

실시간 인력기반 질의응답 시스템

임희석¹, 류기곤^{2*}

Realtime People-powered Question and Answering System

Heui-Seok Lim¹ and Ki-Gon Lyu^{2*}

요약 본 논문은 기존의 정보검색 시스템의 검색 결과에 비연관 문헌이 포함되는 단점과 질의응답 시스템이 가지는 자연어처리 기술의 한계를 극복하고, 사용자의 정보 요구에 실시간으로 정보를 제공하며 사용자의 참여를 적극 활용하여 웹2.0 환경으로의 변화에 적응할 수 있는 실시간 인력기반 질의응답 시스템을 제안한다.

Abstract This research suggests real-time people-powered Q&A system that overcoming limitation of natural language handling technology that Q&A system has and demerits that unrelated documents are included in the results of searching in existing information retrieval system and can adapt to change to Web2.0 environment by actively applying users' participation and providing real-time information to users' request of information.

Key Words : People-powered, Question and Answering

1. 서론

인터넷 기술의 발전과 함께 축적된 방대한 양의 정보를 검색하기 위해 정보검색(information retrieval) 시스템은 매우 유용한 도구가 되었다. 정보에 대한 요구(information need)가 있는 사용자는 정보검색 시스템을 통해 빠른 시간 안에 대용량의 연관된 문헌을 얻을 수 있다.

정보검색 시스템을 사용하는 사용자는 정보의 요구에 대한 답(answer)을 기대하지만 현재 우리가 사용하고 있는 정보검색 시스템은 질의와 연관된 문헌들의 집합(relevant document set)만을 제공한다. 정보검색 시스템의 입력으로 들어오는 질의는 몇 개의 단어로 조합되며 단어를 이용하여 검색한 연관문헌 집합에는 비연관 문헌이 상당수 포함되어 있어 사용자로 하여금 만족스럽지 못한 결과를 제공하기도 한다. 따라서 사용자는 본인이 원하는 정보나 지식을 얻기 위해서 정보검색 시스템이 출력한 문헌 집합의 문헌들을 일일이 살펴봐야 한다[1, 2, 3].

정보검색 시스템을 이용하는 사용자들이 양질의 검색

결과에 대한 기대가 커지면서, 사용자의 질의에 상응하는 답을 생성하는 자동 질의응답 시스템(automatic question and answering system)에 대한 관심이 높아졌다. 사용자의 질의를 파악하고 답을 생성하기 위해서는 고성능의 질의 의미 분석, 문서의 의미 분석 및 문서에서의 정보추출(information extraction) 등이 요구된다. 하지만 아직까지의 자연어처리 기술과 관련 기술의 한계로 사용자가 만족할만한 질의응답 시스템의 개발은 미흡한 실정이다 [2, 3].

현재의 많은 정보검색 시스템과 자동 질의응답 시스템은 컴퓨터로 하여금 사용자의 질의를 분석하여 유사한 문헌이나 답을 제공할 수 있도록 성능을 개선하였다. 하지만 정보검색 시스템이나 자동 질의응답 시스템은 정확성(accuracy), 효율성(efficiency), 실시간성(real-time), 그리고 상호성(interactivity)에 있어 사용자가 만족할만한 성능을 보여주지 못함에 따라 사용자의 정보 요구에 정답이나 문제 해결 방안을 가장 잘 제시할 수 있는 사람을 활용하는 방안에 대한 필요성이 자연스럽게 제기되었다 [4, 5, 6].

실제 정보의 수요가 있는 사용자에게 정답이나 가장 올바른 문제 해결 방법을 제공할 수 있는 방안은 그 분야에 대한 전문 지식이나 경험을 가지고 있는 사람과의 직접적인 의사소통을 통해서라고 말할 수 있다. 따라서 사용자의 정보 요구에 정답이나 문제 해결 방법을 제공할

본 논문은 2008년도 한신대학교 학술연구비 지원에 의하여 연구되었음

¹한신대학교 컴퓨터공학부

²고려대학교 컴퓨터교육학과

*교신저자: 류기곤(gon0121@korea.ac.kr)

수 있는 적합한 사람을 검색하고 이들을 연결시켜주어 의사소통을 할 수 있도록 지원함으로써 사용자가 원하는 정보를 획득할 수 있도록 제공하는 시스템이야말로 어느 시스템보다 더 효율적인 정보검색, 또는 질의응답 시스템이라고 할 수 있다.

