

자동초록을 위한 지식기반시스템 설계에 관한 연구

A Study on the Knowledge-Based System for Automatic Abstracting

최인숙*

초 록

본 연구에서는 자연언어 텍스트 분석을 통해 초록을 생성하기 위하여 특정 주제분야의 지식에 입각하여 운영되는 지식기반시스템을 설계하였다. 주제영역은 상표법 관계 대법원 판결서를 대상으로 하였으며 지식베이스는 판결서의 내용을 담을 수 있는 틀과 상표법조문지식, 상품구분지식을 프레임과 의미네트웍으로 표현하여 구성하였다.

초록을 생성하는 과정은 (1) 텍스트를 이해하기 위하여 필요한 주제분야지식을 미리 체계적으로 조직하여 지식베이스를 구성하고 주요 정보가 들어갈 자리를 괄호로 둔 채 그 정보들을 연결하는 일반적인내용만 기술한 예비초록을 각 경우별로 준비한 다음 (2) 이러한 지식에 입각하여 입력테스트를 분석하고 초록에 필요한 주요 정보를 추출하며 (3) 입력테스트에 해당하는 예비초록을 호출하여 그의 괄호 안에 필요한 정보를 채워 초록을 완성하는 3 단계로 구성된다.

ABSTRACT

The objective of this study is to design an automatic abstracting system through the analysis of natural language texts. For this purpose a knowledge-based system operating on the basis of domain knowledge was developed.

The procedure of generating an abstract consists of three steps:

- (1) A knowledge-base containing domain knowledge necessary to understand a text is constructed using frame and semantic network structures, and preliminary abstracts are prepared for various cases.
 - (2) Input text is analysed on the basis of domain knowledge in order to extract information filling slots of the abstract with.
 - (3) A Preliminary abstract corresponding to the input text is called and filled with the information, completing the abstract.
-

I. 서 론

1. 연구의 목적 및 방법론

초록은 이용자의 정보 선별을 도와주고 정보를 배포하는 기능 외에도 정보검색 시스템에서 정보추적 및 탐색을 위한 키워드의 추출에 이용된다.¹⁾ 특히 2차정보 화일을 통해 원정보자료에 접근하는 서지정보검색시스템이나 본문검색시스템에서는 색인이나 초록 작성과 같은 주제분석작업이 매우 중요하다.

자동초록에 관한 종래의 연구는 대부분 문헌을 대표할 수 있는 중요 문장을 선정하여 원문헌에 출현한 순서대로 나열하는 방법을 쓰고 있는데 이는 초록이라기보다는 발췌문(extracts)의 성격이 강하고 문맥이 자연스럽게 못하여 이용하기에 불편한 점이 많다.

본 연구는 자연언어로 된 텍스트를 분석하여 발췌문이 아닌 순수한 초록을 생성해 내는 시스템을 설계하는 목적을 지니고 있으며 기존의 초록기법과 인공지능(artificial intelligence)의 여러 기법을 적용해서 특정 주제 분야의 지식에 입각하여 운영되는 지식기반시스템(knowledge-based system)을 개발함으로써 시스템의 확장이 용이하도록 하였다.

본 논문에서 초록을 생성하는 과정은 다음의 3단계로 구성된다.

(1) 텍스트를 이해하기 위하여 필요한 주제분야지식(domain knowledge)을 미리 체계적으로 조직하여 지식베이스(knowledge base)를 구성하고, 주요 정보가 들어갈 자리를 괄호로 눈 채 그 정보들을 연결하는 일반적인 내용만 기술한 예비초록을 각 경우별로 준비해 둔다.

(2) 이러한 지식에 의존하여 입력텍스트를 분석하고 초록에 포함될 주요 정보를 추출한다.

(3) 입력텍스트에 해당하는 예비초록을 호출하여 그의 괄호 안에 필요한 정보를 채워 완성된 초록을 생성한다.

본 시스템의 지식베이스 구성은 지식표현기법 중 프레임(frame)과 의미네트워크(semantic network)을 이용하여 표현하였다. 그리고 텍스트 분석시 초록에 포함될 정보가 나타나는 문장을 암시하는 단서어와 텍스트의 구조적인 특성 등을 이용하여 정보를 추출하였으며 생성된 초록에 문맥수정규칙을 적용하여 부자연스러운 문맥을 수정하였다.

본 연구의 실험을 위해 이용된 컴퓨터 시스템은 IBM-PC AT호환기종인 ACE이며 프로그래밍 언어는 Microsoft-C(버전 4.0)언어를 사용하였다.

2. 연구의 범위 및 제한점

본 연구는 상표법 관계 대법원 판결서²⁾를 대상으로 초록을 생성하는 시스템을 설계하였는데 판결서를 대상으로 한 이유는 텍스트가 일정한 양식에 따라 기술되므로 구조적 특성을 파악할 수 있고 판결의 결과를 관련 법조문에 의거하여 예측할 수 있다는 특성이 있어 본 연구에서 채택한 방법론으로 텍스트를 처리하기가 비교적 용이하기 때문이다.

본 연구에서는 상표법 제 8조, 제 9조, 제 24

1) C.Guinchart and M.Menou, 정보관리론, 사공철·김태수 공역, 서울: 구미무역, 1987, p.176.

2) 판결서는 판결을 표시하기 위하여 확정적으로 최초로 작성한 서면으로 판결원본과 같은 뜻으로 쓰이며 흔히 판결문이라 부른다.

조, 제 25 조, 제 26 조, 제 43 조, 제 45 조와 관계된 판결서를 처리할 수 있도록 주제분야지식을 구성하였다.

본 연구의 제한점은 시스템 구현시 다양한 경우를 모두 처리하지 못한 점과 컴퓨터시스템 환경상 한글처리 조건이 되어 있지 않아서 텍스트를 로마자로 번차 처리한 점이다.

II. 이론적 배경

1. 텍스트 분석과 자동초록

초록은 텍스트의 내용을 반영하므로 초록작성을 위해서는 우선 텍스트의 내용분석이 필요하다. 자동초록을 위한 텍스트 분석 방법은 초기에는 통계적 방법과 낮은 수준의 언어학적 방법이 주류를 이루었으나 점차 구문분과 의미론을 응용한 방법이 도입³⁾되었고 최근에는 자연언어를 이해하기 위해서는 단어와 문장의 구조에 대한 지식 뿐만 아니라 실세계에 관한 일반적인 지식까지 필요하다는 점이 지적되어 미리 조직된 지식베이스를 이용하는 지식기반시스템이 대두되고 있다.⁴⁾

대부분의 자동초록은 원문으로부터 중요하다고 판단되는 문장들을 그대로 또는 약간 수정하여 발췌한 것으로 초록이라기보다는 발췌문의 성격을 띠고 있다. 따라서 자동으로 작성된 초록은 발췌된 문장들을 순서적으로 나열하여 놓은 것으로 사람이 작성한 초록과는 달리 문장의 흐름이 부자연스럽고 문장 간의 연결에 상당한 무리가 있게 된다.⁵⁾

자동초록에서의 문장 선택 기준은 통계적 기준, 언어학적 기준, 문장구조적 기준의 세 가지로 구분할 수 있다.⁶⁾ 통계적 기준은 단어의 출현빈도를 이용한 것이며 언어학적 기준을 사용

한 것으로는 특정한 단어가 갖는 의미를 이용한 단서어기법이 대표적이다. 문장구조적 기준은 위치적 기준과 편집적 기준으로 세분되는데, 표제에 나타난 단어를 단서어로 삼는 표제어기법과 논문의 특별한 위치에 나타나는 문장을 선택하는 소재지기법은 위치적 기준에 속하고 이탤릭체나 고딕체로 씌여진 단어와 구두점을 이용하여 강조한 단어를 단서어로 선택하는 방법은 편집적 기준에 속한다.

최초로 자동초록을 시작하였던 룬(Luhn)⁷⁾은 입력테스트에서 기능어를 제외한 나머지 단어들의 출현빈도를 조사하여 빈도가 일정치 이상인 단어들만 내용을 대표할 수 있는 주요어로 보고 출현빈도가 높은 다수의 단어들이 서로 인접해서 나타나면 그 문장이 담고 있는 정보는 문헌의 내용을 대표할 수 있다고 보았다.

룬의 방법을 수정, 확대한 것으로 ACSI-Matic 연구⁸⁾가 있다. 이 연구에서는 문헌을 형

- 3) G. Salton, Dynamic Information and Library Processing, Englewood Cliffs; Prentice-Hall, 1975, p.106-109.
- 4) A. Barr and E.A. Feigenbaum, The Handbook of AI, vol. I, Los Alto; Kaufmann, 1981, p.225-226.
- 5) 정영미, 정보검색론, 서울; 정음사, 1987, p.174.
- 6) R. E. Wyllys, "Extracting and Abstracting by Computer," In: H. Borko ed. Automated Language Processing, New York: Wiley, 1968, p.127-179.
- 7) H. P. Luhn, "The Automatic Creation of Literature Abstracts," IBM J. of Research and Development, 2(2), 1958, p.159-165.
- 8) International Business Machines Corporation, Advanced Systems Development Division, ACSI-Matic Auto-Abstracting Project, Final Report, vol. I, III, Yorktown Heights, New York, 1960, 1961. (H. Borko and C. Bernier, Abstracting Concepts and Methods, New York; Academic Press, 1975, p.169-171에서 재인용)

태적인 특성에 따라 3 종류로 구분하여 각각 다른 언어 선정 방법을 적용하고 문장내 대표단어의 총수와 비대표단어의 패턴에 따라 문장의 접수를 계산한 다음 문헌의 길이에 따라 초록에 포함될 문장의 수를 계산하였다.

에드먼슨(Edmundson)과 윌리스(Wyllis)⁹⁾는 문의 절대빈도기법이 매우 일반적인 용어나 고도로 전문적인 용어에는 적용되지 않는다 하여 이를 해결하기 위한 상대빈도기법을 제안하였다.

박센데일(Baxendale)¹⁰⁾은 주제문장을 이용하여 문헌의 대표단어를 선정하는 소재지 기법을 최초로 제시하였다.

에드먼슨¹¹⁾은 문장 선정을 위해 단서어기법, 주요어기법, 표제어기법, 소재지기법을 제시하였는데 단서어기법과 표제어기법, 소재지기법을 통해 문맥상 추론(contextual inference)을 가능케 하였다.

그러나 이러한 방법을 통해 생성된 초록들은 문장과 문장 간에 사상의 흐름이 단절되고 연결이 자연스럽지 못한 경우가 종종 있고 선정된 문장이 주제문장이라 할지라도 전후 문장에서 논리가 충분히 드러나지 못하며 원문헌의 유용성에 관해 주관적 견해나 편집적 비평을 담지 못하는 단점이 있었다.¹²⁾

이런 단점은 구문론과 의미론을 고려하여 응집성(coherence) 기준을 채택한 연구들보다 소 완화되었다. 클리멘슨 등(Climenson, Hardwick, Jacobson)¹³⁾은 문헌에서 필수불가결한 내용을 추출하기 위해 구문분석적인 방법을 사용하였다.

