

자연언어에 의한 질의응답 시스템의 설계

○ 김 영길, 강 석훈, 우 요섭, 김 한우, 최 병욱
한양대학교, *인천대학교

Design of Q/A System By Natural Language Processing

○Y. K. Kim, S. H. Kang, *Y. S. Woo, H. W. Kim, B. U. Choi
Hanyang Univ., *Inchon Univ.

요 약

본 논문에서는 자연언어 처리를 기반으로 한 질의응답 시스템을 설계하며 그 대상을 호텔 예약 시스템으로 한다. 시스템은 크게 언어 해석부, 문맥 처리부, 화제 처리부 및 한국어 생성부로 구성한다. 언어 처리부에서는 단일화 문법을 적용하여 사용자가 입력한 질의문을 해석하며 그 해석결과로부터 이후 질의응답에 사용되는 정보들인 술어, 각 명사구들의 의존 관계치 및 의미 제약치를 얻어 낸다. 문맥처리를 수행하여 생략문 등에 의해 결핍된 정보들을 추출함으로써 정보의 손실을 보완한다. 이 정보들을 이용하여 사용자와 시스템간의 자연스러운 대화를 진행하기 위한 화제처리를 수행하며, 최종적으로 화제 처리부에서 파악한 화자의 의도를 기초로 사용자가 요구한 정보를 제공하고 계속된 대화진행을 위해 적절한 한국어를 생성해 주는 한국어 생성부를 거치게 된다.

본 논문에서는 호텔 예약이라는 정해진 영역하에서 자연언어처리에 의해 추출한 술어, 각 명사구들의 의존관계치 및 의미제약치 정보들을 이용하여 문맥처리와 화제처리를 수행함으로써 더욱 효율적인 질의응답 시스템을 구현할 수 있음을 보인다.

I. 서론

자연언어에 의한 질의응답 시스템은 질문자가 그 내용을 형식에 구애받지 않는 일상의 회화문으로 입력하면 의미를 파악하여 원하는 정보를 제공해 주는 시스템이다. 시스템의 목표(goal)는 사용자(user)에게 원하는 정보를 제공하는 것이므로 질의응답 도중 사용자가 요구하고 있는 정보가 무엇인지 그 정확한 의도를 파악할 필요가 있다. 일반적으로 자연언어에 의한 질의응답 시스템은 정형적인 텍스트가 아닌 일반 회화문을 그 처리 대상으로 하며 전화 통화문 등이 이에 해당한다.

질의응답 시스템에서 자연스러운 대화가 이루어지기 위해서는 화자의 의도나 대화의 배경이 되는 상식 등을 시스템 내에 형식화하여야 하고 대화의 전후 관계나 지시사 문제, 화제의 관리 및 일관성 유지 등이 필요하다. [1,6] 대화의 예측 처리 및 불연속적인 대화상황의 변화를 수용하는 시스템을 목표로 한다면, 적절한 소규모의 적용 영역을 대상으로 하는 연구가 현실적일 것이다.

본 논문에서는 입력된 한국어 문장의 처리에 의하여 해석된 술어와 각 명사구들의 의미제약치 및 그들간의 개념적인 의존관계를 사용하여 화제를 유도 및 관리해 나가는 질의응답 시스템을 제안한다. 이때 화제처리를 위하여 화제추출 테이블 및 화제관리 지식베이스를 정의하고, 질의응답 시스템의 일례로 호텔예약 시스템을 구축한다.

II. 시스템의 설계

시스템의 처리 개요는 그림 1과 같다. 시스템은 크게 언어 해석부, 문맥 처리부, 화제 처리부, 한국어 생성부의 4가지 부분으로 구성한다. 시스템의 유용성 또는 실용성을 고려할 때 이식성은 매우 중요하므로 본 시스템에서는 언어처리부를 질의응답 대상영역과 독립적으로 구현하는 방식으로 설계하였다.

