

## 데이터베이스 검색을 위한 자연 질의어 변환 시스템

○  
신누미, 최준혁, 이정현  
인하대학교 전자계산학과

### Natural Query Translation System for Database Retrieval

Numi Shin, Junhyeog Choi, Junghyun Lee  
Department of Computer Science, Inha University

#### 요 약

본 논문에서는 대형 데이터베이스에서의 정보검색을 위한 자연언어 인터페이스 시스템을 구현하기 위한 방법을 제안한다.

질의문의 특성을 고찰하고 이를 일반적인 문장의 관점에서 수용하여 구문분석시에 반영한다. 구문분석 결과는 다음의 후처리 절차를 통해 정형 질의어인 SQL로 변환된다. 명사의 의미소성과 도메인의 어휘적 형태를 이용하여 질의문 내에 명시적으로 나타나지 않은 정보를 추출한다. 또한 질의문 내의 애틀리뷰트, 릴레이션, 상수의 관련성을 규명한다. 이 두 절차를 통해 기존의 질의어 변환 시스템에서 지식베이스화하여 사용했던 자료들을 구축할 필요가 없어지므로 데이터베이스의 변경, 삽입, 삭제에 의한 영향을 받지 않으며, 자료구조 생성에 따른 부담을 없앨 수 있다.

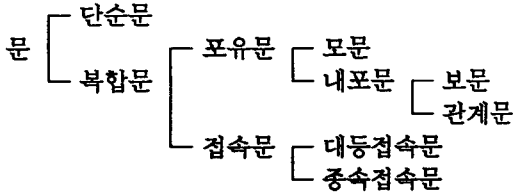
#### I. 서 론

현재까지 개발된 한글 질의어 변환 시스템으로는 NHI[2], HIDES[6], NAULI[11], K-NLQ[17] 등이 있다. 데이터베이스 자연어 인터페이스 시스템은 크게 형태소분석, 구문분석, 정형질의어 생성의 3 단계로 구성되어 있으며, 특히 구문분석은 사용자의 의도와 일치하는 올바른 질의문 생성에 있어서 가장 중요한 단계라고 볼 수 있다. 이러한 이유로 기존의 시스템에서는 질의문의 유형을 일반 문장과는 구별하여 정의하거나 질의어 처리를 위한 별도의 규칙을 만들어 사용하고 있는데, 전자는 구문분석의 필요성을 의심하게 하고, 후자는 질의문 처리만을 위한 구문분석기라는 점에서 일반성이 결여된다. 또한 다양한 의미 표현과 정확한 정보의 처리를 위해 여러 가지 사전과 사상 규칙들로 이루어진 방대한 지식베이스를 참조하고 있으나, 경우에 따라서는 불필요한 정보까지 포함하고 있는 예가 많다. 이런 문제점의 해결을 위해 본 시스템에서는 한글 질의문의 특성을 고찰하고, 구문분석에 반영하여 질의문은 물론 일반 문장의 분석도 동일하게 수행되도록 고려하였다. 형태소분석과 구문분석에 필요한 사전을 제외한 지식베이스는 해당 정보들의 관련성을 찾아 시스템 내에서 처리하였다.

## II. 한글 질의문의 고찰

### 2.1 구문유형

한글 질의문의 고찰에 앞서 한국어의 구문유형을 다음과 같이 분류한다[4].



위의 분류 역시 내포문과 종속접속문의 구조가 같다거나( 동사구절 내포문은 문장부사와 같은 성격을 보인다.) 대등접속문과 종속접속문의 구분이 모호하다는 등의 문제점을 안고 있으나 이런 사항에 대한 논의는 피하기로 한다.

