

# 자동색인을 위한 한국어 형태소 분석 알고리즘

이 영 주  
한국방송공사

(Korean Morphological Analysis Algorithms for Automatic Indexing)

(Young Joo Lee)  
( K B S )

## 요 약

자동색인이라 함은 기존의 수작업에 의한 색인어 선정 대신 컴퓨터에 의해서 자동화하는 것을 말한다. 한국어는 색인어가 될 수 있는 어근에 조사 및 어미가 붙어서 한 어절을 이루는 언어학적인 특성을 갖고 있다. 지금까지는 어근을 분리하기 위해 어근에 대한 사전을 구축하고 이를 Top-down 방법에 의해 처리하는 것이 통례였다. 그러나 이러한 방법은 외래어나 고유명사 등 새로 발생하는 어휘가 많은, 뉴스 원고와 같은 보도자료에는 쉽게 적용할 수가 없으며, 자연어를 다루는 타 분야에서도 미등록어에 대한 처리 방안이 시급한 실정이다.

본 논문은 어휘사전 없이 조사 및 어미의 생성 규칙을 이용한 Bottom-up 방식으로 처리하여 후보 색인어를 추론하고, 어절 상호간의 관계를 밝히는 구문분석을 통하여 이를 확정하는 알고리즘을 제안하였다.

## I. 서 론

기하급수적으로 늘어나는 정보의 홍수속에서 필요할 때마다 효용성 있는 정보를 제공하는 일이 무엇보다도 중요하다. 더우기, 문자인식 시스템의 개발로 다량의 자료 입력 방법이 시도되면서, 수작업에 의한 색인처리 방법에는 한계를 느끼기 시작했다.

이러한 연유로 얼마전부터 자동색인에 대한 관심이 일기 시작했으며, 이와는 별도로 인공지능 분야의 자연어 이해 시스템에 대한 많은 연구가 진행되고 있다.

최근에는 언어의 구조 및 의미 분석에 컴퓨터를 응용하는 컴퓨터 언어학(Computer Linguistics)이라는 새로운 연구 분야의 출현을 보기에 이르렀다.

지금까지는 어휘사전 혹은 색인어사전(Thesaurus)이 구축된 상태에서 불용어 처리에 의해 색인어를 추출하는 방법이 제시되었으나 색인어사전의 구축이라는 난제로 인해 효율적인 방법이 되지 못했다.

더우기, 신생어, 외래어 및 고유명사의 처리에 문제가 있어, 이들의 구성 비율이 많은, 보도 자료 및 여러 전문 분야에 포괄적으로 이용될 수가 없었다. 또한, 한국

어의 언어학적인 특성 때문에, 주어진 문장에서 필요한 어휘를 분리하는 일이 쉽지 않으며 이러한 문제는 기계 번역 시스템이나 자연어 이해 시스템에서도 대두되고 있다.

따라서, 어휘사전 없이 자연어의 연구들 통해서 이런 문제를 해결하는 방안이 요구되기에 이르렀다.

본 논문에서는 형태소 분석 단계에서 어휘사전을 이용하는 지금까지의 방법에서 벗어나, 한국어의 통사론에서 중요한 역할을 차지하는 조사와 어미의 생성 규칙을 구성하여 어휘사전 없이 형태소 분석을 수행하고, 통사 및 구문구조 분석을 통해 이를 확정하는 방안을 제안하였다. 형태소 상호 관계를 명확히 하기 위해 문법적 기능을 가진 기계사전을 구성하였으며 제시된 알고리즘에 의해 시뮬레이션을 하고, 그 결과에 대하여 기술하였다.

## II. 이론적 배경

### 2.1 자동색인의 이론

자동색인(Automatic Indexing)이라 함은 수작업으로 색인어를 추출하던 방법 대신 컴퓨터를 통하여 언어학

및 통계학적인 분석으로 색인어의 추출을 자동화하는 방법을 말한다.

자동색인 처리를 위해서는 어절에서 접사형태소와 조사 및 어미를 제거시킨 어근형태소를, 봉사 및 구문분석을 위해서는 조사 및 어미형태소를 이용하여야 한다. 한국어의 구조는 서구 언어에 비하여 한 어간을 중심으로 파생되는 어휘가 무수히 많다. 따라서 방대한 양의 체언과 용언을 전부 컴퓨터에 기억시킬 수 없으며, 처리 효과도 떨어진다. 또한 색인어의 추출에서 제외되어야 할 불용어에 대한 사전(Lexicon)을 구성하고, 체계적인 형태소 분석을 통해 이를 처리하여야 한다.

2.2 생성 규칙의 이용

구축된 사전을 통해 지금까지 이용된 방법에는 최장일치법과 Head-Tail 구분법이 있다.

전자는 이미 구축된 어휘사전을 통해 어절을 분리하고 그 중 최장매칭이 되는 것을 후보색인어로 보는 방법이다. 그러나 어절에 대한 두가지 이상의 분석이 불가능하며 이의 문제점 해결 방안으로 등장한 것이 후자이다.