사용자가 입력한 한국어 문장은 전처리 과정, 형태소 해석, 구문 및 의미 해석을 거쳐 의미표현식이 추출된다. [7-8] 의미표현식의 성분들을 사용하여 생략(Ellipsis) 및 대용(Anaphora)현상과 같은 문맥현상을 고려하는 문맥처리부를 거침으로써 결핍된 정보가 보완되고 화제처리에 필요한 완전한 정보가 얻어진다. 이 의미표현식의 정보와 화제추출 Table을 이용하여 사용자(user)의 의도파악과 대화진행에 있어서 필수적인 정보인 화제를 추출한다. 추출된 화제는 지식베이스에 미리 설정되어 있는 상황지식으로 부터 시스템이 예상하고 있는 화제와 비교되고, 이전에 입력된 다른 화제들과 함께 일관성 있게 관리되며 대상 영역에 대한 데이터베이스를 검색하여 적절한 응답을 사용자에게 제공한다.

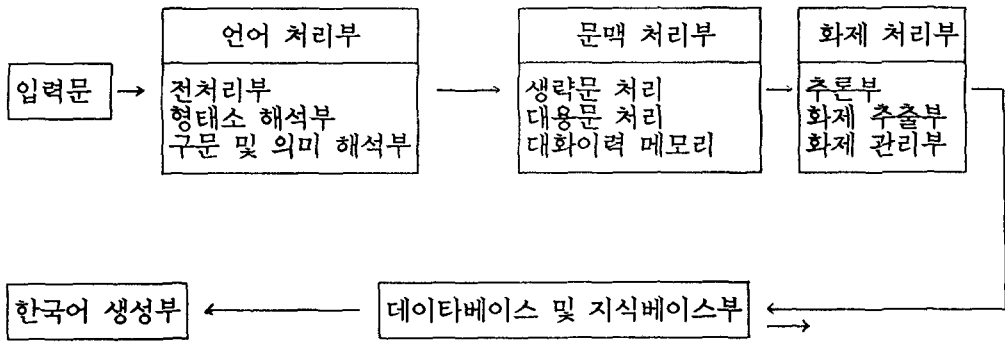


그림 1. 질의응답 시스템의 구성도
Fig. 1. Q/A System Configuration

Ⅲ. 언어 처리부

본 시스템에서의 언어 처리부는 사용자가 입력한 한국어에 대하여, 전처리, 형태소 해석, 구문 및 의미해석을 수행하여 의미표현식을 얻어내는 부분이다. 전처리부에서 각종 부호와 영어 등의 자종을 분류해 내고 그 후 형태소 해석과정에서는 어간과 어미를 구분한다. 형태소 해석에서는 조사나 어미등을 하나의 독립된 형태소로 취급하지 않고 어간부분에 특정한 소성을 제공하는 부속성분으로 고려함으로써 실제 계산기 처리상의 효율을 꾀하였다. 그림 2는 사용자의 입력문에 대한 형태소 해석을 거친 결과를 나타내고 있다. 그리고 하나의 어절을 하나의 단어로 처리하는 방식을 택하고 있다.

입력문 : 3일간 방을 예약했으면 합니다.

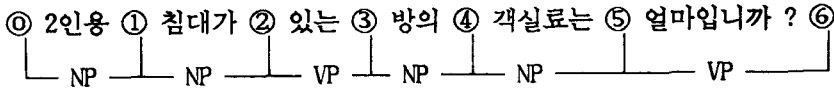
out_sentence :

```
((STEM (CAT NP)
  (SYN (HEAD (SEMK TIM)))
  (SEM (REF 3일)))
 (SUFFIX 간))
((STEM (CAT NP)
  (SYN (HEAD (SEMK ROC)))
  (SEM (REF 방)))
 (SUFFIX 을))
((STEM ((CAT VP)
  (SYN (HEAD (SUBCAT (ACC ROC))))
  (SEM (PRED 예약하다)))
 (SUFFIX 겠으면))
((STEM ((CAT VP)
  (SYN (HEAD (SUBCAT (VERB NON))))
  (SEM (PRED 하다)
  ((SUFFIX ㅂ니다)))
SEN_FORM : DECL
```

그림 2. 전처리 및 형태소 해석 결과
Fig. 2. Result of Preprocessing and Morphological Analysis

구문 및 의미해석은 어절내 해석과 어절간 해석으로 나누어 수행한다. 어절내 해석에서는 어간 이외의 부분들이 가진 정보를 어간내에 포함시키며 어절간 해석에서는 단일화 문법 규칙을 이용하여 최종적인 언어 해석을 행한다. 본 시스템의 언어 해석 알고리즘은 그림 3과 같이 일반적으로 사용하는 차트(chart)방식을 이용하여 상향(bottom up)해석을 수행한다. [4. 5]

