

Process-Based Quality Control Techniques in Crowdsourcing

by

Ria Mae Harina Borromeo

A dissertation submitted in partial fulfillment for the
degree of Doctor of Philosophy in Engineering

to the

School of Science for Open and Environmental Systems
Graduate School of Science and Technology

Keio University

August 2017

Thesis Abstract

No. _____

Registration Number	<input checked="" type="checkbox"/> "KOU" <input type="checkbox"/> "OTSU" No. _____ *Office use only	Name	BORROMEO, Ria Mae Harina
Thesis Title Process-Based Quality Control Techniques in Crowdsourcing			
Thesis Summary <p>Crowdsourcing has become a popular approach to complete tasks that are tedious using manual methods or difficult for automatic ones. As crowdsourcing taps the capacities of humans, its possibilities are endless. However, the unpredictability of human behavior and the actuality of human error make it difficult to consistently achieve high-quality outcomes, posing quality control as a major challenge in crowdsourcing. Although many strategies have been proposed to optimize data quality, their effectiveness is dependent on each particular crowdsourcing application. To solve this, more techniques and experiments from which humans and algorithms can learn from are needed to be able to build a recommendation system that proposes quality management techniques depending on the attributes of the crowdsourcing application.</p> <p>In this dissertation, I approach quality management in crowdsourcing based on the sub-processes involved, specifically: task design, task deployment, and task assignment. I first experimented on factors affecting task design. In particular, I tested the effect of task complexity on a data extraction task and crowd type on a sentiment analysis task. Experiments show that there is no significant difference in the quality achieved from simple and more complex versions of a data extraction task and that the performance of paid unpaid workers are comparable in a sentiment analysis task. For task deployment, task deployment strategies were proposed along three dimensions: work structure, workforce organization, and work style. To semi-automatically implement these strategies in a crowdsourcing platform, a deployment tool was designed and developed. The effectiveness of the strategies when applied to text creation tasks were then studied and recommendations were drafted for both crowdsourcing researchers and practitioners. Finally, for task assignment, a fuzzy clustering-based method for building a personalized summary of tasks, also known as composite tasks, for crowd workers was validated. As observed from the experiments, personalization improves the workers' overall experience and that diversifying tasks can improve the workers' output quality.</p>			