정보관리연구, vol.42, no.2 2011, pp.43-65

ISO 25964를 통해 살펴본 시소러스 구축 표준에 관한 연구*

The Study on the Standards of Thesaurus Construction Based on ISO 25964

박 옥 남**
Ok-Nam Park

차 례

1. 서 론
2. 시소러스 표준 개요
3. 방법론
4. 비교분석
5. 결론 및 제언
· 참고문헌

초 록

시소러스 제작과 개발을 위한 지침을 제공하기 위하여 여러 표준이 발표되고 있으며, 지난 수십 년간 여러 시소러스가 구축되었음에도 불구하고, 시소러스 구축을 위한 가이드라인 및 표준 발전 연구의 계속성은 이루어지지 않고 있다. 최근 ISO 25964 표준 제정 프로젝트는 변화하는 정보기술 환경에 맞게 시소러스 표준을 개정하여, 표준 사이에 각기 다르게 사용되던 용어 및 약어의 사용을 일관성 있게 함으로써 시소러스 표준의 실효성을 높이고자 하는 목적으로 진행되고 있다. 이에 본 연구는 ISO 25964를 소개하고 ISO 25964가 범용되기 위하여 향상되어야 할 점을 제시하였다. 이를 위해 시소러스 표준 ISO 2788: 1986, ANSI/NISO Z39.19-2005와의 비교분석이 이루어졌다. 본 연구를 통해 시소러스 구축표준을 살펴봄으로써 표준을 기반으로 한 시소러스 구축 실제를 조력하고자 하였다.

키 워 드

시소러스 구축, 시소러스 표준, ISO 표준, ISO 25964

* 이 논문은 2011년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음.
** 한남대학교 문헌정보학과 전임강사
(Fulltime Lecturer, Dept. of Library and Information Science, Hannam University, ponda@hnu.ac.kr)
• 논문접수일자: 2010년 9월 6일
• 최종심사(수정)일자: 2010년 12월 15일
• 게재확정일자: 2011년 2월 28일

ABSTRACT

Despite a long tradition and rich discussion of thesaurus construction, more research is needed to make standards for thesaurus construction applicable to practice. Recently, in order to reflect changing information technology environment, and provide consistency of terms and abbreviations of standards, ISO committee has been planning to publish ISO 25964. Therefore, the study introduces ISO 25964, reveals main focal points of ISO 25964, and finally discusses suggestions to improve ISO 25964. The study employs qualitative content analysis of ISO 25964, ISO 2788: 1986, and ANSI/NISO Z39.19 - 2005.

KEYWORDS

Thesaurus Construction, Thesaurus Standards, ISO Standards, ISO 25964

1. 서 론

시소러스 제작과 개발을 위한 지침을 제공하기 위하여 여러 표준이 1970년대와 1980년대에 발표되었다(Clarke 2010). 국제표준으로 ISO 2788: 1986 Documentation - Guidelines for the establishment and development of monolingual thesauri(문헌 - 단일어로 제공되는 시소러스 제작과 개발을 위한 지침)와 ISO 5964: 1985 Documentation - Guidelines for the establishment and development of multilingual thesauri(문헌 - 다국어 시소러스 제작과 개발을 위한 지침), 미국에서 발표한 표준으로 ANSI/NISO Z39.19: 2005 Guidelines for the construction, format, and management of monolingual controlled vocabularies (단일어로 제공되는 통제어휘 개발, 포맷, 관리를 위한 지침), 마지막으로 영국 표준인 BS 8723 Structured vocabularies for information retrieval(정보검색을 위한 구조화된 어휘)가 대표적인 것이다.

이렇게 많은 표준들이 제공되고 있으며, 효율적인 정보검색을 위해 지난 수십 년간 많은 시소러스가 구축되었음에도 불구하고, 시소러스 표준의 한계가 지적된다. 첫째, 각기 다른 표준 사이에 용어와 관점에 있어 상이한 점이 발견된다. 둘째, ISO 2788이 1986년 2차 개정되어 발표되고, ISO 5964가 1985년에 발표된 이후 표준의 개정이 적시에 이루어지지 않아 현재 시소러스 기반 정보검색의 요구사항을 제대로 반영하지 못하고 있다. 특히 시소러스 기반 정보검색은 단순한 사전적인 용어 검색뿐 아니라 장서, 데이터베이스, 시스템 간 다른 시소러스로 색인된 정보간의 상호운용성을 요구하고 있어, 이러한 사항을 반영하는 시소러스 표준 개정이 요구된다. 셋째, 표준 가이드라

인이 제공되고 있음에도 불구하고 서지 데이터베이스 정보 전문가 사이에 표준의 사용이 제한적이며, 데이터베이스 벤더와 시스템 개발자들은 심지어 이러한 표준의 존재에 대해서 인식하지 못하고 있는 경우도 많다고 지적한다. 즉, 표준이 이들의 정보요구를 제대로 반영하지 못하고 있다는 것이다(Wiliamson 1996).

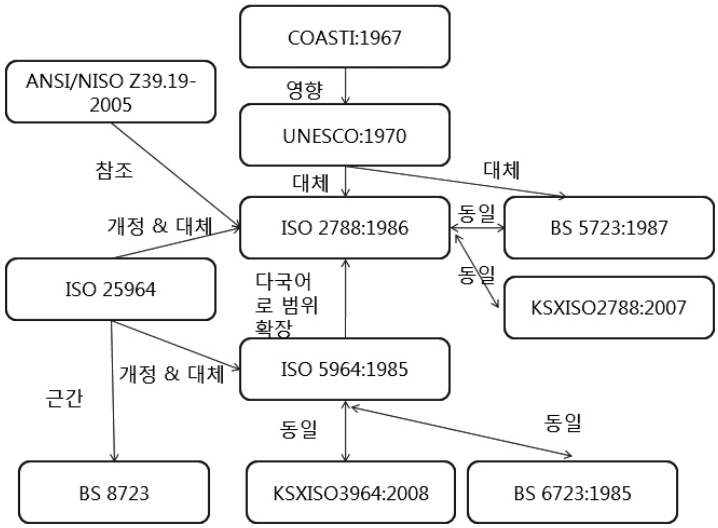
이에 2008년부터 ISO는 개정요구를 반영하며 서로 각기 다른 표준에서 언급되어온 시소러스 구축의 주요 논의사항 및 상이한 용어의 사용을 일관성 있게 재구성하여 국제적으로 통용될 수 있는 표준을 발표할 목적으로 프로젝트를 진행 중인데 그 표준이 ISO 25964이다. 이에 본 연구는 ISO 25964 표준을 소개하고, 기존 시소러스 구축의 주요 표준들과의 차이점을 분석하여 ISO 25964 표준에서 시소러스 구축에 초점을 두고 있는 바를 논의하며, ISO 25964가 범용되기 위하여 향상되어야 할 점을 제시하고자 한다.

본 연구는 가장 최근의 표준인 ISO 25964와 기존 시소러스 표준을 살펴봄으로써 시소러스 구축 논의 동향을 살펴볼 수 있으며, 최근 표준을 살펴봄으로써 표준에 근거하는 시소러스를 구축하는데 도움이 될 것이라 기대된다.

2. 시소러스 표준 개요

시소러스 구축을 위한 가이드라인 및 표준의 발전에 대해서는 Lancaster(1986)의 연구에서 간단하게 언급되었으며 Krooks and Lancaster (1993)의 연구와 Williamson(1996)의 연구에서 조금 더 심층적으로 논의되었으나, 연구의 계속성은 이루어지지 않고 있다. Krooks and Lancaster(1993)의 연구는 1959년 Dupont 시소러스에서 언급된 원칙부터 1992년 초안으로 발표된 ANSI/NISO Z39.19-1993까지 표준의 주요 배경에 대해 언급하고 있으며, Williamson (1996)의 연구는 ANSI/NISO Z39.19-1974와 ISO 2788:1974에 대해서 간단하게 언급하고 있다. 이후 많은 표준이 발표되었음에도 불구하고 이후 표준을 통해 시소러스 구축의 표준이 어떠한 특징을 가지고 발전해 왔는지에 대해서는 논의된 바가 많지 않다. 이에 본 연구는 시소러스 구축에서 활용되는 주요 표준의 배경과 관계를 논의하고자 한다(<그림 1> 참조).

