

플라보노이드 정보의 효율적인 검색을 위한 그리드 웹 포탈*

서해국^o 전세길 정갑주 김현명 정선호
건국대학교 생명/분자 정보학 센터^o

{haegook^o, sgjeon}@imc.konkuk.ac.kr jeongk@konkuk.ac.kr
swisdom@empas.com shjung@konkuk.ac.kr

Grid Web Portal for the Efficient Retrieval of Flavonoid Information

Haegook Seo^o Segil Jeon Karpjoo Jeong Hyunmyung Kim Seunho Jung
Bio/Molecular Informatics Center, Konkuk University^o

요 약

최근 생물학 문제를 해결하기 위하여 생명 정보 분야에서 많은 컴퓨터 기술들이 응용되고 있다. 플라보노이드 정보는 항암, 항산화 효과가 있는 유용한 물질인 플라보노이드의 연구와 관련된 정보를 분석하고 관리하는 컴퓨터 기술 기반의 응용 기술이다. 현재 플라보노이드 정보 관련 연구는 일부에서 수행되고 있으며 플라보노이드의 다양한 정보를 효율적으로 저장 검색 할 수 있는 시스템은 부족한 실정이다. 플라보노이드 물질에 관한 정보를 더욱 효과적으로 저장 관리하고 관련된 연구자들 간에 정보를 쉽게 공유할 수 있도록 본 논문에서는 플라보노이드 정보를 위한 그리드 웹 포탈 시스템 구조와 관련된 데이터 모델을 제안한다. 본 논문에서 제안한 시스템과 데이터 모델의 유용함을 보이기 위해 웹 기반으로 플라보노이드 정보 그리드 포탈을 구현한다.

1. 서 론

플라보노이드는 나무나 꽃의 잎, 줄기, 또는 과일 등 식물의 여러 부분에 함유되어 있는 천연의 항균 물질이다. 플라보노이드는 식물을 외부 자극에서 보호하고 색소의 형성, 항균물질 등의 기능을 할 뿐만 아니라 사람에게 있어 항균작용, 항바이러스작용, 항산화작용, 항암작용, 항 방사작용과 같은 많은 생리화학적 활성을 가지고 있다[1]. 플라보노이드에 대한 다양한 연구가 활발히 수행되고 있지만 현재까지 플라보노이드 정보(플라보노이드 분류, 생화학 생물학적 기초 정보, NMR 구조, 응용 분야, 문헌 데이터)에 관한 종합적으로 데이터베이스화한 사례는 전무한 상태라고 볼 수 있다.

본 논문에서는 플라보노이드 정보의 효율적인 저장 및 검색을 위한 웹 기반의 그리드 포탈 시스템 아키텍처를 제안한다. 본 논문에서 제안한 시스템의 목적은 플라보노이드 실험 관련 연구자들에게 그리드 미들웨어 기반 웹 포탈을 이용하여 플라보노이드 데이터베이스에 대한 쉬운 접근을 제공하고 연구자들 간의 정보 공유를 통해 효율적으로 협력 할 수 있는 환경을 제공하는 것이다.

플라보노이드의 정보는 다양한 형태의 정보를 포함하고 있으며 크게 완전(complete) 데이터와 불완전(incomplete) 데이터로 분류 할 수 있다. 완전 데이터는 일반적인 데이터 형식을 따르는 정형 데이터로 볼 수 있고, 불완전 데이터는 기존의 데이터 형식으로 표현할 수 없고 계속적으로 변화하는 데이터이다. 본 논문에서는 완전 데이터의 경우 복잡한 유형의 데이터 검색을 요구하는 경우가 많으므로 관계형 모델을 사용하고 NMR 스펙트럼(Nuclear Magnetic Resonance spectroscopy)

