

고전 용어 시소러스의 분류 체계에 관한 연구*

A Study on Classification System of Korean Literatures Thesaurus

유 영 준(Yeong-Jun Yoo)**

목 차

- | | |
|---|----------------------|
| 1. 연구의 목적 | 4. 고전 용어 시소러스의 분류 체계 |
| 2. 연구의 방법 | 4.1 분류 기준 |
| 3. 분류 체계의 이론적 배경 및 시소러스에
적용된 분류 체계간 비교 | 4.2 분류 체계 |
| 3.1 열거형 분류 체계와 패시 분류 체계 | 4.3 분류 체계의 구조적 특성 |
| 3.2 시소러스에 적용된 분류 체계 비교 | 4.4 분류 체계의 적용 |
| | 5. 결 론 |

초 록

우리 고전 문집에 나타난 용어로 작성된 디스크립터들을 분류하기 위해서 분류 체계를 개발하는 것이 이 연구의 목적이다. 고전 용어 시소러스의 분류 구조는 6개의 패시로 범주화를 시작하여 고전 분야의 지식을 근거로 연역적으로 분류 체계를 구조화하였다. 그리고 기존의 인문학 분야의 다른 시소러스들의 분류 체계와 비교하였고, 이러한 비교를 통해서 얻은 것은 고전 용어 분류 체계만이 갖는 여러 장점들 즉 패시 기법을 적용한 장점 등을 확인할 수 있었다. 이러한 장점들로 인해서 범주 설정의 일관성과 분류 구조의 복잡성을 줄일 수 있었다. 또한 시대나 지역을 구분하기 위한 공통 구분을 독립적으로 설정하여 분류표의 크기를 줄일 수 있었다. 이 분류 체계는 디스크립터들을 배정하는 과정을 통해 보다 나은 분류 체계로 발전해 나갈 것이다.

ABSTRACT

This study aim to develop a classification system to classify the descriptors, which is been in korean literatures. Firstly this classification structure is categorized on six facets and the classification system is constructed on a deductive method based on korean literature knowledge. The study compared the classification system with various thesaurus's classification system in humane studies and by the comparison, the classification system of korean literature's terms find out having some merits as using the facet method. On account of these merits the classification system has achieved a consistency of categorization independently and reduced a complexity of classification structure. And by categorizing the common categories, the study has reduced the size of schedules. Finally, the classification system has advanced the structure in the process of classifying the descriptors.

키워드: 분류 체계, 고전 용어, 시소러스, 패시 분석

Classification System, Korean Literature's Term, Thesaurus, Facet Analysis

* 이 연구는 2006년도 민족문화추진회의 『고전 용어 시소러스 개발 사업』의 일부분으로 수행되었음.

** 연세대학교 문헌정보학과 강사(yj6370@hanmail.net)

논문접수일자 2006년 5월 15일

게재확정일자 2006년 6월 10일

1. 연구의 목적

시소러스는 특정 주제 분야의 지식을 색인하거나 조직할 때 적절한 색인어나 용어의 선택을 위해서 사용될 뿐만 아니라 검색 시 효과적인 검색어의 선택을 위해서도 사용된다. 즉 시소러스에서는 등가관계, 계층관계, 연관관계에 의해 개념간의 관계를 확인할 수 있고, 이를 바탕으로 이용자가 자신에게 적합한 검색어를 선택하여 문헌을 검색하도록 지원한다. 그리고 시소러스에서 검색을 지원하는 다른 하나의 방법은 분류 체계를 사용하는 것으로, 분류 체계에 의해 디스크립터들의 계층 관계를 확인하여 검색어의 의미 범주를 확장하거나 한정하는 방식으로 검색을 지원할 수 있다.

그러나 시소러스에서 분류 체계가 검색을 지원하는 목적을 제대로 수행할 수 있도록 범주화하는 것이 중요한데, 범주화는 의미상 유사한 용어들은 동일 범주에 모으고, 범주 내에서 용어간의 계층 관계를 제시하기 위한 수단을 말한다. 용어들이 어떠한 개념적 속성을 가지고 있는지를 확인하기 위해서는 시소러스를 구축할 때 용어들의 개념적 속성에 의해 범주화하는 것이 필수적이다.

이 연구에서는 용어들의 개념적 속성을 구분하기 위해서 패시 기법을 적용하여 우리나라 고전 문집에 나타난 용어들을 범주화 하려고 한다. 패시(facet) 이란 한 주제 범주 내에서 하나의 특성을 분할 원칙으로 하여 생성되는 하위 범주들과 그 구성요소의 집합을 말한다. 즉, 사람은 인종, 나이 교육 등의 특성에 따라 분할될 수 있는데, 나이라는 특성을 기준으로 하여 분할하면 어린이, 성인 노인 등과 같은 하위 범

주들과 그 범주를 구성하는 실제 구성원들이 생성되며 이것을 패시이라고 일컫는다. 그리고 이러한 특성을 분석하여 하위 범주들로 범주화하는 과정을 패시 분석이라고 한다.

범주화 기법과 함께 우리 고전 문집의 원문에 사용된 용어를 효율적으로 분류할 수 있는 체계를 개발하기 위해서 고려해야 할 중요한 관점 즉 우리 고전 분야만이 가지고 있는 특성이 패시를 파악하여 이러한 고유한 패시에 맞는 분류 체계를 개발하는 것이 중요하다. 따라서 모든 주제 분야를 포괄하는 보편 분류 체계인 듀이십진분류체계(Dewey Decimal Classification)나 한국십진분류체계(Korean Decimal Classification)와 같은 기존의 열거형 분류체계를 적용하기 보다는 고전 분야의 지식 구조가 갖는 특성을 제대로 반영할 수 있는 패시 기법을 적용한 특수 분류 체계를 개발하는 것이 더 효율적이라고 할 수 있다.

우리 고전분야의 용어들을 분류하기 위한 구조를 개발하는 목적은 고전 용어 시소러스에 포함된 용어들에 관한 주제 정보를 제공하기 위한 것이다. 그리고 실제 검색환경에서 고전 분야의 용어들을 검색하여 고전 문집의 원문과 관련 내용에 접근하려는 이용자들을 지원하는데 기여하는 것이 궁극적인 목적이라고 할 수 있다.

2. 연구의 방법

일반적으로 분류 체계를 개발하기 위해서는 연역적 방법과 귀납적 방법을 고려할 수 있다. 연역적 방법은 이미 밝혀진 해당 주제 분야의

지식 구조를 바탕으로 최상위 범주로부터 최하위 범주로 전개해 나가는 방식이고, 귀납적 방법은 주제 분야의 문헌에서 추출한 색인어의 의미 분석을 통해서 공통적인 특성을 갖는 개념들을 추출하여 그 개념들을 대표하는 범주를 구성하는 방식이다. 방법상으로는 귀납적 방법이 이상적이기는 하지만 많은 시간과 인력이 필요하다는 단점을 가지고 있다. 이 연구에서는 연역적 방법을 사용하되, 필요에 따라서는 귀납적 방법을 병행하는 것으로 고전 용어 분류체계를 개발하였다.

고전 용어 시소러스의 대상이 된 민족문화추진회의 『한국문집총간』은 역대 주요한 인물의 문집 총 663종을 350 책으로 정리한 원전자료이다. 우리 고전 용어 범주화의 첫 단계는 최상위 범주 즉 패시를 설정하는 것으로, 여기에서는 최현배(1941)가 『우리말본』에서 두루 이름씨 즉 보통 명사를 구분하기 위해 용어 분류에 적용한 패시를 적용하였다. 우리 고전 문집에 사용된 용어를 분류함에 있어 출발점이 되는 최상위 범주를 설정하는 데, 우리말의 명사를 분류하기 위해서 사용한 최현배의 패시가 가장 유용한 참조 범주가 되리라고 판단하였다.

