KIPS Tr. Software and Data Eng.
Vol.3, No.4 pp.149~154 pISSN: 2287-5905

http://dx.doi.org/10.3745/KTSDE.2014.3.4.149

Question and Answering System through Search Result Summarization of Q&A Documents

Dong Hyun Yoo† · Hyun Ah Lee††

ABSTRACT

A user should pick up relevant answers by himself from various search results when using user participation question answering community like Knowledge-iN. If refined answers are automatically provided, usability of question answering community must be improved. This paper divides questions in Q&A documents into 4 types(word, list, graph and text), then proposes summarizing methods for each question type using document statistics. Summarized answers for word, list and text type are obtained by question clustering and calculating scores for words using frequency, proximity and confidence of answers. Answers for graph type is shown by extracting user opinion from answers.

Keywords : Question and Answering, Question Clustering, Answer Summarization

Q&A 문서의 검색 결과 요약을 활용한 질의응답 시스템

유 동 현† · 이 현 아††

요 약

지식iN과 같은 사용자 참여 질의응답 커뮤니티에서 원하는 질문에 대한 답을 찾기 위해서는 검색 결과로 제공되는 다양한 문서를 일일이 확인하여 판단하는 과정이 필요하다. 만일 사용자가 원하는 답변을 자동으로 정제하여 제시할 수 있다면, 질의응답의 사용성이 크게 향상될 수 있다. 본 논문에서는 질의응답 데이터 분석을 통해 사용자의 질문의 유형을 단어, 목록, 도표, 글의 4가지 유형으로 분류하고, 문서 내 통계적 특성을 활용하여 각 분류별 답변을 자동으로 제시하기 위한 방식을 제안한다. 단어, 목록, 글 유형은 질의어에 대해 검색된 질문을 군집화하고, 군집 내 빈도와 질의어에 대한 근접도, 답변 신뢰도 등으로 계산된 답변 내 어휘의 적합도를 활용하여 요약한 답변을 사용자에게 제시한다. 도표형은 답변들에서 사용자의 의견 정보를 추출하여 의견 통계를 도표로 제시한다.

키워드 : 질의응답, 질문 군집화, 답변 요약

1. 서 론

Q&A 게시판 등을 이용한 질의응답 서비스는 지식정보를 교류하는 사용자 참여 서비스로, 사용자의 질문에 대한 다른 사용자들의 답변이 축적된 지식 데이터에 대한 검색 기능을 제공한다. 국내에서 가장 많이 사용되는 질의응답 커뮤니티인 지식iN은 수억 개의 질의응답 데이터를 수집하여 다양한 정보를 제공하고 있다. 하지만, 동일한 사용자 참여 지식 서비스인 온라인 백과사전 위키피디아와 대조적으로 답변에 대한 수정이나 삭제가 자유롭지 않아 신뢰성이 떨어지고 중복된 질문과 답변들에 의해 올바른 답변을 파악하는 데 오랜 시간이 소요된다는 단점이 있다. 이를 보완하기 위해 지식iN은 답변자의 등급 등을 제공하고 있으나, 신뢰성이 보장된 답변의 비율이 낮아 여전히 문제를 가지고 있다. 그럼에도 불구하고 지식검색 서비스는 축적된 질의응답이 위키피디아나 기타 웹 서비스보다 방대하고 다양한 장점 때문에 일반상식부터 전문지식에 이르기까지 다양한 분야에 활용되고 있다.

자동 질의응답 서비스에 대한 기존 연구는 대량의 웹문서에서 질의에 대한 응답을 찾는 문제를 중심으로 접근되고 있다[1, 2, 3]. TREC의 Question Answering Track에서는 질문 유형을 사실(factoid), 목록(list), 기타(others)로 분류하고, 자동 질의응답을 위한 다양한 방법이 시도되었다[4]. 국내에서는 지식iN과 같은 질의응답 문서는 존재하였으나, 질의응답의 성능평가나 질의어 추천, 질문 분류를 중심으로 연구들이 이루어져, 실용적인 결과를 보여주는 질의응답 시스템 사례를 찾기 어렵다. 질의응답 서비스의 성능 평가[5, 6, 7]에서는 질문에 대한 답변의 적절성을 평가하였고, 질의어 추천 시스템[8]에서는 질문 간 단어 유사도를 통해 얻은 유사한 질문으로 검색 결과를 확장하여 시스템 활용도를 높

※ 본 연구는 금오공과대학교 학술연구비에 의하여 연구된 논문임.
† 준 회 원 : 금오공과대학교 컴퓨터공학부 학부생
†† 종신회원 : 금오공과대학교 컴퓨터소프트웨어공학과 부교수
논문접수 : 2013년 11월 13일
수 정 일 : 1차 2014년 1월 21일
심사완료 : 2014년 2월 10일
* Corresponding Author : Hyun Ah Lee(halee@kumoh.ac.kr)