>> 어절내 해석에 의한 초기차트 <<



>> 어절간 해석에 의한 최종차트 <<

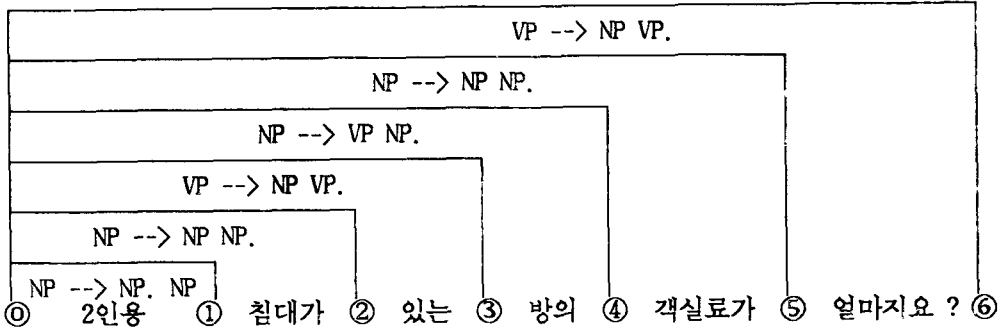


그림 3. 차트에 의한 해석의 일례
 Fig. 3. Example of Parsing by Chart

언어 처리부에서의 해석결과로 부터 이후의 대화상황 처리를 고려하여 추출한 의미표현식은 그림 3과 같이 술어의 의미제약과 각 명사구들간의 의존관계를 표시하고 있다.

입력문 : 2인용 침대가 하나 있는 방의 객실료가 얼마지요 ?
 언어 해석 결과 ::

```
(SEM (DEF (SUBJ 2인용_침대<BED>
            (ADV 하나)
            (PRED 있다)
            (LOC $V_1))
      (SUBJ (REF 객실료<MON>)
            (GEN $V_1 방<ROO>)))
      (PRED 얼마이다)
      (SEN_FORM QUESTION))
```

그림 4. 언어해석부에서의 의미표현식
 Fig. 4. Semantic Representation of Language Analysis

IV. 문맥 처리부

질의응답에서 사용자(user)가 문장의 일부를 대응어구로 대치하거나 불필요한 문장의 반복을 피하기 위해 문장 성분의 일부를 생략하는 경우가 빈번히 발생한다. 본 시스템에서는 언어 처리부의 의미 표현식에서 술어, 의존관계 및 의미제약 정보를 추출하여 대화도중 빈번히 발생하는 문맥적 현상인 생략문 및 대응문을 처리한다.

4.1 생략문 처리

생략문은 크게 문장간(intrasentential) 생략 및 문장내(intersentential) 생략으로 구분한다. 일반적인 생략문에 있어서 문장내의 경우보다는 문장간 생략이 주를 이루고 있으며¹⁾ 1.3) 본 시스템에서는 문장간 생략을 그 처리대상으로 하고 있다.

생략문의 판별은 언어 처리부의 의미 표현에서 얻은 성분들을 이용한다. 우선 술어성분의 존재를 확인한다. 술어성분이 존재할때 그 술어에 해당하는 결합가 패턴에서 필수격 성분의 유무를 조사하고 결핍된 필수격 성분들을 이전 대화의 이력 메모리에서 추출한다. 술어성분이 존재하지 않을시에는 이전 문장에서 술어를 찾아 동일한 과정을 거친다. 그림 4는 간단한 생략문의 처리과정을 표시하고 있다.

