

구문구조를 이용하여 정답을 추출하는 질의응답 시스템

충북대학교 컴퓨터공학과
이 대 연[†] · 서 영 훈

A Question Answering Using Syntactic Structure for Answer Extraction

Dae-Yeon Yi, Young-Hoon Seo

Department of Computer Engineering, Chungbuk National University, Chungbuk, Korea

요 약

본 논문에서는 질의문 내에 포함된 동사를 중심으로 한 질의어 확장 및 정답 추출 기법을 이용한 질의 응답 시스템에 대해 기술한다. 질의 응답시스템 전체의 과정에서 동사는 하나의 정보를 표현하는 중요한 요소로 활용하며, 동사에 대한 활용은 구축된 동사구문 사전의 정보를 이용한다. 동사구문 사전은 동사의 일반적인 표층형태와, 각 문장 성분들의 의미속성, 유의동사 등의 정보를 담고 있다. 또한 동사 구문사전의 활용에서의 동사 모호성을 배제하고, 효율을 높이기 위해 약 3만 어휘의 명사 의미 사전을 사용한다. 명사 의미사전은 구문사전 내에 사용된 의미분류로 나누어져 있으며, 유의명사 및 국어사전 상의 뜻 풀이말을 포함하고 있다. 질의문 및 각 후보 문장에 대한 구문분석은 구문사전 내에 나타난 품사 별 의미속성과, 문법 형태소의 격 정보를 이용한 격 구조를 활용하였다. 논문 중에는 일반적인 질의 응답 시스템의 3단계에 맞추어 구문사전 활용 및 구문분석의 수행 단계를 보이고 마지막에 각 기법의 정확도를 보였다.

서 론

현재의 일반적인 정보검색 시스템들은 사용자가 제시하는 단어 또는 논리 연산자와 결합된 단어집합을 입력으로 하여 정답이 포함되어 있을 가능성이 높은 문서를 제시하고 사용자가 제시된 문서 내에 정답이 포함되어 있는지의 여부를 판단하도록 하고 있다. 그에 반해 질의 응답 시스템의 경우는 일반 사용자들에게 익숙한 자연어 문장 형태의 질의를 받아, 정답 또는 정답이 포함된 문장 등을 제시한다. 그로 인해 사용자에게 보다 높은 편의를 제공한다는 점에서 그 요구가 증가하고 있다.

이러한 요구의 증가에 따라 최근 국내외 연구도 활발히 진행되고 있다. TREC(Text Retrieval Conference)에서는 1999년 TREC-8을 시작으로 질의응답시스템들에 대한 평가대회를 열고 있으며, 국내외 많은 연구소와 대학에서도 관련된 연구를 수행하고 있다.

질의응답시스템은 크게 질의문 분석과 정답추출이 시스템의 핵심을 이루고 있다. 기존 연구 중 질의문 분석에 관하여서는 의문사 등의 키워드 추출 및 가중치 부여와 개체 인식 기법에 관한 연구가 이루어져왔으며,¹⁻³⁾ 정답추출 부분에서는 고유명사에 대한 인식과 추정명사 등에 대한 의미속성 결정 등을 위한 대규모 지식베이스 구축^{4,5)} 및 활용 방법과 함께 정확한 정답추출을 위한 구문분석에 관한 활발한 연구가 이루어지고 있다.⁶⁻⁹⁾ 영어권에서는 워드넷을 이용한 지식베이스 관련연구¹⁰⁻¹²⁾가 이루어 지고 있으나 국내에서는 워드넷과 같은 언어자원의 부족으로 영어권에 비해 낮은 수준의 실험결과를 보이고 있다.

