

KEGG API: A Web Service Using SOAP/WSDL to Access the KEGG System

Shuichi Kawashima¹

shuichi@kuicr.kyoto-u.ac.jp

Yoko Sato³

ysatoh@fqs.fujitsu.com

Toshiaki Katayama²

k@bioruby.org

Minoru Kanehisa¹

kanehisa@kuicr.kyoto-u.ac.jp

¹ Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

² Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku 108-8639, Japan

³ Fujitsu Kyushu System Engineering, Momochihama 2-2-1, Sawara-ku, Fukuoka 814-8589, Japan

Keywords: KEGG API, web service, programming, database

1 Introduction

We have been making the KEGG system available at GenomeNet [2]. KEGG is a suite of databases including GENES, SSDB, PATHWAY, LIGAND, LinkDB, etc. for genome research and related research areas in molecular and cellular biology. These databases and associated computation services are available via WWW and the user interfaces are built on web browsers. Thus, the interfaces are designed to be accessed by humans, not by machines, which means that it is troublesome for the researchers who want to use KEGG in an automated manner. Besides, from the database developer's side, it is impossible to prepare all the CGI programs that satisfy a variety of users' needs.

In recent years, the Internet technology for application-to-application communication referred to as the web service is improving at a rapid rate. For example, Google, a popular Internet search engine, provides the web service called the Google Web API [1]. 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, the web service has attracted an increasing attention [3]. For example, the web service called DAS (distributed annotation system) has been used on several web sites, including Ensembl, Wormbase, Flybase, SGD and TIGR.

With the background and the trends noted above, we have started developing a new web service called KEGG API [5]. Using KEGG API, the user can access to all the resources stored in KEGG as a batch processing style without any limitations that the interactive CGI programs will suffer. This also means that the user can embed the KEGG system into in-house programs to analyze his/her own data with KEGG. The program using KEGG API always accesses the up-to-date databases on the KEGG server, no database update or synchronization is required on the client side to use the latest database.

2 Method and Results

Currently, we provide 40 APIs for accessing SSDB, PATHWAY and GENES databases in KEGG as summarized in Table 1. APIs for the SSDB database provide the user to search pre-calculated networks of orthologous/paralogous genes in the universe of protein-coding genes stored in the KEGG

GENES database. As most of the organisms in GENES are completely sequenced, SSDB can provide best hit and bi-directional best hit information in pairwise genome comparisons. This information on the direction of sequence similarity useful for annotating orthologous and paralogous gene groups. The user is also able to obtain sequences and annotations for a specified gene or all genes in an organism using the APIs for the GENES database. Statistics on the KEGG database such as the list of organisms, the number of genes in each organism, the update information of the database, etc. are also available. To analyze gene function in a cellular context, the user can search pathways concerned with any gene or enzyme group. APIs for the PATHWAY database also provide methods to search genes, compounds and enzymatic reactions on the pathway. The result can be shown as a colored image of the KEGG PATHWAY diagram where the user can specify any combination of colors for the objects on the map such as genes, enzymes and compounds.

By combining these APIs, starting from a gene of interest, the user can search orthologs in other species or find a paralogous gene group, list up pathway diagrams which contain the gene, color boxes of the gene on the pathway diagram according to the gene expression level obtained from micro array experiments. The user can build up a pipeline program and a graphical interface for these tasks embedding the KEGG system.

Table 1: The number of the KEGG APIs.

Types of API	Number of APIs	Database entries
SSDB	11	190,900,825
PATHWAY	20	14,128
GENES	5	580,843
KEGG information	4	

The service has been tested with Ruby (SOAP4R version 1.4.8.1) and Perl (SOAP::Lite version 0.55) languages. BioRuby project [4] already prepared a Ruby library to handle the KEGG API effectively. KEGG API should work with every language which can treat SOAP/WSDL including Java and Python without any difficulties.

3 Acknowledgments

This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, The Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

References

- [1] Calishain T. and Domfest R., *Google Hacks*, O'Reilly, 2003.
- [2] Kanehisa M., Goto, S., Kawashima, S., and Nakaya, A., The KEGG databases at GenomeNet, *Nucleic Acids Res.*, 30:42–46, 2002.
- [3] Stein, L., Creating a bioinformatics nation, *Nature*, 417:119–120, 2002.
- [4] <http://bioruby.org/>
- [5] <http://www.genome.ad.jp/kegg/soap/>