Perspectives in Semantic Adaptive Social Web

Federica Cena, University of Turin, Italy
Antonina Dattolo, University of Udine, Italy
Pasquale Lops, University of Bari Aldo Moro, Italy
Julita Vassileva, University of Saskatchewan, Canada

The Social Web is now a successful reality with its quickly growing number of users and applications. Also the Semantic Web, which started with the objectives of describing Web resources in a machines-processable way, is now outgrowing the research labs and is being massively exploited in many web sites, incorporating high-quality user generated content and semantic annotations. The primary goal of this special issue is to showcase some recent research in the intersection of Social Web and Semantic Web, that explores the benefits that adaptation and personalization have to offer in the Web of the future, the so-called Social Adaptive Semantic Web. We have selected two papers out of fourteen submissions based on the quality of the papers and we present the main lessons learnt from the overall analysis of these submissions.

1. INTRODUCTION

The Semantic Web, born with the goal of describing Web resources to be machine-understandable and easily sharable and re-used among different applications, is already massively exploited in many web sites, which incorporate user generated content and semantic annotations. The recent advances in the Semantic Web field, and specifically the use of weak semantic techniques, such as microformats for adding semantics to content, provide new standardized ways to represent, process and share information on the Web. This creates a basis for using other more complex methods of processing, such as natural language processing and personalization.

At the same time, the Social Web is growing daily in the number of users and applications. Nowadays, users heavily contribute to Web content, by tagging, commenting, rating and uploading resources, in an architecture of participation, where user contribution and interaction add value. Moreover, users are involved in many social activities such as creating new relationships among other users, recommending and sharing resources, suggesting friends, creating groups and communities, commenting friends activities and so on.

The primary goal of this special issue is to showcase some recent research in the intersection of the Social Web and the Semantic Web, that explores the benefits of adaptation and personalization for the so-called Social Adaptive Semantic Web.

In this paper, we first provide some discussion about the impact and the opportunities of Semantic Web on Social Web (Section 2) and of Social Web on Semantic Web (Section 3). Then, we outline the opportunities for adaptation offered by Semantic Social Web (Section 4) and provide a discussion of the special issue topics and papers (Section 5). Conclusions and open issues are finally presented in Section 6.

2. HOW CAN SEMANTICS ENRICH THE SOCIAL WEB?

Social systems offer a democratic way to store and share information and knowledge. The massive use of tags represents a flexible way to freely associate meta-knowledge to data, and induces unstructured, collaborative 'classification' schemes that are commonly known as folksonomies. Unfortunately, the use of an uncontrolled vocabulary gives rise to problems such as polysemy, synonymy, basic level variation, ambiguity, and multiword expressions [Dattolo et al. 2010]; on the other hand, the data of the Social Web are independent and there is not interoperability among them. Creating value from this data coming from different applications is one of the main roles for the Semantic Web: in fact, while there are plenty of ways to create value by aggregating user contributions, there are few that go beyond summarizing or sorting the data. To sum up, the Semantic Web offers two major ways to enrich Social Web [Gruber 2006]:

(1) *Adding value to user contributions by structuring existing data.* There are three basic approaches to do this:
   — expose data that is already stored in databases to generate current application data. An example in this direction is provided by [Celik et al. 2011], where the authors propose strategies for inferring facets and facet values on Twitter by enriching the semantics of individual messages. They present different methods, including personalized and context-adaptive ones, for making faceted search on Twitter more effective.
   — extract the data retrospectively from user contributions. An example is provided by the work presented in [Baruzzo et al. 2009], where the authors propose PiRATES (Personalized Intelligent Recommender and Annotator TEStbed), a general framework that analyzes documents from the Web, extracts meaningful key-phrases from them, and suggests new abstract and more standard concepts by navigating through ontologies, classification schemata, thesauri, and lexicon.
   — capture the data as people share their information. An example in this direction is shown by Folkview [Dattolo and Pitassi 2011], an innovative way to conceive 'structured' folksonomy, where the data is directly organized using structured semantic descriptions, called dimensions, and the system automatically infers contextual relationships among them.
(2) *Enabling data sharing and computation across independent, heterogeneous Social Web applications*, by combining structured and unstructured data. Semantic Web methodologies and technologies provide the necessary substrate for discovering new knowledge that is not contained in any source, and provide the solution to problems that were not anticipated by the creators of individual Web sites. An example is given in [Kreiser et al. 2009], where the authors present an approach allowing the user community to collaboratively model relations between tags.

