



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)

Periodic Management Report

Period covered: from 01/06/2008 to 31/08/2009

Date of preparation: 15/09/2009

Start date of project: June 1st, 2006

Duration: 39 months

Due date of deliverable: October 15th, 2009

Actual submission date: September 15th, 2009

Distribution: Public

Status: **Draft**

Project coordinator: Vittorio Loreto

Project coordinator organisation name: Sapienza Università di Roma

Lead contractor for this deliverable: Sapienza Università di Roma

Contents

1	Justification of major cost items and resources	3
1.1	PHYS-SAPIENZA	4
1.2	SONY-CSL	6
1.3	UNI KO-LD	7
1.4	UNIK	9
1.5	UNI-SOTON	10
1.6	Tabular overview of budget and actual person-months	12
1.7	Cost budget follow-up table	14
1.8	Major deviations from cost and person-month budget	15
1.8.1	Major deviations from Cost Budget	15
1.8.2	Major deviations from Person-Month Budget	15
2	Form C Financial statement per activity for the contractual reporting period, to be completed by each contractor	16
3	Summary financial report	19

Chapter 1

Justification of major cost items and resources

In this report we provide a justification of the major costs incurred and resources deployed by the contractors, linking them to activities implemented by each contractor.

Section 1.6 provides a tabular overview of budgeted and actual person-months.

Section 1.7 includes a tabular overview of budgeted and actual costs.

Section 1.8 describes the major deviations from cost and person-month budget (see also the deviation sections of each WP and section 3 of the Periodic Activity Report).

Section 2 includes the financial statement (FORM C) and Audit certificate of the contractors.

Section 3 includes the summary financial report of the total costs claimed by each contractor.

1.1 PHYS-SAPIENZA

Brief description of the work performed

WP1: (Planned man-months: 5, actual person months: 7.5)

In order to test our Semantic Walker Model (see WP4 below), we collected, analyzed and reformat the South Florida Free Association Norms database (freely available at <http://w3.usf.edu/FreeAssociation/> website) that includes roughly 60000 associations. Though this is not precisely a folksonomy database its adoption has revealed crucial as a proxy for a latent shared semantic graph.

WP3: (Planned man-months: 6.92, actual person months: 15.5)

We studied the bursty tagging activity occurring in a folksonomy in order to understand whether different bursty events, which are the result of the apparently uncorrelated action of users, might lead to any loose tag categorization.

We examined, by using the tool of the Inverse Participation Ratio, whether users reach a spontaneous consensus on the description of resources in terms of significant tags.

We studied whether social interaction carries any influence on the semantic relatedness between users, i.e. whether users supposed to be in close social relationship are also prone to use more similar tags. To this aim, we analyzed the evolution in time of the semantic assortativity in the social networks hosted by the Flickr folksonomy, based both on the contact data and on the group membership data provided by the users themselves.

In order to understand whether tag order in a post carries any semantic meaning we analyzed a directed version of the cooccurrence network of tags.

WP4: (Planned man-months: 28, actual person months: 40)

We extensively analyzed the Semantic Walker Model (SWM), mainly focusing on its robustness and dependence on the underlying network, as also suggested by the reviewers of the project. We also tested the SWM with the real network of associations available at the South Florida Free Association Norms database, described above in the WP1 paragraph.

We have studied the statistical properties of the post structure in real folksonomies. The better characterization of the post structure has been used in a further improvement of the Epistemic Model and as a reference for the SWM.

Together with UNIK, we set up the theoretical basis of the user recommendation system in Bibsonomy, by studying various tag-cloud based semantic distances.

WP5: (Planned man-months: 1, actual person months: 3)

In connection with UNI-SOTON we carried the Live Social Semantics Experiment at the European Semantic Web Conference 2009 and at HyperText Conference 2009, with the objective to illustrate possibilities of utilising various TAGora technologies for the analysis of social connectivity of conference participants.

We equipped our dissemination activity with publications on international journals, conference and workshop presentations and organization.

WP6: (Planned man-months: 3, actual person months: 3)

As coordinator of the project, we: (a) were responsible for the day-to-day condition of the project, (b) we organized the TAGora periodic meetings, (c) we acted as the main interface between the contractors and the European Commission, (d) we distributed the financial contribution to the partners, (e) we coordinated the preparation of deliverables and Periodic Reports.

Description of major cost items

The following numbers are rough estimated, and can only be confirmed once the Auditing is completed.

