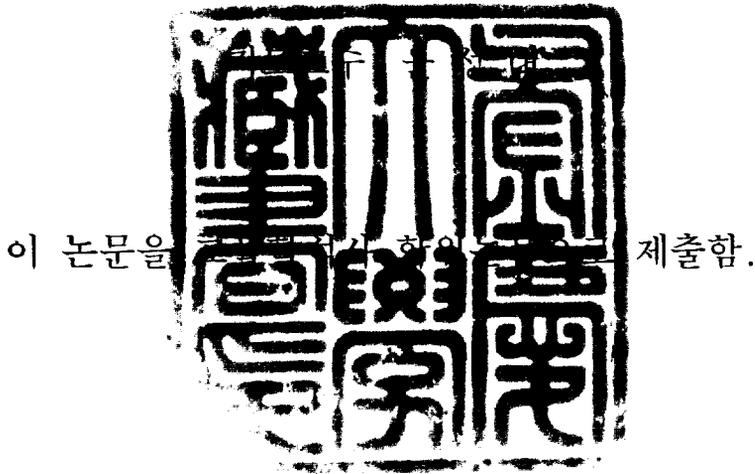


7M  
002  
156  
13

교육학석사학위논문

# 자연어처리를 이용한 교육행정의 질의응답시스템



2006년 2월

부경대학교 교육대학원

전산교육전공

이 미 나

# 이미나의 교육학석사 학위논문을 인준함

2006년 2월 24일

주 심 이학박사      박    홍    복



위 원 공학박사      김    창    수



위 원 이학박사      윤    성    대



## [ 차례 ]

차례 .....	i
표 차례 .....	ii
그림 차례 .....	iii
Abstract .....	iv
I. 서 론 .....	1
II. 관련연구 .....	4
2.1 자연어처리 시스템 .....	4
2.2 자연어처리를 이용한 질의응답시스템 .....	7
2.3 행정처리 시스템 .....	11
III. 시스템 구성 .....	14
3.1 사용자 인터페이스 .....	15
3.2 업무관리자 인터페이스 .....	16
3.3 내부처리엔진 .....	18
IV. 실험 및 평가 .....	28
4.1 실험 환경 .....	28
4.2 비교 실험 .....	30
4.3 성능 평가 .....	32
V. 결 론 .....	36
참고문헌 .....	38

## [ 표 차례 ]

<표 1> 자연어처리 시스템의 종류 .....	6
<표 2> 행정기관 사이트 정보제공의 문제점 .....	11
<표 3> 테이블 current_duty의 구성 .....	17
<표 4> object_base 테이블 .....	23
<표 5> object_match 테이블 .....	23
<표 6> topic_base 테이블 .....	23
<표 7> topic_match 테이블 .....	23
<표 8> 히스토리 테이블 .....	25
<표 9> 생성된 뷰 테이블 .....	26
<표 10> 실험 환경 .....	28
<표 11> 사용자 질의세트 .....	29
<표 12> 비교질의 세트 .....	30
<표 13> 시스템 적중률의 비교 .....	33

## [ 그림 차례 ]

(그림 1) 질의응답의 과정 .....	8
(그림 2) LAASO 시스템의 구성 .....	9
(그림 3) AnyQuestion 1.0 시스템 구성도 .....	10
(그림 4) 키워드 검색 (부경대학교) .....	12
(그림 5) 게시판식의 질의응답 (부산광역시 교육청) .....	13
(그림 6) 시스템 구성도 .....	14
(그림 7) 사용자 인터페이스 .....	15
(그림 8) 관리자 인터페이스 .....	16
(그림 9) 내부처리엔진 흐름도 .....	18
(그림 10) 형태소분석기의 구조 .....	19
(그림 11) 형태소 분석기 .....	21
(그림 12) 복합어 구성 알고리즘 .....	22
(그림 13) 정답제공 화면 .....	27
(그림 14) 단일키워드의 응답 분포도 .....	34
(그림 15) 이중키워드의 응답 분포도 .....	34
(그림 16) 자연어질의의 응답분포도 .....	35

# Question and Answering System of Educational Administration Using Natural Language Processing

Mi Na Lee

*Graduate School of Education  
Pukyong National University*

## Abstract

Nowdays, almost companies and public institutions can offer necessary information to their clients through their website with the development of information and communication. As this result, almost commercial web sites are using variable approaching methods what make users to search information. However, information offering system used in the field of administration has given information, until now like that users get the information personally through Keyword Searching System.

We suggested Question and Answering System using Natural Language Processing in order to improve this problem in this paper. The system offer suitable answers, after comparing the general database and the intention of personal questions as forming correct answer data View.

Moreover, History that adaptable FAQ(frequently asked question) management function, can answers promptly about the

questions already answered.

We applied the system to Education Administration web page to verify its usability and possibility. We have shown that it provided more accurate answer than Keyword Searching System by the result.

# I. 서 론

정보통신 기술과 인터넷이 발달하면서 사용자는 인터넷을 통해서 더 많은 정보를 찾고, 가상공간에서 서로의 의견을 교환하여 기존의 정보를 확대재생산하고 있다. 이러한 인터넷의 발달은 정보의 폭발적인 증가를 가져왔고, 사용자는 거대한 정보의 홍수 속에서 자신에게 필요한 정보만을 찾기 위해서 많은 시간과 노력을 투자해야 하는 문제를 가지게 되었다. 이를 해결하기 위해서, 전자상거래 시스템에서는 사용자와 사업자 모두의 요구를 충족시킬 수 있는 효과적인 정보 검색 도구의 필요성이 제기[1,2] 됨에 따라 여러 가지 검색 도구들이 개발되어 졌고, 초기의 복잡한 메뉴 기반 네비게이션에서 지루한 키워드 기반검색을 거쳐 사용자와 친숙할 수 있는 자연어 처리를 이용한 검색도구까지 발달하게 되었다[3].

자연어 처리를 사용한 검색도구는 사용자의 편의가 우선적으로 적용되는 전자상거래 사이트와 학술검색용 사이트에서 주로 사용되고 있다. 이러한 사이트들은 특정 시스템 관리자에 의해서 내부 시스템이 관리되고, 시스템은 사용자의 자연어 질문의 의도를 파악하기 위하여 정밀한 자연어 처리 기술과 정보추출 기술을 필요로 한다.

관리자에 의해 구축된 정보 검색의 방법 외에 현재는 사용자들끼리 정보를 교환하여 원하는 정보를 찾을 수 있는 지식검색이 국내 검색 시장의 대표적인 서비스로 자리매김하고 있다. 지식검색은 별도의 관리자 없이 사용자의 질문에 다른 사용자가 답을 달고 많은 사용자들이 이 지식의 유용성에 대해 평가함으로써 네티즌들이 지식을 공유하는 수단으로 크게 각광받고 있다. 지식검색의 장점으로서는 시사에 민감한 네티즌의 특성상 최근에

이슈화 되고 있는 정보를 쉽고 빠르게 찾을 수 있다는 점과 네티즌들의 경험에 의한 실용적인 질문과 답변이 많다는 것이다. 반면에 제공되는 정보의 상당수는 신뢰성 있는 근거를 기반으로 하지 않아 이것이 총체적인 정보 신뢰성의 약화를 가져온다는 단점이 있다[6].

인터넷 상에서는 전반적으로 자연어를 사용한 정보검색이 확대되고 있지만 아직 일반 행정기관 및 행정업무를 담당하는 곳에서는 이러한 시스템이 미비한 실정이다. 사용자가 자신의 원하는 행정정보를 알기 위해서는 행정기관의 웹 사이트를 방문하여 복잡한 경로를 통해 정보를 얻거나 혹은 행정기관에 직접 문의하는 경우가 대부분이다. 이러한 방법은 사소한 정보를 얻기 위해 사용자가 많은 시간과 노력을 투자해야하는 한계를 지닌다. 따라서 본 논문에서는 우리 주변에서 가장 많이 접할 수 있는 교육기관의 행정업무에 관한 정보제공 형태를 살펴보고 교육행정업무에서 보다 쉽고 정확하게 사용자질의에 응답할 수 있는 효율적인 정보제공자로서의 질의응답 시스템을 제안하고자 한다.

자연어처리를 사용한 질의응답시스템은 자연어로 질문을 하고 대답을 얻을 수 있기 때문에 정확하고 친밀하게 사용자의 질의에 응답할 수 있어 효과적인 정보 제공의 필요성이 증대되고 있는 웹사이트의 가상 대리자로서의 역할에 적합하다. 본 논문에서는 사용자 질의 의도에 따른 답변 매칭 방법과 동적 FAQ 관리기능인 히스토리를 제안하여 사용자 의도에 알맞은 정답을 신속하고 정확하게 제공하고, 시스템 구현을 통하여 제안한 시스템의 유용성을 검증하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구로서 자연어처리와 자연어처리를 적용한 질의응답시스템에 대하여 간략하게 소개를 하고 행정처리 시스템의 정보제공의 문제점과 사용자의 접근성에 있어서의 한계를 지적한 다음 개선방안에 대하여 알아본다. 3장에서는 본 논문에서 제안하는 시스템의 구성

과 핵심내용에 대하여 서술한다. 4장에서는 제안한 시스템에 대한 실험 및 평가를 하고, 끝으로 5장에서는 결론 및 향후 연구 과제를 제시한다.

## II. 관련 연구

이 장에서는 본 논문에서 기반 환경으로 하고 있는 자연어처리 시스템과 자연어 처리를 적용한 질의응답시스템에 대하여 살펴본다. 또한 현재의 교육행정처리 시스템의 실태와 문제점을 지적하고 개선사항에 대해서 알아본다.

### 2.1 자연어처리 시스템

자연어의 이론적인 연구만이 아니라, 이것을 응용하여 실제로 사용하는 자연어처리 시스템을 만들려는 시도도 활발하다. 현재까지 개발된 자연어 시스템은 아래와 같이 분류될 수 있다[7].