관형절의 피수식어가 그 관형절 내부의 한 성분이 되는 구성을 관계절(relative clause) 혹은 관계문(relative sentence)이라 한다. 관계절은 모문의 필수적인 문장 성분 구실을 하지 못하며, 이러한 특성이 관계절과 보문을 구별해 주는 역할을 한다. 보문은 문장의 필수적인 성분으로, 모문에 삽입될 경우 종결어미가 원래의 형태를 유지하는지 여부에 의해 완형보문과 불구보문으로 구분되기도 하고, 명사를 보충해 주느냐 동사를 보충해 주느냐에 따라 명사구 보문과 동사구 보문으로 분류되기도 한다. 한글 질의문은 내포문 중에서도 관계문의 형태를 가장 많이 취한다.

접속문의 경우 대등접속과 종속접속을 나누는 데 있어서 선행문과 후행문의 의미 관련이 독립적이냐 의존적이냐 하는 것은 절대적인 기준이 될 수 없으며, 동일한 접속의 기능을 수행하는 형태소가 경우에 따라 대등접속어미가 되기도 하고, 종속접속어미가 되기도 한다. 이와 같은 문제점을 고려하지 않더라도 자연 질의문에서 문장의 접속관계는 그다지 큰 의미를 갖지 못하지만 접속문의 처리는 필요하며 2.2.2에서 다룬다.

### 2.2 한글 질의문의 특성

질의문에 [7]를 근거로 분석한 결과 자연 질의문의 특성은 다음의 네 가지로 요약된다.

예) 3학년 학생의 학번과 이름을 구하라.

-----

(1)

예) 학번이 88001인 학생이 수강한 과목의 번호와 성적을 구하라.

-----

(3)

(2)

-----

(4)

- (1) 명사구
- (2) 공동격 조사의 접속기능으로서의 출현
- (3) 이중주어 구조
- (4) 명사 술어 구문

### 2.2.1 명사구

명사구는 보통 속격조사에 의한 둘 이상의 명사의 결합으로 조사의 출현 여부와 결합형태에 따라 다음과 같이 구분할 수 있다.

- (1) 명사 + 속격조사 + 명사 ( 예 : 인형의 집 )
- (2) 명사 + 복합조사 + 명사 ( 예 : 자유에의 갈망 )
- (3) 명사 +  $\emptyset$  + 명사 ( 예 : 송아지 눈, 네 이름 )

(1)은 속격조사만으로 명사가 결합된 형태이고, (2)는 다른 조사가 속격조사와 복합조사를 형성하여 명사구를 이루는 경우로 한국어의 조사는 속격조사의 도움 없이는 명사구를 만들 수 없다. (3)은 단순히 두 명사가 조사 없이 결합된 경우와 앞 명사에 속격 조사 '의'가 포함되어 있는 두 가지 경우로 나타난다.

### 2.2.2 접속

접속은 구문분석시에 반드시 처리되어야 할 현상인데도 불구하고, 기존의 구문분석기에서는 구체적인 설명이 없거나[13,14,16] 언어학적인 입장에서만 기술하고 있다[9]. 여기서는 전산학적 관점에서 접속을 설명한다.

#### 1) 명사구 접속

명사구 접속의 기능을 수행하는 접속조사에는 '와, 하고, 이니, 이며, 이다, 에(다가)' 등이 있으나 여기서는 '와/과'만을 고려하기로 한다.

'와/과'는 '와 함께'의 의미를 가지는 공격(committative)조사로서 보다는 단순히 명사를 접속시켜 주는 접속조사로 더 널리 쓰인다.

- 예) 인생은 뜬 구름과 같다.                   -- 공격  
    강아지가 옷과 산발을 물었다.       -- 접속격

문장의 표면 형태가 명사구 접속인 문장을 보는 입장은 문접속과 구접속으로 구분되는데, 본 연구에서는 구접속으로 보는 입장을 취하기로 한다[4].

명사구 접속문은 접속된 명사구의 문장에서의 역할 즉, 주어 성분인가 목적어 성분인가에 따라 통사적 특징을 달리 한다. 두 경우 모두 문장 내에서의 위치 이동이 자유롭지만 명사구가 목적어 성분일 경우는 다른 언어 성분의 개입이 허용되지 않고 한 요소를 떼어 다른 곳에 위치시킬 수도 없다.

