

시소러스의 부가관계에 관한 연구

A Study on Added-Term Relationship of Thesaurus

한상길(Sang-Kil Han)*

목 차

1 서 론	3.3 부가관계의 문제점
2 시소러스의 용어관계와 부가관계확장의 필요성	4 부가관계의 확장
2.1 시소러스의 용어관계	4.1 범위주기
2.2 부가 용어관계확장의 필요성	4.2 한정어
3 기존 시소러스의 부가관계 분석	4.3 용어 관련정보
3.1 분석대상	5 결 론
3.2 부가관계 분석	

초 록

이 연구의 목적은 새로운 정보검색 환경에 맞는 시소러스 부가관계의 확장 방안을 제시하는데 있다. 이를 위해 본 연구는 ISO 2788과 ANSI/NISO Z 39.19 기준, 그리고 현재 사용되고 있는 20개 시소러스의 부가용어관계를 비교 분석하여 문제점과 한계점을 도출하고, 새로운 정보검색환경에 맞는 부가관계 확장방안을 제시하였다. 이러한 문제점을 해결한 용어관계의 확장방안은 다음과 같다. 첫째, 범위주기에서 정의를 별도로 분리하고, 범위주기를 유형별로 구분했다. 둘째, 한정어 사용에 대한 원칙을 제시하였다. 셋째, 디스크립터 용어 정보의 확장방안을 제안했다.

ABSTRACT

The purpose of this study is to present solutions to expanding added-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 added-term relations currently used to find out problems and limitations. Based on findings of the study, this report suggests how to expand thesaurus added-term relations to accommodate changes in information retrieval environment. In order to solve those problems, this report presents solutions to expanding added-term relationships as follows: First, definition was separated from scope note and scope note is categorized by type. Second, principles for the use of qualifiers are presented. Third, principles for expanding of descriptor term information are presented.

키워드: 시소러스, 용어관계, 부가관계, 범위주기, 한정어

* 대림대학 문헌정보과 조교수

■ 논문 접수일 : 2000년 5월 18일

1 서론

시소러스란 문헌정보의 축적과 검색에서 색인작성자와 검색자가 사용하는 용어를 표준화된 어휘로 통일한 어휘집으로, 용어간의 개념관계를 동등관계, 계층관계, 연관관계 등의 측면으로 결합하여 체계적으로 배열해 놓은 용어통제표이다. 시소러스는 전통적인 도서관에서 주제명목표를 위해 편찬한 「주제명표목표」의 개념에서 시작하였다. 그러나 정보환경의 변화에 따라 컴퓨터 정보검색을 위한 정보생산자, 입력자, 이용자 사이에 용어통일의 역할뿐만 아니라 용어에 대한 정보를 제공하는 용어사전의 역할까지를 수행하는 용어 데이터베이스로 그 개념이 바뀌고 있다.

정보기술의 발달에 따라 정보 제공이 온라인 환경이 보편화되고 정보이용자들의 이용환경도 Windows를 기본으로 하는 GUI(graphic user interface) 환경으로 바뀜에 따라 기존의 인쇄형 시소러스에서 사용할 수 없었던 다양한 용어관계의 새로운 설정이 가능하게 되었다. 온라인 환경에서의 시소러스는 단순히 색인과 검색에서 용어통제 역할만을 수행하는 것이 아니다. 시소러스는 색인에서의 동의어 통제와 더불어 개념의 축소나 확장검색의 도구로써 사용할 수 있어야 한다. 또한 시소러스가 개념의 체계도로서의 역할을 수행하기 위해서는 검색자에게 다양한 용어정보를 제공해야 한다. Schmitz-Esser(1991)는 시소러스가 도서관과 문서보관소의 좁은 분야를 넘어서, 언어와 지식공학과 같은 보다 넓은 영역으로 영향을 미칠 것이며, 그렇게 될 때 그 모양이나 형태뿐만 아니라 심지어 명칭도 변할 것이다. 이것은 새로운 용어관계 형성을 특징으로 하며 탐색자나 색인자가

직접하기 보다는 점점 기계가 수행할 것이며, 그들은 새로운 종류의 다른 기술과 합쳐서 나타날 것이라고 주장하며 정보검색 환경변화에 따른 시소러스의 기능 확대와 형태 변화를 예고하였다.

그러므로 시소러스의 이러한 기능을 만족시키기 위해서는 현재 사용하고 있는 용어관계보다는 더 다양한 용어관계의 설정이 요구되고 있다.

그럼에도 불구하고 기존 시소러스의 작성지침으로 사용되고 있는 국제 표준인 ISO 2788-1986(E) (*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 정보검색환경의 변화에 부응하지 못하고 있다. 특히 시소러스의 부가용어관계는 탐색자에게 디스크립터를 안내하는 중요한 용어관계로 다양한 용어관계를 지시할 수 있음에도 불구하고 범위주기와 한정어 사용만으로 제한하고 있어 시소러스를 이용하는 이용자에게 보다 많은 정보를 제공하지 못하고 있다.

본 연구는 시소러스 구축을 위한 국제 표준 지침으로 사용되고 있는 ISO 2788과 미국 국가표준인 ANSI/NISO Z 39.19-199X (*National Information Standards Organization. Guidelines for the Construction, format, and Management of Monolingual Thesauri*. 1991. 이하 Z 39.19로 표기함)의 기준, 그리고 현재 각종 시소러스에서 사용하고 있는 부가용어관계를 비교 분석

하여 문제점을 도출하고 이를 토대로 새로운 부가용어관계를 확장하는 방안을 제시하는데 목적이 있다.

2 시소러스의 용어관계와 부가관계 확장의 필요성

2.1 시소러스의 용어관계

시소러스에서 사용하는 용어관계는 두 가지로 대별할 수 있다. 첫 번째는 거시적 레벨(macro level)로 각 용어간의 의미적 관계인 등가관계(equivalence relationship), 계층관계(hierarchical relationship), 연관관계(associative relationship)라는 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(Aitchison & Gilchrist 1987). 다시 말해 전체 시소러스 용어들의 기본적인 관계를 나타내는 것을 거시적 레벨로, 이들 각각의 관계에서의 하위관계를 미시적 레벨로 보는 관점을 말하는 것이다.

본 연구에서는 등가, 계층, 연관관계를 시소러스의 기본관계로 정의하고, 범위주기를 포함하여 한정어 사용 등의 나머지 관계를 시소러스의 부가관계로 구분한다.

2.1.1 기본관계

1) 등가관계

등가관계는 동등관계라고 부르기도 하는 데, 색인과 검색에서 디스크립터로 사용하는 용어와 디스크립터로 사용하지 않는 비디스크립터

용어와의 관계를 말한다.

등가관계의 관계기호는 ISO 2788과 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. 등가관계의 처리 대상은 ISO 2788과 Z39.19에서는 동의어와 유사동의어, 업워드포스팅으로 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.

2) 계층관계

계층관계는 시소러스 디스크립터 용어를 상하위 개념으로 연결시켜주는 관계를 말한다. 일반적으로 계층관계는 상위어와 하위어로 나누어지며, 최상위 관계와 고립어관계도 계층관계에 포함시킬 수 있다.

계층관계는 용어의 특성에 따라 다양하게 나타날 수 있다. ISO 2788과 Z 39.19에서는 하위어(NT: narrow term)는 속관계(NTG: narrow term generic)와 전체부분관계(NTP: narrow term partitive), 사례관계(NTI: narrow term instance)로 구분하고, 상위어(BT: broader term)는 포괄적인 하위어를 나타내는 속관계(BTP)와 부분 상위어인 전체부분관계(BTP), 사례관계(BTI)로 구분 표시해 줄 수 있다고 정의하고 있다. 또한 Z 39.19에서는 계층관계 지시기호로는 BT1(1차 수준 상위 개념어), BT2(2차 수준), BTn(n차 수준), NT1(1차 수준 하위 개념어), NT2(2차 수준),

NTn(n차 수준)와 같은 관계기호를 사용할 수도 있다.

