

시소러스의 기본 용어관계에 관한 연구

A Study on the Basic Term Relationship of Thesaurus

한 상 길 (Sang-Kil Han)*

〈 목 차 〉

- I. 서론
- II. 기존 시소러스의 용어관계 분석
 - 1. 정보검색환경변화와 시소러스
 - 2. 정보검색환경변화에 따른 시소러스의 기능변화
- III. 기존 시소러스의 용어관계 분석
 - 1. 등가관계
 - 2. 계층관계
 - 3. 연관관계
 - 4. 용어관계의 문제점
- IV. 용어관계의 확장
 - 1. 새로운 용어관계 지시기호의 설정
 - 2. 동등관계 지시기호의 확장
 - 3. 패싯지시자의 사용
- V. 결론

초 록

본 논문의 목적은 색인과 검색기법의 변화와 사용자 위주에서 이용자 위주로 바뀌고 있는 정보검색환경의 변화에 따라 지금까지의 시소러스 구축 국제 표준 지침과 기존 시소러스에서 사용하고 있는 기본 용어관계에 문제점이 있음을 지적하고 새로운 환경에 맞는 용어관계의 확장 방안을 제시하는데 있다. 이를 위해 본 연구는 ISO 2788과 ANSI/NISO Z 39.19 기준, 그리고 현재 사용되고 있는 20개(외국 14개, 국내 6개) 시소러스의 기본 용어관계를 비교 분석하여 문제점과 한계점을 도출하였다. 이를 토대로 정보검색환경의 변화에 따라 필요한 시소러스의 기본 용어관계의 확장 방안을 제시하였다.

Abstract

With the development of indexing and retrieval techniques and the trends of information retrieval environment becoming more and more user-friendly, existing guideline for international thesaurus establishment standards and basic term relations currently used in thesaurus have become no longer adequate. The purpose of this study is to present solutions to expanding term relations which will fit new information retrieval environment.

This report reviews standards for ISO 2788 and ANSI/NISO Z 39.19, and compares and analyzes 20 thesaurus basis term relations currently used both at home and abroad(14 overseas and 6 in Korea) to find out problems and limitations. Based on findings of the study, this report suggests how to expand thesaurus basic term relations to accommodate changes in information retrieval environment.

* 대림대학 문헌정보과 조교수

I. 서 론

정보기술의 급속한 발전은 시소러스의 기능에도 많은 변화를 가져왔다. 용어통제를 위해 시소러스를 도입하였던 초기에는 수작업 서지 데이터베이스가 대부분이었지만, 현재는 쉐문 데이터베이스가 추적을 이루는 자동색인 환경으로 변화되었다. 쉐문 데이터베이스시스템은 자연 언어처리를 기본으로 하기 때문에 수작업 색인에서 사용하던 전통적인 시소러스로는 대응할 수 없는 부분이 많다. 또한 정보이용자들의 이용환경도 GUI(graphic user interface) 환경으로 바뀜에 따라 기존의 인쇄형 시소러스에서 사용할 수 없었던 다양한 용어관계의 설정이 가능하게 되었다.

이러한 정보환경에서 시소러스를 자동색인에 이용할 경우 시소러스가 제 기능을 수행하기 위해서는 다양한 용어관계의 설정이 요구된다. 특히 동일한 개념이 다양한 용어에 의해 표현될 수 있으므로 관련문헌이 여러 개의 색인어 아래 분산되는 것을 막아 일관성 있는 색인을 위해서는 동등관계의 정의가 매우 중요하다. 이를 위해서는 동일한 개념이 다양한 용어로 표현되는 용어관계를 동등관계로 정의하여 색인어 선정에 도움을 주고 탐색 시에도 탐색자가 이들의 용어관계를 인식할 수 있도록 해야 한다.

또한 계층관계에서 동일한 계층관계에 속한 형제관계 용어들과 연관관계에서 동일용어에 포함된 여러 관련관계 용어들이 다양한 속성을 지니고 있으므로 검색 시 잡음이 많이 나타날 수 있다. 그러나 이러한 변화와 문제점에도 불구하고 기존 시소러스의 작성지침으로 사용되고 있는 국제 표준인 ISO 2788-1986 [International Organization for Standardization. *Documentation : Guidelines for the Establishment and Development of Monolingual Thesauri*. 2nd ed. 1986. 이하 ISO 2788로 표기함]에서는 용어간의 관계를 BT, NT, RT의 기본관계와 범위주기 등으로 단순히 나타내고 있어 이용자의 요구를 만족시키지 못할 뿐만 아니라 정보검색환경의 변화에 부응하지 못하고 있다.

이 점에 착안하여 본 연구는 시소러스 구축을 위한 국제 표준 지침으로 사용되고 있는 ISO 2788과 미국 국가표준인 ANSI/NISO Z 39.19-199×[National Information Standards Organization. *Guidelines for the Construction, format and Management of Monolingual Thesauri*. 1991. 이하 Z 39.19로 표기함]의 기준, 그리고 현재 사용되고 있는 시소러스의 등가관계, 계층관계, 연관관계로 정의되는 기본 용어관계를 비교 분석하여 문제점을 도출하였다. 이를 토대로 기존의 시소러스 구축을 위한 국제 표준지침으로 사용되고 있는 ISO 2788에서 제시되지 않은 새로운 용어관계의 설정과 기존의 용어관계 확장을 통하여 정보검색시스템의 효율성 향상에 기여할 수 있도록 하는데 본 연구의 목적이 있다.

II. 정보검색환경변화와 시소러스

1. 정보검색시스템에서의 시소러스

생성된 정보를 정보전문가가 분석 가공하여 축적해둔 축적매체에서 이용자의 요구에 적합한 정보만을 탐색해내는 일련의 과정을 정보검색이라고 한다. 일반적인 정보검색시스템에서 사용하는 자연언어색인은 같은 개념이 여러 개의 다른 용어로 표현될 수 있기 때문에 재현율의 저하를 가져오게 되며, 또한 동형어의어를 조절하지 않는 한 정확률의 저하를 가져오게 된다. 이러한 자연어색인의 언어적 문제를 해결하고 보다 효과적인 탐색을 하기 위해서는 시소러스와 같은 통제어휘집이 필요하다.

전통적인 시소러스의 정의는 “어떤 개념에 대하여 그것을 가장 적절히 표현 할 수 있는 표목을 선택하기 위하여 만들어진 어구의 집대성”¹⁾이라는 개념으로 나타났다. 다시 말해 시소러스란 용어를 알고 내용을 찾는 일반사전과는 반대로, 뜻을 알고 있으나 그 개념에 해당하는 적당한 용어를 찾는 데 사용되는 용어집이라고 할 수 있다.

그러나 보다 최근에는 시소러스를 지식베이스를 구성하는 하나의 용어 데이터베이스로 보는 관점이 많이 나타나고 있다. Susan Jones는 “시소러스는 특별한 주제분야에서 문헌을 색인하고 검색하는 데 사용된 단어와 구(句)들간의 관계에 대한 정보의 寶庫이며, 문헌에서 발생하는 생생한 텍스트 단어와 질의에서 사용된 단어간의 매핑을 향상시키기 위해 전문가들이 구축해 놓은 용어 데이터베이스이다”²⁾라고 정의해 용어 데이터베이스라는 개념으로 시소러스의 개념을 확장시켰다.

다시 말해 시소러스는 색인어를 선정할 때 비디스크립터를 디스크립터 용어로 변환시켜 색인어를 통일시키고, 용어의 계층관계를 이용하여 가중치를 부여하는 역할을 수행한다. 또한 검색시에 적절히 질의어 선정, 용어의 안내, 용어간 상호참조를 이용하여 효율적인 탐색을 수행할 수 있도록 하는데 이용된다. 그러므로 시소러스는 정보검색시스템에 필수적인 요소로, 색인어 통제와 다양한 검색기능으로 재현율과 정확률을 높이는 중요한 도구라 할 수 있다.

2. 정보검색환경변화에 따른 시소러스의 기능변화

지금까지의 일반적인 시소러스의 기능과 목적은 단순한 용어 통제기능과 참조 기능 등으로

1) 日本ドキュメンテーション協會. 『シソ-ラス』. 東京:同協會, 1970, p. 2.

2) Susan Jones. "A thesaurus data model for an intelligent retrieval system." *Journal of Information Science*, Vol.19, No.3, 1993, p. 167.

4 한국도서관·정보학회지(제 30권 제 4호)

제한되어 정의되어 왔다. Z 39.19에서는 ① 저자, 색인자, 이용자가 사용한 자연언어를 색인작성과 검색에 사용할 통제어휘로 번역하는 수단을 제공한다. ② 색인어 부여의 일관성을 보증한다. ③ 용어간의 의미관계를 지시한다. ④ 문헌 탐색에서 탐색보조도구가 된다³⁾고 정의하고 있다.

그러나 이와 같은 시소러스의 목적과 기능은 수작업색인에서 타당한 견해라고 할 수 있다. 정보환경이 급격하게 변화하는 오늘날에 있어서의 시소러스 기능은 더욱 확장될 필요가 있다. 다시 말하면, "현재의 정보검색환경에서 시소러스는 이용자/서버구조에서 분산된 데이터베이스관리, 시소러스 관계의 새로운 확장, 즉 개념적 접근, 어의범주, 언어적 기능, 그리고 멀티미디어 표현을 위한 확장 등이 필요하다"⁴⁾고 할 수 있다. 따라서 시소러스의 목적과 기능은 용어통제와 표준화에 그칠 것이 아니라 그 자체를 하나의 지식베이스로 보는 관점이 추가되어야 할 것이다.

J. Rowley는 시소러스의 새로운 이용으로 "이용자 인터페이스는 GUI 또는 윈도우 환경에서 검색된 레코드나 탐색 프로파일을 디스플레이 하는 동시에 스크린에서 시소러스를 찾거나 통제어휘의 리스트를 보는 것이 가능하다는 것이다. 또 다른 발전은 자연어 탐색의 이용으로 지능 인터페이스를 제공하는 지식베이스로서 시소러스를 이용한다. 시소러스는 단어들 사이의 관계를 정의하고, 확장하거나 축소하는 등 다른 방법으로 이용자 탐색을 발전시키는 시스템에 의해 자동적으로 이용될 수 있기 때문이다"⁵⁾고 주장하였다. B. H. Weinberg도 "시소러스의 혁신과 미래에서의 적용은 색인단계에서 보다 오히려 탐색단계에서의 이용에 있다고 주장하고, 인공지능과 하이퍼텍스트를 지적하고 시소러스의 개념은 대부분 지식베이스의 기초가 되며, 한 영역의 지식표현을 구성한 것이므로 인공지능에 시소러스를 적용할 수 있다"⁶⁾고 주장하였다.

