

Thesaurus에 관하여

연세대학교 의학도서관

김 나 원

Thesaurus

Na Won Kim

Yonsei Medical Library

서 론

정보화시대를 맞이하여 각 분야의 각종 정보를 축적, 검색하기 위한 database의 개발이 절실히 요청되고 있고, 그러한 요구에 따라 문헌 database가 대량으로 구축되면서 효율적인 정보검색도구에 대한 필요성이 증대되고 있다.

이런 추세에 따라 database를 검색함에 있어 동일한 주제의 분산과 상이한 주제의 혼합을 방지함으로써 시스템의 용어와 이용자의 용어를 연결시켜주는 역할을 하는 시소러스의 이용이 불가결하게 되었다.

시소러스는 색인작업시와 정보검색시 필요한 참고도구로서, 현재에 와서는 후자의 역할에 대한 중요성이 증대되고 있다.

즉, 용어의 통제 외에도 용어간의 계층적 관계를 이용하여 보다 포괄적인 탐색어를 선택하거나 아니면 보다 특정한 탐색어를 선택함으로써 검색되는 문헌의 수를 조절할 수 있으며, 또는 관련된 다른 용어를 탐색어로 사용하여 보다 광범위한 탐색을 할 수도 있다.

그러나 현재 우리나라에서는 교육, 노동경제, 농업경제 등 극소수 분야에서 외국의 시소러스를 번역한 것 외에는 시소러스의 개발 사례가 전무한 실정이다. 연세대 의학 도서관 역시 고유 시소러스를 아직 구축하지 못하였으나, 의학도서관의 특수성에 맞추어 의학, 건강, 간호 분야와 관련된 내용으로, 문헌의 색인작업·목록작성·검색에 사용되고 있는 시소러스인 MeSH를 참조하고 있다.

전산화가 이미 이루어진 환경에서 색인 작업시 뿐만

아니라 검색에 있어서 그 효율성을 증대시키기 위해서는 도서관 전산프로그램을 이용한 시소러스의 구축이 시급한 실정으므로, 이러한 요구에 따라 기본 작업으로서 시소러스에 대한 이론적 연구를 시도하게 되었다.

II. 본 론

1. Thesaurus의 정의

회랍어를 어원으로 하여 사전·백과사전과 같은 지식의 보고를 의미하며, 1852년 P.M. Roget의 "Thesaurus of English word and phrases"에서 'thesaurus'라는 단어가 처음 사용되었다. 그후 1950년대 후반부터 60년도에 이르러서는 주로 미국에서 색인, 초록지가 컴퓨터 database화 됨에 따라 각종 시소러스가 작성되기 시작하면서 시소러스가 정보검색 시스템의 어휘를 통제하기 위해 본격적으로 사용되기 시작하였다.

구조적인 면에서보면 상위개념의 용어와 하위개념의 용어를 의미론적으로 밝힌 어휘집이며, 기능적인 면에서는 색인자 또는 정보이용자가 자연어를 통제어로 변환하는데 사용되는 어휘집으로서 database, 책자형 색인, 목록류 등 정보검색시스템에서 이용된다.

2. Thesaurus의 내용과 구성

Thesaurus의 내용이 색인이나 검색에 미치는 영향을 살펴보면 첫째, 컴퓨터에 의한 검색을 통해 조합에 의한 검색을 즉시에 할 수 있으며 둘째, 검색의 특정성을 결정하여 주며, 셋째로는 특정성이 높은 디스

