

# 분류체계 일치를 통한 과학기술정보 상호 교환 방법에 관한 기초 연구

## A Preliminary Study on Interchange of Science and Technology Information through Harmonization of Classification Schemes

홍성화\*, 서태설\*\*  
Sung-Wha Hong, Tae-Sul Seo

### 차 례

- |                  |                                |
|------------------|--------------------------------|
| 1. 서론            | 5. 분류일치를 통한 과학기술정보 상호<br>교환 방안 |
| 2. 관련 연구         | 6. 결론 및 향후연구                   |
| 3. 분류체계의 종류와 특징  | • 참고문헌                         |
| 4. 과학기술분야의 분류 현황 |                                |

### 초 록

과학기술정보의 의미적 상호운용성 문제는 빈번하게 발생한다. 잘 만들어진 분류체계는 상이한 데이터베이스 간에 의미상 불일치 없이 정보를 교환하기 위한 도구로 사용될 것이다. 하지만 각 데이터베이스가 취하고 있는 분류체계가 상이함으로 인해서 여전히 현실적인 장벽이 존재한다. 따라서 분류체계간의 일치 및 조화는 매우 시급한 문제이다. 본 논문의 목표는 다른 분류체계('국가과학기술표준분류'와 'KISTI 표준 분류')를 갖는 데이터베이스간의 정보 교환 시에 발생할 수 있는 의미적 불일치를 해결하는 것이다. 이를 위해서 과학기술의 개념적 체계 분석을 수행하였고 다섯 가지의 일치/불일치 유형을 사례에 기반하여 분석하였다.

### 키 워 드

과학기술, 분류 체계, 상호운용성, 정보 교환

\* 한국과학기술정보연구원 해외정보실 선임연구원  
(Senior Researcher, Overseas Information Dept. KISTI, shong@kisti.re.kr)  
\*\* 한국과학기술정보연구원 정보표준화실 실장  
(Manager, Information Standardization Dept. KISTI, tsseo@kisti.re.kr)  
• 논문접수일자 : 2004년 8월 20일  
• 게재확정일자 : 2004년 9월 15일

## ABSTRACT

The problem of semantic interoperability in science and technology information is frequently raised. Well-established classification scheme will be used as a tool to interchange information between different databases without semantic inconsistency. However, there is still a practical barrier due to different classification schemes each database adopts. Accordingly, it is urgent to harmonize or reconcile those classifications with each other. This paper aims to solve semantic inconsistencies occurred when interchanging information between databases having different classification schemes, the Standard National Sci-Tech Classification and the Standard KISTI Classification. For the purpose a conceptual analysis of science and technology are performed and five consistency/inconsistency types are analyzed based on some examples.

## KEYWORDS

Science and Technology Information, Classification Schemes, Interoperability, Information Interchange

### 1. 서론

과학기술정보의 개발, 관리, 서비스 등에서 과학기술분류에 대한 필요성이 자주 제기된다. 따라서 과학기술 분야의 연구관리 기관, 자원 관리 기관, 정보서비스 기관들은 나름대로의 분류체계를 만들어 적용하고 있다. 국가적으로도 여러 가지 목적으로 과학기술분류체계를 제정하여 사용하고 있는데, 여기에는 KISTEP의 '국가과학기술표준분류', 학술진흥재단의 분류, 과학재단의 분류 등이 있다. 이렇게 여러 가지 분류가 존재하다보니 자연스럽게 이러한 유사분류들을 어떻게 조화시킬 수 있을 것인가의 문제가 자연스럽게 제기되고 있다.

한편 과학기술은 시간이 지남에 따라 계속해서 변화하고 있기 때문에 새로운 분야의 발생, 용어 개념의 변화 및 세분화를 분류체계에 정기적으로 반영해 주어야 한다. 또한 똑같은 분류표라 할지라도 분류체계를 사용하는 사람이 누구냐에 따라서 그 체계가 적절하거나 부적절하게 느끼는 경우도 발생할 수 있다. 즉, 분류의 관점과 레벨의 차이로 인해서 똑같은 영역(domain)에 대한 분류체계가 다르게 구성될 수도 있다.

이와 같은 문제를 해결하고, 정보개발업무를 보다 체계적으로 수행하기 위해서는 이러한 분류들은 일관된 체계로 재편성하거나 새롭게 개발할 필요가 있다. 이것은 과학기술계 뿐만

아니라 정보서비스 분야의 전반적인 요구이기도 하다. 하지만 이러한 필요가 발생할 때마다 새로운 분류체계를 만든다면 그것은 또 하나의 분류체계가 추가되는 결과를 초래하여 문제를 더 어렵게 만들 것이다.

