CrowdTerrier: Automatic Crowdsourced Relevance Assessments with Terrier

Richard McCreadie, Craig Macdonald, and Iadh Ounis
firstname.lastname@glasgow.ac.uk

School of Computing Science
University of Glasgow
G12 8QQ, Glasgow, UK

Categories and Subject Descriptors: H.3.3 [Information Storage & Retrieval]: Information Search & Retrieval

General Terms: Experimentation, Performance

Keywords: Crowdsourcing, Relevance Assessment, Terrier

ABSTRACT

Information retrieval (IR) systems rely on document relevance assessments for queries to gauge their effectiveness for a variety of tasks, e.g. Web result ranking. Evaluation forums such as TREC and CLEF provide relevance assessments for common tasks. However, it is not possible for such venues to cover all of the collections and tasks currently investigated in IR. Hence, it falls to the individual researchers to generate the relevance assessments for new tasks and/or collections. Moreover, relevance assessment generation can be a time-consuming, difficult and potentially costly process. Recently, *crowdsourcing* has been shown to be a fast and cheap method to generate relevance assessments in a semi-automatic manner [1]. In this case, the relevance assessment task is outsourced to a large group of non-expert workers, where workers are rewarded via micro-payments.

In this demo, we present *CrowdTerrier*, an infrastructure extension to the open source Terrier IR platform [2][1] that enables the semi-automatic generation of relevance assessments for a variety of document ranking tasks using crowdsourcing. The aim of CrowdTerrier is to reduce the time and expertise required to effectively Crowdsource relevance assessments by abstracting away from the complexities of the crowdsourcing process. It achieves this by automating the assessment process as much as possible, via a close integration of the IR system that ranks the documents (Terrier) and the crowdsourcing marketplace that is used to assess those documents (Amazon's Mechanical Turk (MTurk)).

As illustrated in Figure 1, CrowdTerrier is comprised of three components. *CrowdControl* handles the conversion of ranked results from Terrier as well as the administration of the MTurk tasks. The *JSP Interface* is responsible for the presentation of the pages to be assessed and the assessment interface. Finally the *Validator* performs quality assurance on the assessments produced, possibly with human input. Each of these components is fully customisable to facilitate tackling different tasks and collections.

[1] `http://terrier.org`

SIGIR'12, August 12–16, 2012, Portland, Oregon, USA.
ACM 978-1-4503-1472-5/12/08.

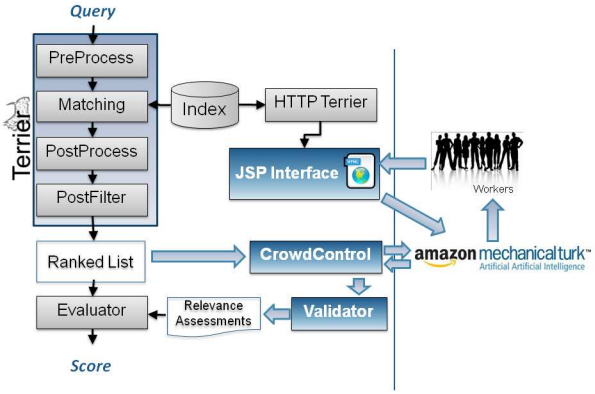

Figure 1: CrowdTerrier relevance assessment generation process.

In comparison to crowdsourcing document relevance assessments from scratch, CrowdTerrier[2] makes the following four key contributions:

- It abstracts away from the assessment task design and implementation, by providing out-of-the-box assessment interfaces driven by Terrier's document presentation capabilities.
- Once configured, it automatically spreads documents ranked by Terrier across multiple assessment tasks, according to best crowdsourcing practises.
- It supplies tools to monitor and summarise the progress of each relevance assessment task once launched.
- It supports both automatic and semi-automatic work validation strategies out-of-the-box for performing quality assurance on the assessments produced.

In summary, we believe that CrowdTerrier provides a useful toolkit for creating relevance assessments by researchers.

1. REFERENCES

[1] R. McCreadie, C. Macdonald and I. Ounis. Identifying Top News using Crowdsourcing. *Information Retrieval*, 2012. DOI: 10.1007/s10791-012-9186-z.

[2] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald and C. Lioma. Terrier: A high performance and scalable information retrieval platform. In *Proc. OSIR Workshop 2006.*

[2] CrowdTerrier is available from `http://terrier.org/crowdterrier/`.