Ricardo Baeza-Yates

Yahoo Labs

Friday, April 10, 2015
3:00 PM
Lindley Hall, Rm. 102

Big Data or Right Data?

Abstract: Big data nowadays is a fashionable topic, independently of what people mean when they use this term. But being big is just a matter of volume, although there is no clear agreement in the size threshold. On the other hand, it is easy to capture large amounts of data using a brute force approach. So the real goal should not be big data but to ask ourselves, for a given problem, what is the right data and how much of it is needed. For some problems this would imply big data, but for the majority of the problems much less data will and is needed. In this talk we explore the trade-offs involved and the main problems that come with big data: scalability, redundancy, bias, noise, spam, and privacy.

Biography: Ricardo Baeza-Yates is VP of Research for Yahoo Labs leading teams in United States, Europe and Latin America since 2006 and based in Sunnyvale, California, since August 2014. During this time he has lead the labs in Barcelona and Santiago de Chile. Between 2008 and 2012 he also oversaw the Haifa lab. He is also part time Professor at the Dept. of Information and Communication Technologies of the Universitat Pompeu Fabra, in Barcelona, Spain. During 2005 he was an ICREA research professor at the same university. Until 2004 he was Professor and before founder and Director of the Center for Web Research at the Dept. of Computing Science of the University of Chile (in leave of absence until today). He obtained a Ph.D. in CS from the University of Waterloo, Canada, in 1989. Before he obtained two masters (M.Sc. CS & M.Eng. EE) and the electronics engineer degree from the University of Chile in Santiago. He is co-author of the best-seller Modern Information Retrieval textbook, published in 1999 by Addison-Wesley with a second enlarged edition in 2011, that won the ASIST 2012 Book of the Year award. He is also co-author of the 2nd edition of the Handbook of Algorithms and Data Structures, Addison-Wesley, 1991; and co-editor of Information Retrieval: Algorithms and Data Structures, Prentice-Hall, 1992, among more than 500 other publications. From 2002 to 2004 he was elected to the board of governors of the IEEE Computer Society and in 2012 he was elected for the ACM Council. He has received the Organization of American States award for young researchers in exact sciences (1993), the Graham Medal for innovation in computing given by the University of Waterloo to distinguished ex-alumni (2007), the CLEI Latin American distinction for contributions to CS in the region (2009), and the National Award of the Chilean Association of Engineers (2010), among other distinctions. In 2003 he was the first computer scientist to be elected to the Chilean Academy of Sciences and since 2010 is a founding member of the Chilean Academy of Engineering. In 2009 he was named ACM Fellow and in 2011 IEEE Fellow.

INDIANA UNIVERSITY NETWORK SCIENCE INSTITUTE

SCHOOL OF INFORMATICS AND COMPUTING
INDIANA UNIVERSITY
Bloomington