이와 같이 패시 기법을 적용하게 되면 동일한 개념적 특성을 갖는 용어들을 최상위 범주에서 구분하여 범주화할 수 있는데, 이러한 방식으로 최현배가 적용한 패시는 다음과 같다. 고전 용어 분류를 위한 최상위 범주로, 사람, 개, 나무 등을 포함하는 ‘자연물’ 범주, 집, 배, 기차 등을 포함하는 ‘인조물’ 범주, ‘시간’ 범주와 ‘공간’ 범주, 그리고 뜻, 마음, 생각 등을 포함하는 ‘정신’ 범주, 일, 울음, 노래 등을 포함하는 ‘행동’ 범주 등을 기본 패시로 하여 최상위 범

주를 구성하였다.

그리고 상위 범주에서 하위 범주로의 범주 구분은 동일 패시 내에서 계층적으로 전개하였다. 즉 고전 문집에 나타난 용어 중에서 ‘자연물’의 특성을 갖는 용어들을 범주화한 후에 ‘인조물’의 특성을 갖는 용어들을 범주화하여 계층적으로 범주들을 전개하였다. 따라서 이 연구에서 사용한 분류 체계는 최상위 범주에서는 패시 기법을 적용하였고, 그 이외 하위 범주의 전개는 계층적인 방식으로 범주화 하였다.

그러나 기존 분류 체계와 구분되는 차이점도 분명히 존재한다. 즉 기존의 분류 체계는 분류 대상을 특정한 한 범주에 배정하는 방식이었으나, 고전 용어의 경우 인물과 같은 분류 대상은 여러 특성 즉 여러 패시를 가지기 때문에, 이러한 여러 특성을 표현해야 할 필요가 있다. 따라서 인물과 같은 분류 대상이 가지고 있는 다양한 특성을 표현하기 위해서 이미 설정된 복수의 범주를 사용해서 구조화하였다. 다시 설명하면, 상이한 범주의 여러 특성을 하나의 용어에 부여할 수 있도록 구조화 하였다. 인물의 경우를 예로 들면, 황진이는 시인이면서 기생인 두 가지 특성을 동시에 가지고, 정약용은 문신, 실학자, 천주교인과 같이 여러 특성을 가지고 있기 때문에 여러 다른 범주의 다양한 특성을 하나의 대상에 동시에 적용하여 표현해야 할 필요가 있는 것이다.

또한 기존의 분류 체계에서 전통적으로 사용해 오던 시대나 지역과 같은 공통적인 특성을 표현하기 위해서 공통 구분 범주를 사용함으로써 기존의 인문학 분야의 시소러스의 분류 체계와 차별화하였다.

3. 분류 체계의 이론적 배경 및 시소러스에 적용된 분류 체계간 비교

3.1 열거형 분류 체계와 패킷 분류 체계

분류 체계의 구조원리에 의해 시소러스에 적용된 분류 체계를 구분하면 열거형 분류 체계와 분석합성형 분류 체계 즉 패킷 분류 체계로 나눌 수 있다. 그러나 여기에서는 시소러스 구축과 관련된 분야와 더불어 분류 체계를 이용하여 웹 자원을 조직한 연구들도 함께 살펴보고 한다.

그 이유는 이 논문에서 적용한 패킷 분류 체계와 전통적으로 지식을 구조화하는 연구에서 주로 사용되어온 열거형 분류 체계간의 구조상의 비교를 통해서 고전 용어 시소러스의 분류 체계 즉 패킷 분류 체계가 갖는 장점에 대해 언급하는 것이 필요하다고 생각했기 때문이다.

그리고 다음 절에서 시소러스에 적용된 분류 체계간의 실제적인 비교를 통해서 보다 구체적으로 분석하려고 한다. 다시 말해서 분류 체계의 구조 원리라는 이론적 관점에서 두 분류 체계를 고찰한 후에 시소러스에 적용된 사례를 분석하려는 것이다.

이 논문의 주제 영역인 고전 분야와 같은 특정한 주제 분야의 분류 체계를 이용한 기존 연구들은 대부분 열거형 분류 체계에 의존하고 있다(김성희 2002, 김정현 문지현 2002, 이란주, 성기주, 양정하 2001, 이창수 2000, 최재황 1998, 최희운 1998). 이러한 현상은 우리 문헌정보학이 기존의 열거형 분류체계로 대표되는 보편분류체계에 대한 연구에 많은 관심과 노력을 기울여 왔으나, 특정한 주제 분야의 지식을

색인하고 구조화하는 특수 분류 체계에 관한 연구에는 소홀해 왔음을 반증하는 것이다. 해당 주제 분야의 지식 구조가 가지고 있는 특성을 제대로 반영한 분류 체계를 만들기 위해서는 그 분야의 지식 구조에 적합한 분류 체계를 고려해야 함에도 기존의 보편분류체계인 열거형 분류체계를 적용함으로써 그 주제 분야의 지식 구조에 적합한 분류체계와는 거리가 생길 수밖에 없었던 것이다.

그리고 열거형 분류 체계는 근본적으로 구조적 결함을 가지고 있으며, 웹 자원을 조직하는데에는 더욱 그러하다. 그 이유는 인간이 사물을 인식하고 분류하는 사고 작용과 열거형 분류 체계가 지식을 분류하고 조직하는 방식이 다르기 때문에 이용자가 원하는 의미 있고 특정한 정보를 정확하게 제공할 수 없다는 것이다.

따라서 위의 선행 연구들에서 사용된 열거형 분류 체계가 가지고 있는 단점을 설명하면 다음과 같다. 첫째, 열거형 분류 체계에서는 지식 구조 전체를 완벽하게 표현하려고 시도하기 때문에 규모가 커지고, 더욱이 지식의 역동적인 속성 때문에 지식의 세계에 존재하는 '모든' 주제를 열거하는 것이 불가능하다. 둘째, 엄격한 계층 구조로 인하여, 이 분류 체계들은 학제적인 주제뿐만 아니라 새로운 주제를 받아들이기 어렵다. 셋째, 웹과 같이 특정 정보를 검색하는 데 이용될 경우, 열거형 분류체계는 매우 특정한 주제를 분류하는데 적합하지 않다. 넷째, 열거형 분류체계를 사용할 경우, 분류자가 정보 패키지에 맞는 체계를 선택하도록 강요당해서, 분류표에 포함되지 않은 해당 항목의 여러 중요한 측면들을 표현할 수 없다. 다섯째, 열거형 분류 체계의 계층적 구조 때문에 오

류가 발견되거나 보다 나은 구조가 개발되었을 때 주제를 재분류하기 어렵다.

그러나 열거형 분류 체계도 다음과 같은 장점을 가지고 있다. 첫째, 대다수 이용자들이 분류표에 익숙하고, 이용하기가 쉽다. 둘째, 분류 체계의 계층구조를 브라우징하기가 쉽다.

이와 함께 고전 용어 시소러스와 유사한 인문학 분야의 시소러스의 분류 체계에서도 열거형 분류 체계를 사용하고 있다. 대표적으로 국사편찬위원회에서 제공하고 있는 한국역사용어 시소러스의 주제 분류를 보면, 용어들을 범주화하는 기준이 모호하고 분류 체계의 구조가 대단히 복잡하여 이용자들이 그 분류 체계 사용 시에 불편을 느낄 수밖에 없는 상황이다. 따라서 기존의 열거형 문헌 분류 체계들이 가지고 있는 구조상의 경직성을 그대로 답습하고 있으며, 특히 기존의 분류 구조에서 사용하고 있는 공통 구분을 설정하지 않음으로 해서 분류표의 크기가 필요 이상으로 커졌고, 이용자가 분류 체계를 이용해서 찾고자 하는 용어들을 검색하기가 매우 어렵다.

