NLM의 데이터베이스와 통합의학언어시스템(UMLS)

황 말 례
(의학도서관)

미국 국립의학도서관은 미국 Maryland주의 Bethesda에 위치하고 있으며, 미국 보건교육 후생부 산하 기관인 국립보건원의 한 기관으로, 세계적으로 발간되고 있는 각국의 의학관련 자료를 비치하여, 이용을 원하는 사람들에게 공개하고 있다. NLM은 120년의 오랜 역사를 가진 인쇄자료인 Index Medicus를 1964년 기계가독형으로 전산화 작업을 시작하여 1965년에는 MEDLARS(Medical Index Medicus Literature Analysis and Retrieval System)라는 종합적인 의학문헌 검색도구를 이용할 수 있도록 하였다. 1966년 MEDLINE(MEDLARS-online)으로 개칭하여 생의학 문헌 정보를 쉽고 빠르게 접근할 수 있도록 하였다. 또한 1971년에는 전문 분야별로 데이터베이스가 보강되어 MEDLINE외에 25개의 데이터베이스로 확대 발전하였고, 분야가 점점 세분되어 지금은 40여개가 넘는 보건 의료관계의 주요 데이터베이스로 발전시켰으며, 근래에는 원문 제공 자동화시스템인 DOCLINE과 LoanSome.Doc을 이용하여 미국 내 뿐만 아니라 국제 메들라스 센타가 설치되어 있는 각 국가에서도 원문 제공 서비스를 하고 있어 세계 의학도서관을 주도하고 있다.

미국 국립의학도서관의 의학정보 데이터베이스에 그 기원을 둔 MEDLINE은 Index Medicus와 Index to Dental Literature, International Nursing을 합친 데이터베이스로 미 국립의학도서관이 개발한 의학, 간호학, 치의학, 수의학뿐 아니라 건강관리 시스템과 임상 과학을 망라하는 최고의 서지데이터베스로, 7,300여종의 학술지, 900만의 참고서지, 1,100만 여개의 의학논문들이 주제 표목(subject heading), 부표목(subheading), 저널의 형태(publication type), 연령군(age groups), 부집합(subsets) 등의 검색용 인덱스를 체계적으로 정리하여 상호 연결한 복잡한 입체적인 구조물이다. 이들 검색어의 가능한 조합은 실제로는 무한에 가까워 매번 사용할 때마다 그 복잡성과 풍부함에 놀라게 되는데, 이는 MEDLINE이 의학 주제명 표목(MeSH)과 통합언어시스템(UMLS)으로

구성되어 있기 때문이다. 보건전문인, 과학자, 사서, 일반인들의 MEDLINE 검색량은 하루에 35만건에 이른다. 비록 현재의 MEDLINE이 1966년 이후 지금까지의 논문을 다루고 있지만, 다른 데이터베이스인 OLDMEDLINE은 좀 더 과거의 자료들을 다루고 있다. OLDMEDLINE에서는 1960년대에서 1965년대 사이의 자료는 Cumulated Index Medicus에서 1958년부터 1959년 사이의 논문은 Current List of Medical Literature형태로 NLM Gateway를 통하여 제공하고 있다.

1. NLM의 데이터베이스

1.1. PreMEDLINE(http://www.nebi.nim.gov/PubMed)

완전한 서지정보가 작성되어 메들라인에 수록되기 전의 기본적인 서지사항과 초록을 포함하고 있는 데이터베이스로 출판형태, 유전자은행회번호, Mesh 등의 색인어 부여 후 1주일 간격으로 메들라인에 추가 등록 되고 PreMedline 에서는 데이터가 삭제된다.

1.2. PubMed/MEDLINE (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed)

NLM에 있는 NCBI에서 개발된 것으로 세계적으로 사용되는 의약학, 간호학, 치과, 수의학, 치료시스템 및 임상 전 의학 분야의 인용 및 초록 데이터베이스, 1950년부터 소장된 5천 여종의 생의학 저널의 초록정보를 수록한 참고자료 등이 Entrez 검색시스템에서 이용 가능하다.