이고자 하였다. 어휘 연관성을 이용한 질문 분류[9]에서는 어휘 연관성을 통해서 질문을 분류하여 질문-질문 쌍에 의한 질문 분류가 질문-답변 쌍에 의한 질문 분류보다 더 뛰어난 분류 결과를 나타냄을 보였다. 확장 나이브 베이지언 분류기를 활용한 질문 분류[10]에서는 질문들을 질문의 목적에 따라 정보, 제안, 의견으로 분류하고, 각 유형으로 수동 분류한 질문 데이터를 학습하여 질문 분류를 시도하였다. 질문-답변의 품질 측정 방법[11]에서는 질문-답변의 길이나 구두점의 수와 같은 텍스트 정보와 답변자의 등급, 추천수, 조회수 등의 내용 외적인 정보로 답변 품질을 측정하였으며, K-Means 알고리즘을 활용한 Yahoo! Answer의 카테고리 분류 연구[12]에서는 질문 간 답변의 개수, 사용자 간의 상호작용 패턴 등의 특성에 따라 질문들을 토론, 상식-조언, 사실 방식으로 질문을 분류하였다.

이처럼 질의응답 서비스에 대한 다양한 연구가 이루어지고 있으나, 질의응답 서비스 성능 평가나 질문 분류에 그치고 있어 사용자가 필요로 하는 답변 추천에 대한 연구는 미비한 실정이다. 본 논문은 어휘 빈도에 기반한 답변 추천 연구[13, 14]를 기반으로 질의응답 커뮤니티에서의 답변 추천 시스템을 제안한다. 본 논문에서는 질의응답 서비스의 질의를 답변의 유형에 따라 네 가지로 분류하고, 질의응답 콘텐츠의 낮은 신뢰성과 중복된 답변의 문제점을 해결하기 위한 답변 요약 시스템을 제안한다. 이를 위해 질의어에 대해 얻어진 검색 결과를 실시간으로 군집화하고, 군집 내 빈도와 질의어에 대한 근접도, 포함된 답변의 신뢰도 합, 역문서빈도로 계산된 답변 내 어휘의 적합도를 이용하여 사용자의 질문에 적합한 답변을 추천하고자 한다.

본 논문은 다음과 같이 구성된다. 2장에서는 질의응답 문서의 특징을 분석하고, 3장에서는 제안하는 답변 요약 방식을 소개한다. 4장에서는 실험과 평가 결과를 보이며, 5장에서 결론을 맺는다.

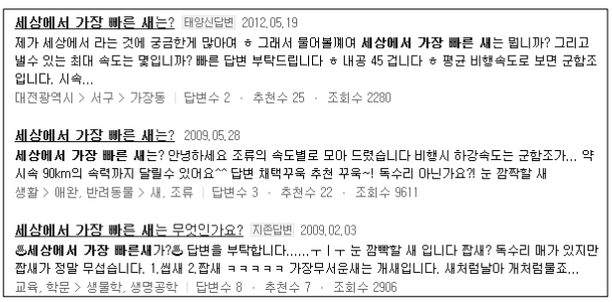
2. 질의응답 문서의 특성 분석

본 논문에서는 질의응답 문서에 대한 답변 추천을 위해 질의응답 콘텐츠의 특성을 분석한다. 데이터 분석에서는 국내에서 대다수의 사용자가 사용하여 대량의 질의응답이 축적된 네이버 지식iN의 데이터를 활용한다.

2.1 중복된 질문과 답변

지식iN에서는 여러 사용자가 별도의 제약 없이 질문과 답변을 작성할 수 있어 이미 등록된 질문과 답변이 중복 작성되는 경우가 매우 많다. 또한 잘못된 답변도 등록될 수 있어, 검색에서 얻어진 여러 질문과 그에 대한 답변에서 올바른 답을 찾기 위해 사용자의 판단이 요구된다. Fig. 1은 “세상에서 가장 빠른 새”라는 질의에 대한 지식iN의 검색결과를 보인다. 동일한 질문들이 중복 등록되어 있고, 정답 단어인 ‘군함조’가 여러 답변에서 중복 발생함을 알 수 있다. 지식iN 검색 결과에서의 정답 단어가 얼마나 중복되어 발생하는가를 확인하기 위해 무작위로 선정된 질문 300개에 대한 답변을 수동 분석하였다. 질의에 포함된 단어를 제외하였을 때, 검색된 전체 답변에 포함된 단어 중 가장 빈도가 높은 단어가 정답 단어인 경우가 34%, 빈도 상위 5위 안에 정답 단어가 포함된 경우가 56%로 분석되어, 단어 빈도수는 정답을 추출하기 위한 자질로 적합한 것으로 나타났다. 본 논문에서는 답변이 중복되어 작성되는 지식iN의 특성을 활용하여, 답변에 포함된 단어의 빈도수를 활용하여 사용자의 질문에 대한 답변을 자동으로 추천하고자 한다.

답변별 길이 특성을 파악하기 위해 지식iN 12만개 답변에 대하여 분석을 시행하였다. 답변의 평균 길이는 5천여 글자였으며, 10%가 1만 글자가 넘는 길이를 가지고 있었다. 답변 내에서 정답이 어느 위치에 발생하는지를 분석해 본 결과, 답변의 초반부 4% 이내의 위치에서 가장 높은 정답 확률을 보였다. 이러한 분석의 결과에 기반하여 본 논문에서는 답변의 앞부분 5천자만을 대상으로 분석하여 처리 속도를 높이고자 한다.