- 1) 질의응답 시스템(컴퓨터와 인간사이의 질의응답을 반복하는 시스템)
- 2) 문제해결 시스템(산수 응용문제 등의 해답을 구하는 시스템)
- 3) 데이터베이스 조작 시스템(데이터베이스를 검색, 갱신하는 시스템)
- 4) 문장해석 시스템(문장의 의미를 이해하는 시스템)
- 5) 기계번역 시스템(언어 사이의 번역을 행하는 시스템)

자연어처리 시스템의 초기 성공의 예로서는 화면상의 나무쌓기를 자연어인 영어로 조작하는 Winograd 의 SHRDLU가 유명하다. 자연어 처리 시스템은 그 후 정신분석을 행하는 ELIZA, 여행계획의 질의응답 시스템인

GUS, 영문해석 시스템인 MARGIE, 대화해석 시스템인 TOPLE 등 여러 가지 자연어처리 시스템이 개발되었다.

자연어로 데이터베이스를 조작하는 시스템의 개발도 활발히 진행되고 있다. SRI가 개발한 대규모 분산데이터베이스 조회 시스템인 LADDER, 관계모델(relational model)의 창시자인 코드(Codd)가 개발한 RENDEZVOUS, 지질학에 관한 질의응답 시스템 LUNAR, IBM이 개발한 질의응답 시스템 REQUEST 그리고 항공편에 관한 대규모 관계형(relational) 데이터베이스를 취급하는 질의 응답 시스템 PLANE 등이 대표적이다.

1970년 이후에는 자연어처리 시스템이 상품화되기 시작했다. INTELLECT는 자연어를 취급할 수 있는 최초의 상품으로, 아티피셜 인텔리전스사(Artificial Intelligence Inc.)가 발매한 상품의 데이터베이스용 자연어 인터페이스이다. 최초에 Q&A라는 종합 소프트웨어 패키지는 워드프로세서, 데이터베이스에 대한 명령어를 자연어인 영어로 할 수 있도록 구성하여 성공한 시스템이다.

한국에서는 한국전자통신연구소(ETRI; Electronic & Telecommunication Research Institute)가 주도하는 정부 국책연구개발과제인, 지능형 컴퓨터에서 자연어 인터페이스에 관한 연구를 한국과학기술원 인공지능센터와 공동으로 한 바 있다.

실용화된 시스템으로서는 한국과학기술원 전산학과 최기선 교수팀에서 개발한 한국어 텍스트의 자동색인검출기 KAIS(Korean Automatic Indexing System)와 자동철자검사 및 교정시스템(Hspe 11)이 자연어처리 시스템의 실용화로서는 효시이다.

KAIS는 1991년에는, Hspe 11는 1989년에 시제품이 완성되었다. 이 외에도 금성소프트웨어의 철자검사 시스템이 상용화되는 등 1990년 하반기부터 이에 대한 연구가 번성하기 시작하였다[8,9].

아래 표 1은 자연어처리 시스템의 종류에 대해 요약한 표이다.

<표 1> 자연어처리 시스템의 종류

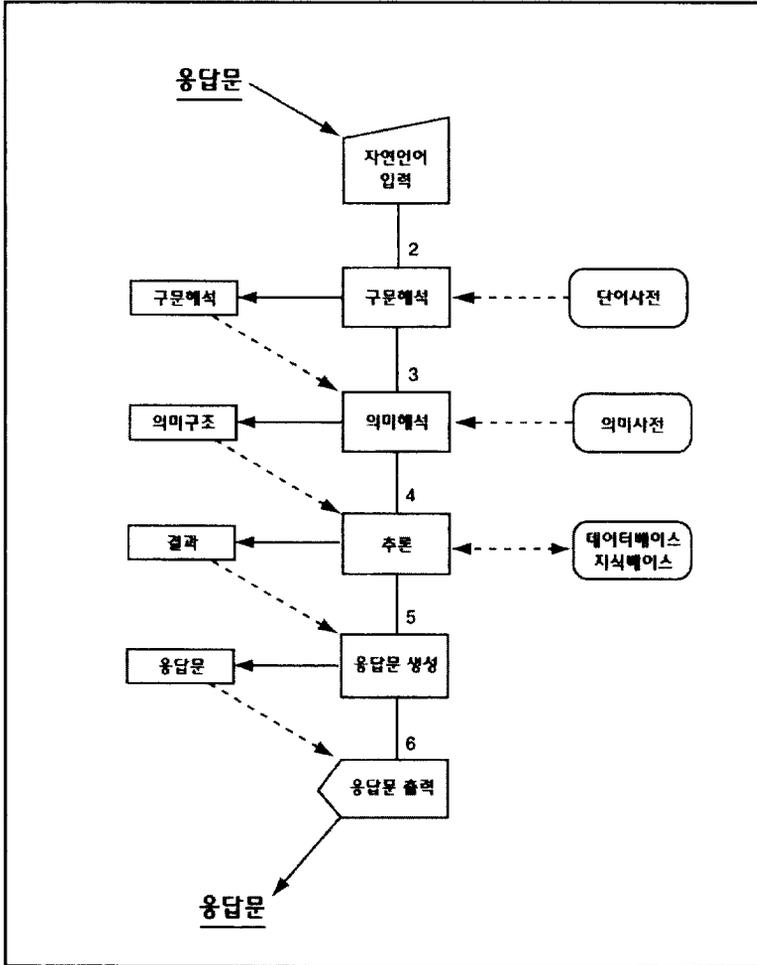
	시스템명	기능
질의응답 시스템	SHRDLU	나무쌓기의 QA시스템
	GUS	여행계획의 QA시스템
문제해결 시스템	STUDENT	산수 문제해결 시스템
	Isaac	지리 문제해결 시스템
문장해석 시스템	MARGIE	영문해석 시스템
	TOPL	대화이해 시스템
데이터베이스 검색 시스템	RENDEZOUS	관계 데이터베이스 조작 시스템
	LUNAR	지질학에 관한 Q&A시스템
	REQUEST	데이터베이스 검색 시스템
	PLANS	항공편에 관한 Q&A시스템
상용시스템	INTELLECT	main frame의 데이터베이스용 자연어 인터페이스
	SAVVY	PC 자연어 인터페이스
	STRAIGHT TALK	WP와의 자연어 인터페이스

## 2.2 자연어처리를 이용한 질의응답시스템

질의응답 시스템은 사용자의 질의와 관련된 문서를 검색하는 정보검색 (Information Retrieval) 시스템과는 달리 사용자의 질의에 대한 답변이 될 수 있는 정답을 문서 집합 내에서 탐색하여 사용자에게 제시해 주는 시스템이다[4]. 인터넷을 통해서 정보를 얻는 일반 사용자에게 보다 편리한 사용자 인터페이스 환경을 제공하기 위하여 질의응답시스템에서는 자연어처리를 이용한 연구가 활발히 진행되고 있다.

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.

그림 1에서 이 단계를 세분화 하여 보여 주고 있다. 질의응답 과정에서 정답추출단계는 구문분석 또는 의미분석 등과 같은 고급의 언어처리 기술을 사용하여 사용자의 질의에 적합한 정답을 추출하는 단계이다. 따라서 정보검색 시스템에서 사용되는 색인 가능한 기본적인 정보 이외에도 색인할 수 없는 다양한 구문 정보 혹은 의미 정보들을 사용하여 정답임을 판별해 내는 분석작업이 수행되어야 한다[10,11]. 색인 되어있지 않은 정보들을 이용하는 특성으로 인해 정답추출 과정을 모든 문서에 대해 일괄적으로 적용하는 방법은 사용하기 어렵다. 이로 인하여 질의응답 시스템에서는 본격적인 정답추출 작업을 수행하는 전 단계로 검색시스템을 사용해서 정답을 포함하고 있을 가능성이 있는 문서들을 선별해내는 후보검색 단계를 수행한다. 즉 후보검색 단계는 정답추출 단계가 적용될 후보를 찾아내는 전처리 단계이다. 따라서 후보검색 시스템의 재현율, 정확도 그리고 검색 결과량은 질의응답 시스템 전체의 재현율, 정확도 그리고 성능에 결정적인 영향을 미친다[12,13,14].

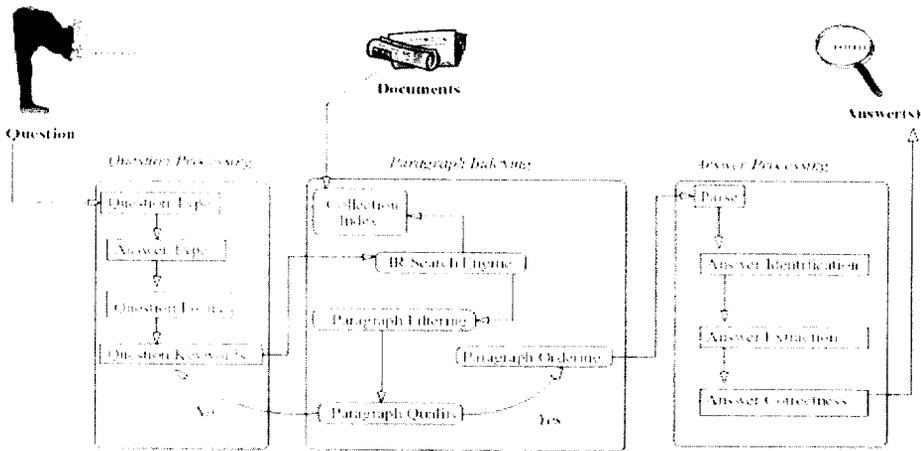


(그림 1) 질의응답의 과정

이 결과를 바탕으로 사용자의 질의 의도를 파악하는 의미 분석의 과정을 거쳐 최종적으로 응답문을 생성하여 사용자에게 응답을 제공한다. 주요 질의응답시스템으로는 국외에서 개발된 LASSO[5]와 국내에서 개발된 MAYA[15], AnyQuestion[16] 등이 있다.

## (1) LASSO

LAASO[5]는 TREC-8 QA Track에서 우수한 성능을 보여준 시스템으로 정답의 추출을 위해 질문 분석을 통해 정답유형을 정의한 뒤 정답을 추출하는 방식을 사용하는 특징이 있다. 질의 처리과정에서 1)질문의 유형정의, 2) 예상되는 정답의 유형, 3)정답의 포커스 선택, 4) 검색엔진을 위해 질문을 질의로 변화하는 역할을 수행한다. 이러한 과정을 거쳐서 얻어진 정보를 통해서 정답후보를 결정하고 정답후보들 중에서 다양한 가중치 함수를 두고 이를 결합하여 정답을 추출한다.



