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 5.1: Project presentation report

Due date of deliverable: 31st October 2006
Actual submission date: 31st October 2006

Start date of project: June1st, 2006 Duration: 36 months

Organisation name of lead contractor for this deliverable: PHYS-SAPIENZA

The vision

TAGora is a project sponsored by the **Future and Emerging Technologies** program of the **European Community** (IST-034721) focussing on the semiotic dynamics of online social communities. A new paradigm is quickly gaining impact in large-scale information systems: Folksonomies. In applications like Flickr, Connotea, Citeulike, Del.icio.us, etc. people no longer make passive use of online resources - they take on an active role and enrich resources with semantically meaningful information. Such information consists of terminology (or “tags") freely associated by each user to resources and is shared with users of the online community. Despite its intrinsic anarchist nature, the dynamics of this terminology system spontaneously leads to patterns of terminology common to the whole community or to subgroups of it. Surprisingly, this emergent and evolving semiotic system provides a very efficient navigation system through a large, complex and heterogeneous sea of information.

Our project proposes visionary and high risk research aimed at giving a scientific foundation to these developments, so contributing to the growth of the new field of Semiotic Dynamics. Semiotic Dynamics studies how semiotic relations can originate, spread, and evolve over time in populations, by combining recent advances in linguistics and cognitive science with methodological and theoretical tools of complex systems and computer science.

The project aims at exploiting the unique opportunity offered by the availability of enormous amount of data. This goal will be achieved through: (a) a systematic and rigorous gathering of data that will be made publicly available to the consortium and to the scientific community; (b) designing and implementing innovative tools and procedures for data analysis and mining; (c) constructing suitable modeling schemes which will be implemented in extensive numerical simulations. We aim in this way at providing a virtuous feedback between data collection, analysis, modeling, simulations and (whenever possible) theoretical constructions, with the final goal to understand and engineer the Semiotic Dynamics of on line social systems.

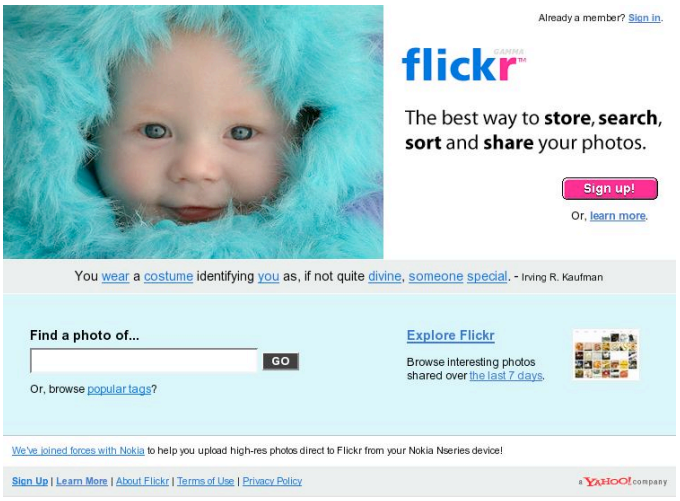
The challenges

To successfully navigate one’s way in a sea of information, one needs a system of fixed points, a coordinate system and maps. In most current day large-scale information systems (e.g. enterprise-resource planning systems, knowledge management systems) the coordinate system is given by some kind of ontology, the fixed points are given by annotations that describe documents or other data, while the map is produced by connecting ontology and annotations.

Unfortunately, this system becomes less and less workable as it concerns ever larger-scale information systems. On the one hand, the top down architecture of the system does not respond flexibly enough to the needs of the

users who plough the sea of information. On the other hand, anarchical approaches mostly do not work either, because without organization of ontologies and annotations one ends up with plenty of information systems that live completely independently from each other and cannot be joined for useful exploitation. For instance the distributed creation of taxonomies and tags and the multiplicity of conceptual schemata generate the well known problem of semantic interoperability. One solution is to standardise. The different users of a collective information system could all agree a priori to use the same taxonomies to structure their data and to use the same conceptual schemata for their data and meta-data. The tags in the owner taxonomies can then act as a shared communication protocol between peers. Unfortunately such a standardisation approach is unlikely to work for truly open-ended collective information systems in rapidly changing domains like music file sharing, picture exchange, medical imaging, scientific papers, etc.

Social bookmark systems are finding increasingly large user communities on the Web in a very short time frame and the new paradigm of Folksonomy is quickly gaining impact. In applications like Flickr (see figures in this page), YouTube, Connotea, Citeulike, Del.icio.us, etc. people do not "annotate" by hyperlinks, but by terminology. The terminology used in these applications can be freely chosen, but it evolves and leads towards patterns of terminology usage in these communities and in subgroups of these communities.