본 논문에서의 질의응답 시스템은 대용량 지식베이스의 사용을 배제하기 위해 구문분석에 비중을 두고 설계되었다. 논문은 문장에 대하여 구문사전을 이용한 부분 구문분석을 수행하고, 그 결과를 질의응답 시스템에 적용한 내용을 기술한다. 구문사전은 동사를 중심으로 한 한국어 문장성분들에 대한 자질 값으로 구성되어 있는 주요 언어자원 및 지식 자원이다. 구문사전은 동사를 중심으로 하여 동사에 종속적인 명사들에 대한 의미속성 정보로 구성된다. 구문구조를 이

[†]E-mail : hyperion@dcentlp.chungbuk.ac.kr

용한 구문분석은 동사와 결합된 어미의 종류에 따라 문장 성분의 문장 내 품사 탐색 범위를 유동적으로 적용시킨다. 구문사전 내에 포함된 품사 별 의미속성 분류를 이용하여 별도의 개체명 인식과정은 생략하였다. 논문 내에서의 실험은 구문사전을 이용한 부분 구문분석의 정확도 및 질의 응답 시스템에서의 응용가능성을 측정하였다.

시스템의 구조

본 논문에서 구현한 질의 응답 시스템은 사용자의 질의문을 받아 문서집합 내에서 순위가 부여된 단답식 정답을 찾아서 제공한다. 정답은 단일문서 내에 하나 이상 존재할 수 있고, 복수의 문서들에 다수가 존재할 수 있기 때문에 하나 이상의 정답들이 추출될 수 있다. 하나 이상의 정답이 추출될 경우 각 정답들은 순위화되어 보여진다. 다수의 정보들로부터 추론되는 정답을 필요로 하는 질의문은 제외되었으며, 또한 정답이 없는 질의문도 제외하여 실험하였다.

시스템에 대한 전체적인 구성은 Fig. 1과 같다.

본 시스템은 크게 3부분으로 나누어진다. 사용자의 질의를 어를 구조화된 구문구조로 변환하고 확장하는 부분, 확장된 질의 구조집합을 이용하여 문서집합으로부터 후보문장을 검색하는 부분, 마지막으로 얻어진 후보 문장으로부터 순위화

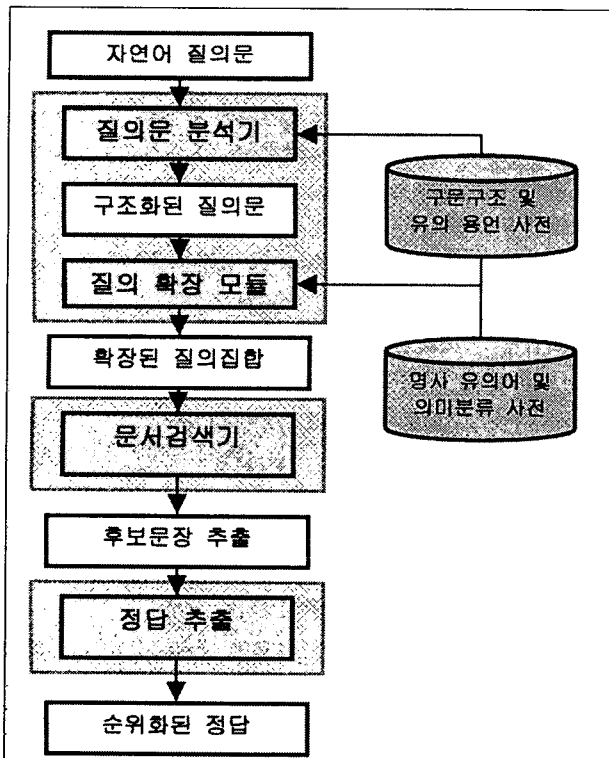


Fig. 1. 전체 시스템의 구성도.

된 정답을 추출하는 부분이다.