초기 시소러스 가이드라인 및 표준은 시소러스가 구축이 되면서 그 필요에 의해 생기게 되었는데, 초기 시소러스는 알파벳순의 색인어 나열로 주제명 표목표와 유사하게 간단한 용어와 구조로 구성되는데, 이러한 초기 시소러스 구축이 COSATI 지침(1967)생성에 영향을 주었으며, COSATI 표준은 초기 시소러스 표준의 발전에 영향을 준 사람인 Derek Austin이 발표한 UNESCO Guidelines for the Establishment and Development of Monolingual Scientific and Technical Thesauri(UNESCO 단일어 과학기술분야 시소러스 구축지침)(1970)에 영향을 주었으며, 이 지침은 후에 ISO 2788:1986과 BS 5723:1987로 대체되었다(Aitchison,



〈그림 1〉 시소러스 표준 관계

Gilchrist, and Bawden 2000). 국제 표준인 ISO 2788과 영국표준인 BS 5723은 동일하다. 미국표준인 ANSI/NISO Z39.19:2005은 전체-부분관계를 계층적 관계로 다루고 있다는 점에서 ISO 2788:1986를 참조하였으며 ISO 2788:1986에 관리 가이드라인을 추가하였다. 그러나 ANSI/NISO Z39.19:2005는 ISO 2788:1986의 부분은 아니다(Krooks and Lancaster 1993). 이 미국표준은 시소러스의 화면 디스플레이와 시소러스 관리 시스템을 다룬 첫 번째 표준이다. 다국어 시소러스를 다루고 있는 국제표준은 ISO 5964이며, 1985년에 초판이 발행되었고, 영국표준인 BS 6723:1985와 동일하다. BS 8723은 영국표준으로 시소러스 구축 시 기술의 발전과 이에 따른 시소러스 구축 시 고려해야 할 사항들이 지침에 반영되지 못하고 있음이 계속적으로 지적되자, 이러한 요구를 반영하고자 만든 표준이며, 2005년부터 부분적으로 발표되고 있으나 그 사용은 제한적이다.

이와 같이 여러 표준이 발표되어 사용되고 있음에도 불구하고 이들 표준이 기술의 발전과 시소러스 구축 시 고려해야 할 사항들을 포함하지 않고 있으며(Aitcheson 2000), 패싯 시소러스 구축에 대한 지침(Williamson 1996; Spiteri 1997)이 포함되지 않거나, 표준간 상이한 원칙, 약어나 지시어의 사용 등이 나타나고 있으므로 ISO 위원회에서 시소러스 표준을 통합적으로 대체할만한 국제적 수준의 시소러스 표준의 필요성을 인식하고 표준제정을 위한 노력을 기울이고 있는데, 그것이 바로 ISO 25964이다. ISO 25964는 BS8723을 ISO표준으로 채택하기로 결정하고, ISO 2788:1986과 ISO 5964:1985를 대체하기 위한 것으로, ISO에 의해 프로젝트가 승인되어 프랑스, 독일, 캐

나다, 유럽, 미국 등 13개국이 참여하여 진행 중인 프로젝트이다.

즉, 시소러스 구축 발전과정을 통해 현재 많이 참조되는 표준은 ISO 2788:1986, ISO 5964:1985, ANSI/NISO Z39.19-2005이며, 현재 시소러스 구축과 관련된 환경의 변화에 맞게 가이드라인을 제공하기 위해 진행되고 있는 표준은 ISO 25964이다. 이 주요 표준의 내용을 조금 더 상세히 살펴보고자 한다(〈표 1〉 참조).

2.1 ISO 2788:1986

ISO 2788:1986 Documentation - Guidelines for the establishment and development of monolingual thesauri(문헌-단일어 시소러스 제작과 개발을 위한 지침)은 단일 색인기관 또는 다른 색인 기관 사이의 일관성 있는 업무를 지원하기 위한 것을 목적으로 하며, 단일어 시소러스 구축을 위한 지침을 제공하고 있다. 주요 내용으로는 어휘통제, 복합어, 관계어 정립, 디스플레이, 시소러스 관리 등을 포함한다.

ISO 2788은 1974년 국제표준기구(International Organization for Standardization, 이하 ISO)에 의해 처음 발표되었으며, 1986년에 그 개정판이 발표되었다. 1986년에 발표된 개정판은 1967년에 발표된 미국 과학 기술 분류표(American Thesaurus of engineering and scientific terms, 이하 TEST) 시소러스의 접근법인 알파벳순 표현법과 유럽 컴파일러(Euro-

〈표 1〉 국제표준

출처	표준	비고
미국표준	ANSI/NISO Z39.19:2005 Guidelines for the construction, format and management of monolingual controlled vocabulary	2005년에 제4판이 발행되었음.
국제표준	ISO 2788:1986 Guidelines for the establishment and development of monolingual thesauri	표준 유효하나 개정된 지 오래되었음.
	ISO 5964:1985 Guidelines for the establishment and development of multilingual thesauri	표준 유효하나 개정된 지 오래되었음.
국제표준 초안	ISO/FDIS 25964-1 Thesauri and interoperability with other vocabularies	ISO2788과 ISO5694를 개정하는 것으로, 영국표준인 BS 8723을 ISO표준으로 채택하기 위함이며 시소러스 표준 개정요구를 반영하기 위함.
영국표준	BS 8723: Structured vocabularies for information retrieval	5개 부분으로 구성되었으며 ISO 25964에 의해 국제표준으로 재구성됨.
한국표준	KSXISO2788 문헌-단일어 사전의 제작과 개발을 위한 지침	2002년에 처음 고시된 후 2007년에 변경되었음.
	KSXISO5964 문헌-다국어 시소러스의 제정과 개발 지침	2003년에 처음 고시된 뒤 2008년에 변경되었음.

pean compiler) 분류상의 접근법을 혼합한 것이다(Lancaster 1986).

2.2 ISO 5964:1985

ISO 5964: 1985 Documentation - Guidelines for the establishment and development of multilingual thesauri(문헌 - 다국어 시소러스 제작과 개발을 위한 지침) 표준은 ISO 2788과 기본적인 지침은 같이 하고 있으나, 다국어 시소러스 개발 지침에 초점을 두고 있다. ISO 2788에서 제시하고 있는 용어의 형태, 시소러스 관계, 평가와 유지를 위한 관리지침과 같은 일반적인 시소러스 구축 절차나 지침이 ISO 5964에도 동일하게 적용되며, 디스크립터와 비디스크립터를 구분하는 지침이 ISO 2788과 구분된다.

2.3 KSXISO2788

한국표준인 KSXISO2788은 단일어 사전의 제작과 개발을 위한 지침으로 ISO 2788:1986을 번역한 것으로 내용 역시 ISO 2788:1986과 일치한다.

2.4 KSXISO5964

KSXISO5964는 한국표준으로 다국어 시소러스 제정과 개발지침을 제공하고 있으며, ISO 5964의 번역본으로 내용 역시 ISO 5964:1985와 일치한다.

2.5 ANSI/NISO Z39.19:2005

ANSI/NISO Z39.19:2005 Guidelines for the construction, format and management of monolingual controlled vocabularies(단일어로 제공되는 통제어휘 개발, 포맷, 관리를 위한 지침)은 미국에서 발표한 표준으로, 현재 4번째 개정판이다. 본 표준은 단일어 통제어휘 구축 및 유지에 관한 지침을 제공하며, 주요 내용으로 콘텐트, 디스플레이, 테스트 등을 포함한다.

ANSI/NISO Z39.19 초판은 시소러스 구축 분야 전문가들로 구성된 위원회에 의해 1974년과 1980년에 차례로 개정되었다. 1988년부터 사도요한 대학(St. John's University)의 문헌정보학 교수인 Weinberg의 주도하에 세 번째 개정이 시작되어 1993년에 고시되었는데, 주요 논의사항은 급변하는 전자정보환경을 반영하기 위함이었다. 1993년에 발표된 개정판은 이용자에 의해 가장 자주 참조되어오던 미국 정보표준화기구(National Information Standards Organization, 이하 NISO)에서 발표한 표준 중 하나로 이후에 여러 환경적인 변화를 조금 더 적극적으로 반영하고자 4번째 개정이 진행되었다.

제4판의 주요 내용은 첫째, 시소러스가 사용되는 환경을 문서(documents)에서 콘텐트 객체(content objects)로 변형하였다. 둘째, 용어

의 유형을 기존에는 시소러스만으로 한정시켰던 것에 반해 통제어휘 리스트, 동의어 리스트, 탁사노미로 확대시켰다. 셋째, 표준 역시 프린트 형태는 물론 웹 문서의 형태로 발행하여 표준에 대한 접근성 및 활용성을 높였다. 마지막으로, 단일어 용어뿐 아니라, 다국어 용어, 상호운용성, 패싯 등의 이슈를 다루고 있다. 2005년 개정은 미국의 윌슨 재단(The H.W.Wilson Foundation), 게티 기금 프로그램(The Getty Grant Program), 국립 의학도서관(the National Library of Medicine)에 의해 지원되었다.

