Emerging Sciences of the Internet:
Some New Opportunities
(Extended Abstract)

Ron Brachman

Yahoo! Research, New York, NY 10011, USA
`rjb@yahoo-inc.com`

Semantic Web technologies have started to make a difference in enterprise settings and have begun to creep into use in limited parts of the World Wide Web. As is common in overview articles, it is easy to imagine scenarios in which the Semantic Web could provide important infrastructure for activities across the broader Internet. Many of these seem to be focused on improvements to what is essentially a search function (e.g., "list the prices of flat screen HDTVs larger than 40 inches with 1080p resolution at shops in the nearest town that are open until 8pm on Tuesday evenings" [`http://en.wikipedia.org/wiki/Semantic_Web`]), and such capabilities will surely be of use to future Internet users. However, if one looks closely at the research agendas of some of the largest Internet companies, it is not clear that the staples of SW thinking will intersect the most important paths of the major broad-spectrum service providers. Some of the emerging trends in the research labs of key industry players indicate that SW goals generally taken for granted may be less central than envisioned and that the biggest opportunities may come from some less obvious directions. Given the level of investment and the global reach of big players like Yahoo! and Google, it would pay us to look more closely at some of their fundamental investigations.

While not all companies see the future in the same way, there are some trends and priorities that are instructive to contemplate. While Web search will continue to play a large role, and services composed from piece-parts offered by multiple vendors will be important, some relatively novel ideas may come to dominate. In one view, there will emerge a set of new sciences that are fundamental to future generations of Internet businesses. By understanding the imperatives of those new areas of thought, we may be able to get a better assessment of the true ultimate impact of SW technologies on the broader Internet. At the very least, framing future Semantic Web directions and examples in terms more aligned with some of these ideas may encourage more attention from some of the major Internet companies, which to date have shown only lukewarm interest.

Search will continue to play a significant role for users of the Internet. While from a user's perspective the most essential thing is the ability to find a website or document or element of data relevant to a task at hand, it is critical to the service provider that, in addition to producing relevant results, it be able to expose the user to advertisements well-suited to his or her immediate and longer-term needs and interests. While not restricted to search settings, high-quality matching of ads is critical to the continued success and growth of the major

Internet search and content providers. In the long run, a better understanding of users in general and customers in specific is essential to providing a better experience and providing better opportunities through advertising. This is an area where a huge amount of investment is being made. Does it provide any interesting opportunities for SW technology?

One interesting element here is that with the scale of the Web, companies are increasing their use of machine learning technologies to improve the performance of their search engines and other computing systems, since it is simply impossible to process information at the scale and speed of the Web manually. Machine learning technology, based ultimately on richer knowledge representation technologies, will play an extremely prominent role in the infrastructure that makes the Web successful in the future. Will this development provide interesting opportunities for the Semantic Web community? Can it avoid the relative lack of mutual interest that has plagued the KR and ML communities in the recent past?

Another emerging element in what we might call a new Science of Search is a social one. Products like Yahoo! Answers add the human element back into the process of finding relevant information. A similar human element pervades tagging-based services like Flickr. It's not clear what the success of large-scale popular web services based on ad hoc human tagging says about the future of the Semantic Web - are the two incompatible? Will ontologies matter or do folksonomies rule? Others have begun to address the substantial differences between the social Web world and the Semantic Web world. While sometimes portrayed as diametrically opposed, the sides may benefit from each other if we look more deeply. My intuition is that there is room for synergy, and it would behoove us to investigate.

The social element is evident in many areas other than Search. Community-oriented services are already extremely popular (witness the scale of MySpace and FaceBook, and even older services like Yahoo! Groups). On the face of it, this direction would seem not to hold much promise for SW thinking. But if one looks at some of the underlying infrastructure needed to make new community systems succeed, one sees opportunities for robust search and information extraction and organization technologies, among others.

Another exciting and somewhat unexpected set of developments in new Internet science involves collaboration between Economics and Computer Science. Driven in part by underlying elements of advertising auctions that account for extraordinary revenue streams at Google, Yahoo!, and other companies, technical approaches to constructing novel market mechanisms, deploying them, and understanding their practical ramifications are becoming critical. Prediction markets and other novel mechanisms based on large-scale interactions among masses of people are creating never before seen opportunities because of the Internet, and substantial resources are going into the understanding of fundamental principles governing such interactions. This whole line of thinking is not yet prominent in the worldview of the Web community, but it is likely to grow in importance, and it may provide some novel opportunities for SW research.

Finally, as has been pointed out by many others, text-based documents are not the only media available on the Web, and there has clearly been astronomical growth in user-generated non-textual content over the last few years. Podcasting, Flickr, YouTube, Yahoo! Video, and other media-sharing opportunities have shown that there is extraordinary latent interest in the creation and spread of audio, image, video and other complex content. While not yet receiving as much attention, the study and understanding of how users experience media, both individually and in social settings, is likely to lead to an explosion of novel technologies and perhaps even devices for consuming and sharing media. How these new scientific directions will relate to the SW world is anyone's guess, but it makes sense to take a close look to see what opportunities they present.

It is clear that some very large players on the Internet are making substantial - in some cases huge - investments in novel forms of search and information navigation, social and community systems, economic mechanisms for auctions and advertising, and media experiences. With the scale of public interest and the level of economic investment we are now seeing, it is probable that these directions will have a dramatic impact on the evolution of the Internet. While many of the arenas traditionally targeted by the Semantic Web community will continue to matter in the next generation of the Web, it is conceivable that the really big interest, including the majority of investment and opportunity for broad influence on the world, will go elsewhere. It would be nice if the Semantic Web were not left behind. A quick look indicates that there could be some interesting opportunities, and it would appear to be very much worth the community's while to stop to consider the emerging sciences and their implications for future generations of the Internet.