

색인언어의 어의적 관계 및 구문적 관계*

윤 구 호**

<목 차>

- | | |
|------------------|---------------------|
| I. 서 론 | 1. 주제분석 |
| II. 색인언어 | 2. 색인언어로의 변환 |
| 1. 색인언어의 의미와 필요성 | 3. 서지사항의 기술 |
| 2. 색인언어의 구성요소 | IV. 색인표목 및 참조의 자동생산 |
| 3. 색인언어의 어의적 관계 | 1. 색인표목의 자동생산 |
| 4. 색인언어의 구문적 관계 | 2. 참조의 자동생산 |
| III. 색인작성법 | V. 결 론 |

I. 서 론

색인은 정보탐색자(정보이용자)와 정보자료를 연결시켜 주는 중간 매체로서 정보 검색을 위한 핵심적 도구이다. 즉, 색인은 정보자료와 정보이용자 사이에 위치하여 동일한 주제의 정보자료들을 선별하여 주고, 선별된 자료의 소재위치를 지시하여 주는 역할을 수행한다.

오늘날과 같은 정보화사회에서 정보자료의 기하급수적 증가현상과 이용자의 정보 요구의 다양화 및 세분화 현상등은 고도의 정보검색기법을 요구함으로써 색인의 중요성이 새롭게 인식되고 있으며, 특히 주제색인의 효용성이 날로 고조되어 그에 대한 연구개발이 꾸준히 지속되고 있다.

* 본 연구는 재단법인 정부학술장학재단의 1995년도 학술연구비에 의하여 수행되었음

** 계명대학교 문헌정보학과 교수

문헌의 내용을 제시하는 주제색인은 색인대상문헌의 정확한 내용분석(주제분석)을 필수적으로 요구하기 때문에 특정주제에 대한 분석자의 지식 및 주관성이 개입되는 문제점을 내포하고 있다. 주제색인은 또한 주제분석을 통해 추출된 특정개념을 나타내는 색인어의 선정, 선정된 색인어의 다양한 어의적 관계, 단일어로서 표현할 수 없는 복합주제의 경우에는 다단어 색인어(multiword indexing terms)의 논리적 조합 및 배열에 관한 구문적 관계등으로 인하여 효율적인 색인작성이 비교적 어렵다. 그러나 주제색인은 주제를 나타내는 용어로서 검색한다는 점과 또한 색인어의 다양한 어의적 관계를 활용하여 검색의 확장내지는 축소를 조절하여 검색효율을 향상시킬 수 있다는 점에서 제일의 검색매체로서 널리 이용되고 있다.

주제색인의 검색효율을 제고하기 위한 노력은 색인언어의 어의적 관계에 대한 연구뿐만이 아니라 구문적 관계에 대한 연구도 병행되어야 한다. 다단어 색인어의 잘못된 조합으로 인한 부적합 문헌의 검색은 주제색인의 효율성을 저하시키는 요인이 된다. 따라서 과거에는 별로 관심의 대상이 되지 못한 구문적 관계의 연구가 중요시 되고 있다. 비교적 최근에 개발된 다양한 용어열색인은 검색효율의 제고를 위해 특히 언어학적 측면의 연구가 시도된 색인어의 구문관계에 중점을 두고 있다.

그러나, 대부분의 주제색인법이 영어를 위시한 서구어를 중심으로 한 이론과 기법으로 우리나라에서 사용하기에는 문제점이 많다. 한국어는 영어와는 다른 특성을 갖고 있기 때문이다.

문헌자료의 서지적 정리법은 방법론에 있어서는 국제성이 강하면서도 실제면에서는 지역성(locality)이 크게 작용하는 학문의 하나이다. 서지적 정리는 그 대상이 되는 자료의 내용과 형태, 쓰여진 언어 및 문자와 밀접한 관계가 있고, 그 서지목적의 이용자가 쓰는 언어, 문자의 구조, 그들의 사고방식 및 검색습성과 불기분의 관계가 있기 때문이다.

본 연구는 주제색인의 필수요소가 되는 색인언어의 어의적 관계 및 구문적 관계를 상세하게 고찰하여, 한국어의 어법(語法)과 표현방식에 합치되는 효율적인 주

제색인 작성기법을 제시함으로써 도서관을 비롯한 모든 정보서비스기관의 검색효율을 제고하는데 일치적 목적이 있다. 또한 본 연구는 주제색인작성의 필수 도구인 통제어휘집이 매우 부족한 우리나라의 실정에서 조속히 수행되어야 할 주제명표목표나 시소러스의 개발에 일조하고자 함에 부차적 목적이 있다.

II. 색인언어

1. 색인언어의 의미와 필요성

정보검색시스템에서 사용하는 일종의 인공언어를 검색언어라고 하는데, 검색언어는 색인언어와 탐색언어로 구성된다.¹⁾ 이들은 각각 정보자료의 색인과 탐색이라는 특수한 목적을 위해 작성되거나 사용되는 인공적인 언어라고 정의할 수 있다. 여기서 인공이라는 말은 언어의 구성이나 사용에 있어서 어느 정도의 인위적 통제가 취해진다는 의미이다.

Soergel²⁾은 색인언어를 문헌내용의 표현과 정보요구자의 탐색질문에 사용되는 용어라고 정의함으로써 색인언어를 검색언어라는 의미로 사용하고 있다. 일반적으로 정보검색의 효율을 높이기 위해 탐색언어를 색인언어와 일치하도록 하고 있으므로 결국 탐색언어는 색인언어에 의해 결정된다. 따라서 색인언어와 탐색언어는 같은 언어로 인식되고 있으며, 대부분의 경우 검색언어라는 표현 대신 색인언어라는 용어를 사용하고 있다.

색인언어란 색인의 표목을 조절하기 위하여 사용하는 기호군이다.³⁾ 즉, 검색의

1) B. C. Vickery, "Structure and Functions in Retrieval Languages," *Journal of Documentation*, Vol. 27, No. 2 (1971), pp. 69-82.