세상에서 가장 빠른 새는? 태양신답변 2012.05.19
제가 세상에서 라는 것에 궁금한게 많아여 ㅎ 그래서 물어볼께여 세상에서 가장 빠른 새는 뭡니까? 그리고 낼수 있는 최대 속도는 몇입니까? 빠른 답변 부탁드립니다 ㅎ 내공 45 겁니다 ㅎ 평균 비행속도로 보면 군함조 입니다. 시속...
대전광역시 > 서구 > 가장동 | 답변수 2 · 추천수 25 · 조회수 2280

세상에서 가장 빠른 새는? 2009.05.28
세상에서 가장 빠른 새는? 안녕하세요 조류의 속도별로 모아 드렸습니다 비행시 하강속도는 군함조가... 약 시속 90km의 속력까지 달릴수 있어요^^ 답변 채택꾸욱 추천 꾸욱~! 독수리 아닌가요?! 눈 깜짝할 새
생활 > 애완, 반려동물 > 새, 조류 | 답변수 3 · 추천수 22 · 조회수 9611

세상에서 가장 빠른 새는 무엇인가요? 지존답변 2009.02.03
ㅎ세상에서 가장 빠른새가?ㅎ 답변을 부탁합니다.....ㅜㅣㅜ 눈 깜빡할 새 입니다 잡새? 독수리 매가 있지만 잡새가 정말 무섭습니다. 1.씹새 2.잡새 ㅋㅋㅋㅋㅋ 가장무서운새는 개새입니다. 새처럼날아 개처럼물죠...
교육, 학문 > 생물학, 생명공학 | 답변수 8 · 추천수 7 · 조회수 2906

Fig. 1. Search Result Example of 지식iN

2.2 검색순위 연관성

지식iN에서 제공하는 정확도순 검색의 유효성을 검증하기 위해 2.1에서 사용한 300개의 평가데이터를 활용하여 검색 순위별로 정답 여부를 판단하여 정확도를 분석하여 Fig. 2를 얻었다. 그림에서 x축은 순위, y축은 정확도를 나타내며, 검색순위가 낮아질수록 정확도가 감소되는 것을 확인할 수 있다. 검색순위 최상위 문서 중 56%가 정답 답변을 포함하고, 이후 낮은 순위로 갈수록 정확도가 감소하였다. 본 논문에서는 이 점에 착안하여 정확도가 25% 이상인 검색 순위 30위 이내 문서들을 대상으로 답변 요약을 시도하여 처리 속도를 높이고자 한다.

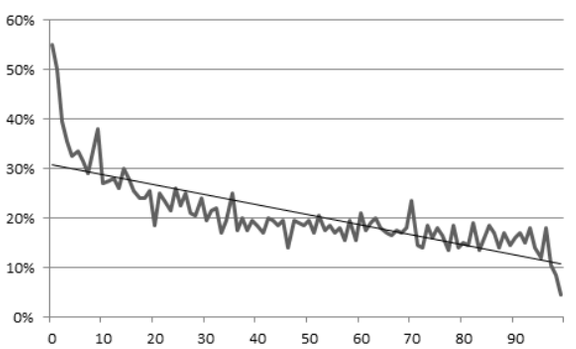


Fig. 2. Precision of search for each rank

지식iN은 질의어에 대해 질문-답변 쌍을 검색 결과로 제시한다. 검색 결과 30위 이내에는 질의어에 적합하지 않은 질문이 포함되기도 한다. 예를 들어 “대구에서 가장 높은 곳”에 대한 검색 결과로는 Fig. 3과 같이 “정기적금 이율이

가장 높은 곳"이 30위 이내의 결과로 나올 수 있다. 이처럼 30위 이내에서도 적합하지 않은 질문과 답변이 포함될 수 있고, 이 중 일부는 정답을 포함하고 있지는 않지만 질문자가 관심을 가질 수 있는 내용일 수도 있다. 본 논문에서는 Fig. 3과 같이 사용자의 질의에 대해 검색되는 다중 질문을 군집화하여, 대표 질문을 기준으로 답변을 추천하여 사용자 편의를 증진시키고자 한다.

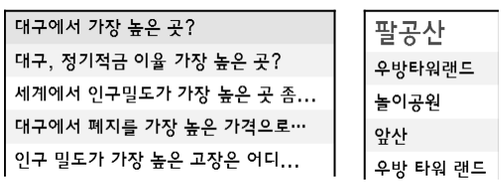


Fig. 3. Example of Q&A by question clustering

2.3 질의 유형

질의응답 문서에서의 검색을 통한 질의응답 시스템 구축을 위해서 사용자들의 질의 특징을 수작업으로 분석하였다. 요구하는 답변의 특징에 따라 분류하였을 때에 사용자들의 질의 유형을 Table 1과 같이 단어, 목록, 글, 도표로 나눌 수 있었다. 아래에서는 각각에 대해서 설명한다.

