통제어휘 표준의 변화 및 시사점에 대한 연구*

A Study on the Changes in Standards Related to Controlled Vocabulary and Their Implications

김 성 원(Sungwon Kim)**
김 정 우(Jeongwoo Kim)***

목 차

1. 서 론
1.1 연구의 목적 및 배경
1.2 연구의 방법 및 제한점
2. 이론적 배경
2.1 시소러스 효용 및 활용현황
2.2 선행연구 개관
3. 표준의 제·개정 현황
3.1 ISO 동향
3.2 BSI 동향
3.3 ANSI 동향
3.4 기타 표준
4. 표준 개정의 주요내용 및 시사점
4.1 적용 대상자료
4.2 통제어휘 도구의 유형
4.3 통제어휘 표현
4.4 다국어 수용
4.5 상호운용성 확보
5. 결 론

초 록

대표적인 통제어휘인 시소러스는 지난 50년간 정보의 색인과 검색을 위한 도구로 널리 활용되어 왔다. 이러한 흐름에 발 맞추어 다양한 주제분야에서의 시소러스 개발을 위한 지침을 제공하기 위해 국제표준 및 국가표준이 개발되었으며, 최근 들어 시소러스 관련 표준의 개정작업이 활발히 이루어지고 있다. 이러한 시소러스 관련 표준으로 ISO 25964와 BS 8723이 대표적인 사례이다. 본고에서는 이러한 표준의 개정 현황을 살펴보고 변화된 내용과 그 시사점을 도출하여 보았다. 이러한 고찰 결과를 기반으로 오늘날의 새로운 정보환경 아래에서 시소러스가 제공할 수 있는 새로운 기능과 그 기능을 수행하기 위해 필요한 요소들을 제시하였다.

ABSTRACT

Thesaurus, a well-known form of controlled vocabulary, has been widely used for indexing and searching of information during the last 50 years. There also have been developments of international and national standards to provide guidelines for developing thesaurus in diverse subject areas. In recent years, the revisions of thesaurus-related standards have been made. Among them are ISO 25964 and BS 8723. This article examines the current status of revision of these standards, and discusses its implications. Based on this examination, it suggests functional requirements of thesaurus in the present information environment, and also proposes elements needed for the development of these functions.

키워드: 통제어휘, 어휘통제, 시소러스, 표준, 표준화, 정보검색, 색인, 상호운용성
Controlled Vocabulary, Vocabulary Control, Thesaurus, Standard, Standardization, Information Retrieval, Indexing, Interoperability

---

* 이 논문은 2010년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2010- 32A-H00006).
** 충남대학교 문헌정보학과 부교수(sungwonk@cnu.ac.kr) (제1저자)
*** 경남대학교 인문학부 국어국문학전공 교수(kjwn@kyungnam.ac.kr) (교신저자)
논문접수일자: 2011년 1월 17일 최초심사일자: 2011년 1월 19일 게재확정일자: 2011년 2월 10일
한국문헌정보학회지, 45(1): 211-232, 2011. 〔DOI:10.4275/KSLIS.2011.45.1.211〕

www.kci.go.kr

1. 서 론

1.1 연구의 목적 및 배경

정보검색의 영역에서 시소러스를 대표로 하는 통제어휘 도구가 지난 50년간 활발히 사용되어 왔다. 시소러스는 다양한 개발주체에 의해 다양한 주제영역에서 다양한 언어로 구축되고 활용되었다. 다수의 개발 주체들이 시소러스를 구축하는 과정에서 준용할 수 있고 작업의 결과물인 시소러스의 질적 수준을 확보할 수 있도록 하기 위해 합리적인 표준(standard)을 제정할 필요성이 대두되었다. 시소러스 구축과 관련된 표준 제정의 필요성 대두에 따라 국가표준기관 및 국제표준화기구에서 관련 표준을 제정하고 개정한 바 있다(ANSI 1974, ISO 1985). 이러한 표준의 제정은 70년대에 시작되었고 후속 개정작업은 80년대 중반까지 활발히 이루어졌으나, 90년대 들어서는 1993년 ANSI/NISO Z39.19 개정 3판이 발간된 것을 제외하고는 20년 가까운 휴지기를 맞게 된다(Clarkes 2010).

2000년대 들어 시소러스는 기존의 색인 및 검색 지원기능에 더하여 다양한 조직에서 구축한 지식관리시스템(knowledge management system)의 활성화에 따라 인터넷과 인트라넷을 통해 유통되고 있는 네트워크 정보자원의 관리와 소통을 지원할 수 있는 도구로 재인식되면서 부흥기를 맞게 된다(Nielsen 2004). 이런 상황적 변화에 따라 관련 표준의 개정작업이 활발히 추진되고 있다. 2000년대에 제・개정이 추진된 시소러스 관련 표준의 대표적인 것으로 국제표준인 ISO 25964-1, 국가표준인 미국의 ANSI/NISO Z39.19-2005와 영국의 BS 8723-2005 등이 있다. 언급한 관련 표준들은 대개 80년대 중반 제・개정되었던 표준들을 오늘날의 기능적 요구사항을 반영하여 개정한 통제어휘 관련 표준들이다. 개정된 시소러스 및 통제어휘 관련 표준들은 기존의 시소러스 관련 표준에서 다루지 않았던 내용들을 포함하고 있다. 이러한 변화는 오늘날의 정보환경이 시소러스 및 통제어휘 도구에 새로이 요구하는 기능들을 반영한 결과라 할 수 있다.

본고에서는 시소러스 관련 표준의 변화를 분석하고 변화의 주요내용을 파악해 봄으로써 오늘날의 정보환경에서 통제어휘 도구가 제공해야 하는 기능적인 특성과 그러한 기능을 효과적으로 제공하기 위해 필요한 속성을 파악해 보고자 한다. 본 연구는 구체적으로 두 가지 목적을 가지고 있다. 그 하나는 시소러스를 대표로 하는 통제어휘와 관련하여 현재 주요 표준들이 어떻게 개발되고 있는 가를 중심으로 제・개정 현황을 살펴보는 것이다. 또 다른 하나는 이러한 통제어휘 관련 표준 제・개정 과정에서 발생한 주요 변화를 파악하고 정보유통 현장에서 참고할 수 있는 통제어휘 활용 관련 시사점을 도출해 보고자 하는 것이다.

이러한 고찰의 의의는 분석을 통해 통제어휘 도구를 구축, 활용하는 과정에서 고려해야 하는 시사점을 얻을 수 있다는 점이다. 시소러스 이용자의 관점에서 시소러스로 대표되는 과거의 통제어휘 도구가 전문가 집단의 사용을 전제로 하였다면 오늘날의 통제어휘들은 비전문가인 일반 이용자들의 직관적인 이용과 기계의 추론에 활용될 수 있도록 변화하여야 한다는 점도 시소러스 및 통제어휘에 대한 고찰을 필

www.kci.go.kr

요로 하는 요인이라 할 수 있다. 또한 정보의 양이 크게 증가한 현재의 정보환경에서 적합 정보를 일괄 검색하는 과정에 활용할 수 있는 통제어휘 도구와 관련된 연구는 그 의의가 매우 크다고 할 것이다.

1.2 연구의 방법 및 제한점

본고의 연구 목적을 달성하기 위해 관련 표준 및 선행연구결과를 중심으로 문헌연구를 수행하였다. 구체적인 연구과정으로 첫째, 선행연구의 분석을 통해 통제어휘 관련 주요 연구주제를 파악하여 보았다. 둘째, 관련 주요 표준을 대상으로 조사하여 표준의 제・개정 현황을 파악하였다. 셋째, 각각의 주요 표준 별로 개정내용을 분석하여 변화된 내용을 파악하였다. 넷째, 개별 표준의 변화된 내용을 중심으로 여러 표준들에서 공통적으로 파악할 수 있었던 변화를 종합적으로 정리하고 이를 통해 시사점을 도출하였다. 이러한 과정을 통해 주요표준의 제・개정 현황과 주요 변화를 분석하고 현재의 환경에서 통제어휘 도구를 적용하는 과정에서의 시사점을 정리하였다.

본 연구는 연구 수행의 시점으로 인해 야기된 제한점을 갖는다. 분석의 대상으로 채택한 일부 표준의 경우 아직 개발 중으로 미완성 상태인 경우가 있어 일부의 내용에 있어서는 관련 표준의 제・개정이 완료된 이후 추가적인 후속연구가 필요한 부분이 있다는 것이 그 제한점이다.

2. 이론적 배경

이 장에서는 본격적인 논의에 앞서 시소러스의 목적 혹은 시소러스 등 통제어휘 도구가 정보의 조직화 및 검색과정에서 제공하는 효용 및 활용현황에 관해 정리해 보고자 한다. 또한 관련으로 어떠한 선행연구들이 진행되었는지도 정리하고자 한다.