2) D. Soergel, *Indexing Languages and Thesauri: Construction and Maintenance*, Los Angeles, Melville, 1974, p. 35.

3) 長澤雅男, *情報檢索入門*, 東京, 森出版社, 1974, p. 142.

대상이 되는 문헌의 내용을 표현하기 위하여 선정한 기호군으로서 여기에는 용어뿐만 아니라 숫자나 부호 등 모든 기호가 포함된다. 이러한 색인언어는 문헌의 표제나 본문에서 사용된 자연언어 형태를 그대로 사용할 수도 있으나(자연언어색인), 대부분의 경우 인공적인 통제를 가함으로써 검색의 효율성을 높이고 있다(통제언어색인). 따라서, 문헌의 내용을 나타내는 개념을 일관성있게 표현하기 위하여 인공적으로 규정한 언어를 색인언어라고 할 수 있다.

자연언어색인은 색인의 효율성과 관련된 두 가지 중요한 문제를 안고 있다. 첫째 문제는 동일한 개념이 다양한 용어에 의해 표현될 수 있으므로 관련문헌들이 여러 개의 색인어 아래 분산되는 점이다. 예를 들면 ‘도서관 자동화’와 ‘도서관 기계화’ 또는 ‘도서관 전산화’ 등은 동일한 개념을 나타내는 용어지만 저자에 따라서 사용하는 용어가 다를 수 있으며 또한 한 저자의 동일 문헌내에서도 이들 용어가 혼용될 수 있을 것이다. 따라서 자연언어 색인시스템에서는 동일한 개념을 나타내는 동의어나 유의어를 모두 탐색어로 사용하지 않으면 관련문헌을 전부 검색할 수 없게 된다. 즉, 검색효율의 재현율이 저하된다. 이와 같이 동의어로 인해 발생하는 검색효율의 저하는 통제된 색인언어를 사용하여 용어를 통제함으로써 방지할 수 있다. 즉, 항상 ‘도서관자동화’를 색인어로 채택하고 ‘도서관 기계화’나 ‘도서관 전산화’ 등은 ‘도서관 자동화’로 참조해 주는 것이다.

둘째 문제는 색인하고자 하는 주제가 복합주제일 때 각 개념에 해당되는 여러개의 단어를 결합하여 이 주제를 표현하는 경우에 발생한다. 예를 들어 ‘출산후 산모의 건강관리’, ‘산모의 출산후 건강관리’나 또는 ‘기계가독형 목록’, ‘컴퓨터 목록’, ‘온라인 목록’ 등은 동일한 주제를 다르게 표현한 것으로, 이런 경우 표현방법의 차이때문에 관련문헌이 분산되는 결과를 가져온다. 이러한 문제는 다양하게 표현되는 자연언어의 구문을 통제하여 색인언어가 일정한 구문구조를 갖게 함으로써 해결된다. 즉, ‘출산후 산모의 건강관리’와 ‘온라인 목록’을 색인표목으로 채택하고 다른 표현을 이들 표목으로 참조해 줌으로써 같은 개념의 복합주제는 언제나 일정

한 구문구조를 사용하여 색인하도록 통제하는 것이다.

이와 같이 색인언어란 문헌과 정보탐색자간의 효과적인 커뮤니케이션을 위해 사용하는 일종의 인공적인 언어로서 정보검색의 효율성을 높이는 데 그 목적이 있는 것이다.

2. 색인언어의 구성요소

색인언어는 우리가 일상적으로 사용하는 자연언어와 마찬가지로 어휘(vocabulary)와 구문(syntax)의 두가지 기본요소로 구성된다.⁴⁾ 어휘란 언어에서 사용하는 단어의 집합을 말한다. 색인언어에서는 분류시스템에서 사용하는 분류기호의 집합이나, 주제명색인에서의 주제명의 집합, 또는 키워드색인에서의 통제키워드나 자유키워드의 집합이 어휘에 해당된다.

구문이란 어휘를 구성하는 요소들을 결합하여 의미있는 구문단위(syntactic unit)를 형성하는 일련의 규칙을 말한다. 색인언어에서는 문헌의 주제를 나타내는 색인어가 둘 이상이며 그들이 문맥속에서 구문적 특성을 가질 때 색인어들의 조합에 관련된 특정한 규칙들이라고 할 수 있다.

인공적인 색인언어에서는 어휘를 구성하는 각 요소 즉, 개념을 분명하게 정의하고, 또한 일정한 구문적 규칙을 적용하도록 하고 있다. 이를 위해 포괄적 의미에서의 사전이라고 볼 수 있는 통제어휘집을 사용하는데, 통제어휘집에는 일반적으로 색인어만이 열거되어 있는 것이 아니라 색인어간의 어의적 관계가 제시되어 있어서 색인어의 선택을 통제하고 있다.

자연언어에서는 일반사전이나 동의어사전 등이 일종의 통제어휘집에 해당하나 색인언어에서는 색인언어의 종류에 따라 각각 다른 명칭을 사용하고 있다. 예를 들

4) C. T. Meadow, *The Analysis of Information Systems*, 2nd ed., Los Angeles, Melville, 1973. pp.21-22.

면, 분류색인시스템에서의 분류표, 주제명색인 시스템에서의 주제명표목표(주제명표라고도 함), 키워드색인시스템에서의 시소러스 등이 통제어휘집에 해당된다.

주제명표목표는 사전체목록시스템을 채택하고 있는 도서관에서 주제명목록 작성을 위해 전통적으로 사용하고 있는 통제어휘집이며, 시소러스는 주제명표목표가 주로 단행본 도서의 주제목록을 위해 사용되는 점을 감안하여 논문이나 보고서 등의 주제색인을 위해 주제명표목표를 발전시킨 통제어휘집이라고 볼 수 있다.

