# REVIEW ON BUG-REPORT ANALYSIS TECHNIQUES AND APPROACHES

**Romanova U.V., Liutenko I.V.**
*National Technical University «Kharkiv Polytechnic Institute», Kharkiv*

Abstract. The paper presents a review on the existing works on bug-report analysis. In particular, this paper provides some background on bug-reports, describes main approaches to bug-report deduplication, prioritization and triage.

Introduction. Bug report processing is a key element of bug fixing in modern software maintenance. As programmers can hardly write programs without any bugs, it is inevitable to find bugs and fix them in software development. Moreover, it is costly and time-consuming to find and fix bugs manually. Bug reports provide crucial information to developers, however, these reports widely differ in their quality.

Relevance. Programs designed for automated bug-report processing are in great demand among software professionals who are continuously striving to achieve high quality software development by fixing various types of software bugs. During the software development and maintenance stages, software bugs are the major factor that can affect the cost and time of software delivery. To efficiently fix a software bug, open bug repositories are used for identifying, classifying and prioritizing bug reports. Owing to a lack of resources such as time and manpower, these bug-report deduplication, prioritization and triage processes are extremely important in software development.

In general, bug-report processing involves the processes from submitting the bug report [1], to fixing the bug. To collect and assign bug reports automatically, some works focus on how to automate bug-report triage [2] by identifying duplicate bug reports on the same bug, predicting the priority of bug reports, and assigning bug reports with high precision. Furthermore, to help developers fix the bugs reported by bug reports, some works focus on fixing bugs based on these bug reports as well.

In this work, in the case of duplicate bug report identification, topic modeling, natural language processing, information retrieval, and clustering based on similarity techniques are analyzed. In the case of improving bug report triage accuracy such classification techniques as k-nearest neighbors (KNN), Naive Bayes (NB), and support vector machine (SVM) as well as latent Dirichlet allocation (LDA) in combination with other methods [3] are investigated.

Reviews of each method analyzed in this work consist of a description, used techniques, experiments, and comparison results.

**References:**

1. T. Dal Sasso and M. Lanza, "In-bug: visual analytics of bug repositories," Antwerp, Belgium, 2014

2. Runeson P, Alexandersson M, Nyholm O.,"Detection of duplicate defect reports using natural language processing", Minneapolis, 2007.

3. Zhang J., Wang X., Hao D., Xie B., Zhang L., Mei H., "A survey on bug-report analysis" China, 2014.