

THEMATIC OASES

Towards a Social Semantic Web

Maria Teresa Pazienza, Savino Sguera and Armando Stellato
DISP, University of Tor Vergata, Via del Politecnico 1, Rome, Italy

Keywords: Semantic Web, Social Web, Social Frameworks.

Abstract: A few years have passed from the first alarming yells about the unmanageable growth of data which is literally exploding from the Web. While Web2.0 technologies, born and grown from the crowd of the web community, are reaching their full maturity, and with W3C eventually managing to give concreteness to Berners-Lee Semantic Web Vision through a plethora of new languages and protocols, the same problem is still a living matter. Lots of vendors and providers offer social services with more-than-overlapping aspects, but with no intentions of sharing their data. RSS aggregators as well as blog and mailing list scrapers, erupt tons of data which are irritatingly replicated by search engines indexes. At the same time, most of the information services (wikis, blogs, mailing lists, forums, newsgroups etc...) still maintain their traditional functionalities and move no step forward reaching any kind of interoperability. In this paper we analyze the current scenario and propose our personal view on how new Semantic Web technologies could be employed to give life to a new generation of social, heterogeneous and coordinated informative services: Thematic Oases.

1 INTRODUCTION

The original vision of the Web, as Tim Berners-Lee exposed it in his book: "Weaving the Web" (Berners-Lee, 2000), seems now very close to its realization: a web equally participated by major publishing entities as well as from ordinary people willing to make public their thoughts, artworks, opinions and ideas. Today technologies and approaches coming from new Web2.0 paradigms have helped this dream come true: now we have millions of people blogging, contributing to social networks by producing and publishing huge amounts of digital stuff in several multimedia formats, and sharing their knowledge through Wikipedia or in domain specific wikis. Yet the several possibilities that new Web era is giving to the mass, are producing a lot more information than before, but are not sensibly improving the way we access it: the first alarming yells about the crescent unmanageable grow of data which were characterizing the growth of the traditional Web are still an important warning to take into account. Semantic Web technologies and standards fostered by the W3C are trying to address this issue, by providing vocabularies and methodologies for organize web data in a decentralized and neutral way. The process of

realization of the Semantic Web layer cake of language and protocols is near to the end, though, paradoxically, the adoption of these standards is hampered by those which autonomously generated from the Web2.0 stream of innovation. It appears evident as these new standardized and open technologies should make their way through the resistance of existing service providers, possibly beneficiating from the success of new open semantic applications and frameworks. In this paper we analyze the current scenario, considering state-of-the-art on social/semantic organization of data, and propose our personal view on how Semantic Web technologies could be employed to give life to a new generation of social, heterogeneous and coordinated informative services: *Thematic Oases*

2 SCENARIO AND RELATED WORKS

Most recent works and discussions on the web made clear that the need for actual data portability and shareability has become the top priority for web and Software as a Service (SaaS) applications providers.

There is an amazingly increasing number of differently flavoured social web applications, which

inherently lead to data replication all over the web: every time a user joins a new social service she probably has to sign up, invite friends, add/remove friends, generally ask for email addresses too, requiring people to send out address verification emails, not even citing the tedious “lost email/password” issues. Negative implications of having users’ data tied to a proprietary platform – worthwhile to name Facebook (Facebook, ©) here, given the exponential growth in terms of interest and users it has been experiencing (100.000 new users per-day, mostly in the golden over-25-years-old market share) – with its own markup language and its own set of API, are evident: user’s experience is based upon a specific framework and set of enabling technologies, while data portability is – often deliberately – not granted. On the other side, developers are forced to master the n-th set of REST API, to write applications that are not even close to the “write once, run everywhere” paradigm which underlies enterprise software engineering principles, sparkling programmers’ and managers’ interests.

A solution to this problem has been proposed by Google with the Open Social (Google ©) API set, which is a specification for widgets and applications deployable on social networks. Open Social defines three broad areas of specification: Widget/Application, Friends, Activity.

All of these still have a long way to evolve but, yet being not standard at all, they bring powerful concepts of openness and interoperability into the social network marketplace. Personal data, however, are not limited in any way in scope and practice to the usual profile-related information: depending on the service being used, personal information span from pictures to videos, from wikis to blog posts, from forums to discussion groups; the list would go a long way. Heterogeneous information sources continuously change in nature and content, moving around highly dynamic centroids, *topics*, which attract people sharing interests or just the desire of publishing something: personal data, pictures, artworks etc... To name a few, Facebook, Myspace (MySpace.com, ©) and Flickr groups (Yahoo, ©), YouTube channels (YouTube, LLC ©) and LinkedIn (LinkedIn Corporation ©) or web sites aggregating similar feeds from different sources. It is also the case of newsgroups, or wikis. None of the above, however, gives the user a thorough understanding nor a total access to topic-related information.

Thus, building virtual communities of people sharing the same areas of interest, and moving onto *topic-driven* web surfing and information sharing is a key aspect in (re)organizing world’s information.

