RGC Ref.: M-HKUST602/12
_____
*(please insert ref. above)*

# The Research Grants Council of Hong Kong
## SRFDP & RGC ERG Joint Research Scheme
## <u>Completion Report</u>

*(Please attach a copy of the completion report submitted to the Ministry of Education by the Mainland researcher)*

## Part A:  The Project and Investigator(s)

### 1.  Project Title

Crowdsourcing via Social Media Platforms

### 2.  Investigator(s) and Academic Department/Units Involved

| | Hong Kong Team | Mainland Team |
|---|---|---|
| Name of Principal Investigator *(with title)* | Prof. Lei Chen (陈雷) | Prof. Yunhao Liu (刘云浩) |
| Post | Professor | Professor |
| Unit / Department / Institution | Department of Computer Science and Engineering, Hong Kong University of Science and Technology | School of Software, TsingHua University |
| Contact Information | leichen@cse.ust.hk | yunhao@Tsinghua.edu.cn |
| Co-investigator(s) *(with title and Institution)* | N.A. | N.A. |
| PhD student(s) (with period of involvement) | Name: Zhao Chen Institution: HKUST Period from 01/03/2013 to 29/02/2016 Name: Xinglin Zhang Institution: HKUST Period from 01/03/2013 to 31/07/2014 | |

*Note:  The Hong Kong project team must involve at least one research postgraduate student pursuing a Doctor of Philosophy degree at the UGC-funded university (PhD student) at any time throughout the project period.*

### 3.  Project Duration

| | Original | Revised | Date of RGC/ Institution Approval *( must be quoted)* |
|---|---|---|---|
| Project Start date | 01/03/2013 | | |
| Project Completion date | 29/02/2016 | | |
| Duration   *(in month)* | 36 | | |
| Deadline for Submission of Completion Report | 28/02/2017 | | |

## 5.  Project Objectives

5.1  Objectives as per original application

1.  *Develop a general framework for crowdsourcing using social media platforms.*
2.  *Develop techniques for automatic social media user profiling based on message graphs, follower graphs and hybrid-graphs.*
3.  *Develop effective solutions for aggregating crowdsourcing results based on pay-as-you-go models and predication market models.*
4.  *Develop novel data quality control mechanisms for controlling the quality of the aggregated crowdsourcing results.*
5.  *Develop a prototype system to demonstrate the research outputs with example applications*

5.2  Revised Objectives

Date of approval from the RGC:  _____

Reasons for the change: : N.A_____

    *1.*
    *2.*
    *3. ....*

## 6.  Research Outcome

Major findings and research outcome

*1.* We have developed a general framework for crowdsourcing over a social media platform.
*2.* We have developed a solution to profile users on social networks.
*3.* We have designed solutions for aggregating the crowdsourced results via predication market models.
*4.* We have designed quality control mechanisms, including selecting the proper workers and assigning suitable tasks to workers based on their profiles.
*5.*  We have developed a prototype system, which we have released on the open source platform, GitHub.

Potential for further development of the research and the proposed course of action
*(maximum half a page)*

There are several future research directions based on the outcome of this project
1. Based on the collected user profiles on social networks, we will design solutions to categorize workers into clusters and develop algorithms to assign tasks to workers.
2. We will design different voting mechanisms to collect the crowdsourced answers, which can help reduce possible errors in crowdsourced results.
3. In addition to handling single worker tasks, it is quite challenging to develop solutions for crowdsourcing campaign tasks on social media, we will work on that direction as well.

**7. The Layman's Summary**
*(describe <u>in layman's language</u> the nature, significance and value of the research project, in no more than 200 words)*

In this project, we developed a general framework to enable crowdsourcing tasks on social media platforms. Specifically, we designed various techniques to implement the framework, including user profiling, answer aggregation, quality control and incentive design. The outcomes of this project shed light on how to best utilize the power of social media for crowdsourcing tasks. The developed algorithms together with the prototype system lay a solid foundation for crowdsourcing over social media theoretically and experimentally.
.

