

ABSTRACT

ANALYSIS OF CLUSTERING IN JAVA PROGRAM

Andimas Muhammad Bagaswara

14/368447/PA/16285

Program comprehension is an important cognitive process in software development, it has many potential applications like reverse engineering and program collaboration. However, program comprehension has been known as time-consuming activity and very hard to accomplish without the knowledge of the program domain.

This research tries to apply clustering method using *k-prototypes* to facilitate program comprehension on the method level. This research uses two application as case studies: Ghost - Professional Blogging App and Protein – an atypical Dribbble App. The clustering result then analyzed to get the nature of method creation in a Java program.

The clustering evaluation from both case studies shows that using Silhouette Measure $k = 2$ is the best result while using Sum of Squared Errors $k = 6$ has the best cluster validity result. Although some of the clusters have revealed the similarities of behaviors, others have not formed a certain similar behavior.

Keywords: data mining, cluster analysis, *k-prototypes* clustering, program comprehension, Java.

INTISARI

ANALISIS CLUSTERING DALAM PROGRAM JAVA

Andimas Muhammad Bagaswara

14/368447/PA/16285

Program Comprehension adalah proses kognitif yang penting dalam pengembangan perangkat lunak, proses ini memiliki banyak potensi pengaplikasian seperti *reverse engineering* dan *program collaboration*. Namun, *program comprehension* telah dikenal sebagai aktivitas yang memakan waktu dan sangat sulit dicapai tanpa pengetahuan mengenai domain program.

Penelitian ini mencoba menerapkan metode *clustering* menggunakan *k-prototypes* untuk memfasilitasi pemahaman program pada tingkat *methods*. Penelitian ini menggunakan dua aplikasi sebagai studi kasus: *Ghost - Professional Blogging App* and *Protein – an atypical Dribbble App*. Hasil pengelompokan kemudian dianalisis untuk mendapatkan sifat pembuatan *methods* dalam program Java.

Evaluasi pengelompokan dari kedua studi kasus menunjukkan bahwa dengan menggunakan Silhouette Measure $k = 2$ adalah hasil terbaik semestara dengan menggunakan Sum of Squared Errors $k = 6$ memiliki hasil validitas kluster terbaik. Meskipun beberapa kluster telah mengungkapkan kesamaan perilaku, beberapa yang lain belum membentuk perilaku yang sama.

Kata kunci: data mining, cluster analysis, *k-prototypes* clustering, program comprehension, Java.