Preface: Introduction to the Special Issue

Recent advances in technology are challenging the traditional business model of publishers, as individuals are now able to inexpensively publish and distribute content themselves. Yet these same advances do offer publishers a rich opportunity to develop novel offerings for scientists. In response to the wealth of opportunities, the idea of the Elsevier Grand Challenge on Knowledge Enhancement in the Life Sciences was conceived by Gully Burns at the University of Southern California and Anita de Waard and colleagues at Elsevier, and formulated as a contest to elicit innovative ways in which scientific information might be communicated and used. The contest invited members of the scientific community, primarily in life sciences, semantic web, and natural language processing, to describe and prototype tools to improve the interpretation and identification of meaning in journals and databases relating to the life sciences. The objective was to generate useful new ideas that could have a widespread impact on scientific publishing in general.

Specifically sought were new ways to:
1) improve the process/methods/results of creating, reviewing and editing scientific content,
2) interpret, visualize or connect the knowledge more effectively, and/or
3) provide tools/ideas for measuring the impact of these improvements.

Over 70 ideas were submitted to the Grand Challenge. After a rigorous review by an international panel of judges that included leading scientists from academia and industry as well as representatives from Elsevier, 9 semi-finalists were invited to present their ideas in person to the panel in December 2008. Four of the ideas were selected for the final round in April 2009. The first and second prize winners received awards of $35,000 and $15,000 respectively.

This special issue of the Journal of Web Semantics contains the 9 semifinalists’ papers, selected according to the judging criteria that focused on the overall vision, novelty, usability, and technical feasibility of the work.

Thoma et al. describe work that is being undertaken at the U.S. National Library of Medicine to create an Interactive Publication. This work describes the concept, and outlines the benefits and challenges of self-contained, multimedia-rich interactive publications. They outline tools that they have created for authoring and visualizing such publications.

Ahmed et al. describe the SLIF project that combines text mining of figure captions and image processing to extract structured information from biomedical literature. The results can either be queried online or accessed through web services. A usability study is presented that compares the results of SLIF to those of typical search engines.

The paper by Garcia et al. describes the concept behind the Living Document and provides details regarding its implementation. A Living Document enables concepts that are described within a paper to be interconnected with other sources of information regarding those concepts using the web. The technical approaches utilize tagging and semantic web.

In the article by Greenacre and Hastie, a variety of tools and prototypes are presented that have been developed to add motion to scientific graphics, allowing high-dimensional data to be visualized dynamically thus aiding interpretation and understanding.
Newman et al. introduce the use of topic maps for visualizing the semantic content of collections of documents, which has applicability in aiding scientists in understanding the span of their search results. Techniques are described for assessing the validity and accuracy of topic maps.

In the tradition of the semantic web, Novacek et al. have built a prototype of a tool called CORAAL that combines asserted publication metadata with knowledge extracted from the text in the form of triples. The results are then made available through a search and browse interface. This work won second prize in the Challenge.

O’Donoghue et al. introduce a tool called Reflect, which augments web browsing by allowing end users to view semantic annotations about small molecules, genes and proteins. The paper describes the tagging process, and outlines how additional entity types can be added to Reflect. This work won first prize in the Challenge.

The paper by Page describes a proof of concept for interlinking structured content within papers to data within external databases. A web interface was designed to enable the useful interrogation of the enriched data environment.

Wan et al. introduce the Citation-Sensitive In-Browser Summariser (CSBIS), a natural language tool designed to aid scientists in the task of browsing the literature. CSBIS presents users with both a generic and tailored preview about a citation, dynamically generated with respect to the context in which it is referenced in a paper.

The number and quality of submissions to the Elsevier Grand Challenge was very high. We would like to congratulate all the authors of papers in this special issue for having reached the final rounds.

We would like to also acknowledge the panel of judges who brought a diversity of knowledge to the competition and who were very dedicated to thoroughly reviewing the large number of submissions:

- Eduard Hovy (Chair), Director of the Natural Language Group, Information Sciences Institute, University of Southern California, Los Angeles, CA, U.S.
- Bernard Aleva, Global Director of Publishing, Life Sciences, Elsevier, Amsterd am, The Netherlands
- Emily Marcus, Editor of Cell, Cell Press, Cambridge, MA, U.S.
- Susie Stephens, Principal Research Scientist, Eli Lilly, Boston, MA, U.S.
- Alfonso Valencia, Protein Design Group, National Centre for Biotechnology, CNB-CSIC, Madrid, Spain

We would like to highlight the far-sightedness manifested by Elsevier in setting up this contest. It demonstrated that their keen interest in cutting edge technology for enabling scientists to gain maximum value from literature.
The diversity of ideas submitted to the Challenge, ranging from visionary to already in part implemented, was truly exciting. There was also great diversity in subject matter ranging from visualization techniques, to text mining, and greater integration of literature with other data sources. The response illustrated clearly the potential for much richer and in-depth interactions between the producers and consumers of scientific content. We hope that the Challenge, this special issue, and the further ongoing discussions that have been initiated by these events, lead to a rethinking of the nature of scientific publishing.