2.1 시소러스 효용 및 활용현황

문헌정보학 영역의 대표적인 통제어휘 도구로 주제명표목표를 들 수 있다. 주제명표목표는 이차원의 선형구조에 주요 주제어들을 전개한 것으로서 선택된 주제간의 관계만을 제공할 뿐이며 관련 표목어간의 다차원적인 관계 표현과 연결고리 제공에는 제한적이라는 한계를 갖는다. 이러한 한계로 인해 검색보다는 색인 도구로서의 기능에 강조점이 두어져 있다(조현양, 남영준 2004). 시소러스는 그 구조에 있어 개념간의 다차원적인 관계를 포함함으로써 보다 효과적인 검색의 도구로서 기능할 수 있다는 특성이 있다. 검색의 도구로서 시소러스의 기능을 좀더 상세히 설명하면 확장의 탐색, 대체적 검색어의 제시, 군집화(clustering) 혹은 탐색 정련화 지원, 일반적 오탈자의 판별, 그리고 자동색인의 지원 등의 효용을 가지는 것으로 정리할 수 있다(ISO 2010).

시소러스를 구축하기 위해서는 특정한 하나의 개념을 표현하는 표준 용어를 선정할 필요가 있으며 이 과정에서 관련 어휘를 검토하여 통제하는 어휘통제의 과정을 거치게 된다. 즉 통제어휘 도구인 시소러스를 구축하기 위해서

www.kci.go.kr

는 어휘통제의 과정이 전제되는 것이다. 어휘통제의 필요성은 기본적으로 하나의 개념이 상이한 용어로 표현될 수 있는 동의어(synonym), 그 반대로 하나의 용어가 여러 개의 개념을 의미할 수 있는 동음이의어(homograph)라는 자연어의 특성으로부터 발생한다. 어휘통제는 몇 가지 방법을 통해 이루어지는데 그 주된 방법은 용어가 표현하는 개념의 범위와 의미를 정의하는 것, 용어를 동의어 및 유사동의어로 연결하기 위한 대등관계를 활용하는 것, 동음이의어간의 구별을 제공하는 것 등이 그 주된 방법이다(ANSI 2005). 모든 통제어휘 개발지침은 이와 같은 어휘통제의 기본적인 방법론을 공통적으로 포함하고 있다.

통제어휘 도구는 정보검색의 효율을 향상시킬 수 있는 도구로 인정되어 널리 활용되고 있다. 최근 정보기술의 발전에 따라 시스템의 정보처리의 능력이 크게 향상된 정보환경하에서 비용대비 효과성의 관점에서 그 효용성에 대한 의문이 제기되기도 한다. 통제어휘 도구의 효용성에 대한 회의적 반응의 주요 원인은 통제어휘 도구가 가시적으로 사용되기도 하지만 일부의 경우 이용자가 인식하지 못하는 시스템의 백 엔드에서 기능하는 것에서도 원인을 찾을 수 있다. 예를 들어 이용자들이 자연어 키워드 검색엔진으로 인식하고 있는 인터넷 검색 포털에서 특정한 검색어로 검색을 수행한 경우, 이용자는 입력한 검색어와 상이하나 유사한 의미의 관련어가 포함된 검색결과를 제공받게 된다. 이것이 가능한 것은 자연어 검색으로 보이는 검색엔진의 뒤에서 다양한 형태의 통제어휘 도구가 기능하고 있기 때문이다. 이렇게 보이지 않는 곳에서 동작하고 있는 통제어휘 도구의 역할을 이용자가 가시적으로 볼 수 없는 것이 통제어휘 도구의 중요성과 효용에 대한 인식 부족의 한 원인을 제공한다.

현재까지의 다양한 논의를 통해 얻어진 결론은 정보의 조직화와 검색의 과정에서 통제어휘 도구를 사용하는 것이 검색효율 향상에 상당한 효과가 있다는 것이다(Clakes 2010). 실제 다수의 정보검색시스템에서 통제어휘 도구를 활용하고 있는 것이 그 효용성을 반증하고 있다고 할 것이다. 한 관련 기관의 조사에 따르면 현재 73개 주제영역에서 39개 언어로 작성된 670여 개의 시소러스가 구축되어 운용중인 것으로 집계되고 있다(Taxonomy Warehouse 2009).

시소러스가 주로 특정 주제 영역의 전문 정보원을 효율적으로 검색할 수 있도록 지원하는 도구로 활용되었던 것에 더하여, 오늘날 네트워크 기반 지식관리시스템의 활성화에 따라 특정 조직이나 전문영역에서의 효율적인 정보관리와 소통을 지원하는 도구로 활용되고 있다(Nielsen 2004). 또한 시소러스는 정보환경의 개선을 위한 기반도구로도 활용되고 있다. 새로운 정보환경을 위해 제안된 시맨틱웹과 이를 구현하기 위한 도구인 온톨로지(ontology)의 구축과정에서 기존에 개발된 시소러스를 활용하는 방안과 관련한 다양한 연구가 수행되고 있는데(정도헌, 김태수 2003) 이는 시소러스가 새로운 정보환경 구현에 필요한 도구를 구축하는 것에도 기여할 수 있음을 의미한다.

2.2 선행연구 개관

지난 50여 년간 문헌정보학 및 정보처리 분야에서 시소러스를 대표로 하는 통제어휘 도

www.kci.go.kr

구들이 색인 및 정보검색의 지원도구로 활용되어 왔음은 주지의 사실이다(Kwasnik 1999). 시소러스는 다양한 주제분야에서 해당 영역의 전문적 정보검색을 효율적으로 수행할 목적으로 구축, 활용되어 왔고 그에 따라 다양한 관련 연구가 수행되어 개별 연구 하나하나를 소개할 수 없을 정도로 방대한 선행연구결과들이 있다. 이러한 상황을 반영하여 개별 연구를 소개하기 보다 주요한 연구분야로 어떤 세부분야에서 연구가 이루어졌는지를 개괄해 보고자 한다.

우선 가장 기본적인 연구영역으로 시소러스 개념 및 효용성과 관련한 근본적인 문제에 대한 연구들을 들 수 있다. 시소러스 개발 초기부터 2000년대 부흥기에 이르기까지 시소러스의 효용성과 역할에 대한 다양한 연구가 진행된 바 있다. 두 번째 연구분야로는 시소러스의 구축과 관련한 연구분야를 들 수 있다. 이 분야의 연구주제는 개념 및 용어의 수집에 관련한 연구, 개념과 용어의 정의와 선정에 관한 연구, 그리고 개념과 용어의 조직화에 관련한 세부 영역을 포함한다. 특히 이 분야의 연구결과는 시소러스 구축관련 지침을 제공하는 단행본으로 출판된 예를 많이 찾아볼 수 있다. 시소러스 관련 세 번째 연구분야는 시소러스의 구축 지침과 관련된 연구영역이다. 시소러스의 구축과정인 개념 및 용어의 수집/정의 및 선정/조직화는 일정한 규칙들을 필요로 했고 이러한 규칙을 확립하고 이를 기반으로 구축과정에서의 의사결정 기준을 수립하는 것과 관련한 연구가 진행되어 왔다. 네 번째 연구영역으로는 시소러스의 자동구축과 관련된 연구영역을 언급할 수 있다. 이 영역에 해당되는 세부 연구로는 용어를 자동추출하고 이를 통계적으로 처리하여 군집화를 통해 조직화하는 연구영역, 자동구축 과정에서 인간 지능의 개입을 필요로 하는 과정의 정의와 개입의 정도에 대한 연구, 그리고 시소러스의 구축 및 관리를 지원할 수 있는 소프트웨어 도구의 활용과 그 선정기준 등의 연구영역을 들 수 있다. 위와 같은 기존의 연구영역에 더하여 최근 들어서는 시소러스를 통해 다양한 정보원에 대한 일괄 검색 기능을 구현하기 위한 목적으로 상이한 시소러스 및 통제어휘간 상호운용성을 확보하는 방안에 대한 연구가 활발히 진행되고 있으며 시소러스를 활용하여 온톨로지를 구축하는 것과 같이 시소러스의 새로운 활용과 관련된 연구도 진행되고 있다(정도헌, 김태수 2003). 이렇게 다양한 시소러스 관련 주요 연구결과들을 영역별로 나누어 소개하고 있는 리뷰 논문들도 다수 정리되어 있는 바, 관심 영역에서의 선행연구를 파악하고자 하는 경우에 활용할 수 있을 것으로 판단된다. 세부 관심 연구영역으로의 접근을 제공할 수 있는 서지도구로 활용할 수 있는 포괄적 자료로 Nielsen(2004)의 논문과 ANSI/NISO Z39.19(2005), Will(2010)의 서지자료를 대표적인 예로 들 수 있고, 시소러스 관련 연대기적 개관을 제공하는 자료로 Aitchison and Clarkes (2006)의 저술을 제시할 수 있다.

3. 표준의 제·개정 현황

이 장에서는 다양한 통제어휘 도구와 그 구축을 위한 주요 표준의 제·개정 현황에 대해 살펴보고자 한다. 통제어휘와 관련된 표준 모

www.kci.go.kr

두는 기본적으로 모호성의 제거, 동의어 통제, 용어간의 관계 설정, 그리고 용어의 평가 및 승인 절차 등을 공통적으로 다루고 있다. 대표적인 통제어휘 관련 표준으로 국제표준인 ISO, 국가표준인 BSI와 ANSI, 관련기관의 지침인 IFLA의 가이드라인(guideline) 자료 그리고 최근 웹의 발전과 더불어 활발한 활동을 하고 있는 W3C(World Wide Web Consortium)의 사실표준(de facto standard) 등을 들 수 있다.

