Plagiarism Detection Tool

Harsh Gupta   Alankrit Chona

Indian Institute Of Technology, Delhi

April 11, 2015

Project Facilitator
Huzur Saran,
Professor,
Department of Computer Science & Engineering,
Indian Institute of Technology, Delhi

Objective

- Identify similarity of code written in a variety of languages.

MOSS:Measure Of Software Similarity

- "We give an upper bound on the performance of winnowing, expressed as a trade-off between the number of fingerprints that must be selected and the shortest match that we are guaranteed to detect"

- "The service currently uses robust winnowing, which is more efficient and scalable than previous algorithms we have tried."

Activity in this Field

XPlag

- "To detect plagiarism involving multiple languages using intermediate program code produced by a compiler suite."
- "Relies on the components of an existing generic compiler suite."

Proposed Approach

FRONT END
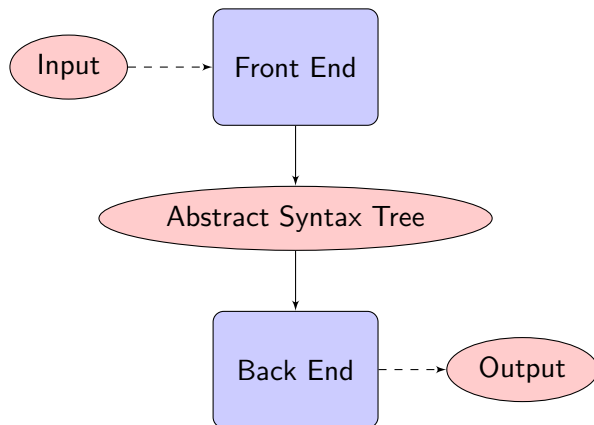
- Language dependent.
- Takes input as the source code.
- Creates an intermediate abstract syntax tree.

BACK END

- Independent of input language.
- Uses the Longest Common Subsequence Algorithm(LCS).
- Uses a similarity metric that can be configured.

OUR EXPERIENCE SO FAR

- Analyzed source code plagiarism dispute (Delhi High Court Case)
- Problems faced:
  - Unrealistic runtimes
  - High rate of false positives

A NEW APPROACH

- Use of LCS algorithm.
- Retaining the syntactic structure throughout computation.
- No loss of information by selection criterion (as in the case of MOSS and others).
- Configurable Similarity Metric.

Phases of the project

Timeline

1. **Initial Phase:** Development of tool for foxpro language.
2. **Phase 1:** Implementation of algorithm in easily extensible manner.
3. **Phase 2:** Improvement of the user-interface.

Budget

No financial requirements are anticipated.

Thank You