2.6 BS 8723

영국표준인 BS 8723 Structured vocabularies for information retrieval(정보검색을 위한 통제어휘 지침)은 총 다섯 부분으로 구성되는데 첫 번째는 정의, 약어를 다루고 있으며, 두 번째는 시소러스의 개념 및 정의를 다루고 있고, 세 번째 부분은 분류시스템(classification schemes), 주제명 표목표(subject heading lists), 탁사노미(taxonomies), 온톨로지(ontologies), 시맨틱 네트워크(semantic networks)의 주요 특징 및 구축 방법을 다루고 있다. 네 번째는 다른 구조화된 어휘간의 상호운용성을 다루고 있다. 마지막 다섯 번째는 상호운영을 위한 포맷 및 프로토콜을 다루고 있으므로, 있으므로, 검색엔진, 콘텐트 관리 시스템, 웹 출판 소프트웨어 등과의 상호운용성, 또는 Z39.59, SKOS, Topic Map, MARC 등의 데이터 포맷 및 프로토콜에서 데이터 교환을 위한 가이드라인을 다루고 있다(Zeng 2005).

2.7 ISO 25964

ISO 25964 Thesauri and interoperability with other vocabularies(시소러스 및 상호운영성에 대한 문서)는 기존에 국제표준으로 사용되어오던 ISO 2788 및 ISO 5964를 개정하기 위하여 2008년 5월부터 진행되어온 프로젝트이다. ISO 25964는 과거의 시소러스는 인간에 의해 유지되는 시스템인 것에 반해 현재의 시소러스는 콘텐트 관리 시스템, 검색엔진, 소프트웨어에 의해 개발·유지되는 등 시소러스 환경이 변화하고 있다는 인식에서 출발하고 있다. 이를 위해 본 표준은 1판은 여전히 인적 자원이 색인어 선택 및 검색어 선택에 관여하고 있음을 고려하여 시소러스 개발 및 유지를 위한 지침을 제공하고 있으며, 2판에서는 다른 시스템이나 구조적 용어체계를 정의하고, ISO 25964 제1판에 기반하여 구축된 시소러스와 다른 구조적 용어체계 사이의 상호운용성을 확보하기 위한 지침을 제공하기 위하여 표준 제정 작업이 진행 중이다.

ISO 25964 제정은 정보와 문서(Information and documentation)를 담당하고 있는 ISO 기술위원회 46분과 중 식별 및 기술(Identification and description)을 담당하고 있는 9 하위 분과위원회에서 진행하고 있으며, 미국, 영국, 유럽, 한국 등의 13개 국가의 전문위원들

이 참여하고 있는 프로젝트이다. ISO 25964는 이를 위해 BS 8723에 기반을 두고 있는데, BS 8723의 다섯 부분 중에서 정의, 약어를 다루고 있는 첫 번째 부분, 시소러스의 개념 및 정의를 다루고 있는 두 번째 부분을 재구성하여 하나의 부분으로 구성하고, 시소러스 및 시소러스 외 다른 어휘를 다루고 있는 세 번째 부분, 어휘간의 상호운용성을 다루고 있는 네 번째 부분, 상호운영을 위한 포맷 및 프로토콜을 다루고 있는 다섯 번째 부분을 재구성하여 다른 하나의 부분으로 구성하여, 총 두 부분으로 구성된 표준을 계획하고 있다. 제 1판은 현재 국제표준 최종 초안(Final Draft International Standard, 이하 FDIS)으로 출판되어 여론을 수집・반영하고 있으며 제2판은 현재 초안(Working Draft) 작업이 진행 중이다.

3. 방법론

본 연구는 시소러스 표준 간의 유사점 및 차이를 분석하기 위하여 질적 내용분석(Qualitative Content Analysis)방법을 채택하였다. 분석대상으로는 ISO 25964가 개정하고자 하는 표준 ISO 2788:1986와 미국에서 발표한 표준인 ANSI/NISO Z39.19-2005, ISO 25964를 수집하여 비교 분석하였다. 이 세 가지 표준으로 분석대상을 한정한 것은 가장 국제적으로 통용되고 있는 표준이기 때문이다. 또한 본 연구는 단일어 시소러스 구축에 대한 지침만을 비교 분석하였다.

질적 내용분석을 위해 지시적 내용분석(directed content analysis)을 채택하였다. 지시적 내용분석은 연구자가 분석하고자 하는 항목을 연구목적에 따라 문헌이나 이론을 기반으로 도출하고 도출된 항목에 따라 데이터를 분석하는 것이다(Hsieh and Shannon 2005). 지시적 내용분석은 기존 문헌이나 이론의 총체적인 분석을 바탕으로 도출된 항목을 기반으로 한다는 점에서 체계적이고 일관성 있는 데이터의 분석이 이루어질 수 있다는 것이 장점이다. 질적 내용분석을 위해 두 가지 방법을 선택할 수 있는데, 첫 번째 방법은 전체 문서를 비교하는 방법(text-by-text)과 두 번째는 비교될만한 대상을 비교(point-by-point)하는 것인데, 본 연구에서는 대상비교법을 채택하였다. 비교대상을 선택하기 위하여 표준 간 명백하게 유사점 및 차이점을 비교할 수 있는지, ISO 25964에서 표준제정의 초점으로 논의되었던 사항은 무엇이며, 그 전 표준과는 어떠한 차이가 있는지를 고려하였으며 또한 비교분석을 위해 ISO 25964 및 기존 표준의 목차를 분석한 결과 다음의 8가지 사항이 채택되었다 - 서론(introduction), 범위(scope), 약어 및 기호(abbreviation and symbols), 절차(steps), 시소러스의 개념화(thesaurus), 목차(table of contents), 상세사항(details).

본 연구는 양적 내용분석은 포함하지 않았다. 그러므로 단어의 빈도수, 문서 내에서 단어의 위치 등으로 파악할 수 있는 문서 내에서 단어

의 중요성, 단어 사이의 관계 등은 조사항목에서 제외되었으며 이는 향후 연구과제로 남는다.

4. 비교분석

4.1 서론(Introduction)

ISO 2788은 서론에서 주제색인의 주요 기능을 문서의 주제를 정하고, 주제를 설명할 수 있는 용어를 선택하고, 용어로 표현되는 개념 사이의 관계를 표현하는 것이라고 명시하였다. 이 중에서 ISO 2788은 용어선택과 관련된 사항을 다루고 있는데, 용어통제 및 색인어 사이의 용어관계를 표현하는 것 등이 그 대표적인 것이라고 명시한다. 즉, ISO 2788은 서론에서 검색의 효율성을 향상시키기 위하여 문서색인에 사용될 시소러스의 기본지침을 제공하는 것이라고 밝히고 있다. ANSI/NISO Z39.19도 그 맥락을 함께하고 있으며 본 표준은 정보저장 및 검색 시스템, 웹 내비게이션 시스템 등의 효율성을 향상시키기 위한 것이며, 색인의 일관성을 성취하기 위한 지침이라고 명시하고 있다. ISO 2596은 시소러스 제작 및 사용환경이 인쇄본 중심에서 전자적 도구로 변화하고 있으므로, 이를 반영하기 위한 지침을 제공하는 것이라고 밝히고 있다. 즉, ISO 2788에서 ANSI/NISO Z39.19, ISO 25964로 변화할수록 시소러스 구축의 전자적 환경을 강조하고 있는 것을 파악할 수 있다.

4.2 범위(Scope)

ISO 2788은 통제어휘의 원칙 정도만 설명했던 것에 비해 ANSI/NISO Z39.19는 통제어휘의 목적, 개념, 구조 등 조금 더 상세하게 부연한 것이 그 특징이다. 특히 개념부분에서 ANSI/NISO Z39.19와 ISO 2788은 조금 다르게 접근하고 있는 것을 볼 수 있는데, ISO 2788은 문서(document)를 색인 하는 것으로 통제어휘의 개념을 언급하고 있다. 여기서 문서(document)는 "목록이나 색인에 적용될 수 있는 인쇄본이나 다른 형태의 아이템"이라고 정의하고 있다(ISO 2788 1986, 2). 이에 반해 Z39.19는 통제어휘의 개념을 콘텐트 객체(content objects)와 관련된 메타데이터를 표현하기 위하여 부여되는 용어로 정의하고, 콘텐트 객체를 정보 검색 시스템에 포함되는 아이템으로 인쇄자료, 녹음자료, DVD, 영화, 또는 전자형태 등까지 포함하는 일차 콘텐트 객체(primary content objects)와 일차 콘텐트 객체를 기술하는 메타데이터인 이차 콘텐트 객체(secondary content objects)까지 포함하는 개념으로 확장시켰다.

