



질의응답 기술 개발

황 이 규* 김 현 진* 장 명 길**

목 차

1. 서 론
2. 질의응답 기술 동향
3. 주요 질의응답 시스템
4. 백과사전 인물 질의응답 시스템
5. 결 론

1. 서 론

인터넷 기술의 발전과 함께 축적된 방대한 정보에 대하여 사용자는 이전보다 더욱 정확하고 빠르게, 그리고 원하는 정보를 더욱 편리하게 얻을 수 있는 새로운 정보검색 기술을 기대하고 있다. 최근 사용자의 질문 의도를 파악한 후, 검색 대상 문서로부터 답을 찾아 제공하는 질의응답형 정보검색 기술(이하 질의응답 기술)에 대해 많은 관심이 집중되고 있다.

질의응답 기술은 사용자의 자연어 질문과 검색 대상 문서의 의미를 파악하기 위한 고정밀 자연어 처리 기술과 대상 문서로부터 답을 추출하기 위한 정보추출 기술을 필요로 하며, 많은 후보 문서들로부터 답을 포함하는 문서를 걸러주는 역할을 위해 기존의 문서검색 기술도 활용한다. 질의응답 기술은 현재 정형 데이터베이스에 담겨 있는 상품 정보나 기업 정보를 찾아주는 데이터베이스 질의응답과 게시판 혹은 전자메일 정보를 찾아주는 FAQ(Frequently Asked Question) 질의응답에 부분적으로 활용되고 있다. 하지만, 고정밀 자연어 처

리, 정교한 정보추출 등의 텍스트 마이닝 기술이 요구되는 백과사전 지식정보나 전자 매뉴얼 정보에 대한 질의응답 기술은 현재 국내외적으로 연구가 진행 중이다.

본 논문에서는 먼저 질의응답 기술의 동향을 살펴보고, 현재 개발된 주요 질의응답 시스템을 간략히 소개한다. 특히, ETRI의 백과사전 인물분야 질의응답 시스템 개발을 통해 국내 질의응답 기술 개발의 현 주소를 살펴본다.

2. 질의응답 기술 동향

2.1 정보검색 기술의 발전 동향

기존의 정보검색은 대량의 문서에서 사용자의 질의에 적합한 문서를 찾아내는 문서검색에 초점을 맞추어 왔다. 검색 대상 문서의 양이 많아짐에 따라, 검색의 결과로 나타나는 문서의 양은 검색 시스템 사용자에게 큰 부담이 되었다.

정보검색의 기술적 한계에 직면한 많은 검색 포털 업체들이 새로운 검색 서비스 모델로 "지식검색"이라는 개념을 도입하였는데, 2003년도 국내 검색시장의 대표적인 서비스로 자리매김하게 되었다. 지

* 한국전자통신연구원 지식마이닝연구팀 선임연구원

** 한국전자통신연구원 지식마이닝연구팀 팀장

식검색은 사용자의 질문에 다른 사용자가 답을 달고, 많은 사용자들이 이 지식의 유용성에 대해 평가함으로써, 네티즌들이 지식을 공유하는 수단으로 크게 각광을 받고 있다. 대부분의 인터넷 포털 사이트 업체는 지식검색을 중요한 서비스의 하나로 제공하고 있는데, 네이버(www.naver.com)의 '지식iN', 엠파스(www.empas.com)의 '지식거래소', 세이클럽(www.sayclub.com)의 '세이테마', 야후(www.yahoo.co.kr)의 '야후!지식검색' 등이 널리 알려져 있다.

지식검색 서비스의 장점은 실생활에 도움이 되는 실용적인 질문과 답변이 많고, 정보의 시사성에 민감하다는 것이다. 이런 장점에도 불구하고 지식검색이 가지고 있는 문제점은 다음과 같다[16]. 첫째, 네티즌 다수가 인정하면 '답'으로 채택된다는 것이다. 또한, 질문자가 선택한 답변이 마치 실제 정답인 것처럼 왜곡될 수 있다. 둘째, 방대한 양의 질문과 답변 중에서 신변잡기적인 내용이 많으며, '핼(피움)'과 '핼'을 거듭하면서 출처가 불분명해진 정보들은 지식검색의 신뢰도를 크게 떨어뜨린다. 셋째, 흥미를 끌기 힘든 주제의 질문에 대해서 충실한 답변을 기대하는 것에는 무리가 따른다.

