## 1. Understanding Language Selection in Multi-language Software Projects on GitHub

Li, Wen (1); Meng, Na (1); Li, Li (1); Cai, Haipeng (1)
**Source:** *Proceedings - International Conference on Software Engineering*, p 256-257, May 2021, *Proceedings - 2021 IEEE/ACM 43rd International Conference on Software Engineering: Companion Proceedings, ICSE-Companion 2021*; **ISSN:** 02705257; **ISBN-13:** 9781665412193; **DOI:** 10.1109/ICSE-Companion52605.2021.00119; **Conference:** 43rd IEEE/ACM International Conference on Software Engineering: Companion, ICSE-Companion 2021, May 25, 2021 - May 28, 2021; **Publisher:** IEEE Computer Society
**Author affiliation:** (1) Washington State University, United States
**Abstract:** There are hundreds of programming languages available for software development today. As a result, modern software is increasingly developed in multiple languages. In this context, there is an urgent need for automated tools for multi-language software quality assurance. To that end, it is useful to first understand how languages are chosen by developers in multi-language software projects. One intuitive perspective towards the understanding would be to explore the potential functionality relevance of those choices. With a plethora of publicly hosted multi-language software projects available on GitHub, we were able to obtain thousands of popular, relevant repositories across 10 years from 2010 to 2019 to enable the exploration. We start by estimating the functionality domain of each project through topic modeling, followed by studying the statistical correlation between these domains and language selection over all the sample projects through association mining. We proceed with an evolutionary characterization of these projects to provide a longitudinal view of how the association has changed over the years. Our findings offer useful insights into the rationale behind developers' choices of language combinations in multi-language software construction. © 2021 IEEE. (8 refs)
**Main heading:** Software design
**Controlled terms:** Application programs - Computer software selection and evaluation - Modeling languages - Quality assurance
**Uncontrolled terms:** Automated tools - Evolution - Functionality relevance - Language selection - Multi languages - Multi-language software - Multiple languages - Potential functionality - Software project - Software quality assurance
**Classification Code:** 723 Computer Software, Data Handling and Applications - 723.1 Computer Programming - 723.5 Computer Applications - 913.3 Quality Assurance and Control
**Database:** Compendex
**Data Provider:** Engineering Village