

색상에 기반한 감성시소러스 구축*

The Construction of Sensibility Thesaurus Based on Color

남 영 준**

Young-Joon Nam

차 례

- | | |
|-------------|---------------|
| 1. 서 론 | 4. 감성시소러스의 구축 |
| 2. 시소러스의 특성 | 5. 결 론 |
| 3. 감성시소러스 | · 참고문헌 |

초 록

본 연구의 목적은 멀티미디어 시대에 필요한 새로운 검색도구를 조사하는 것이다. 시소러스가 유용한 검색도구이나 디스크립터가 명사로 이루어져 인간의 감성을 검색하기 위해서는 형용사의 사용이 필요하다. 따라서 본 연구에서는 IR사의 형용사 이미지 스케일을 사용하여 감성적 의미를 갖고 있는 색상을 이용하여 감성 시소러스를 실험적으로 구축하였다. 구축의 규모는 261개로 이루어졌으며, 관계설정은 스케일의 거리와 주요 이미지 스케일의 중첩도를 활용하였다. 구축에는 후보 디스크립터를 보관하고 구조화할 수 있는 용어풀 기능이 있는 감성 형용사 전용 구축 프로그램을 사용하였다.

키 워 드

시소러스, 감성시소러스, 색상시소러스, 형용시소러스

* 본 연구는 한국과학재단 목적기초연구(R01-2003-11588-0) 지원으로 수행되었음.

** 중앙대학교 문헌정보학과 부교수

(Associative Professor, Dept. of Library and Information Science, Chung-Ang University, namyj@cau.ac.kr)

· 논문접수일자 : 2003년 11월 20일

· 게재확정일자 : 2003년 12월 12일

ABSTRACT

The aim of this article is to study the new searching tool for era of multimedia. Thesaurus is a useful tool but is constituted with noun phrase including the adjective for retrieving the human Sensibility. Therefore, I experimentally construct the Sensibility thesaurus using the color scale which contains the Sensibility meanings. Terms are 261, Relation standards are distance and ratio of reiteration between the terms. I would use an exclusive program of the thesaurus construction for Sensibility adjective.

KEYWORDS

Thesaurus, Sensibility Thesaurus, Color Thesaurus, Adjective Thesaurus

1. 서 론

주제명 표목표나 혹은 시소러스와 같은 정보검색 사전에 등재된 디스크립터는 대부분 명사나 명사구로 구성되어 있다. 이는 전통적으로 제한된 장서를 구조화하기 위해 도서관에서 활용하는 분류표 표목의 형태로도 편목과 같은 색인작업을 충분히 수행할 수 있었기 때문이었다. 한편 인터넷에 존재하는 데이터의 형태는 책자형태의 자료를 포함하여 사진 및 동영상 파일까지 과거에 비해 대단히 다양해지고 있다. 또한 정보의 양은 수작업으로 통제할 수 있는 범위를 벗어나는 수준에 이르게 되었다. 이러한 외부적 환경의 변화에 따라 도서관에서도 수작업에 의존하는 검색에서 정보통신기술과 컴퓨터의 적극적인 활용을 모색하기에 이르렀다.

이용자도 직접 방문과 같은 전통적인 정보탐색 방법에서 웹과 같은 새로운 커뮤니케이션 채널을 이용하여 필요한 정보를 검

색하는 방법에 더욱 익숙하게 변화하였다.

현재 웹에서 이루어지는 정보검색은 비문자 형태인 동영상 자료부터 문자로 구성된 자료에 이르기까지 자료 종류나 형태에 관계없이 키워드 검색이 이루어지고 있다. 즉, 검색대상이 되는 새로운 형태의 자료적 특성을 고려하지 않고 모든 정보검색 행위는 키워드 위주의 전통적 검색방법을 사용하고 있다.

대표적 검색도구인 시소러스도 다른 검색도구에 비해 효과적인 검색도구이지만 대부분 키워드 검색을 예상하고 개발한 것이다. 따라서 국내 시소러스의 디스크립터는 대부분 명사(구)이며, 형용사와 같은 용언들을 표목이나 디스크립터로 채택하지 않고 있다.

한편 소리정보를 비롯하여 이미지, 색상, 영상정보와 같이 비문자형태로 이루어진 비정형 멀티미디어 정보를 검색하기 위해서는 인간의 감성을 수용할 수 있는 새로운 검색도구가 필요하다. 형용사는 수식어

로써 뒤에 나타나는 명사에 의존하여 명사가 갖고 있는 개념을 보다 명확하게 제한하는 특성을 갖고 있다. 또한 물리적인 실체 이외에도 보이지 않는 인간의 느낌이나 감성과 같은 것을 효과적으로 표현할 수도 있다. 즉, 형용사로 이루어진 구조화된 지식체계가 구축되면 인간의 감성을 검색할 수 있는 새로운 정보검색시스템이 구현될 수 있을 것이다.

본 연구는 인간의 감성을 수용할 수 있는 형용사로 이루어진 감성 시소러스를 개발하고자 한다. 형용사는 주제나 상황에 따라 여러가지 관점으로 해석할 수 있기 때문에 포괄적인 형태의 감성시소러스(Macro Sensibility Thesaurus)는 일반 주제의 전통적 시소러스를 구성하는 것과 같이 어려운 작업이 될 수 있다. 따라서 효과적인 검색 효율성을 보장하기 위해서는 특정 주제분야에 대한 마이크로 시소러스가 필요하다. 본 연구는 이러한 점을 고려하여 인간의 감성과 색상(color), 용어의 의미(meanings)간의 관계를 고려한 감성 색채 시소러스를 구축하고자 한다. 연구방법은 일련의 색이 갖고 있는 의미를 문자화한 색상 데이터베이스를 기본 데이터로 선정하기 위해 색채학에서 연구된 선행 연구 자료를 비교·분석한다. 수집된 자료를 구조화하기 위해 형용사 디스크립터 구축 전용시스템을 사용한다. 이 시스템은 감성 형용사와 색 이름에서 나타날 수 있는 의미의 애매성과 다양성의 구조화에 중점을 두고 개발하였다.

본 연구결과는 색상 외에도 성상 혹은 성질을 표현하는 형용사로 이루어진 매크

로 감성 시소러스를 위한 기초자료로 활용될 수 있을 것이다.

2. 시소러스의 특성

2.1 디스크립터의 형태

최근에 대부분의 검색엔진이나 검색시스템은 자연언어와 통제언어를 병행하여 사용하기 때문에 어휘통제를 위한 시소러스의 역할이 절대적으로 필요하다(남영준 1998). 왜냐하면 시소러스에 수록된 용어는 해당 자료의 색인과 검색에 실제 활용되고 있는 용어들로 구성되어 있기 때문이다.