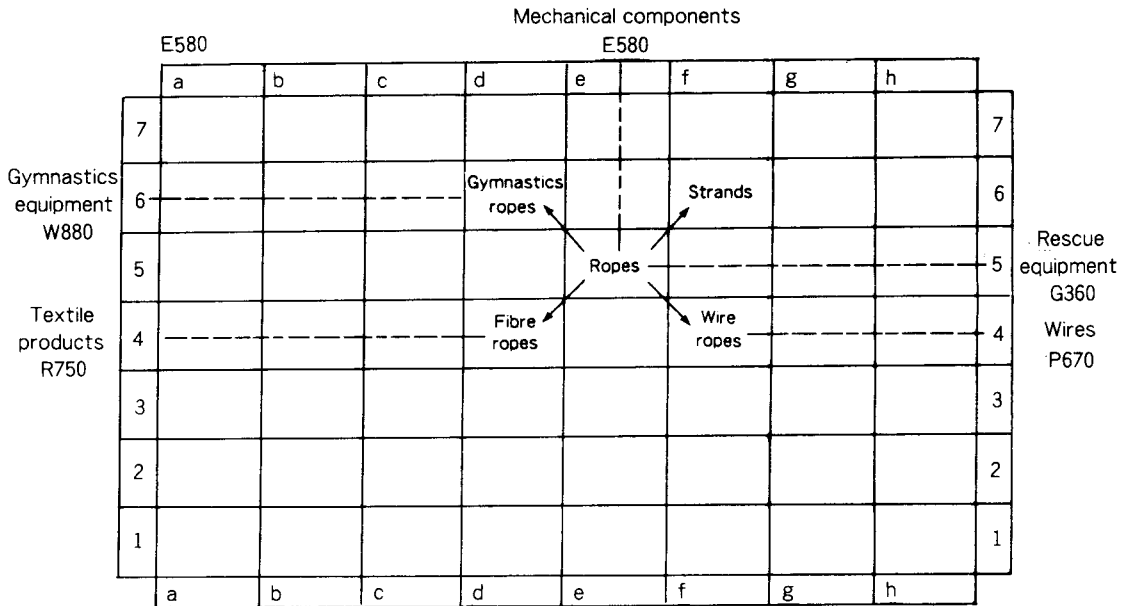
여기에 이어 인쇄되고, 관계는 연결선에 의해 지시된다. 일반적으로 도식상에서는 상위어나 하위어, 관련어를 구별하지 않는다. 그리고 화살표 그래프는 특정의 주제분야와 그 하위분야를 나타내는 차트형식의 표시로서 최상위어는 굵은 문자로 차트의 중심위치에 놓여

진다. 터미노그래프(박스차트)경우 우선어는 굵은 선으로 둘러싸여 있고, 이러한 우선어는 박스내의 최상위어 밑에 계층적으로 기재되어 있고, 박스내의 계층은 하위의 용어를 인덴션으로 표시하며, 선굵기로 최상위어·우선어 관계 표시하여 준다.

예)

<圖 23a>

圖式表示: 화살표 그래프



4. Thesaurus의 구성 언어

시소러스는 용어집이기 때문에 그 구성언어에 대해 알아보는 것이 매우 중요하다. 구성언어로는 색인어휘와 도입어휘·식별어가 있다. 색인어휘는 디스크립터 혹은 우선어를 의미하며, 도입어휘는 디스크립터로 채택되지 않은 언어 즉 비디스크립터 혹은 비우선어를 말하며, 마지막으로 식별어는 색인으로서 가치가 있다고 판단되는 인명·지명·기관명·선박명·제품명 등 고유명사가 주로 선택된다.

보다 자세히 시소러스의 색인 용어에 대하여 살펴보면, 단일개념의 용어로 이러한 용어들은 한 개의 단어, 명사와 형용사로 이루어진 두개 또는 세개의 단어로 된 구로 연결된 두 개의 단어, 복합구, 그리고 개인·단체·지명들로 구성된다. 개념들은 일상적인 용어와 호환성을 가지되 될수록 간단하게 묘사한 것으로 한정

어나 주기 등을 사용하여 보다 자세하게 정의하기도 한다.

이러한 시소러스 언어들은 용어통제의 과정을 거치므로써 본격적인 구성언어가 되어진다. 그 방법으로서 첫째, 표현방법이 다양한 자연어를 일정하게 약속된 언어로 변화시킴으로써 일관성을 유지하고, 둘째로 용어의 관계를 나타내줌으로써 전체적인 체계를 조직화하여 준다. 따라서 용어를 통제하면 용어의 부정확성이나 용어의 변화 및 다양성에 대처하게 되고, 본문 중의 용어와 검색자의 용어를 조정하여 검색의 정확률을 향상시킬 수 있다.

용어통제의 주요기능으로는 일관성있는 주제표현을 제공하여 주는 동의어의 통제와 망라적 탐색 수행을 촉진하기 위하여 어의적으로 관련된 용어들을 서로 연결하여 탐색자가 필요한 모든 용어를 스스로 생각해야 하는 것을 도와주는 역할이 있다.

용어를 통제하는데 있어서 먼저 이루어져야 할 작업은 용어간의 관계를 지정하여 주는 것으로서, 주제를 어떠한 방법으로 취급하여도 그 취급방법에 있어서 시소러스 중에서는 일관되게 적용하는 것이 가장 중요하다.

그 관계로는 등가관계, 계층관계, 연관관계가 있다.

먼저, 등가관계라는 것은 색인 작업시 복수의 용어가 동일 개념을 나타내고 있다고 인정되는 경우의 우선어와 비우선어간의 관계로서, 표시방법은 “A는 사용하지 말고 대신에 B를 사용하십시오”(A: 비우선어, 비디스크립터, B: 우선어, 디스크립터)이다. 이는 동의어, 반의어, 동형의어를 갖고 있는 용어를 사용하여 정보를 탐색할 때 도움을 준다. 이때 동의어 선정시 우선하는 용어로서는 도서관을 이용하는 집단에 가장 친밀한 표목과 다른 목록류의 이용을 통해 가장 친밀한 표목, 용어가 내포하는 의미의 수가 적은 표목, 그 주제와 관련된 다른 주제로 가깝게 인도할 수 있는 표목을 우선으로 한다. 보다 자세히 동의어 개념을 살펴보면 다음과 같다.