시소러스는 색인작업시에 적절한 색인표목의 선택과 색인어의 통제를 위해 필요할 뿐만 아니라 검색시에는 적절한 탐색어의 선택을 위해 필요하다. 시소러스는 용어의 통제목적 이외에도 탐색어의 확장이나 축소를 통해 검색효율을 조절하는데에도 사용된다. 즉, 용어간의 계층적 관계의 용어나 기타의 관련용어를 사용하여 보다 포괄적인 탐색을 함으로써 검색문헌의 수를 늘릴 수 있으며, 또한 보다 특정한 하위 개념의 용어를 사용하여 탐색을 제한함으로써 검색문헌의 수를 줄일 수 있다.

이와 같이 색인언어는 색인어들의 집합인 색인어휘와 채택된 색인어간의 어의적 관계 및 복합적 색인어들의 구문적 관계로 조직된다.

3. 색인언어의 어의적 관계

색인언어의 어의적 관계(semantic relationships ; 일명 priori relationships 또는 syntagmatic relationships이라고도 함⁵⁾)는 동등관계(equivalence relationship), 계층관계(hierarchical relationship) 및 연상관계(associative relationship, 또는 친근관계(affinitive relationship) 라고도 함)로 구분된다.⁶⁾

5) A. C. Foskett, *The Subject Approach to Information*, 4th ed. London, Clive Bingley, 1982. p.91.

6) *Ibid.*, p. 72.

이들 모든 관계의 용어들은 적절한 관계표시기호에 의해 채택된 색인어와 상호 연결되어야 한다.

1) 동등관계

동등관계는 색인어와 동일하거나 유사한 개념의 용어들로서 일반적으로 '보라' 참조(시소러스에서는 'USE' 나 'UF' 사용)로 연결된다.

'보라' 참조는 특정개념을 나타내기 위해 채택된 색인어를 언제나 일관성있게 사용함으로써 동일개념의 분산을 방지하기 위한 기법이다. 따라서 동일개념을 나타내는 용어지만 색인어로 채택되지 않은 탐색 가능한 모든 용어는 채택된 색인어로 안내되는 '보라' 참조를 작성하여 검색효율의 재현율을 향상시켜야 한다.

동등관계에 속하는 범주를 살펴보면 다음과 같다.

(1) 동의어

노고지리 —————> 종달새
 도서관 기계화 —————> 도서관 자동화
 도서관 전산화 —————> 도서관 자동화

'보라' 참조 :

- 노고지리는 종달새를 보시오.
- 도서관 기계화는 도서관 자동화를 보시오.
- 도서관 전산화는 도서관 자동화를 보시오.

(2) 약어

전산기 —————> 전자계산기
 United Nations —————> U. N.
 International Standard Organization —————> ISO

(3) 두분자

DB —————> database

PC ————— personal computer

LA ————— Los Angeles

(4) 상이한 철자어

colour ————— color

labour ————— labor

sulphur ————— sulfur

(5) 상이한 어순

linguistics, descriptive ————— descriptive linguistics

chemistry, analytical ————— analytical chemistry

psychology, educational ————— educational psychology

(6) 번역어

컴퓨터 ————— 전자계산기

(7) 특수한 개념을 디스크립터가 결합된 표목으로 지시할 경우

ferromagnetic film ————— ferromagnetic materials and
film

microwaves amplifiers ————— microwaves and amplifiers

(8) 표목으로 채택되지 않은 특수한 용어로 부터 표목으로 채택된 상위용어로
참조하는 경우

planet waxes ————— waxes

(9) 유의어

근사동의어(near synonyms)나 유사동의어(quasi synonyms)등으로 표현되고
있는 용어들을 총칭하는 유의어란 서로 밀접하게 관련된 용어로서 정보검색의 목적
을 위해서는 동의어로 취급되는 용어라고 정의하고 있으며⁷⁾ 이에 속하는 범주를

7) J. Aitchison and A. Gilchrist, *Thesaurus Construction : a Practical Manual*, London, Aslib, 1972, p. 27.

살펴보면 다음과 같다.

① 속성에 관한 것으로 관점의 차이를 나타내는 용어 즉, 반의어에 가까운 용어

resistance ————— conductivity
 roughness ————— smoothness
 intemperance ————— temperance
 unemployment ————— employment

이와 같은 유의어의 참조는 일반적으로 긍정적 반의어(positive antonyms)를 색인어로 채택하고 부정적 반의어(negative antonyms)로부터 참조를 작성한다.⁸⁾

② 의미가 중복되는 용어

heredity ————— genetics

③ 특정한 개념을 표현한 용어와 이 용어 대신 표목으로 사용할 용어

secondary batteries ————— storage batteries

2) 계층관계

계층관계는 색인어의 상위개념이나 하위개념에 해당하는 용어들로서 일반적으로 '도보라' 참조로 연결된다.

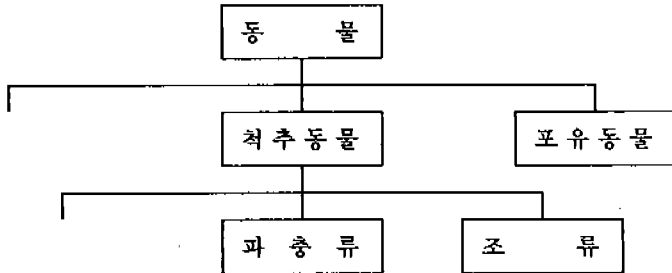
'도보라' 참조는 일반적으로 상위개념어에서 하위개념어로 작성되는데 반드시 두 개념어가 색인어로 채택된 경우에만 작성될 수 있다.

계층관계에 속하는 범주를 살펴보면 다음과 같다.

(1) 진종속관계(true generic relationships)

분류학적 개념의 류(genus)와 종(species)의 계층관계이다.

