Guest Editorial

Innovations in Web communications infrastructure

The Web dominates, and will continue to dominate, our lives in ways that were unimaginable just a few years ago. Any business, organization, group, cause or individual that seeks to reach beyond its own immediate locality recognizes the value of a Web presence. Undeniably, the Web has radically reshaped the way we communicate with the world. Among the many astonishing things about the Web is the fact that it has been achieved with the aid of what are essentially basic technologies: a client, a server, and a network, together with a simple text-based mark-up language and a basic addressing scheme to locate resources. What we have done with these basic elements is truly amazing.

This achievement does not stop us, however, from designing, developing and testing new technologies in order to add value to the Web, to enhance the way we use it or to move the Web in new directions. Our experience of the Web is still far from perfect in many ways, and our expectations of the Web have changed radically – simply as a result of its presence in our lives. In addition, the basic technology that we use to access and describe resources has changed dramatically. For example, mobile and embedded Internet-ready devices are growing in use.

The aim of this special issue is to present research that aims to investigate how ways in which our experience, usage and understanding of the Web can be enhanced. One of our primary goals has been to look at how changes in the underlying network layer can affect what features can be added to the Web: for example, a better way to find and tag Web pages. The Web is now so vast that the task of finding relevant information is difficult and time-consuming, and hence the popularity of search

engines and latterly of Web-based social networking. In addition, we are interested in understanding further the nature of the new forms of social networking built on top of the Web. Models of how the Web is evolving help us understand how people are using the Web whilst eliminating low-level detail.

We also note that the current Web is based on a static network and static resources. We can start changing these assumptions to move into a ubiquitous network environment. If we do, then rather than a client-server architecture, the environment is one of autonomous peers that collaborate in providing resources and performing services. The problems facing a distributed repository with dynamic characteristics are radically different from those of a distributed repository with an essentially static topology and static placement of resources.

Four of the six papers that are presented here are directly concerned with the development of new distributed communications architectures for the Web.

The paper by Michlmayr, Pany and Kappel examines the topic of content-based search in distributed systems, specifically peer-to-peer networks. Searching for information in distributed environments by content is hard because we are not sure where in the distributed environment we should search. While we could flood the network with a query, certain in the knowledge that it will reach every peer and eventually be matched against the appropriate indexes, this is not an efficient strategy. Network congestion is almost guaranteed under such a scheme and response times are likely to be poor since routes to indexes are not designed to be optimal. The authors' research considers how metadata, in the form of taxonomies, can be exploited to

improve the efficiency of routing compared to flat namespaces of keywords.

Vazquez and Lopez-de-Ipina's paper also looks at discovery in distributed environments, this time from the perspective of locating resources in ubiquitous or mobile ad-hoc networked environments. The additional problem posed to resource discovery by these environments is that resources and their locations can change over time, perhaps rapidly. For example, a user's PDA needs to be continuously aware of resources in its current location, as well as their network organization, as it is moved from place to place. This kind of dynamism may lead to a device being used in locations and with resources that were unexpected at the time the device was designed. In such situations, lexical approaches to resource discovery, based on text pattern mining, for example, perform poorly. Vazquez and Lopez-de-Ipina consider, therefore, the use of semantic queries to discover relevant resources. They show that, although in performance terms, lexical approaches to resource discovery outperform semantic queries because of, for example, the overhead imposed by continuously reasoning about the environment, semantic queries are more powerful because they are able to explore relationships between entities and interpret and assign meaning to information about them.

Gu, Pung and Zhang's paper addresses a related topic, that of efficient lookup of RDF-based information in dynamic environments. In terms of information retrieval in such environments, there appears to be a trade-off between efficiency of information retrieval and the ease with which information nodes can be dynamically relocated in the environment. This is illustrated, for example, by the comparison between Gnutella, which allows nodes to relocate easily but which must flood the network with queries, and DHT-based P2P network architectures such as Chord and CAN, which easily locate information within the network but struggle with node relocation. The authors address the problem posed by this trade-off by partitioning the P2P network into *semantic clusters*: data and queries that are semantically similar, according to an ontology-based schema, are mapped to the same cluster of nodes within the P2P network, with each cluster responsible for its own self-organization. Experimental results show that this approach combines search efficiency with low maintenance cost.

In the paper by Wang et al., the way in which resources are retrieved on the Web is examined, with the aim of speeding up time to access, particularly over long distances. The authors describe the Grid-Jet toolkit, which employs grid-based technologies to speed up data transfers through parallel data streams. GridJet, though similar in this regard to other browser download tools and download managers, goes further. Whereas download tools such as BitComet allow fast, parallel download from multiple sources, each source is considered separate from the others. GridJet allows a common view of the aggregated storage space, in other words, a single system image, managed by clusters of P2P-networked Web servers.

The final two papers in this special issue address issues concerning social interaction through the Web.

Benz et al. describe a means of collaboratively classifying bookmarks into a hierarchically-organized space. The Web can sometimes seem chaotic and complex, and bookmarks have traditionally been one of the ways in which users remember their browsing experience and maintain a view of what is important to them in the Web. However, the cost of maintaining and organizing bookmarks can sometimes seem to outweigh their benefits. Social bookmarking has become popular as a way of reducing the cost of this exercise, but suffers from the problem of choosing which keywords to use to *tag* pages, a problem compounded by the fact that different groups of users may choose different keyword sets that relate to their differing perspectives of the content. The authors consider, therefore, techniques of automatically organizing bookmarks using similarity metrics over user interests (as represented by previous bookmarks). These are shown to compare favourably with other bookmark classification techniques and to outperform content-based analyses.

Finally, the paper by Fenner et al. presents a stochastic model of social networks. In the last couple of years, social networks in the form of websites such as MySpace and Bebo have driven much of the growth of the Web, but social networks have a much more pervasive influence as Internet-capable devices become more prevalent and personal, allowing users to find and connect to each other in different and changing environments. Clearly, as the Internet adapts to embrace the growing mobility of its users (including machines as well as human users), our understanding of the way social networks grow and evolve will be key to the development of new protocols and architectures that allow efficient use of bandwidth and resources.

As a final note, it is interesting to conclude that the papers presented here are part of a trend that acknowledges that the Web is a far from complete product. Just as the structure of the Web is itself constantly evolving, so the nature of the Web, what it is and what we can do with it, are also evolving.

We thank the many individuals who made this special issue possible. We would like to especially thank Harry Rudin for hisadvice and support, as well as his patience. Our thanks, too, to the many researchers who submitted their work for consideration, and to the reviewers for their dedication and whose contribution helped shape this special issue.

Simon Courtenage
University of Westminster, School of Informatics,
115 New Cavendish Street, London W1W 6UW,
United Kingdom
E-mail address: courtes@wmin.ac.uk

David Lewis
Trinity College Dublin, Knowledge and Data
Engineering Group, Dublin 2, Republic of Ireland
E-mail address: david.lewis@cs.tcd.ie

Thanassis Tiropanis
Athens Information Technology Center (ATT),
Markopoulo Avenue, P.O. Box 68, 19002 Peania,
Athens, Greece
E-mail address: ttir@ait.edu.gr

Available online 31 July 2007