A fielded wiki for personality genetics

Finn Årup Nielsen
DTU Informatics
Technical University of Denmark
Kongens Lyngby, Denmark
fn@imm.dtu.dk

ABSTRACT

I describe a fielded wiki, where a Web form interface allows the entry, analysis and visualization of results from scientific papers in the personality genetics domain. Papers in this domain typically report the mean and standard deviation of multiple personality trait scores from statistics on human subjects grouped based on genotype. The wiki organizes the basic data in a single table with fixed columns, each row recording statistical values with respect to a specific personality trait reported in a specific paper with a specific genotype group. From this basic data hard-coded meta-analysis can compute individual and combined effect sizes. The meta-analytic results are displayed in on-the-fly computed hyperlinked graphs and tables. Revision control on the basic data tracks changes and data may be exported to comma-separated files or in a MediaWiki template format.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: On-line Information Services—*Web-based services*; H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia—*Architectures*; J.3 [**Life and Medical Sciences**]: Biology and genetics

Keywords

Wikis, structured data, tables, meta-analysis, bioinformatics, neuroinformatics

1. INTRODUCTION

Neuroinformatics deals with representing neuroscience data and develops tools for analysis and visualization of such data. Most data in the neuroscience field is only reported in scientific papers and database curators have difficulties in keeping up with the increasing amount of data being generated. We have constructed Brede Database[1] based on XML

---

[1]http://neuro.imm.dtu.dk/services/brededatabase/

and Matlab for results of neuroimaging data. However, neuroinformatics databases cannot keep up with the papers published in the neuroimaging field [3]. I have advocated for a more collaborative and wiki-oriented approach to overcome the problem of database entry [5], and set up the MediaWiki-based Brede Wiki to record text and numerical information from neuroscience [4]. Neither the Matlab entry interface for the Brede Database nor the raw edit field of the MediaWiki interface present a convenient interface for adding data, and therefore we are exploring alternative ways of data entry. Here, I describe a field-based wiki interface (i.e. a highly structured wiki) to add and edit data from a very specific area of neuroinformatics—personality genetics—that investigates the association between personality traits and specific genetic variations The researchers typically compute average personality scores across subjects within a genotype group and list the values in tables in their published papers. AlzGene [2], SzGene and PDGene record data from genetic association studies on Alzheimer, schizophrenia and Parkinson diseases, respectively. The original data from the scientific studies are entered and presented in a Web-based environment with meta-analysis results and visualizations. Inspired by these systems the fielded wiki also includes meta-analysis and visualization.

2. THE WIKI

The basic data structure of the wiki is a table where the columns are: row identifier, gene, polymorphism, genotype, inventory, trait, mean value, standard deviation, number of subjects, recruitment information and document identifier, see Figure 1. A revision table has the same columns and further adds columns for revision time, revision user as well as a revision identifier. New data is entered in a form displayed as an extra row in the table. Through JavaScript an autocomplete feature copies the corresponding value from the previous row to the input field if the input field is left empty For the trait field a simple autoreplace expands, e.g., 'N' to 'Neuroticism'. There is little constraint on what the user can enter in the fields, e.g., it is possible to enter a value in the gene field that is not a valid human gene symbol. Apart from adding new data the wiki interface also allows a user to go back and edit a specific row. A separate form is used for that purpose. On this form predefined options for the most common choices guide the user to select appropriate values.

For gene, polymorphism, inventory, trait, recruitment and PMID hyperlinks are automatically constructed so filtered

Figure 1: The main window with the basic data of the wiki as organized in a table. Field values are hyperlinked to filtered views or to external wikis: Wikipedia and the Brede Wiki.

views of the data are presented when the links are followed. Also gene, polymorphism, inventory and trait are hyperlinked to their corresponding pages in Wikipedia and the Brede Wiki, and the PMID is linked to the PubMed database. The wiki can generate so-called forest plots showing each row in the table as a dot with whiskers: The mean determines the position of the dot, the standard deviation the whiskers, and the number of subjects the size of the dot.

The wiki applies a standard statistical framework for meta-analysis by computing an *effect size* for each genetic polymorphism and each trait in a paper. These effect sizes are combined to form a meta-analytic mean. The wiki generates a forest plot with the effect sizes and the combined meta-analytic effect. Through a form interface the user can filter which effect sizes should be included in the visualization and the on-the-fly computation of the combined effect sizes, see Figure 2.

Data can be exported in comma-separated files or in a MediaWiki template format, that can be included in the Brede Wiki (or any other MediaWiki). The wiki exports each row as a MediaWiki template, and in the Brede Wiki these templates are formatted to rows in HTML tables. A simple cookie-based authentication system may allow a user to log in. Otherwise the edit activity is recorded with the IP number in the revision table. The wiki is implemented in a single Python script using SQLite, SciPy and hyperlinked Scalar Vector Graphics

3. DISCUSSION

Some of the features of the wiki could be implemented with other structured collaborative systems, e.g., with MediaWiki templates, MediaWiki table extensions, Semantic MediaWiki [6] with form interface, OntoWiki [1], Google Fusion Tables

or online spreadsheets, such as the spreadsheet in Google Docs. As an example SNPedia is a wiki that uses Semantic MediaWiki to store information about polymorphisms and genetic associations, i.e., data very much like the data in the present wiki. Although such systems may represent the data, most

Figure 2: Forest plot of effect sizes in SVG controlled by a form interface with filtering.

of these systems cannot directly perform the automated hyperlinking, the specialized meta-analysis and visualization implemented in the wiki. Furthermore, the wiki can tailor its interface to support fast entry.
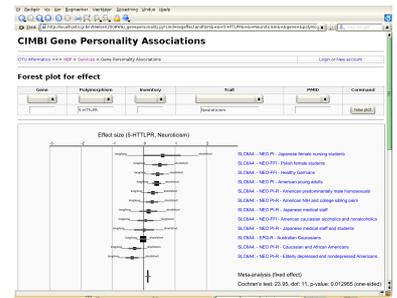
The wiki was designed so the basic data would be represented in an as simple structure as possible. More tables would be needed if, e.g., bibliographic information and subject demographics should be recorded in detail.

The wiki lacks a flexible interface for meta-analysis. However, the user may export the data as comma-separated files and make a more tailored analysis in an external program. Alternatively, the functionality of the wiki should be extended, e.g., with a programming-oriented interface or with an interface to specify which values should enter the meta-analysis.

The wiki can be regarded as a system of open data in science, where the results included in a meta-analysis are readily available for re-use, and the wiki has potential for large-scale meta-analysis across multiple traits and genetic variants.

4. ACKNOWLEDGMENTS

5. REFERENCES

[1] S. Auer, T. Riechert, and S. Dietzold. OntoWiki — a tool for social, semantic collaboration. In *The Semantic Web — ISWC 2006*, pages 736–749, 2006.

[2] L. Bertram, et al. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature Genetics*, 39:17–23, 2007.

[3] J. Derrfuss and R. A. Mar. Lost in localization: The need for a universal coordinate database. *NeuroImage*, 48:1–7, 2009.

[4] F. Å. Nielsen. Brede Wiki: Neuroscience data structured in a wiki. In *Proc. 4th Workshop on Semantic Wikis*, pages 129–133, 2009.

[5] F. Å. Nielsen. Lost in localization: A solution with neuroinformatics 2.0. *NeuroImage*, 48:11–13, 2009.

[6] M. Völkel, et al. Semantic Wikipedia. In *Proc. 15th Int. Conf. World Wide Web*, pages 585–594, 2006.