

객체지향 기법을 이용한 시소러스 관리 시스템의 개발에 관한 연구

Development of a Thesaurus Management System based on the Object-Oriented Technique

박계숙(Kay-Sook Park)*

목 차

1. 서 론	3.2.4 고유명사 테이블
2. 시소러스 구조	3.3 GUI 화면 설계
3. 객체지향 기법을 이용한 시소러스 관리 시스템 개발	3.3.1 단어관리
3.1 시스템 개발 환경	3.3.2 그룹관리
3.2 데이터베이스 구조	3.3.2.1 그룹 윈도우
3.2.1 사전 테이블	3.3.2.2 계층 윈도우
3.2.2 계층 테이블	3.3.3 시소러스 관리
3.2.3 시소러스 테이블	3.3.4 고유명사 관리
	4. 결 론

초 록

시소러스를 구축하기 위해서는 새로운 단어의 입력, 수정, 삭제 및 단어간 새로운 관계생성 등의 동적인 변화에 신속 정확히 대응할 수 있도록 시소러스 관리 시스템이 개발되어야 한다. 본 연구에서 개발한 시소러스 관리 시스템은 객체지향적 구조로 데이터베이스를 설계하였으며, GUI화면을 제공하여 이용자 인터페이스를 향상시켰다. 또한 시소러스를 구축하는 데 있어서 동일한 영문 용어의 다양한 한글표기로 인한 정보검색 효율 저하 문제를 해결하기 위해 동의어의 확장에 역점을 두었다.

ABSTRACT

For the construction of thesaurus, a thesaurus management system is needed which can process dynamic variations fast and exactly such as input, correction and deletion of words, and definition of new relationship between words. In this paper, I developed a thesaurus management system based on the object-oriented technique and GUI(graphic user interface) screen, and to enhance the effectiveness of information retrieval. I put emphasis on the expansion of synonym, English and Korean words containing the same concept.

* 한국전자통신연구소 기술정보실
 ■ 논문 접수일 : 1996년 10월 22일

1. 서 론

전통적인 개념의 시소러스는 색인 및 검색 작업에 있어서 색인자와 검색자 사이에 정보 표현 용어의 통일을 기함으로써 정보의 원활한 유통을 도모하고 높은 검색효율(재현율)을 보장하는 기능을 갖는다. 즉 시소러스 용어는 일종의 통제어로서 시소러스 용어를 주제어로 채택한 통제어 필드에 한해서 높은 검색 효율을 보장하고 있다. 그러나 시소러스를 이용한 수작업 방식의 색인작업은 데이터베이스 구축에 보다 많은 시간을 요하며, 전문(Fulltext) 데이터베이스의 증가는 자동색인방식의 적용을 불가피하게 하는 요인이 되고 있다. 이러한 현실 속에서 시소러스는 수작업에 의한 통제된 색인어의 선정도구로써보다는 자동색인 및 검색에 있어서 색인어나 탐색어의 자동확장 등 지적인 향상을 도모하기 위한 지식 베이스로서의 활용성이 보다 더 인식되고 있다.

최근 들어 한글 자동색인 및 능형 정보검색 기술에 관한 연구가 활발히 진행됨에 따라 국내에서도 시소러스 구축에 대한 관심이 증대되고 있으나, 주로 외국의 시소러스를 그대로 번역한 결과 용어의 용례가 우리 실정에 맞지 않거나 시소러스 관리 시스템의 개발이 미흡해 시소러스의 지속적인 갱신이 이루어지지 않은 채 책자형의 시소러스를 발간한 데 그치고 있는 실정이다.

그러나 특정 주제분야의 전문용어는 시대적인 흐름 및 학문의 발달에 따라 새로운 단어의 생성, 구(舊)단어의 도태, 단어 간 새로운 관계 생성 등의 변화가 예측된다. 정보통신 분야는 이러한 변화가 매우 활발한 분야 중의

하나로, 시소러스의 구축 및 지속적인 유지관리를 위해서는 단어의 추가, 수정, 삭제 등의 작업을 정확하고 융통성 있게 처리할 수 있는 시소러스관리시스템이 개발되어야 한다.

본 고에서는 본 연구를 통해 구축한 정보통신 분야 시소러스의 구조에 대해서 간략히 언급한 후 시소러스 관리 시스템의 개발에 관하여 기술하고자 한다.

2. 시소러스 구조

시소러스에서는 계층관계 표시나 정의, 범위주기 등을 통하여 용어의 의도된 의미를 지시하고 있으며, ISO 2788에서는 이러한 용어간 관계표시에 SN(Scope Note : 범위주기), USE(우선어), UF(Used For : 비우선어), TT(Top Term : 최상위어), BT(Broader Term : 상위어), NT(Narrower Term : 하위어), RT(Related Term : 관련어)와 같은 기호를 사용할 것을 제안하고 있다. 시소러스마다 정의하고 있는 용어간 관계개념은 ISO 2788에서 제안하고 있는 바와 유사하나 이들 관계를 표시하기 위한 기호의 배치 및 사용법 등 시소러스의 구조와 내용은 다양하다. 시소러스 관리 시스템의 기능 및 설계는 시소러스 구조에 의해 결정되며, 동일한 구조를 갖고 있는 시소러스의 생성, 운용에만 활용될 수 있다.