대부분의 시소러스에서 계층표시는 알파벳순 시소러스에서는 특정 디스크립터 다음에 BT와 NT를 사용하여 계층관계를 나타내고, 계층 순 시소러스 디스플레이를 갖는 시소러스에서는 점(·)으로 나타낸다. 이 경우 점의 수에 따라 계층관계의 특정성을 나타낸다.

3) 연관관계

연관관계란 계층적 관계에 속하지 않으며, 개념적으로 밀접하게 관련되어 있으나 등가관계에는 포함되지 않는 용어 관계이다. 즉 색인 작성과 탐색에서 이용될 가능성이 있는 대체용어를 제시하는 것을 말한다.

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 수단체, ⑤ 특성에 관한 개념, ⑥ 기원에 관한 개념, ⑦ 인과관계에 의해 연결된 개념, ⑧ 사물과 그것의 반대 행위자, ⑨ 개념과 그것의 측정단위를 포함시켰다.

ISO 2788과 Z 39.19, 그리고 대부분의 시소러스에서는 연관관계를 나타내는 지시기호로 RT(related term)를 사용하고 있다.

2.1.2 부가관계

1) 범위주기와 정의

용어의 부가관계는 시소러스의 동등, 계층, 관련관계의 기본관계 이외의 용어관계를 말하

며, 여기에는 시소러스 디스크립터에 대한 용어 정보를 제공하는 범위주기, 한정어, 주제구분, 이력주기, 주제구분, 포스팅 노트와 같은 것들을 포함한다.

범위주기(SN: scope note)란 색인어 내에서 색인의 의미를 나타내기 위해서 부속된 설명, 다시 말해 범위주기는 색인목적으로 사용되는 제한적 의미를 지시하여 다른 가능한 의미들을 제거하기 위하여 용어에 부가될 수 있는 것을 말한다. ISO 2788에서는 범위주기에 지시되는 내용으로는, ① 그 용어가 채택된 혹은 그 범위가 제정된 날짜, ② 신어인 경우 그 용어의 근원이나 부상하는 개념을 표현하는 용어의 새로운 제정, ③ 주어진 색인어에서 용어의 조합 허락 여부에 관한 사항을 색인자에게 의도적으로 지시한 내용을 포함하고 있다. Z 39.19에서는 범위주기에 포함되는 내용으로 디스크립터의 적용제한, 적용확장, 중복 디스크립터의 식별, 용어 사용범위를 들고 있다.

범위주기는 동형이의어에 부여된 한정어와는 달리 용어의 일부는 아니다. 또한 범위주기는 사전에 기재되어 있는 것과 같이 완전한 정의를 제시하는 것이 아니라 시소러스의 구조에 의해 제시된 의미를 보완하기 위하여 한정된 정의가 필요한 경우에 사용된다. 그러나 만약 용어가 일반관례에 있어 막연하게 해석되어지거나 사전들이 다양한 의미를 줄 때 범위주기는 완전한 정의를 제시할 필요가 있다.

정의(definition)는 디스크립터로 사용되는 용어의 의미를 나타내는 것을 말하는 데, ISO 2788과 대부분의 시소러스에서는 정의에 관련된 별도의 지시기호를 갖지 않고, 범위주기에 포함하고 있다.

2) 한정어

시소러스에서 한정어의 사용은 다의성이 있는 용어를 색인어로 사용했을 때 탐색자의 요구와 관계없는 주제가 검색되게 되어 잡음률이 높아지게 된다. 이를 방지하기 위하여 디스크립터 용어에 세부분야나 구분자를 추가시키는 용어를 한정어라고 한다.

다시 말해 한정어란 일반적으로 동형의어를 구별하기 위해 사용하는 부연관계라 할 수 있으며, 한정어는 동일어의 단수형과 복수형의 의미가 다른 경우 필요하다면 양쪽에 한정어나 한정구를 추가할 수 있다. 한정어는 대부분의 경우 디스크립터 용어에 ()로 추가된다.

3) 용어관련 정보

시소러스에서 사용할 수 있는 용어관계에는 동등관계, 계층관계, 연관관계의 기본관계와 범위주기 외에도 많이 있다. 특히 특정 데이터베이스에 종속된 시소러스일 경우에는 더 많은 정보를 탐색자에게 제공하여 탐색의 효율을 높일 수 있다. 그러기 위해서는 지금까지 시소러스에서 사용하지 않았더라도 가능한 한 다양한 관계기호를 사용할 필요가 있을 것이다.

이력주기는 용어가 시소러스에 사용된 내역을 나타내는 용어관계이며, 주제구분은 시소러스에 포함된 용어를 주제 카테고리로 구분하여 검색이 주제 한정을 할 수 있도록 하는 용어관계를 말하며, 여기에는 분류번호가 많이 사용된다. 또 포스팅노트는 시소러스 용어의 사용에 대한 정보를 나타낸다.

본 연구에서는 시소러스 구축 지침과 기존 시소러스에서 사용하고 있는 시소러스의 문제점을 도출하기 위해서 부가관계를 중심으로 살펴본다.

2.2 부가관계확장의 필요성

정보검색환경은 빠른 속도로 변화하고 있다. 따라서 시소러스의 기능도 변화하는 정보검색 환경에 맞게 바뀌어야한다. 초기 시소러스는 주로 수작업 환경에서 서지 데이터베이스의 색인과 검색에서 사용할 수 있도록 만들어 졌다. 그러나 현재의 정보검색환경은 全文 데이터베이스가 주축을 이루는 환경으로 변화되었다.

시소러스는 특별한 주제분야에서 문헌을 색인하고 검색하는 데 사용된 단어와 구(句)들간의 관계에 대한 정보의 寶庫이다. 또한 시소러스는 문헌에서 발생하는 생생한 텍스트 단어와 질의에서 사용된 단어간의 매핑을 향상시키기 위해 전문가들이 구축해 놓은 용어 데이터베이스이다 (Jones 1993)라고 정의해 용어 데이터베이스라는 개념으로 시소러스의 개념을 확장시켰다.

全文 데이터베이스의 색인은 자연언어색인을 기본으로 하고 있어 수작업 색인에서 사용하던 전통적인 시소러스로는 대응할 수 없는 부분이 많이 있다. 따라서 시소러스의 목적과 기능은 위에서 제시한 용어통제와 표준화에서 그칠 것이 아니라 멀티미디어나 하이퍼텍스트 기능을 수행하는 지식베이스로 보는 관점이 추가되어야 할 것이다. 시소러스 기능은 더욱 확장될 필요가 있다. 현재의 정보검색환경에서 시소러스는 이용자/서버구조에서 분산된 데이터베이스 관리, 시소러스 관계의 새로운 확장, 즉 개념적 접근, 어의범주, 언어적 기능, 그리고 멀티미디어 표현을 위한 확장 등이 필요하다(Stern & Rischette 1991)고 할 수 있다.

따라서 정보검색시스템이 DOS 환경에서 이용자에게 편리한 GUI 환경으로 바뀔 때 따라 기존의 시소러스에서 사용할 수 없었던 다양한

관계기호들을 사용하여 색인자나 검색자에게 도움을 줄 수 있게 되었다. 그러므로 시스템 설계자는 계속적으로 정보검색시스템의 인터페이스를 개선해야 한다.