정보와 같은 불완전 데이터는 확장가능한 반구조 형태의 문서 형식인 XML 형태로 표현하는 혼합형 데이터베이스 모델을 사용한다. 본 논문에서 제안한 혼합형 데이터 모델과 그리드 웹 포탈 기술을 플라보노이드 연구 분야에 적용하면 분산된 플라보노이드 정보에 관한 공유의 어려움을 해결 할 수 있으며, 관계형 데이터베이스 모델과 XML 모델의 혼합형 모델은 복잡한 질의에 대한 처리와 NMR과 같은 불완전 데이터에 대한 처리가 모두 가능하여 효율적으로 정보를 검색할 수 있다. 본 논문에서는 또한 제안한 시스템의 유용성을 보이기 위한 사용자 인터페이스를 제시하고 웹을 기반으로 프로토타입을 구현한다. 본 논문의 구성은 다음과 같다. 제 2장에서는 관련연구에 대해서 기술하고 3장에서 플라보노이드 정보와 웹 포탈 시스템에 대해서 설명한 후 4장에서 사용자 인터페이스와 구현에 대해서 기술한다. 제 5장에서는 결론 및 향후연구에 대해서 기술한다.

2. 관련연구

현재 다양한 분야에서 플라보노이드에 대한 연구가 활발히 진행되고 있다. 또한, 이러한 연구 결과를 데이터베이스화 하려는 작업도 꾸준히 모색되고 있으며, KEGG 와 USDA 플라보노이드 데이터베이스가 그 대표적인 사례이다. KEGG(Kyoto Encyclopedia of Genes Genomes)는 유전자나 분자 사이의 상호작용으로 구성되는 정보전달 경로를 이용해 분자와 세포의 생물학 지식을 전산화한 웹사이트로 플라보노이드 관련정보는 대표적인 생합성 경로 시스템과 플라보노이드의 구조 및 반응 경로에 대한 정보를 일부 포함하고 있다[2]. USDA(United States Department of Agriculture's) 플라보노이드 데이터베이스는 미 농무부 산하 농업 연구소의 데이터베이스 시스템으로서 다른 플라보노이드 화합물에 대한 항량 분석 결과만을 대상으로 정보를 제공하고 있다[3]. 플라보노이드가 이차대사 산

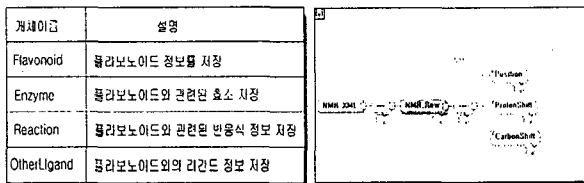
*본 연구는 학술진흥재단의 중점연구소 사업(KRF 2004-F00019)의 지원에 의해 수행되었음.

물로서 가지고 있는 중요한 생리활성들에 관한 정보는 거의 전무한 상태이다. 본 시스템은 KEGG의 기존 플라보노이드 관련 데이터를 포함한 기초에서 추가적으로 NMR 정보와 MS 구조 파일, 분자 구조 파일을 3차원 디스플레이, 플라보노이드 응용 분야 및 관련된 문헌 데이터베이스 시스템을 구축하고 효율적인 검색 방안을 제시한다. 그리드스피어(Gridsphere)는 독일 그리드 랩 프로젝트를 통해 OGSA(Open Grid Service Architecture) 그리드 서비스를 기반으로 개발되어진 웹 포털 어플리케이션 개발 툴이다[4]. 그리드스피어는 포털 개발자들이 쉽고 빠른 웹 포털 개발을 위한 프레임워크를 제공하며 또한 그리드 랩의 GAT API(Grid Application Toolkit)를 이용하여 그리드 리소스 자원에 접근할 수 있다. 본 논문에서 제안한 시스템 구조에서는 그리드스피어 프레임워크를 이용함으로써 위와 같은 장점을 포털 개발자들에게 제공하게 된다.