ISO 25964는 정보검색을 위한 시소러스 개발 및 유지를 목적으로 하고 있으나, 정보자원(information resources)의 정보를 검색하기 위한 것으로 개념화하고 있는데, 정보자원은 "미디어에 상관없으며(텍스트, 소리, 이미지, 멀티미디어와 같은) 지식기반 시스템 및 포털, 서지 데이터베이스, 박물관, 멀티미디어 컬렉

션, 이들을 기술하는 아이템을 포함한다(ISO 25964 2009, 16)"고 규정하고 있어, 단지 물리적인 형태뿐만 아니라, 활용 영역까지 언급하고 있는 것이 특징이다. 이렇게 통제어휘의 개념이 문서, 콘텐트 객체, 정보 자원 및 정보로 확장된 것은 시소러스 의 활용이 정기간행물 기사, 기술 보고서, 뉴스기사 등 일반적으로 인쇄된 자원의 색인에 주로 사용되던 것을 벗어나 문서의 개념이 특허, 화학구조, 비디오, 음악자료 등 여러 다양한 형태로 확장되었으므로 이러한 변화를 반영하고자 한 것이며, 동시에 시소러스가 사용되는 환경을 확장하여, 시소러스 표준의 활용영역의 확장을 도모하고자 한 것이다.

4.3 약어 및 기호 (Abbreviation and Symbols)

〈표 2〉와 같이 ISO 2788에서는 한정된 약어를 사용하여 시소러스의 관계를 표현하고 있는데, 용어의 정의, 이력주기, 사례관계 상위어, 사례관계 하위어의 관계를 표시할 수 있는 약어를 제시하지 않고 있는 것이 그 예이다. 이에 반해 ANSI/NISO Z39.19에서는 약어의 사용을 확대하여 시소러스에서 풍부한 용어의 관계를 표현하는 것을 가능하게 하였는데, 예를 들어, 이력주기를 표시하도록 하여 용어 변화에 대한 근거를 남기게 하였으며, 범위주기 참조를 표시하도록 하여 용어사용 범위의 근거를 표시하도록 하였다. 또한 사례관계 하위어나 상위어를 표시하도록 하여, 보통명사로 표현되는 사물이나 사건의 범주와 개별 예 사이의 관계를 계층적 관계로 표시하는 것을 가능하게 하였다. 그 예는 아래와 같다.

예) 사례 관계
산악지역
NTI 알프스
NTI 히말라야

ISO 25964에서는 기존의 표준에서는 제시하지 않았던 약어를 정의함으로써 용어의 기술이나 관계표시의 다양성을 추구하였으며, 이는 기존의 표준에서는 제시되지 않았지만, INSPEC 시소러스, JICST 시소러스, ROOT 시소러스, Women's 시소러스 등 시소러스 구축에서 상이하게 제시되던 용어의 기술이나 관계표시의 약어를 통일하여 실제의 일관성을 도모하기 위함이다. 또한 ISO 25964와 기존의 표준과의 차이점을 자세히 살펴보면, ISO 25964는 기존의 표준에서 약어를 나열했던 것과 대조적으로, 약어를 기술요소(descriptive elements), 코드(codes), 관계(relationships) 범주로 구분하였다. 기술요소는 하나의 용어 자체를 설명하기 위한 약어를 범주화하기 위해 사용되었고, 관계범주는 용어 간 관계를 표현하기 위한 약어를 구분하기 위함이며, 코드는 개별 개념이나 개념의 그룹화를 범주화하기 위함이다.

기술요소 군에서 ISO 25964가 기존 표준들

〈표 2〉 표준간 약어 비교

약어	정의	ISO 2788	ANSI/NISO Z39.19	ISO 25964
Descriptive Elements	**기술요소**			○
SN	범위주기(Scope Note)	○	○	○
DEF	정의(Definition)			○
HN	이력주기(History note)		○	○
X SN	범위주기참조(See Scope Note for...)		○	
Codes	**코드**			
SC	주제구분(Subject Category)			○
CC	개념 코드 표기법(Concept code or notation)			○
Relationships	**관계**			
USE	우선어 안내	○	U(Use) ○	○
UF	비우선어 안내	○	○	○
USE+	2개 이상의 우선어 사용지시			○
UF+				○
TT	최상위어(Top Term)	○	○	○
BT	상위어(Broader term)	○	○	○
BTG	속관계 상위어(Generic Broader term)	○	○	○
BTI	사례관계 상위어(Instantial Broader term)		○	○
BTP	부분관계 상위어(Partitive Broader term)	○	○	○
NT	하위어(Narrower term)	○	○	○
NTG	속관계 하위어(Generic Narrower term)	○	○	○
NTI	사례관계 하위어(Instantial Narrower term)		○	○
NTP	부분관계 하위어(Partitive Narrower term)	○	○	○
RT	관계어(Related term)	○	○	○

과 구분되는 점은 범위주기와 정의를 구분하고 있는 것이다. ISO 25964에서 범위주기는 상·하위 관계만으로는 개념의 의미가 충분히 전달되지 않을 때 개념의 의미나 용법을 명확하고 일관성 있게 유지하기 위하여 사용된다. 즉, 범위주기는 개념의 사용 영역을 명확히 하기 위하여 사용되므로 용어의 용례를 제공하여 색인자나 검색자의 활용을 도울 수 있다. 이에 반해 정의는 용어 자체의 개념을 정의하기 위한 것으로, 범위주기만으로 충분히 그 의미를 반영할 수 없다고 판단될 때, 사전적 정의를 제공하기 위하여 사용되며 정의의 근거를 밝혀야 한다. ISO 25964에서 제공하고 있는 예는 다음과 같다.

예)

일루미네이션(Illuminations)

SN장식용 전등과 옛날 책에 손으로

한 삽화는 물론 필사본의 삽화를 포함한다.

명암법(chiaroscuro)

DEF 빛과 그림자로만 표현되는 회화 스타일(옥스포드 영어사전)

또한 ANSI/NISO Z39.19와 마찬가지로 이력주기를 유지하여 용어의 변화를 추적하도록 하였다. 그러나 ANSI/NISO Z39.19에서는 이력주기에 대한 설명이 생략한 것에 반해, ISO 25964에서는 이력주기 정의, 사용법 및 용례를 자세히 설명하여 표준 사용자들의 편의를 도모하였다.

코드군은 기존의 표준에서 존재하지 않던 것으로, SC는 주제구분 약어로 특정한 주제에 적합한 개념군을 표현하기 위해 적용되는 코드를 지칭하며, CC는 콘셉트 코드로 특정 개념에 적용되는 코드를 지칭한다. 예를 들어, 교육이나 사회과학 등의 주제를 구분하기 위하여 주제구분코드를 적용할 때 사용되는 약어로 SC를 제시하였으며, 우선어에 특정 코드를 부여하기 위하여 사용되는 약어로 CC를 제시하였다. 이 코드군은 기존의 표준에서는 존재하지 않았으나, INSPEC 시소러스, JICST 시소러스, ROOT 시소러스, Women's 시소러스에서 사용되어오던 것으로(이화여자대학교 2000) 개별 시소러스에서 상이하게 사용되던 약어를 통일하여 실제의 일관성을 도모하였다. 다음은 Women's 시소러스에서의 코드군의 사용약어와 예이다.

예)

이타주의(altruism)

SG(주제구분) 언어, 문학, 종교, 철학(Language, Literature, Religion, and Philosophy)

DG(개념코드) 문화와 철학운동(Cultural and Political Movements)

마지막으로, 관계군에서 기존의 표준과 구분되는 것은 USE+와 UF+의 사용인데, USE+는 복합어가 우선어로 표현되는 것이 적절하지 않을 때 두 개 이상의 우선어로 표현하기 위하여 사용되며, UF+는 그 역의 관계를 표현하기 위한 것이다. 표준에서 제시하는 용례는 아래와 같다.