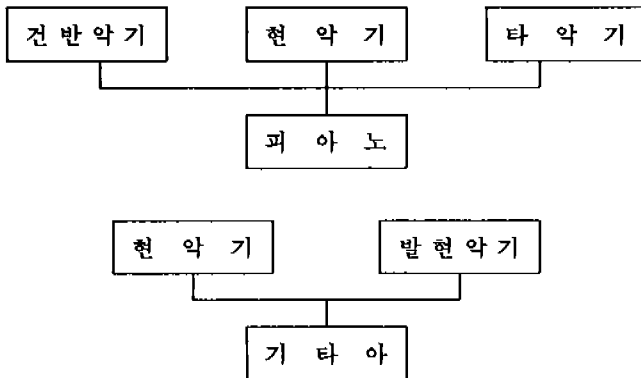
8) M. J. Ramsden, *PRECIS: a workbook for students of librarianships*, London, Clive Bingley, 1981, p. 92.



‘도보라’ 참조 :

- 동물은 척추동물도 보시오
- 척추동물은 파충류도 보시오 등등

어떤 경우에는 하나의 개념이 두개 이상의 상위개념에 속하는 계층관계를 가질 수도 있다.



피아노는 건반악기에도 속하고 현악기에도 속하며 또 타악기에도 속한다. 기타아는 현악기에도 속하며 또 발현악기에도 속한다.

‘도보라’ 참조 :

- 건반악기는 피아노도 보시오.
- 현악기는 피아노도 보시오.
- 타악기는 피아노도 보시오.

- 현악기는 기타아도 보시오.
- 발현악기는 기타아도 보시오.

그러나, 유사종속관계(quasi-generic relationships)는 '도보라' 참조를 작성할 수 없다. 진종속관계와 유사종속관계는 다음과 같이 구분될 수 있다.

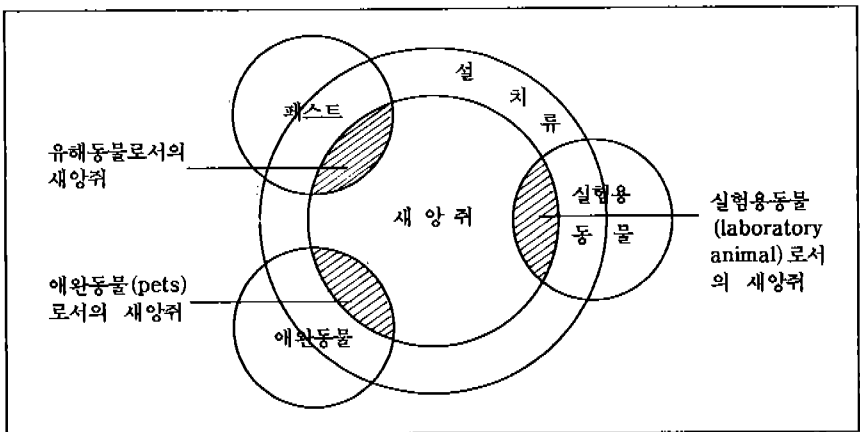
진종속관계 : 불변적 류-종의 관계

모든(all) 새양귀는 설치류의 동물이다. (모든 새양귀는 설치류 동물에 속한다는 새양귀와 설치류와의 불변적 류-종의 관계)

유사종속관계 : 가변적 류-종의 관계

어떤(some) 새양귀는 애완동물이다. (모든 새양귀가 애완동물이란 의미가 아니고 다만 어떤 새양귀는 애완동물이 될 수 있다는 새양귀와 애완동물과의 가변적 류-종의 관계)

이와 같이 새양귀는 모두가 설치류에 속하는 동물이다. 따라서 '설치류는 새양귀도 보시오' 라는 '도보라' 참조는 작성될 수 있다. 그러나 모든 새양귀가 애완동물이 될 수는 없고 다만 어떤 새양귀 즉, 일부의 새양귀는 애완동물이 될 수도 있기 때문에 '애완동물은 새양귀도 보시오'라는 '도보라' 참조는 논리상 작성될 수가 없다.



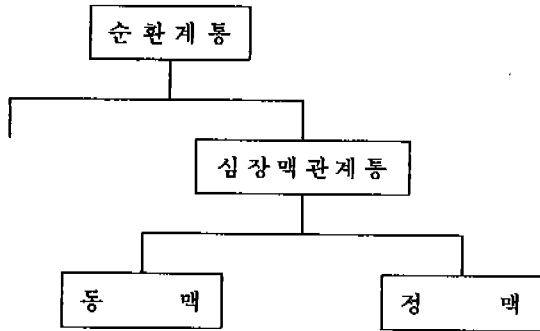
이들 두 관계를 좀 더 자세히 도식화하면 앞의 그림과 같다.

이와 같이 모든 새앙쥐는 설치류에 속하는 동물이고, 어떤 새앙쥐는 애완동물일 수도 있고, 인간에게 해로운 페스트일 수도 있고, 실험용 동물일 수도 있다.

(2) 계층적 전체-부분관계(hierarchical whole-part relationships).

전체와 부분의 계층관계로서 다음의 세 종류를 들 수 있다.

① 신체의 기관(organs)이나 계통(systems) 등의 전체와 부분관계



‘도보라’ 참조 :

- 순환계통은 심장맥관계통도 보시오
- 심장맥관계통은 동맥도 보시오

② 지식의 영역 또는 학문분야 등의 전체와 부분관계

‘도보라’ 참조 :

- 동물학은 조류학도 보시오
- 생물학은 식물학도 보시오

③ 지리적 장소의 전체와 부분의 관계

‘도보라’ 참조 :

- 캐나다는 토론토도 보시오
- 인도는 뉴델리도 보시오

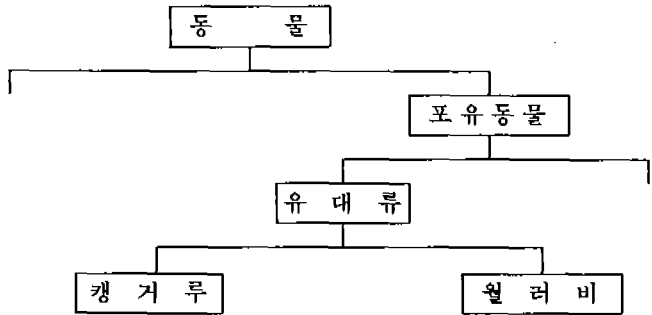
3) 연상관계

연상관계는 동등관계나 계층관계 이외에 어의적 관련성을 갖는 용어들로서 개념간에 은유적으로(mentally) 어떤 관련성을 갖는 용어들의 관계이다.”