## Part C:   Research Output

**8. Peer-reviewed journal publication(s) arising <u>directly</u> from this research project**
*(Please attach a copy of each publication and/or the letter of acceptance if not yet submitted in the previous progress report(s).   All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.))*

| The Latest Status of Publications | | | | Author(s) *(bold the authors belonging to the project teams and denote the corresponding author with an asterisk\*)* | Title and Journal/ Book *(with the volume, pages and other necessary publishing details specified)* | Submitted to RGC *(indicate the year ending of the relevant progress report)* | Attached to this report *(Yes or No)* | Acknowledged the support of this Joint Research Scheme *(Yes or No)* | Accessible from the institutional repository *(Yes or No)* |
|---|---|---|---|---|---|---|---|---|---|
| Year of publication | Year of Acceptance *(For paper accepted but not yet published)* | Under Review | Under Preparation *(optional)* | | | | | | |
| 2014 | | | | **Xinglin Zhang**\*, Zheng Yang, Zimu Zhou, Haibin Cai, **Lei Chen**, Xiangyang Li | Free Market of Crowdsourcing: Incentive Mechanism Design for Mobile Sensing. TPDS. 25(12),3190-3200, (2015) | 2014 | Yes | Yes | Yes |
| 2014 | | | | Lei Zou\*, M. T. Özsu, **Lei Chen**, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao | gStore: a graph-based SPARQL query engine. VLDB J. 23(4): 565-590 (2014) | 2014 | Yes | Yes | Yes |

5

| 2015 | | | | | Mingxuan Yuan*, **Lei Chen**, Philip S. Yu, Hong Mei | Privacy preserving graph publication in a distributed environment. World Wide Web 18(5): 1481-1517 (2015) | 2017 | Yes | Yes | Yes |
|---|---|---|---|---|---|---|---|---|---|---|
| 2015 | | | | | Xiaofei Zhang*, **Lei Chen**, Min Wang | Efficient Parallel Processing of Distance Join Queries Over Distributed Graphs. IEEE Trans. Knowl. Data Eng. 27(3): 740-754 (2015) | 2017 | Yes | Yes | Yes |
| 2015 | | | | | Shaoxu Song*, Aoqian Zhang, **Lei Chen**, Jianmin Wang | Enriching Data Imputation with Extensive Similarity Neighbors. PVLDB 8(11): 1286-1297 (2015) | 2017 | Yes | Yes | Yes |

9. **Recognized international conference(s) in which paper(s) related to this research project was/were delivered** *(Please attach a copy of each delivered paper. All listed papers must acknowledge RGC's funding support by quoting the specific grant reference.)*

| Month/Year/Place | Title | Conference Name | Submitted to RGC *(indicate the year ending of the relevant progress report)* | Attached to this report *(Yes or No)* | Acknowledged the support of this Joint Research Scheme *(Yes or No)* | Accessible from the institutional repository *(Yes or No)* |
|---|---|---|---|---|---|---|
| 06/2014 Salt Lake City, U.S.A. | CrowdMatcher: crowd-assisted schema matching. | SIGMOD, 2014 | 2014 | Yes | Yes | Yes |
| 04/2014 Chicago, IL, U.S.A. | C-DMr: Crowd-powered Decision Maker for real world Knapsack Problems | ICDE, 2014 | 2014 | Yes | Yes | Yes |
| 04/2014 Chicago, IL, U.S.A. | CrowdCleaner: Data cleaning for multi-version data on the web via crowdsourcing. | ICDE, 2014 | 2014 | Yes | Yes | Yes |
| 08/2013 Chicago, Illinois, U.S.A. | WiseMarket: a new paradigm for managing wisdom of online social users. KDD 2013: 455-463 | KDD, 2013 | 2014 | Yes | Yes | Yes |
| 08/2013 Riva del Garda, Trento, Italy | Reducing Uncertainty of Schema Matching via Crowdsourcing. | VLDB, 2013 | 2014 | Yes | Yes | Yes |

| | | | | | | |
|---|---|---|---|---|---|---|
| 08/2013 Riva del Garda, Trento, Italy | Bitlist: New Full-text Index for Low Space Cost and Efficient Keyword Search | VLDB, 2013 | 2014 | Yes | Yes | Yes |
| 10/2013 San Francisco, U.S.A. | Causality and responsibility: probabilistic queries revisited in uncertain databases | CIKM, 2013 | 2014 | Yes | Yes | Yes |
| 11/2013 Orlando, Florida, USA | GeoTruCrowd: trustworthy query answering with spatial crowdsourcing. | SIG GIS 2013 | 2014 | Yes | Yes | Yes |
| 06/2015 Melbourne, Australia | Online Video Recommendation in Sharing Community | SIGMOD 2015 | 2017 | Yes | Yes | Yes |

## 10. Student(s) trained *(Please attach a copy of the title page of the thesis.)*

| Name | Degree registered for | Date of registration | Date of thesis submission/ graduation |
|---|---|---|---|
| Chen Cao | Ph.D. | Fall 2011 | July 27th, 2014 |
| Xinlin Zhang | Ph.D. | Fall 2011 | July, 2014 |

## 11. Other impact *(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, etc.)*

Best paper award, APWEB 2013, "Privacy Preserving Graph Publication in a Distributed Environment"