

질의문 자동생성방식의 질의응답시스템의 설계 및 구축

김민경⁰, 김한준[†]

서울시립대학교 전자전기컴퓨터공학부

comengs@gmail.com⁰, khj@uos.ac.kr

Design of Question Answering System with Automated Question Generation

Min-kyoung Kim⁰, Han-joon Kim[†]

Dept. of Electrical & Computer Engineering, Univ. of Seoul

요약

질의응답시스템에서 사용자 질의로 입력된 자연어문장을 완벽하게 분석하는 것은 쉬운 일이 아니며, 사용자의 질의 의도의 불명확성으로 키워드 여러 개의 질의문이 존재할 수 있다. 본 논문에서는 질의를 하기 전에 사용자가 안게 되는 자연어 질의문의 작성 부담감을 줄이고, 키워드만으로 자신이 원하는 질의문을 선택할 수 있는 시스템을 제안하고자 한다. 제안 시스템은 평서문에서 자동으로 질의문을 생성한다. 질의문 생성은 장문형질의문생성과 단문형질의문생성으로 구분하며, 장문형질의문은 문장의 전체형태를 유지하면서 특정고유명사를 질의하는 것이고, 단문형질의문은 주어진 고유명사를 질의하는 최소한의 요소를 갖춘 단순 형태의 질의 문장이다. 또한 제안 시스템은 생성된 질의문이 유한 해답을 도출할 수 있는 의미있는 질의문을 선별하는 과정을 포함한다. 본 논문에서 제안한 방식이 사용자에게 의미있는 질의문을 제시하여주고 사용자가 원하는 질의문을 선택하게 함으로써 검색의 시간단축과 자연어문장 질의어 자체에 대한 고민을 해소시킬 수 있다. 또한 이는 자연어 문장처리의 한계를 극복할 수 있는 시스템을 구현할 수 있는 기반을 마련한 것이다.

1. 서론

최근 기존 정보검색 엔진의 한계를 극복하기 위해 자연어검색에 의한 질의응답시스템(Question Answering System)기술을 병합하기 위한 노력이 진행되고 있다.[1][2] 이는 기존 정보검색 엔진이 문서 단위의 결과를 반환하기 때문에 사용자의 구체적인 정보 요구에 대해서 정확한 답변을 제공하지는 못하기 때문이다. 질의응답시스템은 자연어로 된 문장 질의어 입력에 대하여 문서수준이 아닌 구체적 답변을 주는 시스템이다. 질의응답시스템에 관한 연구는 자연어 질의문의 해석, 답변 검색 등에 있어서 그 정확도를 높이기 위해 다양한 시도가 이루어져 왔지만, 기본적으로 정확한 질의문의 작성 부담 또는 질의문 해석의 오류 가능성을 최소화 하기는 어려운 문제이다. 예를 들면 사용자가 'Albert Einstein'이 누구이고, 언제, 어디서 태어났는지, 혹은 그가 어떤 일을 했는지에 대한 명확한 정보만을 얻고 싶고자 할 때, "Who is Albert Einstein?", "When(or Where) was Albert Einstein born?", "What theory did Albert Einstein propose?" 등의 의문사를 포함한 특정 형식의 질의문을 작성할 것이다. 이와 같은 자연어 질의문 작성은 해당하는 답변을 검색하기 위해 질의문 분석이 정확해야 하기 때문에, 사용자는 질의문 작성에 있어 자유도가 크지는 못한다. 즉

질의문의 형식이 복잡하거나 정형성이 부족하다면 문장 분석이 어렵다는 것이다. 질의응답 시스템은 동일한 답변에 대하여 가능한 여러 질의문을 미리 준비하거나, 또는 입력된 자연어 질의를 분석하여 해당 답변을 찾아내야 한다. 하지만, 이러한 시스템은 사용자의 임의의 자연어 질의문장과 일치하지 않거나, 또는 질의문 분석의 오류가 발생할 수 있다. 사용자 입장에서도 자기의 검색요구에 합당한 질의문을 완성하는 것은 쉽지가 않을 수 있다.

