ISICIL: Semantics and Social Networks for Business Intelligence

Michel Buffa, Nicolas Delaforge, Guillaume Erétéo, Fabien Gandon, Alain Giboin, Freddy Limpens

▶ To cite this version:

HAL Id: hal-01171260
https://hal.inria.fr/hal-01171260

Submitted on 3 Jul 2015

ISICIL: Semantics and Social Networks for Business Intelligence

Michel Buffa,Nicolas Delaforge, Guillaume Erétéo, Fabien Gandon, Alain Giboin, Freddy Limpens

[1]Wimmics, Inria, I3S, University of Nice,
2004 route des Lucioles, BP 93, 06902 Sophia Antipolis
fabien.gandon@inria.fr

Abstract. The ISICIL initiative (Information Semantic Integration through Communities of Intelligence onLine) mixes viral new web applications with formal semantic web representations and processes to integrate them into corporate practices for technological watch, business intelligence and scientific monitoring. The resulting open source platform proposes three functionalities: (1) a semantic social bookmarking platformmonitored bysemantic social network analysis tools, (2) a system for semantically enriching folksonomies and linking themto corporate terminologies and (3) semantically augmented user interfaces, activity monitoring and reporting tools for business intelligence.

Keywords: social semantic web, business intelligence, social network, social network analysis, folksonomies, semantic wiki

1 Introduction

Recently, online communities of interest have emerged and started to build directories of references in their domains of interest at an impressive speed. One of the main strengths of the tools enabling these communities is their ability to turn usually passive users into active participants and producers. The diversity and the mass of users are used to tackle the diversity and the mass of information sources. Monitoring market, science and technological changes is a vital ability of today's organizations, yet the growing diversity of sources to track in each domain of interest remains a challenge for any organization. Therefore there is a growing interest in importing inside corporate information systems the tools and practices that made the success of these online communities. But, Web 2.0 tools exhibit limits when it comes to automating tasks or controlling processes. On another hand, more structured information systems often suffer from usability and knowledge capture issues. Thus a challenge of the ISICIL project is to reconcile viral new web applications with formal representations and processes and to integrate them into corporate practices. More specifically, we studied and experimented with the usageof new tools for assisting corporate intelligence tasks. These tools rely on web2.0 advanced interfaces for interactions and onsemantic web technologies for interoperability and information processing.

2 Overview of the ISICIL project

Over the past four years, the goal of the ISICIL project has been to combine social networks and semantic web in a user-friendly platform supporting corporate intelligence. The project started with a multidisciplinary design methodology to specify a new form and a new platform for corporate intelligence. From the technical point of view, the main challenge of this project was to reconcile the new viral web applications with formal representations of business processes and to integrate them into communities of practice of the company. We explored new scientific developments of the notion of epistemic cooperation (human interaction oriented toward the development and transmission of knowledge) and we identified usable technological solutions. An ergonomic approach, combining impregnation of ground truth data and freer technological inspirations from bibliographic and webographic sources, was followed and evaluated. The results of this study allowed us to specify,design, experiment and evaluate new tools to support collaborative tasks in business intelligence by leveraging Web 2.0 interfaces (blog, wiki, social bookmarking) for interactions and semantic web technologies for interoperability and information processing.

All the models proposed and formalized in the project are typed graphs.These models capture structures and semantics underlying epistemic communities, their networks, their resources and their interactions.ISICIL relies on a unifying model based on RDF graphs to represent resources and community stakeholders. These models are integrated with bookmarking or "web scraping" tools. Theoutputs of these tools are tagged and the tags are collectedto form folksonomies and analyzed to semi-automatically structure these folksonomies. The user feedback on this structure is capture when they use the search engine which offers tags related to their keywords.Users can accept, reject or adjust these suggestions and enrich or correct the structure as a side-effect of refining their queries. User profiles and links are processed by a series of operators to propose a semantic analysis of social networks e.g. centrality metrics parameterized by an ontology. Merged graphs of structured folksonomies and of social networks finally allow the detection and labeling of epistemic communities.

ISICIL is a proof of concept of the compatibility of the Semantic Web formalisms, practices and models of Web 2.0 and the philosophical framework of social epistemology. An open-source platform is available under CeCILL-C and was tested at ADEME and Orange, and all the deliverables are online[1]. In the following sections we zoom on three types of results from the project: semantic social network analysis (section 3.1), folksonomy enrichment (section 3.2) and community and interest detection (section 3.3). We explain the software architecture and implementation in section 4. We provide a guided tour of some of the users' the interfaces in section 5. Finally we position the project within the state of the art (section 6) and we conclude and discuss the results and evaluation (section 7).

---

[1] http://isicil.inria.fr

3 Semantics and Social Metrics

3.1 Social Network Analysis (SNA)

ISICIL integrates a bookmark-centric social network and Semantic Network Analysis (SNA) tools. We used popular ontologies to model a social graph in RDF (linking people, resources, tags).We implemented the computation of the main SNA indices using Sparql and post-processing in a few cases, see [1] and[3] for details.We designed stack of tools (Fig.1) to conduct a semantic social network analysis. The goal of this stack is to provide a framework that enables us to consider not only the network structure embedded in social data, but also the schemas that are used to structure, link and exchange these data. This stack is composed of (1) tools for building, representing and exchanging social data and (2) tools for extracting social network analysis metrics and leveraging social graphs with their characteristics.
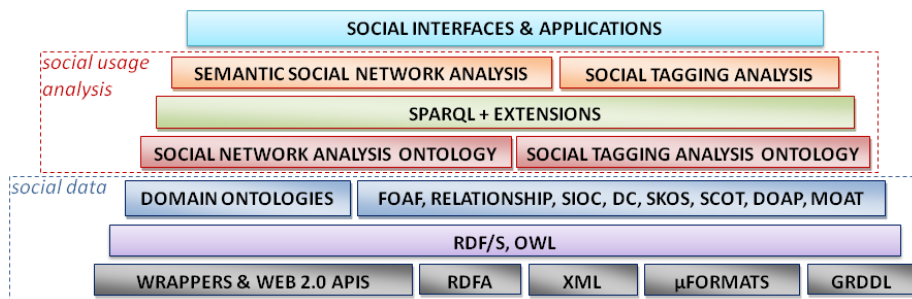

Fig.1. Abstraction Stack for Semantic Social Network Analysis.