1. 구문구조 사전의 구축

본 논문에서 구현한 시스템의 가장 중요한 자원으로 사용되는 구문사전은 의미정보를 포함한 한국어 동사의 구문구조 사전이다. 이 사전은 한국어 동사의 통사적 의미적 용례에 따라 구축되었다. 기본적으로 한국어 동사 구문사전(홍재성 등, 1997)을 참고하였으며, 거기에 유의 동사목록을 첨가하는 형태로 구축되었다. 동사 구문 사전은 동사의 중의성을 해소하기 위해 단일 동사일 지라도 주어 또는 목적어 등 각 품사가 갖는 명사의 의미분류가 다르면 각각을 다른 엔트리로 구축하였기 때문에, 구문사전에서 하나의 구문을 검색하기 위해서는 동사키워드가 필수적으로 필요하며, 하나 이상의 품사에 대한 의미분류가 선택적으로 필요하다.

Fig. 2는 하나의 동사구문의 내용을 나타낸다. 하나의 동사구문을 정의하기 위한 구성요소는 구문사전을 참고하여 다음 몇 가지로 정의하였다.

- 동사 : 구문구조의 동사.
- 용례 : 표현에 사용되는 품사와 어순.
- 품사정보 : 문장의 각 품사가 갖는 의미속성.
- 유의동사 : 질의응답 시스템의 확장성을 위한 추가 정보

유의 구문구조에 대한 정보를 구축할 때, 동사가 바뀔 때 따라 어순이 바뀔 수 있는 점에 대해서는 구문구조간 품사 매칭 테이블을 추가함으로써 처리의 효율을 높였다.

Fig. 2에서 '키워드' 필드는 각 구문의 동사를 나타내고, '품사시퀀스'는 가장 일반적인 형태의 품사 출현 순서를 나타낸다. 이 시퀀스는 한국어 동사구문 사전에서 인용하였으며, 실제 문장에서 변형되어 나타날 수도 있지만, 보다 낮은 가중치를 받게 된다. '품사 별 의미분류'는 구문의 각 품사에 올 수 있는 명사의 의미 속성을 정의한다. 각 명사에는 하나 이상의 의미 속성이 적용될 수 있으며, 그 명사의 분류는 한국어 동사 구문사전에 나타나는 의미분류를 사용하였

키워드:	품사 시퀀스			
짓다	NO N3 에 N1 을 V			
품사별 의미분류				
주어-N0	직목-N1	간목-N2	부사1-N3	
인물 단체	건물	X	장소	
유의동사			피동사	
새우다	지어지다			
건조하다	N1 NO N3			
건축하다	(NO N2 에 의해 N1에 V)			

Fig. 2. 구문사전의 예.

다. 여기에 사용된 의미분류는 구문구조에 의존적인 의미분류를 적용하였다. 마지막으로 유의동사 및 피동사 필드가 있다. 이 두 필드에는 키워드의 동사의 유의 동사들이 들어간다. 유의 동사는 2개의 형태로 나누었다. 유의 동사 필드에 있는 동사들은 시퀀스와 품사 별 의미분류를 공유하는 동사들이지만, 피동사 필드에 나타난 동사들은 원래의 동사와 다른 품사시퀀스를 갖는다. 품사시퀀스가 다른 두 동사의 품사 매칭을 위하여 품사 매칭 테이블을 포함하였다. 그림에 나타난 정보 “N1 NO N3(N0 N2 에 의해 N1에 V)”가 나타내는 의미는 동사 ‘짓다’를 ‘지어지다’로 확장하는 경우, ‘짓다’의 N1, NO, N3가 ‘지어지다’의 N0, N2, N1으로 매치된다는 것을 의미한다.

‘짓다’의 경우 크게 3가지 외형적 구문 형태로 나뉘어지며, 각 형태별로 3, 4, 5개의 의미 집합으로 구성되어 있다. ‘짓다’라는 동일 표제어 아래 총 12개의 구문 구조가 들어가게 된다.