관련 시스템으로서 IBM사의 Jennifer ChuCarroll 연구원이 개발한 'Piquant' 시스템[3][4]은 추론과정을 자동화 하여 사용자의 직접적인 질의에 대한 대답을 가능하게 해준다. 예를 들면 Piquant 시스템에 "Who is Canada's prime minister?" 라고 질의했을 때 본문에서 수상의 이름이 나오지 않았음에도 질의문에 대한 대답으로 수상의 이름이 나왔다고 한다.[5] 이 시스템은 질의문의 의미를 분석하여 추론을 통해 그 해답을 찾는 방법을 제시하였다. 이 방법은 자연어 문장처리를 하는데 있어 기존의 시스템보다 인공지능적인 방법으로 접근 한다는데 의의가 있지만, 이 방법 역시 사용자가 느끼는 자연어 질의문장에 대한 어려움을 해소 시켜 주지는 못한다.

본 논문에서는 위 문제를 극복하기 위하여 답변을 포함한 문장 자체로부터 가능한 질의문을 자동 생성하고자 한다. 본 논문의 관점은 질의문을

† 교신저자: 김한준 (khj@uos.ac.kr)

자동생성하기 위한 규칙과 이를 적용해 ‘의미있는’ 질의문의 패턴을 찾는 것이다. 본 논문의 의의는 미리 생성된 의미있는 질의문을 통해 사용자가 보다 쉽게 질의응답 시스템을 이용할 수 있는 시스템의 발판을 설계하여 구축하는데 있다.

2. 문제정의

질의문을 자동으로 생성하기 위해서는 우선 평서문에서 질의가 가능한 고유명사를 찾아야 한다. 이러한 방법이 ‘고유명사인식기술(Named Entity Recognition)’ [6] 이다. 고유명사인식²은 인물의 이름을 나타내는 <Person>, 위치나 지위를 나타내는 <Position>, 지역을 나타내는 <Location>, 시간을 나타내는 <Time>, 기관을 나타내는 <Organization>, 기타 고유명사를 나타내는 <Miscellany> 등을 식별하는 것이다. 예를 들어 아래와 같은 문장이 있다고 하자.

“George Walker Bush is the forty-third and current President of the United States of America”

이 문장에 고유명사인식기술을 적용하게 되면 아래와 같은 결과를 생성할 수 있다.

George Walker Bush : <Person>
President : <Position>
the United States of America : <Location>

이렇게 분별된 고유명사를 이용하여 질의문을 자동으로 생성할 수 있다. 위의 결과로 자동생성 할 수 있는 질의문은 <Person>을 물어보는 질의문이다. 이때 질의문의 대답은 <Position>+<Location>이 된다. 예를 들면 <Person>을 질의하는 문장 “Who is George Walker Bush?” 라는 질의문을 생성하면 이 질의문에 대답은 “Present of the United States of America”가 된다. 그렇다면 질의문 자동 생성을 위해서 필요한 문제들은 무엇이고, 그 해결책에는 어떤 것이 있는지에 대해 살펴보도록 한다.

질의문 자동생성을 위해 두 가지 중요한 문제가 있다. 첫번째 문제는 평서문에서 자동으로 질의문을 생성하는 것이고 두번째 문제는 자동생성된 질의문에서 ‘의미 있는 질의문’을 선별하는 것이다. 이들 문제는 아래와 같다.

- ① 질의문을 자동으로 생성하는 문제는 ‘장문형 질의문’ 생성과 ‘단문형질의문’ 생성으로 나뉘 볼 수 있다. 장문형질의문은 원문의 내용과 형태를 유지하면서 질의문을 생성한 문장이다. 이에 반해, 단문형질의문은 원문의 내용을 완전하게 유지하지 않고, 추출한 고유명사를 이용하여 그 고유명사 자체에 대한 기본적인

형식의 간단한 형태의 질의문을 생성한 문장이다.

- ② 자동생성된 질의문에서 ‘의미있는’ 질의문을 골라내는 문제는 질의문의 의미성을 규정하는데서 출발해야 한다. 본문에서 규정하는 의미있는 질의문은 생성된 질의문이 단독으로 사용되었을 때 그 의미가 명확한 문장이다. 명확한 의미전달은 일반적이거나 포괄적인 의미가 아닌 특정인물, 특정 시간, 특정 장소 등을 포함하여 무엇을 질의하는지가 분명한 문장을 의미한다. 이러한 의미성을 사람이 직접 인식하는 것은 문제가 없지만, 컴퓨터가 질의문의 ‘의미성여부’를 판단하는 것은 어려운 문제이다.

