

통합의학언어 시스템(UMLS)의 구성 및 적용에 대한 고찰

A Review of Structure and Application of Unified Medical Language System(UMLS)

김혜선*
Hye-Sun Kim

차례

- | | |
|-------------|---------------|
| 1. 서론 | 4. UMLS의 적용사례 |
| 2. UMLS의 개요 | 5. 결론 |
| 3. UMLS의 구성 | • 참고문헌 |

초록

효과적인 정보검색 도구로 활용되는 시소러스, 분류표 등의 통제어휘집들은 동일한 개념이나 의미를 표현하는데 서로 다른 어휘를 사용하고 있다. 이러한 동일 개념에 대한 용어 표현 차이로 인한 정보의 검색 및 통합문제를 해결하기 위하여 미국국립의학도서관은 통합의학언어 시스템(Unified Medical Language System ; 이하 UMLS)을 개발하였다. UMLS 개발은 1986년부터 시작된 장기 프로젝트로서, 2001년판 UMLS는 메타시소러스, 의미망, 전문가 사전의 세 가지 요소로 구성되어 있다. 본고는 UMLS의 개발배경과 구성요소 및 PubMed, NLM Gateway에서의 적용사례에 대해 고찰하였다.

키워드

통합의학언어 시스템, 메타시소러스, 의미망, 전문가 사전, 어휘통제, 시소러스, 통합개념체계, 미국국립의학도서관

* 콘텐츠개발실 연구원(Researcher, Contents Development Dept., KISTI)

ABSTRACT

Various controlled vocabularies such as thesaurus and classification used for effective information retrieval contain different terms in expressing the same concept or meaning. National Library of Medicine has developed the Unified Medical Language System(UMLS) to solve the problems of information retrieval and integration resulted from the difference of concepts between different sources. The UMLS development was initiated in 1982 as a long-term project, and the 2001 edition of the UMLS consists of three parts : Metathesaurus, Semantic Network, and SPECIALIST Lexicon. This paper reviews background and structure of the UMLS including applications in PubMed, NLM Gateway.

KEYWORDS

Unified Medical Language System, UMLS, Metathesaurus, Semantic Network, SPECIALIST Lexicon, vocabulary control, thesaurus, NLM, PubMed, NLM Gateway

1. 서 론

정보량의 증가와 다양한 정보검색 시스템의 출현은 정보에 대한 접근성을 확대시킨 반면 적합한 정보의 검색을 더욱 어렵게 하고 있다. 따라서 원하는 주제, 개념에 효과적으로 접근할 수 있도록 지원하는 통제어휘집이 탑재된 정보검색 시스템이 개발, 활용되고 있다. 통제어휘집은 동의어, 동음이의어 등을 통제하여 하나의 개념을 하나의 고유한 용어로 색인, 검색하도록 하는데, 분류표, 주제명표목표, 시소러스 등이 대표적 사례가 된다.

과거에는 시소러스, 분류표, 용어집 등이 특정 데이터베이스에 한정되어 사용되었지만, 복수 데이터베이스에 대한 통합 검색이 가능해지면서 시소러스와 같은 주제접근 도구의 분산에 따른 호환성 문제가 대두되었다. 이를 해결하기 위하여 개

발된 대표적인 것으로 미국국립의학도서관(National Library of Medicine : 이하 NLM)의 통합의학언어 시스템(Unified Medical Language System : 이하 UMLS)을 들 수 있다. UMLS는 60여개 의학관련 시소러스, 분류표, 용어사전에 수록된 개념을 연계한 통합개념체계로서, 기본 엔트리로 단어, 용어가 아닌 개념을 이용하고 있다. UMLS와 유사한 개발사례로는 미국 프린스턴대학교의 WordNet, 일본 전자화사전연구소의 EDR 전자사전이 있다.

본고는 UMLS의 구성, 적용방법 등에 대해서 고찰함으로써, 이종의 데이터베이스에 대한 정보 통합검색 수단으로 활용되는 통합개념체계에 대한 이해를 돕고 향후 국내에서의 유사 연구에 도움이 되 고자 한다.

2. UMLS의 개요

2.1 개발배경

생물의학 분야의 다양한 정보검색 시스템은 정보검색의 효율향상을 위하여 시스템마다 고유한 통제어휘집을 활용하고 있다. 이런 통제어휘집들은 동일한 개념을 표현하는데 서로 다른 용어를 채택하고 있으므로, 이용자는 각 시스템에서 사용되는 용어로 검색해야만 필요한 정보를 누락없이 찾을 수 있다. 이러한 전자정보원간 개념의 표현 차이로 인해 발생하는 문제를 해결하기 위하여 NLM은 1986년부터 UMLS 개발에 착수하였다.

UMLS 프로젝트는 생물의학 분야의 다양한 정보 시스템간의 정보검색과 통합을 촉진시키기 위한 것으로, 지적중개자 역할을 하는 Knowledge Source를 개발하였다. Knowledge Source의 목적은 각종 응용프로그램에 적용되어 이용자와 정보 시스템이 사용하는 개념의 표현 차이로 인한 문제를 해결하고, 서지 데이터베이스, 사실 데이터베이스, 환자기록 시스템 등 이종의 시스템간 정보를 연계시키는데 활용되고자 한다.

UMLS는 1990년에 초판이 발행된 이후 매년 주기적 갱신을 통하여 2001년 현재 제 12판이 발행되었으며, 전세계 1,000여 개 이상의 개인과 기관에서 무료로 사용되고 있다.

2.2 이용 신청방법

웹, CD-ROM 형태의 UMLS를 무료로 사용하기 위해서는 "UMLS 제품 이용을 위한 2001년도 라이선스 협정"¹⁾을 맺어야 한다. 협정서는 저작권 보호를 위해서 이용자가 준수해야 할 협정내용과 이용신청서, 소스 어휘집 제공기관 리스트로 구성되어 있다.

협정내용의 특이사항은 첫째, 이용권한을 부여받은 후 연말 30일 이전에 UMLS의 유용성에 관한 간략한 보고서를 제출해야 한다는 것으로, 이런 이용자 피드백을 기반으로 UMLS의 지속적인 갱신이 이루어진다. 둘째, 일부 어휘의 경우 NLM 이외의 소스 어휘집 제공자에게서 별도로 이용허가를 받아야 사용가능하며, 이중 일부는 유료일 수 있다. 셋째, 소스 어휘집에 따라 1~3단계 범주의 이용제한 수준이 존재한다.

(1) 범주 1

소스 어휘를 다른 언어로 번역하거나 이차저작물로 생산하는 것을 금지한다.

(2) 범주 2

범주 1의 제한과 함께 소스 어휘를 포함한 레코드나 정보를 생산하는 응용 프로그램에서 이용할 수 없다.