2. 명사 의미사전

각 문장에 대한 구문의 적용을 결정하기 위해서는 표제어로 사용되는 동사 외에도 기타 품사에 대한 추가적인 의미 정보가 필요할 수 있다. 이러한 의미정보를 얻기 위해 명사의 의미사전이 사용된다. 명사의미사전은 약 30,000여 개의 일반 명사들을 8개의 의미분류로 나누어 놓았으며, 각 명사들은 하나 이상의 의미 집합에 포함될 수 있다.

명사 의미사전에 포함된 명사의 의미분류는 인물, 단체, 장소, 건물, 사물, 추상, 교통, 학문, 음식, 질병의 8가지이며, 이 분류는 한국어 동사구문사전(홍재성 등, 1997)에서 발췌하였다. 또한 시스템의 확장성을 위해 그 이상의 의미분류는 포함하지 않았다. 추가적으로 질의 확장을 위한 동의명사 목록과 국어사전에서 추출된 뜻 풀이말이 포함되었다.

질의 분석, 확장 및 문서검색

1. 질의 분석 및 확장

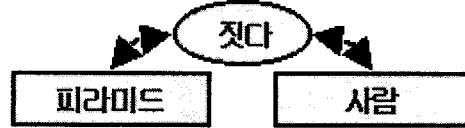
본 논문에서는 자연어 질의문으로부터 크게 두 가지 성분을 추출하여 사용한다. 하나는 사용자가 원하는 정답의 패턴(의미속성)이고, 다른 하나는 그 정답의 범위를 한정하는 요소(키워드)들의 집합이다.

다시 말해서 질의문의 경우, 정답의 패턴을 나타내는 명사 또는 대명사가 나타나고 그 명사의 범위를 한정 짓기 위

피라미드를 지은 사람은 누구인가?		
정답패턴	사람	
한정요소	피라미드	짓다

한 품사집합이 나타나게 되는데, 그 품사집합은 동사와 명사의 형태로 나타날 수 있다. 위의 질의문을 바꿔서 생각해 보면 정답패턴 ‘사람’과 한정요소 ‘피라미드’가 ‘짓다’라는 의미 관계로 결합된 형태로 볼 수 있다.

피라미드를 지은 사람은 누구인가?



질의문에 나타나는 사람을 정의하기 위해서는 명사, ‘피라미드’가 ‘짓다’ 또는 그 유사 의미로 연결된 ‘사람’을 정의해야만 함을 의미한다. 각 단어들의 의미관계를 ‘쓰다’에 초점을 맞추어 구문 구조로 표현할 수 있다.

위의 질의문을 구조화된 구문구조로 변환하면 다음과 같다. 질의문 구조의 대표단어는 동사 ‘짓다’가 되며, 각 중요명사들은 피라미드와 사람이 된다. 이 중에서 사람이 정답이라는 정보는 패턴으로부터 추출 가능하다.

Fig. 3은 질의문의 구조화된 표현이다. 동사 ‘짓다’의 구문사전에서 유도된 구조에 각 단어들을 할당하고 정답패턴에는 정답표지를 한다.

‘짓다’의 총 12개의 구문 구조 중 어느 것을 결정할 것인가의 문제는 명사 ‘피라미드’의 의미분류를 결정할 수 있는가에 달려있다. 피라미드가 건축물이라는 의미분류라는 것을 의미사전으로부터 얻어 낼 수 있다면 위에 보인 구문구조를 적용할 수 있지만 그렇지 않은 경우, ‘짓다’의 구문 12개 모두에 구문구조를 생성하게 된다.

생성된 구문구조들에 대하여 구문사전의 유의어 사전을 이용해 구문구조 확장을 실시한다. 위에 보인 ‘짓다’ 구문 구조의 경우 유의어 구조로 ‘세우다’, ‘건조하다’, ‘건축하다’, ‘지어지다’의 구조를 얻을 수 있다. 각 구조에는 질의 구조 원형의 명사들(피라미드)이 할당된다. 또한 질의 구조 원형에 할당되어있던 명사들에는 각각의 동의어가 연결되게 된다.