3. 플라보노이드 정보 웹 포털 시스템

3.1 정보화 요구 사항 및 데이터 모델

플라보노이드에 관한 연구가 활발히 진행됨에 따라 발생하는 대량의 연구 결과를 체계적으로 저장하고 검색하기 위한 요구가 증가하였다. 이런 정보들에는 플라보노이드와 관련된 생화학적, 생물학적 기초자료, 단백질 3차원 구조 규명을 위한 NMR 데이터, 생합성 관련 새로운 유전자 정보, 기존 선별된 플라보노이드의 건강기능 효과 (항암, 항산화, 항균, 신경계질환 예방, 콜레스테롤 저하 효과 등) 및 기존 선별된 플라보노이드 물질의 화학 및 물리적 속성 (용해도, 안전성, 전기화학적) 등이 포함된다. 이러한 플라보노이드 정보의 경우 복잡한 질의를 요구하며 비정형 데이터와 불안정한 특성을 포함하고 있다. 본 논문에서는 플라보노이드 정보의 다양성을 고려하여 데이터 모델을 설계 할 때 관계형 데이터베이스 모델과 XML 데이터베이스 모델을 혼합하여 사용하였다.



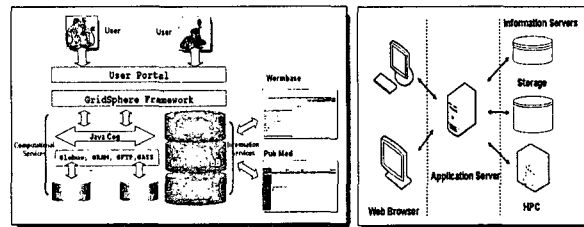
(a) 관계 모델의 개체 리스트 (b) NMR XML 스키마
그림 1. 플라보노이드 혼합형 데이터 모델

그림 1 (a)에서 관계형 모델은 복잡한 유형의 데이터 검색에 사용 되었으며 문헌 검색, 플라보노이드 정보에 관한 고급 검색에 사용 되었다. 그림 1 (b)에서 XML 모델은 NMR 스펙트럼 정보와 같은 불안정 데이터의 효율적인 검색에 사용하였다.

3.2 그리드 웹 포털 시스템 및 데이터 검색

그리드 포털은 다양한 목적을 가지고 구축되어 사용되고 있다. 플라보노이드 포털은 범용적 목적으로 만들어진 것이 아니고 플라보노이드 정보공유 및 검색을 위한 어플리케이션 포털이다. 또한, 플라보노이드 연구자들에게 정보 공유를 통해 효율적으로 협력 할 수 있는 환경을 제공 하는데 목적이 있다.

그림 2(a)에서와 같이 본 논문에서 제안한 시스템은 그리드 웹 포털, 인포메이션 서비스, 계산(computation) 서비스 세 부분으로 구성 된다. 이 구조에서 그리드 웹 포털은 통합 사용자 인터페이스를 이용하여 서로 다른 소스로 부터 제공되는 콘텐츠를 통합 해주는 웹 어플리케이션이다. 그리드 웹 포털의 기능은 야후, 네이버 등의 기존 웹 포털과 의미상 유사하며, 연구자들과 과학자들에게 가상 커뮤니티를 지원하며 사용과 설치가 편리한 장점이 있다. 또한, 그리드 어플리케이션에 쉽게 접근하여 작업을 실행할 수 있는 환경을 제공해 주는 것을 목적으로 한다. 본 논문에서 제안한 시스템 구조에서는 그리드스피어 포털 프레임워크를 사용하여 그리드 포털 시스템을 구축하였다.