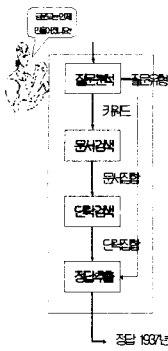
질문에 대한 정답을 사람이 제시하는 지식검색과는 달리, 질의응답 기술에서는 사용자의 질문에 대해 질의응답 시스템이 문서 등으로부터 정답을 추출하여 사용자에게 제시한다. 질의응답 시스템의 목적은 질문에 대해 이를 만족하는 문서의 리스트를 찾기보다는 정답을 추출해내는 것이다[13]. 이는 단순히 키워드에 기반한 정보 검색이 아니라, 사용자 질문의 정교한 분석, 검색엔진을 통한 문서 및 정답후보단락 추출, 단락에서 원하는 정답을 추출하는 과정으로 이루어져 있다. 또한, 질의에 대한 정답이 사실에 기반한 정확한 정답인지를 검증하기 위한 방법도 연구되고 있다.

질의응답 기술은 현재 인터넷 쇼핑몰에서 데이

터베이스에 담겨있는 상품정보 등을 찾아주거나 고객의 소리와 같은 게시판 또는 FAQ 등을 검색하는데 부분적으로 이용되고 있다. 문서검색이 문서로부터 필요한 정보를 찾는 것임에 반해, 데이터베이스에서 사용자가 원하는 자료를 검색하는 데이터베이스 자연어 질의처리 기술은 자연어를 DB의 SQL과 같은 질의처리 언어를 대신하도록 하는 것이다. 초창기의 대표적인 DB를 위한 자연어 인터페이스로는 BASEBALL[6]과 LIFER[7]가 있다. BASEBALL은 야구 통계에 대한 질의처리 시스템으로 "Who did the Red Sox lose to on July 5?"와 같은 자연어 질문을 처리할 수 있었으며, LIFER는 인사통계에 대한 질의처리 시스템으로 "How many professors are there in the computer science department?" 등에 대한 질문을 의미적 수준에서 파싱하고 이를 직접 DB 질의어로 변환하였다. 현재는 DB 질의처리 기술이 사용자의 자연스러운 대화체 질의 문장을 형태소, 의미, 구문 분석하여 사용자의 요구에 맞는 최적의 상품을 검색해주는 상품검색 및 추천 등에 이용되고 있다.

2.2 일반적인 질의응답 기술

질의응답 시스템과 정보검색 시스템의 큰 차이점 중 하나는 자연어를 입력하여 문서를 검색하는 것이 아니라 정답을 찾는 것에 있다. 이를 위해 질문의 처리과정에서 사용자가 원하는 정답이 무엇인지 질의의도를 파악할 수 있는 질의유형이나 키워드 등의 정보를 질의로부터 추출한다. 또한, 기존의 정보검색 방법에 의해 질의와 유사한 문서를 추출하고, 문서에서 다시 정답을 포함할 가능성이 있는 단락을 추출한 후, 단락에서 질의유형과 동일한 개체를 찾아내어 사용자에게 정답으로 제시한다. 일반적인 질의응답 시스템의 구조는 (그림 1)과 같다.