Screenshot of Flickr

Hence, one observes the emergence of terminologies providing a navigation system through a large, complex and heterogeneous sea of information. This phenomenon indicates a currently ongoing grass-root creation of knowledge spaces on the Web which is closely in line with "the 2010 goals of the European Union of bringing IST applications and services to everyone, every home, every school and to all businesses".

Screenshot of Flick

Example of a tag cloud, a commonly used way to visualize tags

In this framework a recent point of view gaining increasing consensus is that of Semiotic Dynamics which studies how populations of humans or agents can establish and share semiotic systems, typically driven by their use in communication. Semiotics studies the relation between conventional representations (e.g. language, gestures, pictures), the conceptualisations expressed by these representations, and the real world in which the conceptualisations are grounded. Linguistics, as a subfield of semiotics, focuses more specifically on language. A language is beyond doubt the most complex form of a conventional representation system that is shared by large groups of human users, and its collective dimension is more important than for other forms of representations (like visuals). Recently, emphasis has been shifting from viewing language as a static body of rules that are uniformly shared by all members of a language community towards viewing language as a complex adaptive system that arises through self-organisation out of local interactions, and is constantly shaped and reshaped by language users in order to maximise communicative success and expressive power while at the same time minimising effort. This viewpoint has been adopted both by researchers empirically studying human semiotic interactions in dialog, and by researchers interested in language change and the origins of linguistic form.

The already mentioned new developments in Information Technology (such as Flickr or Del.icio.us) fall precisely in this perspective and they can be seen as examples of Semiotic Dynamics at play. For instance social tagging sites, through which tens of thousands of web users share information by tagging items like pictures or websites and thus develop the so-called folksonomies, exhibit a

dynamics similar to the one observed in human languages, such as a struggle between competing tags until one dominates. As a second example there are information systems with emergent semantics, in which tags or other forms of meta-data are related to the actual data through signal or image processing algorithms. Other examples of new forms of Semiotic Dynamics occur in grounded interactions among robots whose communication systems are not pre-programmed but have to evolve in situated embodied interactions.

Semiotic Dynamics is a new field, still in its infancy, which tries to give a scientific foundation to these developments. It studies how semiotic relations can originate, spread, and evolve over time in populations. It combines recent advances in cognitive science with advances in complex systems. Cognitive science helps to provide insights into the grounding of conceptualisations through perception in the world, the highly complex cognitive processes that are implicated in the production and interpretation of representations, the processes underlying the invention and learning of concepts and their representation, and the way humans align concepts and language to co-ordinate their semiotic systems. Complex systems science is a precious aid to model and understand the collective dynamics whereby conventions can spread in a population, how conceptual and linguistic coherence may arise through self-organization or evolution, and how concept formation and expression may interact to co-ordinate semiotic systems of individuals.
The past decade has seen the first important contributions in the study of Semiotic Dynamics but most of the key questions remain unanswered. In particular the main question can be cast as follows: how do agents establish symbolic conventions between meaning and form (or sign) when there is no global control nor any prior (innate) convention, and only peer-to-peer interaction? This project moves precisely in this direction and it draws a certain number of important steps to reach a global understanding of these phenomena.

The opportunity

This project aims to contribute to the economic development of the Community by advancing the state of the art in Complexity Science as relevant to IT and by exploring highly novel IT applications based on Science of Complexity. In particular, the project takes at its starting point that there has been a massive increase in the amount of autonomy and information flow and in the number of people that are participating in IT processes. Moreover there is a pressing need that information systems become ever more adaptive to user needs and rapidly expanding infrastructures. Consequently, there is a much higher interdependence of actors than in the past and various properties observed in complex systems (such as self-organised criticality) are now also observed in information systems. In particular, the focus of this project bears direct relevance to fundamental and applicative problems that are currently attracting the attention of researchers from different fields, ranging from multi-agent systems to knowledge-management, from the semantic web to peer-to-peer content

distribution. This is possible because of two general traits explicitly addressed and investigated by this project:

- **The role of metadata.** The widespread use of electronic content and the availability of high-speed networking have created the opportunity for unprecedented levels of information sharing, at the same time bringing forth the problem of managing such a vast amount of information in ways that allow its effective and efficient use. The traditional approach of top-down, centralized categorization becomes challenged both in scalability and effectiveness, and gives way to new paradigms characterized by bottom-up, distributed, intrinsically collaborative categorizations, hinged on *emergent metadata* . In these systems - for example folksonomies - *metadata* lose the simple role of passive "labels" and take on a new role as evolving entities, displaying their own (semiotic) dynamics. That is, the complex dynamics of the system arising from the interaction of its many parts leads to hierarchies and emergent dynamical features.