이에 본 논문은 사용자의 정보 요구에 사람(휴먼 에이전트)의 힘을 이용하여 실시간 상호 커뮤니케이션에 의해 정답이나 해결방안을 제공할 수 있는 실시간 인력기반 질의응답 시스템(People-Powered Question and Answering: P2QA) 시스템을 제안한다.

2. 관련연구

인력기반 정보검색 시스템은 검색 과정에 사용자의 참여를 적극 활용하여 질의와 연관된 문헌 집합의 결과를 정제하거나, 사용자의 정보 요구에 적합한 다른 사용자를 연결시켜 주어 정보를 제공하는 시스템을 말한다.

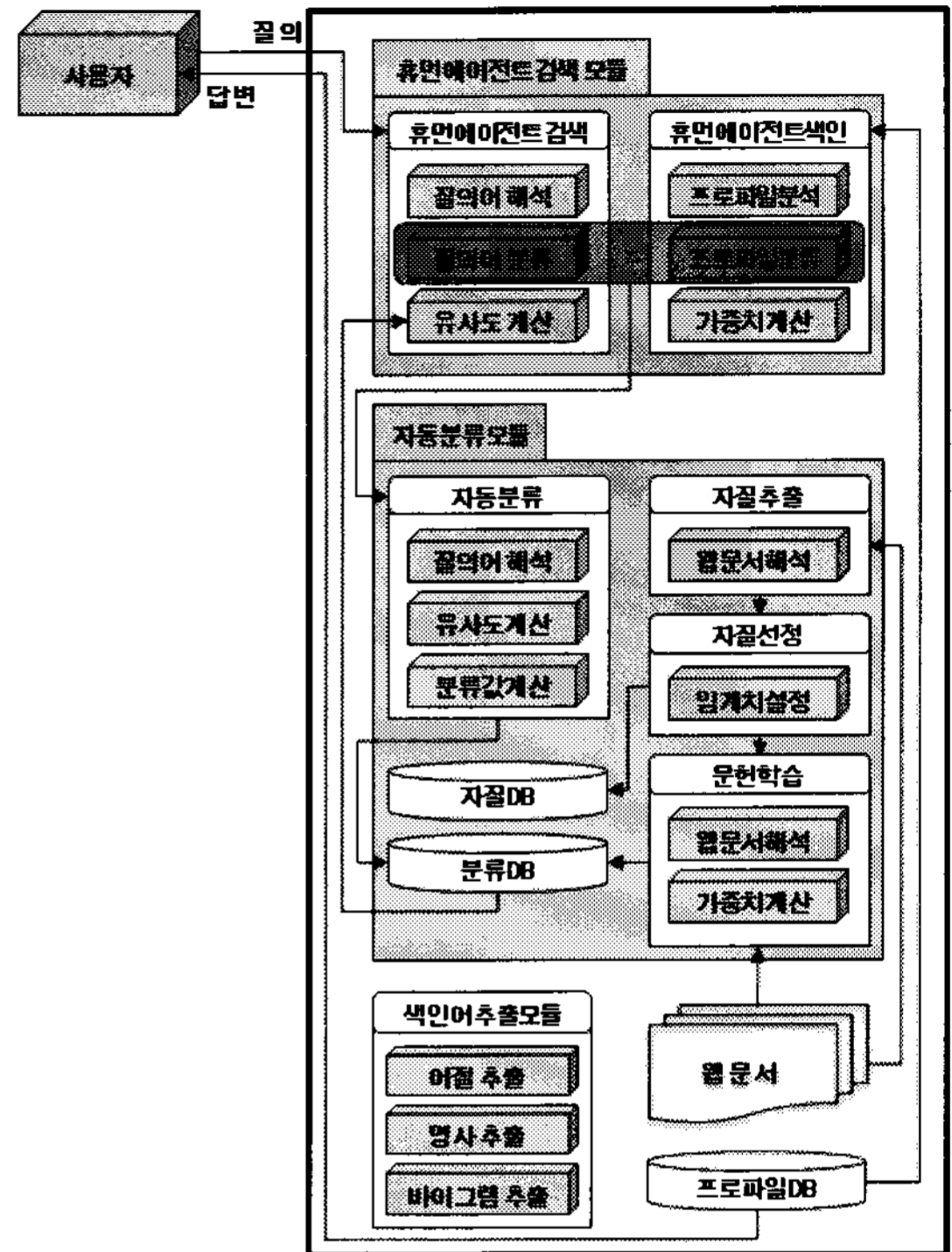
사용자의 참여 방식에 따라 크게 두 가지로 분류된다. 첫째는 사용자의 의견을 반영하여 검색 결과를 정제함으로써 일반적인 검색 시스템에서 원하지 않는 결과가 출력되는 것을 방지하고, 문헌의 순위를 결정하여 질의와 관련된 문헌을 상위로 이동시키고 관련 없는 문서를 제거하여 정제된 결과를 사용자에게 제공한다. 둘째는 본 논문이 제안하는 시스템과 유사한 문제 해결 능력이 있는 다른 사용자(guide)를 검색하고 연결시킴으로써 가이드가 대신 검색하고 결과 정보를 사용자에게 제공하는 것이다. 대표적으로 미국의 ChaCha.com, Anoox.com 등이 있다.

