

시소러스의 표시형식에 관한 연구

A Study on Display Format of Thesaurus

최 석 두 (이화여대 문헌정보학과)

Suk Doo Choi
(Dept. of Lib. & Inf. Sci., Ewha Womans Univ.)

시소러스는 다양한 형식과 순서로 표시되며, 인쇄매체인 경우와 화면인 경우에 따라 형식이 달라지게 된다. 온라인시소러스에서는 여러 가지 표시방법을 제공할 수 있으며 즉시 관련색인과 연관시켜 참조할 수 있으나 인쇄시소러스는 시간과 공간의 제약이 있으므로 표시방법을 달리 고려하여야 한다. 특히 인쇄시소러스는 어떤 개념을 찾았을 때 그 개념과 관련된 모든 정보를 한꺼번에 볼 수 있어야 한다. 관련부분을 다시 참조하는 것은 온라인시소러스에 비해 매우 어렵기 때문이다.

본 연구에서는 이와 같은 관점에서 인쇄시소러스의 각 표시방법에 대한 장단점을 고찰하고 보다 바람직한 표시형식을 제안하고자 한다.

1 서론

인쇄시소러스에는 일반적으로 본표(자모순시소러스)와 본표의 보조도구로서 순환색인, 주제색인, 계층색인을 갖는 것이 보통이다. 순환색인은 주로 KWIC이나 KWOC색인을, 주제색인은 상세한 분류번호 혹은 포괄적 분류번호를, 계층색인은 최상위개념어구조나 양방향 계층구조를 사용하여 이용자의 검색편의를 고려하고 있다.

인쇄시소러스를 이용하는 경우에는 가능한 한 보조색인을 이용하지 않고 직접 본표만으로 검색작업을 끝낼 수 있는 형식이 좋을 것이다. 본고에서는 기존 인쇄시소러스 표시형식의 장단점을 고찰하고, 보다 바람직한 인쇄 및 온라인시소러스의 표시형식을 제안하고자 한다.

2 기호 및 관계지시기호

시소러스에서 사용하고 있는 기호 및 관계지시기호는 매우 다양하다. 일반적으로 문자와 기호를 사용하는 두 가지 경우로 대별할 수 있다.

문자를 사용하는 경우를 보면, BT/NT(상위개념어/하위개념어), BTG/NTG(상위개념어(속)/하위개념어(속)), BTI/NTI(상위개념어(사례)/하위개념어(사례)), BTP/NTP(상위개념어(전체-부분)/하위개념어(전체-부분)), GS(속구조), HN(이력주기), RII(관련어), SEE(use), SN(범위주기), TII(최상위개념어), U(use), UF(used for), UFA(used for...and...), USE+(use...and...), X(see from-UF), PT(prior term), SA(도보라), CC(분류번호), DI(입력일자), DC(디스크립터코드) 등을 사용하고 있다.

기호를 사용하는 경우를 "ROOT시소러스"(ROOT, 1988)에서 주로 사용하며, =(UF), <(상위

개념어), >(하위개념어), -(관련어 혹은 하위개념어를 갖는 용어), *<(상위개념어(다른 부분)), *>(하위개념어(다른 부분)), *(관련어(다른 부분)), → (USE, 혹은 신규등록어), +(합성어), =*(합성어의 동의어), α(array term) 등을 사용하고 있다.

이들 기호 및 관계지시기호의 배열순서는 시소러스에 따라 매우 다양하다. 디스크립터 다음에 <분류번호, 입력일자, 포스팅수, 이력주기, 디스크립터코드 등>에 대한 정보를 기술하는 것이 바람직하다. 그 다음에 SN, UF류, BT류(BTG, BTI, BIP 혹은 BTh, ..., BTI), TT, NT류(NTG, NTI, NTP 혹은 NT1, ..., NTn)를 나열하며(대신 CS를 사용해도 좋다), 이어서 RT를 나열한다. 그리고 필요하다면 SA, PT를 기술한다. 기호를 사용하는 경우에도 이와 같은 기준에 따라 배열하는 것이 바람직할 것이다. USE류는 독립적으로 사용되므로 순서와는 관계가 없다.

