

SGML 기반 비교 가능 코퍼스 구축

The Construction of the Comparable Corpus Based on SGML

이창열(Chang-Yeol LEE)*, 김용순(Yong-Soon KIM)**, 김성혁(Sung-Hyuk KIM)***

목 차

- | | |
|-----------------------------|--------------------|
| 1. 서 론 | 4. MLCC 자료 구조 |
| 2. 언어간 정보검색에서 비교 가능 코퍼스의 역할 | 5. 한국어 코퍼스 |
| 3. 사용되는 DTD | 5.1. 대상자료 |
| 3.1. TEI | 5.2. 신문기사 구조화 |
| 3.2. CES | 5.3. 신문기사 태깅 |
| 3.3. MLCC Newspaper DTD | 5.4. SGML기반 코퍼스 구축 |
| | 6. 결론 및 향후 연구과제 |

초 록

대규모 문헌 자료는 정보원으로써 다양한 목적에 사용될 수 있다. 언어간 정보검색에서 한 언어의 단어가 다의미적(polymorphic)일 경우 대상 언어로 정확한 번역을 위하여 언어간 대응 구조를 제공하는 다국어 코퍼스가 필요하다.

본 논문에서는 언어간 정보검색에서 대응 구조로 사용될 수 있는 비교 가능한 코퍼스를 구축하였다. 구축된 코퍼스(KFCM)는 유럽 6개 국어로 구축된 금융 관련 기사 자료인 MLCC 코퍼스에 대응되는 코퍼스로, 다양한 목적에 사용되며, MLCC 코퍼스로부터 독립적으로 구축한 SGML 기반 코퍼스이다. 본 논문에서는 KFCM의 응용과 코퍼스 구축 과정에서 발생하는 기술적 사항을 기술하였으며, 구축된 자료를 웹에 공개하였다.

ABSTRACT

The large scale documents of the data repository are utilized to the diverse applications. In the cross-language information retrieval, if the words of a query contain polymorphic meanings, the system needs multilingual corpus to exactly translate to the target words.

We constructed the financial comparable corpus, called KFCM(Korean Financial Corpus corresponding to MLCC Corpus), comparing to the MLCC Polylingual Documents which consisted with the 6 European languages. It is independently constructed under the DTD of MLCC comparable corpus, and can be utilized to the cross-language information retrieval. In this paper, we discussed about the application and construction procedures of KFCM which is public domain data.

키워드 : SGML, 코퍼스(Corpus), 언어간 정보검색(CLIR), 비교 가능 코퍼스(Comparable Corpus), 신문 기사

* 첨단학술정보센터 선임연구원

** 첨단학술정보센터 연구원

*** 첨단학술정보센터부장(숙명여자대학교 교수)

■ 논문접수일 : 1998년 10월 26일

1. 서 론

인터넷의 발달은 정보의 자유로운 이동을 허가하고 있으며, 이들 정보 자원에 대한 구축 방법론과 효과적인 검색 방법이 DL(Digital Library)에서 이슈화되고 있다. ERCIM(European Research Consortium for Informatics and Mathematics)의 DELOS Working Group과 DLI(Digital Library Initiative) 연합에서 연구되는 5개의 연구 분야(DLWG 98)를 포함하여 많은 연구가 인터넷에서 언어간 독립적이고, 문헌의 표준화를 지원하는 방향으로 연구되고 있다. LDC(Linguistic Data Consortium)나 ELRA(European Language Resources Association)는 다양한 문헌 정보를 소유하고 있으며, 표준화의 선도적 역할을 함으로써 DL 산업의 중요 정보원의 역할을 하고 있다. 이들 기관은 인터넷에서 다국어 처리의 테스트 베드와, 통계적 자료, 다국어 코퍼스, 사전을 포함하는 정보의 하부 구조 기능을 제공하고 있다.

언어간 정보 검색(CLIR : Cross-Language Information Retrieval)은 개별 언어와 문화에 대한 세계적인 공유를 제공하는 방향으로 발전하고 있다. 본 논문에서는 이러한 언어간 정보 검색을 위하여 필요한 하부 구조로서, 기 구축된 정보 저장소의 자료와 비교 가능한 한국어 비교 가능 코퍼스(Comparable Corpus)에 대하여 논의하기로 한다.

기존의 코퍼스가 단일 언어에 대한 정보의 구축으로 자원의 재창조(어휘 사전), 실제 언어의 통계적 자료 수집, 그리고 테스트 베드로 활용되고 있다면, 인터넷 상에서 언어간의 상호 연관성과 다국어 응용에 필요한 기능을 지원하는 다국어 코퍼스(Multilingual Corpus)의 등장은 언어

간에 다양한 응용을 지원하는 하부 구조로 활용되고 있다.

다국어 코퍼스는 다양한 소오스로부터 균형 있는(balanced) 자료를 통한 비교 가능 코퍼스나 병렬 코퍼스(Parallel Corpus)의 형태로 구현될 수 있다. 비교 가능 코퍼스는 유사한 시기, 유사한 주제, 같거나 유사한 DTD(Document Type Definition)를 가지고 여러 언어로 구축된 다국어 코퍼스이며, 병렬 코퍼스는 같은 내용, 같은 DTD를 가지고 여러 언어로 구축한 다국어 코퍼스이다. 병렬 코퍼스는 특정 문서에 기계 번역을 사용하여 여러 언어로 구축하는 방법이 보편적이다.

다국어 코퍼스는 다국적으로 활용하기 위하여 기존의 단일어 코퍼스에 비하여 표준화된 기능에 의한 구현, 재 사용성을 보장하는 기능, 코퍼스 처리와 마크에 통합된 플랫폼을 가지고 기술되어야 한다(Nan, 98).

ECI(European Corpus Initiative), ICAME(International Computer Archive of Modern and Medieval English), BNC(British National Corpus), TELRI(Trans-European Language Resources Infrastructure) 등 여러 기관에서는 다국어간 언어적, 문화적 비교를 위하여, 표준 DTD를 통한 다양한 비교 가능 및 병렬 코퍼스를 제공하고 있다.

문헌에 대한 일관성 있는 인코딩 스킴을 디자인하는 것은 쉬운 문제가 아니다. 우선 문헌의 특징과, 속성, 구조적, 논리적, 언어적 관계를 표현하는 기능, 다양한 주석 달기(annotation) 기능과, 증가적(incremental) 인코딩을 지원하는 구조를 가져야 하기 때문이다.

SGML (Standard Generalized Markup Language) (ISO 86) / XML (eXtensible

Markup Language)을 사용한 언어가 가지는 특징에 대한 세부적 기술 사항으로 EAGLES와 MULTTEXT는 다국어 언어 자원 기술에 적합한 CES(Corpus Encoding Standard)를 개발하였다. CES는 NEH(National Endowment for the Humanities)의 지원 하에 개발한 TEI(Text Encoding Initiative)의 사양을 따라 기술한 DTD이다.

본 논문에서 구축 대상으로 하는 코퍼스는 유럽 6개국어로 기술된 MLCC(Multilingual Corpora for Cooperation)의 비교 가능 코퍼스에 대응되는 한국어 금융 관련 기사 코퍼스다. MLCC에서 구축한 비교 가능 코퍼스는 SGML을 사용한 코퍼스 인코딩 표준 DTD인 CES를 기준으로 구축되었다. MLCC 금융 관련 비교 가능 코퍼스는 NSF(National Science Foundation)가 지원하는 Vassar/CNRS 공동체(VAS 98)와 유럽 프로젝트로 EAGLES, MULTTEXT (LRE), MULTTEXT-EAST(Copernicus)에 의한 국제 협력 프로그램인 LRE(Linguistic Research and Engineering) 작업(LRE61-101)의 일부로 구축한 6개국(네델란드, 영국, 프랑스, 독일, 이태리, 스페인)어로 구성된 다 언어 문헌(PDC: Polylingual Document Collection)이다. 현재 MLCC 자료는 1995년 완료되었으며 최종 보고서가 1997년 제출되었고, ELRA에서 학술 및 상업용으로 배포하고 있다.

