

# 대사경로 탐색용 전문가 시스템 설계

최영상\*, 최한석\*, 양기철\*, 김동욱\*, 조광문\*\*

\*목포대학교 정보공학부

\*\*목포대학교 전자상거래학과

e-mail: choi7040@mokpo.ac.kr

## Design of an Expert System for Pathway Search

Young-Sang Choi\*, Han-Suk Choi\*, Gi-Chul Yang\*,

Dong-Wook Kim\*, Kwang-Moon Cho\*\*

\*Dept. of Multimedia Engineering, Mokpo National University

\*\*Dept. of Electronic Commerce, Mokpo National University

### 요 약

현재 생물정보학에서는 무수히 많은 생물들에서 엄청난 양의 EST(expressed sequence tag) 정보가 나온다. 본 논문에서는 그런 정보들을 가지고 대사경로를 유추할 수 있는 전문가 시스템을 개발하고자 한다. 현재까지의 대사경로 탐색은 발견해 놓은 데이터를 가지고 전부 수작업으로 이루어진다. 그러한 과정을 자동화 시키고자 하는 연구가 많이 진행되고 있다. 본 논문에서는 전문가 시스템을 활용하여 생물정보학의 EST(expressed sequence tag) 데이터에서 효율적으로 대사경로를 찾을 수 있도록 하는 방안을 제시한다.

### 1. 서론

전문가 시스템은 데이터 처리, 계산, 정보 검색이 주 기능 이었던 종래의 컴퓨터 시스템과는 달리 사실과 사실사이의 관계를 나타내는 규칙과 결합을 함으로써 인공지능 형태의 사고를 연는다. 그리고 획득된 지식을 어떻게 효율적으로 표현·저장하는가 하는 것도 전문가 시스템의 필수 구성요소이다. 또한 전문가 시스템을 방법론적 관점에서 보자면 지식 기반 시스템(Knowledge Based System)으로 볼 수 있다.

이러한 전문가 시스템을 바탕으로 미국의 NCBI(The National Center for Biotechnology Information)[1]나 일본의 KEGG(Kyoto Encyclopedia of Genes and Genomes)[2] 등의 엄청나게 많은 EST(Expressed Sequence Tag) 데이터들을 바탕으로 대사경로를 탐색하고 유추하고자 한다. 본 연구에서 참고한 대사경로를 보자면 각각의 링크를 모두 수작업으로 만들었다. 이런 수작업의 과정을 자동화 시키고자 하는 많은 연구가 진행 중이다. 많은 양의 논문과 실험을 바탕으로 대사경로를 구성했다. 현재 연구가 진행되고 있는 것으로 시멘틱 웹 기반으로

논문의 의미검색을 통해 대사경로를 자동으로 찾아 주는 시스템 개발이 연구 중에 있다. 이러한 시스템을 더욱더 효과적으로 대사경로를 찾기 위해서는 전문가 시스템의 이용이 필수적이 된다.

### 2. 전문가 시스템과 대사경로 탐색

전문가 시스템이란 충분히 복잡한 문제 영역에 대하여 전문가의 전문지식을 사용하여 추론을 하여 전문가와 동등한 수준에서 문제를 해결할 수 있는 지적인 소프트웨어를 말한다. 전문가 시스템은 전문가의 지식이 저장되는 지식베이스, 저장된 지식을 이용하여 추론을 하는 추론기관, 지식 습득을 지원하는 시스템, 추론결과를 설명해 주는 설명 시스템, 사실과 추론의 중간 결과를 저장하는 작업 메모리 등으로 구성된다.

전문가 시스템의 대사경로를 응용을 보자면 전세계에서 발표된 수많은 논문들에서의 대사경로를 지식베이스에 자장을 하고, 이러한 논문들을 바탕으로 대사경로를 추론하는 시스템, 대사경로를 설명해 주는 시스템, 일련의 과정까지 발견된 대사경로를 저장하고 새로운 대사경로 발견 시 새로 저장하는 시

시스템이 있다. 이러한 전문가 시스템의 기능을 생물들의 대사경로를 찾고 효율적으로 표현하는 방법에 대하여 연구하고자 한다.