이 가운데 ISO, BSI, ANSI 등의 국제 및 국가표준은 공인된 표준화기구에서 제정하는 표준으로서 대표성이 있는 국제표준화기구나 정부기관이 제정하는 공식표준(de jure standard)에 해당하며, W3C의 표준은 시장에서 실제적 표준으로 인정받거나 필요에 따라 업계를 중심으로 결성된 사실표준화기구에서 제정하는 표준으로서 강제사항은 아니나 관련 업계 대부분이 준수하는 사실표준에 해당된다

1970-1980년대에 거쳐 다양한 시소러스가 특정한 데이터베이스 검색을 위한 지원도구로서 구축되었고 이를 위한 관련 표준이 개발되어 왔다. 그러나 80년대 중반 ISO 2788과 5964의 제정 이후 20여 년간 관련 표준화 활동이 정체되었다(Clarkes 2010). 미국의 ANSI만이 1993년에 ANSI Z39.19를 개정하였을 뿐이다. 이렇게 정체되었던 통제어휘 관련 표준들이 2000년대에 들어 개정되고 있다. 이러한 개정의 배경에는 기존의 특정한 데이터베이스를 검색하기 위한 도구로서가 아닌 검색요구에 적합한 다양한 데이터베이스와 리포지터리, 그리고 정보자원을 일괄 검색할 수 있는 도구로서 통제어휘를 활용할 수 있는 가능성이 대두되었기 때문이다(Nielsen 2006). 또한 변화된 정보환경 아래에서 새로운 활용을 통한 시소러스 등 통제어휘의 유용성이 재인식되고 있는 것에 기인한다.

3.1 ISO 동향

ISO 표준은 국제표준화기구인 ISO(International Organization for Standardization)에서 제정하는 공식적인 국제표준이다. ISO의 활동에는 국가의 대표기관이 참여할 수 있고 해당국의 담당 기관에서 선정한 전문가가 대표로서 참여할 수 있다. 우리나라를 대표하여 국제표준기구 ISO에 참여하고 있는 기관은 기술표준원이며, 각 분야의 표준화에 참여하기 위해서는 기술표준원의 각 분과에서 심의과정을 거쳐 전문가로 등록하여야 한다.

통제어휘 관련 ISO 표준을 관장하는 분과는 ISO TC46 SC9 WG8이다. 기술위원회(TC; Technical Committee) 46은 'Information and Documentation' 분야를, 분과위원회(SC; subcommittee) 9는 'Identification and description'을, 그리고 작업반(working group) 8은 'structured vocabulary'를 담당하는 분과이다. 특히 TC 46 SC 9의 경우 우리나라의 오삼균이 의장(chairperson)을 맡아 기여하고 있다(ISO 2011).

ISO의 통제어휘관련 표준은 ISO 25964로서 파트 1, 파트 2의 두 부분으로 구성되어 있다. 파트 1은 '정보검색을 위한 시소러스'이며 파트 2는 '여타의 통제어휘와의 상호운용성'이다. 이 표준은 기존의 관련 표준인 ISO 2788:1986과 ISO 5964:1985를 개정한 것이다. ISO 2788과 5964는 각기 '단일어 시소러스 구축 및 개발 지침'과 '다국어 시소러스 구축 및 개발 지침'을

www.kci.go.kr

규정하고 있었다. 새로운 표준인 ISO 25964에서 이 두 표준이 하나의 표준으로 통합되었으며 기존 표준의 주요 구성요소들은 ISO 25964 파트 1에 포함되어 있다. 통제어휘 관련 ISO 표준의 변화를 정리하면 〈표 1〉과 같다(ISO 2010).

전술한 바와 같이 기존 ISO 표준인 ISO 2788, 5964는 모두 시소러스의 구축과 개발에 관련된 것으로 다만 단일어 시소러스냐 혹은 다국어 시소러스를 위한 것이냐 하는 차이를 가지고 있었을 뿐이며, 새로이 제정된 ISO 25964 표준에서는 다국어 시소러스 구축을 전제로 하나의 표준인 ISO 25964 파트 1로 통합되었다. 그리고 새로운 표준인 ISO 25964의 두 번째 파트는 이전의 ISO 표준에서 고려하지 않았던 복수 시소러스간의 혹은 택소노미(taxonomy), 온톨로지 등 시소러스와 다른 형태의 통제어휘와 시소러스간의 상호운용성 확보를 위한 내용을 담을 예정으로 그 개발이 추진되고 있다. 따라서 새로이 개발되고 있는 ISO 25964에는 ISO 2788, 5964 등 기존의 표준에서 포함하지 않았던 상호운용성 관련 내용들을 포함하고 있다. 또한 이미 완성 단계에 있는 ISO 25964의 파트 1에서 데이터모형(data model), 타 시소러스와의 통합문제, 그리고 프로토콜 등 기존의 표준에서 다루지 않았던 내용을 포함하고 있다(ISO 25964-1 2010).

ISO 25964의 개발 진행 단계에 대해 살펴보면, 시소러스의 구축을 다룬 파트 1은 최종국제표준안(FDIS; Final Draft International Standard)의 단계이다. ISO의 단계코드(stage code) '50.00'으로 이는 공식적인 최종승인을 위한 표준안의 등록이 완료된 최종 승인단계를 의미한다. 상이한 시소러스 및 통제어휘 도구와의 상호운용성을 규정한 파트 2의 경우에는 표준안의 준비단계인 ISO 단계코드 '20.00'에 있다. 이를 기반으로 향후 일정을 예상하여 보면, 파트 1의 경우에는 2011년 중에 표준으로 발간될 것으로 예상되며, 파트 2의 경우에는 프로젝트가 등록되었을 뿐 구체적인 표준안의 작성이 시작되지 않은 단계로서 향후 표준안의 작성, 전문위원회의 검토, 전문가 자문, 관련자의 의견수렴 그리고 승인절차를 남겨 놓고 있어 상당한 시일이 소요될 것으로 사료된다.

〈표 1〉 통제어휘 관련 ISO 표준 현황

구분	표준번호	제·개정년	제목	비고
개정전	ISO 2788	1986	Documentation－ Guidelines for the establishment and development of monolingual thesauri	
	ISO 5964	1985	Documentation－Guidelines for the establishment and development of multilingual thesauri	
개정후	ISO25964-1	2011 (예상)	Thesauri and interoperability with other vocabularies－ Part 1: Thesauri for information retrieval	FDIS (처리일: 2010.12.23)
	ISO25964-2	미정	Thesauri and interoperability with other vocabularies－ Part 2: Interoperability with other vocabularies	(처리일: 2010.8.9)

www.kci.go.kr

3.2 BSI 동향

영국의 국가표준인 BS(British Standard)는 국가표준화 단체인 BSI(British Standards Institution)에 의해 개발된다. 영국의 경우 국제표준화기구인 ISO에서 개발한 ISO 표준인 ISO 2788과 5964를 통제어휘 관련 국가표준을 채택하여 각각 BS 5723과 BS 6723으로 제정하였으나, 관련 표준의 개정 필요성이 대두됨에 따라 2005년 이후 독자적인 표준인 BS 8723을 개발하고 있다. 개발방향 논의 과정에서 많은 내용을 담아야 할 필요성이 대두되었던 이유로 5개의 파트로 나누어 개발하기에 이른다(Clarkes 2010). BS 8723은 기존의 표준인 BS 5723, 6723에 포함되었던 내용을 포괄할 뿐만 아니라, 분류체계, 주제명표목, 택소노미, 온톨로지 그리고 전거화일까지를 포함한다. 또한 다양한 통제어휘간의 상호운용성 확보를 지향하여 통제어휘 도구간의 데이터 교환을 지원할 수 있는 틀로서 새로운 데이터모형과 XML 스키마(schema)를 포함하고 있다.

이렇게 기존의 시소러스 이외의 다양한 내용을 담아야 했던 이유로 BS 8723은 모두 5개 파트로 구성되어 있는데, 기존의 표준이 담고 있던 내용은 파트 1, 2, 4에 포함되어 있으며 파트 3, 5의 내용은 최근의 정보통신기술 접목에 따라 새로이 규정할 필요가 발생한 내용을 담고 있다. BS8723의 5개 파트 가운데 4개의 파트는 확정된 규격으로 발간이 완료되었고 마지막 부분인 파트 5의 경우에는 최종원고(DD: Draft for Development)의 상태로 우선 활용되고 있다(BSI 2005a, BSI 2008) (표 2 참조).