본 논문에서는 MLCC 코퍼스에 대응하여 구축한 한국어 관련 코퍼스는 같은 시기, 같은 주제를 가진 자료로 1992년 한국 9개 신문사의 금융, 경제, 국제 관련 기사(article)를 대상으로 비교 가능 코퍼스를 구축하였으며 세부 과정은 (1)과 같다.

- 1) - 초기 자료에 대한 전처리,
 - 금융, 경제, 국제 관련 기사 추출,
 - 기사 구조에 대한 BNF(Backaus Normal Form) 정의,

MLCC의 Newspaper DTD에 대한 한국 기사 자료 Mapping 구조 정의

SGML 기반 코퍼스 구축

2. 언어간 정보검색에서 비교 가능 코퍼스의 역할

비교 가능 코퍼스는 다국어 처리 시 다양한 응용에 적합한 기초 자원으로 역할을 한다. 특히 CLIR에서 중의성을 내포하는 하나의 단어를 대응되는 타 언어로 번역할 때 잘못 번역이 가능하므로, 번역시 정확성을 향상시키기 위한 언어간 대응 구조를 제공한다(Aki 98).

〈그림 1〉에서는 일반적인 비교 가능 코퍼스를 사용한 다국어 정보 검색의 개념 구조를 나타내었다.

만약 “스키에 적합한 눈을 가진 산을 알려주십시오?”와 같은 한국어 질의어에 대하여 한국어 정보 검색 엔진을 통하면 질의어 전체에 적합한 한국어 문서를 추출할 수 있다. 그러나 질의어에 포함된 단어인 ‘눈’이라는 단어는 다국어(영어에 한정) 사전에서 ‘snow’에 해당되는 것인지 ‘eye’에 해당되는 것인지 알 수 없다. 그러므로 번역된 영어 질의 리스트에는 ‘snow’와 ‘eye’가 둘 다 포함되어 있다.

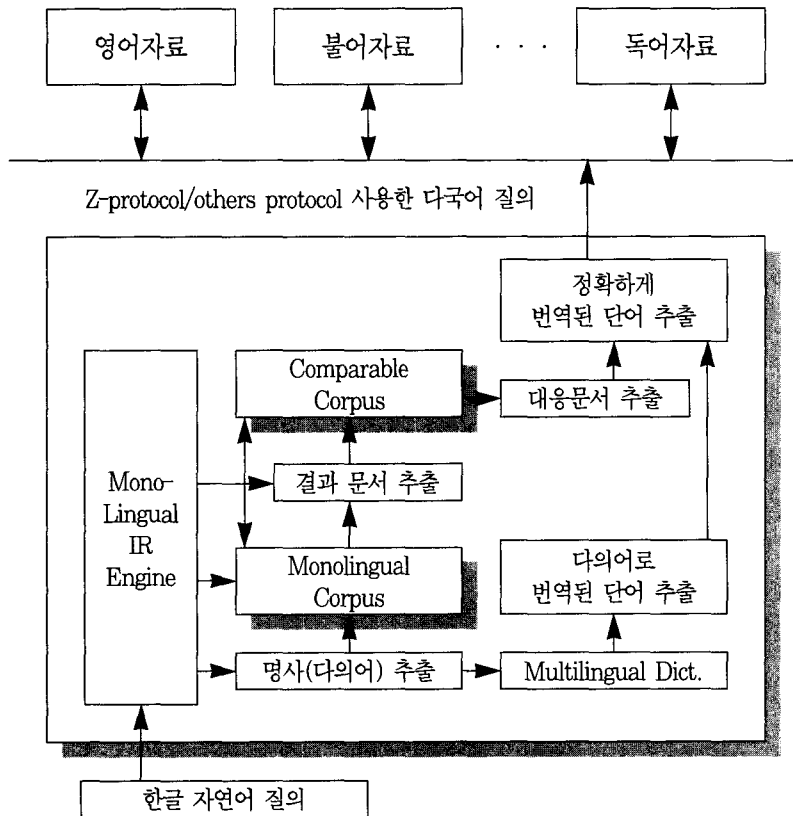
‘눈’이라는 단어가 ‘snow’에 해당되는 것인지 ‘eye’에 해당되는 것인지는 비교 가능 코퍼스를 통하여 결정할 수 있다. 즉 추출된 한국어 문헌에

대응하여 구축된 비교 가능 코퍼스는 'snow'가 'eye'보다 문헌의 구별성(discrimination)이 높은 키워드로 존재할 가능성이 높다. 다시 말하면, 추출된 한국어 문헌 내용이 'snow'에 관련된 문헌이기 때문에(한국어 IR 엔진이 정확하게 검색 시) 대응되는 비교 가능 코퍼스는 같거나 유사한 주제로 구축되어서, 그 문헌에서는 'snow'가 'eye'에 비하여 상대적인 문헌의 대표성을 가질 것이다. 그러므로 질의에서 '눈'은 비교 가능 코퍼스를 거치면 'snow'로 번역된다. <그림 1>에는 본 예제와 같이, 다국어 검색에서 발생하는 번역

의 오류를 비교 가능 코퍼스를 통하여 처리하는 과정을 기술하였다.

이와 같은 방법으로 정확하게 번역된 질의어는 인터넷을 통하여 각 목적 언어를 가진 시스템에 목적 언어로 질의를 가능하게 한다.

물론 목적 언어로 정확히 번역하기 위한 대응 구조로써, 비교 가능 코퍼스 이외에 다국어 시소러스, 병렬 코퍼스 등이 있으나, 비교 가능 코퍼스는 유사한 DTD와 같은 주제 분류 하에서 타 코퍼스와 독립적 구축 가능한 편리성을 가지고 있다.



<그림 1> 비교 가능 코퍼스를 사용한 CLIR 구조

3. 사용되는 DTD

일반적으로 코퍼스는 SGML을 사용하여 구축한다. SGML은 1986년에 ISO(International Organization for Standardization)에서 문헌 교환 및 출판에 위한 표준으로 제정되었고, 출판 분야와 매뉴얼 분야를 주축으로 이용되어왔으며, 최근 들어 디지털화 된 문서의 양이 급속도로 증가하면서 SGML은 문헌의 효율적 관리를 가능하게 하는 표준 포맷으로 부각되었다(윤소영 96; 이원석 98).

SGML은 구조를 기반으로 문헌의 형을 정의할 수 있는 메타언어(meta language)로 포맷팅과 문헌 구조를 분리하여 구조화된 문헌의 특성을 모두 살릴 수 있으며, 1990년대 CALS/EC 의 문헌 표준으로 제정되면서 여러 분야에 활용되기 시작하였다.

SGML을 이용한 문헌의 구조화 과정은, 구조화할 문헌에 대한 구조와 태그 명을 정의하는 DTD 개발, 그리고 이를 기반으로 하여 문헌을 태깅(Tagging)하는 단계로 이루어진다. SGML에 의하여 태깅 된 문헌은 구조화된 형태를 가지고 있어서 다양한 응용에 적절한 정보를 제공할 수 있는 응용 독립적인 체계를 가지는 장점이 있다.

MLCC 비교 가능 코퍼스는 SGML을 사용하여 정의한 CES DTD를 기반 구축되었다. CES는 TEI DTD와 가이드 라인에 따라 구축된 다국어 문헌(다국어 코퍼스, 사전)에 대한 적절한 마크업을 제공하는 표준 DTD이다. 그러므로 본 논문에서 구축할 비교 가능 코퍼스에서 사용하는 DTD는 TEI DTD, CES DTD, 그리고 신문 기사와 관련된 MLCC의 Newspaper DTD의 계층 구조를 가진다. 이들 각각 DTD에 대하여 간

략히 살펴보자.