- 언어학적 어원이 상이한 용어(예: Annuals; Yearbook)
- 속칭과 학술상의 명칭(예: Mercy killing; Euthanasia)
- 일반명칭과 상품명(예: Vacum flasks; Thermos flasks)
- 새로 발생한 개념의 다양한 명칭
(예: Continuing; Life-long education)
- 현재 사용되거나 선호되는 용어와 시대에 뒤떨어지거나 가치가 떨어지는 용어
(예: Developing contry; undeveloped contry)
- 어간의 변형과 불규칙적인 복수형을 포함하는 변형된 철자
(예: Geese; Goose)
- 공통언어를 사용하나, 문화가 상이하야 발생하는 용어
(예: Lifts; Elevators)
- 간략형과 완전형(완전형이 우선)
(예: ISBD; International Standard Bibliographic Description)
- 복합어의 분해형과 비분해형
(예: Milk and hygiene; Milk hygiene)
- 용어의 직접형과 도치형(직접형이 우선)
(예: Electric power plants; Power plants, Electric)
- 명사형과 동사형(예: Applying; Application)
- 단축형(예: Autometric doffers; Auto doffers)

그리고, 유사동의어란 일반적인 사용에서는 상이하다고 간주되는 의미를 갖고 있는 용어나 색인작업의 목적상 동의어로 취급되는 용어로서, 특정 용어의 집합을 유사동의어로 취급할 지 여부는 그 시소러스가 취급하는 주제분야에 좌우된다. 그런 용어로는 동일한 특성의 연속체에 대한 것이나 관점의 상이를 나타내는 용어나 의미가 상당히 중복되어 있는 개념을 나타내는 용어가 있다.

예) Dryness; Wetness
Urban area; Cities

이렇듯 동의어나 유사동의어를 선정함으로써 등가관계를 명확히 하는 이유는 첫째, 정보검색시스템에서 시소러스를 이용하여 정보를 검색하는 이용자의 불필요한 노력을 제거시키는데 도움을 주기 위함이며, 둘째로는 시소러스의 호환성에 관한 연구에도 중요하다.

표시기호로는 USE(보라 참조의 기능으로서 동의어를 통제로서 동일개념의 복수의 동의어로 표시될 수 있는 경우에는 이중의 하나를 색인작업에서 일관되게 사용되는 우선어로서 선택)와 UF(Used For: 색인작업과 검색작업에서 사용할 수 없는 특수한 용어 대신으로 사용할 수 있는 용어를 지시하는 기능)가 있다.

관계의 설정과 그에 대한 표시의 목적으로는 첫째, 우선용어의 범위를 정하는데 도움을 주며 둘째, 이용자가 불필요가 없는 용어를 표시해 주며 셋째, 특정 용어 삭제시 이와 관련된 용어의 참조를 삭제해 주므로써 시소러스의 개정 등과 같은 유지에 필요하다.

두번째로 계층관계에 대하여 살펴보면 “A를 사용하지만 B 또는 C 또는 ... N 등도 참고하라”는 내용으로, 그 표시형태는 BT(광의의 언어, 상위용어), NT(협의의 언어, 하위용어)가 있다.

예) PRIVATE ENTERPRISES
BT Enterprises
ENTERPRISES
NT Private enterprises

세번째로 연관관계는 색인용어간의 보다 더 연결된 관련성을 나타내는 것으로서, 계층관계와 유사한 점이 있으나, 차이점은 연관관계는 계층관계보다 조직적이지 못한 관계로서 상이한 계층의 용어로 의미나 개념이 밀접하게 관련되어 있는 용어·혹은 용도에 따른

관계와 같이 관계가 상호관계에 있는 용어·부분전체의 관계에 있는 개념을 나타내는 용어를 의미한다. 그 표시방법은 RT(Related Term)로 나타낸다.

예) Laser RT Holography
Alcohols RT Solvents

5. Thesaurus의 구축

구축에 있어서 가장 먼저 해야 할 작업은 구축방법을 선택하는 것이다. 그 방법에는 3가지 정도로 나누어 볼 수 있는데, 첫번째 방법으로는 같은 주제분야의 시소러스가 이미 만들어져 있는 경우 최소한의 수정을 가한 후 그대로 사용하는 방법이며, 두번째 방법으로는 기존의 일반적인 시소러스나 관련분야의 시소러스 또는 주제명표·분류표 등의 어휘집을 전체적인 틀로 사용하되 핵심주제의 용어는 별도로 수집하여 상세한 시소러스를 개발하는 방법이며, 마지막 방법으로 새로이 용어를 수집하여 완전히 새로운 체계의 시소러스를 개발하는 방법이 있다.