시소러스의 새로운 이용으로 이용자 인터페이스는 GUI 또는 윈도우 환경에서 검색된 레코드나 탐색 프로파일을 디스플레이 하는 동시에 스크린에서 시소러스를 찾거나 통제어휘의 리스트를 보는 것이 가능하게 되었고(Rowley 1994) 색인단계에서 보다 오히려 탐색단계에서의 이용에 있다고 주장하고, 인공지능과 하이퍼텍스트를 지적하고 시소러스의 개념은 대부분 지식베이스의 기초가 되며, 한 영역의 지식 표현을 구성한 것이므로 인공지능에 시소러스를 적용할 수 있으므로(Weinberg 1995) 시소러스의 기능을 기존의 기능뿐만 아니라, 멀티미디어 환경에서의 인공지능 지식베이스와 용어 데이터베이스의 기능까지를 확장한 개념으로 사용해야 한다. 따라서 이러한 멀티미디어 환경과 하이퍼텍스트 기능 등에 따른 색인과 검색 기법의 변화와 더불어 사용자 중심에서 이용자 중심으로 바뀌고 있는 정보검색 환경의 변화에 따라 시소러스도 정보환경과 이용자요구를 만족시킬 수 있는 방향으로 변해야 한다.

이외에도 시소러스는 개념을 연결해 놓은 용어관계표 혹은 용어지도라는 관점에서 볼 때 정보검색자가 특정 용어에 대한 상하관계, 계층관계뿐만 아니라, 그 용어에 관련된 용어의 정의를 별도의 파일로 제공해 하이퍼텍스트로 연결시킬 수 있다. 이렇게 연결된 시소러스는 전자사전의 기능을 수행할 수 있어 검색자에게 도움을 줄 수 있을 것이며, 필요하다면 그 용어에 대한 이미지나 동영상, 소리까지도 들을 수 있게 하여 용어에 대한 관련정보를 확장해서

제공할 수 있을 것이다.

그러므로 새로운 환경에 적합한 시소러스의 기능을 수행하기 위해서는 시소러스는 동등, 계층, 관련관계의 기본관계뿐만 아니라 범위주기, 용어정의, 한정어의사용, 용어관련정보표시 등의 부가관계도 지금까지의 단순한 용어관계로만 만족할 것이 아니라 새로운 환경에 맞는 용어관계의 설정이 필요하다고 할 수 있을 것이다.

3 기존 시소러스의 부가관계 분석

3.1 분석대상

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

분석 대상으로 삼은 외국 시소러스는 『*Root Thesaurus*』(ROOT), 『*Construction Industry Thesaurus*』(CIT), 『*Thesaurus of Engineering and Scientific Terms*』(TEST), 『*INSPEC Thesaurus*』(INSPEC), 『*Thesaurus of ERIC Descriptor*』(ERIC), 『*Ei Thesaurus*』(EI), 『*Library of Congress Subject Headings*』(LCSH), 『*JICST 科學技術用語ソソーラス*』(JICST), 『*NASA Thesaurus*』(NASA), 『*INIS Thesaurus*』(INIS), 『*UNESCO Thesaurus*』(UNESCO), 『*Medical Subject Heading*』(MESH), 『*A Women's Thesaurus*』(Women's), 『*Thesaurus of Sociological Indexing Terms*』(TSIT로 표기함)의 개별 주

〈표 1〉 분석대상 국내 시소러스

	주제분야	개발주제	개발기관	연도
과학기술용어 시소러스	일반과학기, 화학물질	시스템공학연구소	시스템공학연구소	1992
신문기사 시소러스	신문기사 종합(전 주제)	한국언론연구원	한국언론연구원	1993
국방과학기술 시소러스	과학기술, 국방	국방과학연구소	국방과학연구소	1994
주제명요어집	전 주제분야	국립중앙도서관	한국문헌정보학회	1994
한국통신 시소러스	총류, 문화, 전기전자	한국통신 연구개발원	한국과학기술원	1995
법률시소러스	법률 분야	법원도서관	한국정보관리학회	1998

제 분야에서 용어수가 많고 인지도가 높은 14종을 주제분야별로 선정하였다.

한편 분석 대상으로 삼은 국내 시소러스 6개는 현재 국내에서 개발된 시소러스 중 ERIC 시소러스를 번역한 『KEDI 교육 분야 Thesaurus』와 인쇄시소러스로, 출판하지 않은 중앙일보사의 『JOINS 시소러스』를 제외한 전체를 포함하였다.

3.2 부가관계 분석

3.2.1 범위주기

ISO 2788과 Z 39.19, 그리고 대부분의 시소러스에서 디스크립터 적용제한이나 용어사용 범위안내 등을 위해 범위주기 기호인 'SN'을 사용하고 있으며, ROOT에서는 범위주기를 위한 기호 대신에 [] 사용하도록 하고 이탤릭체로 표기하고 있다.

JN/JO Petroleum products
[Products or by-products of crude petroleum refining]
 *)Liquefied petroleum gas JQG
[For general process synthesise, for example]
 JNG Petroleum products storage
 →Petroleum products JN/JO
 + Fuel storage QRV

또 INIS는 용어관계기호 없이 ()에 정의나 특수한 사용제한을 표기하고 있다.

CANONICAL DIMENSION[138 : 138]

(Scale dimension of quantum fields obeying canonical equal-time commutation relations.)
 BT1 scale dimension
 RT commutation relations

그러나 INSPEC에서는 ()한정어 부기를 범위주기에서 하도록 하고 있고, MESH는 주기(annotation)를 사용하고 있고, TSIT에서는 정의와 이용자 지시 그리고 이 둘을 함께 사용하는 범위주기와 시소러스에 나타나지 않은 특별한 개념을 위해 더 정적인 엔트리를 사용하거나 다른 용어와 조합하도록 지시하는 문맥의 존용어(context-dependent terms)를 범위주기에 포함하고 있다.

국내 시소러스의 경우는 한국통신 시소러스와 과학기술용어 시소러스는 범위주기를 사용하지 않는다. 법률 시소러스에서는 범위주기는 인쇄형태에서는 제외했다.

주제명요어집에서는 용어 적용의 제한이나 확장, 요어 식별, 유의사항 지시, 사용법 안내를 위해서 SN 기호로, 신문기사 시소러스도 범위주기를 위해 SN 기호를 사용하고 있다. 그러

나 실제 시소러스에서 사용하고 있는 범위주기는 미미한 실정이며, 국방과학기술 시소러스에서는 표제어 사용오류를 방지하기 위한 설명을 위해 SN을 사용한다고 규정하고 있지만 실제 시소러스에서 사용된 예는 없다.

3.2.2 정의

ISO 2788에서는 정의를 사용하는 경우를 막연한 용어나 한 용어가 다양한 의미를 가지고 있을 때 사용한다고 정의하고 있으나 정의에 대한 별도의 지시기호가 없다. 범위 주기와 정의를 명확히 구분해서 사용하는 시소러스는 없으며, 시소러스 표준에서도 이를 명확하게 구분하고 있지 않다.

INSPEC, ERIC, UNESCO, Women's, TSIT에서는 정의를 범위주기에 포함한다고 명확히 규정하고 있다. NASA는 정의를 범위주기에 포함하고 있으면서도 정의에 사전적 의미는 포함시키지 않고 있으며, Women's는 정의를 범위주기에 포함하지만 사전적 정의는 포함하지 않는다. UNESCO에서는 범위주기에 포함된 정의에 대한 출전을 나타내고 있다.

그러나 NASA의 경우는 디스크립터에 포함된 용어의 대부분에 대한 정의를 별도의 장으로 구성하여 정의에는 기준용어와 정의 그리고 정의의 소스, 그리고 시소러스에 추가된 연도를 기록하고 있다. 그러나 일반 장소명, 화학요소, 컴퓨터의 특정 모델, 비기술적 용어와 같은 일반 용어는 정의에서 제외하고 있다.

3.2.3 한정어

시소러스 구축 표준과 대부분의 시소러스에서 동형의어를 구분하기 위해 한정어를 부기하도록 하고 있으며, []로 표시하는 JICST와

분류번호를 []로 한정하는 LCSH를 제외하고는 대부분 () 한정어를 사용하고 있다. 그러나 JICST에는 동음이의어의 디스크립터가 있는 경우 용어 앞에 ※ 표시를 한다.