현재 나와 있는 대부분의 과학기술분류는 Tree 구조를 띠고 있으나 계층형과 비계층형이 혼재되어 있고, 분류의 깊이와 폭이 균형적이지 않다. 이는 과학기술이라는 범주가 너무 넓어서 여러 분야의 사람이 개입해야 하기 때문에 모든 사람의 의견을 일치시키기가 어렵기 때문이다.

경우에 따라서는 순수 학문적 분류가 되어야 할 과학기술분류가 이해관계에 얽매어서 구조가 왜곡되기도 한다. 또한 특허분류나 특정 기관의 분류는 현재 활발한 정보 발생이 있는 분야로 치우쳐서 균형적이지 못한 측면이 있다.

결과적으로 과학기술분류 상에 나타난 왜곡과 불균형을 어디에선가 분석하고 해결방안을 모색하는 것이 필요하다. 이것은 앞으로 다가오는 지식사회와 차세대 웹서비스를 대비해서도 반드시 해결해야 할 문제인 것이다.

따라서 본 논문에서는 다양한 과학기술분류체계를 아우를 수 있는 기본 개념체계를 만들고 이를 통해서 분류체계간의 상호 운용성을 체크할 수 있는 방법론을 개발하였다. 그리고 국가적인 분류 표준인 '국가과학기술표준분류'와 하나의 기관 표준인 'KISTI 표준 분류'를 사례로 상호 일치성 분석을 수행하였다.

## 2. 관련 연구

인터넷을 통한 정보서비스가 증가함에 따라서 정보서비스를 위한 분류체계에 대한 연구가 관심을 모으고 있다.

먼저 김정현 등은 분류업무 중에서도 다른 주제분야에 비해 독특하고 복잡한 형식요소로 구분되고 있는 DDC 20판의 문학류(800)를 대상으로 구체적인 분류작업 과정을 분석하여 시스템화함으로써 분류전문가 시스템 개발을 위한 기초 자료를 제시하였다.

김성희는 5개의 인터넷 검색엔진을 선정하여 인터넷상의 문화컨텐츠 분류 현황을 분류체계 조건, 주제의 논리성을 기준으로 분석한 후 새로운 분류 방안을 마련함으로써 정보 및 문화컨텐츠의 효율적인 검색 및 관리를 위한 기초 자료를 제공하였다.

문지현 등은 KDC, 야후! 코리아, 엠파스, 농업정보디렉토리(한국농림수산정보센터), Agri\_Directory(농민넷) 등에서의 농학분야의 분류체계를 비교분석하여 농학분야의 인터넷 자원을 효율적으로 관리하기 위한 새로운 분류체계 모형을 제시하였다

김정현 등은 KDC, DDC, UDC, NDC 등의 문헌분류법에서 농학분야 주제를 전개하고 있는 강목표에 대한 비교분석 및 NAL의 AGRICOLA SCC 분석을 통하여 농학분야의 정보자료를 효율적으로 관리하기 위한 새로운 문헌분류표 모형을 작성하였다.

지금까지의 분류체계에 대한 연구는 특정분

야의 분류를 통합하여 새로운 분류를 만드는 것과 분류전문가시스템 개발 등이 주류를 이룬다고 볼 수 있다. 하지만 서로 다른 분류체계간의 정보호환에 관한 연구는 찾아보기 힘들다.

### 3. 분류체계의 종류와 특징

#### 3.1 분류체계의 개요

분류체계라(classification scheme)는 것은 “유사 특성(형식, 주제, 품목, 관계 등)을 갖는 개체들을 그룹으로 배열(clustering) 또는 구분(distinguishing)하기 위한 체계”라고 할 수 있다. 이러한 분류의 기능과 용도는 지식 표현, 정보 관리, 정보 검색, 통계 분석, 식별 등이 있을 수 있다<sup>1)</sup>.

분류는 기본적으로 다음과 같은 요건들을 만족시킬 때라야 좋은 분류라고 할 수 있다.

- ① 분류 항목의 유일성: 분류 항목의 명칭과 의미가 유일하여야 한다.
- ② 분류 체계의 일관성: 계층, 항목 정의 방법 등에서 일관된 원칙과 기준을 가져야 한다.
- ③ 분류 체계의 포괄성: 분류하고자하는 대상이 모두 분류될 수 있어야 한다.
- ④ 분류 체계의 연속성: 기존의 분류 체계를 무리없이 계승할 수 있어야 한다.

