

NCBO-Galaxy: bridging the BioPortal web services and the Galaxy platform

Jose Antonio Miñarro Giménez^{1*}, Mikel Egaña Aranguren², Jesualdo Tomás Fernández-Breis¹ and Erick Antezana³

¹School of Computer Science, UM, Spain

²Ontology Engineering Group, Department of Artificial Intelligence, School of Computer Science, UPM, Spain

³Department of Biology, NTNU, Norway

1 INTRODUCTION

BioPortal (Noy *et al.* (2009)) is a web-based application for searching, sharing, visualizing, and analyzing bio-ontologies. It has become one of the major, centralised bio-ontologies repositories. BioPortal not only hosts a considerable number of important biomedical ontologies (currently almost 300 ontologies covering various life science domains) but also provides access to its contents via RESTful web services, which are a flexible means for programmatically exploiting the stored ontologies. Consequently, its usage should be promoted in bioinformatics environments for facilitating the usage of bio-ontologies.

The lack of integration of bio-ontologies and semantic tools with traditional bioinformatics suites is a major reason for the limited usage of bio-ontologies by bioinformaticians. Galaxy (Goecks *et al.* (2010)) is a web-based platform offering a one-stop-shop of common bioinformatics tools enabling biological data analyses. The so-called Galaxy tools are executed within an environment that keeps an execution history as well as the output of each executed tool, which can be easily shared and reproduced. Even though Galaxy offers a wide range of tools, and recently, some efforts have provided a few tools for ontology manipulation: ONTO-toolkit (Antezana *et al.* (2010)), OPPL-Galaxy (Aranguren *et al.* (2012)), and Blast2GO (Conesa *et al.* (2005)). Each of them offers a complementary functionality, but none of them provides a mechanism to exploit directly a repository of bio-ontologies, such as BioPortal, without having to upload them prior their exploitation.

Therefore, providing Galaxy users with direct access to the BioPortal ontologies seems an interesting option. In this work, we describe the development of a set of Galaxy tools, called NCBO-Galaxy, which provide BioPortal functionalities via its set of RESTful web services (Whetzel *et al.* (2011)). Such a coupling enables the development of advanced analysis workflows, which eventually could improve data curation and management processes.

2 NCBO GALAXY

NCBO-Galaxy has the following components (depicted in Figure 1):

- The Galaxy platform facilitates developing and sharing new tool definitions through a common web interface.
- The NCBO-Galaxy tools provide Galaxy users with the functionality of the BioPortal services.

- The NCBO RESTful services, which belong to NCBO BioPortal, allow to access BioPortal content.

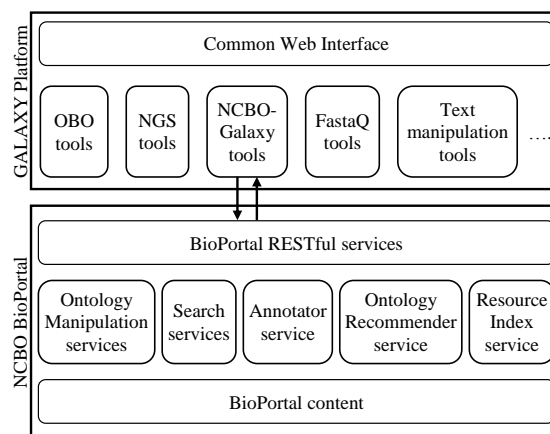


Fig. 1. Component architecture of NCBO-Galaxy.

Each NCBO-Galaxy tool provides the functionalities according to its RESTful service. All the tools have a web interface through which the user can provide the values and preferences for the execution of such service. For example, an excerpt of the interface of the NCBO Galaxy tool for annotating a text with bio-ontology terms is displayed in Figure 2. The list of tools included in the current version of NCBO Galaxy are:

- Get ontology by its identifier,
- Extract a branch from an ontology,
- Get a concrete view of an ontology,
- Annotate a text with bio-ontology terms,
- Recommend a bio-ontology depending on annotations of a text,
- Search for terms in bio-ontologies depending on the text provided,
- Search for resources matched up with terms in bio-ontologies.

All these tools can be combined in Galaxy workflows, as it will be illustrated in the next section.

