

패싯분류체계를 이용한 시소러스작성에 관한 연구

A Study on Thesaurus Construction Using Facet Classification

손대형, 첨단학술정보센터 연구개발부
김태수, 연세대학교 문헌정보학과

Son Dae-Hyung, Research & Development Dept. KRIC
Kim Tae-Soo, Dept. of Lib. & Inf. Sci., Yonsei Univ.

본 연구에서는 기존의 분석합성형 분류체계(BC2)로부터 개념과 관계를 추출하여 작성한 자모순 시소러스에 패싯도구를 삽입하여 용어의 개념과 관계를 정확하게 이해할 수 있도록 제시하는데 그 목적이 있다. 그리고, 이러한 과정을 통해 기존의 분류체계나 시소러스가 지니고 있는 한계를 개선하고자 한다.

1. 서론

전통적인 도서관의 분류에서는 복합주제 중 분류과정에서 선택되지 않은 주제로는 검색이 불가능하다. 전통적인 분류법에서는 속종관계로 모든 개념을 유취하려고 하였다. 그러나 이러한 계층분류는 하나의 계층관계만을 표현하므로 지식에 포함된 개념이나 단어 관계를 적절히 표현하지 못한다.

지식조직의 다른 주된 유형인 패싯분류의 기본적인 특성은 주제의 구성요소를 패싯으로 분석하고, 일정한 원칙에 따라 합성하는 것이다.

기존의 평면적 구조의 시소러스는 여러 계층의 관련된 용어들을 자모순으로 배열함으로써 그 수준을 제시하지 못하였다. 시소러스를 작성하는데 있어 분석합성분류처럼 개념을 특성에 따라 구분하고 개념간의 관계구조를 제시함으로써 효과적으로 주제접근을 지원할 수

있을 것이다.

본 연구에서는 기존의 분석합성형 분류체계로(Bliss Bibliographic Classification Second Edition)부터 개념과 관계를 추출하여 작성한 자모순 시소러스에 패싯도구를 삽입하여 용어의 개념과 관계를 정확하게 이해할 수 있도록 제시하는데 그 목적이 있다.

2. 이론적 배경

일반적으로 도서관의 분류체계는 크게 열거형분류(비패싯분류)와 분석합성형분류(패싯분류)로 구분할 수 있지만 모든 분류체계는 최소한 보조표를 통해 어느 정도 주제간의 합성기법을 도입하고 있다.

분류는 개념간에 존재하는 복합적인 상관관계를 표현할 수 있는 상세한 분석방법이 요청되고, 한번 분류된 자료는 새로운 자료의 입수시 융통성 있게 조합될 수 있는 안정성이 필요하다. 그러나 기존의 분류법은 분류원칙이 제

한되어 있고 융통성이 적어 주제의 정확한 표현을 기하지 못하고 있다.

패싯분류의 기초는 개념을 패싯에 따라 분석하고 합성하는 것이다. 즉, 패싯은 공통의 특성을 가진 개념 범주이다. 각 패싯은 패싯의 순서를 보여주기 위한 표시기호(notation)와 패싯지시기호(facet indicator)를 갖는다.

일반적인 패싯분류체계의 특징은 다음과 같이 설명된다.

1. 분석과 합성 개념이 분류의 기초이다.
2. 주제를 패싯으로 분석하기 위해 포괄적인 수준에서 정해진 범주를 이용한다.
3. 인용순서(citation order), 즉 패싯의 표현 방법에 관한 규칙을 제공한다.
4. 인용순서를 통해 문헌의 집단화가 이루어진다.
5. 열거형분류와 같이 패싯분류도 각 패싯과 그 개념을 식별하기 위해 기호체계를 이용한다.
6. 패싯체계는 매우 유연하다. UDC같은 분류표는 어느 패싯도 결합될 수 있으며, 분류자에게 열거형분류보다 더 많은 자유를 준다.

열거형분류와 분석합성형분류는 다음과 같이 비교될 수 있다.

첫째, 패싯분류는 단일개념만을 나열하기 때문에 열거형분류보다 간략하다.

둘째, 패싯분류는 새로운 주제를 기존의 체계에 삽입하는 것이 아니라 패싯내의 개념을 조합하기 때문에 새로운 주제를 쉽게 수용할 수 있다.

셋째, 패싯분류에서 표시기호는 합성기능 때문에 상대적으로 복잡하다.

3 분류와 시소러스의 통합

분류체계는 특정 개념에 대해 하나의 계층관계만을 제시하여 다중계층관계를 효과적으로 표현하지 못한다. 패싯화된 시소러스는 시소러스와 결합된 분류체계로서 서로를 보완하여 이

와 같은 분류체계의 문제점을 극복하기 위한 것이다. 분류표는 속종관계와 같은 계층관계를 표현하고, 시소러스는 관련어와 상·하위어를 제시하여 자료를 후조합으로 검색하는데 사용될 수 있다.