최근 정보검색 시스템은 웹2.0의 특징인 사용자들의 집단지능(collective intelligence)과 적극적인 참여를 활용하려는 방안이 시도되고 있다. 정보의 수요가 있는 사용자들이 양질의 정보 획득에 대한 기대가 커지고, 사용자의 의견이 적극적으로 반영되는 웹2.0 환경으로 변화하면서 인력기반 시스템은 활발한 연구가 진행 될 것으로 예상된다.

3. P2QA 시스템

본 논문이 제안하는 P2QA 시스템은 사용자로부터 정보의 요구를 자연어로 입력받아 질의에 대한 정답을 제공할 수 있는 휴먼 에이전트를 선별하여 사용자-휴먼 에이전트간 실시간 질의응답을 통한 정답을 제공하며, 그림 1은 제안하는 시스템의 전체 시스템 구성을 도식화한 것

이다.



[그림 1] P2QA 전체 시스템 구성도

그림 1과 같이 제안하는 시스템은 크게 언어처리모듈, 자동분류모듈 그리고 휴먼 에이전트검색모듈로 구성된다. 색인어추출모듈(index term extractor)은 가공하지 않은 문장에서 어절을 인식하여 추출하는 어절(word) 추출, 추출한 어절의 형태소 분석 결과 중 명사인 것만 추출하는 명사(noun) 추출, 그리고 문장을 순차적으로 2음절씩 구분하여 추출하는 바이그램(bi-gram) 추출로 구성된다. 자동분류모듈(automatic classification)은 기 수집된 문헌에서 자질 후보를 추출하고 자질을 선정하는 자질추출과 자질선정, 선정된 자질로 문헌을 학습하는 문헌학습, 그리고 입력받은 문장을 자동 분류하는 자동분류로 구성된다. 휴먼 에이전트검색모듈(human agent retrieval)은 휴먼 에이전트의 프로파일(profile)을 색인하고 자동분류모듈을 통해 전문분야를 추출하여 가중치를 계산하는 휴먼 에이전트색인과 질의를 입력받아 질의를 자동 분류하고 분류된 값과 기 색인된 프로파일간의 유사도를 계산하여 정답 제공 가능성이 높은 휴먼 에이전트를 선별하는 휴먼 에이전트검색으로 구성된다.

3.1 색인어추출 모듈

기존의 한국어 색인어 추출 방식은 크게 형태소 분석, 명사 추출, 단어 인식, 그리고 바이그램 추출 방식이다. 제안하는 시스템은 대용량 문서를 처리하기 위해 고속의 색인 과정이 필요하므로 어절 추출, 명사 추출, 그리고 바이그램 추출기를 이용한 색인을 수행하였다.

어절 추출기는 입력된 문장을 어절 단위로 인식하여 그대로 색인하는 방식으로, 키워드 검색을 통해 사용자의 질의와 정확히 일치하는 정보를 검색할 때 유용한 정보로 사용된다.