3.3 ANSI 동향

ANSI(American National Standards Institute)는 미국의 국가표준기구로서 다양한 분야의 표준을 개발하고 있으며 이 기구에서 제정한 국가표준인 ANS(American National Standards)는 'ANSI'라는 기구명으로 식별된

〈표 2〉 통제어휘 관련 BS 표준 현황

구분	표준번호	제·개정년	제목	비고
개정전	BS 5723	1987	Guidelines for the establishment and development of monolingual thesauri	ISO 2788 채용
	BS 6723	1985	Guidelines for the establishment and development of multilingual thesauri	ISO 5964 채용
개정후	BS 8723		Structured vocabularies for information retrieval-Guide	
	Part 1	2005	Definitions, symbols and abbreviations	
	Part 2	2005	Thesauri	
	Part 3	2007	Vocabularies other than thesauri	
	Part 4	2007	Interoperability between vocabularies	
	Part 5	작업중	Exchange formats and protocols for interoperability(DD: Draft for Development)*	최종원고(DD)

* DD(Draft for Development)는 정식 표준은 아니지만 상당한 작업이 이루어진 최종 원고로서 관련 지침의 제시가 긴급히 필요한 경우 발행되는 표준자료임.

www.kci.go.kr

다. 다양한 분야에서 관련 표준을 제정하는 업무를 수행하다 보니 표준개발과정의 효율성 제고를 위해 전문분야에 관련된 표준은 220여 개의 전문기관을 표준의 개발자로 인증하여 표준개발을 추진하고 있다(ANSI 2011). 이러한 표준개발 협력기관 가운데 정보관리와 관련된 표준개발은 NISO(National Information Standards Organization)에서 담당하고 있으며, NISO는 ISO TC46 등 관련 국제기구에 ANSI를 대표하여 참여하고 활동하고 있다. 이에 따라 정보관리 관련 미국 표준은 개발협력기관의 명칭을 부기하여 'ANSI/NISO'와 같이 표기된다.

정보관리영역에 포함되는 통제어휘 관련 ANSI 표준 역시 NISO에 의해 구체안이 제·개정되어 왔다. 미국의 국가표준 ANSI의 통제어휘 관련 표준인 Z39.19는 NISO의 여러 표준 가운데에서도 그 활용도가 높았던 표준으로 1974년 최초로 제정된 이후 1980년과 1993년 두 차례의 개정을 거쳤고 2005년 제 4판이 제정되었다. ISO와 BS의 경우 80년대 개정 이후 관련 표준의 개정이 이루어지지 않았던 것에 비교해 ANSI의 경우에는 1993년 개정이 이루어졌음을 주목할 만하다. 1993년 제 3판 개정 이후 1999년 관련 워크샵에서 개정의 필요성이 대두되었고, 2002년부터의 개정작업을 통해 2005년 새로운 표준을 제정하게 되었다(ANSI 2005)(표 3 참조).

ANSI Z39.19의 특징으로 ISO나 BS와는 달리 표준의 표제에 시소러스라는 표현 대신 통제어휘(controlled vocabularies)라는 용어를 채용하고 있는 점을 들 수 있다. 내용에 있어 오늘날의 정보검색 환경에서 사용되고 있는 시소러스 이외의 통제어휘 도구인 선택목록(pick list), 동의어 순환목록(synonym ring), 그리고 택소노미 등도 포함하고 있다.

3.4 기타 표준

국제표준이나 국가표준 이외에도 관련 기관이나 사실표준기구에서 제정하는 통제어휘 관련 표준들도 있다. 도서관계의 대표적인 국제기구인 IFLA에서도 전문보고서(professional report)로서 다국어 시소러스 지침인 'Guidelines for Multilingual Thesauri'을 발간한 바 있다(IFLA 2009). 이 지침은 1976년에 발간되어 20년 이상 경과된 유네스코의 관련 지침인 'Guidelines for the Establishment and Development of Multilingual Thesauri'을 대체하기 위한 것으로 다국어 시소러스 구축을 위한 두 가지 접근법을 중심으로 간략한 지침을 제시하고 있다. 구체적으로는 특정 시소러스의 일부분을 채용하여 비대칭적(non-symmetrical) 시소러스를

〈표 3〉 통제어휘 관련 ANSI 표준 현황

구분	표준번호	제·개정년	제목	비고
개정전	ANSI/NISO Z39.19	1974 1980 1993	Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies	1판 2판 3판
개정후	ANSI/NISO Z39.19	2005	Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies	4판

www.kci.go.kr

구축하는 방안과 둘 이상의 시소러스를 연결하여 다국어로 시소러스를 구축하는 방안을 제시하고 있다. 이러한 지침은 특정 언어만이 아닌 다양한 언어로 생산된 정보자원을 일괄 검색하여 접근하기 위한 도구로서 다국어 시소러스를 구축하기 위한 지침을 제공하기 위한 것이다.

시소러스 관련 표준과 관련하여 새로이 주목할 만한 기관으로 웹의 발전과 더불어 관련된 사실표준을 제정하고 있는 W3C를 들 수 있다. W3C의 성격은 웹의 장기적인 성장을 보장하기 위한 표준을 개발하는 것을 목적으로 하는 기구로서 개선된 웹 환경인 시맨틱웹을 구현하기 위한 각종의 관련 표준을 제정하고 있으며 특히 정보조직화와 검색의 도구로 제안된 온톨로지 관련 표준도 제정하고 있다. W3C의 통제어휘 관련 표준으로 대표적인 것은 'SKOS Simple Knowledge Organization System Reference'로서 이 표준은 시소러스, 택소노미, 분류표, 그리고 주제명표목과 같은 기존의 정보조직도구에서 공통적으로 찾을 수 있는 요소 및 관계 유형을 정의함으로써 웹 상에서 이들 정보조직도구를 공유하고 연결할 수 있는 공통된 데이터 모형을 제시하고 있다(World Wide Web Consortium 2009). 이 데이터모형은 그 자체로 새로운 정보조직도구로 활용될 수 있으며, 시소러스와 같은 기존의 정보조직도구를 활용하여 시맨틱웹으로 이전할 수 있는 지침으로 활용될 수도 있다.

이상 살펴본 바와 같이 통제어휘 관련 표준은 국제표준기구, 국가표준기구 그리고 관련기구와 사실표준기구 등에서 활발히 검토되고 제·개정이 이루어지고 있다. 이러한 관련 표준의 개발 노력은 첨단의 정보통신기술이 정보검색과 결합된 오늘날의 환경에서도 대표적인 통제어휘 도구인 시소러스가 여전히 중요한 도구로 활용될 수 있음을 반증하고 있다고 할 것이다.

4. 표준 개정의 주요내용 및 시사점

이 장에서는 통제어휘 관련 표준의 주요 변화내용을 개별 표준이 아닌 관련 표준 전체의 종합적인 관점에서 파악해보고 시사점을 정리하여 보았다. 분석의 대상으로는 ISO, BSI, 그리고 ANSI의 표준을 중심으로 하였다. 이들 표준을 분석 중심으로 삼은 이유는 기존의 표준과 개정된 표준간의 비교를 통해 변화를 파악해 볼 수 있었기 때문이다. 이러한 분석을 통해 정보의 조직화 및 검색에 있어 통제어휘 도구를 활용하는 것과 관련한 변화와 의미 그리고 통제어휘 도구를 현재의 시점에서 적용하는 것과 관련된 시사점을 얻을 수 있을 것이다.

4.1 적용 대상자료

우선 주요 관련 표준의 공통된 변화로 통제어휘 도구를 적용하는 대상자료의 확장을 볼 수 있다. 관련 국제표준인 ISO의 25964-1의 경우 기존의 표준에서 대상으로 하는 자료의 대표적 유형을 도큐멘트(document)로 정의했던 것에 비해 새로운 표준에서는 전달 매체와 상관없이 정보자원(information resource)으로 확대하여 정의하고 있다. 물론 기존의 도큐멘트의 개념이 단순히 인쇄된 자료만을 의미하였던 것은 아니고 비 인쇄자료까지를 포함하는 것으로 정의되고 있었으나 그 중심은 인쇄매체를

www.kci.go.kr

중심으로 하고 있었다. 그러나 ISO 25964-1에서는 네트워크를 통해 입수할 수 있는 인터넷 및 인트라넷 리소스 자료까지를 포함하여 지식베이스(knowledge bases), 웹 포털(portals), 멀티미디어(multimedia) 등 모든 형태의 정보자원으로 확장하고 있으며 다양한 형태의 정보자원을 'information resources'라는 용어로 표현하고 있다.

ANSI의 경우에도 기존의 표준에서 대상으로 하는 정보자원을 'document'로 표현하였던 것에 비해 이를 데이터와 정보를 담고 있는 객체로 정의되는 내용객체(contents object)로 확장함으로써 좀더 다양한 전자적 자원을 적용대상자료로 포함하게 되었음을 반증하고 있다. BS의 경우 개정 이전에는 ISO 표준을 국가표준으로 채택하였음은 전술한 바와 같다. 이에 따라 통제어휘의 적용대상으로 상정한 대상 정보자원은 도큐멘트가 중심을 이루었음은 ISO와 동일하다. 개정된 BS 8723의 경우 표준개정이 필요한 이유로 인쇄형태에서 전자적 활용으로의 확장이 필요한 점을 제시하고 있어 앞에서 언급한 다른 표준과 그 궤를 같이 하고 있다(Clarke, S. D., Gilchrist, A., & Will, L. 2004).

사실상의 표준화기구인 W3C가 제정한 통제어휘 관련 표준은 W3C 권고(recommendation)인 'SKOS Simple Knowledge Organization System Reference'이다. W3C가 기본적으로 웹의 장기적인 발전을 위한 표준을 제정하는 기관임은 전술한 바와 같다. 웹의 성장을 보장할 수 있는 프로토콜과 지침을 개발함으로써 웹의 가능성을 최대화하는 것에 목적을 두고 있는 만큼, 웹을 통해 유통되는 모든 정보자원을 대상으로 하고 있다.