얼(Earl)^{14) 15)}은 빈도와 통사론을 결합한 새로운 방법을 사용하였는데 문헌내의 명사구들이 포함하는 단어 중에서 기능어를 제외한

단어들의 빈도를 계산하여 고빈도 단어를 규명하고 이 고빈도 단어들을 3개 이상 포함하는 문장들을 선정하였다.

ADAM(Automatic Document Abstracting Method)시스템¹⁶⁾은 단어나 문장들 간의 문맥을 고려하여 문맥상 추론을 실현한 점 이외에 문장간 상호참조(intersentence reference)를 나타내는 단서어들(these, they, it, above, presented earlier, stated above 등)을 미리 소장하였다가 이들을 포함하는 문장을 선정할 때는 단서어들이 참조로 하는 선행문장도 포함시키는 방법과 동

- 9) H.P. Edmundson and R.E. Wyllis, "Automatic Abstracting and Indexing-Survey and Recommendations," Comm of the ACM, 4(5), 1961, p. 226~234.
- 10) P.B. Baxendale, "Machine-Made Index for Technical Literature-An Experiment," IBM J. of Research and Development, 2(4), 1958, p. 354~361.
- 11) H.P. Edmundson, "New Methods in Automatic Extracting," J. of ACM, 16(2), 1969, p. 265~285.
- 12) G. Salton, op. cit., p. 109.
- 13) W.D. Climenson, N.H. Hardwick and S.N. Jacobson, "Automatic Syntax Analysis in Machine Indexing and Abstracting," American Documentation 12(3), 1961, p. 178~183.
- 14) L.L. Earl, "The Resolution of Syntactic Ambiguity in Automatic Language Processing," Information Storage and Retrieval, 8, 1972, p. 277~308.
- 15) _____, "Use of Word Government in Resolving Syntactic and Semantic Ambiguities," Information Storage and Retrieval, 9, 1973, p. 639~664.
- 16) J.E. Rush, R. Salvador and A. Zamora, "Automatic Abstracting and Indexing. II. Production of Indicative Abstracts by Application of Contextual Inference and Syntactic Coherence Criteria," J. of ASIS, 22(4), 1971, p. 260~274.

일한 주요어가 인접한 문장에 나타나면 두 문장은 연관성이 있다는 전제하에 둘 다 선정하거나 둘 다 제거하여 사상의 흐름을 자연스럽게 하는 방법을 통해 구문적 응집성을 가능케 하였다. 특히 의미있는 문장을 선정하는 방법 뿐만 아니라 초록에 포함되지 않아야 할 문장을 규명하여 제거하는 규칙 및 절차를 개발한 점과 문장 선택시 아주 부분적이기는 하나 구문분석을 하여 문장에 수정을 가하였다는 점을 특성으로 들 수 있다. 폴록(Pollock)과 자모라(Zamora)¹⁷⁾는 단어가 자주 나타나면 그 정보량이 감소한다 하여 텍스트에서 천 단어 당 4번 이상 나타나는 용어는 가중치를 낮추는 방법으로 초록의 크기를 축소시켰다.

마티스 등(Mathis, Rush, Young)¹⁸⁾의 연구는 ADAM시스템에 본격적인 구문분석 방법을 도입한 것으로 인접한 문장들에서 반복되는 요소들을 병합하여 문장을 수정하거나 여러 문장 중의 구와 절을 한 문장 내에 결합하여 새로운 문장을 생성함으로써 발췌문이 아닌 순수한 초록이라 할 수 있는 결과를 얻었다.

그러나 이러한 방법들이 어느 상황에서나 일관성있게 적용되는 원칙이라 할 수는 없고 초록을 체계화하는 데 있어 그 길이나 문장의 수가 정해져 있으므로 관계없는 문장이 반복되어 포함되거나 정작 관계있는 문장이 누락되는 경우가 생기며 문장과 문장을 논리적으로 자연스럽게 연결하기가 어려워서 이용에 불편한 점이 많다.¹⁹⁾

스코로코호드코(Skorokhod'ko)²⁰⁾는 텍스트의 구조에 따라 초록의 방법론은 달라져야 한다고 주장하여 문장내 단어들의 의미적 관계에 근거한 문장들 간의 관계를 이용하여 텍스트를 그래프로 표현하였고 다수의 문장과 의미

적으로 연관되어 있어 그 문장을 생략하면 의미가 심히 손상되는 문장들을 선정하였다.

테일러(Taylor)와 크롤리(Krulle)²¹⁾는 격문법의 관계를 이용하여 텍스트를 의미네트웍으로 변환하여 가장 연결성이 높은 하부그래프를 규명하고 그래프 내에서 가장 영향력있는 개념들을 찾아냄으로써 초록으로서 자격이 있는 단일한 하부그래프를 추출하였으며 이를 다시 텍스트 형태로 변환하여 초록을 생성하였다.

파이스(Paice)²²⁾는 문장이 주제와 관련된 주요 요소를 포함하고 있음을 나타내는 지시자(The principal aim of this paper

17) J. J. Pollock and A. Zamora, "Automatic Abstracting Research at Chemical Abstracts Service," *J. of Chemical Information and Computer Sciences*, 15, 1975, p. 226 ~ 233.

18) B. A. Mathis, J. E. Rush and C. E. Young, "Improvement of Automatic Abstracts by the Use of Structural Analysis," *J. of ASIS*, 24(2), 1973, p. 101 ~ 109.

19) H. Borko and C. Bernier, Abstracting Concepts and Methods, New York : Academic Press, 1975, p. 161 ~ 186.

20) E. F. Skorokhod'ko, "Adaptive Method of Automatic Abstracting and Indexing," In : IFIP Congress 71, Amsterdam ; North-Holland, 1972, p. 1179 ~ 1182.

(R. N. Oddy S. E. Robertson, C. J. van Rijsbergen and P. W. Williams ed., Information Retrieval Research, London : Butterworth, 1981, p. 173에서 재인용)

21) S. L. Taylor and G. K. Krulle, "Experiments with an Automatic Abstracting System," In ; Proceedings of the ASIS Annual Meeting, Chicago, 14, 1977.

22) C. D. Paice, "The Automatic Generation of Literature Abstracts : An Approach Based on the Identification of Self-indicating Phrases," In : R. N. Oddy et al. ed., op. cit., p. 172 ~ 191.

is to investigate……, In the present paper, a method is described for……등)를 이용하여 구체적이고 화용론적인 처리를 하였다. 즉, 규칙을 이용하여 지시자를 규명한 후 단서어리스트를 만들고 이들을 포함하는 문장을 선정한 다음 기타 문장참조(exophoric reference)여부를 고려하여 지시된 문장에 인접하거나 선행하는 문장을 추가하는 방법으로 초록을 생성하였다.

리디 등(Liddy, Bonzi, Katzer, Oddy)²³⁾도 텍스트에서 한 대상을 처음 언급할 때는 완전한 표현을 하지만 다시 반복할 때는 수식어나 명사 자체를 사용하지 않고 대용어(anaphora)를 사용한다는 인식에 기초하여 대용어가 나타나는 것을 탐지하고 계수하여 그것이 가리키는 바가 무엇인지를 파악할 수 있는 알고리즘을 규칙으로 개발하였고 이를 자동초록 뿐만 아니라 자연언어이해, 질문응답, 질의어분석, 서지검색 등에 이용할 수 있다고 하였다.

일찌기 살톤(Salton)²⁴⁾은 변형문법을 이용하여 원문헌의 어휘와 문법적 상호관계를 그대로 유지하는 새로운 문장을 생성하는 말바꿈(paraphrasing)시스템을 통해서 내용이나 길이, 형태에서 수작업 초록과 견줄만한 초록을 기대할 수 있다고 하였다.

자동초록의 궁극적인 목표는 컴퓨터가 문헌을 읽고 그것에 대해 자연언어로 된 요약문을 쓸 수 있는 방법을 개발하는 것이다.²⁵⁾ 클라인(Klein)²⁶⁾의 연구에서 컴퓨터가 요약문을 작성하는 초보적인 시도를 볼 수 있다. 그는 일관성 있는 담화를 생성하기 위한 자동 말바꿈시스템을 시도하였는데 문장들을 작은 요소로 분해하고 필요시 더 큰 단위로 말바꿈하여 새로운 문장을 생성해 내었다.

최근의 인공지능에서 연구된 텍스트 이해시스템들은 텍스트 이해를 통해 말바꿈하거나 요약문을 제시한다. 이런 시스템들은 대부분 지식베이스에 포함된 상황정보를 이용하여 주제분야 텍스트를 분석하는 지식기반시스템으로서 입력문장 각각의 의미를 표현하는 작업 외에도 문장들 간의 연결을 원활히 하고 각 문장의 의미를 조합하여 텍스트의 총체적인 의미를 구성한다.

SAM(Script Applier Mechanism)²⁷⁾은 스크립트(script)²⁸⁾를 이용하여 이야기를 이해하는 시스템으로 말바꿈이나 요약문을 생성하는 기능 외에 질문에 응답하거나 다른 언어로 번역하는 기능도 갖고 있다. 스크립트가 등장인물들의 상대적 중요성을 규명해내면 요약기는 중요 등장인물들에 준거해서 중요도가 높

23) E.Liddy, S.Bonzi, J.Katzer and E.Oddy, "A Study of Discourse Anaphora in Scientific Abstracts," *J. of ASIS*, 38(4), 1987, p.225 ~ 261.

24) G.Salton, Automatic Information Organization and Retrieval, New York: McGraw-Hill, 1968, p.362 ~ 364.

25) R.E.Wyllis, *op.cit.*, p.176.

26) S.Klein, Automatic paraphrasing in essay format, Report No.SP-1600/001, Santa Monica: System Development Corporation, 1964 (G.Salton, Dynamic Information and Library Processing, p.109에서 재인용)

27) R.Cullingford, "SAM", In: R.C.Schank and C.K.Riesbeck ed., Inside Computer Understanding, Hillsdale: LEA Publishers 1981, p.75 ~ 119.

28) 스크립트는 특정한 문맥에서 전형적으로 발생하는 연속적인 사건을 기술하기 위한 구조로서 여러번 경험한 일이 완전히 굳어진 형태로 된 지식구조를 표현하기 위한 방식이다.(R.C.Schank and R.Abelson, Scripts Plans Goals and Understanding, New York: John Wiley & sons, 1977, p.36 ~ 68 참조)

은 개념화된 사실 (conceptualization)²⁹⁾ 들을 선정하여 그에 부착된 상황정보를 요약문에 포함시키고 상황변화를 설명하기 위한 매개적 사고들을 추가하여 요약문을 완성한다.

