

# GOssTo & GOssToWeb: user-friendly tools for calculating semantic similarities on the Gene Ontology

Horacio Caniza Vierci<sup>1</sup>, Alfonso E. Romero<sup>1</sup>, Samuel Heron<sup>1</sup>, Haixuan Yang<sup>1</sup>, Marco Frasca<sup>2</sup>, Marco Mesiti<sup>2</sup>, Giorgio Valentini<sup>2</sup>, Alberto Paccanaro<sup>1</sup>

<sup>1</sup>Department of Computer Science and Centre for Systems and Synthetic Biology, Royal Holloway, University of London, Egham, TW20 0EX, UK.

<sup>2</sup>Dipartimento di Informatica, Università degli Studi di Milano, via Comelico 39/41, 20135 Milano, MI, Italia.

---

## 1 INTRODUCTION

Semantic similarity measures have become important in bioinformatics as they quantify relatedness between genes in a way that is complementary to both experimental information and sequence-based approaches. Several semantic similarity measures have been proposed for the Gene Ontology.

Recently, Yang *et al.* (2012) introduced the Random Walk Contribution, a kind of “add-on” for any existing semantic similarity measure: a measure is modified to take into account the structure of the GO DAG beneath the terms as well as the uncertainty of GO annotations. Yang *et al.* were able to show a large increase in the performance of existing semantic similarity measures when they were combined with their Random Walk Contribution.

In this poster we present GOssTO (Gene Ontology semantic similarity Tool) and its online counterpart GOssTOWeb, two new tools for calculating semantic similarities. GOssTO and GOssTOWeb support both graph-based as well as term-based similarity measures and implement Yang’s *et al.* (2012) Random Walk Contribution.

## 2 METHODS

GOssTo is bundled with six commonly used semantic similarity measures: the ones proposed by Resnik, Lin, Jiang and Conrath, as well as GraSM (Couto *et al.*, 2007.), simUI and simGIC (Pesquita *et al.*, 2008). All these measures are extended with Yang’s *et al.* random walk based procedure.

The guiding principles for GOssTo’s design aimed at producing a fast and flexible software package. This resulted in a highly modularized system architecture with very low coupling between the individual modules, which can be readily removed or replaced without affecting the overall behavior of the system.

The user can interact with GOssTo either through the command-line or an application programming interface (API). The command-line interface provides UNIX-like console parameter options as well as an interactive menu; the API provides access to all functionalities in the different modules through a set of well-defined functions.

GOssTo includes a powerful extension mechanism to add new semantic similarity measures. After a new measure is independently compiled, it can be dynamically linked to GOssTo’s application core. The new measure can then be used in the same way as the ones bundled with GOssTo.

GOssTo is available online as GOssToWeb from our server at [www.paccanarolab.org/gosstoweb](http://www.paccanarolab.org/gosstoweb). Through a clean web interface, GOssToWeb provides access to the same functionalities of the stand-alone application, allowing extensive configuration of the experiments through a user-friendly web form.

GOssToWeb provides the user with a searchable list of UniProt-GOA proteomes that are constantly updated to the latest version. The user can select the desired GOA evidence codes, ontological relationships and the similarity measure.

Results are provided by redirecting the user to a page from which they can be downloaded. For large calculations, the user can choose to be notified with an email containing a link to the result download page. Results are downloadable in plaintext CSV files for offline use.

GOssTo was developed using the Java programming language. The Apache Commons Library provides the internal data types and the required mathematical routines. GOssTo’s source code is freely available from github at <http://github.com/pwac092/gossto> and is released under the GPLv3 license.

## REFERENCES

- Couto, F.M. *et al.* (2007) Measuring semantic similarity between gene ontology terms. *Data Knowl. Eng.*, 61, 137-152
- Pesquita, C. *et al.* (2008) Metrics for GO based protein semantic similarity: a systematic evaluation. *BMC Bioinformatics*, 9 (Suppl. 5), S4
- Yang, H., Nepusz, T. and Paccanaro, A. (2012) Improving GO semantic similarity measures using downward random walks. *Bioinformatics*, 28(10), pp. 1383-1389