국내 과학기술 분야 문헌에서는 전문용어의 한글화 부진에 따라 한글 및 영문 용어의 혼용은 물론 동일한 영문용어에 대한 다양한 한글표기 단어가 사용되고 있다. 이로 인해 한

글 데이터베이스를 검색하는 데 있어서 검색어 선정에 혼란을 야기시키고 있으며 검색효율이 저하되는 결과를 초래하고 있다. 또한 국내에서는 데이터베이스 제작에 다수의 전문 색인자를 투여하고 있지 않으므로 시소러스에 근거한 통제된 색인어를 수작업에 의해 부여하는 일은 현실적으로 어려운 실정이다. 따라서 본 연구에서는 시소러스를 제작하는 데 있어서 시소러스의 용도를 수작업에 의한 색인어 통제보다는 서로 다른 표기형태로 사용되고 있는 동일한 개념의 단어(동의어)를 통합 관리함으로써 검색시 탐색어의 확장을 돕기 위한 탐색용 시소러스(Search-aid Thesaurus)를 구축하는 데 역점을 두었다.

본 연구에서는 시소러스 구조를 정보통신 분야의 대표적인 시소러스인 IEE(영국전기학회)의 *INSPEC Thesaurus* 구조를 기본으로 삼되, 디스크립터(Descriptor)와 비디스크립터(Nondescriptor)의 관계를 두지 않고 동일 개념을 갖는 다양한 표기형태의 용어를 모두 동의어(Synonym) 관계로 정의함에 따라, 용어간 관계표시 기호로 동의어(ST), 최상위어(TT), 상위어(BT), 하위어(NT), 관련어(RT)를 적용하였다.

동의어란 동일한 개념을 표현하는 두 가지 이상의 용어로서 사실상 대체가 가능하나 시소러스 표준에서는 이 중 하나를 디스크립터로 선택하고 나머지는 비디스크립터로 정한 후 디스크립터만을 색인어나 검색어로 사용할 것을 권하고 있다. 그러나 이용자는 디스크립터나 비디스크립터의 구분 없이 아무 단어나 검색어로 사용할 것을 희망하므로, 본 연구에서는 예외적으로 디스크립터와 비디스크립터

를 구분하지 않고 동의어 관계로 정의하였다. 이외의 단어관계는 *INSPEC Thesaurus*에서와 마찬가지로 NT는 동일 단어그룹 내에서 자신의 위치보다 1단계 낮은 계층의 단어를, BT는 자신의 위치보다 1단계 높은 계층의 단어를, TT는 그룹의 최상위에 위치해 있는 그룹 대표어를, 그리고 RT는 상하 계층관계가 아닌 횡적인 상호 연관관계의 단어를 나타내도록 하였다.

3. 객체지향 기법을 이용한 시소러스 관리 시스템 개발

3.1 시스템 개발 환경

시소러스는 새로운 단어의 입력, 기입력된 단어의 수정, 삭제, 단어의 중복입력 방지를 위한 중복단어 점검, 단어간 새로운 관계 생성 등의 지속적인 변화가 예상된다. 시소러스 관리 시스템은 시소러스 데이터베이스를 효율적으로 관리하기 위한 인터페이스 시스템으로, 본 연구에서는 이러한 동적인 변화에 신속 정확히 대응할 수 있도록 시소러스 관리 시스템을 개발하는 데 있어서 아래의 4가지 개발전략을 채택하였다.

첫째, 시소러스의 특수성에 적합한 객체지향적 구조의 데이터베이스 설계를 통해 모든 단어 및 계층그룹은 하나의 객체로 인식되며 이용자가 선택하는 접근점이 곧 객체가 되도록 하였다.

둘째, 객체 지향적 시스템의 흐름을 위해 쉽고 편리한 접근법을 제공하는 하이퍼텍스트 기법을 적용하였다.

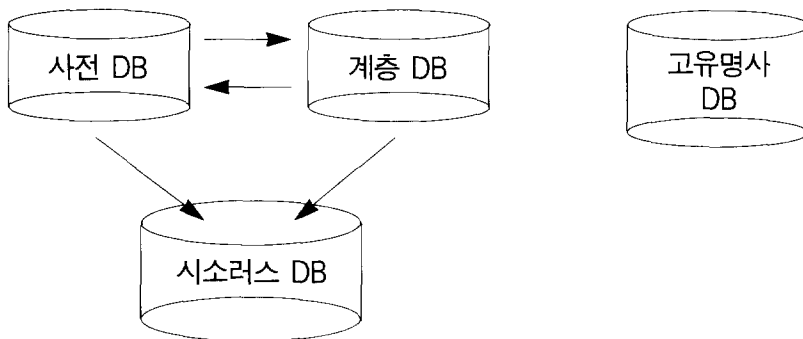
셋째, GUI(그래픽-사용자 인터페이스)를 이용한 사용자 환경을 제공하여 사용자가 특별한 사용법을 익히지 않아도 손쉽게 이용할 수 있도록 하였다.

넷째, 수치체계를 적용하여 체계적이고 빠른 검색과 색인 생성의 용이성을 확보하였다. 시소러스 관리 시스템은 이러한 개발전략에 따라 윈도우 환경에서 운영할 수 있도록 윈도우용 Visual Basic 3.0과 Visual C++를 이용해 개발했으며 DBMS는 Microsoft ACCESS 2.0을 이용하였다. Visual Basic은 쉽게 사용할 수 있는 개발툴로 차후의 시스템 수정, 관리 차원에서 보다 쉽게 작업할 수 있으며, ACCESS는 처리속도가 뛰어난 윈도우용 DBMS로 구축된 데이터베이스에 대해 별도의 시스템 없이도 모니터링, 수정 등을 할 수 있는 장점을 지니고 있다. 또한 본 시스템은 Standalone 시스템으로 개발하였으나 클라이언트/서버 환경의 시스템으로 발전시켜 활용할 수 있으므로 그 기대효과는 증대될 것으로 보인다.