PAM(Plan Applier Mechanism)³⁰⁾은 스크립트와는 달리 양식화되지 않은 이야기를 처리하는 시스템으로 영어원문을 받아서 이를 개념의존 (conceptual dependency)³¹⁾ 표현으로 바꾼 뒤에 모든 등장인물들의 의도를 파악하기 위해서 목표(goal)³²⁾ 와 관(theme)³³⁾ 계획상자(planbox)³⁴⁾ 에 관한 지식을 이용하고 그들의 의도를 행위에 연관시킴으로써 이야기를 이해한다. 이 과정이 끝나면 PAM은 분석대상이 되는 이야기의 요약문을 생성할 수 있고 목표가 성취되는 방법에 관한 지식을 이용하여 문장과 문장 간을 연결하기 위한 추론을 하고 말바꿈과 질문에 대한 응답을 한다.

FRUMP(Fast Reading Understanding and Memory Program)³⁵⁾는 스케치 스크립트(sketch script)³⁶⁾ 라는 데이터 구조를 사용하여 실제계지식을 조직하고 신문 기사를 요약하는 시스템이다. 기사의 각 형태별로 찾기를 원하는 기대사실들(expected facts)의 리스트가 스케치 스크립트를 구성하고 있으며 텍스트 처리시 적절한 기대사실들을 선정하여 그의 스크립트 변수들을 텍스트에서 찾아낸 사항으로 대치시킨다. 모든 변수들을 다 만족시키면 이해작업이 완성되는 것이며 요약문은 기사의 형태별로 미리 만들어 둔 템플릿의 슬롯들을 스크립트 변수의 속성들로 대치시킴으로써 생성한다.

이 시스템은 짧고 간단한 이야기는 이해할 수 있지만 복잡한 이야기를 이해하는 능력은 없으며³⁷⁾, 현재로는 인공지능 시스템들은 극히

한정되고 잘 정의된 주제분야에서만 가능하다고 한다.³⁸⁾

2. 텍스트 분석을 위한 지식베이스 구성

언어이해를 위한 인공지능의 방법론은 2가지 부류로 발전되어 왔는데 첫째는 문장을 구문분석하여 각 단어마다 적합한 소성(feature)을 부여하는 프로그램을 개발하는 것이고 둘째는 텍스트 해석에 있어서 문맥이 압도적으로 중

29) 개념화된 사실이란 언어의 기저에 깔려있는 의미명제를 가리킨다.(R.C.Schank and R.Abelson, op.cit., p.12 참조).

30) R.Wilensky, "PAM", In : R.C.Schank and C.K.Riesbeck ed., op.cit., p.136 ~ 179.

31) 개념의존이론은 문장의 의미를 표현하는 이론으로 그 기본적인 원리는 의미가 동일한 두 문장은 어느 언어로 쓰여졌든지 관계없이 그에 대한 의미표현도 같아야 한다는 것이다. 따라서 문장을 개념의존이론을 사용하여 표기하기 위해서는 주어진 문장이 어떤 요소로 기술되었는지 그 과정은 무관하게 단어의 뜻을 구성할 수 있는 개념기본요소(conceptual primitive)가 사용된다. (R.C.Schank and R.Abelson, op.cit., p.11 ~ 17 참조).

32) 목표는 사람의 목적과 욕구의식에관계되는 지식 구조이다. (Ibid., p.101 ~ 130 참조).

33) 관은 사람에 따라 왜 다른 목표가 생기는가에 대한 배경지식으로 각 행동에 대한 근거를 제공해주는 지식구조이다.(Ibid., p.131 ~ 149 참조).

34) 계획상자는 계획의 각 단계를 실현하는 방법을 가리킨다. (Ibid., p.88 ~ 97 참조).

35) C.F.DeJong, " Prediction and Substantiation : A New Approach to Natural Language Processing," *Cognitive Science*, 3, 1979, p. 251 ~ 273.

36) R.C.Schank and R.Abelson, op.cit., p.176.

37) T.E.Doszkoacs, " Natural Language Processing in Information Retrieval," *J. of ASIS*, 37(4), 1986, p.191 ~ 196.

요하다는 인식하에 실세계지식을 조직하여 문장의 의미추출에 이용하는 것이다.³⁹⁾ 문법적 지식만으로 문장의 구조는 어느 정도 알 수 있겠지만 그 의미는 파악할 수 없다. 따라서 문장을 이해하는 데 필요한 배경지식 즉, 문장 내에 언급되는 않았으나 그 문장의 필자와 독자가 익히 알고 있는 기본적인 공통지식이 요구된다.⁴⁰⁾

일반적으로 시스템의 문제해결 능력은 단지 그것이 채택하고 있는 형식(formalism)과 추론체계(inference scheme)에만 기인하는 것이 아니라 시스템이 보유하고 있는 지식에 기인하므로 시스템을 지적으로 만들기 위해서는 문제분야에 관한 양질의 특정한 지식을 많이 제공해 주어야 한다.⁴¹⁾ 이러한 맥락에서 전문가시스템 혹은 지식기반시스템이 개발되기 시작했으며 이 시스템들은 모두 문제를 해결하기 위한 특정 전문분야의 사실적 지식(factual knowledge)을 표현하고 응용하도록 고안되어 있다.

지식기반시스템은 문제분야에 관한 특정한 지식 즉, 주제분야지식을 갖는 시스템⁴²⁾이라 정의할 수 있고 지식을 수집하고 표현하며 이 지식을 적절히 이용하여 일련의 논리를 구축하고 설명하는 기술적 방법들이 지식기반시스템의 설계시 중요한 점이라 할 수 있다.

최근의 자연언어 연구에서는 언어를 이해한다는 것은 대개 메모리기반 작업이라는 관념이 널리 퍼져있다.⁴³⁾ 즉, 구문적 분석은 문장에서 용납되는 구조를 인식해내는 것이고 의미분석은 특정 대상물이 지칭하는 바를 인식하는 것이라 할 수 있으며 텍스트를 이해하는 일은 정형화된 기술문과 현상학을 나타내는 기술문을 비교하여 유사성과 차이점을 인식함으로써 이루어진다는 것이다. 따라서 자연언어처리 분야

에서도 1970년대에 들어오면서 적용분야에 대한 광범위한 지식을 저장하고 이 지식에 입각하여 자연언어를 이해하는 데 도움을 주는 지식기반시스템들이 많이 고안되었으며⁴⁵⁾이런 시스템들은 앞으로도 증가될 것이다.⁴⁶⁾

주제분야지식은 시스템의 다른 지식들 즉, 문제를 어떻게 해결할 것인가에 대한 일반적인 지식이나 이용자와 어떻게 소통할 것인가에 관한 지식 등과는 분리되어 조직된다. 이 주제분야지식의 집합체는 지식베이스라고 불리우며 기존의 데이터베이스와 근본적으로 다른 점은 지식베이스는 확실적인 내용의 방대한 화일이 아니라 지식이 될만한 사실이나 추론, 과정의 집합이므로 주어진 지식으로부터 새로운 지식을 만들 수 있다는 점이다.⁴⁷⁾

지식을 조직하는 방법은 의미네트워크, 프레임, 스키마(schema), 생성규칙(production

39) R.C.Schank and R.Abelson, op.cit., p.9.

40) 최기선, "자연언어이해," 정보과학회지, 4(3), 1986, p.21~30.

41) D.A.Waterman, A Guide to Expert Systems, Massachusetts: Addison-Wesley, 1986, p.3~4.

42) Ibid., p.18.

43) E.A.Feigenbaum, "The Art of Artificial Intelligence," In: Proceedings of the 5th International Joint Conference on AI, 1977, p.1014~1029.

44) H.Tennant, Natural Language Processing, New York/Princeton: PBI, 1981, p.220~221.

45) A.Barr and E.A.Feigenbaum, op.cit., p.226~229.

46) R.C.Schank, E.Charniak, Y.Wilks, T.Winograd and W.Woods, "Panel on Natural Language Processing," In: Proceedings of the 5th International Joint Conference on AI, 1977, p.1007~1013.

47) A.Barr and E.A.Feigenbaum, op.cit., p.34~35.

rule), 절차적(procedural) 표현방식 등이 널리 쓰이고 있다. 일반적으로 어느 대상을 표현하기 위해서는 그 대상에 맞는 표현방법이 있다. 지식을 표현하는 데 있어 가장 중요한 점은 주제와 관련된 지식들이 그 의미를 잘 나타내면서 또 신속히 검색될 수 있도록 표현, 조직하는 것이라 할 수 있다.⁴⁸⁾

의미네트웍은 기본적으로 의미점(node)과 연결호(arc)로 구성되는데, 의미점은 사물, 행위, 사건, 기능, 사물의 속성 등의 개념을 나타내고 연결호는 이러한 개념 간의 계층적이거나 문맥적인 관계 또는 개념과 관련된 정보를 나타낸다.⁴⁹⁾

프레임은 거실에 있을 때나 어린이의 생일 파티에 갔을 때 등 정형화된(stereotyped) 상황을 표현하는 데이터 구조로서 각 프레임에는 그 프레임을 어떻게 사용할 것인가에 대한 정보나 다음에 발생할 상황의 예측에 대한 정보, 그리고 이런 예측이 확인되지 않을 경우 어떻게 조치할 것인가 등 여러가지 정보가 부착되어 있다. 다시 말해서 프레임은 한 대상에 대한 의미속성이나 그 대상에 대한 속성을 알아내기 위한 규칙들이 한 틀에 뭉쳐져 있는 것이다.⁵⁰⁾ 프레임에서는 슬롯(slot)이라 불리는 속성들 즉, 프레임이 나타내고자 하는 개념 기술의 한 측면인 슬롯 1, 슬롯 2, 슬롯 3 등이 모여서 한 개념을 표현하고 개념을 표현하는 각각의 의미점들은 그들 간의 관계와 함께 계층구조로 조직되어 최상위의 의미점은 일반적인 개념을 나타내고 하위 의미점들은 좀 더 특정한 실례들을 나타낸다. 슬롯은 값(value) 디폴트 값(default value), 절차(procedure) 등 여러가지 패싯(facet)을 가질 수 있다.

텍스트 분석을 위해 사건과 배경에 관한 지식을 미리 표현하고 이 구조를 이용하는 시스템으로는 1절에서 설명한 SAM, PAM, FRUMP 외에 MARGIE⁵¹⁾, IPP⁵²⁾와 같은 언어이해시스템이 있다.

MARGIE (Memory, Analysis, Response Generation and Inference on English)는 분석기와 추론기, 생성기로 구성되는데 입력문장을 개념적 구조로 변환하면 추론기는 실세계지식을 기반으로 특정한 형식에 맞추어 저장하고 추론을 하게된다.

IPP(Integrated Partial Parser)는 단어사전과 미리 축적된 지식을 이용하여 테러리즘과 연관된 뉴스기사를 이해하는 시스템이다. 이 시스템은 단어를 유형별로 조직하여 통사론적 분류와는 무관하게 처리방법이 결정될 수 있도록 했고 대부분의 처리지식을 생성규칙으로 표현되는 요구조건(request)의 형태로 구

48) A. Barr and E.A. Feigenbaum, op.cit., p.143~152.

