Project Topic 1: Comments in GitHub Projects

Comments are important to understand and maintain source code. The goal of this project is to compare commenting styles in Java and Python projects by analyzing comments in Java and Python projects on GitHub.

Empirical Method: **Data Mining Study - Mining Software Repositories**

Work Description

1. Define research questions and suitable metrics to compare comments in Java and Python projects.
2. Develop an analysis strategy based on GitHub data with Python.
3. Implement a data extraction and measurement strategy for GitHub data.
4. Perform the data extraction.
5. Perform the data analysis and visualization.
6. Prepare a written report and a presentation as indicated in the description of Assignment 8.

References

Bissyandé, T. F., Thung, F., Lo, D., Jiang, L., & Réveillere, L. (2013). *Popularity, interoperability, and impact of programming languages in 100,000 open source projects*. In Computer Software and Applications Conference (COMPSAC), 2013 IEEE 37th Annual, 303-312. IEEE.

Guzman, E., Azócar, D., & Li, Y. (2014). *Sentiment analysis of commit comments in GitHub: an empirical study*. In Proceedings of the 11th Working Conference on Mining Software Repositories, 352-355. ACM.

Vasilescu, B., Filkov, V., & Serebrenik, A. (2013). *StackOverflow and GitHub: Associations between software development and crowdsourced knowledge*. In Social computing (SocialCom), 2013 international conference on, 188-195. IEEE.

GitHub Archive: https://www.githubarchive.org/

Sheoran, J., Blincoe, K., Kalliamvakou, E., Damian, D., & Ell, J. (2014). *Understanding watchers on GitHub*. In Proceedings of the 11th Working Conference on Mining Software Repositories, 336-339. ACM.