A Grid-Based Flavonoid Informatics Portal*

HaiGuo Xu¹, Karpjoo Jeong^{2,**}, Seunho Jung³, Hanku Lee², Segil Jeon⁴,
KumWon Cho⁵, and Hyunmyung Kim³

¹ Department of Advanced Technology Fusion, Konkuk University, Seoul, Korea
haegook@ricl.konkuk.ac.kr

² School of Internet and Multimedia Engineering, Konkuk University, Seoul, Korea
(jeongk, hlee)@konkuk.ac.kr

³ Department of Microbial Engineering, Konkuk University, Seoul, Korea
shjung@konkuk.ac.kr, swisdom@empal.com

⁴ BioMolecular Informatics Center, Konkuk University, Seoul, Korea
sgjeon@ricl.konkuk.ac.kr

⁵ Korea Institute of Science and Technology Information
ckw@kisti.re.kr

Abstract. Recently new techniques to efficiently manage biological information of biology have played an important role in the area of information technology. The flavonoids are members of a class of natural compounds that recently has been the subject of considerable scientific and therapeutic interest. This paper presents a Grid-based flavonoids web portal system. We designed relational schema, XML schema for flavonoids information and their user interfaces and proposed interoperable web service components for an efficient implementation of flavonoids web portal.

1 Introduction

The need for efficient management of biological information of biology on complex and various data is rapidly increasing in the area of information technology. Flavonoids are a class of plant pigments and found in a wide range of foods. Recent interests by the biochemical community in flavonoids information are dramatically increased [1, 2, 3]. These include antioxidative, antiallergic, anticarcinogenic effects etc. To address these needs a database and user-friendly application of the flavonoids information (i.e. flavonoid name, mass, structure, NMR, activity, related literature etc.) was developed. Today flavonoids information system deals with many different data types. But legacy database systems can't properly manage (i.e. modeling, storing, and querying) flavonoids information that has been recognized as a key component of today's flavonoids research.

Grid web portals make it possible to provide seamless access to heterogeneous information via a browser-based user interface. It provides a complete set of open, productive, and self-service tools for publishing information, building applications, and deploying and administering enterprise portal environments. These portals are typi-

* This work was supported by the Bio/Molecular Informatics Center at Konkuk University (KRF2004-F00019).

** Corresponding author. Ph: 82-2-450-3510.

cally built around several services including data management and user session management. These services may be built on top of Grid technologies such as Globus [4]. However, developing applications utilizing the Grid technologies remains very difficult due to the lack of high-level tools to support developers. Web service applications have recently been popularized through various domains. The development of web services protocols such as the Simple Object Access (SOAP), Web Services Description Language (WSDL) and the Universal Description, Discovery and Integration (UDDI) protocol has simplified the integration and the interaction between other organizations [5, 6, 7]. However, existing web services protocols have not been designed for the use in science experimental domain or knowledge intensive infrastructures. The representative databases for flavonoids are the United States Department of Agriculture's (USDA) flavonoids database that provides the information of the flavonoids composition in food, and the part of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database that provides the information of some flavonoids related to flavonoids biosynthesis [8, 9, 10, 11]. However, any well-defined flavonoids database systems for managing of research results and related knowledge (e.g. flavonoids information, literatures etc.) are not systematically well-developed yet. An important shortcoming of existing databases is their lack of interoperability and reusability.

In this paper, we propose a data model for storing and retrieving of the flavonoids information, and design and develop a Grid-based flavonoids web portal system. We also propose web service components for efficient implementation of the portal. The main goal of the flavonoids web portal is to collaboratively work with flavonoids researchers, to provide reusable services, to provide hybrid data model for flavonoids information, and to show how effectively flavonoids information is shared and retrieved in a Grid web portal environment using web services. By adopting grid web portal technology into flavonoids research domain, we can overcome the difficulties caused by a large amount of distributed and shared flavonoids information. Moreover, compared to legacy querying models, the web portal can more effectively support complex queries and incomplete data (i.e. NMR spectroscopy data) by combining ER-model and XML-model.

2 Related Works

The flavonoids are members of a class of natural compounds that recently has been the subject of considerable scientific and therapeutic interest. These include antioxidative, antimicrobial, anticarcinogenic, and cardioprotective effects [12]. To represent these effects, flavonoids database systems started to develop. But each system only works for specific queries and data. So far only a few databases have developed. The flavonoids database developed by USDA provides the contents of flavonoids. KEGG pathway database provides the information of biosynthesis for flavonoids. KEGG mainly deals with the ligand of an enzyme involved in biosynthesis, whereas the physiological properties and chemico-physical characteristics were neglected. But these databases are inadequate for the fields of natural and medical science. Consumption of flavonoids, biologically active polyphenolic compounds found in plants,

has been associated with decreased risk for some age-related and chronic diseases in humans. KEGG consists of three databases: PATHWAY for representation of higher order functions in terms of the network of interaction molecules, GENES for the collection of gene catalogs for all the completely sequenced genomes and some partial genomes, and LIGAND for the collection of chemical compounds in the cell enzyme molecules and enzymatic reactions. To the further investigation of flavonoids intake and health, the USDA published the database for the flavonoids content of selected foods in 2003.