49) G. Salton and M. J. McGill, Introduction to Modern Information Retrieval, New York: McGraw-Hill, 1983, p.265.

50) M. Minsky, "A Framework for Representing Knowledge," In: P.H. Winston ed., The Psychology of Computer Vision, New York: McGraw Hill, 1975, p.211~277.

51) R.C. Schank, N. Goldman, C.J. Rieger and C. Riesbeck, "MARGIE: Memory, Analysis, Response Generation and Inference on English," In: Proceedings of the 3th International Joint Conference on AI, 1973, p.255~261.

52) R.C. Schank, M. Lebowitz and L. Birnbaum, "Parsing Directly into Knowledge Structure," In: Proceedings of the 6th International Joint Conference on AI, 1979, p.772~777.

현하여 요구조건의 조건부분이 만족되면 그에 상응하는 행위가 실행되도록 하였다.

언어이해시스템 외에도 자동색인시스템에 사전지식을 응용한 사례를 볼 수 있다. 험프리(Humphrey)와 밀러(Miller)⁵³⁾는 프레임 기반 지식표현 언어를 사용하여 MEDLINE 데이터베이스의 의학 관계 정기간행물의 색인작업을 보조하기 위한 지식기반시스템을 개발하였다.

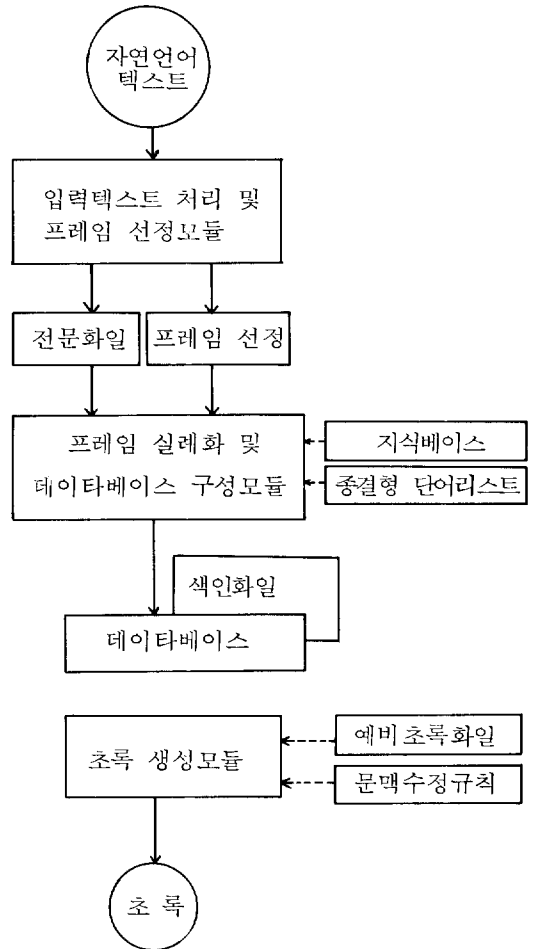
블레두츠-스토콜로프(Vleduts-Stokolov)⁵⁴⁾는 BIOSIS에서 주제색인을 위한 자동보조장치로서 생명과학분야 문헌의 표제에서 600여개의 생물학적 개념을 자동으로 인식해내는 자연언어 텍스트 처리시스템을 고안하였다. 이 시스템은 생물학 문헌에 빈번히 사용되는 15,000여종의 용어와 그에 대응하는 개념기본요소, 애매성 제거규칙, 색인정책에 관한 규칙 등의 방대한 지식을 조직하여 자연언어로 된 논문의 표제를 개념기본요소라는 매개적 표현을 거쳐 그에 상응하는 개념 표목(concept heading)들로 변환하는 데 이용하였다.

Ⅲ. 시스템 설계

1. 시스템 개요

본 연구에서 개발한 시스템은 자연언어로 된 텍스트를 입력한 후 분석하여 초록을 생성하는 시스템으로, 프로그램과 지식베이스, 종결형 단어리스트, 예비초록화일, 문맥수정규칙 등을 이용하여 설계하였다. 본 시스템은 입력텍스트 처리 및 프레임 선정모듈, 프레임 실행화 및 데이터베이스 구성모듈 그리고 초록 생성모듈의 세 부분으로 구성된다. 시스템의 전반적인 구성도는 <그림-1>과 같다.

세개의 모듈은 서로 연관성을 가지며 운영된



<그림-1> 시스템 구성도

다. 입력텍스트 처리 및 프레임 선정모듈은 입력된 텍스트에서 빈 칸을 단서로 토른을 분리

53) S.M.Humphrey and N.E.Miller, "Knowledge-Based Indexing of the Medical Literature : The Indexing Aid Project," *J. of ASIS*, 38(3) 1987, p.184~196.

54) N.Vleduts-Stokolov, "Concept Recognition in an Automatic Text-Processing System for the Life Science," *J. of ASIS*, 38(4), 1987, p.269~287.

한 뒤 순서대로 번호를 부여하여 임시기억장소에 전문(全文)화일을 만든 다음 텍스트를 어떤 프레임에 맞추어 처리해야 할 지를 선정한다. 프레임 실례화 및 데이터베이스 구성모듈은 앞의 모듈에서 선정된 프레임의 슬릿들을 지식베이스에서 탐색하여 슬릿값을 채우기 위한 패킷을 처리한 후 슬릿명—슬릿값 쌍들을 모아 데이터베이스를 구성하고 색인화일을 만든다. 초록 생성모듈은 예비초록화일에서 해당 프레임의 예비초록을 불러와 앞의 모듈에서 찾아낸 슬릿값들로 예비초록의 괄호 안을 채워 완성된 초록을 생성하고 데이터베이스에 저장한다.

본 시스템은 텍스트를 이해하기 위해 필요한 주제분야지식을 프레임과 의미네트웍으로 표현하였으며 이러한 지식에 의존하여 텍스트를 분석하고 초록을 생성하였기 때문에 일종의 지식기반시스템이라 할 수 있는데, 종래의 자동초록에서 문장 선택의 기준이었던 통계적 기준, 언어학적 기준, 문장구조적 기준 중에서 언어학적 기준과 문장구조적 기준을 응용하였고 인공지능분야의 텍스트 처리 기법 중 주제분야지식의 언어학적, 텍스트 구조적 특성을 이용하여 지식베이스를 조직하는 기법과 형태소 분석 기법을 채택하였다. 따라서 본 시스템만이 갖는 유일한 특성을 들자면, 텍스트 처리를 판결서에 적용할 수 있도록 판결서의 언어학적, 텍스트 구조적 특성을 분석하여 필요한 정보를 찾아내는 알고리즘을 개발하였기 때문에 복잡한 구문분석을 하지 않고도 자동초록을 만들 수 있다는 점을 들 수 있겠다.

2. 지식베이스 구성

지식베이스를 조직하는 일은 지식기반시스템에서 가장 중요한 작업이라 할 수 있다. 상표

법 관련 판결서를 대상으로 하는 본 시스템의 지식베이스를 구성하는데 있어 가장 중요한 점은 판결서 처리에 필요한 지식을 효율적으로 조직하는 방법과 또 꺼내어서 사용하고 제어하는 방법론이라 할 수 있다. 따라서 판결서의 내용을 담을 수 있는 틀을 프레임으로, 상표법조문과 상품구분지식을 의미네트웍으로 조직하였고 사용과 제어의 방법에 대한 정보는 슬릿의 각 패킷에 담아두었다.

2.1 「판결서」 프레임의 조직

본 연구에서는 상표법 제 8 조, 제 9 조, 제 24 조, 제 25 조, 제 26 조, 제 43 조, 제 45 조와 관계된 판결서를 처리할 수 있는 시스템을 설계하기로 하였다. 상표법 제 8 조는 상표등록의 요건을 기술한 조문이고 제 9 조는 등록을 받을 수 없는 상표를, 제 24 조는 타인의 의장권 등과의 관계를, 제 25 조는 상표등록의 범위를, 제 26 조는 상표권의 효력이 미치지 아니하는 범위를, 제 43 조는 심판의 청구를, 제 45 조는 상표등록의 취소사유를 각각 기술한 것으로 관계되는 판결서를 분석한 결과 5 범주로 나눌 수 있었다.

각 범주의 판결서들은 참조조문에 의거하여 볼 때 그 내용을 예측할 수 있었으므로 각 범주별로 예상되는 정보의 종류와 그 내용을 알아내기 위한 절차들을 프레임구조로 표현하였다. 프레임을 사용한 이유는 처리할 주제분야의 텍스트에서 일어날 상황을 미리 예측할 수 있으므로 미리 지식베이스를 만들어 둬으로써 처리를 용이하게 할 수 있다는 장점이 있고 주제분야 지식과 처리 결과를 계층적 구조로 조직할 수 있으므로 지식의 소장과 탐색에 있어 효율성이 있기 때문이다. 뿐만 아니라 절차를 부착할 수 있으므로 텍스트 처리에 아주 효과적으로 사용할 수 있다. 즉, 하나의 프레임은 여러 개의 슬

릿으로 구성되며 각 슬릿은 한 개 이상의 패킷을 갖는데 각 패킷은 슬릿을 채우는 데 필요한 정보를 담고 있다. 패킷의 종류는 값, 텍스트 분석 패킷이라 불리는 특수목적의 절차, 디폴트 값 등이며 텍스트 분석 절차는 '조건-행위'로 표현되는 생성규칙을 2~3개씩 포함하고 있다. 본 시스템에서는 슬릿이 포함하고 있는 패킷들을 순서적으로 실행하여 슬릿값을 찾도록 하였다. 슬릿값은 대부분 텍스트분석절차를 통해 찾도록 했는데 만약 찾지 못했을 경우에는 디폴트 값을 갖도록 하였다.

그리고, 각 프레임 간에 서로 중복되는 슬릿이 많았기 때문에 프레임 집단을 효과적으로 운영하기 위해 서로 일반성/특별성 링크를 주어 계층관계로 조직했으며 상위요소슬릿에다 상위 단계의 프레임과 연결하는 포인터를 주어 구현하였다. 이렇게 조직한 결과 의미네트워크와 마찬가지로 속성계층이 가능하게 되어 상위 프레임이 갖는 슬릿과 부속 정보들을 하위프레임이 이어받을 수 있게 되었다. 그 외에도 소장 장소의 절약으로 지식베이스의 크기가 너무 커지는 것을 방지할 수 있다.

본 연구를 위해 설계한 프레임시스템은 < 그림-2 >와 같다.

각 프레임별로 구조를 자세히 살펴보면 다음과 같다.

(1) 「상표등록가부Ⅰ」프레임은 상표등록을 위해 출원한 상표나 이미 등록된 상표가 과연 등록의 자격요건을 갖추었는지 판단하는 내용의 판결서를 처리한다.