연상관계에 속하는 범주를 살펴보면 다음과 같다.

(1) 동위관계(sibling, or collateral relationships)

두개 또는 그 이상의 용어가 하나의 상위개념 아래 동위관계를 가질때 상위개념의 용어와 하위개념의 용어간에는 물론이려니와 형제, 자매간이라고 볼 수 있는 동위개념의 용어간에도 간혹 ‘도보라’ 참조가 필요한 경우가 발생한다. 이는 대부분의 경우에는 배제되는 매우 제한된 경우로서 꼭 필요하다고 판단될 때에만 작성한다. 예컨대, 특정주제의 자료가 최소한 경우 등이 해당될 것이다.



‘도보라’ 참조 : 상위용어에서 하위용어로

- 유대류는 캥거루도 보시오
- 유대류는 월러비도 보시오

동위관계에 있는 용어간에

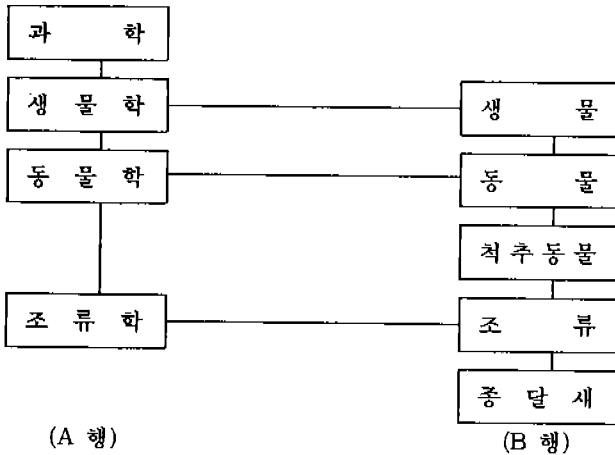
- 캥거루는 월러비도 보시오
- 월러비는 캥거루도 보시오

9) D. Austin, *PRECIS: a manual of concept analysis and subject indexing*, London, The Council of the British National Bibliography, 1974. p. 279.

그러나 '문학'의 하위용어로서 서로가 동위관계라고 볼 수 있는 '소설', '시', '희곡' 등과 같은 용어간에는 이와 같은 '도보라' 참조물 일반적으로 작성하지 않는다.

(2) 교차-범주관계 (cross-category relationships)

서로 다른 범주에 속하는 두 용어간에 하나의 용어가 다른 용어의 개념이나 범위 등을 정의하거나 또는 설명하는 역할을 함축하고 있을 때의 상호의존적 관계라고 할 수 있다.



A행과 B행의 서로 다른 범주에 속하는 '동물학'과 '동물', '조류학'과 '조류'간의 관계는 각각 전자가 후자를 연구하는 행위(활동)를 내포하고 있다. 따라서 다음과 같은 참조가 요구된다.

- 동물학은 동물도 보시오
- 조류학은 조류도 보시오

이와 같은 경우, '도보라' 참조는 일반적으로 범주가 다른 용어의 설명을 도와주는 용어로 일방적 '도보라' 참조가 작성되는 것이 보통이나 경우에 따라서는 다음과 같이 쌍방적 '도보라' 참조가 작성될 수도 있다.

- 동물학은 동물도 보시오
- 동물은 동물학도 보시오

시소러스, 주제명표목표, 분류표에서는 이상과 같은 색인어간의 어의적 관계를 제시해 줄 뿐만 아니라 또한 특정한 색인어의 사용범위를 기술해 줌으로써 색인어의 의미를 보다 명확히 정의해 주고 있으며 동시에 적절한 색인어의 선택이 가능하도록 하고 있다. 또한 어의적 잡음이 발생할 가능성이 있는 동형어의어(homographs)의 경우에는 괄호안에 주기를 삽입하여 색인어의 선택을 돕고 있다.

Bases (Military)	배 (과일)
Bases (Mathematics)	배 (인체)
Bases (Lamps)	배 (선박)

4. 색인언어의 구문적 관계

색인언어의 구문이란 문헌의 주제를 나타내는 주요개념이 둘 이상이며, 그들이 문맥속에서 구문적 특성을 가질 때 개념간의 조합에 관련된 특정한 규칙들로서 색인어의 중요한 요소가 된다.

자연언어에서 특정한 개념을 표현하기 위하여 문구나 문장안에서 사용한 단어들의 상대적 배열위치는 용어간의 구문적 관계를 의미하며, 이러한 구문관계는 색인언어에서도 필요한 것이다. 즉, 용어간의 구문적 관계를 일정한 형식으로 표현함으로써 하나의 복합주제를 항상 일관성있게 색인하도록 인공적으로 통제할 필요성이 있는 것이다.

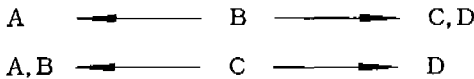
특히 중요한 각각의 개념을 단언어로 변환한 후 검색시에 이들을 조합하는 후조합색인이나 색인시 단어를 조합하는 전조합색인에서는 문맥이 자연언어처럼 표현되도록 하는 것은 어려우며 이로 인해 검색시 부적합 문헌이 검색되는 경우가 발생한다.

1) 색인표목의 형식

문헌의 주제가 단일개념이 아닌 여러개의 복합개념으로 구성되어 있을 때 개념간의 구문적 관계를 나타내는 색인표목(색인기입이라고도 함)의 형식은 단선기입형식(single-line entry format)과 이선기입형식(two-line entry format)으로 크게 구분될 수 있다. 전자로 작성되는 색인표목의 문제점과 후자로 작성되는 색인표목의 합리성을 살펴보면 다음과 같다.