⑤ 미래 환경변화 수용성: 새로운 기술의 등장 및 환경의 변화에 적응 및 확장이 가능한 체계를 가져야 한다.

⑥ 타분야 호환성: 유사 분야의 분류 체계와 조화를 이룰 수 있어야 한다.

또한 분류는 구조적인 면에서 아래와 같이 몇 가지 형태로 구분된다<sup>2)</sup>.

- ① List 형: 무차원 list(ordered/unordered), 차원 list
  - ② Array 형: table(unordered array), matrix(ordered array)
  - ③ Tree 형: Single root(taxonomy), Multiple roots
  - ④ Network 형: Directed graph, Map 등
- 이중에서 과학기술분류에 가장 많이 사용되는 구조는 Tree형이다. Tree형에는 포함관계(is a)를 기반으로 하여 상속이 이루어지는 계층(hierarchy)형과 포함관계와 상속이 없고 하나의 차원에 따라 지식을 표현하는 비계층(non hierarchy)형이 있다.

#### 3.2 계층형 Tree 구조의 특성<sup>3)</sup>

계층구조의 분류체계는 기본적으로 다음과 같은 특성을 갖는다.

- ① 포함관계: 상위 클래스(class)는 하위 클래스

1) 서태설, 2003. 분류통합에 의한 과학기술분류의 개발. 『KOSTI Workshop 2003』.  
 2) ISO/IEC 11179 part 2 Information Technology: Metadata Registry -- Classification 관련 자료를 저자가 분석정리한 것임.  
 3) Barbara H. Kwasnik, 1999. "The Role of Classification in Knowledge Representation and Discovery". Library Trends, 48(1): 24-30.

- 스(subclass and sub subclass)를 포함한다.
- ② 클래스간의 관계: 계층구조는 상위 클래스와 하위 클래스의 사이에 is a 관계만으로 표현된다. 즉, 하위 클래스는 상위 클래스의 한 종(species)이 되는 관계를 갖는다.
- ③ 상속: 상위 클래스에 대해서 참인 것은 그 클래스에 속하는 모든 하위 클래스에게도 참인 관계가 성립된다.
- ④ 공통속성: 하나의 클래스에 속한 모든 엔티티(entities)는 동일한 속성으로 그룹화되며, 동일한 속성으로 형제 클래스와 구분된다.
- ⑤ 상호배타성: 하나의 엔티티는 하나의 클래스에만 속할 수 있다.
- ⑥ 필요충분조건: 어떤 클래스에 속하기 위해서는 하나의 엔티티가 필요 속성과 충분조건을 만족해야 한다.
- 이와 같은 기본 특성은 다음과 같은 장단점을 갖게 하므로 이러한 장단점을 잘 활용해서 분류체계에 적용하여야 좋은 분류체계를 만들 수 있다.
- ① 그룹핑 및 구별 규칙이 선형적으로 만들어지기 때문에 계층분류는 포괄적인 분류가 된다. 즉, 구조가 결정되기 전에 설계자는 엔티티의 범위, 그들의 속성, 유사도 판단기준에 대해서 많은 부분을 알고 있어야 한다.
- ② 계층분류가 가지는 상속 특성은 분류의 표시를 간편하게 만들어준다.
- ③ 결과적으로 불완전한 증거로부터 추론을 가능하게 한다.
- ④ 계층분류는 실제적인 정의를 가능하게 한다.
- ⑤ 계층분류는 전체적인 조망을 가능하게 한다.
- ⑥ 계층분류는 하나의 관점만을 허용하기 때문에 다양한 관점이 적용될 경우에 중복이 발생하게 된다.
- ⑦ 복합적이고 다양한 기준이 적용될 경우 체계가 매우 복잡해지며 상호배타성의 원리를 준용하기가 매우 어렵다.
- ⑧ 완전하고 포괄적인 지식이 없는 경우는 계층분류를 정확히 적용할 수 없는 경우가 발생한다.
- ⑨ 척도가 다른 경우는 상속이 어렵다.
- ⑩ 애매모호성이 있는 경우는 필요충분조건을 만족시키기 어렵다.

### 3.3 비계층형 Tree 구조의 특성<sup>4)</sup>

비계층구조형 분류체계는 외형적인 구조는 계층구조형과 유사하나 그 특징은 매우 다르다.