3.2 데이터베이스 구조

시소러스 관리 시스템의 최종 목표는 가나다순 시소러스 데이터베이스를 구축하는 것으로서, 본 시스템의 주요 기능으로는 단어관리(단어의 입력, 수정, 삭제), 단어그룹 관리(그룹의 생성, 삭제) 및 그룹 내 계층구조 관리(계층구조 생성, 삭제), 시소러스 생성, 단어/그룹/시소러스 인쇄, 그리고 이를 기반으로 한 데이터베이스 구축 기능이 요구된다.

시소러스는 일반적으로 전문용어를 중심으로 구축되나 외래어 형태의 고유명사가 색인어나 검색어로 채택되는 빈도가 적지 않은 데 반해 한글표기가 난해한 점을 고려해 본 연구에서는 시소러스에 고유명사도 포함시켰다. 이를 위해 본 연구에서는 데이터베이스를 <그림 1>에서 보는 바와 같이 단어관리를 위한 사전 테이블, 고유명사 관리를 위한 고유명사 테이블, 단어그룹 내에서의 단어간 계층관계 관리를 위한 계층 테이블, 시소러스 생성을 위한 시소러스 테이블의 4개 테이블로 구성하였다. 시소러스 테이블은 사전 테이블과 계층



<그림 1> 데이터베이스 구성도

〈표 1〉 사전 테이블의 구성

필드명	Word_No	Word_H	Word_E	Synonym	RWord_No
데이터 타입	Integer(Long)	Text	Text	Memo	Memo
데이터 크기	4 Byte	100 Byte	200 Byte	-	-
인덱스	Primary Key	-	-	-	-
비고	단어 번호	한글 단어	영문 단어	동의어	관련어

테이블로부터 생성되나, 고유명사 테이블은 단순히 동의어 관리를 위해 만든 테이블로 계층관계를 정의하지 않으므로 독립적으로 존재한다.

3.2.1 사전 테이블

사전 테이블은 시소러스에 포함되는 모든 단어를 등록하는 테이블로, 〈표 1〉과 같이 고유 단어번호(Word_No), 한글 단어(Word_H), 영문단어(Word_E), 동의어(Synonym), 관련어 번호(RWord_No)로 구성하였다. 단어번호는 Integer형으로 8자리(99,999,999)까지 유효하도록, 그리고 한글단어와 영문단어는 각각 100자, 200자까지 입력할 수 있는 텍스트형(Text Type)으로 설계하였다. 동의어와 관련어는 여러 개가 존재할

것을 고려해 메모형(Memo Type)으로 정의하였으며, 단어 사이를 “.”로 구분한 후 시스템에서는 “.”를 이용해 단어를 분리한 뒤 화면에 표시하도록 하였다. 이들 데이터 요소 중에서 단어번호, 한글단어, 영문단어, 동의어 값은 고유(unique)하여야 하며, 단어번호와 한글단어 값은 반드시 존재하여야 한다.

3.2.2 계층 테이블

계층 테이블은 사전 테이블에 등록된 단어 상호간의 계층관계를 정의해 놓은 테이블로, 사전 테이블에 등록된 단어번호와 계층 테이블에 등록된 단어번호를 일치시킴으로써 사전 테이블 내 단어에 수정, 삭제 등의 변화가 발생했을 경우 이 단어가 계층 테이블의 어느 그룹에 속해 있건 간에 동일하게 수정, 삭제

〈표 2〉 계층 테이블의 구성

필드명	G_L_No	Word_No	Word H	Level
데이터 타입	Currency	Integer(Long)	Text	Integer
데이터 크기	8 Byte	4 Byte	100 Byte	2 Byte
인덱스	Primary Key	-	-	-
비고	수치체계 번호	단어 번호	한글 단어	레벨 번호

가 일어날 수 있도록 설계하였다. 계층 테이블은 <표 2>와 같이 수치체계번호(G_L_No), 단어번호(Word_No), 한글단어(Word_H), 그리고 단어의 계층레벨(Level)로 구성하였다.

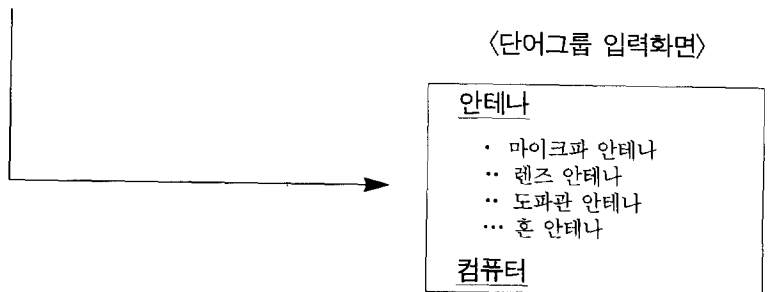
수치체계번호는 각 단어가 속하는 단어그룹 및 단어그룹 내에서의 소속 계층레벨을 표시한 숫자로, 특정 단어그룹 내에서 개개 단어가 속하는 계층레벨을 “.”의 갯수로 표시해 입력하면 계층 테이블 생성시 시스템이 개개 단어의 레벨 값을 수치체계번호로 자동연산해 등록하도록 설계하였다. 수치체계번호는 16자리(9999.99.99.99.99.99.99)까지 유효하며, 여기서 처음의 4자리(9999)는 단어그룹번호를, 그리고 다음의 각 2자리는 그룹 내에서의 단어레벨을 나타내는 수치로 앞에서부터 레벨

1, 레벨 2, ..., 레벨 6까지 표시할 수 있다. 즉 한 그룹 내에서의 계층레벨 수는 최대 6단계까지 전개할 수 있도록 하였으며, 한 어군에 속하는 단어의 수에는 제한을 두지 않았다. 예를 들어, 그룹번호가 1000이고 그룹명이 안테나일 경우, 마이크로파 안테나가 이 그룹명의 하위레벨(NT)이고 렌즈 안테나가 마이크로파 안테나의 하위레벨(NT)이라 하면 <그림 2>와 같은 수치체계번호가 연산된다.