반면에 패킷 분석을 이용한 분류 체계의 장점은 다음과 같다(유영준 2004). 첫째, 학제적인 주제에 대해서 유연성을 제공한다. 즉 패킷 분석에 의한 색인 기법에서는 기본 개념만을 전개하기 때문에 학제적 관점을 가진 문헌도 두 개 이상의 개념들을 조합하여 표현할 수 있으며, 패킷 간의 조합은 규칙적이고 예측 가능하다. 둘째, 새로운 개념을 편입하기가 용이하다. 열거형 분류 체계를 사용하는 색인자는 새로운 주제를 표현할 수 있는 용어를 분류 체계에서 제공할 때까지 기다려야 한다. 그러나 패킷 분석을 사용하면, 새로운 주제를 기존의 용

어들과 조합해서 표현할 수 있으며 분류 체계에 명시되지 않은 복잡한 개념도 구조화할 수 있다. 셋째, 주제의 특정성을 높일 수 있다. 즉 패킷 분석을 이용함으로써 표현할 수 있는 주제의 내용이 기존 분류 체계를 활용한 색인 기법보다 더 세세하고 특정하다. 넷째, 새로운 패킷을 추가할 수 있다. 이것은 이용대상 언어 특정한 관점과 같은 문헌의 범위를 나타내는 패킷들을 용이하게 추가할 수 있다는 의미이다. 예를 들면 웹 자원의 멀티미디어와 관련된 특성을 독립된 패킷으로 범주화할 수 있다.

그러나 패킷 분류 체계도 다음과 같은 단점을 가지고 있다. 첫째 열거형 분류 체계와 달리 이용자 대다수에게 매우 낯설다. 둘째, 매우 복잡하고 경험이 없이는 이용하기 불편하다. 셋째, 분류표의 계층구조를 브라우징할 수 있도록 통합하기가 어렵다.

이러한 연구 경향과 차별화한다는 면에서 이 연구에서는 고전 용어 시소러스의 분류 체계를 구축하기 위해서 채용한 구조화 방법은 패킷 분석과 계층적인 개념 전개 방식을 혼합한 방식으로 위에서 제시한 보편분류체계인 열거형 분류 체계와 특수분류체계 구축에 유용한 패킷 분류 체계의 장점을 함께 수용하려고 시도하였다. 그래서 범주화의 첫 단계에서 6개의 패킷을 적용하여 범주를 구분하고, 이어서 고전 분야의 지식구조에 근거해서 하위 범주들을 구분하였다.

3.2 시소러스에 적용된 분류 체계 비교

3.2.1 시소로패킷(Thesaurifacet)

앞 절에서 언급한 패킷 분석의 방법론을 적

용하여 시소러스의 용어들에 분류 체계를 적용한 시스템으로는 시소로 패킷을 들 수 있다. '시소로패킷'이란 말은 시소러스와 패킷 분류를 합성한 것으로 분류표와 시소러스의 장점을 통합하여 시소러스에서는 개념간의 관계 즉 상위어, 하위어, 관련어 등을 확인할 수 있고 분류 체계에서는 계층 구조를 확인할 수 있다.

시소로패킷은 과학기술 전 분야를 수록 범위로 하고 용어 수는 16,000여개에 달하며 7,000여 색인 표목을 수록하고 있다. 그리고 개념은 시소러스와 분류표에 각각 수록되어 있으며 시소러스는 전통적인 분류표의 자모순 주제색인을 대체하는 기능을 하며, 분류 기호를 통해 상호 연결되어 있다. 따라서 각각 독립적으로 사용하거나 상호 보완적으로 사용될 수 있다(김태수 2000).

즉 시소러스의 디스크립터에 부여된 분류 기호를 통해 분류표와 직접 연결되고, 아울러의 미구조에 따라 상위개념과 하위개념을 확인할 수 있으며 이 연결망을 통해 색인파일의 작성을 지원하게 된다. 시소로패킷은 단일 개념 뿐만 아니라 합성주제와 그 분류기호도 동시에 제시하고 있으며 시소러스와 분류 체계가 완전히 결합된 형태라고 볼 수 있다.

이러한 시소로패킷의 장점을 고려해 볼 때, 고전 용어 시소러스의 용어들을 개념적 특성에 따라 범주화하는 현재의 연구를 더 발전시켜야 한다. 그리고 고전 용어 시소러스의 모든 디스크립터를 계층적으로 범주화하여, 이용자들이 고전 문집에 나타난 용어를 보다 효율적으로 검색할 수 있도록 지원하기 위한 모형으로 시소로패킷을 고려할 수 있다.

3.2.2 예술 및 건축 시소러스(Art & Architecture Thesaurus: AAT)

예술과 건축 분야의 용어를 대상으로 패킷 분류를 적용한 전형적인 시소러스로서, 7개의 패킷과 33개의 계층으로 개념간의 관계를 제시하여 지식베이스로 기능한다(Molholt and Peterson 1993). 7개의 패킷은 연상 개념(associated concept), 물리적 속성(physical attributes), 양식/시대(style and periods), 행위자(agents), 행위(activities), 재료(materials), 대상(objects)인데, 최현배가 우리말의 명사어를 분류 할 때 적용한 패킷과 일치하는 패킷이 5개 있음을 알 수 있다.

이 패킷을 이용하면 주제 분야의 체계와 내용의 변화와 관계없이 분류 체계를 그대로 적용할 수 있다. 이 시소러스에서는 디스크립터에 컴퓨터로 작성된 행번호를 부여하여, 새로운 디스크립터가 추가될 때 마다 자동으로 갱신이 가능하다(Art & Architecture Thesaurus 1994).

고전 용어 시소러스에서 분류 체계를 도입한 것은 단일 용어로는 표현할 수 없는 개념의 계층 수준을 표현할 수 있는 특성을 분류가 지니고 있기 때문이다. AAT와 같이 패킷 분류를 적용함으로써 해당 주제 분야의 특성을 보다 정확하게 파악하고 그 지식 구조를 제대로 범주화할 수 있는 것이 고전 용어 시소러스가 패킷 분류를 채택한 이유라고 할 수 있다.

3.2.3 국회도서관 시소러스

국회도서관 시소러스에는 정치, 경제, 법률, 행정, 사회학, 사회복지, 여성, 교육, 심리, 신문 방송, 문헌 정보, 정보통신·교통, 역사·지리, 인류·민속 등 사회과학분야 14개 주제의 55,678

개 디스크립터와 비디스크립터 용어를 포함하여 용어를 자모순 및 계층 순으로 수록되어 있다(국회도서관 2004).

그리고 검색 시에 시소러스를 활용하기 위해

서 '시소러스 참조검색'을 검색 메뉴에 별도로 제시하고 있다. 국회도서관에서 제공하는 시소러스 참조검색의 예는 <그림 1>과 <그림 2>와 같다.



<그림 1> 국회도서관 시소러스 참조 검색의 예



<그림 2> 국회도서관 시소러스 참조 검색의 예

〈그림 1〉은 '정약용'이라고 입력하여 검색해 본 예로서, 검색 대상인 정약용의 동의어, 상위어, 관련어를 제시하고 있으며 각각의 용어를 선택하여 '검색실행'을 누르면 해당 용어와 연결된 문헌에 접근할 수 있다.

그리고 〈그림 2〉는 '시소러스'를 입력한 예로서, 시소러스의 하위어와 관련어를 보여주고 있고, 개별 용어들을 선택하여 검색할 수 있고 해당 용어와 관련 있는 문헌에 접근할 수 있다.

그러나 국회도서관 시소러스는 검색 대상이 되는 용어를 중심으로 유의어나 상위어, 하위어, 관련어 등은 확인할 수 있지만 해당 용어가 자신이 속하는 범주 내에서 어떤 계층에 속하는지를 알 수 없다. 즉 시소러스에 분류 체계를 적용함으로써 디스크립터들의 계층 관계를 확인하고, 검색어의 의미 범주를 확장하거나 한정하는 방식으로 검색을 지원하지 못한다. 그리고 국회도서관 시소러스의 용어를 사회과학 분야로 범위를 제한한 것은 아쉬운 점으로 남는다. 국립중앙도서관과 함께 우리나라의 대표적인 도서관으로서 모든 주제 분야를 포괄하는 시소러스와 분류 체계를 개발하여 이용자들에게 제공하는 것이 더 바람직하다고 생각한다.

위에서 언급한 단점에도 불구하고 국회도서관 시소러스는 앞에서 언급한 시소로페식과 동일하게 용어들의 모든 관계 유형을 구조화하여 해당 분야에 대한 지식이 없는 이용자들이 시소러스 참조검색에서 제시된 그 분야 용어를 활용하여 편리하게 검색할 수 있는 서비스를 제공하고 있다. 이와 같이 시소러스에 포함되어 있는 용어들의 관계를 활용해서 이용자들의 검색을 지원하는 것이 고전 용어 시소러스가 나아가야 할 방향을 결정하는 데 도움이 되리

라고 생각한다.