MEDLINE에 있는 모든 인용정보는 Medical Subject Headlings (MeSH)와 UMLS의 통합어휘 통제로 색인되어 검색작업의 정확도를 높여주고 있다.

1.3. MEDLINE*plus*(http://medlineplus.gove)

United States Pharmacopoeia Drug Informatioin(USPDI) 같은 환자나 소비자를 대상으로 하는 무료약물 데이터베이스, 소비자 건강 관련 웹정보, 건강주제, 의학사전, 약품정보, 최근보건소식, 주소록 등의 정보를 1998년 10월에 개발하여 제공하고 있다.

1.4. *ClinicalTrials.gov*(http://clinicaltrials.gov)

미국 보건복지부(NIH)에서 환자와 가족 구성원들, 치료전문가들에게 최신 임상 조사 연구 자료를 제공하기 위해 NLM을 통하여 NIH · FDA가 공동개발한 데이터베이스로 2000년 2월 29일 개설이후 임상 연구 프로토콜의 초록, 환자의 치료결정을 도와줄 치료정보를 수록하고 있다.

1.5. Visible Human 프로젝트(http://www.nlm.nih.gov/research/visible/visible_human.html)

NLM에서 인체 내부에 대한 심도 있는 연구를 위해 1989년에 수립된 프로젝트로 남자데이터(Visible Human Man)의 경우 1994년 11월에 데이터가 제작되었고, 여자 데이터(Visible Feman data set)는 1995년 11월에 제작되어 인간 해부학 연구에 참고자료로 사용되고 있으며 의학 이미지 알고리즘, 이미지 라이브러리의 시험대 및 모델로 사용되고 있다.

1.6. LocatorPlus(http://locatorplus.gov)

NLM의 온라인 목록으로 120만개 이상의 단행본, 저널, 시청각자료, 컴퓨타파일 등의 레코드를 수록하고 있다.

1.7. PMC(Public Medical Central), http://www.pubmedcentral.nih.gov/about/intro.html)

미국 국립보건원(NIH)의 생의학, 생명과학저널을 디지털화하고 NLM의 NCBI에서 개발·관리하고 있다. Open Access Repository로 무료 이용 가능하나, 최근 12개월 또는 3년 이내에 발표된 아티클은 이용이 제한되고 있다. 2006년말에 70만건의 기사들이 수록되었고, 매년 7%의 증가율을 보이고 있다.

1.8. MeSH(Medical Subject Headings)

미국국립의학도서관의 통제어휘집으로 논문을 색인하거나, 단행본 목록, Medlars 데이터베이스 검색시 사용된다. MeSH 용어는 각기 다른 용어를 하나의 동일한 개념으로 사용하여 정보를 검색하게 되어 있으며, 이것은 단순한 표목표와는 달리 알파벳 리스트, Tree Structure, Subheading으로 구성되어 있어 MeSH 브라우져가 찾고자 하는 주제 분야의 디스크립터를 찾아 다른 디스크립터와의 관계를 알아보는 데 사용되고 있다.

목적, 구조, 특성에 따라 100여개의 의료용어체계표준이 통용되고 있는데, 예를 들어 간호분야에서는 NANDA(North Amercian Nursing Diagnosis Association), 치의학에서는 CDT(Current Dental Terminology), 종양학에서는 PDQ(Physicians Data Query)가 통용된다. 그래서 1986년 NLM에서는 현재 의학 및 의료분야에서 사용되는 각종 용어체계를 공통개념으로 통합하기 위하여 UMLS 프로젝트를 시작했다. 참고로 2007년도에 MeSH에 수록될 MeSH 데이터를 살펴보면 22개의 descriptor가 폐기, 494개의 신규, 99개가 새 MeSh로 추가될 예정이며, 전체 수록량은 descriptor 24,357개, Subheadings/Qualifiers 83개, Supplement Concepts

Records(SCRs) 164,331개로 http://www.nlm.nih.gov/mesh/에 접속하여 이용할 수 있다.