<그림 2>에 나타난 계층관계를 보면, 렌즈 안테나는 마이크로파 안테나의 NT에 속하는 반면 마이크로파 안테나는 렌즈 안테나의 BT에 속한다. 이와 같이 BT와 NT는 상호 대응관계로서 계층 테이블에 수치체계번호를 들으러서 시소러스 테이블 생성시 상하계층의 단

<계층 테이블>

G-L-No							Word-No	Word-H	Level
1000	00	00	00	00	00	00	000001	안테나	0
1000	01	00	00	00	00	00	000004	마이크로파 안테나	1
1000	01	01	00	00	00	00	000002	렌즈 안테나	2
1000	01	02	00	00	00	00	000005	도파관 안테나	2
1000	01	02	01	00	00	00	000003	혼 안테나	3
1001	00	00	00	00	00	00	000006	컴퓨터	0



<그림 2> 계층 테이블에서의 수치체계번호 연산 예

〈표 3〉 시소러스 테이블의 구성

필드명	Word_No	TT	BT	NT
데이터 타입	Integer(Long)	Memo	Memo	Memo
데이터 크기	4 Byte	-	-	-
인덱스	Primary Key	-	-	-
비고	단어 번호	최상위어	상위어	하위어

어를 자동적으로 추출해낼 수 있을 뿐만 아니라 그룹 내에서 단어의 삽입 또는 삭제가 일어날 경우 신속 정확히 처리될 수 있도록 하였다. 이와 같이 계층 테이블은 단어번호(Word_No)와 수치체계번호(G_L_No)로만 구성되어도 무방하나 계층 테이블의 검색속도 향상을 위해 한글단어(Word_H)와 레벨번호(Level) 필드를 중복적으로 두었다.

3.2.3 시소러스 테이블

시소러스 테이블은 최종적으로 단어별 시소러스가 구축되는 테이블로서, 〈표 3〉과 같이 단어번호(Word_No), 하위어(NT), 상위어(BT), 최상위어(TT)로 구성하였다. 시소러스 테이블은 시소러스 생성 명령에 의해 계층 테이블내에서 각 단어의 위치를 검색해 개개 단어에 대한 NT, BT, TT를 추출함으로써 구축된다. 이때 단어 그룹의 0 레벨에 위치한 단어가 TT, 자신의 위치보다 1단계 아래에 위치한 단어가 NT, 1단계 위에 위치한 단어가 BT로 인식된다. 특정 단어는 하나 이상의 그룹에 속할 수 있으며 따라서 복수개의 TT, BT, NT를 가질 수 있으므로 데이터 타입을 메모형으로 하고 “.”를 이용해 구분하였다.

시소러스 관리 시스템의 최종 결과물인 시소러스는 시소러스 테이블 내 TT, BT, NT 단어와 사전 테이블 내 ST, RT와 더불어 한글단어(Word_H)의 가나다 순으로 배열되어 출력된다. 본 연구 결과 생산될 시소러스 형태는 아래와 같다.

산업용 로봇 (industrial robot)

- ST 산업용 로봇
- NT 산업용 매니플레이터
- BT 로봇
- TT 자동화
- RT 산업용 제어
- 제조 과정
- 이동 로봇
- 로봇 시각

3.2.4 고유명사 테이블

고유명사 테이블은 기관, 단체명, 시스템명 등의 고유명사에 대한 한글 및 영문 표기 단어를 모아 동의어 관리를 하기 위한 것으로, 테이블 구조는 〈표 4〉와 같다. 앞서 언급한 바와 같이 고유명사 테이블은 계층관계를 정의하지 않으므로 계층 테이블이나 시소러스 테이블과는 무관하게 독립적으로 존재한다.