Architecture of the LASSO Q/A System

(그림 2) LAASO 시스템의 구성

## (2) MAYA

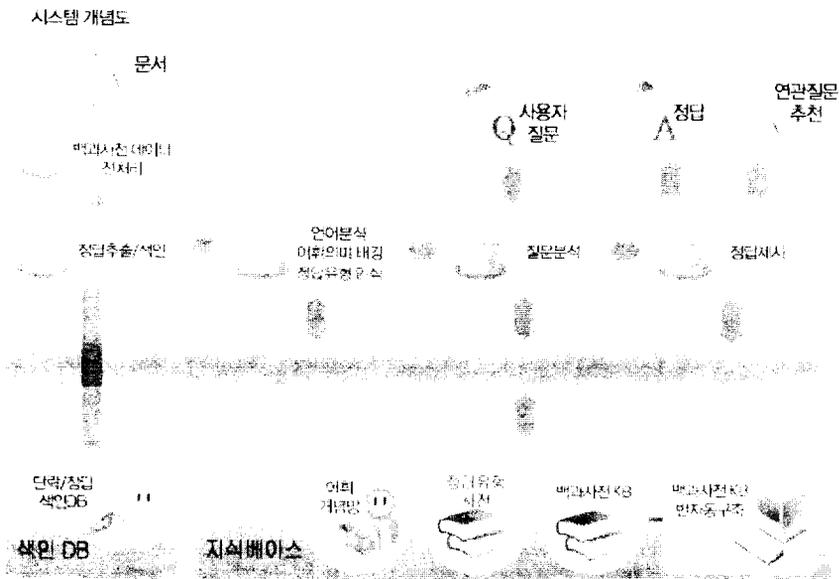
MAYA[15]라는 질의응답 시스템에서는 개체명 사전과 LSP(Lexico-Syntactic Patten)를 이용하여 개체명을 인식하고, 이를 질의응답 시스템이 정답 가능한 후보로 미리 색인 하였다. 사용자 질의 유형을 105가지의 의미범주로 구분하고, 이에 따라 정답 유형을 분류하였다.

분류된 정답 유형은 Lexico-Syntactic Parser를 이용하여 사용자의 질의 유형을 분석하여 색인된 정답 DB에서 정답 후보를 순위화하고 이를 정답으로 제시 하였다.

### (3) AnyQuestion 1.0

국내에서 개발한 백과사전 인물 질의응답 시스템인 AnyQuestion 1.0[16]은 인물분야 약 25,000 표제어에 관한 사용자 질문에 대해 질문의도를 파악하여 단답형의 정답을 제시하는 형태이다. 정답이 본문의 문장을 통해서 추출되었을 경우에는 해당문장을 함께 제시하여 사용자가 정답의 정확성 유무를 확인할 수 있게 하였다.

그림 3은 색인 DB와 지식 베이스를 정보의 원천으로 사용자에게 정답을 제공하고 있는 AnyQuestion 1.0의 시스템 구성도를 보여주고 있다.



(그림 3) AnyQuestion 1.0 시스템 구성도

## 2.3 행정처리 시스템

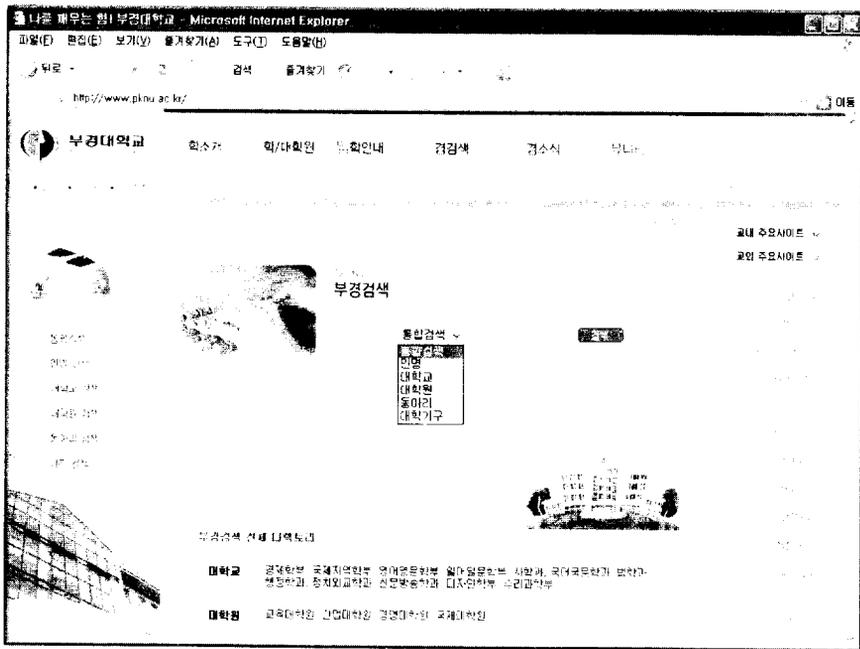
행정처리 시스템은 크게 행정정보를 제공하는 업무담당자와 정보를 제공 받는 사용자의 입장으로 나눌 수 있다. 현재의 행정처리 시스템의 연구는 이 두 가지 입장 중에서 주로 업무담당자에게 초점이 맞추어져 있다. 현재 활성화되어 있는 행정기관 사이트 중 중앙행정기관 20곳, 지방행정기관 15곳, 교육행정기관 20곳에서 행정에 관한 가장 일반적인 정보를 검색해 본 결과 표 2와 같은 문제점을 가지고 있었다.

다음의 표 2에서 보는 것과 같이 대부분 교육행정기관 사이트들은 행정정보의 검색을 지원해 주고 있지만 단일 키워드 검색과 게시판 형태의 일차원적 관리가 이루어지는 사이트가 대부분이었다. 또한, 웹사이트 상에서 찾을 수 없는 정보를 얻기 위해서는 직접 전화를 하거나 게시판에 글을 남긴 후 답변을 기다리는 일반적인 질의 서비스만을 제공하여, 사용자의 편의를 보장하고 있지 않는 사이트가 많았다.

<표 2> 행정기관 사이트 정보제공의 문제점

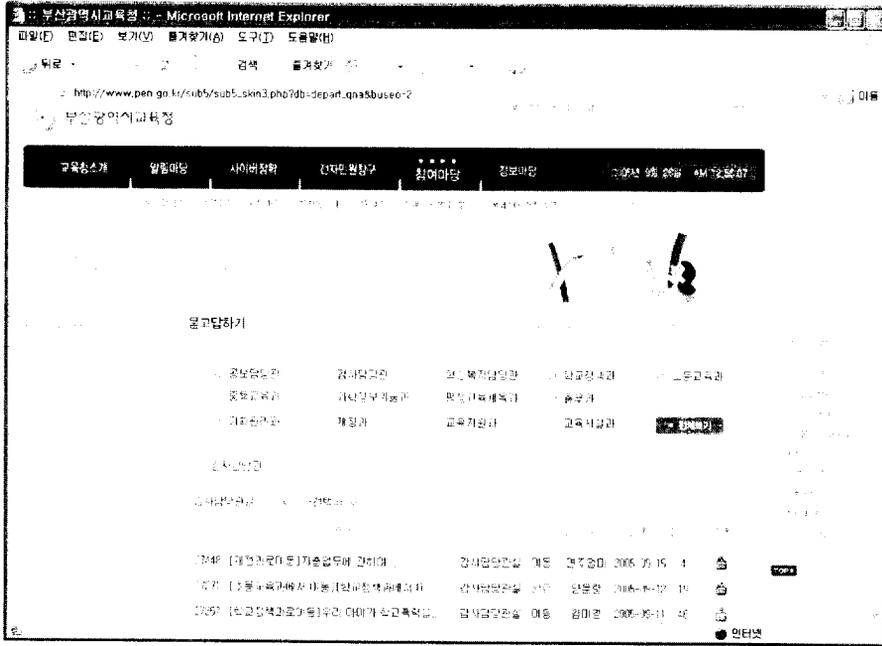
번호	문제점	설명
1	검색의 제공 형태(키워드)	단순키워드 검색으로 자연어 질의가 지원되지 않음.
2	정보제공 형태(게시판식)	게시판의 묻고 답하는 방식으로 답변을 바로 제공 받을 수 없음.
3	정보제공 형태(정보리스트)	관련 답변의 리스트를 제공해줌으로서 2차적으로 다시 검색해야 함.
4	업무담당자 확인 불가	추가적 정보가 필요할 때 업무 담당자를 쉽게 찾을 수 없음.

사용자의 질의는 정확한 정보를 요구하는 것이지만 행정기관 사이트들은 모두 사용자 질의에 대하여 검색된 정보의 리스트를 제공해 주고 있기 때문에 사용자는 이 리스트에서 다시 자신이 원하는 정보를 찾아야 하는 번거로움도 가지고 있었다. 특히 모든 국민을 대상으로 하지 않고 특정사용자(학생 또는 학부모)를 대상으로 하는 교육행정기관의 홈페이지에서 이러한 문제점을 더 많이 발견할 수 있었다.



(그림 4) 키워드 검색 (부경대학교)

그림 4와 그림 5는 현재 교육행정기관에서 제공되고 있는 정보검색의 상황을 보여주고 있다. 그림 4는 교육행정기관 사이트에서 가장 많이 사용하고 있는 키워드 검색을 통한 정보 제공 형태를 보여주고 있는 부경대학교 홈페이지이고 그림 5는 게시판식의 질의 응답 형태를 사용하고 있는 부산광역시 교육청 홈페이지이다.



