BBMS++ – Basic Bioinformatics Meta-searcher

Márcio S. Carocho[1] and Sérgio Deusdado[2]

[1]Superior Agrarian School, Polytechnic Institute of Bragança, 5301-855 Bragança, Portugal

[2]CIMO – Mountain Research Center, IPB, 5301-855 Bragança, Portugal,
`sergiod@ipb.pt`

Abstract   In this paper we present a Basic Bioinformatics Meta-searcher (BBMS), a web-based service aiming to simplify and integrate biological data searching through selected biological databases. BBMS facilitates biological data searching enabling multiple sources transparently, increasing research productivity as it avoids time consuming learning and parameterization of different search engines. As a complementary service, BBMS provides insight and links to common online bioinformatics tools. Users' feedback when evaluating BBMS in terms of usability, usefulness and efficiency was very positive.

Keywords: meta-searching, searching biological databases

Introduction

Due to the increasing amount of biological information available on the Internet, online data and services integration is crucial to foster bioinformatics research and improve searching methodologies in biological databases [1], [2], [3]. Although modern biological databases are well organized and allow different types of manipulation, the searching processes frequently involve loading several databases portals requiring the user to learn each interface peculiarities.

Meta-searching automatically replicates the searching process across multiple databases by using user queries to reach selected searching engines, subsequently integrating, arranging and presenting the multiparty retrieved results [4], [5]. The need for a quick search through several databases has encouraged academics and companies to develop meta-searchers based solutions [6], [7].

Different meta-searchers are present in literature [8],[9],[10],[11] and inspired us to develop a similar but simpler project. Our application was named Basic Bioinformatics Meta-searcher (BBMS) and aims to be a simple and user friendly tool to retrieve biological information transparently from the major biological databases publicly available. The presented version of BBMS was enhanced from its first version [12], and so we refer the new version as BBMS++.

Related Work

Bioinformatics meta-searchers are used to reach biological information spread throughout different databases and display all the relevant results in a unified site. The Bioinformatic Harvester [9] (harvester.fzk.de/harvester/), the Bioextract Server [10] (www.bioextract.org/) and WAVe [11] (bioinformatics.ua.pt/WAVe/) are, currently, three representative examples of meta-searching web-based services.

Regarding the Bioinformatic Harvester, it provides a complete and well organized search across many different aspects of molecular biology, specially genomics and proteomics. The Harvester home page displays all the databases it searches through. It is mainly oriented to search for information related to specific organisms, such as: human, mouse, rat, zebrafish and arabidopsis. Additively, there is an input box that extends the search across the "entire Internet".

The Bioextract Server is an advanced meta-searcher that allows searching through many databases, hierarquically organized and displayed as a tree. Moreover, it provides many tools for alignments, phylogeny and nucleic acids, as well as other bioinformatics services. The Bioextract interface is simple but complete. We may choose where our query will be propagated as it allows choosing specific databases by ticking checkboxes in the desired leafs of the tree. All the actions performed on the Bioextract Server are recorded and can be accessed on the workflow tab.

The web-based application WAVe - acronym for Web Analysis of the Variome, was developed at the University of Aveiro, Portugal. It counts on an tree based interface to explore biological data from several specialized locations in the web, and is described as a variome integration application, focused on providing a centralized access to online available locus-specific databases and genomic variants. WAVe presents essential gene-related information in a simple tree, combining information gathered from available locus specific databases with the most relevant gene-related data available.

Developed Work

Our application was developed to search through specific biological databases for any type of organic material information, such as genes, nucleotides, proteins, biological pathways and species, as well as for related scientific literature. BBMS++ can quickly retrieve listings of biologic material from the most representative biological databases, simply by introducing a query and parameterizing a basic interface. Invoking the selected databases, BBMS++ retrieves the results and displays them on the results frame. The response time of our meta-searcher is the same as any search on the database itself as no significant overhead was introduced. If the translate option was chosen the result page is processed by Google Translate web service to provide translated results.

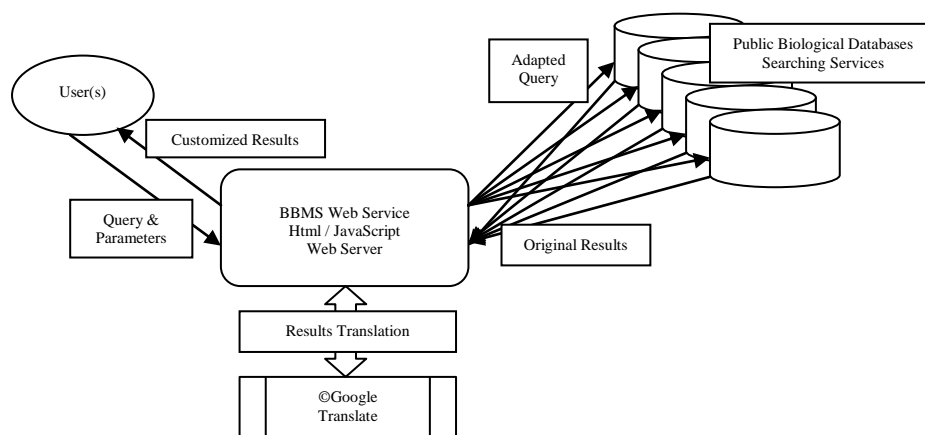The functional architecture of BBMS is depicted in Fig. 1.

