

# 專門情報 檢索시스템의 시소러스 開發에 관한 研究

A Study on Development the Thesaurus for  
the Specialized Information Retrieval System

이상렬

(성균관대학교 문헌정보학과 석사과정)

Lee, Sang-Loel

(Student of Graduate School of Sung Kyun Kwan Univ.)

본고에서는 專門情報 檢索시스템의 檢索效率을 높일 수 있는  
방안에 하나로서 해당주제분야별 專門情報의 시소러스 開發  
을 제안하고, 이를 통해 利用者 중심의 專門情報 檢索시스  
템을 구축하는데 도움을 줄 수 있는 방안을 모색하고자 한다.

## I. 序論

情報技術의 발전은 情報檢索시스템에도 많은  
영향을 끼쳤다. 특히 1960년대부터 컴퓨터에  
의한 情報檢索시스템이 소개되기 시작했고,  
1970년대 전후로 情報通信技術의 발전과 더불어  
온라인 情報檢索시스템이 주로 이용되기 시  
작하였다. 또한 채널을 통해 전송되는 情報의  
내용도 日常生活情報에서부터 專門情報  
(Specialized Information)에 이르기 까지 매  
우 다양하다 [고영만, 1993].

본고에서 언급하는 專門情報 檢索시스템의  
「전문」이라는 의미는 「Full Text(全文)」의 개  
념이 아니고, 「Specialized(專門)」의 개념으로  
사용한다. 그러므로 專門情報 檢索시스템이란  
각 분야마다 세분화되고, 전문적인 主題를 다  
루고 있는 情報檢索시스템을 말한다. 오늘날은  
학문의 特殊化 및 自律化가 이루어 지고 있으  
며, 따라서 情報要求에 있어서도 個別化, 多樣

化가 되어 研究者나 利用者는 專門情報를 要求  
하는 경우가 많아졌다. 그래서 法律情報, 化學  
情報, 女性學情報, 醫學情報 등 다양한 專門情  
報를 다루는 檢索시스템이 늘어가고 있는 실정  
이다. 이러한 경향은 情報檢索시스템의 設計段  
階에도 파급되어 情報檢索시스템의 데이터 項  
目, 敘述言語, 敘述形態 등 여러 부분에 걸쳐  
多様化가 이루어 지고 있다.

이런 다양한 專門情報 檢索시스템의 效率性  
提高를 위해서는 색인과정은 필수적이며, 이때  
적절한 用語統制가 있어야만 한다. 시소러스는  
統制語彙集 性格을 띠고 있는데, 이는 利用者  
들이 디스크립터로써 文獻과 情報請求의 主題  
를 세분하여 索引에 도움이 되도록, 엮어진 일  
종의 표준참고사전의 개념이다 [최성진,  
1976]. 情報檢索시스템에 있어 시소러스는 크  
게 두가지 부분에서 이용가능한 道具的 機能을  
하고 있다.

첫째, 索引段階에서 적절한 索引語의 選擇

과 索引語의 統制에 利用되며, 둘째, 실제 검색과정에서는 효율적인 探索用語의 選定(質疑語 選定)이나 用語間의 相互參照 등에 필요하다.

전세계적으로 500여종의 시소러스가 여러 분야별로 존재하고 있으나, 현재 國內의 實情은 前者의 1/100수준인 4개분야 5개의 시소러스가 있을 뿐이다. 國內의 시소러스 構築은 1980년대 초반에서야 비로소 이루어졌기 때문에 그 수준에 있어서도 미흡하다.

현재 國內의 데이터베이스 數는 1993년 12월 말 기준으로 700여개에 이른다. 이 중에서 모든 데이터베이스가 檢索시스템의 機能을 갖고 있는 것은 아니지만, 대부분의 데이터베이스는 檢索機能을 가지고 있다. 특히 人物情報, 言論情報, 教育情報, 法律情報, 科學技術情報, 醫學情報, 規格情報, 特許情報, 貿易情報 등 다양한 專門情報를 취급하고 있는 데이터베이스의 數는 420여개에 이른다 [한국데이터베이스진흥센터, 1994]. 이와같이 專門情報 보급이 普遍化되고 있는 추세이고, 專門情報 檢索시스템의 效率성과 信頼度 향상을 위해 해당 主題 分野別 專門시소러스 개발은 중요한 의미를 가진다.

본고에서는 專門情報 檢索시스템의 檢索效率를 높일 수 있는 방안에 하나로서 해당주제분야별 專門情報의 시소러스개발을 제안하고, 이를 통해 利用者 중심의 專門情報 檢索시스템을 구축하는데 도움을 줄 수 있는 方法을 모색하고자 한다.