대사경로(pathway)란 세포안에서 발생하는 화학적 반응이다. 대사경로의 원리는 화학적 반응을 통해 계속해서 변화가 된다는 것이다. 이러한 반응은 효소(enzyme)에 의해 빠르게 촉진시켜진다. 식용미네랄, 비타민과 다른 요소들은 효소가 이러한 과정을 하기위해 필요 되어 진다. 대부분의 대사경로를 복잡하다[3].

여러 종류의 신진대사(Metabolic) 대사경로는 세포안에서 신진대사 네트워크를 가진다. 이렇게 반응을 통해서 사람의 몸이 성장하고, 질병이 생기고, 각기 다른 형질이 생겨나고, 식물의 경우 동일한 식물이지만 지역에 따라 성분이 다르게 나타난다. 이러한 대사경로는 너무도 복잡하고 아직까지 발견되지 않는 것이 너무나도 많다, 이러한 현상들을 체계적으로 정리를 해 놓은 것이 대사경로 지도(pathway map)이다. 이러한 대사경로를 통해서 식물의 약효를 만들어내는 반응을 찾을 수 있고, 각 지역마다 그 식물만의 특성이 있어서 원산지 판별이 가능하고, 식물이 성장하는 과정에서 상하는 반응을 찾아내서 상하지 않게도 하다. 사람의 경우 사람의 성장 및 외형 체질을 다르게 만드는 반응을 찾아낼 수 있고, 질병을 일으키는 반응을 찾아내서 바꿀 수가 있게 된다. 이렇게 탐구되고 연구되어진 대사경로를 체계적으로 정리하고, 연구된 논문들에서 쉽게 찾아낼 수 있는 방법을 전문가 시스템을 통해 구현하고자 한다.

3. 관련 연구

현재까지 일본 KEGG의 대사경로는 대사경로 이미지를 만들어서 단순히 링크만 걸어서 대사경로를 찾아가게 되어 있다. 그림 1이 KEGG에서 정보를 제공해주는 대사경로 지도이다. 그림을 보자면 KEGG에서 현재까지 발견한 경로를 모두 제공해 주고 있다. 효소들의 반응을 상세하게 보여주고 있다. 이러한 반응들이 사람의 몸과 식물의 몸을 변화시키는 요인이다.

이러한 대사경로 지도(pathway map)를 체계적으로 관리하고 표현하고자 하는 많은 연구가 진행되어지고 있다. 그중에서 Joanne S Lucianop가 연구한 논문을 보자면 시멘틱 웹 기반에서 대사경로 지도 표현기법에 대해 제시했다[4]. BioPAX(Biological

pathway Exchange)라는 시스템은 시멘틱 기반으로 웹에서 대사경로를 쉽게 접근하고 검색할 수 있는 시스템이다.

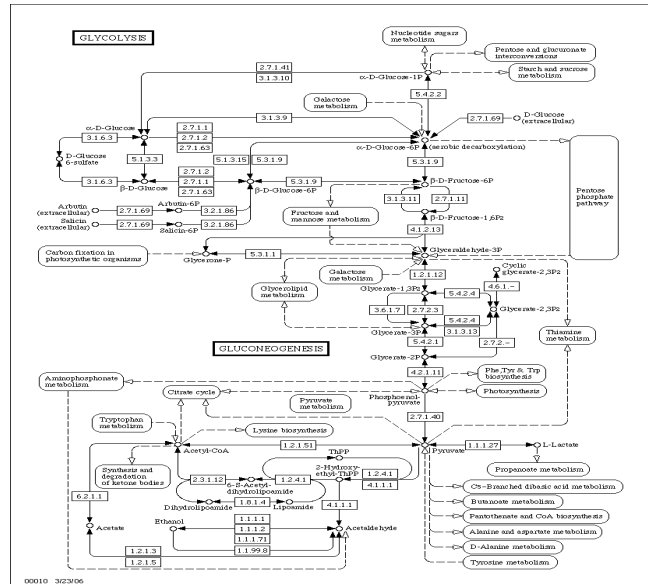


그림 1. KEGG의 대사경로 지도

Kinase Pathway Database는 일본의 동경대학에서 제공해 주는 대사경로 검색 시스템이다[6]. 이곳에서는 대사경로 검색, 온톨로지 데이터 검색, 단백질 구조 검색의 기능을 제공해 주는 곳이다. 대사경로를 검색할 수 있게 해 주는 것에서 이 시스템이 우수하다. 한 단백질에서 많은 단백질의 작용들이 있는데 이러한 과정을 한 단계씩 검색이 가능하게 해 놓았다.