그럼 그 구축에 있어서 단계적인 과정을 차례로 살펴보자.

첫번째 단계로 시소러스의 목적 결정을 결정하는 것이다. 즉, 주제분야, 문헌의 종류, 문헌의 양, 정보축적 시스템의 종류, 정보시스템의 자원, 이용자와 관련되어 결정을 한다.

두번째 단계로서는 주제영역의 설정이다. 주제분야를 설정하고, 핵심영역과 주변영역을 정한 뒤 사용이 가능한 기존 시소러스가 있는지를 확인하여 본다.

세번째 단계는 시소러스의 특성을 결정하는 것으로서, 시스템이 요구하는 사항의 전체를 조사하여, 어떠한 시소러스를 작성할지 명확히 해야 한다. 즉, 통제어와 자연어 중 사용언어를 결정하며, 용어의 특정성·복합어 및 보조장치 사용·계층관계와 연관관계의 표시정도 등 시소러스의 특성을 결정하며, 알페벳순 외에 도식표시나 체계표시 수록 여부나 시소러스의 형태(자모순·체계적 시소러스)와 같은 시소러스 표시를 결정하는 단계이다.

네번째 단계로는 그 편찬에 있어서 연역적 방법과 귀납적 방법 중 어떠한 방법을 사용할 것인지를 결정한다. 연역적으로 편찬하는 경우는 충분한 수의 용어를 수집한 후에 선택한 용어를 검토하여 구조를 발견

하고 어휘를 통제하며, 귀납적으로 편찬하는 경우는 용어가 문헌상에 출현함과 동시에 시소러스에 추가하여 색인작업에도 사용되는 것으로 어휘통제를 처음부터 실시하여 그 용어를 하나 또는 그이상의 상위 카테고리 부여하는 방법이다.

다섯번째 단계로서는 용어의 수집과 선택 작업을 하는 용어의 수집과정으로서, 용어 수집에 앞서 대상영역을 소주제로 먼저 구분하는 것이 편리하다. 용어의 수집원으로서의 전문용어의 정보원과 이용자와 전문가 혹은 편찬자의 경험과 지식을 이용하는 방법이 있다. 먼저 전문용어의 정보원으로는 표준화된 형태의 용어 정보원과 문헌의 스캐닝을 통한 방법이 있다. 표준화된 형태의 용어정보원으로서의 어형의 통제나 관계의 정확성은 결여되어 있을지 모르나 전문어는 풍부한 분류표와 동의어·상위어·하위어·관련어의 정보원인 시소러스와 용어리스트, 그리고 백과사전, 사전, 용어집과 정의·동의어 제공을 하는 전문용어의 데이터뱅크, 주제분야의 전문용어에 관한 논문, 잡지와 초록지의 색인, 마지막으로 해당분야 출판물의 색인이 있다. 이를 보충하기 위해 사용하는 방법으로는 문헌의 스캐닝이 있는데, 이는 특히 급격하게 발전하고 있는 분야에서는 기존의 정보원에만 의존하여 용어를 선택하는 것은 불충분하기 때문에 표준화되어 있지 않은 자연언어의 용어를 조사하기 위함이다. 그 방법으로서 수작업에 의한 선택과 컴퓨터의 지원을 받아 문헌의 제목·초록·전문으로부터 출현빈도에 의해 이루어지는 용어의 자동선택 방법과 이용자로부터 이미 제시된 질문 및 탐색과 이용자 프로파일의 기록에서 모집하는 질문의 스캐닝 방법이 있다.

다음 단계로서 여섯번째 과정으로는 수집한 용어들을 표준화하는 과정이다. 즉, 수집된 용어는 품사의 선정, 단복수형 선택, 기입형식, 표기법, 약어사항, 구두점 사용, 동음이의어 표기 등에 있어서 일정한 규칙에 따른 통제를 가하는 과정을 의미한다. 용어의 형태로는 명사형의 사용을 원칙하기 때문에 동사는 명사화하여 사용하여야 하며, 단수형과 복수형의 경우 가산명사는 복수형을, 집합명사·공정·특성·조건을 나타내는 용어는 단수형을 선택한다. 동형이의어의 경우는 수식어를 추가해주거나 괄호 속에 한정어를 부여하여 이를 디스크립터에 포함하여 구분을 해주고, 어미를 변화하거나 한정어로도 구분이 불충분한 경우 보다 완

전한 scope note를 추가하여 디스크립터의 사용방법을 지지하여준다. 구두법은 표기의 동일성을 유지하고 기계처리를 용이하게 하기 위하여 쉽표·마침표·어포스트로피 및 하이픈·생략형은 피한다. 용어의 배열에 있어서는 단어간의 모든 공백은 무시하고 왼쪽 괄호-숫자-영자의 순으로 배열하는 알파벳순 배열을 한다.