3.1. TEI

TEI는 전자 문서 작성과 교환을 위한 가이드라인으로 1987년 뉴욕 Poughkeepsie에서 개최된 회의에서 제시한 몇 가지 원칙(PLO 87)에서 출발하여 1994년 5월 3번째 안(proposal)으로 TEI P3 DTD가 제안되어서 현재까지 사용되고 있다.

TEI는 기존에 전자 문서 작성 시 서로 상이한 인코딩 방식에 따른 낭비를 제거하고 다양한 유형의 문서에 적절한 코딩 집합을 기술하기 위한 전자문서기술 방식이다. 현재 제공되는 문서의 유형은 산문(prose), 운문(verse), 드라마(drama), 스피치(speech), 사전(dictionaries), 용어사전(terminology database), 텍스트 분석(text analysis), 텍스트 비평(text criticism) 등으로서, 약 400여 개 이상의 원소(elements)를 사용하여 다양한 유형의 문헌을 인코딩 할 수 있게 개발되었다.

TEI 구조는 TEI main DTD, TEI Lite, Bare Bone TEI로 구별될 수 있다. TEI main DTD는 전체 TEI 구조를 기술한 DTD이고, TEI Lite는 그 중에서 많이 사용되는 일부 DTD를 모아서 정의한 DTD이다. Bare Bone TEI는 HTML (HyperText Markup Language) (Dav 98)과 같은 간단한 구조를 표현한다.

3.2. CES

CES(CES 96)는 언어 자원에 대한 최소 인코딩 수준을 제공하는 표준 DTD다. TEI가 제공하는 코퍼스와 사전 관련 DTD에서 부족한 구조적,

의미적 기능에 대한 추가 기능으로 단일어, 다국어의 병렬, 비교 가능 코퍼스 등을 기술하는 스킴을 지원한다. CES는 마크업 최소화를 코퍼스의 중요도 정도에 따라 취급하며, 언어간의 병렬 배치에 관한 기능을 제공한다. 제공하는 마크업은 (2)와 같이 3단계로 정의한다.

- (2) - Level 1 : 전체 문헌 구조를 위한 마크업으로 절(paragraph) 단위로 마크업을 제공
- Level 2 : 하부 절(sub-paragraph) 단위 마크업과 인용부호를 포함한 모든 특수글자의 변환, 표현 (rendition) 정보 제공
- Level 3 : 세부 단위에 대한 마크업으로 축약, 숫자, 이름, 외래 단어 등에 대한 태깅 정보 제공.

CES는 원 자료에 증가적으로 마크업되면서 생기는 복잡성과 비 계층적 문제를 해결하기 위하여 원 문헌과 마크업을 분리하는 TIPSTER model을 채택하며 HyTime에 기반한 TEI를 사용하여 구현하고 있다. 예를 들어 (3)을 살펴보자.

(3) <p><s>According to the visiting leader,

the economy of the country is <q>"better than ever.</s> <s>It is in fact in very good shape.</s>"</q></p>

(3)에서 <p>와 <q>는 문장에 대한 구조적 분리인 반면에 <s>는 의미적 구분에 해당되어 계층적 구조에서 SGML 마크업이 잘못된 표현을 가진다. 그 이유는 SGML이 의미적 구조가 아닌, 구문적 계층 구조에 대한 마크업을 제공하기 때문이다. 이와 같은 의미적 구분에서 발생하는 비 계층적 문제 해결을 위하여 기본 문헌과 분리 문헌으로 구성된 remote markup 구조를 CES는 (4)와 같이 제공한다.

(4) 기본 문헌 :

<p id=p1>According to the visiting leader, the economy of the country is <q id=q1>"better than ever. It is in fact in very good shape."</q></p>

분리 문헌 :

<s from="ID(p1) STRLOC(1)" to="ID(q1) STRLOC(18)">
<s from="ID(q1) STRLOC(20)" to="ID(q1) STRLOC(52)">

한 언어로 구축된 문헌	대응되는 언어로 기술된 문헌
<x id=x1>A</x>	<y id=y1>B C</y>
<x id=x2>B</x>	<y id=y2>A</y>
<x id=x3>C</x>	<y id=y3>D</y>
<x id=x4>D E</x>	<y id=y4>E</y>
<x id=x5>F</x>	<y id=y5>G</y>

<그림 2> 병렬 코퍼스의 배치

```

<tei.2>
  <teiheader> ... </teiheader>
  <text id="ID" lang=en>
    <body>
      <div0 type=storylist>
        <div1 type=article n=N id=ARTICLE_ID> ... </div1>*
      </div0>
    </body>
  </text>
</tei.2>

```

〈그림 3〉 파일의 전체 구조

또한 병렬 코퍼스의 배치(alignment)시 언어 간 구조적 특징을 지원하는 대응 구조를 가진다. 예를 들어 〈그림 2〉와 같은 병렬 코퍼스는 (5)와 같이 CES를 사용하여 배치될 수 있다.

- (5) <link targets=' x1 y2' >
- <ptr id=p1 target=' x2 x3' >
- <link targets=' p1 y1' >
- <ptr id=p2 target=' y3 y4' >
- <link targets=' x4 p2' >
- <link targets=' x5' >
- <link targets=' y5' >

3.3. MLCC Newspaper DTD

MLCC 자료는 병렬 코퍼스와 비교 가능 코퍼스로 구성되어있다. 여기서는 비교 가능 코퍼스로서 6개국어로 된 금융 신문 또는 일반 신문의 금융 관련 기사로부터 추출한 자료인 Polylingual Document Collection(PDC)에 대하여만 고려한다. 〈그림 3〉에는 Newspaper DTD를 사용하여 기술한 전체적인 구조가 기술되었으며, 〈그림 4〉에는 The Financial Times지의 기사에 대한 샘플 구조가 Newspaper DTD에 따라 기술되었다. 각 기사에 대하여 〈그림 3〉의 'div1'이 추가되는 형태로 기술된다.

```

<div1 type=article n=N id=ARTICLE_ID><div2 type=articletext>
<head> Unesco aims at resources shift</head>
<opener>Publication <date>931027FT</date>Processed by FT<date>940714</date></opener>
<byline>By REUTERS<name type=place>Paris</name></byline>
<p> ...</p><p> ...</p> </div2>
<index> <list type=country><item>FR France, EC</item></list>
<bibl><publisher>The Financial Times</publisher><edition>International</edition>
<biblScope>Page 8</biblScope><extent>128</extent></bibl>
</div1>

```

〈그림 4〉 〈그림 3〉의 <div1>에 대한 자세한 기술

4. MLCC 자료 구조

PDC 자료에 대한 정보는 (6)에 기술되었다. 이들 자료의 소오스 자료는 잘 정의된 형태여서 인코딩에 대한 대응 구조는 간단하다. <그림 5>에는 샘플 자료로 Le Monde지의 원 자료 형태에 대하여 기술을 하였다.

(6) 네델란드어

- Het Financieel Dagblad
- 1992-1993
- 8천5백만 단어
영어
- The Financial Times 1993
- 3천만 단어
프랑스어
- Le Monde
- 1992-1993
- 1천만 단어
독어
- Handelsblatt
- 1986 - 1988
- 3천3백만 단어

이태리어

- Il Sole 24 Ore
- 1992-1993 - 1백9십만 단어
스페인어
- Expansion
- 1994
- 1천만 단어

원 자료에서 금융 관련 코드인 ECO, MDE, INI로 분류된 자료를 추출하면 관련 자 1천만 단어 분량이다. 이들 원 자료는 <그림 5>과 같이 SGML과 유사한 형태로 태깅 되어 있으며 Newspaper DTD에 따른 변환은 간단히 구축된 PERL 스크립트를 사용하여 변환하였다.

