

Data Science Institute



Bias on the Web

Ricardo Baeza-Yates

is currently CTO of NTENT, a search technology company based in Carlsbad, California, since June 2016; as well as Director of Computer Science Programs (part-time) of Northeastern University, Silicon Valley campus, since January 2018. Previously, he was VP of Research at Yahoo Labs, based in Barcelona, Spain, and later in Sunnyvale, California, from January 2006 to February 2016. Between 2008 and 2012 he also supervised Yahoo Labs Haifa and between 2012 and 2014 Yahoo Labs London. 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 and founder of the Web Science and Social Computing Research Group (formerly Web Research Group) at the Dept. of Information and Communication Technologies of Universitat Pompeu Fabra in Barcelona, Spain. He maintains ties with both mentioned universities as a part-time professor. Finally, he is also an adjunct professor at the CS department of the University of Waterloo, Canada. His research interests include algorithms and data structures, information retrieval, web search and data mining, and data science and visualization. He is ACM Fellow and IEEE Fellow.

The Web is the most powerful communication medium and the largest public data repository that humankind has created. Its content ranges from great reference sources such as Wikipedia to ugly fake news. Indeed, social (digital) media is just an amplifying mirror of ourselves. Hence, the main challenge of search engines and other websites that rely on web data is to assess the quality of such data. However, as all people has their own biases, web content as well as our web interactions are tainted with many biases. Data bias includes redundancy and spam, while interaction bias includes activity and presentation bias. In addition, sometimes algorithms add bias, particularly in the context of search and recommendation systems. As bias generates bias, we stress the importance of debiasing data as well as using the context and other techniques such as explore and exploit, to break the filter bubble. The main goal of this talk is to make people aware of the different biases that affect all of us on the Web. Awareness is the first step to be able to fight and reduce the vicious cycle of bias.

November 13, 2018

Time: 5-6 p.m. Reception | 6-7 p.m. Talk
Location: Engineering Lecture Hall L2D2



DSI

Data Science Institute

uh.edu/datascience