3.2.4 국사편찬위원회 한국역사용어 시소러스

한국역사용어 시소러스는 열거형 분류 체계를 적용한 예로서 2005년 12월말 현재 총 6만여 개의 용어를 수록하고 있고, 용어 추출과 선정은 중·고등학교 국사 교과서 및 역사 사전류, 국사편찬위원회에서 전산 입력한 자료와 역사정보통합시스템에 구축된 원전(原典) 자료 등을 대상으로 이루어졌다. 그리고 정확한 역사용어를 알지 못하더라도, 이 주제 범주들을 이용하여 이용자들이 한국역사용어에 접근할 수 있도록 하였다.

국사편찬위원회(2006)에서 제시된 상세한 주제별 분류 체계는 다음과 같다. 역사일반, 정치·행정·법제, 국방·군사, 외교·국제관계, 경제·산업, 재정·금융, 교통·통신, 사회·생활, 사회운동·독립운동, 문화·예술, 전통사상·종교, 학술·과학기술, 교육 등 13개 주제 범주와 그 외에 인명, 지명, 서명, 문화재 등의 범주의 용어로 구성되어 있다. 그리고 국사편찬위원회(2004)에서는 순서대로 13개 범주를 최상위 항목으로 기재하고 이 범주들을 '한국역사용어 주제 분류'라는 용어로 표현하였으며, 인명, 지명, 서명 등은 '전거데이터 주제 분류'라는 용어를 사용하였고, 문화재 범주는 포함시키지 않았다.

한국역사용어 시소러스의 분류 체계는 열거형 분류 체계의 구조에 의해 개념을 계층적으로 전개하는 구조를 가지고 있다. 따라서 전체 분류 구조에서 몇 가지 문제점을 안고 있으며 그 상세한 내용은 다음과 같다. 첫째, 기존의 열거형 분류 체계와 동일한 문제점을 가지고 있

어서 분류 체계에 유연성이 떨어지고 새로운 범주를 편입해야 할 필요가 있을 때, 전체 범주를 재배열해야 하는 문제가 발생할 수 있다. 또한 동일한 주제가 여러 범주로 분산되는 결과를 낳게 된다. 예를 들어 최상위 범주의 '정치·행정·법제' 아래에 '법·법령' 범주가 있고 하위 개념에 각종 법들이 제시되어 있으나 '군사법령'은 '국방·군사' 범주에 포함되어 법령이라는 동일한 주제가 여러 범주에 분산되어 이용자들이 법령을 검색할 때 모든 법령을 동시에 검색하기가 어렵다. 이외에도 '재정·금융'의 하위 범주에 '재정' 그리고 '재정'의 하위 범주에 다시 '재정정책·법령'이 하위 범주로 설정되어 있다. 이렇게 주제가 분산될 뿐만 아니라 분류 체계가 대단히 복잡하고, 범주를 설정하는 기준에 일관성이 결여되어 있다.

둘째, 최상위 범주의 대부분이 복합표목 즉 상이한 주제를 표현하는 두 개 이상의 용어들이 하나의 범주를 지칭하는 표목으로 구성되어 있어서 앞으로 분류 체계 내에서 상이한 범주간의 관계를 설정할 필요가 있을 때 심각한 문제를 야기시킬 수 있다. 즉 시소로패시이나 국회도서관 시소러스의 경우처럼 시소러스의 모든 용어의 관계를 설정하여 이용자들의 검색을 지원하는 수단으로서 시소러스를 활용할 때 검색 효율에 악영향을 미치는 요인으로 작용할 것이다. 결국 이러한 문제점은 분류 체계 구조의 개선과 발전을 저해하는 요소가 될 것이다.

셋째, 동일한 범주가 반복적으로 설정되어 있다. 이러한 현상은 시대를 구분하기 위해서 설정하는 시대 범주뿐만 아니라 분류 체계 전반에서 나타나는 현상이다. 예를 들어, '정치·행정·법제' 아래에 '사법', '입법', '선거', '인사'

등의 범주가 있음에도 '정치' 범주의 하위 범주에도 '사법', '입법', '선거', '인사' 등이 반복적으로 범주로 설정되어 있다. 즉 범주를 설정하는 기본적인 틀과 범주들의 의미를 정확하게 규정하지 못한 결과라고 할 수 있다. 이러한 상황으로 말미암아 분류 체계의 계층 수준이 대단히 복잡하고, 분류표의 크기가 지나치게 커져서 분류표를 관리하는 데 어려움을 겪을 수밖에 없다.

그리고 시대 구분 범주를 분석한 내용은 다음과 같다. '정치·행정·법제' → '법·법령' → '고려시대법', '조선시대법' 등의 하위 범주로 범주화하거나, '재정·금융' → '재정' → '재정정책·법령' → '고려시대재정', '조선시대재정' 등으로 범주화하고 있다. 이와 같이 시대 범주를 여러 범주와 계층 수준에서 반복적으로 적용하는 것은 분류 체계를 복잡하게 만들고 이용자들이 분류표를 사용하여 원하는 용어를 검색하는데 장애 요인이 될 수 있다. 뿐만 아니라 시대 범주를 동일 수준에 있는 다른 범주들과 동일한 개념으로 전개한 것은 분류 체계에서 범주화의 일반 원칙에 어긋나는 범주화라고 할 수 있다. 따라서 시대나 지역을 구분하는 표준 구분 즉 공통 구분을 별도의 표로 작성하여 시대나 지역을 구분할 필요가 있는 범주에만 적용하도록 개선해야 한다.

더불어 '역사일반' → '시기·시대' → 시대구분은 가나다순으로 배열 즉 고대, 고려시대, 근대, 근세, 남북국시대, 삼국시대, 선사시대, 역사시대, 원삼국시대, 일본시대명, 중세, 지질시대, 철기시대, 현대, 후삼국시대로 배열하였는데, 시대를 가나다순으로 배열하는 것은 오히려 이용자의 검색에 불편을 줄 수 있으므로 시

대순으로 배열하는 것이 바람직하다.

넷째, 계층관계와 같은 용어간의 관계 유형을 적용하여 모든 용어들을 구조화 한 것이 아니라, 시소러스의 용어들을 최하위 수준에만 제시하여 단순히 용어들의 범주만을 설정한 것으로 보인다. 궁극적으로는 시소러스의 모든 용어들의 계층 관계를 적용하여 용어들을 범주화하여 이용자들의 검색을 지원하는 방향으로 시소러스의 분류 체계의 유용성을 제고시켜야 할 것이다.

결론적으로 한국역사용어 시소러스는 열거형 분류 체계의 경직성에 의해서 파생되는 구조적인 문제와 모든 용어의 계층 관계를 구조화하여 이용자들의 검색을 보다 효율적으로 지원하기 위한 방향으로의 분류 구조의 개선과 함께 시대나 지역을 구분하기 위한 공통 구분표를 개발해야 할 과제를 안고 있다고 말할 수 있다.

3.2.5 고시조 시어 시소러스

고시조 시어 시소러스에서는 데이터베이스로 구축된 고시조 5,180수와 주요 시가집 13종에 수록된 7,183수를 대상으로 어절 단위의 어휘를 추출한 후, 이를 기반으로 기초 어휘를 수집하고 유의어와 동의어를 범주화하였다. 그리고 동음이의어나 동형이의어, 복합어 등에 대한 어휘 통제를 거쳐 시소러스의 기반이 되는 2,000개 용어로 구성된 시소러스를 구축하였다(우응순 권순희, 한창훈 2002).

이 용어들을 대상으로 고시조 시어 시소러스의 분류 체계를 개발하였는데 그 구조는 아래와 같다. 먼저 패킷 분석에 의하여 대분류 항목을 '時間', '空間', '人物', '事物', '觀念', '事件',

'言語'로 구분하고, 고시조의 주제적 특성을 반영하여 중분류 범주 13개와 소분류 범주 60개를 설정하였다. 그리고 각 범주마다 분류기호를 주었으며, 최상위 범주를 패킷을 적용한 것은 고전 용어 분류 체계와 유사한 점이라고 볼 수 있다.