〈표 4〉 고유명사 테이블의 구성

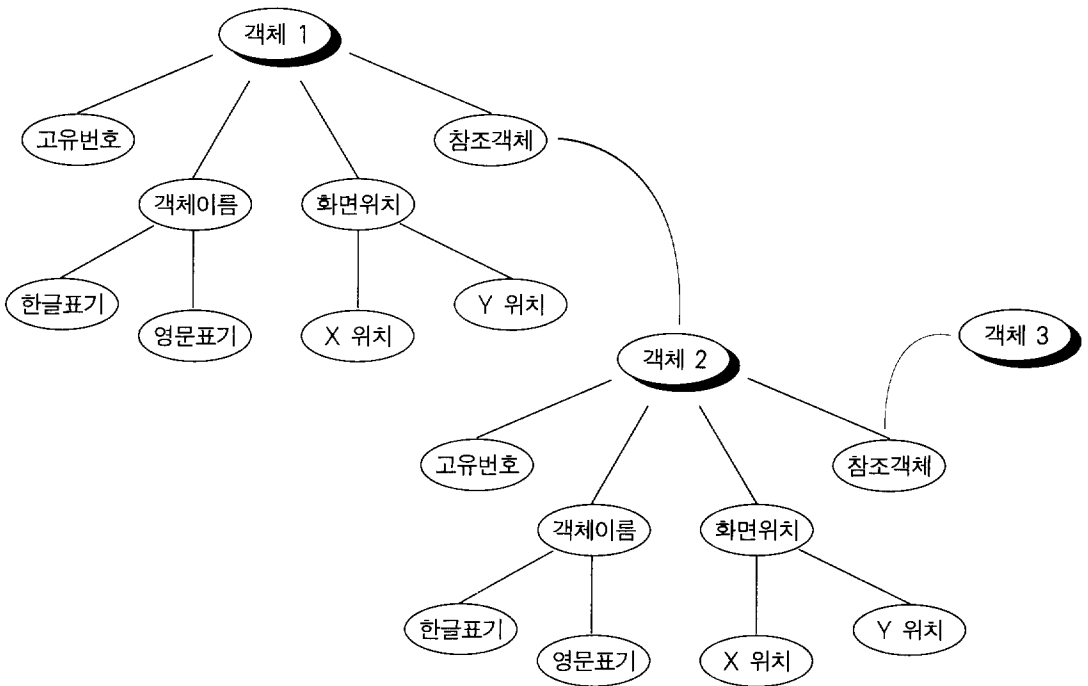
필드명	Word-No	Word-H	Word-E	Synonym
데이터 타입	Integer(Long)	Text	Text	Memo
데이터 크기	4 Byte	100 Byte	200 Byte	-
인덱스	Primary Key	-	-	-
비고	단어 번호	한글 단어	영문 단어	동의어

3.3 GUI 화면 설계

시소러스 관리 시스템은 이용자 인터페이스를 향상시키기 위하여 윈도우의 GUI 화면으로 설계하였으며, 데이터베이스에 대한 동적인 변화에 신속 정확히 대응할 수 있도록 <그

림 3>과 같이 하이퍼텍스트 기법을 이용한 객체 지향적 데이터 구조로 설계하였다.

본 시스템은 객체 지향적 데이터 처리를 위해 이용자에 의해 선택된 단어 또는 그룹을 하나의 객체로 인식하여 객체의 위치 및 관계와는 상관없이 객체의 모든 처리(추가, 수정,



〈그림 3〉 하이퍼텍스트 기법을 이용한 객체 지향적 데이터 구조

삭제, 위치 이동)에 대해 마치 하나의 객체가 처리되듯이 정확히 추출하여 처리되도록 하였다. 이를 위해 화면에 제공되는 활성 객체는 이용자의 마우스를 이용한 선택에 의해 그 객체가 지칭하거나 참조하고 있는 다른 객체로 신속 정확히 이동할 수 있도록 모든 객체는 객체 자신이 지칭할 또다른 객체에 대한 주소 또는 참조 방법 등을 갖고 있다. 따라서 화면에 어떤 객체가 표현되면 시스템은 마우스 사건(event)을 참조하여 사건의 위치와 이미 시스템에 등록되어 있는 객체의 화면 위치를 비교하며, 이때 서로 위치 범위가 일치하면 참조 객체와 현 객체의 고유번호를 시스템에 알리고, 시스템은 다음 객체의 시소러스를 검색하여 화면에 출력한다.

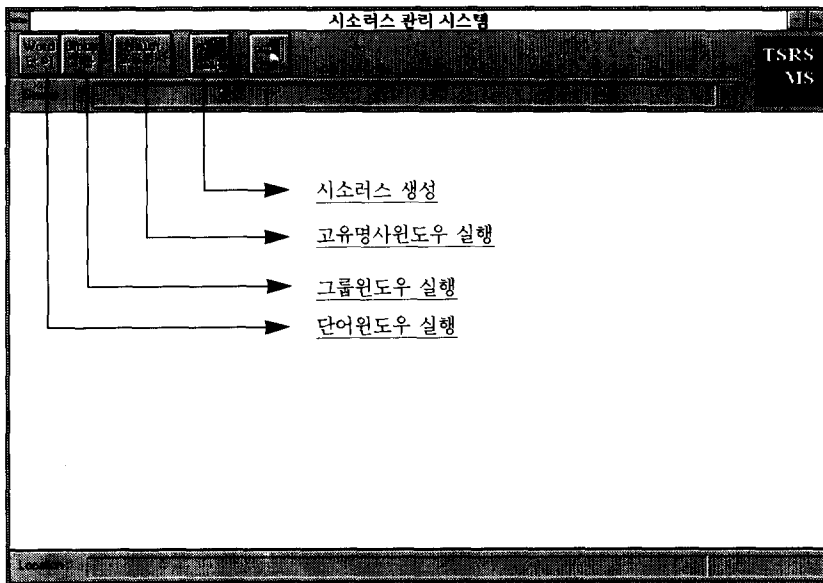
본 시스템에서 사용자가 실제 작업을 수행하게 될 윈도우(화면)의 종류는 단어관리를

위한 단어 윈도우, 단어그룹 관리를 위한 그룹 윈도우, 그리고 고유명사 관리를 위한 고유명사 윈도우로 구성하였다. <그림 4>는 이들 윈도우를 실행하기 위한 시스템의 기본화면을 보여주고 있다.

