Summary

Localization and diagnosis of faults in large computer systems cannot function without effective anomaly detection methods. These methods are widely researched and are gaining popularity; however, their application in commercial projects remains limited. This is due to the fact that the results achieved on example benchmark datasets often significantly differ from those on commercial datasets. Another reason is that localization and diagnostic methods are often based on specific features of a given system.

Aiming to ensure the universality of the solution, I focused on a common source of information about execution, namely logs. Logs are widely used, but their lack of regular structure presents a barrier to their effective utilization in anomaly detection and fault localization. Although natural language processing (NLP) methods are widely applied to log content processing, unsupervised word segmentation methods have yet to gain wider popularity and use in log sequences.

In my doctoral work, I proposed log segmentation and analysis methods that enable better anomaly detection and more precise localization of software fault sources. Experiments were largely conducted on production log datasets from real faults provided by Nokia. Their anonymized form, as well as the algorithm, are available online.

Wrocław, 25.09.2024 Wojciech Dobrowolski