

용어 정의를 도입한 시소러스 개발 연구*

A Study on the Development of Thesaurus Using Terminological Definitions

김태수(Tae-Soo Kim)**

초 록

특정 주제영역의 시소러스에 수록되는 디스크립터의 의미범위를 한정하고, 개념의 본질적인 속성을 제공하기 위해 정의를 도입하고, 정의를 구성하는 데이터요소를 정의모델에 따라 분석한 후, 이를 시소러스 관계구조로 표현하기 위한 연구이다. 이를 위해 한국산업규격(KS)의 정보산업분야에 수록된 용어정의 327개를 정의규칙과 정의모델에 따라 분석한 다음, 각 속성을 디스크립터간의 관계구조로 표현한 시소러스를 실험적으로 구현하였다. 지금까지 시소러스 구축시 개발주체의 성격에 따라 디스크립터의 수록범위와 관계구조의 설정을 달리해 왔다. 용어의 정의를 시소러스 구조에 도입하게 되면 각 디스크립터의 의미범위를 한정할 수 있고, 동시에 정의에 사용된 용어간의 의미관계를 디스크립터 관계구조로 표현할 수 있어, 의미범위와 관계구조의 표준화 가능성을 기대할 수 있다.

ABSTRACT

As contemporary thesauri have become large and complex, it is increasingly difficult to assess the intended meaning of each one of the terms. Thereby the meaning of many descriptors seems to be very similar, and it is often not possible to distinguish among them and to identify the term correctly. Purpose of this article is to induce definitions of descriptors in thesaurus by specifying the characteristics of each concept, locating it in the domain and providing clear and prescriptive information on the meaning of each descriptor in the form of a standardized terminological definition. In this study, a small prototype thesaurus using definition of term in the field of information industry in Korean Standards has been developed. In this thesaurus definitions are written for each descriptor with the help of a proposed defining model and in accordance with defining rules borrowed from the field of terminology. In addition, elements of analyzed definition have been included in the relation structure of descriptors. It is revealed that terminological definition added to thesaurus may permit extraction of separate items of information from definitions for the representation of knowledge structures and makes it easier to confine the scope of descriptors to be included in thesaurus in a given subject field.

키워드 : 시소러스, 정의, 디스크립터 관계구조, 용어
thesaurus, terminological definition, descriptor

* 이 논문은 1999년도 한국학술진흥재단의 연구비에 의하여 지원되었음(KRF-99-041-C00530).

** 연세대학교 문헌정보학과 교수(btrees@yonsei.ac.kr)

- 논문 접수일 : 2001년 5월 21일
- 게재 확정일 : 2001년 6월 5일

1 서론

1.1 연구의 필요성 및 목적

시소러스는 특정 주제영역에서 사용되는 용어(terms)와 이들 용어간의 의미관계를 체계적으로 제시한 색인어휘집으로서, 색인과 검색 과정에서 디스크립터와 검색어를 선정하기 위한 필수도구로 사용되고 있다. 특정 주제영역에서 사용되는 개념을 표현하기 위한 용어를 일관되고 정확하게 사용하고, 모든 용어간의 의미관계를 분명히 제시할 수 있다면 검색효율을 개선하는데 크게 기여할 수 있기 때문에 정보검색시스템에서는 색인도구로서의 시소러스 개발에 많은 노력을 기울여 왔다.

그런데 용어간의 의미관계를 분명히 제시하기 위해서는 무엇보다도 각 용어가 지닌 의미가 명확해야 하는데, 지금까지 대부분의 시소러스는 용어가 지닌 의미속성을 직접 제시하는 대신, 용어간의 관계구조나 범위주기(scope notes)를 통한 간접적인 방식으로 제시해 왔다. 그런데 이러한 관계구조나 범위주기를 통해서 개념이 지닌 본질적인 속성을 이해하기 어렵고, 더욱이 시소러스에 수록되는 용어의 범위를 한정할 수 없다는 한계를 지니고 있다(김태수 2000, 319). 시소러스에 수록되는 용어가 증가함에 따라 그 의도된 의미를 평가하기 어렵고, 개념간의 충돌로 인해 관계구조의 설정이 어렵게 된다. 일부 시소러스에서 용어 분류와 범주를 부차적인 수단으로 제시하고 있긴 하지만 용어의 중의성을 발견할 수 있을 뿐만 아니라 특정한 상황에 필요한 용어를 정확하게 식별하기 어려운 실정이다.

그 동안 국내에서는 시소러스의 구축과 관련된 이론적 접근보다 실무지침(manuals)에 따른 시소러스 구축에 치중함으로써, 동일한 주제영역의 시소러스일지라도 개발주체의 성격에 따라 수록된 정보의 질과 구조화의 수준을 달리하여 데이터를 신뢰하기 어렵고, 특정 용어의 의미 속성을 확인하기 어려운 실정이다. 따라서 시소러스에 수록되는 용어의 선정기준과 선정된 용어의 범위를 결정짓고, 아울러 의미구조를 완벽하게 표현할 수 있는 기법이 요구된다. 이를 위한 한가지 수단으로 본 연구에서는 디스크립터가 지닌 본질적인 속성인 정의(definitions)를 시소러스에 도입하고자 한다.

이러한 정의를 시소러스에 도입하는 주된 근거는 특정 주제분야의 개념이 바로 그 분야에서 인정되고 있는 지식구조를 표현하고 있다는 점과, 그 개념이 속한 지식분야의 다른 모든 개념과 관련하여 그 개념의 위치를 한정할 수 있고, 이에 따라 디스크립터를 선정하는데 일관성을 유지할 수 있다는 점 때문이다.

시소러스는 앞으로도 기본적인 검색도구로서 역할을 할 것이며, 따라서 개개의 디스크립터가 지닌 특성과 이들간의 관계를 일정한 기준에 따라 정확하게 제시하는 것이 중요하다. 이를 통해 다양한 주제분야의 시소러스 개발에 적용 가능한 이론적 틀을 확보할 수 있고, 색인과 검색효율을 개선하는데 기여할 것으로 판단된다.

이 연구는 용어의 정의를 분석하여, 개념이 지닌 본질적 속성을 기본 데이터요소로 하는 정의모형을 제시하고, 이 모델에 사용된 용어의 속성과 관계구조에 기반한 용어정의 시소러스를 구축하기 위한 것이다.

1.2 연구의 내용 및 방법

표준화된 정의모델에 기초한 정의를 시소러스에 도입하여, 각 디스크립터의 의미를 명확히 제시함과 동시에 이 정의모델에 사용된 속성을 디스크립터의 관계구조에 적용하여 용어간의 관계를 체계적으로 제시하기 위해 다음과 같은 방법을 적용하였다.

첫째, 시소러스에 정의를 도입하기 위해, 시소러스 개발과 관련된 기준과 정의규칙, 그리고 기준에 제시된 정의모델을 검토하였다.

둘째, 시소러스에 도입할 정의를 수집하기 위해, 국내에서 제정된 한국산업규격(KS)의 16개 부문 중 정보산업분야(X)의 규격에 수록된 정의를 대상으로, 정의규칙과 정의모델에 의거하여 부분적으로 재정의하였다.

셋째, 특정 주제분야의 지식구조에서 정의의 기술방식을 표준화하기 위한 정의모델은 Sager와 L'Homme(1994)가 제시한 모델을 기본으로 하고 필요한 부분을 확장하였다.

넷째, 정의모델의 구성요소 중 시소러스 구축시 필요하다고 판단되는 요소를 추출하여 이를 디스크립터간의 관계구조에 반영한 시소러스를 실험적으로 구축하였다.

2 시소러스에서의 용어정의

2.1 용어와 정의

용어는 특정 주제분야에서 특정 개념을 표현하는 하나 이상의 단어로 구성된 의미단위로서(Wright and Budin 1997, 64), 지식의 단위

요소인 개념(Dahlberg 1989, 13)을 언어기호를 이용하여 표현한 것이다. 이 언어기호를 통해 우리는 지식을 지시할 수 있는 것이다(Sager 1990, 18). 그런데 이 용어는 기호의 임의성이 인정되는 일반 언어와 달리, 체계적인 원칙이나 규칙에 따라 명명된다. 이를 통해 개념의 명료성과 일관성이 유지되고, 나아가 용어로 표현되는 개념의 본질적 속성을 용어구조 속에 반영할 수 있는 것이다(Sager 1990, 57). 이처럼 용어는 체계화된 단일주제영역의 구조조직을 반영하기 때문에 개념분석을 토대로 조직된 현실적인 정의체계를 구성한다. 개념수준에서 단어가 용어가 되기 위해서는 다른 용어와 구분되고 아울러 전체 용어체계의 한 요소이어야 한다. 이와 같은 상호 식별체계를 공식화하는 유일한 방법은 정의라는 과정을 통해서만 가능한 것이다(Wright and Budin 1997, 65).

