Tutorials

SA1: The Social Semantic Web: How Web 2.0 and the Semantic Web Can Strengthen Each Other

Presented by John Breslin and Alexandre Passant

Web 2.0, a platform where people are connecting through their shared objects of interest, is encountering some boundaries in the areas of information integration, portability, search, and demanding tasks like querying. The semantic web is an ideal platform for interlinking and performing operations on the diverse data available from web 2.0 data silos, and has produced a variety of approaches to overcome some of the limitations with web 2.0. In this tutorial, Breslin and Passant aim to educate attendees on the applications of semantic web technologies to web 2.0. They will also describe how web 2.0 can serve as a very valuable source of semantically-rich data for Semantic Web applications, solving the oft-quoted chicken and egg problem with the semantic web. The tutorial is intended for professionals, researchers, graduates, practitioners and developers who wish to learn how semantic web technologies can be applied to and leveraged by social software and social media systems.

John Breslin is a lecturer in engineering and informatics at NUI Galway. He leads the Social Software Unit at DERI, NUI Galway, and is creator of the SIOC project, which aims to interlink online communities with semantic technologies. He is cofounder of boards.ie Ltd. and an advisor to CrowdGather LLC.

Alexandre Passant is a postdoctoral researcher and adjunct lecturer at DERI, NUI Galway. His research focuses on the semantic web and social software. He is the author of MOAT, a framework that allows people to tag their content using semantic technologies, and he is also coeditor of the SIOC specification.

SA2: Introduction to Social Media Network Analysis

Presented by Marc Smith and Derek Hansen

Social networks are the defining data structure of social media, created as people reply, link, click, favorite, friend, retweet, coedit, mention, or tag one another. In this tutorial, we review the core concepts and methods of social network analysis and apply it to the collection, analysis, and visualization of social media networks. Using the free and open NodeXL application, learn how to extract a social media network and generate metrics and visualizations that highlight key people and positions within streams of tweets, videos, photos, or emails.

Marc Smith is a sociologist specializing in the social organization of online communities and computer mediated interaction. Smith leads the Connected Action consulting group, which applies social science methods in general and social network analysis techniques in particular to gain insights into enterprise and internet social media usage. Smith also contributes to the open and free NodeXL project, which adds social network analysis features to the familiar Excel spreadsheet.

Derek L. Hansen is an assistant professor at Maryland's iSchool and Director for the Center for the Advanced Study of Communities and Information, a multidisciplinary research center focused on harnessing the power of social technologies to support the needs of real and virtual communities. He is also an active member of the Human Computer Interaction Lab. His research focuses on mass collaboration, consumer health informatics, alternate reality games (ARGs), and social network analysis and visualization of online interactions. Hansen completed his PhD from the University of Michigan's School of Information where he was an NSF-funded interdisciplinary STIET Fellow focused on understanding and designing effective online sociotechnical systems.
SP1: Large-Scale Social Media Analytics with Hadoop

Presented by Jake Hofman

Over the last several years there has been a rapid increase in the number, variety, and size of readily available social media data. As such, there is a growing demand for software solutions that enable one to extract relevant information from these data. For sufficiently large data sets (for example, networks with hundreds of millions of nodes) distributed solutions are often necessary, as the storage and memory constraints of single machines are prohibitive. Hadoop, an open source implementation of the map/reduce paradigm, is an increasingly popular framework for processing such data sets at the terabyte and petabyte scale using commodity hardware.

In this tutorial we will discuss the use of Hadoop for processing large-scale social data sets. We will first cover the map/reduce paradigm in general and subsequently discuss the particulars of Hadoop's implementation. We will then present several use cases for Hadoop in analyzing example data sets, examining the design and implementation of various algorithms with an emphasis on social network analysis. Accompanying data sets and code will be made available.

Jake Hofman is a member of the Human Social Dynamics group at Yahoo! Research. His work involves data-driven modeling of complex systems, focusing on applications of machine learning and statistical inference to large-scale data. He holds a B.S. in electrical engineering from Boston University and a Ph.D. in physics from Columbia University.

SP2: The Social Psychology of Social Media through the Lens of Text Analysis

Presented by Cindy Chung and James W. Pennebaker

How do the words we use reflect our personality, psychological states, and social groups? In this tutorial, we discuss computerized text analysis tools widely used in the social sciences. Particular attention will be devoted to word counting programs such as Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007) have been used to efficiently classify texts along psychological dimensions and to predict behavioral outcomes. By analyzing commonly used words such as pronouns, prepositions, and other function or junk words, it is possible to identify the language patterns and personalities of speakers and writers. We describe the rationale of our word counting approach, and highlight several applications — such as a lie detector, status decoder, or as remote sensor of group dynamics. Finally, we present novel ways in which the LIWC program is being applied for computer-mediated conversations and for thematic analyses for online social media. Participants will have the opportunity to test the LIWC software for the day on their own data and laptops, and to discuss issues specific to their research questions.

Cindy K. Chung is a post doctoral Fellow at the University of Texas, Austin. Her research focuses on the development and application of text analysis tools and methods for research questions in psychology. She examines how personality, psychological states, and social dynamics are reflected in word use.

James W. Pennebaker is a professor and chair of the Department of Psychology at the University of Texas, Austin. He studies the links between natural language and physical and mental health. More recent research focuses on the nature of language and range of psychological dynamics in the real world, including emotion, deception and social status.