National Conference on "Advanced Technologies in Computing and Networking"-ATCON-2015
Special Issue of International Journal of Electronics, Communication & Soft Computing Science and Engineering, ISSN: 2277-9477

A Survey on Semantic Mining on Social Web

Asrar Ahmad **Prathamesh V. Phadke** **Dr. V.M.Thakare** **Nupur G. Nanoti**

Abstract- **At present, the research community recognizes a complementary relationship between the semantic and the social web. The merging of these web instances could play an essential role in different knowledge domains. In this study, the authors promote a social–semantic web paradigm using software engineering as the knowledge domain specifically. Architecture is proposed for enhanced resources search, combining the strengths of the social (social annotations) and semantic (semantic metadata) technologies, which has been designed considering the search style of the information seekers.**

Index Term- **Semantic web, social web, tagging system, semantic metadata, search optimization.**

I. INTRODUCTION

Most distance and online educators would agree that Web-based teaching necessarily involves four main steps. First, educators prepare learning content in the form of reusable *learning objects*. Next, they organize these objects in accordance with a chosen instructional approach to provide a cohesive, Web-based course that includes various assessment mechanisms. As students access the course, educators receive statistics about students' interactions, although usually only in a coarse-grained tabular form.

Finally, when the course ends, educators can modify it based on the observed student interactions. Currently, learning content management systems (LCMSs) largely support these activities.

A significant issue with this workflow is that the educator is an abstract role; often many kinds of people are involved, including subject-matter experts, content authors, instructional designers, and teachers. No individual actor is fully involved in every step of the process. In many cases, subject-matter experts work with content authors (who have specific technology skills in Web content creation) to develop learning objects; instructional designers sequence, set learning outcomes, and package the content; and the teacher delivers the course to students, offering some mentorship or help. The last step in the workflow and the one we're most interested in supporting is content modification, which typically works best when all educational actors are involved, although this task often falls to only one person for cost reasons **[1].**

A huge amount of data and metadata emerges from Web 2.0 applications which have transformed the Web to a mass social interaction and collaboration medium. Collaborative Tagging Systems is a typical, popular and promising Web 2.0 application and despite its adoption it faces some serious limitations that restrict their usability. These limitations (no structure on tags, tags validation, spamming and redundancy) are more evident in the case of multimedia content due to its challenging automatic annotation and retrieval requirements.

"Web 2.0" term is used to describe a group of technologies and web frameworks in which collaborative methods of information creation and organization are applied. The key factor of its success is its constant update and continuous evolution, realized by users, who are treated as co-developers, since they provide data and metadata themselves dynamically in a continuous pace. As a result, the knowledge in these systems is built incrementally (by many users) in an evolutionary and decentralized manner, yielding in Emergent Semantics. A typical Web 2.0 application that has recently gained widespread popularity is the Collaborative Tagging Environments, where users label digital resources, by using freely chosen textual descriptions (tags). The simplicity and the user-centered design of those systems have encouraged many web users to annotate their data by using tags which have proven to be very advantageous, especially, for search and retrieval in non-textual Web sources, such as photos, videos, etc. As a result, rapidly and in a short time, a huge amount of data and metadata became available in the Web. While social data (i.e. folksonomies) seem very promising sources of information, they have some serious limitations that restrict their usability. First of all, users are prone to make mistakes and they often suggest invalid metadata (tag spamming). Additionally, the lack of (hierarchical) structure of information results in *tag* ambiguity (a tag may have many senses), tag synonymy (two different tags may have the same meaning) and granularity variation (users do not use the same description level, when they refer to a concept), which restricts the retrieval ability of such systems.

People tend to use redundant tags, in order to tackle low recall, but this worsens the precision of the system, as it causes many irrelevant objects to be fetched to the users. **[2].**

However, the semantic web 'still lacks approachable interfaces allowing contributions from non-specialists' hence, the need for applying social web technologies in software projects becomes evident. In this way, regular users can contribute content, generating a 'collaborative and innovating' ecosystem.

One of the software Engineering (SE) activities that requires a social–semantic approach is the management of software requirements. From this point of view, the software requirements and specification documents are created in a collaborative way by the project stakeholders **[3].**

II. BACKGROUND

State-of-the-art use of semantic and social technologies in software engineering is one of the reasons for which social web or Web 2.0 became so popular is that it is focused on contents, relations and knowledge and not precisely on technology. The tools and services based on social philosophy have been used by a large number of normal web users and domain experts, to generate their own resources, collaborate on a product development or to tag and classify their web resources and later share with other people. On the other hand, the semantic web, through its technologies allows one to structure and semantically enrich the content – to define and use common vocabulary, to generate new knowledge or to solve word–meaning problems. These and other capabilities made it possible to improve or automate certain tasks that a human agent would not be able to perform. The semantic web has an inner capability for processing a large amount of information.

In Social-semantic technologies driven to software engineering the relation between semantic web and social web or Web 2.0 is that, 'these two approaches are complementary and that each field can and must draw from the other's strengths **[3].**

Many earlier research efforts have focused on exploiting knowledge stored and often "hidden" in folksonomies and they have dealt with the following topics:

i) Clustering techniques based only on tagging information and tag co-occurrence to derive semantically-related groups of tags and resources.

ii) Ontology-driven tagging organization and mining, by combining Web 2.0 and Semantic Web ideas. Such efforts include building of an ontology that formalizes the activity of tagging, so as to enable the exchange, comparison and reasoning over the tag data acquired from varied tagging applications, and a number of approaches which have focused mainly on the exploration of the tag space and the detection of emergent relations in social data, which will be exploited for ontology building and/or evolution.

iii) Content-based analysis on tagging-related sources, such as a method which is introduced for exploiting both tags and visual features (in a supplementary manner) for browsing and retrieving of semantically related images **[2].**

In Initial Text Mining and preprocessing stages

A Rapidminer (www.rapidminer.com) data collection and processing tree was developed to look for the most common positive and negative words, and their term-frequency-inverse document frequency (TF-IDF) scores within each post.