예)

- 탄광업(coal mining)

  USE 석탄(coal)

  \+ 광업(mining)
- 석탄(coal)

  UF+ 탄광업(coal mining)
- 광업(mining)

  UF+ 탄광업(coal mining)

기호의 사용에서는 기존의 표준에서는 〈표 3〉에서와 같이 기호 제시가 이루어지지 않았으나, ISO 25964는 우선어(->), 비우선어(=), 상위어(<), 하위어(>), 관계어(-)정도로 제한적인 기호 제시가 이루어진 것이 비교할만하

〈표 3〉 표준간 기호 비교

약어	정의	ISO 2788	ANSI/NISO Z39.19	ISO 25964
Descriptive Elements	**기술요소**			
SN	범위주기(Scope Note)			
DEF	정의(Definition)			
HN	이력주기(History note)			
X SN	범위주기참조(See Scope Note for...)			
Codes	**코드**			
SC	주제구분(Subject Category)			
CC	개념 코드 표기법(Concept code or notation)			
Relationships	**관계**			
USE	우선어 안내			-〉
UF	비우선어 안내			=
USE+	2개 이상의 우선어 사용지시			
UF+				
TT	최상위어(Top Term)			
BT	상위어(Broader term)			〈
BTG	속관계 상위어(Generic Broader term)			
BTI	사례관계 상위어(Instantial Broader term)			
BTP	부분관계 상위어(Partitive Broader term)			-〈
NT	하위어(Narrower term)			〉
NTG	속관계 하위어(Generic Narrower term)			
NTI	사례관계 하위어(Instantial Narrower term)			
NTP	부분관계 하위어(Partitive Narrower term)			〉-
RT	관계어(Related term)			-

다. 이러한 기호의 사용은 기존 표준에서는 제시되지 않았으나, 구축되어온 시소러스에서 상이하게 제시되던 것에 대해 통일된 기준을 제시했다는 점에서 그 의의가 있다 하겠다. 그럼에도 불구하고 최상위어, 사례관계어, 속관계어 등에는 기호를 제시하고 있지 않으므로 이에 대한 제시가 보완되어야 할 것이다.

4.4 시소러스 구축 절차(Steps)

ISO 2788은 시소러스 구축 절차에 대해서는 언급을 하지 않고 있으며, ANSI/NISO Z39.19는 11장에서 시소러스 전반과 관련된 주의사항을 제안하고 있는데, 이를 통해 시소러스 절차를 파악할 수 있다. 이 표준은 구축, 테스트 및 평가, 유지, 관리 시스템 운용의 순서로 제시하고 있는데, 각 절차에 따라 고려해야 할 사항을

제시하고 있다. ANSI/NISO Z39.19의 경우는 조금 더 세부적인 방법론을 이론적으로 제시하고 있는 것이 특징이다. 예를 들어, 구축방법으로 하위 용어부터 수집하여 상향식으로 계층적 용어를 구축해나가는 상향식 방법과 상위 용어부터 하향식으로 계층적 용어를 구축해나가는 하향식 방법을 제시하고 있다. 또한 시소러스 평가방법으로 전문가를 활용하는 발견적 접근법(Heuristic evaluation) 및 유용성 평가를 제시하고 있으며, 평가항목으로 정의의 명확성, 용어 형태의 일관성, 용어의 최신성, 필요한 패싯 포함 여부, 동의어의 완전성 등을 간단하게 제시하고 있다.

ISO 25964는 시소러스 구축을 계획, 구현 초기, 구현, 시소러스 공개, 업데이트로 나누어 제시하고 있는데, 실제적으로 적용할 수 있는 구체적인 방안을 제시하고 있다. 예를 들어, 시소러스 구축을 위해서는 계획단계에서 목적과 대상, 업무 및 책임분담, 사용하고자 하는 소프트웨어를 명확히 할 것을 제시하였으며, 시소러스 구축을 위하여 이용자를 비롯한 이해관계자 및 시소러스 전문가를 참여시킬 것을 제시하였으며, 시소러스 공개 시 주제분야, 언어, 참고한 표준, 약어나 특수문자에 대한 설명, 총 용어 수, 용어 선택 및 복합어, 우선어 선택 등 시소러스 전반에 중요하게 적용된 규칙 등에 대해 설명해야 한다고 권고한다. 또한 배포를 위해 전자적 발행이나 전자 시스템에서 시소러스를 운용할 것을 제시하고 있으며, 시소러스 변경이 발생할 수 있는 상황으로 새로운 개념의 추가, 비우선어의 추가, 우선어나 비우선어의 수정, 우선어 삭제, 우선어와 비우선어의 교체, 사용범위 주기 및 정의의 변경 등을 제시하여 실제 업무를 조력하려고 노력하였다.

시소러스 구축 절차는 ANSI/NISO Z39.19나 기존 가이드라인에서 계획, 구축, 평가, 유지의 단계로 언급되어 왔었던 것에 비해 ISO 25964는 기존에 주의를 기울이지 않았던 시소러스 구축 계획 시 고려사항, 시소러스 업데이트 상황, 시소러스 공개 시 제공해야 할 정보 등 구축업무와 관련된 실제적인 정보를 제공하고 있는 것이 특징이다. 그러나, ANSI/NISO Z39.19에서 언급한 방법론적인 측면에 대해서는 간과하고 있어 시소러스 구축 연구에서 취약점으로 지적되어온 방법론 제공에 대해서 ISO 25964에서 추후 보완되어야 할 것이다.

4.5 시소러스 개념화(Thesaurus)

ISO 2788은 표준에서 시소러스라는 단어를 사용하고 있으며, 시소러스를 개념 사이의 관계를 명확히 하기 위하여 구조화된 통제 색인언어 어휘집이라고 간단하게 정의하고 있다. 또한 이 표준은 자동화 색인, 수학 또는 화학공식을 조직하거나 표현하는 것이나 권말색인에는 활용될 수 없음을 밝히고 있다. ANSI/NISO Z39.19는 통제어휘라는 단어를 사용하고 있으며 조금 더 정의 및 개념화를 구체화하고 있다. ANSI/NISO Z39.19는 통제어휘의 목적은 정

보를 조직하기 위한 수단이라고 정의하고 있으며, 목적, 원칙, 다른 정보조직 시스템과의 관계 등을 제시하고 있다.

첫 번째, 목적으로 다음의 다섯 가지의 세부 목적을 명시화하고 있다.

- 변환(Translation): 저자, 색인자, 이용자의 자연어를 색인 및 검색에 사용될 수 있는 언어로 변환하는 수단을 제공
- 일관성(Consistency): 용어의 형태나 용어부여에 있어서 일관성유지
- 관계표시(Indication of relationships): 용어 사이의 의미적 관계 표시
- 레이블 및 브라우징(Label and browse): 이용자가 필요한 내용 객체(content objects)를 찾을 수 있도록 내비게이션 시스템에서 관계 제공
- 검색(Retrieval): 내용객체(content objects)를 찾을 수 있는 검색 도구 제공

두 번째, 시소러스 구축의 주요 원칙으로 다음의 네 가지 주요 원칙을 제시하고 있다.

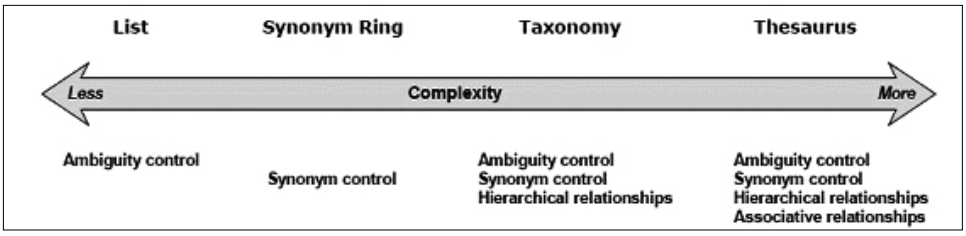
- 모호성 제거: 하나 이상의 의미를 가지고 있는 단어가 주는 모호성을 제거하기 위하여 하나의 용어는 단지 하나의 의미만을 갖는다.
- 유사어 통제: 하나의 개념은 하나의 우선어에 의해서 표현되나, 검색의 효율성을 위하여 여러 동의어나 비우선어를 가질 수 있다.
- 용어 간 적절한 관계 성립: 용어관계는 동등관계, 계층관계, 연관관계를 통해 표현될 수 있다.
- 패싯 분석: 통제어휘는 계층적 구조 외에 패싯 분석을 통해 조직될 수 있다.
- 용어선택근거: 용어선택의 근거로 내용객체를 기술하기 위해, 자연어, 이용자의 용어, 조직의 필요나 우선순위를 고려할 수 있다.

세 번째, 본 표준에서 네 가지 다른 유형의 통제어휘와의 관계를 정의하고 있다.