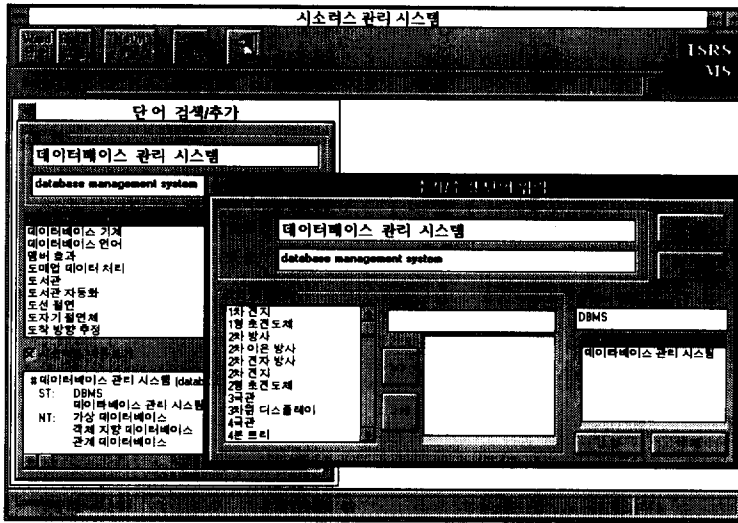
3.3.1 단어관리

시소러스에 포함되는 모든 단어는 사전 테이블에 등록되어야 한다. 이러한 단어관리를 위해 요구되는 기능은 단어의 입력(추가), 수정, 삭제 등의 기본기능과 사전 테이블에 입력된 단어의 조회(검색)기능, 그리고 특정 단어의 시소러스 내용을 볼 수 있는 기능으로, 단어 윈도우를 <그림 5>와 같이 설계하였다.

단어 윈도우에서 사전 테이블에 등록된 단어를 조회하기 위해 조회하고자 하는 단어(한



<그림 4> 시소러스 관리 시스템의 기본 화면



〈그림 5〉 단어 윈도우

글 또는 영문단어)의 일부분을 입력하면 스트링 검색에 의해 입력한 단어 주변의 단어 리스트를 가나다순 또는 알파벳순으로 정렬하여 보여주며, 이 중에서 특정 단어를 클릭하면 이 단어의 한글 및 영문표기를 확인할 수 있다. 이어서 시소러스 내용보기를 클릭하면 특정 단어의 시소러스 내용(NT, BT, TT, RT, ST 관계)을 볼 수 있다.

단어 윈도우에서 추가 또는 수정 기능을 클릭하면 우측에 서브윈도우가 나타난다. 이 윈도우에서 보는 바와 같이 사전 테이블에 등록되는 사항은 한글단어, 영문단어, 동의어, 관련어이다. 사전 테이블을 관리하는 데 있어서 가장 중요한 점은 중복단어의 입력을 방지하는 것으로, 이를 위해 단어가 입력될 때마다 한글 및 영문단어와 동의어에 대해서는 중복 검사 실시 기능을 두었다.

관련어는 사전 테이블에 등록된 단어에 한

해 등록되어야 하므로 서브윈도우 내 단어 리스트 박스를 통해 사전 테이블을 조회함으로써 관련어를 등록할 수 있도록 하였다. 정보 검색시 관련어에 대한 확장검색이 그다지 효과적이지 못할 뿐만 아니라 관련어의 등록작업은 사전 테이블이 완성된 단계에서 처리해야 하므로 본 연구에서는 관련어를 최소한도로 등록하였다.

3.3.2 그룹 관리

사전 테이블에 등록된 단어는 계층 테이블로 옮겨져 특정 그룹에 소속되어야 한다. 이러한 단어그룹 관리를 위해 요구되는 기능은 그룹명의 등록 기능과 그룹에 속하는 개개 단어들간의 상호계층관계 설정을 위한 계층관리 기능으로, 전자는 그룹 윈도우에서, 후자는 계층 윈도우에서 작업하도록 설계하였다. 그

러나 단어간 관계설정은 그룹명이 등록된 후에 수행되어야 하므로 계층 윈도우를 그룹 윈도우의 서브윈도우로 구성하였다.

사전 테이블에 등록된 단어를 그룹(계층) 테이블에 등록하는 과정에서 단어의 반복입력 작업을 없애고 단어 표기의 일관성과 정확성을 기할 수 있도록 단어 윈도우내 단어를 마우스로 드래그(drag)하여 그룹(계층) 윈도우에 드롭(drop)시킴으로써(이하 Drag and Drop 기능이라 칭함) 단어를 등록하는 방법을 사용하였다.

