

archy, therefore the terms that are selected through associative relationship aren't well controlled.

This study examines the relationships of existing thesauri, especially the types and ranges of associative relationship, and suggests the adequate type of associative relationship.

KEYWORDS

thesaurus, association, associative relationship type, term relationship, equivalence, hierarchy

I. 서론

시소러스는 정보검색시스템의 필수적인 요소로서, 자동색인시에는 시소러스의 디스크립터로 색인어를 추출하고 검색시에는 이용자의 질의어를 디스크립터로 바꾸어줌으로써 재현율과 정확률을 높이는 역할을 수행한다. 특히 훌륭한 시소러스의 사용은 유용한 단어를 찾는 것 이상의 도움을 주는데, 이용자가 잘 알지 못하는 분야를 탐색할 경우 시소러스는 그 주제를 보다 잘 이해할 수 있도록 도와주며, 이용자가 질의어로 생각하고 있는 용어들간의 관계를 다양하게 보여줌으로써 정보검색에서 흔히 일어나는 착오검색(false drop)의 발생을 감소시킬 수 있다.

그러므로 시소러스가 정보검색에 유용하게 적용되기 위해서는 용어사이의 정확하고 다양한 관계유형의 설정이라 할 수 있을 것이다. 그러나 기존의 시소러스는 용어사이의 관계를 단순하게 표현하며, 특히 연관관계의 경우는 그 정의나 범위가 모호해서 색인, 검색시 용어 통제가 잘 되지 않으며 방대한 지식을 표현하는 시스템에서는 적절하지 못한 것으로 지적되고 있다.

따라서 본 고에서는 전반적으로 시

소러스의 관계유형을 분석하였으며, 특히 연관관계의 표준, 유형, 범위를 중심으로 문제점들을 살펴보고, 다양한 관계 유형을 제시한 시소러스 관련 연구를 분석함으로써 효율적인 연관관계 유형을 제안하고자 한다.

II. 시소러스 관계유형 연구

밀스테드(J. L. Milstead)는 1960년대 이후부터 간단한 동의어 리스트에서 의미네트워크와 지식베이스까지 의미적 접근을 보조하기 위한 시도로 시소러스에 다양한 방법이 사용되고 있지만 시소러스의 기본적인 설계와 목적은 변하지 않고 있음을 지적하였다.

또한 슈미츠 에세르(W. Schmitz - Esser)는 새로운 시소러스의 확장방안으로 전문가 시스템, 인터페이스 시스템, 객체지향 설계 및 프로그래밍, 하이퍼텍스트 시스템, 기계 번역, 기계 초록 등을 고려하였다. 최근에 시소러스는 언어학, 인공지능, 지식 조직화에 사용된 관련 기술을 받아들임으로써 시소러스의 적용범위가 확대되고 있어서 시소러스 관계유형을 더 정교한 방법으로 확장하고 재정리해서 표준화할 필요가 있음을 주장했다.

시소러스에 대한 새로운 접근과 기능이 요구되고 있지만 시소러스의 가장 중요한 관계유형은 동등, 계층, 연관관계로 단순화되어 있으며, 특히 연관관계는 비체계적으로 되어 있어서, 오히려 정보검색에 좋지 못한 결과로 작용하는 경우가 많다. 이에 대해 시소러스의 전반적인 관계유형을 다양하게 적용하는 연구가 계속 있었으며, 여기서 관계유형은 실세계 지식을 개념간의 관계로 할당하는 것으로 보았다.

패러데인(J. E. L. Faradane)이 개발한 관계색인(Relational Indexing)은 길포드(J. P. Guilford)의 지식 구조도의 개념 조직을 응용한 색인법으로 개념과 이들 개념간의 관계를 조직할 수 있는 방법이 도입되었다.

관계색인은 구문적으로 고려되었다기보다 의미에 따라 달라지므로 특히 시소러스의 상황에 적합하다. 패러데인은 사고과정의 분석결과 인간에게는 개념을 서로 연결시키는 두 가지 기본 기능이 있는데 하나는 '연상'이고 다른 하나는 '식별'로서 이에 따라 9개의 관계로 구성된 관계유형을 제시하였다. 9개의 관계유형은 동시, 자동행위, 연상, 동등, 차원, 종속, 식별, 행위, 기능적 의존관계이다.

윌렛츠(M. Willetts)는 패러데인의 관계색인을 시소러스의 관계형으로 범주화했다. 먼저 패러데인의 관계유형중 동등관계나 계층관계로 확실히 표현될 수 있는 것을 선정한 후, 그것들을 제외한 나머지를 연관관계로 규정하였다.

연관관계는 크게 두 가지로 구분되는데 분명한 관계를 지니는 것과 불분명한 관계를 지닌 것이 있다. 불분명한 관계에는 동시관계의 일부(연상 일부)가 해당된다. 분명한 관계는 식별관계

의 일부, 행위관계, 기능적 의존관계, 자동행위관계의 일부, 연상관계의 일부, 차원관계의 일부가 해당된다.

패러데인의 관계색인은 구문관계와 개념관계를 사용한 구조적 접근으로 볼 수 있다. 그러나 이러한 구문관계와 개념관계를 동시에 사용하면 문헌의 내용을 완벽하게 표현하기 어렵다는 점이 있다.

이러한 점을 고려하여 "어휘관계를 축소시키고 편리한 형태로 필요한 의미정보를 캡슐화"한 어휘-의미관계로 접근하는 시도가 있다. 다양한 분야에 적용될 수 있는 어휘-의미관계는 질문-해답 작성, 실제 지식의 표현, 텍스트 생성을 위해 의미를 지닌다.

폭스(E. A. Fox)는 질의어에 어휘-의미관계를 갖는 용어를 추가해서 확장된 질의어로 검색효율성을 향상시켰다. 어휘-의미관계를 가진 용어와 용어링크를 포함한 리스트를 11개의 그룹으로 범주화했다.

그 뒤를 이어서 왕(Y-C. Wang), 반덴도르프(J. Vandendorpe), 에번스(M. Evens)는 43개의 '어휘-의미관계'를 5개의 집단으로 포괄한 관계형 시소러스(Relational Thesauri)를 설계하였다.

샤핀과 헤르만(R. Chaffin & D.J. Herrmann)은 관계를 주제별로 모아줌으로써 관계의 분류를 실제로 구축했고 이와 관련된 관계요소이론(Relation Element Theory)을 제시했다. 관계자체는 31개로 되어 있으며 다수의 기본적인 요소로 분해할 수 있다. 31개의 관계를 5개 집단, 즉 대조(Contrasts), 유사(Similars), 클래스 포함(Class Inclusion), 격관계, 부분-전체로 구성하였다. 31개 관계자체는 주제를 표현하는데 얼마나 유사하냐에 따라 요소별로 분해될 수 있다.

사용된 요소의 일부로 양극성(Bipolarity : 중간지점을 두고 반대측면), 차원(Dimension : 단일 차원의 공유)과 같은 항목을 포함하고 있다. 관계 유형중 비양립성 관계(예, 솔직-위선)는 차원적 양극성(DIM(BIP)으로 표기)으로 볼 수 있다. 그러나 이 연구는 의미있는 계층을 개발하는데 상당한 시간이 소요되는 단점이 있다.

III. 시소러스 연관관계 유형 분석

1. 기존의 시소러스 연관관계 유형 분석

시소러스 설계시 문제가 되는 것 중의 하나는 용어간의 관계유형이다. 대부분의 시소러스에서 용어사이의 의미와 구조를 정확하게 표현하는 것은 어렵다. 특히 동등관계와 종속관계는 논리를 기반으로 간단하고 확실하게 표현될 수 있지만, 연관관계는 주제를 표현하는 용어의 의미로부터 문헌의 적합한 영역의 경계를 정하는 것이 어렵다. 특히 후조합 색인에서는 용어의 결합으로 무한한 주제를 표현할 수 있으므로 연관관계의 범위를 정하기 어렵다.