명사추출기는 추출된 어절을 한글, 한자, 영문자 및 기호 등으로 나누는 전처리기와 한글의 경우 형태소 분석을 수행하여 명사를 추출하는 명사 추출기, 추출된 색인어를 출력하는 과정으로 구성되어 있다. 전처리를 통해 분리된 한글 어절은 기 구축된 불용어 사전을 통해 불용어를 삭제한 후, 조합형 코드로 변환하여 단일 형태소 및 후절어 분석을 통하여 명사를 추출하거나 미등록어를 추정하여 명사를 추출한다.

바이그램 추출기는 입력된 문장을 2음절 단위로 색인어를 추출하는 것을 말한다. 예를 들어, “한국 산학기술학회”라는 문자열이 입력되면 “한국”, “국”, “산”, 산학“, ”학기“, ”기술“, ”술“, ”학“, ”학회“ 로 색인어를 추출하는 방식이다. 바이그램 추출 방식은 2음절 단위로 추출하기 때문에 1음절 검색 시 앞뒤 단어의 관계를 이용하여 검색할 수 있는 장점이 있다.

3.2 자동분류 모듈

자동분류 모듈은 사용자의 정보 요구를 명확하게 파악하고, 사용자의 질의에 잠재적으로 답변 가능한 휴먼에이전트를 검색 대상에 포함하도록 검색 범위를 확장하여 사용자의 정보 요구를 명확하게 파악하기 위해 사용된다.

자동분류 모듈은 웹에서 수집한 대용량의 문서에서 분류 자질을 추출하고 선정하여 학습기를 통해 웹문서를 학습하고 질의와 휴먼에이전트의 프로파일을 학습된 자질에 의해 자동으로 분류한다. 분류 정보는 인터넷 국내 대형 포털 5곳의 카테고리를 합집합 하였고 대분류 17개 중분류 226개로 구성되어 있다. 각 분류 당 약 65건의 데이터를 수집하였으며 전체 수집데이터는 15,000건이다.

3.2.1 자동 학습

본 논문은 분류하여야 할 분류의 수가 매우 많아 기계 학습 방법 중 kNN(k-nearest neighbor) 방식의 학습 방법을 사용하였다. kNN은 비모수(nonparametric) 기계학습 방법으로 각 개체를 이루고 있는 확률분포를 가정하지

않아 실제로 데이터의 확률분포가 널리 알려진 정규분포나 이항분포, 혹은 포아송 분포를 따르지 않는 많은 영역에서 매우 우수한 성능을 보여주는 학습방법이다. kNN 방식의 학습은 다른 기계 학습방법과 달리 학습 데이터에 대한 일반화과정을 수행하지 않고 학습 데이터의 자질을 추출하여 이를 저장하는 것으로 학습 과정이 끝나게 된다.

본 논문은 신문기사 10,000건과 네이버 지식iN의 질의응답 내용 5,000건, 총 15,000건의 웹문서를 수집하여 색인어추출 모듈로 자질후보를 추출하였고, 자질선정을 위해 문헌빈도(document frequency)를 사용하였다. 문헌빈도는 특정 단어 i 가 나타난 문서들의 개수를 측정한 것으로서, 드물게 나타나는 단어들은 문서의 범주 예측에 제공할 수 있는 정보량이 적다는 가정을 한다. P2QA 시스템은 휴먼에이전트의 프로파일의 전문분야 카테고리를 찾아내고, 검색에 중요한 인수(parameter)로 사용하기 위하여, 웹 문헌에서 출현한 자질의 중요도를 용어 가중치 할당 기법 중 식 1과 같이 tf-idf(term frequency - inverse document frequency) 기법을 이용해 문헌 d_j 에서의 자질 k_i 의 출현빈도수 $f_{i,j}$ 와 역문헌빈도수 idf_i 로 가중치 $w_{i,j}$ 를 구하고 학습하였다[7].

$$w_{i,j} = f_{i,j} \times idf_i = \frac{freq_{i,j}}{\max_l freq_{l,j}} \times \log \frac{N}{n_i} \quad (1)$$

