BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 17 2010, pages 2210–2211
doi:10.1093/bioinformatics/btq348

Databases and ontologies

Advance Access publication June 30, 2010

Biological knowledge bases using Wikis: combining the flexibility of Wikis with the structure of databases

Sylvain Brohée[1,*,†], Roland Barriot[1,2,†] and Yves Moreau[1]

[1]ESAT-SCD. Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium and [2]Université de Toulouse, UPS, Laboratoire de Microbiologie et Génétique Moléculaire, Centre National de la Recherche Scientifique, F-31000 Toulouse, France

Associate Editor: John Quackenbush

ABSTRACT

Summary: In recent years, the number of knowledge bases developed using Wiki technology has exploded. Unfortunately, next to their numerous advantages, classical Wikis present a critical limitation: the invaluable knowledge they gather is represented as free text, which hinders their computational exploitation. This is in sharp contrast with the current practice for biological databases where the data is made available in a structured way. Here, we present *WikiOpener* an extension for the classical MediaWiki engine that augments Wiki pages by allowing on-the-fly querying and formatting resources external to the Wiki. Those resources may provide data extracted from databases or DAS tracks, or even results returned by local or remote bioinformatics analysis tools. This also implies that structured data can be edited via dedicated forms. Hence, this generic resource combines the structure of biological databases with the flexibility of collaborative Wikis.

Availability: The source code and its documentation are freely available on the MediaWiki website: http://www.mediawiki.org/wiki/Extension:WikiOpener.

Contact: sbrohee@esat.kuleuven.be

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on March 4, 2010; revised on May 25, 2010; accepted on June 25, 2010

1 INTRODUCTION

Biological databases are essential tools for many biologists in all research areas. The number of these online resources is constantly increasing. Indeed, as an illustration, the issue of the *Nucleic Acids Research* journal dedicated to biological databases has already described about 1200 online repositories which only represent a small part of all the existing ones (Galperin and Cochrane, 2009).

Maintenance is the key issue for many biological databases. Indeed, with the large number of new papers published each day, it is increasingly difficult to efficiently update smaller databases, maintained by a small group of people. Moreover, often small databases will stop being maintained when the project funding comes to its end or when the main curator leaves the lab.

A new way to tackle this problem consists in building collaborative databases which are curated by all the people in the field wanting to share their knowledge with the community according to their area of expertise. The implementation of some of these databases is based on the MediaWiki engine which is used in well-known collaborative projects, such as Wikipedia. Among these, in biology, we can cite the cases of WikiPathways for the curation of biological pathways (Pico *et al.*, 2008), Proteopedia, for 3D structures of biochemical compounds (Hodis *et al.*, 2008) or Metabolomics.jp for the annotation of metabolites (Arita, 2009). MediaWiki presents numerous advantages: fast and easy set up of a new collaborative database, user rights and statistics management, simple and appealing for users, etc.

However, MediaWiki in itself presents some limitations. First, the web site can only be based on one unique underlying database. Second, classically, the data of this database are not structured and are generally represented in a free text format, which makes difficult any query different from the built-in queries. These limitations are somewhat contradictory with the current common sense practice for biological databases where the available data come from well-structured repositories. Moreover, a lot of web accessible biological databases mainly consists of a set of entries extracted from relational database tables. Third, a lot of online biological databases offers bioinformatics tools to the scientific community (e.g. BLAST), which is a functionality not supported by the MediaWiki engine. These limitations prevent from using only MediaWiki to develop collaborative databases that would be comparable with current biological databases. Some solutions, including Semantic web strategies, have already emerged to tackle these limitations. We followed a totally different direction by developing *WikiOpener*, an extension to MediaWiki, not linked to concepts and closer to the original Wikipedia look and feel.

2 APPROACH

MediaWiki extensions are pieces of code that allow the users to adapt the engine to their needs by improving the existing functionalities or by adding new ones. With *WikiOpener*, the inclusion of external data in a wiki page is done by including a given tag within the page. The options within this tag specify (i) which function (*WikiOpener* component) should be called to obtain the data for inclusion into the Wiki page and (ii) the way this data have to be displayed on the web page (layout which is being stored as another article of the Wiki) (Fig. 1, Supplementary Fig. 1).

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First authors.

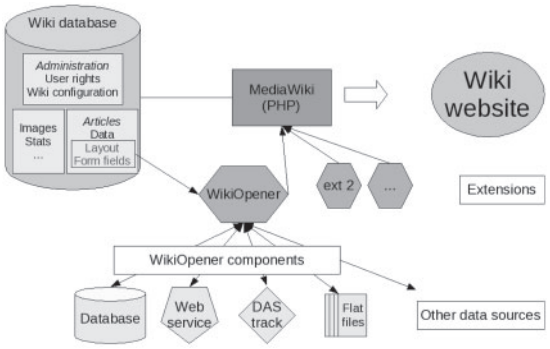

Fig. 1. *WikiOpener* principle. Classically, the wiki data are extracted by the MediaWiki engine from a unique database which also, e.g. stores the usage statistics, and the user rights management. *WikiOpener* allows the extraction and edition of data from any other external resource. The way to extract the data is encoded in the components and the layout of the *WikiOpener* elements (data, forms) are stored as simple Wiki pages.