(그림 5) 게시판식의 질의응답 (부산광역시 교육청)

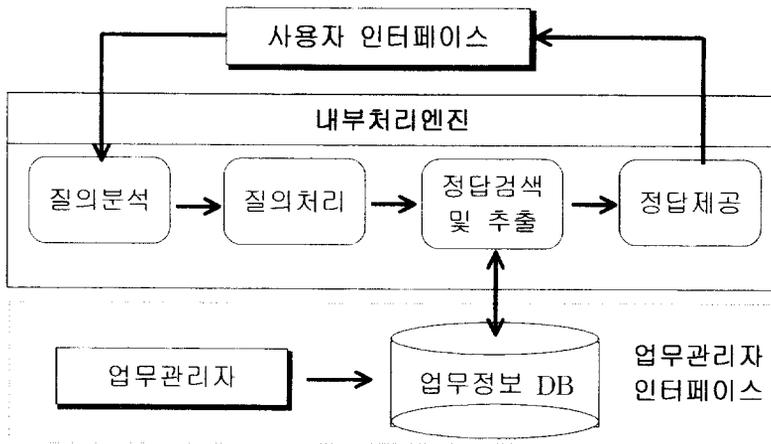
교육행정기관은 교육적 활동을 담당하는 교육기관임과 동시에 행정업무를 담당하고 있다. 교육활동과 달리 행정업무는 특성상 다수의 업무담당자가 각자의 업무를 분담하여 운영되기 때문에 특정 행정정보에 대해서는 업무담당자가 가장 정확한 정보를 가지고 있다. 그래서 정보를 알고자 하는 사용자는 홈페이지를 통해서 정보를 획득하지 못하였을 경우에는 해당정보의 담당자를 알고 있어야만 정보를 빠르고, 정확하게 획득 할 수 있다.

하지만 현재로서는 온라인상에서나 오프라인상에서도 업무담당자를 알기가 쉽지 않을 뿐 아니라, 교육행정분야에서 정보를 제공받는 사용자입장에서 이루어진 연구가 미비하여 검색된 사이트 중에서 질의응답시스템을 사용하고 있는 곳은 없었다. 따라서 현재의 교육행정 시스템에서는 담당자와 사용자가 보다 효율적으로 함께 사용할 수 있는 None-Stop Service의 필요성이 증대되고 있다.

### Ⅲ. 시스템 구성

이번 장에서는 자연어 처리를 통하여 사용자 질의에 대하여 응답을 제공하는 시스템을 제안하고, 시스템 구성에 대해서 알아본다.

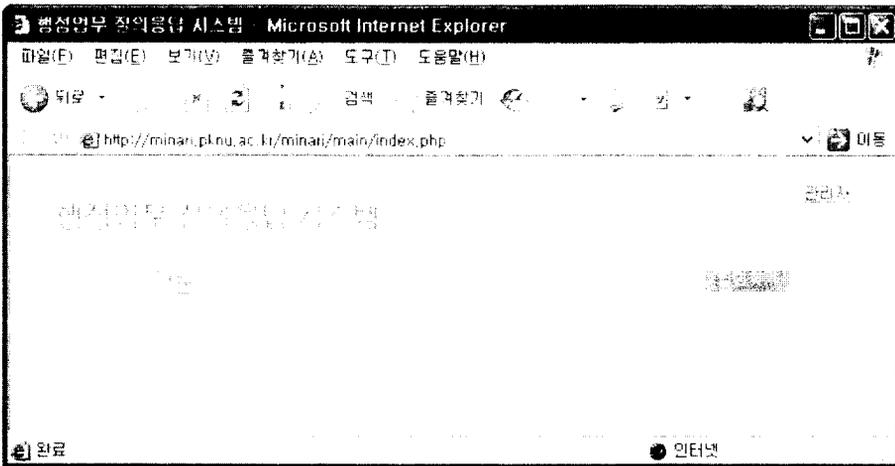
제안하는 시스템은 크게 세 부분으로 나누어 볼 수 있다. 그림 6에서 보는 바와 같이 사용자가 스스로 자연어를 통해 질문을 하고 답변을 받는 사용자 인터페이스, 사용자 질의를 처리하여 답변을 제공해주는 내부처리엔진진과 마지막으로 업무담당자가 데이터베이스의 내용을 관리하는 업무관리자 인터페이스로 이루어져 있다.



(그림 6) 시스템 구성도

### 3.1 사용자 인터페이스

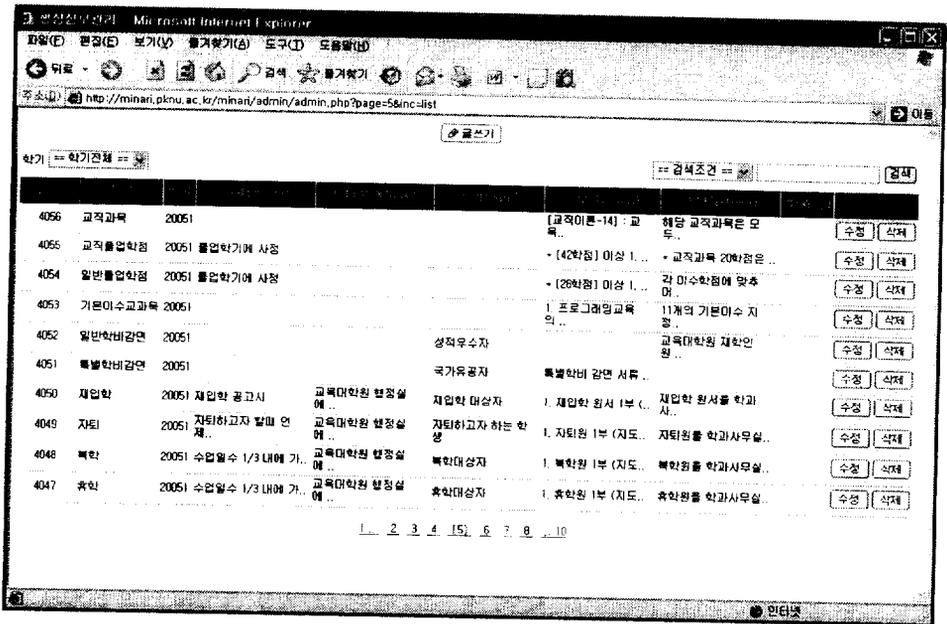
사용자 인터페이스는 그림 7과 같이 사용자가 직접 접할 수 있는 환경으로서 사용자가 질의를 하고 답변을 받을 수 있다. 사용자 인터페이스를 통해서 입력된 사용자 질의는 내부 처리엔진을 통하여 정형화된 뒤 이를 이용하여 업무관리자 인터페이스의 업무정보 데이터베이스에서 정답을 추출하고 최종적으로 사용자 인터페이스에서 질의에 대한 응답을 확인할 수 있다.



(그림 7) 사용자 인터페이스

### 3.2 업무관리자 인터페이스

사용자에게 정보를 제공해주기 위해서 대단위의 행정정보가 필요하다. 이 대단위의 정보를 사용자에게 제공해 주기 위해서는 행정정보를 관리할 수 있는 시스템 관리자가 필요하다. 하지만 한 사람의 시스템 관리자가 이 대단위의 모든 정보를 추가, 삭제, 수정 등 일괄 관리하는 것은 비효율적이다. 그래서 본 논문에서는 이 행정업무정보를 그림 8에서처럼 업무담당자가 직접 데이터베이스에 저장하고 확인 할 수 있도록 하였다.



(그림 8) 관리자 인터페이스

처음 시스템을 접하는 담당자에게는 별도의 관리방법에 대한 교육이 필요하지만, 업무담당자가 직접 해당정보를 관리할 경우에는 다음과 같은 장점을 가질 수 있다. 첫째, 업무담당자는 업무에 대한 변경사항에 대해서 가

장 민감한 대상이기 때문에 정보의 변경사항을 즉시 갱신할 수 있고 사용자는 최신의 정보를 획득할 수 있다. 둘째, 각 업무담당자는 사용자의 요구 정보를 직접 관리하기 때문에 정보에 대한 신뢰도가 높아지게 된다.

저장된 행정정보는 테이블 'info'에 저장되고 질의 처리를 통하여 정형화된 질의로부터 정답후보를 추출하기 위해 사용된다. 행정업무에서 행정정보 이외에 행정업무 담당자를 아는 것 또한 중요하다. 사용자가 원하는 행정정보에 대하여 추가의 정보를 얻기 위해서는 관련 행정업무 담당자를 알아야 할 필요가 있기 때문이다. 하지만 현재의 교육행정업무에서는 업무담당자가 보통 2년 단위로 업무가 변경되고 있는 경우가 많으므로 행정정보를 입력할 당시의 담당자 정보가 필요하다.

그래서 본 논문에서 제안한 시스템에서는 행정정보가 저장될 때 테이블 'topic\_base'에서 행정정보를 테이블 'staff'에서 담당자 정보를 가져와서 현재 업무담당자를 나타내는 테이블 'current\_duty'에 자동 저장이 된다. 즉, 관리자 이름으로 행정정보를 저장함과 동시에 현재의 업무관리자를 자동으로 구성하여 업무내용만으로도 업무관리자를 쉽게 찾을 수 있도록 하였다.

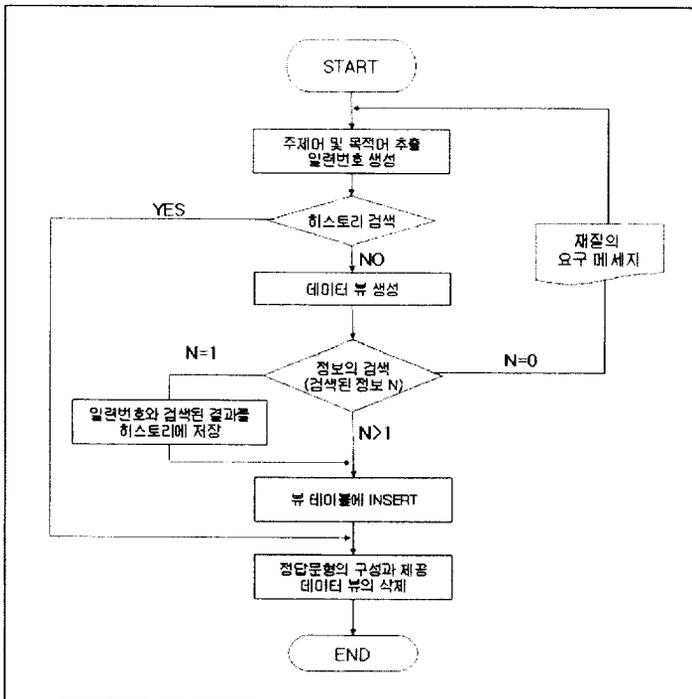
아래의 표 3에서는 행정정보를 나타내는 topic 과 담당자의 정보를 나타내는 staff 으로 구성된 테이블 'current\_duty'의 대해서 보여주고 있다.