본 논문은 위에서 제기한 문제들을 해결하기 위해서 문장의 형식 또는 패턴을 찾는 데서 출발한다. 발견된 패턴마다 질의성 생성규칙을 정의하였으며, 생성된 질의문은 ‘의미있는’ 질의문형식인지를 검사함으로써 최종 질의문을 완성한다. 의미있는 질의문을 생성하기 위해 평서문 문장에서 질의문을 자동생성하는 기법에 대해 알아보도록 한다.

3. 질의응답시스템을 위한 질의문 자동생성기법

의미있는 질의문 자동생성을 위한 첫단계로서, 평서문에 고유명사인식기술을 적용하여 질의문을 자동 생성한다. 예를 들어 다음과 같은 평서문 문장이 있다고 하자.

“Dr.Martin Luther King,Jr. was assassinated on April 4, 1968, in Memphis, Tennessee.”

위 문장에 고유명사인식기술을 적용하면 아래와 같은 결과를 생성할 수 있다.

Dr.Martin Luther King,Jr : <Person>
April 4, 1968 : <Time>
Memphis, Tennessee : <Location>

결과를 살펴보면 평서문 문장에서 고유명사만을 찾아낸 모습이다. 이 결과로 나올 수 있는 질의문은 세 가지이다. <Person>을 질의하는 ‘Who’, <Time>을 질의하는 ‘When’, <Location>를 질의하는 ‘Where’ 의문사를 이용해 질의문을 자동생성할 수 있다.

3.1 질의문 생성방안

고유명사인식기술을 적용하여 나온 결과를 보고 해당 고유명사마다 관련된 의문사를 적용하여 질의문을 생성한다. <Person>에는 ‘Who’, <Time>에는 ‘When’, <Position>에는 ‘Where/What’, <Location>에는 ‘Where’, <Miscellany>에는 ‘Which/What’의 의문사를 적용하여 질의문을 생성한다. 자동질의문생성은 장문형 질의문 생성과 단문형질의문생성으로 나누어 규칙을 규정할 수 있다.

3.1.1 장문형질의문의 생성

² 본 논문에서는 NER 도구는 영국 Sheffield대학에서 개발한 GATE(General Architecture for Text Engineering)를 사용하였다.

장문형질의문은 원문의 전체내용을 유지하면서 질의문을 생성한다. 질의문을 이루는 주어, 동사를 포함한 목적어,보어를 동반하여 문장을 구성한다. 장문형질의문 생성규칙은 아래와 같다.

- ① 고유명사인식기술의 결과 중 질의하려는 고유명사를 선택하여 관련된 의문사로 변환하고 의문사를 문장의 앞으로 보낸다. 이때 의문사 뒤에 주어, 동사 위치를 변경하되 문장을 이루는 목적어, 보어는 원문과 그대로 사용한다.

표 1. 장문형질의문 생성규칙을 적용한 예

원 문	고유명사인식기술 적용
Dr.Martin Luther King,Jr. was assassinated on April 4, 1968, in Memphis, Tennessee.	Dr.Martin Luther King,Jr : <Person> April 4, 1968 : <Time> Memphis, Tennessee : <Location>



자동생성된 질의문	대 답
When was Dr.Martin Luther King,Jr. assassinated in Memphis, Tennessee?	April 4, 1968
Where was Dr.Martin Luther King,Jr. assassinated on April 4, 1968?	Memphis, Tennessee

표 1의 상단테이블 좌측은 평서문의 원문내용이고, 우측은 고유명사인식기술을 적용한 예이다. 상단테이블 위 우측에서 확인한 바와 같이 이 원문의 고유명사 결과는 <Person>, <Time>, <Location> 세 가지이다. 이를 이용 해 하단의 표에서 고유명사를 유도하는 질의문을 생성하였다. 하단테이블의 자동생성된 질의문을 살펴보면 각각 고유명사에 맞게 의문사를 적용하였고, 질의문은 원문의 전체 내용을 유지하면서 질의문을 생성한 것을 확인할 수 있다. 이와 반면, 단문형질의문은 인식된 고유명사만에 충실하여 문장의 일부를 이용한 간단한 형태의 질의문이다.

