先端科学技術研究科　修士論文要旨

所属研究室 (主指導教員)	ソフトウェア工学 (松本　健一　(教授))		
学籍番号	2311320	提出日	令和 7年 1月 21日
学生氏名	田中　温人		
論文題目	Identifying Versions of Reused Libraries through Directory Comparison Using HyperLogLog++		
要旨			

In software development, developers sometimes develop software by copying and reusing source code from existing repositories. Since this practice has challenges such as vulnerabilities and license issues, various code clone detection methods have been proposed. In identifying versions of reused libraries, file-level comparisons have been employed. However, this approach is computationally expensive due to the large number of comparisons required. To overcome this limitation, this study proposes a method for identifying versions of reused libraries based on directory-level comparisons. The method summarizes the source code and documents as a HyperLogLog++ Sketch, a probabilistic data structure that calculates cardinalities. By comparing the target sketch with the sketches of each version, the method efficiently identifies the corresponding version. This approach does not require prior knowledge of directory structures, unlike previous studies that relied on narrowing down files. Furthermore, the number of hash functions used has been reduced from 1024 to just one. The method was evaluated using three libraries accurately recorded in Firefox, achieving 100%, 96%, and 100% precision, respectively. These results demonstrate that the method not only improves efficiency but also maintains high accuracy.