- **Online communities.** The widespread use of the Internet prompted the creation of new models of user interaction, mediated by the concept of a "virtual community". As the supporting IT infrastructure evolves, the metaphor of the "virtual community" becomes more and more sophisticated, and online social networks acquire many of the *complex* aspects pertaining to the structure and dynamics of actual social networks.

Today, the *interplay of Semiotic Dynamics and online communities* is creating novel collaborative scenarios (folksonomies, social tagging, semantic approaches to peer-to-peer systems,) that are truly "complex". That is, these systems exhibit emergent properties that challenge our ability to explain them in terms of the underlying behavior of their elementary components. Two decades ago, physical systems exhibiting the same kind of challenging behaviors prompted the development of a new branch of Physics known as the "Physics of Complex Systems". Since then, the science of complexity has evolved and has developed several analytical and theoretical tools to model and understand "complex" systems. Nowadays, our technological systems have acquired comparable levels of complexity and we believe that the time has come to approach them by using the same tools developed within Physics. We believe that a unique window of opportunity is opening right now to leverage synergies between IT, computer science and the science of complexity, the unifying paradigm being the idea of emergent dynamics in a system composed by many interacting "agents". This justifies the composition of our collaboration, with members having different but complementary areas of expertise. Such complementarity must not mislead - this proposal is focused on the *Semiotic Dynamics* of *emergent metadata* in online social communities. Concepts and tools from various areas of expertise will be employed to study such dynamics, but the problem itself is well defined and very hot at the moment. Indeed, given the ever increasing deployment of collaborative tagging and resource sharing technologies, both on the Internet and in corporate

environments, any fundamental breakthrough in understand and/or controlling the emergent properties of these complex systems will have a significant technological impact.

Expected impact

The project includes partners who have a proven record of excellence in developing and publishing research results in their respective scientific areas. One of the primary objectives of this project is to bring together two communities: researchers in various domains of complex systems and researchers in various areas of Information Technology (computer science, web technologies, ubiquitous computing), in particular those facing the challenge of Semiotic Dynamics in on line social communities. The impact that we seek on the scientific community is potentially enormous because it goes beyond the specific scientific objectives and technologies focused on in this project. We want to foster a general movement towards the interrelation of complex systems and Information Technology by showing successful examples of cooperation and by posing and solving concrete non-trivial problems. It is only by seeing clear examples that we expect the scientific research community to follow suit. These clear successful research examples will be documented and presented through regular scientific communication channels so that they are accessible to the community at large.

We expect however that this project will contribute in a very significant way to develop the technological basis for a totally new generation of on line social communities. The results that come out from these developments will be disseminated as widely as possible so that the maximum number of IT actors can benefit. Because of the high level of exposure of web-based activity, we plan to impact the very phenomenon we are studying by developing and deploying a new generation of enriched collaborative social platforms which could reshape the social approach to sharing online information. Moreover, through the web-portal we are going to develop in the project, we expect to establish a framework where experts, scholars and simple users can meet and share and develop new ideas.

Scientific and technological objectives

The project is articulated in four main areas whose activities are strongly intertwined. The initial phase of the project will deal with collecting actual data from existing, live systems and analyzing them with a variety of formal tools, eventually **inferring models** that are able to capture the essential features of the emergent dynamics, and explain how they might arise from the interactions of single agents. The inferred models of the emergent dynamics will be subsequently used to develop simulations that will allow the **formulation of design strategies** targeted at attaining a specific global behavior.

Emergent metadata

The initial phase of the project will deal with collecting actual raw data from existing, live systems. By "raw data" we mean the **emergent metadata** that arise because of agent interactions in online social communities, as described in the introduction. Several online communities are readily accessible over the web: for a selected set of these systems, tools will be developed and deployed to harvest the relevant data, metadata and temporal dynamics, and to store the acquired information in a form amenable for data analysis.

Data analysis of emergent properties

Examining quantitative aspects of folksonomy is a highly important area of research. Our objective is the set up of several protocols of data analysis to be performed on the **raw data sets**. A data analysis protocol is defined by: (1) indicating a specific quantity / observable / estimator suitable of a quantitative measure on the raw data sets; (2) acquiring the existing software tools, or developing new specific tools, needed to perform the measure; (3) extracting the relevant statistical information characterizing the analyzed data sets.