이상에서 살펴본 바와 같이 통제어휘 관련 표준 모두 기존의 인쇄물 중심의 정보조직화 및 검색의 도구로 인식하였던 통제어휘 도구를 다양한 형태의 정보자원으로 확장하여 적용하는 방향을 잡고 있음을 알 수 있다. 관련 표준들이 그 적용대상으로 상정하고 있는 대상자원이 인쇄자료 형태에서 웹을 통해 유통되는 다양한 자원까지로 확장된 것을 확인할 수 있다.

4.2 통제어휘 도구의 유형

관련 표준의 변화에서 파악할 수 있는 또 다른 변화로서 시소러스로 대표되던 통제어휘 도구가 다양한 형태로 확장되고, 표준의 내용도 그러한 도구의 활용까지를 포괄하여 확장되고 있음을 알 수 있다. 시소러스 이외의 통제어휘 도구로는 ANSI에서는 선택목록, 동의어 순환목록, 택소노미 등을 제시하고 있으며, ISO에서는 택소노미, 전거목록(authority lists), 온톨로지 등을 예시하고 있다. 영국의 국가표준인 BS 8723의 경우에는 시소러스 이외의 다양한 통제어휘 도구들과 관련된 'Vocabularies other than Thesauri'라는 파트 3을 별권으로 구성하여 다양한 유형의 통제어휘 도구에 대한 지침을 제시하고 있다(BSI 2007a).

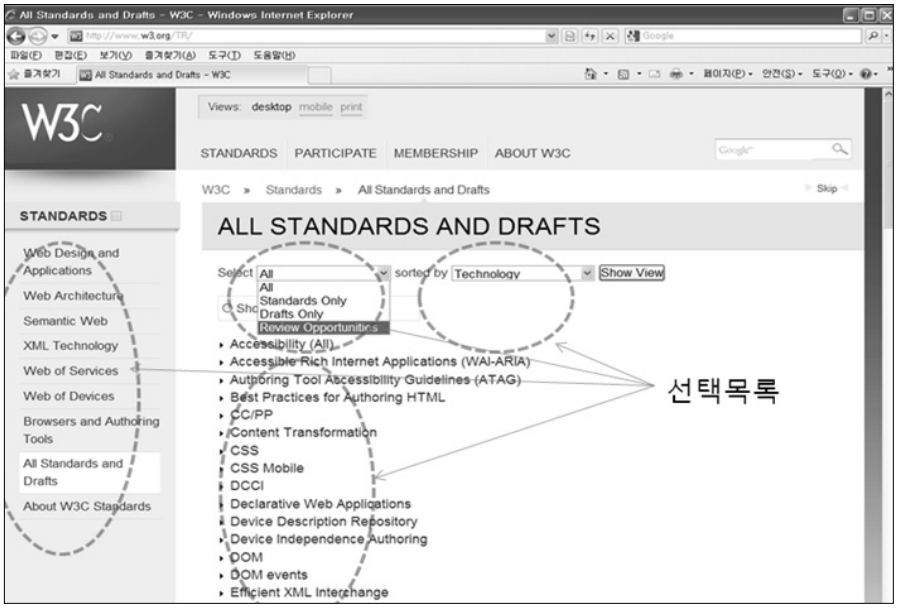
이렇게 제시되고 있는 다양한 통제어휘 도구 가운데 가장 간단한 형태의 통제어휘 도구는 ANSI에서 제시하고 있는 '선택목록(pick list)'으로서 웹 환경에서 제공되는 분류목록이나 팝업 메뉴를 통해 이용자가 선택할 수 있도록 제공되는 형태를 의미한다. 선택목록의 사용 예는 웹 이용 시 주소나 언어의 선택 창, 혹은 특정 항목의 세부선택 항목과 같은 경우에서 쉽

www.kci.go.kr

게 찾아볼 수 있다. 선택목록은 개념을 표현하기 위해 선정된 표목어만을 제시할 뿐, 표목어의 정의와 범위와 같은 주기(notes)를 포함하지는 않는다. 〈그림 1〉은 선택목록의 예시이다. 예에서 점선의 원으로 표시된 부분이 선택목록의 예로서 이용자는 제시된 항목 가운데 관심있는 항목을 클릭하여 선택할 수 있으며, 제공자가 해당 목록을 사전에 통제하여 제공한다는 의미에서 간단한 형태의 통제어휘 도구에 포함된다.

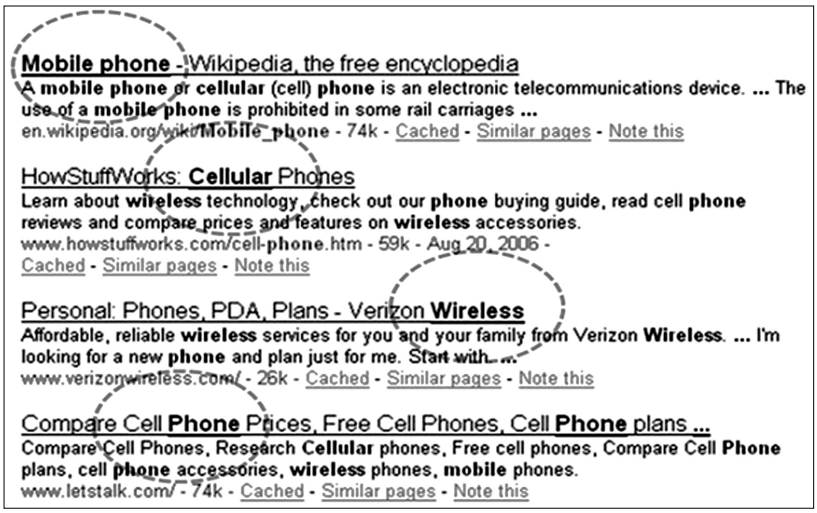
또 다른 통제어휘 도구인 '동의어 순환목록(synonym ring)'은 검색 시 이용자의 검색어를 동의어 및 유사동의어 관계가 있는 용어까지로 확장하여 주는 기능을 하여 검색의 확장 목적으로 활용되며 색인시 활용되지 않는다는 점에서 여타의 통제어휘 도구와 기능상의 차이를 갖는다. 아래 〈그림 2〉는 구축된 동의어 순환목록 동작한 구글검색의 결과를 보여주는 예이다. 예에서 이용자는 'mobile phone'이라는 검색어로만 검색을 하였으나 동의어 순환목록에 등록되어 있던 'cellular', 'wireless' 등의 색인어를 포함하는 정보자원도 검색된 것을 볼 수 있다. 이러한 동의어 순환목록은 웹 포털에서 일반적으로 제공되고 있는 연관검색어를 제시하는 과정에서도 활용된다.

이와 같이 특정한 검색어와 관련된 동의어와 유사동의어까지로 검색을 확장하기 위해서는 검색시스템에 동의어 순환목록이 구축되어 있어야 한다. 〈그림 3〉은 ANSI에서 언어장애(speech disorder)를 예시로 하여 제시하고 있는 동의어 순환목록 개념도이다(ANSI 2005). 동의어 순환목록의 구축을 위해서는 포함될 용어의 선정 등 어휘통제가 이루어지므로 통제어휘 도구의 하나로 볼 수 있다.

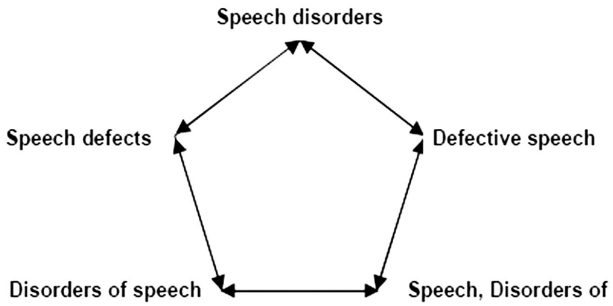


〈그림 1〉 선택목록의 예시화면(W3C의 주제별 표준 선택 화면)

www.kci.go.kr



〈그림 2〉 동의어 순환목록이 적용된 구글의 검색결과



〈그림 3〉 동의어 순환목록 개념도(ANSI 2005)

동의어 순환목록은 이용자에게 가시적으로 보여지지 않고 시스템의 백 엔드에서 동작한다. 오늘날의 정보검색 행태가 자연어 키워드 중심으로만 이루어지고 있다고 생각하기 쉬우나 이용자가 알지 못하는 영역에서 이와 같이 다양한 형태의 통제어휘 도구가 활용되고 있다. 자연어 키워드 중심의 검색환경에서도 어휘통제를 통해 구축된 통제어휘 도구가 정보검색의 효율성과 이용자의 편의성을 개선시킬 수 있음을 보여주는 좋은 예라 할 수 있다.