3. HOW CAN SOCIAL WEB ENRICH SEMANTIC WEB?

While editing meta-data is technically easy in RDF and tools like Protégé, the process of semantic annotation of resources requires knowing the dictionary and understanding the structure of the domain ontology. Even though a number of ontologies have already been developed for many domains, even the simplest ontology (a nomenclature, or just a list of agreed terms) requires that the annotators adjust their viewpoint to it. Designing ontologies is hard. It requires highly trained experts in the domain who have a detailed and comprehensive knowledge of properties, functions at

different levels of abstraction and understand the concepts and their relationships according to the common view of the community. Automatic translation between two different views of the same domain is hard. As soon as the domain gets more realistic, complex, and interesting, people emphasize different relationships and start interpreting things in different ways. Different communities use different naming agreements, emphasize different relationships and concepts [Al-Jabari et al. 2009; Iorio et al. 2009]. In reality, most of the systems that claim to use ontologies use taxonomies, hierarchies, or concept maps. These simpler semantic representations allow straightforward mapping of terms using applications such as WordNet, but exclude the semantics of the relationships among objects. Research on advanced structural ontology mapping is ongoing, but the practical application of such mapping is still limited.

The Social Web offers solutions to some of these problems. User tags provide an alternative to semantic annotation by experts, by letting users annotate content with words they find personally useful. This results in many discrete (not related to each other) tags, some of which could be irrelevant, misspelled, or wrong. Yet the more people tag a certain resource, the tags that have meaning to more people get used more often. Thus, through crowdsourcing, a set of tags, representative of the semantics of the resource as defined by the group of users who created the tags can be obtained, which is in fact semantic annotation without a pre-defined ontology or taxonomy. Such crowdsourcing is used by data and user-rich corporations, like Google for semantic annotation of their repositories with images, old printed text (reCaptcha), audio and videos, so that they are searchable by various keywords.

The tags can express different semantic dimensions. Through the use of data-mining techniques, it is possible to categorize the tags, and to create hierarchical organizations of categories, resembling ontologies. Automatic ontology generation based on tagged documents may be a promising direction of research [Brooks and Montanez 2006; Fernandez-Tobias et al. 2011]. Yet, there is no guarantee that the automatically generated ontology will be understandable for humans and that humans will agree with it. On the other hand, agreement may not be necessary, since for practical reasons, the ontology is better used by the machine and hidden away from human eyes [Gruber 2006; Weinberger 2008]. The humans can deal with tags, which are user-friendly; the ontology should stay in the background to support machine reasoning and more complex inference and adaptation.

In summary, the social web helps to overcome two problems that the semantic web community has been working on:

— the bottleneck in semantic annotation of web resources, due to the necessity to use consistent and systematic annotations adhering to predefined ontologies, through user-tagging, and
— the bottleneck in creating domain ontologies that are meaningful for particular user communities, through crowdsourced social tagging and data-mining/clustering of tags to derive higher level categories, concepts and relationships among them.

4. ADAPTATION AND PERSONALISATION IN THE SOCIAL SEMANTIC WEB

The social web is characterized with the explosion of user-generated content. In such abundance, it becomes very hard for users to find relevant content, including user annotations (tags). Personalized information filters and recommender systems provide a way for users to stay current in selected areas of interest and has become an active area of research [Durao and Dolog 2009; Tandukar and Vassileva 2011; Wang et al. 2010]. Many of these works are based on lexical analysis of the user-generated content, which is the first step towards semantic analysis. However, they rarely use explicitly semantic web approaches by mapping to ontologies. They typically use statistical approaches (machine learning or data mining) to derive clusters of semantically related items.