We represented the social graphs in RDF, which provides a directed typed graph structure. Then we leveraged the typing of nodes and edges with the primitives of existing ontologies together with specific domain ontologies when needed. With this rich typing, semantic engines are able to perform type inferences from data schemas for automatically enriching the graph and checking its consistency.

For the analysis of the network, we designed SemSNA[2] (Fig.2) that defines different SNA metrics ranging from the annotation of strategic positions and strategic actors (like degrees or centralities), to the description of the structure of the network (diameter, etc.). With this ontology, we can abstract social network constructs from domain ontologies to apply our tools on existing schemas by having them extend our primitives.We can also enrich the social data with the SNA metrics that are computed on the network. These annotations enable us to manage more efficiently the life cycle of an analysis, by pre-calculating relevant SNA indices and updating them incrementally when the network changes over time. Moreover they can be used during the

---

[2]http://ns.inria.fr/semsna/2009/06/21/voc.rdf

querying of social data for ordering and filtering the results and reused by other applications.
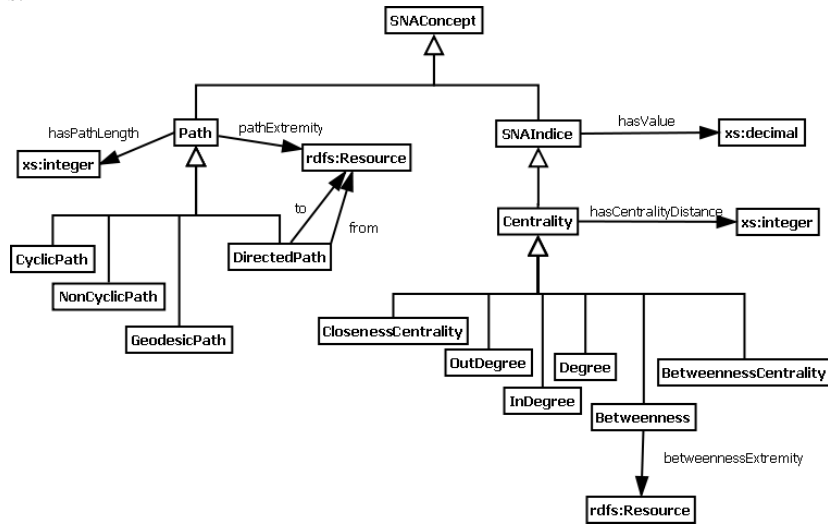

Fig.2. Subset of SemSNA ontology: paths and strategic positions.

Fig.3shows an example of enrichment of a semantic social graph with SemSNA. We use the property `hasSNAConcept` to link `Guillaume` to a `Degree`, the property `semsna:isDefinedForProperty`specifies that this degree has been computed in the `colleague` sub-network (taking into account sub-relationships like `supervisor`), the property `hasValue` describes the value of this degree, and `hasCentralityDistance`defines the path length that was considered.
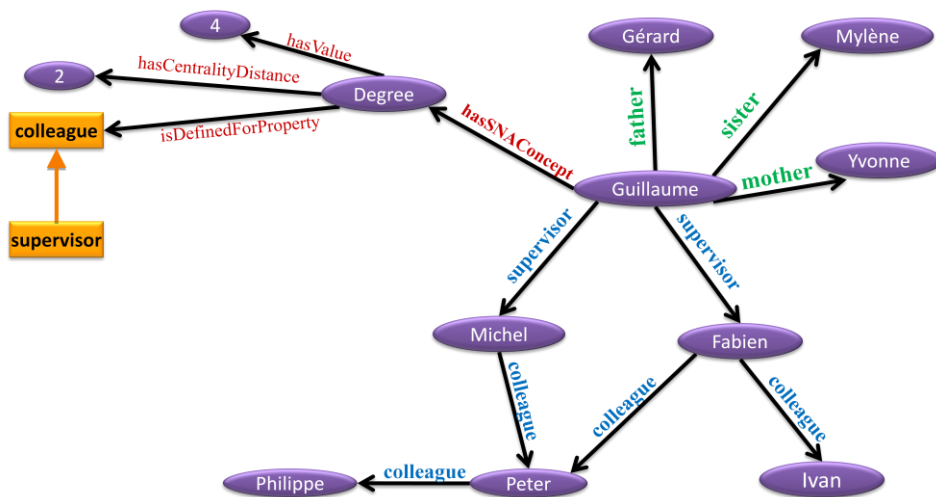

Fig.3. Example of a social graph enriched with social network analysis results.

On top of SemSNA we proposed SPARQL formal definitions of SNA operators handling the typing of the semantic representations of social networks through parameterized queriesfocusing automatically on specific path patterns, involving specific resource or property types. The SPARQL queries that we designed are based on extensions of the SPARQL language that are implemented in the semantic graph engine CORESE [4]. In particular, the property path extension of CORESE [5] enables us to extract paths in RDF graphs by specifying multiple criteria such as the type of the properties involved in the path with regular expressions, or edge directions or constraints on the vertices that paths go through.

3.2 Semi-automated semantic structuration of folksonomies

Social tagging is a successful classification means to involve users in the life cycle of the content they exchange, read or publish online. However, folksonomies resulting from this practice have shown limitations (spelling variations, lack of semantic relationships, etc.) that significantly hinder the navigation within tagged corpora.One way of tackling these limitations is to semantically structure folksonomies. This can help navigate within tagged corpora by (1) enriching tag-based search results with spelling variants and hyponyms, or (2) by suggesting related tags to extend the search, or (3) by semantically organizing tags to guide novice users in a given domain more efficiently than with flat lists of tags or occurrence-based tag clouds, or (4) by assisting disambiguation. We designed a tagging-based system that integrates collaborative and assisted semantic enrichment of the community's folksonomy. We proposed formal models and methods to support diverging points of view regarding the semantics of tags and to efficiently combine them into a coherent and semantically structured folksonomy, see [6] and [7] for details.Our approach consists in creating a synergistic combination of automatic structuring methods to bootstrap the process, and of users' contributions at the lowest possible cost through user-friendly interfaces to improve the results. The system *supports conflicting points of view* regarding the semantic organization of tags, but also helps online communities *build a consensual point of view* emerging from individual contributions.