대사경로는 문서를 바탕으로 제작이 된다. 표 1은 텍스트 기반에서 대사경로를 찾아내는 시스템과 데이터베이스를 나타낸다[5, 7, 8, 9].

표 1. 텍스트 기반 대사경로 시스템 사이트들

이름	설명	주소
AbGene	Protein name tagger	<a href="http://ftp.ncbi.nlm.nih.gov/pub/tanabe">http://ftp.ncbi.nlm.nih.gov/pub/tanabe</a>
NLProt	Protein name tagger	<a href="http://cubic.bioc.columbia.edu/service/nlprot">http://cubic.bioc.columbia.edu/service/nlprot</a>
PubMed	updating and alerting service	<a href="http://www.pubgene.org">http://www.pubgene.org</a>

논문 및 실험에 의한 대사경로를 나열해 주는 시스템들은 세계적으로 많은 연구가 되었다. 표 1에

제시되지 않는 많은 사이트 들이 존재하지만 본 연구와 가장 관련이 있는 사이트만 나열하였다.

**4. 대사경로 탐색을 위한 전문가 시스템 구조**

온톨로지 기반의 시맨틱 웹으로 표현을 하면서 생물학자의 전문가적인 지식과 전산학자들의 전문가적인 지식을 통합해 대사경로 탐색을 위한 전문가 시스템을 설계하고자 한다.

많은 생물정보학자들의 연구에 의해 대사경로 지도를 표현하는 많은 시스템들이 나왔지만 생물학자의 전문가적인 지식과 전산학자들의 전문가적인 지식으로 설계된 시스템이 거의 없는 실정이다.

본 논문에서 제안하는 대사경로 탐색을 위한 전문가 시스템의 구조는 그림 2와 같다.

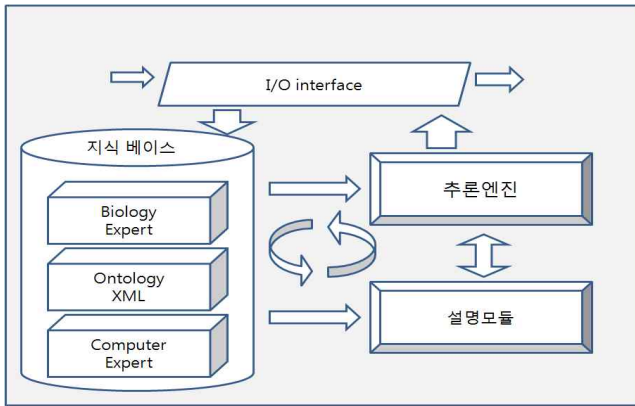


그림 2. 대사경로 탐색을 위한 전문가 시스템

시스템 세부 구조에 대한 설명을 하자면 입력으로는 여러 단백질명을 넣는다. 시작하는 단백질과 대사경로 과정을 알고 싶은 단백질명을 입력한다. 그러면 결과 값으로 어떠한 효소(enzyme)에 의해 몇 번의 과정을 거쳐서 최종단백질이 나오는 모습이 이미지 형식으로 나타나게 된다. 현재의 과정은 모든 단계마다 단백질 이름을 입력을 하여 한 단계씩 보여지게 되지만 본 시스템에서는 해당하는 모든 과정이 나오고 필요에 따라서 원하는 단계만 보여주는 기능도 있다. KEGG의 대사경로를 예로 설명을 하자면 KEGG의 모든 대사경로는 수작업으로 이루어진 결과이다.

대사경로가 알고 싶은 단백질을 그림 3과 같이 입력을 하게 되면 그림 4와 같이 대사경로가 나타나게 된다. 그림 4는 한 페이지로만 결과물이 보이지만 실제 시스템에서는 원하는 부분까지 보이게 된다. 단백질 Acetyl-Coa에서부터 Decanoly-Coa까지의

모든 경로가 보이게 된다.