(3) 범주 3

연구, 제품개발, 통계분석을 위해서만 어휘를 사용할 수 있고 사용자 사이트 내부에서만 이용해야 한다.

협정내용을 검토한 후 이용신청 양식을

1) "2001 License Agreement for Use of UMLS Products"에 대한 상세정보는 웹사이트에서 볼 수 있다 (<http://www.nlm.nih.gov/research/umls/license.html>).

작성하여 NLM에 발송하면 UMLS Knowledge Source Server에 대한 이용권한을 부여받을 수 있다.

3. UMLS의 구성

UMLS Knowledge Source는 메타시소러스(Metathesaurus), 의미망(Semantic Network), 전문가 사전(SPECIALIST Lexicon)으로 구성되며, 이 세 가지 요소가 상호 유기적으로 운영됨으로써 UMLS의 효과가 극대화된다.

(1) 메타시소러스

생물학적 개념과 개념의 다양한 명칭, 개념간의 관계에 대한 정보를 포함한다.

(2) 의미망

메타시소러스 개념의 의미유형과 의미 관계에 대한 정보를 포함한다.

(3) 전문가 사전

메타시소러스 개념명에 속하는 주요 용어들의 구문론적 정보를 포함한다.

3.1 메타시소러스(Metathesaurus)

메타시소러스는 UMLS의 중심 어휘요소로서 생물학분야의 통제어휘집, 분류표 등에서 한번 이상 나타나는 개념에 대한 정보 데이터베이스이다. 개념의 의미, 속성, 계층적 연관성, 개념과 용어간의 관계, 동의어 체계를 포괄하며 80만개의 개념과 190만개의 개념명을 수록하고 있다.

메타시소러스는 NLM 주제명표목표인 MeSH를 포함하여 NCI thesaurus, National Drug Data File, International Classification

of Disease 등 60여개 이상의 생물학 분야 시소러스, 분류표 등의 통제어휘집을 기초로 하여 구성되었다. 2001년판 메타시소러스는 영어 외에 불어, 핀란드어, 독일어, 이탈리아어, 포르투갈어, 러시아어, 스페인어판 MeSH를 포함하고 있다.

3.1.1 개념 표현

메타시소러스는 개념이나 의미에 따라 조직된다. 즉, 메타시소러스는 식별자(identifier)를 이용하여 한 개념(concept)에 대한 서로 다른 용어(terms), 문자열(strings)을 연계하여 같은 개념으로 인식하고 있다.

(1) 개념 식별자(unique concept identifier; 이하 CUI)

메타시소러스의 각 개념이나 의미는 고유한 CUI를 가진다. CUI는 개념에 관한 모든 정보를 연계하고, 속성을 규명하는데 이용된다.

(2) 공통 용어 식별자(common term identifier; 이하 LUI)

용어는 어휘가 변형된 모든 문자열의 그룹으로, 각 문자열은 LUI에 의해 변형된 모든 어휘와 연결이 가능하다.

(3) 문자열 식별자(unique string identifier; 이하 SUI)

문자열은 실제 자료에 나타나는 형태로서 철자변형, 단복수 변형 및 언어변형에 따라서 고유한 SUI가 부여된다.

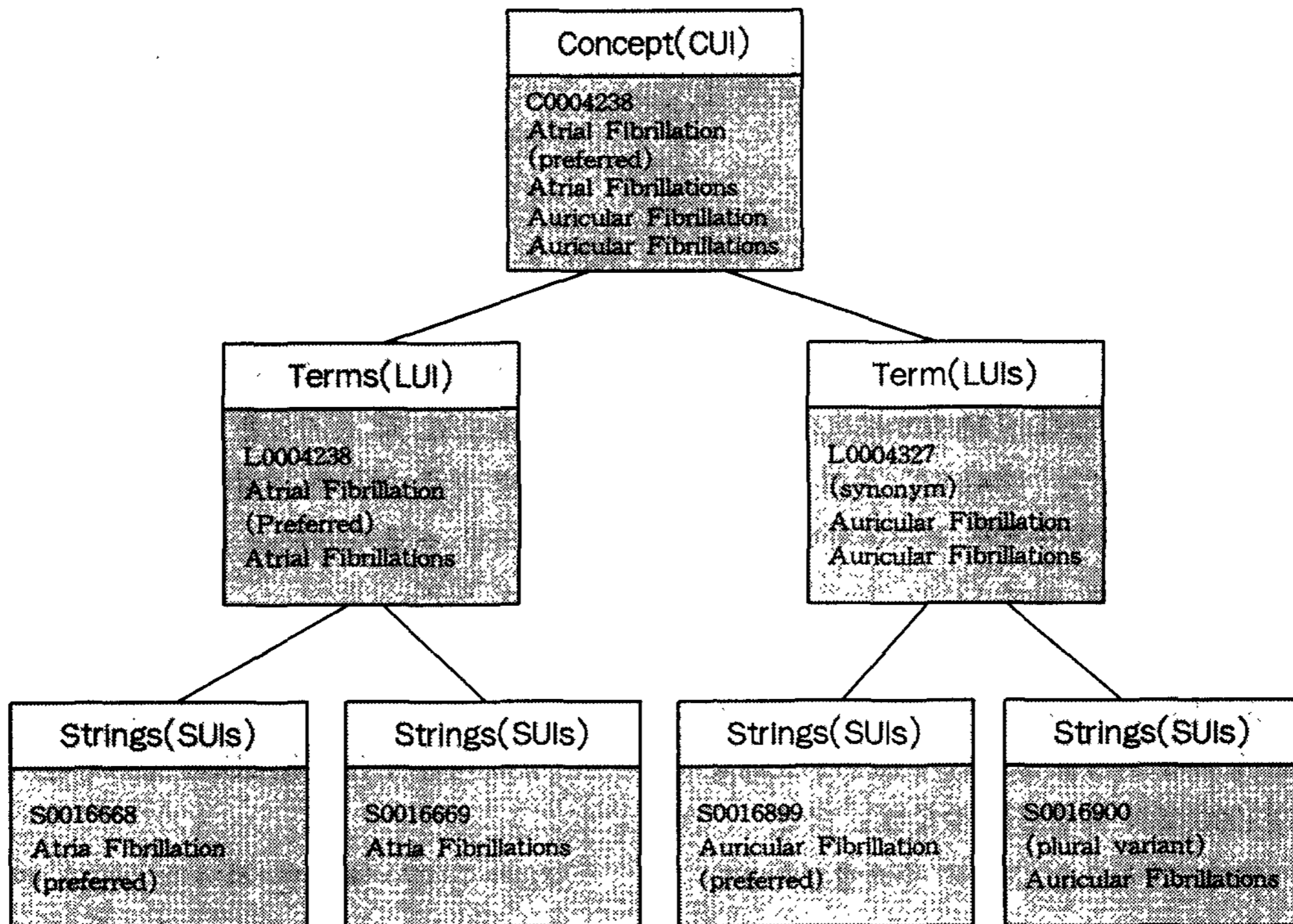
SUI와 LUI는 최소한 하나의 CUI에 연결되며, 동일한 의미를 지닌 다양한 용어들은 동일한 CUI에 의해 서로 연결된다.

