Parametrized String Matching Implementation for Software Plagiarism Check : *CopyDog*

Amit Tomar (MT2013008)
Siddhesh Dosi (MT2013150)
Srinivas R. Vaidya (MT2013152)

International Institute of Information Technology, Bangalore

Sep 24, 2014

Objective

- This project aims at developing a Parametrized String Matching Implementation for Software Plagiarism Check, that given a collection of files which contain code in some programming language, will show a set of possible duplications of parts of the code among these.
- Comparing pieces of software will require discounting comments (optional and language dependent), extra/blank lines and spaces, variable renaming etc.
- The theory of parametrized string matching will be used to implementation this project.
- System will have an easy-to-use UI for selecting files/folders and shall report the plagiarism related information (matches found) in the UI in a nice manner.

Functionality

- Plagiarism check among two given files with direct *copy-paste*.
- Plagiarism check among two given files with variable renaming after *copy-paste* [for the *smarter* ones].
- Pairwise plagiarism check among all the files in a given folder.
- Ignoring a code snippet for plagiarism check.
- Checking over the internet for plagiarism.

Trie data structure

- Quering multiple strings.
- Why do we need one - Dictionary with hash table ?
- Construction of Tries.
- Space-Time complexity of Tries.
- Applications - Dictionary, Phonebook, Auto complete, Spell checking.

Suffix Tree

- Quering single string.
- Searching a substring in a given string.
- Problem with brute force ?
- Construction of suffix tree.
- Space-Time complexity of suffix tree.
- Algorithm of the year - Knuth.

CopyDog

- Problems : White spaces, variable renaming.
- Python - Abstract syntax tree, P-Code.
- Demo of *CopyDog*.
- Pending issues - Internet and ignoring strings.
- Results with actual usage.

References

1 Peter Vamplew, Julian Dermoudy, *An Anti-Plagiarism Editor for Software Development Courses*, Proceeding ACE-05 Proceedings of the 7th Australasian conference on Computing education - Volume 42 Pages 83- 90 , 2005.