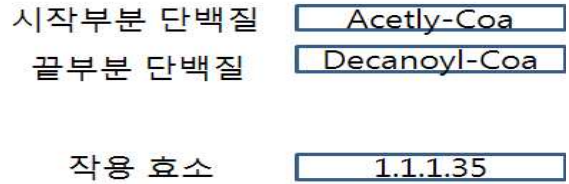


그림 3. 단백질 입력 부분

지식베이스 부분은 데이터 부분은 동일하므로 온톨로지 와 XML로 이루어진 부분의 데이터를 사용하고 생물학의 전문지식과 컴퓨터 전문지식을 저장하는 부분이다. 각 전문지식과 기존 온톨로지와 XML로 되어 있는 부분은 혼합이 되어서 각 추론시스템과 설명시스템으로 전송이 된다.

추론엔진에서는 지식베이스에서 나온 결과를 바탕으로 최종결과물들의 모든 과정을 계산한다. 모든 과정을 전문가적인 지식으로 단계별로 계산을 진행한 후 설명모듈로 넘어가 다시 그 과정에 대한 설명을 하고 결과 값이 부족하다면 다시 지식베이스에서 데이터를 다시 가져온다. 이러한 과정을 거치고 나면 최종 결과 값이 인터페이스에 의해 사용자에게 보여지게 된다.

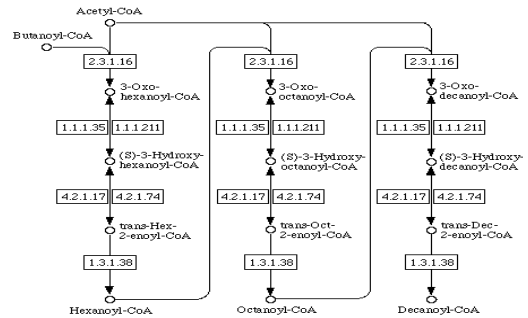


그림 4. 검색 결과 일부 모습

그림 4에서는 작은 원이 단백질이고 네모난 칸이 효소이다. 단백질에서 단백질로 변하는 과정에서 효소에 의한 모든 작용이 전부다 보여주었지만 본 시스템에서는 필요한 부분까지만 보여준다. 시작 단백질에서 끝나는 단백질까지 일련의 모든 과정들을 사용자가 원하는 만큼만 결과 값으로 보여지게 된다.

본 연구에서 제안하는 전문가 시스템은 사실로만 이루어진 기존의 시스템을 기반으로 전문가의 지식을 추가해서 새로운 시스템을 만드는 것이다. 단순 사실만의 결과에서 대사경로를 찾아내고 대사경로

지도를 제작하는데 생물전문가의 관점과 전산 전문가의 관점이 모두 포함된 그런 시스템이다. 전문가적 관점이 들어간 시스템은 초보자가 사용하더라도 그 분야에서 전문가의 관점으로 결과를 확인할 수가 있게 된다.

## 5. 결론

대사경로 탐색용 전문가 시스템은 대사경로 탐색 시 각 대사경로 맵을 작성할 때 일반적인 관점이 아닌 전문가 관점을 부여해서 대사경로를 탐색하는 것이다. 전문가의 지식으로 사용자가 원하는 값까지만 계산을 해서 결과물로 보여 준다. 생물정보학 입문자들도 쉽게 생물 및 전산 전문가의 지식에서 탐색하고, 결과물을 연구할 수 있게 된다. 단백질의 이름만 입력하게 되면 효소와의 반응과 반응 후 결과까지 나오게 되는 시스템이다.

## 참고문헌

- [1] <http://www.ncbi.nlm.nih.gov/>
- [2] <http://www.genome.jp/kegg/pathway.html>
- [3] [http://en.wikipedia.org/wiki/Metabolic\\_pathway](http://en.wikipedia.org/wiki/Metabolic_pathway)
- [4] Joanne S Luciano "e-Science and biological path-way semantics"
- [5] Martin krallinger "Text-mining and information-retrieval services for molecular biology"
- [6] <http://kinasedb.ontology.ims.u-tokyo.ac.jp/>
- [7] <http://ftp.ncbi.nlm.nih.gov/pub/tanabe>
- [8] <http://cubic.bioc.columbia.edu/service/nlprot>
- [9] <http://www.pubgene.org>