Table 1. 4 Types of Q&A of Knowledge-iN

Type	Question Example	Answer Type
Word	세상에서 가장 빠른 새	Word or phrase
List	BB크림 추천	List of words or phrases
Text	임신 초기 증상은?	Sentence or text
Graph	전화영어 괜찮을까요?	Opinion

1) 단어 유형

단어 유형은 사용자의 질문에 대한 답변이 단답형 단어나 어구로 제시할 수 있는 유형이다. 예를 들어 "세상에서 가장 빠른 새", "우리나라 최초 기차역"과 같은 질문에 대해서는 '군함조'와 '노량진역'이 그 답이 될 수 있다. 단답형 답변에서 항상 한 단어가 답변이 되지는 않는다. "세상에서 가장 높은 탑"의 답은 'CN Tower'이며, "봉준호 감독의 데뷔작"의 답은 '플란다스의 개'이다. 따라서 답변 요약 시스템은 복합어가 답변이 될 수 있는 경우를 처리할 수 있어야 한다. 단어 유형 답변은 신뢰도가 높은 다수의 답변에서 가장 자주 발생하는 단어나 어구일 가능성이 높다.

2) 목록 유형

"BB크림 추천"이나 "대구 맛집 추천"과 같은 질의에서는 '아비드 BB크림', '티엔 BB크림', '비오템 BB크림' 등 다수의 단어 혹은 복합어가 그 답으로 제시되어야 하며, 이를 목록 유형으로 분류한다. 목록 유형 답변은 신뢰도가 높은 다수의 답변에서 자주 발생하는 단어를 순위로 제시하는 형태가 사용자 편의에 만족시킬 수 있을 것으로 기대된다.

3) 글 유형

"임신 초기 증상은"이나 "이순신 장군의 업적"과 같은 질의에 대한 답변은 "임신 초기 증상은 … 입니다."와 같이 문장형으로 제시되어야 하며, 이를 글 유형으로 분류한다. 글 유형에 대한 답변은 중복되는 질문과 답변에서 가장 적합한 요약문을 제시하는 것이 바람직하나, 본 논문에서는 검색된 질문-답변들을 전체적으로 검토하여 가장 정답 확률이 높은 답변을 제시하고자 한다.

4) 도표 유형

"전화 영어 괜찮을까요"나 "갤럭시 s4와 아이폰5중에서 어떤 것이 좋아요"와 같은 질문은 전화 영어의 유용성에 대한 답이나 두 스마트폰에 대한 사용자 선호 분포를 답으로 제시하는 것이 바람직하다. 이러한 질문에 대한 답변은 "전화영어는 … 괜찮아요", "전화영어는 … 쓸만해요" 등의 문장형 답변들보다, 감성 분석(sentiment analysis)을 사용하여 의견 정보들의 통계치를 그래프 형태로 제시하는 것이 사용자 요구에 적합하다.

위 4가지 유형의 타당성을 입증하기 위해서 무작위로 추출한 지식iN의 질문 100개를 수동으로 분석하였다. 단어, 목록, 도표형이 아닌 모든 질의는 글 유형 분류할 수 있었으며, 결과에서 글 유형이 67.68%로 가장 높았으며, 목록 유형은 15.15%, 단어 유형은 9.09%였으며, 도표형이 8.08%로 가장 낮았다. 하지만, 글 유형은 36% 정도가 개인적인 질문을 포함하는 등, 등록된 질문-답변의 분포는 지식 제공 효과를 반영하지 못하는 것으로 분석되었다. 사용자의 검색 질의에서는 글 유형 외에 유형들이 좀 더 높은 분포를 가질 것으로 보인다.

본 논문에서는 사용자의 질의에 따른 답변을 각 유형에 맞추어 제시함으로써, 시스템의 효율성을 높이는 동시에 사용자에게 검색의 편의성을 제공하고자 한다.

3. 질의응답 시스템 주요기능

본 논문에서는 사용자 질의로 검색되는 여러 개의 질문-답변 문서를 질의 유형에 따라 각기 다른 방법으로 분석하여 가장 적합한 답변을 사용자에게 유용한 형태로 제시하고자 한다. 시스템은 문서 수집, 분석 및 군집화 단계, 정답 후보 단어에 대한 평가 단계와 출력부로 구성된다. 도표형 질의는 감정 사전에 기반하는 의견 사전 추출과 의견 추론부를 추가적으로 사용한다.

3.1 문서 수집, 분석 및 군집화

문서 수집부는 NAVER Open API[15]를 통해서 지식iN의 검색 결과 문서를 수집하고 형태소 분석을 수행한다. 문서 수집에서는 2.1과 2.2에서 제안한 바와 같이 검색 결과의 30페이지를 대상으로 답변의 앞부분 5천자만을 수집한다.

문서 분석 단계에서는 답변에 포함될 수 있는 복합어를 탐지한다. 아래 표는 시스템의 합성처리 탐지 패턴을 나타낸다. 질의응답 문서에서 자주 발생하는 복합어를 5가지 유형으로 분류하고, 해당 유형의 어구가 발생하면 어구에 포함되는 단일어와 복합어를 모두 후보 단어로 추출한다. 검색된 답변 문서에서 후보 단어의 빈도수를 구한 뒤, (복합어 빈도수)/max(구성 단일어 빈도수) > 0.5를 만족하는 복합어는 의미 있는 복합어로 보고 하나의 단어로 취급한다.