This is what Radar Networks promises with the forthcoming Twine (Radar Networks ©) which promises to set as the first mainstream Semantic Web application. Twine will construct a RDF graph mapping relationships among people and topics as well, giving the user full control over information organization, providing a mean to share knowledge with like-minded people. Twine follows successful past experiences from both the industrial (see the examples so far) and research worlds: consider past Semantic Browsers emerged from the research community, geared towards *personal semantic bookmarking*, like Semantic Turkey (Griesi, Pazienza, & Stellato, 2007), *social semantic annotation*, as for Piggie Bank (Huynh, Mazzocchi, & Karger, November, 2005), or *Web Services composition* (Dzbor, Motta, & Domingue, 2004).

3 THEMATIC OASES

Industrial and research worlds are thus sharing the same views and aims, colliding towards a new web vision, where “knowledge” is no more a huge amount of (semi)structured text but, turning into a cloud of overlapping concepts, topics and domains. At the same time, it emerges the need for a better organization of the huge amount of information erupting from the stream of available technologies.

To find the way through the plethora of information sources, differentiating in content, presentation and accessibility, a viable approach would be to make these layers explicit, be able to recognize them as several possible manifestations of the same knowledge, and organize them accordingly.

It makes sense, then, to make the jump from single (and in-interoperable) specific services (blogs, wikis, forums, discussion groups and so on...) to huge collectors of information on a open and global scale, which we dare to call *Thematic Oases*. Thematic Oases (TOs, from now on) should provide the main intellectual stream of interests around which knowledge should be organized (and different services be offered). By adopting Semantic Web standards, TOs would be developed around ontological repositories of conceptual knowledge, which will be used as reference vocabularies for accessing contents of federated (or simply annexed) services and (socially) bookmarked web pages.

In this scenario, traditional services will be still reusable and will coexist with their new semantic counterparts, with the former being semantically annotated with respect to the ontologies adopted in given Thematic Oases, and the latter natively

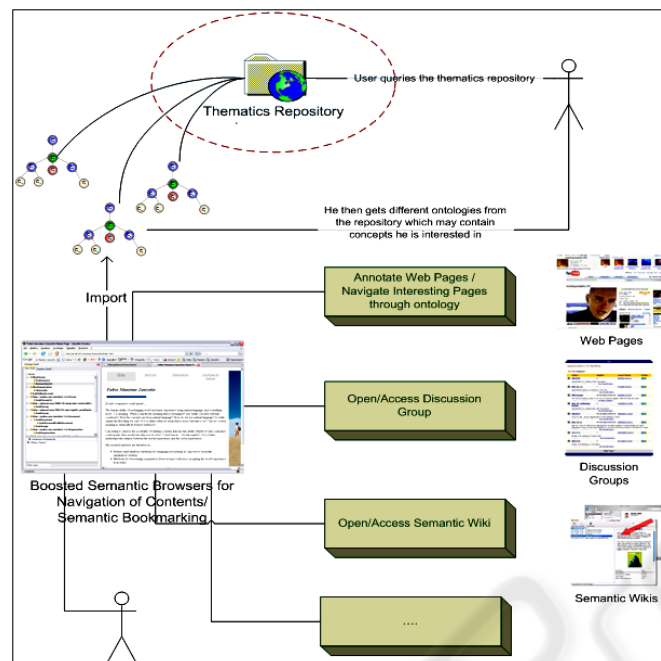


Figure 1: Conceptual Architecture and Use Case of a Repository of Thematic Oases.

supporting a semantic organization of their content. The main principles of TOs should be:

- *Affordable setup*: no more heavy bulked Social Networks held by major company titans. Much the same way a normal web user can now start a forum or a blog using third party (often free) software, Thematic Oases should be at the hand of any user with the availability of a web host or of an hosting service
- *Accessible by (Semantic?) Search Engines*: In our vision, this is surely something related to the open nature of TOs, but would beneficiate at the same time of some commitment from search engines, which will be able to improve quality of searches through proper indexing of semantic annotations publicly exposed by the oases
- *Scalable open architecture*: a given service may be explicitly built upon a TO, committing to its ontologies and content organization. Vice-versa, in an even more open view, independent services may be linked by a given TO. This would allow users to tag the content of these services according to the oasis' reference ontologies, thus easily putting traditional (non semantic-driven) services immediately into play. The same would hold for standard web pages. People could write web pages directly connected to a TO making explicit reference to its vocabulary, as embedded RDFa (Adida & Birbeck, 2007), or could semantically bookmark an external web page (or annotate part of its content) against that same vocabulary.

The above principles should promote a new interpretation of today social networks, where people gets back the ownership of their own data, being able to publish them autonomously and freely move them according to their hosting possibilities. At the same time, these services could be completely defined by the users, according to their specific interests and exigencies, addressing important themes and coordinating different services around their explication, whereas current social networks offer nothing more than well-cooked showcases for exhibiting our personal data, multimedia, and social contacts. Thematic Oases could become just mere aggregators of already existing services, by providing the possibility of storing semantic annotations in their internal repositories.