(그림 1) 일반적인 질의응답 시스템의 구조

- 질문분석: 질문이 요구하는 정답의 유형을 결정하고, 검색엔진을 위한 키워드를 추출
- 문서검색: 키워드를 이용하여 관련 문서를 검색엔진을 통해 검색
- 단락검색: 검색된 문서에서 키워드와 관련이 있는 단락을 검색
- 정답추출: 단락에서 질문이 요구하는 정답유형과 관련이 있는 개체를 정답으로 추출하여 사용자에게 제시

2.3 질의응답 시스템 평가

<표 1> TREC-8의 질문유형 분포

질문유형	질문수	질문유형	질문수
What	64	Which	10
Who	47	What	4
How	31	Why	2
Where	22	Whom	1
When	19		

질의응답 시스템을 평가하는 대표적인 것이 미국 NIST(National Institute of Standards and Technology)가 주관하는 질의응답 평가대회 TREC(Text REtrieval Conference)의 QA Track이다. 1999년부터 시작된 QA Track에서는 질의응답 시스템을 평가하기 위해 대량의 평가셋을 구축하고, 성능평가를 위한 평가척도를 개발하였다. 첫 대회인 TREC-8에서는 약 53만 문서의 코퍼스로부

터 수작업으로 작성한 200질문에 대해서 50바이트 및 200바이트 크기의 정답을 포함한 단락을 제시하도록 하였다. TREC-8에서 제시된 질문을 유형에 따라 분류하면 <표 1>과 같은 분포를 가진다.

TREC-8에 제시된 실제 질문의 예는 다음과 같다.

- "Who is the author of the book, 'The Iron Lady: A Biography of Margaret Thatcher'"
- "What was the monetary value of the Nobel Peace Prize in 1989?"
- "What does the Peugeot company manufacture?"
- "How much did Mercury spend on advertising in 1993?"

TREC-9에서부터는 검색엔진의 실제 로그를 이용하여 질문을 작성하였는데, TREC 2001의 경우, 약 25% 정도가 정의(definition)에 관련된 질문이었다. 질의응답 시스템의 성능을 평가하기 위한 척도로 MRR(Mean Reciprocal Rank)을 사용하였다. MRR은 질의응답 시스템에서 결과로 나타난 것에서 정확한 대답이 포함되어 있는 것이 몇 번째 순위에 나타나는지를 계산해서 그 순위의 역수를 평균한 것이다. 여기서 rank_i는 i번째 질문에 대한 응답으로 제시한 것들 중에서 첫 번째로 정답인 것의 순위이고, n은 질문의 수다.

$$MRR = \frac{1}{n} \left(\sum_i \frac{1}{rank_i} \right)$$

TREC 2002에서는 정답을 포함한 문자열을 제시하는 것이 아니라 실제 정답을 제시하도록 요구하였고, CWS(Confidence-Weighted Score)를 시스템에 대한 새로운 평가척도로 사용하였는데, 이의 정의는 아래와 같다.

$$\frac{\sum_{i=1}^n ic/i}{n}$$

(ic: 첫 i번째 질문에서의 정확한 정답수)

TREC QA 로드맵[2]에 따르면 정답이 단순히 문장의 일부분을 제시하는 것이 아니라 문장들 사이의 추론이 필요한 질문, 새로운 문장을 생성하여 이를 정답으로 제시하거나, 주어진 정답에 대한 배경 설명, 정답의 정당성 검증, 정답의 모호성 해결, 전문가 수준의 의견 제시가 필요한 질문, 이질적 정보의 통합을 통한 정답의 제시 등 점점 질의응답 시스템의 난이도를 높여 갈 계획이다. 이때 가능한 질문의 형태는 아래와 같다.

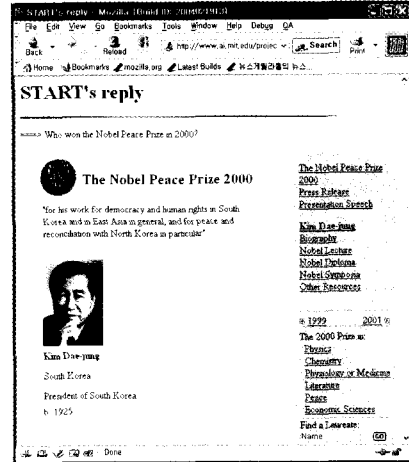
- “AMTRAK의 금융적인 상황은 어떠한가?”
- “유로화에 대한 덴마크인들의 의견은 무엇인가?”
- “최근에 많은 사람들이 사륜구동차량을 구입하는 이유는 무엇인가?”
- “연방준비제도가이사회가 다음 모임에서 금리를 인상할 것 같나요?”