3 기존 표시형식

3.1 평면시소러스구조 : 이 형식은 전통적으로 사용되고 있는 형식이다. 모든 디스크립터가 자모순으로 정렬되고, 각 디스크립터 마다 범위주기(SN), UF참조, 1차수준 상위개념어(BT), 1차수준 하위개념어(NT), 관련어(RT)를 표시하고 있다. 이 형식의 예는 ERIC시소러스(ERIC, 1990)에서 볼 수 있다. 다음 예와 같이 일반적으로 하위개념어는 자모순으로 정렬되고 노드레이블은 용어들을 카테고리별로 그룹화하는 데 사용된다.

```

시소러스
BT 통제어휘
NT <표시형식>
    자모순시소러스
    체계시소러스
    폐쇄시소러스
    <크기>
    메트로시소러스
    미니시소러스
    <언어수>
    다언어시소러스
    단일언어시소러스
NTP 관계지시기호
    기입어
    디스크립터
RT 전거파일
.....
    
```

이 형식은 공간을 적게 차지하는 장점은 있으나 관련계층을 완전히 알기 위해서는 한번에 하나씩 모든 디스크립터들을 찾아보아야 한다.

3.2 속(屬) 혹은 계층구조 : 이 형식은 BTh, ..., BTI(n차, ..., 1차수준 상위개념어) 및 NT1, ..., NTn(1차, ..., n차수준 하위개념어)과 같은 특수한 기호법을 채택한다(INIS, 1987). 이 형식은 BT/NT기호법을 사용하지 않고 들이쓰기와 마침표, 콜론과 같은 구두점을 이용하여 계층의 수준을 표현하기도 하며(NASA, 1988), 이를 "속구조"(generic structure)라고 한다.

```

data
  BT1 information
  NT1 numerical data
  NT2 compiled data
  NT2 evaluated data
  .....
  RT circa
    comparative evaluations
  .....

hydrodynamics
  GS fluid mechanics
    . fluid dynamics
    .. hydrodynamics
    ... elastrohydrodynamics
    ... electrohydrodynamics
    ... magnetohydrodynamics
    . hydromechanics
    .....
  RT ballast (mass)
  α dynamics
  .....
    
```

이 형식은 용어와 관련된 최상위개념어부터 가장 세분된 개념들을 한꺼번에 식별할 수 있는 장점이 있다. 다만 이 구조내에서의 계층은 계속 반복되기 때문에 많은 공간을 차지하게 된다. 인쇄시소러스에서는 바람직하지 못하나 계층이 3~4수준 이내이며 디스크립터의 수가 적으면 고려할 만한 형식이다. 그러나 용어가 같은 수준내에서 자모순으로 배열되기 때문에 계층이 복잡한 경우에는 특정 용어의 위치를 파악하기가 매우 어렵다.

3.3 트리구조 혹은 폐쇄시스템 : 이 형식에서는 각 용어가 분류기호나 행번호를 가지고 있으며, 이것이 이용자를 자모순표시로부터 다른 방식으로 배열된 전체 계층표시로 안내하게 된다. "Medical Subject Headings"(MeSH, 1989)는 분류기호를 갖는 형에 속한다. "Art and Architecture시소러스"(Art, 1990)는 분류기호도 가지고 있으나 행번호를 갖는 계층표시형에 속한다.

nervous systems diseases	C10
autonomic nervous systems disease	C10.120
barre-leeou syndrome	C10.120.253
dysautonomia, familial	C10.120.309
.....	
central nervous system disease	C10.228
brain diseases	C10.228.140
.....	
RK197 <communications structures>	
RK198 communications buildings	
RK199 computer centers	
RK200 publishing offices	
RK201 <telecommunications buildings>	
RK202 broadcasting stations	
RK203 radio stations	
.....	