(a) 전체 시스템 구조 (b) 포털 3-tier 구조
그림 2. 전체 시스템 구조 및 포털 구조

그리드스피어는 포털 개발자들이 쉽고 빠르게 웹 포털 개발을 할 수 있도록 프레임워크를 제공하며 OGSA 그리드 서비스를 기반으로 개발된 포털 웹 어플리케이션 툴이다. 이러한 그리드 포털 시스템은 그림 1(b)에서와 같이 3-tier 구조로 구성되어 있다. 첫번째 층은 사용자 웹 브라우저이며, 두번째 층은 웹 어플리케이션 서버이다. 본 논문의 시스템에서는 어플리케이션 서버로는 톱캣을 사용하고 그 위에 그리드스피어 포털 프레임워크를 설치하고 실행 하였다. 그리드스피어 포털은 자바 기반의 웹 구성 요소인 포틀릿(portlet), 포틀릿들의 콘텐츠로 구성된 포털 및 포틀릿에 런 타임 환경을 제공하고 관리하는 포틀릿 컨테이너로 구성 되어있다. 세번째 층은 리소스 자원을 의미하는데 대용량 데이터 저장소 및 컴퓨팅 리소스 자원 등이 있을 수 있다. 그리드스피어 기반의 본 시스템은 MVC 디자인 패턴을 사용하여 개발 하여 비즈니스 로직과 프리젠테이션 부분을 분리 하므로 유지보수와 확장성이 용이하다. 또한, 그리드 포털은 Single Sign-on으로 글로버스(Globus)로 묶여져 있는 여러 기관의 자원에 접근 할 수 있으며 사용자로 하여금 자신만의 뷰(view)를 정의 할 수 있어 필요에 따라서 페이지를 만들 수 있다[5]. 인포메이션 서비스는 문헌 데이터베이스, 플라보노이드 정보 데이터베이스, NMR XML 데이터베이스로 구성 되어있다. 문헌 데이터베이스는 플라보노이드 관련된 논문정보 저장 및 검색을 위한 것이며, 플라보노이드 정보 데이터베이스는 플라보노이드 관련 생화학적, 분자 생물학적 기초자료 저장 및 검색을 위한 것이다. NMR XML 데이터베이스는 단백질 의 3차원 구조를 알아내기 위한 정보인 NMR 데이터 정보의 저장 및 검색을 위한 데이터베이스이다.

계산 서비스는 그리드 리소스 자원을 접근하고 사용

할 수 있는 서비스를 제공해 준다. 글로벌스 툴킷을 사용하여 글로벌스로 묶여 있고 이질적으로 분산되어 있는 각 기관들의 리눅스 클러스터 또는 슈퍼 컴퓨팅 자원에 접근하고 사용할 수 있어 쉬운 정보의 공유와 공동 연구가 가능해진다.

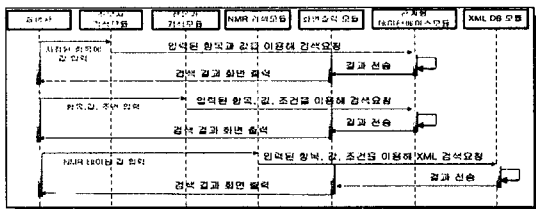
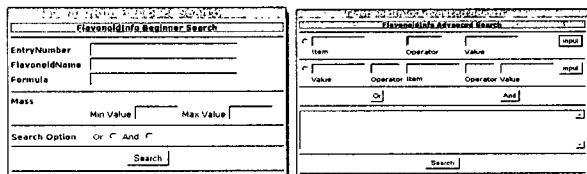


그림3. 데이터 검색 흐름도

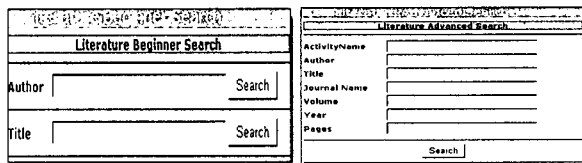
그림3은 본 논문에서 제안한 모델의 데이터 검색 흐름도이다. 이 흐름도는 플라보노이드 및 문헌 정보에 관한 초보자, 전문가 검색 모듈과 NMR 스펙트럼 값 검색을 위한 XML 검색 모듈의 흐름도를 보여주고 있다.

4. 사용자 인터페이스 및 구현

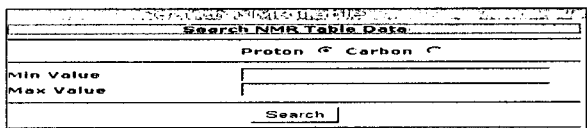
본 논문에서 구현한 사용자 인터페이스는 데이터 입력, 데이터 화면출력 및 데이터 검색 세 부분으로 구성되어 있다. 데이터 입력은 플라보노이드 기본 정보, 효소, 반응관계, 기질, NMR 데이터 및 활용(activity)등에 관련된 정보를 입력 한다. 효소 정보는 플라보노이드 물질의 생성 및 합성과 관련된 효소의 정보이며, 반응 관계는 반응전과 후에 참여하는 플라보노이드와 기질들의 정보 및 관련된 효소들의 정보를 입력한다. 활용은 플라보노이드와 관련된 응용분야 및 논문 정보를 입력하는 것이다.



