

약물 데이터베이스 검색 모듈 구현

이환구*, 임상택*, 강승희**, 김규원**, 김성훈***, 차재혁*
*한양대학교 정보통신대학원
**서울대학교 종합약학연구소
***서울대학교 약학대학 단백질합성효소네트워크 연구단
e-mail : prodog@ihanyang.ac.kr

Implementation of Drug Database Search Module

Hwangu Lee*, Sangteak Lim*, Seung-Hui Kang**, Kyu-Won Kim**,
Sunghoon Kim***, Jaehyuk Cha*

*The Graduate School of Information & Communications, Hanyang University

**Research Institute of Pharmaceutical Sciences, Seoul National University

***National Creative Research Initiatives Center for ARS Network,
College of Pharmacy, Seoul National University

요 약

현재까지 개발된 약물들은 그 종류가 매우 다양하며 그 작용기전 또한 매우 다양하다. 약물이 작용하는 기전에 따라 같은 약물이 다른 용도로 사용되기도 하고, 약물이 어떤 Target 들에 작용을 하게 되면 생체 내에서 이 Target 이 다른 생체 내 분자들과 상호작용을 함으로써 우리가 원하는 약효나 원치 않았던 부작용들이 나타나게 된다. 이렇듯 약물과 그 약물이 작용하는 Target 에 대한 정보의 구축은 약물의 매우 중요한 정보이며, 이러한 정보의 데이터베이스화는 차후 신약개발에 있어 매우 중요한 정보로 사용될 수 있다. 본 논문에서는 이렇게 구축된 약물 데이터베이스에 대하여 이의 사용자 편의성을 극대화하는 효율적인 검색 모듈을 제시한다.

1 서 론

1.1 연구개발의 목표와 의의

약물의 개발은 여러 가지 방식으로 이루어지게 된다. 기존에 알려진 약물의 구조를 바탕으로 새로운 약물을 개발하기도 하며, 질병의 원인이 되는 유전자나 단백질, 혹은 수용체를 발견하여 이것의 구조로부터 새로운 약물을 개발하기도 한다. 또한 구조 연관성이 없어도 기능을 알고 있는 물질들로부터 약물을 개발하기도 한다. 이렇게 개발되는 약물들은 간단한 화학적 소독약에서부터, 유전자 공학 기술을 이용한 최첨단의약품까지 그 종류가 매우 다양하다. 인간게놈프로젝트의 완성이 다가오면서 개개인의 유전자에 맞는 개별 맞춤 약까지 등장하고 있는 실정이다. 하지만 하나의 약물 개발에는 많은 시간과 시행착오가 있으며, 이 기간을 단축하려는 것은 신약개발에 있어

꼭 필요한 일이다. 그 노력 가운데의 하나가 컴퓨터를 이용한 약물 작용점의 발굴이다. 약물은 생체 내 Target 에 작용하여 약효를 나타낸다. 이러한 약물의 Target 으로는 DNA, RNA 및 각종 Enzyme, Receptor 등의 단백질 및, 생체 내 여러 분자들이 존재한다[1, 2]. 약물은 이 Target 들에 작용하여 Target 의 기능을 촉진시키거나, 저해하는 등의 결과를 초래하고, 이 Target 들은 다른 단백질이나 유전자와 상호작용을 함으로써 우리가 원하는 약효나 원치 않았던 부작용들이 나타나게 되는 것이다. 따라서 약물과 그 약물이 작용하는 Target 에 대한 정보의 구축은 약물의 매우 중요한 정보이며, 이러한 정보의 데이터베이스화는 차후 신약개발에 있어 매우 중요한 정보로 사용될 수 있다.

본 논문에서는 이렇게 구축된 약물 데이터베이스에 대하여 이의 사용자 편의성을 극대화하는 효율적인 검색 모듈을 제시한다.

1.2 연구개발 현황

현재 개발된 국내 약물 DB 로는 가장 널리 쓰이는 eKIMS[3], BIT Drug Info[4] 등의 상용 약물 정보 데이터베이스가 있다. 이 약물 데이터베이스들은 약물에 대한 종합적인 정보를 제공하는데, 대상은 일반인 및 보건의료인을 대상으로 하여 그 범위가 상당히 폭넓다. 약품의 성분명, 상품명, 작용기전, 제조회사, 보험수가, 약품 식별, 부작용, 복약지도 등의 실제 임상에서 필요한 정보들을 담고 있다. 그리고 한국의약품정보센터[5]에서 제공하는 약물검색, 숙명여대 의약품정보센터[6]에서 제공하는 의약품 정보 DB 등이 있다. 그러나 이러한 약물 데이터베이스들은 그 특성상 어떤 약물(성분)에 대한 구체적인 작용점 및, 성분의 화학적 특성 및 적용 질환에 대해선 구체적으로 검색하기가 힘들다. 또한 이러한 목적으로 개발되어 있지 않기 때문에 약물과 target 을 중심으로 한 검색과 network 작성이 어려운 실정이다.