TSIT에서는 알파벳 표시에서 비디스크립터일 경우 이전에 디스크립터로 사용된 기간을 이력주기와 별도로 괄호 한정어로 나타내고 있다.

Acquaintance/Acquaintances(1965-1985)
HN DC 009500, deleted 1986. See now
Friendship or Social Contact.

NASA는 용어가 항공우주관계에서 2가지 이상의 의미로 사용되는 경우 혹은 용어간의 구별을 명료하게 할 필요가 있을 경우에 자신의 일부로 되는 용어는 한정어를 삽입하고, 용어의 일부로 되지 않는 경우는 범위주기를 사용한다.

sizing(shaping) ————— 디스크립터의 일부
sing(sing(surfacing)) ————— 디스크립터의 일부

spectroscopic analysis
(use of spectroscopic tools in chemical analysis) — 범위주기

주제명요어집에서는 명사의 의미가 불명료하거나, 또는 여러 주제분야에서 공통적으로 사용되는 용어는 그 의미를 한정하는 말을 원 괄호 () 속에 기재한 형식을 사용하고 동음이의어를 한정하기 위하여 원 괄호 속에 한자나 영어를 부기하고 있다.

가치(경제학) 조류(鳥類)
가치(철학) 조류(潮流)
가트(GATT) 조류(藻類)
사실주의(미술)
사실주의(문학)

또한 주제명요어집에서는 동음이의어나 애매

한 용어의 한정뿐만 아니라 지명아래에서의 주제구분, 주제의 지리구분, 시대구분에서도 괄호 한정어를 사용한다.

법률시소러스에서는 동형이의어가 있는 경우에는 한글/외국어를 불문하고 한글한정어로 구분하고 있으며, 한자가 있는 용어일 경우에는 괄호 한정어 다음 ()에 한자를 부기하고 있다. 이때 부기하는 한자는 참조 정보일 뿐 식별 정보는 아니다.

과실(범죄){過失}
과일(수익){果實}
사법(법률){私法}
사법(법원){司法}

한편 신문기사 시소러스는 동음이의어를 구별하기 위해 괄호 한정어를 사용하는데, ()를 부기하고 식별구를 넣는다. 괄호를 포함하는 식별구는 색인어의 일부가 된다. 한국통신 시소러스는 괄호 한정어 대신에 디스크립터에 영문명을 부기하고 있다.

과학기술용어 시소러스는 동음이의어를 구분하기 위해 괄호 한정어를 사용하고, 한정어로는 각각의 주제분야 명을 ()에 병기하고 주제분야가 동일한 동음이의어와 동일한 한글학명은 영어학명을 ()에 병기한다.

간장(기관)	비둘(다리뻐)
간장(식품)	비둘(코뻐)
격막(조직)	들쥐(Guerrillas)
격막(화학)	들쥐(Soricidate)

그러나 INSPEC은 한정어를 범위주기에 포함시키고 있다.

3.2.4 이력주기

용어가 시소러스에 사용된 내역에 관한 이력

주기는 일반적으로 HN(history note)로 표시되는데, ISO 2788에서는 이력주기를 범위주기에 포함시키고 있으며, Z 39.19에서는 HN으로 표시하고 이전 용어를 USE 참조로 두도록 제시하고 있다.

INSPEC은 이전 용어를 지시하는 용어로 PT(prior term)와 입력일자(DI : date of input)를 나타내는 기호를 갖고 있다.

all-pass filters

BT filters
TT networks(circuits)
CC B1270
DI January 1981
PT filters

EI는 디스크립터가 색인어로 처음 사용되어진 날짜를 DT(entry date)로 나타내고 있으며, 이전에 사용한 용어를 구별하기 위해 *표시를 하고 있으며, 새로운 시소러스 용어로 USE 참조를 제공하고 있다.

Microscope, Electron* (이전 용어)

USE : Electron Microscope(새로 디스크립터로 사용하는 용어)

Electron Microscope

UF : Microscope, Electron*

INIS는 개정판에서 삭제되거나 추가된 모든 디스크립터들은 시소러스의 앞에 리스트로 나타내고 있고 시소러스 자체에서 용어의 왼쪽에 화살표로 나타내고 있다. 또 용어 각각은 시소러스 용어로 등록된 날짜를 표시하고 있다.

→ INTERLOCKS(4: 4) May 86
RT control systems
RT switches

TSIT는 용어의 내역을 담은 이력주기를 가

지고 있으며, ERIC은 Add date(추가일자 표시), 이전용어 사용기간을 지시하고 있다. EI는 색인에 사용된 날짜를 나타내는 지시기호로 DT를 사용하고 있다. MESH 주석판은 디스크립터의 상황에 첨가한 상호참조에 변화의 상세한 역사를 제공하고 있다(Weinberg 1995).

그러나 국내 시소러스는 법률 시소러스만 HN과 DI를 사용하고 있으며, 나머지는 사용하지 않는다.

3.2.5 주제구분

ISO와 Z 39.19는 시소러스 용어의 주제구분을 할 수 있다고 지시하고 있으며, ROOT에서는 주제표시기호를, CIT는 참조번호(code), TEST와 JICST는 주제 범주를 사용하고 있다. JICST의 주제범주 색인은 주제분야에 속하는 전체 디스크립터를 파악하는 데 이용하기 위해 알파벳 리스트에 포함된 모든 디스크립터와 비디스크립터를 대상으로 주제 범주코드에 의해 구분하여 배열하고 있다.

또 INSPEC와 법률 시소러스는 분류코드(CC), 주제명요어집은 KDC분류번호, 국방과학 시소러스는 COSATI 분류번호를 사용하고 있으며, UNESCO도 분류기호를 사용하고 ERIC은 디스크립터코드 번호인 GE를 사용한다.

ROOT의 주제표시기호는 주제표시부에서는 각 용어의 앞에, 알파벳순표시에서 각 용어 다음에 알파벳으로 이루어져 있으며, 다른 주제범주에 속하는 용어일 경우에는 알파벳순 표시에서 이들을 '*' 기호로 나타내고 있다.

Women's는 알파벳 표시와 계층적 표시부 외에 주제그룹표시(subject group)와 딜리미터 그룹(delimiters group)코드 디스플레이를 갖고 있다. 딜리미터는 온라인 검색의 한계를 정하는

것, 즉 범위를 축소 또는 재정의 하거나 수동 자료처리시스템을 표준화하기 위해 사용될 수 있는, 종종 몇 개의 주제그룹에 속하는 포괄적인 검색용어나 고유명사를 포함하고 있다.

EI는 엔지니어링 문헌의 색인과 EI COMPENDEX** Plus, COMPENDEX 등의 데이터베이스 색인을 위한 도구로 사용되기 때문에, 별도의 분류코드를 가지고 있지만, 이 분류코드는 시소러스 용어관계에는 제시되지 않는다.

신문기사 시소러스는 정치(A), 경제(B), 산업(C), 사회(D), 사건사고(E), 문화(F), 과학(G), 국제(H)와 같이 A-H의 8개 대 분류와 영문자 두 자리로 구성된 44개의 분류코드를 가지고 있으며, 모든 색인어는 반드시 1-2개의 분류 코드를 가지고 있다.

3.2.6 포스팅 노트

ROOT는 데이터베이스운영에 필요한 정보를 참조정보, 계층수준, 데이터 범주와 용어출처코드를 포함하고 있으며, CIT는 다계층, 출현빈도, 최신성, 활용요소, 운영방법과 적용방법에 관한 정보를 갖고 있다. 또 ERIC과 INIS는 각각 포스팅 노트(색인횟수)와 디스크립터 사용빈도 정보를 갖고 있으며, MESH는 색인 대상을 표시하고 있다. INIS에서 사용하는 디스크립터의 사용빈도는 디스크립터 오른쪽에 나타나며, [] 안에 2가지 숫자로 표시된다.