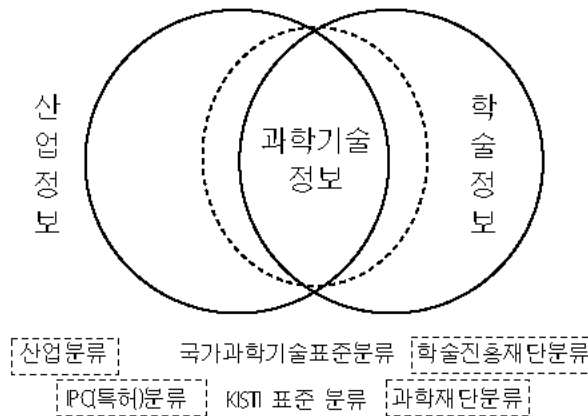
- ① 계층구조와 마찬가지로 그룹핑 및 구별 규칙이 선형적으로 만들어지는 포괄적인 분류가 된다. 따라서 이경우도 무엇이 엔티티를 구성할지 결정되어 있어야 한다.

4) Ibid., pp.30-35.

- ② 구조는 엔티티간의 관계에 의해서 결정된다.
- ③ 분류항목의 순서를 합리적으로 하여야 한다.
- ④ 엔티티간의 관련성의 깊이를 알 수 있다.
- ⑤ 엔티티의 범주에 균형이 있어야 한다. 너무 넓거나 너무 좁아서는 안 된다.
- ⑥ 구조가 경직되어 분류체계의 수정이 어렵다.
- ⑦ 정보의 흐름이 한 방향으로만 이루어진다.

#### 4. 과학기술분야의 분류 현황

과학기술정보는 산업정보와 학술정보의 중간적 위치에 있다고 볼 수 있다. 산업정보의 분류에 사용되는 분류로는 '산업분류'와 '특허분류'가 대표적이며, 학술정보의 분류로는 학술진흥재단 분류와 과학재단 분류가 대표적이다. 과학기술 자체를 대상으로 하는 분류로는 '국가과학기술표준분류'와 'KISTI 표준 분류'가 있다. 여기서는 과학기술분류에 해당하는 두 분류에 대해서만 살펴보고 비교하도록 한다.



〈그림 1〉 과학기술정보와 분류

##### 4. 1 국가과학기술표준분류

‘국가과학기술표준분류’는 과학기술기본법 제27조에 따라 과학기술 관련 정보, 인력, 연구

개발사업 등의 효율적 관리를 위해서 2002년 12월에 과학기술부 장관 명의로 제정되었다<sup>5)</sup>.

이 분류는 공포된지 얼마 안 되어서 아직 사용 사례가 없지만, 향후 과학기술부의 의지여하에 따라서는 널리 활용될 가능성이 높다. 이

5) 정근하, 최문정, 고대승. 2003. 국가 연구개발 사업의 효율적 기획관리를 위한 과학기술 표준분류 체계에 관한 연구. 『기술혁신학회지』, 6(2): 265-277.

분류는 3단계로 되어 있으며, 대분류 19개, 중분류 160개, 소분류 1,023개로 구성된다. <표 1>은 국가과학기술표준분류의 개요를 나타낸다.

#### 4.2 KISTI 표준 분류

한국과학기술정보연구원(KISTI)은 2002년에 30여년 동안 과학기술문헌정보를 분류하여

왔던 분류를 정비하여 'KISTI 표준 분류'를 자체적으로 제정하여 사용하고 있다. 이 분류는 일본의 JICST 분류를 우리나라의 실정에 맞게 재편성하여 만들어진 것으로 과학기술 논문정보의 분류에 적합하도록 되어 있다. 이 분류도 3단계로 되어 있으며, 대분류 37개, 중분류 217개, 소분류 904개로 구성되어 있다. <표 2>는 KISTI 표준 분류의 개요를 나타낸다.

<표 1> 국가과학기술표준분류의 개요

분류단위	코드체계	비고
대분류	(영문 1자리)-총 19개 A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S	
중분류	(숫자 1자리) 1~0	- 0은 '기타'
소분류	(숫자 1자리) 1~0	- 0은 '기타'
분류사례	(대분류) J : 정보 (중분류) J2 : 시스템 소프트웨어 (소분류) J22 : 데이터베이스 관리시스템	

<표 2> KISTI 표준 분류 개요

분류단위	코드체계	비고
대분류	(영문 2자리)-총 37개 AA, AB, AC, AD, AE, AF, BA, BB, BC, BD, BE, BF, BG, BH, BL, BM, CA, CB, CC, CD, CE, EC, EJ, ET, LA, MA, MB, MC, MD, NA, NB, NC, ND, PA, PB, RA, RB, RC	- 첫 자리에 의미부여
중분류	(숫자) 2자리 01~99	
소분류	(숫자) 2자리 01~99	
분류사례	(대분류) ND : 정보가공 (중분류) ND09 : 데이터베이스 및 DBMS 일반 (소분류) ND0903 : DBMS	

### 4.3 비교

국가과학기술표준분류와 KISTI 표준 분류는 과학기술이라는 동일한 주제분야를 다루고

있지만 그 용도와 관점의 차이로 인해서 내용 상으로는 상이한 모양새를 하고 있다. <표 3>에 이 두 분류를 비교하였다.