5. 한국어 코퍼스

5.1. 대상 자료

구축 대상 자료는 1992년 1월 1일부터 동년 12월 31일까지 발행한 국내 9개 신문 기사 내용이다. 대상 신문은 '경향신문(KHM)', '국민일보

원 자료 일부 :

```
<#FIELD NAME = ACC>224745</#FIELD>
<#FIELD NAME = DAT>920101</#FIELD>
<#FIELD NAME = DOC>BHL</#FIELD>
<#FIELD NAME = DOS>EXT,ECO</#FIELD> ...
```

변환 결과 :

```
<div1 type=article n=1 id=224745>
<opener><date>920101</date> DOC=BHL ... SEC=ECO ... </opener>
<index><list type=DOS><item>EXT</item><item>ECO</item></list> ..</index>...</div1>
```

<그림 5> Le Monde지의 마크업 자료의 변환 예제

〈표 1〉 구축 대상 원 자료의 크기

기 준	전처리후 자료	금융관련 자료
크기(size)	326Megabytes	194Megabytes
단어(works)	10,305,475	6,103,828
기사(articles)	219,801	105,329

(KME), '동아일보(DAE)', '서울신문(SUM)', '세계일보(SGM)', '조선일보(CSM)', '중앙일보(JAE)', '한국일보(HKM)', '한겨레신문(HRM)'이다. 이들 자료에 대한 크기는 〈표 1〉과 같다.

원 자료에 대한 전처리로써 일부 완성형 코드 단어의 ASCII 변환, 기사 작성 구조에 어긋나는 부분 수정을 하였으며, 이들 자료로부터 금융, 경제, 및 국제 관련 단어를 포함하는 자료를 추출하였다. 추출된 기사의 크기는 〈표 1〉에, 샘플은 〈그림 6〉에 각각 기술하였다. 〈그림 6〉의 첫라인에서 '***'는 기사의 시작을, 'CSM'은 조선일보를, '920823'은 1992년 8월 23일자 신문을, '07면'은 신문의 7번 페이지를, '135322'는 총 자료에서 135,322번째 자료를, '63652'는 금융 관련 자료의 63,652번째 자료를 나타낸다.

5.2. 신문 기사 구조화

〈그림 6〉의 자료를 일반화하여 BNF(Ronald 91)로 표현하면 〈그림 7〉과 같다. 신문 기사는 첫 라인의 정보와 제목, 부제목, 통신사, 내용, 리포터 순으로 기술되고 있으며, 부제목, 내용, 리포터는 반복되어 나타날 수 있다.

PDC를 정의한 Newspaper DTD의 구조에 따라 사용하는 헤더는 〈그림 8〉에 기술하였다. Newspaper DTD의 헤더 부분은 TEI 헤더에 비하여 간단하게 정의하고 있다. 즉 대부분의 자세한 정보는 'p'를 사용하여 기술하며, 〈그림 8〉에서 진하게 기술한 부분은 BNF 표기에서 정의된 항목에 대응 정보를 가지는 항목이다.

그렇지 않은 부분은 본 비교 가능 코퍼스의 모든 화일에서 동일하게 기술되는 부분이다.

*** CSM 920823 07면 135322 63652
 일 무기력장세 여파/ 세계증시 "구름긴 날씨" &
 @ 재할인을 인상도 영향 &
 한국 증시는 물론 세계 증시가 동반하락세를 보이고 있다.
 세계 경제가 서로 얽히고 설켜 국제화대가고 가국적 기업들은 주식을 해외 각국 증시에 동시 상장하는 경
 우도 많아 각국 증시의 시황은 점차 '동조화' 되는 경향을 보이고 있으며 특히 최근에는 일본 주가하락의
 영향을 동시에 받고 있다. &
 동경 증시에서 닷케이주가평균이 1만7천선이하로 내려간 지난 6월15일이후 22일 현재까지 4.34%의 하
 락률을 보이는 동안 한국도 18.68%의 하락률을 기록했다. &
 <이지훈기자> &

〈그림 6〉 샘플 기사

```

Newspaper_Article := *** Article_Info Article
Article_Info := Newspaper_name Newspaper_Date Newspaper_Page Serial_number_1 Serial_number_2
Article := Header Content
Header := Title Title_desc Column Subtitle Telecom_Name
Title := Str_Title
Title_desc := | / Str_Title_desc | Title_desc
Column := | (Str_Column)
Telecom_Name := | (Str_Telecom_Name)
Content := | Subtitle Rest | Content Subtitle Rest
Subtitle := | & @ Subtitle_list
Subtitle_list := Str_Subtitle_list | Subtitle_list / Str_Subtitle_list
Rest := Paragraph Reporter_name &
Paragraph := & Sentence | Paragraph & Sentence
Reporter_name := | <Str_Report_Name> | <Location = Str_Reporter_Name>
...
    
```

〈그림 7〉 신문 기사의 BNF 표기

5.3. 신문기사 태깅

기사 헤더 정보는 TEI 독립 헤더(Independent

Header)와 같이 필요시 헤더 부분을 파일로 부
터 분리하여 정보 검색에 활용할 수 있게 구현하
였다.

```

<teiheader>
  <filedesc><titleStmnt><title>조선일보(CSM)</title></titleStmnt>
  <editionStmnt><edition>버전1.0</edition></editionStmnt>
  <extent>60KiloBYTES</extent>
  <publicationStmnt><p>화일명 : csm0823.sgml</p> <p> 전자 화일 : 이창열(lcy@kric.ac.kr, 02-3460-5632)</p></publicationStmnt>
  <sourcedesc><p>조선일보 신문 기사</p><p>1992년 08월23</p></sourcedesc></filedesc>
  <encodingdesc><projectdesc><p>본 문서는 MLCC의 Polylingual Document Collections에 대응되는 comparable corpus 구축에 관한 것으로 한국어 버전이다.</p><p>한국어 비교 가능 코퍼스는 1992년 한국 9개 일간지의 금융관련 정보를 CES level-1수준에서 마크업을 하였다</p></projectdesc>
  <editorialdesc><p>원자료의 숫자, 영문자, 기호는 대부분 KSC5601로 표기되어 있어서 ASCII 코드로 변환하였다.</p><p>기사에 대한 통일된 구조를 기술하기 위하여 일부 구조에 대한 변경과, 조정작업이 있었다.</p></editorialdesc></encodingdesc>
  <profiledesc><language><language id=iso-639>kr</language></profiledesc>
  <revisiondesc><change><date>1998-10-16</date><respStmnt><name type=person>이창열(LEE, ChangYeol) </name>
  <resp>Constructor of KFCM(Korean Financial Corpus coresponding to MLCC Corpus)</resp></change></revisiondesc>
</teiheader>
    
```

〈그림 8〉 한국 신문 기사의 헤더 부분

각 신문 기사 내용에 대한 정보는 <그림 9>와 같이 사용할 수 있다. <그림 6>의 자료를 포함하는 파일 정보가 <그림 8>의 헤더로, 내용에 관한 정보가 <그림 9>에 각각 기술되었다. 본 예제(<그림 9>)에서 기술되지 않은 사항은 특히 기사 내용을 파악하여야 기술할 수 있는 부분으로(<list>를 사용하여 인코딩 되는 부분 중 일부), 내용에 대한 키워드, 내용에서 지칭하는 국가, 기관, 장소 등에 대한 정보 기술은 본 논문에서 제공할 수 없는 부분으로 기술하지 않았다.

<그림 9>에서 진하게 기술한 항목과 밑줄친 부분이 <그림 6>의 기사에서 추출한 부분이다. <div1>의 'n'은 파일에서 시리얼 번호를, 'id'는 금융 관련 시리얼 번호(<그림 7>의 신문기사에 대한 BNF에서 Serial_number_2 항목)를, <index> <list type=DOS>에서 <item>은 본 파일의 주제 분야를 나타낸다.