<그림 1>을 보면 문자열 Atrial Fibrillation (S0016668)과 이의 복수형 변형 문자열인 Atrial Fibrillations(S0016669)은 각각 별도의 SUI를 지니지만, 어휘가 변형되었을 뿐 동일한 용어이므로 공통용어 식별자(LUI)에 연결된다. 한편 Atrial Fibrillation (L0004238)과 Auricular Fibrillation (L0004327)은 “심방세동”이라는 동일한 의미를 지니는 용어이므로 동일한 개념식별자 C0004238로 연결된다. 동일 개념에 대한 우선어(preferred)는 임의로 정해진 소스 어휘집의 순서에 의해서 결정되며, 어휘집 순서는 이용기관의 필요에 따라서 변경할 수 있다.

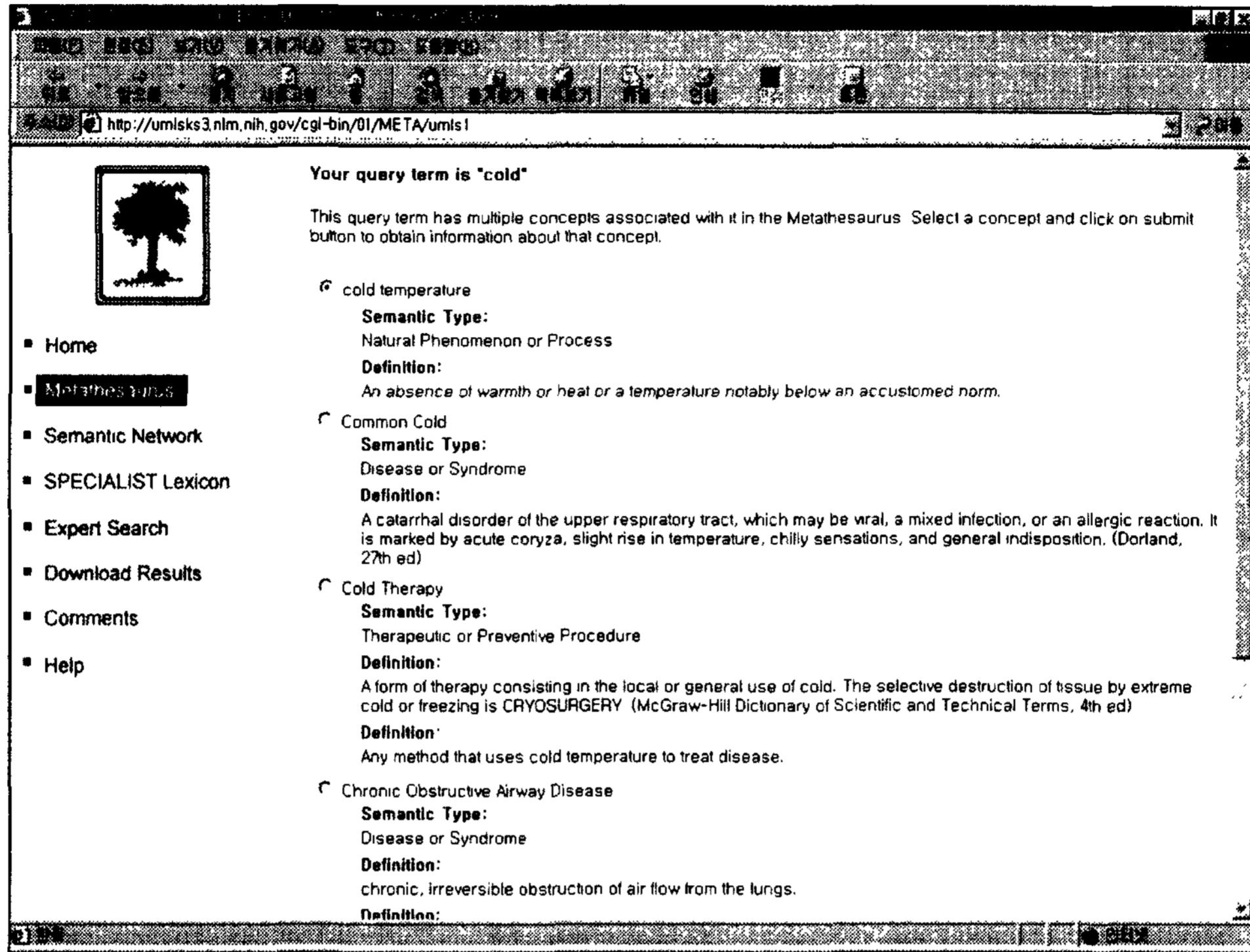
시소러스 등 일반 통제어휘집은 용어를

디스크립터와 비디스크립터로 구분한 후 디스크립터로 색인, 검색하도록 통제하고 있다. 반면에 메타시소러스는 특정한 용어를 디스크립터 또는 통제어휘로 지정하지 않고 동일한 의미를 가진 문자열, 용어들을 서로 연계시켜줌으로써, 이용자가 다양한 시스템에서 용어에 구애받지 않고 검색을 실시할 수 있게 해준다.

한편 동일한 문자열이 복수의 의미를 가질 경우 메타시소러스는 다의성 구분을 통하여 별도의 개념을 부여하고 있다. 예를 들어 문자열 cold는 온도, 감기, 만성폐쇄성 폐질환(chronic obstructive lung disease)의 두문자어 COLD 등 최소한 세 가지 다른 개념의 명칭이 될 수 있다. 이 경우에는



<그림 1> 개념에 의한 용어, 문자열 연계



〈그림 2〉 메타시소러스의 다의성 구분

〈그림 2〉와 같이 cold temperature, Common Cold, Cold Therapy 등의 문자열과 이의 의미유형(semantic type), 정의(definition)를 제시해줌으로써, 정확한 개념으로 찾아갈 수 있다.

