

Systems biology

PATIKAwEB: a Web interface for analyzing biological pathways through advanced querying and visualization

U. Dogrusoz*, E. Z. Erson, E. Giral, E. Demir, O. Babur, A. Cetintas and R. Colak

Bilkent Center for Bioinformatics (BCBI), Bilkent University, Ankara 06800, Turkey

Received on September 9, 2005; revised on November 8, 2005; accepted on November 10, 2005

Advance Access publication November 15, 2005

Associate Editor: Charlie Hodgman

ABSTRACT

Summary: PATIKAwEB provides a Web interface for retrieving and analyzing biological pathways in the PATIKA database, which contains data integrated from various prominent public pathway databases. It features a user-friendly interface, dynamic visualization and automated layout, advanced graph-theoretic queries for extracting biologically important phenomena, local persistence capability and exporting facilities to various pathway exchange formats.

Availability: The Web site is freely accessible to non-commercial users from <http://web.patika.org>

Contact: patikaweb@cs.bilkent.edu.tr

Supplementary information: Detailed information is available on the tool Web site (<http://www.cs.bilkent.edu.tr/~patikaweb>), including key features, screenshots and User's Guide.

Data on biological systems have been increasing on an exponential rate for the last decade. Although there has been an enormous amount of effort for creating ontologies, standards and tools, current bioinformatics infrastructure is far from being able to cope with this quantity of data. The PATIKA Project (<http://www.patika.org>) aims to provide the scientific community with an integrated environment for modeling, analyzing and integrating cellular processes.

Central to the PATIKA Project is an ontology (Demir *et al.*, 2002), which is based on a qualitative two-level model. At the entity level, interactions and relations can be addressed in an abstract manner, where the exact state of the related parties is unknown, such as protein–protein interactions, inferred relations and literature-derived information. At the state/transition level, each entity is associated with a set of states interacting with each other via transitions. This level can capture more detailed information such as compartments, molecular complexes and different types of biological events (e.g. covalent modification, transportation and association). This two-level ontology elegantly covers most biological pathway-related phenomena and is capable of integrating information present in the literature and molecular biology databases. Additionally, the PATIKA ontology uses the notion of compound graphs to represent abstractions, which are logical groupings that may be used to handle the complex and incomplete nature of the data.

The PATIKA Server acts as a central database and provides XML-based Web services for querying and integrating pathway models.

Currently the server contains data integrated from several public databases including Reactome (Joshi-Tope *et al.*, 2005), and interfaces with major databases and ontologies such as Entrez-Gene (Maglott *et al.*, 2005), UniProt (Bairoch *et al.*, 2005) and GO (Harris *et al.*, 2004).¹ A CVS-like concurrent versioning system is used for integration logic, allowing continuous and concurrent updates to the server via other PATIKA tools. The latest numbers on the database contents can be obtained from the tool's welcome page but initially several thousands states of many different biological entities and a few thousand transitions/reactions were integrated into the database.

Perceiving the increasing demand on easily accessible tools, we developed PATIKAwEB to provide a Web-based service with a user-friendly interface without requiring any registration or local installation: simply point your browser to <http://web.patika.org>. Since we expect that the majority of PATIKA users will only be interested in knowledge acquisition, PATIKAwEB focuses on providing a simple yet powerful interface for querying and visualizing the PATIKA database.

PATIKAwEB fully supports the PATIKA ontology, including a multiple-view schema for bioentity and mechanistic levels, compartments and compound graphs for visualizing molecular complexes, pathways and black-box reactions.

Local graph editing functionality such as scrolling and zooming, do/undo and delete as well as advanced features such as expand/collapse nested abstractions are available to manage the current pathway model. A specialized algorithm which can handle both compound graphs and compartments is used to lay out pathways automatically (Dogrusoz *et al.*, 2004). This is critical for on-the-fly generation of pathway drawings, say as a result of a user query.

An overview window aids in navigating over large pathway views. Properties of each object of a pathway view may also be inspected via possibly multiple pop-up windows. Furthermore, links to external databases are provided through this window.

The querying component of PATIKAwEB both supports SQL-like queries (Fig. 1) and an array of graph-theoretic queries for finding feedback loops, positive/negative paths, common targets and regulators or 'interesting subgraphs' based on the user's genes of interest (Fig. 2).

Once retrieved from the database, the query results may be merged to the user's current view and highlighted to provide an

*To whom correspondence should be addressed.

¹Latest versions or releases (as of May 2005) of these databases and ontologies have been used.