2. 문서검색 및 후보문장 추출

질의문의 확장된 구조집합을 가지고 문서를 검색하고 확장된 동사가 사용된 문장을 추출한다. 문서검색에 사용되는 정답문서 집합은 MATEC 99(MATEC 99, 1999)에서 사용된 품사태그 부착된 말뭉치를 사용하였다. 일반 웹 문서

짓다		
주어 (인물/단체)	목적어 (건물)	부사 (장소)
ANSWER SLOT	피라미드	IGNORE

Fig. 3. 질의문의 구문구조.

는 HTML 문서의 특성상 구조화된 비율이 높다. 또한 구조화된 많은 문서들의 경우, 문장 내의 주요품사들에 대한 생략이 빈번하여, 실험을 위한 정답문서로 적합하지 않았다. 위에 보인 질의어에 대해 문서검색을 해본 결과, 다음의 2개의 문장이 추출되었다.

질의문	피라미드를 지은 사람은 누구인가?
후보문장1	마야족들은 저지대에서 생활했기 때문인지, 여러 지역에 흙을 높이 쌓아 피라미드를 세웠다.
후보문장2	그들은 피라미드의 본체를 쌓기 전에 일정 기간을 두고 항상 소형 피라미드를 먼저 세우고, 그것이 어느 정도 견고해지면 다시 재료를 겹쳐 피라미드 본체를 완성하는 특이한 방법을 썼다.

후보문장에서의 정답추출

질의문의 문서검색결과로 얻어진 3개의 후보문장으로부터 정답을 추출하기 위해 Fig. 4와 같이 각 문장을 구문 구조로 변환한다.

변환된 정답후보 문장과 질의문 구문 구조와 비교 함으로써 정답을 추출할 수 있다. 구문구조 '짓다'의 유의어 구문 정보에서 유도된 구문구조 '세우다'의 구문구조에 질의원형에서 유도된 명사 '피라미드'가 결합된 형태의 구문구조는 Fig. 5와 같다.

질의문의 구문구조에서 나타난 ANSWER는 이 SLOT에 들어가는 어휘가 질의문의 정답이라는 뜻이며, 부사 필드의 IGNORE는 질의문에 장소 또는 때에 대한 단서가 주어지지 않았으므로 때나 장소를 나타내는 부사정보는 무시하라는 의미가 된다.

후보문장의 구문구조와 질의문의 구문구조를 비교하여 질

후보문장 1 : 세우다		
주어 (인물 단체)	직목 (건물)	부사 (장소)
마야족	피라미드	여러지역
후보문장2 : 세우다		
주어 (인물 단체)	직목 (건물)	부사 (장소)
그들	피라미드	IGNORE

Fig. 4. 후보문장의 구문구조.

질의문 : 세우다		
주어 (인물 단체)	직목 (건물)	부사 (장소)
ANSWER	피라미드	IGNORE

Fig. 5. 확장된 질의문의 구문구조.

의문 구문 구조 내에 정답 위치에 부합되는 어휘는 '마야족'과 '그들'이 되므로, 두 단어가 정답후보 단어로 얻어지게 된다. 얻어진 정답후보들에 대하여서는 가중치 계산을 통해, 낮은 가중치를 갖는 대명사 정답후보인 '그들'은 낮은 순위를 부여하고, 상대적으로 '마야족'은 높은 가중치를 받아 정답으로 추출되게 된다.

1. 구문구조내의 품사 할당

구문구조에 각 품사를 할당하는 과정은 완전한 형태의 구문분석 없이 문법형태소와 동사-명사간의 품사별 거리, 또한 동사가 결합한 어미의 종류에 따른 품사 탐색 범위 결정에 의한 부분 구문분석으로 수행된다.

구문사전 내에는 각 품사들이 결합하는 고 빈도 조사가 표시되어 있으며, 그 조사들은 문장 내에서 동사에 종속된 품사들을 결정하는데 중요한 단서로 사용된다.