- 예) 그가 모자와 우산을 들고 나갔다.  
    \*모자와 그가 우산을 들고 나갔다.

이와는 달리 주어 성분일 경우에는 명사구들의 위치 이동의 제약이 있으나 명사구 사이에 목적어 성분의 개입이 허용된다.

- 예) 철수가 영희와 빵과 우유를 먹었다.  
    철수가 빵과 우유를 영희와 먹었다.

이 특성은 구문분석시 접속조사를 갖는 명사의 문장 내에서의 역할 규명의 어려움을 초래한다.

## 2) 동사구 접속

질의문에서 나타나는 동사구의 접속은 단문의 접속으로 보아도 생략 성분의 복원이 가능하므로 문접속으로 가정하되 대칭서술어의 경우는 구접속을 따른다[12].

예) 철수는 착하고 정직하다.

→ [ 철수는 착하다  
    철수는 정직하다

문장에서 생략어를 찾기 위해서는, 생략이 일어나는 방향에 주의할 필요가 있다. 동일한 명사구를 갖는 문장의 접속에서는 순방향으로 뒤에 오는 동일 명사구가 생략되며, 동사구의 경우는 이와 반대로 앞에 오는 동일 동사구가 생략된다[13]. 본 논문에서는 동사 위주의 파싱을 수행하므로 명사구의 생략만을 우선 처리하였으며, 동사구의 생략은 단순한 모듈의 삽입을 통해 확장이 가능하다.

### 2.2.3 이중주어

서술어를 제외한 나머지 성분의 위치가 자유롭고 상황에 따라 자명하다고 생각되는 성분이 쉽게 생략된다는 점 외에도, 국어에는 주어나 목적어가 여럿이 잇달아 나타나는 특성이 있다. 특히 주어가 들인 문장구성을 자주 볼 수 있는데, 이들 주어 사이에는 소유주와 소유물, 전체와 부분의 관계가 있다[5].

예) 영희가 마음씨가 곱다.

    그 옷이 색깔이 마음에 든다.

명사들간의 의미적 포함관계를 정의하여 포함관계의 성립 유무에 따라 구문 분석 단계에서 단순한 관형절인지 이중주어 문장인지를 구별할 수 있다. 질의문에서는 이중주어를 갖는 문장이 관형절로 내포되는 경우가 많으며, 이중주어 문장의 구분은 도메인 포함관계 검사과정을 내포하므로 대명사의 경우 관련 항목을 결정하는데 중요한 역할을 한다. 구문분석 단계에서 단순 관형절과 이중주어 문장의 분석 결과는 동일하게 표현되나 필요한 경우 이중주어의 주제를 추출할 수 있다.

### 2.2.4 명사술어구문

명사가 종결형서술격조사 '이다'와 서술어구를 이루는 문장을 명사술어 구문이라고 하는데, 질의문에서 가장 많이 나타나는 문형이다. 명사술어구문은 의해 다음의 5가지로 구분된다[1].

- 제1형 < N이 + N이다 >
- 제2형 < N이 + N에 + N이다 >
- 제3형 < N이 + N와 + N이다 >
- 제4형 < N이 + N보다 + N이다 >
- 제5형 < N<sup>1</sup>이 + N<sup>2</sup>이 + N이다 >

이 중 가장 출현 빈도가 높은 제1형과 제4형을 취하여 구문분석시에 처리하도록 하였다.

## III. 구문분석기

축소된 HPSG의 자질구조와 통합의 개념을 이용한 구문분석기를 제시한다.

### 3.1 자질구조

정보의 표현을 위해 HPSG의 자질구조(Feature Structure)를 간략화하여 이용한다.

자질의 종류는 다음의 세 가지로 구분할 수 있다[15].

