Automatic Classification of Software Related Microblogs

Philips Kokoh Prasetyo, Singapore Management University
pprasetyo@smu.edu.sg
David Lo, Singapore Management University
davidlo@smu.edu.sg
Palakorn Achananuparp, Singapore Management University
palakorn@smu.edu.sg
Yuan Tian, Singapore Management University
yuan.tian.2012@smu.edu.sg
Ee-Peng Lim, Singapore Management University
eplim@smu.edu.sg

28th IEEE International Conference on Software Maintenance
(ICSМ’12), September 2012

LARC Accepted Conference Papers: http://smu.edu.sg/centres/larc/conference-papers/
**ABSTRACT**

Millions of people, including those in the software engineering communities have turned to microblogging services, such as Twitter, as a means to quickly disseminate information. A number of past studies by Treude et al., Storey, and Yuan et al. have shown that a wealth of interesting information is stored in these microblogs. However, microblogs also contain a large amount of noisy content that are less relevant to software developers in engineering software systems. In this work, we perform a preliminary study to investigate the feasibility of automatic classification of microblogs into two categories: relevant and irrelevant to engineering software systems. We extract features from the textual content of the microblogs and the titles of any URLs mentioned in the microblogs. These features are then used to learn a discriminative model used in classifying relevant and irrelevant microblogs. We show that our trained model can achieve a promising classification performance.