다양한 유형의 통제어휘 도구는 그 유형에 따라 포함하는 속성이 상이하다. ANSI에서는 상이한 통제어휘 도구가 포함하고 있는 속성에 대해 아래 표와 같이 정의하고 있다(ANSI 2005). 〈표 4〉에서 선택목록은 선택된 표목만을 포함하며, 동의어 순환목록은 표목에 더해 동의어와 유사동의어를 포함하고, 택소노미는 개념을 나타내는 표목과 개념간의 계층관계를 포함하고 시소러스는 동등, 계층, 연관관계 등 다양한 관계까지를 포함하여 제공하는 차이가 있음을

www.kci.go.kr

〈표 4〉 통제어휘의 속성(ANSI 2005)

Property	List	Synonym Ring	Taxonomy	Thesaurus
Types of Terms				
Preferred terms	Yes	No	Yes	Yes
Entry terms	No	Yes	No	Yes
Candidate terms	No	No	No	Optional
Provisional terms	No	No	No	Optional
Deleted terms	No	No	No	Optional
Relationships	No	Yes	Yes	Yes
Equivalence		Yes	No	Yes
Hierarchy		No	Yes	Yes
Part/Whole		No	Yes	Yes
IsA		No	Yes	Yes
HasA		No	Yes	Yes
Classification		No	Optional	Optional
Related terms		No	No	Yes
Facet		No	No	Optional
Notes	No	No	Optional	Optional
Scope note			No	Optional
History note			No	Optional
Other notes			No	Optional

볼 수 있다.

4.3 통제어휘 표현

관련 표준의 변화에서 파악할 수 있는 통제어휘의 또 다른 변화로서 표현매체와 표현형식(display)의 변화를 들 수 있다. 우선 매체의 관점에서 정보통신기술의 도입 이전에는 인쇄형 책자형태가 대표적인 통제어휘 도구 표현매체였으며, 기본적으로 시소러스는 인쇄된 책자로 발간되는 것이 일반적인 형태였다. 표현형식의 관점에서는 기존에도 통제어휘의 자모순 배열이나 주제순 배열과 더불어 개별 통제어휘의 상세내역이나 KWIC(Key Word In Context), KWOC(Key Word Out of Context) 색인 등 순열표현(permuted display)방식이 제시되기도 하였으나 다양한 형식으로 제공하기 위해서는 방대한 작업 및 자료의 분량이 전제되므로 그 실제적인 제공에는 한계가 있었다.

정보통신기술의 발전에 따라 표현매체가 인쇄매체에서 컴퓨터 스크린과 웹으로 확장되었고 컴퓨터 스크린과 웹을 통해 다양한 통제어휘 표현형식의 제공이 가능하게 되었다. 이에 따라 개정된 통제어휘 관련 표준들에서는 정보통신기술의 발전에 기반하여 통제어휘 도구의 표현매체를 웹 화면 같은 매체로 확장하였다. 또한 기존의 인쇄형 책자형태에서 가능했던 표현형식에 더하여 컴퓨터 스크린과 웹을 통해 제공이 가능하여진 다양한 통제어휘 표현형식에 대한 내용을 포함하고 있다. ANSI의 경우 인쇄매체를 통한 기존의 표현형식과 더불어 통제어휘를 그래픽 형식으로 표현하는 것에 관련

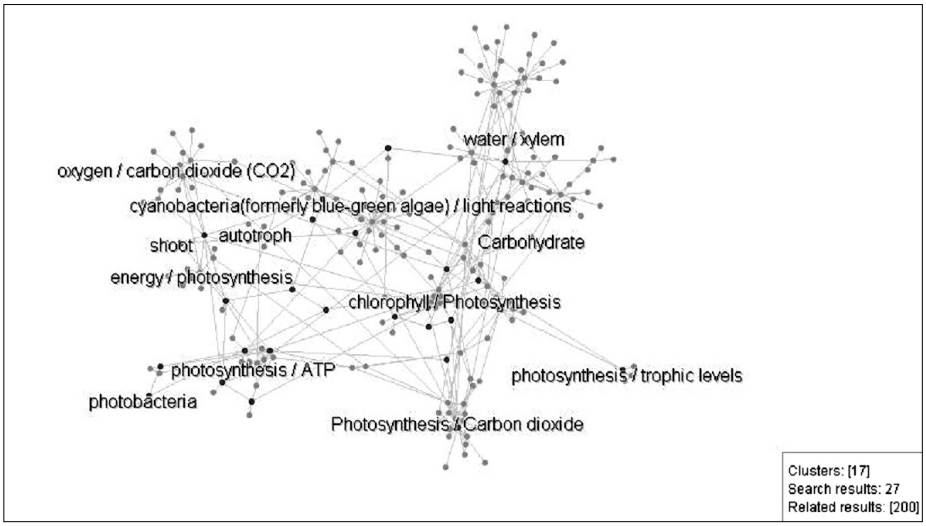
www.kci.go.kr

된 지침을 제시하고 있다. ISO의 경우에도 그래픽 표현형식을 포함한 다양한 표현형식을 제시하고 있음을 확인할 수 있다.

이와 같이 다양한 표현형식을 이용자가 요구하는 시점에서 즉시 제공하는 것이 가능해진 것은 발전된 정보통신기술을 적용하였기 때문이다. 통제어휘의 관리과정에 소프트웨어 보조도구를 사용하게 됨으로써 관련 정보가 전자적 데이터 형태로 존재하게 되었으며, 이러한 데이터를 기반으로 컴퓨터 스크린과 웹 화면과 같은 매체를 통해 시소러스 등 통제어휘 도구에 포함되어 있는 내용을 이용자가 요구하는 다양한 형식으로 보여줄 수 있게 되었다. 관련 기관의 조사에 따르면 현존하는 시소러스의 65%가 디지털형태로 생산되고 있으므로 관련 표준에서 제시하고 있는 표현형식의 지원이 가능한 환경이 구축되어 있다고 할 수 있다(Taxonomy Warehouse 2009).

표현형식 유연성은 기본적으로 다양한 이용자 요구를 즉시 만족시킬 수 있다는 점에서 통제어휘의 활용도를 제고시킬 수 있을 것으로 사료된다. 표현형식의 다양화는 기존의 인쇄된 책자형 매체의 특성에 따라 2차원의 선형구조로 표현할 수밖에 없었던 개념간의 관계를 다차원적인 그림으로 표현할 수 있게 해주므로 관계 표현에 있어 좀 더 직관적인 이해를 제공할 수 있게 해준다. 개념지도(concept map)를 통해 시각적인 구조로 개념간의 관계를 표현하는 아래의 〈그림 4〉에서 관련된 개념간의 관계를 좀 더 직관적으로 파악할 수 있음을 볼 수 있다(ANSI 2005).

디지털 전자정보 형태로 저장된 통제어휘의 경우 시각화 도구를 활용하면 개념간의 관계를 도식화하여 보여줄 수 있는 잠재력을 가지고 있다. 또한 보조 소프트웨어 도구의 활용을 통해 전자적 정보형태로 통제어휘를 관리하는 것은 다음 개정판의 출간까지 갱신하기 어려웠던 기존의 책자형 시소러스와는 달리 새로운 통제



〈그림 4〉 개념간의 관계 표현을 위한 개념지도 예시 화면(ANSI 2005)

www.kci.go.kr

어휘의 도입, 기존 통제어휘의 개정 및 폐기 등 갱신이 이루어진 시점에서 바로 반영되므로 통제어휘 도구의 최신성이 개선되는 효과를 수반할 것이다.

4.4 다국어 수용

통제어휘 관련 80년대의 대표적인 표준이었던 ISO 2788, 5964 및 BS 5723, 6723이 각각의 표준 체계 내에서 단일어 시소러스와 다국어 시소러스 구축을 위한 표준으로 나뉘어 있던 것을 통해 볼 수 있듯이, 기존의 통제어휘는 기본적으로 단일어 시소러스의 구축을 전제로 하고 특수한 경우 다국어 시소러스를 구축하는 것이 일반적이었다. 이러한 단일어 통제어휘를 기본으로 하였던 점은 미국표준 ANSI Z39.19-2005의 경우 여전히 표준의 표제에 단일어(monolingual) 통제어휘를 위한 지침임을 명시하고 있는 것으로도 확인이 가능하다. 개정된 ISO 25964-1의 경우에도 별도의 표준으로 제정되어 있던 단일어 시소러스와 다국어 시소러스 관련 표준을 하나로 통합하였음은 전술한 바와 같다. 다국어 시소러스를 시소러스의 기본형태로 상정하였기 때문에 발생한 변화라 할 것이다.

다국어 시소러스의 대상언어 범위에 있어서도 확장된 모습을 확인할 수 있다. 다국어 시소러스 구축을 위한 국제표준이었던 ISO 5964의 경우 다국어 시소러스 구축을 위한 표준이었음에도 불구하고 본문 중의 예시에 포함하고 있는 언어는 영어, 독일어, 그리고 불어의 3개 국어가 중심을 이루고 있다. 다국어 시소러스를 지향하는 표준이었음에도 불구하고 실제 고려한 언어는 소수에 그쳤음을 반증한다. 이러한 제한은 표준의 제・개정에 참여한 전문가의 모국어에 영향을 받았을 것이라는 점을 고려하더라도 그 범위에 있어 매우 제한적이었다고 할 것이다. 새로이 개정된 ISO 25964-1을 살펴보면 각각의 언어로 작성된 통제어휘 도구에서 채용하고 있는 기호(tag)의 의미를 10개 국어로 매핑하고 있어 포괄하는 다국어의 범위가 크게 확장되었음을 볼 수 있다(ISO 2010).

