

UNITEX를 이용한 로봇 주제의 자동 질의응답 시스템

정병호* 박충식* 우영운**

*영동대학교 컴퓨터공학과

**동의대학교 멀티미디어공학과

*Automatic Query Answering System Using The UNITEX for Robots Domain

Byung-Ho Jung* Choong-Shik Park* Young-Woon Woo**

*Dept of Computer Eng, Youngdong University

**Dept of Multimedia Eng, Dong-Eui University

E-mail : nauthiz@nate.com, leciel@youngdong.ac.kr, ywwoo@deu.ac.kr,

요 약

기존의 자동 질의응답 시스템은 복잡한 구문처리와 의미분석으로 인하여 그 구현에 많은 어려움이 있다. 본 논문에서는 대화의 주제를 특정 분야로 제한할 경우 구문구(syntactic phrases)와 동의구(synonymous phrases)를 용이하게 처리하는 UNITEX를 사용해 문법을 처리하는 자동응답시스템을 제안한다. 자동응답시스템의 지식표현과 추론, 응답문 생성은 자체 개발한 지식처리시스템인 NEO를 사용한다.

키워드

UNITEX, local grammar, Query Answering System, NLP, Natural Language

I. 서 론

지식사회, 정보사회가 대두되면서, 컴퓨터로 하여금 사용자의 질의에 의한 정확하고 의미 있는 정답을 보여주는 지식기반 질의응답 시스템을 요구하게 되었다.[1] 질의응답 시스템이란 사용자의 질의와 관련된 문서를 검색하는 정보검색 시스템과는 달리 사용자의 질의에 대한 답변이 될 수 있는 정답을 문서 집합 내에서 탐색하여 사용자에게 제시해주는 시스템이다. 일반적으로 질의응답 시스템은 사용자의 질의에 관련된 문서를 검색하는 후보검색 단계(candidate retrieval phase)와 검색된 문서 내에서 정답을 생성하는 정답추출 단계(answer extraction phase)로 구성된다. 그러나 현재 질의응답시스템이 제공하는 것은 언어 처리 기술을 이용하여 질의를 분석하여 특정한 문서를 제시하는 수준에 그치고 있다. 이런 의미에서 보다 정밀적이고 도메인 제약적인 효율적인 시스템에 대한 연구가 활성화 될 필요가 있다. 즉 질의응답 시스템이 하나의 소규모 대화형 시스템이라는 점을 감안한다면, 사용자와 컴퓨터간에 지식을 서로 주고받을 수 있는 의미 있는 시스템이 개발되어야 한다.

본 논문에서는 단순히 질의에 관련된 문서를 질의 분석을 통해 특정문서를 제시하는 수준이 아니라, 소규모의 '로봇'이라는 도메인을 제시하여 일반적인 질의응답시스템과는 달리 로봇에 관련된 지식베이스를 구축하여 사용자에게 정확한 응답을 습득하게 해주고, 사용자의 인지 체계 속에 어렵듯이 내포되어 있는 개념적 지식을 더욱더 표면적으로 확장해 나갈 수 있는 질의응답시스템의 구축 방안을 제시하고자 한다.

II. 관련 연구

질의응답 시스템은 사용자의 질의와 관련된 문서를 검색하는 정보검색(Information Retrieval) 시스템과는 달리 사용자의 질의에 대한 답변이 될 수 있는 정답을 문서 집합내에서 탐색하여 사용자에게 제시해주는 시스템이다.

기존의 질의 응답 시스템을 보게 되면 각종 방법이 존재하고 있다. 그중 한가지인 지능적 질의 처리가 있는데 살펴보자면 지능적 질의 처리란 사용자의 질의에 대해 직접적인 응답만을 사용자에게 제공하는 것이 아니라, 시스템이 보여주고

있는 정보를 활용해서 질의의 응답과 함께 보다 많은 정보를 제공 해 주는 것이다. 관련 연구로는 협조적 질의 응답과 내재적 응답이 있다.

또한 본 논문에서는 사용자 질의를 처리하기 위하여 어휘 문법과 부분 문법을 사용하는 UNITEX 프로그램을 사용한다. 부분 문법은 프랑 스 전산언어학자인 모리스 그로스(Maurice Gross)가 제안한 방법론이다. 언어학적 관점으로 부분문법을 살펴보면 어휘문법 방법론을 좀 더 유연하게 보완하는 장치이다. 간단하게 어휘 문법과 부분 문법을 언어학적 의미로 본다면, 어휘문법이란 언어 연구의 기본대상이 단문으로 이루어 지며, 기본 문형의 다양한 통사적 변형 구조로 기술된다. 이 어휘문법을 보완하기위해 부분문법은, 문장보다 더 작은 단위(부분적 현상)에 대해 기술 하고, '통사적' 관계가 아닌 형태, 어휘, 의미, 분 야별 연관 관계가 나타날 때 이를 기술한다. 이 부분 문법은 전산학적 관점으로 언어적 기술 결 과를 효율적으로 적용할 수 있는 형식 모델의 역 할을 한다고 볼 수 있다. 부분 문법은 유한상태 문법을 지향하고 방향성 그래프의 형태로 표상하 며 유한 오토마타/트랜스듀서로 변환하게 된다. 그리고 UNITEX란 이런 부분 문법의 그래프의 효 율적인 구현과 전산학적 변환을 위한 전용도구이다.[2]