시소러스에서 연관관계는 그 유용성이 불분명하고, 관계범위를 설정하는데 직접적인 규칙도 없이 계속 사용되어 왔다. 이러한 이유로 시소러스의 용어 관계를 의미에 따라 세분화하는 경향이 있으나 이론적으로 어려움이 있으며 타당하지 못하거나 신뢰할 수 없는 경우도 있다. 설사 실험적으로 좋은 성과를 얻었다 하더라도 실제로 적용하는데는 어려움이 따른다.

따라서 본 장에서는 시소러스의 연관관계 표준과 실제 사용되고 있는 시소러스의 연관관계 유형과 범위를 살펴봄으로써 그 각각의 문제점을 알아본다.

1.1 시소러스 연관관계 표준

베르직과 노벨링(G. Wersig & U. Noveling)은 “계층관계가 아닌 개념이나 포괄적인 주제간의 관계”로 연관관계를 정의하였다.

프랑스 AFNOR Z47-100 표준에서는, 연관관계를 가진 용어는 정보검색 목적상 필요하다면 시소러스에 수록되어야 하지만 그 기준을 정하기 위한 규칙은 없다.

ANSI/NISO Z39.19에서는 연관관계의 특성을 명확히 하고, 주관적인 판단을 배제할 것을 강조하면서 그 유형을 다음과 같이 범주화하였다.

첫째, 같은 범주에 속하는 디스크립터간의 관계로서 배타적인 형제용어간의 관계를 연관관계에서 제외시키고 파생관계나 혈통관계를 연관관계에 포함시킬 것을 권고하였다.

둘째, 다른 범주에 속하는 디스크립터간의 관계로 다른 계층에 속하는 디스크립터간을 관련짓기 위한 여러가지 기준을 만들 수 있다. 구체적인 내용은 ISO 표준과 유사하다. 그렇지만 한가지 다른 사항은 실제로 상위어가 아닌 명사구의 경우로서 해당 용어가 중심어(head noun)의 하위어에 해당되지 않는 형용사나, 실제로 그 주제에 속하지 않는 형용사를 포함하는 경우이다. 예를 들어 rubber ducks은 duck의 하위어가 아니며, 이와 같은 용어로는 artificial limbs, paper flowers 등이 해당된다.

ISO 2788에서는 연관관계를 개념적으로 밀접하게 연관되어 있으나, 계층관계나 동등관계에 속하지 않는 관계로 규정하고 있다. 연관관계에 있는 용어는 색인작성과 탐색에 이용될 가능성이 있는 대체용어를 제시한다는 관점에서 볼 때, 연관된 용어간을 명시하는 것이 좋을 것이라는 심리적인 연상에 따라 만들어진 것이다.

연관된 용어의 선택에 있어서 엄격한 통제를 수행하는 것과 주관적 판단을 금하는 것이 중요하다. 색인이용자의 사고의 범위에 근거하여 한쪽의 용어가 색인어로서 채택되는 경우 항상 다른 쪽의 용어가 강력하게 암시되는 것을 관련어로 제안하고 있다. 그리고 표준규격은 몇 개의 용어가 밀접하게 관련되어 있을 때에는 한쪽의 용어가 다른 쪽의 용어를 정의하거나, 설명하기 위하여 필요한 구성요소인 경우가 많다고 기술하고 있다. 예를 들면 “새”는 “조류학”을 정의하기 위하여 필요하나 그 반대는 필요하지 않다.

상이한 범주에 속하는 연관어를 선정하기 위해 많은 노력이 행해져 왔으며 아래 부류는 실제 전형적인 연관관계의 예를 제시하고 있다.

- 학문분야, 연구영역, 연구의 대상과 현상 (임학-산림학)
- 조작, 과정과 행위자 또는 도구(속도측량-속도계)
- 행위와 그 행위의 결과(도로건설-도로)
- 행위와 그 수동체(교육-학생)
- 특성에 관한 연구(독약-독성)
- 기원에 관한 개념(물-우물)
- 인과관계를 가진 개념(사별-죽음)
- 사물과 반대 행위자(범죄-범죄방지장치)

- 개념과 측정단위(전류-암페어)

이상에서 살펴보았듯이 연관관계 정의는 불분명하며, 다양하게 표현되어서 한가지로 통합하기가 어렵다. 앞서 정의 내린 연관관계의 내용을 종합해보면 연관관계는 비계층관계이면서, 개념적으로 밀접하게 연관될 경우이어야 한다. 그리고 색인작성과 탐색에 이용될 가능한 대체용어를 제시하기 위해서 연관관계는 엄격하게 통제되어야 한다. 그러므로 연관관계의 표준은 이러한 사항을 모두 포괄하면서 연관관계 특성을 명확히 나타냄으로써, 연관관계의 정의뿐만 아니라 유형 및 범위를 선정하는데 기반이 되어야 할 것이다.

1.2 시소러스 연관관계 유형

실제 시소러스에서는 연관관계 유형이 어떻게 설정되었는가 살펴보기 위해 시소러스를 분석하였다. 분석대상 시소러스는 Thesaurofacet(TF), Information Retrieval Thesaurus of Education Terms(IRTE), TEST Thesaurus(TEST), Thesaurus of Common Topics(TCT)이다. 그리고 이론상 시소러스 관계유형을 제시한 포스켓(A. C. Foskett)과 마니즈(J. Maniez)의 시소러스 관계유형을 추가하였으며, 다른 시소러스를 비교하기 위해 국제 표준인 ISO를 검토하였다.

TEST에서는 연관관계 유형을 4가지로 구분하고 있다. 첫째, 상이한 계층의 용어로 의미나 개념이 밀접하게 관련되어 있는 용어로서 “레이저 RT 레이저사진”이다. 둘째, 유사동의어로서 “정밀성 RT 정확성”, 셋째, 관점에 따

라 상호관계에 있는 용어로서 “술 RT 감정 완화”, 넷째, 부분-전체의 관계에 있는 용어로서 “기어 RT 치차(齒車 : 톱니바퀴)”이다.

TEST 연관관계 유형은 다른 시소러스보다 계층관계, 동등관계를 엄격하게 분리해서 부분/전체관계와 명확하지 않은 유사동의어의 경우는 모두 연관관계로 포함시켰다. 다른 시소러스에서 볼 수 있는 연관관계 유형은 세 번째 부분인 관점에 따른 상호관계 유형뿐이다.

TF은 상세한 패싯분류체계를 갖는 대표적인 시소러스로서, 연관관계의 유형은 다음과 같다.

- 사물 / 부분 : 보일러 RT 증기관
- / 성질 : 레이저 RT 밀착
- / 과정 : 배 RT 조선
- / 속성으로서 사물 : 아크(arcs) RT 아크로(arc furnaces)
- / 응용 : 여파기(濾波器) RT 신호처리
- 성질 / 과정 : 전하(電荷) RT 전하 측정
- / 속성으로서의 성질 : 비대칭 RT 비대칭 대들보
- 과정 / 사물 : 착륙 RT 착륙 장치
- / 성질 : 폭발 RT 폭발 파동

TF는 다른 연관관계 유형과는 달리 실세계에서 가질 수 있는 관계유형을 체계적으로 제시하고 있다. 그러나 관계 유형을 크게 사물, 성질, 과정으로 구분함으로써 시간, 공간 등과 같은 관련 유형을 고려할 수 없으며, 관계 유형이 구체적이지 못해 적용하기 어렵다는 단점이 있다.

랭카스터는 정보검색의 용어통제를 위해 IRTET에서 연관관계 유형을 다음

과 같이 제시하고 있다.