어원상으로 보면 정의와 용어는 긴밀히 관련되어 있다. 정의를 내리는 일을 용어의 기본적인 의미를 제시하는 과정, 다시 말해 용어의 의미범위를 확정짓는 행위라고 한다면 이 행위의 결과가 용어로 표현된다. 개념적으로 표현하면 하나의 단어가 용어가 되기 위해서는 정의를 통해 그 용어의 경계가 분명하게 규정되어야 한다는 점이다(Sager and Ndi-Kimbi 1995, 62). 따라서 정의는 개념이 지닌 본질적인 속성을 제시하는 수단이면서, 이 정의를 통해 특정 개념이 속한 개념체계와 이 체계 속에서 특정 개념을 고유하게 식별하고 분류하는 기능을 지닌다.

정의란 어떤 개체가 지닌 개념적 특성(내포적 의미)을 밝히는 논리적이고 분석적인 방법

으로서, 정의를 사용하는 주된 목적은 1) 용어와 개념간을 등식으로 규정하고, 2) 독립된 정의를 통해 용어를 식별하고, 3) 개념의 의미를 제시하기 위한 것이다(Sager 1990, 45). 즉 특정한 추상화 수준에서 개념을 기술하고, 그 개념을 관련된 개념과 구분하며, 개념체계에서 특정 개념의 위치를 결정하기 위해 그 개념과 다른 개념간의 관계를 설정하고, 개념의 범위를 한정하여 용어의 이용을 표준화하기 위한 것이다(ISO 1987, 5 ; Dahlberg 1989, 21 ; Hudon 1996, 364 ; Wright 1997, 66). 따라서 정의는 용어를 선택하고 분석하기 위한 출발점으로, 특정 개념에 대한 용어를 선택할 때는 분명한 개념정의를 가져야 한다.

정의의 표현형식을 구체적으로 보면, '피정의항 = 정의항(종차 + 상위개념)'의 등식으로 표현된다(이병근 1992, 14-17). 예컨대 '이름'이라는 용어에 대한 정의는 다음과 같이 구성된다.

이름 = 인간을 위한 객체와 개념에 대한 식별의 기본 수단		
①	②	③

여기서 ①은 피정의항으로서 개념을 규범적인 표현형식으로 나타낸 것이고, ②③은 정의항으로 피정의항의 의미를 명시하는 명사구 형태로 구성된다. 부연하면 ①은 종개념, ②는 종차(differentiae), ③은 유개념으로서, 종차는 유개념에 속하는 종개념들간의 특성을 나타낸다.

위의 등식구조에서 피정의항과 정의항은 주어와 술어로 구분되며, '피정의항=정의항'이라는 형식으로 표현되는데, 일반적으로 간결한 문체로 구성하기 위해 연결동사는 생략하고 명

사구 형식으로 구성된다.

비록 다양한 유형의 정의가 사용되고 있지만 개념의 본질적 속성과 종차를 확인하고 식별하여 개념을 정의하는 분석적 정의가 대부분의 용어시스템의 기초가 되고 있다(Hudon 1996, 364). 그런데 분석적 정의에서 요구되는 전제는 두 가지 이다. 하나는 피정의항이 다른 용어나 개념을 지시해야 한다는 점이다. 따라서 개념을 둘러싼 외부 공간을 한정하거나 다른 개념과 관련하여 공간상에서 그 개념의 상대적 위치를 한정해야 한다. 다른 하나는 다양하고도 관련된 다른 개념과 피정의항을 구분할 수 있는 식별기준, 즉 특성을 제시해야 한다(Sager and Ndi-Kimbi 1995, 62). 이 두 가지 전제조건을 통해 피정의항은 다른 개념과 연결되며, 특정 개념의 위치를 한정함으로써 이에 근거한 개념구조를 분석할 수 있게 된다.

2.2 정의의 구성요소

개념은 용어가 지닌 의미로서, 용어 이전에 생성되며 어느 특정 순간에 인간이 관찰한 내부, 외부 세계의 개체(individual object)를 분류하기 위해 사용되는 지적 구성체이다. 이 개념을 분류하기 위한 기반이 특성인데, 특성은 개념의 성질을 표현한 것으로 특정 개체의 개념을 형성하고 한정하며, 특정주제분야에서 개념 상호간을 식별하는데 필수요소이다(ISO 1987, 1-2 ; Sager and Ndi-Kimbi 1995, 64 ; ISO 1990). 이 특성은 흔히 내적 특성과 외적 특성으로 구분되는데, 일반적으로 특정분야에서 새로운 개념이 형성되기 위해서는 모양, 크기, 재료, 색깔 등의 내적 특성이 기원,

기능, 위치, 발견자, 발명자, 상태 등의 외적 특성에 우선하여 사용된다. 또한 특성은 본질적(essential) 특성과 비본질적(inessential) 특성으로도 구분되는데, 본질적 특성은 일정한 관점에 따라 특정분야에서 개체의 본질을 반영한 특성으로서, 개별개체에서 관찰 가능한 비본질적 특성과 구분된다(ISO 1987, 2).

한편 특정 개념과 다른 개념과를 구분하는 특성 전체를 내포라고 하며, 동일한 추상화 수준에 속하는 모든 종(species)이나 특정 개념의 특성을 지닌 개체 전체를 그 개념의 외연이라고 한다. 내포와 외연의 관계를 예시하면 다음과 같다.

① “평행사변형”이란 개념의 내포는 “사각형”, “평행으로 마주보는 변”이라는 특성으로 구성되며, 이의 외연은 평행사변형의 가능한 모든 유형인 “마름모”, “정사각형” 등이다.

② “대등서명 : 본서명을 다른 언어나 문자로 적은 서명”에서 “대등서명”은 상위개념(“서명”)과 특성(“본서명을 다른 언어나 문자로 적은”)으로 구성된다.

③ “태양계행성 : 수성, 금성, 지구, 화성, 목성, 토성, 천왕성, 해왕성, 명왕성”과 같은 정의에서 동일 수준의 모든 하위개념을 외연이라고 하며, 시간이 지남에 따라 새로운 종이 추가될 수 있기 때문에 타당성이 지속되지 않을 수도 있다.

이들 개념은 항상 다른 개념과 연관되어 고려되는데, 실제로 개념간의 관계는 계층관계와 비계층관계로 구분된다. 계층관계는 상위개념과 하위개념으로 구성되는 포섭(generic)관계와 부분(partitive)관계로 나눌 수 있는데, 포섭관계는 하위개념범주에 속하는 모든 개념이

상위개념의 외연이 되며, 하위개념은 상위개념의 모든 특성을 지니면서 동시에 최소한 하나 이상의 식별 특성을 추가로 가지게 된다. 이 추가된 식별 특성이 동일한 추상화 수준에 있는 하위개념을 식별하게 된다. 부분관계는 전체와 부분간의 관계로서, 부분개념은 전체개념에 의존하기 때문에 부분개념이 전체개념보다 우선하여 정의될 수 없다. 예컨대 “자동차”를 정의하기 앞서 “자동차 엔진”을 정의할 수 없는 것과 같은 이치다. 비계층관계는 시간, 공간, 인과관계를 말한다(ISO 1987, 3).

결과적으로 개념은 각 주제분야에서 그 개념의 특정한 위치를 점유하는 체계를 형성하는데, 이 개념체계는 지식의 지적 체계화의 수단으로 사용되며, 개념간의 관계를 분명히 하며, 통일된 표준 용어를 사용하여, 상이한 언어로 표현된 용어간의 등가성을 구현할 수 있다(ISO 1987, 4).

2.3 용어의 정의규칙

정의규칙은 대개 사전편찬자가 관심을 가져왔는데, 정의에 대한 전통적인 규칙은 주로 정의의 표현형식을 다루고 있으며, 그 주된 내용은 다음과 같다(Ndi-Kimbi 1994, 329-330).

① 정의항은 피정의항과 동등해야 한다.

② 정의항은 피정의항에 있는 표현을 포함해서는 안된다.

③ 피정의항이 부정적이지 않다면, 정의항을 부정적으로 표현해서는 안된다.

④ 정의항은 불분명하거나 비유적인 언어로 표현되어서는 안된다.

이 정의규칙에서 ① 항은 동의어와 의역에

의한 정의방법을 인정하고 있으며, ② 항은 용어의 반복을 피하기 위한 의도로 보이나, 복합 용어에 대해 적용하기 어려워 용어학적인 면에서 다소 부적절한 것으로 판단된다.