3.1.2 개념간의 관계

메타시소러스에서 개념간의 관계(relationship)는 일반 시소러스와 마찬가지로 계층, 연관, 동등관계를 가진다. 이러한 관계는 소스 어휘집에서 직접적으로 추출

〈표 1〉 메타시소러스 개념간의 관계

Broader (RB)	상위어 관계
Narrower (RN)	하위어 관계
Other related (RO)	동의어, 상위어, 하위어 이외의 관계
Like (RL)	유사한 두 개념간의 관계. 동의어로서 메타시소러스 향후 개정판에서는 동일한 개념식별자를 갖게 됨
Parent(PAR)	소스 어휘집에서 부모관계
Child(CHD)	소스 어휘집에서 자식관계
Sibling(SIB)	소스 어휘집에서 형제관계
AQ	소스 어휘집에서 개념을 위하여 허용된 한정어
QB	소스 어휘집에서 개념에 의하여 한정될 수 있음

되거나, 소스 어휘집이 다를 경우에는 메타시소러스 편집자에 의해 부여된다. 9가지 유형으로 이루어진 메타시소러스 개념간의 관계는 다음과 같다(〈표 1〉 참조).

3.2 의미망(Semantic Network)

의미망은 의미유형(semantic type)과 의미관계(semantic relationship)를 이용하여 메타시소러스의 각 개념에 대하여 일관된 범주를 부여하고, 개념간의 유용한 관계를 제공하는데 목적이 있다. 의미망은 134개 의미유형과 54개 의미관계로 구성되어 있다.

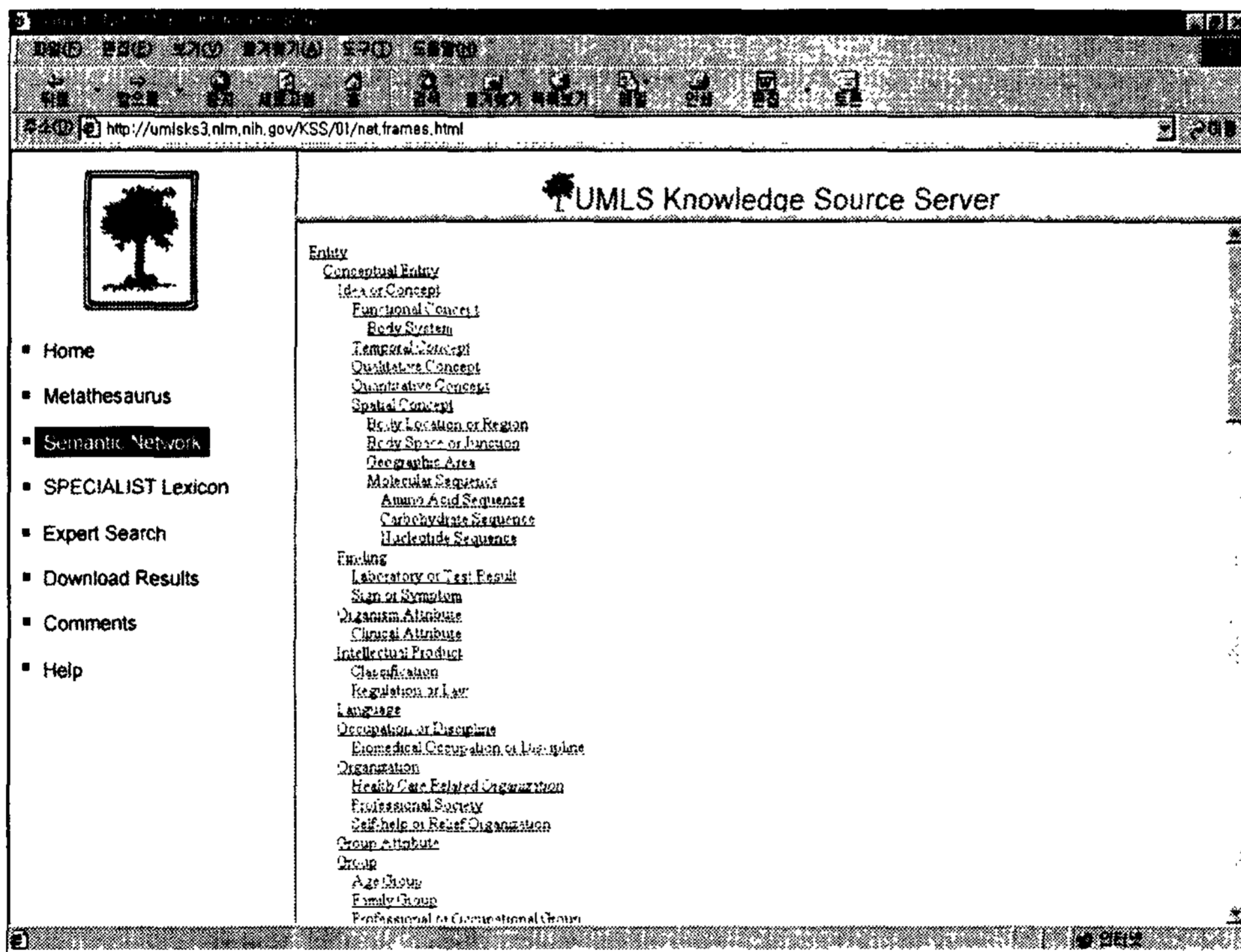
3.2.1 의미유형

계층적 구조로 되어 있는 의미유형은

개념을 크게 개체(Entity)와 사건(Event)으로 구분하고 있다(〈그림 3〉 참조). 개체는 99개 유형으로 물리적 대상, 개념적 실체를 그룹화한 것이며, 사건은 활동(activity), 과정(process), 상태(state)와 관련된 개념들을 그룹화한 것으로 35개 의미유형으로 구성된다. 하나의 개념에는 최소한 하나 이상의 의미유형이 할당된다.

3.2.2 의미관계(Semantic Relationship)

의미관계는 의미유형 사이에 부여되는 관계로, 계층관계(isa)와 비계층관계(associated with)로 구분된다(〈표 2 참조〉). 의미망에서 기본적인 연결은 isa 링크로 이루어지며, 비계층관계는 물질, 공간, 시간, 기능, 개념의 다섯가지 범주로 구분된다.