2 본 론

2.1 약물 데이터베이스

DTNet(Drug-Target Network) 데이터베이스는 현재 시판 중인 1300 여종의 약물들에 대해 약물과 그것의 Target 및 여러 정보들을 체계적으로 정리한 데이터베이스이다. 이를 위하여 먼저 각종 약물 DB 들로부터 약물 List 를 확보하고, 약물들을 카테고리화할 수 있게, 약물들이 주로 관여하게 되는 질병, 약물작용에 따른 분류, 약물의 작용방식에 따라 각각 Disease List, Drug Application, Mechanism of Action 을 정리하여, 각각의 약물에 해당하는 정보들을 List 에서 추출하여 데이터를 작성하였다. Disease List 로는 주요 질환에 대하여 441 가지의 카테고리 나뉘었고, Drug Application 은 약물의 주요 용도에 따라 19 가지의 대분류로 나누고 각 큰 분류 내에 좀 더 자세하게 소분류로 나뉘어 146 가지로 나누었다. Mechanism of Action 은 약물이 target 에 작용하는 방식으로, activator, agonist, alkylator, antagonist, antibody, antisense, blocker, coenzyme, cofactor, immunotoxin, inactivator, inhibitor, opener 의 13 가지로 나누었다. DTNet 은 Drug Name, Chemical Composition, IUPAC Name, Target, Mechanism of Action, Related Disease, Drug Application 의 정보가 이루어져 있다.

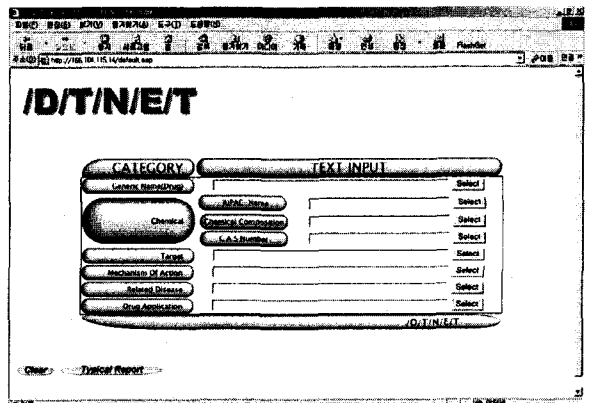
Drug Name 은 약품의 일반명으로 성분명을 지칭하고, Chemical Composition 은 각 원자가 몇 개씩 있는지의 화학식을 뜻한다. Ethyl Alcohol 의 경우 C₂H₅OH 이므로 C₂H₆O 가 된다. IUPAC Name 은 IUPAC 에서 정한 화학품의 공식 이름이고, 구조 정보로 2D, 3D mol file 을 이용한 구조 데이터도 있다. Target 은 약물

의 작용지점이 되는 분자의 이름이고 이것이 DNA 인지, Protein 인지, RNA 인지에 대한 정보도 같이 들어간다. 그리고 Mechanism of Action 은 Inhibit, Activate, Agonist, Antagonist 등의 약물의 Target 에의 작용에 대한 정보이며, Related Disease 는 약물의 용도를 말한다. 예를 들어 Ulcer 등의 위궤양 등에 쓰이는 Ranitidine 이나 Famotidine 등은 위의 Proton Pump 를 Inhibition 하여, 위산분비를 저해시키므로 Antacid(제산제)로 분류된다. 하지만 하나의 약물이 하나의 Target 에만 적용되는 것이 아니므로, 하나의 약물은 여러 Target 정보를 가질 수 있고, 또한 여러 Mechanism of Action, Disease, Drug Application 등의 정보를 가질 수 있게 된다.

2.2 검색 모듈 구현

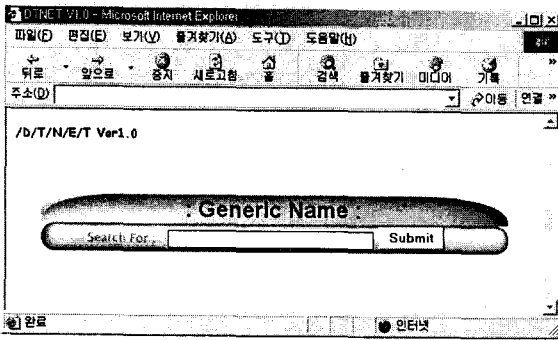
사용자는 DTNet 데이터베이스에 대하여 장소에 구애 받지 않고 검색하기를 희망할 것이며, 우선적으로 Drug, Target, Mechanism of Action, Related Disease, Drug Application 의 항목으로 검색하고자 할 것이다.