따라서 이러한 멀티미디어 환경과 하이퍼텍스트 기능 등에 따른 색인과 검색기법의 변화와 더불어 사용자 중심에서 이용자 중심으로 바뀌고 있는 정보검색 환경의 변화에 따라 시소러스도 정보환경과 이용자요구를 만족시킬 수 있는 방향으로 변해야 한다. 그러므로 새로운 환경에 적합한 시소러스의 기능을 수행하기 위해서는 지금까지의 단순한 용어관계로만 만족할 것이 아니라 새로운 환경에 맞는 용어관계의 설정이 필요하다고 할 수 있을 것이다.

3) ANSI. Z39.19-1974 : *American National Standard Guidelines for Thesaurus Structure, Construction, and Use*. New York : American National Standards Institute, 1974. 『시소러스 개발 지침』(최석두, 정동열 공역. 서울 : 문헌정보처리연구회, 1994), p. 199.

4) Alice Stern and Norbert Rischette. "On the construction of a super thesaurus based on existing thesaurus." In *Tools for knowledge organization and the human interface*. Frankfurt : Indeks Verlag, 1991, p. 144.

5) J. Rowley. "The controlled versus natural indexing languages debate revisited : a perspective on information retrieval practice and research." *Journal of Information Science*, Vol.20, No.2, 1994, pp.115-116.

6) B. H. Weinberg. "Library classification and information retrieval thesauri : comparison and contrast." *Cataloging and Classification Quarterly*, Vol.19, No.3/4, 1995, p. 39.

Ⅲ. 기존 시소러스의 용어관계 분석

1. 분석대상의 선정 및 범위

1) 분석대상

본 연구는 시소러스 구축을 위한 국제 표준 지침으로 사용되고 있는 ISO 2788과 미국 국가 지침인 Z 39.19의 2개 지침과 14종의 외국 시소러스와 6종의 국내 시소러스를 포함 전체 20개 시소러스를 분석대상으로 하였다.

시소러스 구축지침을 ISO 2788과 Z 39.19만으로 한정된 이유는 ISO 2788은 영국의 BS 5273(1979), UNESCO(1981), 독일의 DIN 1463(1976), 프랑스의 AFNOR Z 47-100 지침을 통합한 것으로 전 세계에서 보편적으로 사용되고 있기 때문이다. 또한 Z39.19는 ISO 2788에 영향을 미치지 않았기 때문에 분석 대상으로 삼았다.

시소러스 용어관계 분석 대상으로 삼은 외국 시소러스는 『Root Thesaurus』, 『Construction Industry Thesaurus』, 『Thesaurus of Engineering and Scientific Terms』, 『INSPEC Thesaurus』, 『Thesaurus of ERIC Descriptor』, 『Ei Thesaurus』, 『Library of Congress Subject Headings』, 『JICST 科學技術用語シソラス』, 『NASA Thesaurus』, 『INIS Thesaurus』, 『UNESCO Thesaurus』, 『Medical Subject Heading』, 『A Women's Thesaurus』, 『Thesaurus of Sociological Indexing Terms』(이하 TSIT로 표기함)의 14종이다.

『Encyclopedia of Library and Information Science』⁷⁾에서는 시소러스의 주제분야를 항공학, 농학, 천문학과 천체물리학, 대기과학, 사회과학, 생물학 및 의학, 화학, 지구과학, 전기전자공학, 에너지, 재료의 11가지로 구분하고 있다. 또한 『Thesaurus Guide : Analytical Directory of Selected Vocabularies for Information Retrieval』⁸⁾에서는 종류 일반, 정보 영역, 수학과 물리화학, 물리화학 기술, 천문학 및 기하학, 농학과 영양학, 생물의학, 종교와 환경과학, 사회과학, 문화와 예술의 10개 범주로 구분하고 있다.

본 연구에서 분석대상으로 삼은 외국에서 개발된 14개의 시소러스를 『Thesaurus Guide』에서 제시한 주제분야별로 구분하면 <표 1>과 같다.

7) D. J. Foskett. "Thesaurus." In *Encyclopedia of Library and Information Science*. Vol.30. New York : Marcel Dekker, 1980, p.423.

8) GID. *Thesaurus Guide : analytical directory of selected vocabularies for information retrieval*. Amsterdam : North-Holland, 1985. p. xxxvi.

〈표 1〉 분석대상 외국 시소러스

영역		분석대상 시소러스		
기본영역	세부영역	시소러스명	주제분야	연도
총류 일반	도서관, 과학기술 일반	UNESCO	교육학, 사회과학, 문화, 인문과학, 자연과학, 기술공학, 커뮤니케이션 등	1977
		LCSH	전 주제	1992
정보영역	정보학, 커뮤니케이션, 컴퓨터	ROOT	공학, 에너지, 전기, 커뮤니케이션, 컴퓨터 등	1988
수학 및 물리화학	수학, 물리학, 엔지니어링, 기계공학, 전자기술, 에너지, 원자력	INSPEC	물리학, 전기전자 공학, 컴퓨터 공학	1991
		JICST	정보학, 과학, 공학, 의학 등	1987
물리화학 기술	화학, 야금술	TEST	물리, 화학 등 과학기술 분야 전체 (전자, 유전공학 분야 제외)	1967
천문학 및 기하학	천문학, 기하학, 지리학, 광물학	NASA	천문학, 항공우주 공학	1988
농학 및 영양학	농학, 식물학, 임학, 영양학	INIS	물리학, 화학, 재료학, 지구과학, 생물학, 방사선학, 핵, 원자력 등	1987
생물의학	생물학, 의학, 수의학	MESH	의학, 생물학, 약학, 수의학	1997
종교 및 환경과학	환경, 도시공학, 종교, 교통	CIT	토목, 건설분야	1976
		EI	도시공학, 광물학, 기계공학, 전자공학, 화학공학, 엔지니어링 공학	1992
사회과학	사회학, 심리학, 교육학, 경영학, 정치학, 법학, 경제학, 군사학	ERIC	교육학	1990
		TSIT	건강, 결혼, 가족, 종교 등 사회학 분야	1989
문화 예술	문화, 역사, 인종학, 박물관학, 미술, 언어학, 음악, 스포츠	Women's	커뮤니케이션, 경제, 교육, 역사, 언어 문학, 법률학, 자연과학, 과학 기술, 사회과학 문화, 예술 등	1987

이들 14개 시소러스는 개별 주제 분야에서 용어수가 많고, 인지도가 있는 시소러스를 중심으로 선정하였다. 이중 범위가 넓은 일반영역에서는 UNESCO와 LCSH 시소러스, 수학 및 물리화학에서는 INSPEC과 JICST 시소러스, 종교 및 환경과학에서는 CIT와 EI 시소러스, 사회과학에서는 ERIC과 TSIT 시소러스와 같이 각각 2개 시소러스를 분석대상으로 하였다.

한편 국내 시소러스는 『과학기술용어 시소러스』(시스템공학연구소, 1992) 『신문기사 시소러스』(한국언론연구원, 1993), 『국방과학기술』(국방과학연구소, 1994), 『주제명요어집』(국립중앙도서관, 1994), 『한국통신 시소러스』(한국통신연구개발원, 1995), 『법률시소러스』(법원도서관, 1998)의 6개를 분석대상으로 삼았다. 이는 국내에서 개발된 시소러스 중 ERIC을 번역한 『KEDI Thesaurus』와 온라인 시소러스인 중앙일보사의 『JOINS 시소러스』를 제외한 전체를 포함한 것이다.

2) 분석범위

시소러스에서 사용하는 용어관계는 여러 가지로 구분할 수 있다. 우선 거시적 수준과 미시

적 수준의 두 가지로 대별할 수 있다.⁹⁾ 첫 번째는 거시적 레벨(macro level)로 각 용어간의 의미적 관계인 등가관계, 계층관계, 연관관계라는 기본적인 관계이며, 두 번째는 미시적 레벨(micro level)로, 등가관계에 속한 용어의 집합과 계층관계에 있는 용어, 연관관계에 있는 용어의 범주의 상호관계 및 이러한 사항과 주제분야 전체와의 관계이다.

다음으로는 기본관계와 부가관계로 나눌 수 있다. 등가관계, 계층관계, 연관관계를 시소러스의 기본관계로 정의하고, 범위주기를 포함하여 한정어 사용 등의 나머지 관계를 시소러스의 부가관계로 구분하는 방법이다.

본 연구에서는 등가관계, 계층관계, 연관관계를 시소러스의 기본관계로 정의하여 ISO 2788과 Z39.19의 시소러스 구축 지침과 기존 시소러스에서 사용하고 있는 용어관계를 비교분석하였다.

2. 등가관계

1) 관계기호

등가관계를 나타내는 관계기호는 ISO 2788에서와 대부분의 시소러스가 'USE'와 'UF(Used for)'를 사용하고 있다. 그러나 Z 39.19에서는 'USE' 대신 'U' 혹은 'SEE'라는 관계기호를 사용하고 있으며, 'SEE'의 상호참조인 'X(See from)'를 사용하고 있다.

ROOT 시소러스에서는 영어 약어 대신 '='(동의어, 유사동의어), '→'(USE)' 같은 기호를 사용해 이들을 지시하고 있으며, CIT 시소러스에서는 비우선어에서 우선어로의 안내는 'USE'를 사용하지만 우선어 참조는 'UF'를 사용하지 않는다. 그러나 계층관계 표시에서는 동의어는 '='로 나타낸다. 신문기사 시소러스에서는 우선어 참조에 대한 관계지시기호는 UF(used for)를 사용하고 있으나, 우선어 'use'대신 다음과 같이 '⇒'를 사용하고 있다. 또한 ROOT 시소러스에서는 각 용어관계에서 다른 주제 분야의 용어일 때는 '*'기호로 구분해서 표시하고 있다.

JICST 시소러스에서는 자모순 리스트에서 표목으로 사용되었으나 비디스크립터인 경우 '*'기호를 사용하고, UNESCO 시소러스에서는 시소러스 용어로 사용할 수 없는 개념은 용어 앞에 '+'기호를 부가하고 있다.