그러나 이 역시 문법적인 규칙에 대한 정보가 적어 임의의 어절에 대한 좋은 효과를 기대할 수 없다.

한국어의 어형 변화가 아무리 많아도 조사와 어미의 형태소는 전체 자연어의 어휘 수에 비하면 극히 적은 편이다. 자연어를 일반화된 규칙으로 표현하기는 어려운 일이지만 어절의 생성 단계에서 볼 수 있는 규칙 및 제약 조건을 형태소 분리 단계에서 이용함으로써 위 두가지 방법에서의 문제점 해결을 시도할 수 있다. 물론 제안된 분리 방법이 한국어 어순의 복잡성 관계로 완전한 분리 방법이라고 볼 수는 없지만, 다른 방법에서 필수적으로 요구되는 어휘사전에 대한 제약을 피할 수 있다.

본 연구에서는 조사 및 어미사전을 Table화 하고 어미의 생성 제약 조건을 규칙화하였으며 형태소 분석을 거친 용언에 대해서는 기본형화함으로써 통계적 분석을 가능하게 한다. 또한 색인의 역할을 할 수 없는 불용어에 대한 정보를 사전화하여 초기 단계에서 이를 제외시킴으로써 불필요한 시간을 줄이고자 한다.

III. 형태소 분석 및 생성

3.1 조사의 분석 및 생성

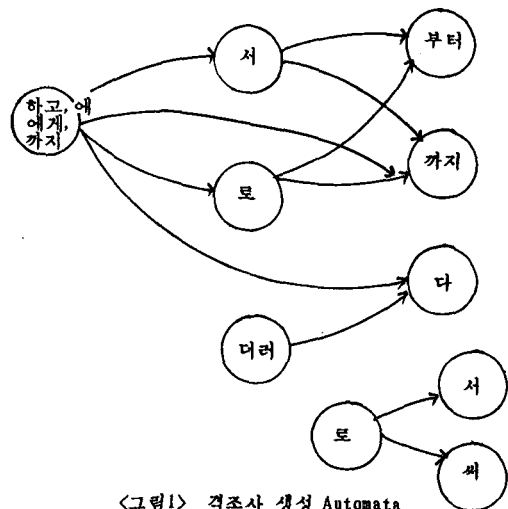
지금까지 조사에 대한 여러 연구가 있었으나 조사 상호간의 배열 순위와 그 배열에 따른 규칙을 추출하는데는 어려움이 많았다. 그 주된 원인은 조사 기능의 상관관계를 중심으로 한, 자연어 처리 방법이 활발하게 논의되지 못했기 때문이다.

따라서 기존의 분류 방법과 달리 배열 순위와 기능에 따라 조사를 재분류한 방안을 이용한다.

<표1> 조사의 기능에 의한 분류

| 조사의 분류 | 등급 | 조 사                   |
|--------|----|-----------------------|
| 격 조사   | 1  | 에, 예게, 한테, 께, 로, 더러   |
|        | 2  | 토, 서, 씨               |
|        | 3  | 다, 부터, 까지             |
| 연결조사   | 접속 | 와, 하고, 탕, 서건          |
|        | 이접 | 고, 나, 든               |
|        | 열거 | 예, 며, 하며              |
|        | 대등 | 만, 만큼, 처럼, 같이, 마냥, 대토 |
|        | 차등 | 만, 보다                 |
| 한정조사   | A  | 부터, 까지, 뿐             |
|        | B  | 밖에, 초차, 마저, 만         |
|        | C  | 나마, L 들, 나            |
|        | D  | 는, 도, 라도, 야말로         |
| 주어조사   | 9  | 가                     |
| 목적어조사  |    | 를                     |
| 소유조사   | 10 | 의                     |

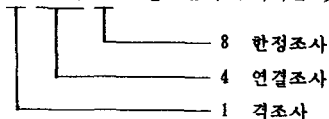
격조사의 어휘항목들은 1등급에서 3등급까지 분포를 보이는데 격조사들간의 상호 배열 유형을 위한 생성 Automata를 다음과 같이 표기한다.



<그림1> 격조사 생성 Automata

격조사는 조사의 복합배열상 <표1>과 같이 우선순위로 본다.

예) 네 게 보 다 는 / 철수한테 부탁하는 것이 좋겠다.



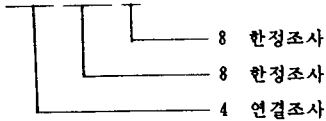
만일 이러한 배열 순위를 무시한다면 비문법적인 문장이 되거나 조사가 아닌 형태소로 추론할 수 있다.

예) 영 일 만 에서 / 가스층이 발견됐다.

위 문장에서 '만'이 연결조사나 한정조사로 사용된다면 격조사 '에'나 '서'의 뒤에 와야 한다. 따라서 어휘사건이 없어도 '영일만'이 체언의 역할을 하고 있음을 추론할 수 있다. 일반적으로 어휘사건에 고유명사를 일일이 갖고 있기는 어려운 까닭에 이러한 생성 규칙이 효과적으로 이용될 수 있다.

이러한 관계는 다른 조사류에서도 찾아볼 수 있다.