1.9. Entrez(http://www.ncbi.nlm.nih.gov/Entrez)

유전자 염기 서열, 단백질 서열, 단백질의 결정 구조, NMR data 그리고 이와 관련된 논문들의 초록 database인 MEDLINE 등을 통합하여 쉽고 편리하게 이용할 수 있도록 구성되어 있으며 검색 및 이용은 NCBI 홈페이지 위쪽에 7개의 버튼들 중 하나인 Entrez나 왼쪽 메뉴의 Database Services내 Entrez Search System을 선택하여 다음의 5가지 데이터베이스를 검색할 수 있다.

① Nucleotides : GenBank, EMBL, DDBJ의 염기 서열들의 검색이 가능하다.

② Proteins : Swiss-Prot, PIR, PRF, PDB에 있는 단백질 서열들과 GenBank, EMBL, DDBJ에 있는 염기 서열들을 translation해서 얻은 예상 단백질 서열들의 검색이 가능하다.

③ 3D structures : PDB와 NCBI의 Molecular Modeling Database(MMDB)에 있는 단백질의 3차 구조 및 검색 결과의 Structure summary를 선택하여 Cn3D로 검색 가능하다..

④ Genomes : genome과 염색체의 mapping data 검색이 가능하다.

⑤ Taxonomy : NCBI/GenBank에서 운영되는 taxonomy database 검색이 가능하다.

1.10. TOXNET Files (http://toxnet.nlm.nih.gov)

주로 독물학, 위험 화학 물질, 환경보관 관련 영역 정보로 다음의 데이터베이스들이 있다.

① HSDB(Hazardous Substances Data Bank) : 산업 위생 등의 위험 가능성이 존재하는 화학 물질 데이터를 수록하고 있다.

② IRIS(Integrated Risk Information System) : 인간의 건강 평가를 지원하는 데이터로 주로 EPA 발암물질 분류표, 단위별 위험도, 경사 요인, 경구 복용량, 흡입물 표준농도등의 데이터를 수록하고 있다..

③ ITER(International Toxicity Estimates for Risk) : 국제 위험평가 정보를 각 기관별로 추출하여 제공하는 데이터베이스로 600개 이상의 화학 레코드를 수록하고 있다.

④ CCRIS(Chemical Carcinogenesis Research Information System) : 발암성, 돌연변이 유도, 종양촉진, 종양활동 억제 실험 결과에 관한 9,000개의 이상의 화학정보를 수록하고 있다.

⑤ GENE-TOX(Genetic Toxicology) : 미국 환경보호국에서 만들어진 데이터베이스로 3,200개의 화학물질의 실험결과를 수록하고 있다..

⑥ TOXLINE : 생화학과 약리학, 생리학, 약제와 다른 화학물의 영향을 다루는 서지정보를 모아놓은 광범위한 데이터베이스이다.

⑦ DART/ETIC (Development & Reproductive Toxicology / Environmental Teratology Information Center) : 미국 환경보호국, 환경 건강 과학 국제기구, 독물학 연구센터, NLM에 의해 창시된 DART와 ETIC은 TOXNET의 서지 데이터베이스로써, 1965년 이후의 기형학을 포함한 재생성 및 개발 독물학 문헌들을 수록하고 있다.

⑧ TRI(Toxic Release Inventory) : 미 환경보호국이 수집한 환경과 폐기물에 의해서 년간 발생되는 독성화학물질에 관련된 정보를 제공하고 있다.

⑨ ChemIDplus : NLM의 데이터베이스에서 인용되는 화학 물질명을 식별하기 위해 이용되는 화학물질 구조와 명칭의 전거파일로 380,000개가 넘는 화학 레코드와 263,000개가 넘는 화학 구조식을 포함하고 있다.

⑩ DIRLINE(Directory of Health Information) : 온라인상의 건강관계 정보원의 주소, 전화번호, 서비스 내용등을 수록한 디렉토리이다.

⑪ Haz-Map : 작업중의 노출된 화학물질 관련 데이터를 수록하고 있다.

⑫ Household Products : 가정용품에 노출된 유해 물질 관련 데이터를 수록하고 있다.