<표 3> 테이블 current\_duty의 구성

term	topic_no	topic_data	staff_no	staff_name	staff_tel
2005-1	4002	입학식	201	이미나	620-6390

### 3.3 내부처리엔진

내부처리엔진은 사용자 질의에 대한 분석을 통하여 정확한 정보를 제공해주는 역할을 한다. 그림 9에서 보는 바와 같이 먼저 사용자 질의를 행정 업무에 맞게 일반문형으로 변형한 후 질의분석과정을 거쳐 주제어를 추출한다. 추출된 주제어는 사용자에게 정보를 제공해 주기 위하여, 시스템의 정보제공 형태에 맞게 구성된 후 업무정보 데이터베이스에서 정답의 검색 및 추출에 이용된다. 최종적으로 검색된 정보를 동적 정답 문형의 구성을 통하여 사용자에게 정답을 제공해준다.

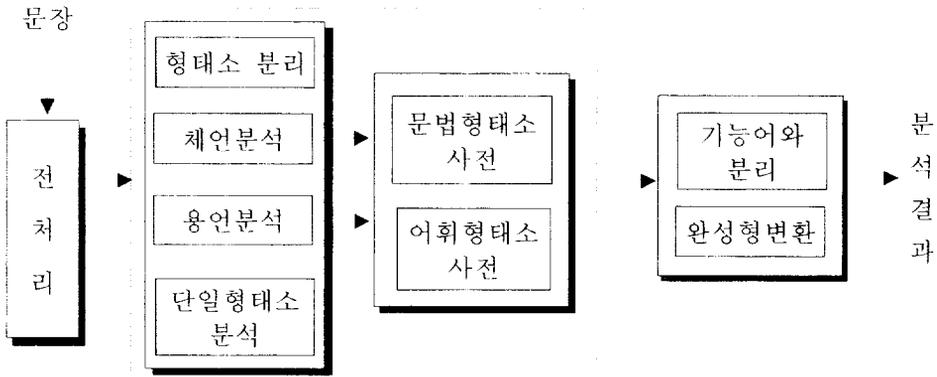


(그림 9) 내부처리엔진 흐름도

### 3.3.1 질의분석

질의응답시스템에서는 정형화되지 않은 질의를 분석하여 사용자의 의도를 추출해 내는 것이 중요하다. 왜냐하면 자연어의 표현은 무한하여 사용자 질의는 정형화된 형태가 있는 것이 아니라, 개인의 특성에 따라 여러 형태로 표현될 수 있어 사용자의 의도가 정확히 추출되지 않는다면 정답의 정확성을 보장 할 수가 없기 때문이다. 본 논문에서 사용자 질의 분석은 PHP로 제작된 공개된 형태소 분석기[17]를 사용하였다.

본 논문에서 참고한 형태소 분석기는 그림 10과 같은 구조를 가진다.



(그림 10) 형태소분석기의 구조

한국어의 단어 유형은 [명사 + 조사], [용언 + 어말어미], [용언 + 선어말어미 + 어말어미], [용언 + 명사형 어미 + 서술격조사 + 선어말어미 + 어말어미] 등으로 구성되어 있다. 따라서 한국어 형태소를 분석하기 위해서는 각 품사와 기능어의 분석이 무엇보다 중요하다.

형태소 분석기는 문장이 입력되면 전처리 과정에서 띄어쓰기 단위로 단어를 추출한 후 배열에 넣고 해당 단어별로 분리하여 분석을 한다.

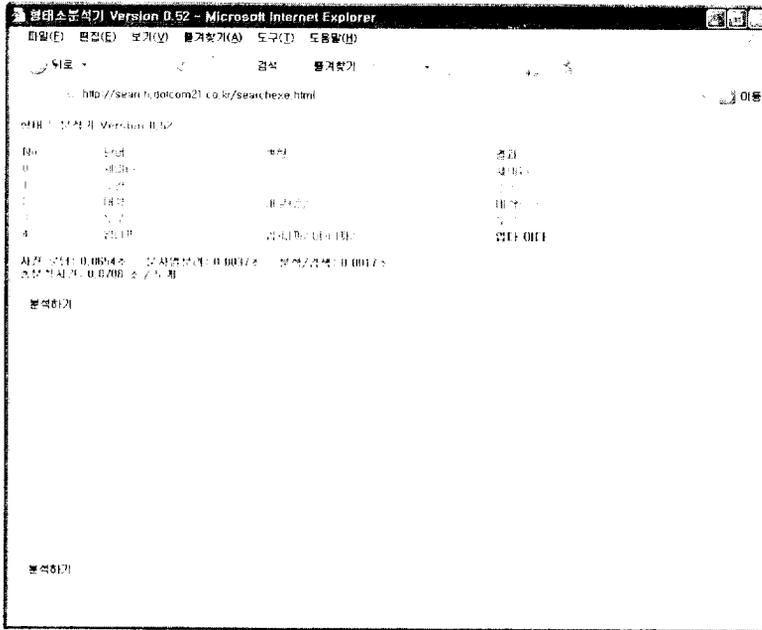
형태소 분리 단계에서는 단어를 구성하고 있는 각 형태소들을 분리한다. 한국어의 단어형성규칙에 따라 문법형태소인 조사와 어미를 분리하고 어미가 분리된 경우에는 선어말어미를 분리한다. 이 때 어휘형태소와 문법형태소간의 결합관계를 검사하여야 한다. 어미가 분리되어 용언으로 추정되는 단어에 대하여 불규칙 활용이 일어났을 가능성이 있는 것은 문법형태소 사전을 사용하여 용언활용의 기본형으로 분석한다. 문법형태소 사전은 용언의 활용형 뿐만 아니라 어미의 변환, 조사의 분석 등 총 4개의 개체 사전을 포함하고 있다.

체인 분석은 조사가 분리된 분석후보에 대하여 체인 분석을 행하는 과정이다. 형태소 분리 결과 조사가 분리된 단어는 체언으로 추정되므로 어휘형태소 사전에서 이를 확인하여 체언 분석을 시도한다. 어근부가 체언이 아닌 경우에는 접미사를 분리하고 다시 체언 분석을 시도하며, 명사형 전성어미가 발견된 단어는 용언 분석 단계에서 처리한다.

형태소 분리 단계에서 조사나 어미가 발견되지 않은 단어는 단일형태소로 이루어진 단어라고 추정할 수 있다. 이러한 유형의 단어는 어휘형태소 사전을 참고하여 접미사가 분리되는 경우에는 접미사를 분리한 후에 어휘형태소 사전을 확인한다.

이렇게 처리된 형태소는 품사와 기능어를 분리한 후 한글 완성형으로 변환하여 나타낸다. 마지막 단계에서는 형태소 분석기의 출력양식에 맞추어 분석결과를 재구성 하여 출력한다.

그림 11은 공개된 형태소 분석기를 실행시킨 화면이다. 분석된 형태소는 표의 형식으로 출력이 되며 분석결과는 사전로딩시간, 문자열 분리시간, 분석/검색시간, 총 분석시간, 형태소의 총 개수로 나타낸다. 또한 분석결과를 쉽게 구별하기 위하여 명사, 서술부, 불용어 등을 다양한 색으로 구분하여 나타내었다.



(그림 11) 형태소 분석기

하지만 공개된 형태소 분석기는 조사가 분리되어 체언으로 추정된 어근부가 체언으로 분석되지 않은 단어는 어근부가 복합명사로 인식될 수 있는지 확인하여야 하는데 이 과정을 거치지 않고 있다.

즉, 형태소의 원형 추출에 중점을 두고 있기 때문에 복합명사들은 각 명사 단위로 분리하여 추출하고 있다. 하지만 행정정보를 나타내는 용어들은 대부분 복합명사로 이루어져 있기 때문에 본 논문에서는 참고한 형태소 분석기에 논문의 특성에 맞게 불용어 제거와 복합어 구성을 추가하여 적용하였다.

복합어를 구성하기 위하여 총 분석된 형태소 개수에서 분석된 형태소의 수가 3개 이상일 때는 복합어 구성단계를 거친다. 이 단계에서는 서술부를 제외한 나머지 형태소를 어근부가 분석되지 않은 체언으로 판단하여 하나의 복합어로 구성한다.

분석된 형태소의 개수가 2개 이하일 때는 형태소가 단일 형태소로 분류될 수 있기 때문에 추가적인 복합어 구성 단계를 거치지 않는다.

```
/* Store a number of analyzed morphemes to variable cnt */
    $cnt = count( $words );

/* if cnt>2, store extracted morphemes to array except predicative
keyword */
    if( $cnt > 2 ){
        for( $i=0; $i<$cnt-1; $i++ ){
            $w[0] = $words[$i];
        }
        $w[1] = $words[$cnt-1]; /* store predicative keyword */
    }

/* else if cnt =1, store first extracted morphemes to cnt */
    else if( $cnt == 1 ){
        $w[0] = $words[0];
    }

/* else if cnt =2 , store extracted morphemes to array */
    else if( $cnt == 2 ){
        $w[0] = $words[0];
        $w[1] = $words[1];
    } else{ /* not extracted morphemes */
        $w[0] = "";
        $w[1] = "";
    }
}
```

(그림 12) 복합어 구성 알고리즘

이러한 복합어를 구성하기 위한 알고리즘은 그림 12와 같다.

### 3.3.2 질의처리

사용자 질의가 질의분석 단계에서 질의의도에 맞게 추출되면 질의처리 단계에서는 이 추출된 질의의 정답을 제공해 줄 수 있는 형태로 변형한다. 분석된 질의는 사용자가 행정정보에서 장소, 시간, 방법 등 사용자 의도를 의문사로 나타낸 목적어 OBJECT와 사용자가 원하는 행정정보를 나타내는 주제어 TOPIC으로 구성된다. OBJECT와 TOPIC은 표 4, 표 5와 표 6, 표 7과 같이 두 테이블의 쌍으로 이루어져 있다.