이 때 시소러스에 수록된 용어가운데 현재 사용되고 있는 용어들을 디스크립터(descriptor)라고 하며, 각 디스크립터와 관련되어 있으나 주 검색어로 채택되지 않은 유사어 개념의 단어를 비디스크립터(non-descriptor)라고 한다. 즉, 디스크립터는 유사 개념어군에서 추출한 통제된 대표 키워드이고 유사어사전에 등록된 키워드는 비디스크립터라 할 수 있다. 따라서 디스크립터의 형태와 구조는 정보검색시스템의 성능과 검색효율의 향상에 절대적인 영향을 미친다.

디스크립터의 어형은 대부분 명사나 명사구, 명사절로 이루어졌으며, 명사절의 경우에는 형용사를 비롯하여 영문 시소러스의 경우 접속사도 포함되어 있다. 그 가운데 국내에서 구축된 시소러스의 경우는 디스크립터 어형이 명사나 명사구로 이루어져 있으며 이를 구성하고 있는 형태소들은

Infrared bombing 1507

BT Bombing

RT Night bombing

Radar bombing

Strategic bombing

전기자동차 electric automobiles 1306

BT 객차

자동차

전기차량

RT 전기트럭

<그림 1> TEST 시소러스 예

<그림 2> 국방 과학 기술 시소러스 예

대부분 명사에 국한하고 있다. 즉, 영문 시소러스의 경우에는 명사를 제외한 그 밖의 품사들도 발견할 수 있으나, 국내 개발 시소러스의 경우에는 오직 명사로만 디스크립터가 구성되어 있다. <그림 1>은 미국의 국방성에서 1967년에 개발한 TEST 시소러스(Office of Naval Research 1967)의 디스크립터 예이고, <그림 2>는 국내에서 1994년에 국방과학연구소(1994)에서 구축한 국방과학기술시소러스의 디스크립터 예이다.

위의 예에서 확인할 수 있는 것처럼 영어 명사절로 이루어진 디스크립터의 경우는 형용사와 동명사를 사용하여 동작을 나타내는 개념을 표현하고 있다. 한글 디스크립터의 경우는 동작을 표현하는 것의 대부분은 형용사를 사용하지 않고 복합명사를 사용하여 이를 대신하고 있다. 이는 한글이 갖는 조어 특성에 기인한 것으로 판단된다. 예를 들면, '아름다운 여성'과 같이 형용사가 포함된 개념도 '미인'이라는 명사로 변환이 가능한 것과 형용사의 끝에 명사형 어미 '미음(口)'을 붙여 명사로 변환할 수 있는 특성이 있다. 실제 한국심진분류표 20판에 등재된 색인어를 분석한 결과 전체 표목 33,76개 가운데 형용사가 포함된 표목은 존재하지 않으며, '적'과 '성', '및'

과 같은 명사가 아닌 품사가 포함된 표목은 전체대비 6%미만이었다. 또한 주제명 표목표(이재철 1961)의 경우에 형용사가 포함된 표제항은 '아이보는 이'와 '큰 사슴' 등 2개에 불과하였다(남영준 1994).

이상의 연구결과에서 기존의 정보검색과정에 사용하는 검색어 구조는 해외 자료의 경우 형용사가 포함된 검색이 가능하지만, 국내 자료를 대상으로 한 검색과정에서는 명사나 복합명사로 이루어지는 것이 가장 효율적인 검색방법이었다. 왜냐하면 색인 도구라 할 수 있는 시소러스나 주제명 표목표, 분류표의 상관색인이 명사나 복합명사로 대부분 이루어졌기 때문이다.

2.2 질의어(query)와 디스크립터 구조

최근에는 전문정보뿐만 아니라 시각정보, 영상정보 등의 멀티미디어 서비스를 제공하는 시스템이 다양하게 개발되고 있다. 문헌정보학 분야에서 사용하는 전통적인 정보검색의 도구와 방법이 일반 상용 웹검색엔진에서도 원용되고 있다. 예를 들면, 야후(코리아)를 비롯하여 한미르 등 대부분의 웹검색엔진은 웹사이트 초기화면에 키워드 검색창과 함께 분류표와 유사한 목구조 형태로 이루어진 카테고리 서비스를

제공하고 있다. 즉, 카테고리의 항목은 분류표의 표목 혹은 시소러스의 디스크립터로 활용될 수 있으며, 유사어 사전도 사용하고 있다. 특히, 유사 계층 구조를 갖는 개념의 경우에 '@' 표시를 활용하여 '보라'와 '~도 보라'가 참조표시 원리를 수용한 것도 문헌정보학적 방법이라 할 수 있다.

왜냐하면 웹검색엔진의 목적이 정보를 검색하는 용도로 개발되었기 때문에 전통적인 도서관을 통한 정보검색 과정과 그 목적이 매우 유사하다고 판단한다.

한편 정보통신기술의 발달로 일반 이용자들은 웹을 통한 커뮤니케이션에 익숙해짐에 따라 정보검색을 위한 채널을 물리적 도서관에서 웹검색엔진을 보다 적극적으로 활용하게 되었다. 이러한 이용자 변화에 부응하기 위해 대부분의 도서관은 웹이라는 가상공간에 새로운 디지털 도서관을 구축하여 보다 적극적인 서비스를 제공하고 있다.

따라서 실제 이용자들이 사용하는 질의어의 패턴을 찾아내기 위해서는 웹검색엔진과 디지털 도서관에서 정보검색을 위해 사용하는 질의어(query)의 형식을 분석할 필요가 있다.

이를 위해 2003년 11월 20일을 기점으로 야후(코리아)에서 제공하는 지식검색서비스에서 전문가에게 질의한 항목 가운데 '과학박사 3인방'이라는 게시판에 있는 204개의 질문의 유형을 분석하였다. 이 주제 분야를 선정 한 이유는 '과학' 관련 질의어가 내용적인 것이나 질의어 형태적인 것을 고려할 때, 실제 참고사서에게 의뢰하는 질문

의 형태와 가장 유사하였기 때문이다. 분석 결과, 질의어의 형태는 전통적 도서관의 참고 사서에게 행해지는 질의 형태로 자연어 문장과 혹은 명사(복합 명사구 포함)의 형태였다. 전체 질의문에 대해 형태소 해석기를 사용한 결과 형용사가 포함된 질의어는 27개였다. 해당 형용사 가운데 후치하는 명사를 수식하여 실질적으로 개념의 범위를 제한하거나 분명하게 한 문장은 7개 질의문이었다. 불용어 사전을 이용하여 색인어가 될 수 있는 명사나 명사구를 추출한 결과 '질문이 있습니다'라는 질의어를 제외한 203개 질의어 문장에서 색인어로 활용될 수 있는 (복합)명사(절)를 추출할 수 있었다. 즉, 자연어 입력 형태로 검색 질의문을 구성한 경우나 혹은 적절한 불린식을 이용한 키워드 검색이나 유사한 검색결과가 나타날 수 있음을 확인하였다. 또한 검색어으로써 형용사의 중요성을 확인할 수 있었다. 왜냐하면 형용사를 포함한 모든 질의문이 다른 품사는 제거되어도 '상치난 부위'와 같은 어구에서 해당 형용사를 제외할 경우에 이용자의 검색의도가 왜곡되었기 때문이다.

