

# 시소러스 용어관계에 관한 연구

## A Study on Term Relationship of Thesaurus

한상길, 대림대학 문헌정보과

이두영, 중앙대학교 문헌정보학과

Sang-Kil Han, Dept. of Lib. & Inf. Sci., Daelim College

Too-Young Lee, Dept. of Lib. & Inf. Sci., Jung Ang University

시소러스는 정보검색과정에서 용어를 통제하여 색인과 검색에서 중요한 기능을 수행하는 도구로 사용되고 있는 통제용어집이다. 그러나 정보검색시스템 환경의 변화에도 불구하고 기존의 시소러스와 시소러스의 작성지침으로 사용되고 있는 국제 표준인 ISO 2788이나 Z39.19에서 제시하고 있는 기준은 용어간의 관계를 단순히 표현하고 있어 색인이나 검색시 용어통제의 어려움을 겪고 있다. 본 연구는 시소러스의 용어관계를 국제 표준과 현재 사용되고 있는 각종 시소러스를 비교 분석하여, 시소러스 용어관계의 확장방안을 제시한다.

### 1. 서론

시소러스는 정보검색시스템에서 용어를 통제하여 효율적인 색인작업과 검색의 효율성을 높이는 중요한 도구로 색인시에는 동일한 개념이 각기 다른 용어로 색인되는 것을 방지하고, 검색시에도 용어간의 관계를 표시하여 이용자가 빠른 시간 내에 정확하게 정보를 검색할 수 있도록 하는 기능을 수행한다.

시소러스가 단순히 색인에서 용어통제역할만을 수행하는 것이 아니라, 이용자에게 용어정보를 제공하고 용어확장검색의 도구로써 사용할 수 있게 하려면, 용어간의 다양한 관계 유형의 설정이 필수적이라 할 수 있을 것이다. 그러나 기존의 시소러스와 시소러스의 작성지침으로 사용되고 있는 국제 표준인 ISO 2788에서 제시하고 있는 기준은 용어간의 관계를 단순히 표현하고 있어 멀티미디어 환경에서의 이용자 요구를 수용하지 못하고 있다.

본 연구는 정보검색환경의 변화를 살펴보고, 시소러스 구축 표준과 현행 시소러스에서 사용하고 있는 용어관계 문제점을 고찰하여 멀티미디어 환경에 맞는 시소러스 용어관계 확장방안을 제시하고자 한다.

### 2. 정보검색 환경의 변화와 현행 시소러스의 문제점

#### 2.1 정보검색환경의 변화

전통적으로 시소러스는 탐색자보다 색인자를 안내하기 위해 정보검색에서 사용되어 지금에 이르고 있지만, 색인이나 검색환경이 시소러스가 처음 나왔던 시기와는 많은 변화가 있다. 다시 말해 시소러스는 도서관과 문서보관의 좁은 분야를 넘어서, 언어와 지식공학과 같은 보다 넓은 영역으로 파고들고 있으며, 그렇게 될 때 시소러스의 모양이나 형태가 변할 것이며, 심지어 이름도 변할 것이다[Schmitz-Esser, 1991]라고 예견되고 있다. 또한 이용자/서버구조에서 분산된 데이터베이스관리, 시소러스관계의 새로운 종류 확장 즉, 개념적 접근, 어의카테고리, 언어적 기능, 그리고 멀티미디어 표현을 위한 확장 등이 중요 이슈로 등장하고 있다[Stern & Rischette, 1992].

이러한 멀티미디어 환경과 하이퍼텍스트 기능 등에 따른 색인과 검색기법의 변화와 더불어 사용자 중심에서 이용자 중심으로 바뀌고 있는 정보검색환경의 변화에 따라 시소러스도 변화해야 할 것이다.