- 과정과 과정수행에 필요한 도구나 장치 : 컴퓨터 RT 프로그래밍 교육
- 측정도구와 측정대상 : 시력 RT 눈
- 과정과 제품 : 직조 RT 옷
- 원재료와 제품 : 고령토 RT 도자기
- 사물, 과정, 상태와 그와 관련된 장소 : 외국어 RT 외국어 학원
- 두 개의 상보적 활동 : 가르치는 것 RT 배우는 것
- 사물과 강하게 연관된 성질 : 고무 RT 탄력성
- 사물과 그 부분 : 비행기 RT 동체, 기체
- 특정 계층에서 BT, NT일 경우 그 의미가 제한되는 다중 의미를 가진 용어를 가진 용어와 다른 한 용어의 의미 : 인식 BT 기억 RT 칭찬
- 활동과 관련된 성질 : 굽히는 것 RT 가소성
- 결과와 원인 : 범죄 RT 적의
- 반대 : 생명 RT 죽음
- 활동과 활동의 행위자 : 흡연 RT 흡연자

IRTET의 경우는 부분/전체관계, 반의관계, 상보관계 등을 연관관계에 두었다. 그리고 체계적인 관계유형이 아닌 필요에 의한 직관적 연관관계 유형으로 이루어져 있다.

TCT의 용어 관계는 자연언어 시스템에 출현한 어형변화적 관계나 외부세계를 구성하는 객체(전 우주의 요소) 사이에 나타나는 언어 외적 관계(extralingual relationship)를 반영한 것이다.

어형변화적 관계 사이에서는 동의어와 종속(종 → 속, 속 → 종)관계로 구별할 수 있고, 언어 외적 관계는 연관

관계와 단순존재론적(mereologic) 관계로 되어 있다.

그리고 연관관계는 다음과 같이 특정한 관계의 범위로 되어 있다.

- 원인→결과 : 피해→정체
- 결과→원인 : 정체→피해
- 활동→특징 : 자료검증→통계적 착오
- 활동→활동도구 : 피해예방→경보기구

TCT의 경우, 관계유형에서 외부세계 수준을 자연언어단계와 구분하여 고려했다는 점이 특이하며 그러한 고려로 인해 종/속관계를 부분/전체관계와 엄격하게 구분하고 있다. 그러나 연관관계의 경우 원인-결과 관계, 활동-특징 등의 관계로 단순화해서 그 범위나 표현력이 너무 협소하다는 단점이 있다.

포스켓은 시소러스에 사용된 개념을 영구개념, 의미개념으로 구분하고 이 의미관계를 동등관계, 계층관계, 연관관계로 구분하였다.

연관관계는 다시 동위관계, 발생관계, 동시관계, 원인과 결과, 도구, 재료, 유사성으로 구분하였다.

포스켓이 제시한 연관관계 유형의 경우 동위관계, 발생관계는 계층관계와 더불어 명확하게 정의되어야 하며, 동시관계와 유사성은 좀더 구체적으로 정의되거나 세분될 필요가 있다.

마이니쯔는 3가지 범주로 연관관계의 유형을 제안했다. 3가지 유형은 유사 동의어관계 (near-synonymy relatedness), 의미의 중복관계, 여분의 의미 (extrasemantic)관계이다.

첫번째 관계유형인 유사동의어의 경우 ISO에서는 유사동의어를 동등관계로 본다. 물론 TEST에서 “accuracy RT

precision”과 같이 유사동의어인 디스크립터를 연관관계로 두는 경우도 있으나 대부분의 시소러스에서는 유사동의어를 동등관계로 취급하는 것이 일반적이다.

두 번째 관계유형인 의미가 중복되는 용어관계는 ISO에서 제안한 노드레이블과 유사하다.

세 번째 관계유형인 여분의 의미관계가 실제로 연관관계에 해당된다.

그러나 이 관계유형이 실제적으로 연관관계를 나타내는 것은 여분의 의미관계뿐이기 때문에 이런 방식으로 구분하는 것은 합리적이지 못하다.

1.3 시소러스의 연관관계 범위

앞에서 연관관계 유형을 고찰한 결과, 먼저 이루어져야 할 것은 연관관계 유형보다는 연관관계의 범위 설정이다. 연관관계의 범위는 계층관계(부분/전체관계)나 동등관계(동의어, 유사동의어, 반의어 등)와 관련을 지니기 때문에 이들 관계의 범위중 범위를 확실히 구분하지 않으면, 이로 인해 연관관계의 범위와 유형이 복잡해지게 된다. 연관관계 유형을 [표 1]과 같이 상호 개념관계, 동의어, 유사동의어, 부분/전체, 유사반의어로 구분하여 실제 시소러스 연관관계의 범위를 알아보고, 그 범위를 확실히 정한 후, 연관관계 유형을 고려해야 할 것이다.

[표 1]을 보면 부분/전체관계를 연관관계의 범주에 포함하는 경우(TEST)도 있으며, 확실한 부분/전체관계는 계층관계에 두고 불확실한 부분/전체관계는 연관관계에 두는 경우(FOSKETT, IRTE T)도 있다. 그리고 계층관계에만 두는 경우(ISO, TCT)도 있다.

그러나 ISO의 계층관계는 종/속관계와 부분/전체관계를 포함하므로 부분/전체관계를 계층관계의 범주로 둘 것을 권고하고 있다. 그리고 필요하다면 종/속관계, 부분/전체관계, 사례관계 각각을 BTG(Generic)/NTG, BTP(Partitive)/NTP, BTI(Instance)/NTI로 구별할 수 있어야 하고, 상이한 계층에 속하는 용어나 상이한 수준의 용어를 인정하고 이들 용어를 취급할 수 있는 규정이 필요하다고 지적하고 있다.

그리고 동의어의 경우는 연관관계에 포함되지 않지만 유사동의어와 유사반의어의 경우는 연관관계에 포함되는 경우(MANIEZ, IRTET, TEST)도 있다. [표 2]는 시소러스에서 동의어와 반의어가 동등관계와 연관관계중 어디에

속한지를 나타낸 것이다.

유사동의어 경우 대부분이 동등관계로 보며, 연관관계로 보는 경우는 3개의 시소러스가 있다. 여기서 TEST는 유사동의어의 정도를 구별해서 강한 유사동의어의 경우는 동등관계로 보고, 약한 유사동의어는 연관관계로 두었다. IRTET의 경우 동의어는 동등관계로, 유사동의어는 연관관계로 두었다. FOSKETT의 경우는 ISO와 동일하게 동의어와 유사동의어를 다 동등관계로 두고 있다.

반의어의 경우 FOSKETT은 동등관계로, IRTET는 연관관계로 사용하고 있다.

패러데인의 관계색인에서는 동의어, 유사동의어를 동등관계로 인정하면서

[표 1] 연 관 관 계 분 석

	ISO	IR	MANIEZ	FOSKETT	IRTET	TEST	TCT
상호개념관계	✓		✓	✓		✓	✓
동 의 어							
유사동의어			✓		✓	✓	
부 분 / 전 체				✓	✓	✓	
유사반의어					✓		

[표 2] 동의어와 반의어의 관계유형

	ISO	IR	MANIEZ	FOSKETT	IRTET	TEST	TCT
동 의 어 ER	✓	✓	✓	✓	✓	✓	✓
유사동의어 ER	✓	✓	✓	✓		✓	
동 의 어 AR							
유사동의어 AR			✓		✓	✓	
반 의 어 ER	✓		✓	✓			
반 의 어 AR					✓		

※ 동등관계(ER), 연관관계(AR)

덧붙혀 사물의 다른 측면을 고려한 경우도 동등관계로 보았다. 즉 두 개념이 동일 개념선상에 있으므로 동등관계는 반의어에 대해서도 적용될 수 있다는 것이다. 물론 동일개념선상의 경우와 달리 상대적인 차이점을 인식해야 할 경우는 식별관계로 두어서 두 관계를 구분하였다.