## II. 시소러스의 理論的 考察

利用者が 원하는 情報과 精確하게 연결시키는 기능을 하는 索引語는 單語, 意味構造 등의 統制 有無에 따라 統制方式과 非統制方式으로 나뉘어 진다. 統制方式은 다시 두가지로 分될 수가 있다. 分類表, 코드집과 같은 코드 統制가 있으며, 또한 시소러스, 키워드집 등이 포함되는 用語統制가 있다. 살펴본 바와 같이 시소러스는 統制方式에 의한 索引語에 속한다.

시소러스는 그 사용과정에 있어 다양한 機能을 가진다. 앞에서 언급했듯이 索引段階에서 시소러스의 기능은 적합한 索引語를 選定하고 또한 統制를 위한 補助道具의 역할을 하며, 情報檢索의 段階에서는 시스템내에서의 檢索效率性 提高를 위한 探索用語의 選定, 用語 相互間의 關聯性을 밝혀 檢索의 正確率을 높이는 등

의 機能을 한다. 또한 시소러스는 同義性, 多義性을 統制하여 색인자와 검색자가 일치하도록 하여 關聯語를 연계시킴으로써 主題檢索을 전반적으로 해 낼 수 있도록 하는 기능을 가진다 [Lancaster, 1986].

시소러스의 相互 關係規定을 살펴보면 크게 3가지로 나눌 수 있다. 첫째, 같은 개념이나 유사동의어를 나타내는 同等關係, 둘째, 개념의 종속관계나 부분.vs.전체의 관계를 나타내는 從屬關係, 셋째, 동의개념도 아니고 종속적인 關聯性은 없지만 그 용어간의 關聯性이 있는

는 關聯關係 등으로 구분된다. 또한, 시소러스의 構成은 디스크립터(descriptor)와 비디스크립터(non-descriptor)로 구분된다. 디스크립터는 索引段階나 情報檢索의 과정에서 실제 사용되는 우선키워드의 역할을 하는 것이며, 비디스크립터는 索引段階나 情報檢索段階에서 실제 사용되는 우선키워드의 의미는 아니지만 비디스크립터는 해당하는 디스크립터를 지정해 줌으로써 檢索의 接近點을 제공해 주는 補助役割을 한다.

이러한 기능을 하는 시소러스는 情報檢索시스템에 있어 중요한 의미를 가진다. 특히 앞에서 설명한 後者의 機能과 밀접한 관계가 있는데, 情報檢索시스템에서 探索語나 質疑語 構成에 있어 시소러스는 固요한 道具이다. 또한 컴퓨터에 의한 自動索引의 과정에서도 지식베이스구조를 형성하기 위해서 꼭 필요한 요소이기도 하다. 앞으로 知識의 推論機能을 갖춘 전문가시스템이 발전된다면 시소러스의 중요성은 더해 갈 것이다.

살펴본 바와 같이 시소러스는 情報檢索시스템에서 필수적인 요소이며, 自動索引시스템, 利用者의 探索語나 質疑語 形成에 있어 유용한 道具일 뿐만아니라 적절한 디스크립터로 바꾸어 주어 檢索의 效率性을 높인데 중요한 역할을 하는 것이다.

## III. 國內의 시소러스 現況과 問題點

각 主題別 領域들이 專門情報로 세부화되고 있는 상황에서 專門情報를 효율적으로 檢索하기 위해서는 시소러스의 개발은 필수적이다. 그러나 國內의 시소러스분야는 초보단계에 머무르고 있는 실정이다. 그러므로 施行錯誤의 과정을 거치기 전에 좀 더 精確한 시소러스의 개발에 관한 概念定立을 해 놓을 필요가 있다.

國內의 시소러스는 4개분야 5개의 시소러스가 개발되어 있다. 한편으로 정부나 기업체에

서 自動索引을 위한 시소러스구축이 계획, 추진중에 있기도 하다.

國內의 시소러스의 발전을 살펴보면 다음과 같다.

초창기 시소러스에 해당하는 것으로 韓國教育開發院에서 개발한 「KEDI교육 Thesaurus」(1981年)가 있다. 이는 美國의 「Thesaurus of ERIC descriptors」를 번역하여 편찬한 수준이다.

農業分野의 시소러스인 「농업경제문헌검색어집」(1985年)이 있다. 이는 韓國農村經濟研究院에서 개발한 것이다. 또한, 科學技術分野의 시소러스로는 시스템공학연구소(SERI)에서 개발한 것으로 「과학기술용어시소러스」(1992年)가 있다.