- (1) 양분값의 자질( Binary - valued feature )  
: + 혹은 -를 값으로 취하는 자질
- (2) 다원값의 자질( Multi - valued feature )  
: 여러 값들의 집합에서 하나의 값을 취하는 자질
- (3) 범주값의 자질( Category - valued feature )  
: 통사범주를 값으로 취하는 자질

KPSG등의 한국어 문법에서는 조사 뿐만 아니라 어미들까지도 중심어라 할 수 있는 명사나 동사에 종속되는 구성 성분이 아닌 독립된 범주로 분류하고 있다. 그러나 본 논문에서는 조사와 어미를 각각 중심어인 명사와 동사에 속하는 성분으로 간주하여 속성화하지 않고, 하나의 명사구와 동사구로 간주하여 처리한다.

하위어의 범주는 문장 성분의 필수적인 요소인 보충어(COMPLEMENT)와 수의적인 요소인 부가어(ADJUNCT)로 구분할 수 있다[15].

부사어나 관형어는 부가어로 구분되는데, 부사는 용언을 수식하는 기능을 수행하지만 일부 용언은 부사를 필수적인 성분으로 요구하기도 한다. 이렇게 필연적으로 요구되는 경우를 제외하고는 부사어는 부가어 성분으로 처리되며 하위범주화의 틀에서 배제될 수 밖에 없다.

### 3.2 통합

통합은 다음과 같이 정의된다[18].

#### 1) 통합A

두 범주 X, Y는 양쪽 모두 어떤 자질 F에 대하여 특정한 값을 가지고 다음과 같으면 서로 통합할 수 있다.

- ① F가 원자적 자질이면, X의 F값과 Y의 F값이 동일할 때
- ② F가 범주를 값으로 가지는 자질이면, X의 F값과 Y의 F값이 통합할 때

#### 2) 통합B

두 집합의 범주는, 두 집합간에 어떤 일대일 대응관계가 성립하고, 한 집합의 구성원이 다른 집합의 구성원과 통합하면 서로 통합한다고 말할 수 있다.

그러나 이 개념만으로는 구문분석에 약간의 어려움이 있으므로 다음과 같이 확장하도록 한다.

#### 1) 계층속성의 처리

올바른 구문분석을 위해서는 명사의 의미소성[3] 처리가 필요한데, 의미소성은 계층구조를 가지며 의미소성의 통합은 하위 개념을 취해야 한다.

예) 철수가 영희를 사랑한다.

이 문장에서 '철수'는 '사랑하다'의 의미역 AGT 통합될 때 '철수'의 의미소성은 JAHM이고 동사 '사랑하다'의 의미역 AGT는 JA의 값을 갖는데, 이 두 의미소성은 하위 개념인 JAH로 단일화된다.

2) 접속조사를 갖는 명사는 의미역 할당 후 의미소성의 통합을 시도하지 않는다.

예) 아이가 사과와 빵을 먹었다

이 경우 '사과'는 동사 '먹었다'의 의미역 AFF로 통합되거나 의미소성의 통합은 실패한다. 이는 '사과'는 자연물로 JAN의 의미소성을 갖지만 '빵'은 인공물로 JIR의 의미소성을 갖기 때문이다. 그러므로 의미역과의 통합만 성공하면 두 명사의 의미소성은 통합하지 않는다.

### 3.3 하위범주화

중심어 동사의 SUBCAT값은 KPSG나 JPSG와 마찬가지로 순서를 문제시하지 않는 무순집합 (unordered set)으로 구성된 동사의 하위분류를 따른다[8].

### 3.4 문법규칙

명사등의 수식어 제약을 모듈에 포함시켜 자질구조를 단순화한 것과 마찬가지로 다음의 한국어의 수식구조의 특성과 단순한 어순 제약에 근거하여 구문분석을 수행한다. 수식구조의 특징은 다음과 같다[13].

- (1) 수식어는 항상 피수식어의 앞에 위치한다
- (2) 수식구조에는 비교차성이 있다.
- (3) 어순은 수식구조를 한 단위로 하여 바뀐다.
- (4) 수식구조는 수식의 범위를 최소화하려 한다.