3. 주요 질의응답 시스템

3.1 START

START[10, 11]는 1993년부터 제공된 온라인 질의응답 시스템으로 구조화 또는 반 구조화된 웹 데이터들을 통합한 가상의 데이터베이스인 Omnibase를 이용한 자연어 질의응답 시스템이다. 이 시스템은 다수의 웹사이트에 대한 단일한 질의응답 인터페이스로 볼 수 있는데, [4]를 정보의 원천으로 하여 각국에 대한 정치, 경제, 지리적 정보에 대한 질의응답을 제공하며, [1]로부터 25,000여 명의 인물에 대한 정보를, [8]로부터 수십만편의

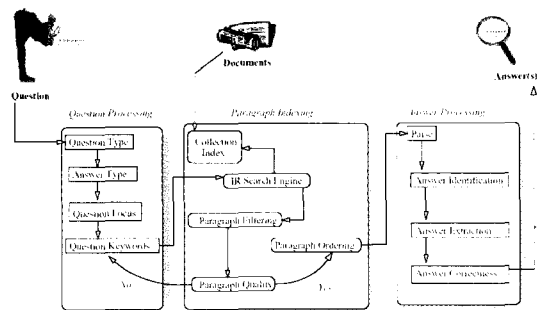
영화에 대한 전반적인 정보, [12]를 이용하여 각종 정의와 관련된 정보 등을 제공한다. START에서 질문 “Who won the Nobel Peace Prize in 2000?”에 대한 정답으로 (그림 2)를 제시하고 있다.



(그림 2) START의 정답제시 화면

3.2 LASSO

LASSO[5]는 TREC-8 QA Track에서 우수한 성능을 보여준 시스템으로 (그림 3)과 같은 구조를 가지고 있다.



(그림 3) LASSO의 구조

이 시스템의 주요 특징은 정답의 추출을 위해 질문분석을 통해 정답유형을 정의한 것인데, 질의처리 과정에서 1) 질문의 유형 정의, 2) 예상되는 정답

의 유형 정의, 3)정답의 포커스 선택, 4)검색 엔진을 위해 질문을 질의로 변환하는 역할을 수행한다. 예를 들어, "Which city is has the oldest relationship as sister-city with Los Angeles?"라는 질문의 분석 과정에서 얻어진 정보는 다음과 같다.

- 1) 질문유형: Which-Where
- 2) 정답유형: Location
- 3) 포커스: city
- 4) 키워드: city, oldest, relationship, sister-city, Los Angeles

정답처리과정에서는 정답유형의 대상이 되는 개체명, 화폐단위, 날짜, 시간/장소 표현, 상품명 등을 인식하기 위해 Lexico-Semantic 정보를 이용하였다. 이를 통해 얻어진 정답후보들을 대상으로 다양한 가중치 함수를 두고 이를 결합하여 정답을 추출하였다.

3.3 국내 주요 질의응답 시스템

[15]는 한국어 질의응답을 위한 시스템으로, 질의해석을 통해 날짜, 범위, 핵심어, 중요어, 질의유형을 파악하고, 단락검색에서 질문해석 과정에서 얻어진 정보를 이용하여 문서에서 대답을 포함할 것 같은 단락을 검색하며, 대답추출에서 여러 언어 자원을 이용하여 개체명을 추출한 후, 비사실 문맥에서 나타나는 대답제거, 대답확인 및 대답해석을 통해 정답을 찾아낸다. 미리 정의한 대답유형에는 '사람', '기관', '장소(국가, 도시, 장소)', '시간(년, 월, 일, 계절)', '숫자', '숫자+단위', '방법', '이유' 등이 있다. 주요한 특징으로 메타-사실 프레임을 통해 대답을 확인하는 부분인데, 질문에 나타난 어휘를 모두 포함하지 않더라도, 메타-사실 프레임을 만족하는 패턴을 이용하여 정답에 대한 확신을 증가시킨다. 예를 들어, "허준의 동의보감"은

'〈PERSON〉의 〈BOOK, PEOM, WRITING〉'이라는 프레임을 통해 "허준"이 "동의보감"을 쓴 저자임을 검증하는 것이다.