⑬ LactMed(Drugs and Lactation Database) : 수유시 약물투여 관련 정보를 수록하고 있다.

2. NLM의 상호대차 시스템

2.1. DOCLINE

1985년 미국 NLM에서 개발한 도서관 상호대차 시스템으로 도서관간 소장정보를 연결하고, 신속하고 능률적인 문헌제공 및 상호협력을 위해 구축된 시스템으로 보건관련 전문인들에게 의학도서관 국가망(NNLM)을 통하여 도서관 상호대차의 빠른 루트를 제공해 주고 있다. 필요로 하는 원문을 신속하게 받아볼 수 있는 상호대차 서비스는 2000년 7월에 웹버전으로 개발되어 이용하고 있고, 우리나라의 경우 한국 메들라스 센터의 DOCLINE을 이용하여 원문신청 할 수 있다.

2.2. LoanSome Doc

그 지역 메들라스 센타를 통하지 않고 PubMed나 NLM gateway를 통해 개인이 직접 원문신청을 하는 것으로, 서비스를 등록하기 전에 그 지역 메들라스센타의 중간 승인을 거친 후에 신청가능하며 사용요금 역시 그 지역 메들라스센터로 지불한다.

2.3. 한국메들라스센터

한국 메들라스센터는 1992년 4월, 서울대학교 의과대학과 미국 국립의학도서관과의 협약으로 설립되었다. 현재 메들라스센터는 미국 국립의학도서관(NLM)에 본부를 두고 세계 22개 국가에 메들라스센터가 설립되어 운영되고 있다. 국제메들라스센터는 미국 국립의학도서관이 보유하고 있는 최신의 의학관련 정보를 직접 이용자에게 제공하고 있으며 이용자가 요구하는 각종 참고봉사업무를 서비스하고 있다. 또한 가입되어 있는 국내 보건의료 분야 및 관련 전문가들에게 미국 국립의학도서관의 MEDLARS 시스템을 직접 연결하여 문헌을 검색할 수 있도록 하고, NLM으로부터 부여받은 아이디/패스워드를 이용자에게 부여해 주며, 데이터베이스 사용요금을 NLM에 송금하는 등의 서비스를 DOCLINE과 LoanSome Doc을 통하여 제공하여 주고 있다.

3. NLM의 데이터베이스를 검색하는 소프트웨어

3.1. Internet Grateful Med(IGM)

IGM은 웹상의 15개의 NLM 데이터베이스를 검색할 수 있도록 보조하는 소프트웨어이다. NLM의 주요 시스템과 데이터베이스가 점차로 발전되어 나감에 있어, 이용자들을 다양한 정보원들로 연결시켜 상호작용 할 수 있는 도구가 필요하게 되었고, IGM은 바로 그런 역할을 담당하는 지능형 검색 도구였다. IGM은 MEDLINE (PREMEDLINE 포함), AIDSLINE, AIDSDRUGS, AIDSRIALS, DIRLINE, HealthSTAR, HSRPROJ, HISTLINE, SDILINE, OLDMEDLINE, SPACELINE, BIOETHICSLINE, POPLINE, TOXLINE과 CHEMID를 통합하는 검색이 가능하도록 해 준다. MeSH 통제 어휘로 변환하여 MeSH 표목과 기사명 키워드로 동시에 검색한다. 그리고 검색어를 자동적으로 증폭시켜 그 모체가 되는 언어는 물론이고 세부적인 하부의 언어들까지 포함한 검색을 실행할 수 있어 편리한 기능을 가졌으며, 검색 용어에 제한어를 더하고 부표목 자격어를 더함으로써 검색에 강조점(focus)을 둘 수 있도록 하였다. 다양한 종류의 제한 검색을 할 수 있으므로 이를 잘 이용하면 검색 결과가 무용하게 확대되지 않고 목적에 맞는 검색 결과만을 찾을 수 있다. IGM의 독특한 기능인 "Analyze Search"기능은 상호 참조 할 수 있 는 수천 개의 관련어를 제공한다. 예를 들어 "Organization & Administration" 또는 "Therapy"라는 의미로 동시에 쓰일 수 있는 모호한 단어인 "Management"라는 단어로 검색한다고 할 때 IGM은 그 유사어들 까지 검색하여 모호한 검색을 명확하게 하는 연산식을 제공한다. 또한 IGM은 검색식을 형성하고 정제하는데 가장 중요한 자원이라고 할 수 있는 NLM의 통합 의학 용어시스템(Unified Medical Language System:UMLS) Metathesaurus을 이용한다. Metathesaurus 1998년도 판은 40개의 다른 생물학 어휘집, 용어 사전, 분류집에서 추출한 476,000개의 용어를 정의하는 1,052,000개 이상의 단어를 포함한다. 또한 MeSH의 프랑스, 독일, 스페인, 포르투갈, 러시아 역어까지 포