<표 3> 국가과학기술표준분류와 KISTI 표준 분류의 비교

구분	KISTI 표준 분류	국가과학기술표준분류
주제분야	과학기술	과학기술
주요용도	정보개발 및 서비스	국가예산지원 연구관리
사용범위	KISTI 내부	대한민국 국가단위
분류체계 코드구조	대(영2)-중(숫2)-소(숫2)	대(영1)-중(숫1)-소(숫1)
확장성	확장가능(자릿수가 여유있음)	확장어려움(자릿수가 불충분함)
역사	30년 이상	2년
갱신주기	필요시	3년

## 5. 분류일치를 통한 과학기술정보 상호교환 방안

분류체계 구축은 개념화, 언어화, 기호화의 3단계를 거쳐서 개발된다. 여기서 가장 중요한 것은 개념화이다. 서로 다른 두 분류체계가 개념적으로 같은 체계를 가지고 있다면, 언어화와 기호화 단계에서 차이가 있더라도 통합을 하는 것이 그리 어렵지는 않을 것이다. 따라서 본 논문에서는 과학기술 분야의 개념적 체계를 만들어으로써 모든 분류체계의 개념적 상호운영의 기반을 만들고자 한다.

### 5.1 과학기술 개념체계 분석

과학기술 개념체계 분석을 위해서는 top down과 bottom up 방법이 있을 수 있지만 여기서는 후자의 방법을 사용하도록 한다. 즉, 국가과학기술표준분류와 KISTI 표준 분류의 중분류 이상을 가지고 유사분류의 그룹핑을 하였다. 그 결과 <그림 2>와 같이 4개의 계층과 22개의 그룹으로 분류할 수 있었다.

먼저 계층으로는 정책(policy), 소재 기술(materials technology), 제품 기술(product technology), 순수과학(science)의 4계층으로 나눌 수 있다. 먼저 순수과학 계층에는 수학, 물리학, 화학, 생명과학, 지구과학 등 5개의 대표적 분류가 존재한다. 이것은 학술진흥재단과 과학재단의 대분류와도 일치하는 것이다. 이러한 5대 순수과학에 기초한 응용분야로서 제품



Policy	과학 기술정책								
Materialc Technology	화공재료기술 금속재료기술				약품기술 생화학재료기술			건설재료기술 에너지기술	
	정 보 기 술	전 산 학	전 기 제 어 통 신 기 술	기 계 기 술	화 학 화 공	농 림 수 산 기 술	생 물 공 학	건 설 기 술	환 경 자 원 기 술
Production Technology									
Science	수학		물리학		화학		생명과학		지구과학

〈그림 2〉 과학기술체계

기술이 존재한다. 이 계층에는 정보기술, 전산학, 전기제어통신기술, 기계기술, 화학화학, 농업수산기술, 생물공학, 건설기술, 환경자원기술 등의 9개 분야가 포함된다. 다음으로 모든 응용기술의 근간이 되는 소재 기술 계층에는 화공재료기술, 금속재료기술, 약품기술, 생화학재료기술, 식품기술, 건설재료기술, 에너지기술 등의 7개 분야 존재한다. 마지막으로 정책 계층에는 과학기술정책 1개가 존재한다.

### 5.2 분류의 상호운용방안

앞서 얻어진 과학기술체계는 국가과학기술 표준분류와 KISTI 표준 분류의 중분류를 기반으로 한 것이므로 중분류 이상에서는 두 분류가 이와 같은 체계 속에서 통합되는 데는 근간이 될 수 있다. 따라서 본 연구에서는 이 과학기술체계에서 나타난 22개 그룹별로 양 분류체

계를 그룹핑하고 각 그룹 내에서 상호운용 가능성을 타진하도록 하였다. 이렇게 하면 서로 상이한 분류체계를 가지는 기존에 구축되어 있는 정보들을 최소한의 수정으로 정보를 상호 호환할 수 있게 될 것이다.