오스틴(D. Austin)은 유사동의어를 두 가지 유형으로 구분하였다. 하나는 동일 개념의 연속선상에서 다른 지점을 표현한 것으로 여기에는 반의어를 포함한다. 다른 하나는 의미상에서 중복되는 용어(예, 고도 : 높이)이다.

이상을 종합하면 동등관계의 경우는 ISO에 따라 동의어와 유사동의어가 포함되며, 반의어는 동일 개념선상에서 다른 지점을 표현하는 유사동의어로 동등관계에 포함된다.

2. 관계 유형의 실험적 연구 분석

앞서 살펴보았듯이 기존의 시소러스에서는 연관관계의 유형이나 범위에 있어서 일관성이 없으며, 직관적으로 이루어져 있어서 각 시소러스 연관관계를 하나로 통합할 수 있는 공통점을 찾을 수 없었다. 본 장에서는 다양한 연관관계 유형을 제시한 관계유형의 실험적 연구를 통해 관계 유형의 설정 방법과 특징, 그리고 연관관계의 유형을 중심으로 분석한다. 연관관계뿐만 아니라 전반적인 관계유형을 고려하는 이유는 앞에서 분석한 결과 연관관계는 다른 관계유형과 상호관련되면서 구성되므로 전반적인 관계유형을 살펴보고, 특히 연관관계에 초점을 맞춰 분

석하고자 한다.

분석대상의 선정은 문헌조사에 근거하여 본 연구목적에 알맞은 기준을 다음과 같이 설정하였다.

첫째, 전통적인 시소러스의 관계유형과는 달리 관계유형이 다양하게 접근되었는지에 따라,

둘째, 연관관계가 직관적으로 이루어지지 않고 정확한 실험과 평가를 통해서 유형화되었는지에 따라,

셋째, 다양한 연관관계 유형뿐만 아니라 다양한 연관관계 기호로 사용되거나 접근점이 다양하게 되어있는지에 따라 분석대상을 선정하였다.

이 기준에 따라 선정된 관계유형의 실험 연구로는 왕, 반덴도르프, 에번스의 관계형 시소러스(Relational Thesauri), 루(X. Lu)의 구조화된 문헌의 자동 검색모델에서 어휘-의미관계 (Lexical-Semantic Relation), 미국 국립의학도서관(National Library of Medicine : NLM)의 통합의학언어시스템(Unified Medical Language System : UMLS)에서 메타시소러스(Metathesaurus), 최석두의 한국문화예술진흥원의 의미구조 시소러스이다.

2.1 왕, 반덴도르프, 에번스의 관계형 시소러스

왕, 반덴도르프, 에번스는 관계형 시소러스 설계를 제안해서, 검색에서 질문을 강화시키는데 사용하였다. 43개의 관계를 5개의 집단으로 클러스터화하여 질의어 형성에 보조로 작용함으로써 검색 효율에 영향을 미치며 궁극적으로 새로운 유형의 용어분류를 위해 구축하였다. 5개의 집단은 부분/전체관계, 연어(collocation)관계, 어형변화적(paradigmatic) 관계, 분류-동의관계, 반의관계이다.

부분/전체관계는 부분-전체, 요소-집합, 사물-재료 등을 연결하는 관계로 되어 있으며, 이 관계는 라파엘(B. Raphael)의 의미정보검색시스템(Semantic Information Retrieval)을 통해서 성공적으로 적용되었다.

언어관계는 같은 구나 문장에 빈도상 동시출현한 단어사이의 관계를 포함한다. 통계적인 연관성 측정은 용어 출현에 의존하는데 이런 이유 때문에 언어관계는 분리해서 처리된다.

어형변화적 관계는 '태양-태양의', '새끼고양이-고양이'와 같이 동일한 의미적 중심을 공유하는 다른 형태와 연관된다. 어형변화적 관계 그룹은 언어 정규화 과정에서 유용하다.

분류-동의관계는 전통적인 관계유형이다. 워너(O. Werner)는 동의어를 분류의 제한된 형태로 다루면서 이 두 관계를 함께 그룹짓는 것을 정당화했다.

관계형 시소러스의 많은 유형은 정보검색시스템의 효율성을 향상시키는데 유용함을 보여준다. 이 연구에서 보여주는 어휘-의미관계의 이론을 기반으로 한 관계 시소러스는 통계적 속성보다 오히려 의미정보가 기본으로 되어 있다는 장점을 지녔다.

왕은 관계형 시소러스를 구성하는 5개의 관계 집단을 가지고, 각 관계의 결합방식에 따라 실험함으로써 검색시스템에 어떤 영향을 주는지를 살펴보았다.

시스템에서 가장 좋은 결과는 반의관계를 제외한 모든 관계를 결합하는 것이다. 그러나 이 경우는 일괄처리 시스템에서 가능하며, 실제적으로 상호작용하는 검색시스템의 경우에는 개별적인 관계가 더 유용하다. 상호작용 시스

템은 관계형 시소러스로부터 색인용어를 제시할 수 있으며, 이용자가 의도하는 의미를 가장 잘 전달할 용어를 선택하도록 허용한다.

부분과 전체관계는 용어를 특정하게 해주고, 분류관계는 체계적인 분류기준을 가지고 용어관계를 나타냄으로써 질의어를 더 일반적으로 해준다.

언어관계는 같은 의미분야에서 다른 주제로 안내해 주는 역할을 한다. 그러나 언어관계와 어형변화관계는 검색성능에 거의 영향을 미치지 않은 것으로 나타났다.

관계형 시소러스의 경우 연관관계를 지칭하는 관계 그룹을 따로 두지는 않았지만, 새로운 5개의 관계 그룹을 형성해서 43개의 관계유형을 제시함으로써, 실질적으로 다양한 연관관계 유형을 제시해 주고 있다. 특히 언어관계는 두 용어가 관련되어 있는 경우이므로, 연관관계 유형(예를 들어 결과, 도구, 요소, 위치 등)을 어느 정도 포함하고 있다.

부분/전체관계의 물질(substance)관계와 기원(provenance)관계는 의미상 연관관계에 가깝기 때문에 부분/전체관계에 포함되는 것은 좋지 않다.

반의관계는 관계형 시소러스의 관계유형을 가지고 시스템 내에서 자동으로 용어제안을 할 경우 부정적인 결과를 낳는다. 그러나 이용자가 질의어 형성시 관계유형을 제시한다면 오히려 개념의 이해나 설정에 도움이 되며, 검색에서 더 좋은 결과를 낼 수 있다. 그리고 반의관계는 동등관계와 함께 그룹화하는 것이 바람직하다.

