Viewing the Web as a Distributed Knowledge Base

Serge Abiteboul (Professor at Collège de France and Senior researcher
at Inria Saclay)

Abstract: Information of interest may be found on the Web in a variety of forms, in many systems, and with different access protocols. A typical user may have information on many devices (smartphone, laptop, TV box, etc.), many systems (mailers, blogs, Web sites, etc.), many social networks (Facebook, Picasa, etc.). This same user may have access to more information from family, friends, associations, companies, and organizations. Today, the control and management of the diversity of data and tasks in this setting are beyond the skills of casual users. Facing similar issues, companies see the cost of managing and integrating information skyrocketing.

We are interested here in the management of such data. Our focus is not on harvesting all the data of a particular user or a group of users and then managing it in a centralized manner. Instead, we are concerned with the management of Web data in place in a distributed manner, with a possibly large number of autonomous, heterogeneous systems collaborating to support certain tasks.

Our thesis is that managing the richness and diversity of user-centric data residing on the Web can be tamed using a holistic approach based on a distributed knowledge base. All Web informations are represented as logical facts, and Web data management tasks as logical rules. We discuss Webdamlog, a variant of datalog for distributed data management that we use for this purpose. The automatic reasoning povided by its inference engine, operating over the Web knowledge base, greatly benefits a variety of complex data management tasks that currently require intense work and deep expertise.