그러나 '공간' 범주의 하위 범주인 '자연의 공간' 범주에 '속세'를 배열하고, 관념의 공간에 '고사', '소설' 등을 배열하여 작위적인 범주를 포함시킴으로써 개념이 합리적으로 범주화되지 못했고, '관념' 범주의 하위 범주인 '의식'과 '감정' 범주가 지나치게 주관적인 관점에서 범주화하였다. 이러한 점은 이용자들이 분류 체계를 활용하는데 장애 요인이 될 수 있다. 그리고 다른 인문학 분야 시소러스의 분류 체계와 동일하게 시대와 지역 범주를 공통 구분으로 설정하지 않는 등의 문제점을 드러내었다.

그럼에도 고시조 시어 시소러스에서는 용어들의 계층관계와 연관관계를 적용하여 용어들을 구조화한 사례를 제시함으로써 고전 용어 시소러스의 발전 방향에 시사하는 바도 있다고 말할 수 있다.

4. 고전 용어 시소러스의 분류 체계

4.1 분류 기준

앞에서 언급한 기존의 선행 연구의 연구 경향과 구별되고 현재까지 시소러스를 구축하여 서비스하고 있는 시스템들의 구조상의 단점을 보완하여 보다 나은 검색 서비스를 제공하려는 목적으로 이 연구에서 제시한 분류 기준은 다

음과 같다.

첫째, 최현배가 『우리말본』에서 두루 이름씨 즉 보통 명사를 구분하는 용어 분류에 패짓 기법을 적용하였는데, 이를 고전 용어 분류 체계의 최상위 범주 즉 기본 패짓으로 설정하였다. 사람, 개, 나무 등을 포괄하는 '자연물' 범주, 집, 배, 기차 등을 포괄하는 '인조물' 범주, '시간' 범주와 '공간' 범주, 그리고 뜻, 마음, 생각 등을 포괄하는 '정신' 범주, 일, 울음, 노래 등을 포괄하는 '행동' 범주 등을 기본 패짓으로 하여 최상위 범주를 구성하였다. 이외의 하위 패짓들과 구체적인 패짓의 수준은 분류 체계와 분류표에서 상세하게 제시하였다.

둘째, 기본 패짓은 하위 범주들의 개념적 속성을 구분하기 위한 범주로 이용자에게 제시할 것인지를 결정하는 것은 어려운 과제라고 할 수 있다. 그 이유는 패짓을 적용한 최상위 범주의 의미를 이용자가 제대로 이해하지 못하거나 혹은 익숙하지 않은 이유로 인해 분류 체계를 이용하지 않는 상황이 발생할 수 기 때문이다. 이러한 점은 패짓 분류가 가지는 단점 중 가장 해결하기 어려운 것으로서 이용자에게 친숙하지 않아서 발생할 수 있는 장애요인이다. 따라서 분류 체계에 대한 이용자들의 이용 상황에 따라 패짓을 적용한 최상위 범주를 보여 줄 것인지를 결정할 필요가 있다고 생각한다.

셋째, 기초 수준 범주는 검색 화면에서 이용자가 직접 선택하는 범주 중에서 중간에 위치하는 범주를 지칭한다. 즉 범주의 구성원이 다양한 계층 관계로 연결되어 있을 때 보편적으로 우리가 사물을 지각하고 개념화하는 수준이고, 전반적으로 기초 수준에서 개념을 가장 쉽게 이해할 수 있다. 따라서 범주의 여러 수준

가운데 특정 수준에 정보가 집중되고 이 수준에서 지식이 가장 용이하게 조직되며, 대개는 전체 계층 수준의 중간에 위치한다.

넷째, 역사 시대를 구분하는 시대 범주와 행정 구역을 구분하는 지역 범주를 공통 구분으로 규정하고, 시대와 지역 구분은 분류 체계를 구성하는 모든 범주의 시대나 지역을 구분할 때 공통적으로 적용한다. 그리고 지역 범주의 구분은 『新增東國輿地勝覽』에서 제시한 행정 지명을 기준으로 작성하였다.

기존의 문헌분류체계에서는 보편적으로 널리 사용되고 있는 공통 구분이 인문학 분야 시소러스의 분류 체계에서는 지금까지 적용된 바가 없다. 따라서 특정 주제 분야의 분류 체계를 개발하는 연구에서는 처음으로 시도한 사례라고 할 수 있다.

4.2 분류 체계

분류 체계에서는 패짓을 적용한 최상위 범주부터 최하위 범주까지 계층 수준별로 전개하여 표현하되, 패짓을 적용한 최상위 범주는 이용자에게 제시하지 않는 범주이므로 분류 체계상에서는 포함시키지 않았다.

4.2.1 인물

인물 범주는 왕실, 관인, 문사, 예술인, 종교인, 기술인, 집단, 기타의 하위 범주로 구분한다. 그리고 각 범주의 하위 범주 구분은 다음과 같다. 왕실 범주는 국왕(황제, 추존왕), 제후, 왕비(황후), 후궁, 왕자, 공주, 종친(군), 외척·인척(부원군, 부마)으로 전개하고, 관인 범주는 문신, 무신, 경대부, 아전(이서), 환관, 역관,

궁녀(의녀)로, 문인/학자 범주는 문인, 학자, 유학(幼學)으로, 예술인 범주는 서화가 음악인, 광대로, 종교인 범주는 승려 무격 천주교인, 동학교인으로, 기술인 범주는 지관, 장인, 상인, 의원, 기생, 백정으로, 집단 범주는 성씨, 종족, 조직(당파, 결사), 기타로, 기타 범주는 성현, 은자, 도사, 공신, 양인(良人), 귀화인, 외국인, 도적, 역적, 효자, 열녀, 노비, 기타인물로 각각 세분한다.

4.2.2 자연물

자연물 범주는 동물, 식물, 광물, 천체, 신체, 기타자연물의 하위 범주만으로 구분하고, 그 이하의 하위 범주로는 전개하지 않는다.

4.2.3 문헌

문헌 범주는 서명, 편명, 작품명, 지도, 서화, 약보, 판본, 기타문헌으로 구분하고, 그 이하의 하위 범주로는 전개하지 않는다.

4.2.4 유적

유적 범주는 건물과 시설의 하위 범주로 구분한다. 그리고 건물은 궁궐, 사찰, 사제(私第)(재실, 정사), 사당, 누정(정자, 누각), 서원·향교, 기타건물(관, 원, 사)로 시설은 성·진·책, 창고, 능묘(능, 제단, 비각), 기타시설(도요지, 저수지, 정원)의 하위 범주로 세분한다.

4.2.5 물품

물품 범주는 의복, 식약품, 생활용품, 기타용품의 하위 범주로 구분한다. 그리고 각 범주의 하위 범주 구분은 다음과 같다. 의복은 관복, 예복, 상복, 장신구, 신발, 포복으로, 식약품은 식

품, 음료, 식재료, 탕약, 약재로, 생활용품은 그릇, 집기, 가구, 화폐, 공예품으로, 기타용품은 의료용품, 의례용품, 공구, 농기구, 악기, 무기, 형구, 활자, 선박과 기타물품으로 각각 세분한다.

4.2.6 법제

법제 범주는 관청, 관직, 법률, 제도, 의례, 기타 법제의 하위 범주로 구분한다. 그리고 관청, 관직, 법률 제도 범주는 적합한 하위 범주를 차후에 설정하기로 하고, 의례 범주는 국가의례, 향촌의례, 가례(관, 혼, 상, 제), 기타 의례로 세분한다.

4.2.7 시간

시간 범주는 국명(왕조), 연호, 간지, 절기, 월일시의 하위 범주로 구분하고, 그 이하의 하위 범주로는 세분하지 않는다. 그리고 시간 범주는 공통 구분의 시대별 범주와는 구별되는 범주이다.

