



Project no. 34721

## **TAGora**

# **Semiotic Dynamics in Online Social Communities**

<http://www.tagora-project.eu>

Sixth Framework Programme (FP6)

Future and Emerging Technologies of the Information Society Technologies (IST-FET Priority)

---

## **Deliverable D5.4: Website and Portal**

---

Period covered: from 01/06/2007 to 31/05/2008	Date of preparation: 31/05/2008
Start date of project: June 1 <sup>st</sup> , 2006	Duration: 36 months
Due date of deliverable: July 15 <sup>th</sup> , 2008	Actual submission date: June 20 <sup>th</sup> , 2008
Distribution: Public	Status: Final

Project coordinator: Vittorio Loreto  
Project coordinator organisation name: "Sapienza" Università di Roma  
Lead contractor for this deliverable: "Sapienza" Università di Roma

## Project website and portal

The website [www.tagora-project.eu](http://www.tagora-project.eu) has been enriched with many features, in the path leading to the delivery of a real portal at the end of the project. [www.tagora-project.eu](http://www.tagora-project.eu) already share some of the characteristics of a portal, including a large number of resources and links to external ones. However, it has to be reminded that our research object, that is, social tagging communities, is an intrinsically distributed one. Therefore, to gather many resources in a centralized manner is a somewhat contradictory task which has to be tackled with caution.

Moreover, due to global business strategies, many portals already exist, with a strong integration with social tagging websites. For example, the portal Yahoo! owns two major sites such as Flickr and del.icio.us. A portal on this matter built from scratch would compete with such giant portals. Therefore, TAGora will explore in the third year the possibility of delivering a portal with a stronger focus on control strategies and leading-edge research in the field, with a distributed approach based on platform developed within the project such as Bibsonomy. However, this delivery needs a full deployment of techniques and algorithms studied and developed during the TAGora program. Thus, an autonomous portal, surviving beyond TAGora lifetime, would be possible only at the end of the project. In the meanwhile, the transitional phase will be carried on by gradually enforcing new tools and capabilities on the [www.tagora-project.eu](http://www.tagora-project.eu) website.

So, the functions already included in the TAGora project website - described in the "Pla for using and disseminating the project", are now being expanded as reported in the following sections.

## Datasets

One of the new features of the [www.tagora-project.eu](http://www.tagora-project.eu) website is the data repository. Here, general public and interested researchers can find datasets reporting the tagging activities in some major online communities. Collected data are generally presented as time-ordered lists of tag assignments, i.e. a table of 4 columns providing the time of a tag assignment, the tagged resource, the tagging user and the employed tag. Additional data and more refined summaries of such information are also provided, such as lists of tags, users and resources with their rank according to several criteria. Each dataset is presented and annotated in order to be fully employed by an interested user. Since the data included in such bases regards the activity of individual users, raising many privacy issues about personal data, a special care has been devoted in anonymizing the datasets as deeply as possible.

The repository so far includes data related to del.icio.us, Bibsonomy, Flickr, Last.fm plus narrower surveys of particular aspects inside them. As far as the project advances, new datasets will be added to it.

## Simulators

The "Data" section contains a subsection called "Simulators", which includes links to software tools producing artificial tag streams and networks.

The first one is the "Epistemic Dynamical Tagging Model" developed at the University of Koblenz, Germany. It produces artificial tag stream built following some accepted model of the dynamical evolution of folksonomies. The resulting stream is statistically similar to real one, and can be used for both the analysis of tag streams and for the development of control strategies to be applied in real systems. The program, which is written in Java, can be downloaded for free from the TAGora website and from the University of Koblenz website, where an extended documentation is also provided.

Second, a software called “NET” is also presented and made available to the public. It is a software tool with a larger scope, written in C and developed by Vito D.P. Servedio of the University “Sapienza” of Rome, Italy. NET provides tools to analyze and simulate random graphs with several algorithms, including the growing scale-free network model introduced by Bárábási and Albert Barabasi and Albert (1999) and the static fitness model introduced by Caldarelli *et al.* Caldarelli *et al.* (2002). Network analysis is strictly related to the study of tagging website, because a number of relevant properties have been detected by projecting the datasets on a tri-partite graph Hotho *et al.* (2006) or, for instance, by building co-occurrence weighted networks, where links connect tags which have been assigned to the same resources Cattuto *et al.* (2007). Therefore, new analysis tools of NET have been developed, and now include the measurements of quantities which are relevant to the study of tagging systems represented by network structures.

## Tutorials

Using social tagging tools is easy and intuitive. However, each tagging websites provides slightly different features in order to serves the users’ community in the most useful and pleasant manner. Thus, many tagging sites now show their capabilities through video-tutorial, short movies in digital format which explain in a few minutes how the basic funtions of each website work. The advantage of such tutorial, beyond their explicative power, is their portability across websites, which allows a website to embed such video-tutorial with great convenience.

We have included a number of them as a resource for the general public, along with a list of nine major social tagging website appeared in the last few years, to give a user a perspective of the state-of-the-art in this area of the World Wide Web.

## Outlook

As mentioned above, the TAGora project foresees to deliver a portal on the matter of semiotics and semantics dynamics taking place on the Internet, living beyond TAGora. In the spirit of the field, the portal will rely heavily on user-generated content. To do this, we will employ the control strategies and tools developed within the project, which at the end will be ready to be implemented and integrated in a single website. For example, both academic and non-academic resources about the field will be provided by a suitable integration of the portal with Bibsonomy.org, where many useful tools are already operating Cattuto *et al.* (2008).

## Bibliography

- A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439): 509–512, October 1999. ISSN 0036-8075. URL <http://view.ncbi.nlm.nih.gov/pubmed/10521342>.
- G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz. Scale-free networks from varying vertex intrinsic fitness. *Phys Rev Lett*, 89(25), December 2002. ISSN 0031-9007. URL <http://view.ncbi.nlm.nih.gov/pubmed/12484927>.
- Ciro Cattuto, Christoph Schmitz, Andrea Baldassarri, Vito D. P. Servedio, Vittorio Loreto, Andreas Hotho, Miranda Grahl, and Gerd Stumme. Network properties of folksonomies. *AI Communications*, 20(4):245 – 262, 2007. URL [http://www.kde.cs.uni-kassel.de/hotho/pub/2007/aicomm\\_2007\\_folksonomy\\_clustering.pdf](http://www.kde.cs.uni-kassel.de/hotho/pub/2007/aicomm_2007_folksonomy_clustering.pdf).
- Ciro Cattuto, Dominik Benz, Andreas Hotho, and Gerd Stumme. Semantic analysis of tag similarity measures in collaborative tagging systems. In *Proceedings of the 3rd Workshop on Ontology Learning and Population (OLP3)*, Patras, Greece, July 2008. URL <http://olp.dfki.de/olp3/>.
- Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Folkrank: A ranking algorithm for folksonomies. In *Proc. FGIR 2006*, page to appear, 2006. URL <http://www.kde.cs.uni-kassel.de/stumme/papers/2006/hotho2006folkrank.pdf>.