함하여 국제적으로도 인정받고 있다.

IGM의 메타시소러스 키를 이용하면 1.6 기가 바이트 이상의 메타시소러스 파일을 사용하여 이용자가 정확히 찾고자 하는 용어로의 접근점을 가깝게 할 수 있다. 또한 검색 용어에 대한 결과 리스트와 용어 정의, MeSH 주석을 제공하고 MeSH의 트리 구조를 보여주어 이용자가 선택한 검색 용어의 상하 구조를 볼 수 있다. 이 트리 구조에서 이용자가 관심 있는 분야를 발견했다면 이를 검색식에 가져올 수 있다. 이렇게 발견한 검색용어를 클릭함으로써, 거의 즉각적으로 검색식을 조합할 수 있다. 이렇게 유용한 검색 소프트웨어인 IGM은 2001년 NLM Gateway 서비스로 대체되면서, 기존 IGM에서 검색되었던 데이터베이스들이 NLM Gateway에서 검색하게 되었다.

3.2. NLM Gateway

Gateway는 NLM에 무엇이 있는지, 찾고자 하는 것을 검색할 수 있는 가장 좋은 방법이 무엇인지 전혀 알지 못하는 이용자를 위한 검색 도구이다. NLM Gateway는 단 하나의 인터페이스만을 제공하여 다양한 검색 시스템을 이용할 수 있도록 한다. 이러한 신 개념의 검색도구인 NLM Gateway는 기존의 상용되던 Internet Grateful Med를 대체하게 되었는데 NLM Gateway에 관해서 서술하면 다음과 같다.

3.2.1. 시스템 목표

① NLM의 여러 정보원들을 한 곳에서 이용할 수 있게 한다.

② 서지정보, 원문, 비디오, 오디오와 이미지 정보를 제공한다.

③ 이용자들을 위해 NLM 데이터베이스는 물론 외부자료까지 링크를 제공한다.

④ 통합된 의학용어 시스템인 UMLS를 최대한 활용할 수 있도록 한다.

3.2.2. NLM Gateway에서 현재 이용할 수 있는 데이터베이스

① 서지관련 데이터베이스

- MEDLINE/PubMed : 저널 인용색인, 초록
- NLM Catalog : 도서, 저널, 시청각 자료
- TOXLINE Special : 독물학 인용색인
- DART : 개발 및 재생 독물학
- Meeting Abstracts : 초록
- Bookshelf : 생의학 서적 전문

② 소비자 건강관련 데이터베이스

- MedlinePlus : 건강관련 주제
- MedlinePlus : 약품 정보
- MedlinePlus : 의학사전
- MedlinePlus : 최근 건강 뉴스
- MedlinePlus : 기타 다른 자원들
- ClinicalTrials.gov : 임상 실험 정보
- DIRLINE - 의학기관 주소록
- Genetics Home Reference : 유전학 참고정보
- Household Products Database : 가정용품에서 유발되는 독물 데이타베이스

③ 기타 정보 자원

- HSRProj : 의학 서비스 연구 프로젝트
- OMIM : 멘델 유전을 보이는 사람의 유전자 및 병명 검색
- HSDB : 위험물질 데이터뱅크
- ITER : 위험물질에 대한 국제 독물
- GENE-TOX : TOXNET의 독물학 데이터 파
- CCRIS : TOXNET의 독물학 데이터 파일