GEOLOGIC STRUCTURES(779: 1,376) Nov 75
UF basins(sedimentary)
UF sedimentary basins

여기에서 첫 번째 숫자는 실제로 색인작성자에 의해 할당된 횟수를 나타내고, 두 번째 숫자

는 디스크립터로 사용된 횡수를 나타낸다.

ERIC에서의 포스팅 노트는 알파벳표목 각각의 디스크립터 범위주기 위에 부여되어 있으며, RIE(Resources in Education)나 CIJE(Current Index to Journals in Education) 색인에서 디스크립터나 식별어로 사용된 횡수를 나타낸다.

그러나 국내에서 개발된 시소러스는 온라인 시소러스로 사용되고 있는 시소러스가 없기 때문에 포스팅노트를 사용하고 있는 것은 없다.

이상에서 살펴본 용어의 부가관계를 종합해서 정리하면 <표 2>와 같다.

3.3 부가관계의 문제점

3.3.1 범위주기

범위주기는 색인의 목적으로 사용되는 제한적 의미를 지시하여 다른 가능한 의미들을 제거하기 위해 용어에 부가하는 설명을 말한다. 범위주기는, ① 색인어 의미를 제한하고, ② 주제의 범위를 제시하며, ③ 복합어의 용법을 지시하고, ④ 체계 표시의 구조를 나타내기 위한 더미텀을 나타내기 위해 사용하며, ⑤ 용어변천으로 용어가 채택된 시기, 특정용어 범위가 변경된 시기를 나타내는 기능을 수행한다(Aitchison & Gilchrist 1987).

주제명요어집에서는 범위주기의 설명이 모호한 경우가 많아 명확히 사용하기 어려우므로 사용방법까지 상세히 설명할 필요가 있다(한상완 외 1995)고 주장하고 가능한 상세하게 설명하고 예시를 포함하였다.

[예] 문학(文學)

SN 1) 각국 문학은 각 국어명 아래 사용한다.

(예 : 영문학, 한국문학)

2) 종교문학은 관련 종교명 아래 사용한다.

(예 : 기독교문학, 불교문학)

3) 특정 형식의 문학작품은 각국 문학 아래 사용한다.(예 : 영국소설, 한국소설)

4) 특정 시대의 문학이나 문학형식은 시대로 한정한다.(예 : 영국소설(20세기), 한국시(고려시대))

UF 문예

NT 기독교문학

RT 관능주의

그러나 범위주기의 기술을 단어로 하는지 문장으로 하는지, 문장으로 한다면 몇 단어 이내로 제한하는지 등에 관한 분명한 기준이 제시되지 않고 있으며 현재 사용 중인 시소러스에서 여러 가지 형태로 나타나고 있다.

3.3.2 한정어 사용기준

동형이의어는 하나이상의 의미를 가지는 문자열이나 발음을 갖는 용어를 말하고, 동음이의어(homonym)는 발음은 같으나 한 가지 이상의 의미를 갖는 용어를 말한다. 또한 동음이형이의어(homophone)는 발음(cell과 sell)은 같으나 뜻이 다른 용어를 말한다. 동음이의어는 정보검색에서 문제가 되지 않으나 동형이의어는 검색에 어려움이 있다(Lancaster 1986).

여러 학문분야의 문헌이 통합되어 색인될 때 발생하는 동형이의어의 빈도는 매우 높으며 이를 구분하는 방법이 없으면 검색에서 잡음률은 높아질 것이 틀림없다. 따라서 색인자가 브라우징 해야하는 어휘사전이나 탐색자가 브라우징 할 수 있는 색인파일에서는 이들을 쉽게 구분할 수 있어야 한다(김태수, 최석두 1997).

ISO 2788에서는 동형이의어 색인 시 한정어나 구를 부가한다. 색인어는 한정어와 구별하기 위해서 다른 활자를 사용하거나 괄호 속에 넣어서 용어 뒤에 추가하여 두 개 이상의 의미를

〈표 2〉 시소러스의 부가관계

	범위주기	정의	한정어	이력주기 (이전용어)	주제구분	기타
ISO표준	SN 용어 채택 날짜 신어의 근원 용어의 조합 지시	SN에 포함 (막연한 용 어나 다양한 의미일 때)	()한정어 동일어의 단 복수 형 동형이의어	SN에 포함	없음	
Z 39.19	SN 디스크립터 적용 제한 디스크립터 적용 확장 중복 디스크립터 식별 용어사용법 안내 상호 범위주기 X SN		()한정어 동형이의어 애매한 용어	HN 디스크립터 정보원 근거가 되는 전거 이전용어- USE참조		
UNESCO	SN	SN에 포함	()동형이의 어 구분	SN에 포함	계층적주제 분류번호	
LCSH	SN 표목이 적용되는 주제 문제의 영역 표시 표목구별 사용제한 약 4,000개		[]분류번호 한정		분류번호 (LC)	범례(표목을 장소 및 분 류번호로 세 분)
ROOT	[] 이태릭체로 표시 범위주 안내주 합성어 안내				주제 표시 기호	DB운영에 필요한 정 보(참조정보, 계층수 준, 데이터 범위와 용어출처코드)
INSPEC	이태릭체로 표시 우선어의 범위설정 의미를 명확히 할 때 이전용어 표시 see also	SN에 포함	()한정어부가 일부는 SN에 포함	DI(입력일) PT(이전용어) HN(이력주기)	CC (분류기호) DC 디스크립 터코드)	
JICST			[]한정어 (의미표기) ※(동음이의어 표시)		주제 범주 코드	
TEST	이태릭체로 표시 디스크립터의 사용지시		()한정어		COSATI 분류번호	주제별 용어 리스트 포함
NASA	SN 용어의 의미제한 다의성 구분	별도의 책 으로 구성 됨	()한정어 일부는 SN에 포함	연월 표시		보유판을 가 짐
INIS	()에 표시 용어정의 설명 용어의 사용기간 지시	SN에 포함	()한정어	Date기호 →(새 디스크 립터 표시) 연월표시		[]디스크립 터 사용 빈 도(포스팅노 트)
MESH	주기(Annotation) 용어선택 주의 사항 온라인 주기 색인주기 목록작성 주기			History note (해당연도 표시)		tree번호, 온 라인주기 색인대상표 시

CIT	사용하지 않음		()한정어 동형이의어일 때 참조번호로 구분		facet사용 참조번호 (code)	출현빈도, 최신성 활용요소 운영방법과 적용방법
EI	SN (용어의 의미나 사용이 명확하지 않을 경우)			DT(처음색인에 사용된 entry date) * (이전 용어 구별위해 용어 다음에 표시)	분류코드 (4단계)	날짜가 명시되지 않은 용어는 93년 이후부터 사용된 용어임
ERIC	SN 디스크립터 용어 설명 사용범위 제한 범주지시 이전시소러스지시 디스크립터 그룹코드 안내	SN에 포함	한정어 (동형이의어) 식별어 (주제구분, 출판형식)	Add Data (추가일자 표시) 이전용어 (용어 다음에 사용기간 지시)	GE (디스크립터 그룹코드:3자리)	포스팅 노트 (CLJE와 RIE에서의 색인 횟수) 신규, 변경, 삭제 디스크립터 리스트 있음)
TSIT	SN 의미한정 사용범위 정의 이용지시 문맥의존용어	SN에 포함	()한정어 ()이전 디스크립터 사용기간:비디스크립터일 경우	HN(이력주기) 형식이 다양함	DC 디스크립터 코드	
Women's	SN 정의 사용영역 표시 의미를 명확히 할때 속어, 비공식 용어 이탤릭체로 표시	SN에 포함 사전적 의미는 제공 않음			SG (주제그룹) DG(delimiter group)	주제별 용어 리스트 있음
과학기술 용어	사용하지 않음		()한정어		COSATI 분류기호	
신문기사	SN		()한자 한정어		분류코드	분류별 계층색인가 분류별 자모순 색인 있음
국방과학기술	SN (표제어 사용 오류를 방지하기 위한 설명주)		()한정어 주제분야, 전문용어, 한자어, 외래어	사용하지 않음		SN은 실제 사용된 예가 없음
주제명 요어	SN 용어 적용의 제한이나 확장 요어식별 유의사항 지시 사용법 안내		()한정어 ()지명아래 주제구분 ()지리구분 ()시대구분		KDC분류기호	
한국통신	사용하지 않음		디스크립터에 영문부기	사용하지 않음		
법률	SN			HN(이력주기) DI(입력일자)	CC (분류기호)	SN은 인쇄형 시소러스에서는 표기하지 않음