구문구조 '짓다'의 경우, 장소를 나타내는 명사가 조사 '에'와 함께 사용되며, 건축물을 나타내는 직접목적어는 '을' (를)과 함께 사용됨을 나타낸다. 물론 조사의 생략도 발생할 수 있으나 조사가 생략된 경우, 명사의 의미정보와 품사의 순서가 품사 추정의 근거로 사용된다.

또한 질의문과 후보문서에 대하여 부분 구문분석만을 수행하므로 문장 내에 나타난 구 구조가 품사 추출의 가장 큰 걸림돌이 될 수 있다. 이러한 문제를 해소하기 위해 같은 의미속성을 지니고, 동일한 조사와 결합된 두 개의 명사가 경쟁할 경우 동사와의 거리가 가까운 쪽을 우선적으로 결정하게 하였다.

마지막으로 동사와 결합된 어미의 종류에 따라 문장 내에서 품사 탐색의 범위를 다르게 설정하였다. 동사가 종결 및 연결 어미와 결합된 경우, 품사 탐색은 동사의 왼쪽에 대하여서만 수행하고, 동사가 관형사형 전성어미와 결합된 경우에 한해서 동사의 우측 한 어절까지 범위를 확장하여 수행한다.

시스템의 확장

1. 동사의 유도

질의문 중에는 동사를 포함하지 않는 경우가 있지만 그러한 질의문의 경우, 문장 내의 명사로부터 동사를 유도하여 구문구조 기반의 질의 응답시스템에 적용시키는 것이 가능하다. 다음은 질의문 내에 동사가 나타나지 않는 경우의 예이다.

동의를감의 저자는 누구인가?

다음의 질의문에서 얻을 수 있는 중요단어는 ‘동의보감’과 ‘저자’이며, 국어사전의 정의를 이용하여 동사를 유도하는 것이 가능하다.

저 : 작—자(著作者)[—자][명사]
 1. 책을 지은 사람. 지은이. (준말)작자·저자.
 2. (저작권법에 따른) 저작물을 작성하는 사람.

다음의 두 뜻 풀이말에서 얻어진 문장에는 각각 ‘짓다’, ‘작성하다’ 라는 동사가 포함되어 있으며, 각 문장은 문서 검색 및 정답추출이 가능하도록 질의 확장을 수행할 수 있다. 이와 같은 방법이 적용되지 않는 질의문 형태도 있지만, 이에 해당하지 않는 질의문 형태는 명사의 패턴 매칭을 통해 후보 문장을 추출할 수 있다.

명사 의미사전 내에는 각 의미정보에 맞추어 국어사전에서 추출된 뜻 풀이말이 담겨있다. 질의문 분석할 때, 질의문 내에 동사가 나타나지 않는 경우에는 뜻 풀이말을 이용하여 다른 질의문을 유도한다. 동사가 포함되어 있지 않거나 유도되지 않는 대부분의 명사는 ‘A는?’ 형태의 질의문이 대부분을 차지하며, 이러한 질의문 형태는 패턴 매칭에 의해 해결될 수 있다.

2. 가중치 부여

시스템은 각 부분에서 다단계로 가중치를 부여할 수 있으며 누적 계산된 가중치는 최종단계에서 계산을 통해 정답후보 및 정답후보 문장의 정확도로 반영되고, 정확도가 높다고 판단되는 정답이 우선적으로 보여지게 된다. 시스템 내에서는 각 단계에서 가중치를 부여할 수 있다.