4.2.8 지명

지명 범주는 자연 지명, 행정 지명, 기타 지명의 하위 범주로 구분한다. 그리고 자연 지명 범주는 토와 수로 세분하고, 땅 범주를 구성하는 용어들은 산, 산맥, 고개, 바위, 평야, 도서 등 이고, 물 범주를 구성하는 용어들은 바다, 강, 하천, 호수, 저수지, 포구, 항만, 나루 등 이다. 행정지명과 기타지명 범주는 그 이하의 하위 범주로 세분하지 않는다.

4.2.9 사상

사상 범주는 이론, 기법, 현상, 술어, 기타 사상의 하위 범주로 구분한다. 그리고 각 범주의

하위 범주는 다음과 같다. 이론 범주는 학설과 교리(불교, 도교, 천주교, 천도교)로, 기법 범주는 기술과 방법으로, 현상 범주는 자연 현상과 신체 증상으로 술어 범주는 고사성어, 관용어, 핵심어로 각각 세분한다.

4.2.10 사건

사건 범주는 전쟁, 사회, 역모(반란), 송사, 풍속, 기타 사건의 하위 범주로 구분한다.

4.3 분류 체계의 구조적 특성

고전 용어 시소러스 분류 체계의 구조적 특성 중의 하나는 패킷 분석을 적용하여 최상위 범주를 6개로 범주화하여 최상위 범주에 속하는 하위 범주들을 보다 명확하게 구분하고 해당 범주에 속하는 개념들을 보다 용이하게 배정할 수 있다는 점이다. 즉 시소러스에 포함되는 고전 용어 전체를 6개의 특성 즉 자연물, 인조물, 시간, 공간, 정신, 행동에 의해 구분하고, 해당 특성을 갖는 개념들만을 분리하여 해당 범주에 배정한다. 따라서 해당 개념을 해당 범주에 보다 정확하고 용이하게 범주화할 수 있다는 점에서 기존의 인문학 분야에서 개발된 시소러스의 분류 체계와 비교해 볼 때 두드러진 장점이라고 할 수 있다.

그리고 패킷 분석을 적용함으로써 우리 고전 분야의 용어들이 가지고 있는 특성을 제대로 반영할 수 있는데, 그 이유는 다음과 같다. 패킷 분석은 분석의 대상이 되는 특정 주제 분야의 고유한 특성을 반영하여 그 주제 분야의 용어들을 분류할 수 있는 체계를 구조화 할 수 있다는 것이 가장 큰 장점이기 때문이다. 따라서 패

킷 분석은 우리 고전 분야의 용어들을 분류함으로써 고전 분야의 지식 구조를 가장 잘 표현할 수 있는 지식 구조화 방법 중 하나라고 할 수 있다.

고전 용어 시소러스 분류 체계의 또 다른 구조적 특성은 기존의 문헌분류체계에서 사용해진 공통 구분을 적용한 것이라고 할 수 있다. 고전 용어 시소러스에 포함된 용어들이 지역적 혹은 시대적 특성을 포함하고 있을 때, 즉 용어에 지역이나 시대와 관련된 개념적 특성을 추가할 필요가 있을 때, 지역이나 시대를 구분하는 공통 범주의 개념들을 추가하여 범주화하는 것이다. 기존의 인문학 분야에서 개발된 시소러스들은 공통 구분을 전혀 사용하지 않았기 때문에 대단히 복잡한 계층구조와 동시에 하위 범주를 전개함에 있어서 일관성 있게 분류 기준을 적용하지 못했다. 특히 시소러스에서 분류 체계를 적용하는 이유는 이용자가 검색하려는 관련 주제의 전체적인 내용과 구조를 파악하여 보다 정확한 검색어를 선택하기 위해서 해당 용어가 속한 범주의 계층 관계를 확인하기 위한 것이다. 기존의 인문학 분야 시소러스에서는 모든 범주에서 필요에 따라 무원칙하게 혹은 일관성 없게 시대적이거나 지역적인 특성을 구분하여 범주를 전개함으로써, 오히려 이용자가 분류 체계를 활용해서 원하는 문헌을 검색하는 데 필요한 검색어를 찾아내어 질의를 확장하는 것을 더 어렵게 만들었다.

위의 세 가지 특성과 다르게 단점이라고 지적할 수 있는 점은 이용자가 분류 체계를 사용해서 보다 쉽게 용어를 검색할 수 있는 인터페이스를 구축하는 것과 관련된 것이다. 이러한 점은 패킷 분석을 적용하여 구조화한 분류 구

조에 대한 이용자의 실제적인 활용과 밀접하게 연관된 문제이다. 예를 들면 6개의 기본 패킷을 이용자에게 제시하는 것이 이용자의 검색에 어떤 영향을 미칠 것인가에 대한 것이다.

그리고 이 6개의 패킷을 적용한 최상위 범주는 고전 용어와 직접적으로 연관된 범주라고 할 수는 없지만, 고전 용어 시소러스의 분류 체계의 구조에서는 하위 범주들을 특성에 따라 구분하는 중요한 역할을 하기 때문이다. 따라서 이용자에게 익숙하지 않은 범주를 제시함으로써 시소러스의 분류 체계를 활용하는데 있어서 오히려 부정적인 영향을 미칠 수 있다. 이러한 이용자와의 상호작용과 관련된 문제는 실제로 이용자의 검색 행태를 파악한 후에 결정하는 것이 바람직하다.

4.4 분류 체계의 적용

연역적인 방법으로 개발된 분류 체계의 구조 안에서 시소러스에 포함된 용어들이 어떻게 조직되고 범주화되는가를 설명하기 위해서 디스크립터 작성 사례를 제시하고, 이 분류 체계를 이용해서 디스크립터들을 실제로 분류하였고, 분류 체계에 의한 검색 서비스가 이용자들에게 어떤 방식으로 제공될 것인지를 간략하게 제시하였다. 그리고 이러한 과정을 통해서 개발된 분류 체계가 시소러스의 용어들을 분류하는데 얼마나 적합한지를 검토하고, 최종적으로 분류 체계의 범주들을 수정해 나감으로써 고전 용어를 분류하는데 보다 나은 분류 체계로 그 구조를 개선해 나갈 것이다. 그리고 디스크립터 작성 사례는 아래와 같다.

- [우선어] 강희안(姜希顔)
- [범위주기] 조선 세종(世宗)-세조(世祖) 때의 문신. 시·글씨·그림에 모두 뛰어나 삼절(三絶)이라 불리었음.
- [유의어] 경우(景愚)>>삼절(三絶)>>인재(仁齋)>>청천자(靑川子)
- [관련어] 강석덕(姜碩德)>>강희맹(姜希孟)>>+고사관수도(高士觀水圖)>>+동국정운(東國正韻)>>+양화소록(養花小錄)>>+인재집(仁齋集)
- [시대분류] 조선전기
- [지역분류] 경상도>진주
- [유형분류] 인명>관인>문신

위에 제시된 사례는 우선어로 선정된 강희안(姜希顔)'에 대한 범위주기, 유의어, 관련어를 작성하였고, '유형분류'가 고전 용어의 주된 분류 체계이고, 디스크립터가 인물이므로 공통 범주인 시대와 지역을 적용한 것을 확인 할 수 있다.

- [우선어] 토란(土卵)
- [초집의미] 천남성과(天南星科)의 다년초(多年草). 추석 같은 명절에 먹는 전통 음식 중 하나로, 남쪽 지방에서 많이 재배되며 전통 민화나 문양에서 무병장수(無病長壽)를 상징함
- [유의어] 군자우(君子芋)>>우경(芋莖)>>*토련(土蓮)>>*우자(芋子)>>*토지(土芝)>>*우(芋)
- [관련어]
- [시대분류]

[지역분류]

[유형분류] 자연물>식물

그리고 두 번째 사례는 자연물 범주의 토란(土卵)'으로서 시대나 지역을 구분할 필요가 없는 용어이고, 자연지명이나 행정지명 범주에 속하는 용어들도 동일한 개념적 특성을 갖는다.

[우선어] 퇴계집(退溪集)

[초점의미] 조선 전기의 대유(大儒) 이황(李滉)의 문집(文集). 문인(門人) 조목(趙穆)의 주도로 도산서원(陶山書院)에서 간행하였으며 여러 종류의 판본이 전함.