- 리스트(List): 용어의 알파벳 또는 논리적 정렬
- 동의어 리스트(Synonym ring): 검색의 효율성을 높이기 위하여 표시된 동의어 관계
- 탁사노미(Taxonomy): 우선어로 구성된 통제어휘로, 계층적 또는 다계층적으로 구성되어 있음.
- 시소러스(Thesaurus): 다양한 용어관계로 정렬된 통제어휘

즉, 통제어휘의 유형을 다음 〈그림 2〉와 같이 네 가지로 규정하고, 통제어휘의 표준이나 지침이 동일하게 이 유형에 적용될 수 있다고 밝히고 있으나, 온톨로지, 시맨틱 네트워크와 같은 어휘시스템에 적용되기에는 한계가 있음을 언급하고 있다. 즉, ISO 2788에 반해 ANSI/NISO Z39.19는 표준 적용범위를 조금 더 명확하고 포괄적으로 제시하고 있는 것이 특징이다.

ISO 25964는 기존 표준에서보다 조금 더 구체적으로 시소러스의 개념을 정의하고 있으며, 목적, 다른 정보조직 시스템과의 관계 등을 설명하고 있다.



〈그림 2〉 ANSI/NISO Z39.19 통제어휘 유형(p.17)

우선 시소러스의 전반적인 목적으로 검색확장, 대안적인 검색어 제시, 클러스터링이나 검색을 제한할 수 있는 수단제공, 철자오류 식별, 자동색인조력 등을 제시하고 있다. 또한 어휘통제의 목적으로 어휘와 개념간의 연결, 일관성 있는 어휘의 사용, 동의어 제시를 통해 정보접근점 제공확대 등을 제시하였으며, 관계의 목적으로 이용자가 한 개념을 표현하기 위하여 문맥에 대한 이해를 통해 가장 적절한 용어를 선택하기 위한 것이라고 제시하고 있다. 즉, ISO 25964는 ANSI/NISO Z39.19와 마찬가지로 일관성, 관계표시, 검색 등 전통적 색인의 목적을 제시하고 있으나, 검색에 대하여 대안적인 검색어 제시나, 검색을 제한할 수 있는 수단 제공, 철자오류 식별 등 조금 더 세부적 목적을 제시하고 있으며, 자동색인조력을 제시하여 자동화 시스템적인 측면을 부각한 것이 그 특징이다. 또한 우선어와 비우선어, 관계어의 세부 목적을 밝혀 시소러스 요소의 기능을 명확히 하고 있는 것이 특징이라 할 수 있다.

다른 정보조직 시스템과의 관계에서는 표준의 적용 영역으로 분류시스템, 파일링이나 비즈니스 분류시스템, 탁사노미, 주제명 표목표, 온톨로지나 토픽맵과 같은 시맨틱 네트워크뿐만 아니라 용어사전, 동의어 관계, 이름 전거리스트 등을 포함하고 있으며, 각각의 시소러스 유형의 특징 및 통제어휘 구축의 특징을 설명하고 있는 것이 특징이다. 이것은 ANSI/NISO Z39.19나 다른 표준에 비해 시소러스 표준 활용의 영역을 넓힌 것이라고 해석할 수 있다.

4.6 목차 및 세부사항(Table of contents and Details)

4.6.1 용어 선택 범위 및 형태 (Thesaurus terms)

ISO 2788은 용어 선택, 범위 및 형태와 관련하여, 용어 선택, 단수 및 복수형태, 범위주기 작성 설명 등을 간단하게 다루고 있는 반면, ANSI/NISO Z39.19는 용어 선택 시 근거, 그룹화 될 수 있는 개념의 유형, 명사, 형용사, 동사와 같은 용어의 문법적 형태, 우선어 선택 시 고려사항, 대소문자 사용 시 고려사항에 대해 추가 설명하고 있다. 용어 선택, 범위 및 형태에 있어서는 ANSI/NISO Z39.19와 ISO 25964

간에 큰 차이가 없으나, ANSI/NISO Z39.19에서 용어 선택 시 근거로 도메인(Organizational Warrant), 이용자(User Warrant), 언어적 근거(Literary Warrant), 어휘의 특수함의 정도, 다른 통제어휘와의 관계 등을 제시하고 있는 반면, ISO 25964에서는 이러한 부분이 생략된 것이 특징이다.

4.6.2 패싯분석(Facet analysis)

ISO 2788에서는 패싯형 디스플레이에 대한 언급이 생략된 것에 반해 ANSI/NISO Z39.19는 디스플레이에 대해 설명한 장에서 여러 유형의 디스플레이 중에서 계층형 디스플레이의 예로 패싯 디스플레이를 간단하게 설명하고 있다. 즉, "통제어휘에 따라서는 계층적으로 광의의 카테고리 또는 패싯에 따라 용어를 디스플레이 할 수 있다(ANSI/NISO Z39.19 2005, 68)"라고 간단하게 언급하고 있다. 이에 반해, ISO 25964는 패싯 분석을 하나의 장을 할애하여 설명하고 있는데, 패싯 선택 지침, 패싯 적용 예, 나열 방법에 대하여 설명하고 있다.

4.6.3 동등어(The equivalence relationship)

ISO 2788은 동등관계의 유형을 동의어와 준동의어으로 제시하고 있는 반면, ANSI/NISO Z39.19와 ISO 25964는 동등관계를 동의어, 준동의어, 어휘적 변형, 복합어를 여러 언어의 조합으로 표현하고자 할 때 사용한다고 제시하고 있다. 두 표준에서는 유사한 예를 제시하고 있는데, 그 예는 다음과 같다.

예)

동의어: 자유(liberty) UF 자유(freedom)

준동의어 관목(bushes) UF 관목숲(shrubs)

복합어를 여러 용어의 조합으로 표현: 탄광업(Coal mining) USE 석탄(coal) + 광업(mining)

4.6.4 계층어(The hierarchical relationship)

ISO 2788, ANSI/NISO Z39.19, ISO 25964 모두 계층어의 유형을 파생, 전체-부분, 사례, 다계층으로 제시하고 있으며 그 예 역시 유사하다.

예)

파생관계: 새(birds) NT 앵무새(parrots)

전체-부분관계: 캐나다(Canada) NT 온타리오(Ontario)

사례관계: 산악지역(mountain regions) NT 알프스(Alps)

다계층 관계: 생물학(biology) NT 생화학(biochemistry), 화학(chemistry) NT 생화학(biochemistry)

4.6.5 연관어(The associative relationship)

ISO 2788과 ANSI/NISO Z39.19는 연관관계에 속하는 대상을 동일 범주에 속하는 용어와 상이한 범주에 속하는 용어로 구분하였다. 동일 범주에 속하는 용어는 의미가 중복되는 것(예: 배(Ship), 보트(Boat)), 서로 배타적인 형제어(예: 당나귀, 말, 노새)가 포함된다. 상

이한 범주에 속하는 예로 과정/행위자(사냥/사냥꾼), 행위/속성(여론조사/여론), 행위/상품(직조/의류), 행위/목표(제본/책), 원인/결과(죽음/사별), 객체 또는 개념/속성(독/독성), 객체 또는 개념/기원(벨루가 캐비어/카스피해), 객체 또는 개념/측정단위(온도/온도계), 원재료/상품(밀/밀가루) 등을 언급하고 있다. ISO 25964도 기본 방침은 동일하나, 상이한 범주에 속하는 예를 더 추가하여 연관관계의 활용을 명확히 하였다. 추가된 예는 다음과 같다 - 학문분야/객체(신경학/신경시스템), 과정/도구(온도조절/온도조절장치), 물체/부분(광학계기/렌즈).

4.6.6 데이터모델(Data model)

데이터모델은 기존의 표준에서 언급되지 않았던 부분으로 ISO 25964는 15장에서 설명하고 있다. 시소러스를 컴퓨터 시스템에서 구현할 때 요구되는 지침을 제시하기 위하여, UML(Unified Modeling Language)을 사용하여 도식화하고, 더블린코어 요소를 사용하여 시소러스를 시스템에 기록할 때 필요한 요소를 제시하였다.