<표 4> object\_base 테이블

object_no	data
501	WHEN
502	WHERE
503	WHO
504	WHAT
505	HOW
506	WHY

<표 5> object\_match 테이블

object_no	data
501	시간
501	때
501	기간
501	언제
501	일자
...	...

<표 6> topic\_base 테이블

topic_no	data
4001	오리엔테이션
4002	입학식
4003	학생증발급
4004	수강신청
4005	논문제출
...	...

<표 7> topic\_match 테이블

topic_no	data
4001	오리엔테이션
4001	오티
4002	입학식
4002	입학
4002	신입생입학식
...	...

추출된 사용자 질의에서 사용자 의도를 나타내는 OBJECT는 사용자의 특성에 따라 다르게 표현될 수 있기 때문에 미리 정의된 사용자 표현인 표 5의 테이블 'object\_match'와 비교 한 후에 컬럼 (object\_no)의 값을 가진 후에 표 4의 테이블 'object\_base'의 정형화된 형태의 값 (data)를 가지게 된다.

이와 마찬가지로 추출된 주제어 TOPIC은 표 7의 테이블 'topic\_match'에서 사용자가 질의한 TOPIC과 매칭하여 (topic\_no)의 값을 가지게 되고 이 값을 표 6의 테이블 'topic\_base'에서 정형화된 형태의 값 (data)를 가지게 된다.

이렇게 각각 OBJECT와 TOPIC으로 분석된 질의가 각 (\_no) 와 (data)를 가진 후에는 웹사이트의 FAQ와 같은 기능을 가진 히스토리를 검색하기 위해서 별도의 일련번호생성이 필요하다. 일련번호는 3.3.3절에서 설명할 테이블 'history'에 정의 된 것으로 result\_no로 표현된다.

본 논문에서 result\_no는 OBJECT 와 TOPIC을 동시에 포함하여 행정정보의 특징을 나타내도록 [object\_no · topic\_no]와 같이 구성하였다.

예를 들어 [입학식 언제인가요?]라는 사용자 질의에 대해서는 다음과 같은 값을 가지게 되는 것이다.

<p>사용자 질의 : 입학식 언제인가요?</p> <p>분석된 질의 : 입학식(TOPIC) , 언제(OBJECT)</p> <p>처리된 질의 : <b>TOPIC</b> (no: 4002, data: 입학식)</p> <p style="text-align: center;"><b>OBJECT</b> (no: 501, data: WHEN)</p> <p>일련번호생성 : <b>result_no</b> = 5014002 (object_no · topic_no)</p>
--

### 3.3.3 정답의 검색 및 추출

정답의 검색 및 추출 과정에서는 사용자에게 정답을 정확하고 빠르게 제공하기 위하여 데이터 뷰의 생성을 통한 정답 제공과 히스토리를 통한 정답제공의 두 가지 과정을 거친다.

히스토리는 한번 검색된 내용이 저장되어 있는 곳으로서 질의처리과정을 거쳐서 생성된 각 질의의 고유 일련번호를 통해서 히스토리에 있는 데이터들의 정답 여부를 검색한다. 행정업무는 각 시기마다 특정적인 업무가 있고 그 시기에는 반복된 질문들이 많기 때문에 히스토리에 한번 질의된 내용은 반복질의 때 빨리 응답해 줄 수 있다는 장점이 있다.

<표 8> 히스토리 테이블

history_no	result_data	count	insert_date
5014002	3월 2일	1	2005.08.27

히스토리는 자주 묻는 질문에 대해서 빠르게 응답을 제공하기 위한 것이 목적이기 때문에 항상 최신의 정보를 담고 있어야 한다. 그래서 히스토리 테이블은 표 8과 같이 한번 검색된 후 저장될 때 날짜를 (insert\_date)에 저장하게 되고 현재 시스템 날짜 시간과 비교하여 일주일이 지난 정보 중에서 count값이 1이하인 것은 테이블에서 삭제하여 항상 최신정보를 유지한다.

즉, 히스토리 테이블에서 (count)값은 검색된 횟수를 나타내는 값으로서 기본값은 1을 가진다. (count)가 1 이하라는 것은 사용자가 한번 질의한 후에 일주일 동안 한번도 검색되지 않았다는 것을 의미하므로 테이블에서 삭제하게 된다.

한번도 질의 되지 않은 최초의 사용자 질의는 데이터 뷰의 생성을 통해서 정답이 추출된다. 뷰는 물리적 공간을 차지하지 않을 뿐 아니라 자료검색을 단순화하여 원하고자 하는 데이터만을 볼 수 있기 때문에 사용자 의도에 알맞은 정답을 제공해 줄 수 있다.

본 논문에서는 뷰 생성 과정에서는 검색된 정답의 개수를 파악하여 정답 제공과정에 영향을 준다. 검색된 정답의 개수를  $N$  이라고 하면,  $N=0$ 일 때는 검색된 정보가 없기 때문에 시스템은 사용자에게 재질의 요구 메시지를 보낸다.  $N=1$ 일 때는 검색된 정보를 뷰에 저장하고 정보가 하나라는 것은 정확한 정보라는 것을 의미하기 때문에 히스토리에 저장을 한다.  $N>1$ 일 때는 검색된 정보를 모두 뷰에 저장을 하지만, 히스토리에는 추가하지 않는다.

<표 9> 생성된 뷰 테이블

no	term	staff_name	staff_tel	objec_data	topic_data	result_data
1	2005-1	이미나	620-6390	when	입학식	3월 2일

뷰에는 단순히 행정정보만 저장되어 있는 것이 아니라 사용자에게 추가 정보를 제공해 주기 위한 업무담당자의 정보까지 포함하고 있다.

### 3.3.4 정답제공

질의응답시스템들은 정답을 제공해 주기 위하여 미리 일정유형의 정답문형을 정의해 놓고 그 유형에 맞추어 정답을 제공해 준다. 하지만 본 논문에서는 이러한 정답문형의 정의가 필요하지 않고 데이터 뷰의 컬럼명을 통해서 정답 문형을 구성하여 사용자에게 제공해준다. 표 9에서 보듯이 데이

터뷰에서 (topic\_ data) , (result\_data) , (staff\_name), (staff\_tel)을 사용하여 다음과 같은 정답 문형을 구성한다.

- 정답문형의 형성

[ "topic\_ data"은(는) "result\_data"입니다.]

[ ※ 담당자: "staff\_name" (" staff\_tel") ]

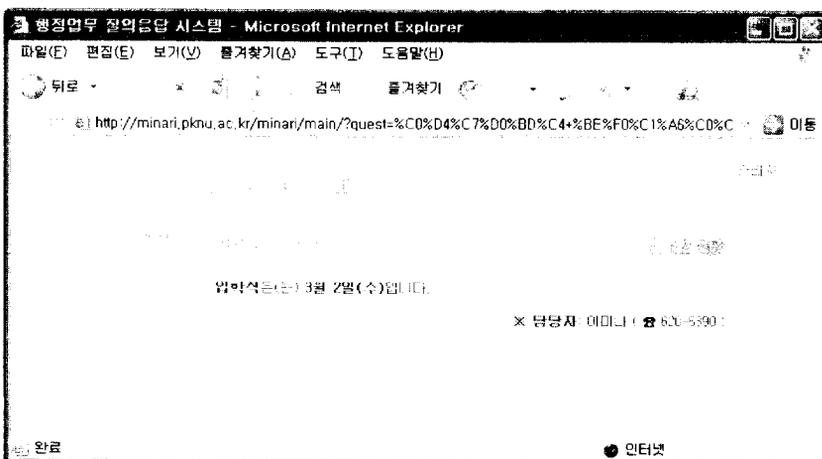
- 정답의 제공

[ 입학식은(는) 3월 2일입니다.]

[ ※ 담당자: 이미나(620-6390) ]

한번 검색된 내용을 다른 사용자가 다시 질의를 하였을 경우는 히스토리에 있는 정보를 가져오게 되는데 이러한 경우도 같은 형식으로 정답을 제공해 준다.

데이터 뷰를 통해 제공된 정보는 사용자에게 정답을 제공한 후 그 사용자 질의에 맞추어 생성된 데이터 뷰는 삭제를 한다.



(그림 13) 정답제공 화면

## IV. 실험 및 평가

이 장에서는 본 논문에서 제안하고 있는 자연어 처리를 이용한 질의응답 시스템 유용성을 평가하기 위하여 시스템을 구현하고 제안하는 시스템의 우수함을 검증하고자 한다.

### 4.1 실험 환경

실험을 위한 시스템의 운영체제는 Windows XP Professional에서 IIS(V.5) 윈도우 웹서버를 구축 하였고, 웹페이지는 PHP 5.0 과 데이터베이스 Mysql 5.0을 사용하여 시스템을 구현하였다. 행정업무정보는 2005학년도 1학기 부경대학교 교육대학원 전산교육전공의 업무내용을 참고하여 데이터베이스를 구축 하였다.

<표 10> 실험 환경

구 분	구축 환경
OS	Windows XP Professional
DBMS	Mysql 5.0
Language	PHP 5.0
Web Server	MS IIS 5.0
Web Browser	MS Internet Explorer 6.0
행정정보	2005-1년도 부경대학교 교육대학원 전산교육전공 업무내용

평가에 활용하기 위한 사용자 질의는 30자이내의 단문형식으로 자주 묻는 질의 100개를 선정하여 질의세트를 구성하였다.