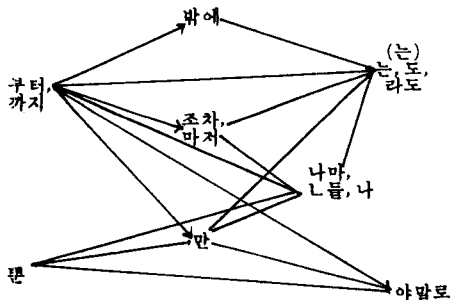
예) 너 하 고 조 차 도 / 즐기 싫다.



예) 한반도물 /무 단 조 차 까 지/ 한 일본은 중국을 넘봤다.

형태소 '조차'와 '까지'의 복합 배열 순서가 바뀌어 '무단조차'가 체언임을 추론할 수 있다.

한정조사의 상호 복합관계에 대한 Automata는 <그림2>과 같다.



<그림2> 한정조사 복합관계 Automata

주어조사와 목적어조사는 격조사, 연결조사 및 한정조사의 뒤에 나타나는 것이 복합조사의 배열 순위이다.

예) 집 예 틀 / 가야겠다.

후보색인어: 집

예) 이 / 고 울 에서 / 쉬어 가자.

후보색인어: 고울

소유조사는 조사의 복합배열에 있어서 제일 뒤에 오며 이 순서가 파괴되면 비문법적인 문장이거나 형태소 '의'가 조사가 아님을 추론할 수 있다.

예) 깊은 이 와 의 / 대화를 갖자.

후보색인어: 깊은이

예) 전문 의 예 계 / 진단을 받는 편이 낫겠다.

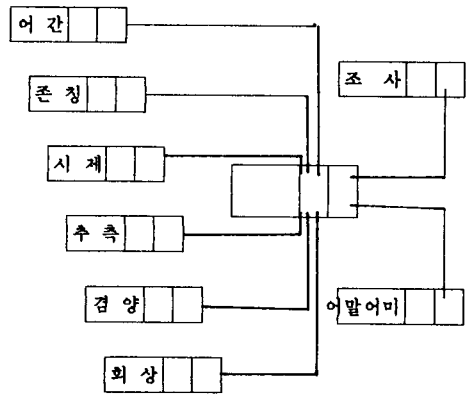
후보색인어: 전문의

### 3.2 어미의 분석 및 생성

용언의 형태론적 구성은 어휘적 형태소와 문법적 형태소로 결합되어 있다. 어휘의 의미를 나타내는 형태소가 어간, 문법기능을 나타내는 형태소가 어미이다.

어미는 다시 문장내의 위치와 순서에 따라서 선어말어미와 어말어미로 구분할 수 있다. 선어말어미는 어말어미를 반드시 필요로하지만 어말어미는 그 자체만으로도 단어를 이룰 수 있다. 선어말어미에는 경칭, 시제, 회상, 추측이 있으며 이들 각각의 순서가 고정되어 있어 그 위치를 바꿀 수 없다.

용언의 생성 규칙을 List구조로 나타내면 <그림3>과 같다.



<그림3> 용언의 생성 규칙 List 구조

### 3.3 명사 및 동사의 어미 생성 제약

명사는 기능 수행의 자립성 여부에 따라 자립명사와 의존명사로 나뉜다. 기능 수행의 자립성 여부라 함은 관형어를 필수적으로 가지는 것과 그렇지 않은 것을 말하는데 이러한 분류 기준에 따라 전자들 의존명사, 후자를 자립명사로 구분한다.

(예1) 영희가 그런 일을 할 리(가) 없다.

(예2) 영희가 그런 일을 하는 리(가) 없다.

(예3) 영희가 그런 일을 한 리(가) 없다.

위의 (예2)와 (예3)에서 처럼 의존명사 '리'는 관형화 어미 '-은', '는'과 함께 쓰일 수 없어 비문법적인 문장이 된다.

또한 어미들 중에는 계속되는 동사와 긴밀한 통합 관계를 유지하여야만 문법적인 구성이 되는 생성 제약을 볼 수 있다.

예) ---고 (싶다, 있다, 나다, 말다, 하다)

----지 (아니하다, 말다, 못하다)

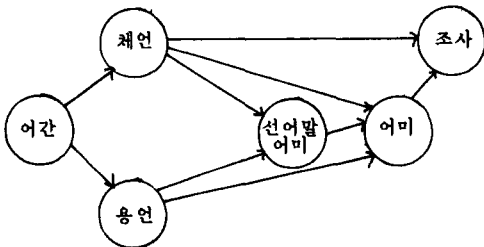
우리말에서 의존명사 및 동사의 수요는 제한되어 있어 이들에 대한 Table 구성으로 형태소분석에 이용될 수 있다.