5.4. SGML 기반 코퍼스 구축

한국어 코퍼스를 인코딩하기 위하여 문장단위로 전처리 과정을 거친 후 CES의 Level 1으로 마크업을 한다. Level 1으로 마크업 된 자료를 Level 2나 3로 마크업을 하기 위하여는 반-자동적 작업(예를 들어 인용 부호는 자동, 표현은 수동, 품사 태깅은 반자동)을 필요로 하며, 응용에 따라 다를 수 있으므로 본 논문에서는 CES 마크업 (2)에서 기술한 Level 1 단계를 사용하여 구축하였다.

원 자료를 SGML 자료로 변환하는 구조는 <그림 6>의 내용을 <그림 8>과 <그림 9>로 바꾸는 작업으로 각 신문사의 기사 형태를 분석하여 적절한 SGML 태그를 삽입하기 위하여 파서(parser)를 사용하였다.

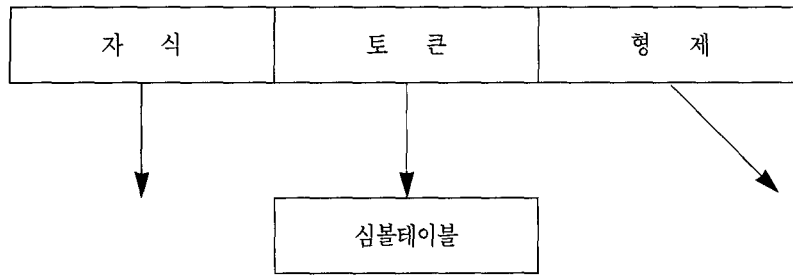
파싱(parsing)하는 과정은 입력된 기사의 형태

```

<text id="ID" lang=kr><body><div0 type=storylist>
<div1 type=article n=01 id=63652F>
  <div2 type=articletext>
    <head type=superheading>일 무기력장세 여파/ 세계증시 "구름긴 날씨</head>
    <head type=subheading>재할인율 인상도 영향</head>
    <opener>인쇄<date>920823kr</date></opener>
    <p>한국 증시는 물론 세계 증시가 동반하락세를 보이고 있다. 세계 경제가 서로 얽히고 설켜 국제화돼가고 가
    국적 기업들은 주식을 해외 각국 증시에 동시 상장하는 경우도 많아 각국 증시의 시황은 점차 '동조화' 되는
    경향을 보이고 있으며 특히 최근에는 일본 주가하락의 영향을 동시에 받고 있다.</p>
    <p>동경 증시에서 닛케이주가평균이 1만7천선이하로 내려간 지난 6월15일 이후 22일 현재까지 4.34%의 하락
    률을 보이는 동안 한국도 18.68%의 하락률을 기록했다.</p>
    <byline>기자<name type=person>이지훈</name></byline>
  </div2>
  <index> <list type=DOS> <item>EXT</item> <item>ECO</item></list> <list type=code> <item>I35322-
  63652</item></list></index>
  <bibl><publisher>조선일보사</publisher> <edition>조선일보 편집국 <bibscope> 07면</bibscoep> </bibl>
</div1>
</div0></body></text>

```

<그림 9> 한국 신문 기사의 내용 부분



〈그림 10〉 구문 트리 노드

에 대하여 각 기사의 단어, 어휘에 대한 어휘 분석 과정과 구문 분석으로 구성되어 있고 입력된 기사의 형태가 SGML로 태깅할 수 있는 정형화된 형태를 가지고 있다면 파싱 결과로 구문 트리 (abstract syntax tree)를 생성하고 틀린 경우는 삽입되어야 할 기사 형태에 대하여 부가 메시지를 출력하는 것으로 하였다.

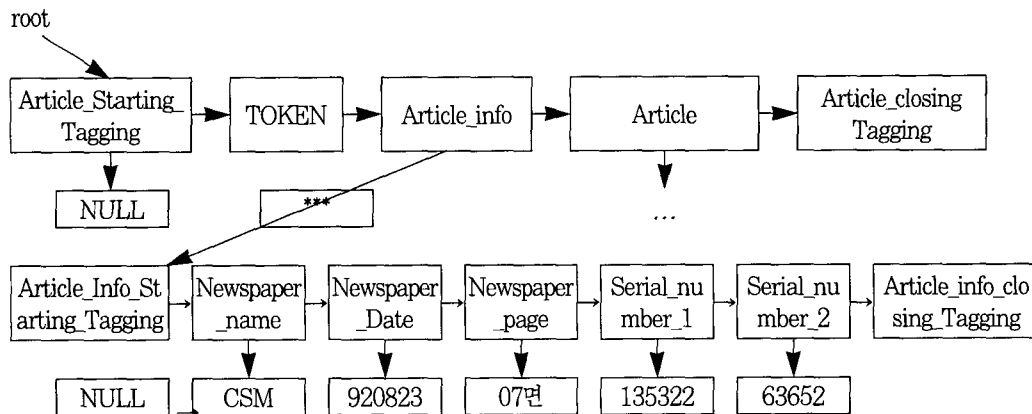
어휘 분석을 위한 과정에 Flex를, 구문 분석 과정에 Bison을 사용한 것은 기존의 lex와 yacc 보다 처리가 빠르기 때문이다(John 92). 신문 기사의 기사 날짜, 기자, 컬럼 작가, 신문사, 기사에 대한 본문 및 문단 등의 정보는 Flex를 이용한 어휘 분석기에 미리 정의하여 하나의 문장 중 최

소의 의미를 가지는 어휘로 정의하고, Bison을 이용하여 기사를 분석한다.

구문 트리의 기본 노드는 〈그림 10〉처럼 왼쪽 자손 오른쪽 형제(left child right sibling tree) 구조를 사용하였다. 어휘분석과 구문 분석의 결과가 옳으면 태깅 할 수 있는 정보와 함께 왼쪽 자손 오른쪽 형제구조의 노드를 사용하여 하나의 트리가 생성된다.

〈그림 10〉에서 실제 토큰으로 인식된 기사는 심볼 테이블이라는 문자열 배열 구조에 저장되어 있다. 예를 들어, 〈그림 6〉에 대한 일부 구문 트리 구조는 〈그림 11〉과 같다.

〈그림 11〉을 통하여 인식된 기사는 실행 트리



〈그림 11〉 〈그림 6〉에 대한 구문 트리의 일부

를 통하여 <그림 8>과 <그림 9>처럼 태깅된다.

6. 결론 및 향후 연구 과제

비교 가능 코퍼스는 분야별, 주제별, 아니면 세부 내용별로 유사한 내용의 자료를 독립적으로 구축한 자료로, 다국어 정보 검색에서 사용된다. 여기서 기술한 금융, 경제, 국제 관련 자료와 같이 계속적인 비교 가능 코퍼스가 구축되어야 CLIR에서 효과적으로 사용될 수 있다. 그러나 모든 주제에 대하여 비교 가능 코퍼스를 한군데서 제작할 수가 없음은 분명한 사실이므로 자료의 공유가 매우 중요하다.

비교 가능 코퍼스로 금융, 경제, 국제 관련 1992년 신문 기사를 선택한 것은 MLCC에서 6개국어로 된 같은 분야, 같은 시기 자료를 비교 가능 코퍼스로 구축하였기 때문이다. CES는 다국어 코퍼스 자료에 대한 적절한 구조적, 의미적

마크업을 제공하기 때문에 본 논문의 비교 가능 코퍼스에 매우 적절한 DTD로 사용될 수 있었다.

MLCC Newspaper DTD는 메타데이터 정보 및 기사 정보에 대한 최소 인코딩 스키를 제공하고 있다. 그러므로 본 논문의 코퍼스를 위하여 추가적 DTD를 정의하지 않았다. 본 논문에서는 한국 9개 신문 기사 파일에 대한 전처리 작업, 구조화, 그리고 Newspaper DTD에 대응되는 구조를 정의·구현하였다.