관계형 시소러스의 관계유형은 구문지향관계(예, 완료-부정사)를 의미적인 관계(예, 동의어)와 실용적인 관계(예,

[표 3] 어휘-의미관계

1. 동등관계		
인지(cognitive) 동등	c-s	바이올린(비공식적) : 바이올린(공식적)
유사(near)동등	n-s	인용 : 참고
어형(morphology) 변형	m-v	남자 : 남자들
2. 분류관계		
분류	t-x	말 : 성장한 숫말
동시-분류	t-x	성장한 암양 : 성장한 숫양
3. 부분/전체관계		
분의(分意 : meronymy ¹⁾)	p-w	팔 : 손
동시 분의(co-meronymy)	p-w	손바닥 : 손가락
그룹-일원	p-w	의회 : 국회의원
계급-일원	p-w	프롤레타리아 : 노동자
집합-일원	p-w	산림 : 나무
개체와 그 특징	p-w	옷 : 사이즈
재료	p-w	타이어 : 고무
출처	p-w	우유 : 소
조각	p-w	각설탕 : 설탕
4. 비분기(non-branching) 계층관계		
체인형	n-b	어깨 : 상박 : 팔꿈치 : 하박
나선형(helices)	n-b	봄 : 여름 : 가을 : 겨울
정렬-용어	n-b	1배 : 2배 : 3배
등급-용어	n-b	작은 : 큰 : 거대한
단계-용어	n-b	낙제 : 합격 : 이수증명 : 우등
5. 기타 의미관계		
개체 : 연관된 과정	e-pc	도서관 : 대출
개체 : 연관된 개체	e-e	차 : 차고
개체 : 연관된 사람	e-ps	병원 : 환자
과정 : 연관된 개체	pc-e	자동화 : 컴퓨터
과정 : 연관된 과정	pc-pc	색인 : 의사결정
과정 : 연관된 사람	pc-ps	색인 : 색인작성자
사람 : 연관된 과정	ps-pc	이용자 : 탐색
사람 : 연관된 개체	ps-e	사서 : 책

1) 분의는 “part”를 의미하는 라틴어 ‘meros’에서 따서 만든 말로서 흔히 말하는 부분-전체 관계를 나타내는 것이며 meronym ↔ holonym의 관계가 된다.

전체/부분)로 혼합하여 얽혀 있는 계층을 만들었다. 이는 이용자가 질문 강화를 위해 사용했을 때 성능이 향상되었지만, 기존의 관계유형의 틀을 벗어난 복잡하고 비체계적인 관계유형 때문에 실제 이용에서는 혼란을 가져올 수 있다는 단점이 있다.

2.2 루의 구조화된 문헌의 자동검색 모델에서 어휘-의미관계

루(Lu)는 색인용어사이에 어휘-의미관계(Lexical-Semantic Relation)를 기반으로 설계된 구조화된 문헌의 자동검색 모델을 제시했다. 이 연구에서 사용된 모든 어휘-의미관계를 5개의 범주로 그룹화하여 동등관계, 분류관계, 부분/전체관계, 비분기(non-branching)계층관계, 기타 의미관계로 구분하였다. 관계유형을 너무 많이 사용하여 이용이 어렵고, 어휘의 의미 차별성을 흐리게 할 수 있기 때문에 왕이나 폭스가 제안한 것보다는 그 범위를 제한하였다. 왕이나 폭스가 제안한 연어관계와 어형변화관계는 검색성능에 거의 영향을 미치지 않으므로 제외시켰다. 제안된 모델은 색인과정에서 자동색인시스템의 수행을 가능하게 함으로써 주관성의 개입을 피할 수 있다. 특히 기계가독형 사전의 정의부분에서 어휘-의미관계를 자동으로 추출할 경우 유용하다.

분류관계, 부분/전체관계, 비분기계층관계의 어휘-의미관계에서 중심적인 어휘 환경은 계층이다. 이 계층을 두 가지 구조 형태로 볼 수 있는데, 가지를 치는지의 유무에 따라 “체인구조”와 “나무구조”로 나뉘어진다.

루는 이 연구 모형의 개발을 통해서 순수 의미관계(예, t-x, p-w)에서 순수

어휘관계(예, e-pc, ps-pc)를 구별해서 나무구조나 네트워크와 같은 상이한 구조로 두 개의 관계 그룹을 표현하는 것이 훨씬 더 의미있는 일이라 제안했다.

이 관계 유형은 크게 시소러스 관계 유형인 동등, 연관, 계층관계로 되어있다. 계층관계의 경우 분류, 부분/전체, 비분기계층관계로 세분화하였다. 그러나 계층관계의 경우 관계기호가 분류관계(t-x), 부분-전체관계(p-w), 비분기계층관계(n-b)로 크게 구분되어서 동등관계의 인지 동등(c-s), 유사 동등(n-s), 어형 변형(m-v) 등과 같은 수준으로 다루어지기 때문에 실제로 적용하는데 있어 그리 세분화된 것은 아니다. 그리고 연관관계 유형이라 할 수 있는 기타 의미관계는 개체, 과정, 사람으로 구분, 세 부분의 상호조합으로 관계유형을 형성한다. 9가지 관계유형은 필수적이지만, 다른 가능한 관계유형이 제외되어서 폭넓은 연관관계를 형성할 수 없다는 제한이 따른다. 구조화된 문헌의 자동검색모델을 자동구축하기 위해서 추출하기 용이한 관계로 관계유형을 형성했다. 그래서 세부적인 관계유형을 설정하는 것이 어렵지만, 연관관계 유형을 RT로만 정의하던 기존의 정의와는 달리 e-pc, e-e, e-ps 등의 다양한 관계기호를 사용함으로써 더 정확하게 용어에 접근할 수 있다는 장점을 지니고 있다.

2.3 미국 국립의학도서관 통합의학 언어시스템의 메타시소러스

라다(R. Rade)는 메타시소러스를 시소러스보다 우수하다고 하였고, 루린(C. Roulin)은 계층적으로 구조화된 개념과 색인어간에 설정된 관계를 통해

서 부합된 다양한 색인어와 참조 시소러스의 집합이라 정의하였다. 이는 메타시소러스와 통합어휘집이 전통적인 시소러스와는 달리 한가지 형태로 존재하는 것을 의미한다

1986년, 린드버그(D. Lindberg)은 NLM에 의해 지원된 UMLS를 보고하였다. 이 시스템은 정보 요소 맵(Information Sources Map)과 메타시소러스, 의미네트워크로 구성되어 있다.

메타시소러스는 기존 어휘와 분류시스템을 사상화할 규범적인(canonical) 형태에서 의학개념과 용어를 축적한 것이다. 그리고 개념정의를 포함하며 MeSH의 계층관계와 화학부록으로부터 약 16,000의 의학 주제명표목을 포함한 208,559의 용어와 다른 생의학에서 나온 용어를 포함하고 있다.

메타시소러스의 각 개념은 네트워크로부터 하나 이상의 의미유형으로 부여되며 특정개념에 대한 모든 정보를 메타시소러스에서 볼 수 있다. 네트워크는 기본적인 의미 유형이나 범주화에 대한 정보를 제공한다. 의미네트워크의 의미유형과 의미관계는 개발된 이래로 계속해서 증가하고 있으며 1997년 현재 135개의 의미유형과 51개의 의미관계를 포함하고 있다.

의미관계는 네트워크내에 크게 계층관계와 비계층관계로 나뉘어 있는데, 계층관계는 'isa'링크이며, 비계층관계는 '물리적(physically)', '시간적(temporally)', '기능적(functionally)', '개념적(conceptually)', '공간적(spatially)' 5가지 주요 범주로 그룹화되어 있다. 51개의 관계는 다음과 같이 계층구조로 제시되어 있다.

H. isa (계층관계)

R. associated_with (연관관계)

1. physically_related_to 물리적 관련성

1. part_of 일부
2. consists_of 구성
3. contains 포함
4. connected_to 연결
5. interconnects 상호연결

2. spatially_related_to 공간적 관련성

1. location_of 위치
2. adjacent_to 근접
3. surrounds 주변
4. traverses 교차

3. functionally_related_to 기능적 관련성

1. affects 영향
 1. manages 관리
 2. treats 취급
 3. disrupts 분열
 4. complicates 심화
 5. interacts_with 상호작용
 6. prevents 예방
2. bring_about 성취
 1. produces 생산
 2. causes 원인
3. performs 수행
 1. carries_out 달성
 2. exhibits 제시
 3. practices 습관
4. occurs_in 발생
 1. process_of 과정
5. uses 사용
6. manifestation_of 징후
7. indicates 지시
8. result_of 결과

4. temporally_related_to 시간적 관련성

1. co_occurs_with 동시 발생
2. precedes 선행

5. conceptually_related_to 개념적 관련성

1. evaluation_of 평가
2. degree_of 정도
3. analyzes 분석
 1. assesses_effect_of 결과 평가
4. measurement_of 측정
5. measures 측정단위
6. diagnoses 진료
7. property_of 성질
8. derivative_of 파생
9. developmental_form_of 개발 형태

- 10. method_of 방법
- 11. conceptual_part_of 개념의 일부
- 12. issue_in 문제

이 관계유형은 크게 계층관계와 비계층관계로 구분되며 비계층관계 즉, 연관관계는 5개의 범주로 구분하여 각 범주와 그 관계명에 대한 정의와 도치 관계 등을 명시적으로 나타내고 있다. 1997년 이전에는 '공간 관련성'을 제외

으로 링크된다. 이 관계명은 방향성(양방향관계)을 가지며, 관계정의의 틀(layer)이 된다. 즉 하나의 관계류가 하나의 틀을 형성하게 된다. 개념관계(동의, 계층관계 등)나 논리관계(인과관계 등)는 명시적으로 구분하지 않고 관계명에 의해 암시적으로 구분하고 있다.

