OXFORD JOURNALS
Bioinformatics

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

Journal:	*Bioinformatics*
Manuscript ID:	BIOINF-2010-0309
Category:	Applications Note
Date Submitted by the Author:	04-Mar-2010
Complete List of Authors:	Brohée, Sylvain; Katholieke Universiteit Leuven, ESAT - SCD - SISTA - Bioinformatics group Barriot, Roland; Centre National de la Recherche Scientifique, Laboratoire de Microbiologie et Génétique Moléculaires; Université Paul Sabatier, Laboratoire de Microbiologie et Génétique Moléculaires Moreau, Yves; Katholieke Universiteit Leuven, ESAT - SCD - SISTA - Bioinformatics group
Keywords:	Data integration, Databases, Data analysis, Wiki, Data annotation

scholarONE
Manuscript Central

Bioinformatics

Vol. 00 no. 00 2010
Pages 1–2

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

Sylvain Brohée [1,+,*] Roland Bariot [1,2,3,+] and Yves Moreau [1]

[1]ESAT-SCD. Katholieke Universiteit Leuven. Kasteelpark Arenberg 10, B-3001 Leuven. Belgium.
[2]Université de Toulouse; UPS; Laboratoire de Microbiologie et Génétique Moléculaire; F-31000 Toulouse ; France.
[3]Centre National de la Recherche Scientifique; LMGM; F-31000 Toulouse; France.
[+]Both authors contributed equally to this work.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Summary: In recent years, the number of knowledge bases developed using Wiki technology has exploded. Unfortunately, next to their numerous advantages, 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 *WikiOpener* extension is licensed under the GNU-GPL. 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

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 (Rhee and Crosby (2005)).

A new way to tackle this problem consists in building collaborative databases that are no longer curated by one unique annotation team but 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 or Wiktionnary. Among these, we can cite the cases of Wikiproteins for textually annotating proteins (Mons *et al.* (2008)), WikiPathways for the curation of biological pathways (Pico *et al.* (2008)), or Proteopedia, for 3D structures of biochemical compounds (Hodis *et al.* (2008)). Using the MediaWiki engine to build a collaborative database has numerous advantages. For example, MediaWiki gives the possibily to biologists with limited programming skills to rapidly set up a new collaborative database. Moreover, user rights (edition, consultation only) are managed via this engine. Finally, because MediaWiki is widely used and aimed to be simple, the biological users will access a system that is user friendly and has a structure and syntax with which they are already familiar.

However MediaWiki presents some limitations. First, the web site can only be based on one unique underlying database. Second, the data of this database are not structured and are only represented in a free text format, which makes impossible 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 to current biological databases. As such, all current Wiki-based biological knowledge bases have had to address this problem in an *ad hoc* fashion.

Aware of these limitations, we have developed, *WikiOpener*, an extension for the MediaWiki engine that allows Wiki web sites to combine both the advantages of Wikis and classical biological databases.

*to whom correspondence should be addressed

© Oxford University Press 2010.

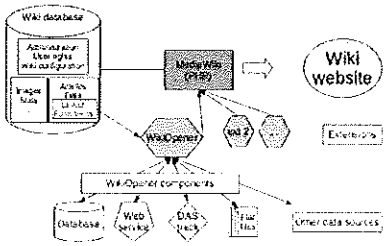


Fig. 1. Principle of the *WikiOpener* extension. Classically, the wiki data are extracted by the MediaWiki engine from a unique underlying database. It is also responsible for the article layout, the usage statistics, and the user rights management. The *WikiOpener* extension allows to extract and edit 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.

2 APPROACH

MediaWiki extensions are pieces of PHP code that allow the users to adapt the MediaWiki 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, on the one hand, which PHP function (*WikiOpener* component) should be called to obtain the data for inclusion into the Wiki page, and on the other hand, 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).

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). 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)) and YTPdb (Van Belle and Andre (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).

4 CONCLUSION

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)). Some extensions to the MediaWiki PHP engine run by the majority of the Wiki knowledge bases have already been developed in order to circumvent the limitations inherent to the wiki structure. It is for example the case of the *Metabolomics.jp* wiki database which displays both editable unformatted information as well as formatted information in which relational operations can be performed (Arita and Suwa (2008)). However, the *WikiOpener* extension that we present in this paper provides database administrators with a tool able to combine the advantages of Wiki-like web sites and of more structured databases. Indeed, wikis running *WikiOpener* can display and edit both wiki unformatted and formated data coming from external databases, create pages automatically and propose bioinformatics tools.

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.

ACKNOWLEDGEMENT

SCD-SISTA is funded by grants provided by the (1) Research Council KUL: GOA MaNet, CoE EF/05/007 SymBioSys, (2) IWT: SBO-MoKa, (3) the Belgian Federal Science Policy Office: IUAP P6/25 (BioMaGNet); (4) the European Union: FP7-HEALTH CHeartED.

REFERENCES

Arita, M. and Suwa, K. (2008). Search extension transforms wiki into a relational system: A case for flavonoid metabolite database. *BioData Min*, **1**(1), 7.

Barriot, R., Breckpot, J., Thienpont, B., Brohée, S., Van Vooren, S., Coessens, B., Tranchevent, L.-C., Loo, P. V., Gewillig, M., Devriendt, K., and Moreau, Y. (2010). Collaboratively charting the gene-to-phenotype network of human congenital heart defects. *Genome Med*, **2**(3), 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**(Database issue), D1–D4.

Hodis, E., Prilusky, J., Martz, E., Silman, I., Moult, J., and Sussman, J. L. (2008). Proteopedia - a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. *Genome Biol*, **9**(8), R121.

Hu, J. C., Aramayo, R., Bolser, D., Conway, T., Elsik, C. G., Gribskov, M., Kelder, T., Kihara, D., Knight, T. F., Pico, A. R., Siegele, D. A., Wanner, B. L., and Welch, R. D. (2008). The emerging world of wikis. *Science*, **320**(5881), 1289–1290.

Mons, B., Ashburner, M., Chichester, C., van Mulligen, E., Weeber, M., den Dunnen, J., van Ommen, G.-J., Musen, M., Cockerill, M., Hermjakob, H., Mons, A., Packer, A., Pacheco, R., Lewis, S., Berkeley, A., Melton, W., Barris, N., Wales, J., Meijssen, G., Moeller, E., Roes, P. J., Borner, K., and Bairoch, A. (2008). Calling on a million minds for community annotation in wikiproteins. *Genome Biol*, **9**(5), R89.

Pico, A. R., Kelder, T., van Iersel, M. P., Hanspers, K., Conklin, B. R., and Evelo, C. (2008). Wikipathways: pathway editing for the people. *PLoS Biol*, **6**(7), e184.

Rhee, S. Y. and Crosby, B. (2005). Biological databases for plant research. *Plant Physiol*, **138**(1), 1–3.

Van Belle, D. and Andre, B. (2001). A genomic view of yeast membrane transporters. *Curr Opin Cell Biol*, **13**(4), 389–398.