Combination	Example
noun + noun	손난로
alphabets + alphabets	CN TOWER
noun + alphabets	아반떼 XD
alphabets + numbers	iPHONE 4S
noun + alphabets + numbers	아이폰 4S

2.2의 Fig. 3에서 제안한 바와 같이 본 논문에서는 사용자 질의에 대한 검색 결과에 포함된 질문들을 군집화하여 사용자에게 제시한다. 이를 위하여 검색된 질문과 이에 대한 답변에 발생하는 어휘 벡터에 대한 유사도를 기준으로 질문에 대한 군집화를 수행한다. 군집화를 위해 Leader-Follower, K-means, Hierarchical 3가지 유형의 클러스터링 방법의 결과를 적용해 본 결과, Leader-Follower가 가장 좋은 결과를 보여 이를 이용하여 군집화를 구현하였다. K-means는 적절한 결과를 유추하지 못하였으며 Hierarchical은 계층적 구조를 만들기 때문에 지식iN의 검색결과 군집화에 적합하지 않은 것으로 보였다.

3.2 후보 평가

후보 평가에서는 군집별 답변에 포함된 후보 단어의 정답 적합도를 계산한다. 점수 계산에서는 단어 빈도 등의 후보 단어의 통계 정보와 함께 단어가 포함된 답변의 메타데이터를 동시에 활용한다.

단어별 점수를 구하기 위해서 군집 내 빈도수, 근접도, 답변 신뢰도, 역문서빈도(IDF)를 사용한다. 군집 내 빈도수는 군집별 답변에서 많이 발생할 단어일수록, 근접도는 질의에 포함된 단어와 답변에서 가까운 거리에 나타난 단어일수록, 답변 신뢰도는 신뢰도가 높은 답변에 많이 나타난 단어일수록, 역문서빈도는 희소성이 높은 단어일수록 정답일 가능성이 높다는 점을 반영하기 위한 척도이다.

군집 내 빈도수 CTF(Cluster Term Frequency)는 식 (1)으로 구한다. 식에서 C는 군집을, ALL은 모든 군집을, $d \in C$는 군집 C에 포함된 문서 d를, $freq(w, d)$는 문서 d에 포함된 단어 w의 빈도를 의미한다. 식에서는 대상 군집에 포함된 문서들에서의 단어 w의 빈도의 합을 문서 전체에서의 단어 빈도로 나눈다. 전체 문서에서의 빈도로 나누어 값의 범위를 [0, 1]으로 정규화하고 대상 군집에서 상대적으로 많이 나타나는 단어에 높은 점수를 부여한다.

$$CTF_C(w) = \frac{\sum_{d_j \in C} freq(w, d_j)}{\sum_{d_i \in ALL} freq(w, d_i)} \quad (1)$$

문장 내에서 질의어와 근접하여 발생하는 단어는 질의어와 구문적 연관성이 높을 것으로 예상되며, 질의어와 후보 단어 간의 거리가 가까울수록 후보 단어의 점수를 높게 부여하기 위해 근접도(proximity)를 사용한다. 지식iN의 경우 답변 길이에 대한 제한이 없어 짧은 답변에 의한 거리 값의 영향력을 조정하기 위해 단어 간 거리의 최대치를 7로 하여 근접도를 계산한다. 아래 그림에서 질의어 "세상에서 가장 큰 나라"에 대한 답변에서 후보 단어 '유럽'은 질의어 '나라'와의 사이에 명사 '러시아' 하나를 가지므로 7-1=6의 근접도를, 질의어 '큰'과는 두 개의 명사를 사이에 두고 있으므로 7-2=5의 근접도를 가진다. 최종적으로 '유럽'은 두 값의 합산한 결과인 5+6=11의 근접도를 가진다. 군집 C에 포함된 문장 s에서의 단어 w의 근접도를 $prox_s(w)$라고 했을 때, 식 (2)는 군집 C에서의 단어 w의 근접도를 계산한다. 단어 w의 근접도 합을 군집에 포함되는 모든 단어의 근접도 합으로 나누어, 해당 군집에서 질의어와 가까운 거리에서 자주 발생하는 단어에 높은 점수를 부여한다.

	제일	큰	나라는	러시아	(유럽 +	아시아)에요
근접도:				6 + 7	5 + 6	4 + 5

$$PRX_C(w) = \frac{\sum_{s \in C} prox_s(w)}{\sum_{w_i} \sum_{s \in C} prox_s(w_i)} \quad (2)$$

단어를 포함하는 답변의 개수가 많다면 해당 단어는 정답일 가능성이 높다. 예를 들어 "비료의 3요소"와 같은 질문에 대해서는 '질소'가 최상위로 18개의 답변에서, '식물'이 14개, '인산', '성분'이 13개, '필요'가 12개, '칼륨'이 11개로 나타나, 많은 답변에서 나타난 단어가 유용함을 알 수 있었다. 또한, 신뢰도가 높은 답변, 즉 높은 등급의 답변자나 전문가가 답변하거나 채택이 된 답변에 포함된 단어는 더 유용할 것으로 기대할 수 있다. 본 연구에서는 식 (3)을 통해 단어가 포함된 답변의 신뢰도의 합이 클수록 해당 단어에 높은 점수를 부여한다. 식에서 분모는 군집 C에 포함된 답변들 중에서 단어 w를 포함한 답변인 ans들의 신뢰도 $conf(ans)$의 합을, 분자는 군집내 모든 답변의 신뢰도의 합을 취하여, 군집 내에서 해당 단어의 상대 답변 신뢰도 합을 구한다.