Fig. 1. BBMS functional architecture.

A beta version of the BBMS++ is publicly available at http://www.esa.ipb.pt/bbms. The web-service is available in English or Portuguese interface. Its functionalities were developed using the JavaScript language, and the source code was made available. In this way, BBMS is platform independent and all the meta-search functions are open source. The main interface is composed of an input box, three combo boxes and a group of radio buttons. The input box is used to introduce the search criteria. The first combo box is related to the biological database types and it displays five search options: (*i*) primary databases, (*ii*) protein databases, (*iii*) motifs databases, (*iv*) scientific papers and, finally, (*v*) a global search through the Web. The specific choice of the data provider and their database divisions are available in the subsequent combo boxes, working in an interdependent manner, i.e, by choosing an option on the first combo box, the second one is activated and filled with the appropriated options, presenting the different databases related to the first choice. Finally, the third combo box allows the user to choose the type of biologic material, such as proteins, sequences, enzymes, biological pathways, etc.

After the input process, our application forms a string of parameters, according to the specificities of each biological database engine search, in order to append it to the basic URL, forming the complete URL to obtain the desired response from the database. If required by the user, the formed link uses Google Translation services to provide the results in the selected language.

Usability, simplicity and efficiency were mandatory in the development of BBMS and, in the next section we present some results from users' evaluation. A snapshot of the BBMS main interface is presented in Fig. 2.
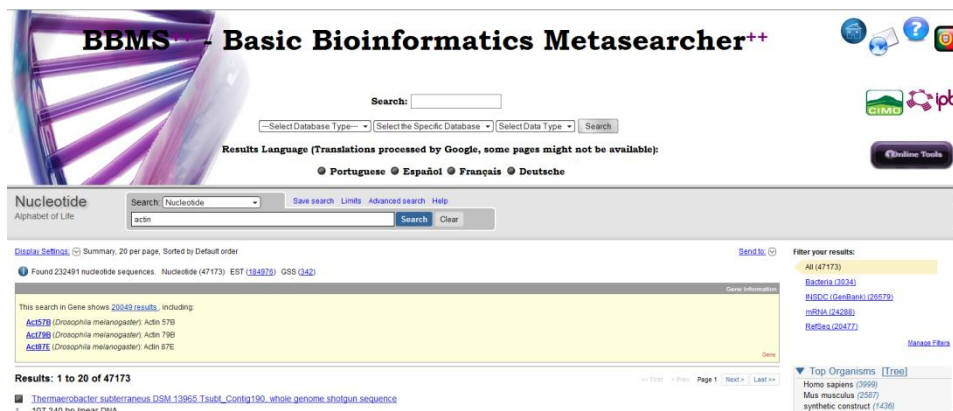
Fig. 2. BBMS interface snapshot.

The results appear on the bottom frame and, by default, are displayed in English. The developed meta-searcher allows the results to be translated into three alternate languages: Portuguese, Spanish and German, simply by selecting that option in the respective radio buttons. The BBMS also includes a help section that provides insight on the specific terminology used.

As a complimentary service, BBMS++ web page also provides access to online tools to solve common bioinformatics needs such as: Blast, ORF-Finder, or protein modelling.

Results and Discussion

The BBMS was meant to be a quick, easy and light web-based service. It was developed having in mind the academic community devoted to the biosciences, especially those without informatics or bioinformatics technical knowledge. Major public biological databases were included in BBMS. However, some databases like DDBJ (DNA Database of Japan) could not be incorporated because the searching commands accepted by that database are not open, not allowing external applications to inquire the database.

Additionally, we considered important to overcome language barriers and, thus, introduced the possibility of automatic translation services, at the interface level and also in the results presentation language.

In order to assess the usability, usefulness and efficiency of our application we selected 30 testers among the students of our institution, 20 of them having little or no experience dealing with biological databases. The rest are frequent biological databases users. After their user experience with BBMS++ they answered to an opinion pool to express their evaluation focusing three factors: usability, usefulness and efficiency. All the questions were answered having the same scale of five levels, ranging from 1 to 5, where 1 means the worst evaluation and 5 the best.