(표-1) 「상표등록가부Ⅰ」프레임의 구조

슬릿명	패킷의 종류
본원상품	1. 텍스트 분석절차 : 단서어기법과 상품구분지식과의 대조를

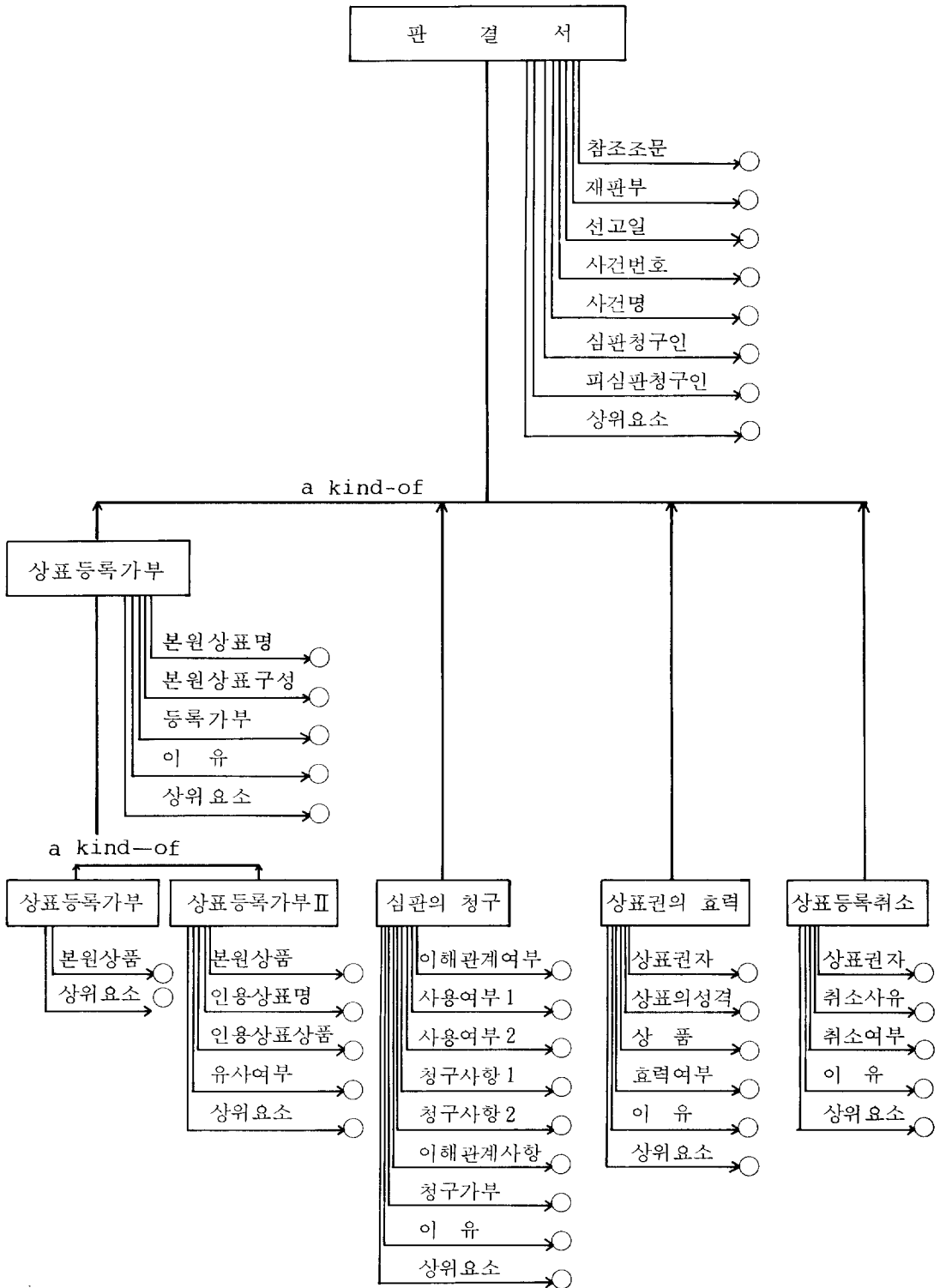
	이용하여 분석 2. 디폴트값 : 심판청구인의 제조 상품
상위요소	1. 값 : 「상표등록가부」프레임에 대한 포인터

(2) 「상표등록가부Ⅱ」프레임은 출원상표나 등록상표가 이미 등록되어 있는 상표와 유사하거나 동일한지를 판단하여 등록자격요건이 되는지를 결정하는 내용의 판결서를 처리한다.

(표-2) 「상표등록가부Ⅱ」프레임의 구조

슬릿명	패킷의 종류
본원상품	1. 텍스트 분석절차 : 단서어기법과 상품구분지식과의 대조를 이용하여 분석 2. 디폴트값 : 심판청구인의 제조 상품
인용 상표명	1. 텍스트 분석절차 : 단서어기법과 편집적기준을 이용하여 분석
인용상표 상품	1. 텍스트 분석절차 : 단서어기법과 상품구분지식과의 대조를 이용하여 분석 2. 디폴트값 : 본원상품과 유사 혹은 동일한 상품
유사여부	1. 텍스트 분석절차 : 단서어기법을 이용하여 분석 2. 디폴트값 : 외관이 유사
상위요소	1. 값 : 「상표등록가부」프레임에 대한 포인터

(4) 「상표권의 효력」프레임은 등록상표의 권리범위가 어느 정도인지, 타 상표에까지 그 효력을 미쳐 타 상표의 등록을 거절 혹은 무효로 만들 수 있는지 여부를 판단하는 내용의 판결서를 처리한다.



<그림-2> 프레임 시스템

(표-4) 「상표권의 효력」 프레임의 구조

슬러트명	패킷의 종류
상표권자	1. 텍스트 분석절차: 단서어기법을 이용하여 분석 2. 디폴트 값: 서보
상표의 성격	1. 텍스트 분석절차: 단서어기법과 상표권자의 값, 경험적 지식을 이용하여 분석
상 품	1. 텍스트 분석절차: 단서어기법과 상품구분지식과의 대조를 이용하여 추론 2. 디폴트 값: 각각의 지정상품
효력 여부	1. 텍스트 분석절차: 상표의 성격의 값과 경험적 지식을 이용하여 분석
이 유	1. 텍스트 분석절차: 단서어기법과 형태소 분석, 참조조문의 값을 이용하여 분석
상 위 요소	1. 값: 「판결서」 프레임에 대한 포인터

(5) 「상표등록취소」 프레임은 이미 등록된 상표가 취소사유에 해당하는지를 판단하는 내용의 판결서를 처리한다.

(표-5) 「상표등록취소」 프레임의 구조

슬러트명	패킷의 종류
상표권자	1. 값: 피심판청구인
취소사유	1. 텍스트 분석절차: 단서어기법을 이용하여 분석
취소여부	1. 텍스트 분석절차: 단서어기법과 위치적기준, 경험적지식을 이용하여 분석

이 유	1. 텍스트 분석절차: 단서어기법과 형태소 분석, 참조조문의 값을 이용하여 분석
상 위 요소	1. 값: 「판결서」 프레임에 대한 포인터

(6) 「상표등록가부」 프레임은 「상표등록가부」 프레임과 「상표등록가부Ⅱ」 프레임의 상위프레임으로 두 가지 프레임에 공통되는 슬러트들로 구성되며 하위 프레임은 본 프레임의 속성을 계승하게 한다.

(표-6) 「상표등록가부」 프레임의 구조

슬러트명	패킷의 종류
본원상표명	1. 텍스트 분석절차: 단서어기법과 편집적기준을 이용하여 분석
본원상표성	1. 텍스트 분석절차: 단서어기법과 본원상표명의 값을 이용하여 분석
등록가부	1. 텍스트 분석절차: 단서어기법과 위치적 기준, 경험적 지식을 이용하여 분석
이 유	1. 텍스트 분석절차: 단서어기법과 형태소 분석, 참조조문의 값을 이용하여 분석 2. 디폴트 값: 일반수요자에게 혼동을 줄 우려가 있다.
상 위 요소	1. 값: 「판결서」 프레임에 대한 포인터

(7) 「판결서」 프레임은 「상표등록가부」 프레임과 「심판의 청구」 프레임, 「상표권의 효력」 프레임, 「상표등록취소」 프레임의 상위 프레임으로 이 4가지 프레임에 공통되는 슬러트로 구성되며 하위 프레임들은 본 프레임의 속성을 계승하게 된다.

(표-7) 「판결서」프레임의 구조

슬릿명	패킷의 종류
참조조문	1. 텍스트 분석절차 : 단서어기법을 이용하여 분석 2. 디폴트 값 : 텍스트의 해당 프레임이 ① 「상표등록가부Ⅰ」프레임이면 ‘상표법 제 8조’ ② 「상표등록가부Ⅱ」프레임이면 ‘상표법 제 9조’ ③ 「심판의 청구」프레임이면 ‘상표법 제 43조’ ④ 「상표권의 효력」프레임이면 ‘상표법 제 26조’ ⑤ 「상표등록취소」프레임이면 ‘상표법 제 45조’
재판부	1. 텍스트 분석절차 : 위치적기준을 이용하여 분석 2. 디폴트 값 : 대법원
선고일	1. 텍스트 분석절차 : 위치적기준을 이용하여 분석
사건번호	1. 텍스트 분석절차 : 위치적기준을 이용하여 분석
사건명	1. 텍스트 분석절차 : 위치적기준을 이용하여 분석
심판청구인	1. 텍스트 분석절차 : 위치적기준과 단서어기법을 이용하여 분석
피심판청구인	1. 텍스트 분석절차 : 단서어기법과 위치적기준을 이용하여 분석
상위요소	1. 값 : 없음

2.2 상표법조문지식의 조직

상표법조문은 텍스트의 내용분석에 중요한 역할을 하는 지식이다. 따라서 상표법조문전문을 입력하여 자동색인을 하고 그 색인어들을 사용

함이 원칙이나 자동색인은 본 연구의 범위에 속하지 않으므로 수작업으로 색인어를 선정하였고 판결서에 주로 쓰이는 주요어와 동의어를 추가하였다. 그리고 색인어(주요어와 동의어 포함)의 의미점과 상표법의 각 조, 각 항, 각 호 의미점 간에 관계를 주어 의미네트워크로 조직하였다.

상표법조문지식은 판결서의 내용 중에서 판결 결과에 대한 이유를 찾는 데 매우 유용하게 쓰인다. 판결 결과는 참조조문에 의거하여 결정되므로 참조조문을 구성하는 색인어들이 많이 나타나는 단어열이나 문장이 이유슬릿의 값이 되기 때문이다.