- **Personnel costs:** 1 Full Professor (4.8 man-months, 57.606,08 EURO), 2 Associate Professors (20 man-months, 116.957,60 EURO), 3 Post-Docs (44 man-months, 128.166,71 EURO).
- **Personal computers:** (depreciation for the server for data analysis bought in the 1st year and a Computer bought in the third year, 5.320,27 EURO).
- **Other minor costs** (including: travel expenses, 24.280,19 EURO approx.; Contribution to the organization of the Hypertext 2009 Conference in Turin, 8.681,82 EURO; Cost for the audit certificate for the II year, 2.702,70 EURO; Cost for the audit certificate of the III year, 2.500 EURO approx.)
- **Overheads:** (167.167,72 EURO)
- **Total costs:** (513.383,09 EURO)
- **Requested Contribution:** (276.236,40 EURO)

1.2 SONY-CSL

Brief description of the work performed

WP1: (Planned man-months: 4.1, Actual man-months used: 2.5)

The work includes the continuation of the gathering of data of the Phenotypes/Limited Forms project and the start of the data gathering for the project Noisetube about tagging noise pollution.

WP2: (Planned man-months: 8.47, Actual man-months used: 10.3)

Most of the time was used to consolidate and extend the Zexe.net platform into Noisetube.net. To adapt it for noise pollution and the urban context, we extended the mobile application with a real-time signal processing algorithm to measure noise level and evaluated its accuracy. To aggregate tags by urban element (e.g. street, district) we improved the localization component by developing algorithm to correct the localization of tags using urban geographical data. Finally we created a new visualization using Google earth with new features. Finally to ensure the reusability of Ikoru, we developed a lightweight API for Flickr images.

WP3: (Planned man-months: 15.75 Actual man-months used: 7.1)

The time was mostly spent on 1) automatic tag prediction from acoustic signals, and 2) so-called Description-Based Design. We also analysed the data of the Armin Linke exposition.

WP5: (Planned man-months: 3.38, Actual man-months used: 8.70)

Dissemination activities consisted mainly in publications: 3 papers for Noisetube, 4 papers for music-related works. The Phenotypes/Limited Forms expositions in different places involved also a lot of works.

Description of major cost items

The following numbers are rough estimated, and can only be confirmed once the Auditing is completed.

- Personnel costs:

- Senior Researcher and Researchers: 8,31 man-months; 69.595,25 EURO (estimated)
- Associate Researchers: 17,50 man-months; 80.438,52 EURO (estimated)
- PhD Student: 2,78 man-months, 3.880,13 EURO (estimated)
- Total: 153.913,90 EURO (estimated)

- Personal computers: 2.117,09 EURO (estimated)

- Travel costs: 2.599,05 EURO (estimated)

- Overheads: 136.195,65 EURO (estimated)

- Audit costs: 1.500 EURO (estimated)

- Total costs: 296.325,68 EURO (estimated)

- Requested Contribution: 126.326,97 EURO

1.3 UNI KO-LD

Brief description of the work performed

WP1: (Planned person months: 1, actual person months: 1.0)

After the major folksonomy crawling activity was finished in the second year, we collected some data from our MyTag system, i.e. query logs and the generated personomy data of the users, in order to see whether we can find similar properties. Additionally, in order to make the data sets available to the public, we checked them for consistency and did some format transformations where necessary.

WP2: (Planned person months: 3.6, actual person months: 5.7)

We continued the work on the MyTag system by improving the result ranking mechanism and the general implementation to achieve better performance and usability. In cooperation with Southampton, we integrated a search assistant in MyTag which provides search term disambiguation based on concepts extracted from wikipedia. The work on Tagster was discontinued as suggested by the reviewers.

WP3: (Planned person months: 4, actual person months: 6.5)

Based on the observation that the sparseness of tagging data is hampering the search result quality in tagging system we have been investigated ways to overcome this problem by enriching the vector space model with inferred data, e.g. based on user or tag co-occurrence. The results were evaluated with respect to precision and recall.

WP4: (Planned person months: 3, actual person months: 10.2)

The epistemic dynamic model was further extended to integrate the simulation of complete tag postings. A more fine-grained evaluation was done to measure the difference between simulated an real tag frequency in the tag streams. Furthermore, we investigated the classification of landmark photos in the Flickr dataset. An extensive evaluation was done with a large photo dataset.

WP5: (Planned person months: 1, actual person months: 2.0)

The dissemination activity was mainly concentrating on publications at different international conferences and presentations of the MyTag system to attract more users.

Description of major cost items

The following numbers are rough estimated, and can only be confirmed once the Auditing is completed.

- **Costs funded by UNI KO-LD:** (46 person months, 302.400 EURO)

- **Personnel costs:** PhD Student and Junior Staff (25.4 person months) 96.973,82 EURO
- **Personal computers:** 1.569,59 EURO (depreciation)
- **Other minor costs** ~16.300 EURO (including: travel and consumables)
- **Overheads:** ~25.000 EURO
- **Total costs:** ~140.000 EURO
- **Requested Contribution:** ~140.000 EURO

- **Permanent Staff costs:** Professor (0 person months)

1.4 UNIK

Brief description of the work performed

WP1: (Planned man-months: 1, actual person months: 1.5 PhD students, 1 technician, 1 junior staff)

We continued publishing BibSonomy dumps every half year. For the ECML PKDD Discovery Challenges 2008 and 2009, we generated additional test and training data sets.

WP2: (Planned man-months: 5, actual person months: 6 PhD students, 2 technician, 2 junior staff)

The functionality of BibSonomy was further enhanced, as described in Deliverable 2.5. In particular, we implemented a logging framework and a spam framework, and set up the design of the recommender framework. The latter had to be implemented outside of Tagora, financed by a national follow-up project, due to the limited Tagora resources. The frameworks were the technical platform for the experiments performed in WP4.

WP3: (Planned man-months: 0.5, actual person months: 0)

Following the recommendations of the reviewers, the resources were shifted to WP4.

WP4: (Planned man-months: 3, actual person months: 7 PhD students, 2 technician, 4 junior staff)

Our work in WP4 covered the following topics: measures for the semantic similarity of tags, tag recommendations, user recommendations, the analysis of user behavior, and spam detection.

WP5: (Planned man-months: 2, actual person months: 2.25 PhD students, 2.5 technician, 4.75 junior staff)

The BibSonomy data were the basis for the ECML PKDD Discovery Challenges 2008 and 2009, as well as for the Vizards session of the Sunbelt 2009 conference. BibSonomy has thus reached a broad recognition in the research community.

Description of major cost items

The following numbers are rough estimated, and can only be confirmed once the Auditing is completed.

- **Personnel costs:** PhD Students (16.75 man-months), Technician (7.5 man-months), Junior Staff (11.75 man-months), \approx 108.250 EURO).

- **Personal computers:** (\approx 2600 EURO depreciation)

- **Other minor costs** (including: travel expenses, consumables, \approx 17.500 EURO; etc.)

- **Overheads:** (\approx 25.700 EURO)

- **Total costs:** (\approx 154.050 EURO)

- **Requested Contribution:** (\approx 125.400 EURO)

- **Permanent Staff costs:** Professor (1 man-months, \approx 9.000 EURO)

1.5 UNI-SOTON

Brief description of the work performed

WP1: (Planned man-months: 3, actual man-months: 4)

Effort on data collection was focused in the last year of the project on supporting our work on cross-folksonomy integration and analysis, and on feeding into the Live Social Semantics application (D4.5). Data We have also generated much data by collecting and processing Wikipedia pages to support our Tag Sense recommendation service. Additionally, data was gathered on-the-fly by our several open services for generating profiles of interests.

WP2: (Planned man-months: 0, actual man-months: 1.5)

Building the Live Social Semantics application was not planned for in the proposal. However, designing and deploying the services that LSS is built one is core to our work in WP3 and WP4. Since we were asked to chair the Semantic Technologies track of ESWC 2009, it was only natural to put our services together to realise LSS, and to use it as a great dissemination opportunity for TAGora technologies. Approximately one and a half months in total were spent on architecting and implementing the back end of LSS which integrates our many services, and on building the LSS front end website.

WP3: (Planned man-months: 1, actual man-months: 2.5)

We have focused our efforts this year on developing and opening access to several services for cross-folksonomy integration and analysis. These services include Tag Filtering, Sense Matching, and Profile Building. We have also extended our tool for generating profiles of interests. These tools and services were used in MyTag as well as in LSS. Please refer to D4.5 for further information and detail.

WP4: (Planned man-months: 4, actual man-months: 4)

Several recommendation services were developed and deployed; for recommending interests, tag senses, and conference talks. These services builds on the one mentioned in WP3 above to provide various recommendations, and were deployed and tested in MyTag and LSS.

WP5: (Planned man-months: 1, actual man-months: 2)

In the third and last year of the project, we have published 1 workshop paper, and 1 conference paper. We have coorganised one workshop at HT2009 with all other TAGora partners. We are also coorganising a workshop with SMI, Stanford University, on collaborative knowledge that is accepted at ISWC 2009. The workshop paper was coauthored with the Facultad de Informática, Universidad Politécnica de Madrid. Additionally, we have launched a major dissemination event at ESWC 2009 and HT 2009 where we deployed the Live Social Semantics experiment.

Description of major cost items

The following numbers are rough estimated, and can only be confirmed once the Auditing is completed.

- **Costs funded by UNI-SOTON:** (36 man-months, 252.840 EURO)

- **Personnel costs:** Post Doc (12 man-months, 36.917,64 EURO), Senior Researcher (2 man-months, 8.588,18 EURO).

- **Personal computers:** (558,40 EURO)
- **Other minor costs** (including: travel expenses, 15.337,00 EURO)
- **Overheads:** (16736,14 EURO)
- **Total costs:** (78.137,36 EURO)
- **Requested Contribution:** (78,137.36 EURO)
- **Permanent Staff costs:** Professor (0.5 man-months, 7.000 EURO), 2 Senior Research Fellow (3.5 man-months, 45.000 EURO).

1.6 Tabular overview of budget and actual person-months

Tabular overview of budget and actual person-months for the third year of activity. In the last line the figures in brackets correspond to the estimate of the number of man-months considering the project extension from 36 to 39 months.

Contract n.	34721	Partner Person-months (PM)										AC - own staff			
		Acronym:	TAGora	PHYS-SAPIENZA			SONY-CSL			UNI KO-LD	UNIK	UNI-SOTON	AC TOTAL	UNI KO-LD	UNIK
Reporting Period:	3														
		TOTALS	PHYS-SAPIENZA	SONY-CSL	UNI KO-LD	UNIK	UNI-SOTON	AC TOTAL	UNI KO-LD	UNIK	UNI-SOTON				
Workpackage 1	Actual PM:	18.5	7.5	2.5	1	3.5	4	2	0	0	2				
Emergent Metadata	Planned PM:	14.1	5	4.1	1	1	3	1	0	0	1				
Workpackage 2	Actual PM:	27.5	0	10.3	5.7	10	1.5	0.5	0.25	0.25	0				
Applications	Planned PM:	17.1	0	8.5	3.6	5	0	1.25	0.25	1	0				
Workpackage 3															
Data Analysis or emergent system properties	Actual PM:	31.6	15.5	7.1	6.5	0	2.5	0.5	0.25	0.25	0				
	Planned PM:	28.2	6.9	15.8	4	0.5	1	1.25	0.25	1	0				
Workpackage 4	Actual PM:	67.2	40	0	10.2	13	4	2.5	0.25	0.25	2				
Modeling and simulations	Planned PM:	38	28	0	3	3	4	2.25	0.25	1	1				
Workpackage 5															
Dissemination and exploitation	Actual PM:	25.2	3	8.7	2.0	9.5	2	0.5	0.25	0.25	0				
	Planned PM:	8.4	1	3.4	1	2	1	1.25	0.25	1	0				
Workpackage 6	Actual PM:	3	3	0	0	0	0	0	0	0	0				
Management	Planned PM:	3	3	0	0	0	0	0	0	0	0				
Total Project	Actual Tot.	173	69	28.6	25.4	36	14	6	1	1	4				
	Planned Tot.	108.7	45.9	31.7	10.6	11.5	9	7	0	4	2				
		(136.1)	(57.8)	(36.7)	(14.4)	(15.3)	(12)								

1.7 Cost budget follow-up table

THE "COST BUDGET FOLLOW-UP TABLE" WILL BE INSERTED BY PHYS-SAPIENZA IN THE FINAL VERSION OF THE MANAGEMENT REPORT.

1.8 Major deviations from cost and person-month budget

1.8.1 Major deviations from Cost Budget

PHYS-SAPIENZA: none

SONY-CSL: none

UNI KO-LD: none

UNIK: none

UNI-SOTON: none

1.8.2 Major deviations from Person-Month Budget

PHYS-SAPIENZA: More person months than originally planned were spend due to project extension as well to a stronger effort devoted to the implementation of the feedback from theoretical approaches and modelling to control in applications. The additional person-month did not change the overall project budget.