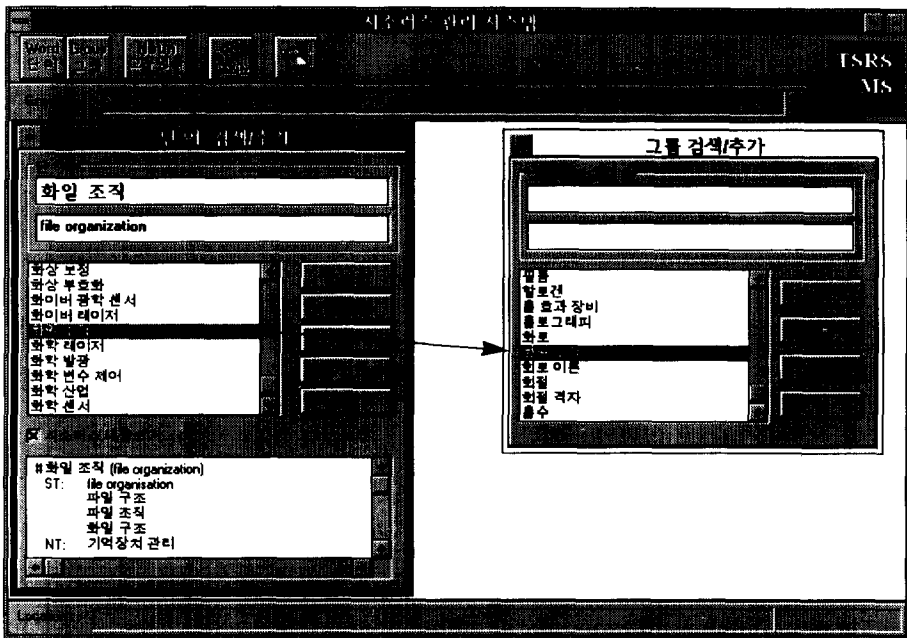
3.3.2.1 그룹 윈도우

그룹 윈도우는 단어 그룹 관리를 위한 그룹명의 등록, 삭제, 조회(검색) 기능을 수행할

수 있는 화면으로, <그림 6>과 같이 단어 윈도우와 그룹 윈도우를 화면에 모두 열어 놓은 상태에서 그룹명에 해당되는 단어를 단어 윈도우로부터 그룹 윈도우로 Drag and Drop 기능에 의해 이동하여 그룹명을 등록한다.

그룹 윈도우에서 그룹명은 등록 및 삭제만 가능하도록 하였다. 그러나 그룹명을 삭제하면 그 그룹에 속하는 모든 단어가 함께 삭제되므로 신중을 기하여야 한다. 이러한 위험을 방지하기 위해 그룹 대표어로 등록된 단어를 사전 테이블에서 삭제하고자 할 경우는 먼저 그룹 윈도우에서 해당 그룹명을 삭제한 후에 사전 테이블 내 단어를 삭제할 수 있도록 하였다.

3.3.2.2 계층 윈도우



<그림 6> 단어 윈도우와 그룹 윈도우의 관계

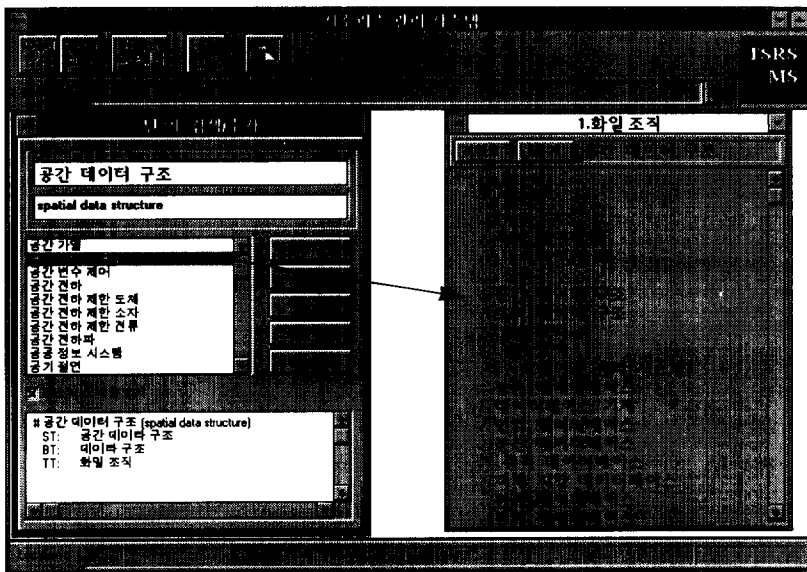
계층 윈도우는 그룹에 속하는 개개 단어들의 계층레벨을 등록하고 관리하는 화면으로, 그룹 윈도우에서 해당 그룹명을 선택하면 계층 윈도우가 열린다. 그룹에 속할 단어들의 계층레벨 설정은 <그림 7>과 같이 단어 윈도우와 계층 윈도우를 화면에 모두 열어 놓은 상태에서 Drag and Drop 기능에 의해 단어 윈도우 내 단어를 계층 윈도우로 이동하되 이 단어를 자신의 계층레벨보다 1단계 위에 속하는 단어 위치에 드롭시키면 이 단어의 하위어로 등록되도록 하는 방법을 사용하였다. 이러한 방법을 통해 단어의 계층레벨을 자유롭게 설정할 수 있도록 하였으며, 계층 윈도우에 등록된 단어의 계층레벨이 “.”의 갯수로 표시되도록 하여 그룹 내 단어간 계층 식별이 용이하도록 하였다.

이와 같이 계층 윈도우에서의 그룹 내 단어

관리는 단어 윈도우를 통해 이루어지므로 단어 윈도우에서 단어를 수정/삭제하면 계층 윈도우 내 단어에도 수정/삭제가 일어난다. 그러나 특정 단어를 삭제할 경우 이 단어의 하위어로 등록된 모든 단어가 동시에 삭제되므로 단어 윈도우에서 단어를 삭제하고자 할 경우는 반드시 계층 윈도우나 단어 윈도우 내 시소러스 내용보기 박스를 통해 하위레벨의 단어가 있는지 확인한 후 삭제하여야 한다.