그리고, 國內에서 가장 활발하게 시소러스개발을 하고 있는 분야는 言論分野이다. 韓國言論研究院에서 만든 「신문기사종합시소러스」(1993年)를 제작하였다. 이는 언론종합정보은행(KINDS) 온라인 서비스의 효율적인 索引과 檢索을 위해 만든 것으로, 주로 日本의 유니치신문사의 뉴스시소러스를 참고한 것이다. 또한 중앙일보사에서 개발한 「JOINS시소러스」(1991年)가 있다. 이는 시소러스파일이 5만여개의 디스크립터와 비디스크립터가 있다. 특히, 「JOINS시소러스」는 지금까지의 시소러스가 索引의 초보단계로 책자형태로만 이용되어 오던 것을 情報檢索시스템의 下部시스템인 索引시스템에 이용했다는 점에서 큰 의미를 지닌다.

위의 소개된 시소러스중에서 言論分野의 시소러스는 종합적으로 망라적인 主題內容을 담고 있다고는 하지만, 이들 시소러스들은 言論媒體에서 많이 언급되는 用語中心으로 구성되어 있을 뿐이다. 즉, 실질적으로 종합시소러스로서의 기능은 다 할 수가 없다. 세상의 존재하는 知識을 體系化하여 하나의 틀 속에 집어넣는다는 것은 어쩌면 불가능할 수도 있다. 또한 오늘날은 專門情報가 다양하게 제공되고, 요구하고 있기 시기이므로, 종합시소러스의 개발은 한계가 있기 마련이다.

또한 위에 소개된 대부분의 시소러스는 책자형태의 統制語彙集 수준에 머물고 있다는 것이다. 이는 앞으로 발전하게 될 전문가시스템이나 自動索引에 의한 情報檢索에는 수준에 미치지 못하고 있다.

#### IV. 專門情報 檢索시스템과 시소러스開發을 위한 提言

세부화되고 전문적인 주제를 담고 있는 專門情報 檢索시스템은 앞으로 계속적인 증가를 보일 것이다. 그래서, 각기 세분화된 분야별 專門情報의 體系化나 利用의 效率性 증대를 위해 專門情報 檢索시스템에 적합한 해당 主題領域別 시소러스의 개발이 시급하다.

앞에서도 지적했듯이 國內의 시소러스 수준은 단순한 索引言語의 參考道具에 불과한 것이기 때문에 초보수준의 探索者나 情報檢索시스템에 접근기회가 적었던 利用者에게는 利用의 便利性이나 情報檢索의 效率性에는 별다른 도움을 주지는 못하고 있다.

축약된 主題領域을 가진 專門情報 檢索시스템도 그 분야별로 상호간에는 시스템 구조적인 측면에서나 검색절차의 유사성은 최소한 가지고 있다. 이는 상호간에 主題內容에 있어 互換性을 가지고 있다는 의미와도 통한다. 그러한 專門情報 檢索시스템의 下部시스템인 索引시스템을 완벽하게 互換性을 가질 수 있는 방법의 하나가 바로 해당분야의 專門情報 檢索시스템의 索引言語의 統一, 특히 統制語에 의한 방식을 가지고 있을 경우는 더욱 절실히 요구되는 부분이다.

또한, 기존의 國內 시소러스는 계속적인 갱신(updating)과정이 없어, 급속하게 발전하게 있는 현대사회에서 나타나는 새로운 학문분야, 新造語 등의 用語들은 재빨리 수용을 하지 못하기 때문에 信賴度가 낮게 된다. 이를 해결하기 위해서는 지속적으로 情報를 수용하고 갱신하는 과정에 적극성을 가지고 있어야 할 것이다. 물론 재정적인 측면의 지원을 간과해서도 안 될 것이다.

情報通信의 발전과 國際化 추세에 의해 다른 국가들의 情報를 通信網, 人工衛星 등을 통해 쉽게 접근할 수 있는 기회가 늘어났다. 이런 國際化 상황속에서 각 국가의 言語의 障壁을 극복하는 것이 최선의 情報를 수용할 수 있는 지름길이 될 것이다. 그래서 현재의 우리말로 구성된 시소러스만으로는 해결하기 힘들다. 그래서, 전세계적으로 널리 통용되고 있는 國際語(英語, 中國語, 獨語 등)를 함께 사용할 수 있는 시소러스를 구축하여야 한다. 하지만, 國際語는 서로 다른 言語構造의 特殊性을 가지고 있기 때문에 相互 國際語를 연계시킬 수 있는 研究도 반드시 선행되어야 한다.

