

PANEL

Crowds, Clouds, and Algorithms: Exploring the Human Side of “Big Data” Applications

Panelists: Sihem Amer-Yahia (Yahoo! Research), AnHai Doan (Wisconsin),
Jon Kleinberg (Cornell), Nick Koudas (U. Toronto and Sysomos, Inc.)

Moderator: Michael J. Franklin (UC Berkeley and Truviso, Inc.)

1. OVERVIEW

The creation, collection, analysis, curation, and dissemination of data have become profoundly democratized. Social networks spanning 100's of millions of users enable instantaneous discussion, debate, and information sharing [8]. Streams of tweets, blogs, photos, and videos identify breaking events faster and in more detail than ever before [10]. Global, *ad hoc* collaborations addressing scientific, commercial, political, and even mathematical problems make progress where individual investigators or small groups cannot [6,11,12,13,14].

This sea change is the result of a confluence of information technology advances in areas such as intensively networked systems, cloud computing, social computing, and pervasive devices.

The connectivity of billions of device-enabled people to massive cloud-computing infrastructure has created a new dynamic that is moving data to the forefront of many human endeavors, changing the way that data-centric systems must be envisioned and architected. While Human-Computer Interface and Data Visualization research has long investigated how information can be presented to and manipulated by users, the emerging crowd/cloud infrastructure is changing this relationship in a more fundamental way. People not only enter and consume information, they also play a central role, individually and in groups, throughout the entire information lifecycle.

Human participation can be *direct* such as when entering User Generated Content in blogs, microblogs, and review sites, or when knowingly participating in a crowdsourcing marketplace such as Amazon Mechanical Turk. People can also participate in *indirect* ways, simply by going about their on-line lives, when searching [2], reading content, shopping, or playing on-line games [1].

The development of hybrid Crowd/Cloud systems will be a major driver of information technology innovation going forward. Progress will require breakthroughs in Machine Learning, Query Processing, Data Integration, Distributed Computing Infrastructure, Security, Privacy, and Social Computing.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGMOD '10, June 6–10, 2010, Indianapolis, Indiana, USA.

Copyright 2010 ACM 978-1-4503-0032-2/10/06...\$10.00.

2. THE PANEL

This panel explores recent results and emerging opportunities in hybrid Crowd/Cloud computing. The panelists represent a mix of viewpoints from both academia and industry. Their recent work spans a range of related topics: Recommendation Systems [2], Mass Collaboration [4], Social Network Structures [5], Social Media Analysis [9], and User-Assisted Data Integration [7].

The panel is intended to be wide-ranging and interactive, but we expect to address at least some of the following issues: Social structures and incentive schemes; collaborative data management, analysis and filtering; scaling issues for algorithms, machines and people; and new application and research opportunities.

3. REFERENCES

- [1] L. von Ahn, Games with a Purpose, *IEEE Computer*, June 2006
- [2] S. Amer-Yahia, *et al.*, Group Recommendation: Semantics and Efficiency, *PVLDB* 2(1): 754-765 (2009)
- [3] H. Choi, H. Varian, Predicting the Present with Google Trends, *Google Technical Report*, 2009.
- [4] A. Doan, R. Ramakrishnan, A. Halevy, Mass Collaboration Systems on the World Wide Web, *CACM*, to appear.
- [5] D. Easley, J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010, in press.
- [6] R. Hotz, "More Scientists Treat Experiments as a Team Sport", *Wall Street Journal* 11/20/09.
- [7] S. Jeffrey, M. Franklin, and A. Halevy, "Pay-as-you-go user feedback for dataspace systems," in *Proc. SIGMOD* 2008.
- [8] C. Li, J. Bernoff, *Groundswell, Winning in a world transformed by social technologies*. Harvard Business, 2008.
- [9] M. Mathioudakis, N. Koudas, P. Marbach, Early Online Identification of Attention Gathering Items in Social Media, *Intl. Conf. on Web Search and Data Mining (WSDM)*, 2010.
- [10] E. Ramstad, "Gulags, Nukes and a Water Slide: Citizen Spies Lift North Korea's Veil", *Wall Street Journal* 5/22/09
- [11] H. Rheingold. *Smart Mobs*. Perseus Publishing, 2003.
- [12] C. Shirky, *Here Comes Everybody: The Power of Organizing Without Organizations*. Penguin, 2008.
- [13] J. Surowiecki, *The Wisdom of Crowds*. Anchor Books, 2004.
- [14] A. Williams, *Wikinomics, How Mass Collaboration Changes Everything*, Portfolio, 2006.