2.3 디스크립터의 관계성

2.3.1 전통적 관계성

전통적으로 시소러스내에서 디스크립터의 관계는 디스크립터의 속성(attributes)에 따라 계층관계(hierarchical relationships)와 연관관계(associative relationships), 유사관계(equivalence relationships)로 구분하고 있다. 관계설정의 기준은 시소러스의 특성과

함께 디스크립터가 내포하고 있는 의미를 직관하는 국면(facet)에 따라 변화될 수 있다. 예를 들면 A라는 용어와 B라는 용어가 완전하게 내삽되어 있는 경우에 이 두 용어의 관계는 계층관계이거나 혹은 유사관계라 할 수 있다. 즉, 자동차라는 개념내에는 디젤자동차가 완전하게 내삽되었다고 할 수 있어 이는 계층관계로 표현될 수 있다. 다만 해당 시소러스가 매크로 시소러스일 경우에 디젤자동차가 특정성이 높아 디스크립터로 채택되지 않을 경우에는 자동차의 유사어 혹은 대등어으로써 디젤자동차가 채택될 수 있다. 그렇지만 이를 계층관계로 설정하였다고 검색의 오류나 왜곡현상은 발생하지 않으며, 다만 검색결과외의 정도에 영향을 미칠 것이다.

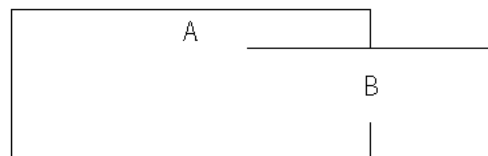
한편 A라는 용어와 B라는 용어가 <그림 3>과 같이 의미상 부분 공유일 경우에 서로 간의 관계성을 설정하기가 상대적으로 어렵다. 공유의 범위가 클 경우에 이를 계층관계로 설정하는 것이 일반적 견해이나 (남영준 2002) 해당 용어가 느낌이나 감정을 표현하여 수치적으로 측량하기 어렵기 때문에 정확한 관계설정이 어려운 것이다.

정영미(1997)는 연관관계를 '상호 계층적 관계는 아니면서 개념적으로 어떤 관련성을 갖는 용어들의 관계를 연관관계라고 말하며.. 종략.. 이 관계의 설정을 정확하게

정의하기 어렵다'라고 정의하여 관계설정의 애매성을 표현하고 있다. ISO에서도 연관관계를 '계층적이 아니라 개념적으로 밀접하게 관련되어 있으나, 등가집합에는 포함되지 않은 용어간의 관계이다. 연관관계로 이루어진 용어는 색인작성과 탐색에 이용될 가능성이 있는 대체용어를 제시하는 관점에서, 용어간의 연결을 시소러스 중에서 명시하는 것이 보다 좋을 것이라는 심리적인 연상관계에 따라 만들어진 것이다. 이 관계는 상호보완적인 개념간에 이루어진다(ISO 1986)'고 정의하고 있다.

이러한 관계 설정에 대한 명확한 설명이 없기 때문에 Aitchison, Gilchrist, and Bawden (2000)은 연관관계의 설정이 다른 관계설정보다 어렵다고 주장하며 잘 정돈된 연관관계설정은 보다 효율적인 정보검색효율을 유지할 수 있다고 하였다.

이와 같은 연관관계 설정의 어려움은 다른 관계설정의 경우 유사한 개념집단의 용어에 한하여 유사관계나 계층관계를 설정하나, 연관관계는 상이한 개념어군에서도 관계설정이 이루어질 수 있기 때문이다. 즉, 대등관계나 계층관계의 경우 유사 용어군에서 색인자 편의에 따라 각각의 관계가 유연하게 이루어지고 그 설정결과를 객관화할 수 있다. 이에 비해 연관관계는 유사 개념군 외에 서로 다른 개념군의 용어간에



<그림 3> 의미 내삽과 외연현상의 예



<그림 4> 용어간 개념 불합치의 예

도 연관관계가 설정될 수 있기 때문에 시소러스 구축자나 기관의 성격에 따라 주관적인 행위가 될 수 있다.

<그림 4>와 같이 별개의 개념들이 특정한 상황에 따라 연관관계로 표시된 예이다. A와 B는 아무런 유전적 연관성이 없음에도 불구하고 사회의 관례 혹은 인간의 인지 형태에 따라 서로를 연관관계로 규정짓는 관계이다. 예를 들면, '입시와 한파'라는 용어간의 관계에서 '입시'와 '한파'는 개념적으로 아무런 유전적 관계가 없음에도 불구하고 사회적 인지도 때문에 연관관계로 표현될 수 있다.

2.4 시소러스의 구축

시소러스의 구축 과정은 학자마다 조금의 차이는 있으나 대략적인 절차는 매우 유사하다. 시소러스의 형태나 언어, 범위에 크게 좌우되지 않으며, 용어의 수집 대상과 범위, 평가 등의 모든 과정에 대한 적용 기법만 차이가 있을 뿐이다. 크게 용어의 수집을 비롯하여 용어의 개념 구조화, 평가 및 유지보수 등 세 단계로 이루어진다.

2.4.1 용어의 수집

시소러스 구축의 첫 번째 작업은 주제 분야에 대한 주요 용어를 수집하는 것이다. 수집은 전문가와 같이 수작업에 근거한 방법과 컴퓨터를 이용하여 통계값에 근거한

방법으로 구분할 수 있다. 특히 기계를 이용한 용어 수집은 크게 용어 추출 정보원의 수집과 자연어처리방법을 이용한 후보 디스크립터의 추출과정으로 다시 구분될 수 있다.

시소러스에 등재될 후보 디스크립터를 선정한 후에 디스크립터와 비디스크립터 이외에 후보 디스크립터를 구분하는 것이 바람직하다(남영준 2002). 후보 디스크립터는 용어 풀에 입력하여 실제 시소러스에는 등재되지 않지만 필요에 따라 유연하게 시소러스에 업/다운로드될 수 있는 용어를 의미한다.