한국어 어순상의 특징은 주부가 술부 앞에 놓이고, 수식어가 피수식어 앞에 놓인다는 두 원칙만 뚜렷할 뿐 나머지 문장 성분들의 순서는 자유롭다. 이러한 자유어순의 특성과 전산학적 측면에서 분석의 계층을 줄이고 평면 구조를 갖도록 하였다.

### 3.5 파싱 알고리즘

#### □ Algorithm

STACK과 CONJ\_STACK를 초기화

```
while ( 어절 ) {
  if ( 어절 ≠ 동사 )
    if ( 어절의 조사 = 접속격 ) part를 CONJ_STACK에 저장
      else part를 STACK에 저장
    else {
      공동격 명사 처리
      STACK으로부터 한 어절씩 읽어 부가어와 필수격 처리
      CONJ_STACK의 접속격 처리
      관형절의 처리
    }
}
```

생략성분 처리

## IV. 질의어 변환부

### 4.1 정보 추출

구문 분석결과로부터 정형 질의문 생성에 필요한 애트리뷰트와 릴레이션 이름을 추출한다. 이러한 정보는 문장 내에 명시적으로 나타날 수도 있고 생략되어 있을 수도 있다. 명

시적으로 나타나 있는 경우에도 필요없는 정보까지 추출되는 일도 있으므로 다음의 과정을 거쳐서 올바른 정보를 얻도록 하였다.

#### < 정보 추출 알고리즘 >

- (1) 문장 내에 명시적으로 나타나 있는 모든 애트리뷰트와 릴레이션 이름을 추출
- (2) 발견된 릴레이션에 모든 애트리뷰트가 속하면 (5)를 수행
- (3) (2)에서 발견된 애트리뷰트가 주키 애트리뷰트이면 (5)를 수행
- (4) 그 애트리뷰트가 속한 릴레이션을 찾아서 발견된 릴레이션의 리스트에 추가
- (5) 모든 애트리뷰트가 하나의 릴레이션에 속하는 경우 나머지 릴레이션을 삭제

여기서 추출된 정보들로 SELECT절과 FROM절이 이루어진다.

## 4.2 조인

### 4.2.1 조인 연산의 필요성 검사

보통은 발견된 릴레이션이 둘 이상인 경우 조인 연산이 필요하지만 동일 테이블에 대해 조인 연산이 필요한 특수한 경우가 있다. 전자는 단순히 릴레이션의 수를 셴으로써 판별이 가능하고, 후자는 릴레이션의 수가 하나이고 관형절의 서술어가 '같다'인 경우이므로 이 조건을 검사하면 조인 연산의 필요성 여부를 알 수 있다.

### 4.2.2 조인 조건 생성

조인 조건은 릴레이션들의 주키 애트리뷰트를 찾아서 서로 같은 경우만 조건으로 생성해 주면 된다. 동일한 테이블에 대한 조인의 경우는 모든 튜플이 상이한 값을 갖는 애트리뷰트를 찾아야 하는데, 도메인의 어휘적 형태는 알 수 있으나 값의 구체적인 형태는 알 수 없으므로 논외로 한다.

## 4.3 where predicate 생성

구문분석 결과인 연결리스트 형태의 단문들로부터 다음의 처리를 통해 정형 질의문 처리에 필요한 정보를 얻는다.

### 4.3.1 명사구 처리

속격조사는 고려하지 않고 명사구의 형태는 4가지로 요약된다.

- (1) 상수 + 애트리뷰트이름
- (2) 상수 + 릴레이션이름
- (3) 릴레이션이름 + 애트리뷰트이름
- (4) 상수 + 애트리뷰트이름 + 릴레이션이름 + 애트리뷰트이름

(1)은 동격관계이므로 where절의 predicate으로 직접 변환된다. (2)의 경우는 릴레이션 내의 애트리뷰트 중 도메인의 어휘적 형태가 상수와 같은 것을 추출하여 (1)과 마찬가지로 처리할 수 있다. (3)은 '릴레이션이름.애트리뷰트이름' 으로 변환되며 (4)는 (1)과 (3)의 결합 구조로 순차적으로 나누어 처리할 수 있다.