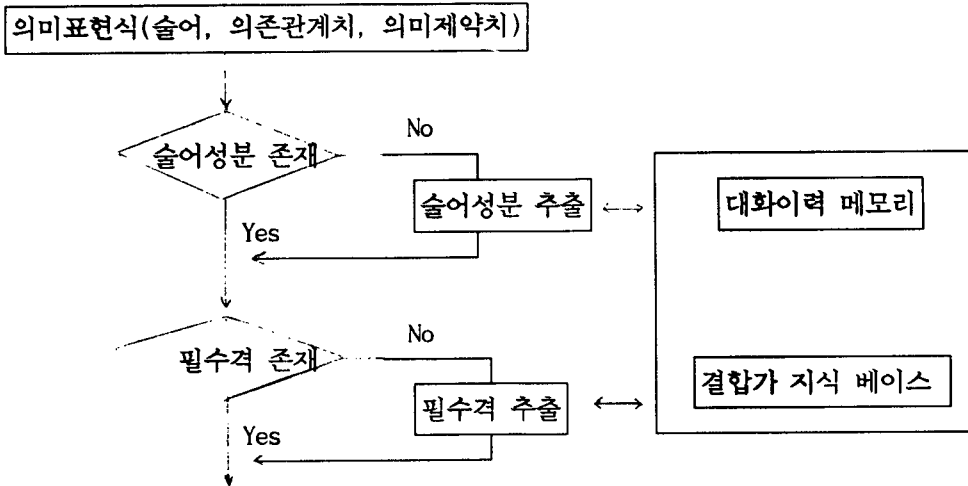


그림 5. 생략문 처리과정
Fig 5. Processing of Elliptical Sentence

user> Single Room을 예약했으면 합니다. --- (1)
hotel> 언제로 예약하시겠습니까? --- (2)
user> 5일부터 3일간요. --- (3)

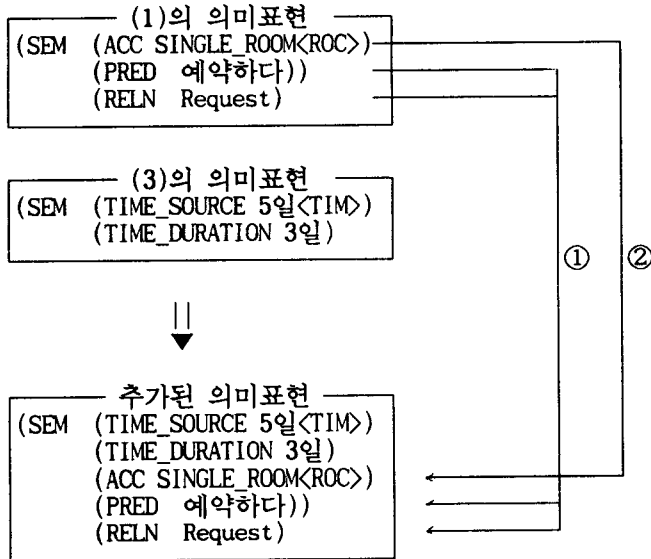


그림 6. 생략문 처리의 일례
Fig. 6. Example of Processing of Elliptical Sentence

4.2 대용문 처리

본 논문에서 대용문은 한 발화문 내 또는 연속된 두 개의 발화문에서, 앞 문장의 체언과 동일한 체언이 다음 발화문에서 나타날 때 사용하는 것으로 한정하기로 한다. 대용문의 처리는 대용어가 지시하는 이전 문장의 명사를 추출하여 대치함으로써 처리한다. 본 시스템에서는 대용문의 술어성분의 결합가와 대용어의 의미제약치를 이용하여 처리한다.

user> 5일부터 3일간 Single Room을 예약했으면 합니다. --- (1)
 hotel> 예, 예약하실 수 있습니다. --- (2)
 user> 그것의 객실료는 얼마입니까? --- (3)