또한 이렇게 함으로써 새로운 기술이 발생할 경우 과학기술체계의 관점에서 분석을 하여 분류를 수정하게 되기 때문에 분류의 일관성이 흐트러지지 않으면서 새롭게 확장할 수 있다는 장점이 있다. 뿐만 아니라 분류의 요건인 유일성, 포괄성의 면에서도 장점을 가질 수 있을 것이다.

다음으로 서로 다른 체계를 갖는 분류체계를 상호운용하기 위해서는 각 분류항목의 의미적인 일치여부를 확인해 보아야만 한다. 그러나 의미적으로 완전히 일치하지 않는 경우가 발생하는 데 그러한 경우는 상황에 따라 해결이 가능한 경우가 있고 불가능한 경우가 있다.

〈표 4〉 의미적인 일치/불일치 유형

구분	설명과 해결 방법	예시
① 완전 일치 (A1 ≡ B1)	분류 항목명이 의미상으로 서로 일치하는 경우로서 <b>기계적으로 일대일(1:1) 매칭</b> 을 하면 된다.	
② 포함 관계 (A1+A3 - B1)	하나 이상의 다른 분류 항목과 함께 하나의 분류와 일치되는 경우로서 이것도 <b>기계적으로 자동 다대일(n:1) 매칭</b> 이 가능하다.	
③ 분할 관계 (A2-B1+B2+B5)	하나의 분류 항목이 두 개 이상의 분류와 일치하는 경우로서 <b>상위 분류항목에서 매칭이 가능</b> 한 경우가 있을 수 있으나, 완전한 일치를 위해서는 인간이 개입을 하거나 인공지능을 활용하여야 한다.	
④ 중첩 관계 (A1 ∩ B1)	분류 항목간에 부분적으로만 일치하는 경우로서 이 경우도 <b>상위 분류항목에서 매칭이 가능</b> 한 경우가 있을 수 있으나, 완전한 일치를 위해서는 인간이 개입을 하거나 인공지능을 활용하여야 한다.	
⑤ 완전 불일치 (A1 ≠ B1)	새로 발생된 분야이거나 타 분야의 유입에 의해서 하나의 분류항목과도 관련이 없는 것으로서, 완전한 일치를 위해서는 <b>인간이 개입을 하거나 인공지능을 활용</b> 하여야 한다.	

〈표 4〉는 두 분류 체계를 상호운용할 때 발생할 수 있는 5가지 유형과 대처방법을 정리한 것이다.

### 5.3 분류의 상호 일치/불일치 분석 사례

여기서는 ‘화학’ 분야에서 ‘KISTI 표준분류’와 ‘국가과학기술표준분류’ 간의 의미상 일

치여부를 확인하고 검증을 하도록 하였다. 분류 실험은 실제 'KISTI 표준분류' 수년간 사용해 본 경험이 있는 화학분야의 전문가를 통해서 이루어졌다.

- ① 완전 일치: 이름의 표기는 차이가 있지만 의미상으로 동일한 분류항목을 8개 찾을 수 있었다(〈표 5〉 참조).
- ② 포함 관계: KISTI 분류에서는 세분화되어

있는 것이 국가과학기술표준분류에서는 하나의 분류항목으로 그룹화된 2개의 사례가 발견되었다(〈표 6〉 참조).

- ③ 분할 관계: KISTI 분류에서는 하나의 분류항목으로 되어 있는 것이 국가과학기술표준분류에서는 세분화된 3개의 사례가 발견되었다(〈표 7〉 참조).
- ④ 중첩 관계: 관점의 차이에 의한 것으로 의미

〈표 5〉 완전 일치 분류 항목

KISTI 표준 분류		국가과학기술표준분류	
CA0201	열역학, 열화학	C11	열역학, 통계열역학
CA0205	분광화학	C13	분광학
CA0209	계면화학, 콜로이드화학	C15	표면, 계면화학
CA0210	결정화학	C36	결정학
CA0301	분리분석	C43	분리분석화학
CA0304	전기분석	C42	전기분석화학
CA0407	유기금속, 유기半金屬化合物	C26	유기금속화학
CA0408	천연유기화합물	C21	천연물화학

〈표 6〉 포함 관계 분류 항목

KISTI 표준 분류		국가과학기술표준분류	
CA0204	전기화학	C19	전기물리화학
CA0208	전기, 자기현상		
CA0402	지방족화합물	C22	유기합성 · 전합성
CA0403	지방족고리화합물		
CA0404	방향족고리화합물		
CA0405	탄소축합여러고리화합물		
CA0406	이종원자고리화합물		