IV. 시스템 구성 및 알고리즘 설계

4.1 시스템 구성

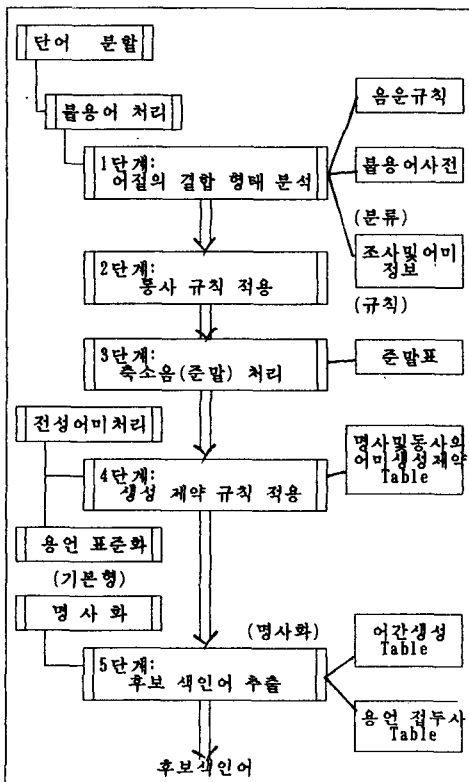
언어의 표현은 적어도 하나 이상의 문장으로써 나타나게 되는 까닭에, 문장이 언어 표현의 단위일 수밖에 없다. 그 문은 짧게는 단 한개의 어절로도 되고, 긴 것은 수십 행에 이르는 장문으로 되기도 한다. 여기서 문 구성의 단위를 어절이라고 한다.

어절에는 어미나 조사가 붙어서 다양한 형태를 이루는데 이를 Automata로 표시하면 <그림4>과 같다.



<그림4> 어절의 생성 Automata

자동색인을 목표로 한 한국어 형태소 분석 시스템 구성은 <그림5>과 같다.



<그림5> 형태소 분석 시스템 구성도

조사의 생성 규칙은 조사 및 어미 Table 구성시 각 조사별로 배열 순위에 따른 등급을 부여하며, 조사 생성의 복합관계를 규명함으로써 미등록 어휘에 대한 기능을 추론할 수 있다. 따라서, 별도의 Rule 정보를 요구하지 않는다.

어미의 생성 제약 규칙은 관계되는 어휘나 품사별로 어절 상호간의 제약조건을 구축한다. 명사의 경우 주로 관형화어미의 생성 제약에 대한 정보를, 동사의 경우 어미의 동사 생성제약에 대한 정보를 Linked list로 구성한다.

4.2 알고리즘 설계

1 단계: 어절의 결합 형태 분석

문장내에서 어절을 분리하여 어절의 결합 형태를 분석한다.

어휘변화가 심한 한국어의 특성 때문에, 초기 단계에서 분석된 어절의 형태는 다음 단계에서 변형될 수 있다.

INPUT : / \* 문장내에서 하나의 어절을 입력 단위로 한다. \* /

$W_{ij}$  :  $n$  개의 어절로 이루어진 문장내에서  $i$  번째 어절 중 뒤에서  $j$  번째 음소;

$P$  : 조사, >사, 대명사등 색인어치리에 중요하지 않은 정보를 갖고 있는 사전;

$T$  : 조사와 어미에 대한 정보를 갖고 있는 Buffer(자모순 Reversed Data);

OUTPUT : / \* 1 단계에서 분리된 어절의 임시 형태를 보관 \* /

$S_{ik}$  :  $i$  번째 어절을 결합수식  $k$  번로 분류;

METHOD :

```

FOR i = 1 TO n
    어절분리 : / * Space 를 기준으로 하며 길이는 m * /
    어절 전체를 Key로 하여 사전 P 탐색;
    IF (사전에서 발견되면) THEN
         $S_{i1}$  <= 형태소 및 품사;
    ELSEIF ( $W_{i1}$  = 명분이나 숫자) THEN
        / * 불용어처리를 위한 품사로 간주 * /
         $S_{i1}$  <= 형태소 및 품사;
    ELSE
         $k = 0$ ;
        WHILE (j > 1 &  $S_{i1}$  Not = 어말어미) DO
            / * 어말어미는 어절내에서 한가지만 존재한다고 본다 * /
             $k = k + 1$ ;
            IF ( $W_{ij}$  한글 자모) THEN
                WHILE (조사 및 어미 Record 마지막음소 =  $W_{ij}$ ) DO
                    조사 및 어미를 Key로 하여 어절과 Match 시킨다;
                    음운규칙에 맞는 내용만 임시 Buffer에 보관한다;
                    조사 및 어미사전의 다음 형태소를 수색한다;
                WHILE -END
    
```

```

IF (Not Matched) THEN
    Sij <= 형태소 및 품사 (용언 or 체언);
ELSE
    Sij <= 최장일치된 형태소 및 품사;
ELSE
    j = j + 1; /* 한발자도 이치는 Skips */
WHILE-END
DO

```

/\* 어절의 형태 분석을 출력한다. \*/

2 단계 : 품사규칙

1 단계에서 만들어진 어절의 결합 형태에 대해 품사규칙을 적용시키는 단계이다. 조사 및 어미의 생성 규칙을 적용시켜 새로운 결합 형태를 만든다.

