

# Streszczenie

Lokalizacja i diagnostyka błędów w dużych systemach komputerowych nie mogą funkcjonować bez skutecznych metod wykrywania anomalii. Metody te są szeroko badane i zyskują na popularności, jednak ich zastosowanie w komercyjnych projektach pozostaje ograniczone. Wynika to z faktu, że wyniki osiągnięte na przykładowych zbiorach referencyjnych często znacząco odbiegają od wyników na zbiorach komercyjnych. Drugim powodem jest to, że metody lokalizacji i diagnostyki nierzadko opierają się na specyficznych cechach danego systemu.

Dążąc do zapewnienia uniwersalności rozwiązania, skoncentrowałem się na powszechnym źródle informacji na temat przebiegu wykonania, jakim są logi. Logi są szeroko wykorzystywane, jednak na drodze do ich efektywnego wykorzystania w wykrywaniu anomalii i lokalizowania błędów stoi brak regularnej struktury. Chociaż metody przetwarzania języka naturalnego (ang. *Natural Language Processing* - NLP) są szeroko stosowane do przetwarzania treści logów, metody nienadzorowanego segmentowania wyrazów nie zyskały jeszcze większej popularności i użycia do sekwencji logów.

W doktoracie zaproponowałem metody segmentacji i analizy logów, które pozwalają na lepsze wykrywanie anomalii i dokładniejsze lokalizowanie źródeł błędów w oprogramowaniu. Eksperymenty były przeprowadzone w dużej części na zbiorach logów produkcyjnych pochodzących z prawdziwych błędów firmy Nokia. Ich zanonimizowana postać jak i algorytm są dostępne online.

Wrocław 25.09.2024

Wojciech Dobrowolski