<표 11> 사용자 질의세트

번호	질문
1	입학식 장소가 어디예요?
2	입학식에 무엇을 하나요?
3	졸업시험 언제집니까?
4	수강신청 기간을 알려주세요.
5	교육 실습비는 언제 돌려줍니까?
6	등대 장학생은 무엇을 지원해주나요?
7	휴학하고 싶은데 무엇을 제출해야 하나요?
8	기본 이수 교과목에는 무엇이 있나요?
9	선수 과목이란 무엇입니까?
10	세미나 수강 대상은 누구입니까?
~~~~~	
99	담당 조교 누구인가요?
100	언제부터 방학 입니까?

표 11에서 보는 바와 같이 사용자 질의세트는 단순 정의를 요구하는 질의부터 시간, 장소, 사람을 묻는 질의 등으로 다양하게 구성하였다. 또한 질의 4번과 같이 평서문도 포함하여 제안하는 시스템이 의문문에서 뿐만 아니라 다양한 질의형식에 대해서도 응답을 제공할 수 있는지를 확인하고자 하였다.

## 4.2 비교 실험

제안한 시스템을 비교 실험하기 위하여 자연어로 구성된 사용자 질의를 3단계로 세분화 하였다. 3단계 비교질의 세트는 기존의 자연어 질의에서 추출한 키워드 질의와 더 정확하게 검색하기 위한 이중키워드 질의로 구성하였다.

<표 12> 비교질의 세트

번호	자연어 질의	키워드	이중키워드
1	입학식 장소가 어디예요?	입학식	입학식, 장소
2	입학식에 무엇을 하나요?	입학식	입학식, 내용
3	졸업시험 언제칩니까?	졸업시험	졸업시험, 시간
4	수강신청 기간을 알려주세요.	수강신청	수강신청 기간
5	교육 실습비는 언제 돌려줍니까?	교육 실습비	교육 실습비, 반환
6	등대 장학생은 무엇을 지원해주나요?	등대 장학생	등대 장학생, 지원
7	휴학하고 싶은데 무엇을 제출해야 하나요?	휴학	휴학, 방법
8	기본 이수 교과목에는 무엇이 있나요?	기본 이수과목	기본 이수과목, 종류
9	선수 과목이란 무엇입니까?	선수	선수과목, 정의
10	세미나 수강 대상은 누구입니까?	세미나	세미나, 대상
99	담당 조교 누구인가요?	담당 조교	담당조교, 이름
100	언제부터 방학 입니까?	방학	방학, 일정

키워드질의는 자연어 질의에서 가장 중심이 되는 주제어를 사용하였다. 자연어 질의와 키워드 질의로 비교해 볼 때, 키워드 질의는 사용자 질의에 대한 분석 없이 바로 주제어인 TOPIC은 추출할 수 있지만 사용자 의도인

OBJECT를 추출할 수 없다. 그래서 이중키워드 질문세트를 추가하여 사용자 의도인 OBJECT도 키워드로 추출할 수 있도록 하여 보다 정확하게 제안한 시스템의 성능을 비교 평가하고자 하였다.

### 4.3 성능평가

질의응답시스템을 평가하기 위하여 질의응답시스템의 성능평가를 위한 척도로 사용되고 있는 응답의 평균 역순위( $MRR$ :Mean Reciprocal Rank)와 단일키워드, 이중키워드, 자연어질의로 비교한 시스템 적중률( $HR$ :Hit Rate)을 사용한다.

응답의 평균 역순위( $MRR$ )에 대한 정의는 식 1과 같다.

$$MRR = \frac{1}{n} \left( \sum_{i=1}^n \frac{1}{rank_i} \right) \quad (1)$$

식 1에서 사용된  $rank_i$  는  $i$  번째 질문에 대한 응답으로 제시한 것들 중에서 첫 번째로 정답인 것의 순위이고  $n$  은 질문의 수이다.

시스템 적중률( $HR$ )에 대한 식은 식 2와 같으며  $r_i$ 는  $i$  번째 질문에 대한 응답 중 정확한 응답의 포함 확률이다.  $n$  은 식 2와 같이 질문의 수를 나타낸다.

$$HR = \frac{1}{n} \sum_{i=1}^n r_i \quad (2)$$

시스템 적중률은 정확하게 응답을 제공한 경우에는 1의 값을 주고 응답을 제공하지 못한 질의에 대해서는 0의 값을 가진다. 만약 가능한 응답을 모두 제공한 경우에는 총 응답 가능한 질의 중에서 원하는 정답이 하나는 포함되어 있을 것이라 판단되므로 전체 응답 중 정답의 포함 확률을 계산하여 나타내었다.

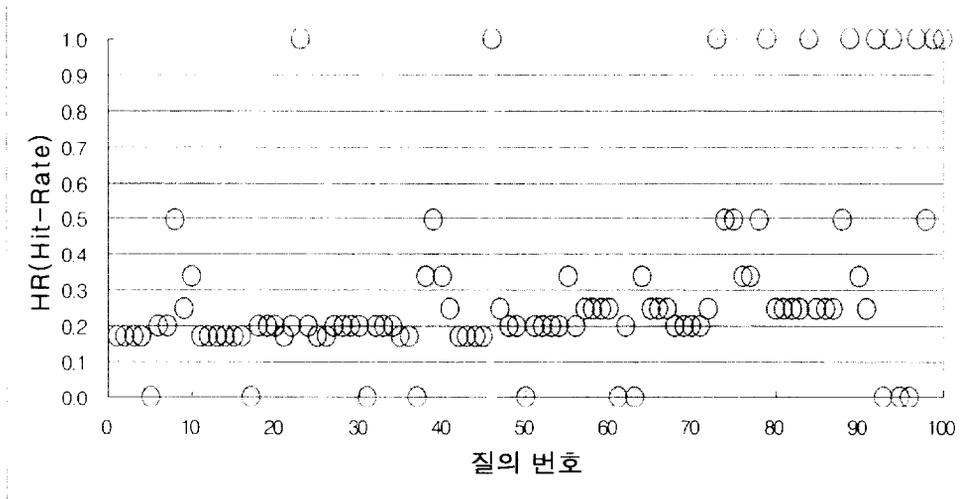
표 13에서는 표 12의 세분화된 질의세트에 대한 시스템 적중률을 비교하여 나타내고 있다.

<표 13> 시스템 적중률의 비교

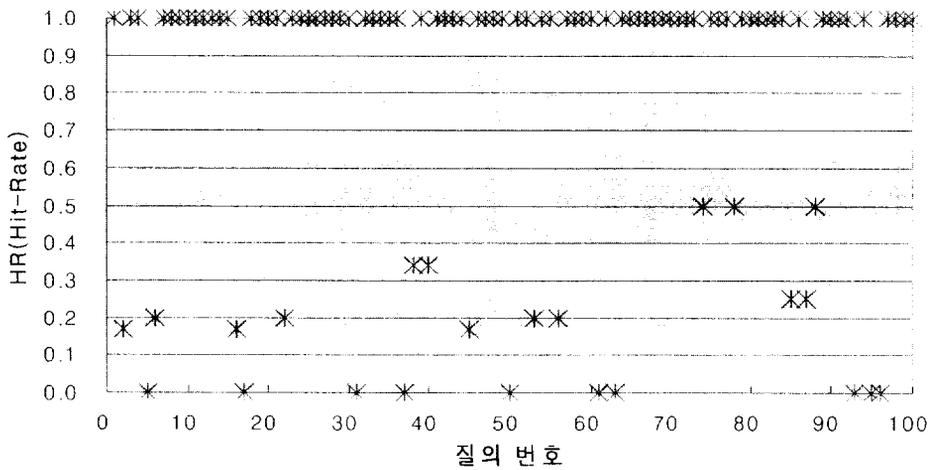
비교질의 번호	자연어질의 검색	키워드 검색	이중키워드 검색
1	1.00	0.17	1.00
2	1.00	0.17	0.17
3	1.00	0.17	1.00
4	1.00	0.17	1.00
5	0	0	0
6	1.00	0.20	0.20
7	1.00	0.20	1.00
8	1.00	0.50	1.00
9	1.00	0.25	0.17
10	1.00	0.34	1.00
~~~~~			
99	1.00	1.00	1.00
100	1.00	1.00	1.00
<b>평균</b>	<b>0.88</b>	<b>0.30</b>	<b>0.79</b>

평가 결과 MRR은 0.87를 얻었고 시스템 적중률(HR)은 표 13와 같은 결과를 얻을 수 있었다. 표 13에서 보듯이 단일키워드와 이중키워드에 대한 시스템 적중률(HR)은 각각 0.30과 0.79를 보였다. 자연어 질의의 시스템 적중률(HR)이 0.88인 것을 비교해 볼 때 본 논문에서 제안한 시스템이 단일 키워드에 비해서 34%, 이중키워드에 비해서는 9%가 뛰어난 시스템 적중률(HR)을 보임을 알 수 있었다.

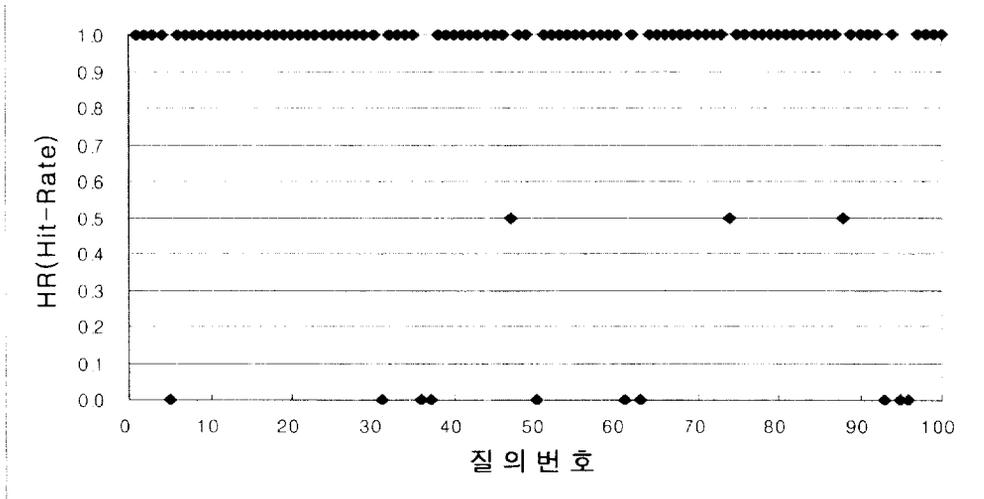
또한 총 100개의 3단계 비교 질의에 대한 응답 분포도를 그림 14, 그림 15, 그림 16에서 보여주고 있다.