2.3 상품구분지식의 조직

상품구분지식은 판결문의 내용 중에서 상품슬릿의 값을 찾을 때 쓰며 상표법 시행규칙 부칙의 별표Ⅰ에 나오는 상품구분과 별표Ⅱ에 나오는 서서비스구분을 의미네트워크로 조직하였다.

상품슬릿의 값을 찾는 제 1원칙은 해당 단서어 근처에서 ‘*품’을 찾는 것인데 만약 이 원칙이 실패했을 경우에는 전문화일의 각 토큰들을 상품구분지식에 수록된 상품명이나 각 상품류와 대조하여 일치하면 그 단어를 슬릿값으로 선택하는 원칙을 사용한다.

3. 종결형 단어리스트 구성

종결형 단어리스트는 문장 형태로 된 이유슬릿의 값을 찾을 때 사용하기 위한 언어학적 지식으로 ‘어간+어말어미’ 혹은 ‘어간+선어말어미+어말어미’의 형태로 구성된 단어들을 포함한다. 어간은 동사와 형용사 중에서 판결서에 자주 나타나는 것으로 ‘되-’, ‘하-’, ‘가-’, ‘이-’, ‘같-’, ‘다르-’, ‘있-’, ‘없-’, 의 8가지를 선정하였다. 이 종결

형 단어리스트는 형태소 분석에서 요구되는 단어와 그의 구문정보를 가진 사전의 일종이라 할 수 있다. 조직된 종결형 단어리스트는 (표-8)에 나타내었다.

(표-8) 종결형 단어리스트

간다	*된다*	*이다*	*한다*
가다	*되다*	*이더라*	*하다*
가더라	*되더라*	*이러라*	*하더라*
가라	*되라*	*이리라*	*하라*
갔다	*뒀다*	*있다*	*하였다*
같다	*되었다*	*있더라*	*해라*
다르다	*없다*	*있었다*	*했다*
다르더라	*없더라*	*이었다*	
다르리라	*없었다*	*였다*	

4. 프레임 선정

처리할 텍스트와 관련되는 프레임을 선정하기 위해서는 우선 텍스트 내에서 발견되는 ‘제 * 조’라는 토큰을 이용하고 만약 이런 토큰들이 발견되지 않으면 사건명(토큰번호가 8인 단어로 ‘거절사정’이나 ‘상표등록취소’ 등을 가리킨다)이나 단서어(각 프레임과 관련된 텍스트에 고유하게 나타나는 단어들로 예를 들어 「심판의청구」 프레임과 관련된 텍스트에는 ‘이해관계’, ‘이익’, ‘소익’, ‘손해’, ‘심판’, ‘청구’ 등이 나타나고 「상표권의 효력」 프레임에는 ‘권리범위’, ‘상표권’, ‘효력’, ‘상표권자’ 등이 나타난다)를 이용한다.

본 연구를 위한 프레임시스템은 7종의 프레임들이 서로 연관성을 가지면서 조직된 것이다. 「상표등록가부Ⅰ」 프레임과 「상표등록가부Ⅱ」 프레임은 상표법 제 8 조나 제 9 조와 관계된 것이고 「심판의 청구」 프레임은 상표법 제 43 조

와 관계된 것이며 「상표권의 효력」 프레임은 상표법 제 24 조나 제 25 조 혹은 제 26 조와 관계된 것이고 「상표등록취소」 프레임은 상표법 제 45 조와 관계된 것이다. 이 5종의 프레임은 계층구조로 조직된 프레임시스템에서 가장 하위 단계에 있는 프레임들이고 「상표등록가부」 프레임은 「상표등록가부Ⅰ」 프레임과 「상표등록가부Ⅱ」 프레임을 포함하며 「판결서」 프레임은 가장 일반적인 개념으로 하위의 모든 프레임들을 포함한다.

5. 프레임 실례화

프레임 실례화란 슬릿에 특정한 값을 채워 프레임을 일반적인 개념 유일한 개념으로 변환시키는 것을 의미한다.

프레임을 실례화하기 위해서는 슬릿을 채우기 위한 조건을 만족시켜야 하므로 각 슬릿의 텍스트 분석절차를 실행시키는 것이 가장 중요한 작업이다. 처리할 텍스트의 프레임이 선정되면 해당 프레임에 소속된 슬릿들을 탐색하고 상위 요소슬릿에 의해 연결되는 상위 프레임의 슬릿들을 탐색하게 된다.

본 시스템에서는 초록에 포함될 정보가 나타나는 문장을 암시하는 단서어와 텍스트의 구조적 특성등을 이용하여 슬릿값을 검색하였다. 단서어는 프레임 구조만이 가지는 특성, 즉 예측을 알 수 있다는 특성을 이용하여 참조조문과 주제지식의 분석을 통해 추출한 것이고 텍스트의 구조적 특성은 판결서를 기술하는 양식에서 처음부터 9번째 단어까지는 항상 동일한 구조를 갖는 사실에서 추출한 것이다. 만약 슬릿값을 텍스트에서 검색하지 못했을 때는 디폴트값으로 슬릿을 채운다.

6. 데이터 베이스 구성

본 시스템에서는 텍스트 처리가 끝나고 슬릿값을 모두 찾으면 슬릿명과 슬릿값을 저장한 후 바로 초록 생성시스템으로 연결되도록 설계하였고 슬릿명과 슬릿값의 쌍, 생성된 초록을 이후의 검색시스템에서도 쓸 수 있도록 데이터베이스를 구성하였다.

데이터베이스는 주파일(main file)인 슬릿화일과 초록화일 그리고 색인화일로 구성된다.

7. 초록 생성

초록을 생성하기 위해서는 예비초록과 텍스트 처리에서 얻은 슬릿명-슬릿값 쌍들을 이용해야 한다. 예비초록이란 주요 정보가 들어갈 자리를 괄호로 둔 채 그 정보들을 연결하는 일반적인 내용만 기술하여 미리 작성해 둔 불완전한 초록을 말한다. 예를들어 「본 건은 청구인()가 피청구인()를 대상으로~」와 같이 기술된 예비초록의 괄호를 해당 슬릿값으로 채우면 「본 건은 청구인 박 선우가 피청구인 동서가구산업주식회사를 대상으로~」와 같이 완성된 초록이 되는 것이다. 예비초록은 프레임의 종류별로 5가지를 만들어 프레임명을 키로 하여 소장해 두고 이를 예비초록화일이라 하였다. 초록이 생성되고 소장되는 과정을 순서적으로 살펴보면 다음과 같다.

(1) 해당 프레임의 예비초록을 임시기억 장소에 기억시킨다.

(2) 텍스트 처리결과 얻어진 슬릿값들을 예비초록의 괄호 안에 넣는다. 이 때 문맥이 자연스럽게 되지 않는 경우가 생길 것이므로 그를 해결하기 위해 문맥수정규칙을 이용한다.

(3) 완성된 초록은 'A사건번호'를 키로 하

고 초록의 내용을 소장하여 초록화일을 만든 다음 데이터베이스에 저장한다.

문맥수정규칙은 슬릿값의 받침유무에 따라 예비초록에 제시된 2가지 선택사항 중에 적당한 하나를 선택하게 하는 것이다. 즉, 예비초록의 괄호 다음에 있는 선택사항이 '이' 혹은 '가', '은' 혹은 '는', '을' 혹은 '를', '으로' 혹은 '로', '이라는' 혹은 '라는', '하다' 혹은 '다'일 때 괄호를 채운 슬릿값의 마지막 음절에 받침이 있으면 '이', '은', '을', '으로', '이라는', '하다'를 선택하고 받침이 없으면 '가', '는', '를', '로', '라는', '다'를 선택하게 하는 것이다.

프레임별로 기술된 예비초록은 다음과 같다.

(1) 「상표등록가부Ⅰ」프레임의 예비초록

(재판부), (선고일), (사건번호), (사건명)

본 건은 청구인 (심판청구인)이 < 혹은 '가' > 피청구인 (피심판청구인)을 < 혹은 '를' > 대상으로 심판을 청구한 것으로 그 내용은 다음과 같다.

본원상표 (본원상표명)은 < 혹은 '는' > (본원상표구성)으로 < 혹은 '로' > 된 상표인데, (본원상품)에 사용하면 (이유)하여 (참조조문)의 규정에 의하여 등록이 (등록가부)하다. < 혹은 '다' >

(2) 「상표등록가부Ⅱ」프레임의 예비초록

(재판부), (선고일), (사건번호), (사건명)

본 건은 청구인 (심판청구인)이 < 혹은 '가' > 피청구인 (피심판청구인)을 < 혹은 '를' > 대상으로 심판을 청구한 것으로 그 내용은 다음과 같다.

본원상표 (본원상표명)은 < 혹은 '는' >

(본원상표구성)으로 <혹은 ‘로’>된 상표로서 (본원상품)을 <혹은 ‘를’> 지정상품으로 하고 있는데, 인용상표 (인용상표명)은 <혹은 ‘는’> (인용상표상품)에 사용되고 있다. 이 두 상표는 전체적으로 비교해 볼 때 (유사여부)하다는 점을 알 수 있다. 따라서 (이유)하여 본원상표는 (참조조문)의 규정에 의하여 등록이 (등록가부)하다.

(3) 「심판의 청구」프레임의 예비초록

(재판부), (선고일), (사건번호), (사건명)

본 건은 청구인 (심판청구인)이 <혹은 ‘가’> 피청구인 (피심판청구인)을 <혹은 ‘를’> 대상으로 심판을 청구한 것으로 그 내용은 다음과 같다.

청구인은 피청구인의 상표와 동일 또는 유사한 상표를 사용한 바 (사용여부 1) 현재 사용하고 (사용여부 2), 피청구인의 (이해관계사항)에 직접적인 이해 관계가 (이해관계여부) (청구사항 1)이라는 <혹은 ‘라는’> 청구를 하였다. 그러므로, (참조조문)에 의거하여 살펴보건대 (이유)하여 청구인의 이러한 (청구사항 2) 청구는 (청구가부)하다 할 것이다.

(4) 「상표권의 효력」프레임의 예비초록

(재판부), (선고일), (사건번호), (사건명)

본 건은 청구인 (심판청구인)이 <혹은 ‘가’> 피청구인 (피심판청구인)을 <혹은 ‘를’> 대상으로 심판을 청구한 것으로

그 내용은 다음과 같다. 청구인의 상표와 피청구인의 상표를 대비하여 보면 (상표권자) (상표의 성격)이들은 (상품)에 사용되고 있다. 그러므로 (참조조문)의 규정에 의거하여 살펴볼 때 (이유)하여 청구인의 상표는 피청구인의 상표에 효력을 미칠 수 (효력여부).

(5) 「상표등록취소」프레임의 예비초록

(재판부), (선고일), (사건번호), (사건명)

본 건은 청구인 (심판청구인)이 <혹은 ‘가’> 피청구인 (피심판청구인)을 <혹은 ‘를’> 대상으로 심판을 청구한 것으로 그 내용은 다음과 같다.