2) 포함 대상

동등관계 처리 대상으로는 ISO 2788과 Z39.19에서는 동의어와 유사동의어, 업워드포스팅으

9) J. Aitchison and A. Gilchrist. *Thesaurus Construction : A practical manual*. 2nd ed. London : Aslib, 1987, p. 34.

8 한국도서관·정보학회지 (제 30권 제 4호)

로 구분하고 동의어에는 어원이 상이한 용어, 속명과 학명, 통칭명과 상품명, 새로운 개념과 경쟁상태의 용어, 빈번히 사용되는 용어와 아닌 용어, 이형표기, 문화차이에 따른 용어, 약어와 완전어, 복합어의 분해형과 비분해형을 포함하고 있다.

그러나 Z 39.19에서는 동등관계에 반의어를 포함시키고 있으며, INSPEC 시소러스와 Women's 시소러스도 반의어를 동등관계에 포함시키고 있으며, LCSH와 TSIT 시소러스는 이 전용어를 동등관계에 포함하고 있다. Women's 시소러스는 별도의 우선어와 사용금지 용어 리스트를 갖고 있다.

3) 외국어 표시

디스크립터 용어의 외국어 표시를 위해서 한국통신 시소러스와 국방과학 시소러스는 한정어와 별도로 디스크립터에 해당하는 영어를 표기해 주고 있으며, 법률시소러스에서는 한자를 부기하고 있다. 또한 법률시소러스에서는 2개 국어 이상의 용어일 때 이를 구분하기 위해서 eng/ken(영어/한글용어), jap/kja(일어/한글용어), ger/kge(독일어/한글용어), fra/kfa(불어/대응 한글용어)와 같이 언어 구분을 해주고 있다.

<표 2> 시소러스의 등가관계

	관계지시기호	대 상	기 타
ISO표준	USE UF(Used for) +(복합어의 분해형)	동의어(어원이 상이한 용어, 속명과 학명, 통칭명과 상품명, 새로운 개념과 경쟁상태의 용어, 빈번히 사용 용어와 아닌 용어, 이형표기, 문화차이에 따른 용어, 약어와 완전어, 복합어의 분해형과 비분해형) 유사동의어 업워드 포스팅	우선어는 고딕체로 표시하도록 장려
Z 39.19	U 혹은 USE, SEE, UF(Used for) UFA(Used for...and) USE+(Use...and) X(See from; see의 상호참조)	ISO 표준과 같음 유사동의어에 반의어 포함	우선어는 고딕체로 표시하도록 장려
UNESCO	USE, UF (*) 시소러스 용어로 사용 불가 개념		우선어는 진하게
LCSH	USE(see) UF(x)	동의어, 철자, 표현형식, 양자택일 이전표목	표목은 진하게
ROOT	=(동의어, 유사동의어), →(use) =*(합성어에 대한 동의어)		우선어는 고딕체
INSPEC	USE, UF	일반적 참조어 철자, 약어, 도치, 상위어, 속어, 반의어	우선어는 진하게

	관계지시기호	대 상	기 타
JICST	USE, UF *(비디스크립터 표시)		표기에서는 우선어와 비우선어의 차이가 없음
TEST	USE,UF 2개 이상의디스크립터 사용지시(+)	유사동의어, 어순변화 최신용어지시	우선어는 진하게
NASA	USE, UF, 복합어 사용	동의어, 유의어, 약어	우선어는 진하게 용어길이 42자로 제한
INIS	USE, SEE UF(Used For), SF(Seen For)		우선어는 고딕체
MESH	use,see,x(역 see) conder also(개념의 언어학적 상관관계)		표목은 진한 고딕체
CTT	USE(비우선어에서 우선어 안내) =(동의어:계층관계에서 사용)	동의어, 약어, 두문자어 유사동의어, 철자통제	우선어는 진하게
EI	USE, UF USE AND, USE OR	정의 없음	글자표기에서는 우선어와 비우선어의 차이가 없음
ERIC	USE, UF #(2개 이상의 기본표목어 동등관계일 때 UF용어다음 부기	동의어, 변형된 형태 일반용어로 색인할 특정용어	우선어는 대문자, 고딕체
TSIT	Use, UF	동의어, 이형표기, 이전용어	우선어는 고딕체 60자
Women's 과학기술용어	USE, UF, (use/do not use 디스플레이)	동의어, 약어, 반의어	우선어는 진하게
신문기사	USE, UF	별도의 정의 없음	우선어는 고딕체, 진하게
국방과학기술	⇒(UF⇒USE), UF	별도의 정의 없음	우선어는 고딕체, 진하게
주제명요어집	USE, UF ST(specific term)	별도의 정의 없음	영어 부기
한국통신	USE, UF	ISO 표준과 같음	표기에서는 우선어와 비우선어의 차이가 없음,
법률시소러스	USE, UF eng/ken(영어/한글), jap/kja(일본어/한글) ger/kgel(독일어/한글), fra/kfr(불어/한글)	동의어, 유사동의어, 업워드포스팅, 다국어 관계	우선어와 비우선어의 차이가 없음, 영어 부기
법률시소러스	USE, UF(Used for) +(복합어의 분해형) eng/ken(영어/한글), jap/kja(일본어/한글) ger/kgel(독일어/한글), fra/kfr(불어/한글)	정의없음(보다 세분된 용어 안내)	우선어는 고딕체, 진하게

4) 속포스팅

국방과학기술 시소러스는 ST(specific term)기호를 두고 표제어가 디스크립터로는 너무 일반적이어서 색인이나 검색에 있어서 이 기호 뒤에 출현한 디스크립터 가운데서 적절한 용어를 디스크립터로 선택하도록 하고 있다.

유도

ST 관성유도

미사일유도

종말유도

여기에서 용어관계기호 'ST'가 지시하는 것은 '유도'라는 용어가 너무 일반적이므로 다른 용

어를 사용하도록 지시하는 방법이다. 이것은 업워드 포스팅의 반대개념이라 할 수 있다. 주제명요어집에서는 어떤 용어들이 너무 세분되는 경우에 클래스명과 그 멤버명 중에서 넓은 클래스명을 우선어로 하고 나머지는 유사동의어로 처리하는 이 유사동의어는 용어가 확장됨으로써 하위개념어로 바뀔 수 있는 속(屬)포스팅을 사용하고 있다. 이러한 시소러스의 등가관계를 정리하면 <표 2>와 같다.

3. 계층관계

1) 관계기호

계층관계는 용어의 특성에 따라 다양하게 나타날 수 있다. ISO 2788과 Z 39.19에서는 하위어(NT)는 속관계(NTG)와 전체부분관계(NTP), 사례관계(NTI)로 구분하고, 상위어(BT)는 포괄적인 하위어를 나타내는 속관계(BTP)와 부분 상위어인 전체부분관계(BTP), 사례관계(BTI)로 구분 표시해 줄 수 있다고 정의하고 있다. Z 39.19에서는 계층관계의 지시기호로는 BT1(1차 수준), BT2(2차 수준), BTn(n차 수준), NT1(1차 수준), NT2(2차 수준), NTn(n차 수준)와 같은 관계기호를 사용할 수도 있다.

ROOT 시소러스에서는 BT, NT대신 <, > 같이 단언어를 위해 Symbol 표시를 하고 있으며, NASA 시소러스는 상위어와 하위어 모두를 계층구조(GS)로 포함한다.

2) 최상위어와 고립어

어느 계층에도 속하지 않는 용어인 고립어에 대해서는 Z 39.19에서는 사용을 권장하고 있다. JICST 시소러스와 신문기사, 법률시소러스에서도 미판정어와 고립어를 사용하고 있다. 신문기사 시소러스에서는 고립어를 자모순 시소러스에서 배열하고 용어의 주제구분을 표시하고, 동등관계가 있을 경우에는 UF로 연결하고 있다. 또 이 용어는 분류별 계층색인과 분류별 자모순 색인에도 나열하고 있다.

또한 최상위어를 위하여 ISO 2788, Z 39.19에서는 TT(top term)를 사용하도록 했으며, INSPEC 시소러스와 UNESCO, 법률 시소러스에서 이를 사용하고 있다.

3) 패싯지시자

패싯지시자는 공통의 특성을 갖는 용어를 함께 그룹화하기 위하여 주제표시에서 사용한다. 동일계층에 속한 형제용어의 구별을 위해서 ISO 2788에서는 패싯지시자(facet indicator or

node label)를 사용할 수 있다고 제시하고 있는데, ROOT, UNESCO, CIT, 법률시소러스 등에서 사용하고 있다.

ROOT 시소러스에서는 공통의 특성을 갖는 용어를 함께 그룹화하기 위하여 주제표시표에서 사용하며 괄호 안에 이탤릭체로 표시한다. ROOT 시소러스에서 패킷은 엄격하게 적용되지는 않으나, 주제 분야의 논리적 배열을 지원하기 위한 유용한 도구로 사용하고 있다. 그러나 명백한 정의가 필요한 모호한 패킷이 많으며, 일관성이 없거나 애매한 것도 많이 있다.¹⁰⁾

UNESCO와 CIT 시소러스도 패킷을 사용하고 있다. UNESCO 시소러스는 분류 시소러스 내에서 각 Class에서 Subclass가 식별되지 않는 곳에서 용어들을 그룹화 하는 데 도움을 주기 위해서 사용한다. CIT 시소러스에서는 계층관계에서 BT, NT 대신 패킷지시자로 계층관계를 구분하고 있다. 유개념에서 종개념의 구분은 패킷지시자의 이용으로 개념이 명백해 진다.

4) 식별어

ERIC 시소러스에서는 식별어(identifier)를 사용하고 있는데, 시소러스에 나타나지 않는 개념 뿐만 아니라 지리적인 위치, 인명, 프로그램명, 법률명을 색인하기 위해 사용된다. NASA 시소러스도 숫자 혹은 알파벳 표시를 포함하는 주제 용어나 특별한 모델이나 아이템에 대한 용어는 정규 용어로 취급하고 완전한 상호참조를 주고 있다.

UNESCO 시소러스는 도시, 강, 산 과 같은 고유명사와 인명, 기구명, 프로젝트명 등은 시소러스 파일에는 포함되지 않고 별도의 식별어 리스트에 기록된다.