(a) 플라보노이드 정보 검색



(b) 문헌 검색

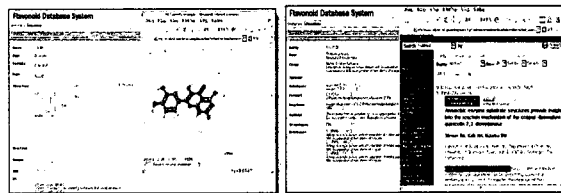


(c) NMR 테이블 정보 검색

그림 4. 플라보노이드 그리드 웹 포탈 검색 인터페이스

데이터 검색은 플라보노이드 정보의 다양성을 고려하여 관계 모델과 XML 데이터 모델을 사용하도록 설계하

였다. 검색을 위한 인터페이스는 검색 속력 정도에 따라 초보자와 전문가 검색 두 가지 검색방법을 제공하고 있다. 그림 4(a)에서와 같이 플라보노이드 정보 검색에서는 물질의 이름, 분자식 및 분자량의 범위에 근거하여 플라보노이드 정보를 검색할 수 있다. 그림4 (b)에서 문헌 검색은 저자 이름, 발표 년도 등 여러 가지 옵션으로 검색을 할 수 있으며 모든 검색 결과는 모두 PubMed 사이트와 링크되어 있다. 그림4 (c)에서 NMR 스펙트럼과 같은 불완전 데이터는 XML 모델로 설계되었으며 XPath 질의를 이용한 NMR 테이블에서 proton, carbon 값의 범위를 이용하여 관련된 플라보노이드 정보 및 NMR데이터 정보를 검색할 수 있다.



(a) 플라보노이드 정보 (b) PubMed 논문 링크
그림5. 플라보노이드 정보와 관련 논문 검색 결과

그림 4(a)의 실행결과는 그림 5(a)와 같으며 플라보노이드 정보를 화면 출력하는 부분에서는 플라보노이드의 기본 정보, 관련효소 정보 및 관련 반응 관계를 보여준다. 플라보노이드 정보 디스플레이에서는 Jmol API를 이용하여 구현된 물질의 3차원 구조를 확인 할 수 있으며, 그림4 (b)의 인터페이스를 이용해 논문 검색한 결과는 그림5 (b)와 같으며, 플라보노이드 관련 논문들을 PubMed 혹은 PDB 등 사이트에 있는 논문들과 링크되어 있어 관련된 논문도 쉽게 확인 할 수 있다[6,7].

5. 결론 및 향후 계획

본 논문에서는 플라보노이드와 관련된 정보화 시스템을 구축하여 응용 과학자들이 정보를 공유하고 공동 연구를 할 수 있는 그리드 웹 포탈 환경을 제공 하였다. 본 논문은 시스템 아키텍처와 데이터 검색에 중점을 두고 연구 하였으며, 플라보노이드와 관련된 데이터는 KEGG에서 제공하는 모든 데이터를 포함하여 입력되어 있다. 향후 연구과제로는 완전한 그리드 기반의 시스템을 구축하여 쉬운 그리드 리소스 자원에 접근하고 정보를 공유할 수 있게 하며, 문헌 데이터베이스는 오토로지 형식의 데이터 시스템을 구축 하여 쉽고 빠른 관련 논문 검색을 할 수 있게 하는 것이다.

참고문헌

[1] Havsteen, B.H., "The biochemistry and medical significance of the flavonoids," *Parmacol. Ther.* 96, 67-202, 2002
 [2] KEGG, <http://www.genome.jp/kegg>
 [3] USDA, <http://www.nal.usda.gov/fnic/foodcomp>
 [4] GridSphere, <http://www.gridisphere.org>
 [5] Globus, <http://www.globus.org>
 [6] Jmol, <http://jmol.sourceforge.net/>
 [7] PubMed, <http://www.ncbi.nlm.nih.gov/>