3.1.2 단문형질의문의 생성

단문형질의문은 문장의 일부를 이용해 질의문을 생성한다. 질의문을 이루는 주어, 동사를 이용하여 장문형질의문보다 단순한 구조로 문장을 구성한다. 단문형질의문 생성규칙은 아래와 같다

- ① 주어가 'be동사'를 동반한 인물이면서 be동사 뒤에 고유명사인식기술의 결과가 <Position>으로 인식하면 주어를 질의하는 단문형질의문 생성이 가능하다.

표 2. 단문형질의문 생성규칙 ①을 적용한 예

원 문	고유명사인식기술 적용
Sejong was the third son of King Taejong and he established the Hall of Worthies.	Sejong <Person> Son of King Taejong <Position> the Hall of Worthies <Location>



자동생성된 질의문	대 답
Who is Sejong?	Son of King Taejong

표 2에서 상단테이블의 고유명사결과는 <Person>, <Position>, <Location> 세 가지이다. 이 중 단문형질의문 생성규칙 ①에 의해 단순히 주어 <Person>이 누군지를 물어보는 단문형질의문 생성이 가능하다. 이때 <Position>은 주어가 누구인지에 대한 답변이 된다. 생성된 질의문은 원문의 모든 내용을 유지하지 않고 '의문사+be동사+주어'의 단순한 형태로 문장을 구성한다.

- ② 문장의 구조가 수동태일 때 목적어/보어를 질의하는 단문형질의문 생성이 가능하다.

표 3. 단문형질의문 생성규칙 ②를 적용한 예

원 문	고유명사인식기술 적용
Dr.Martin Luther King,Jr. was assassinated on April 4, 1968, in Memphis, Tennessee.	Dr.Martin Luther King,Jr : <Person> on April 4, 1968 : <Time> in Memphis, Tennessee: <Location>



자동생성된 질의문	대 답
When was Dr.Martin Luther King,Jr. assassinated?	April 4, 1968
Where was Dr.Martin Luther King,Jr. assassinated?	Memphis, Tennessee

표 3에서 상단테이블 원문은 be동사+과거분사로 수동태 문장이다. 단문형질의문 생성규칙 ②을 적용하여 언제, 어디서를 질의하는 간단한 문장을 생성하였다. 이와 같이 단문형질의문은 문장의 전체내용을 유지하지 않고 특정고유명사만을 질의하는 문장이다. 규칙에 맞게 질의문을 생성한 후에는 의미있는 질의문 자동생성을 위한 의미있는 질의문을 추출하는 작업이 필요하다.

3.2 의미있는 질의문의 선별

지금까지 장문형질의문 생성규칙과 단문형질의문 생성규칙을 적용하여 자동으로 질의문을 생성하는 규칙을 살펴보았다. 그런데 이렇게 생성된 질의문이 모든 질의문으로서 그 의미성이 완전한 것은 아니며, 의미없는 질의문도 생성될 수 있다. 여기서 의미없는 질의문이란 질의문을 생성하였지만 질의문 자체가 해답을 요구할 수 없거나 해답이 무한집합일 수 있는 문장을 의미한다. 예를 들어 다음과 같은 평서문 문장이 있다고 가정하자.

“UOS was established in 1918.”

위 문장에 고유명사인식기술을 적용하면 아래와 같은 결과를 생성할 수 있다.

UOS <Organization>
1918 <Time>

이 결과를 이용해 주어를 물어보는 질의문 “Which was established in 1918?” 라는 문장이 생성 가능하다. 이 질의문은 광범위하고 직접적으로 질의하는 대상에 대한 명시가 존재하지 않는 의미없는 질의문이다. 의미있는 질의문이란 질의하는 대상에 대한 명확하고 분명한 의미가 담겨있는 문장이다. 의미있는 질의문을 선별하는 규칙은 아래와 같다.

① 생성된 질의문의 주어가 인칭대명사나 지시대명사가 아닌 문장

표 4. 의미있는 질의문 ① 특징을 가진 예

원 문	고유명사인식기술 적용
The FIFA collection is held by the National Football Museum in England.	FIFA : <Organization> in England : <Location>
↓	
자동생성된 질의문	대 답
Where is the FIFA held by the National Football Museum?	in England

표 4에서는 의미있는 질의문을 선별하기 위해 장문형 질의문을 생성하였다. 하단테이블에서 자동생성된 질의문을 살펴보면 주어가 'FIFA'로 명확히 명시되어 있다.

