# 시소러스 브라우저의 설계

## A Study on the Design of a Thesaurus Browser

이 나 니
(이화여대 문헌정보학과 대학원)

Nanee Lee
(Dept. of Libr. & Inf. Sci., Ewha Womans Univ.)

정보검색과정에서 온라인 시소러스는 이용자가 생각하고 있는 개념어를 시스템의 색인어로 보다 정확하게 표현하는 동시에, 관련구조를 통해 탐색의 범위를 확장·축소할 수 있게 한다. 온라인 시소러스를 충분히 활용하기 위해서는 이용자가 시소러스의 구조에 쉽게 접근하여 계층·관련구조를 충분히 살피고, 용어 사이를 자유롭게 탐험할 수 있는 장치가 필요하다. 본 연구에서는 기존의 시소러스 표시방법의 장단점 등에 근거하여 정보검색과정에서 이용할 수 있는 시소러스 브라우저를 事前탐색 기능, 시소러스 구조의 표시 기능, 탐색문 형성 기능으로 나누어 설계하였다.

## 1. 서론

시소러스는 색인, 검색자가 사용하는 자연어를 보다 통제된 시스템 언어로 변환하는 어휘통제장치임과 동시에, 디스크립터 상호간의 관계 및 도입어와 디스크립터의 관계를 구조적으로 표현하여 해당 주제분야에 대한 개념관계도, 참조시스템을 제공하고, 이용자의 주제탐색을 안내하는 역할을 한다.

특히, 검색과정에서의 온라인 시소러스는 하부구조, 표시형식, 기능 등이 올바르게 구성되기만 하면, 이용자의 정보검색과정에서 항상 이용할 수 있고, 다양한 표시형식을 통해 관련구조를 쉽게 탐색할 수 있다는 장점이 있다. 본고에서는 이용자의 정보검색행위, 기존의 시소러스 표시형식의 장단점 등에 기초하여 이용자가 정보검색과정에서 이용할 수 있는 시소러스 브라우저를 설계하고자 한다.

## 2. 시소러스 브라우저

### 2.1 설계의 원칙

인쇄 시소러스의 표시형식은 어떤 개념을 찾았을 때 그 개념과 관련된 모든 정보를 한꺼번에 볼 수 있는 형식이 바람직하다. 이 원칙은 온라인 정보검색에 이용되는 시소러스 브라우저에서도 마찬가지이나, 다음과 같은 추가적인 요구사항을 생각할 수 있다.

첫째, 시소러스 브라우저의 초기접근에서 이용자가 입력한 용어에 대하여 그 의도를 반영하는 주제개념어 그룹으로 안내함으로써, 시스템에의 접근실패로 인해 탐색을 중단하지 않도록 한다. 둘째, 이용자의 참조용어를 중심으로 한 다양한 관련, 계층구조를 보여주어 이용자가 원하는 용어를 선택하고 탐색의 범위를 확

장할 수 있도록 한다. 셋째, 시소러스 구조에 나타나는 모든 용어를 하이퍼텍스트 방식으로 연결하여 이용자가 최소한의 키입력으로 참조 용어를 변경하고, 관련구조를 충분히 파악할 수 있도록 한다. 넷째, 시소러스의 구조를 제공할 때 이용자가 관련분야의 개념구조를 이해하고 기억하기 쉽도록 하며, 자신의 탐색경로를 충분히 이용할 수 있도록 한다. 다섯째, 시소러스 브라우저와 정보검색시스템을 연결하여 시소러스 브라우저를 완전히 빠져나가지 않고도 정보검색시스템의 정보를 이용한 탐색을 충분히 할 수 있도록 한다. 또한, 다국어 시소러스, 복수 시소러스를 이용하는 시스템에서도 활용하기 쉽게 한다.