The aim of the data analysis is to identify and quantify emergent properties of the system in study, i.e. properties that can not be simply inferred from the behavior of the single agent. Beyond suggesting the collection of new or more refined raw data, the results of the data analysis will be used to

1. identify general features common to the different systems in study;
2. characterize/discriminate the specific features of different systems in study;
3. orient the modelling phase of the research project (see below);
4. providing benchmarks to test/improve existing systems or to suggest the creation of new more performing systems.

Modeling and simulations

The objectives of this research area are twofold:

- *understanding complexity:* develop models that captures the essence of the emergent dynamics and explain how it might arise from the interactions of single agents;
- *taming complexity:* formulate design strategies that allow controlling the behavior of the system at the emergent level by suitably choosing the microscopic dynamics of the interacting agents.

One of the most important goals is to construct, implement and study specific modeling schemes aiming at reproducing, predict and control the emergent

properties seen in the semiotic dynamics orchestrated in on-line communities. We plan in particular a modeling activity at different scales. On the one hand it will be important to construct microscopic models of communicating agents performing language games without any central control. At a different scale we shall consider more coarse-grained probabilistic models. Several models will be proposed to address specific aspects/scales of folksonomy. The models will allow computer simulation aimed at measuring emergent features to be compared with the results of the data analysis activity. The simulations should give an insight in how users select tags, what kind of categories and category structures underlying the evolving system of tags, how categories and tags are related to the objects being tagged, etc. It will also give information on what kind of more global structures (such as the most frequent tags) can be provided to users to optimize their on-line community infrastructure The models will require components for assigning or adopting tags, categorizing data, and collective dynamics. However the approach will be to keep the models as simple as possible, identifying the minimal ingredients responsible for the emergent properties. The minimal character of the models should make a more analytical mathematical study feasible.

A possible way to tackle the complexity of the systems is to individuate different time scales, which can be separated. For instance, we expect that the dynamics of the social network of the folksonomy could be different from the time scale of the dynamics of the resources and of the tags. In this case one can, as a first approximation, propose a model of tags and/or resource dynamics based on a given, slowly evolving social network topology. This kind of assumption should be tested and corroborated as much as possible with the observations coming from the real data analysis.

Finally, the output of this activity has the potential to feed back into the data collection activity, specifically to the live social tagging system developed as part of it, in order to experimentally verify the devised control strategies and demonstrate the technological advantage achieved by the present project.

New applications

Collaborative tagging originated from the need to manage large collections of data. Tagging data is a means to describe, search, and retrieve objects in an intuitive way, which constitutes an important factor of its success. TAGora will provide experimental systems which are on the one hand intended to further improve navigation possibilities provided by tags, and on the other hand deliver data for the research work of the project. In order to have privileged and controllable data sources for the collaboration, TAGora plans to design and deploy systems - both online systems and actual demonstrations/experiments - for the specific purpose of data collection. The first objective involves building systems that add value to existing tagging sites. One possibility is to enrich navigation based on tags by adding data analysis. The combination of data

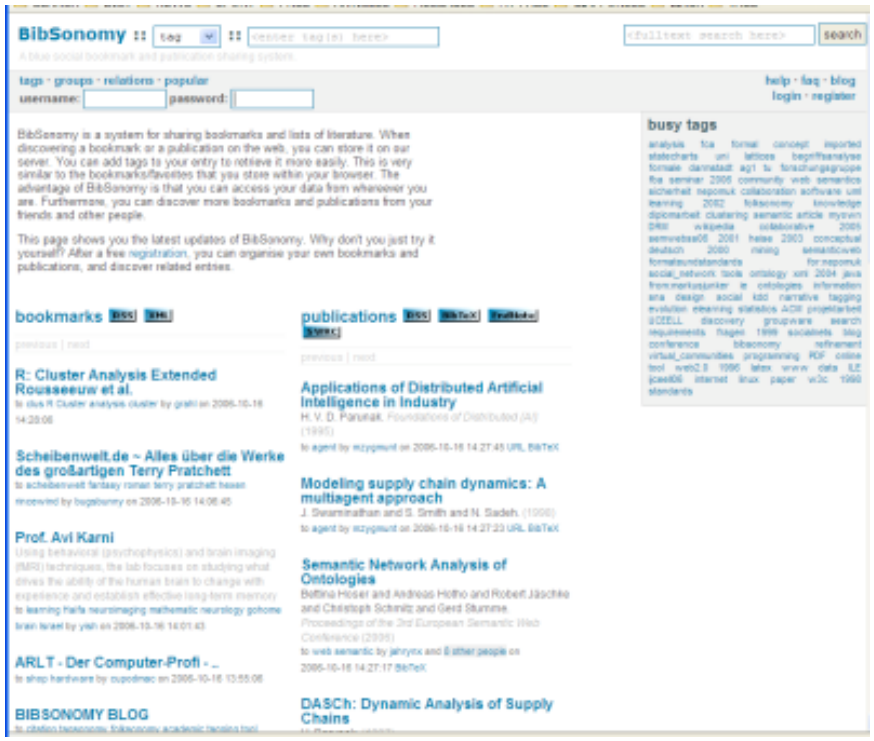
features and tagging allows to overcome shortcomings of tag-based search, such as problems caused by synonymy, homonymy, missing tags, or spelling mistakes. The added value of original TAGora systems is important in order to attract users and thus fulfil our second objective: to serve as a valuable source for data delivery. Moreover the new applications will allow the Consortium to gain unimpeded access to the raw data and will ultimately provide an experimental "clean room" platform that will be employed to validate the understanding of metadata emergence, and to experiment innovative control strategies.

