

Projekt zespołowy

Generowanie słownego podsumowania fragmentów kodu źródłowego w języku C

Opiekunowie: prof. dr hab. inż. Anna Fabijańska, dr inż. Piotr Duch

Obecnie temat automatycznego generowania kodu źródłowego [1] oraz jego tekstowego podsumowania [2] bardzo szybko zyskuje na popularności. Ma to miejsce za sprawą popularnych rozwiązań sztucznej inteligencji wbudowanych w zintegrowane środowiska programistyczne, które dostarczają użytkownikom różnorodnych narzędzi do generowania, optymalizacji oraz poprawiania fragmentów kodu źródłowego. W przypadku pracy w zespołach, kiedy wielu programistów pracuje nad jednym projektem niezbędne jest dokumentowanie pisanego kodu, co jest czasochłonne i monotonne. Przygotowanie algorytmu, który zautomatyzuje ten proces pozwoli zaoszczędzić czas programiście i przeznaczyć go na inne zadania.

Obecnie istniejące rozwiązania dedykowane problemowi generowania i podsumowywania kodu źródłowego wykorzystują modele rekurencyjnych sieci neuronowych oraz modele typu transformer [3][4]. Przy czym, zadanie generowania kodu źródłowego było szerzej rozważane niż problem generowania słownego podsumowania kodu źródłowego.

W związku z powyższym, celem projektu jest opracowanie algorytmu, który na podstawie fragmentów kodów źródłowych w języku C przygotowuje opis funkcjonalności kodu w języku naturalnym. W projekcie można wykorzystać zarówno zewnętrzne bazy danych dedykowane problemowi (np. CodeNet), jak również zasoby zgromadzone w systemie Dante.

- [1] Le, Triet HM, Hao Chen, and Muhammad Ali Babar. "Deep learning for source code modeling and generation: Models, applications, and challenges." *ACM Computing Surveys (CSUR)* 53.3 (2020): 1-38.
- [2] McBurney, Paul W., and Collin McMillan. "Automatic documentation generation via source code summarization of method context." *Proceedings of the 22nd International Conference on Program Comprehension*. 2014.
- [3] MacNeil, Stephen, et al. "Generating diverse code explanations using the gpt-3 large language model." *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 2*. 2022.
- [4] Sarsa, Sami, et al. "Automatic generation of programming exercises and code explanations using large language models." *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1*. 2022.