## 2.2 시소러스 기본관계

### 2.2.1 등가관계

ISO 표준에서는 등가관계를 동의어와 유사동의어, 업워드포스팅으로 구분하고 동의어에는 어원이 상이한 용어, 속명과 학명, 통칭명과 상품명, 새로운 개념과 경쟁상태의 용어, 최근에 빈번하게 사용되는 용어와 그다지 사용되지 않게 된 용어, 이형표기, 문화차이에 따른 용어, 약어와 완전어, 복합어의 분해형과 비분해형으로 구분하고 있다. 유사동의어에는 일반적으로 의미가 상이하다고 인정되나 색인작업에서는 동의어로 취급하는 용어로 정의하고 있다. Z39.19에서는 동의어와 유사동의어로 구분하고 유사동의어를 일반적으로 일상용법에서는 의미가 다르지만 시소러스의 목적에 따라 동의어로 취급하는 용어로 종종 반의어를 포함시키고 있다고 정의하고 있다.

INSPEC 시소러스에서는 반의어를 등가관계에 배치했고, TEST 등 일부 시소러스는 반의어를 유사동의어로 취급 등가관계에 배치하고 JICST와 ERIC시소러스는 반의어를 관련관계에 두고 있는 등 시소러스에 따라 다양하게 나타나고 있다. 그러나 이들 용어의 관계기호를 우선어인 'USE'와 비우선어인 'use for'로만 정의하고 있다.

또한 하위어를 상위어 종이 아니라, 등가어로 취급해 어휘의 수를 감소시키고 동시에 하위개념인 특정용어에서 하위개념을 나타내기 위하여 이 용하는 업워드포스팅의 경우도 특별한 관계지시기를 사용하지 않아 색인시 문제가 발생한다.

### 2.2.2 계층관계

상위개념 및 하위개념을 논리적으로 전개하는 순서를 나타내는 계층관계는 구조화되지 않은 일반용어와 시소러스를 구별하는 기본적인 특징으로 재현성과 적합성을 재고시키는 중요한 요소이다.

계층관계는 일반적으로 속관계, 사례관계, 전체/부분관계의 세 가지로 구분된다. ISO 표준에서는 상위어(BT)와 하위어(NT)로 구분하고, 속관계(BTG/NTG), 전체부분관계(BTP/NTP)와 같이 구분 표시해 줄 수 있다고 정의하고 있다. Z39.19에서는 사례관계를 나타내는 기호로 사례관계 상위개념어(BTI)와 사례관계 하위개념어(NTI)와 같이 지시할 것을 권하고 있다. 또한 계층관계의 지시기호로는 BT1(1차수준 상위개념어), BT2(2차수준 상위개념어), NT1(1차수준 하위개념어), NT2(2차수준 하위개념어)와 같은 관계기호를 사용할 수도 있고 규정하고 있다.

그러나 실제 대부분의 시소러스는 단일 계층 밑에 내용이 다른 용어를 NT로 통합하고 있다. 이

경우 단일계층에 속하는 용어간의 구분관계가 없어 탐색자가 하위어 포괄검색을 할 경우 동일한 수준의 하위어에서 여러 다른 속성을 나타내는 용어가 함께 처리되어 잡음이 많이 발생한다.

이를 해결하기 위한 방안으로 시소러스 표준에서는 Node Label(facet indicator)을 사용 할 수 있다고만 제시하고 있을 뿐, 노드레이블의 종류나 유형을 언급하지 않고 있다. 또한 현재 이를 사용하고 있는 시소러스의 경우도 계층관계에서 노드레이블만 제시할 뿐 실제 검색에서 사용하지 않고 있다.

### 2.1.3 관련관계

관련관계란 계층적 관계에 속하지 않으며, 개념적으로 밀접하게 관련되어 있으나 등가관계에는 포함되지 않는 용어 관계이다.

ISO 표준에서 제시하고 있는 관련관계로는 동일범주에 속하는 용어와 상이한 범주에 속하는 용어를 구분하고, 전자에는 자매어와 가족적인 혹은 파생적인 관계를 후자에는 학문분야·연구영역·연구의 대상과 현상, 조작·과정과 행위자 또는 도구, 행위와 그 행위의 결과, 행위와 그 수동체, 특성에 관한 개념, 기원에 관한 개념, 인과관계에 의해 연결된 개념, 사물과 그것의 반대 행위자, 개념과 그것의 측정단위로 구분하고 있으나 관계지시기호는 'RT'한가지로만 표시하고 있어 용어의 체계도라 할 수 있는 시소러스의 기능을 만족시키지 못하고 있다.