2.4.2 용어의 개념 구조화

용어의 개념 구조화는 용어간의 관련성을 파악하여 관계를 설정하는 과정을 의미한다. 이를 위해서는 특정 개념을 기준으로 관련있는 용어를 군집화하는 과정이 필요하다. 군집화가 이루어진 후에는 개발자나 개발 기관의 목적에 따라 용어를 구조화한다. 이 때 적절한 수의 디스크립터와 비디스크립터를 유지하여 적절한 검색효율을 얻을 수 있는 시소러스가 되도록 해야 한다. 특히 구조화 작업은 완전하게 수작업에 의존하기 때문에 개발자의 용어선정 및 관계설정의 오류를 방지할 수 있는 시소러스 개발 프로그램을 사용하여야 한다. 이 프로그램에는 용어의 관계 충돌 및 동형의어의 중복 채택을 시스템적으로 검출하는 기

능이 반드시 포함되어야 한다.

2.4.3 평가 및 유지보수

구축된 시소러스는 반드시 평가가 이루어져야 한다. 각 디스크립터를 실제 검색어휘로 사용하여 계층 및 유사 관계는 검색 결과의 내포 크기로 결정하며, 연관 관계는 상위어나 계층어의 검색결과 중첩도로 평가할 수 있다. 이 과정에서 의미는 있지만 실제 해당 분야에서 사용되지 않는 사어(死語)와 다른 의미로 사용되고 있는 동형이어 등을 수정하여야 한다. 특히 개념간의 부분적으로 일치된 개념을 공유한 디스크립터의 관계의 배정과 구조화에 대한 적절성 여부는 실제 실험과 함께 해당 분야 전문가의 자문을 받아야 한다.

3. 감성시소러스

3.1 감성 어휘

인간의 감성은 감정과 분명하게 구분되는 심리적 현상이며, 개인의 생활경험에 의하여 발생하는 것이다(이구형 1998). 이에 비해 감정은 일반인들이 갖는 선형적 의식에 의해 얻어지는 것이다. 인간의 감정은 전통적으로 희노애락애오욕(喜怒哀樂愛惡慾)으로 표현하고 있다. 이는 인간의 시간적, 공간적, 지리적 차이를 구분하지 않고 일반적으로 모든 인류에게 적용될 수 있는 선형적 지식이다. 즉, 감성은 경험적 지식에 근거하고 개인에 따라 인지도가 크게 발생하는 지식이라면, 감정은 선형적 지식

에 따라 개인에 따라 인지도가 크게 차이 나지 않는 지식이라 할 수 있다. 예를 들면, 인간이 몸이 편하면 모두 즐거워하지만(감정), 눈이 내리는 바다를 보는 인간의 느낌(감성)은 개인의 경험과 지식에 따라 차이가 발생한다. 감성은 외부의 자극에 대하여 반사적이며 직관적으로 발생하며, 개인성과 함께 애매모호하고 시간과 환경에 따라 변화가 심한 특성을 가지고 있기 때문에 이를 정량화하고 객관화하기가 상대적으로 어려운 작업이다. 또한 감성을 객관적으로 측정하기 어렵기 때문에 형용사 어휘로 인간의 내재된 의미를 파악할 수 있다. 이 경우 다른 문화권의 감성 모형을 그대로 사용하는 데에는 지역차에 따른 감성 반응의 차이 외에도 번역이 어려운 표현들이 있을 수 있다(박수진, 정찬섭 1998). 따라서 외국의 형용사로 표현된 감성의 의미를 국내에 그대로 적용하기 어려운 것이다. 다만 동일 시대의 동일 지역내에서는 개인에 따라 선호하는 감성 표현의 차이가 있을 수 있지만 선호하는 감성 표현의 경우 사용하는 어휘는 다를지라도 어휘들이 나타내는 바는 공통된 감성으로 표현되는 경우가 많다(박미자 외 1998). 즉, 유사한 지식과 경험을 갖고 있는 집단이나 사전에 의미가 정의된 주제분야에서는 감성 어휘들의 의미가 동일하게 적용될 수 있다. 인간의 느낌이나 감성은 언어적으로 어느 정도 표현이 가능하므로 감성어휘의 수집을 통해 보편적인 감성을 추출할 수 있을 것이다. 추출된 대표 어휘는 위계 구조에 대한 분석이 필요하다. 감성 표현 어휘들의 위계 구조에 대한 분석이 선행되어야 감성

범주 대표어를 보다 합리적으로 선정할 수 있을 것이다(박미자와 1998). 또한 선정된 대표어를 중심으로 구조화가 가능하며, 이는 색인 및 검색시 정확한 감성어휘의 사용을 가능하게 할 것이다.

3.2 감성 어휘와 형용사

감성은 인간의 머릿속에서 어떤 이미지 구조를 이루고 있다. 감성의 이미지 구조는 실제로 관찰될 수 없는 것이기 때문에 형용사를 사용하여 유추할 수 있다. 이러한 의미에서 감성을 표현하는 형용사는 감성디스크립터로서 매우 중요한 의미를 갖는다. 따라서 비문자로 표현되어 있는 이미지나 영상 데이터를 색인하는 과정에서 감성의미를 표현하는 언어정보가 중요한 요체가 될 수 있다. 이들은 감성에 근거하고 있기 때문에 감각정보라 할 수 있으며, 이를 용어로 표현한 것을 감성 용어라 할 수 있다.

유명희와 최석두(2002)는 형용사의 개념관계를 표현하기 위하여 형용사의 특성, 관련 개념구조를 고찰하고 아울러 구조화, 관계, 표시방법, 배열 등을 고려하여 형용사 시소러스를 설계하였다. 그들은 이 연구에서는 형용사의 계층관계를 설정하지 않고 관련어 위주로 시소러스를 구축하였다. 또한 동의어와 유사동의어에 대하여 해당 어휘의 속성을 기술하여 서로를 구별할 수 있도록 하여 언어 사전적인 일부 역할을 수행하도록 하였다. 또한 다의성이 있는 어휘에 대하여 패킷을 두어 상이한 의미단위로 별도의 용어구조를 갖게 하였다.

3.3 색채이름과 형용사

색채에서 얻는 이미지는 감정적인 측면보다 감성적인 측면으로 인식된다. 왜냐하면 동일한 색채를 통해 항상 동일한 느낌을 얻지 않기 때문이다. 이는 해당 색채를 인식하는 사람의 과거 경험이나 지식에 영향을 받기 때문이다. 또한 색채는 인간에게 개별적이고 감성적이고(김형민 1996) 상징성이 있기 때문에 인간의 의사 전달의 수단으로 사용될 수 있다(Arneim 1963).