이러한 변화는 인터넷의 보급활성화에 따라 이용자들이 모국어 이외의 언어로 생산된 정보자원에 접근할 수 있게 되었으며 대상 정보자원의 확장에 따라 다양한 언어로 기술된 정보자원까지를 일괄 검색할 수 있도록 해주는 통제어휘가 필요하게 된 상황을 반영한 결과이다. 기존의 통제어휘 도구 활용과 상이한 활용목적이 대두됨에 따라 통제어휘에서 다국어를 수용할 수 있도록 하는 방향으로 변화가 발생하였음을 볼 수 있다.

4.5 상호운용성 확보

통제어휘 관련 표준의 변화된 또 하나의 내용으로 시소러스간의 상호운용성(interoperability)이나 시소러스와 택소노미 등 여타의 통제어휘 도구간의 상호운용성을 확보하기 위한 내용이 표준의 주요내용으로 포함되고 있다는 점이다.

시소러스 관련 상호운용성은 두 가지 관점에서의 요구에 기반하고 있다. 우선 초기 전문가들은 인쇄형 시소러스를 보고 적합 용어를 선정하여 검색시스템에 입력하는 방식으로 시소러스를 활용하였고 이 과정에서의 입력의 번거

www.kci.go.kr

로움과 오류발생을 없애기 위해 시소러스와 검색시스템의 연동에 대한 요구가 발생하였다(Aitchison and Clarkes 2006). 이러한 요구는 시소러스와 검색시스템을 연동하여 검색 시 시소러스의 특정 용어를 복사하여 붙여 넣을 수 있는 'copy & paste' 기능이나 시소러스 상의 특정 용어를 클릭하면 검색시스템이 검색을 실행하도록 하는 방식으로 시스템에 반영되었다. 이러한 상호운용성의 제공은 초기 정보검색시스템과의 연동에서 시소러스의 적용이 다양한 시스템으로 확장됨에 따라 점차 컨텐츠관리시스템(CMS; Contents Management System), 웹 출판시스템(Web Publishing Software) 등 여러 관련 시스템과의 상호운용성으로 확장되었다.

두 번째 관점에서의 통제어휘 상호운용성에 대한 요구는 정보이용 환경의 변화에서 비롯되었다. 초기 시소러스 등 통제어휘는 기본적으로 특정 주제분야에서 특정한 정보원에 대한 조직화와 검색을 위해 독점적으로 활용하고자 하는 것에 개발목적이 있었다. 네트워크와 정보처리기술의 발전에 따라 기존에 독점적 사용을 목적으로 하였던 특정 주제분야의 시소러스를 활용하여 상이한 주제분야의 정보자원을 일괄 검색하고자 하는 요구가 발생하였다(Nielsen 2004). 이에 따라 최근의 통제어휘 관련 표준은 각 영역에서 제한적으로 사용되던 통제어휘를 일괄 검색의 도구로 활용하고자 하는 요구를 지원하기 위해 시소러스 구축 시 상호운용성을 확보하는 방안을 제시하고 있다.

ANSI의 경우에는 통제어휘 상호운용성에 대한 관련 당사자들의 요구를 정리하고 상호운용성에 영향을 미치는 요인으로 주제영역의 유사성과 전문성, 관계 표현 및 동의어 처리 방법 등의 요인을 제시하고 있으며 향후 이러한 상호운용성에 대한 요구가 점증할 것으로 예상하고 있다(ANSI 2005). BS 8723의 경우에는 시소러스간 혹은 시소러스와 다양한 통제어휘 도구간의 상호운용성 확보와 관련하여 'Interoperability between Vocabularies'라는 파트 4를 별권으로 구성하여 상호운용성 확보에 관련된 대한 지침을 제시하고 있다(BSI 2007b). ISO의 경우에는 표준의 표제 자체를 'Thesauri and interoperability with other vocabularies'로 하여 상호운용성의 비중을 크다는 것을 반영하고 있다. 전체 두 파트로 구성된 표준의 둘째 파트를 'Interoperability with other vocabularies'로 계획하고 있어 그 비중이 크다는 것을 확인할 수 있다. 다만 ISO의 관련 내용은 표준화 주제로 공시되었을 뿐 아직 초안의 작성 작업이 진행되고 있지 않아 구체적인 내용은 추후 확인이 가능한 상황이다.

IFLA 지침의 경우에는 상호운용성이라는 용어를 직접적으로 사용하고 있지는 않으나 다국어 시소러스의 구축을 전제로 하여 상이한 주제 영역과 언어로 구축된 시소러스간의 합병(merging), 연계(linking), 매핑(mapping)을 통한 다국어 시소러스 구축방안을 제시하고 있다. 이러한 통합과 매핑 방안은 궁극적으로 상이한 시소러스간 상호운용성 확보의 한 방안이 될 수 있으므로 제한된 범위에서 나마 상호운용성 확보를 위한 내용을 포함하고 있다고 할 수 있다(IFLA 2009).

상호운용성 확보의 필요성은 기존의 대표적 통제어휘 도구인 시소러스의 경우 독립된 특정한 데이터베이스를 검색하기 위한 전용도구로

www.kci.go.kr

활용되었던 것에 비해 하나의 시소러스를 통해 관련된 모든 정보자원을 일괄 검색하고자 하는 요구가 반영된 것을 의미한다고 할 것이다. 즉 통제어휘 도구의 용도가 상이한 정보자원을 일괄 검색하는 과정에서 사용되는 도구로 그 성격이 크게 변화하였음을 의미하는 것으로 해석할 수 있다. 상호운용성의 확보를 위해서는 특정 시소러스에 포함된 내용을 다른 시소러스에서 동일한 의미로 인식할 수 있고, 필요 시 관련 내용을 합병하거나 교환해야 할 필요가 있다. 이러한 요구에 따라 ISO 25964-1은 데이터모형과 통신규약(protocol)을 정의하고 있으며 데이터의 교환을 위한 XML 스키마와 관련된 정보를 부록으로 포함하고 있다.

5. 결 론

본고에서는 정보조직화와 검색과정에서 효율성 제고를 위해 사용되어온 통제어휘 도구와 관련된 주요 표준의 변화를 살펴보고 변화의 공통점을 파악해 보았다. 이러한 과정을 통해 1980년대 중반 이후 정체되어 있던 관련 분야의 표준 제정을 위한 연구가 최근 들어 활발히 진행되고 있음을 확인할 수 있었다. 통제어휘 관련 표준의 제・개정은 기존에 관련 표준을 제정하였던 국제표준화기구 ISO, 국가표준기구인 ANSI, BSI 등과 더불어 웹과 관련된 표준을 제정하고 있는 사실표준기구인 W3C에서도 활발히 추진되고 있다. 2000년대 들어 통제어휘 관련 표준의 제・개정을 위한 노력이 활발히 경주되고 있다는 것은 변화된 정보환경 아래에서도 시소러스를 대표로 하는 통제어휘 도구가 유용하다는 것을 재인식한 결과이다.

관련 표준들의 변화내용에서 몇 가지 공통된 변화를 파악해 볼 수 있었다. 통제어휘 관련 표준의 주요 변화로 1) 전자적 형태의 자원까지를 포함하는 통제어휘 도구 적용 대상자원의 확장, 2) 동의어 순환목록, 택소노미, 온톨로지 등 시소러스 이외의 다양한 통제어휘 도구의 수용, 3) 표현형식의 다양화와 이를 통한 직관적 관계 제시 방안의 제공, 4) 통제어휘의 다국어 수용, 그리고 5) 상이한 통제어휘 도구간의 상호운용성의 확보 등을 확인할 수 있었다. 이외에도 개별 표준에서 구체화 되지는 못하였지만 그 개념이 도입된 항목들로 패싯 주제분석, 용어 전조합의 확대 적용, 그리고 개념간의 다양한 관계 정의 등과 같은 항목도 중요한 변화로 제시할 수 있다.