$$CONF_C(w) = \frac{\sum_{w \in ans_i \in C} conf(ans_i)}{\sum_{ans_j \in C} conf(ans_j)} \quad (3)$$

식 (3)의 답변의 신뢰도 $conf(ans)$는 답변의 내용 외적인 정보인 메타데이터를 활용하여 구한다. 지식iN의 각 답변은 답변 채택 여부, 전문가 답변 여부, 답변자의 등급 등 다양한 메타데이터를 포함한다. 본 연구에서는 지식iN에서 1,000개의 답변과 1,000개의 비정답을 추출하여 WEKA의 RBF(Radial Basis Function) Network[16]를 통해 메타데이터에 의한 신뢰도를 학습하였다. 얻어진 신뢰도는 평가에서 24.13%의 오답률을 보였다. $CONF(w)$는 군집 C에 포함된 답변 ans 중에서 단어 w를 포함하는 답변의 신뢰도 $conf(ans)$의 합산으로 구하여, 신뢰도가 높은 답변에 많이 포함된 단어에 높은 점수를 부여하였다.

식 (4)는 군집 C에서의 단어 w의 점수를 최종적으로 계산한다. 위에서 구한 군집 내 빈도수 CTF, 근접도 PRX, 답변 신뢰도 $CONF$에 가중치를 곱하고 합산한 뒤에 역문서 빈도 IDF를 곱하여 정보성 높은 단어에 높은 점수를 부여한다. 시스템에서는 Google 검색의 결과 문서수에 반비례하는 값을 IDF로 사용하였으며, 가중치로는 실험에서 높은 정확도를 보인 α=0.75, β=1.0, γ=0.25를 적용하였다.

$$Score_C(w) = (\alpha \times CTF_C(w) + \beta \times PRX_C(w) + \gamma \times CONF_C(w)) \times IDF(w) \quad (4)$$

3.3 정답 추론 및 분석

정답 추론 및 분석 단계에서는 각 질문 유형에 맞는 답변을 구성하여 사용자에게 제시한다. 단어, 목록, 글 유형의 정답 추론에서는 역검색 평가로 후보 단어를 검증하고, 도표 유형에서는 감성 분석을 통해 정답을 구성한다.

단어와 목록, 글 유형에 대해서는 3.2에서 구한 신뢰도가 높은 후보 어휘를 순차적으로 지식iN에 질의를 던져서 검색 결과의 요약에서 최초 질의어가 모두 발현하는지의 여부를 확인한다. 만일 최초 질의어가 발현되지 않으면 해당 어휘는 후보에서 제외시키고, 얻어진 최종 후보 목록에 기반하여 유형별 추천 정답을 생성한다. 단어와 목록 유형에 대해서는 단어 점수 순으로 추천 정답을 제시한다. 글 유형은 단어 점수 최상위 후보를 포함하고 나머지 후보를 가장 많이 포함하는 답변을 추천한다.

도표 유형을 위해 의견 추출을 하기 위해서는 일부 단어로 해당 문장의 의견의 긍정과 부정, 중립을 판별해야 한다. 본 시스템에서는 감정을 파악하기 위하여 [17]의 방법에 따라 구축한 극성 사전과 극성 판별 방법을 사용한다. 극성 판별에서는 극성 사전에 포함된 단어와 가장 가까운 대상 명사에 극성이 적용되는 것으로 판별한다. 예를 들어 “아이폰은 무료 애플리케이션이 많아 아주 좋고 세계인들에게 맞게 설계되어 있어서 아주 편하다고 하는군요”에서는 긍정 극성 어휘 ‘좋고’와 ‘편하다’가 발생하여 ‘아이폰’에 대한 긍정으로 판별하고, “갤럭시 s2유저입니다만 솔직히 아이폰4s가 낫다고 생각합니다”에서는 ‘낫다’가 긍정 극성이므로 아이폰 4s는 긍정으로, 갤럭시는 중립으로 판별한다. 또한 접속사 정보를 활용하여 “아이폰도 좋지만 옵티머스도 괜찮아요”의 경우 ‘~지만’이 역접 접속사이므로 이를 기준으로 문장을 분할하여 ‘좋다’는 아이폰에, ‘괜찮다’는 옵티머스에 대한 극성 어휘로 사용한다.