## 2.3 용어해설 및 부가관계

범위주기(scope note)란 색인어 내에서 색인의 의미를 나타내기 위해서 부속된 설명을 말하며, 정의(Definition)는 용어의 사전적 정의를 의미한다. ISO 표준에서는 이 둘은 함께 취급하고 있으며, 어떤 경우에 SN을 사용하는지, 단어로 기술하는지 문장으로 기술하는지, 문장으로 기술한다면 몇 단어 이내로 제한하는지 등에 관한 분명한 기준이 제시되지 않고 있으며, 현재 사용 중인 시소러스에서 여러 가지 형태로 나타나고 있다.

또한 시소러스 표준에서 언급하지 않지만 현재 각종 시소러스에서는 다양한 용어관계지시기호들이 사용되고 있다. 'INSPEC 시소러스'는 이전 용어를 지시하는 용어로 PT(prior term)을 사용하고 있으며, HN(이력주기), SA(도보라), CC(분류번호), DI(입력일자), DC(디스크립터 코드) 등 여러 시소러스에서 다양한 문자로 나타내고 있지만 이들 용어관계에 관한 지시기호가 ISO에서는 마련되어 있지 않다.

### 3. 시소러스 관계기호의 확장

#### 3.1 기본관계의 확장

##### 3.1.1 등가관계

###### ① 등가관계의 세분

시소러스를 사전, 특히 사상에 의하여 형성된 단어집이라는 측면에서 본다면, 동등관계를 단순히 비디스크립터에서 디스크립터로 안내하는 역할 이외에 비디스크립터로 사용된 용어의 정의를 표시할 필요가 있다. 예를 들어 완전동어어, 약어, 이형표기, 외래어 표기 등에 관한 지시기호를 첨가하여 용어사전의 기능을 할 수 있어야 할 것이다.

###### ② 반의관계

인지언어학에는 용어관계를 등위관계, 연어관계, 상하관계, 동의관계의 4가지로 구분하는데[임지룡, 1997], 이를 시소러스에 적용하면, 동의관계는 동등관계, 상하관계는 계층관계, 연어관계는 연관관계로 적용할 수 있다. 그러나 등위어(coordinate)와 대립어에 해당하는 반의관계는 일반적으로 동등관계로 취급하고 있을 뿐 반의관계를 나타내는 관계지시기호를 사용하지 않고 있다.

에치슨(Aitchison)은 '두개의 반대 의미를 갖고 있는 용어간에 확실한 구별이 있다면, 양쪽의 용어를 색인어로서 이용하며, 양자간에 참조어를 작성해야하며 그렇게 하지 않으면, 적합성이 저하된다'[Aitchison & Gilchrist, 1987]고 주장했다.

예를 들어 '흡연'의 탐색에서 흡연과 그 반의어인 금연에 대한 문헌이 항상 필요하다면, 이 두가지의 용어는 유사동의어로 취급하여 등가관계로 할 수도 있지만, 용어관계를 보다 명확히 하기 위해서는 별도의 관계기호를 사용할 수도 있을 것이다. 이 경우 반의관계를 나타내는 관계지시기호(AT: antonym)를 사용하면 필요한 정보를 검색하는 데 도움이 될 것이다.

###### ③ 고립어

Z39.19에서는 어떤 다른 디스크립터와도 관계가 없는 디스크립터인 고립어로 등록되는 디스크립터는 없어야 한다고 제시하고 있지만 실제로 시소러스에는 고립어로 등록되는 용어들이 많다. 이들에 관한 용어관계지시기호(OT: orphan term)도 제시되어야 할 것이다.

###### ④ 고유명사

시소러스에 따라서 다르지만, 특정 시소러스의 경우 인명이나 단체명, 회사명 등의 고유명사 파일을 관리하는 경우가 많다. 이 경우도 각각의 고유명사에 대한 관계지시기호가 필요할 것이다.