주요 표준들에서 발생한 공통된 변화는 오늘날의 정보환경에서 시소러스 등 통제어휘 도구가 수행해야 하는 기능상의 요구와 그 기능들을 수행하기 위한 통제어휘 도구의 속성들을 반영한 결과라 할 수 있다. 통제어휘 관련 표준의 변화를 분석하여 봄으로써 향후 통제어휘 도구의 구축과 활용 과정에서 우리가 염두에 두어야 하는 몇 가지 시사점을 도출할 수 있다. 첫째, 통제어휘 도구의 적용 대상자원을 확장할 수 있다는 점을 염두에 두어야 할 것이다. 이러한 확장은 관련 표준에서 대상자원을 도큐멘트가 아닌 정보자원, 내용객체 등으로 확장한 것으로부터 얻을 수 있는 시사점이라 할 것이다. 둘째, 대상 자원의 확장에 따라 시소러스 이외의 다양한 통제어휘를 활용하는 방안을 고려할 필요가 있다는 점이다. 기존의 대표적인 통제어휘 도구였던 시소러스 이외에도 변화된

www.kci.go.kr

웹 환경에서는 선택목록이나 동의어 순환목록 등 다양한 통제어휘 도구의 활용도 적극 수용할 필요가 있다. 셋째, 개념간의 관계 표현을 강화할 필요성이 있다는 점이다. 기본적으로 관계 표현의 필요성은 개념의 조합만으로는 원하는 검색결과를 얻을 수 없으므로 원하는 검색결과를 얻기 위해서 개념간의 관계를 검색과정에 반영할 필요가 있었기 때문이다. 개념간의 관계 표현 강화는 개정된 표준들이 다양한 관계 표현 방안을 도입한 것과 더불어, 구체화 되지 못했으나 패싯 분석의 개념을 도입한 표준의 변화를 통해서도 유추해 볼 수 있는 사항이다. 넷째, 통제어휘 도구의 용도가 변화하고 있다는 점이다. 기존의 통제어휘 도구가 기본적으로 특정 주제영역 전용으로 독점적으로 활용되었던 것에 비해 이제는 상이한 정보자원을 일괄 검색하는 과정에서 활용되는 것으로 용도가 변화하고 있다. 통제어휘 도구가 일괄 검색 지원도구로 기능할 수 있기 위해서 기존에 관련된 분야에서 개발된 시소러스를 합병, 매핑, 연계할 요구가 발생하였고 이렇게 상이한 시소러스를 활용하기 위해 상호운용성이 강조되고 있는 것이다. 상호운용성 확보 요구는 표준의 내용에 통제어휘 데이터모형, XML 스키마 그리고 프로토콜 관련 항목을 규정한 것에서도 확인할 수 있다. 또한 일괄 검색을 지원하는 도구로의 용도 변화는 분야가 상이한 시소러스간의 상호운용성과 더불어 상이한 언어로 구축된 시소러스간의 상호운용성 확보 필요성을 제기하였고 이에 따라 통제어휘 도구 구축, 활용 시 상이한 다국어를 수용하여야 함을 시사하여 준다. 다섯째, 초기 통제어휘 도구의 이용자는 소수의 전문가 집단에 국한되었던 것에 비해 이제는 통제어휘에 익숙하지 않은 다수의 비전문 이용자가 이용할 가능성이 크다는 점을 염두에 두어야 할 것이다. 한 단계 더 나아가 인간 비전문가뿐 아니라 기계추론의 과정에서도 통제어휘 도구가 활용될 수 있음을 염두에 두고 이를 지원할 수 있도록 해야 한다. 여섯째, 기본적인 시사점으로 통제어휘 도구의 구축, 관리를 위한 보조 소프트웨어 등 정보기술을 최대한 활용할 필요가 있다는 점이다. 한 예로 컴퓨터 화면이나 웹 화면을 통해 다양한 형식으로 도식화한 용어간의 관계를 제공하려면 디지털 형태로 구축된 통제어휘 데이터가 필요하며 이를 위해서는 구축 및 관리 과정에서 소프트웨어 관리 도구의 활용을 전제로 해야 한다. 오늘날의 업무환경을 감안하면 통제어휘 도구를 수작업으로 관리하는 경우는 찾아보기 쉽지 않을 것이지만 관리상의 효율성 제고를 위해서도 구축 및 관리 과정에서 소프트웨어 관리 도구의 활용이 반드시 필요하다. 이를 위해 각 표준에서는 이러한 보조 소프트웨어가 갖추어야 하는 속성과 선택을 위한 평가기준까지 정의하고 있다.

위와 같은 시사점들은 각 항목들이 별도로 독립된 것이 아니고 상호 연관된 부분이 있다. 예를 들어 전자적 자원까지로 적용대상을 확장하는 과정에서 기존의 대표적인 통제어휘 도구인 시소러스 이외에도 동의어 순환목록과 같은 새로운 형태의 통제어휘 도구를 적용한다든지, 동의어 순환목록의 구축에 있어 기존에 구축된 관련 선택목록이나 시소러스를 활용하기 위한 상호운용성 확보가 필요하다든지 하는 것은 이러한 시사점들이 상호 연관되어 있음을 보여주는 한 예가 될 것이다.

본고를 통해 통제어휘 관련 표준의 변화를

www.kci.go.kr

분석하고 변화를 통해 도출해 볼 수 있는 시사점을 고찰해 보았다. 통제어휘를 구축, 운용하는 과정에서 이러한 표준의 변화가 의미하는 통제어휘 도구 관련 요구와 기능상의 변화를 반영하여야 할 것이다. 오늘날 통제어휘에 요구되는 기능과 속성에 기반하여 통제어휘 도구를 구축하고 운용한다면, 통제어휘는 정보의 조직과 검색의 과정에서 더욱 효과적인 도구로 활용될 수 있을 것이다.

참 고 문 헌

[1] 정도헌, 김태수. 2003. 시소러스를 기반으로 한 온톨로지 시스템 구현에 관한 연구. 『정보관리학회지』, 20(3): 155-176.

[2] 조현양, 남영준. 2004. 시소러스와 온톨로지의 상호 호환성에 관한 연구. 『정보관리학회지』, 21(4): 27-47.

[3] Aitchison, J. & Clarkes, S. D. 2006. "The thesaurus: A historical viewpoint, with a look to the future." *Cataloging & Classification Quarterly*, 37(3): 5-21.

[4] ANSI. 1974. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*(ANSI/NISO Z39.19-1974).

[5] ANSI. 2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*(ANSI/NISO Z39.19-2005).

[6] ANSI. 2011. "FAQS." [online]. [cited 2010.12.17]. 〈http://www.ansi.org/about_ansi/faqs/faqs.aspx?menuid=1〉.

[7] BSI. 2005a. "Definitions, symbols and abbreviations." Pt.1 of *Structured Vocabularies for Information Retrieval: Guide*(BS 8723-1:2005).

[8] BSI. 2005b. "Thesauri." Pt.2 of *Structured Vocabularies for Information Retrieval: Guide*(BS 8723-2:2005).

[9] BSI. 2007a. "Vocabularies other than thesauri." Pt.3 of *Structured Vocabularies for Information Retrieval: Guide*(BS 8723-3:2007).

[10] BSI. 2007b. "Interoperabilitiy between vocabularies." Pt.4 of *Structured Vocabularies for Information Retrieval: Guide*(BS 8723-4:2007).

[11] BSI. 2008. "Exchange formats and protocols for interoperability." Pt.5 of *Structured Vocabularies for Information Retrieval: Guide*(BS DD 8723-5).

[12] Clarke, S. D. 2010. "Thesaurus standards on a converging track." *Legal Information Man-*

www.kci.go.kr

agement, 10: 43-45.

[13] Clarke, S. D., Gilchrist, A., & Will, L. 2004. "Revision and extension of thesaurus standards." [online]. [cited 2010.12.16]. 〈http://www2.db.dk/nkos-workshop/pp%20presentationer/NKOS-Stella.pdf〉.

[14] IFLA. 2009. "Guidelines for multilingual thesauri." *IFLA Professional Report*, no.115.

[15] ISO. 1985. *Documentation: Guidelines for the Establishment and Development of Multilingual Thesauri*. 2nd ed(ISO 5964-1985(E)).

[16] ISO. 1986. *Documentation: Guidelines for the Establishment and Development of Monolingual Thesaurus*. 2nd ed(ISO 2788-1986(E)).

[17] ISO. 2010. "Thesauri for Information Retrieval." Pt.1 of *Information and Documentation: Thesauri and Interoperability with Other Vocabularies*(ISO/FDIS 25964-1: 2010(E)).

[18] ISO. 2011. "ISO - Technical committees - TC 46/SC 9 - Identification and description." [online]. [cited 2011.1.15]. 〈http://www.iso.org/iso/standards_development/technical_committees/list_of_iso_technical_committees/iso_technical_committee.htm?commid=48836〉.

[19] Kwasnik, Barbara H. 1999. "The role of classification in knowledge representation and discovery." *Library Trends*, 48(1): 22-47.

[20] Nielsen, Marianne Lykke. 2006. "Thesaurus construction: Key issues and selected readings." *Cataloging and Classification Quarterly*, 37(3): 57-74.

[21] Taxonomy Warehouse. 2009. "About Taxonomy Warehouse." [online]. [cited 2010.12.23]. 〈http://www.taxonomywarehouse.com/Include_about.asp〉.

[22] Will, Leonard. 2010. "Publications on thesaurus construction and use." [online]. [cited 2011.1.23]. 〈http://www.willpowerinfo.co.uk/thesbibl.htm〉.

[23] World Wide Web Consortium. 2009. "SKOS simple knowledge organization system reference: W3C recommendation." [online]. [cited 2010.12.9]. 〈http://www.w3.org/TR//2009/REC-skos-reference-20090818〉.

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

[1] Jeong, Do-Heon & Kim, Tae-Soo. 2003. "A study on the thesaurus-based ontology system for the semantic web." *Journal of the Korean Society for Information Management*, 20(3): 155-176.

www.kci.go.kr

[2] Cho, Hyun-Yang & Nam, Young-Joon. 2004. "A study on the interchangeability between a thesaurus and an ontology." *Journal of the Korean Society for Information Management*, 21(4): 27-47.

www.kci.go.kr