INPUT : /\* 전단계에서 만들어진 어절의 결합 형태 \*/

S<sub>ij</sub> : n개의 어절 형태 Table중 i번째, 위에서 j번째 형태소

OUTPUT : /\* 새로운 어절의 결합 형태 \*/

U<sub>ij</sub> : 새로 만들어진 어절의 결합 형태

METHOD :

```

FOR i = 1 TO n
    WHILE (Sij이 어근이 아닌 형태소) DO
        k = 어근의 위치;
        SWITCH (Sij의 형태소 Type) OF
            CASE 조사
                IF (Sij+1의 형태소 Type = 조사) THEN
                    IF (Sij의 순위 .LE. Sij+1의 순위) THEN
                        FOR i = j + 1 TO k - 1
                            APPEND Si TO Sk;
                        ELSE j = j + 1;
                    ELSEIF (Sij+1이 명사형어미 or 의문형어미) THEN
                        FOR i = j + 1 TO k - 1
                            APPEND Si TO Sk;
                        ELSE
                            j = j + 1;
                CASE 선어말어미;
                IF (j = 1) THEN
                    FOR i = j TO k - 1
                        APPEND Si TO Sk;
                ELSEIF (Sij+1의 형태소 Type이 선어말어미) THEN
                    IF (Sij의 순위 .LE. Sij+1의 순위) THEN
                        FOR i = j + 1 TO k - 1
                            APPEND Si TO Sk;
                        ELSE
                            j = j + 1;
                    ELSEIF (Sij+1의 형태소 Type이 어말어미) THEN
                        FOR i = j TO k - 1
                            APPEND Si TO Sk;
                        ELSE
                            j = j + 1;
                CASE 어말어미
                IF (Sij+1의 형태소 Type이 조사) THEN
                    FOR i = j + 1 TO k - 1
                        APPEND Si TO Sk;

```

```

ELSE
    j = j + 1;
WHILE-END
END
/* 새로운 결합 형태를 출력 */

```

3 단계 : 준말 처리

한국어에서 용언의 형태소가 축소되어 나타나는 경우가 있다. 이것을 그대로 처리할 경우 용언의 기본형을 구하는데 어려움이 있다.

4 단계 : 명사 및 동사의 생성제한 규칙 적용

제 3장에서 기술한 명사 및 동사의 어미 생성 규칙을 적용한다. 조사의 Type을 이용하여 미동어의 수단을 확정하며, 용언을 기본형화 한다.

INPUT :

R : 명사 및 동사의 어미 생성 제한 규칙 Rule

U<sub>ij</sub> : 본디말로 변형된 i번째 어절중 위에서 j번째 Type

OUTPUT :

V<sub>ij</sub> : 수단이 확정된 결합 형태 중 용언은 기본형과 함.

METHOD :

```

FOR i = 1 TO n
    규칙 R 적용;
    SWITCH (Uij의 형태소 내용) OF
        CASE 조사 '을', '에', '에서', '로', '만이', '부터';
            /* 뒤에 연이어 올 가능성이 많은 조사 */
            SWITCH (Uij+1의 내용) OF
                CASE '아', '게', '지', '고';
                    /* 부사형 전성어미 */
                    DELETE & APPEND '다';
                    i = i + 1;
                CASE '나', '르', '는', '을', '던', '고', '고', '(으)면', '기';
                    /* 관형사형 전성어미 */
                    DELETE & APPEND '다';
                CASE '시아', '이저';
                    DELETE & APPEND '다';
                CASE '일';
                    DELETE & APPEND '다';
                CASE '(이)며';
                    IF (i > 2 & Uij+1의 TYPE가 조사)
                        DELETE & APPEND '다';
                    ELSE END
                CASE '인', '께', '등';
                    IF (i > 2 & Uij+1의 TYPE가 조사)
                        DELETE;
                    ELSE END
    END

```

5 단계 : 후보 색인을 추출

어간의 형성요소를 지닌 용언을 명사화시키고, 접두사를 분리시킴으로써 통계적 분석 방법을 적용시킬 수 있다.

V. 시뮬레이션

신문의 자동 편집을 위한 원고 시스템 도입이 진행된 1970년대 말, 방송국을 위한 뉴스 원고 시스템 개발이 뉴스센터의 진산화 일환으로 시작되었다.

그 뒤 ITN, BBC, TV-am(이상 영국), KRON TV, BBC(이상 미국), RAI(이탈리아)등의 각 방송기관에 설비되었으나, 한국에서는 아직 초기단계에 머물고 있는 실정이다.

5.1 분석자료 및 데이터 베이스

시뮬레이션을 하기 위한 대상으로서 KBS 뉴스 원고, 연합통신 기사, 각종 선언문 등을 이용하였다.

본 알고리즘은 고유명사나 외래어 같은 미등록 어휘를 사전 없이, 형태소분석을 하기 위하여 설계되었으므로 이들 자료가 적합하다고 판단되었다.

한국어의 조사 및 어미, 그리고 불용어를 데이터 베이스화 하고 생성 단계에서 발생하는 규칙에 대한 Rule들을 이용하였다.

5.2 시스템

본 시뮬레이션에서 사용하는 한글코드는 음소 단위로 처리하기 위해서 N-byte 한글체계를 사용하였다. 적용 H/W 시스템은 PRIME-9955를 이용하였다.