III. 자동 질의응답 시스템 설계

본 논문에서는 자동 질의응답 시스템을 개발하 기 위하여 NEO 개발 환경에 UNITEX모듈을 결 합하여 문장의 입, 출력이 가능한 기능을 보완하 여 자동 질의응답 시스템 개발 환경을 구현하였 다.

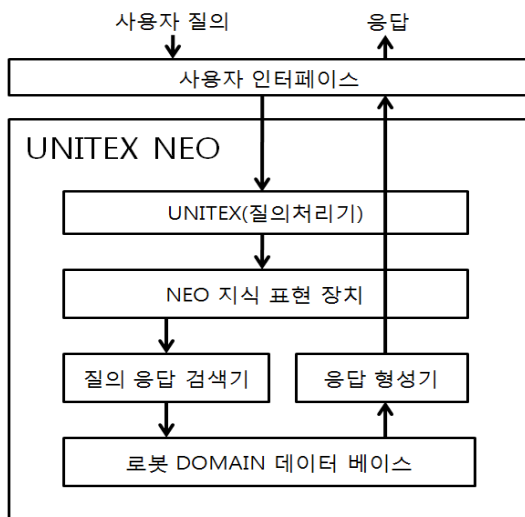


그림 1. 자동 질의 응답 시스템 구조

프로그램 개발도구로는 Dev-C++와 Visual C++을 이용하였다.

개발된 프로그램에서는 부분 문법을 이용한 UNITEX의 파싱을 이용해 입력받은 문자를 처리 하게 되고, 처리 받은 결과를 NEO로 지식표현을 통한 문서 생성을 통해 질의에 대한 응답을 출력 하게 된다.

그림 1.의 구조와 같은 UNITEX NEO를 개발 하여 본 논문에서 지향하고자 하는 자동 질의 응 답 시스템을 구축한다. 실험 및 구현을 하면서 부 족한 점이나 오류가 있으면 그때 마다 수정 및 보완을 해준다.

그리고 그림 1.과 같이 시스템을 구현할 경우, 사용자 질의를 입력하게 된다. 본 논문에서는 사 용자 질의를 로봇이라는 주제로 질의문을 제한하 였고 시스템을 만들었다.

본 연구에서는 질의문의 양이 많아질 경우 많 은 단어가 중복되고 다시 사용되는 경우가 있을 을 볼 수 있다고 생각한다. 우리는 이런 문장의 재사용성을 가지고 어떻게 활용할 수 없을까 라는 생각이 들것이다. 그래서 본 논문의 연구에서 는 Local Grammar 즉 부분 문법을 이용하여 이 런 질의문들을 효율적으로 처리할 수 있는 UNITEX NEO를 개발하였다. UNITEX NEO에서 는 먼저 로봇 관련 질의문을 받게 되면 UNITEX 프로그램에서 미리 정의해둔 부분 문법과 Deco를 이용하여 질의문을 처리하여 NEO 지식 표현 장 치로 전송을 해주게 된다.

이와 같은 질의문을 토대로 UNITEX를 통한 로컬 문법 그래프를 만들도록 한다. 이와 같이 만 든 그래프를 이용하여 질의 처리기에서 질의문에 대한 형태소 분석과 파싱을 통한 질문의 의도를 파악 할 수 있게 한다.

이때 지식 처리는 LISP과 비슷한 NEO 자체의 언 어로 이루어진다. 질의 처리기에서 Deco와 NEO 의 규칙추론으로 지식표현을 하게 되면 질의문이 어떤 의도인지 알 수 있다. 그렇게 알게 된 의도 로 질의 응답 검색기에서 미리 만들어진 로봇 도 메인의 데이터 베이스에서 알맞은 답을 찾게 되 는 것이다. 그렇게 찾게 된 질의문에 대한 응답은 응답 형성기에서 올바른 답안으로 인터페이스에 출력 되게 된다.

IV. 실험 및 구현

본 논문에서는 로봇 도메인의 자동 질의응답 시스템의 실험을 위하여 로봇 도메인의 질의문을 만들었다. 그런데 질의문을 보게되면 많은 수의 동의어 질의문들이 입력될 수 있는데, 이런 다양 한 패턴을 동일한 의미 구조를 가지는 질의문들 로 만들어 보았다.[2][3] 표 1.을 보게 되면 다양한 질문을 볼수있다.