[14]는 MAYA라는 질의응답 시스템을 개발하였는데, 이 시스템에서는 개체명 사전과 LSP (Lexico-Syntactic Pattern)을 이용하여 개체명을 인식하고, 이를 질의응답 시스템이 정답 가능한 후보로 미리 색인하였다. 사용자의 질의유형을 105가지의 의미범주로 구분하고, 이에 따라 정답유형을 분류하였으며, Lexico-Syntactic Parser를 이용하여 사용자의 질의유형을 분석하여 색인된 정답DB에서 정답후보를 순위화하고 이를 정답으로 제시하였다.

4. 백과사전 인물 질의응답 시스템

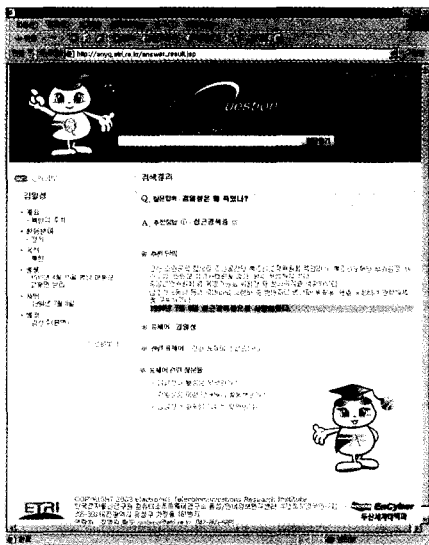
ETRI에서는 2003년부터 백과사전 대상의 질의응답 시스템을 개발하고 있다. 본 장에서는 지난해에 개발한 백과사전 인물분야 질의응답 시범 시스템 AnyQuestion 1.0(<http://AnyQ.etri.re.kr>)을 소개한다.

4.1 AnyQuestion 1.0의 특징

AnyQuestion 1.0은 ㈜두산이 공동연구로 참여하여 제공하는 '두산세계대백과 엔사이버'를 활용하고 있다. 백과사전 인물분야 질의응답 시스템은 인물분야 약 25,000 표제어에 관한 사용자 질문에 대해 질문의도를 파악하여 단답형의 정답을 제시한다. AnyQuestion은 정답이 본문의 문장을 통해서 추출되었을 경우에는 해당 문장을 함께 제시하여, 사용자가 제시된 정답의 정확성 유무를 확인할 수 있게 하며, 현재 질문과 관련된 추가 질문들을 추천하는 기능 및 백과사전 인물에 대한 주요한 정보를 간략히 제시하는 기능 등을 제공하고 있다. 주요한 특징은 다음과 같다.

- 사용자 질문에 대한 3단계 정답 추출 및 제시
- 정보추출을 통한 백과사전 지식베이스(KB) 반자동 구축
- 사용자 질문과 관련이 있는 연관질문 자동 생성 및 제시

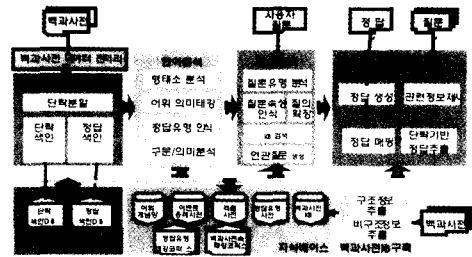
(그림 3)은 사용자 질문에 대한 AnyQuestion의 정답제시 화면이다.



(그림 3) AnyQuestion 시스템 정답 제시 화면

4.2 시스템구성

백과사전 인물분야 질의응답 시스템의 전체 흐름은 크게, 1) 사용자의 질문을 예측하여 백과사전 문서를 분석하여 정답이 될 만한 부분을 미리 색인하고, 정답을 포함한 문장들을 효과적으로 저장하는 부분과 2) 사용자 질문을 실시간으로 입력 받아서, 질문자의 의도를 파악하고, 색인된 데이터베이스를 중심으로 정답으로 추정되는 부분을 추출해서 제시하는 부분으로 나뉠 수 있다.