(상표권자)의 등록상표가 (취소사유)에 해당하므로 그 등록이 취소되어야 한다는 주장을 (참조조문)에 의거하여 판단해 보면 (이유), 따라서 본 등록상표는 취소를 면할 수 (취소여부).

IV. 텍스트 처리의 과정 및 결과 분석

1. 텍스트 처리의 과정

본 시스템의 실험데이터는 「상표등록취소」 프레임과 관계된 6건의 판결서로 하였으며 입력시에 다음과 같은 통제를 가하였다.

(1) 한 레코드는 80 자를 기준으로 하되 하나의 단어는 반드시 한 레코드에 있어야 하며 두 레코드에 하나의 단어가 분리되어 있어서는 안 된다.

(2) 컴퓨터시스템 환경상 한글처리 조건을 갖추지 못하여 실험텍스트를 로마자로 번지하여 처

리하였다.

(3) 띄어쓰기는 한글 맞춤법에 맞도록 고쳐서 입력하였다.

(4) ‘Canon’ 과 같이 2 단으로 표기된 단어는 ‘Canon/케논’ 과 같이 변형하여 입력하였다.

(5) 그림은 생략하였다.

데이터 처리의 과정을 사건번호가 ‘84 후10’ 인 판결서를 예로들어 살펴보면 다음과 같다.

판결서의 예

대 법 원

1986.3.25. 판결선고

1986.3.25 원본영수

제 2 부

판 결

사건 84 후 10 상표등록취소

심판청구인, 상고인 한해수

소송대리인 변리사 김운배, 이범일
 피심판청구인, 피상고인 주식회사 고려당

소송대리인 변리사 임석재

원 심 결 특허청 1984.1.18 자, 1981년 향
 고심판 당 제 89 호 심결

주 문 상고를 기각한다.

상고 소송비용은 심판청구인의 부담으로 한다.

이 유 심판청구인 대리인의 상고이유를 판단한다.

1. 제 1 점에 대하여

회사의 상호로 된 이른바 상호상표에서 회사의 종류를 표시하는 “주식회사” 등은 일반적인 관용어에 불과하고 그 명칭이 지정상품의 제조업체를 표시하는 부분으로서 일반구매자로부터

가장 주의를 끌기 쉬운 상표의 요부라 할 것이므로 전체를 결합하여 표시하지 아니하면 타인의 상표와 혼동될 염려가 있는 경우를 제외하고는 회사의 종류를 표시하는 부분을 제외하고 명칭부분만으로 간략하게 표시하였다 하여 상표법 제 45 조 제 1 항 제 2 호에 해당한다고 볼 수는 없다 할 것이며, 또한 지정상품의 영업과 같이라면 상표의 이전도 가능한 바(상표법 제 27 조), 원심이 적법히 확정된 사실에 의하면 피심판청구인은 부동산임대 및 과자류판매업을 영위해 오다가 1978.3.20 청구위 김지웅으로부터 1945.9.1 경 위 김지웅의 조부 김동환이 “고려당”이라는 상호 및 상표로 양과자 제조, 판매업을 개시한 이래 부 김규욱을 거쳐 위 김지웅으로 대를 이어 지속되어 온 서울 종로 2 가 84 의 6, 소재 “고려당과자점”의 영업을 양수하여 그곳에 본점을 두고 종래의 부동산임대 및 과자류판매업과 아울러 “고려당”이란 상표를 사용한 과자류제조업을 영위하여 왔으며 1979.1.15 피심판청구인의 상호로 된 문장상표 “주식회사 고려당”을 상품구분 제 3 류 과자와 당류를 지정상품으로 하여 등록받고, 1979.9.17 본점을 성남시 상대원동 307 의 2 로 이전하였지만 위 본점이전 후에도 위 종로 2 가 소재 고려당과자점을 피심판청구인 회사제품인 양과류 직매장 및 소비자상담실등으로 사용하면서 위 고려당과자점에서 판매하는 양과류에 “고려당”이란 상표를 계속 사용하였다는 것이므로 이는 등록상표인 “주식회사 고려당” 중 회사의 종류를 표시하는 관용어인 “주식회사”부분을 생략하고 위 상표의 요부인 회사의 명칭 “고려당”부분만으로 간략하게 표시한 것일 뿐 아니라 청구위 김지웅이 사용하던 주지저명상표를 그 영업과 함께 양수하여 계속 사용하는 것이기도 하

며 그 사용으로 인하여 상품의 출처나 품질을 혼동, 오인시킬 염려가 있다고도 보여지지 아니하니 피심판청구인의 위 “고려당”이란 상표의 사용이 상표법 제 45조 제 1항 제 2호 소정 사유에 해당한다고 할 수 없다. 원심이 위와 같은 취지로 판단하였음은 정당하고 거기에 소론과 같은 법리오해의 위법이 있다고 할 수 없다.

2. 제 2점에 대하여

상표는 상품에 대한 표장이므로 상표권자가 상표를 그 지정상품에 대하여 사용하는 이상 본점과 지점뿐만이 아니라 직매장등 어떤 장소에서의 사용도 제한을 받지 아니한다 할 것이므로 상표법 제 45조 제 1항 제 1호의 “타인”에는 위 직매장등이 포함될 수 없는 법리라 할 것이다.

원심은 그 거시증거에 의하여 서울 종로 2가 소재 “고려당제과점”이 피심판청구인 회사의 직매장이라는 사실을 인정한 다음 위 직매장에서 이건 상표 사용이 상표법 제 45조 제 1항 제 1호 소정의 “타인이 사용하는 것을 묵인하거나 또는 사용하게 하였을 때” 해당하지 않는다고 판단하고 있는 바, 기록에 의하여 살펴보면 원심의 위 인정과 판단은 정당하고 그 사실 인정의 경로에 소론과 같은 체증상의 위법이 있거나 상표법의 법리오해의 위법이 있다고 할 수 없다.

3. 그러므로 상고논지는 모두 이유없어 상고를 기각하고, 상고 소송비용은 패소자의 부담으로 하여 관여법관의 일치된 의견으로 주문과 같이 판결한다.

1986. 3.25.

재판장 대법원판사 이 정 우

대법원판사 정 태 균
대법원판사 신 정 철
대법원판사 김 형 기

완성된 프레임의 예

S 84 후 10

상표권자 : 피심판청구인

취소사유 : 혼동될 염려가 있는 경우 그리고 타인이 사용하는 것을 묵인하거나 또는 사용하게 하였을 때

취소여부 : 있다

참조조문 : 상표법 제 45조 제 1항 제 2호, 상표법 제 27조, 상표법 제 45조 제 1항 제 1호

이 유 : 출처나 품질을 혼동, 오인시킬 염려가 있다고도 보여지지 아니하니 피심판청구인의 위 “고려당”이란 상표의 사용이 상표법 제 45조제 1항 제 2호 소정 사유에 해당한다고 할 수 없다. 그리고 상표권자가 상표를 그 지정상품에 대하여 사용하는 이상 본점과 지점 뿐만이 아니라 직매장등 어떤 장소에서의 사용도 제한을 받지 아니한다 할 것이므로 상표법 제 45조 제 1항 제 1호의 “타인”에는 위 직매장 등이 포함될 수 없는 법리라 할 것이다.

재판부 : 대법원 제 2부

선 고 일 : 86.3.25. 판결

사건번호 : 84 후 10

사 건 명 : 상표등록취소

심판청구인 : 한해수

피심판청구인 : 주식회사 고려당

완성된 초록의 예

대법원 제 2부, 86.3.25. 판결, 84 후 10, 상표등록취소.

본 건은 청구인 한해수가 피청구인 주식회사 고려당을 대상으로 심판을 청구한 것으로 그 내용은 다음과 같다.

피심판청구인의 등록상표가 혼동될 염려가 있는 경우 그리고 타인이 사용하는 것 묵인하거나 또는 사용하게 하였을 때에 해당하므로 그 등록이 취소되어야 한다는 주장을 상표법 제 45 조 제 1 항 제 2 호, 상표법 제 27 조, 상표법 제 45 조 제 1 항 제 1 호에 의거하여 판단해 보면, 출처나 품질을 혼동, 오인시킬 염려가 있다고도 보여지지 아니하니 피심판청구인의 위 “고려당” 이란 상표의 사용이 상표법 제 45 조 제 1 항 제 2 호 소정사유에 해당한다고 할 수 없다. 그리고 상표권자가 상표를 그 지정상품에 대하여 사용하는 이상 본점과 지점 뿐만이 아니라 직매장 등 어떤 장소에서의 사용도 제한을 받지 아니한다 할 것이므로 상표법 제 45 조 제 1 항 제 1 호의 “타인” 에는 위 직매장 등이 포함될 수 없는 법리라 할 것이다. 따라서 본 등록상표는 취소를 면할 수 있다.

2. 텍스트 처리의 결과 분석

본 연구에서 설계한 자동 초록시스템은 텍스트로부터 초록에 포함될 슬릿값들을 검색하여 예비초록의 괄호를 채움으로써 초록을 생성하기 때문에 시스템의 성능을 평가하기 위해 초록이 포함하는 각종 슬릿값들의 적합성 여부를 판단해 보았다.

각 초록이 포함하는 슬릿값들은 대체로 적합했으나 취소사유슬릿과 이유슬릿의 값은 부적합한 경우가 종종 있었다. 텍스트의 구조적 특성을 이용하여 단어나 짧은 단어열 형태의 값을

검색하는 재판부, 선고일, 사건번호, 사건명, 심판청구인, 피심판청구인 슬릿과 경험적 지식을 이용하여 값을 판단하는 상표권자, 취소여부 슬릿 그리고 형태상의 특징을 단서로 하여 값을 검색하는 참조조문슬릿의 경우는 텍스트에서 규칙에 적용되는 사항을 검색하기가 용이하여 슬릿값 검색 알고리즘은 100%의 정확률을 보였다. 그러나 수 개의 단서어와 위치정보, 종결형 단어 등 다양한 지식을 이용하여 문장 형태로 된 값을 검색하는 취소사유슬릿과 이유슬릿은 유사한 내용은 검색할 수 있으나 정확하고 세련된 문장을 검색하기가 어려웠고 특히 복수의 슬릿값을 요구하는 경우에 상이한 내용의 해당 사항을 모두 검색해야 하므로 정확성을 기하는 데 어려움이 있었다. 따라서 취소사유슬릿과 이유슬릿 검색 알고리즘은 각각 67%의 정확률을 보였다.

부적합한 값을 갖는 경우를 분석하면 다음과 같다. 86 후 100의 초록과 80 후 118의 초록에서 이유슬릿은 두 가지 값을 가져야 하는데 두 가지 이유 중 한 가지는 참조조문에 포함되는 색인이나 「상표등록취소」프레임에 고유한 단서어들로 문장내용을 구성하지 않았기 때문에 검색되지 않았다. 또 83 후 70의 초록과 77 후 4의 초록에서 취소사유슬릿은 동일한 단서어로 구성된 유사한 내용의 문장들이 동일한 규칙에 적용되어 복수로 검색되었다.