구별한다고 제시하고 있다. 그러나 ISO 2788에서는 한정어와 복합표목 중, 어떤 구조로 표현해야 하는가에 대한 분명한 기준이 제시되지 않고 있으며, 대부분의 시소러스에서도 동형어의 구분방법으로 괄호 한정어를 사용하고 있지만 괄호한정어 사용에 대한 명확한 기준이 마련되어 있지 않다.

3.3.3 용어관련정보 표현

시소러스를 이용하여 자동색인을 할 경우에 주제구분은 가중치 부여에 중요한 역할을 수행한다. 본문에 등장한 용어와 같은 주제일 때 높은 가중치를 적용하여 색인어 선정에 돕기 때문에 시소러스 용어에 주제구분이나 분류기호를 부가할 필요가 있다. 또한 시소러스에 포함된 용어의 입력일자, 입력자, 이전에 사용했던 용어 같은 이력주기도 색인자나 검색자의 용어선택에 도움을 줄 수 있다. 그러나 이들에 관한 용어관계 지시기호가 일부의 시소러스에만 사용하고 있을 뿐 대부분의 시소러스와 시소러스 구축 지침에 포함하고 있지 않다.

이외에도 특정 데이터베이스에 종속된 시소러스일 경우 현재까지 디스크립터가 색인어나 탐색어로 사용된 횟수는 색인자와 검색자의 용어선정에 도움을 줄 수 있으나 대부분의 시소러스와 구축 지침에 설정하고 있지 않다. 그러므로 ISO 2788 지침에서는 제시되지 않고 있지만, 필요하다면 특정한 배열 순서에 의해서 주제구분, 분류번호, 입력일자, 이전용어, 이력주기 등에 대한 정보를 기술하는 것이 바람직할 것이다. 이를 위해서는 이러한 용어관계의 설정에 따른 기준이 마련되어야 한다.

4 부가관계의 확장

4.1 범위주기

4.1.1 DF(용어정의)

ISO 2788에서 범위주기는 동형어의어에 부여된 한정어와 달리 용어의 일부는 아니다. 또한 범위주기는 사전에 기재되어 있는 것과 같이 완전한 정의를 제시하는 것이 아니라 시소러스 구조에 의해 제시된 의미를 보완하기 위하여 한정된 정의가 필요한 경우에 사용되며, 때로는 상세한 정의가 요구된다고 규정하고 있고, Z 39.19에서는 범위주기를 일반적으로 디스크립터로서 선택된 의미를 명시해야 하지만 자연언어에서 인식되고 있는 다른 의미도 명시할 수 있다고 제시하고 있다. 일반적으로 시소러스에는 완전한 정의는 필요하지 않으나, 의미를 보완하기 위해 한정된 정의가 필요한 경우가 있으며, 때로는 상세한 정의가 필요한 경우가 있다는 것이다.

그러나 Soergel(1974)은 시소러스 구축을 위해서는 완전하고, 공식적인 정의가 필요하다. 용어의 의미는 시소러스 구조의 전후관계를 통해서 명백해 진다. 그러나 자주 전후관계가 용어의 의도된 사용을 수행하는 것이 충분치 못하기 때문에 추가적인 범위주기가 필요하다고 주장하고 이의 필요성으로 ① 시소러스 구조에서 적절한 용어를 배치시키기 위해서(시소러스를 구축할 때), ② 색인과 탐색에서 시소러스를 이용하기 위해서라고 주장했다. 또한 Foskett(1996)도 정의의 필요성을 강조하면서 특히 사회과학에서는 하나 이상의 상황을 언급하거나 애매한 용어의 명백한 의미를 제공하기 위하여 범위주기 뿐만 아니라 정의주기(DF:

definition notes)가 필요하다고 주장하였다.

시소러스 구축지침에서는 정의를 위한 별도의 용어관계를 지시하지 않고, 범위주기에 포함시키고 있으며, 현재 구축되어 있는 시소러스 중에서 정의를 별도의 기호로 사용하고 있는 예는 없다. 대부분 용어 정의가 필요하다면, 범위주기에서 함께 사용하고 있다.

그러므로 디스크립터에 대한 애매성을 해결하고, 색인자에게 보다 정확한 용어정보를 제공하기 위해서는 정의가 필요하다. 본 연구에서는 정의를 범위주기와 분리시켜 정의를 위한 용어관계를 DF(definitions)로 제안한다. 또한 현재 시소러스는 자동색인을 위한 지식베이스로 기능을 수행하기 때문에 ISO 2788 지침의 적용 기준은 현실에 맞지 않다. GUI 환경의 온라인 정보검색시스템에서 사용하는 시소러스일 경우에는 시소러스 용어를 사전과 하이퍼링크로 결합시켜 정의를 확인할 수 있도록 할 필요가 있을 것이다. 이것은 개념의 정의가 철저히 분석되고 모든 유용한 상호관련성이 추출되고 형식화되면, 결과적으로 시소러스구조는 독립된 정보도구와 사전에 대한 부차적인 도구로 서비스할 수 있을 것이기 때문이다(Nkwenti-Azeh 1994).

또한 하이퍼텍스트를 고려할 때 그 용어에 대한 이미지, 소리정보까지 제공할 수 있을 것이다. 이 경우 용어사전에 관한 정보는 별도의 파일로 두고 하이퍼텍스트 기능으로 연결시켜 이용자에게 필요한 용어정보를 제공할 수 있을 것이다.

4.1.2 SN(범위주기)

정의란 한정적인 의미를 나타내지만 범위주기는 여러 가지로 이용된다. 범위주기는 색인과

탐색과정에서 용어가 어떻게 사용되는지를 설명한다. 이것은 특히 공동목록시스템에서 사용되는 시소러스나 다국어 시소러스에서 중요하다. 특히 설명은 디스크립터에서 취급하는 용어를 배제하는 어떤 의미를 지적한다.

모든 디스크립터가 범위주기를 필요로 하는 것은 아니지만 시소러스 이용자에게 범위를 명백히 하기 위해서 필요하다. 여기에는 외국어를 포함한 생소한 용어, 최신용어, 일반적인 사용과는 달리 사용되는 용어들이 포함된다. 더 나아가 범위주기는 정확한 정의를 필요로 하지 않는다. 그것도 단지 특별한 용어를 어떻게 사용할 것인지 사용하지 않을 것인지에 대한 것을 안내하는 것이다(Lancaster 1986). Barhydt와 Schmidt는 범위주기를 ① 용어범위에 대한 긍정적 제한(positive limitation) ② 부정적 제한(negative limitation) ③ 정확한 정의 ④ 정의와 긍정적 제한의 조합 ⑤ 예시의 4가지 형태로 구분했다. 그러나 용어가 채택되거나 범위가 지정된 된 날짜에 관한 기록으로는 많은 시소러스에서 이력주기에 포함시키고 있으며, 때로는 다양한 방법으로 이들을 구분하고 있다.