### 4.3.2 명사술어 구문

질의문에서 종결형 서술격조사 '이다'와 결합하는 성분으로는 상수와 대명사가 있다. 보어가 상수인 경우는 주어와 동격으로 처리하고, 대명사인 경우는 주어의 일반화 (generalization)로 주어를 검색 대상으로 처리한다.

## V. 시스템 구성

NATUR(NATURAL qUery tRansformer)는 한글 질의어를 입력으로 하여 동일한 의미의 정형 질의어인 SQL로 변환하는 시스템이다. 입력 문장은 형태소분석기[10]에 의해 최소의 유의적 단위인 형태소로 분리된다. 이 형태소분석 결과는 구문분석기에 의해 단문의 연결리스트 형식으로 분석되고 SQL생성기를 통해 최종적으로 정형 질의문을 생성하게 된다.

기존의 시스템들과는 달리 단어사전과 동사유형사전을 제외한 지식베이스를 구축하지 않으므로 사전구성 방법이 매우 중요하다.

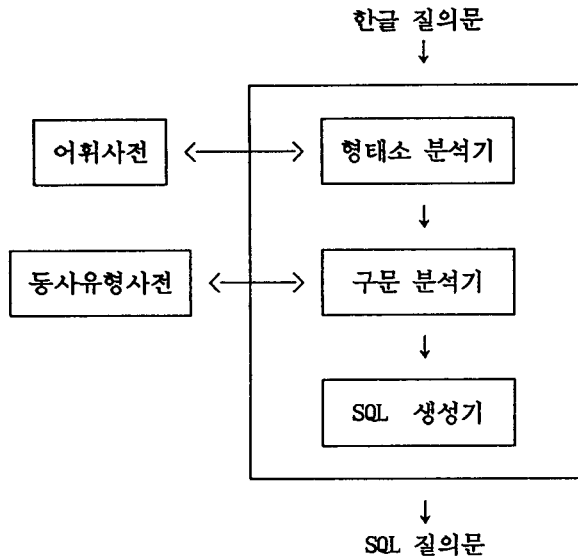


그림1. 시스템 구성도

### 5.1 사전

#### 5.1.1 어휘사전

어휘의 음운값과 품사(Part Of Speech)가 기본적으로 포함되고 명사인 경우는 하나 이상의 의미소성을 갖고 용언은 하나 이상의 하위분류유형의 값과 각 유형에 대해 의미역과 부합되는 의미소성 값을 갖는다. 의미역과 조사의 조합에 대해 가능한 의미소성을 테이블로 구성하면 동사마다 번거롭게 의미역과 의미 소성을 기술할 필요가 없다. 그러나 이와 같은 유형을 갖더라도 의미역이 다른 의미 소성을 취하는 경우가 있으므로 의미를 고려한 정확한 구문분석을 위해 이와 같이 사전을 구성하였다.

예) (도리 (CAT N) (JIR JAPL JAN II))  
 (구하 (CAT V) (VT10(AGT{JA},AFF{HS PS I B},SOR{I B L})))

#### 5.1.2 동사유형분류사전

단문과 내포문을 합하여 73가지 유형으로 동사(편의상 용언을 동사로 칭하기로 한다)를 하위분류하였으며[8] 구문분석시에 참조된다.

예) 단문 : [SUBCAT { !1:PP[이], !2:PP[이], !3:PP[에서] } ; CHD:1!,  
 CHA:2!, SPH:3!]