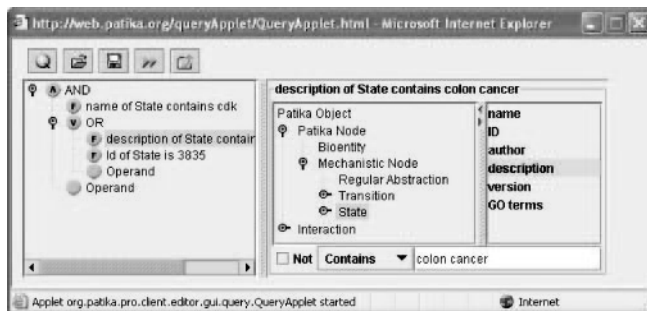


Fig. 1. Find all states whose names contain 'cdk' among all whose descriptions contain the string 'colon cancer' or whose IDs are '3835'.

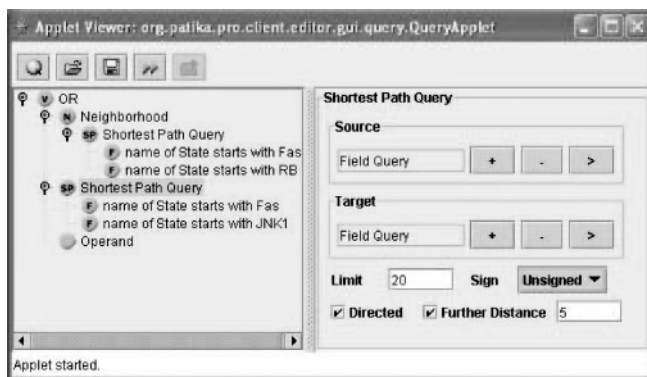


Fig. 2. Find the union of 1-neighborhood of the shortest path from states whose names start with 'Fas' to states whose names start with 'RB' with the shortest path from states whose names start with 'Fas' to states whose names start with 'JNK1'.

incremental user-friendly retrieval and analysis interface. Constructed models may be saved locally in an internal XML-based format, exported to standard formats such as BioPAX (BioPAX, 2003, <http://www.biopax.org>) and SBML (Hucka *et al.*, 2003) or converted to static images (JPEG or SVG).

PATIKAWEB has a distributed architecture, where the server is composed of a database component and an application server. The implementation uses the JSP (JavaServer Pages technology) edition of the Tom Sawyer Visualization technology to handle this highly dynamic and advanced visual content along with Java™, JavaScript™ and DHTML/HTML.

A screenshot of PATIKAWEB is given in Figure 3. We believe that PATIKAWEB's unique visualization and querying features, coupled with its user-friendly Web-based interface fill an important gap in the pool of currently available tools and databases. In parallel to the

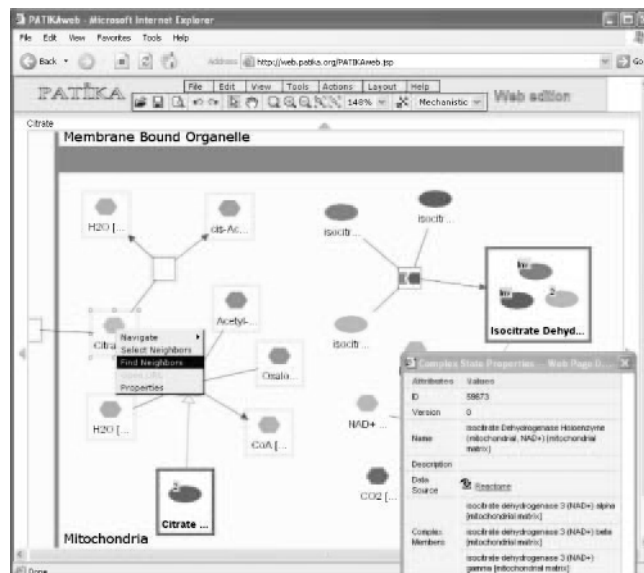


Fig. 3. A screenshot from PATIKAWEB.

dramatic expansion in its data and features, we expect PATIKAWEB to quickly become a popular tool among molecular biologists.

ACKNOWLEDGEMENTS

The authors wish to thank the rest of the Patikans for their contributions and the anonymous referees for their helpful suggestions. This work was supported in part by TUBITAK (The Scientific and Technological Research Council of Turkey), grant number 104E049.

Conflict of Interest: none declared.

REFERENCES

- Bairoch, A. *et al.* (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **33** (Database issue), D154–D159.
- BioPAX (2003), Biological pathways exchange.
- Demir, E. *et al.* (2002) An ontology for collaborative construction and analysis of cellular pathways. *Bioinformatics*, **20**, 349–356.
- Dogrusoz, U. *et al.* (2004) A compound graph layout algorithm for biological pathways. *L.N.C.S.*, **3383**, 442–447.
- Harris, M.A. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32** (Database issue), D258–D261.
- Hucka, M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.
- Joshi-Tope, G. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33** (Database issue), D428–D432.
- Maglott, D. *et al.* (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33** (Database issue), D54–D58.