표 1. 로봇 관련 질의문

로봇 분야 관련 질의문의 예	
▶ 어떤 회사가 아시모를 만들었나요?	
▶ 혼다에서 만든 로봇은 무엇인가요?	
▶ 아시모의 기능은 무엇이 있나요?	
▶ 아시모의 키는 어떻게 되나요?	
▶ 아시모의 몸무게는 어떻게 되나요?	
▶ 아이보를 만든 어떤 회사인가요?	
▶ 소니에서 어떤 로봇을 만들었나요?	
▶ 아이보는 키가 어떻게 되나요?	
▶ 아이보의 몸무게는?	

표 1.의 질의문들을 보게되면 다양한 패턴을 가지게 되는 동일한 의미 구조를 사용하는 것을 볼수 있는데 그것을 표 2.와 같이 유형 별로 나타낼수 있다.

표 2. 유형별 질의문

유형별 질의문 하위 분류	
<Company/Robot>	
▶ 어떤 회사가 아시모를 만들었나요?	
▶ 아이보를 만든 회사는 어디가요?	
<Robot/Company>	
▶ 혼다에서 만든 로봇은 무엇인가요?	
▶ 소니에서 어떤 로봇을 만들었나요?	
<Ability/Robot>	
▶ 아시모의 기능은 무엇이 있나요?	
<Height/Robot>	
▶ 아시모의 키는 어떻게 되나요?	
▶ 아이보는 키가 어떻게 되나요?	
<Weight/Robot>	
▶ 아시모의 몸무게는 어떻게 되나요?	
▶ 아이보의 몸무게는?	

이런 예제를 가지고 본 논문에서는 그림 2. 시스템 흐름도와 같은 방식으로 UNITEX NEO가 구현되었다.

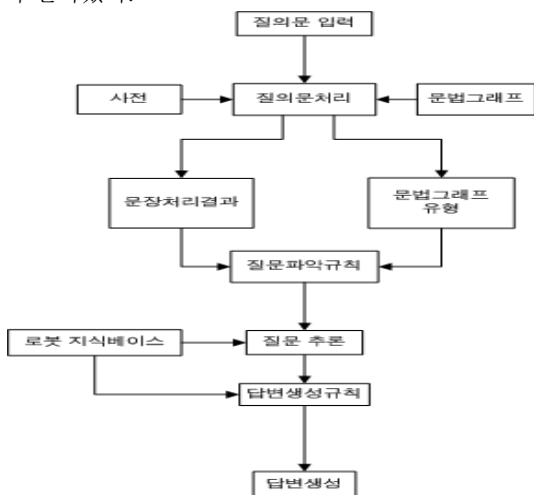


그림 2. 시스템 흐름도

이렇게 나타낸 유형별 질의문을 가지고 UNITEX 내의 문법그래프를 나타내어, 질의문을 입력받아 문법그래프와 사전을 이용하여 문장 처리를 할 수 있게 되는 것이다. 예를 들면 사용자가 “어떤 회사가 아시모를 만들었나요?”라는 질문을 하게 되면 질의문을 사전(표 3.)과 문법그래프(그림 3.)에 의해 처리하게 된다.[4]

표3. 사전

사전	
어떤(What)	를(O+PA)
회사(N+Company)	있(P+PA)
아시모(N+Robot=아시모)	나요(Q+PA)
만들다(V+Make=만들)	?[Q+Sign]
가(S+PA)	

(tag ex : N=명사, O+PA=목적격 조사, P+PA=과거 조사, Q+PA=의문형 조사)

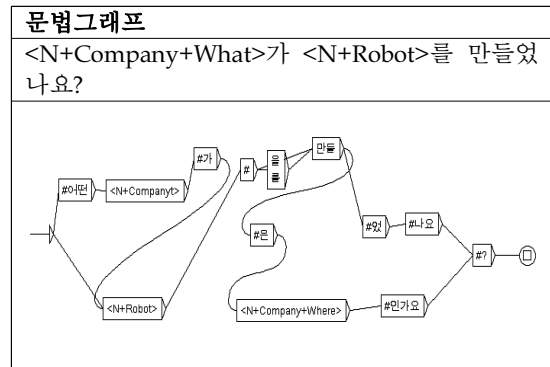


그림 3. 문법그래프

사전을 이용하여 처리를 하게되면 질의문의 문장들이 태깅이 돼서 나오게 된다. 예를 들면 그림 4.의 결과1 같이 “어떤”이란 단어는 “어떤 [what=]”이라는 처리가 되어서 나온다. 각각의 문장의 단어에 대하여 문장이 사전에 의해 태깅이 되어 나온다. 그리고 문법 그래프를 이용하면 결과 2와 같은 결과를 받을 수 있다.

