일’ 등은 의문사의 수식을 받으며 응답문이 되는 검색대상어의 범위를 제한하는 기능을 한다. 예를 들어, ‘몇 일’이라고 물었을 때에는 ‘1997년’이라는 어구는 같은 시간속성의 어구지만 검색대상이 되지 않고 ‘12일’ 등과 같이 일자(日字)를 의미하는 어구만이 검색대상이 될 수 있다. 즉, 의문사에 후접하는 명사인 ‘나라’, ‘사람’, ‘요일’, ‘일’ 등은 의문사는 아니지만 반드시 의문사와 함께 고려되어야 할 요소이다.

- 1류 :
- ①명사+조사+의문사*+명사+(부사격)조사+용언
예)입진왜란은 몇 년에 일어났습니까?
 - ②명사+조사+의문사*+명사+서술격조사
예)불교는 어떤 종교입니까?
- 2류 :
- ①명사+조사+의문사*+용언
예)서울 올림픽은 언제 개최되었습니까?
 - ②명사+조사+의문사*+서술격조사
예)개천절은 언제입니까?
- 3류 :
- ①명사+조사+의문사*+(부사격)조사+용언
예)KOEX는 어디에 있습니까?
 - ②명사+조사+명사+조사+의문사*+서술격조사
예)커피의 원산지는 어디입니까?
- 4류 : 누구, 무엇
- ①명사+조사+의문사*+조사+용언
예)안중근은 누구를 암살했습니까?
 - ②명사+조사+명사+조사+의문사*+서술격조사
예)필리핀의 대통령은 누구입니까?
- 5류 : 왜, 어떻게
- ①명사+조사+의문사*+용언
예)감은 왜 떨어습니까?

그림 4. 5분류에 의한 의문사와 의문의 초점

2류는 부사와 명사로 사용되는 의문사로 시간 속성에 관계된다. 3류 역시 부사와 명사로 사용되며 장소 속성에 관계된다. 4류는 사람 속성(누구)과 사물 속성(무엇)에 관계되며 5류는 어구전체나 문장전체를 검색 대상으로 한다.

5. 의문문 질의 시스템 설계

5-1. 전체 시스템 진행 과정

의문문 질의 시스템이 기타의 키워드 정보검색 시스템이나 자연어 정보검색 시스템과 차별화되는 점은 의문문 질

의어에 대한 분석하는 질의어 분석 프로세서와 의문의 초점에 주목하여 검색하는 응답문 검색 프로세서가 별도로 존재한다는 데 있다. 이러한 차이점에 주목하여 본 의문문 질의 시스템의 진행 과정을 그림 5와 같이 나타내 보았다.

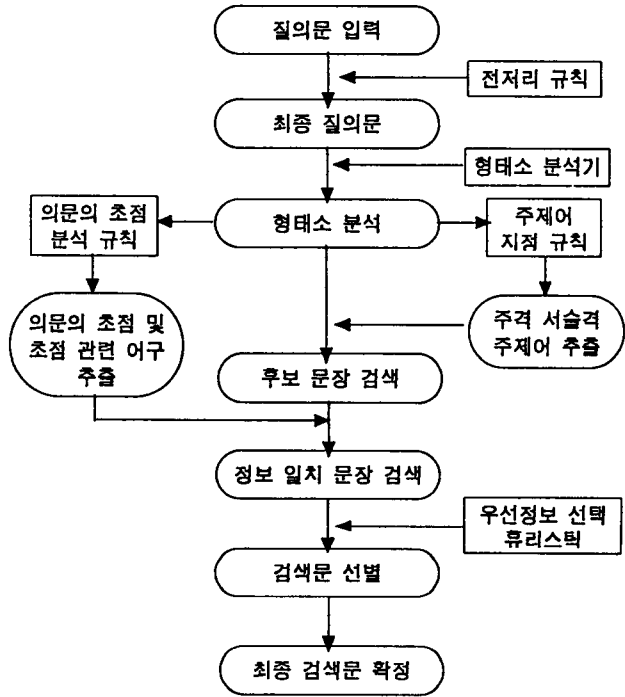


그림 5. 의문문 질의 시스템의 진행 과정

5-2. 질의어 분석 프로세서

의문문 질의 시스템에서 가장 핵심이 된다고 할 수 있는 부분이 바로 이 질의어 분석 프로세서이다. 이 부분에서는 질의어의 내용을 분석하여 주격 주제어와 서술격 주제어를 추출할 뿐만 아니라 응답문 검색시 의문의 초점에 주목하기 위해 의문의 초점과 초점 관련 어구를 추출해 내는 기능을 한다.