그러므로 앞에서 제시했듯이 범위주기와 정의는 분리되어야 하며, 이력주기 등과도 구별할 수 있도록 구분할 필요가 있을 것이다.

한편, 주제명표목표, 분류표 등에서 사용하고 있는 범위주기의 유형을 살펴보면 KDC에는, ① 정의주기, ② 포함주기, ③ 항목전개 유보주기, ④ 특정 서명주기, ⑤ 인명 지명주기, ⑥ 지시주기, ⑦ 양자택일주기, ⑧ 참조주기의 8가지 주기가 있으며, DDC에는, ① 정의주기, ② 범위주기, ③ 이전 표목 주기, ④ 별명 주기, ⑤ 당소 분류 주기, ⑥ 포함 주기, ⑦ 타소 분류주

기, ⑧ 보라참조, ⑨ 도보라 참조, ⑩ 개정주기, ⑪ 미 사용주기, ⑫ 재배치주기, ⑬ 사용금지주기 등의 주기가 있다. 또 LCC(Library of Congress Classification)에는, ① 범위주기, ② 설명주기, ③ 참조주기, ④ 보라주기, ⑤ 포함주기 등이 있으며(Chan, 1980), BC(Bibliographic Classification)에는 분류항목 다음에 별표(*)와 함께 나타나며, 종류는 지시주, 설명주, 참조주의 3가지가 있다. 이외에도 LCSH에서는, ① 정의, ② 다른 표목과의 관련성(표목 범위), ③ 지시, 설명, 참조와 같은 3가지 유형의 범위주기를 사용하고 있다(Chan 1986).

이상의 범위주기에 관한 다양한 정의와 일반 분류표와 시소러스에서 사용하고 있는 유형을 종합해서 구분하면 정의주, 설명주, 지시주, 예시주, 범위주, 포함주, 제외주, 이력주기, 인명주기로 나눌 수 있을 것이다.

ROOT, UNESCO, ERIC, TSIT의 4가지

시소러스에서 사용하는 범위주기를 살펴보면 <표 3>과 같다.

전체적으로 정의와 설명이 68.13%로 가장 많으며, 다음이 지시주기로 20.55%를 차지하고 있다. 그러므로 범위주기에서 정의를 분리시킬 필요가 있으며, 범위주기도 주기의 내용에 따라 다양하게 구분할 필요가 있을 것이다.

따라서 본 연구에서는 앞에서 정의를 별도의 용어관계인 'DF'로, 이력주기는 'HN'으로 제시하여 범위주기와 분리시킬 것을 주장했던 것처럼, 지시, 범위, 포함, 제외와 같은 범위주기 유형을 구분하여 'SN' 이외의 부가기호를 사용할 것을 주장한다. 이를 위해 본 연구에서는 설명[SN(e) : explanatory], 지시[SN(i) : instruction], 범위[SN(s) : scope]와 같이 구분하는 것이 효과적일 것이다.

4.2 한정어

<표 3> 시소러스의 범위주기 사용 예

(단위 : 회)

	정의		지시			예시	범위	포함	제외	이전 용어	문맥 의존	생물년	계
	정의	설명	합성	타용어 지시	안내								
ROOT		180	421	75			14	7	8				705
UNESO	399	218		25	11			71	23				747
ERIC	2,160 *948		13 *33	45 *269	21 *413	*30	*2	*13	10 *13	13 *175			2,262 *1,896
TSIT		616 *11			27 *11	35	9				105	228	1,020 *22
계	3,057	1,025	467	414	483	65	25	91	54	188	105	228	6,202
	4,082		1,364										
비율	57.09		22.85			1.4	.09	3.12	1.24		4.2	9.14	100

(* 표시는 중복되어 사용된 것임)

동형의어를 구별하기 위해서 주로 사용하는 한정어는 범위주기의 축소형으로 고려된다. 그러나 한정어는 디스크립터의 일부분으로 고려되지만 순수한 범위주기는 'SN'에 의해 완전히 분리되어 기술된다.

한글 시소러스의 경우 전산환경에서 한자를 사용하게 되면, 한자를 선정 시간이나 한자만을 한정어로 사용할 수 없는 경우와 예상되는 배열체계의 혼란 등 여러 가지 문제가 있을 수 있으므로 한글을 사용하는 것이 편리하다(김태수, 최석두 1997). 그러나 현재 국내 어휘사전에서 사용되는 한정어로는 한글, 한자, 영문 등을 섞어서 사용하고 있으며, 그 중 한자 부기에 의한 구별방법을 가장 많이 채택하고 있다.

종래 한글 주제명표목표에서 한자를 한정어로 주로 사용하게 된 원인은 우리의 일반적인 사회관습이나 사전에서 한자를 식별수단으로 사용한데서 비롯된 것으로, 이용자 측면을 고려할 때 한자나 로마자보다 한글을 우선 고려하는 것은 아주 자연스러운 일이며, 일찍이 한정어의 한글사용이 제기된 바 있다.

물론 한글시소러스에서는 한글 한정어 사용이 타당할 것으로 보여지지만, 가능하다면, 동일 주제에 속한 용어라면 상위개념을 사용하는 것이 바람직 할 것이다. 왜냐하면, 시소러스를 이용한 자동색인에서 관련 개념이 있을 경우 가중치를 부여하여 한정어에 해당하는 용어를 자동으로 생성해 낼 수 있기 때문이다.

의미에 대한 한정어 이외의 정보를 한정어 형식을 빌어 한정어의 위치에 부기하지 않고 다른 관계지시기호(예를 들면, 범위주기나 새로운 관계지시기호를 정의하여)를 사용하여 표현할 수도 있다. 그러나 이 방식은 시소러스에서 점차 배제하는 추세이며, 색인 및 검색시 색

인자와 탐색자가 편리하게 참고할 수 있도록 디스플레이하기 위해서는 한정어 형식이 적당할 것이다.

그러므로 한글 시소러스를 작성할 경우에는 지금까지 동형의어나 동음의어를 구별하기 위해 사용하던 한자의 부기보다는 한글 한정어를 사용하는 것이 색인의 효율성 측면에서 좋다. 또한 가능하면 한글의 한정어 사용에 있어서도 대상용어의 상위 개념을 사용하는 것이 훨씬 효과적이라고 할 수 있다. 왜냐하면, 상위 개념을 사용했을 경우, 색인과정에서 디스크립터로 사용한 용어뿐만 아니라 한정어로 사용한 용어를 참고로 하여, 색인어 선정에 가중치 부여를 할 수 있기 때문이다.

본 연구에서는 동형의어나 동음의어를 구별하기 위한 한정어 사용원칙을 다음과 같이 제시한다.

첫째, 한글 한정어를 사용하는 것을 원칙으로 한다. 한글 한정어로 구별이 어려울 경우에는 한자어를 부가한다.

둘째 상위 개념어를 사용한다. 가능하면 대상 용어의 상위 개념에 해당하는 용어를 사용하므로 시소러스를 자동색인에 이용할 경우 인접 용어를 이용한 가중치 부여에 이용할 수 있으므로 효과적인 색인어 선정이 가능할 것이다.

셋째, 쉽게 의미를 파악할 수 있는 다의성이 적은 용어를 사용한다.

넷째, 하나의 한정어로 구분이 어려울 경우는 복합 한정어를 부기하고 필요하다면 한자를 부기한다.

4.3 용어 관련정보

4.3.1 CC(분류 및 주제구분)

일반적으로 시소러스와 분류는 다른 개념으

로 정보검색에 이용되어 왔다. 분류는 문헌이 배열 순서를 지시함과 더불어 접근점으로 사용되었고, 시소러스는 주제어를 통제하기 위해서 사용되었다. 그러나 최근에는 시소로 패시(thesaurofacet)과 같이 분류와 시소러스를 통합하는 시스템이 등장하고 있으며, 이 둘의 관계가 상호 보완적으로 작용할 수 있는 방안이 연구되고 있다.

