Community Systems: The World Online

Raghu Ramakrishnan

Yahoo! Research, Santa Clara, CA, USA

ramakris@yahoo-inc.com

EXTENDED ABSTRACT

Online communities are a fundamental and increasingly important part of the Web. Sites such as Y! Answers, flickr, MySpace, and YouTube are well-known examples. Further, online applications in diverse domains increasingly incorporate community tools, and the richness of the baseline community offerings is growing rapidly. We routinely see bulletin boards with presence and real-time chat, rating and reputation systems, and other such capabilities on a wide range of sites. Today, it is possible to integrate modules providing rich community features using a local database instance. Increasingly, hosted application providers will make such capabilities available, with potential advantages such as single authentication, integrated payment systems, and shared personalization.

The growth in online communities offers a number of research challenges. Social scientists seek to understand the dynamics of these communities, and the factors that lead some community sites to emerge from a pack of competing sites, and become dominant in their space. Economists study mechanism designs that provide incentives calculated to further desirable behaviour and discourage anti-social behavior. In fact, we need to understand a range of characteristics of social groups, such as reputation and trust, in the online setting.

A natural question is whether we can exploit shared community interactions to improve other Web activities, in particular, search. We call this *social search*, and there are broadly three ways to use social interactions to improve search: 1) Use shared annotations (tags, comments, ratings, etc.) as metadata to improve search result ranking; 2) Use shared activity profiles to connect users with a mutual interest in being connected, as an extension of search; and 3) Create communities of purpose that are empowered to collect and integrate repositories of data harvested by crawling the Web. We will refer to the third approach as Community Information Management (CIM).

We observe that the trend towards hosted applications extends beyond online communities. The Web has become a powerful and ubiquitous means of delivering a range of end-user (e.g., email, spreadsheets) and collaborative (e.g., IM, social question answering, sales force management) applications to hundreds of millions of users. As online application development moves in the direction of very large-scale hosted applications, the requirements on the backend change substantially.

This has created a radically different approach to developing and distributing applications, disrupting the traditional software distribution model. In turn, it has challenged us to develop new types of service-oriented software platforms, new kinds of customizable application environments, and forced us to think about massively distributed systems with novel quality of service guarantees, fail-over mechanisms, and the ability to manage massive numbers of application instances.

In this talk, I will discuss online communities, social search (in particular, CIM), and application hosting, highlighting the role of data integration and structured data retrieval. The next generation of the Web is likely to require robust database foundations (although it is unlikely to be powered by traditional DBMS software!). It offers an exciting direction for database research, with opportunities for theoretical as well as systems development, and I hope that our community will pursue some of these challenges.

ACKNOWLEDGEMENTS

The ideas discussed here are influenced by, and often are the result of, joint work with colleagues at Yahoo! Research [1] and the DBlife team [2] at UW-Madison.

REFERENCES

[1] The Yahoo! Research Team, Content, Metadata, and Behavioral Information: Directions for Yahoo! Research, *in IEEE Data Engineering Bulletin,,* December 2006.

[2] De Rose et al., DBLife: A Community Information Management Platform for the Database Research Community, *demo report in CIDR 2007*