The Semantic Web challenge was created to show off the benefits of semantic technologies by developing integrated, easy to use applications that could provide new levels of Web functionality for end users on the Web or within enterprise settings. Applications submitted to the competition demonstrate clear practical value that goes above and beyond what is possible with conventional web technologies alone.

In 2008, a second track was added to the challenge to push research in the scaling of Semantic Web technologies. The Billion Triples Track required the participants to make use of a dataset — consisting of about a billion triples — provided by the organizers. The goal of this challenge was not to be a benchmarking effort between triple stores, but rather to demonstrate applications that could scale to a Web size using realistic Web-quality noisy data.

In this issue we present papers by five of the six winning systems: the three winners of the regular challenge and two of the winners of the Billion Triple Challenge.

The first place winner was Paggr, a system for creating Web widgets for linked data queries. Essentially, Paggr is an online tool that provides novel ways to manage and repurpose information on the web when that info is available in Semantic Web formats. Paggr allowed access to a number of different datasets and query types, making it a very flexible way to interact with linked data systems.

The second place winner was DBpedia Mobile, an application for browsing DBpedia information via a mashup with geonames and other location data presented on a map (and available via a mobile client). DBpedia, as most readers know, is a central node in the growing linked data cloud, and this application showed the power of using this Wikipedia-based dataset in conjunction with various geospatial data.

Third place was taken by HealthFinland, a semantic web-based portal for managing health information. Using Semantic technologies, HealthFinland is able to customize health information found on the Web to the needs of a particular user. Thus, a younger person looking for information on, say, diet, would find different guidance than might a more senior citizen. A number of health ontologies are used in producing the semantics used in the system, and interfaces are provided both to the human user, via a Web-based frontend, and to a machine, via a widget-based service interface.

For the Billion Triples Challenge, the winning application was SemaPlorer, a system allowed users to interactively explore and visualize a very large subset of the billion triple data semantic data set in real-time. Its use case was to allow a user to learn about a city, tourist area, or other area of interest. By visualizing the data using a map, media, and different context views, the system goes beyond the simple storage and retrieval of large numbers of triples. SemaPlorer leveraged a number of the different semantic data sources such in the Billion Triple Dataset, including DBpedia, GeoNames, WordNet, and personal FOAF files. It also connected with a large Flickr data set converted to RDF, thus extending the triples provided to make for a map-based browser of a very large dataset.

Third place in the competition was Marvin, a system which used a novel method to generate a sound, but not necessarily complete, closure of the billion triples data. Marvin distributed the data to many processors, materialized locally, and then “reshuffled” the data among processors. In this way, the closure probabilistically approaches within epsilon of the complete system, but with very efficient use of multiple processors in a cluster or other distributed machine.

These five papers demonstrate the diversity of ways that Semantic Web data can be used, and represent some of the best applications developed in the research community.

Peter Mika
Yahoo

Jim Hendler
Rensselaer Polytechnic Institute