###### ⑤ 업워드포스팅

업워드포스팅의 경우도 분류에서 특정분류번호는 '~을 포함한다'고 한정하는 것처럼 특정한 용

어 아래에 지시기호(UP)를 주어 더 이상 세분하지 않는 업워드포스팅을 사용할 필요가 있을 것이다. 계층구조를 보다 적은 공간에 효율적으로 표시할 수 있는 접기기술(fold-in)을 사용할 수도 있다.

##### 3.1.2 계층관계

온라인 시소러스에서 계층관계를 이용해 확장 검색을 효율적으로 수행하기 위해서는 단일계층에 속한 array들을 개념단위로 구분하는 패킷방법을 이용하여 구분할 수 있을 것이다. 적절하게 고안된 패킷 프레임은 한 분야에서 기본적인 개념을 포착할 수 있으며, 색인어에 포함된 개념을 끌어내는 데 도움이 되기 때문이다.

시소러스에 포함된 용어들을 패킷개념으로 구분하기 위해서는 현재 일부 시소러스에서 단일계층에서의 형제 디스크립터를 구별하기 위해 사용하는 노드레이블이나 역할기호, 시소러스 표준에서의 용어구분, 분류에서의 구분카테고리, 인지언어학에서의 용어범주화, 지식표현 측면에서의 용어구분 등 구분방법을 이용한 패킷구분으로 단일 계층의 용어를 구별해 특정 기호로 표시할 수 있을 것이다.

랑가나단(Ranganathan)은 콜론분류표에서 모든 지식을 다섯 개의 기본 카테고리 즉, 속성, 물질, 에너지, 공간, 시간으로 범주화했고, 영국의 분류연구그룹(CRG)은 랑가나단의 PMEST를 확장해 다음과 같이 패킷을 구분했다[Anderson & Rowley, 1992].

- 사물, 실체, 생산물, 관념적 개념
- 부분, 구성요소, 구조
- 재료, 성분
- 속성, 과정, 행위
- 작용
- 작용기관
- 장소, 상태

1997년에 만들어진 브리스(Bliss)의 서지분류표 2판은 완전한 패킷분류 기본 프레임을 개발하기 위해서 만들어 졌다. BC2의 기본 카테고리는 실체, 부분, 속성, 행위, 행위자, 적용으로 구분하고 일반에서 복잡의 논리적 순서로 배열했다.

대표적인 패킷 시소러스인 'ROOT 시소러스'에서는 최소한 600개의 다른 패킷과 1,600개 이상의 패킷 구분이 있다. 그 중에 가장 많이 사용되는 노드레이블은 By equipment(82), material(64), purpose(55), part(52), instrument(52), component(51), application(46), property(43), product(30), form(29), process(28), location(26), design(23), derivative(20), shape(18), personnel(15), motive power(15), mode of operation(15), function(13), method(12) 등 이다.

'ROOT 시소러스'에서 패킷은 엄격하게 적용되는 않으나, 주제 분야의 논리적 배열을 지원하기

위한 유용한 도구로 사용하고 있다.

유네스코 시소러스도 각 관계 내에서 공통적인 특성을 갖는 용어를 한데 모으기 위해서 by application, by form, by material, by part, by use 등과 같은 패싯인디케이터를 사용한다.

그러므로 계층관계의 합리적인 확장을 위해서는 현재 시소러스에서 사용하고 있는 로드레이블과 다양한 용어범주화 기법을 이용해 계층관계 용어를 효과적으로 구분하고, 이들을 표현할 수 있는 관계지시기호가 마련되어야 할 것이다.

### 3.2 용어해설 및 부가관계

시소러스의 표준에서는 용어설명을 위해 범위주기(scope note)를 제시하고 있으며, 기존의 많은 시소러스에서도 이 기능을 사용하고 있다.

ISO 표준에서는 일반적으로 시소러스에는 사전에서 기재되어 있는 바와 같은 완전한 정의는 제시하지 않으나, 시소러스의 구조에 의해 제시된 의미를 보완하기 위해 한정된 정의가 필요한 경우가 있다고 했지만, 정의를 범위주기에 포함시키고 있다.

