Crowdsourced Databases: Query Processing with People

Author: Marcus, Adam; Wu, Eugene; Karger, David R.; Madden, Samuel; Miller, Robert C.

Citable URI: http://hdl.handle.net/1721.1/62827

Department: Massachusetts Institute of Technology. Department of Electrical Engineering and Computer Science; Massachusetts Institute of Technology. Computer Science and Artificial Intelligence Laboratory

Publisher: CIDR

Date Issued: 2011-01

Abstract:
Amazon's Mechanical Turk (MTurk) service allows users to post short tasks (HITs) that other users can receive a small amount of money for completing. Common tasks on the system include labelling a collection of images, combining two sets of images to identify people which appear in both, or extracting sentiment from a corpus of text snippets. Designing a workflow of various kinds of HITs for filtering, aggregating, sorting, and joining data sources together is common, and comes with a set of challenges in optimizing the cost per HIT, the overall time to task completion, and the accuracy of MTurk results. We propose Qurk, a novel query system for managing these workflows, allowing crowd-powered processing of relational databases. We describe a number of query execution and optimization challenges, and discuss some potential solutions.

URI: http://hdl.handle.net/1721.1/62827

Citation: Marcus, Adam, et al."Crowdsourced Databases: Query Processing with People." 5th Biennial Conference on Innovative Data Systems Research (CIDR '11) January 9-12, 2011, Asilomar, California, USA

Version: Final published version

Terms of Use: Creative Commons Attribution

Detailed Terms: http://creativecommons.org/licenses/by/3.0

Published As: http://www.cidrdb.org/cidr2011/program.html


Show full item metadata

Files in this item

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downloadable Full Text</td>
<td></td>
<td>PDF</td>
</tr>
</tbody>
</table>

This item appears in the following Collection(s)

- MIT Open Access Articles

Except where otherwise noted, this item's license is described as Creative Commons Attribution
people seeking answers on micro-blogs through asking or sharing questions with their friends. This can be easily done via smart phones, which disseminate a question to a large number of users through message propagation in microblogs. This trend is important and known as CrowdSearch. In recent years, Crowdsourcing databases [2, 5, 6] have attracted substantial interest in the database research community. Many fundamental infrastructures are proposed to support various kinds of query processing in the crowd.