Nature Precedings : doi:10.1038/npre.2010.4795.1 : Posted 20 Aug 2010

waveomics: bringing experimental data to online collaboration

Neil Swainston

Manchester Centre for Integrative Systems Biology, University of Manchester, Manchester M1 7DN, UK

Abstract

Summary

Systems biology offers an interdisciplinary approach to scientific research that typically involves the collaboration of teams of experimentalists and mathematical modellers. While the importance of data standards has been recognised in facilitating exchange of data between the parties, challenges still remain regarding the practicalities of disseminating experimental data.

The introduction of novel web-based tools aimed at promoting collaborative work has provided a platform upon which scientific applications can be built. The recently released Google Wave protocol provides a facility for real-time collaboration between teams of researchers.

This work introduces a customized Robot that automatically scans text in Google Waves for experimental data identifiers, extracts corresponding experimental data from remote resources associated with such identifiers, and appends charts showing this experimental data to the Wave.

Availability and Implementation

The waveomics Robot is available at waveomics@appspot.com and has the project page http://waveomics.appspot.com/. Source code and associated build files are available under the open source Academic Free Licence v3.0 from http://mcisb.sf.net/. Waveomics is written in Java 1.6 using the Google Wave Robots API, version 2.

Contact

neil.swainston@manchester.ac.uk

Introduction

Systems Biology applies a holistic approach to bioscience, in which an understanding of the system is elucidated from an integrated, cyclical process of observation, modelling and hypothesis generation. Due to its very nature, it is an interdisciplinary science, drawing upon heterogeneous experimental datasets from a range of sources. The ability to disseminate this experimental data in such a way that it can be accessed in a simple and intuitive manner is of particular importance in such an interdisciplinary environment. Systems biology projects commonly operate a distributed approach to collaborative work, where experimentalists, modellers and coordinators may be housed in different departments, and often even different countries.

Google Wave (http://wave.google.com/) has been described as “what email would look like if it were invented today”. A web-based application, it has been developed specifically to facilitate on-line collaboration, and shares much of its functionality with that of e-mail, instant messaging and wikis. Data are organized as “Waves”, which are partly discussions, partly editable documents. Multiple participants can contribute to and edit a Wave, in this sense providing functionality similar to that of a wiki page. Additionally, there exists the facility to both develop and utilise customized plugins, in the form of “Gadgets” and “Robots”. Gadgets are extensible applications that can be shared and used interactively by all participants in a Wave. Examples of these include embeddable calendar and map applications. Robots behave as automated participants in the Wave, and can respond to update events to perform such tasks as automatically checking input for spelling errors, or in a more specialised scientific example, scanning text for chemical formulae.

As such, there has been great interest in the bioinformatics community in the use of Google Wave[1], and while the original excitement has subsided a little, tools such as this (or those that supercede them) provide great promise in facilitating collaborative work. The speed at which the community will adopt such tools is likely to be dependent upon the development of useful and usable tools that will persuade a mass of users to move away from more familiar media such as e-mail.

A Robot is introduced here, waveomics, which provides the first example of allowing experimental data from multiple sources to be queried and shared in the Google Wave environment. While this represents a first step, the work gives an indication of what can be achieved in such a system, and provides a concrete use case as to how research laboratories can benefit by making their data available in standard formats that can be exploited and shared online.

Methods

Robots are automated participants in a Wave. They can perform customised tasks, such as modifying the contents of the Wave, interacting with other participants, or interfacing with external resources, such as other web resources or databases.

The waveomics Robot was written using the Google Wave Robots API. The Robot itself is a Java servlet that extends the `AbstractRobot` base class. Event handler methods are overwritten and these are triggered when a given operation occurs in a Wave to which the Robot has been added. Examples of these events are `onWaveletSelfAdded`, which is fired when the Robot is added to a Wave, and `onWaveletParticipantsChanged`, fired when a new participant enters the discussion.

The waveomics Robot implements the `onBlipSubmitted` event. In Google Wave terms, a “Blip” is described as a single unit of information, which commonly is a single update or message from a participant. Upon submission of such a Blip, the event handler in the Robot scans the text within the Blip, applying one or more regular expressions, to determine identifiers of interest. These may take the form of gene names, protein identifiers, or chemical species.

Robots, like human participants, can also contribute to Waves by adding or updating Blips themselves. Upon detection of an identifier, the Robot updates the Blip by appending a customised Gadget (a plug-in application) that attempts to search for and display any experimental data appropriate to this identifier.