소오스 자료 및 SGML 자료를 공개(이창열 98)하여 앞으로 타 연구에 도움을 주고자 하며, 참고적으로 본 SGML 파일에 대응되는 MLCC 비교 가능 코퍼스(ELRA로부터 각자 구입)는 저작권 문제로 인하여 본인이 언급할 사항이 아님을 밝혀두는 바이다.

본 비교 가능 코퍼스와 더불어 추후 다국어 사전을 SGML을 사용하여 구축하여 언어간 정보 검색에 필요한 정보원으로써 계속적인 정보를 구축할 예정이다.

참 고 문 헌

- | | |
|--|--|
| <p>윤소영 96 "학술지 논문기사의 문헌구조 분석을 통한 DTD 개발", 숙명여자대학교 석사학위논문, 1996.</p> <p>이원석 98 "대량의 구조화 문헌 관리를 위한 SGML 저장 관리기의 설계 및 구현", 충남대학교 석사학위논문, 1998</p> <p>이창열 98 http://www.kric.ac.kr/~lcy/</p> <p>Aki 98 Akitoshi Okumura, Kai Ishikawa, Kenji Satoh, Translingual Information Retrieval by a Bilingual Dictionary and Comparable Corpus, LREC First</p> | <p>International Conference on Language Resources and Evaluation, p26-30, Granada, Spain May31-June 1, 1998</p> <p>BNC 98 http://info.ox.ac.uk/bnc/</p> <p>BNF 96 http://pllab.kaist.ac.kr/~jay/HyperSpec/Body/sec_1-4-1-2.html</p> <p>CALS 95 http://www.oasis-open.org/cover/tr9501.html</p> <p>Eric Severson, SGML Open Technical Research Paper 9501</p> <p>CES 96 http://www.cs.vassar.edu/CES/</p> |
|--|--|

- Dav 98 Dave Raggett, Arnaud Le Hors and Ian Jacobs, "HTML 4.0 Specification", World Wide Web Consortium Recommendation <http://www.w3.org/TR/REC-html40/>
- DELOS 98 <http://ntserv.iei.pi.cnr.it/DELOS/>
- DLI2 98 <http://www.nsf.gov/pubs/1998/nsf9863/nsf9863.html>
- DLWG 98 http://www.si.umich.edu/UMDL/EU_Grant/home.htm
- EAGLES 98 <http://www.lpl.univ-aix.fr/projects/eagles/TR/>
- ECI 98 <http://www.links2go.com/more/http://www.elsnet.org/resources/eciCorpus.html>
- ELRA 98 <http://www.icp.grenet.fr/ELRA/home.html>
- HyTime 98 <http://www.techno.com/TechnoTeacher/hytime.htm>, <http://ftp.sunet.se/ftp/pub/text-processing/sgml/SIGhyper/information>
- ICAME 98 <http://www.hd.uib.no/whatis.html>
- ISO 86 International Organization for Standardization. Information processing text and office system standard generalized markup language (SGML), 1986, ISO/IEC 8879, 1986
- John 92 John R. Revine et al., lex & yacc, O'Reilly & Associates, Inc
- LDC 98 <http://www.ldc.upenn.edu/>
- LRE 98 <http://www2.echo.lu/langeng/en/lre1/lre1.html>
- MLCC 97 Susan Armstrong, et al, "Multilingual Corpora for Cooperation Revised Final Report 3.
- MULTEXT 98 <http://www.lpl.univaix.fr/projects/multext/>, <http://www.lpl.univaix.fr/projects/multext-east/>
- Nan 98 Nancy Ide, Corpus Resources and Multi-Lingual Research: Considerations for the Future, LREC First International Conference on Language Resources and Evaluation, p8-10, Granada, Spain May 31-June 1, 1998
- NEH 98 <http://www.neh.gov/>
- NSF 98 <http://www.nsf.gov/>
- PLO 87 <http://www.tei.uic.edu/orgs/tei/info/pcp1.html>
- Ronald 91 Ronald Mak, Writing Compiler & Interpreters, John Wiley & Sons, Inc.
- TEI 98 <http://www-tei.uic.edu/orgs/tei/>
- TELRI 98 <http://www.idsmannheim.de/telri/telri.html>
- TIPSTER 98 <http://www.tipster.org/>
- Vassar 98 <http://www.cs.vassar.edu/~ide/research/>
- XML 98 <http://www.w3.org/TR/RECxml.html>

<부록> <http://www.kric.ac.kr/~lcy/index.html>에 있는 한국일보 1992년 7월 25일 파일 일부(hkm0725.sgml)

```

<tei.2>
<teiheader>
  <filedesc><titleStmnt><title>한국일보</title></titleStmnt>
  <editionStmnt><edition>버전 1.0</edition></editionStmnt>
  <extent>Marking 이전 화일 크기 : 55 KiloBytes</extent>
  <publicationStmnt><p>화일명 : hkm0725.sgml</p><p>전자 화일 : 이창열(lcy@kric.ac.kr; 02-3460-5632)</p>
  </publicationStmnt>
  <sourcedesc><p>한국일보 신문기사</p><p>1992년 07월 25일</p>
</sourcedesc></filedesc>
  <encodingdesc><projectdesc><p>본 문서는 MLCC의 Polylingual Document Collections에 대응되는 comparable corpus 구축에 관한 것으로 한국어 버전이다.</p>
  <p>한국어 비교 가능 코퍼스는 1992년 한국 9개 일간지의 금융관련 정보를 CES Level-1 수준에서 마크업하였다.</p></projectdesc>
  <editorialdesc><p>원자료의 숫자, 영문자, 기호는 대부분 KSC5601로 표기되어 있어서 ASCII 코드로 변환하였다.</p>
  <p>기사에 대한 통일된 구조를 기술하기 위하여 일부 구조에 대한 변경과, 조정 작업이 있었다.</p></editorialdesc></encodingdesc>
  <profiledesc><language><language id=iso-639>kr</language></profiledesc>
  <revisiondesc><change><date>1998-10-16</date><respStmnt><name type=person>이창열(LEE, ChangYeol)</name><resp>Constructor of KFCM(Korean Financial Corpus corresponding to MLCC Corpus)</resp></change></revisiondesc>
</teiheader>
<text id=ID lang=kr>
<body>
<div0 type=storylist>
<div1 type=article n = 1 id=55715F>
<div2 type=articletext>
<opener>인쇄<date>920725KR</date></opener>
<head type=superheading>'남포조사단' 내달초 파견/ 판문점 통해/ 민관합동 전문가로 구성 </head>
  <p>청 메시지</p>
  <p>노태우대통령은 24일 청와대에서 북한의 김달현 부총리겸 대외경제위원장을 접견, 남북경제협력의 전단계로 남포경공업단지 합작사업에 대한 타당성 조사단을 파견하겠다고 밝혔다.</p>
  <p>노 대통령은 이날 낮 김 부총리 일행과의 오찬에 앞서 김 부총리와 50분간 개별면담한 자리에서 노 대통령의 안부와 함께 남포경공업단지 사업에 대한 협조를 요청하는 김일성주석의 구두메시지를 전달받고 이같이 밝혔다.

```

두메시지를 전달받고 이같이 밝혔다.</p>

<p>노 대통령은 그러나 실질적인 남북경협과 관련, '상호 핵사찰문제가 해결되고 남북 기본합의서의 부속합의서가 타결돼 남북간 공동위원회들이 정식 발족돼야 남북 합작이나 투자사업이 전개될 수 있다는 것이 우리 정부의 기본 방침'이라고 천명했다.