Here are two examples of systems for which a prototypical version is already working.

BibSonomy The Hertie Chair of “Knowledge & Data Engineering” has been developing a Social Bookmark System, called BibSonomy (www.bibsonomy.org), which allows users to upload their bookmarks or bibliographic references and assign them arbitrary labels, denoted “tags”. BibSonomy allows users to share bookmarks and publication references. In general, social resource sharing systems all use the same kind of lightweight knowledge representation, called folksonomy. The word “folksonomy” is a merge of the words “folk” and “taxonomy”, and stands for the conceptual structure of knowledge created by people. By assigning a list of tags to a resource, each user builds up a so called “personomy”. The user can explore his own personomy, as well as the personomies of different users.

In BibSonomy one may address queries to each level: user, tag or resource.
In details:

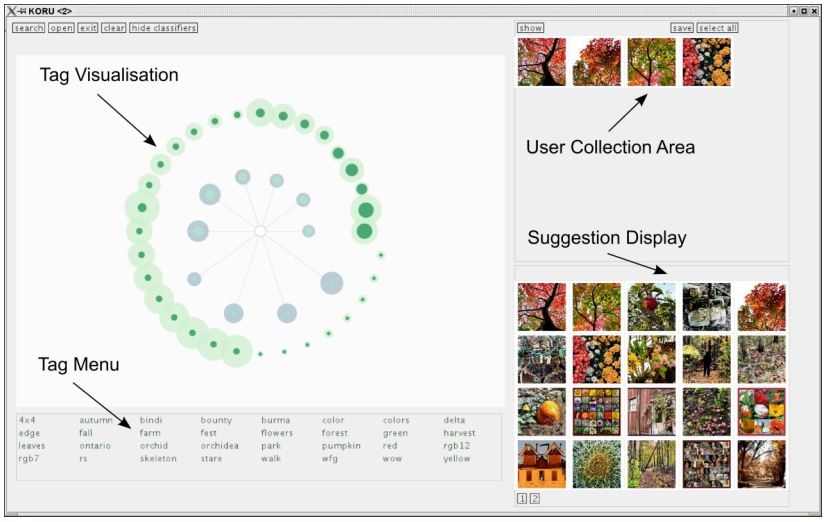
1. for a given user one can see all resources she has uploaded, together with the tags she has assigned to them;
2. for a given resource one can see the users who uploaded it and the relative assigned tags;
3. for a given tag, the resources labelled with it and all users who uploaded it are shown.



