# wimmics

Web-Instrumented Man-Machine Interactions, Communities, and Semantics.

**Summary:** Wimmics[1] is a proposal for a joint research team between INRIA Sophia Antipolis - Méditerranée and I3S (CNRS and University of Nice – Sophia Antipolis). The research fields of this team are graph-based knowledge representation, reasoning and operationalization to model and support actors, actions and interactions in web-supported epistemic communities.

**Main research area:** interaction, knowledge, communities, graphs, semantics, web

**Main application area:** supporting and fostering interactions in online communities

## 1 The research team

**Head (and INRIA contact):** Fabien Gandon

**Vice Head (and I3S contact):** Nhan Le Thanh

**Researchers:**

Michel Buffa, MdC (UNS)
Olivier Corby, CR1 (INRIA)
Catherine Faron-Zucker, MdC (UNS)
Fabien Gandon, CR1, HDR (INRIA)
Alain Giboin, CR1 (INRIA)
Nhan Le Thanh, Pr. (UNS)
Isabelle Mirbel, MdC, HDR (UNS)
Peter Sander, Pr. (UNS)

**Assistants:** Christine Foggia (INRIA) and Marie-Hélène Prosillico (I3S)

**PhD students:**

Pavel Arapov, to start in October (EDSTIC-INRIA)
Adrien Basse, 2nd year (UGB-INRIA)
Franck Berthelon, 2nd year (UNS-EDSTIC)
Ahlem Bouchahda, 3rd year (UNS-SupCom Tunis)
Khalil Riad Bouzidi, 2nd year (UNS-CSTB)
Luca Costabello, 1st year (INRIA-CORDI)
Corentin Follenfant, 1st year (SAP)
Maxime Lefrançois, 1st year (EDSTIC-INRIA)
Nicolas Marie, 1st year (Bell-ALU, INRIA)
Hasan Rakebul, to start 4th of July (ANR-Kolflow)
Oumy Seye, 1st year, (Rose Dieng allocation)
Imen Tayari, 3rd year (UNS-Sfax Tunisie)
Viet-Hoang Vu, 4th year (UNS-Factory)

**Post-doc:** Serena Villata (ANR-Datalift)

**Research engineers:**

Nicolas Delaforge (INRIA, ANR ISICIL)
Erwan Demairy (ADT)
Guillaume Husson (I3S, ANR ISICIL)
Amel Hannech (I3S)

**Apprentice:** Amélie Gyrard (INRIA)

**Interships:** Nemanja Vukosavljevic, Cissé Soda Mareme, Mouhamadou Ba.

[1] wimi is a variety of roses

# 2 Introduction and *raison d'être*: "as we may link"

Vannevar Bush wrote an article called "As we may think" originally published in the July 1945 issue of The Atlantic Monthly, just before the atomic bombs. In this article Bush already identified that "there is a growing mountain of research. (…) The difficulty seems to be, not so much that we publish unduly in view of the extent and variety of present-day interests, but rather that publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships." Vannevar Bush then proceeds with the proposal of an imaginary machine called the Memex (Memory Extension) a mechanized desk that can store books, records, and communications, so that they may be consulted with speed and flexibility and which will form encyclopedias with a mesh of associative trails running through them; a collective memory extension structured by association links.

Twenty years later, Ted Nelson wrote the article "Complex information processing: a file structure for the complex, the changing and the indeterminate" in ACM 65's Proceedings. Directly referring to the article by Vannevar Bush, Ted Nelson proposes a flexible file structure built by links between elements of documents. He coins the terms hypertext and hypermedia already acknowledging the multimedia explosion in computer science 30 years before it happens.

Twenty years later, Tim Berners-Lee allows hypertext to jump through the networks by designing hyperlinks crossing internet paths and weaving a logical web of documents that overlays a physical network of machines.

Twenty years later, the web is crossing a new boundary jumping into our daily life investing many objects and virtually every place. And the challenges Vannevar Bush foresaw are all too real now, revealing how visionary he was when he envisioned a world where we would have cameras with us all the time generating even more multimedia documents to be organized. What we also witnessed since the beginning of this new century is that the initial graph of associative thoughts has been joined by a growing number of other graphs. The graph structure that was weaved by our trails of thoughts is now mixed with sociograms capturing the social network structure, workflows specifying the decision paths to be followed, browsing logs capturing the trails of our navigation, service compositions specifying distributed processing, etc.