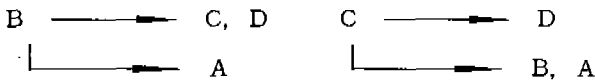
문맥-의존성의 원리(principle of context-dependency)에 따라 조직되는 색인표목은 A > B > C > D와 같이 도식화 될 수 있다.¹⁰⁾¹¹⁾ 즉, A, B, C, D의 4개의 개념이 포함된 주제는 분류이론과 같이 대개념에서 중개념, 소개념으로 점진적으로 전개된다.

색인표목은 이와 같은 복합개념의 연관순서를 합리적으로 유지할 때 그 유용성이 높을 것이다. 그런데 단선기입형식은 각각의 개념을 탐색의 기준어(access point)로 하고자 할 때 합리적인 연관순서를 유지할 수 없는 경우가 생긴다. 예컨대, A-B-C-D에서 B나 C는 상위개념과 하위개념의 두 방향으로 연관이 된다.



따라서 B나 C가 탐색의 기준어가 될 때는 B-C-D-A, C-D-A-B 와 같이 될 수 밖에 없어 두 방향의 연관성이 합리적으로 유지될 수 없게 된다.

그러나, 이선기입형식은 이와 같은 문제를 해결할 수 있다. 즉,



와 같이 기준어의 상위개념과 하위개념의 두방향으로 연관성이 합리적으로 유지될

10) M. J. Ramsden. *op. cit.*, p. 22.

11) D. Austin. *op. cit.*, p. 14.

수 있다.

예를 들어 ‘대학도서관 대출시스템의 전산화’라는 주제는 다음과 같은 3개의 색인표목이 작성될 수 있다.

- ① 대학도서관. 대출시스템. 전산화
- ② 대출시스템. 전산화
대학도서관
- ③ 전산화
대출시스템. 대학도서관

또한 ‘한국 농촌지역 중학생의 학력측정’이라는 주제는 다음과 같은 5개의 색인표목이 작성될 수 있다.

- ① 한국. 농촌지역. 중학생. 학력측정
- ② 농촌지역. 중학생. 학력측정
한국
- ③ 중학교. 학생. 학력측정
농촌지역. 한국
- ④ 학생. 학력측정
중학교. 농촌지역. 한국
- ⑤ 학력측정
중학생. 농촌지역. 한국

따라서, 전통적으로 사용되어 온 단선기입형식의 색인표목은 색인어의 구문관계가 애매하고 비합리적일 수 밖에 없다. 이와 같은 문제점을 해결하기 위해 출현한 대부분의 용어열색인은 이선기입형식의 색인표목을 작성하고 있다.

2) 색인표목의 구문적 장치

주제명색인이나 열거식 분류색인과 같은 전조합색인에서는 일반적으로 문헌의 주

제를 구성하는 중요개념간의 관계가 직접 또는 간접적으로 색인표목속에 이미 정의 되어 나타나 있으며, 키워드를 색인어로 하는 후조합색인에서는 연결기호나 역할기호를 사용하여 개념간의 기능상의 관계를 표시해 주고 있다. 비교적 최근에 개발된 용어열색인에서는 다양한 관계기호를 사용하거나 용어의 조합순서를 문맥에 맞게 정해 주고 있다. KWIC색인과 같은 순열표제색인에서는 키워드와 함께 표제의 나머지 부분이 제시됨으로써 문맥을 구성하는 요소들의 구문적 관계가 포착되며 이를 통해 기준어의 정확한 의미를 파악할 수 있다.

Lynch등이 개발한 분절주제색인¹²⁾이나 Craven이 개발한 NEPHIS¹³⁾ 등은 개념의 연결을 위해 자연언어에서 사용하는 불변화사(전치사, 접속사 등)를 색인표목에 포괄함으로써 색인어의 구문관계를 명확히 해준다.

Farradane의 연관색인¹⁴⁾과 Gardin의 SYNTOL¹⁵⁾ 등은 관계기호를 사용하여 개념의 역할 내지는 개념간의 관계를 표현한다.

개념의 범주에 따른 조합순서에 의해 개념간의 구문적 관계 즉, 색인표목의 문맥을 표현하는 구문적 장치는 일찌기 Kaiser의 체계색인법¹⁶⁾에서 비롯되어 Coates의 영국공학색인(British Technology Index)¹⁷⁾과 Ranganathan의 콜론분류법에 기초한 연쇄색인¹⁸⁾등에서 찾아 볼 수 있다.

개념의 범주에 따른 조합순서와는 달리 자연언어에서의 절대구문에 맞추어 문장

- 12) M. F. Lynch and J. H. Petrie, "A Program Suit for the Production of Articulated Subject Indexes," *Computer Journal*, Vol.16, No.1(1972), pp.46-51.
- 13) T. C. Craven, "NEPHIS: A Nested - Phrase Indexing System," *JASIS*, Vol.28, No.2 (1977), pp.107-114.
- 14) J. Farradane, "Relational Indexing, part I and part II," *Journal of Information Science*, Vol.1 (1980), pp.267-275. pp.313-324.
- 15) J. C. Gardin, *SYNTOL*, Rutgers Series on Systems for the Interlectual Organization of Information, New Brunswick, Rutgers Univ. Press, 1965.
- 16) J. Kaiser, *Systematic Indexing*, London, Pitman, 1911.
- 17) E. J. Coates, *Subject Catalogues: Headings and Structure*, London, Library Association, 1960.
- 18) S. R. Ranganathan, *Colon Classification*, Bombay, Asia Publishing House, 1963.

을 이루는 개념의 순서대로 용어를 결합하여 색인표목의 구문관계를 제시해 주는 용어열색인으로는 <PRECIS>¹⁹⁾와 <한국용어열색인시스템>²⁰⁾이 있다.