4. 실험 및 평가

제안한 방법을 평가하기 위하여 실험을 수행하였다. 단어 유형의 평가에서는 무작위로 100개의 질문을 추출하고 이에 대한 정답을 수동으로 구축하여 사용하였다. 결과에서는 1순위로 추천된 답변은 53%의 정확도를, 5순위까지를 고려한 경우 65% 정확도를 보였다. Table 2는 단어 유형 평가의 예를 보인다. 오답이 나온 경우는 “한국 최초 근대식 병원”에 대해서 ‘중원’과 ‘선교사’가 나온 것처럼 답변에서 자주 발생하는 고유 명사 중에서 정답에 대한 부가 정보로 항상 사용되는 단어가 정답으로 추천된 경우가 많았다. “신라시대 신분제도”에 대해서는 시스템에서는 ‘뼈’가 정답으로 선택되었는데, 이는 정답인 ‘골품제’에 대한 설명에서 “뼈(신분)을 통해 등급을 나눈다”는 문장이 많았기 때문이다. 이러한 문제는 질문 군집화의 정확성 향상과 함께, 중요 정답은 답변 앞부분에 위치한다는 휴리스틱 등의 고도화된 문맥 정보사용으로 개선할 수 있을 것으로 본다. “세계에서 가장 높은 타워”인 경우에 최근에 지어진 버즈두바이 타워가 가장 높은 타워로 알려졌지만, 지식iN의 과거 자료에 CN Tower가 언급된 경우가 많아 시스템 결과는 CN Tower로 추천되어, 답변 작성 일시 등의 추가적인 메타데이터 사용이 필요할 것으로 지적되었다. 이외에도 고유 명사에 대한 형태소 분석 오류, 동의어 미처리 등에 의한 오류도 다수 발견되었다.

Table 2. Evaluation example for word type

Question	System result	Correct answer
세상에서 가장 큰 나라?	러시아	러시아
미국 49번째 주	알래스카	알래스카
우리나라 최초 국문소설	홍길동전	홍길동전
88올림픽 개최국	서울	서울
성조기 별 개수	50개	50개
한국 최초 근대식 병원	중원	광혜원
신라시대 신분제도	뼈	골품제
세계에서 가장 높은 타워	CN Tower	버즈 두바이
세계에서 가장 높은 산?	네팔	에베레스트
2002월드컵 개최지	2개국	한국,일본

목록 유형에 대한 100개의 질문에 대한 평가에서는 해당 질문-정답 평가 데이터와 시스템에서 추천한 정답목록을 비교하여 시스템이 정답 목록에 포함된 단어나 어구를 3개 이상 추천한다면 정확하다고 판단하였다. 평가에서 시스템은 52%의 정확률을 보였다. 오류 분석에서는 단어 유형의 오답 근거와 함께 온톨로지 사용의 필요성이 두드러졌다. 예를 들어 질의어 “아기 신발 추천”에 대한 정답 결과 ‘우미슈즈’, ‘우미’, ‘아이’, ‘유아’ 등의 정답 단어가 발생하였다. 여기서 나타나는 오답인 ‘아이’, ‘유아’와 같은 동의어와 관련된 오답들은 온톨로지(ontology)를 활용하여 해결할 수 있을 것으로 보이며, 온톨로지의 사용은 “제일 높은 산”, “큰 나라” 등의 단어 유형 답변에도 유용할 것으로 기대된다.

글 유형의 평가에서는 100개의 질문을 임의로 선택하여 실험을 수행하였다. 시스템에서 제시한 답변을 피실험자가 살펴보고 필요한 정답이 1순위 답변에 포함되어 있을 경우 올바르다고 평가하였을 때에 85%의 정확률을 보였다.

본 논문에서 사용한 방식이 얼마나 효율적인지를 알아보기 위한 실험으로 빈도나 추천 등의 개별 정보만을 이용한 경우에 대한 실험을 수행하였다. Table 3은 개별 정보를 이용한 경우를 베이스라인으로 보고 결과를 비교한다. 단어 유형의 경우 군집 내 빈도가 최고인 단어를 정답으로 채택하면 27%, 답변 신뢰도 최고인 단어가 24%, 근접도 최고인 단어가 25%의 정확률을 보였으며, 목록 유형의 경우 군집내 빈도는 25%, 답변 신뢰도는 26%, 근접도는 29%의 정확률을 보이는 것에 비해, 각 정보를 결합한 경우 53%와 52%의 정확도를 보여 성능 향상을 이루었다. 글 유형은 채택되었으면서 추천이 가장 많은 답변을 정답으로 보는 경우를 베이스라인으로 보았으며, 이 경우 67%의 정확도를 보여 본 논문의 방식이 큰 성능 향상을 보였다.

Table 3. Precision of each type of questions

	Word	List	Text
Baseline	24~27%	25~29%	67%
Our method	53%	52%	85%

도표 유형의 경우 주관적인 평가가 필요하여 질의어 '아이폰'에 대한 평가만을 수행하였다. 시스템에서 추천한 답변들을 분석한 결과에서는 질의어 '아이폰'에 대하여 긍정 74%와 부정 26%의 결과를 보였다. 각 의견정보에 대한 분석한 결과 아이폰의 답변에 존재하는 갤럭시에 관한 의견이 존재하였다. 일례로, "갤럭시는 아이폰보다 배터리를 교체할 수 있어 좋아요."라는 문장에서 갤럭시에 대한 의견이지만 고유명사 '아이폰'에 대한 구간으로 분리되어 '아이폰'의 의견으로 분류되는 문제점이 있었다. 이와 같은 오류는 구문분석을 통해 문맥정보를 이용한다면 해결할 수 있을 것으로 보인다.