〈표 7〉 분할 관계 분류 항목

KISTI 표준 분류		국가과학기술표준분류	
CA0203	촉매, 반응속도론	C37	반응속도론·반응메카니즘
		C38	촉매화학
CA0211	무기화합물	C31	이론무기화학
		C32	전이금속화학·착물화학
		C33	유기금속화학
		C34	생무기화학
		C35	고체무기화학
CA0401	물리유기화학	C23	유기합성방법론
		C24	이론유기화학

상으로 부분적으로만 일치하는 분류 항목 사례가 존재하였다. 〈표 8〉에서 KISTI 분류는 분석의 특성이 무엇이나에 따라 물리분석과 화학분석으로 나눈 반면, 국가표준분류에서

는 분석의 방법과 대상이 무엇이나에 따라 정량, 분광, 표면, 구조, 질량분석 등으로 세분화 하였다.

〈표 8〉 중첩 관계 분류 항목

KISTI 표준 분류		국가과학기술표준분류	
CA0302	물리분석	C41	정량분석화학
		C42	전기분석화학
		C43	분리분석화학
CA0303	화학분석	C44	분광분석화학
		C45	표면분석화학
		C46	구조분석화학
		C47	질량분석화학

⑤ 완전 불일치: 분류 범위의 차이 및 신학문에 따른 이질적이 분류 항목들이 존재하였다. 〈표 9〉는 그 중의 일부를 나타낸다. KISTI 분류의 CA0202는 국가표준분류에

서 다루지 않는 관점이라서 분류 항목에 반영되지 않았다. 또한 국가표준의 C25, C27, C28과 같은 항목은 국가표준분류가 가장 최근에 만들어졌기 때문에 최신 기술을 분

〈표 9〉 완전 불일치 분류 항목

KISTI 표준 분류		국가과학기술표준분류	
CA0202	상평형, 화학평형, 용액론	C25	초분자화학
		C27	생유기화학
		C28	의약화학

〈표 10〉 분류 일치/불일치성 실험 결과

구분	완전 일치	포함 관계	분할 관계	중첩 관계	완전 불일치
합리성	60% 일치	100% 일치	40% 일치	90% 대응	100% 불일치

류에 반영하였으나, KISTI 분류는 최신화가 되지 않아 이러한 분야를 분류 항목에 반영하지 못한 것이다.

이와 같은 유형의 실제 분석을 위해서 이미 KISTI 표준 분류가 부여된 KISTI의 과학기술 문헌정보(BIST)를 대상으로 간단한 실험을 하였다. 5개 유형마다 각각 10씩 50건의 샘플 데이터를 대상으로 국가과학기술표준분류를 부여한 결과 〈표 10〉과 같은 결과를 나타내었다. 완전일치와 분할 관계의 경우는 60%와 40%의 비교적 낮은 일치도를 나타내었다. 이것은 KISTI 표준 분류의 부여가 잘못된 부분이 그대로 포함되었기 때문인 것으로 사료되기도 하지만, 분류항목의 완전 일치라는 것은 존재하기 어려우며, 의미상으로 두 항목이 대체적인 일치를 보이면 완전 일치로 보기 때문에 오는 오차에 기인한다고 할 수 있다. 분할 관계도 같은 방식으로 설명될 수 있을 것이다.

나머지 포함 관계, 중첩 관계, 완전 불일치의

경우는 100%에 가까운 합리성을 보였다. 포함 관계도 이 경우는 100%를 보였지만 다른 실험으로 하면 이보다 낮은 결과를 보일 가능성이 있다. 그것은 앞서 설명한 완전 일치와 포함 관계의 경우와 마찬가지로이다. 중첩관계의 경우는 실험 결과 KISTI 분류에서 CA0302(물리분석)로 분류된 것이 국가표준에서는 C44(분광분석화학)로 분류되어 상위분류 즉 KISTI 분류의 CA03(분석화학)과 국가표준의 C4(분석화학) 간에 상호 호환이 가능할 것으로 보였다. 마지막으로 완전 불일치의 경우는 실험 데이터가 대상으로 하는 상대방의 분류와 일치하는 것이 하나도 없는 것으로 나타났다. 이 경우도 완전한 불일치라기보다는 대체적인 불일치라고 보아야 할 것이다.