There has been a significant interest in the area of social recommender algorithms, probably because there is a clear methodology for evaluation. Researchers use existing datasets provided by the social networking sites, or create their own datasets by crawling or using the APIs of social networks and focusing on the quality of prediction provided by the algorithms using standard mea-

sures of precision and recall. This allows measuring the improvements. Yet, the usefulness of the proposed algorithms from a user perspective in real world recommender systems is not always clear.

The main reason seems to be the lack of transparency and user understanding of the underlying algorithms. A semantic approach based on recommending items using an explicit model of user interests with respect to semantic categories holds a promise to create this transparency, understanding and increase the user's trust in the recommendations [O'Donovan and Smyth 2005].

Apart from recommender systems and filters, personalization with some semantic elements on the social Web is applied also in learning systems [Liang et al. 2012], for text annotation [Baruzzo et al. 2009], and for improving search in social sites, such as Twitter [Celik et al. 2011]. Similar to recommender systems, they use a combination of lexical analysis and statistical data mining methods.

Excluding the approach of [Baruzzo et al. 2009] that uses ontologies, these approaches are not strictly semantic, as they do not involve ontologies, but they do create representations of basic semantics through tags and keywords (as result of lexical analysis), and these can provide a basis for future sophistication, potentially enable mining of concept structures and simple forms of reasoning. An illustration of this kind of work can be found in [Fernandez-Tobias et al. 2011], where statistical methods combined with lexical analysis are used to find clusters of tags with the purpose of creating user profiles and item profiles that can be used later in recommender systems.

Evaluations of the human acceptance of the generated recommendations, especially in realistic settings and with large sets of users are rare and this is clearly an area where more research needs to be done [Konstan and Riedl 2012].

5. DISCUSSION OF THE SPECIAL ISSUE PAPERS

In this section we provide an overview of the topics covered by the papers submitted to the special issue. We received 14 submissions, covering a wide range of topics. Unfortunately, the thorough peer-review process selected only two of these submissions as mature enough to be published in an archival venue such as TIST. This shows that the Semantic Adaptive Social Web is a very interesting area, attracting the attention of many researchers, but at the same time there are many challenging issues that have not been addressed at sufficient depth and methods that have not been evaluated thoroughly enough. We may conclude that the area is active, but not yet mature. This is due to the following factors:

— the presence of different communities (Adaptive Web, User Modeling, Semantic Web and Social Web) with very different points of view, approaches and techniques, that only recently started to cooperate and share results;
— the limited diffusion of real Semantic Web applications, due to the lack of useful tools and robust platforms to support the development of that kind of applications;
— the success of Social Web applications, which catalyzed most of the researchers' attention, at the expense of other aspects.
— the success of purely collaborative recommender systems as a method used for adaptation and personalization on the Social Web, at the expense of content-based approaches, that rely on semantics.

However, we felt it would be interesting and informative to provide an overview of the "work in progress" in the area. To do this, we analyzed the papers submitted to the current special issue as addressing the following aspects:

(1) **Impact of the Semantic Web on Social Web**. This aspect is related to the use of Semantic Web techniques to enrich Social Web data (using for example weak semantic techniques), to the social navigation, social search and semantic browsing, to the creation of semantic mashups, to the management of folksonomies and creation of a mapping with some ontologies.

(2) **Impact of the Social Web on Semantic Web**. This relates to technologies such as user-generated content (e.g., tags, comments) which can be exploited to automatically create, update or populate ontologies.

(3) **Novel systems and approaches combining semantic, social and adaptive aspects**. We refer to applications and tools using Social Semantic Web technologies, and new theoretical and conceptual approaches.

(4) **New challenges in the Social Adaptive Semantic Web**. We refer to novel trust and reputation models, new privacy approaches, and novel metrics for evaluating Social Adaptive Semantic Web applications.

The main outcome of this analysis is that the only topic that has been adequately addressed is the impact of the Semantic Web on Social Web, while the other topics are not completely or adequately addressed. We can also observe that there were some out-of-topics papers (containing neither social nor semantic aspects), which is a signal that there is not yet a common interpretation by the community of the meaning of topics related to the Social Semantic Web.

