Databases and Ontologies

Dasty2, an Ajax protein DAS client

Rafael C. Jimenez1,2,#, Antony F. Quinn1,#, Alexander Garcia3,#, Alberto Labarga1, Kieran O'Neill1,2, Fernando Martinez2, Gustavo A. Salazar2, Henning Hermjakob1,*

1Proteomics Services Team, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, CB10 1SD, UK.
2National Bioinformatics Network, NBN Central, Raapenberg Rd, Pinelands, 7405, Cape Town.
3Faculty of Linguistics and Literary Sciences, University of Bremen, Bremen, Germany

ABSTRACT

Summary: Dasty2 is a highly interactive web client integrating protein sequence annotations from currently more than 40 sources, using the Distributed Annotation System (DAS).

Availability: Dasty2 is an open source tool freely available under the terms of the Apache License 2.0, publicly available at http://www.ebi.ac.uk/dasty/

Contact: hhe@ebi.ac.uk

1 INTRODUCTION

We have developed Dasty2, a lightweight, modular DAS client for researchers who aim to visualise, share and compare protein sequence annotations using the Distribution Annotation System (DAS). DAS is a lightweight system for integrating data from multiple heterogeneous distributed databases (Dowell et al., 2001). DAS provides a client-server communication protocol to exchange biological annotation in XML format via HTTP requests. It allows integrating DNA or protein annotations from multiple sources in a web browser, rather than in large central databases. DAS allows annotations to be visualised in the DAS client while control over the original source remains with the data provider (Prlic et al. 2005). Unlike SPICE, a Desktop Java-based DAS client for protein sequence and structure views (Prlic et al. 2005), Dasty2 is a web application that only requires a web browser. Compared to other web clients like Ensembl ProtView (Flicek et al. 2007), Dasty2 increases responsiveness and interactivity by providing a rich user experience and mimics the efficiency of desktop applications. Asynchronous loading and local caching of results enabled by AJAX improves usability and system response time (Garrett et al. 2005).

2 RESULTS

Dasty2 retrieves, integrates and visualizes protein sequence information and facilitates rapid searches to find, share and compare annotations for the protein of interest. The graphical layout of Dasty2 is composed of panels which display different related protein information such as sequence details, positional annotations, literature references and the protein structure, if present in PDB. Additionally it includes other panels which complement the user interface by providing additional functionality such as protein ID search and ontology-based filtering. Panels are interconnected, presenting an interactive display. For instance, the action of clicking on a positional feature highlights its coverage on the sequence and on the protein structure and it shows a pop-up window with more details about the feature. Dasty2 improves the rich user interface allowing users to expand and collapse sections, zoom and resize the graphic, display extra features, drag and drop feature lines, as well as to highlight and sort features.

The modular architecture of Dasty2 allows the inclusion of further panels to add functionality as well as the use of only selected panels; for instance, OntoDas (O'Neill et al. 2007) allows users to build conceptual queries using the Gene Ontology (GO). For this purpose an additional panel displaying ontology annotations and enabling query construction was created. As more modules are added the number of panels will increase. Figure 1 illustrates how the initial information is displayed. Some of the panels are:

- Search panel: Dasty2 provides a text box to search for proteins using UniProt identifiers and accession numbers. The search can be restricted to a certain number of DAS servers classified in the DAS Registry by a specific label, for example a large, collaborative annotation project like Biosapiens (Valencia 2005). The search panel includes a drop-down list of registered labels.

- Status panel: Dasty2 typically queries dozens of servers in response to a user query. To optimize the user experience, data will be displayed as soon as the first server replies are received. A progress bar on the left shows the percentage of DAS servers that have been processed, while status, error and warning messages are displayed under the “System information” section on the right. The “Annotation Sources Loaded” section shows the status of each DAS source being loaded, with links to the raw XML data.

- Sequence panel: Displays the protein identifier, amino acid sequence and length.

- Configuration panel: Contains options to control the display of the positional features panel. Users can zoom into the sequence, or show/hide columns and annotations from specific DAS sources.

- Positional features panel: Shows annotations which are located on specific regions of the protein, for example signal sequences and functional domains. Annotations are aligned against the protein sequence, allowing visual comparison of features, for example predicted signal sequences from SignalP (Bendtsen et al. 2004) versus experimentally determined signal sequences from UniProt. The feature colours are defined by the UniProt DAS stylesheet.

To whom correspondence should be addressed.

*These authors contributed equally to this work.
which provides details on how the annotations should be rendered. Separate columns show the feature type and source of a feature. On mouse-over, a popup window displays annotation details on the individual feature, usually also containing a link back to the original annotation source for additional detail. To support visual comparison, “drag and drop” can be used for manual reordering of features.

- Non-positional features panel: Displays annotations with no positional coordinates, for example protein synonyms or literature references.

- Filtering panel: Presents three tree menus allowing customization of the display by filtering positional and non-positional annotations by type, server name and category. The hierarchical organization of feature types allows efficient focus on annotation types of interest, by including/excluding whole annotation groups like structural annotation with a single mouse click.

- Protein structure panel: Provides a 3D view of the protein structure whenever PDB structures are available for the UniProt ID requested. If Dasty2 finds more than one structure, it includes a drop-down list to chose among them. Users can chose three defined sizes to display the structure and have an option to view it in a pop-up window.

![Figure 1](image.png)

Figure 1. A view of Dasty2 and its panels.

Technically, Dasty2 retrieves annotations mapped to protein sequence coordinates from remote DAS servers, taking as input UniProt identifiers or accession numbers (Apweiler et al. 2004). The client establishes an initial connection to the DAS Registry (Prlic et al. 2007) and retrieves a list of available DAS annotation servers. The client then connects to the UniProt DAS reference server (Jones et al. 2005) to retrieve the protein sequence, and then retrieves feature annotation from each DAS annotation server. Dasty2 merges the collected data from all these servers and provides the user with a unified, structured and interactive view of the sequence and its annotated features.

Dasty2 is based on a client-server architecture. The retrieval module on the server side combines annotations from multiple DAS servers and makes them available to the visualisation module. The visualisation module displays the information obtained from the retrieval module. The visualisation module is based on Ajax and runs on all major web browsers. It has been thoroughly tested on Mozilla Firefox 1.5 and 2.0, and Windows® Internet Explorer 6.0 and 7.0. The retrieval module has been written in Perl (CGI) and runs on a HTTP servers like Apache. Dasty2 can be accessed at http://www.ebi.ac.uk/dasty, but can also be downloaded and customised for third party web sites.

3 CONCLUSIONS AND FUTURE WORK

Dasty2 is not just a report of annotations but an interactive display to conveniently manipulate and visualise protein sequence information. This facilitates the interaction between the user and the information stored in DAS servers. The use of Ajax technology improves its efficiency when displaying information as it delivers a richer user-experience; this technology also reduces the server workloads and increases the responsiveness of the application.

As the recognition of the Distributed Annotation System for provision of protein information increases, the load on the Dasy2 client increases through more DAS servers and more features per server. Currently, filtering for relevant results is performed on the client side after querying all available sources. Future versions of Dasty2 will include targeted querying facilities to reduce server and client workload, and increase client responsiveness.

ACKNOWLEDGEMENTS

This work has been supported by the European Commission within its FP6 Programme, contract number LSHG-CT-2003-503265 (BioSapiens); the European Bioinformatics Institute; the National Bioinformatics Network of South Africa; and the Spanish Ministerio de Educacion y Ciencia.

REFERENCES