We proposed an RDF schema (**Fig.4**), SRTag[3], which makes use of named graphs mechanisms[8] and [9] to capture statements and points of views. We encapsulate statements about tags within a named graph typed assrtag:TagSemanticStatementor more precise subclasses.
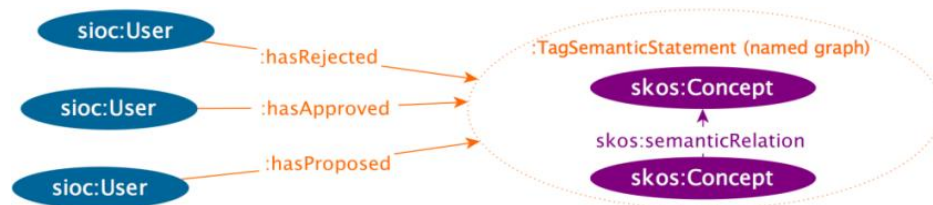

Fig.4.SRTag RDF schema

The relationships between tags can be taken from any model, but we chose to limit the number of possible relations to thesaurus-like relations as modeled in SKOS. Then we modeled a limited series of semantic actions which can be performed by users (represented using `sioc:User`class), namely `srtag:hasApproved`, `srtag:hasProposed`, and `srtag:hasRejected`. This allows us to capture and track users' opinions on the asserted relations, and thus to collect consensus and diverging points of view. We distinguish different types of *automatic* and *human* agents according to their role in the life cycle of the folksonomy. We modeled different subclasses of the class `sioc:User`in order to filter statements according to the users who approve it. This includes `srtag:SingleUser`which corresponds to regular users of the system, `srtag:ReferentUser`(e.g. an archivist) who is in charge of building a consensual point of view, `srtag:TagStructureComputer` which corresponds to the software agents performing automatic handling of tags, and `srtag:ConflictSolver`corresponding to software agents which propose temporary conflict resolutions for diverging points of view before referent users choose one consensual point of view.

As a result, our model allows for the factorization of individual contributions as well as the maintenance of a coherent view for each user and a consensual view linked to a referent user. Furthermore, by modeling different types of agents who propose, approve or reject tag relations, we are able to set up a complete life cycle of enriched folksonomies (**Fig.5**) which starts with a "flat" folksonomy (ie. with no semantic relationships between tags) and can be decomposed as follows:
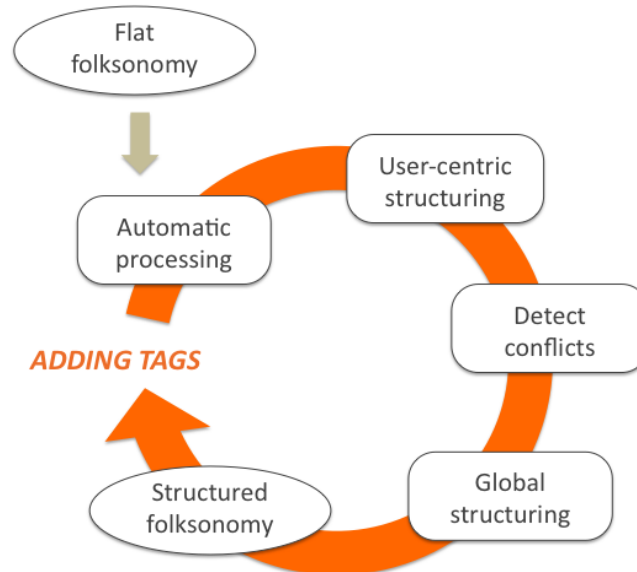

Fig.5. Folksonomy enrichment lifecycle

1. **Automatic processing** is performed on tags using (a) string based comparisons between pairs of tags and (b) methods on the network structure of the folksonomy

(linking tags, users and resources),see [6] and [7] for details.Agents then add assertions to the triple store stating semantic relations between tags. These computations are done overnight due to their algorithmic complexity. We conducted a benchmark to evaluate the ability of string metrics to retrieve other types of semantic relations such as *related* relation, or *narrower* or *broader* relation, also called *hyponym* relation. We also implemented the algorithm described by [10] in order to extract *subsumption* relations which consists in looking at the inclusions of the sets of users associated to a tag. Co-occurrences of tags have also been used for determining missing *related* relations. The zoom in**Fig.6**shows an extract of the structured folksonomy we obtain just after automatic processing.
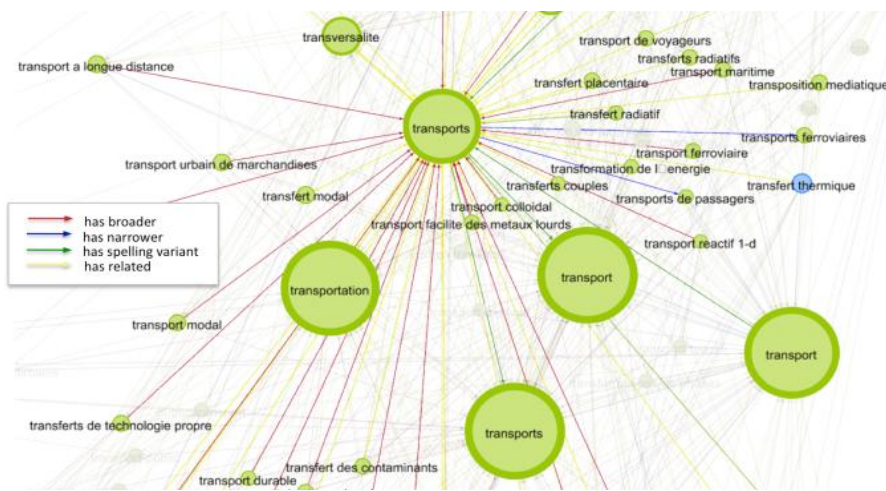

Fig.6. Example of the results of automatic processing with the String Based method showing tags linked with the tag "transports". The size of the nodes indicates the in-degree.

2. **User's contributions**: peoplecontribute through interfaces integrated into tools they use daily by suggesting, correcting or validating tag relations. Each user maintains his point of view, while benefitting from the points of view of other users.
3. **Conflict detection**: as logical inconsistencies arise between users' points of view, a software agent using production rules detects these conflicts and proposes resolutions when possible. The statements proposed are used to reduce the noise that may hinder the use of the system when, for instance, different relations are stated about the same pair of tags.
4. **Global structuring**: the statements from the conflict solver agent are also used to help *referent users*(e.g. librarians) in theirtasksto maintain global and consensual thesaurus with no conflicts. This view is used to filter the suggestions of related tags by giving priority to referent-validated relations over others suggested by computers or individual users.
5. **Structured folksonomy**: at this point of the life cycle we have a semantically structured folksonomy in which each user's point of view co-exists with the con-

sensual point of view. Then a set of rules is applied to exploit these points of view in order to offer coherent navigation and search tools to each and every user.