3.2.3. Gateway의 특징

Gateway는 이용자가 하나의 검색어를 입력하면 이를 자동적으로 다양한 검색 시스템으로 보내 이용자에게 유용하지만 서로 다른 특성을 갖는 정보들을 제공하게 된다. 검색 결과물에는 괄호 속에 정보원이 표시되며 이용자들은 Gateway를 이용하여 NLM 정보원들을 개괄적으로 훑어볼 수 있다. 단 한번의 검색으로 원하는 정보를 찾을 수도 있고, PubMed 나 MEDLINEplus와 같은 검색한 결과의 정보원을 연결하여 원하는 정보를 얻는 등 포괄적인 검색이 가능하다. 기본적인 Gateway의 검색 인터페이스는 간단하며 Gateway에 링크된 데이터 베이스의 기능도 모두 이용할 수 있다. 이용자는 자신에게 맞는 인터페이스를 적용하여 우선 순위를 정할 수 있다. 또한 Gateway의 "locker"키를 이용하여 선택한 검색결과를 저장 및 다운로드, 프린트, E-메일 발송을 할 수 있다. locker를 이용하기 위해서는 로그인을 위한 아이디와 패스워드 등록을 해야하며 이는 무료이다.

4. NLM의 통합 의학언어시스템(UMLS) 개발

4.1. 개요

효과적인 정보검색도구로 활용되는 시소러스, 분류표등의 통제어휘집을 동일한 개념이나 의미를 표현하는데 서로 다른 어휘를 사용하고 있다. 이러한 동일개념에 대한 용어표현 차이로 인한 정보의 검색 및 통합문제를 해결하기 위하여 개발한 것이 UMLS로 NLM이 1986년부터 시작하여 2001년도까지 연구한 장기적 프로젝트이다.

4.2. 개발배경

생물의학 분야의 다양한 정보검색 시스템은 정보검색의 효율성을 높이기 위해 시스템마다 고유한 통제어휘집을 사용하고 있으며, 이용자들은 각 시스템에서 사용되는 용어로 검색해야만 필요한 정보를 누락없이 찾을 수 있었다. 이러한 전자정보원간 개념의 표현차이로 인해 발생하는 문제를 해결하기 위해 NLM은 UMLS 개발에 착수하였다.

또한 Knowledge Source를 개발하여 이용자와 정보시스템이 사용하는 개념의 표현차이로 인한 문제를 해결하고, 서지 데이터베이스. 사실 데이터베이스, 환자기록 시스템 등 이종의 시스템간 정보를 연계시키는데 활용하고자, 1986년부터 NLM에서 개발을 시작하여 1990년 초판이 발행되 었으며 매년 주기적으로 갱신하고, 전 세계 개인 및 기관에서 무료로 사용할 수 있도록 하였다.

4.3. UMLS의 구성

문헌데이타베이스, 의무기록 관리시스템, 지식기반 등의 다양한 서로 다른 정보원의 원천에서 얻은 정보를 사용자의 편의를 돕기위해 개념적으로 연관지으려는 시도로서 만들어진 것으로, UMLS Knowledge Source에는 메타시소러스(Metathesaurus), 의미망(Semantic Network), 전문가사전(Specialist Lexicon)으로 구성하며, 이 세 가지 요소는 상호 유기적으로 운영됨으로써 UMLS의 효과가 극대화 된다.

4.3.1. 메타시소러스

UMLS의 중심 어휘요소로서 생물의학분야의 통제어휘집, 분류표 등에서 한 번 이상 나타나는 개념에 대한 정보 데이터베이스로 개념의 의미, 속성, 계층적 연관성, 개념과 용어간의 관계, 동일어 체계를 포괄하고 있다. 또한 메타시소러스는 NLM 주제명 표목표인 MeSH를 포함하여 NCI thesaurus, National Drug Data File, International Classification of Disease 등 생물의학 분야 시소러스, 분류표 등의 통제어휘집을 기초로 구성되었으며 2001년에는 영어 외에 불어, 핀란

드어, 독일어, 이탈리아어, 포르투갈어, 러시아어, 스페인어판 MeSH를 포함하고 있다.