Also in Cataloging and Tagging Text Data,

Text data contains the highest TF-IDF scores were tagged with a modified NLTK toolkit (http://www.nltk.org/) using MATLAB to ensure that they reflected the negativity of a negative word and the positivity of a positive word in context. This approach was used before using negative tags on positive words **[5].**

III. Current Technologies

i) Social Tagging Systems Basics

A Social Tagging System, STS, is a web-based application, where users assign tags (i.e. arbitrary textual descriptions) to digital resources. The digital resources are either uploaded by users or, are, already, available in the web. The users are either "isolated" or, more commonly, members of web communities (i.e. social networks) and their main motivation (for tagging) is information organization and sharing. The tagging activity inside an STS shows the way users categorize resources and it is known as its folksonomy. Figure 1 depicts the basic structure of a web-based STS. **[2].**

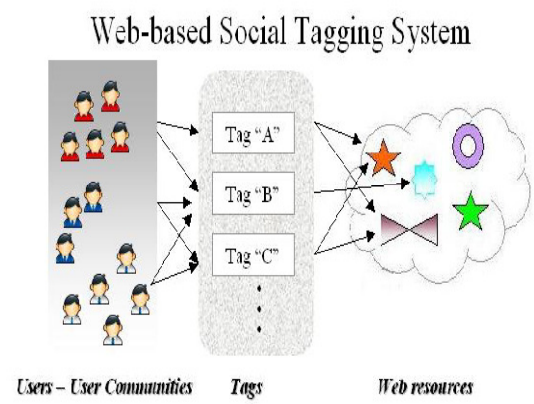

Figure 1. A web-based social tagging system

ii) Social-semantic technologies driven to software engineering

On the direction semantic-to-social, the contribution is given through the semantic enrichment of tags or content created by users, through social tools. It may be possible that, where semantic data are associated to web pages links, based on domain ontology and the user gives opinions about the link content.

The other point of view is social-to-semantic direction; most of the efforts are focused on the usage of social annotations and folksonomies to create and to populate ontologies.

After an analysis of the application of semantic web technologies in software engineering, probably the more common proposals are the ones that use ontologies.

There are some studies about organising the knowledge domain in SE and allowing both tools and developers to share information and work cooperatively **[3]**.

IV. Analysis

The search prototype has been developed with a layered architecture using an evolving development cycle, based on prototypes. The three layers implemented can be seen in below figure 2. The semantic component is represented by the OWL(Online Writing Lab) file and RDF(Resource Description framework) store, which are used to codify knowledge in the requirements domain and are available on a publicly accessible server. Figure 2 also shows the technologies used in the development:

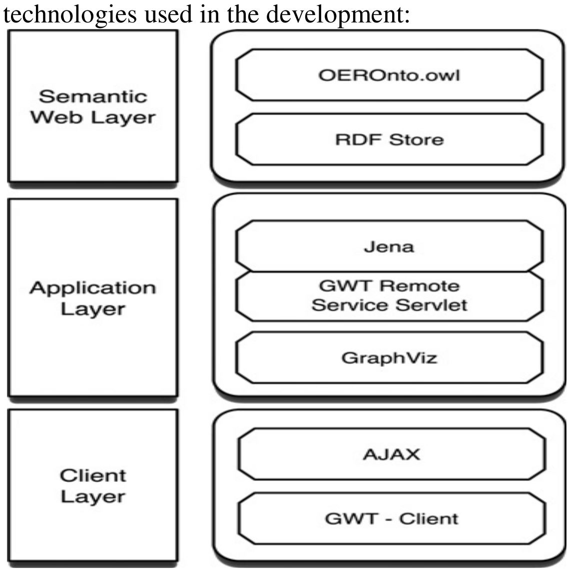

Figure 2: Application layers of the search engine prototype

CONCLUSION

In this we have studied various existing technology on semantic mining on social web. We also take a survey on various technologies.

REFERENCES

1. Jelena Jovanovic and Vladan Devedzˇic, "Using Semantic Web Technologies to Analyze Learning Content,", IEEE Computer Society pp. 45–53, Sep-2007.
2. Eirini Giannakidou, Ioannis Kompatsiaris, Athena Vakali, "SEMSOC: SEMantic, SOcial and Content-based Clustering in Multimedia Collaborative Tagging Systems", The IEEE International Conference on Semantic Computing, pp. 128-135, 2008.
3. J. Chicaiza1 J. Lo´pez, N. Piedra, O. Martı´nez, E. Tovar, "Usage of social and semantic web technologies to design a searching architecture for software requirement artifacts", IET Softw., Vol. 4, Iss. 6, pp. 407–417, 2010.
4. Altug Akay, Andrei Dragomir, Bj¨orn- Erik Erlandsson, "Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, NO. 1, JANUARY 2015.
5. Tao Jiang, Ah-Hwee Tan and Ke Wang,"Mining Generalized Associations of Semantic Relations from Textual Web Content", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 19, NO. 2, FEBRUARY 2007