일곱번째 단계는 용어의 기록 과정이다. 선택한 용어의 기록은 카드나 가능하면 기계가독형식으로 관리하는 것이 좋으며, 용어에 관한 정보로는 정의·동의어·계층관계·연관관계 등을 기록하여 둔다.

여덟번째 과정은 구조의 발견 단계이다. 구조의 발견을 위해서는 먼저 시소러스에 수록할 주제의 예비적 조직화를 하여야한다. 즉, 용어를 수집하기 전이나 수집과정에서 시소러스의 대상이 되는 주제를 주요 카테고리 정리해야 하며, 시소러스의 중심영역과 주변영역을 결정한다. 그후 포괄적 카테고리내의 용어의 분석과 그룹화를 시킨다. 용어를 적절한 포괄적 카테고리에 배치시키고, 알파벳순으로 배열하고, 체계표시를 수록할 경우 이에 대한 기호의 추가나 서체의 선택·상호참조의 점검 등을 하여야 한다.

아홉번째 단계는 체계표시로부터 알파벳순 시소러스의 작성하는 것이다. 알파벳순 시소러스에는 전형적인 알파벳순 시소러스·체계표시를 갖춘 알파벳순 표시·사전 형식의 알파벳순 시소러스가 있다.

열번째 과정은 전문가에 의한 최종점검으로서 주제 전문가와 상담과정을 통해 이루어진다. 그러므로써 보다 적합한 시소러스의 구축이 가능해진다.

열한번째 과정은 시소러스의 서문을 작성하는 것이다. 여기에서 언급하여야 할 사항으로는 시소러스의 목적, 중심분야와 주변분야를 포함하는 주제분야 명시, 색인어의 총수, 우선어와 비우선어의 내역, 색인어의 선택, 형태결정에 대하여 근거가 되는 기준과 채택한 규칙인 어휘통제의 방법, 구조와 상호관계, 시소러스의 레이아웃과 표시, 사용한 배열 규칙, 표준적이 아닌 방법에 따라 작성된 모든 약속사항, 약어, 구두점 기호의 의미 명확화, 시소러스의 사용방법, 갱신과 유지관리를 위한 갱신방침의 세목 및 유지관리에 책임이 있는 기관의 명칭과 주소, 사용한 기록정보원의 사례, 관계자에 대한 사례 등 이상의 사항을 가능한 한 예를 사용하여 기술하여 주는 것이 좋다.

열두번째 단계는 편집과정이다. 편집시 점검사항으로는 상호기입의 점검, 체계표시 중의 기호 점검, 용어의 형태나 첨자 등을 점검한다. 특히 하이픈 사용의 일관성에 주의하여야 하며, 알파벳순 배열과 체계표시 부분의 인덴션이나 공백, 레이아웃, 도식 표시와 체계표시를 살펴보아야 한다.

다음 열세번째 단계는 테스트 과정이다.

마지막으로는 출판을 위한 시소러스 작성을 작성한다.

이러한 열네가지의 과정을 통해 드디어 시소러스의 작성이 이루어진다.

7. Thesaurus 작성시 문제

아무리 완벽한 시소러스를 구축하고자 노력하여도 근본적으로 시소러스는 정보검색에서 발생할 수 있는 모든 언어 문제를 해결하지 못한다. 여기에는 구문적인 문제가 있다. 즉, 어의적 및 구문적 모호성이 원인으로 부적합문헌의 검색이 가능하게 되며, 용어간의 부적합조합으로 인해 문헌이 특정한 두개 이상의 용어로 색인되었거나 또는 이 용어를 포함하고 있기 때문에 검색되기도 하고, 부정확한 용어관계로 인한 문제가 발생할 수도 있다. 그 해결방법으로서는 부적합조합의 경우 색인작성에서 직접 관련되는 용어들을 서로 연결하여 주고 그 연결을 모든 관련된 용어에 보통 문자나 숫자를 할당하여 주고, 부정확한 용어관계의 경우 용어들의 정확한 관계를 나타내는 숫자나 문자코드 등 일종의 역할지시자나 관계지시자에 의해 배제시킨다.