② be동사+과거분사인 수동태에서 생성되고 동사의 결과에 대한 원인이 전치사 뒤에 나오는 문장으로 이 문장의 주어를 질의하는 문장

표 5. 의미있는 질의문 ② 특징을 가진 예

원 문	고유명사인식기술 적용
Pierre was killed in a street in 1906 by carriage	Pierre : <Person> in 1906 : <Time>
↓	
자동생성된 질의문	대 답
Who was killed in a street in 1906 by carriage?	Pierre

표 5에서 원문의 구조는 “의문사+동사+주어+..+by +..?”의 형태로 이루어진 수동태이고 주어를 물어보는 질의문을 생성하였다. 이 문장 뒤에 동사의 원인이 되는 'by carriage'가 나와 의미있는 질의문이 된다. 추출한 의미있는 질의문과 질의문에 해당하는 대담을 함께 매칭시켜 데이터베이스에 저장한다. 만약 이 문장에서 killed의 원인이 되는 by carriage가 없는 문장 “Who was killed in a street in?” 이 된다면 이 질의문은 의미없는 질의문이 된다.

4. 프로토타입 시스템 구현

지금까지 질의문을 장문형질의문과 단문형질의문으로 나누어 생성하고, 이를 바탕으로 의미있는 질의문의 의미와 이것을 찾는 규칙에 대해 알아보았다. 이 장에서는 이 시스템을 테스트할 수 있는 시스템 구조도와 데이터베이스를 설계하였다. 데이터베이스를 활용하여 사용자에게 자동생성된 의미있는 질의응답 시스템을 제공하는 기반을 마련하고자 한다.

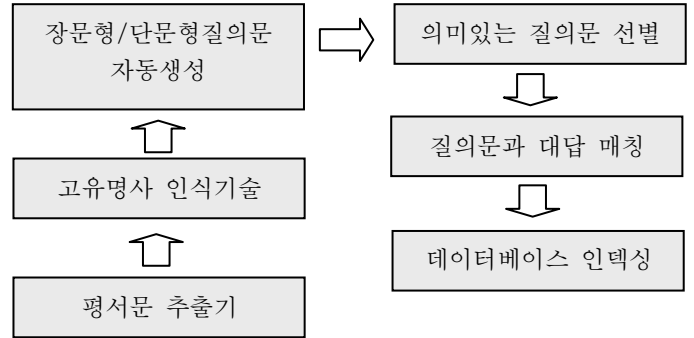


그림 1. 시스템 흐름도

그림 1의 시스템 흐름도에서는 질의어 생성과정과 데이터베이스에 저장하는 모습을 보여준다. 우선, 원본문서에서 평서문을 골라낸다. 이 평서문에 고유명사인식기술을 적용하여 장문형질의문과 단문형질의문을 생성한다. 생성된 질의문 중 의미있는 질의문을 선별하여 이를 질의문과 그에 따른 대담을 매칭시킨다. 이렇게 매칭된 데이터는 데이터베이스에 인덱싱되어 저장된다. 데이터베이스 인덱싱은 데이터 값과 데이터들간의 관계를 인덱스화 하는 의미이다. 데이터베이스에 인덱싱하여 저장하기 위해서는 데이터를 모델링 하는 작업이 필요하다.

4.1 데이터 모델링

데이터 모델링은 생성된 질의문, 질의문에 대한 대담, 원본문서 등의 데이터들을 데이터베이스에 저장하기 위해 데이터베이스를 설계하는 작업이다.³ 실험을 위한 데이터베이스에는 원본문서에 관련된 내용을 관리하는 ‘웹페이지 정보’ 테이블과, 웹페이지로부터 평서문을 추출하여 질의문을 생성하여 질의문과 대담등과 같이 질의문생성에 관한 내용을 관리하는 ‘생성된 질의문 정보’ 테이블로 구성된다.