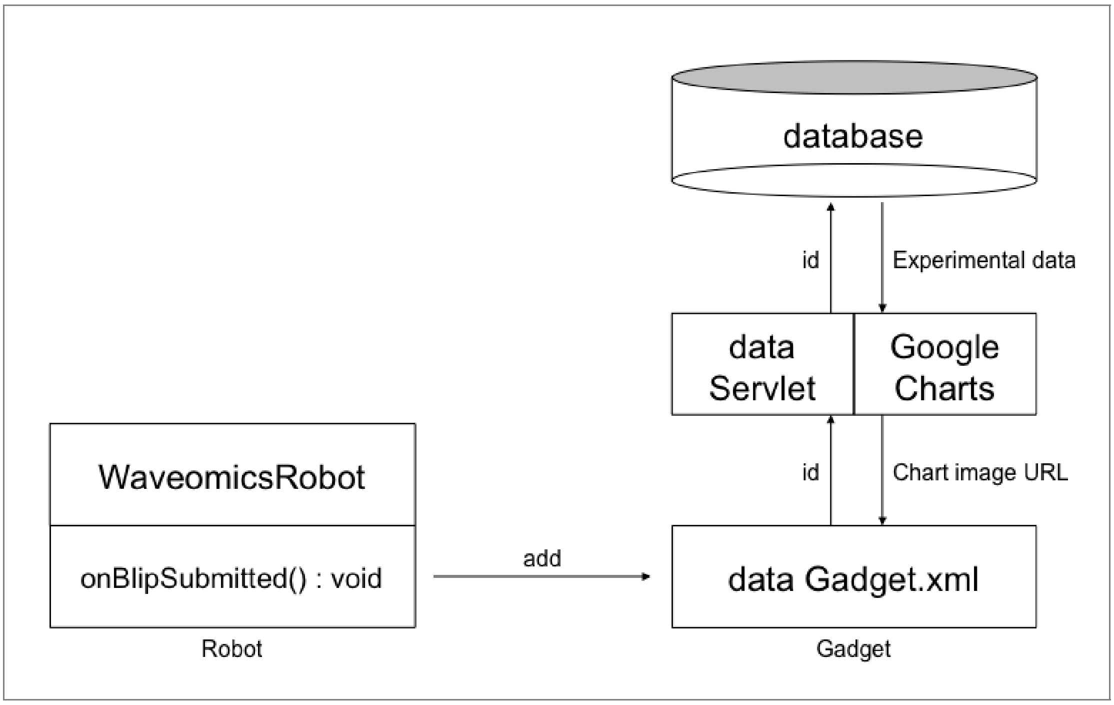


Figure 1. Architecture of the waveomics Robot, associated Gadget and data servlet.