- **질의어 확장** : 질의문에 주어진 동사에는 일반 사용자들이 자주 사용하는 동사가 사용할 확률이 높다고 가정하며, 질의문에 사용된 동사에는 높은 가중치를 그 외에 유의어 목록에서 파생된 동사들에 대해서는 보다 낮은 가중치를 줄 수 있다.
- **유의명사 확장** : 질의문에 사용된 명사들의 경우, 문서 검색 시에 유의명사로 확장하게 된다. 이때 확장된 명사들은 보다 낮은 가중치가 부여된다.
- **정답추출 과정** : 정답추출의 Slot allocating과정에서도 가중치가 부여될 수 있다. 정답패턴은 정해진 의미속성을 의미하고, 정답후보들은 그 의미속성에 부합되어야 정답으로 인정된다. 하지만, 의인화 등으로 인하여 정답의 의미속성이 부합되지 않

는 경우가 발생할 수 있다.

질의문 : 사랑노래를 부른 것은 누구인가?

후보문장 : 바람이 사랑노래를 불렀다.

다음과 같이 질의문과 후보문장이 주어진 경우 ‘노래’를 목적어로 갖는 ‘부르다’ 구문 구조는 주어 ‘인물’ 속성을 지닌 명사를 갖게 된다. 하지만 후보문장에서 추출된 정답 후보는 ‘바람’으로 ‘인물’의 속성을 갖지 않는다. 하지만 이런 경우, 의미속성이 일치하지 않아도 낮은 가중치를 갖는 정답으로 추출할 경우, 보다 높은 가중치(정확도)를 갖는 정답후보가 있지 않은 경우, ‘바람’이 정답으로 추출된다.

실험결과

실험결과는 각 단계별에 적용한 기법들의 정확도를 중심으로 측정하였다. 질의문 분석 단계에서는 질의문 내에 동사의 출현 여부와, 동사의 유도여부를 측정하였으며, 그 밖에 질의어 확장과 문서 검색 및 정답추출 등의 과정은 문장 내에서 부분 구문 분석을 통한 주요 품사 추출을 주요 평가요소로 하여 실험하였다.

질의문에서의 동사 추출 및 동사의 유도 과정은 시스템의 전반적인 실용성 측정에 가장 기초적인 판단 근거로 사용된다. 여기서는 TREC-8에 사용된 질의문을 한글로 번역한 질의문 893개 중 100 개가 무작위로 선택되었다.

후보문장에서 정답추출에 대한 정확도를 측정하기 위한 방법으로 문장 내에서 SLOT ALLOCATING기법을 통한 부분 구문 분석에 대한 정확도를 사용하였다.

구문사전을 통하여 동사의 각 품사를 얻어내는 방법에 대한 검증으로 MATEC 99의 태그부착 말뭉치내에서 고빈도 동사가 사용된 문장 중 임의로 1,000개의 문장을 선택하여 그 정확도를 측정하였다.

고빈도 동사는 한국어 동사 약 4,000개(하다 및 되다 동사 포함)에 대하여 코퍼스 내에서의 사용빈도를 중심으로 추출하였다.

실험 결과 동사에 대한 문장성분의 할당과정은 목적어의 경우 높은 수치를 나타냈지만, 상대적으로 생략이 빈번한

Table 1. 질의문에서의 동사 추출

질의 수	동사포함	동사유도	동사유도율
100개	52	19	71%

Table 2. 구문구조를 이용한 품사결정

품사	주어	목적어
정확도	76.2%(762/1000)	93.3%(933/1000)

주어의 경우, 목적어에 비해 낮은 정확도를 보였다.

결론 및 향후 과제

본 논문에서는 질의 응답시스템 내에서 사용되는 지식베이스의 구축비용을 줄이기 위해 구문분석의 방법론에 대하여 서술하였다. 구문분석을 위해 문장성분 별 의미정보가 포함된 구문사전과 의미속성이 분류된 명사 유의어 사전을 이용하였으며, 그 기법들을 질의 응답시스템에 적용하였다.