### 2.2 사전탐색 기능

주제탐색과정에서 이용자는 자신이 생각하는 개념을 시스템의 언어와 일치시키는 데 많은 어려움을 겪는데, 이 과정에서 많은 어려움을 겪으면 이용자는 좌절하여 탐색을 중단하게 된다(Markey, 1984). 탐색의 초기에 시소러스는 이용자에게 관용적이어야 하며, 이용자가 한두 개념어를 입력하는 것만으로 용어의 의미네트웍에 접근할 수 있어야 한다. 이 때, 시소러스의 용어와의 매치는 단순한 일치가 아니라 이용자의 의도를 반영할 수 있는 용어와의 연결을 필요로 한다.

이를 위해 시소러스 브라우저에서는 순열표시를 이용한 事前탐색기능을 제공하여 시소러스 구조를 직접 탐험하기 전에 이용자가 생각하고 있는 개념을 구체화시키거나 확인하며, 이를 통해 시소러스 구조로 접근할 수 있도록 한다.

순열표시는 KWOC 방식의 키워드 리스트와 KWIC 리스트의 두 가지로 구성되며, 대상어휘는 시소러스의 우선어, 비우선어, 외국어 표기, 약어 및 용어설명, 범위주기에 나타나는 어휘이다. 또한, 정보검색시스템이 어휘사전 등과 연결되어 있는 경우에는 시소러스의 보라 참조에 나타나지 않으나 시소러스 어휘의 어형이 변화된 단어 역시 키워드 리스트의 대상으로 한다. 키워드 리스트에는 대상 어휘에 나타나는 모든 단일어·복합어를 포함시킨다. 특히, 한글 키워드는 띄어쓰기 뿐 아니라 복합개념어 내에서 의미를 가지고 있는 모든 단일개념어 단위로 추출하여 접근점을 제공한다.

브라우저는 이용자가 입력한 어휘와 같거나 가장 근접한 키워드, KWIC 용어를 <그림 1>과 같이 보여주며, 순열색인에서 이용자가 선택한 우선어는 시소러스 구조의 화면표시로 연결된다. 순열표시에서 이용자가 비우선어나 외국어를 선택하였을 때 시소러스로의 연결은 자동적으로 이에 해당하는 우선어로 하며, 연결사실을 이용자에게 알리는 메시지를 표시한다.

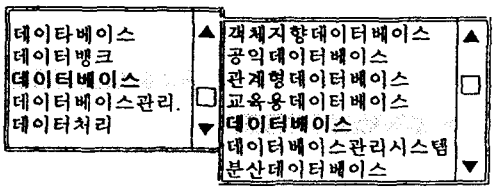


그림 1. 순열색인의 표시

### 2.3 시소러스 구조의 표시 기능

키워드 리스트로부터 시소러스 구조에 접근한 이용자는 다양한 형태로 탐색을 하게 되는데, 다음과 같은 행동을 생각할 수 있다(Bates, 1986).

(a) 같은 주제를 나타내는 다른 용어들을 표시하라(예: 동의어).
(b) 관련되는 다른 개념을 표시하라(예: 관련어).
(c) 이 개념영역에 대한 주제계층을 표시하라(예: 계층구조, 트리).
(d) 이 주제영역에서 상위의 개념들을 표시하라(예: 상위어).
(e) 이 주제영역에서 하위의 개념들을 표시하라(예: 하위어).
(f) 이 용어의 정의를 표시하라.
(g) 이 용어가 문헌의 색인에 어떻게 사용되고 있는지 표시하라(예: 범위주기).
(h) 검색된 문헌의 색인에 사용된 다른 색인어(co-indexing terms)를 표시하라.
(i) 이 주제에 관련된 문헌 중 일부를 표시하라.
(j) 검색된 문헌과 비슷하거나 관련있는 문헌을 표시하라.

(k) 지금까지 탐색·참조한 용어의 리스트를 보이고, 현재 참조어가 속한 시소러스 구조 내의 위치를 표시하라.

(a)~(f)는 시소러스 구조의 표시 기능에 속하며, (g)~(k)는 탐색문 형성을 지원하기 위한 기능에 속한다. 이와 같은 요구는 개별적으로 이루어지기도 하고, 복합적·연속적으로 이루어질 수도 있다.