| System | Author(s) | Kinds of links/graph – Linked items | Linking Goal |
|---|---|---|---|
| MEMEX | V. Bush | Association links<br>Mesh of associative trails | Memory extension with a desk able to remember associations and organize readings. |
| HYPERTEXT / HYPERMEDIA | T. Nelson | Links between elements of documents | Digital structure to virtually organize tracks between fragments of multimedia resources |
| WORLD WIDE WEB | T. Berners-Lee | Link documents across the network. | Expand the structure over the internet to share it among many users. |
| Semantic Web | (several) | Link descriptions of any resources and the schema of the descriptions | Making humans and software agents cooperate through the web |
| Web of Data | (several) | Linked open data on the web | Use the web as a giant blackboard for data exchanges and integration |
| Social Web / web 2.0 | (several) | Link people capture relations | Foster awareness, exchanges and interactions between users |
| Web of things & ubiquitous web | | Link devices, places through their characteristics and services | Allow contextual interaction and web-augmented reality. |

Not only do we need the means to represent and analyze these graphs, we also need the means to combine them to allow multi-criteria analysis, and the means to precisely capture them and in particular the different types of constituent links. Wimmics will attack the question of these changing data structures in a ubiquitous and heterogeneous web: the characterization of typed graphs to model and capture these different pieces of knowledge and of operators to process them.

# 3 **Research Challenges:** Analyzing, Modeling, Formalizing and Implementing Graph-based Social Semantic Web Applications for Communities

The web is no longer the simple documentary system built on a simple protocol (HTTP), a simple addressing scheme (URI) and a simple document formatting language (HTML). It grew to become a huge network of distributed data, applications and users where many flows of heterogeneous messages travel. *The web is an object of science: it is a very complex system that requires a multidisciplinary scientific approach.* INRIA is more and more solicited to understand how these data and interactions can be processed, supported, controlled, exploited or improved. We believe Wimmics can contribute to this understanding in two manners:

(1) by analyzing and modeling, using a multidisciplinary approach, the many aspects of these intertwined information systems, communities of users and their interactions;
(2) by formalizing and reasoning on these models to propose new analysis tools and indicators, and support new functionalities and better management.

## 3.1 Analyzing and Modeling Communities Interactions through Social Semantic Web Applications: interacting with dynamic semantic systems of the web.

The web is a worldwide system never sleeping. The dynamics of the system and its ever-changing contexts make it very difficult to interact with it. The richness of semantics-based systems is more and more used to tackle the diversity of web resources and applications through metadata describing web resources but it also augments the complexity of the web and makes it more difficult to interact with. In all its dimensions the complexity of the web is growing.

*How do we improve our interactions with such an information system that keeps getting more and more complex?* We propose to rely on cognitive studies to build models of the system, the user and the interactions in order to support and improve these interactions.

Semantics in knowledge representation and in computer science in general are fixed (e.g. semantics based on first order logics) while semantics in social context are renegotiated all the time (e.g. semantics based on "natural logics").

*How do we reconcile and integrate the formalized stable semantics of computer science and the negotiable social interactions?* We propose to rely on social studies to build models of the communities, their vocabularies, activities and protocols in order to identify where and when formal semantics is useful.

When users interact with the web they can use a variety of devices (e.g. mobile phone), of modalities (e.g. vocal interaction), of languages (e.g. Chinese) and be in various contexts (e.g. in the bus).

*How do we reconcile local contexts of users and global characteristics of the world-wide virtual machine and information systems that the web has become?* We propose to rely on knowledge representation methodologies and theories (e.g. ontologies) to build models of the contexts, devices and mediums ensuring the effectiveness, quality and precision of the information delivered, to provide proofs and explanations of the processes applied and ultimately foster acceptation and trust from the users.

## 3.2 Formalizing Models and Implementing Social Semantic Web Applications: calculating on heterogeneous joined typed graphs of the web.

The models identified in the previous section need to be formalized in order to automate their analysis and processing when supporting web applications.

*What kind of formalism is the best suited for such models?* We defend that the network nature of linked data, social communities and service compositions on the web and the large variety of types of links that compose them call for typed graphs as formalized in languages like conceptual graphs or RDF/S. We intend to build on our experience with such formalisms to identify, propose and characterize fragments of typed graph formalisms best suited for each type of model identified before.

Each type of network of the web is not an isolated island. Networks interact with each other: the networks of communities influence the message flows, their subjects and types, the semantic links between terms interact with the links between sites and vice-versa, etc.

*How do we analyze these typed graph structures and their interactions?* We believe that type-based inference algorithms (e.g. conceptual graph projection, inference rules) and type-parameterized operators (e.g. parameterized betweeness centrality) provide declarative formalisms to flexibly define operations to monitor, filter, query, mine, validate, protect, etc. these imbricated graph structures taking into account constraints spanning several types of network at once.

These graphs are not available in a single central repository but distributed over many different sources. Some sub-graphs are small and local (e.g. a users' profile on a device), some are huge and hosted on clusters (e.g. Wikipedia), some are largely stable (e.g. thesaurus of Latin), some change several times per second (e.g. social network statuses), etc.

