

The Smart Crowd – Learning from the Ones Who Know

Tova Milo

School of Computer Science, Tel Aviv University, Tel Aviv, Israel
milo@cs.tau.ac.il

Abstract

One of the foremost challenges for information technology over the last few years has been to explore, understand, and extract useful information from large amounts of data. Some particular tasks such as annotating data or matching entities have been outsourced to human workers for many years. But the last few years have seen the rise of a new research field called crowdsourcing that aims at delegating a wide range of tasks to human workers, building formal frameworks, and improving the efficiency of these processes.

In order to provide sound scientific foundations for crowdsourcing and support the development of efficient crowd sourcing processes, adequate formal models and algorithms must be defined. In particular, the models must formalize unique characteristics of crowd-based settings, such as the knowledge of the crowd and crowd-provided data; the interaction with crowd members; the inherent inaccuracies and disagreements in crowd answers; and evaluation metrics that capture the cost and effort of the crowd.

Clearly, what may be achieved with the help of the crowd depends heavily on the properties and knowledge of the given crowd. In this talk we will focus on knowledgeable crowds. We will examine the use of such crowds, and in particular domain experts, for assisting solving data management problems. Specifically we will consider three dimensions of the problem:

1. How domain experts can help in improving the *data* itself, e.g. by gathering missing data and improving the quality of existing data,
2. how they can assist in gathering *meta-data* that facilitate improved data processing, and
3. how can we find and identify the most relevant crowd for a given data management task.

1998 ACM Subject Classification H.2.4 [Database Management] Systems, Query Processing, H.2.3 [Database Management] Languages, Query Languages, H.2.8, [Database Management] Database Applications, Data Mining

Keywords and phrases Data Management, Crowdsourcing

Digital Object Identifier 10.4230/LIPICs.ICDT.2017.3

Category Invited Talk



© Tova Milo;

licensed under Creative Commons License CC-BY

20th International Conference on Database Theory (ICDT 2017).

Editors: Michael Benedikt and Giorgio Orsi; Article No. 3; pp. 3:1–3:1

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany