Editorial

Semantic Web and Web 2.0

The cluster of technologies and design patterns known as Web 2.0 has now emerged as the leading contender for “the next evolution of the Web.” Researchers, developers, and venture capitalists are all flocking toward the banner of Web 2.0, based on its promise of massively increased sharing and participation among web users. At the same time, the technologies of the Semantic Web have been quietly maturing and spreading, and now may provide a clear way to apply a basic level of formal semantics to the infrastructure and pages of the web. In distinctive ways, both the patterns of Web 2.0 and the technologies of the Semantic Web address the fundamental concept of socially shared meaning.

In the past year, it has become clear that these two techniques are natural complements of each other. Further, the combination of Semantic Web technologies with Web 2.0 application design patterns should give rise to a new and exciting offspring: the social-semantic web, also referred to as Web 3.0.

In a social-semantic web, certain formally representable parts of human meaning can be encoded and reasoned about via the tools of the semantic web, but can also be curated and maintained via the social, community-oriented techniques of Web 2.0. The results of this combination would be powerful indeed. The social-semantic web promises that the subtle variations in meaning that characterize different human communities can be managed via the user-friendly collaboration mechanisms of Web 2.0, while still maintaining the expressive precision and reasoning power of the semantic web. This would make possible a new class of applications that could leverage the semantic relations that exist between certain kinds of web-accessible data to automatically locate and fuse information, perform basic reasoning, and pivot and transform representations to meet a wide variety of user needs. This special issue of the Journal of Web Semantics is dedicated to the emerging social-semantic web.

Given the early, experimental stage of the field, the impressive flowering of ideas so far, and the resulting fragmentation of the works we have seen, we faced an interesting editorial challenge in creating this special issue. We elected to depart from the customary Journal of Web Semantics format of three-to-four deep and lengthy research papers, in favor of a larger number of shorter and hopefully more thought-provoking pieces. We did not propose particular topics of interest in our call for contributions, nor did we specify a desired format for contributions. Instead, we posed a set of questions to motivate our contributors in thinking about the promises and pitfalls of combining Semantic Web and Web 2.0 technologies. In particular, we requested that our authors consider two different types of concrete challenges to the vision of the social-semantic web:

- Semantic Web technologies were initially designed before the rise of the Web 2.0 methods for managing large-scale socially contributed content. What lessons should Semantic Web technologies draw from the popularity of tag systems, social networks, mashups, and other Web 2.0 techniques? Does the success of the social, user-oriented contribution models of Web 2.0 impact the way that Semantic Web data should be created, deployed, exploited, managed, and shared? In short, should particular aspects of the Semantic Web be reconsidered in order to maximally leverage the strengths of Web 2.0?
- Web 2.0 applications always depend on some type of shared semantics—for example, between the software components of a mashup, or within the user community that contributes to a particular tagging system. Can the relative precision and rigor of Semantic Web representations and inference add significant value to Web 2.0 applications? Are the various Semantic Web “bridge” technologies (like RDFa, GRDDL, and SPARQL) adequate to the semantic demands of Web 2.0 applications? Can Semantic Web techniques be used to substantially enhance perceived user value in Web 2.0 social networks, for example by linking across communities?

We received contributions that span a range from discussion papers to advances in research technology to demonstrations. We selected ten papers that represent a balance in terms of these specific types of contributions to the quickly emerging research and technology landscape. They fell into three rough categories: discussion papers, technology papers, and systems papers.

1. Discussion papers

Much to our surprise, while we expected debates to emerge among the discussion papers, our authors were united in their belief that the social-semantic web is likely to become a major application area of Semantic Web technology. As Tom Gruber explains in our first paper, Web 2.0 brings massive amounts
of user-generated content and a proven recipe of humans and machines working together in synergy. However, he argues that Web 2.0 stops at what he calls “collected intelligence,” the mere pooling of contributions from individual users. Adding Semantic Web technology into the mix would allow to reach true collective intelligence, which he defines as the higher level understanding constituted by answers, discoveries and other results not found in the original contributions. Gruber suggests that the Semantic Web technology will act in two principal ways: by adding structure to user data and by connecting the existing silos of data that characterize the Web 2.0 landscape. What makes Gruber’s argumentation particularly strong is the wealth of examples drawn from his experiences at RealTravel, a Web 2.0-style travel site employing lightweight semantic technologies. These examples validate the old semantic web cliché that a little semantics goes a long way, and at a fraction of the cost.

