

Call for Papers: [Information Retrieval](#), a Springer journal

## Special Issue on *Crowdsourcing for Information Retrieval*

### Introduction

The advent of [crowdsourcing](#) is revolutionizing information technology by enabling a wide host of new methodology for data collection, annotation, and processing, as well as system design and evaluation. Given the massive datasets that are both ubiquitous and rapidly growing in today's digital age, the field of information retrieval (IR) stands to particularly benefit from such advances as academic researchers and industrial practitioners alike leverage the [wisdom of crowds](#) to more effectively cope with and exploit massive datasets. The novelty of crowdsourcing makes this an especially exciting time to focus the IR community's attention toward it and provide an opportunity for discussion and education on this key emerging area. Traditionally manual-labor intensive processes have been particularly impacted by dramatically reducing the time, cost, and effort involved. In IR, this has driven a disruptive shift in areas like:

*Evaluation:* The Cranfield paradigm for evaluating search engines requires manually assessing document relevance to search queries. Recent work on stochastic evaluation has reduced but not removed this dependence on manual assessment.

*Supervised Learning:* While traditional costs associated with data annotation have driven recent machine learning work (e.g. Learning to Rank) toward greater use of unsupervised and semi-supervised methods, the emergence of crowdsourcing has made labeled data far easier to acquire, thereby driving a potential resurgence in supervised or semi-supervised methods.

*Applications:* Crowdsourcing has introduced exciting new opportunities to integrate human labor into automated systems: handling difficult cases where automation fails, exploiting the breadth of backgrounds, geographic dispersion, and real-time crowd response, etc.

While existing studies on crowdsourcing for IR have been encouraging, a variety of important questions remain open with regard to theory, methodology, policy, and best practices:

- How can alternative labor incentive structures be optimally used individually or in combination with one another: financial (pay real or [virtual money](#) for work), goodwill (support charity via work, e.g. [Samasource](#)), prestige (credit and publicize workers for quantity and quality produced), or fun (e.g. [Games with a Purpose](#)).
- How can we encourage greater participation from the most effective annotators?
- How can spammers be quickly and accurately detected?
- What form of technological interfaces and interaction mechanisms are most useful?

- Should the community create reusable public test collections, and what form should such collections take (e.g. shared [data](#) from the NAACL 2010 AMT Workshop)?
- Automated systems stand to particularly benefit from greater annotation of rare events, yet such events are much more likely to expose users' personal information. How can we effectively crowdsource rare events while still protecting users' privacy?
- Distinguishing broad topics in crowdsourcing vs. those specific to search and data mining contexts, how can we effectively engage each other and those outside our community in addressing the range of open questions before us?
- In comparison to existing work developing effective practices for in-house annotation, managing outsourcing and remote labor, and conducting and interpreting online surveys, what related lessons have we already learned? What lessons do not apply to crowdsourcing, or if some must be adapted or reconceived to be applicable, which and how? What questions did prior work in these areas leave open, and what truly novel questions has crowdsourcing introduced?
- What crowdsourcing policies and practices should we adopt so that cooperation among requestors and workers maximizes reward and fairness for all parties?
- The novelty of crowdsourcing coupled with the decentralized and varied nature of Institutional Review Boards (IRBs) across different institutions has created significant confusion for many researchers regarding IRB-implications of crowdsourcing. What positions and policies have IRBs adopted at different institutions, and how have researchers worked with their local IRBs when appropriate to obtain approval?
- Much of crowdsourcing work today involves fairly simple applications in which workers perform many independent, brief, low-level annotations. Is crowdsourcing fundamentally limited to this simple model, or can we envision richer, more sophisticated application of crowdsourcing ideas without sacrificing key tenets and benefits of the paradigm?

## Call for Papers

The special issue welcomes novel, high-quality manuscripts on the development, evaluation, and theoretical analysis of crowdsourcing for IR. Submissions to the special issue should not be under consideration in any other journal or conference and will be evaluated according to the Information Retrieval Journal reviewing criteria and appropriateness to the special issue. If the submission is a revision of a previous conference paper, the revision must contain significant amplification or clarification of the original material or there is some significant additional benefit to be gained. For more details, please refer to "manuscript submission" on the journal [homepage](#). Submissions should use the Information Retrieval Journal style templates available from the Journal's [homepage](#) and should be submitted through the IR Journal [online submission](#)

[page](#), selecting the “S.I.: Crowdsourcing for Info. Ret.” article type.

Topics of interest include but are not limited to the following areas:

- Theoretical, experimental, and/or methodological developments advancing state-of-the-art knowledge of crowdsourcing for IR
- Novel applications of IR enabled by crowdsourcing (e.g. use of real-time crowd response, using human computation to “close the loop” with automation for hybrid IR systems, etc.)
- Tutorials or best practices on employing crowdsourcing to support different IR scenarios
- Innovative ideas for better using crowdsourcing platforms such as [Amazon’s Mechanical Turk](#), [Crowdfunder](#), [Livework](#), [DoMyStuff](#), etc.
- Description of novel software packages improving ease and/or quality of crowdsourcing, (e.g. [TurkIt](#), [qmturk](#), [Turk Surveyor](#), [Get Another Label](#), etc.)
- Reflective or forward-looking vision on use of crowdsourcing in this area

### **Important dates**

Submissions due: May 6, 2011

Reviewer feedback to authors: June 24, 2011

Revised submissions due: July 22, 2011

Notification of acceptance and final reviewer feedback: August 22, 2011

Final submissions due: September 30, 2011

### **Guest Editors**

All questions regarding submissions should be directed to the special issue Guest Editors:

- [Vitor Carvalho](#), Intelius
- [Matthew Lease](#), University of Texas at Austin
- [Emine Yilmaz](#), Microsoft Research