3.2.2 자동 분류

kNN은 가설공간에서 분류를 위한 최적의 가설을 미리 결정해 두는 것이 아니라, 분류하는 시점에 분류할 개체와 유사한 학습예제들을 선별하고, 선별된 학습예제들만이 가설 결정에 참여하는 지연학습(lazy learning)이다.

kNN은 입력 값과 유사한 k 개의 이웃을 찾아 이 이웃들의 분류 값이나 유사도 값을 이용하여 입력 값의 분류 값을 결정한다. 본 논문은 식 2와 같이 코사인 유사도를 이용하여 각 휴먼에이전트의 프로파일 q 과 유사한 학습 데이터 d_j 를 k 개 찾아 각 학습데이터가 속한 분류를 유사도 크기로 내림차순 정렬하여 프로파일의 카테고리를 결정한다. k 는 실험결과 성능이 가장 좋은 80으로 정하였다.

$$\begin{aligned} sim(d_j, q) &= \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} \\ &= \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} \end{aligned} \quad (2)$$

3.3 휴먼에이전트검색 모듈

휴먼에이전트 검색 모듈은 사용자의 질의를 자연어로 입력받아 사용자의 의도를 분석하기 위해, 자동분류 모듈을 사용하여 질의를 분류하고, 분류된 정보를 검색 범위로 설정하여 분류에 속하는 휴먼에이전트를 검색 후보로 결정한다. 또한 자연어처리 기술을 이용하여 질의를 분석하고 질의 내 색인어와 검색 범주 내 휴먼에이전트 프로파일간의 유사성을 계산하여 검색 후보를 추출한다. 검색 후보로 추출된 휴먼에이전트의 답변 가능 여부를 확인하는 상태 정보와 지식의 신뢰도를 반영한 랭킹 알고리즘을 통해 검색된 휴먼에이전트 결과를 정제하고 정렬하여 사용자에게 실시간 정보 검색을 제공한다.

본 논문이 제안하는 P2QA 시스템의 성능을 크게 좌우하는 랭킹 알고리즘은 질의-프로파일의 자동분류 결과, 질의-프로파일 유사도 결과, 휴먼에이전트의 신뢰도 그리고 상태정보로 구성된다.

질의-프로파일 자동분류 결과가 일치한 경우 검색 대상에 포함되며, 식 3처럼 프로파일을 자동 분류하여 얻은 3순위까지의 분류 값 $Class_{k,j}$ 와 질의를 자동 분류하여 얻은 3순위까지의 분류 값 $Class_{k,q}$ 를 각 순위마다 가중치를 차등 계산하여 두 값의 곱을 자동분류 결과로 사용한다.

$$Class_{q,j} = Class_{k,q} \times Class_{k,j} \quad (3)$$

분류 값의 순위별 가중치를 반복적인 실험을 통해 결정해야 하지만, 본 논문에서는 휴리스틱 접근을 통해 시스템 초기 표 1과 같이 가중치를 차등 계산한다.

[표 1] 자동분류 값에 대한 가중치

	1순위	2순위	3순위
프로파일 자동분류 값	$Class_{1,j}$	$Class_{2,j}$	$Class_{3,j}$
질의 자동분류 값	$Class_{1,q}$	$Class_{2,q}$	$Class_{3,q}$
가중치	1.0	0.7	0.5

질의-프로파일 유사도 결과는 휴먼에이전트 검색모듈에서 계산된 질의-프로파일 간의 코사인 유사도를 사용한다.

휴먼에이전트 신뢰도는 휴먼에이전트의 답변에 대한 신뢰도를 정보의 요구가 있는 사용자가 질의응답을 완료한 후 시행하는 평가 정보를 이용하며, 0부터 5까지의 정수를 이용하여 평가한다. 식 4와 같이 휴먼에이전트 j 의 신뢰도 $Trust_{opt,j}$ 는 평가 받은 신뢰도 $Trust_j$ 를 신뢰도

의 최대값 $Trust_{max}$ 로 정규화 된 값을 사용한다.