5-2-1. 주제어 분석

주제어는 주격 주제어와 서술격 주제어로 나누어 분석한다. 주격 주제어는 통사구조에 따라 2개 이상이 나타날 수 있으므로 보조 주제어를 설정하여 이중주어구문과 겹주어구문들에 대한 처리가 가능하도록 설계하였다. 따라서, 명

사 색인어는 1개의 질의어에 대해 2개까지 사용할 수 있도록 하였지만 필요로 따라서는 3개 이상도 생성이 가능하도록 설계하였다.

```
Q102;S0=="누구";P1=="KJ1*";G1=="KJJs";
P2=="KA*";.....:[ST1]=S2;.....:[DT1]=S4;
// 누가 명왕성을 발견했습니까?
```

그림 6. 주제어 지정 규칙의 예

그림 6에서와 같이 본 의문문 질의 시스템에서는 주제어 지정 규칙을 통해 주제어를 지정한다. [ST1]은 주격 주제어를, [DT1]은 서술격 주제어를 가리킨다. 규칙Q102에 의해 ‘명왕성’은 주격 주제어(ST1)로, ‘발견하다’는 서술격 주제어(ST2)로 추출되었다. 응답문 검색 프로세서에서 이 두 주제어를 키워드로 하여 후보 문장을 검색한다.

5-2-2. 의문의 초점 분석

질의어 분석 과정에서 주격 및 서술격 주제어 추출뿐만 아니라 의문의 초점 및 초점 관련 어구 추출도 동시에 이루어진다. 의문의 초점은 주로 의문사 위주로 추출되며 초점 관련 어구는 문장 구조에 따라 조금씩 다르지만 일반적으로 검색 대상문에서의 검색 초점이 될 어구를 지정하여 추출하게 된다. 단, 초점 관련 어구는 문형에 따라 나타나지 않을 경우도 있다.

```
Q115;P0=="KA*";P1=="KJ1*";G1=="KJJs";
S2=="몇";P3={"KA1*"|"KA4*"};P4=="KJ1*";
P5=="KD*";.....:[QF1]=S2:[QF2]=S3;.....;
// 임진왜란은 몇 년에 일어났습니까?
```

그림 7. 의문의 초점 분석 규칙의 예

그림 7은 의문의 초점과 초점 관련 어구를 추출하는 규칙의 한 예이다. [QF1]은 의문의 초점을, [QF2]는 초점 관련 어구를 가리킨다. 규칙 Q115에 의해 ‘몇’은 의문의 초점(QF1)으로 ‘년’은 초점 관련 어구(QF2)로 추출되었다.

5-3. 응답문 검색 프로세서

응답문 검색 프로세서에서는 질의문 처리 프로세서의 규칙들에 의해 추출된 여러 정보들(주격 주제어, 서술격 주제어, 의문의 초점, 초점 관련 어구)을 통해 응답문을 검색한다. 응답문 검색은 후보 문장 검색, 정보 일치 문장 검색, 검색문 선별 등 3단계를 거쳐 최종 검색문을 확정하게 된다.

5-3-1. 응답문 검색의 과정

‘후보 문장’은 해당 한국어 문서에서 주격 주제어와 서술격 주제어를 동시에 포함하고 있는 문장을 1차적으로 검색한 것이다. 이때 본 시스템은 이미 추출된 주격 주제어와 서술격 주제어를 키워드로 해당 정보를 검색하며 이들 각각의 유의어도 검색 대상에 포함시켜 검색한다.

‘정보 일치 문장’은 1차적으로 검색한 ‘후보 문장’에서 초점 관련 어구에 부합하는 어구가 발견되고 이것이 이미 추출된 의문의 초점과 일치하는 문장을 말한다. 즉, 이 문장은 검색자가 검색하고자 한 정보일 가능성이 매우 높은 문장이다. 그러나, 모든 요소가 일치한다고 해서 무조건 검색자가 원하는 정보는 아니다. 검색자의 검색 요구에 보다 접근하는 정보를 선별하기 위한 ‘우선정보 휴리스틱’을 통해 어떤 것이 보다 검색자에 요구에 맞는 검색문인지 휴리스틱으로 분석하여 우선되는 정보를 최종 검색문로 확정하게 된다.

5-3-2. 예문을 통한 검색 과정 분석

본 절에서는 5-1에서 살펴본 전체 시스템 진행 과정에 대비시켜 실제 질의문이 각 과정에서 어떻게 처리되는지에 대해 살펴보고자 한다. 각 단계는 그림 5의 단계에 기준해서 진행되며 질의문으로는 “임진왜란은 몇년에 일어났습니까?”을 사용하였다.