Screenshot of BibSonomy

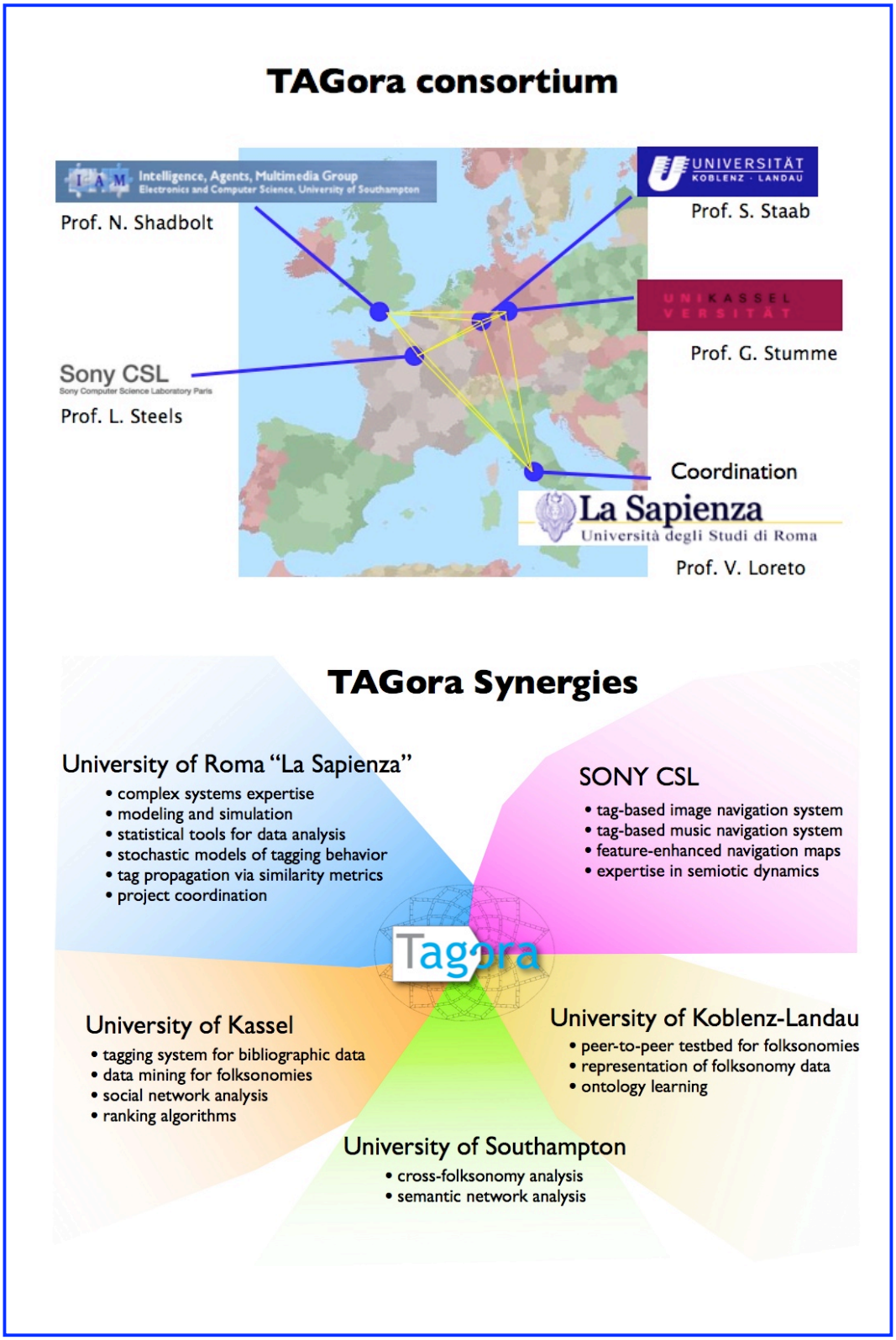
Bibsonomy allows for unfettered access to the complete data, the possibility of tracking the full temporal evolution of the system, from the very beginning, and the possibility to directly observe the influence of different parameter settings. The latter includes the option to analyse the effect of more sophisticated knowledge representations, like hierarchically structured personomies, binary relations between resources (as known from the semantic web), or statistical relationships indicating the degree of similarities of the content of different personomies. This provides the possibility to directly observe on a large natural user group the influence of knowledge representation decisions on the structure and evolution of the social network. An existing prototype offers the basic features for social tagging, but does not yet provide the functionality of the systems above, nor does it provide exclusive functionality. For attracting a larger set of users, and thus obtaining significant benchmark data, we will have to provide extended features, as well as a scalable implementation and a component for systematic data collection. On these data, we will study the structure of the emerging network, in particular its topological properties, dynamical properties**,** and clustering/social properties**.**

Ikoru Ikoru is a prototypical system developed by Sony CSL that unifies browsing by tags and visual features. This combination allows an intuitive exploration of databases and helps to overcome shortcomings of solely tag-based systems. In contrast to traditional image retrieval approaches, Ikoru employs user tags, complemented by image analysis and classification. The image analysis in Ikoru is based on simple global features. Rather than trying to recognize objects or even explain the meaning of an image, Ikoru seeks to measure a certain “atmosphere”, or a vague visual pattern, which can be captured by low level image features such as colour or texture. Sony CSL plans to develop a similar system for music data as well.



Screenshot of the interface of the experimental system for images. User Collection: images selected by the user to the theme “fall”. Suggestion Display: Search results returned for first image of User Collection.

Ikoru also serves as a platform for further experiments. For example, the relation between data features and tags will be investigated. We observe that only a fraction of tags can be grounded, e.g. low-level tags such as "red", or "blackandwhite" for images, and "loud" or "fast" for music. We expect that a simple one-to-one mapping from a tag to a category that can be described in terms of data features is only possible for these low-level tags. High-level tags such as "abandoned", "decay", tags denoting locations such as "Paris", or tags denoting persons will not show the same behaviour. We will thus investigate possibilities to achieve an indirect grounding, e.g. by exploiting the co-occurrence relation between high-level and low-level tags. The findings of these experiments will then be used to implement tag proposition into our systems, in order to assist the user in the tedious process of tagging. This mechanism will allow us to study the influence of tag suggestions on convergence properties as well as on the precision of tagging.