$$Trust_{opt,j} = \frac{Trust_j}{Trust_{max}} \quad (4)$$

휴먼에이전트 상태정보는 식 5와 같이 휴먼에이전트가 답변가능 상태인 경우만 검색이 수행되도록 한다. 본 논문이 제안하는 시스템은 실시간성이 중요한 특징으로써, 사용자의 정보 요구에 실시간으로 답변 가능한 휴먼에이전트를 검색하여 사용자의 만족도를 높이고 시스템의 성능을 향상시킨다.

$$Status_j = \begin{cases} 1 & \text{if answerable} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

랭킹 알고리즘은 앞의 네 가지 인자 중 상태 정보를 제외한 세 가지 인자에 각 α, β, γ 의 가중치 조절 변수를 조합하여 식 6처럼 구성된다. $\alpha + \beta + \gamma = 1$ 이며, 휴리스틱 접근을 통해 시스템 초기에는 $\alpha=0.4, \beta=0.4, \gamma=0.2$ 를 사용하였다.

$$Rank_j = Status_j(\alpha \cdot Class_{q,j} + \beta \cdot Sim(p_j, q) + \gamma \cdot Trust_{opt,j}) \quad (6)$$

4. 실험 및 평가

객관적인 평가를 위해 별도의 실험 컬렉션을 구축하여 시스템을 평가해야 하지만 실험 컬렉션 구축에 많은 시간과 비용이 소요되는 문제가 있기 때문에, 주관적인 방법으로 시스템의 실시간성과 시스템 만족도를 정성적으로 평가하였다.

4.1 실험 평가 방법

실험은 실시간성과 시스템 만족도로 나뉘며, 실시간성의 경우 P2QA 시스템과 유사한 사용자의 질의에 다른 사용자가 답변을 기술하는 형태의 사이트를 선정하여 질의 후 최초 답변 대기 시간을 측정하였고, 시스템 만족도는 휴먼에이전트에게 수신된 질의가 적합한지를 5점 척도로 평가하여 MOS(mean opinion score) 값으로 주관적인 평가를 수행하였다.

4.2 실험 데이터 구성

실시간성을 평가하기 위해 표 2처럼 전문성을 요구하

는 질의 30개와 표 3처럼 실생활에 유용한 지식에 대한 질의 30개로 구분하였다.

[표 2] 전문성을 요구하는 질의

대분류	중분류	질의 예	수
컴퓨터	프로그램	Java에서 this와 super의 차이?	10
경제	대출	아파트 담보대출 받고 등기하는 법?	10
건강	안과	라식과 라섹 중 어떤 수술이 적합?	10

[표 3] 실생활에 유용한 지식에 대한 질의

대분류	중분류	질의 예	수
게임	게임기	닌텐도DS 사용법?	10
쇼핑	핸드폰	핸드폰이 물에 빠졌을 때 대처법?	10
생활	지혜	과일 농약성분 깨끗이 씻는 방법?	10

시스템의 만족도는 표 2, 표 3의 실험 데이터를 사용하고, 대학생 50명을 휴먼에이전트로 선정하여 질의 수신 적합성 여부를 피드백 받아 성능 평가에 사용하였다. 만족도는 표 4처럼 1점부터 5점 사이의 점수를 휴먼에이전트의 개인 의견에 따라 부여하며, 점수의 평균인 MOS 값으로 시스템 만족도를 평가하였다.

[표 4] 만족도에 따른 MOS

만족도	MOS
1	매우 불만족
2	불만족
3	보통
4	만족
5	매우 만족

4.3 실험 및 결과분석

최초답변대기시간은 시스템의 실시간성을 직접적으로 보여줄 수 있는 중요한 요소이다. 국내에서 가장 널리 알려진 네이버 지식iN을 비교대상으로 실험한 결과 최초답변대기시간은 네이버 지식iN이 평균 83.9분, P2QA 시스템은 평균 1.7분이 소요되었고, 제안하는 시스템이 실시간으로 입력기반 질의응답을 지원한다는 것을 보여준다.

시스템 만족도는 시스템의 주관적인 성능을 보여주는 요소이다. 휴먼에이전트 50명에게 “취미”, “특기”, “전공”, “사는곳”, “여행가본곳”을 프로파일에 기술하게 한 뒤 시스템이 학습하게 하였고, 표 2, 표 3의 실험 데이터로 질의 수신여부에 따른 시스템 만족도를 평가하였다.