시뮬레이션의 형태소분석 단계에서는 <표2>에 나타난 항목을 중점으로 하여 시스템을 구성하고 이를 타 알고리즘과 비교함으로써 본 논문에서 제안된 알고리즘에 대한 평가를 얻고자 하였다.

<표2> 알고리즘의 비교 분석

| 비교관점    | 최장일치법    | H-T 구분법  | 본 시스템     |
|---------|----------|----------|-----------|
| 접근방법    | Top-down | Top-down | Bottom-up |
| 어휘사전    | 필수       | 필수       | 불필요       |
| 형태소추출   | 제한       | 제한       | 비제한       |
| 사건 TYPE | 품사별      | 품사별      | 불용어       |
| 사건크기    | 크다       | 크다       | 작다        |
| 기억공간    | 작다       | 작다       | 크다(규칙)    |
| 미등록어처리  | 미흡       | 미흡       | 가능        |

5.3 처리결과

시뮬레이션 실시 결과 TV 뉴스 원고에서는 85% 이상의 자료가 처리되었으며, 띄어쓰기의 오류가 인정되는 신문 기사는 이 보다 낮은 성공율을 보였다. 그러나, 국민교육헌장 및 선언문의 자료는 95% 이상의 성공율을 보이기도 했다. 따라서, 문법이 완벽한 문장일수록 높은 성공율이 기대되며 원시 자료 자체의 Error로 인한 실패율도 높다고 하겠다.

본 논문에서는 형태소 분석에 대한 알고리즘만 다루었으므로, 정보검색에서 요구되는 동어 처리가 뒤따라야 한다. 또한, 통계적인 분석으로 후보 색인어로부터 색인어를 선정하는 작업이 요구되며 보다 높은 성공률 위해서는 인공지능에 의한 완벽한 Rule-Base를 구축하여야 한다.

VI. 결 론

본 논문에서는 자동색인을 목적으로, 조사 및 어미의 생성 규칙에 의한 한국어 형태소 분석을 실시하고, 구문분석으로 이를 확정하는 알고리즘 설계에 관해 기술하였다.

지금까지는 형태소분석을 하기 위해 어휘사전 구축을 필수 조건으로 하였으나, 본 논문에서는 불용어 사전의 구축만으로도 미등록어의 처리가 가능하도록 하였다.

아직 우리나라 문법에 대해 학자마다의 의견이 달라, 문고부에서 제정한 통일문법을 기준으로 하였으나 생성 규칙에 대한 정보가 적어, 이를 위해서 최근에 발표된 논문에서 사용되는 용어를 도입하기도 하였다.

알고리즘 설계에 있어서 언어학적인 전문 지식을 수집하여 이를 적용하는 것이 쉬운 일이 아니었다. 특히 활용이 풍부한 한국어의 특징때문에 입력자료의 성격에 따라 성공률이 좌우되기도 하였다.

이처럼 언어학적인 전문 지식이 요구되는 분야에서는 각 분야의 전문가로 구성된 팀 구성으로 연구가 이루어져야 하겠다.

아직 우리나라에서는 보도 자료에 대한 자동색인 시스템이 도입되어 있지 않은 상태이나, 많은 양의 정보를 효율적으로 관리하기 위해서는 색인작업의 자동화가 필수적으로 요구된다.

따라서, 본 논문에서 제안된 알고리즘에 정보이론을 접목시키고, 더 나아가 인공지능의 추론능력과 Rule 생성능력을 강화하면 더 좋은 시스템이 될 수 있겠다.

참고 문헌

1. Salton, G., McGill, M. J., Introduction to Morden Information Retrieval, pp. 71-78, McGraw-Hill, 1983.
2. Harris, M. D., Introduction to Natural Language Processing, pp. 160-182, Prentice-Hall, 1985.
3. Tennat, H., Natural Language Processing, Petrolcelli Book, 1981.
4. Hovy, E. H., "Language Generation By Computer", Elsevier Science Publishers B. V., 1984.

5. Van Rijsbergen, C. J., Information Retrieval, pp.14-23, Butterworths, 1981.
6. 권재일, "국어의 복합문 구성 연구", pp.74-83, 집문당, 1985.
7. 배태영, 현대 언어학 개론, pp.131-132, 서린문화사, 1986.
8. 김영희, "한국어 조사류어의 연구", 문법연구 제 1집, 문법연구회, pp.271-311, 1986.
9. 김민수, 신국어학, pp.301, 일조각, 1964.
10. Ryo Yoshida, 뉴스센터 운용의 컴퓨터화, 제17차 KBS-NHK 방송 기술 세미나, pp.16, 1987.