설계한 브라우저에서 시소러스 구조의 표시는 <그림 2>와 같이 참조용어의 계층구조, 포스팅수와 참조용어의 관련어를 기본값으로 표시하며, 보다 상세한 관계정보를 요구하는 이용자에게는 외국어표기, 용어의 정의 및 그림, 범위주기 등을 추가적으로 제공한다.

현재 포스팅이 없는 용어는 화면에서 흐리게 표시하여 색인어로 사용되고 있는 용어와 구별되도록 한다. 계층표시는 최상위개념어부터 참조용어의 1수준 하위개념어까지의 전체계층을 표시하되, 일부를 생략한 표시방법을 사용한다. 최상위개념어는 상단의 고정된 위치에 표시하고, 참조용어와 n수준 상위개념어 관계에 있는 용어는 각 용어에 대한 1수준의 하위계층을 전개하며, 참조용어 역시 1수준의 하위계층을 전개한다. 계층구조에 속해있는 나머지 용어는 하위계층을 생략하고, 생략된 하위계층이 있는 용어는 ⊕ 기호를 부기한다. 반대로 하위계층이 펼쳐져 있는 용어는 ⊖ 기호를 부기하여 구분한다. 생략된 계층은 ⊕ 기호를 누름으로써 1수준의 하위계층을 펼쳐 볼 수 있으며, 펼쳐져 있는 계층은 ⊖ 기호를 누름으로써 접을 수 있다.

참조용어와 참조용어의 1수준 하위개념어들은 강조체로 표시한다. 참조용어가 같은 최상위개념어 아래 두 번 이상 나타날 경우에는 계층수준이 깊은 곳의 하위계층만을 전개하고, 나머지는 하위계층은 생략한 채 용어를 강조체로 표시한다. 참조용어가 둘 이상의 계층구조에 속한 경우에는 관련구조를 표시하기 전에 최상위개념어, 1수준의 상위개념어, 참조용어를 표시하여 이용자의 관심대상이 어느 계층에 속해 있는지 결정하게 한다. 이용자가 관심 계층을 선택하면 해당 계층구조를 표시하되 다른 최상위개념어가 있다는 것을 화면에 표시하여, 이용자가 최상위개념어를 변경하고 다른 계층구조를 참조할 수 있도록 한다. 단, 이용자가 최상위개념어 리스트의 확인을 원하지 않는 경

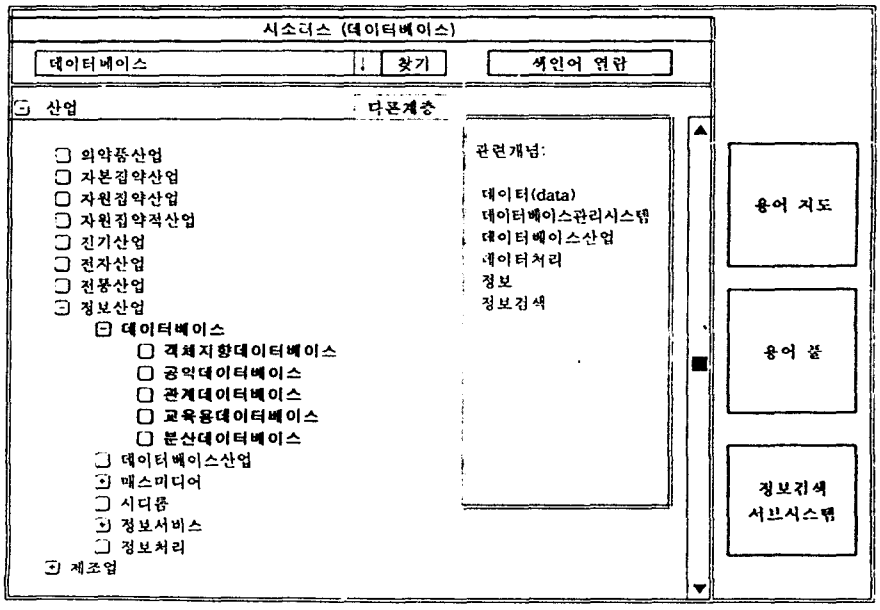


그림 2. 시소러스 구조의 표시방법

우에는 다수의 계층구조 중 처음 만나는 계층을 먼저 참조하고, 필요한 경우에만 다른 계층으로 이동할 수 있도록 한다.