Ⅲ. 색인작성법

주제색인의 작성은 색인대상문헌의 내용을 정확히 분석하여 주제를 나타내는 주요개념들을 추출하는 주제분석단계와 추출된 주요개념들을 시스템이 채택한 색인언어로 변환해 주는 색인언어로의 변환단계 및 서지사항의 기술단계로 나누어 볼 수 있다.

1. 주제분석

주제분석은 내용분석이라고도 하는데 문헌에 포함된 주제를 분석하는 과정이다. 주제분석을 위해 문헌의 내용을 검토하는 과정은 일반적으로 다음과 같다.

① 표제를 주의깊게 읽고 이해한다. 중요한 색인어의 대부분이 표제에서 선택 되는 경우가 많다.

② 목차를 본다. 문헌의 주요개념들을 파악할 수 있다.

③ 초록을 읽는다. 저자초록은 주관적인 것이 있으므로 가능한 한 객관성을 유지하여야 한다.

④ 서론을 읽는다. 연구목적, 연구범위 및 연구방법 등을 파악할 수 있다. 그러나 지금까지의 연구동향 등은 색인대상이 되지않는다.

19) D. Austin, *op. cit.*

20) 윤구호, 한국용어열색인시스템, 서울, 신호무역, 1988.

⑤ 결론을 읽는다. 이것은 연구목적 및 연구방법등 서론 부분에서 서술한 것을 실제로 수행하였는지의 여부를 확인할 수 있다. 그러나 고찰은 색인대상이 되지 않는다.

⑥ 각 장 및 절의 표제, 도표, 사진등의 제목과 설명등을 본다. 도표나 사진은 연구내용을 간결하고 정확하게 표현한다.

⑦ 키워드를 본다. 논문의 키워드가 제시되어 있으면 색인어 선정에 도움이 된다.

⑧ 인용문헌을 검토한다. 저자가 참고한 인용문헌의 목록은 문헌의 주제 및 주요개념을 반영한다.

⑨ 본문을 검토한다. 상기과정(①-⑧)을 거쳐도 주제가 명확하지 않을 경우 실시한다.

주제 분석과정에서 반드시 유의해야할 사항은 색인작성의 망라성과 색인언어의 특정성이다. 전자는 색인작성시 문헌당 색인어의 수를 나타내는 것이고, 후자는 주제를 명확하게 기술하는 색인언어의 능력을 말하는 것이다.

망라성은 검색효율의 재현율을, 특정성은 정확율을 지배하는 요인이 된다. 즉, 망라성이 높은 주제분석에서는 핵심적인 개념뿐만 아니라 개괄적인 개념이나 또는 별로 중요하지 않은 개념까지도 색인어로 채택될 수 있기 때문이다.

색인언어의 특정성은 색인언어 자체의 본질적 질을 나타내는 기능으로서 분류색인 보다는 주제명색인이, 주제명색인보다는 키워드색인이 높다고 본다. 왜냐하면 일반적으로 키워드색인에서는 하나의 개념이 하나의 디스크립터(색인어)로 변환되고, 주제명색인에서는 몇개의 소개념이 합쳐져서 하나의 주제명을 형성하며, 분류색인에서는 문헌에 포함된 모든 개념을 포괄하는 하나의 분류기호가 색인어로 선택되기 때문이다.

2. 색인언어로의 변환

주제분석과정에서 추출된 주요개념들을 시스템의 색인언어로 변환할 때 대부분의 자연언어색인에서는 본문에 나타난 자연어를 그대로 색인어로 채택하지만, 통제언어색인에서는 검색효율을 높이기 위해 색인어를 통제해 준다.

통제언어색인에서는 주제명표목표나 시소러스와 같은 색인어회집과 색인어 선택 규칙을 마련하여 일관성있는 색인어 선택을 가능케 하고 있으며, 자연언어색인에서는 검색시의 검색효율을 높이기 위해 탐색용 시소러스를 사용하거나 또는 용어절단 기법(term truncation)이나 용어집 단화기법(term clustering)을 사용하고 있다.

색인언어로의 변환과정에서 유의해야할 사항은 앞에서 언급한 색인어간의 어의적 관계와 구문적 관계이다.

3. 서지사항의 기술

각각의 색인대상문헌에 적합한 색인어가 할당되면 수록된 문헌에 대한 서지사항의 기술은 가능한 한 국제표준기구(ISO) 등이 제정한 표준화된 규칙에 따라야 한다.

기입되는 서지데이터는 자료유형에 따라 다소 차이가 있다. 예를 들어 잡지논문인 경우에는 문헌번호, 저자, 저자의 소속기관, 논제, 수록잡지명, 권수, 호수, 면수 및 출판년 등을 포함하는 경우가 일반적이다.

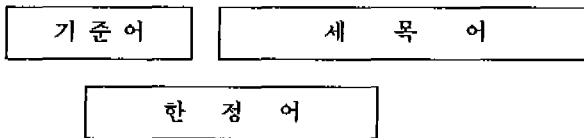
서지사항은 색인표목과 함께 제시되거나, 또는 색인표목은 문헌의 참조기호(연결키)만을 제공하고 이 기호에 의해 이차적으로 탐색되는 별도의 서지파일에 제시된다. Index Medicus나 British Technology Index 등이 전자에 속하고, 유니텀시스템과 같은 후조합색인, KWIC 등의 순열표제색인, SCI 등의 인용색인 등이 후자에 해당된다.

IV. 색인표목 및 참조의 자동생산

1. 색인표목의 자동생산

색인언어의 구문적 관계는 합리적인 색인표목의 작성을 위한 관건이다. 합리적인 색인표목은 개념간의 의미적 연관성을 명료하게 제시함으로써 검색효율을 향상시킬 수 있다.