그러나 온라인 시소러스에서는 다양한 용어관계의 이용이 가능하므로 범위주기(SN) 정의(DF)를 분리시켜 범위주기에는 용어의 사용범위만을 지시하고, 정의가 필요한 경우에는 정의를 별도의 사전파일과 연계시켜 이용할 수 있도록 해야 할 것이다.

멀티미디어 환경에서의 시소러스는 단순히 용어아래에 용어를 설명하는 수준으로 그칠 것이 아니라 용어사전과 연결시켜 자세한 용어설명과 그 용어에 대한 이미지, 음성, 동영상 정보까지 제공할 필요가 있을 것이다. 이 경우 용어사전에 관한 정보는 별도의 파일로 두고 하이퍼텍스트 기능으로 연결시켜 이용자가 검색용어를 확인할 때 필요한 용어정보를 제공해 줄 수 있어야 할 것이다. 이때 각각에 관한 관계지시기호로는 이미지기호를 이용하는 것이 효과적일 것이다.

아울러 INSPEC시소러스에서와 같이 이전용어를 지시하는 PT(prior term), 우선어 등록시기에 관한 이력주기(HN), CC(분류번호, 혹은 분류코드), 입력일자(DI) 등 시소러스 표준에서 사용하고 있지 않은 다양한 문자기호로 용어관계를 표시할 수 있도록 해야 할 것이다.

### 3.3 시소러스가 갖고 있는 정보

시소러스가 본문정보검색시스템과 연결되어 시소러스를 이용한 자동색인을 수행하는 경우에는 색인용어와 시소러스를 매치시켜 시소러스가 갖고 있는 정보를 제공할 수 있다. 예를 들면 자동색인에 필요한 정보 즉 용어가 속한 주제, 용어의 계층

수준 다시 말해 특정성과 전체시스템에서의 용어 정보로 빈도, 관련용어, 특성 등을 가지고 있다면 자동색인에서 훨씬 효과적인 색인어를 생성할 수 있을 것이다.

시소러스 용어에 이러한 정보를 포함하는 것이 자동색인의 작성에 필요하며, 또한 시소러스 용어에 해당되는 색인어의 수를 제시하는 것이 이용자의 효율적인 정보탐색에 필요할 것이다.

## 4. 결론

이상에서 시소러스 구축을 위한 국제표준지침인 ISO 2788을 중심으로 한 시소러스 작성지침을 중심으로 현재 사용되고 있는 시소러스의 용어관계에 대한 문제점을 살펴보고, 온라인 환경에서 보다 효율적인 시소러스를 구축하여 이용할 수 있는 용어관계의 확장에 관해서 살펴보았다.

정보검색 환경의 변화에 따라 시소러스는 색인 과정에서 단순히 용어관계를 지시하는 통제용어집의 환경의 사용자 지향적인 측면에서 벗어나 온라인 정보검색환경에서 이용자 중심의 체제로 변화해야 할 것이다. 이를 위해서는 시소러스의 용어관계를 확장하고, 멀티미디어 기법 등 다양한 기술이 도입되어야 할 것이다.

## 참고문헌

- 임지룡(1997). 인지의미론. 서울 : 탐출판사.  
최석두, 정동열 공역(1994). 시소러스 개발지침, 서울 : 문헌정보처리연구회.  
Anderson, J.D. & F.A. Rowley(1992). "Building End-user Thesauri from Full-text," In *Proceedings of the 3rd ASIS SIG/CR Classification Research Workshop*, Vol.III, Pittsburgh : Learned Information, pp.1-13.  
Aitchison, J. and A. Gilchrist(1987). *Thesaurus Construction : a Practical Manual*, 2nd ed. London: Aslib.  
BSI(1988). *Root Thesaurus*. 3rd ed. British Standards Institution.  
Schmitz-Esser, W(1991). "New Approaches in Thesaurus Application," *International Classification*, Vol.18, No.3, pp.143-147.  
Stern, A. and N. Rischette(1992). "On the construction of a super thesaurus based on existing thesauri", In *Tools for Knowledge organization and human interface*. Vol.2 Frankfurt : Indeks Verl, pp.134-144.