시소러스 브라우저의 화면에 나타나는 모든 용어는 하이퍼텍스트 방식의 연결을 갖는다. 즉, 계층관계에 속한 용어, 관련어, 용어정의·해설에 나타나는 용어 중 시소러스에서 우선어·비우선어로 사용되는 용어는 모두 링크를 가지고 있어, 마우스의 클릭과 같은 방법으로 이들 용어를 선택하면 참조용어를 변경하고 새로운 참조용어의 구조를 살필 수 있다. 현재 화면에 표시되지 않은 용어는 직접 입력하거나 색인어 리스트에서 선택함으로써 참조용어의 변경이 가능하며, 시소러스 구조에 나타나는 용어 중 이전에 참조어로 이용되었던 용어는 색상을 달리하는 방법 등을 통해 탐색이력으로 활용할 수 있게 한다.

시소러스 브라우저의 계층 및 관련어 표시와는 별도로 시소러스의 개념관계도를 그래픽표시를 채용하여 표시한다. 개념관계도는 현재의 참조용어를 중심으로 한 전체 개념구조를 나타내며, 이용자가 탐험해 온 시소러스 용어와 이들 사이의 관계, 현재의 위치는 강조체로 표시한다.

### 2.4 탐색문 형성 기능

정보검색시스템의 데이터베이스와 연결된 시소러스 브라우저에서는 이용자의 탐색이력 기록, 이용자가 참조/선택한 용어 풀의 유지, 이용자가 선택한 용어를 이용한 검색식의 자동생성, 시소러스의 참조용어를 이용한 부분탐색과 같이 탐색문 형성 지원 기능을 제공하여 이용자의 탐색을 돕는다.

## 3. 활용방안 및 결론

시소러스의 원활한 이용을 위해서는 시소러스 브라우저의 대상이 되는 시소러스, 이용대상 등에 따라 전술한 표시형식·기능에 다양한 선택기능을 두어 환경에 맞게 이용하도록 한다. 예를 들어, 계층구조의 표시방법은 전체계층의 일부를 생략한 방법을 기본으로 하지만, 이용자의 환경이나 전문성 등에 따라 1수준의 상하위개념어만을 표시하거나 전체계층을 전개하여 표시할 수도 있다.. 또한, 참조용어의 관련어만을 표시하는 대신 시소러스 용어가 가진 상세한 정보를 표시할 수도 있다.

대상 시소러스의 특성에 따른 활용 역시 가능하다. 예를 들어, 시소러스 브라우저의 대상이 2개 국어 이상을 사용하는 다국어 시소러스인 경우에는 디폴트로 사용할 언어를 설정하고 이에 따라 화면표시를 한다. 즉, 사전탐색을 위한 순열표시는 사용되는 모든 언어로 순열색인을 작성하여 제공하며, 이용자가 선택한 키워드에 대한 시소러스 구조로의 연결은 일반적인 시소러스의 경우와 같이 처리한다. 시소러스 구조의 화면표시는 디폴트로 설정된 언어로 표시하며, 상세한 정보를 표시하는 경우에 한하여 상하위개념어, 관련어, 등가어에 대한 다른 언어를 병기한다. 이 때 관련어, 같은 계층수준의 용어는 디폴트 언어의 자모순으로 배열한다. 이용자가 선택한 용어에 대한 검색식은 해당 용어의 모든 언어 표기를 포함할 수 있도록 생성하는 것을 원칙으로 한다. 이를 위해 시스템의 내부적으로는 시소러스 용어의 모든 언어 표기에 대한 링크를 가지고 있어야 한다.

## 참고문헌

한국문헌정보학회(1994). 『주제명검색요어집』. 서울: 국립중앙도서관.

Bates, M. J.(1986). Subject access in online catalogs: A design model. *Journal of the American Society for Information Science, 37*, 357-76.

Markey, K.(1984). *Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs*. Dublin: OCLC.