이를 간략하게 살펴보면 각 시소러스에 식별자, 기고자, 적용범위, 작성자, 날짜, 설명, 양식, 언어, 발행인, 관계, 권리, 주제, 제목, 유형, 원천, 버전을 부여할 것을 권고하고 있으며, 각 시소러스의 개념에 식별자, 생성날짜, 변경날짜, 상태, 최상위어, 표기법을 표시할 것을 권고하고 있으며, 각 시소러스 용어에 식별자, 어휘 값, 생성날짜, 변경날짜, 원천, 상태, 언어를 표시할 것을 제시하고 있다. 또한 개념범주는 식별자, 개념범주유형(예: 마이크로시소러스와 같은 주제범주), 표기법을 표시하고, 시소러스 배열에는 식별자, 순서의 강제성 유무, 표기법을 표시하도록 권고하고 있다. 이렇게 표시된 각 클래스는 서로 연관관계를 갖는데, 각 시소러스 개념은 다른 개념과의 계층관계, 연관관계 등을 표시해야 하며, 용어는 동등관계, 우선어, 비우선어를 표시해야 한다.

4.6.7 시소러스와 응용프로그램과의 통합 (Integration of thesauri with applications)

ISO 25964는 표준은 16장에서 시스템과 네트워크간 시소러스 상호운영을 위해 필요한 기능 및 특징, 색인 및 검색 시스템에서 시소러스를 운용할 때 요구되는 기능 및 특징을 제시하고 있으며, 이 역시 기존의 표준에서는 언급되지 않던 부분이다. 그 주요 내용을 살펴보면, 먼저 시소러스 상호운영을 위해 필요한 기능으로, 첫째, 시스템 구현 시 시소러스 용어가 갱신될 때마다, 그 전 용어로 색인되었던 데이터의 손실이 일어나지 않아야 함을 고려해야 한다. 이를 위해 용어와 그 용어로 색인된 자원 사이의 링크를 제공하는 것을 권고하고 있다. 둘째, 이용자가 네트워크 상에서 시소러스를 사용하기 위하여 표준 포맷 및 프로토콜의 사용을 제시한다. 셋째, 이용의 편의를 위하여 시소러스는 이용자들이 접근할 수 있는 용어 레

지스트리에 등록해야 한다. 넷째, 색인 및 검색 시스템 상에서 시소러스를 운용하기 위하여 시소러스 데이터를 정보검색 시스템으로 이전하게 되는데, 이 경우 검색 시스템은 다국어나 특수문자를 지원해야 한다. 다섯째, 시소러스의 검색 및 브라우징을 원활하게 하기 위하여 단어조합, 특수문자, 식별자 및 지시기호, 다국어 전환, 계층관계나 관계어 브라우징, 계층형 디스플레이 및 알파벳 디스플레이를 제공해야 한다. 마지막으로, 시소러스 용어별 색인 정보를 검색하기 위하여, 시소러스 디스플레이에서 용어를 선택하는 기능, 시소러스 용어를 사용하여 검색하는 기능, 비우선어가 우선어로 자동적으로 대체되는 기능, 색인된 정보의 수를 표시하는 기능이 제공되어야 한다고 제시하였다. 이렇듯 시스템적인 측면을 표준에 부가한 것은 시소러스 사용이 인쇄본이 아닌 전자적 환경에서 사용되고 있는 추세이며, 이에 따라 시스템적인 운용 지침에 대한 관심이 높아짐을 반영한 것이다.

4.6.8 교환형식(Exchange Formats)

시소러스의 상호운용성을 확보하기 위해 교환형식을 제시하고 있는 것 역시 ISO 25964의 특징이다. ISO 25964는 ISO 2709에 근거한 MARC, W3C에서 시맨틱 웹에 적용을 위하여 RDF/XML 기반 구현을 위해 제시한 SKOS (Simple Knowledge Organization Systems), 시소러스의 데이터 교환을 가능케 하기 위하여 Z39.50의 응용프로파일로 개발된 Zthes, BS 8723에 맞추어 제작된 시소러스의 교환을 가능하게 하는 XML 스키마를 소개하고 있다.

4.6.9 프로토콜(Protocols)

디지털 네트워크상에서 시소러스의 원활한 사용을 위해 ISO 25964는 시소러스를 특수하게 처리하는 프로토콜로 SWAD-E SKOS API, 기타 API, ADL 시소러스 프로토콜을 제시하고 있으며, 웹 데이터베이스 시소러스로 OASIS, SPARQL, Z39.50을 간략하게 소개하고 있다. SWAD-E SKOS API는 웹을 통해 시소러스와 다른 SKOS에 접근을 제공하기 위해 고안된 웹 서비스 API이며, ADL 시소러스 프로토콜은 알렉산드리아 디지털 도서관 프로토콜로 웹 상의 단일어로 구축된 시소러스 접근을 제공하기 위하여 고안된 프로토콜이다. OASIS는 문맥기반 질의 언어(Contextual Query Language)를 활용하여 오픈검색, 웹 기반 탐색 및 검색이 가능하도록 한 웹 데이터베이스 프로토콜이다. SPARQL(Simple Protocol and RDF Query Language)은 W3C에서 2008년 중요한 시맨틱 웹 기술로 발표한 것으로, RDF 기반 질의생성 및 그래픽생성이 가능하게 한다. Z39.50은 가장 잘 알려진 프로토콜 중 하나로 디지털 도서관이 다른 기종을 사용하고 있더라도 데이터호환이 가능하게 하는 표준 프로토콜이다.

본 연구를 통해 ISO 25964와 다른 표준의 차이를 요약하면 〈표 4〉와 같다.

〈표 4〉 표준간 비교요약

항목	ISO 2788:1986	ANSI/NISO Z39.19:2005	ISO 25964
서론	주제색인에 활용되기 위한 통제 및 색인어간 용어관계	정보저장 및 검색 시스템, 웹 내비게이션 시스템의 향상, 색인어의 일관성 향상	전자적 환경에서의 시소러스 제작 및 사용
범위	문서색인에 부여되는 용어	콘텐트 객체와 관련된 메타데이터를 표현하기 위하여 부여되는 용어	정보자원을 검색하기 위하여 부여되는 용어
약어 및 기호	한정된 약어사용	- 이력주기, 범위주기 추가 - 사례관계 하위어, 상위어 등 관계어 추가	- 기술요소, 코드, 관계범주로 구분하여 약어제시 - 범위주기와 정의 구분 - 이력주기 사용법 및 용례추가 - USE+, UF+제시 - 기호제시
시소러스 구축절차	—	- 구축 -〉 테스트 및 평가 -〉 유지 -〉 관리시스템 운용 - 방법론 제시	계획 -〉 구현초기 -〉 구현 시소러스 공개 -〉 업데이트
시소러스 개념화	- 시소러스 사용 - 구조화된 통제 색인언어 어휘집	- 통제어휘 사용 - 정보를 조직하기 위한 수단 - 다른 유형의 통제어휘와의 관계정의	- 시소러스 사용 - 정보검색, 어휘통제 등을 위함 - 자동화 시스템적인 측면 부각 - 다른 정보조직 시스템의 범위확대
목차 및 세부사항	- 간단한 용어선택 원칙 제시 - 동등어를 동의어와 준동의어로 한정	- 용어선택의 근거제시 - 패싯 디스플레이 간략설명 - 동등관계를 동의어, 준동의어, 어휘적 변형, 복합어로 확대	- 패싯분석 상세화 - 동등관계를 동의어, 준동의어, 어휘적 변형, 복합어로 확대
기타	—	—	- 데이터모델 제시 - 시소러스와 응용프로그램과의 통합 (관리소프트웨어 상세사항 제시) - 교환형식 - 프로토콜 - 제2판에서 어휘간 상호운용성 및 매핑가이드라인 제시

5. 결론 및 제언

시소러스 제작과 개발을 위한 지침을 제공하기 위하여 여러 표준이 발표되고 있으며, 효율적인 정보검색을 위해 지난 수십 년간 여러 시소러스가 구축되었음에도 불구하고, 시소러스 구축을 위한 가이드라인 및 표준의 발전에 대해서는 몇몇 연구에서 간단하게 언급되었을 뿐 연구의 계속성은 이루어지지 않고 있다. 최근 ISO 25964 표준 제정 프로젝트는 변화하는 정보기술 환경에 맞게 시소러스 표준을 개정하여, 표준 사이에 각기 다르게 사용되던 용어 및 약어의 사용을 일관성 있게 함으로써 시소러스 표준의 실효성을 높이고자 하는 목적으

로 진행되고 있다. 이에 본 연구는 ISO 2596을 소개하고, 현재 사용되고 있는 시소러스 표준 ISO 2788:1986, ANSI/NISO Z39.19:2005와의 비교분석을 통해 ISO 25964와의 차이점을 분석하고, ISO 25964 표준에서 시소러스 구축 시 초점을 두고 있는 바를 논의하며, 나아가 ISO 25964가 범용되기 위하여 향상되어야 할 점을 제시함으로써, 시소러스 구축 실제를 조력하고자 하였다.