3 The Proposed Flavonoids Grid Web Portal Model

3.1 Data Model

The proposed flavonoids data model is a hybrid model based on Entity Relationship model (i.e. ER-model) and XML-model. We have used ER-model for flavonoids information and XML-model for chemical shift table of NMR spectroscopy. This hybrid data model supports complex query and new types of flavonoids information such as NMR for flavonoids information. Flavonoids information and NMR spectroscopy are two major data types in hybrid data model. Flavonoids information represents a single flavonoids compound such as name, formula, mass, structure, related reaction, related enzyme, NMR, MS, references, etc. NMR spectroscopy, described by XML syntax, represents chemical shift table information for determining the content of a flavonoids as well as its molecular structure.

Fig. 1. E-R diagram for flavonoids information

Figure 1 represents the E-R diagram for the complete flavonoids data model. In E-R diagram, the NMR field is a pdf (i.e. portable document format) file name of the NMR structure, the MS field a pdf file name of the MassSpectroscopy structure, the Coordinate field a text file for the model in the PDB site format, and the Figure field a structure image file name, the PubID field the paperID of the PubMed site.

Fig. 2. An example of a XML Schema of NMR spectroscopy

We have designed a XML schema for NMR chemical shift data. The XML model provides the extensibility for representing incomplete information. Figure 2 shows a textual version and a tree representation of this XML schema. The NMR_XML element is the root element of the XML schema that consists of zero or more NMR_Row elements. The NMR_Row element represents rows with zero or more NMR_Column elements. The NMR_Column element represents columns with zero or more Position, ProtonShift, and CarbonShift elements. The Position element, the ProtonShift element, and the CarbonShift element represent a position value, a protonshift value, and a carbonshift value, respectively, in the chemical shift table.

3.2 The Architecture of Flavonoids Grid Web Portal

The architecture of flavonoids Grid web portal mainly consists of grid middleware, portal, web services and database. Figure 4 shows the architecture of flavonoids Grid web portal. The information service provides metadata of flavonoids. The information service of the proposed web portal consists of three databases: *Literature database*, *Flavonoids content database*, and *NMR XML database*. Literature database stores related journal articles, commercially published books, other references, etc. The proposed literature database provides links of the existing literature databases via the Internet. Flavonoids content database stores KEGG information as well as new information such as NMR, MS, activity of flavonoids, 3D structure, etc. NMR XML database is an XML database for storing of the NML chemical shift table information.

The computational service provides a web-based grid computing environment for molecular simulation. Flavonoids researchers require high-performance computing resources via secure grid web portal interfaces. We have adopted a molecular simulation computational Grid System called MGrid into these computational services. MGrid (the Molecular simulation Grid system) is a very promising research technique for biochemical research areas that inevitably need high-performance environments. For example, simulations for bio-conjugates of protein, DNA, lipid, and carbohydrates definitely need HPC environments. Computational power can't solve the whole problem. It is very difficult to know appropriate simulation settings in advance. Moreover, simulation outcomes of those molecules with three-dimensional structures

are difficult to validate without real experiments. These two critical problems, the requirements of computational power and the validation of simulation outcomes, have decreased the popularity of the molecular simulations. The MGrid system is designed to address these two issues and based on distributed molecular simulation architecture. Currently, the MGrid system is being ported to the test bed (<http://www.mgrid.or.kr>).

Fig. 4. Overall System Architecture

Generally, a portal is consisted of three components: portal, portlet, and portal container. The GridSphere portal framework [13] is installed and executed on the top of the Apache Tomcat that is a servlet container. The GridSphere is a standard portlet framework to build web portals. It is based on advanced web technology, web security and meta-computing technology such as PKI and Globus to provide secure and interactive services.

4 Implementation of Flavonoids Grid Web Portal Using Web Services

We developed flavonoids web services supporting flavonoids insertion and retrieval web services that are a threshold to the GridSphere portlet container. Each web service has only one interface in WSDL, so any flavonoids research groups and developers who want to use the web services can directly implement their own interface in WSDL.

The workflow of the web services is divided into 7 steps as shown in Figure 5. Step 1 uses a Java2WSDL compiler transforming a Java remote interface into a WSDL document. Step 2 uses a WSDL2Java compiler generating tie-based skeletons (i.e. you can extend a tie class and add code to your subclass that sets the target for delegation to be your implementation class). Step 3 and 4 execute insertion or re-

trieval portlet services through the web browser. Then, at next steps, the GridSphere portlet container will manage those portlet applications and call specific web services. Finally, data will be inserted into one of databases or retrieved from it.

