Data crowdsourcing

Tova Milo

Crowd-based data sourcing is a new and powerful data procurement paradigm that engages Web users to collectively contribute data, analyze information and share opinions. Crowd-based data sourcing democratizes data-collection, cutting companies' and researchers' reliance on stagnant, overused datasets and bears great potential for revolutionizing our information world. Yet, triumph has so far been limited to only a handful of successful projects such as Wikipedia or IMDb. This comes notably from the difficulty of managing huge volumes of data and users of questionable quality and reliability. Every single initiative had to battle, almost from scratch, the same non-trivial challenges. The ad hoc solutions, even when successful, are application specific and rarely sharable. In this talk we consider the development of solid scientific foundations for Web-scale data sourcing. We believe that such a principled approach is essential to obtain knowledge of superior

quality, to realize the task more effectively and automatically, be able to reuse solutions, and thereby to accelerate the pace of practical adoption of this new technology that is revolutionizing our life. We will consider the logical, algorithmic, and methodological foundations for the management of large scale crowd-sourced data as well as the the development of applications over such information.

Tova Milo received her Ph.D. degree in Computer Science from the Hebrew University, Jerusalem, in 1992. After graduating she worked at the INRIA research institute in Paris and at University of Toronto and returned to Israel in 1995, joining the School of Computer Science at Tel Aviv university where she is now a full Professor and Department head. Her research focuses on advanced database applications such as data integration, XML and

semi-structured information, Web-based applications and Business Processes, studying both theoretical and practical aspects. Tova served as the Program Chair of several international conferences, including PODS, ICDT, VLDB, XSym, and WebDB. She is a member of the

VLDB Endowment and the ICDT executive board and is an editor of TODS, the VLDB Journal and the Logical Methods in Computer Science Journal. She has received grants from the Israel Science Foundation, the US-Israel Binational Science Foundation, the Israeli and French Ministry of Science and the European Union. She is a recipient of the 2010 ACM PODS Alberto O. Mendelzon Test-of-Time Award and of the prestigious EU ERC Advanced Investigators grant.