〈그림 3〉 의미유형의 구조

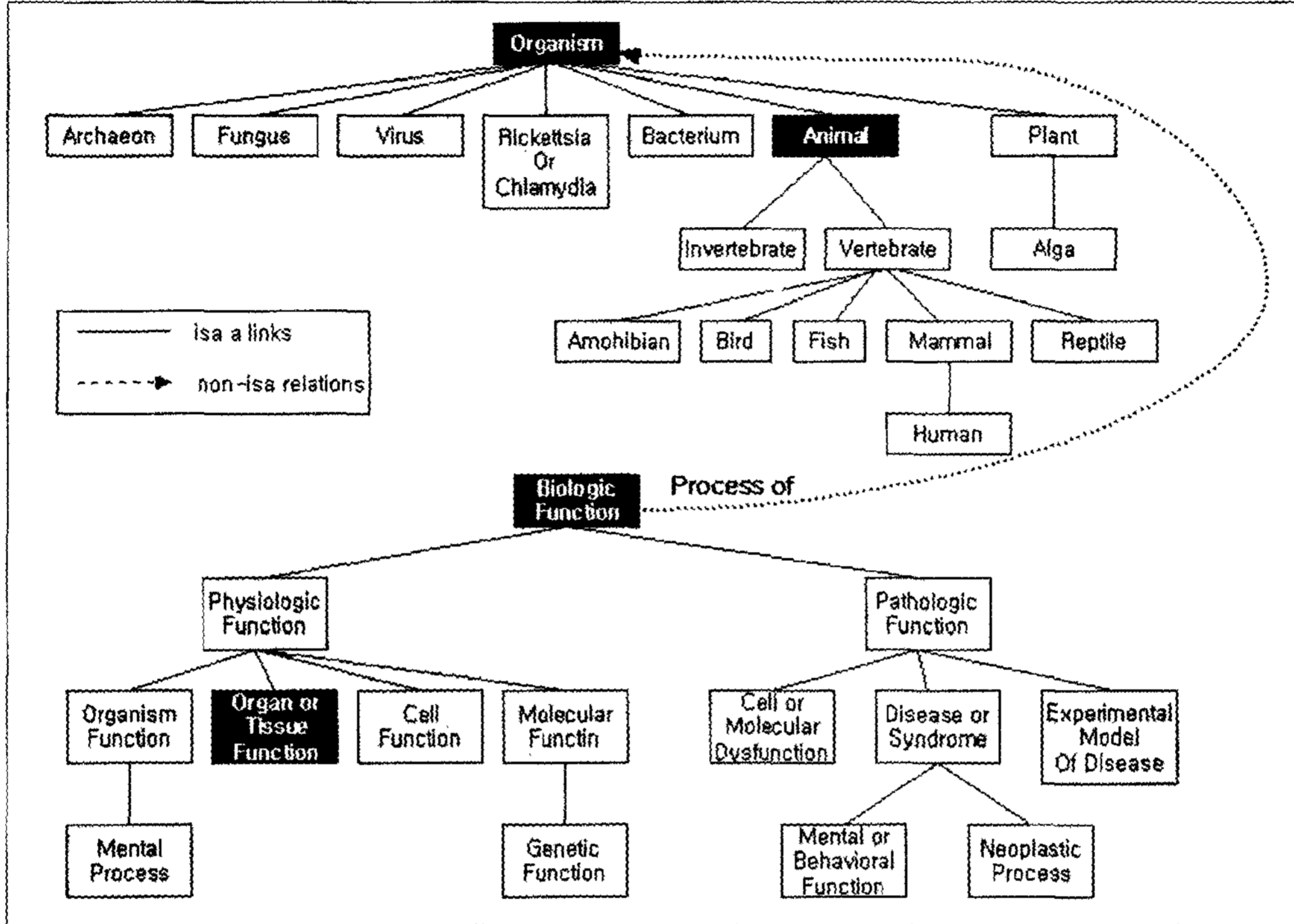
〈표 2〉 의미관계의 구성요소

isa associated with physically related to part of consists of contains connected to interconnects branch of tributary of ingredient of spatially related to location of adjacent of surrounds traverse functionally related to affects manages treats disrupts complicates interacts with prevents brings about produces causes	[associated with] (continued) [functionally related to] (continued) performs carries out exhibits practices occurs in process of uses manifestation of indicates result of temporally related to co-occurs with precedes conceptually related to evaluation of degree of analyze assesses effect of measurement of measures diagnoses property of derivative of developmental form of method of conceptual part of issue in
---	---

일반적으로 관계는 의미망의 상위수준 의미유형에서 명시되고, isa 링크를 통해 그 유형의 모든 자손에게 상속된다. 〈그림 4〉를 보면 의미유형 Organism과 Biologic Function 사이에 비계층관계인 process of 관계가 명시되어 있다. 따라서 isa 링크에 의해서 계층적으로 연결된 Organism의 자손 Animal과 Biologic Function의 자손 Organ or Tissue Function 사이에도 이 process of 관계가 적용된다.

process of 관계는 Organism(유형 1)과

Biologic Function(유형 2) 외에 Individual Behavior(유형 1)와 Social Behavior(유형 2), Natural Phenomenon or Process(유형 1, 2) 의미유형에도 적용된다. 따라서 의미유형 Individual Behavior에 속하는 메타시소러스 개념(812개)과 Social Behavior에 해당하는 개념(1,067개) 사이에는 process of 관계가 존재하지만, 이 의미유형에 해당되는 모든 개념 사이에 반드시 이 관계가 적용되는 것은 아니다.



<그림 4> 의미망의 구성

3.3 전문가 사전(SPECIALIST Lexicon)

전문가 사전은 전문가 자연어 처리시스템에 필요한 모든 사전(lexical) 정보를 제공하기 위하여 개발된 것으로, 생물학적 용어 약 2만 단어뿐 아니라 1만2,000여개의 일반어휘에 관한 정보도 포함한다. 어휘변형 생성 프로그램은 어순, 용어 변형 및 대소문자 변형 등을 자동 관리함으로써, 이용자로 하여금 어휘 변형을 고려하지 않도록 한다.