*To whom correspondence should be addressed: jose.minyarro@um.es

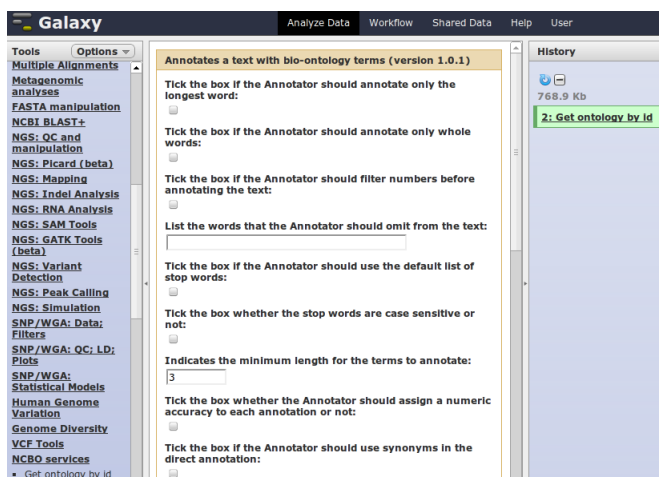


Fig. 2. Excerpt of the web interface of the NCBO Galaxy tool for annotating a text with bio-ontology terms

3 EXAMPLE

Let us describe how we could retrieve the biological entities that have been annotated as participating in both positive and negative regulation processes by combining NCBO Galaxy with other Galaxy tools and, therefore, implemented as a Galaxy workflow. Galaxy permits the graphical definition of the workflow. Our example workflow is shown in Figure 3.

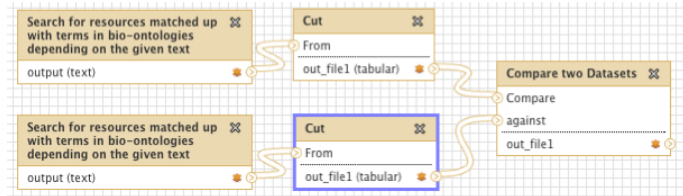


Fig. 3. Description of how different NCBO services can be combined in the Galaxy environment

This workflow consists of two executions of the NCBO Galaxy tool "Search for resources matched up with terms in bio-ontologies", whose input text will be, respectively, positive regulation and negative regulation. The results are two sets of annotations, and an extract of the results for "negative regulation" are shown in Table 1.

Table 1. Some results of the execution of the NCBO service for "negative regulation"

GEO	GSE17559	73.8	44354/obo:INO_0000068
GEO	GSE23920	73.8	44354/obo:INO_0000068
GEO	GSE4724	72.0	44354/obo:INO_0000068
GEO	GSE12205	66.6	44354/obo:INO_0000068
GEO	GSE7849	59.4	44354/obo:INO_0000068

Then, the cut operations allow us to select the first two columns in order to compare the datasets. The final result would contain the ones that appear in both datasets. Both the data used for this example and the workflow are available, respectively, at <http://sele.inf.um.es:9080/u/jesualdo/h/history-demoicbo> and <http://sele.inf.um.es:9080/u/jesualdo/w/demo-ncbo-galaxy>.

4 RESULTS AND FUTURE WORK

NCBO-Galaxy is a prototype aiming at serving ontologists and bioinformaticians who are interested in exploiting controlled vocabularies as part of their data analysis procedures from within the same framework, namely Galaxy, without the hassle of changing tools or working-environments. Moreover, NCBO-Galaxy provides a unique solution to access a major ontology repository (BioPortal). Finally, we plan to extend the tools to access other NCBO services (such as BioPortal SPARQL).

5 AVAILABILITY

NCBO-Galaxy is available at the following public instances: <http://sele.inf.um.es/galaxy> and <http://linkeddata2.dia.fi.upm.es:8080/> and it can be downloaded from the Galaxy Toolshed, available at http://toolshed.g2.bx.psu.edu/repos/mikel-egana-aranguren/ncbo_services.

ACKNOWLEDGEMENTS

Mikel Egaña Aranguren is funded by the Marie Curie Cofund programme (FP7). This work has been possible thanks to the funding of the Spanish Ministry of Science and Innovation through grant TIN2010-21388-C02-02.

REFERENCES

Antezana, E., Venkatesan, A., Mungall, C., Mironov, V., and Kuiper, M. (2010). ONTO-Toolkit: enabling bio-ontology engineering via galaxy. *BMC bioinformatics*, (Suppl 12), S8+.

Aranguren, M. E. n., Fernández-Breis, J. T., and Antezana, E. (2012). OPPL-Galaxy: enhancing ontology exploitation in galaxy with OPPL. In *Proceedings of the 4th International Workshop on Semantic Web Applications and Tools for the Life Sciences, SWAT4LS '11*, pages 12–19, New York, NY, USA. ACM.

Conesa, A., Gtz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy Team (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology*, **11**(8), R86+.

Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A., Chute, C. G., and Musen, M. A. (2009). Biportal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, **37**(suppl 2), W170–W173.

Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., and Musen, M. A. (2011). Biportal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic Acids Research*, **39**(suppl 2), W541–W545.