In the next paper, Hendler and Golbeck also see the man–machine synergy in the marriage of the Semantic Web and Web 2.0. They cast their argument in terms of networks: while Web 2.0 is a rich network connecting users, the Semantic Web is a network connecting data through semantic relations. They argue that the power of linking in networks increases super-linearly in networks in general, and that these two networks are no exception. Interconnecting these networks is thus certain to bring value as new connections become possible within them. As an example, they discuss the existing poor connectivity of collaborative tagging systems as a semantic space (“tags do not create much of a link space”), going almost as far as claiming that most of tagging is selfish (serving only individuals or small groups). These spaces they argue would benefit from minimal semantics in the form of stable URIs for concepts and limited core ontologies. Hendler and Golbeck leave the problem of how to graft ontologies on tag spaces as future work: they reject bottom-up methods or “emergent semantics” as a method that can only reliably determine the semantics of the most common terms (which are not necessarily the ones users may care about), but they also criticize expert-system style approaches or “emergent semantics” as a method that can only reliably determine the semantics of the most common terms (which are not necessarily the ones users may care about), but they also criticize expert-system style approaches to the Semantic Web. They hint that the solution is partly a user-interface challenge, e.g. capturing semantics at the point of metadata creation. Similarly, they see a great deal of work to be done in exploiting the new linkages.

Bojars, Breslin and Finn would certainly agree with Hendler and Golbeck on the power of linking. Their attention is focused on linking Web 2.0-style social spaces, especially the many forums for discussion that exist on the Web. Their criticism of the current state-of-the-art, in particular microformats, is exactly the lack of possibilities for interlinking and their limited extensibility. They describe the work they have carried out in the past years in DERI’s SIOC initiative, which provides a good example of reusing existing core vocabularies in creating an overlay network of semantics. Their work is both technically advanced and pragmatically feasible. In filling SIOC’s semantic network with live data, the authors also do not mind getting their hands dirty: in fact, they have been doing a outstanding work in promoting SIOC in the semantics community, helping others implement it, and stepping in themselves to add SIOC support to existing social platforms in order to capture semantic annotations at the source.

2. Technology Papers

While Bojars et al. are mostly concerned with the external representation of contextual metadata, our first more focused technology paper, Iyad Rahwan’s “Mass Argumentation and the Semantic Web,” goes deeper in exploring how to impose a layer of semantics on discussions themselves. More precisely, Rahwan describes a formal representation of generic argument structures. Although primitive ways of connecting arguments in the social space already exist (e.g. through the use of tags, links and trackbacks as references), Rahwan’s proposal augments these with several much more formal concepts of argumentation theory and rhetoric, such as the different roles that facts can play in debate. To make this work, Rahwan needs the human element in Web 2.0, because extracting argument structures fully automatically is one of the hardest tasks in natural language processing. Participants in complex intellectual debate are often willing to invest the extra time to express their arguments in a careful and structured way. The idea of a rich, ontology-based argument representation certainly holds the potential for interesting applications, such as multi-perspective question answering or argumentative semantic blogging where participants can link precisely to particular aspects of a discussion and provide supporting evidence, endorsements or rebuttals (which can be later retrieved and aggregated with the contributions of other users).

In one of the most formal papers in our collection, Jäschke et al. propose a method for ontology learning in folksonomies, in particular based on frequently co-occurring sets of users, tags and resources. They view ontologies as social constructs, where a concept can be described by a set of tags pertaining to a set of users and used to characterize certain kinds of resources. Adapting results from Formal Concept Analysis, they provide an efficient method of learning the most salient concepts in a folksonomy. Interpreting or even visualizing the results is non-trivial, but their method could lead us to better ways of navigating and comprehending the complex and noisy folksonomies that many Web 2.0 websites are built upon.

Ben Adida’s paper provides a bridge between Web 2.0-style term-based knowledge representation and the more logically demanding formalisms of the Semantic Web. Translation is a theme of many of the successful technologies we have seen in recent months. Adida’s proposal for translating between the emerging worlds of microformats and RDFa is called hGRDDL (lacking vowels, best pronounced as ‘h-griddle’). His idea is to apply GRDDL-style transformations to convert microformat-enabled web pages into semantically equivalent pages where the microformats have been replaced with RDFa markup, while ensuring that the transformation preserves the visual presentation of the page. hGRDDL is a practical proposal that prepares us for the a world where RDFa and microformats will exist side-by-side.