한편 인간 의사전달 과정은 공통적인 지식배경을 전제로 특정한 기호로 전달되는 것이다. 따라서 주관적인 이미지를 어떠한 기호없이 제삼자에게 전달하는 과정에는 잡음(noise)이 생성되기 때문에 반드시 문자와 같이 표현될 수 있는 기호로 의미(이미지)가 전달되어야 한다. 즉, 의사전달이 완전하게 이루어지기 위해서는 개인이 갖고 있는 주관적 감성 이미지를 일반인들이 인식할 수 있는 객관적 문자로 표현해야 한다. 색채 이미지에 관한 커뮤니케이션이 성공하기 위해서는 색채 이미지에 대한 객관화 작업이 필수적이다. 따라서 객관화 작업을 위해 국제 표준 색상명 개발과 함께 각 국에서는 자국의 표준 색상명을 개발하였다.

예를 들면, 우리나라는 한국산업규격 기본색 이름(기술표준원 2003)(한국산업규격 개정안-KSA0011)을 공시하여 국내에서 사용되는 색에 대해 계통색과 관용색 등 두 가지로 구분하여 모든 색이름에 대응되는 적용 색채 표현(형용사 포함)과 색상 범위, 조합색에 관한 정보를 제공하고 있

작성시에 상황에 맞는 적절한 배색결과를 도출하기 위한 데이터로 사용된다.

미국의 Hewlett-Packard사도 색이 갖고 있는 의미를 20가지의 색을 대상으로 정의하고 있다. 이 정보는 파워포인트와 같은 발표 자료 작성시에 활용할 수 있도록 고객 지원용 데이터에 포함되어 있다. <그림 7>은 HP(2003)에서 제안하는 빨강(bright red)의 이미지를 감성 형용사로 표현한 예이다.

이상과 같이 외국에서는 색채가 갖고 있는 감성적 의미를 형용사로 표현하여 각 색채가 갖고 있는 무형적 의미를 객관화된 기호로 변환하는 연구가 활발하게 진행되고 있다. 이러한 연구결과는 대부분 관찰에 의한 통계값으로 이루어지는 것이다. 따라서 감성은 지리적, 공간적, 시간적 차이에 따라 다르게 표현될 수 있기 때문에 외국의 연구결과를 국내 감성으로 활용하기 위해서는 최소한의 수정작업이 필요하다.

국내에서는 색채에 대한 감성의 기호화 작업이 외국에 비해 상대적으로 활발하게 이루어지지 않고 있는 실정이다. 국가차원 지원에 따라 일부 연구소에서 감성의 기호

화 작업이 이루어지고 있다. IRI사는 통상 산업부의 지원을 받아 산업기반 기술사업의 일환으로 우리나라 사람들의 감성을 시각적이며 기호적인 관점에서 기초하여 I.R.I 형용사 이미지 스케일을 개발하였다. IRI의 이미지 스케일(2003)은 추상적 이미지를 색으로써 구체적으로 표현하였으며, ‘부드러운(Soft) / 딱딱한(Hard)’ 과 ‘동적인(Dynamic) / 정적인(Static)’의 두 속성을 이미지 스케일 기준으로 하였다. 색채 이미지 요인 구조는 감성적 요인, 활동성 요인, 명도·채도에 대한 평가 요인, 기호 요인, 복잡요인으로 구성되어 있다. 또한 감성요인, 색채 이미지 공간을 설정하고, 각각의 색을 고유한 이미지 공간상에 배치하였다. 색채의 형용적 의미를 들었을 때 연상되는 색채를 형용사 영역과 같은 맵상에 배치하여 활용할 수 있도록 하였다. IRI 색채 이미지는 ‘맑은’을 비롯하여 색채값이 형용사로 변환된 12개의 기본 영역으로 구성되어 있다. 색상(Hue) & 색조(Tone) 120색을 형용사 의미 분별법과 요인분석법을 이용하여 고유한 이미지 공간상에 배치하였다. “ Hue & Tone 120”과 “ 898 Color System”



Bright Red - Optimistic, dynamic, energizing, exciting, sexy, intense, stimulating, aggressive, powerful, energetic, dangerous

International significance - China = good luck; India = purity; Eastern cultures = signifies joy when combined with white

<그림 7> HP사의 색채 對 감성언어 대조표

지 개념이 포함되기 때문에 주제나 상황에 따라 다차원 국면으로 해석이 가능할 수 있다. 본 연구에서는 한국인 색채 감성으로 주제를 한정하기 위해 국내에서 개발한 형용사 스케일을 사용하였다. 따라서 주 용어 수집원으로 IR에서 개발한 "한국인 색채 감성 척도의 개발에 관한 연구"에서 제시한 단색 이미지 스케일과 형용사 이미지 스케일을 이용하였다.

한국인 색채 감성척도의 개발에서 형용사를 수집하기 위해 조사대상자로 하여금 자유롭게 떠오르는 단어를 제출하도록 하고, 자유연상법(전문 디자이너와 일반인 대상으로 실시)과 디자인 관련 서적, 잡지, 초등학교 국어 교과서, 국어사전 등을 통해 수집하였다. 수집된 형용사는 SD법을 이용한 이미지 평가 및 이미지 공간 구성에 사용하기 위해 체계적으로 정리되었으며, 그 타당도는 관련 학계 자문진으로부터 검증받은 것이다(IRI 1997). 따라서 이 결과값은 외국의 결과에 비해 우리나라 사람의 감성을 가장 잘 표현한 것이라 판단한다. 또한 일반인과 전문디자이너를 통한 설문으로 표본 통계값을 얻고 관련 학계 자문단으로부터 표준화 절차를 거쳤기

때문에 학술적이며 일반적 기초자료로 활용할 수 있다. 색 이름과 실제 색상과의 연결을 위해 IRI의 색채이미지를 먼셀기호로 변환하였다. 변환의 기준은 웹 컬러값으로 하였으며, 최종적으로 먼셀기호를 한국산업규격 기본색이름으로 변환하기 위해 디지털검색팔레트프로그램(한국디자인진흥원 2003)을 사용하였다. 수집된 디스크립터의 수는 117개 였다. 이 외에 Kobayashi와 Haruyoshi의 색채 대 감성 언어 차트에서 개념의 중복이 이루어진 곳에서 144개의 형용사를 채택하였다. 즉, 비디스크립터를 포함하여 261개를 채택하였다. <표 1>은 수집한 색상값과 대응 형용사의 일부에이다.