부 록

일본의 지명인 역사학자인 이이우리가씨는 자신이 저술한 역사고고사흥 2차대전때 일본의 관측행위를 언급한 부분을 일본정부가 삭제 또는 수정하도록 한데 대해 보상을 받아내는 작업을 계속해 나가겠다고 밝혔습니다. 이이우리가씨는 어제 기자회견에서 자신이 22년동안 일본 정부를 상대로 제기해온 3건의 소송이 완전히 해결되지 않았다면서 일본정부가 2차대전중에 일어난 사건의 책임을 승거러 하고 있다고 비난했습니다. 이이우리가씨는 본부상이 약20만명이상이 사망한 것으로 알려진 1937년 중국 남경대학살사건에 대한 자신의 기술을 거부했으며 민주에서 전쟁포로들을 생체실험에 이용한 일본군의 새군부대인 731부대에 대한 언급도 배도록 요구했다고 지적했습니다. 이이우리가씨는 본부상이 오랜 역사고고사흥 사진 감철하는 전향제도를 완전 철폐하려고 촉구했습니다.

●● PHASE 1: ●●

|               |   |         |   |     |
|---------------|---|---------|---|-----|
| 어질 : 일본의      | = | 일본      | + | 의   |
| 어질 : 지명한      | = | 지명한     |   |     |
| 어질 : 역사학자인    | = | 역사학자    |   |     |
| 어질 : 이이우리가씨는  | = | 이이우리가씨  | + | 는   |
| 어질 : 자신이      | = | 자신      | + | 이   |
| 어질 : 저술한      | = | 저술한     |   |     |
| 어질 : 역사고고사흥   | = | 역사고고사흥  |   |     |
| 어질 : 2차대전때    | = | 2차대전때   |   |     |
| 어질 : 일본의      | = | 일본      | + | 의   |
| 어질 : 관측행위     | = | 관측 행위   | + | 를   |
| 어질 : 언급한      | = | 언급한     |   |     |
| 어질 : 부분       | = | 부분      | + | 을   |
| 어질 : 일본정부     | = | 일본 정부   | + | 가   |
| 어질 : 삭제       | = | 삭제      |   |     |
| 어질 : 또는       | = | 또       | + | 는   |
| 어질 : 수정하도록    | = | 수정하     | + | 도록  |
| 어질 : 반대       | = | 반대      |   |     |
| 어질 : 대해       | = | 대해      |   |     |
| 어질 : 보상을      | = | 보상      | + | 을   |
| 어질 : 받아내는     | = | 받아내     | + | 는   |
| 어질 : 작업       | = | 작업      | + | 을   |
| 어질 : 계속해      | = | 계속해     |   |     |
| 어질 : 나가겠다고    | = | 나       | + | 가   |
| 어질 : 밝혔       | = | 밝혔      | + | 습니다 |
| 어질 : .        | = | .       |   |     |
| 어질 : 이이우리가씨는  | = | 이이우리가씨  | + | 는   |
| 어질 : 어제       | = | 어제      |   |     |
| 어질 : 기자회견에서   | = | 기자 회견   | + | 에   |
| 어질 : 자신이      | = | 자신      | + | 이   |
| 어질 : 22년동안    | = | 22년동안   |   |     |
| 어질 : 일본       | = | 일본      |   |     |
| 어질 : 정부       | = | 정부      | + | 를   |
| 어질 : 상대로      | = | 상       | + | 대   |
| 어질 : 제기해온     | = | 제기해온    |   |     |
| 어질 : 3건의      | = | 3건      | + | 의   |
| 어질 : 소송이      | = | 소송      | + | 이   |
| 어질 : 완전히      | = | 완전히     |   |     |
| 어질 : 해결되지     | = | 해결되지    |   |     |
| 어질 : 않았다면서    | = | 않았다면서   | + | 서   |
| 어질 : 일본정부가    | = | 일본 정부   | + | 가   |
| 어질 : 2차대전중에   | = | 2차대전중   | + | 에   |
| 어질 : 일어난      | = | 일어난     |   |     |
| 어질 : 사건의      | = | 사건      | + | 의   |
| 어질 : 책임을      | = | 책임      | + | 을   |
| 어질 : 승거러      | = | 승거러     |   |     |
| 어질 : 하고       | = | 하고      |   |     |
| 어질 : 있다       | = | 있       | + | 다   |
| 어질 : 비난했습니다   | = | 비난했     | + | 습니다 |
| 어질 : .        | = | .       |   |     |
| 어질 : 이이우리가씨는  | = | 이이우리가씨  | + | 는   |
| 어질 : 본부상이     | = | 본부상     | + | 이   |
| 어질 : 약20만명이상이 | = | 약20만명이상 | + | 이   |
| 어질 : 사망한      | = | 사망한     |   |     |