6. **And again:** another cycle restarts with automatic handlings to take into account new tags added to the folksonomy.

3.3   Identifying communitiesand shared interests

Building on top of our results on semantic social network analysis and folksonomy structuring, we proposed a community detection algorithm, SemTagP. This algorithm not only offers to detect but also to label communities of interest by exploiting at the same time the structure of the social graph and the semantics of the tags used. Doing so, we are able to refine the partitioning of the social graph with semantic processing and to label the activity of detected communities. We tested and evaluated this algorithm on the social network built from Ph.D. theses funded by ADEME, the French Environment and Energy Management Agency. We showed how this approach allows us to detect and label communities of interest and control the precision of the clustering and labeling.

SemTagP ([2] and [3]) is an extension of the RAK algorithm [18]in which we turned the random label propagation into a semantic propagation of tags: instead of assigning and propagating random labels, we assign to actors the tags they use and we propagate them using generalization relations between tags (e.g. `skos:narrower/ skos:broader`) to merge over-specialized communities and generalize their labels to common hyperonyms (for instance merging football and handball communities into a sport community).

We use the directed modularity on RDF directed graphs to assess the quality of the community partition obtained after each propagation loop. When a partitioned network has a high modularity, it means that there are more connections between nodes within each community than between nodes from different communities.

SemTagP iteratively propagates the tags in the network in order to get a new partitioning: nodes that share the same tag form a community. During a propagation loop each actor chooses the most used tag among its neighbors, for a tag t we count 1 occurrence for each neighbor using t and 1 occurrence for each neighbor using a skos:narrower tag of t. We iterate until the modularity stops increasing. The penultimate partitioned network is the output of the algorithm.

In our first experimentation, we witnessed that some tags with many `skos:narrower` relations absorbed too many tags during the propagation phase, such as the tag "environnement" (environment), which is ubiquitous in the corpus of the ADEME agency. Such tags grouped actors in very large communities. Consequently, we added an option to refine manually the results: after the first propagation loop we present the current community partition and labeling to a user that can reject the use of `skos:narrower` relations of tags labeling too large communities. Then, we restart the algorithm and repeat this process until no more relation is rejected.

We formalized our algorithm in [2] and [3] and it was implemented on top of the semantic graph engine CORESE-KGRAM [4] that supports RDF/S and SPARQL 1.1 query language. We delegate all the semantic processing performed on the graph to

the semantic graph engine, taking benefits of SPARQL queries to exploit semantic relations between tags. Notice that the pattern matching mechanism of KGRAM's SPARQL implementation is based on graph homomorphism that is an NP complete problem. However, many heuristic optimizations enable us to significantly cut the time calculation of the RDF graph querying.
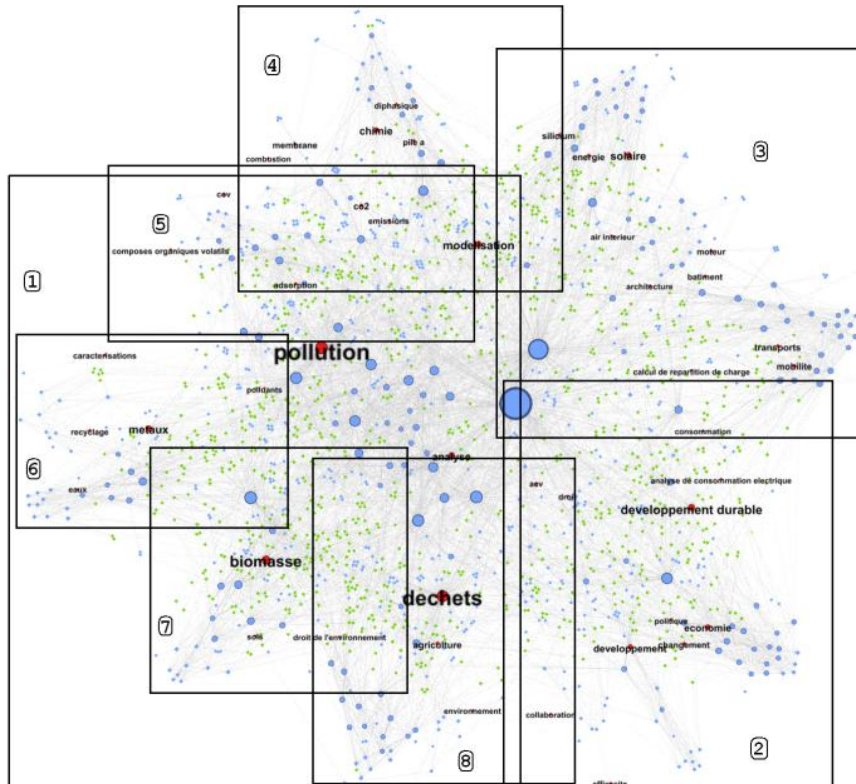

Fig.7. Ph.D. social network of the ADEME with tags labeling the communities.

In Fig.6 we show an overview of the ADEME social network with the labels of the communities identified by SemTagP. We used a graph visualization tool, GEPHI, with a force layout to render the results. The size of the nodes is proportional to their degrees, and the size of the tags is proportional to the size of the labeled communities. Groups of densely linked actors are gathered around few tags, which highlight the efficiency of the algorithm at partitioning the network. Moreover, communities that are labeled with tags representing related topics are close in the visualization, which enables us to build thematic area of the network using the labeling of the communities. In Fig.6, communities displayed in framed area are respectively labeled with tags related to: pollution (1), sustainable development (2), energy (3), chemistry (4), air pollution (5), metals (6), biomass (7), wastes (8). For instance, the area 3 contains tags

related to energy production and consumption with the tags *energie* (energy), *silicium*, *solaire* (solar), *moteur* (engine), *bâtiment* (building) and *transports*. This observation shows that SemTagP labeled closest communities with related labels.

4   Implementing ISICIL

The ISICIL platform integrates all the approaches summarized in the previous section into a web architecture deploying interconnected social semantic tools on an intranet to support business intelligence and technology monitoring activities including watch, search, notification and reporting[4].

4.1   General architecture of the platform

The platform is a typical REST API architecture, builtas a JavaEEwebapp,hosted in a classical servlet container.For practical reasons we split this project into 3 layers:

The core layer embeds the CORESE/KGram library as the main triple store, conceptual graph engine and SPARQL 1.1 compliant interpreter [11]. As such, we make an extensive use of new features of SPARQL 1.1 like update, named graphs, paths. This layer also implements the read/write mechanisms between CORESE/KGram and the ISICIL custom persistence system.