작성된 시소러스를 이용함에 있어서 검색시 실패 유형으로는 첫째, 어휘의 특성성 결여로 인한 것과 둘째, 용어간의 모호하거나 허위적 관계에 기인된 실패가 있으며 셋째, 포괄적 탐색에 필요한 용어나 특정요구에 대한 가장 적합한 용어를 탐색자로 하여금 못보고 지나치게 할 수 있는 적절한 구조의 결여에 기인된 실패가 있다.

8. Thesaurus의 유지와 갱신

일단 시소러스가 작성된 후 '살아 있는' 시소러스를 유지하기 위해서는 모두 정기적으로 갱신해야 한다. 일반적으로 색인작성을 심도있게 행하는 시스템에서 사용하는 특정성이 높은 어휘로 구성된 시소러스는 변

화가 적은 개념이나 광범위한 개념을 이용하도록 되어 있는 시소러스 보다도 빈번하게 갱신할 필요가 있다. 또한 어휘의 성질 및 색인작업의 심도에 관계없이, 색인작업을 행하는 문헌량의 증가에 따라 수정의 필요성도 증대되고 있다.

시소러스의 수정은 새로운 개념의 추가나 용어의 삭제, 기존 용어의 수정이 있다. 즉, 새로운 개념의 추가는 색인작업과 탐색의 과정에서 발생한 개념에 대하여 새로운 우선어로 선정하는 것과 혹은 새로운 용어추가 시 적절한 **scope note** · 동의어 · 계층관계 · 연관관계를 부여하는 경우이다. 그리고 용어의 삭제에 있어서는 오해를 초래하는 용어나 부정확한 용어, 거의 이용되지 않는 용어는 모든 관계정보와 함께 삭제하고, 기존용어의 수정의 경우 철자 수정, 단수형에서 복수형으로의 변경, **scope note** 추가 및 수정, 동의어와 기타 관계정보의 추가 · 삭제를 한다.

시소러스의 유지와 갱신을 위한 일상작업은 조직적인 관리하에서 행해져야 하며, 갱신에 대해서는 가능한 한 신속하게 이용자에게 통지하는 것이 중요하다. 이때 컴퓨터가 지원을 해주는 경우 많은 도움을 준다. 컴퓨터가 유지하고 있는 시소러스 경우는 용어간의 관계의 추가 및 삭제, 정정된 용어의 어형의 확인, 수정과 같은 사무적인 업무는 시소러스중의 어느 곳에 발생하여도 컴퓨터에 의해 자동적으로 수행되며, 용어의 변천에서 획득된 정보를 이용하여 탐색어는 올바른 용어로 자동적으로 부여된다.

9. Thesaurus의 정리와 종합

하나의 시소러스가 완전하기는 어렵기 때문에 이들간의 정리와 종합이 필요하게 된다. 즉, 시소러스의 수가 증가함에 따라 색인언어간의 호환성의 결여라는 문제가 증가하고 있는 현실이기에 이에 대한 보완이 필요하다. 호환성이 없으면 다양한 통제어를 사용하는 시스템 상호간의 **record**의 전송이 불가능할 뿐만 아니라, 다양한 시소러스를 사용하는 **database**간의 일괄탐색도 불가능하게 된다. 또한 정리와 종합을 통해 시소러스 작성용 표준규격이 이용가능하게 된다. 이러한 필요성에도 불구하고 유사한 구조를 갖춘 시소러스의 개발이 촉진되어 왔으나, 시소러스간에는 여전히 호환성이 없다.

동일 주제분야의 시소러스에서 볼 수 있는 상이점

로는 전문용어 포함여부를 나타내는 특정성, 주제분야 영역에 대한 망라성, 복합어 · 동의어의 경우 우선어 선정 기준, 상호관계에서 나타날 수 있다.

이렇듯 호환성이 없는 시소러스를 정리하는 방법으로는 첫째, 변환시스템의 이용하는 것으로서 중간사전의 작성이 요구되며, 이 사전에는 변환하는 색인언어(분류표나 시소러스)가 각각 부여된다. 두번째 방법으로는 병합과 통합으로 동일 주제분야의 시소러스간의 일치나 불일치를 명확히 하여야 한다. 병합의 단계는 일치하는 우선어와 기입어를 알파벳순으로 배열하고 일치한 용어와 이들 용어의 시소러스 관계의 정보를 하나의 기입어하에 병합을 함으로써 이루어진다.

