CSCI8380 (Spring 2012): Paper Review Form

Reviewer Name: M. Usman Nisar

Paper Name: CrowdDB: Answering Queries with Crowdsourcing

Section I. Overview

A. Reader Interest

1. Which category describes this manuscript?
   ___Practice/Application/Case Study/Experience Report
   _X_Research/Technology
   ___Survey/Tutorial/How-To

B. Content

1. Please explain how this manuscript advances this field of research and/or contributes something new to the literature.

This paper brings up an interesting idea to query relational model for 'missing data', to perform computationally difficult functions, and for matching, ranking or aggregating results based on fuzzy criteria. The idea is to use human input via crowdsourcing that neither database systems nor search engines can adequately answer at their own. Plus, this model uses CrowdSQL as query execution engine which is an extension of SQL. The paper explains the design of CrowdDB, pros and cons of different approaches and present a report on initial set of experiments using Amazon Mechanical Turk.

C. Presentation

1. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on?
   ___Yes
   _X_Could be improved
   ___No

2. How would you rate the organization of the manuscript? Is it focused? Is the length appropriate for the topic?
   ___Satisfactory
   _X_Could be improved
   ___Poor

3. Please rate and comment on the readability of this manuscript.
   ___Easy to read
   _X_Readable - but requires some effort to understand
   ___Difficult to read and understand
   ___Unreadable

Section II. Evaluation

Please rate the manuscript. Explain your choice.

___Award Quality
_X_Excellent
___Good
___Fair
___Poor

Section III. Detailed Comments (provide your thoughts/criticism about the ideas in the paper; not only summarize the paper but have a critical look here)

The paper talks about an interesting fusion of two seemingly different fields: Database Systems and CrowdSourcing. In a nutshell, the paper proposes that crowdsourcing can be used to support use cases over relational model that involve missing data and subjective comparisons. One thing that I mentioned below too is how does very slow response times be practical for a database? Plus, there is no mention of how to cleanse the data?

Additional Comments:

1. Provide one aspect that you liked the most in this paper.
   One thing I like the most is that for CrowdSQL, the authors have extended existing SQL syntax/operators and haven't developed a query language from scratch. This obviously makes the integration of this idea for other databases much more easy and transparent to the users

2. Provide one aspect that you disliked the most in this paper.
   The paper doesn't elaborate the cost model of CrowdDB in detail. Plus, what businesses/customers can benefit from such a crowd-sourced data? Considering the delays in response times, what could be the implications if such a system is used for an OLTP system?

Section IV. Discussion Points (provide at least 3 discussion topics/questions related to ideas/techniques described in the paper; these will be used for discussions in the class)

1. *How to measure the quality of the answer?*
2. *How caching can be used to improve the performance/query performance time by CrowdDB?*
3. *Study how spamming can be automatically detected by the system and rejected in real time?*