|               |   |         |   |     |
|---------------|---|---------|---|-----|
| 어질 : 것으로      | = | 것       | + | 으로  |
| 어질 : 알려진      | = | 알려진     |   |     |
| 어질 : 1937년    | = | 1937년   |   |     |
| 어질 : 중국       | = | 중국      |   |     |
| 어질 : 남경대학살사건에 | = | 남경대학살사건 | + | 에   |
| 어질 : 대한       | = | 대한      |   |     |
| 어질 : 자신의      | = | 자신      | + | 의   |
| 어질 : 기술을      | = | 기술      | + | 을   |
| 어질 : 거부했으며    | = | 거부했     | + | 으며  |
| 어질 : 민주에서     | = | 민주      | + | 에   |
| 어질 : 전쟁포로들을   | = | 전쟁포로들   | + | 을   |
| 어질 : 생체실험에    | = | 생체실험    | + | 에   |
| 어질 : 이용한      | = | 이용한     |   |     |
| 어질 : 일본군의     | = | 일본      | + | 군   |
| 어질 : 새군부대인    | = | 새군부대인   |   |     |
| 어질 : 731부대에   | = | 731부대   | + | 에   |
| 어질 : 대한       | = | 대한      |   |     |
| 어질 : 언급도      | = | 언급      | + | 도   |
| 어질 : 배도록      | = | 배도록     | + | 도록  |
| 어질 : 요구했다고    | = | 요구했     | + | 다   |
| 어질 : 지적했      | = | 지적했     | + | 습니다 |
| 어질 : .        | = | .       |   |     |
| 어질 : 이이우리가씨는  | = | 이이우리가씨  | + | 는   |
| 어질 : 본부상이     | = | 본부상     | + | 이   |
| 어질 : 오랜       | = | 오       | + | 랜   |
| 어질 : 역사고고사흥   | = | 역사고     | + | 고   |
| 어질 : 사진       | = | 사진      | + | 시   |
| 어질 : 감철하는     | = | 감철하     | + | 는   |
| 어질 : 전향제도를    | = | 전향제     | + | 도   |
| 어질 : 완전       | = | 완전      |   |     |
| 어질 : 철폐하려고    | = | 철폐하     | + | 려고  |
| 어질 : 촉구했습니다   | = | 촉구했     | + | 습니다 |
| 어질 : .        | = | .       |   |     |

●● PHASE 2: ●●

|             |   |       |   |   |
|-------------|---|-------|---|---|
| 어질 : 나가겠다고  | = | 나가    | + | 고 |
| 어질 : 일본군의   | = | 일본군   | + | 의 |
| 어질 : 역사고고사흥 | = | 역사고고사 | + | 흥 |

●● PHASE 3: ●●

|             |   |     |   |     |
|-------------|---|-----|---|-----|
| 어질 : 계속해    | = | 계속하 |   |     |
| 어질 : 밝혔     | = | 밝히  | + | 습니다 |
| 어질 : 제기해온   | = | 제기하 |   |     |
| 어질 : 승거러    | = | 승거  |   |     |
| 어질 : 비난했    | = | 비난하 | + | 습니다 |
| 어질 : 알려진    | = | 알려진 |   |     |
| 어질 : 거부했으며  | = | 거부하 | + | 으며  |
| 어질 : 요구했다고  | = | 요구하 | + | 다   |
| 어질 : 지적했    | = | 지적하 | + | 습니다 |
| 어질 : 촉구했습니다 | = | 촉구하 | + | 습니다 |

●● PHASE 4: ●●

|            |   |      |   |   |
|------------|---|------|---|---|
| 어질 : 역사학자인 | = | 역사학자 |   |   |
| 어질 : 지명한   | = | 지명한  |   |   |
| 어질 : 저술한   | = | 저술하  |   |   |
| 어질 : 언급한   | = | 언급하  |   |   |
| 어질 : 계속해   | = | 계속하  |   |   |
| 어질 : 나가겠다고 | = | 나가   |   |   |
| 어질 : 일어난   | = | 일어나  |   |   |
| 어질 : 승거러   | = | 승거   |   |   |
| 어질 : 있다    | = | 있    | + | 다 |
| 어질 : 사망한   | = | 사망하  |   |   |
| 어질 : 알려진   | = | 알려지  |   |   |
| 어질 : 대한    | = | 대한   |   |   |
| 어질 : 이용한   | = | 이용하  |   |   |
| 어질 : 새군부대인 | = | 새군부대 |   |   |
| 어질 : 대한    | = | 대한   |   |   |

●● PHASE 5: ●●

|         |      |         |        |
|---------|------|---------|--------|
| 일본      | 지명   | 역사학자    | 이이우리가씨 |
| 자신      | 지술   | 역사고고사흥  | 2차대전때  |
| 일본      | 관측행위 | 언급      | 부분     |
| 일본정부    | 삭제   | 수정      | 보상     |
|         | 책임   | 나       | 밝히     |
| 어제      | 어제   | 기자 회견   | 자신     |
| 22년동안   | 일본   | 정부      | 상대     |
| 제기하     | 소송   | 일본 정부   | 2차대전중  |
| 일어나     | 해결   | 책임      | 승거     |
|         | 일어나  | 이이우리가씨  | 본부상    |
|         | 일어나  | 이이우리가씨  | 알려     |
| 약20만명이상 | 사망   | 남경대학살사건 |        |
| 1937년   | 중국   |         |        |
| 자신      | 기술   | 거부      | 민주     |
| 전쟁포로들   | 생체실험 | 이용      | 새군부대   |
| 731부대   |      | 언급      | 배      |
| 요구      | 지적   | 이이우리가씨  | 본부상    |
| 역사고고사   | 사진   | 감철      | 전향제도   |
| 철폐      | 촉구   |         |        |