세번째 방법으로는 디스크립터 बैं크의 작성을 통하여 엄밀한 등가관계를 명확히 하므로써 이루어지는 방법이다. 즉, 등가관계를 나타내는 참조에 의해 시소러스 중의 용어간의 유사등가나 부분등가를 나타내고, 시소러스 상호간의 계층관계나 연관관계의 일치나 모순에 관한 정보 제공해 주거나 병합된 용어에 대하여 그 정의와 **scope note**, 등가의 외국어도 기재하여주는 방법이다. 병합된 어휘집의 장점으로는 구성부분인 개개의 어휘집보다 용어나 단어의 관계가 풍부하고, 통제언어 시스템 이외에도 이용 가능하며, 통제언어와 자연언어가 혼재된 시스템에서는 병합한 통제언어를 자연어의 탐색에 이용 가능하다. 반면 한계로는 병합된 복수의 색인언어의 모든 용어를 단순하게 알파벳순으로 배열하게 되면 용어만의 어프로치는 놓치게 되어 개념관계를 표시하는 것이 불가능하며, 구성요소인 어휘집에 상호참조에 의한 지시가 없는 경우에는 용어간의 유사등가성과 부분등가성을 표현하는 것 또한 불가능하다. 또한 네번째 방법으로는 매크로 시소러스와 마이크로 시소러스를 이용하는 방법도 있는데, 마이크로 시소러스란 특수한 시소러스로 상위의 시소러스인 매크로로 시소러스의 계층구조상에 부여하여, 완전하게 그 안에 포함되는 것을 의미하며, 매크로 시소러스는 광범위한 주제의 시소러스를 말한다.

이상과 같은 방법으로 상이한 시소러스를 정리 · 종합하여 보다 유용하게 이들을 사용할 수 있다.

10. 컴퓨터의 지원

컴퓨터의 지원을 받는 경우 인간의 지적작업에 의하지 않고서도 자동적으로 생성되는 시소러스의 구축이

가능하다. 즉, 용어관계를 결정하기 위한 통계적 및 언어학적인 data의 처리에 컴퓨터를 이용하는 것이다. 그 방법으로서 용어의 선택과 예비적인 설명사항을 기록하는 것과 시소러스 구조의 처리와 그 이용과 병합에 사용할 수 있다.

컴퓨터를 통해 지원받을 수 있는 작업을 하나씩 살펴보면 다음과 같다.

첫번째로 용어의 선택과 예비적인 설명 사항의 기록에 있어서 표준화된 정보원과 문헌, 질문을 스캐닝하여 수작업으로 선택한 용어의 상세한 내용을 기록하기 위하여 컴퓨터를 이용하는 것이다. 빈도에 관한 data를 포함하는 경우에는 알파벳순 리스트와 용어의 출현 빈도순과 리스트 출력할 수 있으며, 용어가 크게 분류되어 있는 경우는 개략적인 주제의 그룹화가 가능하다. 또한 복합어의 도치기입 리스트를 작성할 수 있으며, 문헌에서 자동적으로 용어의 선택이 가능하다.

두번째로 시소러스 구조의 처리를 할 수 있다. 그러나 먼저 구조의 발견은 편찬자에 의해 행해지고 나서 일단 구조가 발견되면 그 조작과 갱신작업을 컴퓨터로 행할 수 있다. 즉, 기본적인 시소러스 관계를 나타내는 상호기입의 작성을 통해 상호기입이 자동생성되며, 부가적인 표시의 작성도 가능하며, 컴퓨터의 지원에 의해 시소러스의 전부 또는 일부를 인쇄물인 출판물 형태의 시소러스로 출력 또한 가능하다.

세번째로는 삭제·추가·개정과 같은 시소러스의 유지를 위한 작업에서도 컴퓨터의 지원을 받을 수 있다. 즉, 일상적인 이용으로부터 발생하는 갱신이 가능하다. 그러나 지적인 결정은 편찬자와 색인작성자에 의해 이루어져야 하고 그의 사무적인 업무는 컴퓨터가 수행할 수 있다. 또한 삭제 시 모든 상호기입의 링크가 자동적으로 삭제되기도 하고, 용어의 어형변화나 상위어를 관계어로 변환하는 등 관련된 모든 상호기입의 개정을 확실하게 실행할 수 있다.

네번째로는 색인 작성이나 탐색시 시소러스를 이용할 수 있다. 컴퓨터의 지원에 의한 시소러스는 정보의 축적과 검색시스템의 이용에 있어서 어휘마스터 file로서 중요한 역할을 수행할 수 있다는 장점이 있다. 어휘마스터 file란 색인 작성자와 탐색자가 사용하는 색인어를 체크하기 위해 이용하는 것으로 인정할 수 없는 용어는 배제하며 시스템에 따라서는 비우선어를 입력하면 우선어가 자동적으로 치환되거나 이제는 사

용하지 않는 용어를 입력하면 새로운 용어로 자동적으로 변환되는 기능을 한다.