시소러스에서 분류기호의 역할은 색인에서 동일 류에 속하는 개념의 가중치를 부여할 때 중요한 기능을 수행하게 된다. 그러므로 분류나 주제 구분을 시소러스에서 나타낼 필요가 있다.

본 연구에서는 이를 위한 용어관계기호로 CC(classification code or category code)를 제시한다.

4.3.2 HN(履歴注記)

이력주기는 시소러스에 사용되는 용어에 대한 역사를 나타내는 것이다. 시소러스는 일반적으로 도서관 분류표보다 범위가 좁기 때문에 분류표보다 더 자주 갱신되어야 한다. 특히 온라인 시소러스의 경우는 갱신 주기가 빠를수록 좋다. 온라인 시소러스는 소프트웨어의 도움으로 유지되기 때문에 새로운 디스크립터의 제안을 스크린 형식에서 만들 수 있으며, 시소러스 편찬자가 색인에서의 적용을 위해 빨리 인증된다. 그러나 디스크립터의 개정은 탐색에서 중요한 관계를 가진다. 그 이유는 디스크립터를 수정할 때 이전 문헌을 다시 색인할 수 있는 색인 운영은 거의 불가능하기 때문이다. 그러므로 시소러스는 각 디스크립터의 용어 역사를 기록해야 한다. 최소한 용어가 시소러스에 추가된 날짜라도 포함해야 한다(Weinberg 1995). 또 색인에서 시소러스에 디스크립터가 포함된 날짜

와 삭제된 날짜, 혹은 상호참조가 도입된 날짜도 이용자에게 유용하므로(Soergel, 1974)시소러스 디스크립터는 디스크립터로 등록된 날짜를 가지고 있어야 하며, 여러 사람이 용어를 입력하고 관리할 경우에는 용어 입력자의 책임성 부여를 위해 입력자명을 나타내 줄 필요가 있다.

이를 위해 HN(history note : 이력주기), DI(date of input : 입력일자), PI(person of input)를 관계기호로 사용할 것을 제안한다. 그러나 온라인 시소러스에서는 특별한 용어관계 지시기호 대신 디스크립터 옆에 용어정보로 나타낼 수도 있을 것이다.

4.3.3 포스팅 노트(posting note)

시소러스가 본문 정보검색시스템과 연결되어 자동색인을 수행하는 경우에는 색인용어와 시소러스를 매치시켜 시소러스 정보를 제공할 수 있다. 예를 들면 자동색인에 필요한 정보 즉 용어가 속한 주제, 용어의 계층 수준 다시 말해 특정성과 전체시스템에서의 용어정보로 빈도, 관련용어, 특성 등을 가지고 있다면 효과적인 색인어를 생성할 수 있을 것이다. 그러므로 이를 위한 관계지시 기호가 마련되어야 한다.

포스팅노트는 색인시에는 특정 용어의 중요도를 판단하는 기초로 삼고, 검색시에는 이용자가 용어를 확인할 때 그 용어가 가지고 있는 색인 횟수를 참고할 수 있어 검색에 많은 도움을 줄 수 있다.

포스팅 노트를 위해서는 용어관계기호가 필요하다면 PO(posting note)를 사용할 것을 제안한다.

5 결 론

시소러스의 부가용어관계는 탐색자에게 디스크립터를 안내하는 중요한 용어관계이다. 그러나 기존의 시소러스 구축지침에서는 다양한 용어관계를 지시할 수 있음에도 불구하고 범위주기와 한정어 사용만으로 제한하고 있어 시소러스를 이용하는 이용자에게 보다 많은 정보를 제공하지 못하고 있다.

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

첫째, 범위주기에 대한 분명한 기준이 제시되지 않고 있으며, 현재 사용 중인 시소러스에서 여러 가지 형태로 나타나고 있다. 따라서 범위주기는 정의DF(definition notes)와 범위주기(SN)을 분리시켜 사용할 것을 제안하였다.

이때 정의는 시소러스 용어에 사전과 결합시키고 필요하면 이미지나 동화상까지도 보여줄 수 있도록 하는 방안을 제시하였다. 또 범위주기도 유형별로 구분하여 SN 이외의 설명[SN(e)], 지시[SN(d)], 포함[SN(i)]과 같이 부가기호를 사용할 것을 제안하였다.

둘째, ISO 2788에서는 한정어 사용에 대한 분명한 기준이 제시되지 않고 있으며, 대부분의 시소러스에서도 동형이의어의 구분방법으로 괄호 한정어를 사용하고 있지만 괄호한정어 사용에 대한 명확한 기준이 마련되어 있지 않다. 따라서 동형이의어나 동음이의어를 구별하기 위한 한정어 사용원칙을 제안하였다.

셋째, 시소러스 용어에 관련 된 정보인 주제구분, 분류번호, 입력일자, 이전용어, 이력주기 등에 대한 기준이 마련되어 있지 않다. 따라서 용어관련 정보의 확장으로 기존 시소러스에서는 사용하고 있으나, 시소러스 구축 국제 표준 지침인 ISO 2788에서 제시하고 있지 않은 용어관계를 제안하였다.

참 고 문 헌

- 김태수, 최석두. 1997. 동형이의어 구별을 위한 한정어 사용에 관한 연구. 『情報管理學會誌』, 14(1): 107-124.
- 최석두. 1998. 매크로시소러스에서의 용어관리. 『제1회 전문용어 언어공학심포지움. 전문용어언어공학연구센터』, 1998년 12월 12일: 43-53.
- 한상완 외. 1995. 주제명 검색요어집 개발에 관한 연구. 『도서관』, 50(1): 3-23.
- Aitchison, J. & A. Gilchrist. 1987. *Thesaurus Construction : a practical manual*, 2nd ed. London: Aslib.
- Chan, Lois Mai. 1980. *Immortal's Guide to the Library of Congress Classification*. Littleton, Colorado : Lib. Unlimited.
- . 1986. *Library of Congress Subject Heading : Principle and Application*. Littleton, Colorado : Lib. Unlimited.
- Foskett, A. C. 1996. *The Subject Approach*

- to *Information*. 5th ed. London: Library Association Publishing.
- Jones, Susan. 1993. "A thesaurus data model for an intelligent retrieval system." *Journal of Information Science*, 19(3): 167-178.
- Lancaster, F. W. 1986. *Vocabulary Control for Information Retrieval*. 2nd ed. Arlington : Information Resources Press.
- Library of Congress. 1991. *Subject Cataloging Manual: Subject Headings*. 4th ed. Washington: LC, Office for Subject Cataloging Policy.
- Nkweni-Azeh, Blaise. 1994. "The use of thesaural facets and definitions for the representation of knowledge structure." In *Knowledge Organization and Quality Management*. Frankfurt : Indeks Verlag: 374-381.
- Rowley, J. 1994. "The controlled versus natural indexing languages debate revisited : a perspective on information retrieval practice and research." *Journal of Information Science*, 20(2): 108-119.
- Schmitz-Esser, W. 1991. "New approaches in thesaurus application." *International Classification*, 18(3): 143-147.
- Soergel, D. 1974. *Indexing Languages and Thesauri : construction and Maintenance*. Los Angeles : Melville.
- Stern, A. & N. Rischette. 1991. "On the construction of a super thesaurus based on existing thesaurus." In *Tools for Knowledge Organization and the Human Interface*. Frankfurt : Indeks Verlag: 134-144.
- Weinberg, B. H. 1995. "Library classification and information retrieval thesauri : Comparison and contrast." *Cataloging and Classification Quarterly*, 19(3/4): 23-44.