## V. 結論

세분화된 專門情報가 늘어감에 따라, 利用者 중심의 專門情報 檢索시스템에도 관심을 가질 필요가 있다. 利用者 중심의 檢索시스템이란 主題內容의 信賴度, 시스템의 速度, 索引시스템 구축 등 다양한 측면에 걸쳐 利用者에게 보다 편리하게 檢索을 할 수 있게 하고, 檢索效率을 높여 利用者에게 檢索結果에 대한 滿足을 줄 수 있는 것이라야 한다. 이런 과정에 중요한 道具的인 役割을 하는 것이 시소러스이다.

시소러스란 統制語彙集으로써 索引段階에서나 檢索過程, 自動索引過程에는 절대적인 기능을 하는 것임에도 불구하고, 시소러스에 관한

研究와 開發, 投資의 측면에 있어 소홀하게 다루어져 온 것이 사실이다.

본고에서는 이상에서 살펴보았듯이, 情報化社會에서 專門情報 檢索시스템에서 요구되는 시소러스의 필요성에 관하여 살펴보았다.

결론적으로, 專門情報 檢索시스템과 관련시켜 시소러스의 개발방향을 제시하고자 한다.

첫째, 專門情報 檢索시스템의 구축과정에서 정형화된 해당 주제 분야별 시소러스개발을 하나이상 할 필요성이 있다. 왜냐하면, 다양하게 나타나는 專門情報 檢索시스템은 해당 分野別로 특정한 主題를 담고 있으므로, 종합적인 主題를 포함한 시소러스개발은 特定性이 낮아 檢索效率에 도움을 주지는 못할 것이다. 예를 들면, 종합시소러스를 이용한다면, 특정한 主題分野의 專門시소러스보다 再現率을 높일 수 있을지 모르나, 正確率에 있어서는 상당히 저조하게 나타나기 때문이다. 하나의 시스템속에서 세상의 知識을 대표하는 用語들을 모두 포함시킨다는 것은 이상에 불과한 것이다. 반대로 세분화된 主題領域을 가진 專門情報 檢索시스템을 제작할 때마다 각기 다른 關係構造를 시소러스를 구축한다는 것은 人力이나 經濟的인 측면에서 非經濟的이며, 利用의 측면에서도 利用者에게 혼란을 주어 檢索의 效率性 저하를 가져올 수가 있다.

둘째로, 專門情報 檢索시스템에 적합한 시소러스의 개발을 위해서 해당주제분야별 선정용어와 형태의 통일, 디스크립터의 일치 등과 같은 語彙統制의 方法의 根據나 指針을 마련할 필요성이 절실히 요구된다. 이것을 효과적으로 이루기 위해서는 主題專門家, 索引作成者, 言語學者 등으로 구성되는 常任 專門家 그룹을 형성할 필요가 있다.

셋째로, 전국적인 차원에서 시소러스를 관리할 수 있는 클리어링하우스를 마련해야 한다. 하지만, 반드시 국가기관의 성격을 띤 기관일 필요는 없다. 클리어링하우스를 통해 專門情報 檢索시스템의 專門시소러스를 관리하며, 이의 지속적인 갱신(updating)이나 수정, 변경을 할 수 있는 단일 창구를 마련해야 할 필요가 있다. 이를 통해 專門시소러스의 標準이나 指針을 제공하는 기관으로서의 기능도 할 수 있는 것이다. 또한 중복적인 시소러스개발 활동을 미연에 막아서 人力, 經費의 낭비를 막아야 할 것이다.

넷째로, 국제적인 情報네트워크가 형성되고 있음을 직시하고, 우리말 시소러스의 완전한 구축에만 매달리는 것이 아니라, 外國의 專門情報와의 相互利用, 互換이 가능할 수 있도록 國際語로 이루어진 전문시소러스의 개발에도 주안점을 두어야 할 것이다.

끝으로, 시소러스를 研究·敎育하는 文獻情報學 分野에서도 이론적인 시소러스의 考察에서 벗어나서, 각 학교별로 專門情報의 分野別로 主題專門 시소러스를 한 영역씩 담당하여, 지속적으로 개발, 수정, 보완하면서 전문적인 시소러스를 구축에 심혈을 기울여야 한다. 이러한 과정에는 반드시 主題專門家, 실무현장의 專門家와도 연계가 필요한 부분이다. 이를 통하여 자연스럽게 산·학 공동협조체제를 이루는데 도움을 줄 수 있게 된다.

## < 參考文獻 >

- 고영만, 1993. "전기통신의 자유화가 전문정보 서비스에 미치는 영향에 관한 연구". 한국문헌정보학회지 24집 : p.268.
- 최성진, 1976. 정보학원론. 서울 : p.274.
- 한국데이터베이스진흥센터, 1994. 알기쉬운 한국의 데이터베이스 편람. 서울 : pp.7-8.
- ancaster, F. W., 1986. Vocabulary control for information retrieval, Alington : Information Resources Press.