신문기사 시소러스에서는 지방자치 단체명으로 등록된 색인어는 정식명칭으로하며, 지역명은 도, 시, 군을 생략한 명칭으로 사용하고 있다. 해양, 강, 해협, 사막, 만, 평야, 반도, 호수, 섬, 산, 산맥 등의 고유명사를 시소러스 용어로 사용하고 있으나 고유명사임을 나타내는 지시 기호는 사용하지 않고 있다.

5) 계층표시

대부분의 시소러스에서 계층표시는 알파벳순 시소러스에서는 특정 디스크립터 다음에 BT와 NT를 사용하여 계층관계를 나타내고, 계층순 시소러스 디스플레이를 갖는 시소러스에서는 점(·)으로 나타내며, 점의 갯수가 계층관계의 특정성을 나타낸다.

그러나 Women's 시소러스의 경우는 계층관계에서 알파벳순 표시에서 다른 시소러스처럼 BT와 NT를 사용하고 있지만, 계층표시에서는 기준용어의 위아래로 상위개념과 하위개념을

10) Blaise, Nkwenti-Azeh. "The use of thesaural facets and definitions for the representation of knowledge structure." In *Knowledge Organization and Quality Management*. Frankfurt : Indeks Verlag, 1994, p. 379.

12 한국도서관·정보학회지(제 30권 제 4호)

나타내고 있다. 특히 이 경우 상위개념에는 세미콜론(:)으로 하위개념에는 피리어드(.)의 수로 계층관계를 나타내고 있다. 이 디스플레이에서 각각의 상위항목은 한 행 앞으로 나오고 하위 항목은 뒤쪽에 위치한다.

```

: women
:: workers
::: occupations
:: professional occupations
: religious workers
:: occupations
: female intensive occupations
  women religious —main term
    . clergywomen
    . nuns
    .. mother superiors
    
```

이 양방향 계층구조 배열은 ERIC 시소러스에도 같은 방법으로 사용하고 있다.

INIS 시소러스는 계층관계 지시기호를 BT, NT 이외에 이들 관계의 상하위 계층을, 용어관계 기호를 하위어를 NT1, NT2, NT3, · · · NTn으로 구분하여 배열하고, 상위어도 같은 방법을 사용한다.

계층순 시소러스에서 ‘’은 하나의 계층을 의미한다. ‘’이 부여되지 않은 디스크립터는 해당 용어군의 최상위어를 의미한다. 국내에서 개발된 법률, 국방과학, 과학기술, 한국통신, 신문기사 시소러스에서는 계층관계를 점의 수로 나타내고 있다. 또 INIS 시소러스와 ROOT 시소러스에서는 하위어 접기방법을 사용하고 있다. 이를 정리하면 <표 3>과 같다.

<표 3> 시소러스의 계층관계

	관 계 기 호	대 상	패싯지시자	기 타
ISO 2788	TT(Top term) BT(Broader term) BTG(Broader term generic) BTP(Broader term partitive) BTI(Broader term instance) NT(Narrow term generic) NIG, NIP, NII	속관계 전체-부분관계(실체의 조직과 기관, 지리적 위치, 학문분야 또는 연구영역, 계층적 사회구조) 사례관계 다중계층관계	node label 사용할 수 있다	
Z 39.19	TT(Top term), 고립어 GS(Generic structure) BT(BTG, BTP, BTI) NT(NIG, NIP, NII)	ISO 표준과 동일	node label 사용할 수 있다.	복합적인 전체의 부분은 연관관계에 포함

	관계 기호	대 상	패킷지시자	기 타
UNESCO	TT(최상위어), NI, BT Identifier(식별어)	부분/전체관계	node label 사용	*(하위어 접기) 최대 5계층까지 허용
LCSH	BT, NT 미확정세목	상하위 개념		
ROOT	< >동일부분 상하위어 * < * > 다른부분 상하위어	별도의 정의 없음	node label 사용	다언어 위해 symbol 표시 최대 5계층까지 허용
INSPEC	TT(최상위어) BT, NT	별도의 정의 없음		NISO의 용어는 알파벳순 배열
JICST	BT, NI, · (root 어), 고립어			
TEST	BT, NI	상하위 개념		-(하위계층 접기)
NASA	GS(일반적구조: BT, NI 포함) Identifier(식별어)	상하위 개념		α(세분된 용어 사용 지시) GS(최상위어 역할을 함)
INIS	BT1,2,3,4 NT1,2,3,4	상하위 개념		*(하위어 접기) 최대 7계층까지
MESH	see under 상위어 xu(역 see under)		facet 사용	+(하위어 접기) 계층분류코드
CIT	BT, NI 사용 않고 계층배열만	속관계와 전체/부분관계	node label 사용해서 계층관계 구분	다중계층구조 허용 (*로 우선 위치 지시)
EI	BT, NI	계층관계		
ERIC	NI(알파벳순) BT(:: 계층표시), Identifier(식별어) (Invalid Descriptors)	별도의 정의 없음		(2개 이상 BT를 가질수 있음 : #기호로 표시)
TSIT	BT, NI	상하위 개념		
Women's	BT, NI(알파벳 표시) .: (계층표시)	상하위 개념		계층표시 상하위 5계층
과학기술용어	BT, NI, HE	상하위개념		최대 8계층까지 허용 ...으로 계층관계 나타냄
신문기사	BT, NI, 고립어	상하위개념		최대 6계층까지 허용
국방과학기술	BT, NI	상하위개념		최대 8계층까지 허용 ...으로 계층관계 나타냄
주제명요어집	BT, NI, 고립어	상하위 개념		별도의 계층간계 표시 없음
한국통신	BT, NI, 최상위어	상하위개념을 논리적으로 전개한 관계, 전체/부분 관계		최대 6계층까지 허용 ...으로 계층관계 나타냄
법률시소러스	TT, BT(BTG, BIP, BII) NI(NIG, NIP, NII) 고립어, 미관정어	별도의 정의 없음	node label 사용	...으로 계층관계 나타냄

4. 연관관계

1) 지시기호

ISO 2788과 Z 39.19, 그리고 대부분의 시소러스에서는 연관관계를 나타내는 지시기호로 RT(related term)를 사용하고 있다. 그러나 ROOT 시소러스는 -(관련어-동일부분), *(관련어 다른 부분)과 같이 UNESCO 시소러스는 * 기호로 나타내고 있다. 이외에도 CIT 시소러스는 관련용어를 분류기호로 나타내고 있으며, INSPEC 시소러스와 LCSH는 연관관계이외에 SA(일반 참조) 기호를 사용하고 있다. Z 39.19에서는 관련어에도 node label을 사용할 수 있다고 정의하고 있다.

2) 포함 대상

ISO 2788과 Z 39.19에서는 연관관계에 속하는 대상을 동일범주에 속하는 용어와 상이한 범주에 속하는 용어로 구분하고, 전자에는, ① 자매어, ②가족적인 혹은 파생적인 관계를, 후자에는, ① 학문분야, 연구영역, 연구의 대상과 현상, ② 조작, 과정과 행위자 또는 도구, ③ 행위와 그 행위의 결과, ④ 행위와 그 수동체, ⑤ 특성에 관한 개념, ⑥ 기원에 관한 개념, ⑦ 인과관계에 의해 연결된 개념, ⑧ 사물과 그것의 반대 행위자, ⑨ 개념과 그것의 측정단위를 포함시켰다.

그러나 대부분의 시소러스에서 비슷한 내용을 포함하고 있지만, ERIC 시소러스에서는 유사 동의어를 RT로 보고 있으며, JICST에서는 반의어를 RT에 포함시키고 있다. 그러나 Women's 시소러스에서는 반의어는 연관관계에 포함시키지 않는다고 정의하고 있다.

국내 시소러스의 경우에는 주제명 검색어집에서만 RT(연관관계)와는 별도로 SA(도보라 참조)를 두고 있다. 참조되는 검색어가 예시주제명, 생략주제명인 경우에 사용하는데, 관련 용어에서 동일한 명사로 시작하거나 끝나는 용어가 다수 있는 경우에는 연결 참조의 간소화와 사용기관의 편의를 위해 사용하고 있다.

동물(動物) 604.49 ; 668.49

RT 동물사진

동물원

SA 기타 동물의 각 종속명

기타 개개의 동물명

기타 동물이라는 말로 시작하거나 끝나는 주제명요어

이상의 시소러스 연관관계를 종합하여 표로 나타내면 다음 <표 4>와 같다.

〈표 4〉 시소러스의 연관관계

	연관관계	대 상	기 타
ISO표준	RT(Related term)	동일범주에 속하는 용어(자매어, 가족적인 혹은 파생적인 관계) 상이한 범주에 속하는 용어(학문분야, 연구영역, 구의 대상과 현상, 조작, 과정과 행위자 또는 도구, 행위와 그 행위의 결과, 행위와 그 수단체, 특성에 관한 개념 기원에 관한 개념, 인과관계에 의해 연결된 개념, 사물과 그것의 반대 행위자, 개념과 그것의 측정단위)	node label 사용
Z 39.19	RT(Related term) 관련어 Node label 사용	ISO 표준과 같은 진정한 상위개념어가 아닌 명사를 가진 구	node label 사용
UNESCO	*(분류 시소러스) RT(알파벳 시소러스)	부분 처리관계 사물속성관계	
LCSH	RT(,xx) SA(일반참조)	SA(같은 단어로 시작하는 모든 표목의 그룹을 참조, 세구분 안내)	
ROOT	-관련어(동일부분) *관련어(다른부분)	정의 없음	
INSPEC	RT SA(도보라)	부분-전체관계, 대상-성질/현상 대상-과정, 대상-적용 성질/현상-과정, 성질/현상-작용	
JICST	RT(반의어 포함)	부분/전체관계, 종속관계, 반의어, 수단/목적, 원인/결과, 작용/생산물, 상이한 범주에 속하는 유사용어, 사물/특성	
TEST	RT, 범위주기에 see also도 나타냄	상이한 계층이나 의미 또는 개념이 밀접한 것 거의 동의어에 가까운 것 관점에 따라 사용상 상호관계가 있는 것 상호 부분전체의 관계를 갖는 것	
NASA	RT	2용어가 개념적으로 밀접한 관계가 있지만 상하위 개념에 속하지않는 것	
INIS	RT	관련성이 있는 개념관계	
MESH	See reated, (xR)		
CIT	분류기호	종속관계, 전체/부분관계	
EI	RT	비계층관계의 관련된 개념	
ERIC	RT(알파벳순 배열)	부분-전체, 유사동의어, 개념상 관련용어	
TSIT	RT	BT, NT에 속하지 않지만 관련이 있는 용어	
Women's	RT	관련용어	반의어는 포함하지 않음
과학기술용어	RT	도입어와 표시어는 상호연관관계 용어	
신문기사	RT	관련어	
국방과학기술	RT	표제어와 상호연관이 있음을 표시하는 대등기호	
주제명요어집	RT SA(도보라 참조)	관련개념	
한국통신	RT	동등관계, 계층관계에는 속하지 않지만 디스크립터와 관련성을 가진 용어	
법률	RT	관련어	

5. 용어관계의 문제점

1) 동등관계

(1) 지시기호

대부분의 시소러스는 동의어 처리에서 우선어를 선정하고 나머지는 'Used For'로 정의한다. 그러나 최종 이용자 시소러스에서는 모든 용어는 이용할 수 있고 사실상 사용되기 때문에 이러한 방법은 부당하다. 이것은 최종이용자 시소러스에서보다는 색인자 시소러스에 필요하다.¹¹⁾ 그러므로 동등관계의 경우 단순히 'Use'와 'Use for'로만 정의할 것이 아니라 이들을 세분하여 시소러스를 이용하는 색인자와 검색자에게 또는 색인, 검색시스템에 도움을 주어야 할 것이다.