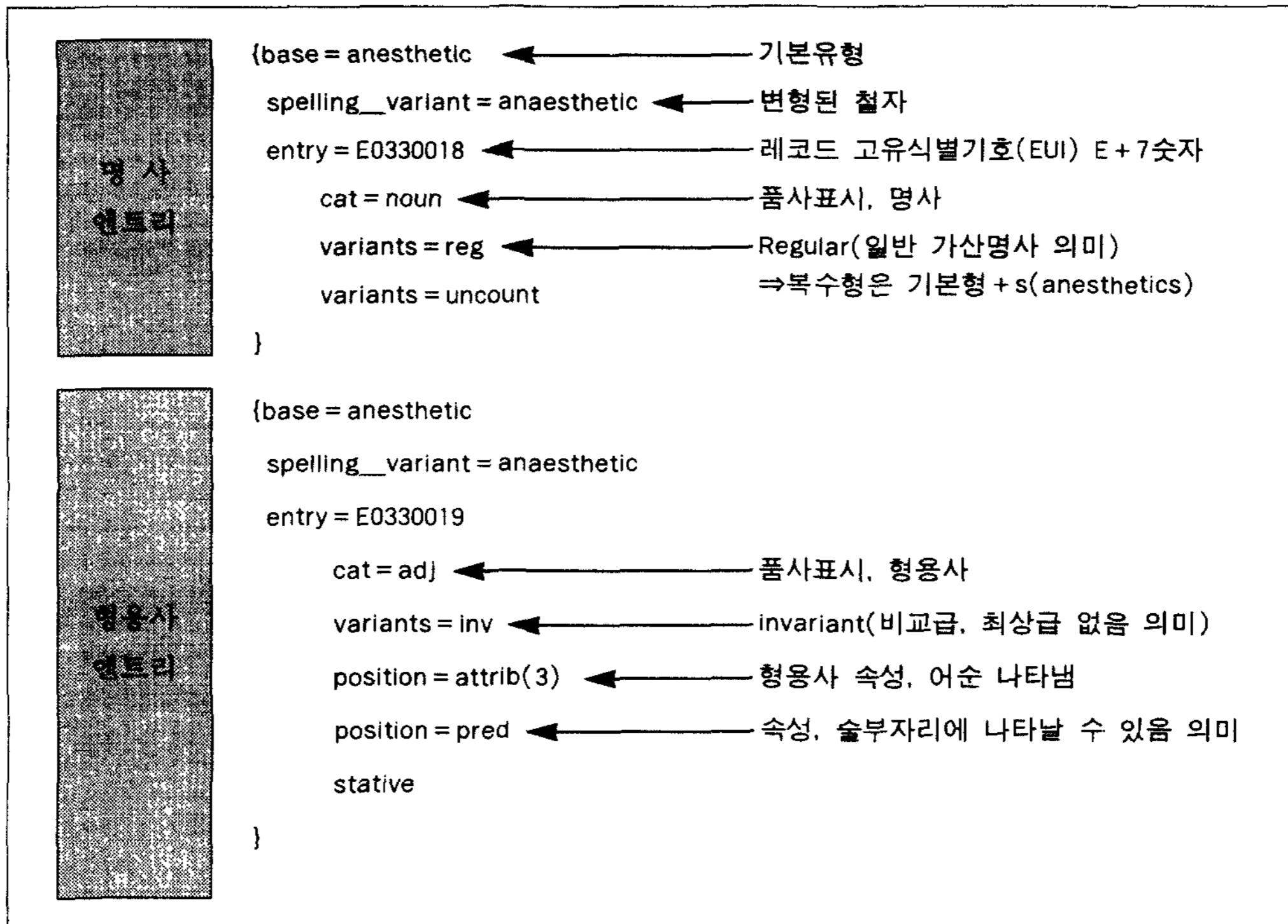
<그림 5>와 같이 사전 레코드는 슬롯(slot)과 필러(filler)²⁾로 구성되며 중괄호({ })로 시작되고 끝난다. 사전 레코드는

어휘의 기본유형, 철자변형, 레코드 고유식별자, 품사, 변형정보, 단어속성 등 어휘의 구문, 어형, 문자소에 대한 종합적인 정보를 포함한다.

전문가 사전은 동일 단어가 복수의 품사를 가질 경우에는 별도의 사전 레코드로 나타낸다. anesthetic은 명사와 형용사로 사용되므로 각 품사에 따라 고유한 사전 레코드를 가지게 된다(<그림 5> 참조).

명사 act와 같이 “행위”와 “법령”의 두 가지 의미를 지니지만, 동일한 철자와 구문 규칙을 가질 경우에는 의미에 따라 엔트리

2) “base=”, “spelling__variant=”, “entry=” 등은 슬롯으로, 이들 각각은 “기본유형(anesthetic)”, “변형된 어형(anaesthetic)”, “레코드 고유기호(E0330018)”를 나타내는 필러를 가진다.



〈그림 5〉 사전 레코드의 구성

를 구분하지 않고, 단일한 사전 엔트리로 나타낸다.

```
{base = Act
  spelling__variant = act
  entry = E0000154
    cat = noun
    variants = reg
}
```

한편 복수 의미를 가진 단어가 의미에 따라서 구문규칙이 다를 경우에도 한 엔트리에 각 특성에 대한 코드를 기입하는 방식으로 나타낸다. 예를 들어, beer는 “맥주”와 “맥주한잔(병)”이란 두 가지 의미를 지니며, 의미에 따라 일반 가산명사(reg)와 불가산명사(uncount)로 사용될 수 있다. 이

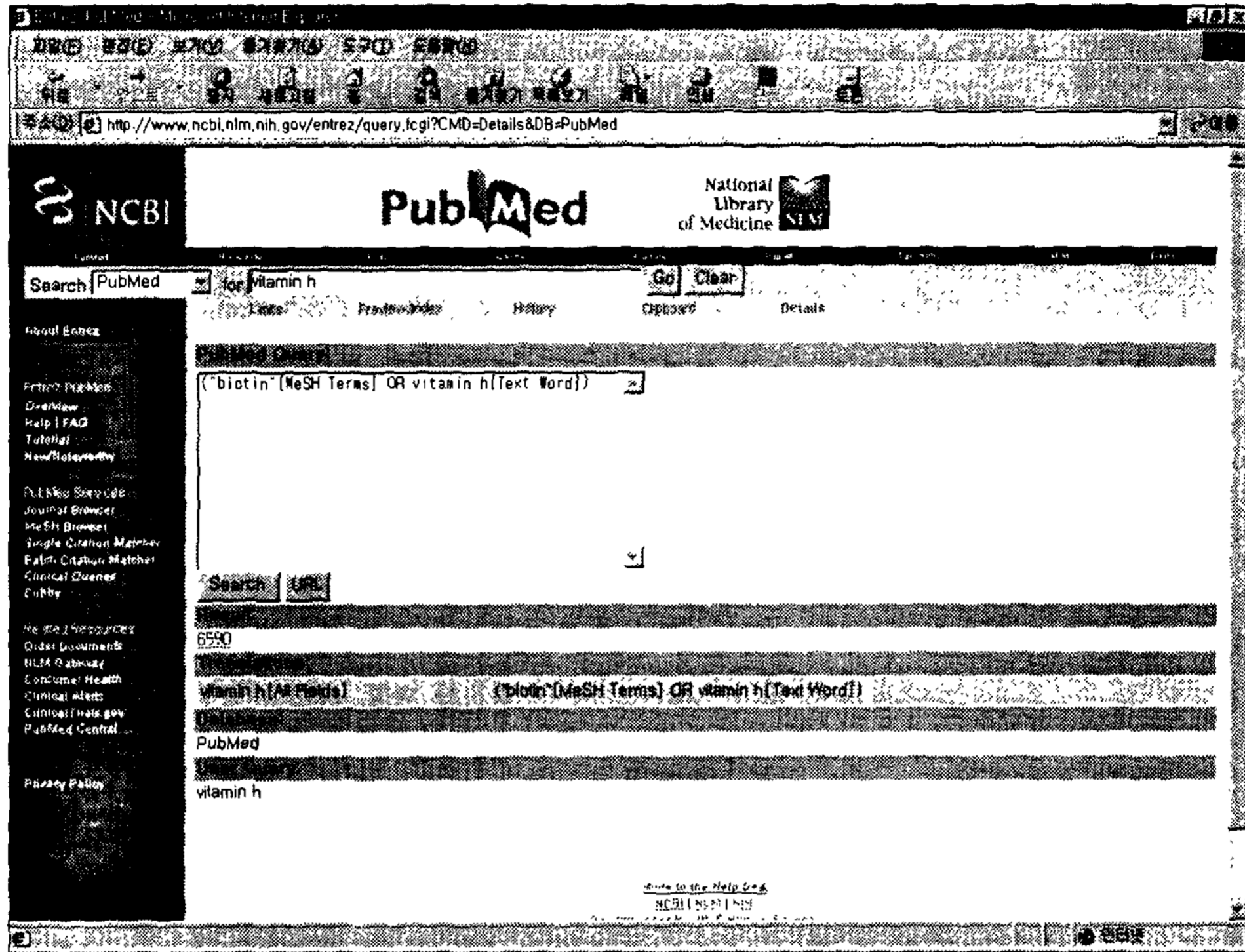
경우 동일한 사전 엔트리에 다음과 같이 “uncount”, “reg”란 특성 코드를 부여한다.

```
{base = beer
  entry = E0012226
    cat = noun
    variants = uncount
    variants = reg
}
```

