BioRuby+ChemRuby: An Exploratory Software Project

Toshiaki Katayama1
k@bioruby.org

Mitsuteru C. Nakao2
n@bioruby.org

Naohisa Goto3
ng@bioruby.org

Nobuya Tanaka4
t@chemruby.org

1 Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
2 Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, AIST Tokyo Waterfront Bio-IT Research Building, 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan
3 Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan
4 Bioinformatics Center, Institute of Chemical Resarch, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

Keywords: open source library, bioinformatics, cheminformatics, BioRuby, ChemRuby, Ruby language

1 Introduction

Integration of bioinformatics and cheminformatics fields is one of the major objectives in post-genomic era. This includes combination of genomic information and chemical information in the context of the biochemical molecular pathway. One of the purposes of this integration is to represent and predict molecular reactions based on the dedicated databases such as KEGG GENES, LIGAND and PATHWAY [1] in combination with various other molecular databases. This will enable researchers to explore which chemical compound affects on the specific biological pathway of the selected organism at a molecular level. However, to make these analyses possible, the integration techniques of these two fields have not been established, although their integration within the databases is being rapidly accomplished. For this purpose, we have developed two open source libraries, BioRuby [2] and ChemRuby [3], for the object oriented scripting language Ruby. Recently, the Ruby language is becoming one of the major computational languages by its clean and powerful nature to deal with complex data structures efficiently. This means researches can easily utilize their own data in their program with the help of BioRuby and ChemRuby libraries. BioRuby and ChemRuby libraries are freely available at http://bioruby.org/ and http://chemruby.org/ respectively.

2 Integration Overview

BioRuby and ChemRuby are a collection of reusable modules implemented in the object oriented scripting language Ruby. As a schematic image is given in Figure 1, both libraries contains basic classes to represent complex biological and chemical data structures, interfaces to other computational applications, various biological and chemical database formats, and computational algorithms to analyze a variety of biochemical molecules from biological sequences to compound structures. Among them, a graph is the common data structure to deal with biological pathways and compound structures. Thus, we have developed a general graph library in Ruby and extended it by adding methods for the biological networks and for the chemical structures. By using BioRuby and ChemRuby, researchers can treat these data simultaneously for their analysis.

2.1 Recent Improvements

In this year, the joint project of BioRuby and ChemRuby is selected as one of the Exploratory Software Project by the Information-technology Promotion Agency, Japan. The goals of this joint project include integration of bioinformatics and cheminformatics, enhancement of the functionality, reinforcement of
To accomplish these tasks, we have implemented graph algorithms that can be commonly used within the two projects, added API documentations in RDoc format which is becoming a standard documentation format in Ruby, and implemented unit tests using a standard Test::Unit framework. We have also developed a command line user interface for our library. This interface enables user to use the functionality of BioRuby and ChemRuby libraries without writing any Ruby script, so that these libraries can be used as a handy tool for the daily analysis.

Acknowledgement

This work is partly supported by Information-technology Promotion Agency, Japan as an Exploratory Software Project.

References