(그림 4) AnyQuestion1.0 시스템 구성도

이러한 기능을 처리하기 위해서, 본 시스템은 다음과 같은 모듈들로 구성되어 있다.

4.2.1 질문분석

질문분석 모듈은 사용자의 질문에 대해, 질문의 유형(5W1H유형)과 사용자가 얻고자 하는 정답의 유형(예: 인명, 지명, 날짜 등)을 인식한다. 또한, 정보추출기술을 통해 백과사전으로부터 추출한 백과사전 지식베이스를 검색하기 위한 지식베이스 질의어(KBQ)를 생성하며, 단락검색을 통한 정답 추출에 용이하도록 질문에서 주요 명사 및 용언에 대해 유사어를 확장한다. 그리고, 입력된 질문과 관련이 있는 연관 질문을 생성하여 사용자에게 추천한다. 이러한 과정을 통해 사용자는 본인이 입력한 질문 외에도 관련된 다른 정보를 부가적으로 얻을 수 있다.

4.2.2 언어분석

본 시스템에서는 형태소분석기술뿐만 아니라, 명사 어휘개념망에 기반한 어휘 의미 태깅, 정답유형 인식을 위한 확장된 개체명 인식, Logical Form(LF)를 만들기 위한 구문분석기술 등의 고정밀 언어분석기술을 적용하였다. 이러한 기술을 통해, 질문분석 및 정답 색인/제시 모듈에서 사용자의 질문이나 백과사전의 본문에 나타난 자연어 문장이 정확하게 분석되도록 하였다.

4.2.3 백과사전 KB 반자동 구축

인물 분야의 백과사전에 대한 질의응답이라는

특성을 고려하여, 사용자가 관심이 있어하는 인물 속성(예: 고향, 수상경력, 출생일, 발명품 등)을 인물의 범주별(과학자, 종교인, 정치인 등)로 미리 정의하고, 이에 따른 템플릿을 정의하였다. 수동으로 백과사전 문서를 대상으로 속성 태깅을 하여 학습 문서를 구축하고, 이 문서를 통해 속성 추출 규칙을 생성한 후, 정보추출기술을 이용하여 템플릿을 완성하는 방법으로 백과사전 KB를 구축하였다.

4.2.4 정답색인

언어분석을 통해서 백과사전 원문을 분석한 후, 단락기반 색인 방식과 LF 기반 정답 색인 방식을 이용하여, 색인DB에 저장한다.

- 단락기반 색인: 각 문장에 의미적인 주제를 부여하여 원문을 의미적 단락으로 재배치하고, 키워드, 명사구 및 용언 등의 단위로 단락을 검색할 수 있도록 색인
- LF기반 색인: 주요 용언을 중심으로 용례에 기반하여 격들을 정의하고, 용언별 정보추출 규칙을 정의하며, 이 규칙을 이용하여 입력문장에서 who, when, where, what, why, with 등의 단위를 색인

4.2.5 정답제시

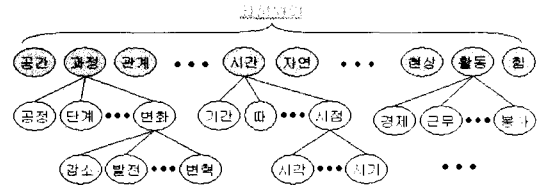
사용자의 질문에 대한 질문분석 결과를 입력으로 색인 DB에서 해당 질문에 맞는 정답을 추출하고 제시하는 기능이다. 정답제시는 다음의 3단계의 정답추출 과정을 단계별로 거치면서 질문에 대한 정답을 추출한다.