다섯번째로는 시소러스를 온라인 화면상에서 이용할 수도 있는데, 이는 용어를 입력하여 알파벳순으로 그 용어의 전후에 있는 용어를 스캐닝하는 것과, 그 용어에 관련된 정보의 전부 또는 일부를 확대표시하는 표준적인 방식을 제공받을 수 있다. 또한 계층표시와 체계표시의 적절한 곳을 스캐닝하는 것과 식별기호를 이용하여 표시된 용어중에서 적절한 용어를 선택하는 기능도 있으며, 하나의 용어와 이에 근접하는 관계뿐만 아니라 계층과 체계표시의 카테고리 전체를 선택하여 탐색전략에 이용하는 것도 가능하다.

마지막으로 컴퓨터의 지원을 받아 기계가독형식의 시소러스를 병합하는 것 또한 가능하다.

III. 결 론

색인 작성 뿐만 아니라 정보검색에 있어서 사용되는 용어의 표준화는 매우 중요한 문제이다. 더욱이 우리 도서관과 같이 외래어의 사용이 빈번한 의학관계 분야 용어의 경우 더욱 그러하다. 그러기에 용어의 표준화 도구인 시소러스의 중요성은 아무리 강조해도 지나치지 않다.

앞에서 살펴본 바를 바탕으로 간단히 언급하자면, 시소러스란 정보검색에 있어서 색인작성자와 이용자 사이에 사용하는 색인 및 검색용어를 표준화하고, 통일적인 사용을 위하여 용어 상호간의 관계를 명시한 용어 통제집을 의미한다.

이러한 역할을 하는 시소러스는 그 구축에 있어서 여러 단계의 과정을 거쳐야 완성된다는 사실을 알 수 있었다. 여기에 필요한 지적노력과 비용 또한 그 성격을 결정하는데 있어서 문제가 된다.

그리고 용어를 다룸에 있어 도서관의 특성과 이용자의 요구를 만족시켜줄 수 있어야 하고, 선택된 주제를 취급하는 방법에 있어서 일관되게 적용하는 것이 필요하며, 또한 용어란 시간이 경과함에 따라 새로운 용어의 출현이나 용어의 정정 등이 발생하므로 용어 시소러스 역시 계속적으로 보완·추가·갱신·유지 작업이 중요하다.

구축방법을 고려함에 있어 기존의 NLM에서 만들어진 MeSH authority file을 이용하여 시소러스를

구축하는 방법도 생각할 수 있는 한 방법이 될 수 있으며, 이와 더불어 다각적인 방면에서 지속적인 연구와 관심과 지원이 수반되어야만 한다.

시소러스 구축에 있어서 시작단계로서 이루어진 이 연구를 마치면서, 우리 도서관 실정에 가장 적합하고 효율적인 시소러스의 작성을 기대해 본다. 그리고 시소러스 작성에 있어서 컴퓨터를 통한 접근 가능성에 대한 연구가 보충되어야 한다고 여겨진다. 그리하여 전산화가 이미 이루어진 도서관 환경을 충분히 이용하였으면 한다. 또한 국내 시소러스의 부재 현실에서 다른 분야들에 앞서 시소러스의 구축에 대한 노력이 의학관련 분야에서 먼저 이루어졌으면 하고, 그러기위해 한국의학도서관협회 차원에서 의학관련 시소러스 구축에 있어서의 통일적인 작성규칙을 마련하였으면 하는 바램을 해본다.

참 고 문 헌

- 1) Lancaster F: 정보검색시스템, 1985, p193-208
- 2) 정영미: 정보검색론, 1993, p105-125
- 3) Atichison Jean: 시소러스의 작성법, 1987
- 4) 전학병: "시소러스에서 용어간의 등가관계에 관한 분석", 국회도서관보 29(4): 55-62, 1992
- 5) 이정일: "Thesaurus의 이용과 최근의 동향", 정보관리연구 14(2): 69-75, 1981
- 6) 강일중: "용어간 관계를 이용한 검색문헌의 순위여부에 관한 연구", 정보관리학회지 8(1): 100-115, 1991
- 7) 이재윤: "동적시소러스의 구축에 관한 실험적 연구", 연세대 석사논문, 1994
- 8) 김은식: "과학기술용어 시소러스 대역 데이터베이스 구축", 정보관리연구, 22(2): 1-28, 1991
- 9) 김정현: "디소러스의 개념과 정보검색에 있어서의 역할", 도서관 47(6): 23-33, 1992
- 10) 전학병: "TEST 시소러스의 내용분석", 정보관리연구 23(1), 63-76, 1992
- 11) 김태중: "우리말 시소러스 작성에 관한 연구", 정보관리연구 22(1), 1988
- 12) 김정현: "MeSH의 구성 및 이용법에 관한 고찰", 도서관 45(2) 30-52, 1990