TREC등에서 제시된 질의 문의 상당수가 동사를 포함하고 있거나, 포함된 명사를 이용해 동사를 유도하는 것이 가능한 것을 보여, 동사가 질의문 분석에서 일반적으로 사용할 수 있는 단서라는 것을 보였으며, 문장성분 별 의미분류가 포함된 구문사전을 통한 품사선택 실험에서도 문장의 각 주요성분이 높은 정확도로 선택됨을 보였다. 또한 이 두 가지 실험내용이 질의 응답 시스템 내에서 시스템의 정확도를 보장하는데 효과적인 방법임을 보였다. 마지막으로 고정적인 지식 베이스에 의존하지 않고 동사 중심의 의미 구문 구조를 사용함으로써, 기존의 지식베이스 기반 정답추출 시스템에서 불가능했던 의인화된 정답까지 추출하는 것이 가능함을 보였다.

하지만 본 논문에서 구현된 시스템의 경우 구 구조에 대한 고려 없이 구현되었기 때문에 문장 성분 추출부분에서 특히 주어의 경우 정확도를 저하시키는 원인이 되었다. 이 문제를 해결하기 위하여, 구 구조 chunking 기법이 시스템에 추가될 경우, 보다 높은 효율을 보일 것이며, 본문 중에 보였듯이 여러 문장에 걸쳐 정답이 나타나는 경우, 정답으

로 추출되는 대명사의 문제가 걸림돌이 되었다.

또한 일반 WEB문서 등과 같이 구조화된 문서 내에서 빈번히 나타나는 품사생략에 대한 처리가 포함된다면, 보다 융통성 있는 시스템으로 발전할 수 있을 것이다.

REFERENCES

- 1) 김수민, 백대호, 김상범, 임해창 : “시소러스 범주를 이용한 질의응답시스템”, 제12회 한글 및 한국어 정보처리 학술대회, pp179-183
- 2) 이경순, 김재호, 최기선 : “한국어 질의응답 시스템에서 개체 인식에 기반한 대답추출”, 제12회 한글 및 한국어 정보처리 학술대회, pp184-189
- 3) 강승식, 이하규, 손소현, 문병주, 홍기채 : “자연언어 질의문장의 용어 가중치 부여기법”, 제14회 한글 및 한국어 정보처리 학술대회, pp223-227
- 4) 장문수, 장명길, 김현진, 오효정, 이재성 : “인터넷 질의/응답을 위한 지식베이스 구축”, 제12회 한글 및 한국어 정보처리 학술대회, pp198-204
- 5) 허 정, 옥철영 : “사전 뜻풀이말에서 추출한 의미정보에 기반한 의미 중심성 해결”, 제12회 한글 및 한국어 정보처리 학술대회, pp269-277
- 6) 최용석, 이주호, 최기선 : “격을 자동구축과 격률평가 방법에 대한 연구”, 제11회 한글 및 한국어 정보처리 학술대회, pp272-279
- 7) 송영빈, 채영숙, 박영일, 이정민, 설가영, 황혜리, 한나리, 최기선 : “동사의 애매성 해소를 위한 구문 의미사전의 구축”, 제11회 한글 및 한국어 정보처리 학술대회, pp280-287
- 8) 조숙환 : “목적어 생략에 대한 동사의 의미표상 및 추론의 역할”, 제13회 한글 및 한국어 정보처리 학술대회, pp457-461
- 9) 정석원, 박의규, 나동열, 윤준태 : “격관계와 상호정보를 이용한 한국어 의존 파서”, 제13회 한글 및 한국어 정보처리 학술대회, pp450-456
- 10) Moldovan D, Adrian N(2002) : “Lexical Chain for Question Answering”, *Coling 2002*
- 11) Eduard H, Hermjakob U, Lin CY, Ravichandran D(2002) : “Using Knowledge to Facilitate Factoid Answer Pinpointing”, *Coling 2002*
- 12) Jimmy L(2002) : “The Web as a Resource for Question Answering”, *LRCC 2002*
- 13) 홍재성 외 공저(1997) : “현대 한국어 동사구문 사전”, 두산동아
- 14) MATEC 99(1999) : ETRI 컴퓨터 소프트웨어 기술 연구소 지식정보 연구부