결국 완전하고 정확한 초록을 생성하기 위해서는 취소사유슬릿과 이유슬릿의 값을 검색하는 알고리즘을 정확하게 구성해야 하며 이를 위해서는 적용 분야 텍스트의 특성을 잘 분석하여 주제지식을 세밀하게 조직하고 문장보다는 단어 형태로 된 값을 요구하는 슬릿들로 프레임을 구성해야 할 것이다.

V. 결 론

자동초록에 관한 종래의 연구는 대부분 문헌을 대표할 수 있는 중요 문장을 선정하여 원문헌에 출현한 순서대로 나열하는 방법을 채택한 것으로 초록하라기보다는 발췌문의 성격이 강하고 문맥이 부자연스러워 이용하기에 불편한 점이 많았다.

따라서 본 연구에서는 발췌문이 아닌 순수한 초록을 생성하기 위해 미리 조직된 주제분야지식을 이용하여 자연언어로 된 텍스트를 분석하고 초록에 포함될 정보를 텍스트로부터 추출하여 예비초록의 괄호를 채움으로써 초록을 완성하는 지식기반시스템을 설계하였다.

본 시스템의 설계에 채택된 기법과 그 효과는 다음과 같다.

(1) 주제분야의 지식을 미리 체계적으로 조직하여 지식베이스를 구성함으로써 어떤 정보를 어떻게 찾아야 하며 그 정보의 특성은 어떠한가를 예측할 수 있도록 하여 정보추출을 쉽게 하였다.

(2) 주제분야지식을 프레임과 의미네트웍으로 표현함으로써 텍스트에서 일어날 상황을 미리 예측할 뿐만 아니라 상이한 개념들이 소유하고 있는 속성을 서로 공유하여 소장과 탐색에 있어 효율성을 얻고 추론이 가능하도록 하였다.

(3) 텍스트 분석시 초록에 포함될 정보가 나타나는 문장을 암시하는 단서어와 텍스트의 구조적인 특성 등을 이용하여 정보를 추출하였다.

(4) 슬릿값 검색 알고리즘을 개발하면서 부분적으로 형태소 분석을 시도하였다.

(5) 주요 정보가 들어갈 자리를 괄호로 둔 채 그 정보들을 연결하는 일반적인 내용만 기술한 예비초록을 각 경우별로 마련하여 텍스트에서

추출된 정보로 괄호를 채움으로써 문맥이 자연스러운 초록을 생성할 수 있도록 하였다.

(6) 어색한 문맥을 수정함으로써 이용자들이 읽고 이해하는데 어려움이 없도록 문맥수정규칙을 적용하였다.

본 논문에서 상표등록취소와 관계되는 대법원 판결서를 대상으로 시스템을 구현시킨 결과 얻은 결론은 다음과 같다.

(1) 사전에 조직된 주제분야지식과 예비초록을 이용하였기 때문에 복잡한 구문분석을 하지 않고도 효과적으로 초록을 생성할 수 있었다.

(2) 텍스트의 구조적 특성을 이용하여 슬릿값을 검색하는 경우와 단어 형태로 된 슬릿값을 검색하는 경우는 비교적 용이했으나 문장 형태로 된 값을 요구하는 슬릿의 경우는 정확한 값을 검색하기 어려웠다. 따라서 구조적 특성을 뚜렷하게 규명할 수 있는 분야의 텍스트에 적용하거나 프레임을 세분화하여 문장 형태보다는 단어 형태로 된 값을 요구하는 슬릿들로 프레임을 구성한다면 효율성은 더욱 증대될 것이다.

(3) 본 시스템은 지식기반시스템이므로 새로운 분야의 주제지식을 조직하여 추가함으로써 시스템을 확장할 수 있다.

결론적으로 본 연구에서 설계한 자동초록시스템은 적용분야 텍스트의 특성을 파악하고 주제 지식과 프레임을 세밀하게 조직하여 응용한다면 경제적이고 효율적인 방법으로 초록을 생성할 수 있으리라 본다.

참 고 문 헌

〈단행본〉

1. 남기삼·고영근, 표준국어문법론. 서울:탑출판사, 1985.
2. 법원행정처 편, 대법원판결요지집, 특별편Ⅱ. 서울:법원행정처, 1984.
3. 이주근, 인공지능, 서울:정문각, 1986.
4. 정영미, 정보검색론. 서울:정음사, 1987.
5. Barr, A. and Feigenbaum E. A., The Handbook of AI, vol. I. Los Alto:kaufmann, 1981.
6. Borko, H. and Bernier, C., Abstracting Concepts and Methods. New York:Academic Press, 1975.
7. Guinchart, C. and Menou, M., 정보관리론. 사공철·김태수 공역, 서울:구미무역, 1987.
8. Oddy, R. N., Robertson, S. E., van Rijsbergen, C. J. and Williams, P. W. ed., Information Retrieval Research. London: Butterworth, 1981.
9. Rich, E., Artificial Intelligence. 유석인 등 번역, 인공지능. 서울:상조사, 1986.
10. Salton, G., Automatic Information Organization and Retrieval. New York:McGraw-Hill, 1968.
11. _____, Dynamic Information and Library Processing. Englewood Cliffs:Prentice-Hall, 1975.
12. Salton, G. and McGill, M. J., Introduction to Modern Information Retrieval. New York:McGraw-Hill, 1983.
13. Schank, R. C and Abelson, R., Scripts Plans Goals and Understanding, New York: Wiley, 1977.
14. Tennant, H., Natural Language Processing. New York/Princeton:PBI, 1981.
15. Waterman, D. A., A Guide to Expert Systems. Massachusetts:Addison - Wesley, 1986.

16. Winston, P. H., Artificial Intelligence, 2nd ed. Massachusetts:Addison - wesley, 1984.

〈논 문〉

1. 최기선, “자연언어 이해,” 정보과학회지, 4 (3), 1986.
2. 정영미, “우리말 신문기사 검색을 위한 질문응답시스템 구현에 관한 연구,” 정보관리학회지, 4(3) 4 (1), 1987.
3. Baxendale, P. B., “Machine - Made Index for Technical Literature - An Experiment,” IBM J. of Research and Development, 2 (4), 1958.
4. Climenson, W. D., Hardwick, N. H. and Jacobson, S. N., “Automatic Syntax Analysis in Machine Indexing and Abstracting,” American Documentation, 12(3), 1961.
5. Cullingford, R., “SAM,” In: Schank, R. C. and Riesbeck, C. K. ed., Inside Computer Understanding, Hillsdale:LEA Publishers, 1981.
6. DeJong, G. F., “Prediction and Substantiation: A New Approach to Natural Language Processing,” Cognitive Science, 3, 1979.
7. _____, “Skimming Newspaper Stories by Computer,” In: Proceedings of the 5th International Joint Conference on AI, 1977.
8. Doszkocs, T. E., “Natural Language Processing in Information Retrieval,” J. of ASIS, 37(4), 1986.
9. Earl, L. L., “The Resolution of Syntactic Ambiguity in Automatic Language Processing,” Information Storage and Retrieval, 8, 1972.
10. _____, “Use of Word Government in Resolving Syntactic and Semantic Ambiguities,” Information storage and Retrieval, 9, 1973.

11. Edmundson, H. P., "New Methods in Automatic Extracting," J. of ACM, 16(2), 1969.
12. Edmundson, H. P. and Wyllys, R. E., "Automatic Abstracting and Indexing - Survey and Recommendations," Comm. of the ACM, 4(5), 1961.
13. Feigenbaum, E. A., "The Art of Artificial Intelligence," In: Proceedings of the 5th International Joint Conference on AI, 1977.
14. Grishman, R., "Natural Language Processing," J. of ASIS, 35(5), 1984.
15. Humphrey, S. M. and Miller, N. E., "Knowledge - Based Indexing of the Medical Literature: The Indexing Aid Project," J. of ASIS, 38(3), 1987.
16. Liddy, E., Bonzi, S., Katzer, J. and Oddy, E., "A Study of Discourse Anaphora in Scientific Abstracts," J. of ASIS, 38(4), 1987.
17. Luhn, H. P., "The Automatic Creation of Literature Abstracts," IBM J. of Research and Development, 2(2), 1958.
18. Mathis, B. A., Rush, J. E. and Young, C. E., "Improvement of Automatic Abstracts by the Use of Structural Analysis," J. of ASIS, 24(2), 1973.
19. Minsky, M., "A Framework for Representing Knowledge," In: Winston, P. H. ed., The Psychology of Computer Vision, New York: McGraw Hill, 1975.
20. Montgomery, C. A., "Linguistics and Information Science," J. of ASIS, 23(3), 1972.
21. Paice, C. D., "The Automatic Generation of Literature Abstracts: An Approach Based on the Identification of Self - indicating Phrases," In: Oddy, R. N. et al. ed., op. cit.
22. Pollock, J. J. and Zamora, A., "Automatic Abstracting Research at Chemical Abstracts Service," J. of Chemical Information and Computer Sciences, 15, 1975.
23. Rush, J. E., Salvador, R. and Zamora, A., "Automatic Abstracting and Indexing II. Production of Indicative Abstracts by Application of Contextual Inference and Syntactic Coherence Criteria," J. of ASIS, 22(4), 1971.
24. Schank, R. C., Charniak, E., Wilks, Y., Winograd, T. and Woods, W., "Panel on Natural Language Processing," In: Proceedings of the 5th International Joint Conference on AI, 1977.
25. Schank, R. C., Goldman, N., Rieger, C. J. and Riesbeck, C., "MARGIE: Memory, Analysis, Response Generation and Inference on English," In: Proceedings of the 3rd International Joint Conference on AI, 1973.
26. Schank, R. C., Lebowitz, M. and Birnbaum, L., "Parsing Directly into Knowledge Structure," In: Proceedings of the 6th International Joint Conference on AI, 1979.
27. Taylor, S. L., and Krulee, G. K., "Experiments with an Automatic Abstracting System," In: Proceedings of the ASIS Annual Meeting, Chicago, 14, 1977.
28. Taylor, S. L., Krulee, G. K. and Henschen, L. T., "Automatic Abstracting of Textual Material," In: Proceedings of the 5th International Joint Conference on AI, 1977.
29. Vleduts - Stokolov, N., "Concept Recognition in an Automatic Text - Processing System for the Life Science," J. of ASIS, 38(4), 1987.
30. Wilensky, R., "PAM," In: Schank, R. C. and Riesbeck, C. K. ed., op. cit.
31. _____, "PAM - A Program That Infers Intentions," In: Proceedings of the 5th International Joint Conference on AI, 1977.

International Joint Conference on AI,
1977.

32. Wyllys, R. E., "Extracting and Abstracting
by Computer," In: Borko, H. ed., Auto-
mated Language Processing, New York:
Wiley, 1968.