The business layer is dedicated to the implementation of ISICIL modelsand publishes services and methods required for interacting with the semantic engine. Each business object has a dedicated service for inserting/updating/deleting annotations. For mobile devices concerns, we also have pre-wired some of the main queries users can have to ask to the system. This work greatly simplifies the client-side interaction with the server and the parsing of the results. For those who need complex queries, a SPARQL endpoint service is also provided.

The RESTful API represents the HTTP translation of the business layer. Almost all of the business services have a corresponding web service. Because of the sensitive aspect of the business intelligence information, we took special care of security issues. An ontology based access control model had been designed and implemented in a prototype based on ISICIL datasets [12].

4.2   Clients of the RESTful API

Two software clientswere also developed during the project. One is a semantic enhancement of anindustrial quality open source wiki[5]whose role is to support web scraping and reporting and the second one is a Firefox add-on whose goal is to propose a semantic augmented browsing experience to the user. Users can access ISICIL information through them, but their activities within these toolsalso feed the ISICIL

---

[4]The platform is open-source, available on the INRIA repository,
    https://gforge.inria.fr/projects/isicil
[5]MindtouchDekiwiki, http://www.mindtouch.com/

social network. For instance, the wiki tracks collaboration between contributors, comments, tags and proposes some reporting features through several dashboard widgets that can be embedded into wiki pages. It also proposes corporate documentation templates to guide users in their editorial activities and to push them to bring this kind of documents out of the office.

The second tool, the Firefox add-on provides a bookmark extension, called *Webmarks*, which allows the user to share his web references through the ISICIL social network. Furthermore, users can scrap (archive) a web page that is uploadedto the wiki and use it in a wiki page to have a backup copy, to illustrate something or to initiate a discussion or a report. The Firefox add-on also embeds anRDFa parser that analyses metadata hidden in the pages that are marked. This feature for instance is used to bootstrap the Webmarking form with some Dublin core, FOAF or OGP metadata[6]. RDFacan also be injected in the PHP templates of corporate legacy tools to use their web front end as integration points and analyze what the user is currently browsing on a corporate intranet tool and start tracking his activity or suggest him some other good readings or good persons to contact, or relevant services for the data he accesses.The Fig.8illustrates the way these tools are connected together.
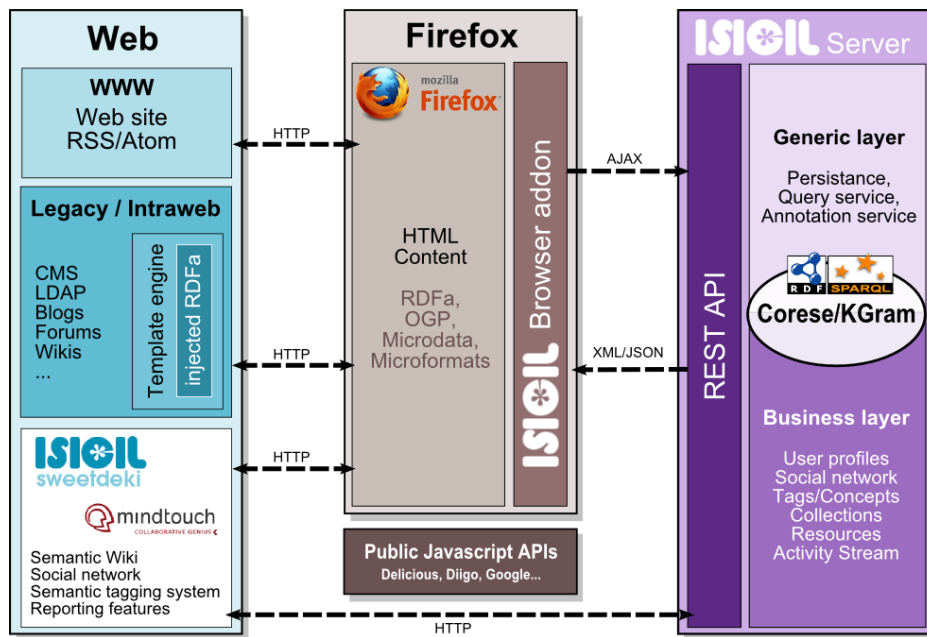

Fig.8. ISICIL General Architecture

---

[6]http://developers.facebook.com/docs/opengraph/

4.3 Data consolidation, metrics computing & connectors

One thing that the three layers architecture allowed was the building of independent, small and easy-to-maintain software agents, sharing the same business layer that the REST API. This is of primary importance as when the model evolves, automatically every dependent module benefits of the business layer update.

In the section 3, we have presented the algorithms that we are using for respectively social network analysis, semantic structuration of folksonomies and community detection. Those resource-consuming processes cannot be done in real-time. For that purpose, those independent software agents can be dispatched through a local network and remotely consolidate the ISICIL database with their results. Their execution can also be scheduled as an entry in a crontab or in any other scheduling system. Our data connectors are following exactly the same design principles.

4.4 ISICIL Ontologies: effective reuse and required adaptations

Modeling the identity was a critical step during the design phase. As a conception guide, we relied mostly on a real use case from the French agency for sustainable development and energy mastery (ADEME). They were very interested in merging their contacts databases into a unified social network. For that task, we set up the ISICIL social network with both of their corporate directory and PHD contact databases, but to do so, many difficult aspects had to be taken into account:

- *Distinguish employees of the company and external persons*.Bootstraping the social network only with ADEME employees was not considered of interest. The most valuable information in such a corporate social network resides in the links between inner experts and industrial or academic partners.
- *Manage the turnover*. Many experts stay only a few years as members of the agency. So the system should be able to disable their account, but also keep track of what they did before and what they are doing next.
- *Bridge existing online identities under a unique URI*.Many of the staff members are contributing to the corporate knowledge under many different online identities on intranet or extranet tools, like blogs or forums. Our first work was to integrate, curate and merge all of these fragments into a single coherent entity so that it is possible to answer to (not so) simple questions such as "who wrote that?", "who's currently working on that topic in the agency?"…
- *Enable online communities*.Social network users like to create online groups and communities to share information. But in our case, groups inherited from the organization chart were not considered representative of the way members wanted to collaborate. So we had to integrate in our model a second kind of groupto model theseparallel online communities.