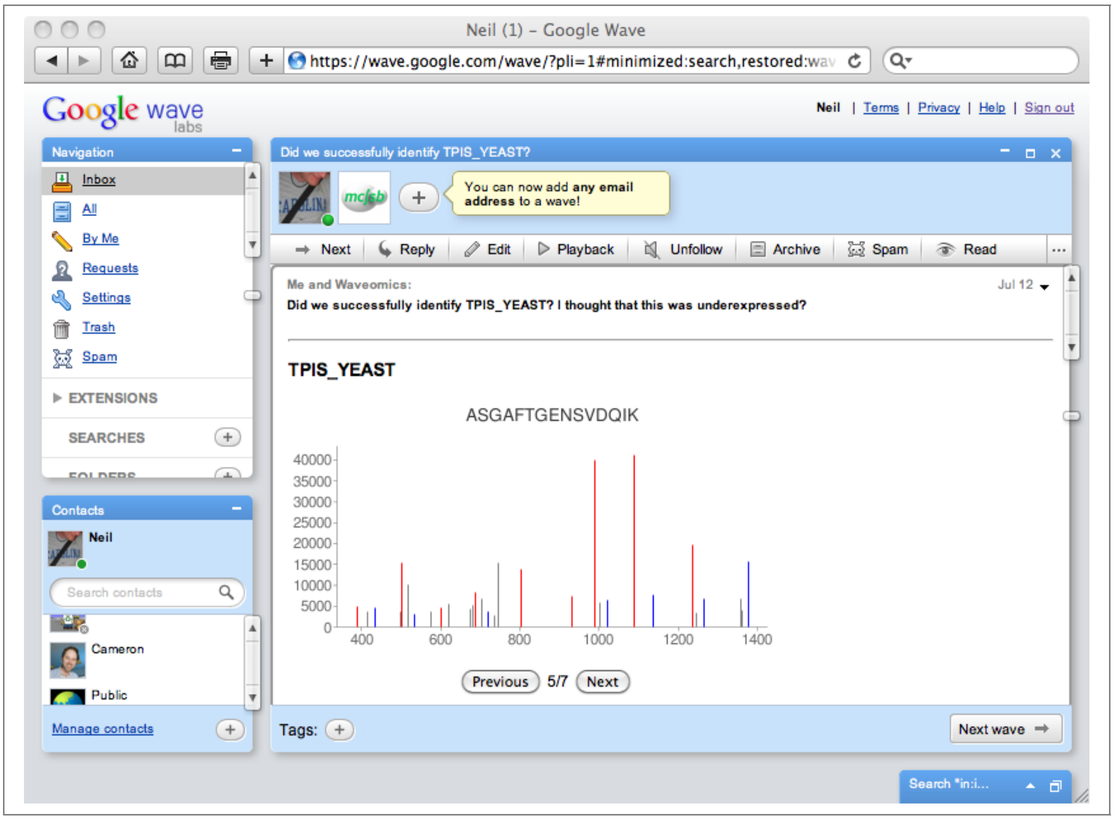
Gadgets are specified in XML files containing HTML markup that can be embedded within a Wave. As such, a Gadget can perform much the functionality of any other web page. The Robot passes the waveomics Gadget the matched database identifier as a query string. The Gadget passes this identifier to an external data retrieval servlet, whose responsibility it is to query an appropriate database for experimental data matching the identifier. Upon extracting data, the data retrieval servlet utilizes Google Charts to generate a chart image, returning a URL specifying the chart image to the Gadget, which then embeds the image within its HTML markup. (See figure 1).

Results

Waveomics supports the querying and visualisation of two heterogeneous experimental data types: mass spectrometry acquired proteomics data, and spectrophotometric enzyme kinetic assays data. It is important to note, however, that the architecture described here may be extended to query for and embed other experimental data types, such as metabolomics data, or even respond to updated Blips in an entirely different manner.

Proteomics

Upon submission of a Blip, the waveomics Robot applies a regular expression matcher to the text within the Blip, attempting to find UniProt identifiers[2]. Upon locating one or more identifiers, a new Gadget is appended to the Blip that submits each to a proteomics data retrieval servlet. This servlet queries a proteomics data resource holding PRIDE formatted XML documents[3,4]. If mass spectra matching the UniProt identifier are extracted, these are converted by Google Charts into image URLs, which are then embedded with the Gadget HTML markup, effectively appending the chart image to the end of the Blip. The Gadget also provides the facility to navigate through multiple spectra if appropriate (see fig. 2).



Did we successfully identify TPIS_YEAST? I thought that this was underexpressed?

Figure 2. Screen-capture showing a mass spectrum Gadget, automatically appended to a Blip by the waveomics Robot upon detection of the UniProt protein identifier TPIS_YEAST.

Enzyme kinetics

A similar process is followed for enzyme kinetics data. Enzyme Classification (EC) numbers[5] are extracted from Blips, with an enzyme kinetics data retrieval servlet querying the MeMo-RK database[6] to returning enzyme kinetics assay data sets that are converted into image URLs as above.

Acknowledgements

The author thanks friendfeed's Life Scientists for fostering an interest in Google Wave within the bioinformatics community.

Funding: This work was supported by the Biotechnology and Biological Sciences Research Council and the Engineering and Physical Sciences Research Council [BB/C008219/1].

[1] Neylon C. Head in the clouds: Re-imagining the experimental laboratory record for the web-based networked world. *Autom Exp*. 2009, **1**:3.

[2] The UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Res*. 2008, **36**, D190-D195.

[3] Jones P, et al. (2006) PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res*. 2006, **34**, D659-D663.

[4] Swainston N, et al. A QconCAT informatics pipeline for the analysis, visualization and sharing of absolute quantitative proteomics data. *Proteomics*. (Submitted).

[5] Webb EC. Enzyme nomenclature 1992. *Academic Press, San Diego, California*.

[6] Swainston N, et al. Enzyme kinetics informatics: from instrument to browser. *FEBS J*. 2010 Aug 3 [Epub ahead of print] DOI: 10.1111/j.1742-4658.2010.07778.x