메타시소러스는 식별자(Identifier)를 이용하여 한 개념(concept)에 대한 각기 다른 용어(terms) 문자열(strings)을 연계하여 같은 개념으로 인식하고 있는데, 예를들어 이용자의 질문을 분석하여 적절한 전문용어로 바꾸어주기도 하고, 동의어와 번역어 등이 연결되어있어 사용하기에 혼란이 없다.

① 개념 식별자(unique concept identifier, CUI) :개념에 관한 모든 정보를 연계하고, 속성을 규명하는데 이용되는 개념이다.

② 공통용어 식별자(common term identifier, LUI) : 용어는 어휘가 변형된 모든 문자열의 그룹으로. 각 문자열이 LUI에 의해 변형된 모든 어휘와 연결이 가능하다.

③ 문자열 식별자(unique string identifier, SUI) : 문자열은 실제 자료에 나타나는 형태로서 철자변형, 단복수 변형 및 언어변형에 따라서 고유한 SUI가 부여된다.

※ 문자열 식별자와 공통용어 식별자는 최소한 하나의 개념 식별자에 연결되며, 동일한 의미를 지닌 다양한 용어들은 동일한 개념 식별자에 의해 서로 연결된다.

4.3.2. 의미망(Semantic Network)

의미망은 의미유형(semantic type)과 의미관계(semantic relationship)를 이용하여 메타시소러스의 각 개념에 대하여 일관된 범주를 부여, 개념간의 유용한 관계를 제공하며, Relational table format과 unit record format의 2가지 포맷으로 제공된다.

① 의미유형 : 계층적 구조로 되어있으며 개체(Entry)와 사건(Event)으로 구분한다. 개체는 99개 유형으로 물리적 대상, 개념적 실체를 그룹화한 것이며, 사건은 활동(activity)과 과정(process), 상태(state)와 관련된 개념들을 그룹화한 것으로 35개 의미 유형으로 구성되며 하나의 개념에는 최소한 하나 이상의 의미유형이 할당된다.

② 의미관계 : 의미유형 사이에 부여되는 관계로, 계층관계와 비 계층관계로 구분되며, 의미망에서 기본적인 연결은 계층관계로 이루어지며, 비계층관계는 물질, 공간, 시간, 기능, 개념의 다섯가지 범주로 구분된다.

4.3.3. 전문가 사전(Specialis Lexicon)

전문가 자연어 처리시스템에 필요한 모든 사전(lexical) 정보를 제공하기 위하여 개발된 것으로, 생물 의학적 용어뿐만 아니라 일반 어휘에 관한 정보도 포함한다. 어휘변형 생성 프로그램은 큰 어순, 용어 변형 및 대소문자 변형 등으로 자동 관리함으로써, 이용자로 하여금 어휘 변형을 고려하지 않도록 해주며, unit record format과 Relation table format으로 제공된다.

4.4. UMLS의 적용사례

4.4.1. PubMed

UMLS 메타시소러스는 MEDLINE 데이터베이스의 검색성능 향상을 위하여 적용, 활용되고 있는데, PubMed에서는 동의어와 변형어 정보를 지난 UMLS를 기반으로 MeSH 변환테이블에 의해서 일반 검색용어를 통제용어로 변환하여 검색을 실시한다. 예를 들어 PubMed 검색화면에 vitamin h를 입력하면, PubMed는 MeSH 변환테이블에 의하여 질의어 vitamin h의 통제어휘인 biotin과 연계하여 검색을 실시한다. PubMed의 독특한 기능은 검색해낸 기사에 관한 관련문헌을 즉시 연결하여 볼 수 있다는 것인데 관련문헌은 사용된 문구의 유사성에 기초해 미리 산출된 것들이다.

4.4.2. NLM Gateway에서의 UMLS적용

NLM Gateway는 NLM의 다양한 정보검색 시스템을 동시에 one-stop으로 검색할 수 있도록 단일 인터페이스를 제공하는 시스템으로 검색에 있어 주요기능은 다음과 같다.