이 형식은 관련개념을 한꺼번에 참조할 수 있으며, 온라인탐색에서 효율성 있게 이용할 수 있다. 그러나 관련개념을 분류기호나 행번호로 찾아가야 한다는 것이 매우 불편하다. 또한 관련개념들이 다른 개념의 상·하개념속에 산재되어 있을 때 이들을 한꺼번에 참조하기가 어렵다.

3.4 그래픽표시 : 이 형식은 논리적으로 트리구조와 같지만 기호법을 갖지 않는다. "Euratom시소러스"(Euratom, 1967)가 이 형에 속한다. 이 형식은 시각적인 효과가 있어서 관계를 이해하기 쉬우나 공간을 많이 차지할 뿐만 아니라 생생하거나 유지관리하는 것이 어렵다. 또한 너무 많은 관련어를 갖거나 너무 많은 계층수준을 갖게 되면 내용을 점점 이해하기 어렵게 되는 단점이 있다.

3.5 최상위개념어 구조 : 이 형식에서는 계층의 최상위개념어라는 코드(TT, top term)를 가지면서 평면형식에서 갖고 있는 모든 관계를 자모순 표시에 포함한다. 최상위개념어란 속관계에서의 가장 넓은 개념어와 같다. 최상위개념어는 자모순으로 다른 곳에 배열되어 있으며, 자신의 모든 하위개념어의 각 수준을 갖고 있는 최상위개념어로 안내한다. "INSPEC시소러스"(INSPEC, 1989)가 이 형식을 채택하고 있다. 다음의 첫번째 예는 자모순 표시로 하위개념 1수준만을 보이고 있으므로 이를 보조하기 위하여 두번째 예와 같은 완전한 계층을 나타내는 표시방법을 병용하는 것이 보통이다.

libraries

- NT library automation
- BT computer applications
- TT computer applications
- RT cataloging
- information centres
- information needs
- information science
- information services
- CC C7210L
- DI January 1972
- PT information services

computer applications

- . administrative data processing
- .. distributive data processing
- ... goods dispatch data processing
- ... stock control data processing
- ... warehouse automation

-
- . libraries
- .. library automation
-

최상위개념어구조는 속구조보다 공간을 경제적으로 사용할 수 있으나 평면구조와 같이 다시 찾아보아야 하는 불편이 있다. 또한 두번째 예에서는 RT를 보여주지 못하고 있다.

3.6 양방향 계층구조 : 이것은 일반적으로 평면형식 시소러스에 첨가하는 구조이다. 일반적으로 관계기호없이 계층을 알아보기 위하여 들이쓰기와 마침표, 콜론 등을 이용하여 상·하위개념어의 모든 수준을 표시한다. 이 형은 "ERIC시소러스"(ERIC, 1990)의 보조색인에서 찾아 볼 수 있다.

- :::: services
- ::: information services
- :: library services
- : library technical processes
- library acquisition**
- . library material selection

양방향계층구조는 논리적으로는 속구조와 같다. 평면형식과 통합하면 이 양방향 계층구조는 최상위개념어 구조보다 더 많은 공간이 필요하게 된다.

3.7 포괄적 계층구조 : 이 형식은 자신의 상위개념어와 하위개념어에 대하여 깊이에 상관없이 모두 자모순으로 배열하는 방법이다. 이와 같은 형식은 TEST시소러스(TEST, 1967)에서 볼 수 있다. 다음의 첫번째 예와 같은 계층구조를 갖고 있을 때 이를 포괄적 계층구조로 표현하면 두번째 예와 같다.

safety

- . accident prevention
- . aviation safety
- . fire safety
- .. fire prevention
- .. fire protection
- . missile safety
- . nuclear reactor safety
- . nuclear weapons safety
- . range safety
- . safety and arming (ordnance)

safety

- UF safety measures
- NT accident prevention
 - aviation safety
 - fire prevention
 - fire protection
 - fire safety
 - missile safety
 - nuclear reactor safety
 - nuclear weapons safety
 - range safety
 - safety and arming (ordnance)
- RT-accidents
 - backup systems
 -

이 형식은 속 혹은 계층구조가 갖고 있는 특정 용어의 확인이 어려운 단점을 보완하고 있으나 계층을 원하는 경우 계층구조를 보아야 한다. 이를 도와주기 위하여 하위계층을 갖는 용어 앞에 하이픈을 부기하고 있다.