여기서 제안하는 의미구조 시소러스에서는 연관관계를 다음과 같이 세분하였다.

<ul style="list-style-type: none"> • 학문분야, 연구영역 및 대상/현상 관계명 : 분야 	<ul style="list-style-type: none"> • 특성에 관한 개념 관계명 : 특성
<ul style="list-style-type: none"> • 조작, 과정/행위자, 도구 관계명 : 도구 	<ul style="list-style-type: none"> • 기원에 관한 개념 관계명 : 기원
<ul style="list-style-type: none"> • 직업과 그 직업에 종사하는 사람 관계명 : 직업 	<ul style="list-style-type: none"> • 사물, 행위/행위자 관계명 : 대항
<ul style="list-style-type: none"> • 행위/결과 관계명 : 인과 	<ul style="list-style-type: none"> • 원재료 - 제작된 제품 관계명 : 재료
<ul style="list-style-type: none"> • 결과와 원인 관계명 : 원인 	<ul style="list-style-type: none"> • 행위 - 이에 연관된 특성 관계명 : 행위성
<ul style="list-style-type: none"> • 행위와 수동체 관계명 : 대상 	

한 4개의 범주로 구성되어 있었으나 최근에 5개의 범주로 구분하여 실세계 지식을 전체적으로 다룰 수 있는 체계를 갖추고 있다.

2.4 한국문화예술진흥원의 의미구조 시소러스

최석두는 새로운 방법으로 각 세부 주제에 대한 용어사전을 선정하고, 그 용어사전에 등록되어 있는 용어(고유명사 포함)와 그 용어에 대한 설명문을 하나의 그룹으로 관계구조를 구축하였다.

개념간의 관계는 관계명(개념 관계 레이블 혹은 관계레이블)으로 관계가 설정되며, 두 가지 개념은 서로 관계명

이 연구에서 전통적인 시소러스의 문제점을 보완하기 위하여 의미구조를 갖는 시소러스를 작성하였다. 그리고 의미구조 시소러스의 경우 관계명의 정의가 일정한 기준을 가질때 보다 정확하고 융통성있는 시소러스가 생성될 수 있으며 고정적인 통제어휘용 시소러스를 먼저 구축하고서 이를 자료의 색인과 검색에 적용할 것이 아니라 자료 자체를 색인함으로써 시소러스를 동적으로 생성, 적용 및 확장하여 가는 방법이 필요하다고 제안하였다.

의미구조 시소러스에서의 관계유형은 ISO 표준의 연관관계 부분중 상이한 범주에 속하는 유형을 기반으로 하

였다. ISO 표준에서의 관계 종류를 기반으로 유형을 결정한 뒤, 적절한 관계명을 두어서 기존의 시소러스에서 RT로만 설정된 용어관계를 분야, 원인, 인과, 특성 등으로 다양하게 표현하였다. 그리고 의미시소러스와 일반시소러스의 대응관계를 만들어서, 검색의 특정성이나 필요에 따라 일반시소러스를 생성하거나 일반시소러스의 기능을 적용할 수 있게 하였다. 특히 연관관계의 경우 세부적으로 나타내느냐, 일반적으로 나타내느냐에 따라 관계명의 표현 정도가 다르므로 색인의 특정성을 달리할 수 있다.

IV. 시소러스의 연관관계 유형 제안

시소러스에서 기존의 연관관계 유형을 대신할 다양하고 의미 있는 연관관계 유형을 선정하기 위해서 기존의 관계유형보다 더 세분화시킬 필요가 있는데, 그 정도가 누구나 공감할 수 있는 정도여야 하고, 논리적이어야 한다. 연관관계 유형 설정시 주의해야 할 점으로는 먼저 너무 많은 관계유형을 사용하면 정확하게 사용하는 것이 어렵고 의미의 차별성이 흐려진다. 그리고 일반적으로 받아들일 수 있는 관계유형이 되려면 중복성, 충돌, 특수성, 모호성, 유사성 등을 피해야 할 것이다.

1. 고려사항

효율적인 연관관계 유형을 제안하기 위해서 다음사항을 고려하였다. 단 연관관계는 시소러스에서 단독으로 구성되는게 아니고 다른 관계유형과 상호

관련되면서 구성되므로 전반적인 관계 유형의 관점에 따라 고려사항을 제시하였다.

첫째, 시소러스 구축방법 즉 수작업이나 자동이냐에 따라서 관계유형의 설정방법과 적용방식이 조금씩 다르게 나타나므로 구축방법을 결정해야 한다.

시소러스를 수작업으로 구축할 경우 의미관계를 자세하게 설정할 수 있다. 그러나 주관적 판단에 따른 일관성 결여로 용어선정과 관계설정이 문제된다. 또한 시간-소모적이며 갱신이 어렵다.

자동으로 시소러스를 구축할 경우 통계적 접근이나 언어적 접근을 사용한다. 의미관계를 잘 표현할 수 없지만 갱신과 관리가 편리하며 일관성을 유지할 수 있다.

둘째, 시소러스 작성과정에 따라서 관계유형의 설정이나 적용방식이 달라질 수 있다. 즉 작성과정중 주제영역의 정의, 시소러스 형태의 결정 등이 시소러스 관계유형에 영향을 미칠 수 있다.

셋째, 시소러스 기능에 따라서도 관계유형이 달라질 수 있다. 시소러스는 용어 브라우징, 질의어 형성, 정보검색의 보조 등의 기능을 수행하는데 그 각각의 기능으로 관계유형이 달라지기도 한다.

넷째, 시소러스 표시형식에 따라, 특히 시소러스 브라우저를 제시해서 이용자가 개념을 형성해 나갈 경우는 다른 관계뿐만 아니라 반의관계를 설정하는 것이 바람직하다. 최근에 하이퍼 텍스트 기반 시소러스나 온라인 시소러스의 경우 반의관계를 제시하는 것이 일반적이다. 그러나 시소러스를 이용자에게 제시하지 않는 시스템의 경우 왕의 관계형 시소러스 실험에서와 같이 반의관계는 부정적인 결과를 낳

으므로 반의관계를 제외시키는 것이 더 효과적일 수 있다.

다섯째, 기본관계를 유지하고 그 이외의 디스크립터에 대한 정보를 다양하게 함으로써 관계유형 이외에도 용어사이의 관계를 다양하게 표현할 수 있다. 기본관계 이외의 디스크립터에 대한 정보로는 분류번호 제공(Women's Thesaurus의 경우 주제그룹으로 분류), 패킷 지시(Unesco와 BSI의 경우), 각 관계의 또 다른 하위개념어 제시(EJC의 경우)나 관계 수 표시(Thesaurus of Information Technology Term의 경우) 등이 있다.