Furthermore, our extension allows the automatic inclusion of data on predefined pages. The inclusion can be done on every page, on a given type of page (namespace), or on a particular page and can consist of simple Wiki text, data obtained from external databases or web services, or even results of analysis tools. As a consequence, if a page has not yet been manually created by the Wiki contributors, external data or tools may already fill it. For example, a Wiki article about a gene can immediately be prefilled with data extracted from any resources (publicly available database, DAS server, user own resources). A page can thus consist of results to any queries preformatted by the wiki administrator. The possibility to run analysis tools on the fly and display their results within the pages of the Wiki is of great interest as it can be used to propose bioinformatics tools (BLAST, network analysis, gene prioritization, etc.), which is common in biological databases.

Edition of external data had to be as transparent and intuitive as the edition of Wiki pages. To this end, *WikiOpener* allows the user to enter and modify data through standard web forms (text fields, pop up menu, etc.). *WikiOpener* components can then make use of the submitted data to update the external databases. The fields and the layout of the forms are stored as Wiki pages.

3 APPLICATIONS

We have currently used the *WikiOpener* extension for the development of two biological databases: CHDwiki devoted congenital heart defects (http://homes.esat.kuleuven.be/~bioiuser/chdwiki/; Barriot *et al.*, 2010, Supplementary Fig. 2) and YTPdb (Van Belle and André, 2001) is dedicated to the classification of the yeast membrane transporters (http://homes.esat.kuleuven.be/ytpdb/). Any expert in these fields is welcome and is able to contribute after registration (simple request to the respective web site administrator).

For testing purposes, we designed a demo wiki running WikiOpener which can be freely edited (http://homes.esat.kuleuven.be/~bioiuser/demo_wiki/).

4 CONCLUSIONS

To solve the recurring maintenance problem of many specialized databases, many research teams choose to open their database to all specialists of the field using collaborative databases or Wikis (Hu *et al.*, 2008). To circumvent the limitations inherent to the wiki structure, some extensions to the MediaWiki engine have already been proposed using the Semantic Web: a general format for data interchange that provides curators with a standardized framework allowing data to be integrated and reused across disciplines. Semantic Web advantages are obvious: unique names for biological entities and consistent standards for knowledge representation and processing. Such standards simplify integration of web resources as they can be queried the same way by linking non-ambiguous concepts and that would lead to a federated wiki landscape. Bioinformatics query systems in diverse fields have started to use this technology [e.g. WikiProteins (Mons *et al.*, 2008)].

Although *WikiOpener* as such does not contribute directly to the unification of different biological web resources, it offers a novel strategy allowing the integration of relational databases (as well as any other data source) to a wiki. Indeed, wikis running *WikiOpener* can display and edit both wiki unformatted and formated data coming from external resources and create pages automatically. It can also thus query Semantic Web triple stores using dedicated components. Moreover, WikiOpener is not only limited to the querying and filling of relational databases but can also be used for the integration of tools in the wiki and can thus be used in combination with Semantic Web technologies. In fact, it realizes a practical solution to integrate any external systems.

This extension has already been used sucessfully for two different web knowledge bases and we are quite confident that it is of interest to the scientific community intending to share their data in a user-friendly way.

Funding: SCD-SISTA is funded by grants provided by the (i) Research Council KUL: GOA MaNet, CoE EF/05/007 SymBioSys, (ii) IWT: SBO-MoKa, (iii) the Belgian Federal Science Policy Office: IUAP P6/25 (BioMaGNet); (iv) the European Union: FP7-HEALTH CHeartED.

Conflict of Interest: none declared.

REFERENCES

Arita,M. (2009) A pitfall of wiki solution for biological databases. *Brief. Bioinform.*, **10**, 295–296.

Barriot,R. *et al.* (2010) Collaboratively charting the gene-to-phenotype network of human congenital heart defects. *Genome Med.*, **2**, 16.

Galperin,M.Y. and Cochrane,G.R. (2009) Nucleic acids research annual database issue and the nar online molecular biology database collection in 2009. *Nucleic Acids Res.*, **37**, D1–D4.

Hodis,E. *et al.* (2008) Proteopedia - a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. *Genome Biol.*, **9**, R121.

Hu,J.C. *et al.* (2008) The emerging world of wikis. *Science*, **320**, 1289–1290.

Mons,B. *et al.* (2008) Calling on a million minds for community annotation in wikiproteins. *Genome Biol.*, **9**, R89.

Pico,A.R. *et al.* (2008) Wikipathways: pathway editing for the people. *PLoS Biol.*, **6**, e184.

Van Belle,D. and André,B. (2001) A genomic view of yeast membrane transporters. *Curr. Opin. Cell Biol.*, **13**, 389–398.