The part of the ISICIL model dedicated to identity, groups and communities is illustrated by **Fig.9**. We made design choices using some parts of the FOAF ontologies, such as `foaf:Person`, `foaf:Group`, `foaf:Organization` to model information from the corporate organization chart while SIOC was used to model online activities.

Though, any user of the ISICIL software has at least two identities, an official one named by a `foaf:Person` URI and an online account represented by a `sioc:UserAccount`.
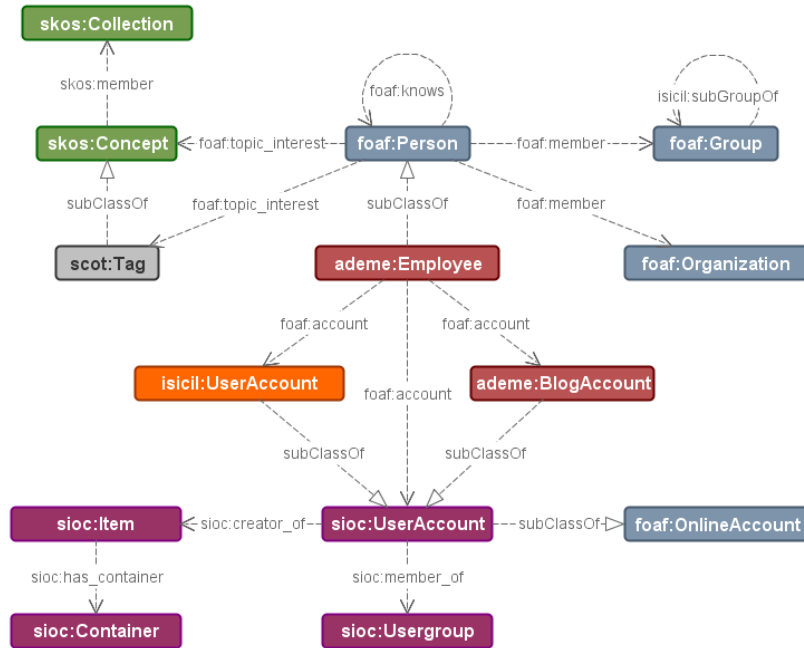

Fig.9. ISICIL Global model

At the beginning, social relations were defined as properties of the *relationship* on-tology[7]. This model proposes a many `foaf:knows` sub-properties enabling the specification of each social relation type. But, we faced a lot of issues when we tried to calculate some basic SNA metrics. Results were false because most of these relations were defined symmetric while we needed to take into account in-degree and out-degree. For that purpose, we had to define our own non-symmetric sub-properties of `foaf:knows`.

Another important part of the design task was the integration of the ADEME controlled vocabulary into the system and its articulation it with other legacy folksonomies inherited from intranet CMS/Blogs/Wikis that provide free form social tagging. The idea was to bootstrap the ISICIL tagging system in such a manner that we can propose some relevant auto-completions and tag suggestions, assisting users in their tagging activities. This approach is also fostered by our semi-automated semantic structuration of folksonomies presented in section 3.2.

To do so, we made an extensive use of the SKOS vocabulary[8], separating "official" controlled terms (`skos:Concept`) from tags created by users. For tags, we are using

---

[7] http://vocab.org/relationship/.html
[8] http://www.w3.org/TR/skos-reference/

the SCOT ontology [13] that defines a `scot:Tag` as a subclass of a `skos:Concept` (**Fig.9**). As explained in the section 3.2, weinferSKOS relations between these concepts.The NiceTag vocabulary [14] is used to model *tagging events*as named graphs including a creator, a date and a targeted resource. Tagging events represent a very important part of the online activity that we are tracking. We analyse the tags an expert is currently using, and tags that are related, to suggest him relevant people to contact. Moreover, a macroscopic usage analysis on tagging events for a given period provides valuable information about scientific and technological trends.

Business Intelligence tasks are not always focusing on the same topics. So, an expert should be able to describe his current field of interest by a set of keywordsthat he can activate or deactivate according to his current activity. For that purpose, we used the `foaf:topic_interest` propertyas a link between an expert and a tag, a concepts or a person. As we can see on the **Fig.9**, the user can organize these topics in labeled lists (`skos:Collection`) that he can share with others.The user who described his topics of interest this way, will be notified of incoming information produced over the ISICIL social network that is related to his active topics.These declarations of interests could also be exploited to improve the suggestions of readings and contactsthe system can provide.

One of the principal concerns of experts at the ADEME is to qualify information. How to distinguish valuable information in a flow of poorly reliable data?Answering such a question requires information about authors.The ISICIL platform manages two kinds of documentary resources: legacy documents coming from corporate databases and pages produced on the ISICIL semantic wiki.

For legacy documents, most of the time, we know nothing about authors and the context where these documents were created. Thus, it is almost impossible to qualify "a priori" their informational value.But it is possible to analyzethe way ISICIL users reference it. For instance, if a well-known expert in the ISICIL social network addsa Webmarkon a report and tag it, there are chances that this document contains something useful. If many users do the same, this probability increases. The ISICIL platform computes ranking metrics based on Webmarksand tagging analysis that give to documentary resource a valuethat represents its social relevance. These values are used to sort suggestions or the results of a full-text search.

For ISICIL wiki pages, we have made an extension of the SIOC ontologyfor wiki pages, revisions, attached documents and comments. This model is mainly based on `sioc:Item` and `sioc:Container` and contributors on wiki pages are identified by their `sioc:UserAccount` and thesioc:`creator_of` property.Then, in addition to the previous analysis made from tagging and webmarking, we also have the social information about the author.

5 User Interfaces

5.1 Capture online activity

For a social network to be alive the social data need to be attractive. None of the ISCIL algorithmsareefficient without good tagging, interesting wiki content, web-

marks or social relations. Thus, providing intuitive tools to produce these data was a primary task.The commercial wiki (**Fig.10**A) we choose to be enhanced with semantic features in ISICIL already provides a friendly WYSIWYG editor and a templating system enabling the wiki pages customisation. We have replaced the original tagging system and we have implemented many hooks to capture editorial events but most of the editing system was left as is.

Finally, the webmarkingextension (**Fig.10**B) allows a user to specify the type of the resource that he is going to mark. He can also specify if and how this resource is geo-localized (e.g. a place), temporally constrained (e.g. an event) or socially linked (e.g. a colleague). The user can also scrap the web page content to preserve and share it on an ISICIL wiki page. The webmarking tool also provides assisted tagging and sharing features.