여섯째, 관계유형을 크게 범주화된 관계유형과 임의적인 관계유형으로 구분함으로써 관계유형의 범위가 달라진다. 범주화된 관계유형은 용어의 출현 빈도와 어휘-의미의 관계유형 구조 등에 따라 여러 관계유형을 설정하지만 한정적이다. UMLS 메타시소러스의 경우 크게 5개 범주로 구분하고 있는데 각각의 관계에 더 세부적인 관계를 설정하고 필요에 따라 관계생성이 용이하도록 되어 있다.

그리고 임의적인 관계유형의 경우는 용어와 용어사이에 있을 수 있는 모든 관계를 인정하는 것이다. 이 경우는 의미를 그대로 가질 수 있는 장점이 있는 반면, 관계유형이 방대해지고 정보 검색시 찾기 힘들다는 단점이 있다. 이 관계유형을 가지는 것으로 CYC가 있는데, CYC는 다양하고 방대한 양의 지식에 대한 표현과 이를 이용한 추론기법을 제시하였다. 이 경우 관계형을 찾기 힘들므로 관계형 기본형태(primitive)를 두어서 관련있는 관계를 묶어주는 시도가 필요하다.

2. 제안

이전의 관계유형을 분석한 결과, 관계유형은 지식분류에 근거하여 모든 개념을 관계로 다 표시할 수 있도록 관계설정이 되어야 한다. NLM의 UMLS에서의 메타시소러스는 연관관계를 5개로 크게 범주화하여 지식의 범위를 전체적으로 다루고 있어서 가능한 모든 개념을 관계로 표시할 수 있게 되어있다. 그리고 5개의 범주화는 콜론분류법의 5개 범주와 유사해서, 메타시소러스에 콜론분류법을 결합시켜 새로운 연관관계 유형을 설정할 수 있으므로, 이것을 고려하여 새로운 연관관계 유형을 제안하고자 한다.

2.1 콜론분류법의 기본범주

문헌을 분류하는데 있어서 대상이 되는 지식 분야는 대단히 다양하다. 곧 특정한 주제의 수는 무한정으로 많고 다차원적이다. 이 수없이 많은 주제는 각각 여러가지 속성을 가지고 있으며 그 속성 중의 일부로 주제계(主題界)의 한 패킷에 항상 포함되며, 랭가나단은 이 패킷에 대하여 "5개의 기본범주만이 존재한다"라고 하면서 지식분류에 근거하여 모든 개념을 표현할 수 있도록 하였다. 그리고 분석·합성의 기법을 문헌분류법에 도입한 최초의 분석합성식분류법인 패킷분류법을 고안하였다. 패킷분류법은 주제와 주제를 연결 또는 결합하는 기법을 도입한 분류체계이다.

랑가나단은 콜론분류를 통하여 지식의 5가지 기본범주라는 개념을 발전시켰다. 기본범주는 개념이 지식영역에서 반복되는 규칙성으로서, 이 지식체계는

[표 3] 콜론분류와 메타시소러스를 이용한 연관관계유형

<p>[P] 개념적 관련성 (P) 일부 추상 개념, 사상, 사고와 관련</p> <p>속성 (P1) 특성, 질 : 인식 - 판단</p> <p>기원 (P2) 요소사이에 구조적으로 관련되거나 제품을 만드는 것으로 기능적 관계를 포함하지 않으며 기원이 동일 : 철학 - 생각</p> <p>발전 형태 (P3) 성숙의 초기 단계 : 오성 - 감성</p> <p>발전 방식 (P4) 행위, 절차에 따른 사건 방식과 순차 : 오성 - 이성</p> <p>[M] 물리적 관련성 (M) 일부 물리적 속성이나 성격의 가치와 관련</p> <p>[MP] 물질-속성</p> <p>속성 (MP1) 물질적 특성, 질 : 고무 - 탄력성</p> <p>연결 (MP2) 물리적 단일체에 직접 연결 : 조류학 - 새</p> <p>상호연결 (MP3) 둘 이상의 물리적 단일체를 함께 연결 : 물감 - 붓</p> <p>[MM] 물질-방법</p> <p>방법 (MM1) 물질의 다양한 방법 : 의약품 - 치료</p> <p>[MMt] 물질-재료</p> <p>구성 (MMt1) 전체 혹은 사물의 일부 구성 : 석탄 - 탄소</p> <p>도구 (MMt2) 활동 수행시 이용, 즉 적용, 활용 : 속도측량 - 속도계</p> <p>[E] 기능적 관련성 (E) 기능이나 활동을 수행하는 것과 관련</p> <p>생산 (E1) 발생, 창조 : 점토성형 - 도자기</p> <p>과정 (E2) 활동, 기능, 상태 : 정제 - 원유</p> <p>원인 (E3) 상황, 결과를 야기 : 질병 - 유기체</p> <p>결과 (E4) 활동, 과정의 결과, 효과, 결론적 상황 : 유기체 - 질병</p> <p>성과 (E5) 실체에 영향을 미치거나 작용 : 제본 - 책</p> <p>영향 (E6) 상황, 상태, 실체에 영향을 미치는 것을 암시 : 공작 - 공구</p> <p>상호작용 (E6.1) 모두 함께 행동, 기능, 작동 : 자동화 - 컴퓨터</p> <p>대항 (E6.2) 활동, 상황을 멈추게 하거나 제거 : 유해생물 - 살충제</p> <p>평가 (E7) 특성이나 과정의 값, 정도를 판단 : 문헌전달서비스 - 비용</p> <p>측정 (E7.1) 치수, 양, 용량 등을 숫자로 표시하는 것 : 눈 - 시력</p> <p>수행 (E8) 실체를 집행하거나 달성 : 배 - 조선</p> <p>발생 (E9) 주어진 상황, 환경, 기간에 발생하는 것 : 원유 - 휘발유</p> <p>[S] 공간적 관련성 (S) 장소나 지역과 관련</p> <p>위치 (S1) 실체의 위치, 장소, 지역, 과정의 위치 : 알프스 - 스위스</p> <p>근접 (S2) 물리적 단위에 가깝거나 결부되거나 밀접 : 파리 - 베르사유</p> <p>[T] 시간적 관련성 (T) 선행, 동시출현, 다음에 오는 것과 관련</p> <p>동시발생 (T1) 동시에 발생 : 후천적 이상 - 교통</p> <p>선행 (T2) 제 시간보다 더 일찍 발생 : 선행자 - 계승자</p>
--

다시 긴밀히 결합된 조각인 주류(main class)로 전달된다. 기본범주는 지식의 세계, 콜론분류법에서 사상화된 지식의 세계, 분류상에서 문헌에 다루어진 전 범위의 주제지식 이 세가지를 동시에 만족시켜야 한다. 랭가나단의 가설에 의하면 핵심주제[P], 물질[M], 에너지[E], 공간[S], 시간[T]이라는 기본범주가 존재한다.

2.2 콜론분류와 메타시소러스를 이용한 연관관계 유형 제안

NLM의 UMLS에서 메타시소러스의 경우 크게 5개로 범주화한 것에 콜론분류법의 5개 범주를 고려할 수 있다. 메타시소러스의 세부적인 관계중 의학과 관련된 특정 관계를 제외시키고 다른 시소러스의 연관관계유형이 이 범위에 없을 경우 5개 범주의 적합한 관계계층에 삽입한다. 관계유형 각각에 개별적인 관계지시기호를 사용할 수 있으며 관계유형이 계층적으로 단계화되어 있어서 용어의 특정성 수준을 조절해서 사용하는 것이 가능하다. 무엇보다도 연관관계유형을 구축하기 이전에 관계유형을 크게 계층과 비계층으로 나눈 뒤 계층에 계층관계와 동등관계를 두며, 비계층관계로 연관관계를 둔다. 제안하고자 하는 연관관계유형은 [표 3]과 같으며, 괄호 안은 관계지시 기호이다.