Fig. 5. Workflow of Web Services for Flavonoids

4.1 Web Services for Insertion and Retrieval of Flavonoids Information

We implemented flavonoids web services for storing and retrieving flavonoids information. The services consist of a flavonoids service, an enzyme service, a reaction service, a literature service, and a NMR data insertion web service. The web services for the insertion of flavonoids and enzyme information deal with fundamental information of flavonoids and enzyme such as mass, molecular formula, molecular name, etc. The reaction service inserts interaction information between flavonoids and ligands into ER-database. The literature service inserts related literature information into ER-database. Through the literature service, flavonoids researchers share flavonoids information.

Moreover, the flavonoids web services can more effectively support complex queries and incomplete data (i.e. NMR spectroscopy data) by combining ER-model and XML-model. For example, a user can search flavonoids information with respect to name, formula, and range of molecular weight. Also a user can send XPath query to search the NMR spectroscopy data and get a result that is presented as XSL style sheets using the NMR Retrieval web service.

4.2 User Interfaces for Grid-Based Flavonoids Web Portal

A user can easily insert or retrieve flavonoids information through the user-friendly flavonoids web portal. Figure 6 shows screen shots of the flavonoids web portal. The inserted data consists of flavonoids contents and NMR XML instances. Items of the flavonoids information are entry number, name, formula, mass, structure, reaction,

enzyme, NMR(PDF format), MS(PDF format), etc. When a user clicks the 3D Structure link on the structure item (refer to c on Figure 6), the system shows the 3D structure of current flavonoids using Jmol API [14]. When a user clicks the PUBID link on the application item (refer to c on Figure 6), the system shows retrieved papers from PubMed site [15].

a. Insertion of flavonoids information

b. Retrieval using keyword and NMR values

c. Results of flavonoids information retrieval

position	proton shift	carbon shift
2	123.1	145.2
3	111.1	553.3

```

<?xml version="1.0" encoding="euc-kr"?>
<NMR_DATA>
  <row>
    <position|2|position|
    <protonshift|123.1|protonshift|
    <carbonsshift|145.2|carbonsshift|
  </row>
  <row>
    <position|3|position|
    <protonshift|111.1|protonshift|
    <carbonsshift|553.3|carbonsshift|
  </row>
</NMR_DATA>
  
```

d. Results of NMR XML retrieval

Fig. 6. Insertion and Retrieval Examples

4.3 Advantages

The proposed Grid based flavonoids web portal contributes to solve the problems of current flavonoids databases by easy-implementing the flavonoids web portal and sharing flavonoids information through the web portal. It is easy for users to customize search preferences and display environments. Moreover, the GridSphere portal framework makes it easy to build and deploy portals without any modifications.

5 Conclusion

In this paper, we proposed a Grid based flavonoids web portal to easily share flavonoids information through the portal. We analyzed existing flavonoids databases such as USDA and KEGG, and designed relational schema, XML schema for flavonoids

information, and their user interfaces. We explained Flavonoinformatics and showed the proposed architecture for efficient implementation of the flavonoids web portal. We implemented the web portal based on the proposed web service components.

There should be further researches on automatic collecting and processing of related literatures from other literature systems (e.g. PubMed etc.). Also, the proposed web service components should be extended to include other functions, such as update and deletion.

References

1. The biochemistry and medical significance of the flavonoids (2002) Havsteen, B.H. *Par-macol. Ther.* 96, 67-202.
2. Anthocyanins and other flavonoids (2004) Williams, C.A., Grayer, R.J. *Nat. Prod. Rep.* 21, 539-573.
3. Peterson J, Dwyer J, "Taxonomic classification helps identify flavonoid-containing foods on a semiquantitative food frequency questionnaire," *J Am Diet Assoc*, 98:682-4, 1998.
4. <http://www.globus.org>
5. SOAP, <http://www.w3c.org/TR/SOAP>
6. WSDL, <http://www.w3c.org/TR/wsdl>
7. UDDI, <http://www.uddi.org/>
8. Kanehisa, M. and Goto, S., "KEGG: Kyoto encyclopedia of genes and genomes", *Nucleic Acids Res*, 1;28(1):27-30, 2000.
9. Reinhold U, Seiter S, Ugurel S, Tilgen W., "Treatment of progressive pigmented purpura with oral bioflavonoids and ascorbic acid: an open pilot study in 3 patients," *J Am Acad Dermatol*, 41(2 Pt 1):207-8, 1999.
10. <http://www.genome.jp/kegg>
11. http://www.ars.usda.gov/main/site_main.htm?modecode=12354500
12. So FV, Guthrie N, Chambers AF, et al., "Inhibition of human breast cancer cell proliferation and delay of mammary tumorigenesis by flavonoids and citrus juices," *Nutr Cancer*, 26:167-81, 1996.
13. <http://www.gridisphere.org/>
14. <http://jmol.sourceforge.net/>
15. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>