The achieved results, expressed in percentage obtained by each level, are presented in the following Fig. 3, which contains the corresponding three pie charts for: (a) usability; (b) usefulness and (c) efficiency.
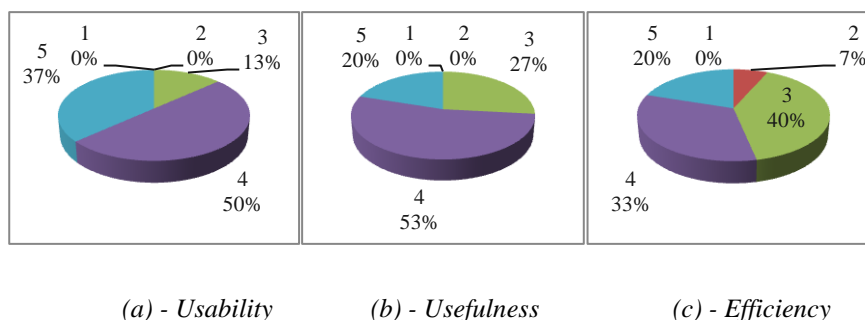

(a) - Usability   *(b) - Usefulness*   *(c) - Efficiency*

Fig. 3. BBMS users' evaluation results.

Even considering it a subjective evaluation, and admitting a reduced number of testers, the results were very satisfactory and encouraging. The majority of the respondents considered the BBMS++ as a very usable, very useful and very efficient web service. The results in the subgroup of the experienced users were even more positive than the general results.

Conclusion

Bioinformatics is a fast growing science producing a huge volume of data, not always well organized or easily accessible. Meta-searchers can simplify the search for biological data, integrating the access to the major databases in one site, avoiding time consuming parameterizations of searching processes in different sites/interfaces. In this way, we developed a Basic Bioinformatics Meta-searcher (BBMS), a web-based service to easily explore the major biological databases publicly available, designed to be very intuitive and requiring minimal parameterization. BBMS++ relief users from using/learning multiple databases interfaces, helping them to increase research productivity. As the presented meta-searcher does not edit the original results, just presents them as obtained directly from the database, the experienced users can interpret them exactly as they are used to. Concomitantly, the useful related links and complimentary services provided by the database, related to the search results, remain accessible as well.

According to testers' experience, BBMS++ is a usable and efficient web service, useful both for beginners and experienced biosciences students or professionals. BBMS is not a finished application and we hope to continue its development in order to improve the features and the integration it provides.

6

References

1. Akula, S., Miriyala, R., Thota, H., Rao, A., Gedela, S.: Techniques for integrating - omics data. Bioinformation. 3, 284-286 (2009).
2. Goble, C., Stevens, R.: State of the nation in data integration for bioinformatics. Journal of Biomedical Informatics. 41, 687-693 (2008).
3. Stein, L.D.: Integrating biological databases. Nature Reviews. Genetics. 4, 337-345 (2003).
4. Dreilinger, D., Howe, A.E.: Experiences with selecting search engines using metasearch. ACM Transactions on Information Systems. 15, 195-222 (1997).
5. Meng, W., Yu, C., Liu, K.: Building efficient and effective metasearch engines. ACM Computing Surveys. 34, 48-89 (2002).
6. Vailaya, A.: An architecture for biological information extraction and representation. Bioinformatics. 21, 430-438 (2004).
7. Lacroix, Z.: Biological data integration: wrapping data and tools. IEEE Transactions on Information Technology in Biomedicine. 6, 123-128 (2002).
8. Diehn, M., Sherlock, G., Binkley, G., Jin, H., Matese, J.C., Hernandez-Boussard, T., Rees, C.A., Cherry, J.M., Botstein, D., Brown, P.O., Alizadeh, A.A.: SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. Nucleic Acids Research. 31, 219-223 (2003).
9. Liebel, U.: 'Harvester': a fast meta search engine of human protein resources. Bioinformatics. 20, 1962-1963 (2004).
10. Lushbough, C., Bergman, M.K., Lawrence, C.J., Jennewein, D., Brendel, V.: BioExtract Server - An Integrated Workflow-Enabling System to Access and Analyze Heterogeneous, Distributed Biomolecular Data. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 7, 12-24 (2010).
11. Lopes, P., Oliveira, J.L.: A Holistic Approach for Integrating Genomic Variation. Information, presented at JBI2010, Málaga, Spain, October, (2010).
12. Carocho, M., Deusdado, S.: BBMS - Basic Bioinformatics Meta-Searcher, presented at JBI2010, Málaga, Spain, October, (2010).