Battle and Benson’s paper bridges the Semantic Web and Web 2.0 at the level of access to data and services. They have
a straightforward, practical view on developments in Web Services. In their paper, they develop two extremely interesting infrastructure elements that will enable Semantic Web applications on top of existing or new services implemented according to the popular REST pattern. The first element, the Semantic Bridge for Web Services, enables Semantic Web developers to execute SPARQL queries against existing web services. It does so by wrapping the WSDL and OWL-S or WADL description of Web service operations, and translating the results returned into the SPARQL query result format. The second element, Semantic REST, is a protocol intended for new Semantic Web applications that need to support REST-style access to query and data manipulation functionality. Battle and Benson show how to extend the capabilities of the existing SPARQL protocol by supporting updates and deletes. The two contributions of the paper are demonstrated by a single example of distributed querying where a SPARQL query is executed against a set of web services, some wrapped using the Semantic Bridge and some supporting the Semantic REST paradigm directly.

3. System papers

The next two papers, by Ankolekar et al. and Heath and Motta, open our collection of system-oriented articles. They have been placed together because they share a common use case of semantic reviewing. Reviewing is a natural fit for work on the boundary of the Semantic Web and Web 2.0, as it requires handling both structured and unstructured information (minimally: the reviews and data about the objects being reviewed) and it also has a strong social dimension. In parallel, both papers start by taking stock of the virtues of Semantic Web and Web 2.0 technology, noting the many complementarities. Ankolekar et al. considers semantic reviewing as an example of how enhanced authoring tools will be able to assist both publishers and readers in producing and consuming content more efficiently. They describe a semantic blogging platform that can be extended by a movie widget when its owner would like to create a review about a movie the she has just seen. The widget helps to create and structure the content, which is subsequently mixed with external data such as showtimes and exposed to the world in both human and machine readable ways.

Reviewing is also the use case targeted by Revyu.com, the web application described by Heath and Motta. In fact, the Revyu system, which just won the ISWC 2007 Semantic Web Challenge, could be considered as an implementation of the Ankolekar’s vision. Revyu also places the emphasis on ease of user interaction, mixing owned content with external data, and in general being a good Semantic Web citizen by opening up its data and services. The difference from the proposal of Ankolekar et al. is that Heath and Motta envision authoring and consuming to be centralized and contextualized to a significant extent.

Finally, our collection of papers on Semantic Web and Web 2.0 would not be complete without discussing semantic wikis. Semantic wikis have been an important and active testbed for ideas on the convergence between Web 2.0 and the Semantic Web. Wiki-oriented ideas for semantics have evolved along two distinct dimensions. First, wikis can provide the collaboration framework and tools that have usually been missing from traditional, single user ontology editors for the Semantic Web. Adding wiki-style collaboration features to ontology editing tools is thus a natural next step in the evolution of ontology editors. In fact, as shown by recent examples such as Freebase, ontology editing and data acquisition can be carried out in parallel and the approach can scale to large numbers of users. Second, specialized extensions for handling semantic and structured information have been added to existing Wiki platforms, such as the popular MediaWiki software that powers Wikipedia. These new extended wiki platforms provide much better control over data, as well as the tantalizing possibility of generating new pages and displays of aggregated content by performing querying and reasoning over the collected data. The paper of Buffa et al. provides a brief, but complete survey of Semantic Wikis. The authors also present their own Semantic Wiki, called SweetWiki. SweetWiki combines ideas from folksonomy and ontology-style representation, sports a rich user interface, and includes many of the latest developments in extensibility, exchange formats, querying, and reasoning.

4. Conclusion

These ten papers provide a comprehensive overview of the best current thinking on the social-semantic web, in its vision, technology, and systems manifestations. We are excited by the prospects of this new evolution of the web, and we look forward to its further development.

Mark Greaves*
Peter Mika
Vulcan Inc., 505 Fifth Avenue South, Suite 900, Seattle, WA 98104, United States

*Corresponding author. Tel.: +1 206 342 2276; fax: +1 206 342 3276.
E-mail address: markg@vulcan.com (M. Greaves)