[유의어] 이선생문집(李先生文集)>>계집(溪集)>>퇴도선생집(退陶先生集)>>퇴계전서(退溪全書)>>퇴계고(退溪稿)>>퇴계일고(退溪逸稿)>>퇴도선생유고(退陶先生遺稿)>>퇴도선생유집(退陶先生遺集)>>퇴도시집(退陶詩集)>>퇴도집(退陶集)>>퇴집(退集)>>도산선생문집(陶山先生文集)>>도산전서(陶山全書)

[관련어] 이황(李滉)>>조목(趙穆)>>퇴계언행록(退溪言行錄)

[시대분류] 조선후기

[지역분류] 경상도>안동

[유형분류] 문헌>서명

세 번째 사례는 서명 범주의 '퇴계집(退溪集)'으로 이황이 저술한 서적이었고, 간행된 연도는 시대 구분을, 간행된 장소는 지역 구분을 이

용해서 분류 대상의 특성을 표현하였다.

[우선어] 교서관(校書館)

[초점의미] 조선시대 삼관(三館)의 하나로, 태조(太祖) 때 설치되어 경적(經籍)의 인쇄·반포와 향축(香祝)·인전(印篆)을 담당하던 관청

[유의어] 전교서(典校署)>>교서감(校書監)>>운각(芸閣)>>운관(芸館)>>예각(藝閣)>>외각(外閣)

[관련어] +홍문관(弘文館)>>+예문관(藝文館)

[시대분류] 조선시대

[지역분류]

[유형분류] 법제>관청

[우선어] 설순빙리(雪筍冰鯉)

[초점의미] 중국 삼국시대 오(吳) 나라의 맹종(孟宗)이 눈 속에서 죽순(竹筍)을 구한 일과 남북조시대 진(晉) 나라의 왕상(王祥)이 얼음을 녹여 잉어를 얻은 일로, 지극한 효성을 뜻함

[유의어] 빙리설순(冰鯉雪筍)>>설순빙리(雪筍冰鯉)>>설순빙어(雪筍冰魚)>>빙리동순(冰鯉冬筍)>>동순빙리(冬筍冰鯉)>>순리(筍鯉)>>순리(筍鯉)>>순어(筍魚)>>어순(魚筍)

[관련어] +맹종(孟宗)>>+왕상(王祥)>>+맹종설순(孟宗雪筍)>>+왕상빙리(王祥冰鯉)

[시대분류] 삼국시대>>남북조시대

[지역분류] 중국

[유형분류] 사상>술어>고사성어

네 번째 사례는 법제>관청 범주의 '교서관(校書館)'으로 조선 태조 때 설치된 관청이고, 다섯 번째 사례는 고사 성어 범주의 '설순빙리(雪筍冰鯉)'라는 용어로 시대는 중국의 남북조시대이고, 지역은 당연히 중국으로 범주화하였다.

이와 같이 작성된 디스크립터들을 분류 체계에 적용하기 위해서 선택한 우선어들은 위에서 제시한 다섯 용어를 제외한 5개의 용어는 다음과 같다.

'건물' 범주의 숭례문(崇禮門), '국명' 범주의 조선(朝鮮), '자연 지명' 범주의 섬진강(滸津江), '공예품' 범주의 설도전(薛濤箋), '사건' 범주의 분서갱유(焚書坑儒)로 분류 체계에 적용하기 위해서 선택한 모든 요영는 상위 10개 범주에서 한 사례씩을 선정하였다.

그리고 이들 우선어를 고전 용어 분류 체계

에 배정해 본 결과는 아래의 <표 1>과 같다.

실제로 작성된 디스크립터들을 분류 체계에 적용한 결과에 의하면, 범주의 특성에 따라서 전개되는 계층의 수가 범주마다 다르며, 용어의 개념적 특성에 따라서 공통 범주의 구분 유무가 결정되었음을 알 수 있다. 그리고 굵고 기울인 글자체가 우선어를 나타내는 것이다.

div>

따라서 분류 체계에 의해 디스크립터들을 실제로 분류함으로써 분류 체계의 구조상의 문제 즉 분류 기준의 비일관성이나 개별 범주의 추가나 통합과 같은 수정 과정을 통해서 고전 분야 시소러스의 용어들을 보다 효율적으로 조직할 수 있는 분류 체계를 구축할 수 있을 것이다. 궁극적으로는 연역적 방법으로 개발된 분류 체계를 활용해서 귀납적으로 즉 고전 용어 시소러스의 용어들을 포괄적인 상위 개념의 용어로부터 구체적인 하위 개념의 용어로 계층적으로 범주화하여 검색을 지원하는 분류 체계로 발전시켜야 할 것이다.

<표 1> 고전 용어를 분류 체계에 적용한 결과

패킷명	범주명	범주명	범주명 및 우선어	우선어	공통 범주
(자연물)	인물	관인	문신	강희안	조선전기 경상도 > 진주
	자연물	식물	토란	-	
(인조물)	문헌	서명	퇴계집	-	조선후기
	유적	건물	기타건물	숭례문	조선시대 서울
	물품	생활용품	공예품	설도전	당 중국
	법제	관청	교서관	-	조선시대
(시간)	시간	국명	조선	-	
(공간)	지명	자연지명	섬진강	-	전라도 > 진안 경상도 > 하동
(정신)	사상	술어	고사성어	설순빙리	남북조시대 중국
(행동)	사건	기타사건	분서갱유	-	진 중국

그리고 분류 체계를 구축하는 최종적인 목적은 시소러스의 용어들을 검색어로 선택하여, 해당 검색어와 관련된 고전 문집의 내용들에 접근하여 이용하기 위한 것이다. 따라서 시소러스의 용어들이 고전 분야의 주제 지식을 구조화한 분류 체계와 일치하는 지를 확인하는 과정을 거치게 되면, 고전 문집의 내용에 대한 검색을 보다 효율적으로 지원할 수 있는 분류 체계의 기능을 제대로 수행하게 될 것이다.

5. 결론

패킷 기법을 고전 용어 분류 체계에 적용한 것은 기존의 열거형 분류 체계에 비해 분류 체계에 유연성과 특정성을 제고하는 장점이 있기 때문이다. 따라서 고전 용어 분류 체계의 최상위 범주는 하위 범주들의 개념적 속성을 구분하기 위한 것으로 자연물, 인조물, 시간, 공간, 정신, 행동 등의 6개의 패킷을 설정하였고, 각 패킷에 속하는 상위 범주 10개, 기초수준 범주 54개, 최하위 범주 100 개로 고전 용어를 범주화하였다.

그리고 주요 범주 외에 시대 구분과 지역 구분 범주를 공통 구분 범주로 범주화하고, 주요 범주를 시대와 지역에 의해 세분할 필요가 있을 때 공통 구분 범주를 적용한다. 공통 구분 범주를 주요 분류 체계에서 분리시킨 이유는 다음과 같다. 각각의 범주를 시대와 지역으로 구분하는 것은 모든 범주에 공통적으로 필요하므로, 각각의 범주에서 시대 구분과 지역 구분을 중복해서 세분하여 전개할 필요가 없기 때문이다.

따라서 각 범주에서 시대와 지역으로 구분할 필요가 있을 때에는 공통 구분의 범주를 사용하면 범주의 중복을 최소화 할 수 있다. 그리고 분류표의 전체적인 크기를 줄이는 효과와 더불어 분류표를 보다 효율적으로 유지할 수 있는 방법이라고 할 수 있다. 이러한 공통 구분을 별도로 범주화하는 것의 타당성은 전통적인 문헌 분류체계에서 그 유용성이 이미 검증되었다는 점에서 그 근거를 찾을 수 있다.

특히 기존의 인문학 분야에서 개발된 한국어 사용자 시소러스와 비교해 볼 때 다음과 같은 면에서 장점을 가지고 있다고 할 수 있다. 첫째, 패킷 기법을 적용하여 동일한 특성을 갖는 개념과 용어들을 일관되게 범주화할 수 있었다. 둘째, 이러한 범주화를 통해 범주 설정의 일관성을 유지하여 분류 구조의 복잡성을 줄였다. 셋째, 한국어사용어 시소러스와 달리 동일 주제의 용어들의 분산을 방지했다. 넷째, 시대나 지역을 구분하기 위한 공통구분을 별도로 설정하여 분류표의 크기를 줄이고 보다 효율적으로 분류표를 유지 및 관리할 수 있게 되었다. 다섯째, 위의 장점들로 인해 이용자들이 분류 체계를 이용해서 보다 편리하게 검색할 수 있다.