5. 결론 및 향후 연구

본 연구에서는 질의응답 커뮤니티의 효용성을 높이고 신뢰성을 증진시키기 위해 기존 질의응답 커뮤니티의 문제점을 보완하여 질문의 유형을 단어, 목록, 도표, 글 4가지 유형으로 분류하여 각 유형에 적합한 답변 추천을 제시하는 방법을 제안하였다. 대상 문서와 평가 방법에 차이는 있으나, 개체명 인식부터 질문 유형, 단어 속성까지의 다양한 정보를 사용하는 TREC-2007의 factoid 유형에 대한 시스템의 최고 70.6%에서 최저 20.6%의 정확도와 비교할 때, 간단한 통계 정보만을 사용한 본 논문 단어 유형에서의 53%의 정확도는 비교적 높은 것으로 볼 수 있으며, 기존 자연언어처리 연구에서 나타난 다양한 방법들을 추가 적용시킨다면 정답률을 개선시킬 수 있을 것으로 기대된다.

향후 연구로 본 논문에서 제시한 4가지 유형을 시스템에 의해 자동 분류하는 방법과 정답률을 개선시키는 방법에 대한 연구를 진행할 예정이다.

Reference

[1] L. Hirschman, and R. Gaizauskas, "Natural Language Question Answering. The View from Here", Natural Language Engineering, 7:4:275-300 Cambridge University Press, 2001.

[2] Mark T. Maybury, "New Directions in Question Answering", AAAI/MIT Press, 2004.

[3] Rivindu Perera, "IPedagogy: Question answering system based on web information clustering", IEEE 4th International Conference on Technology for Education, pp.245-246, 2012.

[4] Hoa Trans Dang, Diane Kelly and Jimmy Lin, "Overview of the TREC 2007 Question Answering Track", TREC, 2007.

[5] Soyeon Park, Joon Ho Lee, Jiwoon Jeon, "Evaluation of the documents from the Web-based Question and Answer Service", Journal of the Korean society for library and information science, Vol.40, No.1, 2006.

[6] Hye-Rhan Chang, Eun-Tae Lee, "Performance Evaluation of the Question and Answer Services in Internet Portals", Journal of information management, Vol.37, No.2, 2006.

[7] Soojung Kim, "Answerers' Strategies to Provide Credible Information in Question Answering Community", Journal of the Korean Society for Information Management, 27(2), 21-35, 2010.

[8] Chan-Min Ahn, Bumghi Choi, Seok-Ju Chun, Ju-Hong Lee, Jung-Sik Lee, "Question Recommendation for Knowledge Search System", Journal of the Korean association of information education, Vol.14, No.3, pp.405-416, 2010.

[9] Xueying Jin and Kyung-Soon Lee, "Question Classification Based on Word Association for Question and Answer Archives", The KIPS Transactions: Part B, Vol.17, No.4, pp.327-332, 2010.

[10] Jongheum Yeon, Junho Shim and Sang-goo Lee, "Modified Naive Bayes Classifier for Categorizing Questions in Question-Answering Community", Journal of KIISE: Computing Practices, Vol.16, No.1, 2010.

[11] E. Agichtein, C. Casillo and D. Donato, "Finding high-quality content in social media", Proc. of the International Conference on Web Search and Web Data Mining, pp.183-194, 2008.

[12] L. A. Adamic, J. Zhang, E. Bakshy and M. S. Ackerman, "Knowledge sharing and yahoo answers: everyone knows something", Proc. of the 17th International Conference on World Wide Web, pp.665-674, 2008.

[13] Hochang Lee, Hyunki Tak and Hyun Ah Lee, "Answer Recommendation for Knowledge Search using Term Frequency", Korea Computer Congress(KCC), Vol.39, No.1, 2012.

[14] Hochang Lee and Hyun Ah Lee, "Answer Suggestion for Knowledge Search", The 24th Annual Conference of Human and Cognitive Language Technology, pp.201-205, 2012.

[15] NAVER Open API, "http://dev.naver.com/openapi"

[16] Machine Learning Lab in The University of Waikato, "Weka", [Online] Available : http://www.cs.waikato.ac.nz/ml

[17] Woo Chul Lee, Hyun Ah Lee and Kong Joo Lee, "Product Evaluation Summarization Through Linguistic Analysis of Product Reviews", The KIPS Transactions: Part B, Vol.17, No.1, 2010.

유 동 현

e-mail : babuluve@nate.com

2007년~현 재 금오공과대학교 컴퓨터공학부 학부생

관심분야 : 자연언어처리, 인공지능, HCI

이 현 아

e-mail : halee@kumoh.ac.kr

1996년 연세대학교 컴퓨터과학과(학사)

1998년 KAIST 전산학과(공학석사)

2004년 KAIST 전산학과(공학박사)

2004년~현 재 금오공과대학교 컴퓨터소프트웨어공학과 부교수

관심분야 : 자연언어처리, 정보검색, 지식공학