3. 다양한 환경에 따른 연관관계 유형 설정

위의 관계유형 이외에 다른 환경에서 시소러스 구축 방법을 다양하게 할 경우 관계유형 설정방법을 달리할 수

있다.

먼저 분류시스템과 결합해서 분류시소러스나 시소로파셋으로 시소러스를 작성하는 것이다. 상호개념적 시소러스의 이용 결과가 관계형을 사용할 경우 일관성이 부족하기 때문에 이 경우 분류시스템과 결합한 시소러스가 유용하다. 이러한 시소러스는 분류시스템을 기반으로 하고 분류구조내에서 시소러스의 여러 기능을 보장함으로써 특별한 연관관계유형이 없어도 효율적인 정보검색도구로서 기능할 수 있다. 분류시스템과 결합한 시소러스는 특히 논리적 기반과 분류구조를 향상시킬 필요가 있다. Art & Architecture Thesaurus(AAT)의 경우 MeSH 시소러스 구조를 사용해서 7개 패킷아래 41개 계층으로 구조화하였다. 계층이 행 번호로 정렬되어 있어서 계층관계를 제시해주고, 분류기호를 절단함으로써 용어의 특정성을 제시할 수 있다. 용어의 계층구조를 통해서 계층관계와 동등관계를 제시해주고 동등관계는 별도로 제시해주고 있다. AAT 용어는 기본적으로 단일 개념을 표현하는 디스크립터로 되어있고, 다양한 구절을 표현할 수 있는 응용 프로토콜이 있어서 복잡한 문자열을 구문적으로 합성할 수 있다.

시소러스 자동구축과 관련해서는 용어의 출현빈도를 이용한 통계적 기법이나 언어적 접근을 통해 일관성있게 구축하고 유지, 갱신할 수 있는 방법이 있다. 그러나 이러한 방법은 개념의 의미를 고려하지 않기 때문에 단독으로 사용하는 것은 그리 좋은 방법이 되지 못한다. 일리노이 대학 디지털 도서관 계획안(Digital Library Initiative)의 용어 제안과 같이 주제 시소러스로서 계층

관계, 동등관계는 기존 방식 그대로 유지하고 연관관계만이 용어의 동시출현 빈도로 구축하는 경우가 있다. 이 용어 제안은 색인, 검색시 주제 시소러스와 용어의 출현빈도를 결합해서 사용된다.

그리고 두 용어, 자극어에 따른 출현 빈도가 높은 용어를 반응어로 두고 반응어의 겹치는 정도를 유사도 측정을 통해서 연관관계를 설정하는 용어관계 시소러스(Term Association- Thesaurus)와 같은 경우도 있다. 그러나 이 경우는 유사도 측정을 위한 계산식을 잘 설정해야 하며, 계산된 수치정도에 따라서 연관관계유형을 두어야 하는 어려움이 있다. 그러나 앞서 언급했듯이 용어의 출현빈도에 따른 시소러스 구축은 자동구축이 가능하다는 장점 이외에 관리, 갱신하는데 유용하며, 일관성을 유지할 수 있다.

V. 결 론

정보의 팽창과 다양화로 인해서 정보를 조직하는 과정과 필요한 정보를 탐색하는 과정은 다같이 중요하다. 정보검색 과정에서 시소러스라는 통제어휘집을 통하여 이용자와 색인자간에 일치된 용어를 사용할 수 있게 되었다. 그러나 정보의 증가에 따라 각 정보간의 관계유형이 복잡해지고 정보의 명확한 의미는 그것이 갖고 있는 관계와 정보의 속성에 따라 다양한 양상을 띄어서 이러한 통제어휘집으로 정보의 관계유형과 정보의 속성 등을 나타내는게 힘들게 되었다. 일반적으로 시소러스의 관계유형은 검색성능에 직접적인 영향을 미침으로써 중요한 역할을 하지만, 시소러스에서 사용하는 용어관

계는 동등, 연관, 계층관계로 단순화되어 있다. 그리고 그러한 용어관계의 유형이나 범위가 분명하게 정의되어 있지 않으며, 논리적인 관점을 기반으로 했다기보다 직관에 의해 연관된 용어를 나열해서 시소러스마다 달리 표현되어 있다. 특히 연관관계의 경우, 관계유형의 유용성에 대한 평가과정 없이 설정되었고, 그 규칙 또한 관계의 범위를 설정하는데 도움이 되지 못하고 있다.

시소러스 관계유형에 대한 이러한 문제점과 필요성에도 불구하고 이제까지는 시소러스 관계유형, 특히 연관관계에 대한 연구가 거의 없었다. 본 고에서는 다양한 관계유형을 가진 시소러스를 분석해서 적절한 연관관계유형을 제안하였다. 앞으로 시소러스의 연관관계에 대한 다양한 분석이 이루어져야 하며 다양한 관계유형을 갖춘 시소러스를 구축해서 보다 정확하고 다양한 탐색결과를 얻을 수 있는 정보검색도구의 역할을 해야 할 것이다.

그러나 이에 앞서 우리말에 대한 이해와 분석을 통해서 용어를 효율적으로 사용할 수 있어야 하고, 문자인식, 국어 형태소 분석 등 현재 개발된 기술을 이용하여 중요 용어를 기계적으로 추출하고 다양한 관계설정이 자동으로 이루어지는 우리말 시소러스가 구축되어야 할 것이다.

참 고 문 헌

- 최석두, 시소러스 개발을 위한 프레임 워크, 서울 : 한국문화예술진흥원, 1994.
- 최석두, 정동열, 시소러스 개발지침, 서울 : 문헌정보처리연구회, 문헌정보처리연구회 시리즈 3, 1994.
- 한국문화예술진흥원, 의미 구조 관리 시스템 : 최종보고서, 서울 : 한국문화예술진흥원, 1994.
- Anderson, J. D., "Building End-User Thesauri from Full-Text," In Proceedings of the 3rd ASIS SIG/CR Classification Research Workshop, Vol.III, Pittsburgh : Learned Information, Inc., 1992, pp.1-13.
- Farradane, J. E. L., "Relational Indexing Part I," Journal of Information Science, Vol.1, 1980, pp.267-276.
- Foskett, A. C., The Subject Approach to Information, 4th ed., London : Clive Bingley, 1982.
- Lancaster, F. W., Vocabulary Control for Information Retrieval, 2nd ed., Arlington : Information Resources Press, 1986.
- Lu, X., "Document Retrieval : A Structural Approach," Information Processing & Management, Vol.26 No.2, 1990, pp.209-218.
- Maniez, J., "Relationships in Thesauri : Some Critical Remarks," International Classification, Vol.15 No. 3, 1988, pp.133-138.
- Milstead, J. L., "Invisible Thesauri : The Year 2000," Online & CD-ROM Review, Vol.19 No.2, 1995, pp. 93-94.
- Myaeng, S. H., McHale, M. L., "Toward a Relation Hierarchy for Information Retrieval," In Proceedings of the 2nd ASIS SIG/CR Classification Research Workshop, Vol. II, Washington, DC : Learned Information, Inc., 1991, pp.101-113.
- Schmitz-Esser, W., "New Approaches in Thesaurus Application," International Classification, Vol.18 No. 3, 1991, pp.143-147.
- Unified Medical Language System (UMLS), http://www.gsf.de/MEDWIS/pg_term/umls_net.html
- U. S. National Library of Medicine, <http://www.nlm.nih.gov/research/umls/META3.HTML>
- Wang, Y. C., Vandendorpe, J., Evens, M., "Relational Thesauri in Information Retrieval," JASIS, Vol.36 No.1, 1985, pp.15-27.
- Willetts, M., "An Investigation of the Nature of the Relation between Terms in Thesauri," Journal of Documentation, Vol.31 No.3, 1975, pp.158-184.