본 연구를 통해 ISO 25964는 다음의 특징을 가지고 있음을 파악하였다.

첫째, ISO 25964는 시소러스 표준의 활용이 기존 분류시스템, 시소러스, 탁사노미, 동의어 리스트 등으로 제한되던 것을 온톨로지, 토픽맵과 같은 시맨틱 네트워크상의 구조적 어휘로 확장하고자 하였다. 이는 표준의 사용범위, 시소러스 개념 등에서 파악된 것으로 이를 통해 표준의 이용자를 문헌정보학 서지데이터베이스 개발자뿐만 아니라, 데이터베이스 벤더와 시스템 개발자들로 확장시키고자 한 것이다. Adam(2002), Jacob(2003), Gilchrist(2001)의 연구에서 분류시스템, 온톨로지, 토픽맵 등의 관계를 밝히고자 하였는데, 용어와 용어 간 관계로 구성된 구조화된 체계라는 점에서 공통점을 가지고 있으며, 중요한 것은 오랜 시간 문헌정보학에서 '시소러스나 통제어휘를 어떻게 구축할 것인가'라는 연구에서 얻어진 지식이나 경험을 어떻게 온톨로지나 시맨틱 네트워크와 같은 분야에 활용할 것인가이다(Park 2008). 이러한 관점이 ISO 25964에 반영되어 표준의 활용도를 높이고자 한 것이다.

둘째, ISO 25964는 기존의 표준은 물론 기존의 표준에서는 제시되지 않았으나, 구축된 여러 시소러스에서 사용되었던 사항들을 반영하여, 시소러스 구축의 표준 활용성을 높임과 동시에 표준과 시소러스 실제 사이의 일관성을 높이고자 하였다. 이는 ISO 25964에서 코드군을 제시하여 주제구분약어나, 콘셉트 코드를 제시한 것과, 기호의 사용을 제시한 부분에서 파악되는데, 이는 INPSEC, JICST, ROOT, Women's 시소러스에서 상이하게 사용되는 것을 반영한 것이다.

셋째, ISO 25964는 실무자들의 표준사용을 향상시키기 위하여 이들의 요구를 반영하고자 하였다. 이를 위해, 시소러스 절차에서 계획, 구현 초기, 구현, 시소러스 공개, 업데이트로 나누어 각 사항마다 개발자들이 점검해야 하는 사항을 제시하였으며, 데이터모델을 구체적으로 제시하여 시소러스 시스템 구축 실제를 조력하고자 하였다.

넷째, ISO 25964는 변화하는 시소러스의 전자적 기술환경을 반영하고자 하였다. 이를 위해 시소러스를 전자적 환경에서 구축하기 위한 데이터모델, 시소러스와 응용프로그램을 통합하기 위한 요구사항, 시소러스의 상호운용성을 향상시키기 위한 교환형식 및 프로토콜을 제시하였다.

이러한 특징에도 불구하고 ISO 25964가 보편적으로 참조되는 표준이 되기 위해서는 다음의 사항이 보완되어야 할 것이다.

첫째, 16장에서 시소러스 검색 및 브라우징을 지원하기 위해 시스템이 갖추어야 할 요소는 비교적 잘 제시되고 있는 반면, 상호운영과 관련하여서는 특징만이 간략하게 설명되어 있어, 시소러스 용어 갱신 시, 데이터 손실을 방지하는 방법, 시소러스를 용어 레지스트리에 등록하기 위한 세부적인 지침이 제공되지 않아 표준의 명확한 사용이 제한되므로 이에 대한 보완이 이루어져야 할 것이다.

둘째, 시소러스 절차에서 기존의 ANSI/NISO Z39.19이나 시소러스 구축 방법론과 관련한 연구에서 언급되어왔던 상향식 및 하향식의 시소러스 구축 접근법, 평가에서 유용성이나 발견적 접근법의 활용, 평가 시 고려사항, 용어의 선택에서 기관, 문헌, 이용자에 따른 근거 제공 등이 ISO 25964에서는 언급되지 않았다. 시소러스 구축에 있어서 개발자들의 가장 주요한 정보요구는 '이용자가 찾고자 하는 정보가 무엇이며, 어떠한 개념과 용어로 그 정보를 찾고자 하는가'이므로(Park 2008), 용어선택의 근거, 평가 및 접근법에서의 개발자들의 정보요구에 맞는 방법론을 제시하는 것 역시 시소러스의 시스템 상에서의 운용 못지 않게 중요한 것이다. 이를 위해 조금 더 상세한 연구조사를 통해 방법론적인 측면의 보완이 이루어져야 할 것이다.

셋째, 패싯 분석에서도 패싯 선택 지침이나 적용 예가 간략하게 언급되었을 뿐 패싯 구축을 위한 상세한 가이드라인, 패싯 선택 시 지침, 패싯 용어 선택 시 지침과 같은 방법론적인 측면은 설명되지 않았다. 패싯형 분류법은 정보기술의 발전에 따라 패싯이 추구하는 정보의 다각적 접근이 가능해짐에 따라 많은 인포메이션 아키텍처 구축에서 사용되어 오던 것으로 이에 대한 상세한 지침을 제공함으로써 표준의 활용영역을 확대할 수 있을 것이다.

위의 언급한 사항이 보완되어야 함에도 불구하고 서로 각기 다른 표준에서 언급되어온 시소러스 구축의 주요 논의사항 및 상이한 용어의 사용을 일관성 있게 재구성하고, 표준의 개정 요구를 반영하여 국제적으로 통용될 수 있는 표준을 제정하고 있다는 점에서 표준의 폭넓은 활용이 기대된다.

참고문헌

기술표준원. 2007. *KSXISO5964. 문헌 - 다국어 시소러스의 제정과 개발 지침.*

기술표준원. 2008. *KSXISO2788. 문헌 - 단일어 사전의 제작과 개발을 위한 지침.*

이화여자대학교. 2000. 『시소러스개발지침』. 서울: 한국데이터베이스진흥센터.

Adams, K. 2002. "The Semantic Web: Differentiating between Taxonomies and Ontologies." *Online*, 26(4): 20-23.

BS 8723 *Structured vocabularies for information retrieval.*

Clarke, S. T. 2010. "Thesaurus Standards on a Converging Track." *Legal Infor-*

mation Management, 10(1): 43-45.

Gilchrist, A. 2003. "Thesauri, Taxonomies, and Ontologies - an Etymological note." *Journal of Documentations*, 59(1): 7-18.

Hsieh, H-F. and Shannon S. E. 2005. "Three Approaches to Qualitative Content Analysis." *Qualitative Health Research*, 15(9): 1277-1288.

International Organization for Standardization. 1985. *ISO 5964: Documentation - Guidelines for the establishment and development of multilingual thesauri*.

International Organization for Standardization. 1985. *ISO 25964: Thesauri and interoperability with other vocabularies*.

International Organization for Standardization. 1986. *ISO 2788: Documentation - Guidelines for the establishment and development of monolingual thesauri*.

Jacob, E. K. 2003. "Ontologies and the Semantic Web." *Bulletin of the American Society for Information Science and Technology*, 9(4): 19-23.

Krooks, D. A. and Lancaster F. W. "The Evolution of Guidelines for Thesaurus Construction." *Libri*, 43(4): 326-342.

Lancaster, F. W. 1986. *Vocabulary Control for Information Retrieval*. Second edition. Arlington, VA: Information Resources Press.

National Information Standards Organization. 2005. *ANSI/NISO Z39.19 - Guidelines for the construction, format and management of monolingual controlled vocabularies*.

Park, O. N. 2008. "Opening Ontology Design: a Study of the Implications of Knowledge Organization for Ontology Design." *Knowledge Organization*, 35(4): 209-221.

Spiteri, L. 1998. "A Simplified Model for Facet Analysis." *Canadian Journal of Information and Library Science*, 23: 1-30.

UNESCO. 1970. *Guidelines for the Establishment and Development of Monolingual Thesauri*. Paris: UNESCO.

U.S., Federal Council for Science and Technology Committee on Scientific and Technical Information. 1967. *Guidelines for the Development of Information Retrieval Thesauri*. Washington, D.C.: Government Printing Office.

Williamson, N. J. 1996. "Standards and Rules for Subject Access." *Cataloging & Classification Quarterly*, 21(3): 155-176.

Zeng, M. L. 2005. "Standards for Controlled Vocabularies." *7TH NKOS Workshop*, JDCL2005. Denver.