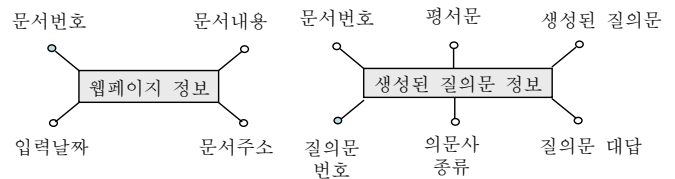


그림 2. E-R 다이어그램

그림 2는 자동응답시스템을 위한 데이터베이스 E-R 다이어그램이다. 웹페이지 정보 테이블은 ‘문서번호’, ‘문서 내용’, ‘입력날짜’, ‘문서주소’를 속성값으로 갖는다. ‘문서번호’는 시스템에서 부여하는 고유번호로 생성된 질의문 정보테이블의 문서번호와 연결된다. ‘문서내용’은 웹페이지의 본문 전체 내용이고, ‘문서주소’는 해당 웹페이지의 URL주소이며, ‘입력날짜’는 데이터베이스에 저장되는 날짜이다. 생성된 질의문 정보 테이블은

³ 데이터베이스관리시스템으로 My-SQL을 사용하였으며, 전체적인 실험의 구현은 JAVA언어와 JDBC(Java Database Connectivity)기술을 사용하였다.

‘문서번호’, ‘평서문’, ‘생성된 질의문’, ‘질의문 대답’, ‘의문사 종류’, ‘질의문 번호’를 속성값으로 갖는다. ‘질의문 번호’는 시스템에서 부여하는 고유번호이다. ‘평서문’은 웹페이지에서 추출한 하나하나의 평서문 내용이고, ‘생성된 질의문’은 평서문에서 자동생성된 의미있는 질의문이다. ‘질의문 대답’은 생성된 질의문에 대한 대답이며, ‘의문사 종류’는 생성된 질의문에 사용된 의문사를 의미한다. 다음 장에서는 저장된 데이터들을 이용하여 사용자가 질의문을 어떻게 선택하는지에 사용자 인터페이스에 대해 살펴보도록 한다.

4.2 사용자 인터페이스

사용자 인터페이스는 자동생성된 질의문을 사용자가 질의를 통해 확인할 수 있는 검색 화면이다. 이 시스템을 개발하기 위해 사용한 데이터는 위키피디아 페이지이며 주로 ‘백과사전’ 성격을 띠는 것이다. 본 논문이 제안하는 시스템은 이러한 ‘백과사전’ 성격의 웹페이지로부터 소스를 제공받아 시스템을 구축하였다. 데이터베이스에 저장한 질의문과 대답은 웹페이지를 통해 검색과 확인이 가능하다.

Question & Answering System

	no	Question	Answer
What	1	What do George Walker Bush do the United States of America?	president
	1	What did George Walker Bush from 1995 to 2000?	Governor of Texas
When	1	When was bush born ?	July 6, 1946
	2	When was bush inaugurated as president?	January 20, 2001
	3	When was George Walker Bush born?	July 6, 1946
	3	When was George Walker Bush served as the forty-sixth Governor of Texas?	from 1995 to 2000
Where	1	Where defeated bush ann richards and was elated in 1994?	Texas
	3	Where is George Walker Bush the forty-third and current President?	United States of America
	3	Where did George Walker Bush previously served as the forty-sixth Governor?	Texas
Who	1	Who is bush ?	president of the united states of america
	3	Who's son is George Walker Bush?	George Herbert Walker Bush
	3	Who was George Herbert Walker Bush?	United States President

그림 3. 웹페이지를 통해 검색결과를 보여주는 화면

그림 3에서와 같이 사용자가 검색어를 입력하면 이미 생성된 질의문이 의문사별로 구분되어 웹페이지에서 보여진다. 사용자는 키워드로 쉬운 질의를 하고 이 질의에 대한 결과로 이미 생성된 질의문을 얻게 된다. 사용자는 이미 생성된 질의문을 보고 자신이 원하는 질의문을 선택할 수 있다. 또한 질의문과 질의문에 대한 대답을 매칭시켜 한 페이지에 보여줌으로써 질의문과 대답을 쉽게 볼 수 있도록 화면을 구성하였다. 사용자가 만약 질의문에 대한 상세한 정보를 알고 싶을 때에는 원하는 질의문을 클릭하면 된다. 해당 질의문의 상세정보에는 질의문을 생성한 평서문문장과, 원본 웹페이지의 전체내용과

원본 웹페이지의 주소까지 알 수 있다. 이러한 정보 들은 위에서 언급한 데이터베이스에 저장되어 있으므로 사용자에게 제공될 수 있다. 지금까지 생성된 의미있는 질의문을 사용자 인터페이스 화면을 통해 보여지는 내용을 살펴보았다. 다음 장에서는 고유명사인식기술을 적용하여 생성된 질의문들의 실험결과에 대해 살펴보도록 하겠다.