이를 충족시키고자 DTNet 데이터베이스에 대한 사용자 검색 인터페이스는 아래 [그림 1]처럼 웹 어플리케이션으로 구현하였으며 초기 화면에 위의 각 항목의 입력란을 만들어 검색할 수 있도록 하였다.

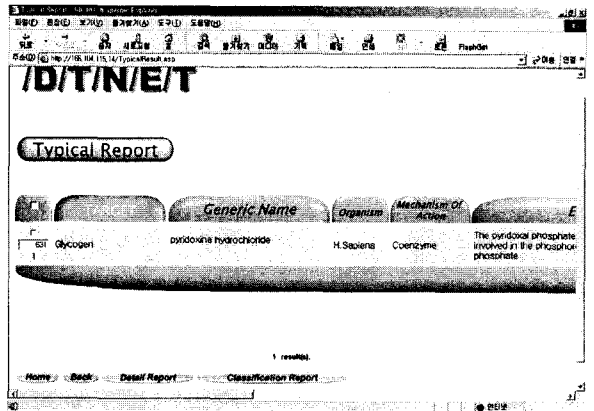


[그림 1] DTNet 검색 초기화면

[그림 1]에서 볼 수 있듯이 각 항목에 검색하고자 하는 키워드를 입력한 후 'Typical Report' 버튼을 누르면 된다. 이때 Drug 는 Generic Name 으로 검색하던지 IUPAC Name, Chemical Composition, CAS Number 로 검색한다. 단일 항목으로 검색하기 위해서는 [그림 1]에서 해당 항목만을 입력한 후 검색하면 된다. 만약 각 항목의 검색 키워드가 잘 생각이 나지 않는다면 입력란 옆의 'select' 버튼을 눌러 [그림 2]의 페이지로 옮긴 후 와일드카드 문자와 조합하여 검색할 수 있다. 즉 현재 데이터베이스에 저장된 것들이 어떠한 것들이 있는지 알 수 있게 된다. 검색 결과는 [그림 3]처럼 나타난다.

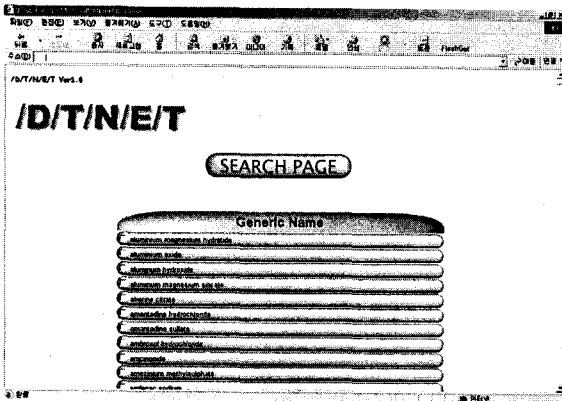


[그림 2] 항목별 키워드 검색

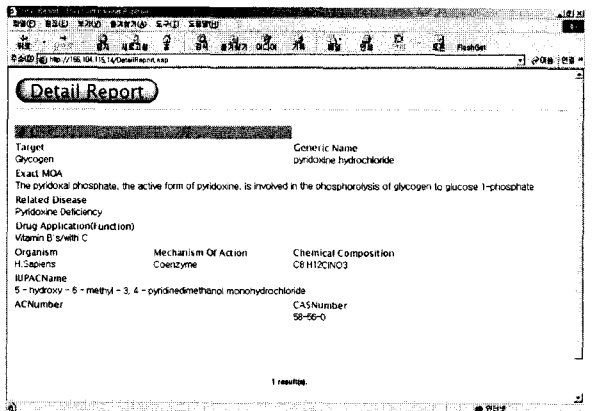


[그림 4] Typical Report

위의 결과 중 데이터베이스에 저장된 다른 항목들을 보고 싶은 레코드가 있으면 해당 레코드에 체크표시를 하고 'Detail Report'를 클릭하면 [그림 5]처럼 그 결과를 출력한다.



[그림 3] 항목별 키워드 검색 결과



[그림 5] Detail Report

다른 한편으로, Drug 와 Target 의 Network 를 규명한 데이터베이스의 특성을 충분히 반영할 수 있게 하기 위하여 각 항목간의 관계를 맺어 검색할 수 있도록 하였다. 이를 위해서는 두 가지 이상의 항목을 입력한 후 검색 버튼을 누르면 해당 항목들이 서로 연관이 된 레코드들만이 검색되어 나온다. 예를 들어 pyridoxine hydrochloride 라는 Drug 와 glycogen 이라는 Target 이 coenzyme 의 Mechanism of Action 관계를 가진 레코드들을 출력하고 싶으면, Generic Name 란에 pyridoxine hydrochloride, Target 란에 glycogen, Mechanism of Action 란에 coenzyme 을 채우고 'Typical Report' 버튼을 클릭하면 된다.