- KB기반 정답 추출: 백과사전에서 KB에 저장되어 있는 정답으로 매핑되는 경우에 정답 추출
- LF기반 정답 추출: 질문이 용언을 기반으로 구조화 될 수 있는 경우, 질문 LF 구조로 변경한 후, LF 기반 정답 색인 DB를 검색하여 정답 추출

- 단락기반 정답 추출: 앞서 두 단계에서 정답을 추출하지 못하는 경우, 기존 단락검색 방식을 이용하여 실시간으로 의미적인 단락을 검색, 언어 분석 및 통계적인 정답 추정 방식을 이용하여 정답 추출

4.2.6 지식베이스

질의응답 시스템 및 그 밖의 다양한 언어처리 시스템을 지원하기 위한 언어 자원을 가리킨다. 지식베이스는 명사 어휘개념망, 하위 범주화 사전, 정답 유형 사전, 이벤트 용례 사전 및 기타 학습 코퍼스 등으로 구성된다. 이 중 명사 어휘개념망은 (그림 5)에서처럼 상하위 관계를 이용하여 트리형태로 구성되어 있으며, 유의어, 동의어, 반의어, 부분-전체 관계 등의 정보가 추가로 명시되어 있다. 총 노드수는 5만여 일반 어휘이고, 개체명 단위의 어휘를 포함한 백과사전 어휘를 합산하면 29만여 어휘를 포함하고 있다.



(그림 5) ETRI 명사 어휘 개념망

4.3 시스템 평가

AnyQuestion 1.0을 객관적으로 평가하기 위하여, 백과사전 질의응답 평가셋(총 200질문)을 구축하였다. 다양한 형식의 질문과 난이도에 따른 정답 추출 능력을 평가할 수 있는 테스트컬렉션을 구축하기 위해서, 일반인 6명이 25개 인물 세부 범주(예: 인물 직업별, 시대별, 지역별 그룹 등)에 고루 분포한 질문을 선정하고 문장 내에서 정답을 추천하여 구축하게 하였다.

질의응답 시스템을 평가하기 위해 MRR과 사용자 정답 만족도를 사용하였다. 사용자 정답 만족도란 정답뿐만 아니라 시스템에서 제공하는 부가정보(요약, 문장, 단락 정보 등)를 통해 사용자가 원하는 정답을 찾을 수 있는지의 여부를 평가한 것이다. 시스템 성능 평가결과는 MRR 0.51, 사용자 만족도 0.74였다.

성능을 분석한 결과, 정답색인 관점에서 KB 및 LF 기반의 정답추출 방법은 정확도 수치는 높은 반면, 전체 질문에 대한 커버리지는 낮았다. 실시간으로 정답을 추정하는 단락기반 정답추출은 다소 낮은 정확도를 나타내기는 하나, 앞의 두 단계 보다 커버리지가 높았다. 향후 빠르고 비교적 정확한 정답 제시 기능을 가진 KB 및 LF 기반의 정답 제시 방법에서의 질문 처리 능력을 향상시키고, 단락기반 정답 추출에서의 정확도를 향상 시킬 수 있는 방법을 연구할 예정이며, 백과사전 전체 지식정보에 대한 질의응답 연구로 확대할 계획이다.

5. 결 론

질의응답은 기존의 문서 검색과는 달리 사용자의 질문에 대하여 정확히 답을 찾아 제공한다. 이를 위해서는 기본적으로 사용자의 다양하고 복잡한 질문 유형을 적절히 파악하고 검색 대상 문서의 의미적 내용을 정확히 분석하기 위하여 방대한 언어지식을 이용한 고정밀 자연어처리 기술이 필요하며, 단답형, 문장형, 나열형 등의 여러 형태의 정답을 제시할 수 있어야 한다.

ETRI의 백과사전 인물 질의응답 시범 시스템 AnyQuestion 1.0은 3단계 정답 매핑 모델을 통해 사용자에게 보다 높은 정답 만족도를 제공하고자 하였다. 특히, 백과사전 지식정보를 미리 지식베이스로 반자동 구축하기 위해 확장 개체명 인식 기술을 이용한 정교한 정보추출 기술을 이용하였다.

앞으로 질의응답 기술은 백과사전이나 전자 매

뉴얼을 대상으로 한 질의응답 외에도, 방대한 양의 웹 정보를 대상으로 한 웹 질의응답, 다국어 텍스트 대상의 다국어 질의응답, 텍스트뿐만 아니라 멀티미디어 정보에 대한 멀티미디어 질의응답, 그리고, 목적 지향의 대화 진행을 통한 대화형 질의응답 등의 종류로 계속 연구가 진행되어, 다양한 분야에서 차세대 정보검색기술로서 사용자에게 보다 큰 만족을 제공할 것으로 기대된다.