- (1)질의문 입력 : 임진왜란은 몇년에 일어났습니까?
- (2)전처리 규칙 : [띄어쓰기 처리]
- (3)최종 질의문 : 임진왜란은 몇 년에 일어났습니까?
- (4)형태소 분석 : [문장종류 분석 포함]
 - 임진왜란[명사]+은[격조사]+몇[관형사]+년[명사]
 - +에[격조사]+일어나다[동사]
- (5)주제어 지정 규칙 :
 - [STP1]=임진왜란, [DTP1]=일어나다

(6)의문의 초점 분석 규칙 :

[QF1]=몇, [QF2]=년

(7)후보 문장 검색 :

- [1]임진왜란은 1592년에 일어나 오랜기간 동안 조선 백성들을 힘들게 했다.
- [2]조선 선조 26년에 일어난 임진왜란은 뒤에 일어난 병자호란과 함께
- [3]1592년 임진왜란이 일어나자 의승병을 모아 순안으로 가서 휴정과 합류하여
- [4]전라남도 수군 절도사로 부임 선조 26년(1592년) 임진왜란이 일어났다.

(8)정보 일치 문장 검색 :

- [1]임진왜란은 1592년에 일어나 오랜기간 동안 조선 백성들을 힘들게 했다.
- [3]1592년 임진왜란이 일어나자 의승병을 모아 순안으로 가서 휴정과 합류하여

(9)검색문 선별 : [우선정보 선택 휴리스틱]

(10)최종 검색문 확정

- [1]임진왜란은 1592년에 일어나 오랜기간 동안 조선 백성들을 힘들게 했다.

6. 결론

본 논문에서는 의문의 초점을 고려한 자연어 기반의 의문문 질의 정보검색 시스템을 제안하였다. 본 시스템은 기존의 키워드 검색 시스템이 가지는 불편함을 해소하고 기존의 자연어 정보검색의 개념을 한층 더 발전시켜 자연어 정보검색의 새로운 모델을 제시하였다.

본 의문문 질의 시스템은 질의어 분석 프로세서와 응답문 검색 프로세서로 구성되어 있다. 질의어 분석 프로세서에서는 1차적으로 주격 주제어와 서술격 주제어로 추출하고 2차적으로 의문의 초점과 초점 관련 어구를 추출한다. 응답문 검색 프로세서에서는 질의어 분석 프로세서에서 추출한 주격 주제어와 서술격 주제어를 이용하여 1차적으로 이들 주제어를 포함하고 있는 문장을 후보 문장으로 간주하여 검색한다. 그 다음 단계로 의문의 초점과 초점 관련 어구를 후보 문장 중에서 각각 검색하여 이들 정보와 일치하는 문장을 검색하고 이들 중 검색자의 검색 요구에 가장 부합하는 문장을 최종 검색문으로 확정한다.

이상에서 살펴본 바와 같이 한국어 의문문 자체를 질의

어로 채택할 경우, 복잡한 처리 과정을 거치게 되지만 검색자 입장에서 매우 편리하고 정확한 검색을 할 수 있다는 장점이 있음을 알았다. 앞으로도 계속해서 복잡한 한국어 문장과 의문사의 다양한 형태 변화, 문장 요소 생략 등에 대응할 수 있는 의문문 질의 시스템에 대한 연구가 필요할 것이라 본다.

참 고 문 헌

- [1] 안운애, 류근호, “지식기반 정보검색 시스템의 질의처리 모델”, 컴퓨터 과학연구, 제3권 제1호, 1995.
- [2] 박세영, “멀티미디어 정보검색에서의 한국어 정보처리”, 한국 정보과학회 논문지, 제12권 제8호, 1994.
- [3] 함호중, “답화문의 질의어 생성시스템의 구현”, 한양여자전문대학교 논문집, 제18집, 1995.
- [4] 채진석, 이석호, “프레임 기반 질의 분해 기법을 사용한 한국어 질의 시스템의 설계 및 구현”, 한국 정보과학회 논문지, 제3권 제4호, 1997.
- [5] 정보시대, “인터넷 검색엔진 주무르기”, 월간 인터넷 7월호, 1997.
- [6] 성시형, “의문의 초점과 의문문의 유형에 대한 연구”, 한양대 석사학위 논문, 1987.
- [7] 서정수, “국어 문법의 연구”, 한국문화사, 1990.
- [8] 서정목, “국어 의문문 연구”, 탑출판사, 1987.
- [9] 서정수, “현대 국어문법론”, 한양대학교 출판원, 1996.
- [10] 장석진, “문답의 화용상”, 어학연구, 제11권 제2호, 1975.
- [11] 김일웅, “국어의 ‘주제-설명’ 구조”, 국어통사론, 진명문화사, 1985.
- [12] 김광해, “국어의 의문사에 대한 연구”, 국어학12, 1983.