

Privacy Restrictions in Social-Semantic Web Research

Introduction: The Semantic Web

The young field of Semantic Web research has been one of the results of a number of different subfields of Computer Science being exposed to the challenges of the Web. Database management systems, computational linguistics, knowledge representation, knowledge-based systems and service-oriented computing are just some of the subfields that are all making contributions to the Semantic vision of the Web, which is a more conscious effort on behalf of the World Wide Web Consortium (W3C) to make the Web 'friendlier' to machines as much as it is friendly to human. While at the moment most of the content in the online world is only accessible to human readers, the Semantic Web would provide additional layers of Web architecture for describing content using shared vocabularies called *ontologies*. Such a machine-understandable Web would make it possible to query and reason over integrated data-sets consisting of separate pieces of information that were never intended to be linked together, yet they can be fruitfully combined and integrated, even by a third party who neither wrote nor owns any of the original pieces of such data.

The Semantic Solution to Web Limitations

According to most Semantic Web researchers, the field is very promising to solve the many limitations of the current Web, especially in the field of Information Retrieval. Many of these researchers argue that even the current advanced Web search engines still suffer from the lack 'understanding' the meanings of natural-language human queries. For examples, the query "*Show me photos of Paris*" could result in retrieving many photos of "Paris Hilton" whereas the intention was to retrieve photos for the city of Paris. Another query "*Find new music that I like*" is at an even higher level of difficulty so much so that most of us wouldn't even think of posing it to a search engine. Studies have diagnosed the current problem of the Web as a *knowledge gap*: what the underlying Web application understands and able to work with is much more limited than the knowledge of the user. Here, Semantic Web research can play an important role to investigate the efficiency of applying *knowledge-feeding* technologies in order to fill this knowledge gap between human and machine, that is, providing knowledge in forms that computers can readily process and reason with. This knowledge can either be information that is already described in the content of the Web pages but difficult to extract, or additional background knowledge that can help to answer natural language queries in some way.

The Emergence of the Social Web

The Web used to be a read-only medium for a majority of users. The web of the 90s was much like the combination of a phone book and the yellow pages. And despite the connecting power of hyperlinks, it instilled little sense of community among its users. This passive attitude toward the Web was broken by a series of changes in usage patterns as well as technology. The result is what now is referred to as Web 2.0. This *new version* for the Web led to an entirely different role of the online world as a platform for intense communication and social interaction. The new introduced Social Web is not only a collection of networks which are maintained and extended online, but it is also a collection of highly-popular global meeting places for dealing with major life situations such as getting

support in case of a major illness, looking for jobs, informing about major investments etc. Social media on Web 2.0 has taken so many different forms, from blogs, wikis, friendship networks, multimedia-sharing environments to other forms of web-based communication and collaboration.

Web 2.0 + Semantic Web = The newest Web 3.0

In practice, the ideas of Web 2.0 and the Semantic Web are not exclusive alternatives. While Web 2.0 mostly effects how users use the Semantic Web opens new technological opportunities for web developers in combining data and services from different sources. It is not surprising that many researchers see very promising opportunities that arise by the combination of ideas from these two major Web developments. Firstly, Semantic Web documents written in RDF or OWL can greatly utilize the structured metadata that users in Web 2.0 are always providing through their interaction and collaboration in social media. Secondly, Semantic Web technology can help in socially matching users with similar interests as well as matching users with relevant Web 2.0 content. Thirdly, the Semantic Web can be a standard infrastructure for building creative combinations of data and services. This Semantic infrastructure can come in many forms, including standard formats for exchanging data and schema information, support for data integration, along with standard query languages and protocols for querying remote data sources provide a platform for the easy development of *mashups*, which are websites based on combinations of data and services provided by third party vendors.

Privacy Restrictions as a Challenge to Web 3.0 Research

In spite of all the promises of the research related to the development of this new promised Social Semantic Web, privacy concerns of Web 2.0 applications, stemming from the use of social networks and the dissemination of social media data, have put a large challenge to the advance of the Web 3.0 research initiatives. Majorly, there are three types of social media privacy concerns: 1) data access-related issues, where both Social-Semantic media researchers and developers are allowed access other users' private data. 2) data publication-related issues, or how the data that is collected from Social-Semantic media is going to be analyzed, evaluated and reviewed in a researching environment with no data privacy violation. And 3) metadata access-related issues, or how the huge amount of user-generated structured data used in collaborative categorization and content management in social media applications, or what is called *Folksonomy*, is going to be utilized by Semantic Web researchers, developers and services to realize the promised Social-Semantic environment.

In summary, since the Social Web is an environment in which millions of users interact with one another on a daily basis, both the dynamic nature of the data as well as the privacy expectations of users need to be taken into account in research studies aiming at designing any novel methodologies that combine the two technologies, Semantic Web and Social Web, to achieve the promised new user and machine understandable Web 3.0

Ahmad Ammari, A.N.Ammari@Bradford.ac.uk
PhD Student in Web and Social Media Mining
School of Computing, Informatics & Media
University of Bradford
United Kingdom