이러한 체계표를 구축하기 위해서는 하나의 관계만을 식별하는 것만으로는 충분하지 못하기 때문에 정확한 개념 분석이 필요하다. 그러므로, 분류와 함께 시소러스를 제공하는 것은 분류표로 제시할 수 있는 것보다 훨씬 많은 용어를 상세한 수준으로 제공할 수 있다.

시소러스에서는 단어와 구의 다양한 형태가 제어되지만 분류에서는 이들이 거의 제어되지 않는다. 그리고, 시소러스에서는 대등관계를 통해 동의어가 제어되어 접근이 가능하다. 분류에서는 특정 개념을 표현하는 여러 용어 중에 대표적인 하나의 용어를 사용하여 표현하고, 계층관계를 제공하여 용어의 의미를 명확히 한다.

최선의 통제언어는 탐색시에 시소러스처럼 사용될 수 있는 분류표일 것이다. 이용자는 탐색어로서 우선어와 비우선어를 모두 사용할 수 있다. 분류표의 계층관계는 탐색어를 확장하고 축소하는데 훌륭한 도구로 사용되며, 시소러스가 지닌 주제영역간의 간격의 공백(gap)을 해소할 수 있다.

이와 같이, 시소러스는 의견상 명확히 드러나지는 않더라도 분류구조를 가져야 하는데, 주제를 조직하기 위한 방법 중 패싯분석을 하는 것이 가장 바람직하다. 패싯분석을 통해 용어의 의미와 용어들의 연관성을 더 명확히 이해할 수 있다. 그러므로, 패싯분석으로 체계화된 분류표를 구축하고 이로부터 전통적인 자모순 시소러스를 구축할 수 있다.

4. BC2로부터 시소러스 작성

시소러스에서 개념관계를 표현하는데 있어서 패싯분석을 적용하는 것은 매우 유용하다. 색인어간의 관계와 색인어와 주제분야와의 관

계를 제시하는 패싯분석을 통해 색인자와 탐색자는 주제분야의 개념구조를 이해할 수 있다. 그런데, 지금까지 패싯화된 시소러스란 디스크 컴퓨터의 자모순배열부분과 개념간의 관계구조로 제시된 두 부분으로 구성되어 있다.

대부분의 구조체계부분은 전통적인 주제영역의 계층관계로 개념을 제시하고 있다. 이는 넓은 주제영역에서 좁은 영역으로 개념에 접근하는 방식에 유용하다고 할 수 있다. 그러나, 자모순 시소러스로도 상하위어의 관계를 특성에 따라 제시함으로써 이와 같은 기능을 할 수 있으며, 한 항목에 관련된 용어의 수가 많을 경우에는 한 계층만을 포함하는 자모순 시소러스가 더 유용할 수도 있다.

4.1 시소러스 작성을 위한 정보원으로서의 BC2

BC의 가장 두드러진 특징은 패싯분석과 소급 기호체계(retroactive notation)이다.

BC2 분류기호의 합성은 직접적인 소급기호체계를 따른다. 특정 주제에 대하여 분류표상에서 앞서 출현한 모든 주제들은 머리글자를 생략한 분류기호를 해당 주제의 분류기호에 직접 추가함으로써 그것을 한정하는데 이용될 수 있다.

BC2는 다음과 같은 이유에서 시소러스 작성에 도움을 준다.

첫째, 완전히 패싯조직되어 있기 때문에 시소러스에 패싯개념을 적용시키기 편리하다.

둘째, 용어의 선택과정과 관계설정과정을 단축시킨다.

셋째, 각 영역별로 발행되고 있어 각각을 변환시킬 경우 전문도서관에서 특정영역에 대한 시소러스로 이용될 수 있다.

넷째, 시소러스의 조합수준은 BC2를 그대로 따르게 되며, 필요한 경우 조절할 수 있다.

다섯째, 주변영역을 추가할 필요가 있을 경우 해당영역을 다루는 권의 필요한 수준 만큼을 시소러스에 추가시킬 수 있다.

여섯째, 시소러스를 이용하는 과정에서 갱신의 필요가 생기는 경우에도 BC2는 다른 분류체계보다 쉽게 변화를 수용할 수 있다.

본 연구에서 작성하는 시소러스는 자모순 시소러스로서 각 표목마다 패싯을 포함하여 계층을 한 단계만 표시함으로써 해서 전체적인 구조 이해를 더 용이하게 만든다.

4.2 용어의 추출

본고에서는 기본적으로 BC2의 용어를 대상으로 하며, 용어의 효과적인 군집화를 위해 사용된 패싯지시기호는 시소러스 용어로 사용하지 않는다.