4.1.2 관계설정

관계설정은 계층·유사관계와 연관관계로 구분하여 이루어졌다. 본 연구에서 사용한 디스크립터는 모두 IR에서 구축한 형용사 이미지 스케일 좌표에 표시된 감성형용사를 기준으로 사용하였다. 또한 유사어의 수집과 연관 관계설정의 데이터를 수집하기 위해 Kobayashi와 Haruyoshi가 구조화한 감성 형용사 스케일을 원용하였다.

<표 1> 수집된 색채·형용사

웹컬러	먼셀기호	계통색명(국문)	계통색명(영문)	형용사	대표형용사
ff0033	5R 4/16	해맑은 빨강	vivid red	선명한	경쾌한
ff9900	5YR 6.5/14	해맑은 주황	strong orange	경쾌한	경쾌한
ff33cc	5RP 4.5/13	해맑은 자주	vivid reddish purple	젊은	경쾌한
cc3333	5R 4.5/12	빨강	strong red	돋보이는	경쾌한
33ffee	5GY 7/9	연두	strong yellow green	자유로운	경쾌한

(1) 계층 및 유사 관계 설정

색채 선호를 기준으로 한 IRI의 군집 분석에 분류된 12개의 이미지 군을 대표하는 형용사를 주요 상위어로 선정하였다. 이에 속하는 것으로는 '귀여운'을 비롯하여 맑은, 온화한, 경쾌한, 내츨릴한, 화려한, 우아한, 은은한, 다이내믹한, 모던한, 점잖은, 고상한 등이 포함된다. 계층의 깊이는 3단계로 하였다.

대등관계는 색채관련 형용사로 주제범위를 제한하여 형용사 어휘와 색상명과 색채기호를 동의개념으로 설정하였다. 형용사에 대응하는 색상을 대등관계로 설정하였으며, 사용된 색상명은 한국산업규격 기본색 이름을 기준으로 선정하였다. 또한 웹 컬러(16진수 표기), 먼셀기호 등에 부기된 색이름도 디스크립터로 추출된 것과 일치할 경우에 모두 대등어로 선정하였다. 또한 유사어(비디스크립터)의 정보원으로 Kobayashi와 Haruyoshi가 구조화한 자료에 주요 상위어가 일치하는 클러스터 내에서 웹 컬러값이 일치하면서 달리 번역된 것은 모두 유사어 처리를 하였다. 이는 감성에 대한 표현이 해당 자료의 번역자와 IRI의 감성 형용사의 이형(異形)이기 때문이다.

(2) 연관관계 설정

연관관계는 하나의 상위개념어를 공유하는 용어이면서 중첩되는 의미를 갖는 용어간에 필요하다. 범주화한 12개의 상위 그룹에 대해 그룹내 직선구조는 계층관계로 표시하고, 계층관계에 속하지 않는 클러스터와 형용사에 대해서는 다른 분류 그룹과의 관계를 고려하여 연관관계를 설정

하였다. 직선구조에 속하는 지의 기준은 형용사 이미지 스케일에 있는 공간의 거리 및 좌표상 위치를 활용하였다. 일차적으로 IRI 형용사 이미지 스케일을 기준으로 Kobayashi와 Haruyoshi 상위어 유사도를 측정하였다.

다음 <표 2>는 상위어간 유사도 측정을 하기 위해 형용사를 매트릭스화한 것이다. 괄호안의 숫자는 해당 영역에 배정된 용어의 개수이다. IRI 형용사의 상위어와 다른 형용사의 상위어가 완전하게 일치하지 않는 것은 제외하였다. 주요 감성 형용사 상위어를 비교한 결과, 의미적으로 완전하게 일치하는 것은 IRI의 기준으로 열두 개의 영역가운데 일곱개 영역이 일치하였다. Haruyoshi와 형용사 이미지 스케일과 IRI의 이미지 스케일은 스물 세 개 영역가운데 여덟 개 영역이 일치하였다. 이미지 스케일의 영역군이 다른 이미지 스케일에 비해 두배 이상 세분되어 있기 때문에 배정된 용어의 수가 기본 영역에 상대적으로 적게 배정되어 있었다.

두 번째, 위치적으로 인접한 용어는 연관관계로 설정하였으며, Kobayashi의 연구에서 클러스터의 중복 영역에 있는 형용사도 연관관계를 설정하는 데이터로 활용하였다. 이 영역에 속하는 용어는 다음과 같다.

- Elegant 영역과 Natural 영역의 중복 : 정숙한, 다정한, 섬세한, 소박한, 자연스런, 정서적인, 마른 (7개 형용사)
- Dandy 영역과 Formal 영역 : 귀중한, 엄숙한, 장엄한, 격조 높은 (4개 형용사)
- Dandy 영역과 Modern 영역의 중복 :

<표 2> 주요 감성 형용사 상위어 비교 표

Kobayashi 형용사		IRI 형용사		Haruyoshi 형용사
Clear(10)	≒	맑은	≒	클리어(10)
-		온화한		
Pretty(3)	≒	귀여운	≒	프리티(5)
Natural(12)	≒	내추럴한	≒	내츄럴(8)
-		경쾌한		
Gorgeous(11)	≒	화려한	≒	고저스(4)
Elegant(18)	≒	우아한	≒	엘라강스(6)
-		은은한		
-		다이나믹한		다이나믹(4)
Modern(14)	≒	모던한	≒	모던(5)
Dandy(15)	≒	점잖은	≒	댄디(4)
-		고상한		노블(6)

송고한, 성실한, 정식의 (3개 형용사)

Haruyoshi의 이미지 스케일은 Kobayashi의 이미지 스케일에서 중복된 부분을 별도의 영역으로 선정하여 용어(형용사)의 중복성을 제외하였다.

<그림 9>는 이상과 같이 자료를 분석한 결과 얻어진 디스크립터 가운데 '성숙한'이라는 감성 형용사의 계층관계를 비롯하여 연관관계, 유사관계를 구조화한 자모순 시소러스의 예이다.

성숙한

UF 5P 3/11

질은 보라 띠 파랑

BT 화려한

RT 시원한

장식적인

<그림 9> 관계 구조화의 예

4.2 감성 시소러스 구축 시스템

기존의 시소러스 구축 프로그램은 명사나 명사구의 형태만을 수용하도록 설계되어 있으며, 용어풀과 같은 용어 저장소에 대한 고려가 없었다.