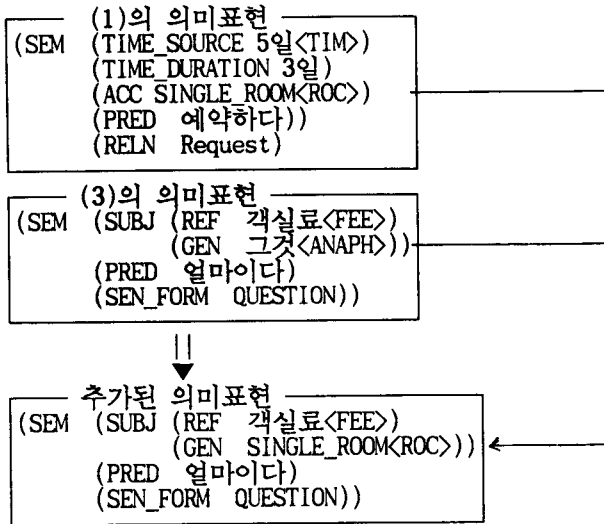


그림 7. 대응문의 처리
 Fig. 7. Processing of anaphoric sentence

V. 화제 처리부

화제 처리부는 크게 화제 추출, 화제 관리부로 나누어 설계한다. 문맥 처리부를 거친 최종적인 의미표현 정보를 이용하여 화제를 문장단위로 추출한다. 추출된 화제는 시스템이 관리하는 상황별 화제와 비교하며 예외적인 화제로 친이되었을 때는 예외 상황으로 규정, 이에 적당한 응답을 사용자에게 제공하여 예외 상황을 처리하고 다시 원래의 화제로 복귀한다. 예상된 화제라면 담화상황 관리부에서 이전에 입력된 다른 화제와 함께 관리하며, 처리 대상에 대한 데이터베이스를 검색하여 적절한 응답을 사용자에게 제공한다.

5.1 화제 추출

본 시스템에서는 대화상황의 명료성과 실용성을 고려하여 호텔예약에 관한 상황을 그 대상으로 설정했다. 호텔상황에 관한 화제는 화제(Topic)와 그것에 종속되는 부화제(Sub-Topic)로 구분한다. 화제는 예약, 예약 문의, 예약 확인, 예약 취소 및 예약 변경으로 구성하며 대화상황이 진행되면 표 1의 화제(Topic)추출 Table의 정보를 참조하여 화제를 추출한다. 이 화제 추출 Table은 입력문 해석에 의해 얻어진 술어, 의존관계치 및 의미제약치와 화제 사이의 관계를 나타내고 있으며 호텔 예약에 관해 수집한 246문장의 발화문들을 기초로 작성하였다. 그 후 추출된 화제에 종속하는 부화제(Sub-Topic)들을 표 2의 부화제 추출 Table을 이용함으로써 각 문장단위로 화제들을 추출한다.

화제	PRED	ACC	SUBJ	TIME_*	TOP	...	RELN	SEN_FORM
예약	예약하다	ROO/ROC		TIM			REQ	
예약	필요하다		ROO/ROC	TIM				
예약 문의	얼마이다		MON		MON			QUESTION
예약 문의	있다		ROO/ROC					QUESTION
:								:
예약 변경	변경하다	RES/ABS				...	REQ	

표 1. 화제 추출 테이블
 Table 1. Table for Obtaining Topic

부화제	PRED	ACC	SUBJ	TIME_*	TOP	...	RELN	SEN_FORM	ORDER
객실 / 시간	예약하다	ROC/ROO		TIM			REQ		1/2
객실 / 시간	있 다		ROC	TIM				QUESTION	1/2
객 실 료	얼마이다		MON		MON			QUESTION	4.1
:									:
예 외 상 황	포함되다		FOD			...			4.2

표 2. 부화제 추출 테이블(화제 : 예약)
Table 2. Table for Obtaining Sub-Topic (Topic : Reservation)

5.2 화제 예측

질의응답 도중 시스템은 화제 관리부에서 사용자(user)의 발화를 어느 정도 예측할 수 있다. 발화에 있어서 상대측이 정보나 어떤 행위를 요구하고 있다는 것을 알고 있다면 그 내용을 충족시키는 발화가 다음에 이어질 것으로 예측할 수 있는 것이다. 또 그 요구의 세부적인 사항에 대해 질문을 하거나 확인을 하는 절차도 고려할 수 있다. 본 시스템에서 이러한 예측은 시스템내에 형식화되어 있는 표 3의 발화간 문장기능의 접속관계에 근거하여 수행하고 있으며 자연스러운 질의응답의 진행이 이루어질 수 있도록 한다.