## 6. 결론 및 향후연구

본 연구에서는 Tree 구조를 갖는 분류의 기

본 이론을 살펴보고, '국가과학기술표준분류'와 'KISTI 표준 분류'를 일치시키기 위한 패러다임으로써 과학기술체계 분석을 시도하였다. 또한 분류를 상호 일치시킬 때 발생하는 전형적인 5개의 유형을 구분하였다. 이렇게 구분된 유형을 '화학' 분야를 대상으로 실제 사례를 들어 분석하였다.

분석결과를 보면 완전 일치와 포함 관계는 평균 80%의 정밀도로 자동일치가 가능한 것으로 나타났다. 분할 관계와 중첩 관계의 경우는 한 단계 위의 분류에 대해서 40%와 90%의 정밀도로 자동일치가 가능한 것으로 나타났다. 마지막으로 완전 불일치의 경우는 사람의 개입 또는 인공지능적인 방법을 통해서만이 완전한 일치를 할 수 있음을 실증하였다. 분류표상에 전개되지 않는 신기술 분야는 해당 항목의 전개가 어렵기 때문에 이에 대한 별도의 분류자리를 고려해야 한다.

결론적으로 본 연구는 서로 다른 분류 체계를 일치시키기 위한 5개의 전형적인 유형을 처음으로 분석하였다는데 의의를 둘 수 있다. 이러한 유형 분석을 근간으로 서로 다른 분류체계는 사람이나 고도의 인공지능을 활용하지 않으면 완벽하게 일치시킬 수 없음을 수치적으로 확인 하였을 뿐만 아니라, 분류체계가 명확하게 구성되어 있다면 자동으로 일치시키는 비율을 제고시킬 수 있을 것임도 확인하였다. 따라서 이 연구결과는 재분류작업이나 다른 분류와의 정보교환을 위한 합리적인 기초 자료가 될 수 있을 것이다.

본 연구는 서로 다른 분류체계를 일치시키기 위한 초기 연구로서, 향후 본 논문에서 제안된 방법은 더 많은 데이터를 가지고 실험하여 증명할 필요가 있다. 더 나아가서 이것을 기초로 과학기술분류의 이론적 연구를 발전시켜 분야별 온톨로지 구축 및 온톨로지 재사용 방법에 대한 추가적인 연구가 필요할 것으로 사료된다.

## 참고문헌

- KISTEP, 2002, 국가과학기술표준분류.
- KISTI, 2003, 과학기술정보 분류 표준.
- KISTI, 2002, KISTI 2002 I 01 KISTI 표준 분류.
- KINITE, 1999, 『과학기술정보 공유와 유통을 중심으로 한 국가혁신체제에 관한 연구』, [서울]: 전산원.
- 서태설, 2003, 분류통합에 의한 과학기술분류의 개발, 『KOSTI Workshop 2003』.
- 서태설, 2003, 과학기술정보 분류 표준 개발, 『제4차 과학기술정보표준화위원회 총회 및 세미나』, 2003년 11월 19일.
- 김성희, 2002, 인터넷상의 디지털 문화컨텐츠 분류방안에 관한 연구, 『한국문헌정보학회지』, 36(3): 181-200.
- 김정현, 이명규, 2003, 농학분야 문헌분류 체계에 관한 연구, 『한국도서관정보학회지』, 34(1): 239-260.
- 김정현, 문지현, 2002, 농학분야 인터넷자료의

- 분류 체계에 관한 연구. 『한국한국도서관정보학회지』, 33(3): 393-413.
- 서태설. 2002. 과학기술정보와 표준화. 『제2차 과학기술정보표준화위원회 총회 및 세미나』, 2002년 10월 10일.
- 설성수, 송충한. 2000. 『지식활동분류의 이론과 실제』, 대전: 한남대학교 출판부.
- 이희상, 유재영, 정의섭. 2003. 부품·소재 정보를 위한 분류 체계 설계. 『기술혁신학회지』, 6(1): 110-124.
- 정근하, 최문정, 고대승. 2003. 국가 연구개발 사업의 효율적 기획관리를 위한 과학기술 표준분류 체계에 관한 연구. 『기술혁신학회지』, 6(2): 265-277.
- 김명옥. 1991. 재분류의 이론과 실제. 『한국문헌정보학회지』, 20호: 127-161.
- 김정현. 1993. 분류전문가시스템 개발을 위한 분류 업무분석. 『도서관』, 48(6): 40-61.
- ISO/IEC 11179 part 2 Information Technology: Metadata Registry Classification.
- Barbara H. Kwasnik. 1999. "The Role of Classification in Knowledge Representation and Discovery". *Library Trends*, 48(1): 24-30.