

SPDBSW: A Service Prototype of SPDBS on the Web*

Tae-Sung Jung¹ and Wan-Sup Cho²

¹ Dept. of Information Industrial Engineering, Chungbuk National University, 361763
Cheongju, Chungbuk, Korea
mispro@chungbuk.ac.kr

² Dept. of Management Information Systems, Chungbuk National University, 361763
Cheongju, Chungbuk, Korea
wscho@chungbuk.ac.kr

Abstract. As the amount of pathway information for various organisms is increasing very rapidly, performing various analyses on the full network of pathways for even multiple organisms can be possible and therefore developing an integrated database for storing and analyzing pathway information is becoming a critical issue. Until now analyzing these networks is not easy because of the nature of the existing pathway databases, which are often heterogeneous, incomplete, and/or inconsistent. We presented a database system called SPDBS to solve this problem. However, application-oriented systems like SPDBS have some limitations on the extension and integration of the heterogeneous databases.

In this paper, we extend previous SPDBS into a web service prototype (SPDBSW) where all functions can be serviced on the web environment. The web services include pathway database integration/search, import/export of SBML documents, pathway reconstruction/visualization. SPDBSW has been implemented by the combination SPDBS and external web services such as OLS, KEGG and NCBI. And user can get more confidential and delicate information from KEGG or NCBI through their web services. The system can be extended or modified immediately by replacing its component web services. We provide SPDBSW at the website <http://database.chungbuk.ac.kr/SPDBSW>.

1 Introduction

Biochemical pathways can be viewed as interconnected processes including an intricate network of interactions between molecular compounds in the cell[1]. There are three kinds of biochemical pathways: metabolic, regulatory, or signal transduction pathways[1]. Metabolic pathways are responsible for carrying out the chemical reactions that provide basic biological functions (DNA, RNA, protein synthesis and degradation, energy metabolism, fatty acid synthesis, and many others). Regulatory pathways are responsible for converting genetic information into proteins (gene products). Signal transaction pathways are concerned with coordinating metabolic

* "This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD)" (The Regional Research Universities Program/Chungbuk BIT Research-Oriented University Consortium).

processes with transcription and protein synthesis. Each of these pathways has been kept in a separate (independent) database with distinct attributes even though they are related each other.

Pathway databases contain data of biochemical pathways which consist of two kinds of information: biochemical components (e.g., substrates, enzymes, products) and their interactions [1]. Most existing pathway databases focus on specific types of pathways rather than an integrated one: e.g., Transpath[2] for protein-DNA interactions, KEGG for metabolic pathways, EcoCyc and MetaCyc[6] for *E. coli* and other organisms' metabolic pathways, BIND[2] for signal transduction pathways, PathFinder[3], WIT[8], PathMAPA[9], BioJAKE[11], and MPW[12] for metabolic pathways. Pathway databases raise many important and challenging computational and bioinformatics issues, such as querying, navigation, and visualization of the pathways; seamless integration/analysis of the heterogeneous pathway data distributed in diverse sources.

We proposed the SPDBS [14] (SBML-based Biochemical Pathway Database System: <http://database.chungbuk.ac.kr/SPDBS/>) for the integration and management of heterogeneous biochemical pathways. *Systems Biology Markup Language (SBML)* is an XML-based language for describing simulations in systems biology. The language is oriented towards representing biochemical networks common in research on a number of topics, including: cell signaling pathways, metabolic pathways, biochemical reactions, gene regulation, and many others[4]. We developed an object database for implementing SBML data model[4]. SPDBS provides dynamic pathway reconstruction or estimation by using an orthologous database[7] for a genome sequence data. However, it is impossible to integrate all of databases in the internet and there are some limitations in the maintainability and extensibility.

Web services provide a standard means of interoperating between different software applications, running on a variety of platforms and/or frameworks [18]. Web services are characterized by their great interoperability and extensibility, as well as their machine-processable descriptions thanks to the use of XML [18]. They can be combined in a loosely coupled way in order to achieve complex operations [18]. Programs providing simple services can interact with each other in order to deliver sophisticated added-value services [18]. For example, Google, a popular Internet search engine, provides the web service called the Google Web API. The service enables users to develop software that accesses and manipulates a massive amount of web documents that are constantly refreshed. In the field of genome research, a similar kind of web service called DAS [17] (distributed annotation system) has been used on several web sites, including Ensembl, Wormbase, Flybase, SGD and TIGR [17]. And some of the conventional systems such as KEGG, NCBI and GeneCruiser are providing the web service already.

In this paper, we propose a service prototype (SPDBSW) which integrates the functions of SPDBS and external systems such as KEGG and NCBI by using web services. Many well-known bioinformatics databases, such as OLS (Ontology Lookup Service) [10], KEGG, NCBI, EMBL, and DDBJ, provide web services to the users. SPDBSW provides the functions of SPDBS on the web and other functions of external web services. At the result, it allows to integrate data from external database such as KEGG, MetaCyc using SBML with semantic lookup service of OLS, reconstruction of metabolic pathway for protein sequences and visualization of the