3.3.3 시소러스 관리

한 단어는 다수의 단어 그룹에 속할 수 있으며 계층 테이블 내 단어에 대한 입력, 수정, 삭제 등의 작업으로 인해 단어간 계층관계에 변화가 발생할 경우, 시소러스 테이블을 갱신시키기 위한 복잡한 작업과정이 수반된다. 따



<그림 7> 단어 윈도우와 계층 윈도우의 관계

라서 본 연구에서는 단어 입력작업의 신속한 처리를 위해 신규 단어(그룹)이 입력되는 즉시 시소러스가 갱신되지 않고, 임의의 시점에 서 시소러스를 일괄 생성할 수 있도록 하였다. 특정 단어의 시소러스 내용(NT, BT, TT, RT, ST 관계)은 단어 윈도우 내 시소러스 내용보기 박스를 통해 확인할 수 있으며, 시소러스 전체에 대한 인쇄기능도 두었다.

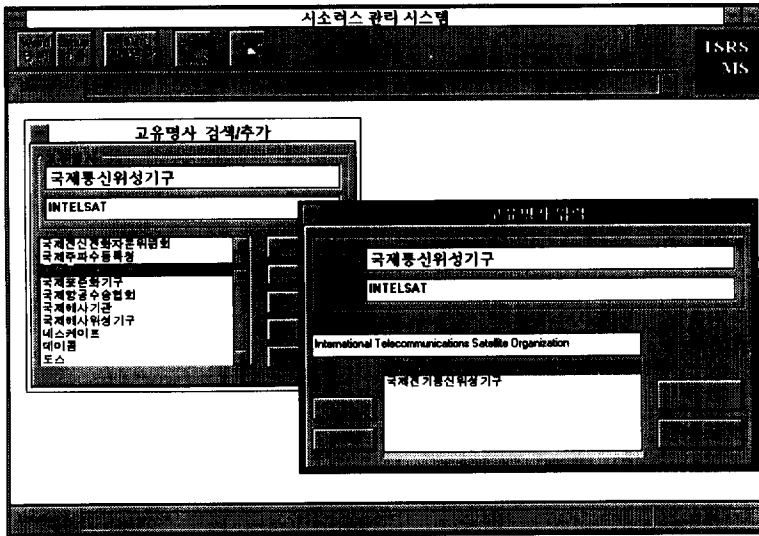
4. 결 론

본 연구에서는 정보통신 분야의 시소러스 구축을 위해 개발한 시소러스 관리 시스템의 데이터베이스 구조 및 기능에 관하여 설명하였다. 본 시스템을 통해 생성될 시소러스는 탐색용 시소러스로서 한국전자통신연구소에서 운영중인 ETLARS 정보검색 시스템에 적용하여 동의어 및 하위 개념어를 이용한 탐색어의 자동확장에 활용할 계획이다. 국산 데이터베이스들의 대부분이 자연어를 색인으로 채택하고 있는 점을 고려할 때 본 시소러스는 정보통신 분야 데이터베이스에 대한 검색 효율을 상당 수준 높일 것으로 예상된다.

3.3.4 고유명사 관리

고유명사 테이블은 단순히 고유명사에 대한 동의어 처리를 위해 마련한 것으로, 단어의 입력, 수정, 삭제와 같은 기본기능과 고유명사 테이블에 등록된 단어의 조회기능, 중복입력 단어의 검사기능 등이 요구된다. 따라서 고유명사 관리를 위한 윈도우는 <그림 8>과 같이 단어 윈도우와 유사하게 설계하였다.

그러나 전문 분야의 시소러스를 제대로 구축하기 위해서는 용어의 선정 및 관계설정 작업에 해당 분야 전문 연구진의 참여가 절실히 요구될 뿐만 아니라 신생 용어를 지속적으로



<그림 8> 고유명사 윈도우

추가해 시소러스를 확장시켜 나가야 할 것이다. 데이터베이스 운영과정에서 입수가 가능한 색인어 및 탐색어의 발생(활용)빈도에 관

한 통계 데이터를 시소러스 용어 선정에 활용하는 것도 시소러스 확장을 위한 좋은 방법이 될 것이다.

참 고 문 헌

- 강현규, 박세영, 한국어 정보 검색, 주간기술 동향, 727 : 15-42, 1995.
- 박영몽, 김민구, 이정태, 지식 기반의 정보 검색 시스템, 한국정보과학회논문지, 21(11) : 2090-2098, 1994.
- 정영미, 국내 문자정보 데이터베이스의 색인에 관한 연구, 정보관리학회지, 13(1) : 19-45, 1996.
- 정영미, 하이퍼텍스트의 개념과 응용에 관한 고찰, 정보관리학회지, 6(2) : 3-20, 1989.
- 최석두, 김영환, 남영준, 하이텔 메뉴검색용 시소러스의 개발에 관한 연구, 정보관리학회지, 13(1) : 227-241, 1996.
- 최석두, 정동열 공역, 시소러스 개발 지침, 서울 : 문헌정보처리연구회, 1994.
- Gauch, S. and Smith, J. B., Search Improvement via Automatic Query Reformulation, ACM Trans. on Information Systems, 9(3) : 249-280, 1991.
- Hares, J. S. and Smart, J. D., Object Orientation. New York : John Wiley & Sons, 1994.
- INSPEC Thesaurus. London : IEE, 1995.
- ISO 2788 (1986). Guidelines for the Establishment and Development of Monolingual Thesauri.
- Kristensen, J., Expanding End-Users' Query Statements for Free Text Searching with a Search-aid Thesaurus, Information Processing & Management, 29(6) : 733-744, 1993.
- Milstead, J. L., Specifications for Thesaurus Software, Information Processing & Management, 27(2/3) : 165-175, 1991.
- Smith, P. J., Shute, S. J. and Galdes, D., Knowledge-based Tactics for an Intelligent Intermediary System, ACM Trans. on Information Systems, 7(3) : 246-270, 1989.