Let's think about a user willing to buy a new monitor for his Pc. He would access a TO about computer hardware, then browse the hardware-ontology looking for video peripherals and getting the pointer to the monitor concept. He could then decide to learn a bit more about monitors before deciding to buy a new one, thus accessing to the wiki page associated to the concept. Thus he learns interesting features which he uses to filter out a few possible models he is interested in. He finally uses the *search comparisons* functionality, pointing to discussions in forums, reports in webpages/wikis, RSS fed discussions which have been tagged by the community as comparisons and instantiated wrt the specific models selected by him.

3.1 The Ontology Commitment Issue

One of the biggest (and most discussed) problems in the realization of the Semantic Web (Dijck, 2003), was about the difficulty in establishing consensus on domain representations. The major concern was about the natural resistance of companies and providers to commit to any kind of knowledge organization which could not reflect their inner nature and characteristics or simply properly address their specific information needs. Also, the different languages and cultures which must be considered on a World Wide scale needed to be kept into account, posing another obstacle towards ontologies seen as acceptable “shared reference vocabularies” in the Semantic Web. These fears, appear, at least in part, as dictated by an old fashioned way of thinking about knowledge organization. Most widely adopted ontologies contain now very simple descriptions of very specific aspects of a domain (or of reality, in general). The FOAF ontology (Brickley & Miller, 2007) contains just very simple data for describing people personal information and for establishing connections between people. The Basic RDF Geo Vocabulary (Brickley, 2007) is composed of just three attributive properties for describing WGS84 standard latitude, longitude and altitude, plus a generic Point concept for addressing points in the space using these properties. These ontologies can be easily imported in any more complex knowledge organization system, with no fear of generating unsolvable inconsistencies, while leaving the possibility of providing ad-hoc domain descriptions for addressing specific needs, by adding arbitrary concepts and relationships to the imported ones. This approach guarantees a desirable degree of shareability of the collected data (at least, on its higher level descriptive units) while preserving the intellectual independence in modeling specific scenarios and domains. Following this approach, TOs, while allowing for ad-hoc ontologies developed for their specific needs, should foster reuse of “standard ontologies”, thus opening to external linkable services which have been developed independently from their suggested vocabulary, as well as enabling peer-to-peering among different Thematic Oases.

4 CONCLUSIONS AND FUTURE WORK

In this paper we have presented our vision about a possible concrete application of Semantic Web

principles to the social paradigms which are characterizing today Web2.0 era. We firmly believe that the Web community is ready for embracing new air breezing from the so called Web3.0 semantic trend, but is currently blocked by the unavailability of solid and impacting killer-applications giving a rightful reason for learning the next step of Web evolution.

REFERENCES

- Adida, B., & Birbeck, M. (2007, October 26). *RDFa Primer*. Retrieved from W3C: <http://www.w3.org/TR/xhtml-rdfa-primer/>
- Berners-Lee, T. (2000). *Weaving the Web*. New York: HarperBusiness.
- Brickley, D. (2007). *Basic RDF Geo Vocabulary*. Retrieved December 10, 2007, from <http://www.w3.org/2003/01/geo/>
- Brickley, D., & Miller, L. (2007, November 2). *FOAF Vocabulary Specification 0.91, Namespace Document 2 November 2007 - OpenID Edition*. Retrieved from [xmlns.com:http://xmlns.com/foaf/spec/20071002.html](http://xmlns.com/foaf/spec/20071002.html)
- Dijck, P. V. (2003, November 15). *Themes and metaphors in the semantic web discussion*. Retrieved from <http://poorbuthappy.com/ease/semantic/>
- Dzbor, M., Motta, E., & Domingue, J. B. (2004). Opening Up Magpie via Semantic Services. *3rd Intl. Semantic Web Conference (ISWC04)*. Hiroshima, Japan: November.
- Facebook, ©. (n.d.). Retrieved December 10, 2007, from Facebook | Welcome to Facebook!: <http://www.facebook.com/>
- Google ©. (n.d.). *OpenSocial - Google Code*. Retrieved December 10, 2007, from Google Code: <http://code.google.com/apis/opensocial/>
- Griesi, D., Pazienza, M. T., & Stellato, A. (2007). Semantic Turkey - a Semantic Bookmarking tool (System Description). *4th European Semantic Web Conference (ESWC 2007)*. Innsbruck, Austria.
- Huynh, D., Mazzocchi, S., & Karger, D. (November, 2005). Piggy Bank: Experience the Semantic Web Inside Your Web Browser. *Fourth International Semantic Web Conference (ISWC05)*, (p. 413-430). Galway, Ireland.
- LinkedIn Corporation ©. (n.d.). Retrieved December 10, 2007, from LinkedIn: Relationships Matter: <http://www.linkedin.com/>
- MySpace.com, ©. (n.d.). Retrieved December 10, 2007, from MySpace: <http://www.myspace.com/>
- Radar Networks ©. (n.d.). Retrieved December 10, 2007, from Twine: <http://www.twine.com/>
- Yahoo, ©. (n.d.). Retrieved December 10, 2007, from Flickr: Groups: <http://www.flickr.com/groups/>
- YouTube, LLC ©. (n.d.). Retrieved December 10, 2007, from YouTube - Broadcast Yourself: <http://www.youtube.com/>