본 시소러스 관리 프로그램은 감성기호인 형용사의 구조와 어절형태를 수용하고, 다양한 의미로 해석될 수 있는 디스크립터를 유연하게 구조화할 수 있도록 설계

되었다. 즉, 형용사는 감성을 표현한 언어이므로 시간·공간적 차이에 따라 다르게 해석될 수 있기 때문에 관계설정이 유연하여야 한다. 예를 들면, 전통적으로 하위어는 반드시 차상위어를 하나만 갖기 때문에 이러한 원칙이 시소러스 구축 프로그램에도 적용되어 계층구조내에서 디스크립터는 복수로 설정되지 않도록 설계되어 있다. 한편 색상 형용사는 국가간 혹은 기관에 따라 감성의 표현이 다르기 때문에 하나의 색이 복수의 색에 배정될 수 있는 특성을 갖고 있다. 따라서 색상에 기반한 감성 시소러스를 구축하기 위해서는 기본 시소러스 구조를 유지하고 중복 색상에 대해 별도로 구조화가 가능하도록 설계하는 것이 필요하다.

본 프로그램의 대표적인 기능을 열거하면 다음과 같다.

- 디스크립터의 입·출력 기능 : 신규 디스크립터 입력과 같은 기본적인 입력 모드 이외에 외부데이터의 일괄 입력과 일괄 반출 기능, 인쇄본 형태(텍스

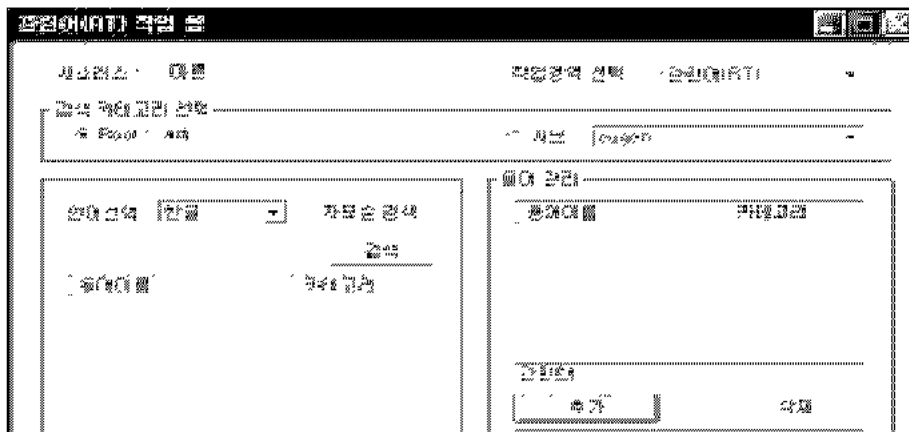
트와 HTML 형태)를 갖고 있다.

- 디스크립터 관계설정 기능 : 기본 관계(계층, 연관, 대등)를 설정하는 기능과 대응 외국어 설정기능도 갖고 있다.
- 관리자 기능 : 원격 접속이 가능하기 때문에 입력자별로 작업 수준을 지정하고, 입력자별 업무 양을 측정할 수 있는 기능을 갖고 있다. 별도의 용어 풀을 지정하여 특정 디스크립터에 대해 별도의 계층 및 연관 구조를 설정하는 기능도 있다.

4.2.1 신규 디스크립터 입력의 예

본 시스템에서는 신규 디스크립터의 입력시에 해당 용어가 기존 데이터베이스에 존재하는 지를 자동 검출하는 기능을 갖고 있다. 또한 신규 디스크립터를 고립어로도 업로드할 수 있으나, 관계설정시에 다른 디스크립터와의 관계 충돌도 자동적으로 검출할 수 있다.

다음 <그림 10>은 신규 용어로 '내추럴한'군의 '마른'이라는 용어를 입력하면서



<그림 10> 신규 디스크립터 입력 예

‘ 간편한’ 용어를 연관 관계로 설정하여 입력하는 예이다.

4.2.2 디스크립터 풀(pool) 운영의 예

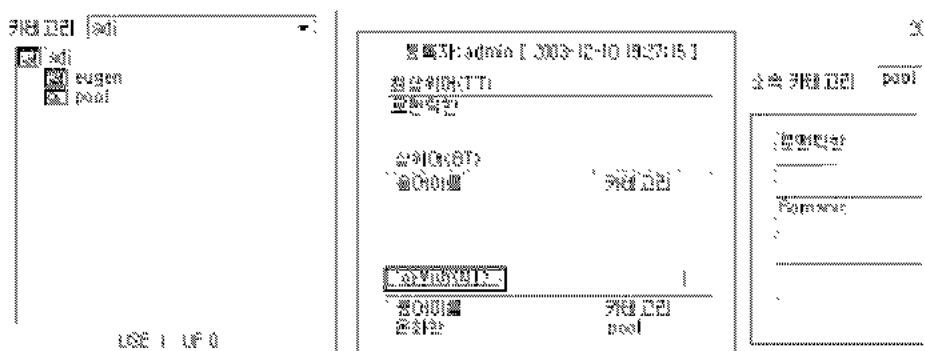
감성 형용사는 앞서 제기한 바와 같이 동일한 감성이 다르게 표현될 수 있으며, 다른 감성이 동일한 감성으로 표현될 수도 있다. 예를 들면, IRI의 상위어에 속하는 ‘ 온화한’이라는 개념이 있으나, Kobayashi의 이미지 스케일에서는 ‘ Romantic’영역에 속해있는 하위어로 설정된다. 따라서 이러한 다양한 감성 표현이 감성 형용사의 특징이기 때문에 시스템은 이를 수용하여 이 감성 논리를 보관할 필요가 있다.

<그림 11>은 ‘ adj’카테고리에 ‘ eugen’이라는 주카테고리와는 별도로 ‘ pool’이라는 카테고리를 구성한 예이다. 이 ‘ pool’에는 단순히 용어의 보관이 아니라 이 카테고리 내에서도 주카테고리와 별도로 새로운 형태의 감성시소러스 구조를 유지할 수 있다. 현재 구축하고 있는 소속 카테고리는 ‘ pool’이라는 공간이며 ‘ eugen’이라는 주영역과 별도로 운영되고 있음을 보여주고 있다. 즉, ‘ eugen’이라는 영역에서 사용하

는 모든 디스크립터를 ‘ pool’이라는 영역에서 사용하면서 관계 및 개념간 충돌이 없이 용어 보관과 구조화가 가능한 것이다.

5. 결 론

시소러스 고유의 역할은 정보검색과정에서 적절한 검색효율을 조절할 수 있도록 하는 것이다. 전통적인 정보검색도구들은 대부분 키워드 검색방법을 대비하여 표목이나 디스크립터들이 명사나 명사구의 형태로 개발되었다. 한편 도서관에서는 비문자 형태로 이루어진 그래픽과 동영상 파일 등도 새로운 장서로서 수집하며, 인터넷과 같은 가상공간에서 실시간으로 이러한 자료들의 다운로드가 가능하게 되었다. 또한 정보의 양이 수작업으로 통제하기 불가능한 수준으로 팽창함에 따라 기존의 정보검색도구를 활용하여 새로운 자원의 형태를 효과적으로 검색하지 못하게 되었다. 따라서 도서관과 이용자들은 보다 정교한 검색도구를 필요로 하였으며, 이러한 이용자 요구를 수용하기 위한 방안가운데 형용사가



<그림 11> 시소러스 프로그램의 용어 풀 운영 예

포함된 시소러스 개발의 당위성은 매우 높다고 할 수 있다.