또한 반의어의 경우 Z 39.19는 반의어를 유사동의어로 취급해 동등관계에 포함시키고 있으며, INSPEC 시소러스와 Women's 시소러스도 동등관계에 포함시키고 있다. ERIC 시소러스에서는 반의어를 연관관계에 포함시키는 등 다양하게 다루어지고 있다.

예를 들어 신문기사 시소러스에서는 반의어를 아래와 같이 연관관계로 취급하고 연관관계로 포함하여야할 흡연을 'UF'로 취급하였다.

흡연(吸煙)
DB 건강
UF 흡연을
NT 여성흡연
청소년흡연
RT 금연

그러나 엄밀히 말하면 반의어는 등가관계가 아니기 때문에 동등관계로 취급해서는 안된다.

(2) 외래어 표기

시소러스의 종류를 대상이 되는 국어의 수에 따라 구분하면 단일어 시소러스와 다국어 시소러스로 나눌 수 있다. 현재 사용되고 있는 대부분의 시소러스는 단일어 시소러스이다. 그러나 단일어 시소러스일 경우에도 외래어는 시소러스 용어로 사용되어야 한다.

국방과학기술 시소러스에서는 일반적으로 외래어는 원칙적으로 영어원음대로 한글로 표기하고, 이의 발음표시 상에 따른 해석의 오류를 방지하기 위해 ()한정어를 부가한다. 과학기술 용어 시소러스에서는 영어로 된 두문자어, 약어, 인명, 화합물명을 영어를 그대로 표기함을 원

11) J. D. Anderson, and F. D. Rowley. "Building end-user thesauri from full-text." In *Proceedings of the 2nd ASIS SIG/CR Classification Research Workshop*. Vol. II. Pittsburgh : Learned Information, Inc., 1992, p.8.

칙으로 하고 관용적으로 사용되는 번역어가 있을 경우에는 번역어로 표기한다.

그러나 법률 시소러스에서는 eng/ken, jap/kja, ger/kge, fra/kfa와 같이 언어구분을 해주고 있다. 법률 시소러스의 외국어 표시는 다음과 같다.

은행	
BT	공탁물보관자
NTi	한국은행
RT	은행발행수표
ENG	bank
FRA	banque
JAP	銀行

이 경우에 있어서 개별 국가 언어도 동등관계로 처리해야 함에도 불구하고 별도의 UF 지시 기호를 사용하지 않았으며, 한자를 일본어로 표기하고 있다. 단일 시소러스일 경우도 외국어 표시가 있을 경우 색인에서 효율성을 높일 수 있을 것이며, 검색한 용어의 정의와 더불어 외국어 표시가 있다면, 시소러스가 용어사전의 기능을 수행할 수 있을 것이다.

그러나 현재 시소러스 구축 국제 표준 지침이 영문위주로 만들어졌기 때문에 외국어 표기에 대한 지침이 전혀 마련되어 있지 않다. 그러므로 외래어를 동등관계로 취급하고, 개별용어에 대한 언어구분을 해 줄 필요가 있을 것이다.

(3) 약어 사용

대부분의 시소러스에서는 일반적으로 약어는 가능한 한 사용하지 않고 있다. ISO 2788에서도 약어와 두문자어는 시소러스가 다루고 있는 주제 내에서 넓게 쓰이거나 쉽게 이해되는 경우 외에는 우선어로 사용될 수 없다. 많은 약어와 두문자어가 하나 이상의 개념을 의미할 수 있으므로 그 완전한 형태의 단어가 용어 형태로부터 상호참조 되면서 우선어로 사용되며, 약어와 두문자어에서 완전한 형태의 용어가 자주 사용되지 않거나 거의 무시되는 경우 우선어로 이용될 수 있다. 물론 완전한 형태와 그의 약어 사이에는 상호참조가 필요하다고 규정하고 있지만, 이 둘의 관계를 명확히 하는 것은 쉽지 않다.

ISO 2788과 Z 39.19에서 동등관계로 약어를 정의하고 있지만 대부분의 시소러스에서 약어를 사용하고 있지 않으며, 약어와 원명의 구분기준도 모호한 경우가 많아 경우에 따라서는 약어가 디스크립터로 사용되기도 한다.

그러나 시소러스를 이용한 자동색인을 고려할 때 약어도 외래어와 마찬가지로 디스크립터에 대한 도입어로 선정되어야 할 것이며, 이를 위해서는 별도의 용어관계 표시가 필요할 것이다.

(4) 속어와 방언의 처리

ISO 2788에서는 특정 소집단이나 사회집단에서 새로운 개념의 용어가 생성되어, 그 용어를 표현할 대체어가 널리 사용되지 않을 경우 속어와 특수계층어가 색인어로 받아들여져야 하며, 속어와 특수계층어가 공존하면서 확립된 대체어로 나타난 경우 확립된 용어를 우선어로, 속어는 색인 이용자의 접근점으로 제공되어질 때만 비우선어로 쓰여진다고 정의하고 있다. 그러나 속어도 문헌에 많이 등장하기 때문에 시소러스를 이용한 자동색인과 용어통제를 이용한 검색에서 중요한 역할을 수행할 수 있기 때문에 속어도 시소러스 용어에 포함되어야 한다.

시소러스 구축에서 디스크립터 용어로 사용되는 용어는 표준어이다. 그러나 우리나라의 경우 표준말과 완전히 달리 표현되는 방언이 많이 존재하며, 특히 남북 분단 이후 북한어는 한글과 표기법에서 많은 차이가 나는 용어들이 존재하기 때문에 방언과 북한어 등도 시소러스 용어로 등록되어야 하며, 이들을 구분해 줄 수 있는 용어관계가 필요하다. 그러나 시소러스 구축 지침이나 시소러스에서는 이를 배제하고 있다.

2) 계층관계 및 연관관계

(1) 업워드 포스팅

업워드 포스팅은 하위어를 상위어 종이 아니라, 등가어로 취급하는 방법이다. 그 목적은 어휘의 수를 감소시킴과 동시에 하위개념인 특정용어에서 하위개념을 나타내기 위하여 이용하는 상위어에 액세스하기 위한 방법을 확보하는 데 있다. 예를 들어 '과일'의 경우는 다음과 같이 나타난다.

- 과일
- UF 사과
- 포도
- 복숭아
- 자두
- etc

이 경우 개별용어인 사과, 포도, 복숭아, 자두 등은 각각의 용어로 색인되는 것이 아니라 대표 용어인 과일로 색인이 된다.

ISO표준에서는 "업워드 포스팅은 주로 시소러스의 주변주제영역에 한하여 사용하고, 특정성이 매우 낮고 광범위한 주제를 대상으로 하고 있는 시소러스에서는 사용해도 좋다고 하고 있고, Z 39.19에서는 업워드 포스팅을 속 포스팅(generic posting)으로 지칭하고 있다.

그러나 ISO 2788에서도 업워드 포스팅을 언제 사용하며 사용할 경우 업워드 포스팅이 아닌 용어와 구별할 수 있는 방법이 마련되어 있지 않고 있다. 그러므로 업워드 포스팅으로 사용하

는 용어와 사용하지 않는 용어를 구별할 필요가 있을 것이다.

(2) 하위어 포괄검색

계층관계는 상위 및 하위개념을 나타낸다. 계층관계는 구조화되지 않은 일반용어와 시소러스를 구별하는 기본적인 특징으로 재현성과 적합성을 재고시키는 중요한 요소이다.

계층적으로 관계가 부여되는 것은 용어가 동일한 기본 범주에 속할 때 가능하며, 모든 하위 용어는 그것의 정의에 해당하는 상위용어를 언급해야 한다. 특히 시소러스는 색인작성단계에서 어휘통계를 적용하는 시스템에서만 한정되는 것이 아니라 색인하지 않은 데이터베이스(예를 들면, 자연언어데이터베이스)를 이용자가 탐색하는 경우에도 유용하기 때문에 탐색자가 하위어 포괄검색을 할 경우 동일한 수준의 하위어에서 여러 다른 속성을 나타내는 용어가 함께 처리되어 잡음이 많이 발생한다.

예를 들어 자동차의 경우 배기량에 따라 소형차, 중형차, 대형차 등으로, 용도에 따라 스포츠카, 경주용차 등으로, 제작사에 따라 현대자동차, 대우자동차 등의 여러 가지 방법으로 구분하여 효과적인 하위어 포괄검색을 할 수 있음에도 불구하고 대부분의 시소러스에서는 하위어(NT)로만 처리하고 있어 검색효율성에 문제가 있다.

이를 해결하기 위해서는 다양한 개념 범주 패킷을 이용, 관계지시기호를 확장하여 시소러스 관계지시기호를 표시하여 효과적인 검색이 가능하도록 할 수 있을 것이다.