The two works accepted for publication exemplify the opportunities offered by applying Semantic Web on Social Web. In more detail, both the accepted papers (Social Semantic Query Expansion [Biancalana et al. 2013] and Web Media Semantic Concept Retrieval via Tag Removal and Model Fusion [Chen et al. 2013]) explore the opportunities given by semantic technologies to improve the Web information retrieval.

The paper *Social Semantic Query Expansion* [Biancalana et al. 2013] describes a system that implements a social and personalized extension of traditional query expansion techniques, generally based on a coarse syntactic analysis which builds two-dimensional matrices basically representing distribution of co-occurring terms in a given collection of documents. The authors propose a novel weak semantic technique for query expansion, which exploits three-dimensional matrices, including as an additional dimension the semantic classes (i.e. categories comprising all the terms that share a semantic property) related to the folksonomy extracted from social bookmarking services such as *delicious* or *StumbleUpon*. The idea is to categorize the user needs and preferences in semantic classes. In the paper, the authors demonstrate the improvement of search results by combining tags and folksonomies (Social Web) with a semantic categorization (Semantic Web). The authors carried out an extensive experimental evaluation with both artificial datasets and real users and showed that their approach outperforms traditional techniques. Then, they also present the results of a question-naire aimed to know the users opinion regarding the system. Finally, they provide the results of a complexity analysis, which demonstrates the efficiency of the system in a real time setting. This kind of evaluation, mixing artificial datasets and studies with real users, quantitative and qualitative techniques, and accuracy and performance evaluation, is a good example of a possible methodology to evaluate the complexity of a social semantic adaptive system, even though it does not propose novel metrics or methodologies.

The paper *Web Media Semantic Concept Retrieval via Tag Removal and Model Fusion* [Chen et al. 2013], proposes an approach for enhancing Web multimedia retrieval integrating content-based and semantic-based models. Since the information on the tags can facilitate media search but they are often imprecise, the authors first propose an approach for noisy tag removal based on a Multiple Correspondence Analysis (MCA) algorithm, which is able to capture relationships among features and identify representative and discriminative tags that hold strong correlations with target semantic concepts. In addition, a novel weighted model fusion method is also proposed to combine scores from both tag-based and content-based models, where MCA is used to estimate the parameters involved in each model. The evaluation of the approach with a big dataset shows that the proposed framework outperforms baseline results and several tag removal algorithms and fusion methods. The weakness of this approach is that it was not evaluated with different datasets, and thus the significance of results cannot be generalized. However, the main strength of the paper relies in the clear demonstration of the real advantages of using user-generated annotations (i.e. tags) for

information retrieval, provided that they are correctly used with the help of semantic technologies. The advantages of social search and semantic browsing are also shown.

Even if the two accepted papers are good examples of the opportunities offered by a Social Semantic Web environment, we cannot claim that they make that step further to exploit those opportunities to improve recommender and adaptive systems, that was in our opinion a key issue of the Semantic Adaptive Social Web.

6. DISCUSSION

After presenting the analysis of contributions to the current special issue we can draw some conclusions and identify the main open issues in the field. The main lessons learnt follow:

— Most of the efforts are concentrated on the use of Semantic Web techniques in the Social Web;
— The potential of using Social Web techniques for solving Semantic Web problems are still under-researched;
— Very few works present Semantic Adaptive Social systems;
— The new challenges which arise in Semantic Social Adaptive Web related to the privacy, trust, as well as new metrics and methodologies for the evaluation have not yet been addressed.

