Crowdsourcing for Information Retrieval: Principles, Methods, and Applications

Omar Alonso  
Microsoft Corp.  
1056 La Avenida  
Mountain View, CA 94043  
omar.alonso@microsoft.com

Matthew Lease  
School of Information  
University of Texas at Austin  
1616 Guadalupe Ste 5.202  
Austin, TX 78701  
ml@ischool.utexas.edu

ABSTRACT
Crowdsourcing has emerged in recent years as a promising new avenue for leveraging today’s digitally-connected, diverse, distributed workforce. Generally speaking, crowdsourcing describes outsourcing of tasks to a large group of people instead of assigning such tasks to an in-house employee or contractor. Crowdsourcing platforms such as Amazon Mechanical Turk and CrowdFlower have gained particular attention as active online market places for reaching and tapping into this still largely under-utilized workforce. Crowdsourcing also offers intriguing new opportunities for accomplishing different kinds of tasks or achieving broader participation than previously possible, as well as completing standard tasks more accurately in less time and at lower cost. Unlocking the potential of crowdsourcing in practice, however, requires a tri-partite understanding of principles, platforms, and best practices. We will introduce the opportunities and challenges of crowdsourcing while discussing the three issues above. This will provide a basic foundation to begin crowdsourcing in the context of one’s own particular tasks.

Categories and Subject Descriptors
H.3.4 [Information Storage and Retrieval]: Systems and software — performance evaluation

General Terms
Measurements, performance, experimentation, human factors

Keywords
Crowdsourcing, human computation.

1. Motivating Application: Search Evaluation
To motivate and ground general discussion of crowdsourcing, we will focus primarily upon applications to evaluating search accuracy (with other examples like blending automation with human computation for hybrid search). While search evaluation is an essential part of the development and maintenance of search engines and other information retrieval (IR) systems, current approaches for search evaluation face a variety of practical challenges. Many Web search engines reportedly use large editorial staffs to judge the relevance of web pages for queries in an evaluation set. This is expensive and has obvious scalability issues. Academic researchers, without access to such editors, often rely instead on small groups of student volunteers. Because of the students’ limited time and availability, test sets are often smaller than desired, making it harder to detect statistically significant differences in performance by the experimental systems being tested. While behavioral data, such as obtained automatically from search engine logs of user activity, is much cheaper than the editorial method, it requires access to a large stream of data, something not always available to a researcher testing an experimental system. These challenges provide an ideal setting for demonstrating both the potential and practical challenges for the crowdsourcing paradigm.

2. Target Audience
The tutorial is designed for those with little to intermediate familiarity with crowdsourcing who want to learn about the capabilities and limitations of crowdsourcing techniques for IR. The tutorial will highlight opportunities and challenges of the crowdsourcing paradigm, emphasize design and execution of experiments, provide practical “how to” knowledge on using existing platforms (Mechanical Turk and CrowdFlower), and present established best practices for achieving efficient, inexpensive, and accurate results with crowdsourcing. Recommended background includes basic familiarity with IR evaluation and experimentation.

3. Tutorial Outline
- Introduce crowdsourcing and human computation
- Survey recent “killer apps” of crowdsourcing principles
- Provide practical “how to” guidance for effectively using Amazon’s Mechanical Turk’s API and user interface
- Summarize recent surveys of crowd worker demographics
- Discuss a variety of incentive structures available to encourage quality and quantity of work
- Emphasize design and human-centric practices with statistical methods to maximize results
- Describe Summarize key best practices for achieving efficient, inexpensive, and accurate work
- Review current opportunities, untapped potential, and open challenges for IR crowdsourcing, particularly evaluation