한편 특정한 주제분야나 목적에 따라 새로운 개념에 대한 정의의 필요성이 증대됨에 따라 전통적인 정의규칙 이외에 각종 표준기관에서는 다음과 같은 정의규칙을 제시한 바 있다.

ISO 704:1987. Principles and Methods of Terminology.

ISO 1087:1990. Terminology-Vocabulary.

ISO 10241:1992. International Terminology Standards-Preparation and Layout.

ISO TC46. 1994. Draft Procedure-annex E : Guidelines for Definitions.

ISO 11179:1994. Information Technology-Specification and Standardization of Data Elements:Part 4. Rules and Guidelines for the Formulation of Data Definitions.

KSX 1517-4 : 1996. 데이터요소-제4부: 데이터요소의 정의를 위한 규칙과 지침.

(한국표준협회 1996).

위에 제시된 정의규칙에서는 정의에 사용된 개념을 일정한 개념체계에서 이미 사용되거나 잘 알려진 용어로 표현하도록 다음과 같은 원칙을 제시하고 있다(ISO 1987, 6-8).

① 정의에 반영된 특성

정의는 특정 개체를 표현하는 개념의 본질적인 특성을 포함해야 하며, 여기서 본질적 특성이란 내적 특성과 외적 특성 모두를 의미한다.

② 정의의 타당성

정의는 특정한 개념체계에서 하나의 개념을 식별하기 위한 본질적이며 직접적으로 관련된 특성을 포함해야 한다.

물(화학) : 산소와 수소의 화합물

물(물리학) : 섭씨 0°에서 였고 섭씨 100°에서 끓는 액체

③ 정의의 체계성

정의는 상위개념의 본질적 특성을 제시하고, 이 특성에 기초한 관계를 통해 개념의 특성 및 개념간의 체계를 제시해야 하고, 아울러 대등한 개념간을 구분할 수 있어야 한다. 부분관계를 정의할 때는 부분개념이 속한 전체 개념을 반드시 언급해야 하고, 전체를 표현하는 개념을 정의할 때는 개체들의 집합임을 분명히 제시해야 한다.

④ 정의의 간결성

정의는 특정 개념의 본질적(내적이나 외적) 특성만을 포함해야 하는 반면, 이러한 특성으로부터 파생된 다른 특성을 포함해서는 안된다.

⑤ 불완전한 정의

정의는 정의의 대상인 개념의 외연과 정확히 일치해야 한다. 그렇지 않으면 불완전한 정의가 되고 이로 인해 등식이 성립되지 않는다.

⑥ 순환추론(circular reasoning)

순환추론은 개념이 다른 개념의 도움으로 정의되는 상황으로, 계속해서 다른 개념의 도움을 필요로 하는 경우이다. 이것은 두 가지 유형의 순환추론을 초래하게 된다.

섬유산업 : 섬유를 생산하는 산업

섬유 : 섬유산업의 제품

⑦ 동의어반복

동의어반복은 순환추론의 한 형태로서, 정의를 내릴 때 용어에 포함된 구분특성을 포함해

서는 안된다.

⑧ 부정적 정의

정의를 내릴 때, 개념이 부정적일 경우에만 부정적 특성을 포함하며, 그 이외에는 긍정문으로 표현되어야 한다.

2.4 시소러스에서 정의의 이용

시소러스의 주된 기능은 개념의 의미를 분명히 제시하고, 개념과 이를 표현한 용어간의 관계를 제시하는데 있다. 모든 디스크립터의 의미가 분명하게 제시된다면 색인자는 색인 대상인 개념을 표현하기 위해 정확하고 적절한 디스크립터를 일관되고 정확하게 선정할 수 있을 것이다.

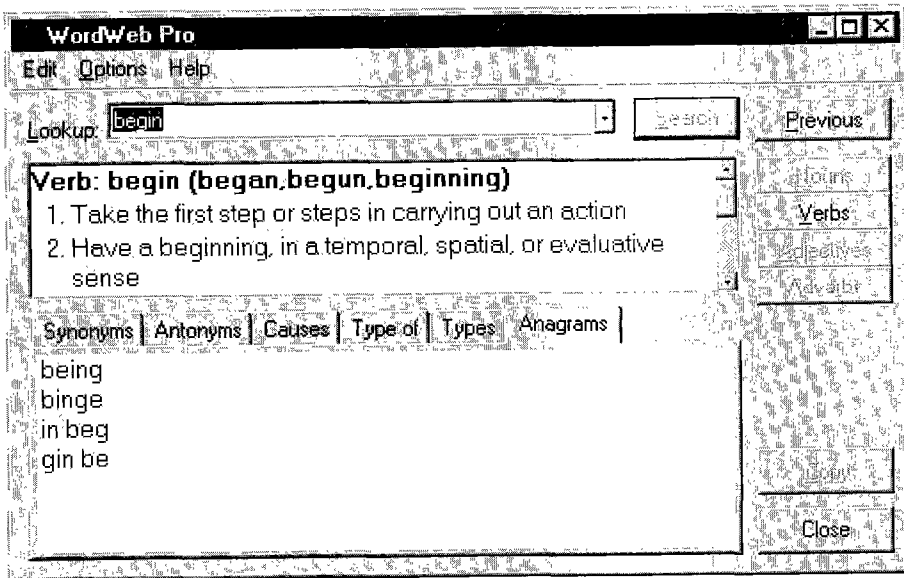
전통적인 시소러스에서는 주로 용어관계나 범위주기를 통해 디스크립터의 의미를 제시해 왔다. 그러나 시소러스에 수록되는 용어가 확장되고 여기에 복잡한 제어도구를 도입함으로써 각 용어의 의미를 평가하는 일이 점차 어렵게 되었다. 그 결과 디스크립터의 중의성을 볼 수 있고 이로 인해 용어가 지닌 의미상의 차이를 구분하기 어렵고, 특정한 상황에 필요한 용어를 정확하게 식별하는 일이 어렵게 되었다. 그 주된 이유는 시소러스 구축에 관한 국제표준 등에서 디스크립터의 정의나 범위주기의 사용을 제한적으로 인정하고 있기 때문이다. 예컨대 시소러스 개발을 규정 한 ISO 2788(1986)에서는 중의성을 해소하기 위한 방안으로 범위주기와 정의의 도입 가능성을 인정하면서도 범위주기는 ① 디스크립터의 이용을 제한하거나 ② 약어나 두문자를 설명하거나 ③ 상이한 주제분야에서 일반적으로 사용되는

용어의 의미를 제한하는 용도로 사용하도록 규정하고 있다(1986, 5). 아울러 정의와 관련하여, 일반적으로 시소러스에서 디스크립터의 개념 범위는 주로 다른 용어와의 관계구조로 표현되고, 디스크립터가 지닌 고유한 해석과 관련하여 의심스러운 경우에만 정의를 사용하도록 규정하고 있다(ISO 1986, 5).

이밖에 시소러스 구축 지침에서도 정의는 해당 용어의 구성요소로 취급되지 않고, 마치 동형이의어의 식별수단인 한정어의 기능으로 이해하여, 특수한 경우에만 사용하도록 규정하고 있으며(Aitchison, Gilchrist and Bawden 1997, 32-33), 범위주기는 ① 의미의 한정 ② 주제범위 지시 ③ 색인자에 대한 지시사항 ④ 용어의 구조체계상 필요하나 실제 색인어로 사용되지 않는 용어의 지시 ⑤ 용어의 내력 ⑥ 범위주기에 대한 참조의 기능으로 규정하고 있다(Aitchison, Gilchrist, and Bawden 1997, 33-34). 다시 말해 정의나 범위주기는 제한된 범위에서만 사용되고 이의 도입을 필수요소로 고려하지 않고 있다.

그런데 시소러스 구축시 디스크립터의 정의를 도입하려는 생각은 일찍이 Moores에 의해 제시된 바 있는데, 그는 색인용어를 디스크립터로 변환함에 있어 이용문맥과 기본적으로 다른 특정한 의미를 정의로 제시할 필요가 있음을 지적한 바 있다(1963).

또한 Soergel은 시소러스 구축시 용어의 의미는 시소러스 구조의 전후관계를 통해 명백히 제시하고, 색인 및 검색시 시소러스를 이용하기 위해서는 완전하고 공식적인 정의가 필요하며, 의도된 의미대로 용어가 사용되지 않을 때 범위주기를 추가해야 한다고 지적한 바 있다



〈그림 1〉 WordWeb 시소러스의 구조

(1974, 145). 이는 일반적으로 시소러스에서는 사전과 같이 완전한 정의를 포함하지 않지만, 의미를 보완하는 역할로서, 제한적이고 확장된 정의를 추가할 필요성을 지적한 것이라 하겠다.