따라서 본 연구에서는 색상에 기반한 시소러스 개발을 위해 일련의 조사를 수행하였으며 조사 내용을 요약하면 다음과 같다.

- 1) 기존에 개발된 국내 검색어집에 등재된 검색어구는 대부분 명사로 이루어져있으며 형용사는 거의 사용되지 않고 있었다.
- 2) 이용자가 시스템을 통해 입력하는 질의문의 형태가 단순하게 키워드와 불린식을 이용한 검색방식에서 자연어 문장으로 검색하는 형태로 변화하고 있었다. 이 때 검색 질의문의 구성은 명사나 복합명사, 명사구가 대부분이었으며, 이 때의 명사구 형태는 형용사가 키워드의 개념을 정교하게 제한하고 있음을 발견할 수 있었다.
- 3) 감성은 지역적, 공간적, 시간적 차이에 따라 서로 다르게 인식될 수 있으며, 경험적 지식이 개념이 갖는 의미를 다양하게 해석한다.
- 4) 따라서 대표적인 감성을 나타내는 색상의 경우도 지식에 따라 의미가 변화될 수 있다.

이러한 기초 연구에 근거하여 본 연구에서는 색상에 기반한 감성 시소러스를 구축하였다. 구축은 형용사와 같이 다양한 의미를 수용할 수 있는 전용 프로그램으로 이루어졌다. 구축의 대상은 우리나라 IR사에서 개발한 형용사 이미지 스케일에 수록된 감성 형용사를 대상으로 하였다. 구축한 디스크립터의 수는 117개이고, 비디스크립터의 수는 144개로 수록된 용어의 수는 261

개였다. 구축과정에서 발견된 결과를 요약하면 다음과 같다.

- 1) 국내에서 개발한 형용사 이미지 스케일과 일본에서 개발된 이미지 스케일의 비교한 결과 상위어는 약 80%가 일치하여 이를 유사어로 처리하였다.
- 2) 관계설정에 있어 색상이 갖는 애매성 때문에 의미를 기준으로 연관관계를 설정하기보다는 휠컬러(16진수)값과 같이 객관적인 기준으로 설정할 필요를 발견하였다.

본 연구의 결과는 감성에 기반한 검색시스템 개발에 중요한 기초자료로 활용될 수 있다. 제한점으로는 구축된 결과에 대한 실험을 수행하지 않아 어느 정도의 검색효율을 얻는지를 확인하지 못한 것이다. 또한 적정한 수의 디스크립터를 보유하고 다양한 색상 형용사가 포함된 시소러스 구축을 위해서는 이와 관련된 대규모적인 조사가 수반되어야 한다.

참고문헌

- Aitchison, J., Gilchrist, A., and Bawden, D. 2000. *Thesaurus Construction and Use: a Practical Manual*. 4th ed. Aslib.
- Arnheim, R. 1963. *Systematic Method for Designers*. 재인용 : 서종환, 『제품디자인에 있어서의 사용자 참여적 프로세스에 관한 연구』 석사학위논문, 한국과학기술원.
- Color-Wheel-Pro. "Color meanings." [cited

2003. 11. 27]. <<http://www.color-wheel-pro.com/color-meaning.html>>.
- Haruyoshi, Nagumo. 김상두 번역. 2000. 『칼러 이미지 차트』 서울 : 조형사.
- Hewlett-Packard. "printing with color - the meaning of color." [cited 2003. 11. 27]. <<http://www.hp.com/sbso/productivity/color/meaning.html>>.
- I.R.I. 1997. 『한국인 색채감성 척도의 개발에 관한 연구』 IRI
- I.R.I. "I.R.I.컬러시스템" [online]. [cited 2003. 11. 28]. <<http://www.iricolor.com/04.colorinfo/colorsystem.html>>.
- ISO. 1986. *Documentation-Guidelines for the Establishment and Development of Monolingual Thesauri. 2nd edition* (ISO 2788-1986(E)). 정동열 역. 1994. 『단일언어 시소러스 제정 및 개발 지침, 시소러스 개발지침』, 문헌정보처리연구회.
- Kobayashi, Shigenobu. 1990. 『Color image scale』. New York : Kodansha America.
- Munsell, Albert Henry. 1915. 『Atlas of the Munsell Color System』. Malden, Mass; Wadsworth, Howland.
- Office of Naval Research. 1967. 『Thesaurus of Engineering and Scientific Terms』, 동연연구소.
- 국방과학연구소. 1994. 『국방과학기술 시소러스』 대전 : 동연연구소.
- 기술표준원. 2003. 『한국산업규격안: 물체색의 색이름-KS A 0011』, 서울 : 기술표준원.
- 김형민. 1996. 『컴퓨터를 활용한 색채 기호 조사에 관한 연구』, 석사학위논문, 한국과학기술원.
- 남영준. 2002. 『고속철도건설공단 시소러스 개발 및 구축』, 서울 : 쓰리소프트.
- 남영준. 1998. 이용자중심의 시소러스관리 프로그램 설계. 『전주대학교 공학기술종합연구소 공학기술논문집』, 14(2): 225-242.
- 남영준. 1994. 『색인어 형태분석에 의한 한국어 자동색인기법 연구』 박사학위논문, 중앙대학교 대학원, 문헌정보학과.
- 박미자, 신수길, 한광희, 황상민. 1998. 감성 측정을 위한 우리말 형용사의 의미 구조. 『한국감성과학회지』, 1(2).
- 박수진, 정찬섭. 1998. 우리말 감성어휘의 범주-차원 모형. 『한국감성과학회지』, 2(1).
- 유명희, 최석두. 2002. 형용사시소러스 설계에 관한 연구. 『제9회 한국정보관리학회 학술대회 논문집』, 203-204.
- 이구형. 1998. 감성과 감정의 이해를 통한 감성의 체계적 측정 평가. 『한국감성과학회지』, 1(1)
- 이재철 편. 1961. 『주제명 표목표』, 서울. 연세대학교 도서관학과.
- 정영미. 1997. 『정보검색론』, 서울 : 구미무역.
- 한국디자인진흥원 (KIDP). [online]. [cited 2003. 11. 5]. <http://www.designdb.com/color/3_00.asp>.