	패턴 1	패턴 2	패턴 3	패턴 4	패턴 5
1. 사용자	정보요구	행위요구	정보요구	행위요구	정보확인
2. 시스템	정보제공	행위수락/거절	재정보요구	재정보요구	확인인정/거절
3. 사용자			정보제공	정보제공	
4. 시스템			정보제공	행위수락/거절	

표 3. 발화간 문장기능 접속관계
Table 3. Connective Relationship Among Utterances

5.3 화제 변화

대화도중 다른 화제로 이야기의 내용이 바뀌는 경우는 흔히 일상적으로 일어나는 상황이다. 각 화제에 따라 시스템내에 형식화되어 있는 상식과 배경이 다르기 때문에 시스템의 화제변화 수용능력은 중요하다. 본 시스템에서는 한 번의 화제추출에 그치지 않고 각 문장에 대한 화제들을 계속적으로 관리해 나감으로서 화제의 변화를 수용할 수 있도록 하였다.

user>> 객실요금에 대해서 알고 싶습니다. ①
system>> single room은 1박에 10만원이고 double room은 15만원입니다. ② } 예약문의
user>> 내일 저녁으로 single room을 예약할 수 있습니까? ③ } 예약
system>> 예, 예약할 수 있습니다. 성함이 어떻게 되십니까? ④

그림 8. 화제 변화의 일례
Fig. 8. Example of Change of Topic

5.4 화제 관리

예상되는 모든 담화상황은 필수사항과 예외사항으로 분리하여 표 4와 같이 지식베이스에 부가한다. 필수사항을 3가지로 정의하였으며 그 밖의 예외사항들을 따로 정의하였다. 담화는 지정된 선행순서에 의거하여 화제를 예측 진행하되 예외상황이 발생하면 순서를 변경한다. 담화의 진행은 필수적인 예약상황을 모두 만족시키면 종료되며 필수사항이 모두 자연스럽게 충족되도록 화제를 유도한다.

	예 약	Feature	Value	
1	객 실 (N_ROOM)	TYPE	Single Room, Double Room, ...	
		PERSON_NUM	1, 2, 3, ...	
2	시 간 (N_TIME)	STIM	6월 26일, ...	
		ETIM	6월 29일, ...	
		DURATION	4일, ...	
3	이 름 (N_NAME)	NAME	김 철수, ...	
		TEL	290-0373, ...	
4	예 외 상 황	4.1	FEE	28만원, ...
			FEE_FOOD	nil, ...
		4.2	FOOD	nil, include
		4.3	TMP_BED	nil, include
		:	:	:

표 4. 화제 관리 지식 베이스
Table 4. Topic Control Knowledge Base

VI. 한국어 생성부

자연스러운 대화가 가능한 질의 응답 시스템을 구축하기 위해서는 진행되고 있는 대화상황과 질문자의 의도에 맞게 적절한 문장을 생성할 수 있어야 한다. 번역 시스템 등의 언어 생성부에서는 입력문 해석에 의해 얻어진 의미로부터 타당한 어휘 선정 및 표현 양태의 결정 등을 통해 문장을 생성하는 일련의 복잡한 과정을 거치는 것이 일반적이다

그러나, 질의응답 시스템에서의 언어 생성부는 입력문에서 얻어진 정보를 형식변경을 통하여 출력하는 과정이 효율적일 것이다. 본 시스템에서는 질의응답 도중 얻어진 화제(topic), 부화제(subtopic) 및 화자의 의도(mind) 등을 참조하여 시스템이 저장하고 있는 다수의 문장패턴들을 연관시키는 방법을 이용한다.

Topic	reservation	Mind	act_request	Subtopic	room & time
SYSTEM >> 예약하실 수 있습니다. 성함과 전화번호를 알려 주십시오.					
Topic	reservation	Mind	information_request	Subtopic	room_fee
SYSTEM >> SINGLE ROOM은 10만원이고 DOUBLE ROOM은 15만원입니다. 어떤 종류의 방을 원하십니까 ?					

그림 9. 한국어 생성 패턴의 예
Fig. 9. Example of Korean Generation Pattern

VII. 실험

그림 10은 호텔 예약에 관한 전형적인 질의응답에 관한 실험내용이다. 우선 ①에서 담화의 전체적인 화제가 "예약문의"라는 정보를 화제추출 Table을 통해 얻을 수 있으며 예약문의라는 화제에 부가되는 부화제들을 계속해서 추출함으로써 사용자에게 정보를 제공해 주고 있다. 그리고 ②에서 예약문의에서 예약으로의 화제의 변화가 일어나고 있다. 이 후 예약에 종속되는 부화제들과 사용자의 의도에 관한 정보를 이용해서 예약에 필요한 모든 필수정보를 얻어내고 있다. ⑤에서 최종적으로 정보확인을 거침으로써 예약에 관한 대화상황이 종료되는 질의응답 과정을 나타내고 있다.