① MeSH가 아닌 용어가 입력될 경우 메타시소러스의 동의어 등 각종 용어 맵(map)이 체크되어 관련된 MeSH 용어를 제시하는데, 만약 적합한 MeSH용어가 없는 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 cold처럼 복수의미를 가질 경우에는 MeSH 용어 선택기능을 이용하여 "cold, common cold, cryotherapy, obstructive lung diseases" 등의 문자열에서 가장 적합한 용어를 선택하게 한다.

4.4.3. 해외 UMLS 유사개발사례

① WorldNet : 1985년 프린스턴대학 인지과학연구실이 개발한 영어어휘 데이터베이스

② EDR : 일본 전자화 사전연구소의 EDR 전자사전

4.4.4. 국내 UMLS 관계 연구현황

① 보건복지부의 지원으로 통계청의 한국 표준 질병사인으로 분류 KCD-3,4에 기반하여 한글의학용어를 UMLS의 메타시소러스체계로 정리하였으며 우리나라가 NLM으로부터 확보한 한글의학용어식별번호(LUIs) 범위는 L9900000-L9999999이b며 한글 의학 문자열 식별번호(SUIs)의 범위는 S9900000 - S9999999이다.

② 한승범, 최진욱. UMLS의 증상용어와 국내의무기록에서 사용되는 증상용어와의 비교연구,

2001년
③ 박수준, 김태현, 이현숙, 장현필, 박선희. UMLS를 기반으로 생물학문헌으로부터 생물학적 개체명을 인식하는 장치 및 그 방법, 한국전자통신연구원. 2003년 10월 출원, 2005년 4월공개
④ 최진욱. UMLS를 이용한 한글진단, 증상용어 Terminology Server 개발, 2003년 5월.

5. 맺음말

이상과 같이 ①NLM이 제공하는 데이터베이스들과 수록 내용 ②검색 소프트웨어 ③ 한국 메들 라스센타의 DocLine과 LoanSome Doc을 이용한 상호대차서비스 ④ 이러한 데이타베이스들의 정보들을 이용자들이 누락없이 검색할 수 있도록 NLM이 개발한 통합의학언어시스템(UMLS)의 구성 및 적용사례 등을 살펴보면서 한국메들라스센타의 업무인 NLM과의 상호대차업무, MeSH 관련업무, NLM과의 국제교류업무 등의 역할을 더욱 증대시키고 적극적으로 수행할 수 있도록 연구하고 노력해야겠다는 것을 느꼈다. 또한 우리나라의 전자의무기록시스템(Electronic Medical Record), 임상의사결정시스템(Clinical Decision Support System)같은 의학정보시스템 구축시 NLM Metathesaurus 표준어휘집를 참조하여 개발해야 될 것으로 사료된다. 이미 한글과 병합된 UMLS를 활용할 수 있는 UMLS Semantic Browser(USB)가 연구 개발되고 있어, 머지 않아 MEDLINE에 한글로 등재된 한국의학 기사들을 검색할 수 있음도 기대해 본다.

끝으로 우리 도서관인들은 환자진료와 연구만으로도 바쁘고 벅찬 일상을 살아가는 학생, 교수 및 전문의료인들의 학습 및 연구활동을 지원하기 위하여 NLM의 데이터베이스를 비롯한 각종 의학 관련 데이터베이스의 내용을 숙지하고, 효율적인 검색법을 습득하여 이용자들에게 신속하고 정확한 최신의학정보를 제공할 수 있도록 노력해야 할 것이다.

◈ 참고문헌 ◈

김혜선. 통합의학언어시스템(UMLS)의 구성 및 적용에 대한 고찰, 정보관리연구 v.32 no.2 (2001)

NLM : http://www.nlm.nih.gov

NLM Gateway : http://www.nlm.nih.gov/gw

NLM Toxnet : http://toxnet.nlm.nih.gov

메들라인 : http://www.nlm.nih.gov/pubs/factsheets/medline.html

한국메들라스센터 : http://medlars.snu.ac.kr

의학주제명 표목 : http://www.nlm.gov/mesh/