실험 결과 시스템 만족도는 표 4와 같이 평균 3.8점으로 전반적으로 “만족”에 근접한 결과를 보여주었다.

5. 결론

본 논문은 사용자로부터 정보의 요구를 자연어로 입력 받아 질의에 대한 정답을 제공할 수 있는 휴먼에이전트를 선별하여 사용자-휴먼에이전트간의 실시간 질의응답을 통해 질의를 해결하는 실시간 입력기반 질의응답 시스템을 제안하였다.

제안된 시스템은 크게 언어처리모듈, 자동분류모듈, 그리고 휴먼에이전트검색 모듈로 구성되어있다. 자동분류모듈은 기수집한 15,000건의 웹문서를 이용하여 학습되었고 휴먼에이전트검색 모듈은 50명의 휴먼에이전트 프로파일을 색인하였다. P2QA는 시스템에 의한 질의응답이 아닌 입력기반 질의응답이 목적이므로 실시간성과 시스템 만족도를 정성적으로 평가하였다. 국내에서 가장 널리 알려진 네이버의 지식iN과 제안하는 시스템의 실시간성에 대한 비교실험 결과 최초답변대기시간이 1.53분으로 실시간성을 보였고, 휴먼에이전트 프로파일 기반의 질의 수신 만족도는 평균 3.8점으로 전반적으로 “만족”에 가까운 결과를 보였으나 MOS 값이 객관적인 성능이라고 말할 수는 없다.

제안된 시스템은 기존의 정보검색 시스템을 대신하여 사용하기에는 당장은 부족하지만 웹2.0으로 변화하면서 사용자의 참여가 적극적으로 변화하는 시대에 사용자들의 집단지능을 활용하여 지식이나 지혜를 찾는 데는 충분히 유용하게 활용될 수 있다고 판단된다. 향후에는 휴먼에이전트의 활동이력에 따라서 지능적으로 프로파일을 갱신하여 시스템 만족도를 향상시키고, 객관적인 평가 콜렉션 구축에 대한 연구를 진행할 계획이다.

참고문헌

- [1] 맹성현(2004), “정보검색 기술의 현황과 발전전망” 정보과학회지 제22권 제4호. pp.16-20.
- [2] 박영진, 송길영, 김경서, 송성환, “웹2.0과 정보검색” 정보통신연구진흥원 주간기술동향 1251호.
- [3] 임희석(2002), “경험적 정보를 이용한 kNN 기반 한국어 문서 분류기의 개선” 컴퓨터교육학회 논문지. pp.37-44.
- [4] 류기곤(2008), “집단지능을 이용한 입력기반 질의응답 시스템 설계 및 구현” 한신대학교 대학원 석사학위

논문. pp.1-59.

- [5] Salton G. and McGill, M. J.(1983), "Introduction to modern information retrieval" McGrawHill Book Co., NY.
- [6] Y. Yang(1994), "Expert network : Effective and efficient learning from human decisions in text categorization and retrieval" In Proceedings of the 17th International Conference on Research and Development in Information Retrieval (SIGIR '94), pp.13-24.
- [7] Ricardo Baeza-Yates and Berthier Ribeiro-Neto (1999), "Modern Information Retrieval" Addison Wesley.
-

임 희 석(Heui-Seok Lim)

[종신회원]



- 1992년 2월 : 고려대학교 컴퓨터학과 (이학사)
- 1994년 2월 : 고려대학교 컴퓨터학과 (이학석사)
- 1997년 9월 : 고려대학교 컴퓨터학과 (이학박사)
- 1997년 9월 ~ 1999년 2월 : 삼성종합기술원 HCI Lab.

- 1999년 3월 ~ 현재 : 한신대학교 컴퓨터공학부 부교수

<관심분야>

자연어처리, 인공지능, 인지신경계산학, 정보검색

류 기 곤(Ki-Gon Lyu)

[정회원]



- 2006년 2월 : 천안대학교 소프트웨어학과 (공학사)
- 2008년 2월 : 한신대학교 소프트웨어학과 (공학 석사)
- 2008년 3월 - 현재 : 고려대학교 컴퓨터교육과 (박사 과정)

<관심분야>

자연어처리, 정보검색