참고문헌

- [1] Biography.com, <http://www.biography.com>
- [2] J. Burger, C. Cardie, et.al. "Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A)", NIST DUC Vision and Roadmap Documents, <http://www-nlpir.nist.gov/projects/duc/roadmapping.html>, 2001.
- [3] J. P. Callan, "Passage-Level Evidence in Document Retrieval", Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 302 - 310, 1994.
- [4] CIA The World Factbook, www.cia.gov/cia/publications/factbook/
- [5] D. Donovan, S. Harabaqiu, M. Pasca, R. Mihalcea, R. Goodrum, R. Girju and V. Rus, "LASSO: A Tool for Surfing the Answer Net", In Proceedings of TREC-8, pp. 65-74, 1999.
- [6] Bert Green, Alice Wolf, Carol Chomsky, and Kenneth Laughery, "BASEBALL: An Automatic Question Answerer", In Proceedings of the Western Joint Computer Conference, 1961.
- [7] Gary G. Hendrix, "Human Engineering for

Applied Natural Language Processing”, In Proceedings of the 5th International Joint Conference on AI(IJCAI-1977), 1977.

- [8] Internet Movie Database, <http://www.imdb.com>
- [9] M. Kaszkiel, J. Zobel, “Passage Retrieval Revisited”, Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 178-185, 1997.
- [10] Boris Katz, Jimmy J. Lin, “START and Beyond”, In Proceedings of 6th World Multiconference on Systemics, Cybernetics, Informatics(SCI 2002), 2002.
- [11] Jimmy Lin, Boris Katz, “Question Answering Techniques for the World Wide Web”, 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2003), 2003.
- [12] Merriam-Webster Dictionary, <http://www.m-w.com>
- [13] Ellen M. Voorhees, “The TREC-8 Question Answering Track Report”, http://trec.nist.gov/pubs/trec8/papers/qa_report.pdf
- [14] 김학수, 서정연, “2-패스 색인 기법과 규칙 기반 질의처리 기법을 이용한 고속, 고성능 질의 응답 시스템”, 정보과학회논문지: 소프트웨어 및 응용, 제29권 제11호, pp. 795-802, 2002
- [15] 이경순, 김재호, 최기선, “KorQuA: 질의응답에서 자료유형을 고려한 대담검색과 대담해석”, 한글 및 한국어 정보처리학술대회, 2000.
- [16] 이강룡, “지식검색이라는 자유공간”, <http://www.hani.co.kr/section-001055000/2004/01/001055000200401081917067.html>

저자약력



황 이 규

1993년 전북대학교전자계산학과(이학사)
 1995년 전북대학교 전산통계학과(이학석사)
 2001년 전북대학교전산통계학과(이학박사)
 2001년 현재 한국전자통신연구원 지식마이닝연구팀
 선임연구원
 관심분야 : 자연어 처리, 텍스트마이닝, 정보검색
 이 메 일 : yghwang@etri.re.kr



김 현 진

1995년 부산대학교 전자계산학과 (이학사)
 1997년 부산대학교 전자계산학과 (이학석사)
 1997년 ~ 현재 한국전자통신연구원 지식마이닝연구팀
 선임연구원.
 관심분야 : 정보검색, 질의응답, HCI
 이 메 일 : jini@etri.re.kr



장 명 길

1988년 부산대학교 계산통계학과 (이학사)
 1990년 부산대학교계산통계학과 (이학석사)
 2002년 충남대학교 컴퓨터과학과 (이학박사)
 1990년 ~ 1998년 5월 시스템공학연구소 선임연구원
 1998년 6월 ~ 현재 한국전자통신연구원 지식마이닝연구팀 팀장
 관심분야 : 정보검색, 질의응답, 자연어처리
 이 메 일 : mgjang@etri.re.kr