전통적으로 사용되어 온 단선기입형식의 색인표목은 앞장에서 살펴본 바와 같이 개념간의 연관성을 명확하게 제시할 수 없는 경우가 발생하여 부적합 문헌의 검색을 배제할 수 없는 결과를 초래한다. 따라서 다음과 같은 이선기입형식의 색인표목을 작성하여 모든 개념간의 명확한 연관성을 제시함으로써 검색효율을 제고할 수 있다.



기준어 : 기준검색어의 준말로서 색인표목에서 탐색의 기준어가 위치하는 곳

세목어 : 세목전개어의 준말로서 색인표목의 모든 구성요소가 최초로 나열되는 곳

한정어 : 색인표목에서 기준어의 범위를 한정해 주는 용어가 위치하는 곳

이와 같은 이선기입형식을 사용하여 한국어의 어법(語法)과 표현방식에 합치하는 색인표목을 수작업으로나 또는 컴퓨터를 활용하여 자동으로 생산하는 기법은 <한국용어열색인시스템>에 상세하게 제시되었다.

2. 참조의 자동생산

앞장에서 살펴본 바와 같이 색인언어의 어의적 관계는 검색의 확장 내지는 축소

를 조절할 수 있는 요소로서 검색효율을 지배하는 요인이 된다.

색인언어의 어의적 관계는 동등관계, 계층관계 및 연상관계로 구성된다. 다양한 참고자료를 이용하여 색인어의 모든 어의적 관계의 용어들을 제시하고 필요한 참조를 제시하여 검색효율을 향상시키는 것은 매우 중요한 업무임에도 불구하고 정보서비스기관에서 흔히 등한시되는 경우가 많다.

일반적으로 동등관계는 '보라' 참조로, 그리고 계층관계 및 연상관계는 '도보라' 참조로 제시된다. 그러나 그 범주와 용법이 다양하여 세심한 주의가 요구된다. 대부분의 경우 동의어에 대한 '보라' 참조는 비교적 용이하게 제시되지만, 기타의 어의적 관계는 간과되는 경향이 있다. 특히 유의어에 해당하는 동등관계와 연상관계의 용어들에 대한 참조는 거의 작성되지 않고 있어 검색효율의 저하를 초래하고 있다.

검색효율의 제고를 위해 필요한 '보라' 참조와 '도보라' 참조는 반드시 제시되어야 한다. 색인언어의 모든 어의적 관계를 위한 다양한 참조의 생산기법, 특히 컴퓨터를 활용한 자동생산기법은 <한국용어열색인시스템>에 상세하게 제시되었다.

V. 결 론

색인은 정보검색을 위한 핵심적 도구이다. 정보자료의 기하급수적 증가현상과 정보요구의 다양화 및 세분화현상 등은 효율성이 높은 색인, 특히 검색효율을 향상시킬 수 있는 주제색인의 연구개발을 촉진하고 있다.

본 연구는 주제색인의 필수요소가 되는 색인언어의 어의적 관계 및 구문적 관계를 상세하게 고찰하여 한국인의 사고방식과 검색습성, 한국어의 어법(語法)과 표현방식에 합치되는 주제색인 작성기법을 제시하였다. 따라서 본 연구는 도서관을 비롯한 모든 정보서비스기관의 효율적인 주제색인 작성에 활용되어 검색효율을 제

고할 수 있을 것이다. 또한 본 연구는 주제색인작성의 필수도구인 통제어휘집이 매우 부족한 우리나라의 실정에서 조속히 수행되어야 할 주제명표목표나 시소러스의 개발에 일조할 수 있을 것이다.

본 연구의 요점을 간추려 보면 다음과 같다.

첫째, 주제색인작성의 필수도구인 통제어휘집, 특히 주제명표목표와 시소러스의 특성과 용법을 고찰하였다.

둘째, 색인언어의 어의적 관계를 상세하게 고찰하였다. 동등관계, 계층관계 및 연상관계의 정확한 개념정의와 각각의 범주등을 명확하게 제시하고, 또한 정보검색 효율의 제고를 위한 '보라' 참조와 '도보라' 참조의 사용법을 상세하게 제시하였다.

셋째, 색인언어의 구문적 관계를 고찰하였다. 특히, 다단어 색인어(multiword indexing terms)의 구문관계를 위한 색인표목의 기입형식을 살펴보았다.

문맥-의존성의 원리에 따라 조직되는 색인표목의 연관성은 단선기입형식의 색인표목보다는 이선기입형식의 색인표목이 훨씬 합리적이어서 부적합 문헌의 검색이 배제될 수 있다.

넷째, 색인어의 어의적 관계를 위한 '보라' 참조와 '도보라' 참조의 컴퓨터에 의한 자동생산기법을 제시하였다.

다섯째, 색인어의 구문적 관계를 명료하게 제시하는 색인표목의 컴퓨터에 의한 자동생산기법을 제시하였다.

〈참 고 문 헌〉

* 각주로 대신함.

Semantic and Syntactic Relationships of Indexing Languages

Koo-ho Yoon*

〈Abstract〉

Indexes, especially subject indexes, are major tools for information retrieval. To enhance the retrieval effectiveness of subject indexes, the semantic and syntactic relationships of indexing languages are very important elements.

This paper examines the afore-mentioned relationships, based on purely the syntax and semantics of Korean language.

The outlines of this study are as follows :

1. The characteristics and usages of controlled vocabularies, particularly subject headings lists and thesauri, are reviewed.
2. The semantic relationships, such as equivalence, hierachical and associative relationships, are defined, and their categories are investigated in detail. Accordingly, the usages of 'See' and 'See also' references are suggested circumstancially.
3. The syntactic relationships are also examined. Particularly, for the syntactic relationships of multiword indexing terms, two kinds of subject entry formats are compared. Since it is more rational for sub-ject

* Professor, Department of Library and Infotmation Science, Keimyung University.

headings organized by the principle of context-dependency, the two-line entry format is recommended for subject indexes.

4. Computerized production techniques of 'See' and 'See also' reference for the semantic relationships of indexing terms are presented.

5. Computerized production techniques of subject indexes representing the syntactic relationships of indexing terms are also presented.