(3) 패킷지시자 사용

디스크립터가 시소러스 내에서 계층관계로 정렬되었을 때 형제 디스크립터간을 구분하기 위해서 사용하는 것을 노드레이블(node label) 혹은 패킷지시자(facet indicator)라고 한다. 그러나 시소러스 구축 지침에서는 노드레이블을 사용할 수 있다고만 제시하고 있을 뿐 노드레이블의 종류나 유형을 언급하지 않고 있다. 또한 현재 이를 사용하고 있는 시소러스의 경우도 계층관계에서 노드 레이블만 제시할 뿐 실제 검색에서 사용하지 않고 있다. 특히 컴퓨터 작성 온라인 시소러스일 경우 용어 디스플레이와 계층관계 용어의 포괄검색과 관련용어 검색에 문제가 있다.

Elaine Svenonius는 “용어색인을 위한 패킷분석의 적용에 대한 분석에서 정확한 패킷 정의에 관한 색인 및 분류전문가들 사이에 동의가 부족하다는 것을 지적하였다. 또 패킷은 자주 잘못 정의되고 귀속관계(membership)에 대한 그들의 상태도 전문화되어 있지 않다는 것이다. 패킷은 주제영역에 대한 개념지도를 제공해야 하며, 개개용어에 대한 어의적, 개념적 배경도 제공해야 한다. 만약 패킷이 잘못 정의된다면 그들이 상호 배타적인지 결정하는 것이 어렵고 패킷 용어를 할당하기도 어려워진다”¹²⁾고 주장하였다.

시소러스 구축을 위한 국제 표준 지침에서 패싯 이용에 대한 강조를 하지 않은 이유는 명백하지 않지만 심지어 출판된 패싯 시소러스에서도 패싯분석에 대한 상세한 언급이 없다.¹³⁾ 결국 이러한 지침의 부족으로 인해 일부 시소러스를 제외한 대부분의 시소러스에서는 노드레이블의 중요성이나 필요성을 인정하면서도 실제로 사용하지 못하고 있는 결과를 초래하고 있다.

(4) 연관관계

연관관계를 나타내는 지시기호는 ROOT시소러스와 MESH를 제외한 나머지 시소러스가 RT를 사용하고 있다. 그러나 INSPEC, LCSH, TEST 시소러스와 주제명요어집에서는 RT 이외에 도보라(SA)참조 기호를 두고 있다.

그러나 실제 많은 시소러스에서 특정 용어와 연관관계를 갖는 용어가 너무 많이 나타나고 있지만, 모든 경우 관계지시기호는 'RT'한 가지로만 표시하고 있어 용어의 체계도라 할 수 있는 시소러스의 기능을 만족시키지 못하고 있다. 또한 ISO와 Z 39.19는 계층관계에서도 패싯지시자를 사용을 제시하고 있으나, 연관관계 용어에서 패싯지시자를 사용하고 있는 시소러스는 없다.

IV. 용어관계의 확장

시소러스 용어관계 비교분석에서 나타났듯이 시소러스 구축 지침이 기존 시소러스에서 사용하고 있는 용어관계들을 완전히 수용하지 못하고 있을 뿐만 아니라 관계 설정이 부족한 부분이 있다. 그러므로 정보환경 변화에 효과적으로 대응하고 정보의 축적이나 탐색과정에서 시소러스가 제 기능을 수행하기 위해서는 지금까지 책자형 시소러스 위주의 용어관계는 온라인 시소러스환경에 적합하도록 확장시킬 필요가 있다.

따라서 본 장에서는 기존 시소러스에서는 사용하고 있으나 ISO 2788에서 제시하고 있지 않은 용어관계를 재정의하여 새로운 용어관계 지시기호로 설정하고, 동등관계와 계층관계 및 연관관계를 확장하는 방안을 제시한다.

-
- 12) Elaine Svenonius. "Facet definition : a case study." *International Classification*, Vol.5, No.3, 1978, pp. 135-141.
- 13) Louise F. Spiteri. "The use of facet analysis in information retrieval thesauri : an examination of selected guidelines for thesaurus construction." *Cataloging and Classification Quarterly*, Vol.25, No.1, 1997, p. 26.

1. 새로운 용어관계 지시기호의 설정

1) AT(반의어)

ISO 2788에서 규정한 등가관계는 언어나, 문화의 차이로 야기하는 다양한 스펠링이나 단어 순서, 약자/두문자어/완전한 형태, 소위 “유사동의어” 같은 것들을 포함하고 있다. 그러나 유사동의어 관계로 취급한 상반관계나 보충관계도 시소러스 용어관계에 반의관계가 설정되지 않았기 때문에 동등관계로 설정했다. 그러므로 이들 모든 반의관계를 포함하는 용어관계 설정이 필요하다.

심지어 “literacy”와 “illiteracy”처럼 두 용어가 반의어일 경우도 등가관계에 포함된다. 규칙을 알고 있는 인간 이용자는 “동의어”의 여러 형태들을 구분할 수 있어 각 예를 적절히 다룰 수 있지만 자동시스템의 경우엔 철자가 다양하거나 두 문자어일 경우로 제한된다.¹⁴⁾ 그러므로 반의관계는 동등관계에 포함시켜서는 않된다.

INSPEC 시소러스에서는 반의어를 등가관계에 배치했고, TEST 시소러스 등 일부 시소러스는 반의어를 유사동의어로 취급 등가관계에 배치하고 JICST와 ERIC시소러스는 반의어를 연관관계에 두고 있는 등 시소러스에 따라 다양하게 나타나고 있다.

인지언어학에는 용어관계를 등위관계, 연어관계, 상하관계, 동의관계의 4가지로 구분하는데,¹⁵⁾ 이를 시소러스에 적용하면, 동의관계는 동등관계, 상하관계는 계층관계, 연어관계는 연관관계로 적용할 수 있다. 그러나 등위어와 대립어에 해당하는 반의 관계는 일반적으로 동등관계로 취급하고 있을 뿐 반의관계를 나타내는 관계지시기호를 사용하지 않고 있다.

그러므로 본 연구에서는 이 경우에 새로운 용어관계인 반의관계를 사용할 것을 제안하면서 관계지시기호를 반의어 AT(antonym term. 이하 AT로 약칭함)로 제안한다. 이 경우 색인을 있어서 보다 정확한 개념으로 색인을 할 수 있을 뿐만 아니라 검색에 서도 보다 정확한 개념을 검색할 수 있어 필요한 정보 검색에도 도움이 될 것이다.

2) OT(고립어)

Z 39.19에서는 어떤 다른 디스크립터와도 관계가 없는 디스크립터인 고립어(orphan term)로 등록되는 디스크립터는 없어야 한다고 제시하고 있지만, 실제로 시소러스에는 고립어로 등록되는 용어들이 많다. INSPEC 시소러스에서는 고립어는 RT로 처리하고 있지만 고립어 자체를 나타내는 지시기호는 없다.

14) Susan Jones. "A Thesaurus data model for an intelligent retrieval system." *Journal of Information Science*, Vol.19, No.3, 1993, p. 173.

15) 임지룡. 『인지의미론』. 서울 : 탐출판사, 1997, p. 308.

22 한국도서관·정보학회지(제 30권 제 4호)

시소러스는 용어들 간의 관계를 나타내는 통제어휘집이기 때문에 특정 용어와 관련되어 있는 용어가 아닐 경우 이러한 용어는 배제되는 경우가 대부분이다. 그러나 용어간의 관계를 가지는 않지만 색인으로 등록하지 않으면 검색어로 사용할 수 없기 때문에 이들 용어는 시소러스 용어로 등록해야 한다.

이 경우 고립어는 주제분야별로 자모순으로 배열하고 별도의 용어관계 기호를 부여해야 할 것이다. 이들에 관한 용어관계 지시기호를 OT(orphan term)로 제안한다.

3) UT(후보어)

“미확정 세목이란 일반적으로 표목의 표현형식과 구조의 일관성을 유지하기 위해서, 표목이나 세목은 책임 있는 기관의 승인을 필요로 하는 세목을 말한다. 이것은 표목이나 세목의 형식과 구조를 목록작성자나 색인자가 임의로 사용하였을 때 오는 혼란을 방지하고, 일관된 표목구조를 유지하기 위한 정책적 수단이다. 그래서 기존의 세목 이외의 세목을 사용해야 하는 경우, 이를 예외적으로 인정하고 이 때 사용할 수 있는 세목”¹⁶⁾을 말한다.

그러므로 시소러스 용어로 등록되지는 않았지만 앞으로 중요한 개념으로 사용될 가능성이 높은 용어를 중심으로 후보어 및 미확정 세목을 위한 용어관계를 제시해야 할 필요가 있을 것이다. 이를 위한 용어관계 기호로 UT(unsettled term)를 제안한다.

4) UP(upward posting)

전통적인 처리에서 비교적 총괄적인 용어에 대한 용어관계로 소위 업워드 포스팅(upward posting)이 있는데, 이러한 처리는 주변적인 주제분야의 경우 광범위한 분류를 함으로써 통제어휘의 규모를 유지하기 위해 의도되었다. 사실 우선어 관계에 대해 안내 할 경우 수작업 색인자는 문맥을 해석할 필요가 있는데, 이것을 인공적으로 재생산하는 것이 어려울 수 있다.¹⁷⁾ UNESCO 시소러스에서도 주제를 세부적으로 정교하게 만들지 못하는 곳과 특정 단어를 보다 상위의 개념으로 유도할 때 많이 사용하고 있다.

X90.89 Sport
UF Games
X83 Ball Games
UF Football
Golf
Tennis

16) 김주성. 『LC주제명표목표와 시소러스의 색인어구조 비교연구』. 석사학위논문, 연세대 대학원, 1995, p. 15.

17) Susan Jones. "A thesaurus data model for an intelligent retrieval system." *Journal of Information Science*, Vol.19, No.3, 1993, p. 173.

이 경우 Ball Games에 속하는 게임의 종류는 이외에도 무수히 많음에도 불구하고 이 세 가지 종류만을 UF로 정의하고 있다. 그러나 사실상 Football, Golf, Tennis는 Ball Game의 하위 개념이므로 개별 종목은 NT로 표시하고 Ball Game에 특정 기호를 추가해서 이들 세 용어는 하위개념으로 사용하고, 이들에 포함되지 않은 구기 운동은 상위개념으로 업워드포스팅시킬 수 있는 방안이 마련되어야 할 것이다.

