

# Email and Trouble Report Analysis for Revealing Context with the Project Replayer

Kimiharu Ohkura<sup>†</sup>, Shinji Kawaguchi<sup>†</sup>, Noriko Hanakawa<sup>‡</sup>, and Hajimu Iida<sup>†</sup>

<sup>†</sup> *Graduate School of Information Science, Nara Institute of Science Technology, Japan*

{kimiha-o, kawaguti}@naist.jp, iida@itc.naist.jp

<sup>‡</sup> *Faculty of Management Information, Hannan Univ., Japan*

hanakawa@hannan-u.ac.jp

**Abstract**— In many software projects, same mistakes tend to be made due to lack of knowledge that should have been gained from past practice. In order to capture such knowledge from past projects, finding underlying contexts as specific phenomena in project data archive is very important. However, such contexts cannot be known directly from software documents or formal reports, and manually finding out valuable phenomena from huge raw archives is very difficult.

To capture invisible contexts, we focused on archives of email and trouble report. Email is widely used by developers to communicate each other during their project. Email archives contain many contexts about the development, such as notifications of program code modification, negotiations for product specification change, and other interactions. Trouble reports are created and recorded by using issue tracking systems, such as GNATS. Trouble report archives also contain various contexts for specific problems and change requests occurred during software development.

Since both emails and trouble reports are produced usually in large number during the projects, we have developed tools to summarize them into number of topics. Those topics are classified based on natural language processing and clustering algorithms. Classified topics are aligned into a time-series chart for intuitive visualization. By overlapping them with other time-series charts such as growth of LoC, some characteristic phenomena in a software project would be visualized and therefore, some underlying contexts can be unveiled. These features for email and trouble report archive analysis and graphic visualization is implemented as a part of the Project Replayer, a tool to review past project data.