한양대 인공지능 연구실 호텔 예약 시스템
 (Automatic Hotel Reservation System of Hanyang AI Lab)
 Tel : (02)290-0000


```

hotel>> 안녕하세요. 한양호텔 예약부입니다.
user>> 객실요금에 대해서 좀 알고 싶습니다. --- ①
hotel>> single room은 10만원이고 double room과 twin room은 15만원입니다.
user>> single room을 예약했으면 하는데요. --- ②
hotel>> 언제로 예약하시겠습니까 ?
user>> 3일부터 5일까지. --- ③
hotel>> 예, 예약하실 수 있습니다. 성함과 전화번호를 알려 주십시오.
user>> 이름은 김철수이고 전화는 200-0000입니다. --- ④
hotel>> single_room의 3일 숙박료는 30만원입니다.
      김철수씨 예약하시겠습니까 ?
user>> 예, 그렇게 하죠. --- ⑤
hotel>> 김철수씨 3일부터 5일까지 single_room 015호실이 예약되었습니다.
      문의전화는 290-0000으로 해주십시오. 감사합니다.
  
```

----- 대화도중 얻어진 정보 -----

```

topic      :: res_inquiry ==> reservation
name       :: 김철수
telephone  :: 200-0000
room       :: single_room
room_number :: 015호실
time.start :: 3일
time.end   :: 5일
  
```

>> 호텔 예약 장부 <<

room_type	room_number	user_name	user_tel	stime	etime	fee	..
single_room	015	김철수	200-0000	3	5	30	..

그림 10. 질의응답 시스템의 실험
 Fig. 10. Experiment of Q/A system

VII. 결론

본 논문에서는 질의응답 시스템으로 호텔예약 시스템을 선정하여 한국어에 의해 처리가 가능하도록 설계하였다. 언어처리부에서 추출한 술어와 각 명사구들의 의존관계 및 의미제약치를 이용하여 대화문 처리에서 필수적인 문맥처리 및 화제처리를 수행하였으며 이러한 처리를 위해 화제추출 Table 및 화제관리 지식베이스를 구축하였다. 대상영역으로써 호텔예약 시스템을 설계함으로써 그 유용성을 보였다.

질의응답 시스템은 자동통역 전화 등과 같은 한국어 해석부의 개발과 유사한 종류의 대상 영역에 대한 응용 시스템의 개발에 있어 기반 요소가 되리라 본다. 본 시스템에서는 복잡한 내포문에서의 화제 추출, 비문법적인 문장의 해석 및 보다 세밀한 화제 정의 등과 같은 사항들이 진행중에 있으며 보다 상세한 문맥적 처리가 지속적으로 연구되어야 한다.

참 고 문 헌

[1] Sandra Carberry : "A Pragmatics-Based Approach To Ellipsis Resolution", in Comput

- ational Linguistics, 1989
- [2] Steven L.Lytinen : "Semantics-first Natural Language Processing", AAAI Press, Vo 1.1, 1991
 - [3] R.E.Frederking : "Integrated Natural Language Dialogue - A Computational Model", Kluwer Academic Pub., 1988
 - [4] 長尾眞 : "畫像と言語の認識工學", コロナ社, 1989
 - [5] 菊井玄一郎, 鈴木雅實 : "表層表現の型を用いた對話文の生成について", 情報處理學會 自然言語處理研究會 研究報告 81-11, 1990
 - [6] 宇津呂 武仁, 松本 裕治, 長尾眞 : "第三者の觀點による對話の理解に關する考察", 情報處理學會 自然言語處理研究會 研究報告 83-17, 1991
 - [7] 우 요섭, 김 정근, 임 명현, 김 한우, 최 병욱 : "데이터베이스를 위한 자연언어 인터페이스 NAULI의 설계 및 구현", 대한전자 공학회 하계학술대회 논문집, 1991, 7
 - [8] 우 요섭 : "데이터베이스의 자연언어 인터페이스", 한양대학교 대학원 박사학위 논문, 1992, 12