4 표시형식의 제안 및 결론

일반적으로 인쇄시소러스의 요건은 우선 "어떤 개념을 찾았을 때 그 개념과 관련된 모든 정보를 한꺼번에 볼 수 있어야 한다"는 것이다. 온라인시소러스에서는 관련용어들의 관계에 대하여 전술한 여러 가지 표시방법을 준비하여 두고 이용자에게 선택하게 할 수 있다. 그러나 인쇄시소러스는 시간과 공간의 제약이 있으므로 표시방법을 달리 고려하여야 한다. 다만 온라인시소러스로 인쇄시소러스를 보완하는 경우나 그 반대인 경우에는 인쇄시소러스의 형식이 달라져도 좋을 것이다.

4.1 분야별 시소러스 : 일반적으로 분야별 시소러스는 용어간의 관계가 밀접하고 거대한 트리구조에 가까운 형태를 형성하고 있다. 이와 같은 시소러스에서는 포괄적 계층구조가 좋을 것이다. 속구조 혹은 계층구조는 용어가 같은 수준내에서 자동으로 배열되기 때문에 계층이 복잡한 경우에는 특정 용어의 위치를 파악하기가 매우 어렵다는 단점이 있기 때문이다. 또한 이용자가 어떤 용어를

찾았을 때 그 용어와 관련된 정확한 계층을 필요로 하는 경우는 드물며, 포괄적으로 하위개념어와 상위개념어라는 사실로서 거의 충분하기 때문이다. 다만 계층을 원하는 이용자를 위하여 하위계층이 있다는 정보를 용어 앞에 하이픈으로 표시할 필요가 있으며, 계층구조색인을 준비해야 할 것이다. 최상위개념어(TT)를 부기함이 바람직 할 것이다.

4.2 범용시소러스 : 일반적으로 범용시소러스는 용어의 수는 많으나 용어간의 관계가 밀접하지 않아 계층의 수준이 깊지 않은 특성을 갖고 있다. 이와 같은 범용시소러스에서는 포괄적 계층구조를 가지면서 계층구조색인까지 본표에 포함시키는 것이 바람직하다고 생각된다. 전술한 포괄적 계층구조의 예를 다시 표현하면 다음과 같다.

safety

- UF safety measures
- NT accident prevention
 - aviation safety
 - fire prevention
 - : fire safety
 - fire protection
 - : fire safety
 - fire safety
 - . fire prevention
 - . fire protection
 - missile safety
 - nuclear reactor safety
 - nuclear weapons safety
 - range safety
 - safety and arming (ordnance)
- RT-accidents
 - backup systems
 -

이 형식은 계층구조색인을 가진 필요가 없으며 한 번 검색함으로써 필요한 정보를 모두 얻을 수 있다. 이 형식에서도 최상위개념어(TT)를 부기함이 바람직 할 것이다.

참고문헌

국립중앙도서관(1994). 「주제명 검색요어집」.
 Art(1990). *Art and Architecture Thesaurus*.
 ERIC(1990). *Thesaurus of ERIC Descriptor*.
 Euratom(1967). *Euratom Thesaurus*. Part II.
 IAEA(1987). *INIS Thesaurus*.
 INSPEC(1989). *INSPEC Thesaurus*.
 MeSH(1989). *Medical Subject Headings*. Suppl.
 NANA(1988). *NASA Thesaurus*.
 ROOT(1988). *ROOT Thesaurus*.
 TEST(1967). *Thesaurus of Engineering and Scientific Terms*.