4. UMLS의 적용사례

4.1 PubMed

UMLS 메타시소러스는 MEDLINE 데이터베이스의 검색성능 향상을 위하여 PubMed에서 적용, 활용되고 있다.



〈그림 6〉 PubMed에서의 UMLS 적용

PubMed는 동의어와 변형어 정보를 지닌 UMLS를 기반으로 한 MeSH 변환테이블에 의해서 일반 검색용어를 통제용어로 변환하여 검색을 실시하게 한다.

예를 들어 PubMed 검색화면에 vitamin h를 입력하면, PubMed는 MeSH 변환테이블에 의하여 질의어 vitamin h의 통제어휘인 biotin과 연계하여 검색을 실시한다. 즉, 검색창에 vitamin h를 입력하면 〈그림 6〉과 같이 "biotin [MeSH Terms] or vitamin h[Text Word]"로 검색 질의문이 자동 변환되어 검색이 수행된다.

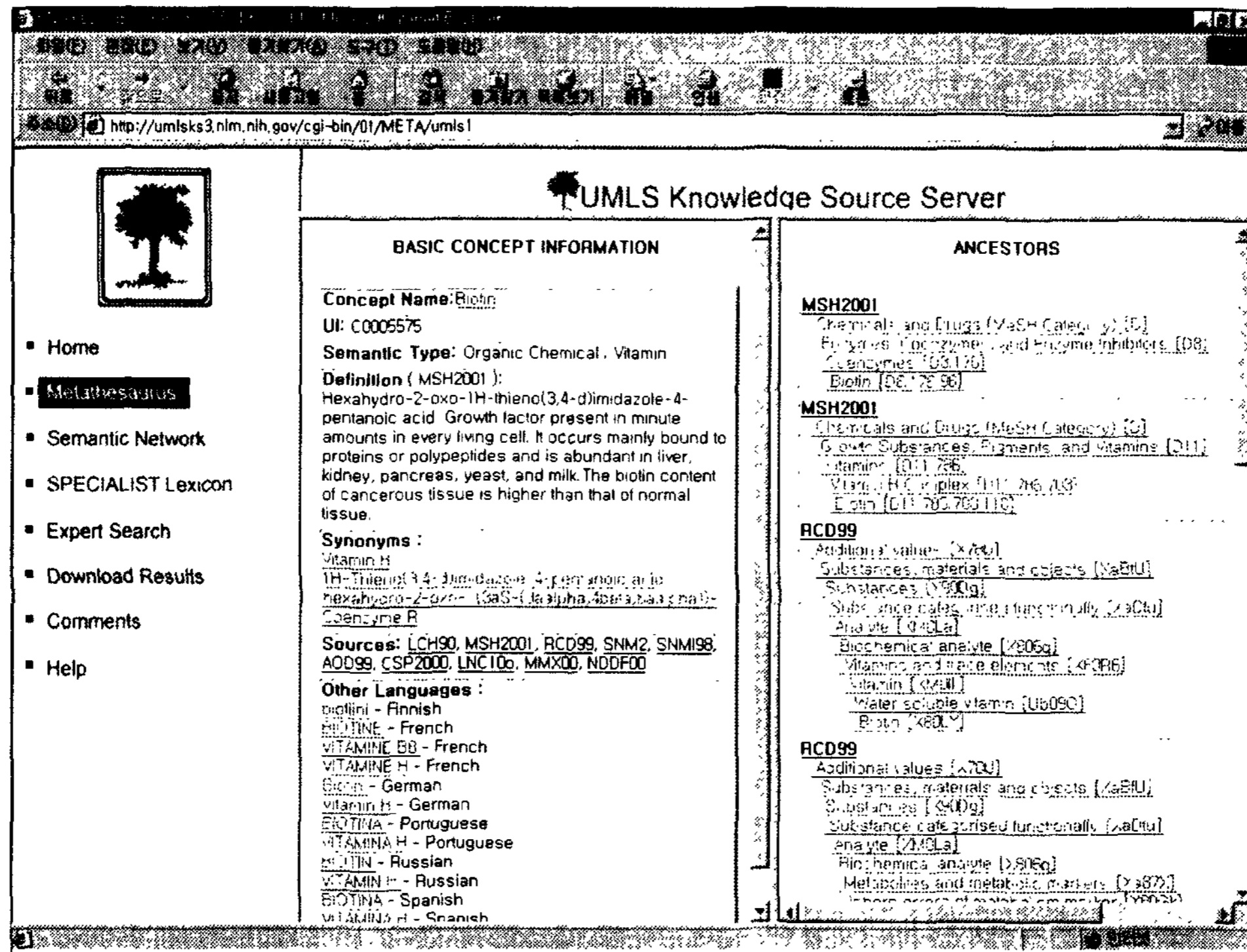
앞서 언급했듯이, 비통제 검색어휘를 통

제어휘로 연계시켜 주는 MeSH 테이블의 기초정보는 UMLS 메타시소러스를 근거로 한다. 〈그림 7〉의 메타시소러스 제공 정보를 보면 개념명 biotin은 vitamin h외에 coenzyme r과 같이 다양한 동의어를 가진다. 따라서 PubMed에서 coenzyme r로 검색하면 vitamin h와 마찬가지로 통제어휘 biotin과 연계되어 검색이 수행될 것이다.