& </p>

<p>노 대통령은 '다만 경제협력의 전 단계로서 우리측 전문가를 북쪽으로 파견해 남포경공업단지의 타당성을 검토하도록 하겠다'면서 구체적인 조사단 파견문제를 25일에 열릴 최각규부총리와 김 북한 부총리의 회담에서 논의토록 배석했던 최 부총리에게 지시했다.</p>

<p>김 부총리는 이날 김 북한 주석의 친서를 휴대하지는 않았으나 '북남 합의서가 발표된 시점에서 정무원 부총리가 남쪽을 방문한 것은 자못 의의가 크며 이번 방문이 성과리에 진행되기를 바란다'는 김 주석의 구두메시지를 노 대통령에게 전달했다.</p>

<p>노 대통령은 이에 대해 김 주석의 안부와 함께 실질적 경협에 앞서 핵문제의 타결을 바란다는 뜻을 김 주석에게 전해달라고 당부했다.</p>

<p>청와대 고위관계자는 노 대통령이 밝힌 남북경협 원칙과 관련, '핵문제 타결과 남북합의서에 따른 3개 분과위(정치, 군사, 교류·협력)에서 부속합의서가 타결돼 3개 공동위원회가 구성되면 교류·협력공동위에서 실질적인 경협문제를 다뤄야 한다'는 뜻'이라고 설명했다.</p>

<p>이날 노 대통령과 김 부총리의 면담에서 우리측에서 최 부총리와 이상연 안기부장 김종희 청와대 외교안보 수석, 북측에서 리성대 중국주재 무역참사와 김동국 정무원 책임지도원이 배석했다.</p>

<p>이어 열린 오찬에는 면담시의 배석자 외에 우리측에서 최영철 부총리겸 통일원장관과 청와대관계 수석비서관, 북측에서 정운업 무역부 삼천리총회사 총사장과 림태덕 대외경제협력 추진위 서기장 등이 참석했다.</p>

</div2>

<index><list type=DOS><item>EXT</item><item>ECO</item></list><list type=code><item>118160 - 55715</item></list></index>

<bibl><publisher>한국일보사</publisher><edition>한국일보 편집국</bibscope>01면</bibscope>

</div1>

<div1 type=article n = 2 id=55716F>

<div2 type=articletext>

<opener>인쇄<date>920725KR</date></opener>

<head type=superheading>최 부총리도 내달말 방북계획/ 오늘 부총리 회담서 구체 논의 </head>

<p>정부는 김달현 북한 부총리의 서울방문에 대응, 내달초 남포경공업단지 합작사업에 대한 타당성 조사단을 민관합동 관계전문가로 구성해 판문점을 거쳐 파견할 예정으로 24일 알려졌다.

특히 이번 조사단은 지난 3월 방북신청을 냈던 대우그룹 뿐 아니라 범재계 차원에서 다른 기업 관계자도 포함시키고 국토개발 전문연구기관이나 정부 투자기관의 관계전문가들도 망라, 민관합동 형태로 구성할 방침이다.</p>

<p>이와 함께 정부는 남포공단 조사단이 서울로 돌아온뒤인 내달 하순께 최각규부총리가 평양을 방문, 경협교류 진전에 대한 우리측 입장을 북측에 통보할 계획인 것으로 알려졌다.

최 부총리는 이번 방북기간중 노태우대통령의 메시지를 김일성 북한 주석에게 전하는 한편 남포를 비롯, 두만강유역의 나진·선봉, 해주 등 남북 합작사업 추진 후보지역을 대부분 둘러볼 예정인 것으로 전해졌다.</p>

<p>이와 관련, 최각규부총리는 25일 상오 김 북한 부총리 일행의 숙소인 힐튼호텔을 방문, 조사단 파견과 방북

일정 등 주요 경협교류 현안에 대한 우리측 입장을 전달한다.</p>

<p>정부의 고위 당국자는 '남포 조사단 파견은 김 북한 부총리가 이번에 제기한 시범사업을 수용한다는 차원은 아닌 만큼 당연히 대우그룹 뿐 아니라 대북 경협교류에 관심있는 대기업 및 업종대표가 골고루 참여하게 될 것' 이라고 밝혔다.</p>

<p>한편 정부는 내주중 전경련 등 경제단체가 중심이 돼 민간업체가 자율적으로 방북 조사단을 구성토록 하는 한편 토지개발공사, 국토개발연구원 등 정부투자 출연기관 전문가들도 조사단 일원으로 참여케 유도할 방침이다.</p>

</div2>

<index><list type=DOS><item>EXT</item><item>ECO</item></list><list type=code><item>118162 - 55716</item></list></index>

<bibl><publisher>한국일보사</publisher><edition>한국일보 편집국<bibscope>01면</bibscope>

</div1>

<div1 type=article n = 3 id=55717F>

<div2 type=articletext>

<opener>인쇄<date>920725KR</date></opener>

<head type=superheading>공영파트너로서의 남과 북(사설) </head>

<p>오늘썬서 일정을 끝내는 북한 김달현부총리 일행의 1주일간에 걸친 남한 방문의 성과와 의의는 한두가지가 아닐 것이다.

첫째는 말로만 들어왔던 남한의 경제발전 현황을 직접 목격한 것이고 다음으로 북한이 갈망하는 경제협력의 파트너로서는 한뼘줄인 남한이 가장 적합하다는 것이며 또 남한의 재계는 물론 대다수 국민들도 북한과의 협력을 환영내지 인정하고 있음을 알게 됐다는 것이다.

하지만 무엇보다 중요한 것은 어떠한 남북간의 경험도 해상호사찰과 기본합의서에 의한 각종 부속합의서 작성 등이 선행되어야만 가능하다는 남한당국과 국민들의 단호한 의지를 확인했다는 점일 것이다.

우리는 김 부총리가 이점을 깊이 새겨 귀환후 김일성주석 부자에게 정확히 건의해서 북한이 획기적 결단을 내리게 될 것을 기대한다.</p>

<p>이와 관련하여 노태우대통령이 지방산업 시찰후 청와대를 예방한 김 부총리에게 강조한 대목은 매우 뜻깊다 하겠다.

즉 탈이념 탈냉전의 화해정신을 바탕으로 무한경쟁시대에 접어든 오늘날 남북한이 적대감을 가진채 대치하는 것은 낭비와 손실만을 초래하는 만큼 진정한 화해와 협력으로 동반자시대를 열어 공존공영을 지향해야 하며 그러기 위해서는 하루빨리 상호 핵사찰로 남북이 평화의 의지와 자세를 확인해야 한다는 점이다.</p>

<p>사실 오늘과 같은 자국이의 위주의 치열한 경쟁상황에서는 외국과 합작 등 경험을 하고 싶어도 결코 쉽지가 않다.

합작사업의 전망이 확실하고 투자에 따른 이익 및 자본 회수의 보장 장치 등이 마련되어야 하는 것이다.

다시 말하면 투자대상국의 신의와 확실한 수용태세가 문제인 것이다.

이것은 북한 스스로도 80년대 중반 소위 합영법을 제정한 이후의 경험으로 잘 아는 내용인 것이다.</p>

<p>이런 점에서 북한이 '뒤통을 잇는 민족경제론'을 내세워 남한과의 경험의사를 적극적으로 표시하고 또 비록 최대요건인 상호 핵사찰이 이뤄지지 않았지만 당국이 남포경공업단지 건설의 타당성을 조사하기 위해 민관합동 조사단을 보내기로한 것은 7·7선언 정신에 따른 민족 동질성 확립과 공영정신의 일단으로 보아 환영한다.