BC2에서 쉽표로 구분되는 일련의 용어들은 동의어로 간주되는데, 가장 앞에 기재된 용어를 대부분의 경우 시소러스에서 우선어로 사용한다. 그러나 해당 용어의 형태나 의미가 분류의 목적에 맞도록 선택되어 시소러스의 용어로 부적절하다고 판단되는 경우에는 시소러스 목적에 맞는 용어를 우선어로 선택한다.

4.3 어형통제

우선어는 자연어순으로 기입될 때와 비우선어 표목 아래에서 우선형태로 참조될 때만 모두 대문자로 표기하고, 그 외의 경우와 비우선어는 첫 글자만 대문자로 표기한다.

BC2의 용어는 단수형과 복수형을 혼합하여 사용하고 있으나 시소러스 목적상 복수형이 가능한 것은 단수형의 용어를 복수형으로 변경한다. 그러나 단수형과 복수형의 의미가 상이한 경우에는 모두 시소러스에 포함시킨다. 필요하다면 한정어나 한정어구를 추가한다. 그리고, 쉽표(.)는 복합어의 도치형을 표현할 경우에만 사용된다.

BC2의 어형을 바꾸거나 우선어를 변경하는 과정에서 분류표상의 의미를 상실하지 않고, 일관성을 유지하고, 이용자가 시소러스를 이용하는데 편리하고, 시소러스 목적에 맞는 용어로 변환해야 하는 것이 중요하다.

본 연구에서는 용어의 자연어순형태를 우선어로 선택하고 있으며, 대부분의 도치형을 포함한다.

4.4 관계의 추출

본 연구에서는 시소러스 용어의 계층표시에 BC2에서 사용되는 패시를 삽입하여, 계층관계에 있는 용어들을 자모순으로 배열하는 것보다 효과적으로 용어간의 관계를 이용자에게 제시하도록 하였다.

표목 아래의 관련 용어들은 반드시 자모순으로 배열되지는 않는다. 하위 배열되는 용어들의 의미가 시간적 순서나 과정과 같은 경우에는 이용자에게 의미적으로 유용하다고 생각되는 순서로 배열된다.

BC2에서 쉽표로 구분되어 나열되는 용어들을 대등관계의 용어들로 인정한다. 일반적으로 가장 앞에 출현하는 용어를 우선어로 그 뒤의 용어들을 비우선어로 간주한다.

계층관계와 연관관계는 패시지시기호가 관계를 표현하는 경우와 패시가 나타나지 않지만 시소러스에서 계층이나 연관관계를 주는 경우, '*'로 시작하는 설명 중 관계를 주어야 할 용어가 나타나는 경우가 있다. 마지막의 경우는 대부분 다음과 같이 다중계층관계가 된다.

일반적으로 동일 클래스에 속하는 용어 즉, 형제어(sibling)를 연관시킬 필요는 없다. 형제어는 체계표시에서 발견할 수 있으며, 자모순 표시에서는 적절한 상위어를 조사하면 발견할 수 있기 때문이다.

4.5 범위주기

BC2는 '*' 표시를 이용하여 많은 설명을 주고 있다. 이들 설명 가운데 'see'나 'see also'를 포함하는 설명 중의 용어는 앞에서 보았듯이 계층 혹은 연관관계로 만들고, 정의나 제한적 용도에 관한 설명은 SN를 주고 시소러스로 옮긴다.

5. 요약 및 결론

이와 같이, BC2로부터 작성된 시소러스를 BC2와 함께 사용함으로써 색인과 검색의 특정성을 높일 수 있으며, 용어 및 관계의 일관성도 유지할 수 있다. 그리고, 두 가지의 도구가 유기적으로 결합됨으로써 분류표를 통한 하향식 접근과 시소러스를 통한 상향식 접근으로 쌍방향의 보완과 갱신이 가능하여 보다 완벽한 체계를 이룰 수 있다.

참고문헌

- 김태수. 1984. "연관색인의 이론과 실제." *정보관리학회지*, 1(1): 25-42.
- Aitchison, Jean and Alan Gilchrist. 1987. *Thesaurus Construction: A Practical Manual*. 2nd ed. London: Aslib.
- Aitchison, Jean. 1986. "A Classification as a Source for a Thesaurus: The Bibliographic Classification of H. E. Bliss as a Source of Thesaurus Terms and Structure." *Journal of Documentation*, 42(3): 160-181.
- Iyer, Hemalata. 1995. *Textbook for Knowledge Organization Vol.2: Classificatory Structures: Concepts, Relations and Representation*. Frankfurt and Maine: Indeks Verlag.
- Maltby, Arthur and Lindy Gill. 1979. *The Case for Bliss*. London: Clive Bingley.
- Mills, J. and Vanda Broughton. 1996. *Bliss Bibliographic Classification Second Edition, Class R Politics and Public administration*. London: Bowker-Saur.