SONY-CSL: none

UNI KO-LD: More person months than originally planned were spend due to the employment of less senior personnel. Because of a lack of experience the more junior personnel needed more hours to finish the tasks that more senior personnel would have finished quicker. This is especially true for undergraduate students who are very cheap, but also take much longer. The additional person month were evenly spend on the work packages. The bigger difference for WP4 is due to the fact that most work in WP4 was done at the end of the project.

UNIK: More person months than originally planned were spend due to project extension as well as the employment of less senior personnel. The people hired very less senior than expected. Because of a lack of experience the more junior personnel needed more hours to finish the tasks that more senior personnel would have finished quicker. This is especially true for undergraduate students who are very cheap, but also take much longer.

UNI-SOTON: The planned person-month for the last year of the project was 10 months. However, we have managed to pull in additional resources at no extra cost to the project, to help with service deployment and LSS implementation. These resources proved extremely valuable and were vital to the success of LSS, as an application as well as a dissemination activity. The additional person-month were free and hence it did effect our overall project budget.

Chapter 2

Form C Financial statement per activity for the contractual reporting period, to be completed by each contractor

Forms C and Audit Certificates (**ALL PARTNERS** please provide 2 original signed paper copies of these documents, and send Form C also electronically).

Please send an electronic copy of Form C (before signature) to Antonella Giampaglia (Antonietta.Giampaglia@roma1.infn.it) so that we can check it and eventually suggest modifications before the official signature.

Quick tips for the compilation of Form C

* **General info:** please fill in the European Commission Form C template attached (cost-analysis-strep.xlt), and available on the EU web-site with all the information required. Instructions for use and frequently asked questions are also available in the website <ftp://ftp.cordis.europa.eu>.

Do not modify the form in any part, do not eliminate any part of the form.

* Declaration of eligible costs

- Direct costs are actual costs occurred during the year. Indirect costs consist of overheads.

- As categories of eligible costs are not identified in FP6, eligible costs must be determined in accordance with the contractor's usual accounting principles, as long as these accounting principles meet recognised standards and are not created purely for the EC contract. Eligible costs need to be actual, economic and necessary for the implementation of the project; incurred during the duration of the project; and recorded in the accounts of the contractor; exclusive of any identifiable indirect taxes, including VAT or duties; exclusive of interest owed; and may not give rise to profit (see Article II.19 of the contract).

- Please divide expenses between the categories: (A) Research and technological development,

(B) Demonstration, (D) Management.

- Calculation of depreciation of durable equipment is based on the partner's own depreciation system.

- Overheads:

* FC contractors apply overheads according to their usual accounting principles.

* FCF contractors apply a flat rate of 20 percent of all direct costs minus costs of subcontracts.

* AC contractors apply a flat rate of 20 percent of all direct additional costs minus costs of subcontracts.

- Duties and taxes are not eligible costs. Any identifiable taxes, including airport taxes, are not eligible costs. However, for certain airlines these extra costs are considered as airport costs and not airport taxes, in which case they are eligible cost.

* Receipts

There are three main kinds of receipts: Financial transfers or their equivalent to the contractor from third parties; Contributions in kind from third parties; Income generated by the project.

* Request of FP6 Financial contribution

Please complete Section 5 of Form C. The contractor should indicate the requested financial contribution, calculated by applying to the eligible costs the reimbursement rate according to activity, cost model and instrument.

In the case of FC and FCF partners Research & Development, Demonstration activities and Management expenses are financed by the commission at a rate of 50, 35 and 100 percent respectively. For AC partners expenses are financed with a rate of 100 percent in all cases (see Article II.25 of the contract). Please notice that cost of equipments (i.e. computers) cannot be entirely claimed during the second year: the cost should be divided by 36 months and only part of the cost (up to 12 months depending on the date of purchase) can be claimed.

* Audit Certificate

Each contractor shall provide an audit certificate prepared and certified by an external auditor, certifying the costs incurred during that period meet the condition required by this contract (See Article II.26 of the contract). Please notice that:

- If external auditors are used for the compilation of the audit certificate then the cost must be charged under "subcontracting".

- Audit certificate costs should be claimed under "Management" category.

* Contractor's certificate

Please be sure to complete the form with:

- Contractor's stamp
- Name of the person responsible for the work, date and signature
- Name of the Financial Officer, date and signature

Chapter 3

Summary financial report

The "Summary financial report" is a summary of the individual contractor's Form C information. It includes the total (direct and indirect) costs - as claimed by each contractor and activity type - incurred during each reporting period.

This summary will be inserted by PHYS-SAPIENZA in the final version of the Management Report.