전문용어와 시소러스를 연결한 용어 시소러스(terminological thesaurus)를 개발하려는 시도는 용어학 분야에서 시작되었다. 이 가운데 Sager는 정보생산과 커뮤니케이션에 관심을 가진 전문가들에게 유용한 다양한 도구를 제시하였으며(1990), 특히 Sager와 L'Homme는 표준화된 정의모형을 제시한 바 있다(Sager and L'Homme 1994). 이와 더불어 Hudon은 Sager와 L'Homme의 모형을 확장한 정의를 사회과학분야의 시소러스에 도입하여, 디스크립터의 의미범위를 정확하게 기술하고, 색인자의 용어 선택을 지원할 수 있음을 제시한 바 있다(1996). 이밖에 특정 주제영역의

시소러스에서도 정의를 통해 용어의 의미를 분명히 제시하고 아울러 용어의 출처와 적용범위를 제시할 수 있다는 연구가 보고된 바 있다(Strehlow 1983, 15). 이러한 성과에 따라 디스크립터에 정의를 포함한 WordWeb 시소러스가〈그림 1〉과 같이 개발된 바 있다(1999),

이밖에 〈그림 2〉와 같이 NASA Thesaurus와 Canadian Literacy Thesaurus : CLT(1996)에서도 일부 정의를 도입하고 있으나, 정의의 구성요소를 디스크립터간의 관계구조에 반영하거나 정의를 기술하기 위한 정확한 지침은 제시되지 않고 있다. 예컨대 CLT에서는 의미의 중의성을 가진 용어의 경우에만 간략한 정의를 제시하고, 디스크립터의 이용범위를 제시할 필요가 있는 경우에는 범위주기를 제한적으로 제시하도록 하고 있다(Hudon 1992, 258-260).

NASA 시소러스

microbursts (meteorology)
 (added January 1993)
 SN (EXCLUDES IONOSPHERIC RADIATION MICROBURSTS)
 DEF A strong, localized downdraft that strikes the ground creating an outflow of severe winds near the ground that diverge radially from the impact point.
 UF *bow echo microburst events*
 GS meteorology
 micrometeorology
 microbursts (meteorology)
 storms

Canadian Literacy Thesaurus

Library literacy programs

Definition:
 programs designed, administered and staffed by a library

French equivalent:
 Programme d'alphabétisation en bibliothèque

Used for:
Library based literacy programs
Library operated literacy programs

Broader Terms:
Literacy programs

〈그림 2〉 NASA 시소러스와 CLT의 구성사례

3 용어정의모델

용어의 정의를 표준화하기 위해 Sager와 L'Homme은 다음과 같이 일곱 가지 요소로 구성된 모델을 제시한 바 있는데, 이것은 전통적인 분석적 정의를 표준형식으로 표현하기 위해 정의기술방식을 범주화한 것이다(1994, 354-356).

정의를 이상의 요소로 분석하게 된 주된 이유는 데이터베이스 관리측면에서 다음과 같은 장점이 있기 때문이다(Sager and L'Homme

1994, 358). 첫째, 자동제어를 통해 데이터 수집을 자동화할 수 있어, 정의의 일관성과 완전성을 확보할 수 있다. 둘째, 정의에 사용된 기존의 문장을 소수의 특성으로 기술할 수 있기 때문에 용어 데이터의 구축에 이용할 수 있다. 셋째, 단일 파일내의 일곱 가지 데이터 요소를 개별적으로 탐색할 수 있고, 또한 복수의 파일에 걸쳐있는 이들 데이터를 조합하여 탐색할 수 있다는 점이다. 예컨대 정의형식과 함께 구성요소의 전체 계층을 검색하거나 복잡한 개체를 생성할 수 있고, 특정 제품의 기원을 원료나 그 원형으로 추적할 수 있다는 점이다.

피정의항 =

- ① 주제분야
- ② 피정의항의 개념범주
- ③ 정의하는 개념이나 정의항
- ④ 정의항의 개념범주
- ⑤ 피정의항과 정의항의 관계
- ⑥ 피정의항의 본질적인 구별 특성(중차)
- ⑦ 기타 특성

3.1 정의모델의 기본 속성

본 연구에서는 Sager와 L'Homme(1994)의 정의모델과 이 모델을 확장한 Hudon(1996)의 정의모델을 기본모델로 적용하였다. 아울러 이 정의모델의 구체적인 적용지침은 다음과 같다(Sager and L'Homme 1994, 359-369).

3.1.1 주제영역

피정의항의 주제영역은 정의를 간결하고 정확하게 표현함과 동시에 정의의 모호성을 해소하기 위한 것이다. 이 요소를 통해 피정의항과 주제영역을 연결할 수 있다. 주제영역을 확인하기 어려운 것은 일반지식영역으로 분류한다.

매트릭스(matrix) :

- 지질학에서 광물을 포함한 암석
- 수학에서 수의 배열
- 건설에서 모래나 자갈 등 큰 골재를 고찰시키는 시멘트

이렇게 주제영역을 제시함으로써 상이한 주제영역에 속하는 개념을 지시할 수 있어, 일반 정의와 구분된다.

3.1.2 피정의항의 개념범주

개념범주는 피정의항과 개념체계를 연결하는 기능을 지닌다. 즉 피정의항의 개념범주는 정의되는 개념이 속한 범주를 확인하는 것으로, 일반적으로 물질적 실체, 추상적 실체, 활

동, 상태, 성질 등으로 구분되며, 기준에 따라 특정 주제영역에서 적절히 하위영역으로 전개될 수 있다(표 1).

실제로 ISO 2788(1986)에서는 색인어에 의해 표현되는 개념을 다음과 같이 구체적 실체, 추상적 실체, 개별적 실체, 고유명사로 표현되는 특정 개체로 구분하고 있다(표 2).

한편 ANSI/NISO Z39.19에서는 디스크립터로 표현된 개념의 일반범주를 다음과 같이 분류하고 있다(ANSI/NISO 1994).

- (1) 사물과 사물의 물리적 부분
 - 예) 조류, 기화기, 마이크로자료, 산, 유화(油畵)
- (2) 재료
 - 예) 접착제, 겨자 가스, 산소, 페인트, 타늄, 물
- (3) 행동 또는 과정
 - 예) 산성화, 빙하작용, 소형 골프, 마케팅, 그림 그리기

〈표 1〉 개념범주

개념범주	사 례	
물질(material)	물체	방간, 압력계
	유기체	세포, 포유동물
	시설	클럽, 학교
추상(abstract)	표현	기계언어
	지식분야	법학, 물리학, 종교
활동(activity)	행위	경주, 증류
	작용(영향)	고통, 성장
상태(state)		은퇴, 고용
성질(property)	질	색깔, 힘, 비틀림
	양	미터, 마하

〈표 2〉 ISO의 개념범주

개념범주	사 례
구체적 실체 ① 사물과 물리적인 부분 ② 재료	조류, 마이크로자료, 산악지역 접착제, 고무, 티타늄
추상적실체 ① 행위나 사건 ② 추상적 실체와 사물, 재료 및 행위의 특성 ③ 학문분야 및 과학 ④ 측정단위	빙하작용, 골프, 마케팅 유연성, 뉴스, 인성 고고학, 화학 헤르츠, 킬로미터
개별적 실체나 고유명사로 표현되는 특정개체	스리랑카, 국제보건기구

(4) 사건

예) 생일, 시민전쟁, 휴가, 혁명

(5) 사람, 사물, 재료 또는 행동의 특성이나 상태

예) 부력(浮力), 의식(意識), 유연성, 인성, 속도, 섬유

(6) 학문분야 또는 주제분야

예) 인류학, 유기화학, 정보학

(7) 측정단위

예) 헤르츠, 킬로미터

3.1.3 정의항

정의항은 피정의항을 개념체계에 포함된 관련된 용어와 연결하는 역할을 한다. 일반적으로 정의항은 류개념으로 제시되며, 관계구를 통해 피정의항과 연결되는 단일 개념이다. 여기에는 포섭관계나 상하개념, 한정된 상위개념, 관계구나 단어가 포함된다.

(1) 정의항은 일반적으로 피정의항과 동일

한 개념범주에 속한다.

팔 : 사지(四肢)의 한 부분

(2) 정의항은 피정의항과 가장 밀접한 개념이어야 한다.

마르(maars) : 화산 분화구

(3) 가장 밀접한 정의개념이 너무 특수하여 이용자가 쉽게 이해할 수 없는 경우, 정의항 자체에 대한 간략한 정의를 포함하여 정의를 확장하는 것이 바람직하다.

방사성동위원소 : 방사성을 지닌 동위원소 원자

화학원소의 최소 입자

(4) 정의하는 개념을 두 가지 이상으로 지시하는 것은 피해야 한다.

방수포(mack) : 코트 혹은 자켓

코트(이것만으로 충분하다)

(5) 하나의 개념으로 정의하는 것이 불가능한 경우, 복수의 정의항을 사용할 수 있다.

① 두개의 포섭개념을 가진 경우

첫 번째(first) : 먼저 오는 사물(something)이나 사람(somebody)

② 다수의 하위개념을 가진 경우

자철광 플라워링(magnetite flooring) : 시멘트와 톱밥의 합성

3.1.4 정의항의 개념범주

정의항의 개념이 속한 개념범주를 제시하기 위한 것으로, 대개 정의항의 개념범주는 피정의항의 개념범주와 동일하다.

3.1.5 피정의항과 정의항의 관계

이 요소는 피정의항과 정의항간의 관계를 제시하기 위한 것으로, 그 관계형태는 다음 중 하나이다.

〈표 3〉 피정의항과 정의항

	피정의항	정의항
용어	재료과학	과학
개념범주	추상적 개체	추상적 개체
하위범주	지식	지식

(1) 피정의항이 정의항에 제시된 상위개념의 유형이거나 범주, 사례인 경우, 이 관계는 생략될 수 있다.

트럼펫 : 금관악기(유형)

포유류(哺乳類) : 척추동물의 한 부류

추골(malleus hammer) : 중이(中耳)의 3개의 소골(小骨) 중 제일 바깥쪽의 골(서열)

(2) 피정의항이 정의항보다 하위개념으로서, 다음과 같이 정의된다.

① 피정의항이 상위개념의 일부분

아랫부리(mandible) : 조류의 부리를

구성하는 요소 중 하나

② 피정의항이 정의항의 한 영역

황반(macula) : 망막의 한 영역

③ 피정의항이 상위개념의 단위(양적 단위)

인시(man-hour) : 한 사람이 한 시간에 하는 일의 량

④ 피정의항이 상위개념의 단계(단계)

중기(metaphase) : 세포분열의 두 번째 단계

⑤ 피정의항이 상위개념의 경향, 관념, 스타일, 흐름

마니에리슴(mannerism) : 건축 양식

⑥ 피정의항은 포괄적 개념의 지시

매스 센터(mass centre) : 신체의 일부

(3) 피정의항이 정의항보다 상위개념인 경우, 다음과 같이 정의된다.

① 피정의항이 하위개념의 군(群)

미포(micelle) : 분자의 집합

아랫부리(mandible) : 짝으로 구성된 부리의 한 부분

② 피정의항이 하위개념의 합성

마그마(magma) : 화성암 덩어리

(4) 피정의항이 정의항과 추상화의 수준이 동일한 경우, 다음과 같이 정의된다.

① 피정의항이 다른 개념과 유사

팔(arm) : 사지와 같은 한 부분

② 피정의항이 다른 개념과 대조

정확성 : 오류 해제

자유 : 선택이나 행동에 제약이 없음

이상의 관계유형은 디스크립터의 기본관계인 등가관계, 계층관계, 연관관계와 관련하여 보면, (1)과 (2)는 계층관계의 상위어, (3)은

계층관계의 하위어, (4)는 관련어로 구분될 수 있다. 따라서 정의에 사용된 용어간의 관계구조를 시소러스 구축시 사용할 수 있는 것이다.

3.1.6 피정의항과 관련개념간의 구분 속성

이것은 전통적으로 종차라고 하며, 개념을 식별하는데 필요하고도 충분한 피정의항의 의미 속성을 제시하는 부분으로서, 이 속성을 통해 동일한 주제영역에 속한 관련된 개념과 구분된다. 이 종차에는 전문적이고 상세한 사항이 포함될 수 있기 때문에 이용자의 수준에 따라 이 필드를 복수로 제시할 수도 있으나, 본 연구에서는 정보전문가라는 하나의 수준으로 가정하였다.

이 속성은 적합한 특성탐색을 위한 기초를 제공할 수 있어 개념정의를 체계화할 수 있다. 그러나 이 속성을 파악하는 데에는 다양한 변수가 존재하기 때문에 체계적으로 공식화하기 어렵지만, 다음과 같은 특성유형을 확인할 수 있으며, 이 연구에서는 이 특성을 폐쇄적으로 사용하였다.

(1) 본질적 특성

① “포함한”, “구성된” 등과 같은 동사로 표현된 특성

마그네슘 석회암 : 백운암을 10%까지 포함한 석회암

말피기소체(malpighian body) : 모양의 끝으로 구성된 신단위(腎單位)의 일부

② “소유”나 “속성”을 표현한 특성

마그네슘 : 대기 중에서 백색 빛을 받으며 타는 금속

자석강(magnet steel) : 자성을 함유한 강철

③ 정의항의 선행수식어이거나, “의”, “인”, “된” 등으로 표현된 성질

마그네슘 : 매우 밝고, 은백색의 금속
큰 세공(macropore) : 0.005mm 이상인 바위 세공

(2) “원인”이나 “기원”을 표현한 특성

마그마 수(magmatic water) : 마그마에 의해 분출된 물

(3) “변형”이나 “수정”, “증가” 등으로 표현된 상태변화 특성

촉매 : 화학공정의 반응속도를 증가시키는 물질

리트머스 용지 : 산(酸)에 노출되었을 때 붉게 변하는 분말로 코딩된 종이

(4) “용도”로 표현된 용도특성

자석강(magnet steel) : 영구 자석으로 사용할 수 있는 강

직장(mace) : 고위공직자나 입법기관에서 사용하였던 지팡이

(5) 기능 또는 행위특성

아코오디온 : 입을 통해 바람을 불어넣어 리드(reeds)를 울려 소리를 내는 악기

대식세포(macrophage) : 박테리아를 섭취하는 식세포

(6) “위치”로 표현된 위치특성

대식세포 : 조직에서 발견된 식세포

황반 : 내이(內耳)의 과낭과 홈에 있는 세포조직

(7) 시간이나 연대특성

마그마분결작용(magmatic segregation) : 마그마 고화과정 동안의 광물 농축

마그마 수(magmatic water) : 마그마

〈표 4〉 정의분석

정의항	교환을 위해 수령자가 분리된 엔티티들을 재구성할 수 있는 방식으로 몇 개의 엔티티들 내에 저장될 수 있는 주문서와 그 관련문서들을 단일 데이터열로 결합할 수 있게 하는 자료구조
정의모델에 의해 분해된 요소	<ol style="list-style-type: none"> 1. 주제분야 : 정보산업-정보기술-SGML 2. 피정의항의 개념범주 : 추상적 실체 3. 정의항 : 자료구조 4. 정의항의 개념범주 : 추상적 실체 5. 관계 : 자료구조(상위개념), 주문서(하위개념), 관련문서(하위개념) 6. 특성 : (본질적 특성) 엔티티들내에 저장될 수 있는 주문서와 그 관련문서들을 단일데이터열로 결합할 수 있게 한다.

과정 동안 분출되는 물
 (8) "유사성"으로 표현된 유사특성
 막(membrane) : 종이와 유사한 구조
 아크릴 수지(acrylic resin) : 유리 와 유사한 플라스틱

터의 의미와 사용된 용어간의 관계를 설정하는데 사용될 수 있다. 예컨대 용어 'SGML 문서교환형식'에 대한 정의는 이 정의모델에 의거하여 〈표 4〉와 같이 분석될 수 있다.

3.1.7 기타 특성

이 특성은 본질적 특성 이외에 피정의항과 관련된 개념을 구별하는데 필요한 특성으로서 선택적으로 사용된다. 일반적으로 기타 특성은 피정의항의 적용범위나 정의항의 범위를 제한하거나 사례를 제시한 것이다.

매카시즘 : 반대에 기반한 이데올로기로서, 특히 확증이 없는 비난에 근거한 것

3.2 정의모델의 적용

정의모델을 적용하여 정의를 작성하거나, 기존의 정의에 적용하여 표준화된 정의를 작성할 수 있으며, 아울러 정의에 포함된 정보를 각 구성요소로 분석하여 시소러스 구축시 디스크립

4 용어정의 시소러스 구축

4.1 시소러스 항목추출

4.1.1 정의의 재정의

정의를 이용하여 시소러스를 구축하기 위해서는 각 용어가 표준정의모델에 따라 정확하게 정의되었음을 전제로 한다. 본 연구의 실험 대상인 한국산업규격에 나타난 정의는 국가표준기관에서 제작된 것이므로 기본적으로는 완전한 정의로 가정하였다(한국표준협회 2000). 그러나 실제로 '정보산업분야'의 규격 중 많은 내용이 국제표준화기구(ISO)의 것을 그대로 도입, 번역하였거나 같은 용어일지라도 각 규격마다 약간씩 달리 정의되어 있어

〈표 5〉 정의의 재정의

데이터열 문자집합(data stream character set)	
원문	변경후
SGML 문서교환형식 데이터열 내의 모든 문서 마크업, 그리고 모든 SGML 문서교환형식 명칭과 SGML 문서교환형식 식별자에 대해서 사용되는 문자집합	SGML의 문서교환형식 데이터열 내의 SGML 문서교환형식 명칭과 문서교환 형식식별자에 대해서 사용되는 문자집합
SGML 문서교환형식 명칭(SDIF name)	
원문	변경후
SGML 문서교환형식 데이터열 전체인 데이터열 또는 SGML 문서교환형식 데이터열에서 문서에 할당된 SGML 명칭	SGML의 문서교환형식 데이터열에서 데이터열 또는 문서에 할당된 SGML 명칭

뢰도가 그다지 높지 않은 것으로 평가되었다. 또한 기술된 내용도 앞서 제시한 정의모델에 근거하여 볼 때 완전하지 못하다고 판단되는 정의는 부분적으로 〈표 5〉와 같은 형식으로 다시 정의하였다.

4.1.2 시소러스 항목 선정

정의모델은 다양하게 기술된 용어의 정의를 표준화할 수 있는 중요한 도구이며, 이에 근거하여 작성된 정의는 시소러스에 '정의항목'으로 부가될 수 있다. 또한 정의를 구성하는 과정이나 기존의 정의를 분석하여 나타난 정의모델의 각 속성은 디스크립터의 관계구조에 반영하여 시소러스를 확장할 수 있다. 본 연구에서는 시소러스 관계구조를 확장하기 위해 정의모델속성 중 필요한 요소를 다음과 같이 추출하였다.

1) 주제분야

한국산업규격(KS)의 정보산업(X)분야 544

개의 규격 중 54개를 임의로 선택하여 이들 규격에 포함된 '용어의 정의' 및 '용어의 뜻'에 제시된 정의 788개를 수집하였다. 따라서 정의의 주제분야는 '정보산업'이 되며, 또한 이 분야는 다음과 같이 6개 부문으로 세분되어 있으므로, 각 부문의 주요어를 하위주제로 선정하고, 규격명칭 중 주요어를 최하위주제어로 선정하였다(표 6).

예컨대 'KS X 6010 : SGML 문서교환형식'에서는 '정보산업'을 주제분야로, 하위주제는 '정보기술', 최하위주제는 'SGML'로 구성하였다. 제3장에 제시한 정의모델보다 여기서 주제분야를 세분한 이유는 동일한 용어에 대한 정의라도 규격마다 다소 상이한 정의가 나타나기도 하거나 용어의 의미를 더 분명히 제시하기 위해서는 세부주제를 정하는 것이 효과적이라 판단되었기 때문이다.

2) 정의항

용어의 정의를 시소러스에 도입하는 것은 용

〈표 6〉 KSX의 세부항목(괄호안의 규격갯수)

정보처리용어 · 일반(KS × 0001~1000) (39개)
문자세트 · 보안 · 부호화(KS × 1001~2000) (84개)
정보기술용 언어 · 소프트웨어 · 컴퓨터그래픽스(KS × 2001~3000) (91개)
데이터 통신망(KS × 3001~5000) (103개)
정보기기 · 데이터 저장매체(KS × 5001~6000) (100개)
정보기술(IT) 응용(KS × 6001~7000) (127개)

어의 의미를 명확히 하여 이용자가 용어의 개념을 쉽게 인지하기 위한 것이다. 보편적으로 정의항은 피정의항을 특정한 개념체계와 관련된 것은 단일개념으로 표현한 것으로, 피정의항과 가장 밀접한 개념을 기술해야 한다. 본 연구에서는 규격에서 정의된 내용을 그대로 수용하되, 일부 재정의하였으며, 이를 통해 피정의항과 정의항에 표현된 중요한 용어간의 상관관계를 볼 수 있다.

(1) 상위 · 하위 개념

응용프로그램 : 파일의 내용을 처리하고, 파일 또는 파일이 기록된 볼륨의 속성 중 일부를 처리할 수 있는 프로그램

(2) 한정된 상위개념

기판 : 광 빔이 투과하여 기록층에 입사하는 디스크 구조층

(3) 관련 단어나 구를 사용한 동일 수준의 추상적 개념

광디스크 : 광 빔에 의해 데이터를 기록할 수 있는 기록층을 가진 평평한 회전원반

3) 피정의항과 정의항내의 용어와의 관계
피정의항과 정의항에 사용된 용어간의 관계

는 정의를 통해 분명하게 파악할 수 있는데, 이를 시소러스의 기본관계에 적용하여 다음 중 하나로 나타내었다(ISO 2788 : ANSI/NISO 1994).

(1) 피정의항이 하위개념

관계가 제시되지 않았거나 피정의항이 정의항의 한 부분으로 파악되는 경우로서, 관계유형은 '피정의항(NT) : 정의항 내의 용어(BT)'로 제시하였다.

① 피정의항이 정의항의 일반적 개념의 특성

오류정정부호 : 오류의 발견과 정정에 사용되는 비트를 만들어내는 수학적 절차

② 피정의항이 정의항의 개념의 부분

기준구멍 : 카세트 테이프를 장치에 부착할 때 사용되는 위치 설정의 기준이 되는 구멍

③ 피정의항이 정의항의 일부영역

외부구역 : 디스크 표면의 기록 영역 중 외부영역

④ 피정의항이 정의항의 개념의 측정방법

기록밀도 : 트랙의 길이 1mm당 기록된 자속반전수(ftpmm)

⑤ 피정의항이 정의항의 개념의 단계

파일저장동작 : 가상파일 저장정의의 일
부로 규정한 동작의 하나

⑥ 피정의항이 정의항의 개념의 경향,
관념, 스타일

파일 : 이름을 붙인 정보의 모임

⑦ 피정의항이 정의항의 포괄적 개념의
일부 참조

기준자계 : 표준 플렉시블 디스크 카트
리지의 전형적 자계

(2) 피정의항이 상위개념

피정의항이 정의항의 용어에 비해 상위개념
으로 파악되는 경우로서, 관계유형은 '피정의
항(BT) : 정의항의 용어(NT)' 로 제시하였다.

① 피정의항이 정의항의 개념균일 때

키 타이틀 : ISDS에서 연속간행물에 부
여한 ISSN과 모든 표제

② 피정의항이 정의항의 합성물일 때

익스텐트 : 연속되는 오름순의 블록번호
를 가진 논리블록의 모임

(3) 피정의항과 정의항의 용어가 동일 수준
의 개념

정의항이 피정의항의 용어와 유사한 개념이
거나 대조적인 다른 개념으로 파악되는 경우로
서, 관계유형은 '피정의항(RT) : 정의항내의
용어(RT)' 로 제시하였다.

① 피정의항이 정의항의 다른 개념과 유사
애퍼처 카드 : 마이크로필름 종이카드와
동어

② 피정의항이 정의항의 개념과 대조적
인 경우

프리 레코드 마크 : 광자기 기법으로는
기록내용을 바꿀 수 없는 마크

4) 구분특성(종차)에 의한 패시 적용

피정의항의 의미특성은 피정의항을 다른 개
념과 구별하기 위한 것으로, 이는 이용자의 수
준을 고려할 때 다양한 수준의 특성이 가능하
다(Nkwenti-Azeh 1994, 379). 그러나 여기
서는 이용자를 전문가 수준에 한정하며, 앞서
제시한 정의모델에 근거하여 다음과 같은 특성
을 파악하였다.

(1) 본질적 특성

① "포함하다", "구성하다", "속하다",
"식별하다", "가리키다" 등의 동사로 된
문장

② "가지다", "소유하다", "참여하는",
"~중의", "~와 함께" 등으로 표현된 문
장

③ "~인", "~이다", "~가 되다" 등으
로 표현된 문장

문서기술자 : 직접 또는 간접적으로 참
조되어 모든 외부 엔티티들에 대한 엔티
티 기술자들을 포함하는 SGML 문서교
환형식 데이터열의 성분

(2) 기원 특성

"~로부터 나타난", "~에 의해 만들어
진", "~의 결과로서" 등으로 표현된 문장
코드값 : 압축 알고리즘에 의해 생성된
0~4095 범위의 수

(3) 상태/변화 특성

"~에서 변형된", "~에서 수정된", "유
지하는" 등으로 표현된 문장

트레일러 : 압축과 패드 비트 추가 후의
블록에 첨가된 데이터

(4) 목적/용도 특성

“~를 위한”, “~를 위해 사용된” 등으로 표현된 문장

데이터열 문자집합 : SGML의 문서교환형식 데이터열 내에서 문서교환형식 명칭 과 문서교환형식 식별자에 대해 사용되는 문자집합

(5) 기능/행위 특성

“~에 적합한”, “~을 생성하는”, “나타 내는” 등으로 표현된 문장

제어 데이터 항목 : 내용의 변화가 제어 단절을 일으키는 데이터 항목

(6) 위치 특성

“~에 위치한”, “~에 존재한”, “~에 있는”, “~에 가까운” 등으로 표현된 문장

트랙피치 : 반지름 방향에 이웃한 트랙 중심선 사이의 거리

(7) 시간 특성

“~동안” 등으로 표현된 문장

경로선택 제어 : 특정 네트워크, 링크 또는 중계를 선택하거나 피하기 위해 경로 선택과정 동안 규칙의 적용

(8) 유사 특성

“~와 같이”, “~와 유사하게” 등으로 표현된 문장

디렉토리 : 책의 목차와 같은 것으로 어떤 필드가 레코드내 어느 위치에서 시작되며, 길이가 얼마인가를 지시해 주는 디렉토리 항목의 집합으로.....

4.2 시소러스 구축 및 검사

정의와 정의에 포함된 요소를 시소러스의 관계구조로 표현하기 위해, 수집된 788개의 정의

를 앞 절의 분석과정을 통해 분석한 결과, 중복되거나 불완전한 정의를 제외한 327개의 개념 정의를 대상으로, <표 7>과 같은 구조로 데이터 베이스를 구축하였다.

<표 7> 용어 데이터베이스

```

<term>용어</term>
<eng>term</eng> <!-- 선택가능-->
<tit>용어의 주제</tit> <!-- 선택가능-->
<sub tit>하위 주제</sub tit>
<def>용어의 정의</def>
<relation> <!-- 복수선택가능-->
    <bt>용어의 상위개념</bt>
    <nt>용어의 하위개념</nt>
    <rt>용어의 관련개념</rt>
</relation>
<facet> <!-- 복수선택가능-->
    <int>본질적 특성</int>
    <ori>기원 특성</ori>
    <sta>상태/변화 특성</sta>
    <usa>목적/용도 특성</usa>
    <fun>기능/행위 특성</fun>
    <loc>위치 특성</loc>
    <tim>시간 특성</tim>
    <sim>유사 특성</sim>
</facet>
    
```

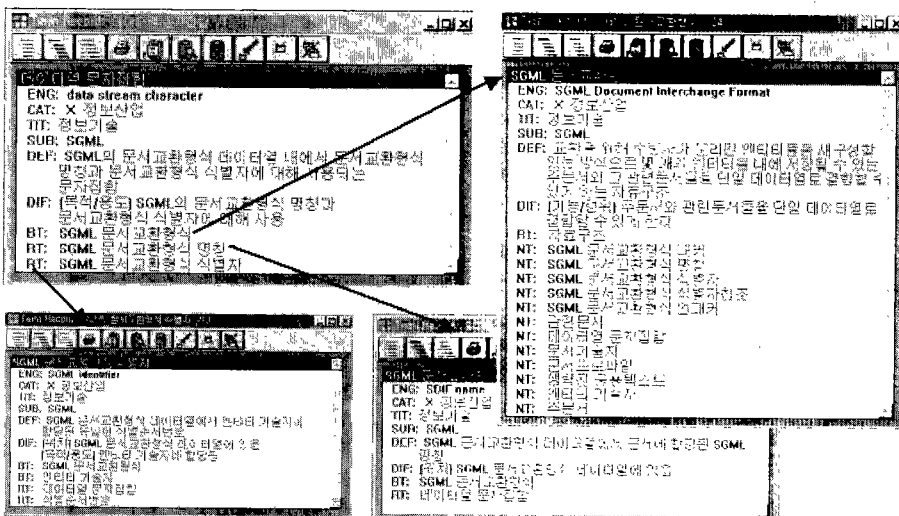
예컨대 ‘데이터열 문자집합’의 정의를 위에 제시한 정의모델에 따라 각 속성으로 분석한 후, 위의 레코드 구조에 따라 <표 8>과 같이 구성하였다.

이 방법으로 구축된 데이터베이스를 용어정의 시소러스로 활용하기 위해 두 가지 방법으로 검증하였다. 첫 번째는 ANSI/NISO Z39.19의 지침과 권고에 따른 상용 시소러스 프로그램인 MultiTes를 이용한 것으로, <그림 3>은 디스크립터와 관련된 용어의 연결구조를 제시한 것이다.

〈표 8〉 '데이터열 문자집합'의 구조

```

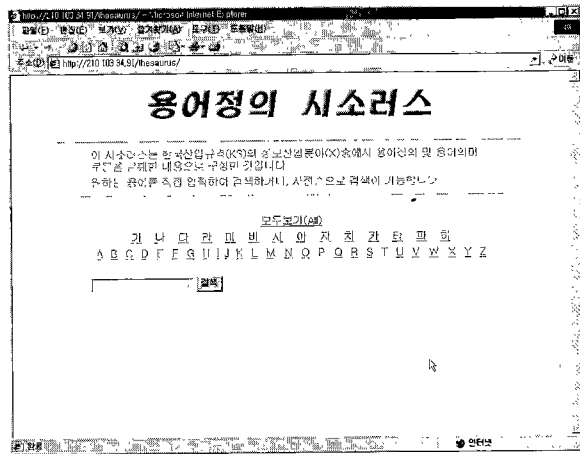
<term>데이터열 문자집합</term>
<eng>data stream character set</eng>
<tit>정보기술</tit>
<sub_tit>SGML<sub_tit>
<def>SGML의 문서교환형식 데이터열 내에서 문서교환형식 명칭과 문서교환형식 식별자에 대해 사용되는 문자집합</def>
<relation>
    <bt>SGML 문서교환형식</bt>
    <rt>SGML 문서교환형식 명칭</rt>
    <rt>SGML 문서교환형식 식별자</rt>
</relation>
<facet>
<usa>SGML의 문서교환형식 명칭과 문서교환형식 식별자에 대해 사용</usa>
</facet>
    
```



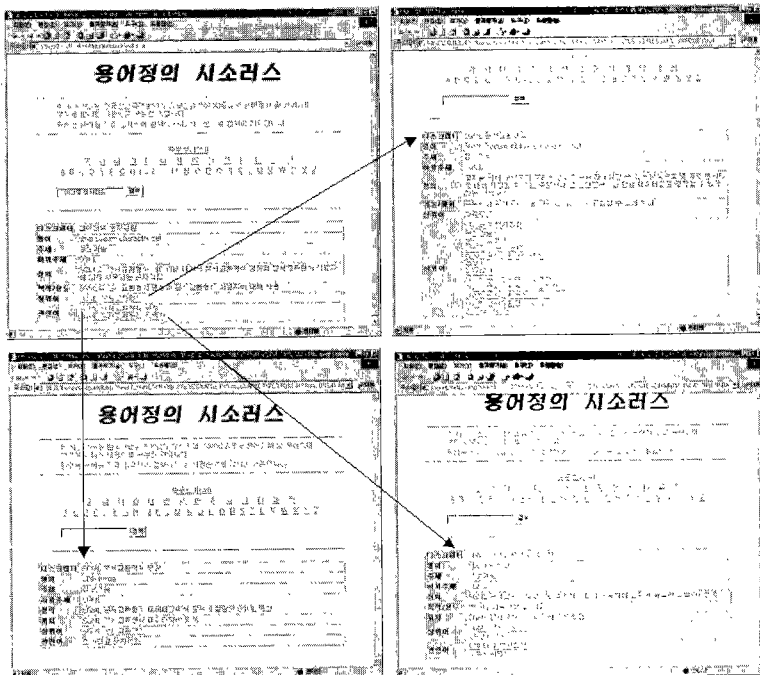
〈그림 3〉 MultiTes를 이용한 용어정의 시소러스

이 프로그램을 이용하여 검사한 결과, 본 연구에서 제시한 주제나 정의항목, 용어관계를 제시하는데는 문제가 없으나, ANSI/NISO Z39.19의 지침에 나타나지 않은 특성(중차)항목을 개별적으로 표현할 수 있는 관계구조를 제시할 수 없으며, 또한 동일 용어에 대한 복수의 정의를 별개로 표현할 수 없는 문제가 나타

났다. 따라서 두 번째 검사에서는 기본적인 시소러스 지침을 수용함과 동시에, 구분특성항목을 개별적으로 제시할 수 있는 관계구조를 만들고, 동일한 용어이면서 개념상 상이한 디스크립터를 동시에 관계구조를 가지도록 Window NT환경에서 ASP로 구축한 프로그램을 수행하였다(용어정의 시소러스 2001). 〈그



〈그림 4〉 용어정의 시소러스 초기화면

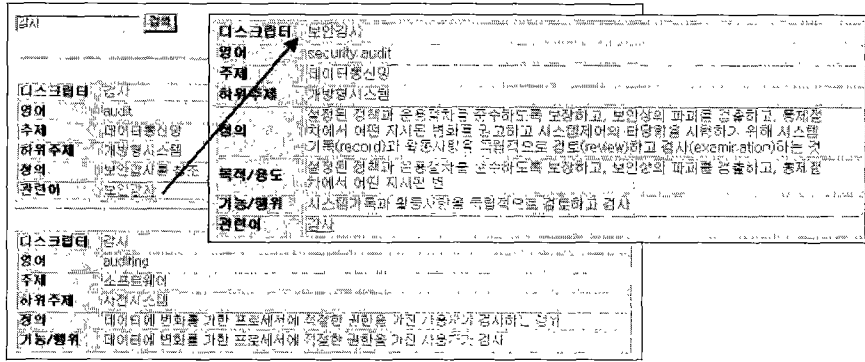


〈그림 5〉 ASP로 구축된 용어정의 시소러스

림 4)는 용어정의 시소러스의 초기화면으로서, 구축된 시소러스의 디스크립터를 검색할 수 있도록 [모두보기], [사전순보기], [직접검색] 형

태의 방식을 제시하고 있다.

또한 〈그림 5)는 원하는 디스크립터를 직접 검색한 결과와 그 디스크립터의 관련어간의 연



〈그림 6〉 복수의 정의를 가진 용어의 연결

결상향을 제시한 것이다. 참고로 이 시스템은 <http://211.223.195.133/thesaurus/>에서 검색할 수 있다.

이와 함께 복수의 정의와 관련된 용어를 한 화면에 제시함과 동시에 각 용어의 관계나 특성을 개별적으로 파악할 수 있도록 하였다. 예컨대 '감사'라는 용어는 적용되는 주제영역에 따라 달리 정의되며 〈그림 6〉은 그 내용과 관계구조를 제시한 것으로, 화살표는 연관어의 시소러스내용을 표현한 것이다.

5 결 론

특정 주제영역의 시소러스에 수록되는 디스크립터의 의미범위를 한정하고, 개념의 본질적인 속성을 제공하기 위해 정의를 도입하고, 정의를 구성하는 데이터요소를 정의모델에 따라 분석한 후, 이를 시소러스 관계구조로 표현하기 위한 연구이다. 이를 위해 한국산업규격

(KS)의 정보산업분야에 수록된 용어정의 327개를 정의규칙과 정의모델에 따라 분석한 다음, 각 속성을 디스크립터간의 관계구조로 표현한 시소러스를 실험적으로 구현하였다.

지금까지 시소러스 구축시 개발주체의 성격에 따라 디스크립터의 수록범위와 관계구조의 설정을 달리해 왔다. 용어의 정의를 시소러스 구조에 도입하게 되면 각 디스크립터의 의미범위를 한정할 수 있고, 동시에 정의에 사용된 용어간의 의미관계를 디스크립터 관계구조로 표현할 수 있어, 의미범위와 관계구조의 표준화 가능성을 기대할 수 있다.

본 연구와 관련하여 후속되어야 할 연구영역으로는 전문용어에 대한 용어학적인 정의를 내리는 연구가 필요하다고 판단되었다. 실제 원시데이터로 사용된 한국산업규격의 정의가 표준화되어 있지 않아 모델에 적용하기 어려운 부분이 많았으며, 아울러 용어간의 의미요소를 식별하기 위한 중차에 대한 연구가 수반되어야 할 것으로 판단된다.

참고 문헌

- 김태수. 2000. 『분류의 이해』. 서울 : 문헌정보처리연구회.
- 용어정의 시소러스. 2001. [2001. 2. 15. 인용] <<http://211.223.195.133/thesaurus/>>
- 李秉根. 1992. 辭典 定義의 類型과 原則, 『새국어생활』, 2(1) : 2-21.
- 한국표준협회. 1996. 『KS X 1517-4 데이터 요소 - 제4부 : 데이터 요소의 정의를 위한 규칙과 지침』. 서울 : 한국표준협회.
- . 1999. [2000. 10. 15. 인용] <<http://standard.ksa.or.kr>>.
- Aitchison, Jean., Alan Gilchrist, and David Bawden. 1997. *Thesaurus Construction and Use : a Practical Manual*. 3rd ed. London : Aslib.
- ANSI/NISO. 1994. *ANSI/NISO Z39.19 : Guidelines for the Construction, Format, and Management of Monolingual Thesauri*. Bethesda, MD : NISO press.
- Canadian Literacy Thesaurus(CLT). 1996. [2000. 9. 10. 인용] <<http://thesaurusalpha.org>>.
- Dahlberg, I. 1989. "Concept and Definition Theory". In *Classification theory in the Computer age : Conversations Across the Disciplines. Proceedings from the Conference, November 18-19, 1988, Albany, New York*. Edited by Batty, David., Ingetraut Dahlberg. Irene L. Travis, and Elaine Svenonius. New York : State University of New York. 12-24.
- Hudon, Michéle. 1992. "Term Definitions in Subject Thesauri: The Canadian Literacy Thesaurus Experience." *Classification Research for Knowledge Representation and Organization*. Edited by Williamson, Nancy J., and Michéle Hudon. Amsterdam : Elsevier. 255-262.
- . 1996. "Preparing Terminological Definitions for Indexing and Retrieval Thesauri : A Model". *Advances in Knowledge Organization*, vol. 5. Edited by Green, Rebecca. Frankfurt/ Main : Indeks Verlag. 363-369.
- ISO. 1986. *ISO 2788 : Guidelines for the Establishment and Development of Monolingual Thesauri*. 2nd ed. Geneva : ISO.
- . 1987. *ISO 704 : Principles and Methods of Terminology*. Geneva :

- ISO.
- . 1990. *ISO 1087 : Terminology - vocabulary*. Geneva : ISO.
- . 1994. *ISO 11179 : Information technology-Specification and standardization of data elements, Part 4 : Rules and guidelines for the formulation of data definitions*. Geneva : ISO.
- Moors, C. N. 1963. "The Indexing Language of an Information Retrieval System". *Information Retrieval Today : papers presented at an Institute conducted by the Library School and the Center for Continuation Study, University of Minnesota*. Ed. by Wesley Simonton. Minneapolis, MN : The Center. 21-36.
- MultiTes. [2000. 8. 20. 인용] <[http : // www.multites.com](http://www.multites.com)>.
- NASA Thesaurus. 1998. [2000. 9. 18 인용] <[http : //www.sti.nasa.gov/thesfirm1.htm](http://www.sti.nasa.gov/thesfirm1.htm)>.
- Ndi-Kimbi, Augustin. 1994. "Guidelines for Terminological Definitions : The Adherence to and Deviation from Existing Rules in BS/ISO 2382 : Data Processing and Information Technology Vocabulary". *Terminology*, 1(2) : 327-350.
- Nkweni-Azeh, Blaise. 1994. "The Use of Thesaural Facets and Definitions for the Representation of Knowledge Structures". *Advances in Knowledge Organization*, vol. 4. Edited by Albrechtsen, Hanne., and Susanne Oernager. Frankfurt/Main : Indeks Verlag. 374-381.
- Sager, Juan C. 1990. *A Practical Course in Terminology Processing*. Amsterdam : John Benjamins.
- Sager, Juan C., and Marie-Claude L'Homme. 1994. "A Model for the Definition of Concepts : Rules for Analytical Definitions in Terminological Databases". *Terminology*, 1 (2) : 351-373.
- Sager, Juan. C., and Augustin Ndi-Kimbi. 1995. "The Conceptual Structure of Terminological Definitions and their Realizations : a Report on Research in Progress". *Terminology*, 2(1) : 61-81.
- Soergel, D. 1974. *Indexing Languages and Thesauri : Construction and Maintenance*. Los Angeles : Melville.
- Strehlow, R. A. 1983. "Terminology and the Well-formed Definition". *Standardization of Technical Terminology : Principles and Practices*, ASTM Special Technical Publication, 806. 15-25. Edited by Interrante, C. G., and F. J. Heymann, Philadelphia, PA : American Society for Testing and Materials.

WordWeb, 1999. [2000. 7. 19. 인용] <<http://www.wordweb.co.uk>>

Wright, Sue Ellen and Gerhard Budin, eds.
1997. *Handbook of Terminology*

Management, volume 1 : Basic
Aspects of Terminology
Management. Amsterdam : John
Benjamins.