4.3 실험결과

이 실험에서 가장 중요한 이슈는 평서문에서 의미있는 질의문을 생성하는 것이다. 이때, 의미있는 질의문은 장문형질의문과 단문형질의문으로 구분되어 생성될 수 있다.

표 6. 의미있는 장문형질의문의 예

no	평서문	질의문	대답
1	<u>Sejong the Great</u> was the fourth king of the <u>Joseon Dynasty of Korea</u> .	Who was the fourth king of the Joseon Dynasty of Korea?	Sejong the Great
2	UOS was established <u>in 1918</u> .	When was UOS established?	1918
3	<u>Korea</u> is currently divided into <u>North Korea and South Korea</u> .	What is currently divided into North Korea and South Korea?	Korea
4	Albert Einstein was born into a Jewish family <u>in Ulm, Württemberg, Germany on March 14, 1879</u> .	Where was Albert Einstein born into a Jewish family in Ulm, Württemberg, on March 14, 1879?	Germany
5	The UN forces crossed into <u>North Korea</u> in early <u>October 1950</u> .	When did UN forces cross into North Korea?	October 1950
6	The <u>Beatles'</u> partner-ship was finally dissolved <u>in 1975</u> .	When was the Beatles' partnership finally dissolved?	1975
7	<u>In 1946</u> , <u>Hermann Hesse</u> received the <u>Nobel Prize in Literature</u>	When did Hermann Hesse receive the Nobel Prize in Literature?	1946
8	One of the best known artifacts of <u>Korea's</u> history of science and technology is <u>Cheom-seongdae</u> , a 9.4-meter high observatory built <u>in 634</u>	What is one of the best known artifacts of Korea's history of science and technology, a 9.4-meter high observatory built in 634?	Cheomseongdae
9	<u>Korean</u> cuisine is probably best known for <u>kimchi</u> , which uses a distinctive fermentation process of preserving vegetables.	What is Korean cuisine probably best known for? It uses a distinctive fermentation process of preserving vegetables.	kimchi

표 6은 의미있는 장문형질의문의 예를 보여준다. 장문형 질의문은 평서문의 내용과 형태를 유지하면서 질의문을 생성한다. 질의하고 싶은 대상에 맞는 의문사를 선택 하여 질의문을 완성시킨다.

표 7. 의미있는 단문형질의문의 예

no	평서문	질의문	대답
1	<u>Albert Einstein</u> was a <u>German-born</u> heoretical <u>physicist</u> .	Who was Albert Einstein?	physicist
2	<u>George Walker Bush</u> is the forty-third and current <u>President of the United States of America</u> .	Who is George Walker Bush?	President of the United States of America
3	<u>Park Ji-Sung</u> is a	Who is Park Ji-sung?	South

	professional South Korean footballer who plays for the English football club Manchester United in the Premier League , as well as the South Korean national football team.		Korean foot-baller
4	Hermann Hesse (2 July 1877 – 9 August 1962) was a German-Swiss poet, novelist, and painter	Who was Hermann Hesse?	German-Swiss poet, novelist, and painter
5	Hesse was born in the Black Forest town of Calw in Württemberg, Germany to a Christian missionary family.	Where was Hesse born ?	the Black Forest town of Calw in Württemberg, Germany
6	The Korean alphabet hangul was also invented during this time	What was the Korean alphabet?	hangul
7	Dr.Martin Luther King,Jr. was assassinated on April 4, 1968, in Memphis, Tennessee.	When was Dr.Martin Luther King,Jr. assassinated?	April 4, 1968
8	Albert Einstein was a German-born theoretical physicist.	Where was born Albert Einstein?	German