내포문: [SUBCAT { !1:PP[이], !2:PP[을], !3:V[ SUBCAT{], FIN[다],  
CMZ[고] } ; EXP:1!, AFF:2!, CTE:3!, DTRV[ SBJ:2! ]]

## 5.2 데이터베이스의 스키마 표현

처음 시스템을 실행시킬 때 데이터베이스 카탈로그를 참조하여 자동적으로 생성되므로 데이터베이스 스키마의 변화에 따라 수정할 필요가 없으며 구조는 다음과 같다.

### ◆ 릴레이션 테이블

릴레이션 이름	에트리뷰트 테이블 포인터	포인터
---------	---------------	-----

### ◆ 에트리뷰트 테이블

에트리뷰트 이름	에트리뷰트 형태	도메인 길이	주기 표시자
----------	----------	--------	--------

에트리뷰트의 주기여부에 따라 조인의 조건을 찾아내고, 질의문에 나타난 릴레이션의 이름 중 필요없는것을 제거한다. 에트리뷰트의 형태와 도메인의 길이 특성은 질의어 변환부에서 에트리뷰트 이름이 생략되고, 도메인 값만 명시적으로 나타나 있는 경우 해당 에트리뷰트 이름을 추출하는데 사용된다.

## VI. 처리 결과

### #1. 학번이 88001인 학생이 수강한 과목의 번호와 성적을 구하라

#### (1) 형태소 분석 결과

: 학번이((학번 (N (BM))) (이 (P))) 88001((88001 (NIL))) 인((이 (P VT7(EXI{ALL},QNT{ALL}))) (ㄴ (E))) 학생이((학생 (N (JAH))) (이 (P))) 수강한((수강하(V VT4(MAG{JAH},MFC{I}))) (ㄴ (E))) 과목의((과목 (N (IP))) (의 (P))) 번호와((번호 (N (BP))) (와 (P))) 성적을((성적 (N (HS))) (을 (P))) 구하라((구하 (V VT10(AGT{JA},AFF{HS PS I B},SOR{I B L}))) (라 (E)))

#### (2) 구문분석 결과

인 -> EXI:1! : 학번이 QNT:2! : 88001

수강한 -> MAG:1! : 학생이 MFC:2! : 과목의

구하라 -> AGT:1! : NIL AFF:2! : 성적을 ( 과목(N) ) 번호와 SOR:3! : NIL

#### (3) SQL 질의문

SELECT 성적 번호

FROM 수강

WHERE 학번 = 88001

### #2. 신장이 170 이상이고 180 이하인 학생의 모든 사항을 구하라

#### (1) 형태소 분석 결과

신장이((신장 (N (BM))) (이 (P))) 170((170 (NIL))) 이상((이상 (N (II)))) 이고((이고 (P VT7(EXI{ALL},QNT{ALL}))) 180((180 (NIL))) 이하((이하 (N (II)))) 인((이 (P VT7(EXI{ALL},QNT{ALL}))) (ㄴ (E))) 학생의((학생 (N (HS))) (의 (P))) 모든((모든 (D))) 사항을((사항 (N (II))) (을 (P))) 구하라((구하 (V VT10(AGT{JA},AFF{HS PS I B},SOR{I B L}))) (라 (E)))

#### (2) 구문분석 결과

이고 -> EXI:1! : 신장이 QNT:2! : 170

( 이상 )

인 -> EXI:1! : 신장이 QNT:2! : 180

( 이하 )

구하라 -> AGT:1! : NIL AFF:2! : 사항을 ( 학생(N) 모든(N) ) SOR:3! : NIL

(3) SQL 질의문

```
SELECT 학생.*
FROM 학생
WHERE 신장 >= 170
AND 신장 <= 180
```

#3. 3학년 학생의 학번과 이름을 구하라

(1) 형태소분석 결과

3((3 (NIL))) 학년((학년 (N (BM)))) 학생의((학생 (N (HS))) (의 (P))) 학번과((학번 (N (HS))) (과 (P))) 이름을((이름 (N (HS))) (을 (P))) 구하라((구하 (V VT10(AGT{JA},AFF{HS PS I B},SOR{I B}))) (라 (E)))