따라서 북한도 이러한 취지에 호응해서 하루속히 해상호사찰로 평화의지와 개방정책을 내외에 선언함으로써 남한이 지원가

능한 모든 분야에서의 협력은 물론 철도 도로 항만 도로건설과 두만강 경제특구 추진 등에 있어 남북경협 시대를 열어야 할 것이다.</p></div>

<p>우리는 여기서 북한이 김 부총리의 남한 방문을 계기로 남북경협에 오해나 착각이 없도록 충고하고 싶다.</p></div>

즉 핵사찰을 지금까지와 같이 국제원자력기구(IAEA)의 제한적인 사찰로 대체하려는 것은 용인되지 않는다는 것과 만에 하나 대북 진출의 과당경쟁을 하는듯한 남한 업계의 움직임을 이용, 개별적인 거래를 추진하려는 생각은 버려야 한다는 것이다.</p></div>

또한 1천만 이산가족이 애를 태우며 고대하고 있는 8·15 노부모 상봉사업을 남북이 그토록 합의했음에도 정치문제와 연결시켜 불발위험을 하는 처사를 중단해야 할 것이다.</p></div>

<p>김 부총리가 평양으로 돌아간뒤 북한이 진정한 화해자세로 남북경협 시대를 스스로 열 것인지, 종래와 같이 적화도도를 감춘채 대남 위장선전과 적대 자세를 견지할 것인지 온겨레가 주시한다는 점을 깊이 인식해 주었으면 한다.</p></div>

<index><list type=DOS><item>EXT</item><item>ECO</item></list><list type=code><item>118166 - 55717</item></list></index></div>

<bibl><publisher>한국일보사</publisher><edition>한국일보 편집국</bibscope>02면</bibscope></div>

</div2></div>

<div1 type=article n = 4 id=55718F></div>

<div2 type=articletext></div>

<opener>인쇄<date>920725KR</date></opener></div>

<head type=superheading>책임질 일이 정말 없나(사설) </head></div>

<p>의혹이 다 풀리지 않았기에 수사는 계속돼야 옳다고 우리는 검찰의 정보사망 사기사건 수사결과 발표에 대해 논평한바 있다.</p></div>

그에 덧붙여서, 우리는 그 사건에 관련된 정부기관의 처리 자세와 태도에 대해, 특히 최세장 국방부장관의 기자회견 내용에 대해 공감할 수 없음을 다시 표시하고자 한다.</p></div>

<p>보도에 따르면 최 장관은 '30여년간 군에 몸담았던 사람이 군사시설을 대상으로 사회적 물의를 야기시켰다는 점에서 유감스럽게 생각하며 사과한다'고 밝혔다고 한다.</p></div>

그는 또 '검찰 수사발표대로 몇사람이 저지른 개인 사기사건이므로 군조직의 책임으로 다루기 힘들다'면서 '이번 사건에 대해 정치적 또는 도의적 책임을 느끼지 않는다'고 말한 것으로 전해졌다.</p></div>

바꿔 말해서 이 사건으로 인한 어떤 형태의 책임을 지는 일도 국방부에서는 없다는 뜻이다.</p></div>

<p>여기서 우리는 이번 땅사기사건에 대한 정부의 처리방식이 일반적으로 축소지향적이었음에 유의하고 있다. 그러나 군의 경우 다른 어느 집단보다도 강한 기율과 연대의식을 지닌 특수집단이라는 점에서, 이번 문제를 김영호 군무원의 개인행위로만 치부하고 지휘책임문제의 외면은 물론 아무런 문책도 없이 넘기겠다는 것은 누구도 공감할 수 없는 일임을 지적하지 않을 수 없다.</p></div>

<p>이번 사건이 6백몇십억원대의 6공 최대 금융부정이기도 하다는 점에서 금융계에서는 대대적인 문책인사로 수습노력을 기울이고 있음에 비추어서도, 국방책임자의 '개인행위'라든가 '유감'이라든가 하는 의견표명과 그 연사는 우선 성실하게 보이지 않았다는 점에서 유감스럽다.</p></div>

<p>성실, 정직의 의무, 품위유지의 의무, 직무유기나 근무지 이탈을 금하는 등의 군인복무 규율을 굳이 쳐들지 않고라도 일반 하급부대에서조차 영내 폭력이나 기타 사고예방을 위해 엄한 규제를 가하고 있음을 국민은 모두 알고 있다.</p></div>

혹시 단 1명이라도 사적인 이유로나마 근무이탈하는 경우를 막기위해 초급 지휘관들조차 온갖 힘을 다해 각종

사고방지에 진력하고 있지 않은가.

이따금 군의 유능한 지휘관들이 부하의 개인사고 때문에 지휘책임을 추궁당하는 사례도 알려져 있는 일이다.</p>

<p>우리는 이번 사건에서 국방부가 거창한 정치적 책임을 져야 한다고는 생각지 않는다.

그러나 사기범종의 상당수는 군용부지 현황을 정도이상 깊이 알고 있었으며 또한 군용부지 처분절차에도 근본적인 문제점이 있다는 사실이 드러난 만큼 그것은 내부사실의 누출경위와 그에 따른 제도적 보완책을 서둘러 밝히고 마련하는 것은 국민에 대한 기본적인 의무이며, 또한 도의적인 책임까지를 부인할 수 없는 일이다.</p>

<p>군용부지가 더이상 토지사기의 대상이 되지 않고 일개 군무원의 범법행위가 엄청난 사회적 물의까지 빚는 일이 다시 생기지 않는다는 것을 국민 모두가 믿을 수 있도록 국방부의 합리적이고 설득력있는 조치가 있어야 한다고 우리는 믿는다.</p>

</div2>

<index><list type=DOS><item>EXT</item><item>ECO</item></list><list type=code><item>118167 - 55718</item></list></index>

<bibl><publisher>한국일보사</publisher><edition>한국일보 편집국<bibscope>02면</bibscope>

</div1>

<div1 type=article n = 5 id=55719F>

<div2 type=articletext>

<opener>인쇄<date>920725KR</date></opener>

<head type=superheading>중기 공제기금 확대 등/ 민주, 종합대책안 발표/ 실태조사위 보고회 </head>

<p>민주당은 24일 국회에서 중소기업 실태 조사보고회를 갖고 중소기업 진성어음에 대한 할인의무화, 중소기업 공제기금의 대폭 확대, 지방 중소기업 전담은행의 설립, 중소기업 취업자에 대한 병역 특례제도 보완·확대 등을 골자로 하는 중소기업 종합대책을 확정 발표했다.</p>

<p>민주당은 이와 함께 일정 조건하에 중소기업의 외국인 취업허용, 중소기업 기술개발 특별기금 조성, 중소기업업청 설립, 농공단지 입주기업 특별지원책 수립 등도 중소기업 종합대책에 포함시켰다.</p>

<p>민주당은 지난 13일 '중소기업 실태조사단' (단장 장재식의원)을 구성, 소속의원 전원을 16개 반으로 나눠 전국의 3천여개 업체를 대상으로 실태조사를 한 결과를 취합해 이같이 결정했다.</p>

</div2>

<index><list type=DOS><item>EXT</item><item>ECO</item></list><list type=code><item>118175 - 55719</item></list></index>

<bibl><publisher>한국일보사</publisher><edition>한국일보 편집국<bibscope>02면</bibscope>

</div1>

<div1 type=article n = 6 id=55720F>

<div2 type=articletext>

<opener>인쇄<date>920725KR</date></opener>

<head type=superheading>주공 사장 김대영씨 내정/ 산업연 원장 황인정씨 내정 </head>

<p>정부는 24일 임기 만료된 주택공사 사장에 김대영 산업연구원 원장(전 건설부 차관)을 내정했다.

후임 산업연구원 원장에는 황인정 전 국제민간경제협의회(IPECK) 부회장이 내정됐다.</p>

</div2>

<index><list type=DOS><item>EXT</item><item>ECO</item></list><list type=code><item>118177 -

55720</item></list></index>

<bibl><publisher>한국일보사</publisher><edition>한국일보 편집국<bibscope>02면</bibsco

</div1>

...

</body>

</div0>

</tei.2>