TAGora consortium
Intelligence, Agents, Multimedia Group
Electronics and Computer Science, University of Southampton
Prof. N. Shadbolt
UNIVERSITÄT KOBLENZ · LANDAU
Prof. S. Staab
UNIKASSEL VERSITÄT
Prof. G. Stumme
Sony CSL
Sony Computer Science Laboratory Paris
Prof. L. Steels
Coordination
La Sapienza
Università degli Studi di Roma
Prof. V. Loreto
TAGora Synergies
University of Roma "La Sapienza"
• complex systems expertise
• modeling and simulation
• statistical tools for data analysis
• stochastic models of tagging behavior
• tag propagation via similarity metrics
• project coordination
SONY CSL
• tag-based image navigation system
• tag-based music navigation system
• feature-enhanced navigation maps
• expertise in semiotic dynamics
Tagora
University of Kassel
• tagging system for bibliographic data
• data mining for folksonomies
• social network analysis
• ranking algorithms
University of Koblenz-Landau
• peer-to-peer testbed for folksonomies
• representation of folksonomy data
• ontology learning
University of Southampton
• cross-folksonomy analysis
• semantic network analysis

Physics Dept., *La Sapienza* University (PHYS-SAPIENZA)

Background of the team

The Phys-Sapienza research group is based in the Physics Department of "La Sapienza" University in Rome, one of the largest physics department in Italy. It includes more than 250 scientists, among professors and researchers, and more than 66 research groups with topics ranging from high-energy physics to condensed matter theory, statistical mechanics and astrophysics. PHYS-SAPIENZA team is very active in the whole area of statistical physics, information theory and complex systems. In the last few years several new projects have been launched among which:

- agent-based modeling in linguistics;
- information theory applied to time-series analysis, linguistics and genomics,
- theory of complex networks in technological, social and biological systems;
- opinion and assimilation dynamics in social systems.

Specific interests in TAGora

Phys-Sapienza team brings into the consortium its research experience on: (a) developing new theoretical tools to collect and analyse data; (b) introducing and studying suitable modeling for complex systems in order to understand the role and the importance of the different factors in a system of communicating agents; (c) constructing theoretical approaches which could provide with different levels of abstraction and a feedback for new experiments and studies.

Team members

- Prof. Vittorio Loreto – Coordinator (http://pil.phys.uniroma1.it/~loreto/)
- Dr. Andrea Baldassarri (http://greco.phys2.uniroma1.it/~andreab/)
- Dr. Ciro Cattuto (http://pil.phys.uniroma1.it/~ciro/)
- Dr. Vito D.P. Servedio (http://pil.phys.uniroma1.it/~servedio/)

Sony CSL

Background of the team

Sony CSL is a basic research laboratory, founded in 1996 by Luc Steels. Research at CSL focuses on four areas: personal music experience, computational neuroscience, developmental cognitive robots, and self-organising communication systems. The team participating to the project is active in two main areas.

- Semiotic Dynamics: (self-organising communication systems): Research in self-organising communication systems investigates through computational simulations and mathematical models how a group of autonomous agents could be able to invent and negotiate a communication system similar to human natural languages.

- Personal music experience: Research in Personal Music Experience focuses on the future of musical listening by building prototypes of interactive devices and ethnographic experiments to see what people find exciting in music and how new ways of listening integrate into their lives.

Specific interests in TAGora

Collaborative Tagging as a form of an evolving language system is of great interest for SONY-CSL. In addition to studying such systems, SONY-CSL contribute to the design of new applications in this field with special emphasis in combining collaborative tagging with features that are automatically extracted from data. Collaborative tagging is considered as an important step towards semi-automatic or automatic music and image categorization, which is one of SONY-CSL key areas.

Team members

- Prof. Luc Steels (http://www.csl.sony.fr/General/People/StaffPage.php?username=steels)
- Dr. Peter Hanappe (http://www.csl.sony.fr/General/People/StaffPage.php?username=hanappe)
- Dr. Melanie Aurnhammer (http://www.csl.sony.fr/General/People/StaffPage.php?username=melanie)

University of Koblenz-Landau (UNI KO-LD)

Background of the team

The Research Group "ISWeb - Information Systems and Semantic Web" concentrates its work on the basic principles and applications of semantic-based technologies and their integration into complex, dynamic information systems.