(2) 구문분석 결과

구하라 -> AGT:1! : NIL AFF:2! : 이름을 ( 3(NIL) 학년(N) 학생(N) ) 학번과 SOR:3! : NIL

(3) SQL 질의문

```
SELECT 이름 학번
FROM 학생
WHERE 학년 = 3
```

## VII. 결 론

본 논문에서는 한글 질의문을 정형 질의문인 SQL형태로 변환하는 시스템을 설계 및 구현 하였다. 질의문의 특성을 고찰하여 구문분석시에 반영하고 후처리를 통해 정형 질의어인 SQL로 변환하였다. 지식베이스를 사용하지 않는 시스템을 설계함으로써 자료구조의 생성, 변경 등의 많은 문제점에 대한 근본적인 해결책을 제시한다. 조인 연산으로의 변환이 불가능한 중첩 질의문, 특별한 의미관계 표현을 위한 여분의 자료구조를 요구하는 명사구접속문의 처리, 구문분석시 둘 이상의 동사 유형을 갖는 동사의 문제, 보조사와 복합조사의 처리에 대한 연구가 진행되어야 하겠다.

### ■ 참고 문헌

- [1] 강은국, 조선어 문형연구, 서광학술자료사, 1993.
- [2] 김성기/이석호, 자연 한글 질의어 처리를 위한 인터페이스의 설계 및 구현, 한국 정보과학회 논문지, 12권, 1호, pp31 ~ 44, 1985.
- [3] 김용호, 용언의 다의어 처리를 위한 SUBSYSTEM의 설계 및 구현, 인하대 공학석사 학위논문, 1993.
- [4] 김진수, 국어 접속조사와 어미연구, 탑출판사, 1987.
- [5] 남기심, 고영근, 표준국어문법론, 탑출판사, 1987.
- [6] 박전규, 한글 자연어로 구성된 질의문 처리 시스템 개발에 관한 연구, 외대 석사학위 논문, 1989.
- [7] 배해영, 데이터베이스론, 상조사, 1989.
- [8] 송영훈, 한국어 내포문 처리를 위한 SUBSYSTEM의 설계, 인하대 공학석사 학위논문, 1993.
- [9] 양재형, HPSG에 기반한 한국어 분석기의 연구, 서울대 공학석사 학위논문, 1990.

- [10] 여상화, 다단계 필터링을 이용한 형태소 분석기의 설계 및 구현,  
인하대 공학석사 학위논문, 1992.
- [11] 우요섭, 데이터베이스의 자연언어 인터페이스,  
한양대 박사학위논문, 1991.
- [12] 이익섭/임홍빈, 국어문법론, 학연사, 1983.
- [13] 이정현, 한국어 처리를 위한 구, 절 문법과 질문 응답 시스템,  
인하대 공학박사 학위논문, 1988.
- [14] 임명현, 단일화 문법에 의한 한국어 문장해석기,  
한양대 공학석사 학위논문, 1991.
- [15] 정희성, 한글 구구조문법, 제1회 자연언어처리 워크숍 발표논문집, pp7 ~ 9, 1987.
- [16] 조혁규, 단일화와 차트를 이용한 한국어 구문 분석 시스템의 구현,  
부산대 이학석사 학위논문, 1990.
- [17] 채진석, 자연 한글 질의어 시스템의 설계 및 구현,  
서울대 공학석사 학위논문, 1992.
- [18] Gunji, Takao, Japanese Phrase Structure Grammar, D. Reidel, 1987.
- [19] Dong-Young Lee, "A Computational Search for a Verb and its Corresponding Subject  
in the Korean Sentence Containing Embedded Clauses",  
Proceedings of the Pacific Rim International Conference on A. I.,  
Vol. 2, pp. 219-225, 1992.
- [20]. Pollard, C. and I. A. Sag, Information-Based Syntax and Semantics: Vol. 1,  
Fundamentals, CSLI Lecture Notes, 1987.