4.2 NLM Gateway

NLM Gateway(<http://gateway.nlm.nih.gov/gw/Cmd>)는 NLM의 다양한 정보 검색 시스템³⁾을 동시에 one-stop으로 검색

3) 검색대상 데이터베이스는 MEDLINE/PubMed, OLDMEDLINE, LOCATORplus, MEDLINEplus, DIRLINE, AIDS Meetings, Health Services Research Meetings, Space Life Sciences Meetings, HSRProj 임.



<그림 7> 메타시소러스 개념 정보

할 수 있도록 단일 인터페이스를 제공하는 시스템이다. Gateway 검색화면에서 사용자가 질의어를 입력하면 서로 다른 특성을 가진 복수 데이터베이스로 질의어가 자동 전달되어 검색이 실시된다. NLM Gateway는 NLM의 다양한 정보에 대한 통합 검색을 제공하는 목적 이외에 용어검색(Find Terms) 기능에 의해 UMLS의 이용을 극대화하고자 개발되었다. 즉, 용어검색기능은 UMLS를 활용하여 질의어와 관련된 MeSH 통제어휘를 제시함으로써 Gateway의 검색효율을 높여준다. 용어검색의 주요 기능을 요약하면 다음과 같다.

(1) MeSH가 아닌 용어가 입력될 경우 메타시소러스의 동의어 등 각종 용어 맵

(map)이 체크되어 관련된 MeSH 용어를 제시한다. 만약 적합한 MeSH 용어가 없을 경우에는 관련된 개념 리스트가 디스플레이 된다.

(2) 상세 검색결과는 용어정의, 관련된 개념정보, MeSH 정보, 부표목 추가기능, MeSH 트리 구조상에서의 용어정보가 나타난다.

(3) 입력된 용어가 메타시소러스에서 cold 처럼 복수의미를 가질 경우에는 MeSH 용어 선택기능을 이용하여 “cold, common cold, cryotherapy, obstructive lung diseases” 등의 문자열에서 가장 적합한 용어를 선택하게 한다.

기타 UMLS는 MEDLINE의 Internet Grateful Med, NLM/AHCPR Large-Scale Vocabulary Test, ClinWeb International, Image Engine, Medical World Search, MedWeaver, Metaphrase 등에서 적용되고 있다.

5. 결 론

UMLS는 동일 개념을 표현하는 데 사용되는 용어의 다양성으로 인한 정보검색의 효율저하 및 통합검색 문제를 해결하기 위하여 NLM이 개발한 통합개념체계로서, 메타시소러스, 의미망, 전문가 사전으로 구성되어 있다. 메타시소러스는 특정 개념에 해당하는 용어, 문자열을 연계하는 UMLS의 중심 어휘요소로서, MeSH를 포함한 60여 개 이상의 통제어휘집에서 추출된 개념들로 구축되었다. 의미망은 개념을 의미유형으로 구분하여 계층 또는 비계층관계로 나타내고 있으며, 전문가 사전은 용어의 기본 유형, 변형어, 품사, 속성 등에 대한 종합적 정보를 가진 사전역할을 한다. 1986년 이래 지속되어온 UMLS 개발 프로젝트는 1990년 초판이 발행된 이후 이용자 피드백 등을 기반으로 오늘날까지 지속적으로 갱신되어 PubMed, NLM Gateway 등 다양한 시스템에서 활용되고 있으며, 향후에 더욱 고도화된 시스템에의 적용 또한 기대된다.

국내에서는 아직 UMLS와 같은 통합개념체계를 적용한 연구결과물은 전무한 실정이다. 대신에 시소러스, 분류표, 용어사전

등에 대한 연구가 개별적으로 진행되어 왔는데, 시소러스의 경우 1992년 시스템공학연구소의 “과학기술용어 시소러스”, 한국언론연구원의 “신문기사 시소러스(1993)”, 법원도서관의 “법률 시소러스(1998)” 등 1990년대에 약 6개의 시소러스가 개발되었다. 그러나 대부분이 일회적인 개발로 끝나고, 관련 시스템과의 연계 부족으로 실효성을 거두지 못하고 있다.

과학기술분야의 경우 분류표는 한국과학재단, 과학기술정책연구원, 한국과학기술정보연구원 등에서 자체적으로 개발한 분류체계가 있다. 전문용어 관련 연구는 1998년부터 문화관광부가 주관해온 “21세기 세종계획”을 근거로 본격화되었으며, 전문용어언어공학연구센터(이하 KORTERM⁴⁾)를 중심으로 연구가 진행되고 있다. KORTERM은 2007년까지 과학기술, 인문사회과학 분야의 전문용어 집성을 목표로 하고 있으며, 현재까지는 물리, 화학, 경제분야의 코퍼스 구축을 통한 목록집성 단계로 향후 많은 연구과제가 남아있는 상태이다.

정보검색 시스템 및 이용자가 다양화될수록 특정개념에 대하여 단일한 통제어휘를 사용하도록 하거나, 시스템마다 별도의 통제어휘집을 사용하는 것은 정보검색, 정보연계, 정보 호환성 확보 차원에서 무리가 있다. 따라서 국내에서도 UMLS와 같이 동일 개념에 대한 유사 표현을 자동으로 연계하여 처리하기 위한 연구가 수행되어야 할 것이다. 특히 국내는 외래어 도입되는 학문의 영향으로 동일 개념에 대한

4) Korea Terminology Research Center for Language and Knowledge Engineering의 약자로 KAIST내에 설치되어 있다.

한글, 영어, 일본어 용어의 표기가 혼용되는 실정이므로, 기본개념을 한글 표준용어로 정립함과 더불어 유사 관련용어를 포괄할 수 있는 통합개념체계에 대한 연구가 활성화될 필요가 있다.

참고문헌

1. 川野惟二, (1998). "統合醫學用語システム (UMLS) 知識 ソースの概要." 『情報管理』, 41(1): 15-23
2. 이재운, 김태수. (1999). "WordNet과 시소러스.", 『언어정보의 탐구』, 1: 232-269
3. National Library of Medicine (2001). *UMLS Knowledge Sources*, 12th.
4. Olivier Bodenreider (2000). "A Semantic Navigation Tool for the UMLS." *JAMIA*, 7: 971
5. Padmini Srinivasan(1996). "Optimal Document-indexing vocabulary for MEDLINE." *Information Processing & Management*, 32(5):503-514
6. Squires, Steven J. (1993). "Access to Biomedical Information: The Unified Medical Language System." *Library Trends*, 42(1): 127-151
7. UMLS <<http://www.nlm.nih.gov/research/umls/umlsmain.html>>
8. UMLS Knowledge Source Server <<http://umlsks3.nlm.nih.gov/KSS/01/index.html>>
9. Willam Hersh, M.D. et al (2000). "Assessing Thesaurus-Based Query Expansion Using the UMLS Metathesaurus." *JAMIA*, 7: 344-348