즉, 검색 시 같은 항목간의 조건들은 OR 연산이 되고 다른 항목간의 조건들은 AND 연산이 된다.

Typical Report 의 결과는 [그림 4]처럼 검색조건에 맞는 레코드의 Target, Generic Name, Organism, Mechanism of Action, Exact MOA, Drug Application, Related Disease 항목을 보여준다.

[그림 5]에서 볼 수 있듯이 Detail Report 는 사용자가 지정한 연관 관계의 Target, Generic Name, Exact MOA, Related Disease, Drug Application, Organism, Mechanism of Action, Chemical Composition, IUPAC Name, AC Number, CAS Number 를 출력한다.

3 검색 결과 분류

3.1 검색 결과 분류

검색 결과가 많다면 사용자는 이를 기준이 될 만한 열을 중심으로 분류하여 좀 더 간략하게 보길 원할 것이다. 이를 충족시켜주는 것이 'Classification Report' 이며, 이것이 본 모듈의 가장 큰 장점 중 하나이다.

사용자는 Typical Report 에서 많은 검색 결과를 Generic Name 과 Target 을 기준으로 나머지 항목들을 묶을 수 있다. 아래 [표 1]처럼 Typical Report 에서 검색 결과가 나왔다면 다음 [표 2]처럼 Target 과 Generic Name 이 같은 것끼리는 같은 행에 출력할 수 있는 것이다.

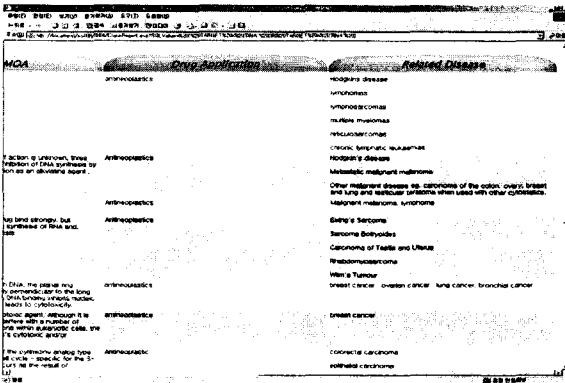
Target	Generic Name	Organism	Exact MOA
A	A	C	E
A	A	D	D
A	C	C	E
R	C	D	A
C	F	C	C
C	F	C	E
A	C	G	E

[표 1] Typical Report 출력 결과

Target	Generic Name	Organism	Exact MOA
A	A	C, D	D, E
A	C	C, G	E
C	F	C	C, E
R	C	D	A

[표 2] Classification Report 분류 결과

아래 [그림 6]은 실제 'Classification Report' 버튼을 눌렀을 때의 결과 화면이다.



[그림 6] Classification Report

3.2 검색 결과 분류 구현 방법

RDBMS 를 가지고 웹 상에서 이를 구현하기 위해서는 몇 가지 고려할 사항이 있다. 웹 어플리케이션 이 가지는 자원의 제약을 극복하는 것과 네트워크 대역폭을 최소화하는 것이 그것이다. 행 중심의 RDBMS 에서 열 중심의 분류를 이루기 위하여 본 모듈에서는 다음과 같은 기법을 사용하였다.

처음 사용자가 지정한 검색조건을 가지고 해당하는 Generic Name 과 Target 을 정렬된 결과 셋으로 받아 이를 배열에 저장한다. 이 배열을 현재 행과 다음 행을 비교하면서 Generic Name 과 Target 의 유일한 순서

쌍 리스트를 얻는다. 다음으로 이 리스트의 Generic Name, Target 순서쌍과 처음 사용자가 지정한 검색조건을 가지고 나머지 각각의 항목에 대하여 SQL 문의 Distinct 처리를 하여 출력한다.

4 결 론

약물 데이터베이스에 대하여 효과적인 검색기법으로 각각의 항목들에 대해 개별적 혹은 조합된 형태의 검색이 가능하고 결과의 개별 출력 및 같은 것끼리는 묶어서 출력할 수 있도록 하였다.

이 모듈은 현재 서비스 중이며 [7] 신약개발에 많은 도움이 될 것으로 보인다.

참고문헌

- [1] Drews, J. (2000) Drug discovery: a historical perspective. Science, 287, 1960-1964
- [2] Ohlstein, E.H., Ruffolo, R.R., Jr and Ellroff, J.D. (2000) Drug discovery in the next millennium. Annu. Rev. Pharmacol. Toxicol., 40, 177-192
- [3] 대한민국의약품정보센터(KIMS), <http://www.ekims.com>
- [4] BIT Druginfo, <http://www.druginfo.co.kr>
- [5] 한국의약품정보센터(KOPAMS), <http://www.kopams.com>
- [6] 숙명여자대학교 의약품연구정보센터(BRIC), <http://dric.sookmyung.ac.kr>
- [7] DTNet, <http://166.104.115.14>