5.2   Search features

Once the social network is fed, BI tasks require global search functionalities to retrieve information. The Firefox add-on (**Fig.10**C) provides a "search and suggest" feature seamlessly integrating tag structuring capabilities (left part). The user was about to drag the tag "energy" towards the "spelling variant" area to state that the tag "energie" (the tag currently searched for) is a spelling variant of "energy". In this interface, as a side effect of searching and refining search results, the users make statements on the structure found for the folksonomy and contribute to its maintenance.

Recently, we have improved our full-text searching system with Fresnel based data selectors [15]. Fresnel lenses were written for each type of resources we want to show. After a full-text query, the system applies the lenses on the list of the matching URIs. Then according to the `rdf:type` of the resource, it automatically selects the adequate properties and output formats as defined in the corresponding lens.

5.3   Reporting and data visualization

As we said before, the ISICIL wiki (called SweetDeki) is an enhancement of a commercial wiki which integrates a custom scripting language, and powerful extension mechanisms. We developed a set of extensions that communicate with the ISICIL REST services like SNA, etc. Some inject RDFa annotations in pages, generate semantic annotations such as collaboration notifications between users for the social network, replace the original tagging system with our own, etc. We also developed a set of extensions for the WYSIWYG editor for inserting dynamic visualizations or data into wiki documents. For example, widgets from the ISICIL social networks services such as the ones presented in **Fig.11** can be integrated in any document. They represent collaboration diagrams between users, users-tag-users relationships, or results from the Semantic Network Analysis tools.
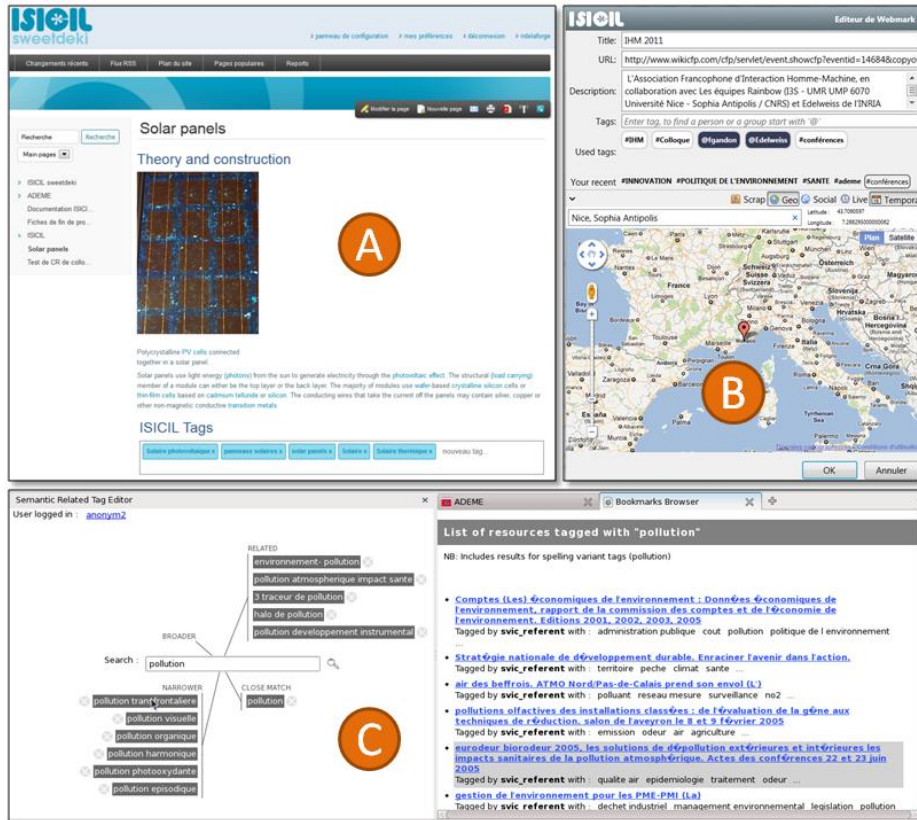
Fig.10. ISICIL GUIsample : (A) the semantic wiki, (B) the webmarking dialog,
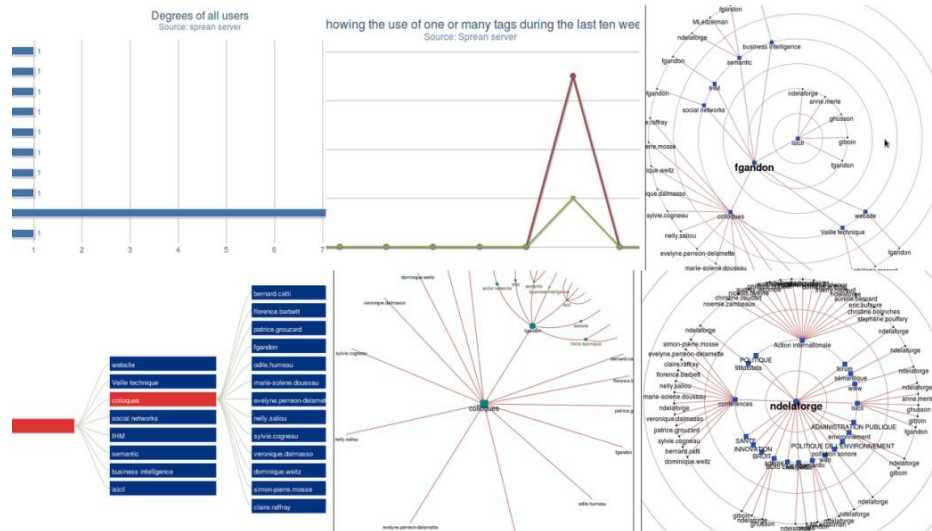(C) the "search and suggest" interface

Fig.11. ISICIL dashboard widgets sample

6   Related works

As this article focuses on the overall ISICIL platform we will restrain this section to this topic too. The KIWI framework[9] has similarities with the ISICIL framework as both have service oriented architecture that provides a set of RESTful web services, both use a fully versioned triplestore, a full text and metadata search index and services to manage content, users and tags. However, where KIWI describes itself as a "generic framework for Semantic Social Media" and is a natural evolution of the semantic wiki IkeWiki into a framework (KIWI means "Knowledge in a Wiki"), ISICIL is a framework dedicated to semantic social networks that provides a social network analysis stack and a system for structuring semi automatically the tags. Moreover, the tools developed on top of this framework have been targeted to business intelligence users: a semantic wiki based on the Mindtouch Core open source wiki that interacts with the social network and acts as a reporting tool, FireFox extensions dedicated to searching and exploring the network in order to locate experts on particular domains or collaborations between members on particular subjects, social bookmarking tools, etc. Other works related to semantic enhancement of content management systems (CMS)share common points with our work (many propose to use ontologies to link resources, people and tags), see [16] for a survey, but are not focused on the social network part or on the semantic enrichment of tags as much as ISICIL.For the years to come, works conducted by members of the IKS European project[10], that address the problem of interaction with semantic contents in CMS and in the Linked Open Data [17], open new perspective for framework like ISICIL.