*How do we support different graph life-cycles, calculations and characteristics in a coherent and understandable way?* We believe that moving to graphs languages with open-world logics, temporal aspects, distributed and loosely coupled algorithms and model-driven programming relying on higher abstractions (e.g. formal ontologies) provides an adequate theoretical framework to allow at the same time the specification and operationalization of the models and algorithms and the opening of these black boxes to be able to explain, document, prove and trace results for the users.

# 4 Wimmics: we mix Edelweiss and Kewi

The research team Edelweiss (INRIA), previously known as Acacia and founded by Rose Dieng-Kuntz, aims at offering models, methods and techniques for supporting knowledge management and collaboration in virtual communities interacting with information resources through the Web and using graph-based and ontology-based formalisms and algorithms. Latest research topics include: multilingual interaction with knowledge bases, mobile access to the web of data, rule-based semantic web formalisms, online presence and resource centric sociality, semantic web in business intelligence, automatic indexing of triple-stores.

The research team Kewi (I3S), is interested in knowledge engineering techniques and web-based applications to capture, extract analyze, organize, store and share knowledge. Latest research topics include: semantic web and graph-based knowledge modeling; semantic security and access control; affective computing and emotion detection; requirement engineering and collaborative work.

The two teams have been collaborating for now more than seven years and, as shown by their descriptions, exhibit both common interests (web, semantic web, graph-based formalisms, ontologies, etc.) and complementarities (interaction design/ affective computing, triple-stores, rules / access control, etc.). Merging the two teams really is a natural acknowledgement of an ongoing collaboration and numerous co-publications.

Moreover this merge will create a group including full-time researchers, assistant professors and professors that could naturally cover all the activities of a research team.

# 5 Statement from the scientific leader

The Wimmics proposal is led by Fabien Gandon, CR1, HDR (INRIA).

## 5.1 Short biography

Dr. Fabien Gandon has PhD in Computer Science from the University of Nice-Sophia-Antipolis. He worked on Knowledge engineering, Ontologies, semantic Web, Multi-Agents Systems, Mobile devices, Context-Awareness and Privacy. He previously worked for the Mobile Commerce Laboratory of Carnegie Mellon in Pittsburgh. He published more than 34 articles in journals, books or conferences, has been reviewer of more than 23 journals, conferences or workshop and was an invited speaker for several events. His professional interests include: Knowledge engineering, Ontologies, semantic Web, Multi-Agents Systems, Context-Awareness and Privacy. He is involved in the W3C working group on GRDDL, RDFa and RDF 1.1. Fabien is currently the leader of the ANR project ISICIL and member of two other ones (Datalift, Kolflow) as well as the supervisor or co-super visor of nine Ph.D. students, two of them having just successfully defended. For more, see http://fabien.info

## 5.2 Motivations

The web is both a set of standards (a technical object) and the largest information system in history (a world-wide social object). This emerging social object touches on every aspects of modern life: more and more jobs depend on it, spanning all activities from journalism to commerce, from education to healthcare, from transportation to arts, from activism to government. The complexity of this information system (more than 20 billion indexed pages) justifies multidisciplinary scientific approaches to the many problems it raises and the idea of a web science gathering these contributions was officially launched in 2006. What is at stake is the need we have to understand and influence this system in order to ensure some desirable properties (e.g. openness) and prevent some undesirable ones (e.g. link farms). This understanding calls upon multidisciplinary contributions: to understand the web we need to go beyond its structure and look at the variety of links, resources, actors and interactions it embeds and supports. We need to call upon psychology to model the users, mathematics to model the structures, sociology to model the dynamics, linguistic to model the texts, etc.

Wimmics positions its activity right in the context of this cross-fertilization and our proposal is centered on the graph structures that are multiplying on the web: social networks, service compositions, suggested or captured browsing, thesauri and types hierarchies, frequent patterns, timelines and workflows, communication networks, information flows, etc. More specifically these structured metadata are typed graphs: nodes and arcs of these graphs are labeled with types that can support inferences and enrich their use; these typed graphs will be our main topic of research.

Building on an extensive experience in particular through a number of projects the members of Wimmics already showed in the context of organizational memories (i.e. on intranets) that these graph-based metadata models not only support reasoning but can also be seen as metric spaces to pilot approximation, or as indexes of knowledge in distributed environments, or as models to make interfaces more intelligible to end-users, or as new frameworks for social structures analysis. We want to take these results much further and target this growing diversity of graphs on the open web and in particular on the social web supporting so many communities of interest. Importantly, the members of Wimmics not only have the needed profile to conduct this line of research, they also have the profile to teach and diffuse that knowledge in universities or to companies.

To summarize, if there is only one point to take back home from this statement it is that I believe that in a near future, he who controls metadata, controls the web. And it will soon be of vital importance for any person or organization to gain and maintain control over the metadata that describes it, its productions its members and virtually every aspect of itself.

Fabien. L. Gandon.