UNITEX 문장 처리 결과	
결과1	결과2
어떤[What+=] 회사[N+Company=] 가[S+PA=] 아시모[N+Robot=아시모] 를[O+PA=] 만들[V+Make=만들다] 있[P+PA=] 나요[Q+PA=] ?[Q+Sign=]	회사[<N+Company>] 아시모[<N+Robot>]

그림 4. 문장 처리 결과

그리고 문법 그래프를 이용하면 결과 2와 같은 결과를 받을 수 있다.

그림 4.와 같이 나온 문장처리 결과로 본 논문에서는 문법 그래프로 무엇을 찾는지 알 수 있게 된다. <Company/Robot>의 경우 "<Robot>"의 만든 "<Company>"가 어디인지를 묻고 있는 유형이다. 그것을 이용하여 우리는 이 질문에 맞는 질문 파악 규칙을 불러올 수 있다. 지금 예제의 질문과 약규칙의 경우 그림 5.와 같다.

<p><Company/Robot>의 질문파악규칙</p> <p>if(?x (N Company 회사)) (?y (N Robot 아시모))</p> <p>then (?x is-a Company) (?y 제작사 ?x) (?y is-a Robot)</p>
--

그림 5. <Company/Robot>질문 파악 규칙

그림 5. <Company/Robot>질문 파악 규칙을 보게되면 "?y"에 <Robot>이 오게 된다. 그렇게 되면 "?x"에 해당하는 객체를 미리 정의된 지식베이스(그림 6.) 안에서 규칙처리로 찾게 된다. 그리고 이렇게 추론된 결과로 <Company/Robot>의 답변 생성 규칙으로 답변을 추출하게 된다.

<Company/Robot>에 대한 답변 생성규칙을 보면 그림 8.과 같다.

<p>지식베이스</p> <p>(혼다 is-a Company) (아시모 제작사 혼다)</p>
--

그림 6. 지식 베이스

<p><Company/Robot>의 답변생성규칙</p> <p>if (?x 제작사 ?y) then (Printf ?y "가" ?x를 만들었습니다)</p>
--

그림 7. <Company/Robot>의 답변생성규칙

이와같은 NEO의 if~ then 을 사용하는 그림 7.과 같은 답변 생성 규칙으로 "if (?x 제작사 ?y)" ?x 즉 아시모 제작사가 ?y 혼다 일 경우 "then (Printf ?y "가" ?x를 만들었습니다)"를 수행하게 되어, "혼다(?x)가 아시모(?y)를 만들었습니다."라는 "어떤 회사가 아시모를 만들었나요?"라는 질의문에 응답을 하게 되는 것이다.[4]

VI. 결론 및 향후 연구 과제

본 논문에서는 로컬 문법을 이용한 UNITEX와 규칙추론엔진인 NEO를 이용하여 질의 응답시스템을 제안하였다. 본 연구의 프로토타입 시스템으로 구현된 자동 질의 응답 시스템인 UNITEX

NEO는 윈도우 상에서 사용자에게 로봇에 관련된 질의를 입력받고 NEO시스템의 규칙 추론을 이용하여 지식 표현을 한다. 그런 다음 질의 응답 검색기를 통하여 알맞은 질의 응답을 출력 했다. 그러나 여러 부분으로 부족한 점이 많이 보였다. 예를 들면 질의문 처리를 위한 UNITEX사전의 부족함을 들 수 있겠고, 또한 이미 정의한 문장이 아닌 다른 문장을 입력할 때의 처리의 문제도 있다. 추후 연구에는 처음 입력하는 문장의 처리도 생각해 보아야 한다고 생각한다.

참고문헌

- [1] 이재홍 외 2명, "개념어의 습득을 위한 지식 기반 질의응답 시스템", 제15회 한글 및 한국어 정보처리 학술대회논문집, pp. 95-100, 2003.
- [2] Tae-Gil Noh, Young-Jin Han, Seong-Bae Pack and Se-Young Pack, "Processing of Korean Natural Language Queries Using Local Grammars", *Lecture Notes in Computer Science Vol 5459/2009 : Computer Processing of Oriental Languages. Language Technology for the Knowledge-based Economy*, pp34-44, 2009.
- [3] 남지순, "자연언어 검색 질의문 인식을 위한 유한 그래프 문법의 구축", 언어과학, 제 15권 pp 39-70, 2009.
- [4] Max Silberztein. "Finite state descriptions of various levels of linguistic phenomena", *Language Research* 28(4), Seoul National University, pp. 731-748. 1992.
- [5] 우영운 외 2명, "게임에서의 지능적 NPC 구현을 위한 자연어 대화 처리 기법", *한국해양정보통신학회논문지*, v.11 no.12 ,pp.2406-2412 2007.