Sport

- UF Games
- NT(UP) Ball Games
- NT Football
- Golf
- Tennis

업워드포스팅을 많이 사용하게 되면 시소러스 분량을 줄이는 데는 도움이 될 수 있으나 중요한 개별 용어를 시소러스 용어로 색인 할 수 없기 때문에 색인어의 특정성이 떨어지게 된다. 온라인 시소러스를 고려했을 때는 효과적인 방법이 되지 못한다. 또한 업워드포스팅에서 대상이 되는 용어를 나열하고 있어 하위개념이 많을 경우 일일이 나열해야 하므로 시소러스의 분량을 많이 차지하게 되는 문제점을 갖고 있다.

그러므로 특정한 개념에서 중요한 용어는 특정성 유지를 위해 하위개념으로 나열하지만, 중요하지 않은 용어는 업워드포스팅으로 만드는 방법을 모색할 필요가 있을 것이다. 이 경우에는 이러한 방법보다는 분류에서 특정분류번호는 '~을 포함한다'고 한정하는 것처럼 특정한 용어아래에 지시기호를 주어 더 이상 세분하지 않는 업워드 포스팅을 사용할 필요가 있을 것이다. 업워드 포스팅을 위한 용어관계로 UP(upward posting)를 제안한다.

2. 동등관계 지시기호의 확장

M. Willetts은 "시소러스의 동등관계는 막연하여 일관성이 없다. 동등관계와 연관관계구조에서 혼란되게 나타나기도 한다. 동등관계는 더 세분되게 정의하고 또 그들은 단지 하나의 상호참조구조에만 관련시켜 동등관계에 모두 포함시켜야 한다¹⁸⁾고 주장했다. 그러나 기존의 시소러스에서는 이들 동등관계에 속한 용어를 모두 'UF' 로만 처리하였는데, 이들을 구별하여 적절한 관계기호로 표시하면 검색자의 용어선정에 도움을 줄 수 있을 것이다.

18) M. Willetts. "An investigation of the nature of the relation between terms in thesauri." *Journal of Documentation*, Vol.31, No.3, 1975, p. 179.

1) UF(외래어)

단일어 시소러스일 경우라도 외국어는 시소러스 용어로 중요하다. 특히 한국어 시소러스를 고려할 때는 한자를 포함한 외국어가 중요한 기능을 수행한다.

그러므로 외국어를 동등관계(UF)로 처리하고 이들 각각에 대한 용어기호를 제시해 주는 것이 필요하다. 본 연구에서는 외국어 표시를 위한 용어관계기호로 EN(영어), JP(일어), GE(독일어), FR(불어), CH(한자)와 같이 사용할 것을 제안한다. 이 경우 동등관계는 UF(e), UF(j), UF(g), UF(f), UF(c)와 같이 나타날 것이다.

외국어의 입력결과는 다음과 같이 나타날 수 있다.

무조음악(無調音樂)

UF(e)	atonal music	영어
UF(j)	むちょうおんかぐ	일어
UF(g)	Atonale Musik	독어
UF(f)	musique atonale	불어
UF(c)	無調音樂	한자

또한 우리말에서 한자가 가지는 특성을 고려하거나 디스크립터에 대한 정보를 제공하기 위해서는 한자를 부기할 수도 있을 것이다. 이 경우 한자에 해당하는 UF(c)를 디스크립터에 참조정보로 부기하면 용어이해에 도움이 될 수 있을 것이다.

2) UF(방언)

우리나라 말 중에서 같은 개념을 나타내는 용어라도 각 지방마다 용어가 다양하게 나타나고 있기 때문에 시소러스 분야에 따라서는 방언이 중요한 용어로 사용될 수 있을 것이다. 또한 남북분단 이후 북한의 용어와 남한의 용어가 완전히 달리 사용되는 경우가 많기 때문에 통일 이후에 대비하기 위해서도 북한용어의 구분이 필요하다.

그러므로 방언을 동등관계 속에서 구분해 주기 위해 이들을 위한 용어관계가 필요하다. 이를 위한 용어관계기호를 DT(dialect term)으로 제안한다. 또한 외래어와 마찬가지로 'UF'를 확장하여 UF(d)나 UF(방)과 같이 나타낼 수 있을 것이다.

3) UF(이전용어)

INSPEC 시소러스에서는 현재는 디스크립터로 사용하지 않지만 이전(以前)에는 디스크립터로 사용되었던 용어를 지시하기 위해 이전용어(PT : prior term)를 사용하고 있다.

그러나 INSPEC 시소러스에서 사용하고 있는 PT는 인쇄형 시소러스에서 용어를 나열하는

데만 사용되고 있다. 결국 이전에 디스크립터로 사용된 용어와 지금 디스크립터로 사용된 용어는 동등관계에 있으므로 이들을 동등관계로 연결하고 이전 용어라는 표시를 해줄 필요가 있을 것이다. 그러므로 본 연구에서는 이전 용어를 PT(prior term)로 정하고 동등관계와 연결시켜 UF(p) 혹은 UF(전)과 같이 사용할 것을 제안한다.

4) UF(약어)

대부분의 시소러스는 디스크립터로 사용하는 용어는 완전어 사용을 원칙으로 하고 있고, 약어(略語)가 일반적으로 더 잘 알려진 경우에는 약어를 디스크립터로 사용할 수도 있다는 원칙을 가지고 있다. 이 경우 약어와 완전어 중 어떤 용어를 디스크립터로 사용할 것이지를 결정하기가 매우 어렵기 때문에 어느 한쪽을 디스크립터로 정하고 나머지 한쪽은 비디스크립터로 등록하는 것이 바람직 할 것이다.

어떤 용어가 디스크립터가 되고 어떤 용어가 비디스크립터로 되는가 하는 것은 온라인 시소러스에서는 큰 의미를 갖지 않기 때문이다. 그러므로 약어와 완전어의 경우에는 완전어를 디스크립터로 하고 약어는 용어관계 기호로 정의해줄 필요가 있다. 본 연구에서는 이를 위한 용어관계로 AB(abbreviation)로 제안한다. 이 경우도 동등관계와 연결시켜 UF(a) 혹은 UF(약)과 같이 표현할 수 있을 것이다.

5) UF(이형표기)

외래어가 우리나라에 들어와서 여러 가지의 형태로 표기되는 경우를 이형표기(異形表記)라고 한다. 외래어 맞춤법이 존재하고 있지만 여전히 외래어를 우리말로 표기하는데는 일관성이 지켜지고 있지 않다. 같은 용어가 달리 나타나는 이형표기 문제는 재현율을 저하시키는 요인이 된다. C. D. Paice는 이러한 문제를 해결하기 위해서는 시소러스 용어는 엄청난 수작업과 용어수의 증가에 따른 시소러스 부피문제가 있다 할지라도 모든 적당한 변형형태를 포함시켜야 한다¹⁹⁾고 주장했다. 그러므로 같은 개념이 다양하게 표기되는 용어들의 표기도 시소러스에 등록시켜야 한다. 본 연구에서는 이형 표기를 위한 용어관계기호로 UF(s)를 제안한다. 이외에도 필요에 따라 속명과 학명, 문화의 차이에 따른 용어 등을 구별해줄 필요가 있을 것이다.

19) C. D. Paice. "A thesaural model of information retrieval." *Information Processing & Management*, Vol.27, No.5, 1991, p. 444.

3. 패싯지시자의 사용

1) 패싯에 의한 검색

시소러스의 계층관계는 용어의 특정성을 표시하고, 용어간의 상하위 개념을 나타낸다. 그 결과 색인에서는 계층수준에 따라 가중치를 부여하여 자동색인의 효율성을 높이고, 검색에서는 개념의 확장과 축소를 통해 재현율과 정확률을 높이는 기능을 수행한다.

계층관계를 나타내는 용어관계는 BT와 NT이다. BT의 경우는 용어관계에 있어서 NT를 결정된 이후 상호관계로 설정되기 때문에 별 문제가 없으나 NT의 경우는 시소러스 구축 지침과 대부분의 시소러스에서는 하위개념의 경우 NT로만 나타내고 있어 동일 계층에 속한 속성이 다른 개념을 한꺼번에 검색하는 노이즈를 발생하는 문제를 가지고 있다.

이를 해결하기 위한 방안으로는 ISO 2788이나 Z 39.19에서는 패싯지시자를 사용할 것을 권장하고 있지만 실제 사용하고 있는 시소러스는 많지 않다. 패싯지시자를 사용하고 있는 시소러스는 ROOT와 UNESCO 시소러스가 대표적인 예이다. 그러나 이들 시소러스에서도 패싯지시자만 설정하고 있을 뿐 실제 검색에서는 이용되지 않는다.

패싯 시소러스를 개발한 Jean Aitchison과 Alan Gilchrist는 “패싯 시소러스란 하나의 알파벳 색인 또는 전체적인 알파벳 시소러스에 의해서 체계적인 선택이나 보충을 함으로써 패싯 분석의 발전에 포함되는 것이며, 색인용어를 디스플레이하는 이런 방식은 분리된 구조에 대해 지도를 그리는 것이 아니라 그들 스스로 패싯분류를 형성하는 것이다”²⁰⁾라고 정의하여 시소러스에서 패싯 도입의 중요성을 강조했다.

그러므로 시소러스의 계층관계에서 한 용어 아래 너무 많은 동위류가 만들어 질 때 기존의 키워드 검색 외에 패싯을 사용하여 이들을 분리하여 검색하면 정확률 향상에 도움을 주기 때문에 시소러스 용어관계에서 패싯지시자의 설정이 필요하다.

2) 패싯지시자의 설정

패싯을 구분하는 방법은 분류의 경우에는 학자에 따라서 다양하게 제시되었고, 시소러스에서도 분야에 따라서 다양하게 나타나고 있다.

이들 다양한 패싯 구분 항목들을 종합적으로 랑가나단이 제시한 PMEST를 기초로 ISO와 Z 39.19의 시소러스 지침, 콜론분류표 7판,²¹⁾ Shera와 Egan, Kyle, Grolrier,²²⁾ CRC(Classification

20) J. Aitchison and A. Gilchrist. *Thesaurus Construction : a practical manual*. 2nd ed. London : Aslib, 1987. p. 88.

21) S. R. Ranganathan. *Colon Classification*. 7nd ed. Mohinder Partap Sataja ed. New Delhi : Sterling Publishers Private Limited, 1993, p. 49.

22) B. C. Vickery. *Faceted Classification : a guide to construction and use of special schemes*. London: Aslib, 1960,

Research Group)²³⁾ Vickery²⁴⁾등의 여러 학자와 PLEXUS, EJC, CIT 등의 시소러스에서 사용하고 있는 정의를 종합해서 이를 26개의 범주로 확장하고, 이를 ROOT와 UNESCO 시소러스에서 사용하고 있는 패싯지시자를 분석하면 <표 5>와 같다.

