The Web: Wisdom of Crowds & a Long Tail

Abstract The Web continues to grow and evolve very fast, changing our daily lives. This activity represents the collaborative work of the millions of institutions and people that contribute content to the Web as well as more than one billion people that use it. In this ocean of hyperlinked data there is explicit and implicit information and knowledge. But how is the Web? Web data mining is the main task to answer this question. Web data comes in three main flavors: content (text, images, etc.), structure (hyperlinks) and usage (navigation, queries, etc.), implying different techniques such as text, graph or log mining. Each case reflects the wisdom of some group of people that can be used to make the Web better. For example, user generated tags in Web 2.0 sites. One important phenomenon of this wisdom is the long tail of the special interests of people. In this talk we cover all these concepts and give specific examples.

Biography Ricardo Baeza-Yates is VP of Yahoo! Research for Europe, Middle East and Latin America, leading the labs at Barcelona, Spain and Santiago, Chile, as well as supervising the newer lab in Haifa, Israel. Until 2005 he was the director of the Center for Web Research at the Department of Computer Science of the Engineering School of the University of Chile; and ICREA Professor at the Department of Technology of the University Pompeu Fabra in Barcelona, Spain. He is co-author of the best-seller book Modern Information Retrieval, published in 1999 by Addison-Wesley with a second edition in 2011, as well as 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 200 other publications. He has received the Organization of American States award for young researchers in exact sciences (1993) and several national awards in Chile. In 2003 he was the first computer scientist to be elected to the Chilean Academy of Sciences. During 2007 he was awarded the Graham Medal for innovation in computing, given by the University of Waterloo to distinguished ex-alumni. In 2009 he was awarded the Latin American distinction for contributions to CS in the region and became an ACM Fellow, followed in 2011 by an IEEE Fellowship.

This is a joint NICTA and RMIT School of Computer Science and IT sponsored talk.