The basics of semantic-based systems include the modeling of ontologies, representation of ontologies, approaches and methods for the design and maintenance of ontologies, as well as the semantic annotation of documents, multimedia data or web services to enable the semantic search and usage of these resources.

Semantic technologies are used for information retrieval, for information integration and for semantic-based peer-to-peer networks or for semantic middleware (Web Services, Grid). Semantic technologies enrich the abilities of information and acknowledgement systems and allow a more efficient and more effective handling of complex, dynamic systems.

Specific interests in TAGora

The contributions of the university of Koblenz-Landau within TAGora are mainly in three research areas, namely the development of a distributed semantic information system, the analysis of tagging data gathered from online social communities and the modeling and simulation of tagging processes. One of UNI KO-LD contributions will the development of the *Semantic Exchange Architecture* (SEA), a peer-to-peer infrastructure for tagging-based organization and sharing of personal information objects, i.e. multimedia data within TAGora. In contrast to common online social communities the idea is not to store the data on a dedicated central server, but locally on each user's computer while still providing the same tag-related navigation functionalities.

Team members

- Prof. Dr. Steffen Staab (http://www.uni-koblenz.de/~staab/)
- Olaf Goerlitz (http://www.uni-koblenz.de/FB4/People/Person/goerlitz)
- Klaas Dellschaft (http://isweb.uni-koblenz.de/People/Dellschaft/)
- Alexander Kubias (http://www.uni-koblenz.de/FB4/People/Person/kubias)

University of Kassel (UNIK)

Background of the team

The research unit "Knowledge & Data Engineering" at the University of Kassel started in April 2004 with the establishment of an endowed chair of the Hertie Foundation. Research in the unit focuses on knowledge engineering, i.e. on discovering and structuring knowledge, on the derivation of new knowledge processes, and better communication of knowledge. In particular, UNIK research is aimed at the development of methods and techniques at the intersection of research areas such as Social Software, Knowledge Discovery, Ontologies/Metadata, Semantic Web, Peer to Peer and Formal Concept Analysis, with the perspective of reaching substantial synergies among them.

Specific interests in TAGora

In the TAGora project, UNIK will develop methods to enable users to manage and share their knowledge in an individual and flexible way. This includes methods for detecting and managing communities. UNIK will monitor the development of community structures in the BibSonomy system (www.bibsonomy.org) over time. Since one expects a large number of BibSonomy users, extended functionality of the system, as well as improved scalability will be needed. In addition to this, UNIK will set up components for data collection and for tracking evolution. Because of the high level of exposure of web-based activity, UNIK plans to develop and deploy a new generation knowledge sharing system which could reshape the social approach to sharing online information.

Team members

- Prof. Dr. Gerd Stumme (http://www.kde.cs.uni-kassel.de/stumme)
- Dr. Andreas Hotho (http://www.kde.cs.uni-kassel.de/hotho)
- Miranda Grahl (http://www.kde.cs.uni-kassel.de/grahl)

University of Southampton (UNI-SOTON)

Background of the team

The School of Electronics and Computer Science at Southampton is a world-leading centre of excellence for research. Within the school, the IAM (Intelligence, Agents, Multimedia) Group focuses on the design and application of computing systems for complex information and knowledge processing tasks. With around 120 researchers, we are international leaders in the three major themes that converge in the Group's tripartite title. UNI-SOTON has a deep expertise in ontology modelling and applications, as well as in social network analysis over the Semantic Web. TAGora will bring these skills together and apply it to various folksonomic domains to help better understand information evolution, as well as influence and dependency across different communities.

Specific interests in TAGora

UNI-SOTON team will be focusing on two main issues: (1) information gathering and integration from various folksonomies (e.g. bibliography, music, images); (2) investigating requirements and approaches for online recommendations over and across folksonomic web sites. Folksonomy web sites are rarely closed worlds. It is quite common for individuals to be active members of several online communities and thus one would expect certain tags to spread across such communities with time. For example one could be adding images to *Flickr*, bookmarking web sites with *Del.icio.us*, creating their music preference profiles in *last.fm*, and tagging articles in *Connotea*. By continuously collecting data from such folksonomy web sites, and monitoring changes and additions, one can cross reference emerging tags between the separate communities to extend and connect their individual networks to create an Integrated Semantic Network.

Team members

- Prof. Nigel Shadbolt (http://www.ecs.soton.ac.uk/~nrs/)
- Dr. Kieron O'Hara (http://www.ecs.soton.ac.uk/people/kmo)
- Dr. Harith Alani (http://www.ecs.soton.ac.uk/~ha/)