표 7은 의미있는 단문형질의문의 예를 보여준다. 단문형 질의문은 평서문의 내용 중 일부만을 이용해 생성되고, 질의하는 대상에 대한 명확한 의미전달이 될 수 있도록 질의문을 생성한다. 표 6과 표 7의 평서문에서 굵은 글씨체는 고유명사 인식을 수행한 결과를 표시한 것이다. 이 중 밑줄 표시한 고유명사는 그 명사가 질의문의 답변이 될 수 있는 고유명사이다. 밑줄 표시한 고유명사를 확인하여 그 명사와 관련된 의문사를 이용하여 질의문을 완성한다. <Time>이나 <Location>, <Miscellany> 등에 포함된 'in', 'by', 'on' 등의 전치사는 대답할 때는 생략 가능하다. 예를 들어, 표 6의 2번에서 보면 고유명사 in 1918은 < Time>을 나타낸다. 이를 이용해 <Time>을 질의하는 질의문은 "When was UOS established?" 이고 이 문장에 대한 대답은 in 1918에서 'in'이 생략된 "1918"이 된다. 또한 질의문에 대한 대답을 할 때 고유명사 인식결과가 동급이고 'of'나 'and', '(, (점표))' 로 연결된 고유명사는 그것들을 합쳐 하나의 대답이 가능하다. 예를 들어, 표 7의 4번에서 보면 고유명사 German-Swiss poet, novelist, and painter는 <Position>을 나타내고 이것들은 ','와 'and'로 연결되어 있어서 'German-Swiss poet, novelist, and painter'라는 하나의 대답이 가능한 것을 확인 할 수 있다.

5. 결론

본 논문에서 제안하는 질의문 자동생성방식에 의한 질의응답시스템은 사용자가 질의하고자 하는 의미있는 질의문을 미리 생성하여 사용자의 편의성을 도모하는 시스템이다. 이 시스템은 장문형질의문과 단문형 질의문으로 구분지어 의미있는 질의문을 생성하였다. 이 시스템에서의 의미있는 질의문은 하나 또는

유한개의 답변을 제공할 수 있는 구체적 의미를 가지는 문장을 의미한다. 본 제안시스템을 활용할 시, 사용자가 질의를 입력하면 미리 생성한 의미있는 질의문을 보여주어 사용자에게 자신이 원하는 질의문을 선택할 수 있게 된다. 이는 일반적인 질의응답시스템의 정확도를 개선하여 사용자에게 질의어 작성에 대한 부담을 줄일 뿐만 아니라 질의어에 대한 정확한 답변을 미리 준비할 수 있어 시스템의 신뢰도를 높이는 데 크게 기여할 수 있다.

향후 연구과제로서, 보다 정확하고 의미있는 질의문을 생성할 수 있도록 더 많은 문장의 패턴과 규칙을 찾는 일이 요구되므로 이에 따른 연구와 시스템 구축을 위한 노력이 진행될 것이다.

6. 감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 육성·지원사업(IITA-2008-C1090-0801-0031)의 연구결과로 수행되었음.

7. 참고 문헌

[1] Rohini Srihari, Wei Li. "A Question Answering System Supported by Information Extraction". *Proceedings of the sixth conference on Applied natural language processing*. Pages: 166 - 172 . 2000.

[2] Susan Dumais, Michele Banko, Eric Brill, Jimmy Lin, Andrew Ng. "Web question answering: is more always better?". *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. Pages: 291 - 298. 2002..

[3] Jennifer Chu-Carroll, John Prager, Christopher Welty, Krzysztof Czuba, David Ferruci. "A multi-strategy and multi-source approach to question answering". *The Eleventh Text REtrieval Conference*. Pages: 281 - 288. 2003.

[4] John Prager, Jennifer Chi-Carroll, Krzysztof Czuba, Christopher Welty, Abraham Ittycheriah, Ruchi Mahindru. "IBM's PIQUANT in TREC2003". *Proceedings of TREC2003*. 2003.

[5] *The New York Times*, December 26, 2004.

[6] R Florian, A Ittycheriah, H Jing, T Zhang . "Named entity recognition through classifier combination". *Proceedings of CoNLL-2003, 2003*. Pages: 168 - 171 . 2003.