



Crowdsourcing for Relevance Evaluation

June 08, 2010

16:00 – 19:00 / Half day

General Description

Length: 3 hours (half-day)

Location: Universidad Carlos III de Madrid (7.0.J05)

Intended Audience: Introductory to intermediate

Background required: familiarity with information retrieval

Relevance evaluation is an essential part of the development and maintenance of information retrieval systems. There are a number of limitations with current approaches for relevance evaluation. 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.

Behavioral data is much cheaper than the editorial method but has limitations as well. It requires access to a large stream of data, something not always available to a researcher testing an experimental system and there are certain tasks for which it does not make sense.

Recently, crowdsourcing has emerged as a feasible alternative for relevance evaluation because it combines the flexibility of the editorial approach at a larger scale.

Crowdsourcing is a term used to describe tasks that are outsourced to a large group of people instead of performed by an employee or contractor. Crowdsourcing is an open call to solve a problem or carry out a task and usually involves a monetary value in exchange for such service.

Amazon Mechanical Turk (AMT) is an example of a crowdsourcing platform that has gained a lot of attention as a tool for conducting different kinds of relevance evaluations. The AMT service is easy to use and has useful features for setting up experiments and collecting results. However, it is important to pay attention to the design of the experiment and its execution to gather useful results.

This tutorial is aimed at those who are interested in using crowdsourcing as another technique for performing different kinds of evaluations and user studies. The goal is to



discuss the value of the crowdsourcing paradigm.

Tutorial Objectives

At the end of the tutorial, participants will understand:

- When to use crowdsourcing for an experiment
- How to use AMT via the user interface and API
- How to set up relevance experiments
- Apply design guidelines to maximize results

Presenter

Omar Alonso (Microsoft Bing.com) is part of the Bing team at Microsoft. He has been working on crowdsourcing for the last two years in industry and as a researcher applying this technique for a diverse set of experiments. Previously he was at A9.com (an Amazon.com company) and Oracle Corp. This past summer he was visiting researcher at Max-Planck Institute for Informatik. He is the workshop co-chair for SIGIR 2010. Omar holds a PhD in computer science from the University of California at Davis.