The potential impact of merging social and semantic aspects is very important, specifically in relation to the improvement of information retrieval on the Web and recommender systems. However, the opportunity offered by Semantic Social Web for realizing intelligent and adaptive systems have not been adequately taken. Moreover, there are a number of technological issues which can influence and motivate future research in different communities (adaptive systems, user modeling, Semantic Web, Social Web, data mining, knowledge representation, . . .):

— *Semantic Web applications:* A big opportunity is to exploit the Social Web, namely the collaboration among people and the user generated content, to improve the development of Semantic Web applications, thus fostering their diffusion. Some questions remain: how can Social Web data be used to improve Semantic Web applications? How can tags be used to create, populate and maintain ontologies, alleviating one of the main bottleneck of the Semantic Web (i.e. ontology development)? This may be the starting point for new social semantic approaches for the Web, which should lead to the design and development of novel applications having both semantic and social features.
— *Adaptive systems:* The main issues relate to the user modeling process in this new and complex kind of environments. Some questions remain: how can we gather user model data in this environment? Which are the most suitable personalization strategies? Which are the most effective recommendation algorithms in the Semantic Social Web?
— *Privacy issues:* In the Social Web users seem very willing to share their personal content with friends and even strangers, but they are not often aware of the consequence of the disclosure. Exploiting user data from the Social Web can lead to an ethical dilemma, as user requirements may be in conflict with the system goals, especially if they are commercial ones. How can a balance between privacy and disclosure be found that reflects the best interests for both the user and the application? In order to foster the development of Semantic Adaptive Social Web applications new approaches for privacy management are needed.
— *Evaluation metrics and methodologies:* The evaluation of a complex system such as a semantic adaptive social system requires the integration of methodologies from different areas (adaptive systems, recommender systems, social sciences) and the use of both qualitative and quantitative approaches. In particular, it is necessary to implement a combination of evaluation methodologies in order to verify all the aspects involved in a semantic social application, as mere qualitative evaluation is not enough (see [Gena et al. 2013]). In this sense, the paper by [Biancalana et al. 2013] is a good example of an exhaustive evaluation, albeit it does not present any novel approach. Some questions remain unsolved: Which are the most useful metrics to evaluate such applications? Do the results from laboratory studies, or studies based on dead data from data-sets,

provide enough information to evaluate new technologies, and how to validate them effectively in field experiments involving real systems with real users engaged in actual use?

To conclude, the fusion of social and semantic aspects in the Web is a highly promising and appealing area for both adaptive systems and information retrieval. Hopefully this special issue will encourage researchers to continue working to further these ideas and support the advancement of this research area.

ACKNOWLEDGMENTS

REFERENCES

Mohanad Al-Jabari, Micheal. Mrissa, and Philippe Thiran. 2009. Handling Users Local Contexts in Web 2.0: Use Cases and Challenges. In *Proceedings of Workshop on Adaptation and Personalization for Web 2.0, UMAP'09, June 22-26, 2009, CEUR Workshop Proceedings*, Vol. 485.

Andrea Baruzzo, Antonina Dattolo, Nirmala Pudota, and Carlo Tasso. 2009. A General Framework for Personalized Text Classication and Annotation. In *Proceedings of Workshop on Adaptation and Personalization for Web 2.0, UMAP'09, June 22-26, 2009, CEUR Workshop Proceedings*, Vol. 485.

Claudio Biancalana, Fabio Gasparetti, Alessandro Micarelli, and Giuseppe Sansonetti. 2013. Social Semantic Query Expansion. *ACM Transactions on Intelligent Systems and Technology, Special Issue on Semantic Adaptive Social Web* (2013).

Christopher H. Brooks and Nancy Montanez. 2006. Improved annotation of the blogosphere via autotagging and hierarchical clustering. In *Proceedings of the 15th international conference on World Wide Web, WWW 2006, Edinburgh, Scotland, UK, May 23-26, 2006*, Les Carr, David De Roure, Arun Iyengar, Carole A. Goble, and Michael Dahlin (Eds.). ACM, 625–632.

Ilknur Celik, Fabian Abel, and Patrick Siehndel. 2011. Adaptive Faceted Search on Twitter. In *Proceedings of International Workshop on Adaptation in Social and Semantic Web (SAS-WEB 2011), UMAP 2011, CEUR Workshop Proceedings*, Vol. 730.

Chao Chen, Qiusha Zhu, Lin Lin, Mei-Ling Shyu, and Shu-Ching Chen. 2013. Web Media Semantic Concept Retrieval via Tag Removal and Model Fusion. *ACM Transactions on Intelligent Systems and Technology, Special Issue on Semantic Adaptive Social Web* (2013).

Antonina Dattolo and Emanuela Pitassi. 2011. Visualizing and Managing Folksonomies. In *Proceedings of International Workshop on Adaptation in Social and Semantic Web (SAS-WEB 2011), UMAP 2011, CEUR Workshop Proceedings*, Vol. 730.

Antonina Dattolo, Francesca Tomasi, and Fabio Vitali. 2010. Towards disambiguating social tagging systems. In *Handbook of Research on Web 2.0, 3.0 and X.0: Technologies, Business, and Social Applications*, San Murugesan (Ed.). IGI Global, Hershey, PA, Chapter 20, pp. 349–369.

Federico Durao and Peter Dolog. 2009. A Personalized Tag-Based Recommendation in Social Web Systems. In *Proceedings of Workshop on Adaptation and Personalization for Web 2.0, UMAP'09, June 22-26, 2009, CEUR Workshop Proceedings*, Vol. 485.

Ignacio Fernandez-Tobias, Ivan Cantador, and Alejandro Bellogin. 2011. cTag: Semantic Contextualisation of Social Tags. In *Proceedings of International Workshop on Adaptation in Social and Semantic Web (SAS-WEB 2011), UMAP 2011, CEUR Workshop Proceedings*, Vol. 730.

Cristina Gena, Federica Cena, Fabiana Vernero, and Pierluigi Grillo. 2013. The evaluation of a social adaptive website for cultural events. *User Model. User-Adapt. Interact.* 23, 2-3 (2013), 89–137.

Tom Gruber. 2006. Where the Social Web Meets the Semantic Web. In *International Semantic Web Conference*. 994.

Angelo Di Iorio, Silvia Duca, Alberto Musetti, Silvia Righini, Davide Rossi, and Fabio Vitali. 2009. Customized Edit Interfaces for Wikis via Semantic Annotations. In *Proceedings of Workshop on Adaptation and Personalization for Web 2.0, UMAP'09, June 22-26, 2009, CEUR Workshop Proceedings*, Vol. 485.

Joseph A. Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User Model. User-Adapt. Interact.* 22, 1-2 (2012), 101–123.

Alexander Kreiser, Andreas Nauerz, and Fedor Bakalov. 2009. A Web 3.0 Approach for Improving Tagging Systems. In *Proceedings of Workshop on Web 3.0: Merging Semantic Web and Social Web 2009, ACM HT 2009, CEUR Workshop Proceedings,*, Vol. 467.

MinEr Liang, Julio Guerra, and Peter Brusilovsky. 2012. Building multi-layer social knowledge maps with Google Maps API. In *Workshop and Poster Proceedings of the 20th Conference on User Modeling, Adaptation, and Personalization, Montreal, Canada, July 16-20, 2012*.

John O'Donovan and Barry Smyth. 2005. Trust in recommender systems. In *Proceedings of the 2005 International Conference on Intelligent User Interfaces, January 10-13, 2005, San Diego, California, USA*, Robert St. Amant, John Riedl, and Anthony Jameson (Eds.). ACM, 167–174.

Udeep Tandukar and Julita Vassileva. 2011. Selective Propagation of Social Data in Decentralized Online Social Network. In *Proceedings of International Workshop on Adaptation in Social and Semantic Web (SAS-WEB 2011), UMAP 2011, CEUR Workshop Proceedings*, Vol. 730.

Yuan Wang, Jie Zhang, and Julita Vassileva. 2010. Personalized Recommendation of Integrated Social Data across Social Networking Sites. In *Proceedings of International Workshop on Adaptation in Social and Semantic Web (SAS-WEB 2010), UMAP 2010, CEUR Workshop Proceedings*, Vol. 590.

David Weinberger. 2008. *Everything is Miscellaneous: The Power of the New Digital Disorder*. Holt Paperbacks: New York.