(그림 14) 단일키워드의 응답 분포도



(그림 15) 이중키워드의 응답 분포도



(그림 16) 자연어질의의 응답분포도

단일키워드에 대한 응답분포도인 그림 14는 응답분포가 정답인 1.0 보다 아래에 많이 분포하고 있어 정확히 응답한 질의의 수가 적음을 보여주고 있다. 반면에 본 논문에서 제안한 자연어질의의 응답분포도인 그림 16은 응답의 분포가 정답인 1.0 에 가장 많이 분포함을 알 수 있다. 그림 15의 이중키워드 응답분포 역시 정확한 응답에 많이 분포하고 있지만 제안한 방법 보다는 아래에 분포하고 있는 질의가 더 많음을 확인할 수 있다.

최종적으로 제안하는 시스템의 자연어 질의가 응답을 제공하지 못한 몇 개의 질의를 제외하더라도 다른 두 질의에 비하여 정답에 분포한 질의가 많음을 알 수 있다.

## V. 결 론

인터넷이 보편화됨에 따라 상업적 용도뿐 아니라 사용자에게 정보를 제공해 주기 위해 많은 홈페이지들이 만들어지고 있으며 행정업무 역시 마찬가지이다. 하지만 현재 행정업무에 사용되고 있는 홈페이지에서는 여러 가지 문제점을 가지고 있다. 첫째, 단순정보만 제공함으로써 사용자가 원하는 정보를 신속히 알 수 없는 경우가 많다. 둘째, 정확한 정답이 아닌 관련정보의 리스트를 제공함으로써 사용자는 다시 정보를 검색해야 하는 비효율적으로 정보를 제공하고 있다. 셋째, 담당업무자를 쉽게 확인할 수 없어 추가 정보에 대한 제공이 제대로 되고 있지 않았다. 따라서 본 논문에서는 주변에서 가장 많이 접할 수 있는 교육행정업무에 적용한 질의응답시스템을 제안하였다.

제안한 시스템은 신속 정확하게 정답을 제공하기 위하여 자연어로 된 사용자 질의의도를 추출하고 각 사용자에게 맞추어 데이터 뷰를 생성한 후 정확한 응답을 제공하도록 하였고 반복적인 질문은 동적인 FAQ 관리기능인 히스토리를 통해서 신속히 제공하도록 하였다. 또한 별도의 시스템 관리자를 두지 않고 업무담당자가 직접 행정정보를 관리하도록 하여 제공하는 정보에 대하여 신뢰성을 높였으며, 항상 최신 정보를 제공하도록 하였다.

제안한 시스템의 효용성을 검증하기 위하여 '부경대학교 교육대학원 전산교육전공'의 행정정보를 가지고 시스템을 구현하였다. 시스템을 평가해본 결과, 자연어 질의를 통해서 친근하게 정답을 제공해주었고, 질의응답시스템의 성능평가를 위한 척도로 사용되고 있는 MRR은 0.87을 얻을 수 있었다.

또한, 사용자 의도를 파악할 수 있었기 때문에 단일키워드 검색일 때보다 34% , 이중키워드에 비해서는 9% 뛰어난 시스템 적중률(*HR*) 보여 제안한 시스템의 성능이 월등함을 보여주었다.

향후 연구과제로는 단순히 정답만을 제공해주는 시스템이 아닌 목적형 대화가 가능한 질의응답시스템의 설계 및 구현이다. 단순 신변잡기식 대화가 아닌 목적형 대화가 가능한 질의응답시스템이 되기 위해서는 질의를 할 때 해당 사용자에게 하나의 세션을 할당해야 하는 추가적 기술이 필요지만 본 논문에서 제안한 시스템의 데이터 뷰의 정보를 활용하여 사용자의 추가 질의에 응답을 해 줄 수 있다.

## [참고 문헌]

- [1] Aggarwal C., Wolf J., and Yu P., *A frame-work for the potimizing of WWW advertising*, In W. Lamersdorf and M. Merz(Eds), Trends in Distributed Systems for Electronic Commerce, LNCS 1402, Berlin: Springer, 1998.
- [2] Muller J. and Pischel M., "Doing business in the information marketplace," In Proceedings of the 1999 International Conference on Autonomous Agents, pp.139-146, 1999.
- [3] Chai J., Lin J., Zadrozny W., Ye Y., Stys-Budzikowska M., Horvath V., Kambhatla N., and Wolf C., The role of a natural language conversational interface in online sales: A case study, *International Journal of Speech Technology*, Vol.4, pp.285-295, 2001.
- [4] Ellen M. Voorhees, Dawn Tice, "The TREC-8 Question Answering Track Evaluation", Proceedings of the 8th Text REtrieval Conference(TREC-8), 1999.
- [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] 이강룡, " '지식검색'이라는 자유공간", 한겨레 칼럼, 2004.
- [7] 김영택외, "자연 언어 처리", 생능출판사, pp407-411 . 2001.
- [8] 롤란트 하우스 저, 장석진 외 8 옮김, "전산언어학의 기초" , 한국문화사, pp.11-28, 2002.
- [9] 정경택, "정보검색을 위한 자연어처리", 전자공학회지, No.9, Vol.24, pp.83-90, 1997

- [10] 강승식, “한국어 정보처리의 현황 및 발전 방향”, 한국음성과학회 제 6차 학술발표대회 학술논문집, 1999.
- [11] 박미화, 원형석, 이근배, “구문 분석에 기반을 둔 한글 자연어 질의로 부터의 불리언 질의 생성”, 정보과학회 논문지(B), 26권, 10호, pp.1219-1229, 1999.
- [12] 윤석환, 황혜정, 김교정, 윤용익, “질의 완화를 이용한 지능적인 질의 응답 시스템” 한국정보처리학회, VOL. 07 , pp.88-98, 2000.
- [13] 이동하, 김성민, 남도원, 이진영, “연관규칙을 이용한 지능적 질의처리 시스템 (Intelligent Query System using Association Rule)”, 한국지능정보시스템학회 학술대회 제1권, pp.171-174, 1998.
- [14] 이영신, 임희석, 임해창, “효율적인 질의응답 시스템 개발을 위한 BM25 기반의 단락 검색 시스템 (A BM25 based Passage Retrieval System for Developing an Efficient Question and Answering System)” : 한국컴퓨터교육학회논문지, 제 6권, 제 4호, pp.23-30, 2003.
- [15] 김학수, 서정연, “ 2-패스 색인 기법과 규칙 기반 질의 처리 기법을 이용한 고속, 고성능 질의 응답시스템”, 정보과학회논문지: 소프트웨어 및 응용, 제29권 제11호, pp.795-802, 2002
- [16] 황이규, 김현진, 장명길, “질의응답 기술 개발”, 한국정보처리학회, Vol.11, pp.48-56, 2004.
- [17] <http://www.mail21.co.kr>

## - 감사의 글 -

남 들보다 한 학기 늦게 시작해서 모든 것에서 한 학기 늦을 수 밖에 없었습니다. 논문 역시 예외가 아닐 수 없었습니다. 처음 논문을 준비할 때는 함께 같이 할 동기들이 없다는 것이 이렇게 큰 부담으로 다가 올지는 몰랐습니다. 그런 부담을 가지고 논문을 시작하면서 정말 수 많은 생각과 고민으로 몇날 며칠 밤을 새어 왔는데 이제 논문을 완성하고 이렇게 감사의 글을 쓰고 있는 것이 실감이 나질 않습니다.

이제 와서 생각해보니 혼자서 준비한다고 생각 했지만 항상 주위에서 격려 해주시고 도와 주신 많은 분들이 계신 것을 제가 모르고 있었습니다.

먼저 항상 저를 돌아 볼 수 있도록 차가운 가르침과 뜨거운 열정으로 교육대학원 생활을 무사히 마칠 수 있도록 지도 해주신 지도교수님 윤성대 교수님께 깊은 감사를 드립니다. 바쁘신 와중에도 논문 심사를 위한 세심한 관심을 보여 주신, 항상 따뜻하게 이름을 먼저 불러 주시던 박홍복 교수님, 조금 어려워하던 저에게 웃음으로 반겨주시던 김창수 교수님께 정말 감사드립니다.

학부과정부터 조교에서 대학원 생활까지 저에게 많은 가르침을 주신 박만곤교수님, 여정모교수님, 정순호교수님, 박승섭교수님, 박홍복교수님, 이경현교수님, 김영봉교수님, 신상욱교수님께도 감사드립니다. 그리고 항상 저에게 채찍과 같은 존재이셨던 하늘에 계신 박지환교수님께도 깊은 감사의 말씀드립니다.

부족한 논문 정성껏 살펴 봐 주셨던 순환 선배, 고석범 선생님, 윤종찬 선생님, 임선자 선생님께도 감사드립니다. 많은 고민과 걱정을 들어주며

해결방안 제시해주던 상일선배, 논문 때문에 옆에서 귀찮게 했던 은아선배, 귀녀선배, 정애에게도 감사하다는 말을 하고 싶습니다. 무엇보다도 옆에서 관심 가져 주고 신경 써준 고등학교 동창이자 대학동기인 지영이에게도 감사의 말 전하고 싶습니다. 그리고 항상 귀여운 웃음으로 미나쌤~ 하던 서영선 선생님, 항상 밝고 당찬 은경이 언니, 집이 먼 저를 노포동까지 태워주시던 김종필 선생님 외 연구실 모든 식구들에게 감사드립니다.

학부 때부터 인연을 쌓아온, 논문을 쓰면서 저에게 가장 큰 힘이자 구세주가 되었던 정훈선배에게 정말 감사한다고 말하고 싶습니다. 또 언제나 제 옆에서 묵묵히 지켜보며 정신적으로 때로는 물질적으로 지원을 아끼지 않았던 정욱이 선배에게도 감사드립니다.

대학을 졸업한 딸이 직장을 그만두고 대학원에 들어간다고 했을 때도 크나큰 성과가 보이지 않는 지금에서도 항상 딸을 믿어 주시고 아껴주시는 어머니, 아버지, 언제나 저의 고민을 해결해주는 하나밖에 없는 언니, 그리고 착한 형부, 철이 없어 보이지만 생각 많은 막내 동생 모두에게 사랑한다고 말씀드립니다.

이 지면을 통해 언급되지 않은 분들을 포함한 저에게 고마움을 주신 모든 분들께 다시 한번 감사하다는 말씀 전합니다.

2005. 12. 15

이 미 나