---

[9]http://kiwi-project.eu

[10]http://www.iks-project.eu

7   Discussion and conclusion

To conclude the project, an experiment was conducted over several months at ADEME. Its aim was to evaluate the usefulness, usability and collaborative practices allowed by the platform for a representative sample of members of the ADEME. This sample was composed of engineers, librarians and PhD students.

The experiment consisted of a series of activities carried out in a spirit of co-design with users: a) a questionnaire on business intelligence practices b) individual sessions of exploration and discovery of the platform c) guided cognitive tracking sessions in the platform d) collaborative online sessions supported by the communication features of the platform (chat and comments), e) focus group.

The data collected during these activities were analyzed in terms of:

- Purpose: to assess whether the platform is useful for achieving business objectives, to accomplish the tasks of business intelligence.
- Usability: evaluating ergonomic three main criteria: effectiveness, efficiency and satisfaction when interacting with the platform.
- Collaborative practices: identify the practices and forms of cooperation induced by the platform.

In general, the ISICIL platform has been validated regarding the usefulness and collaborative practices: users clearly see the collaborative potential of ISICIL. However, the usability of the platform is the criticalpoint. The main point is the lack of unity among the features, it was difficult for the user to understand that the building blocks of ISICIL (browser, wiki and webmarks) form a whole; interfaces are isolated from each other, and users lack a main interface that centralizes and guidesthem; thismade the understanding and learning of the platform too heavy for the user. Despite these negative aspects of usability, multiple users ADEME expressed the wish to continue experimenting with ISICIL, in addition, discussions are underway with ADEME on possible actions for further developments of ISICIL within the agency. To improve the usability of the platform, ergonomic recommendations have been proposed. Many of these recommendations have already been taken into account.

As we have seen, the ISICIL platform is innovative in many aspects of the features it proposes. This truly experimental characteristic has a cost. Some of these features are not optimized enough and can involve sometimes performance issues.

In this paper, we have presented the ISICIL platform as the result of a four years research project. We have seen the core algorithms and the software architecture.In this project, we can say that we have fully embraced the latest technologies of the semantic web and the entire architecture of the software suite relies on semantic web standards. The main difficultyhere resided in the articulation of many lightweight ontologies in a read and write context.Sometimes we had to make some arbitrary choices in the design in order to solve inconsistency issues implied by this aggregation. But, we believe that this experiment proves the effectiveness of semantic web technologies in a "real world" scenario of business intelligence.

References

1. G. Erétéo, M. Buffa, F. Gandon, O. Corby: *"Analysis of a Real Online Social Network Using Semantic Web Frameworks"*, 8th Int. Sem. Web Conf., ISWC'2009. (2009)
2. G. Erétéo, M. Buffa, F. Gandon, *"SemTagP: Semantic Community Detection in Folksonomies"*, IEEE/WIC/ACM International Conference on Web Intelligence, 2011, Lyon.
3. G. Erétéo, *"Semantic Social Network Analysis"*, PhD Thesis Telecom ParisTechInria Orange Labs, April, 2011
4. O. Corby, R. Dieng-Kuntz, C. Faron-Zucker: *"querying the semantic web with the corese search engine"*. ECAI/PAIS2004. (2004)
5. O. Corby: *"Web, Graphs & Semantics"*, ICCS'2008. (2008)
6. F. Limpens., F. Gandon, M. Buffa, *"Helping online communities to semantically enrich folksonomies"*, in proceedings of Web Science 2010, WebSci10. (2010)
7. F. Limpens, *"Multi-points of View Semantic Enrichment of Folksonomies"*, PhD thesis, University of Nice, http://hal.archives-ouvertes.fr/hal-00530714/fr (2010)
8. J. Carroll, C. Bizer, P. Hayes, and P. Stickler. *Named graphs, provenance and trust*. In WWW '05: Proc. of the 14th international WWW conference, p. 613–622, (2005)
9. F. Gandon, V.Bottolier, O. Corby, and P.Durville. Rdf/xml source declaration, w3c member submission. http://www.w3.org/Submission/rdfsource/, 09 2007.
10. Peter Mika. Ontologies are Us: a Unified Model of Social Networks and Semantics. In *ISWC*, volume 3729 of *LNCS*, pages 522–536. Springer, 2005.
11. O. Corby and C. Faron-Zucker, *The KGRAM Abstract Machine for Knowledge Graph Querying*, IEEE/WIC/ACM International Conference. (2010).
12. S. Villata, N. Delaforge, F. Gandon and A. Gyrard, An Access Control Model for Linked Data, OTM Workshops, 454-463. (2011)
13. HakLae Kim, John G. Breslin, Sung-Kwon Yang, Hong-Gee Kim, *Social Semantic Cloud of Tag: Semantic Model for Social Tagging*, KES-AMSTA, 83-92. (2008)
14. A. Monnin, F. Limpens, F. Gandon, D. Laniado, *Speech acts meet tagging: NiceTag ontology*, I-SEMANTICS. (2010)
15. E. Pietriga, C. Bizer, D. Karger, R. Lee, *Fresnel: A Browser-Independent Presentation Vocabulary for RDF*, ISWC 2006, p. 158-171. (2006)
16. Laleci, G.B., Aluc, G., Dogac, A., Sinaci, A., Kilic, O. &Tuncer F. (2010). *"A semantic backend for content management systems"*. In Knowledge-Based Systems, 23(8), 832-843
17. V. Damjanovic, T. Kurz, R. Westenthaler, W. Behrendt, A. Gruber, S. Schaffert, *"Semantic Enhancement: The Key to Massive and Heterogeneous Data Pools"*, In Proc. of the 20th International IEEE ERK Conference. (2011).
18. Raghavan, R.N., Albert, R., Kumara, S.: Near Linear Time Algorithm to Detect Community Structures in Large Scale Network. Phys. Rev. E, 76, 036106 (2007)