<표 5> 패싯 구분 내용

구분1	구분2	포 함 내 용	ROOT	UNESCO	계	평균(수)
속성	속성	Characteristics, Properties, Attribute, Quality, Origin	74	21	95	47.5
분야	분야	Subject, Problem, Field	4	31	35	17.5
부분	부분	Part, Accessory, Group	69	8	77	38.5
형태	형태	Type, Form, Design, Mode, Style, Shape, Configuration	161	63	224	112
	크기	Size, Number, Scale	17	3	20	10
	수준	Label, Grade	8	4	12	6
조건	조건	Ability, Cause, Dimension, Phase	9	2	11	5.5
조직	조직	Organization, Association, Assembly	16	12	28	14
	위치	Order, Stage, Class, Position, Order	30	3	33	16.5
상태	상태	Status, Coverage, Institution, Condition, Structure, Degree	27	22	49	24.5
	현상	Phenomena, Environment,	10	5	15	7.5
재료	재료	Resources, Material, Source	114	2	116	58
	설비	Equipment, Facilities, System, Device	107	32	139	69.5
도구	도구	Instrument, Tool	52	4	56	28
구성 요소	구성 요소	Element, Factor, Element, Composition, Constituent, Compound, Family, Substance	107	16	123	61.5
행위	행위	Activities, Action, Behaviour, Acting, Distribution	41	24	65	32.5
행위자	행위자	Personnel, Ownership, Agency, Agent	18	26	44	22
목적	목적	Policy, Purpose, Planning	55	29	84	42
대상	대상	Age group	4	2	6	3
수령인	수령인	User, Consumer	8	9	17	8.5
과정	과정	Process, Research, Cooperation, Analysis, Procedure	35	25	60	30
방법	방법	Technics, Method, Theory	47	26	73	36.5
	수단	Program, Means	6	9	15	7.5
적용	적용	Application, Administration, Development, Management, Arrangement, Treatment	69	32	111	50.5
작용	작용	Operation, Industry, Movement, Function, Performance	48	23	71	35.5
상호작용	상호작용	Relationship, Network	10	6	16	8
이용	이용	Service, Use	12	3	15	7.5
결과	결과	Result, Effect	6	4	10	5
생산물	생산물	Production, Product, Construction, Manufacture	68	11	79	39.5
측정	측정	Measure, Measurement	12	6	18	9
장소	장소	Location, Place, Country, Station, Area	30	9	39	19.5
시간	시간	Time, History, Period, Date, Frequency, Current	16	10	26	13
계			1,290	482	1,772	

p. 24.

23) J. D. Anderson and F. D. Rowley. "Building end-user thesauri from full-text." In *Proceedings of the 2nd ASIS SIG/CR Classification Research Workshop*. Vol. II. Pittsburgh : Learned Information, Inc., 1992, p. 3.

24) B. C. Vickery. "Knowledge representation : a brief review." *Journal of Documentation*, Vol.42, No.3, 1986, pp. 157-158.

28 한국도서관·정보학회지 (제 30권 제 4호)

이들 패싯지시자 중 재료와 설비를 분리하고, 상태에서 상태와 현상, 조직에서 조직과 위치, 방법에서 방법과 수단은 통합시켜 패싯지시자를 구성 할 수 있을 것이다. 아울러 대상과 수량인, 결과와 생산물 통합하였다.

이를 종합해서 본 연구에서는 속성(a), 분야(b), 부분(c), 형태(d), 크기, 수준(e), 조건(f), 조직, 위치(g), 상태, 현상(h), 재료(i), 설비(j), 도구(k), 구성요소(l), 행위(m), 행위자(n), 목적(o), 대상, 수량인(p), 과정(q), 방법, 수단(r), 적용(s), 작용(t), 상호작용(u), 이용(v), 결과, 생산물(w), 측정(x), 장소(y), 시간(z)의 24개로 제시한다.

〈표 6〉 패싯지시자 설정

구분	패싯지시자	포 함 내 용
속성	속성(A)	Properties, Characteristics, Attribute, Quality, Origin
분야	분야(B)	Subject, Problem, Field
부분	부분(C)	Part, Accessory, Group
형태	형태(D)	Type, Form, Design, Mode, Style, Shape, Configuration
	크기,수준(E)	Size, Number, Scale, Label, Grade
조건	조건(F)	Ability, Cause, Dimension, Phase
조직	조직,위치(G)	Organization, Association, Assembly, Order, Stage, Class, Position, Order
상태	상태,현상(H)	Environment, Status, Coverage, Institution, Condition, Structure, Degree, Format, Performance, Phenomena, Environment
재료	재료(I)	Resources, Material, Source
	설비(J)	Equipment, Facilities, System, Device
도구	도구(K)	Instrument, Tool
구성요소	구성요소(L)	Element, Factor, Element, Composition, Constituent, Compound, Family, Substance
행위	행위(M)	Activities, Action, Behaviour, Acting, Motivation, Distribution
행위자	행위자(N)	Agency, Agent, Personnel, Ownership
목적	목적(O)	Purpose, Policy, Planning
대상 수량인	대상,수량인(P)	User, Age Group, Consumer
과정	과정(Q)	Process, Research, Cooperation, Analysis, Procedure
방법	방법, 수단(R)	Method, Technics, Theory, Program, Means
적용	적용(S)	Application, Administration, Development, Management, Arrangement, Treatment
작용	작용(T)	Operation, Automation, Industry, Movement, Function, Performance
상호작용	상호작용(U)	Relationship, Network
이용	이용(V)	Use, Service,
결과 생산물	결과,생산물(W)	Product, Production, Result, Effect, Construction, Manufacture
측정	측정(X)	Measure, Measurement
장소	장소(Y)	Location, Place, Country, Station, Area
시간	시간(Z)	Time, History, Period, Date, Frequency, Current

V. 결 론

본 연구는 시소러스 구축 국제 표준 지침과 기존의 시소러스에서 사용하고 있는 용어관계의 한계와 문제점을 도출하고, 새로운 환경에 맞는 새로운 용어관계의 확장 방안을 제시하였다. 그 내용을 요약하면 다음과 같다.

첫째, 기존의 시소러스구축 국제 표준 지침인 ISO 2788에서 제시하고 있지 않은 새로운 용어관계로 반의어관계(AT), 고립어(OT), 후보어 및 미확정 세목(UT), 업워드포스팅(UP)을 사용할 것을 제안한다.

둘째, 동일 개념이 다양한 용어로 표현되는 동등관계는 하나의 디스크립터로 모으기 위해 외국어는 영어(en), 일어(jp), 독일어(ge), 불어(fr), 한자(ch)와 같이 구분하고, 방언(dt), 이전용어(pt), 약어(ab), 이형표기(sp)로 구분한다. 또 이들을 동등관계(UF)로 확장하기 위해서는 UF(e), UF(j), UF(g), UF(f), UF(c), UF(p), UF(a), UF(s)와 같이 사용할 것을 제안한다.

세째, 계층관계와 관련관계를 확대하는 방안으로는 동일 계층에 속한 동위개념을 구분하기 위해 사용하는 패싯지시자를 사용할 것을 주장하고 이를 위한 24개 패싯을 제시했다.

참 고 문 헌

- 安倍浩二 “シソ-ラスの問題點.” *Library and Information Science*, No.14(1976), pp. 71-94.
- 임지룡. 『인지의미론』. 서울 : 탐출판사, 1997.
- 최석두. “매크로시소러스에서의 용어관리.” 《제1회 전문용어 언어공학 심포지움》. 전문용어 언어공학연구센터, 1998, pp. 43-53.
- Aitchison, J. and A. Gilchrist. *Thesaurus Construction : a practical manual*, 2nd ed. London : Aslib, 1987.
- Anderson, J. D. and F. D. Rowley. "Building end-user thesauri from full-text." In *Proceedings of the 2nd ASIS SIG/CR Classification Research Workshop*. Vol. II. Pittsburgh : Learned Information, Inc., 1992, pp. 1-13.
- Foskett, A. C. *The Subject Approach to Information*. 5th ed. London : Library Association Publishing, 1996.
- GID. *Thesaurus Guide : analytical directory of selected vocabularies for information retrieval*. Amsterdam : North-Holland, 1985.

- Iyer, Hemalata. *Classificatory Structures Concepts, Relations and Representation*. Frankfurt : Indeks Verlag, 1995.
- Jones, Susan and others. "Interactive thesaurus navigation : intelligence rules OK?" *JASIS*, Vol. 46, No.1, 1995, pp. 52-59
- Jones, Susan, "A thesaurus data model for an intelligent retrieval system." *Journal of Information Science*, Vol.19, No.3, 1993, pp.167-178
- Lancaster, F. W. *Vocabulary Control for Information Retrieval*. 2nd ed. Arlington : Information Resources Press, 1986.
- Nkwenti-Azeh, Blaise. "The use of thesaural facets and definitions for the representation of knowledge structure." In *Knowledge Organization and Quality Management*. Frankfurt : Indeks Verlag, 1994, pp. 374-381.
- Paice, C. D. "A thesaural model of information retrieval." *Information Processing & Management*. Vol.27, No.5, 1991, pp. 433-447.
- Ranganathan, S. R. *Colon Classification*. 7nd ed. Mohinder Partap Sataja ed. New Delhi : Sterling Publishers Private Limited, 1993.
- Rowley, J. "The controlled versus natural indexing languages debate revisited : a perspective on information retrieval practice and research." *Journal of Information Science*, Vol.20, No.2, 1994, pp. 108-119.
- Stern, A. and N. Rischette. "On the construction of a super thesaurus based on existing thesaurus." In *Tools for Knowledge Organization and the Human Interface*. Frankfurt : Indeks Verlag, 1991, pp. 134-144.
- Vickery, B. C. *Faceted Classification : a guide to construction and use of special schemes*. London : Asilb, 1960.
- Weinberg, B. H. "Library classification and information retrieval thesauri : Comparison and contrast." *Cataloging and Classification Quarterly*, Vol.19, No.3/4, 1995, pp. 23-44.
- Willetts, M. "An Investigation of the nature of the relation between terms in thesauri." *Journal of Documentation*, Vol.31 No.3, 1975, pp. 158-184.