그러나 고전 용어 분류 체계가 시소러스를 보완하는 분류 체계로 발전하고 이용자들에게 보다 나은 검색을 제공하기 위해서는, 구축된 분류 체계를 사용해서 디스크립터들을 분류함으로써 보다 나은 구조를 가지는 분류 체계로 개선할 필요가 있다. 그리고 궁극적으로는 고전 용어 시소러스에 포함되어 있는 모든 용어를 계층 관계에 의해서 구조화함으로써 기존의 인문학 분야 시소러스들이 가지고 있던 단점들을 보완하고 검색의 효율성을 극대화 할 수 있

을 것이다. 이러한 분류 구조의 발전으로 말미암아 고전 용어 시소러스 본래의 기능과 역할

을 다 할 수 있을 것이다.

참 고 문 헌

- 국사편찬위원회. 2004. 『한국역사용어 시소러스 자료집』. 국사편찬위원회.
- 국사편찬위원회. 2006. 『한국역사용어 시소러스』. 국사편찬위원회. [cited 2006.2.14].
 <<http://thesaurus.history.go.kr/>>
- 국회도서관. 2004. 『정보검색을 위한 표준 시소러스』. 서울: 국회도서관
- 김성희. 2002. 인터넷상의 디지털문화컨텐츠 분류 방안에 관한 연구. 『한국문헌정보학회지』, 36(3): 181-200.
- 김재훈. 2005. 고전용어 시소러스 개발 사업 개요. 『민족문화추진회보』, 79. 서울: 민족문화추진회.
- 김정현, 문지현. 2002. 농학분야 인터넷 자료의 분류 체계에 관한 연구. 『한국도서관정보학회지』, 33(3): 393-413.
- 김태수. 2000. 『분류의 이해』. 서울: 문헌정보처리연구회.
- 우응순, 권순희, 한창훈. 2002. 고시조 시어 시소러스 개발. 『민족문화연구』, 37. 고려대 민족문화연구원.
- 유영준. 2004. 패킷 분석을 이용한 웹 자료의 조직. 『한국비블리아』, 15(1): 23-41.
- 이란주, 성기주, 양정하. 2001. 여성학분야 인터넷 자료의 분류에 관한 연구. 『한국도서관정보학회지』, 32(3): 397-417.
- 이창수. 2000. 정보통신기술분야 인터넷 자료의 분류 체계에 관한 연구. 『한국도서관정보학회지』, 31(4): 111-138.
- 정영미. 1997. 『지식구조론』. 서울: 한국도서관협회.
- 정영미. 2005. 『정보검색연구』. 서울: 구미무역.
- 최재황. 1998. 인터넷 학술정보 자료의 디렉토리 서비스 설계에 있어서 DDC 분류 체계의 활용에 관한 연구. 『정보관리학회지』, 15(2): 47-67.
- 최현배. 1941. 『우리말본』. 서울: 정음문화사.
- 최희운. 1998. 인터넷 정보서비스의 분류 체계에 대한 비교 연구: 물리학을 중심으로. 『정보관리학회지』, 15(3): 45-71.
- 한국정신문화연구원. 1999. 『한국인물대사전』. 서울: 중앙M&B.
 <<http://www.nanet.go.kr/>>
- Art & Architecture Thesaurus. 1994. 2nd ed. New York: Oxford University Press.
- Molholt, Pat and Toni Petersen. 1993. "The role of the 'Art and Architecture Thesaurus' in communicating about visual art," *Knowledge Organization*, 20(1):30-34.

〈부록〉 분류표

위에 제시한 분류 체계에서 범주명에 패시를 적용하여 배열한 분류표는 다음과 같다. 괄호 안에 있는 용어가 패시를 표시하는 것이고, 범주간의 계층 수준은 들여쓰기와 온점의 개수로 표현하였다. 즉 온점의 개수가 많을수록 분류 체계에서의 계층 수준은 낮아진다는 의미이다.

(자연물)	... 종족
. 인물	... 조직(당파, 결사)
.. 왕실	... 기타집단
... 국왕(황제, 추존왕)	.. 기타인물
... 제후	... 성현
... 왕비(황후)	... 은자
... 후궁	... 도사
... 왕자	... 공신
... 공주	... 양인
... 종친(군)	... 귀화인
... 외척(부원군)	... 외국인
.. 관인	... 도적
... 문신	... 역적
... 무신	... 효자
... 경대부	... 열녀
... 아전(이서)	... 노비
... 환관	... 기타인물
... 역관	
... 궁녀(의녀)	. 자연물
.. 문사	.. 동물
... 문인	.. 식물
... 학자	.. 광물
... 유학	.. 천체
.. 예술인	.. 신체
... 서화가	.. 기타자연물
... 음악인	
... 광대	(인조물)
.. 종교인	. 문헌
... 승려	.. 서명
... 무격	.. 편명
... 천주교인	.. 작품명
... 동학교인	.. 지도
.. 기술인	.. 서화
... 지관	.. 악보
... 장인	.. 판본
... 상인	.. 기타문헌
... 의원	
... 기생	. 유적
... 백정	.. 건물
.. 집단	... 궁궐
... 성씨	... 사찰

<ul style="list-style-type: none"> ... 사제(私第)(재실 정사) ... 사당 ... 누정(정자, 누각) ... 서원·향교 ... 기타건물(관, 원, 사) .. 시설 <ul style="list-style-type: none"> ... 성·진·책 ... 창고 ... 능묘(제단, 비각) ... 기타시설(도요지, 저수지, 정원) . 물품 <ul style="list-style-type: none"> .. 의복 <ul style="list-style-type: none"> ... 관복 ... 예복 ... 상복 ... 장신구 ... 신발 ... 포복 .. 식약품 <ul style="list-style-type: none"> ... 식품 ... 음료 ... 식재료 ... 탕약 ... 약재 .. 생활용품 <ul style="list-style-type: none"> ... 그릇 ... 집기 ... 가구 ... 화폐 ... 공예품 .. 기타용품 <ul style="list-style-type: none"> ... 의료용품 ... 의례용품 ... 공구 ... 농기구 ... 약기 ... 무기 ... 형구 ... 활자 ... 선박 ... 기타물품 . 법제 <ul style="list-style-type: none"> .. 관청 .. 관직 .. 법률 .. 제도 .. 의례 	<ul style="list-style-type: none"> ... 국가의례 ... 향촌의례 ... 가례(관, 혼, 상, 제) ... 기타의례 .. 기타법제 (시간) <ul style="list-style-type: none"> . 시간 <ul style="list-style-type: none"> .. 국명(왕조) .. 연호 .. 간지 .. 절기 .. 월일시 .. 기타시간 (공간) <ul style="list-style-type: none"> . 지명 <ul style="list-style-type: none"> .. 자연지명 <ul style="list-style-type: none"> ... 땅(산, 산맥, 고개, 평야, 바위, 도서) ... 물(바다, 강, 하천, 호수, 저수지, 포구, 향만, 나루) .. 행정지명 .. 기타지명 (정신) <ul style="list-style-type: none"> . 사상 <ul style="list-style-type: none"> .. 이론 <ul style="list-style-type: none"> ... 학설 ... 교리(불교, 도교, 천주교, 천도교) .. 기법 <ul style="list-style-type: none"> ... 기술 ... 방법 .. 현상 <ul style="list-style-type: none"> ... 자연 현상 ... 신체 증상 .. 술어 <ul style="list-style-type: none"> ... 고사 성어 ... 관용어 ... 핵심어 .. 기타사상 (행동) <ul style="list-style-type: none"> . 사건 <ul style="list-style-type: none"> .. 전쟁 .. 사회 .. 역모(반란) .. 송사 .. 풍속 .. 기타사건
---	---