

ADAPTIVE SOURCE CODE FILES CLASSIFICATION USING COMPUTATIONAL INTELLIGENCE TECHNIQUES

Ioannis Komporakis, Giorgos Papadourakis, Spyros Karafyllakis, Ioannis Deligiannis
komporakis@gmail.com, papadour@cs.teicrete.gr, spirosk08@gmail.com, i_-_s@outlook.com

Department of Informatics Engineering
Technological Educational Institute of Crete
Heraklion, Crete 71500, Greece

STATEMENT OF THE **PROBLEM** TO BE ADDRESSED

- The most popular source code management systems use **file extension** to determine programming language
- Absence of textual analysis is a **big limitation**.
- It is supposed that all source code files will always have an **extension**.
- Some proposed solutions are using:
 - Support Vector Machine (SVM) Classifiers (Ugurel et al. 2002)
 - Fuzzy logic (Lerthathairat et al., 2011)
 - Evolutionary Algorithms (Alvares et al., 2014)

INPUTS

Frequency of predefined keywords



36 COMMON KEYWORDS

abstract
break
byte
case
catch
char
class
const
...



14 JAVA UNIQUE KEYWORDS

assert
boolean
extends
final
implements
import
instanceof
native
Package
...



41 C# UNIQUE KEYWORDS

as
base
bool
checked
decimal
delegate
event
explicit
....

DATASET COLLECTION

Process

- Developed a script file for the production of the dataset
- 500 Java source Files and 500 C# source files gathered from GitHub repository in two separate folders
- The script analyzed each file separately
 - Comment removal
 - Keyword isolation
 - Keyword frequency count

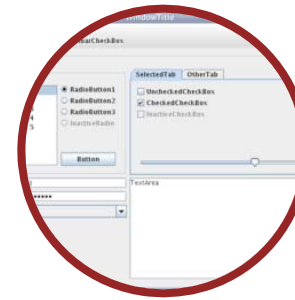
TECHNOLOGIES AND TOOLS



JAVA SE

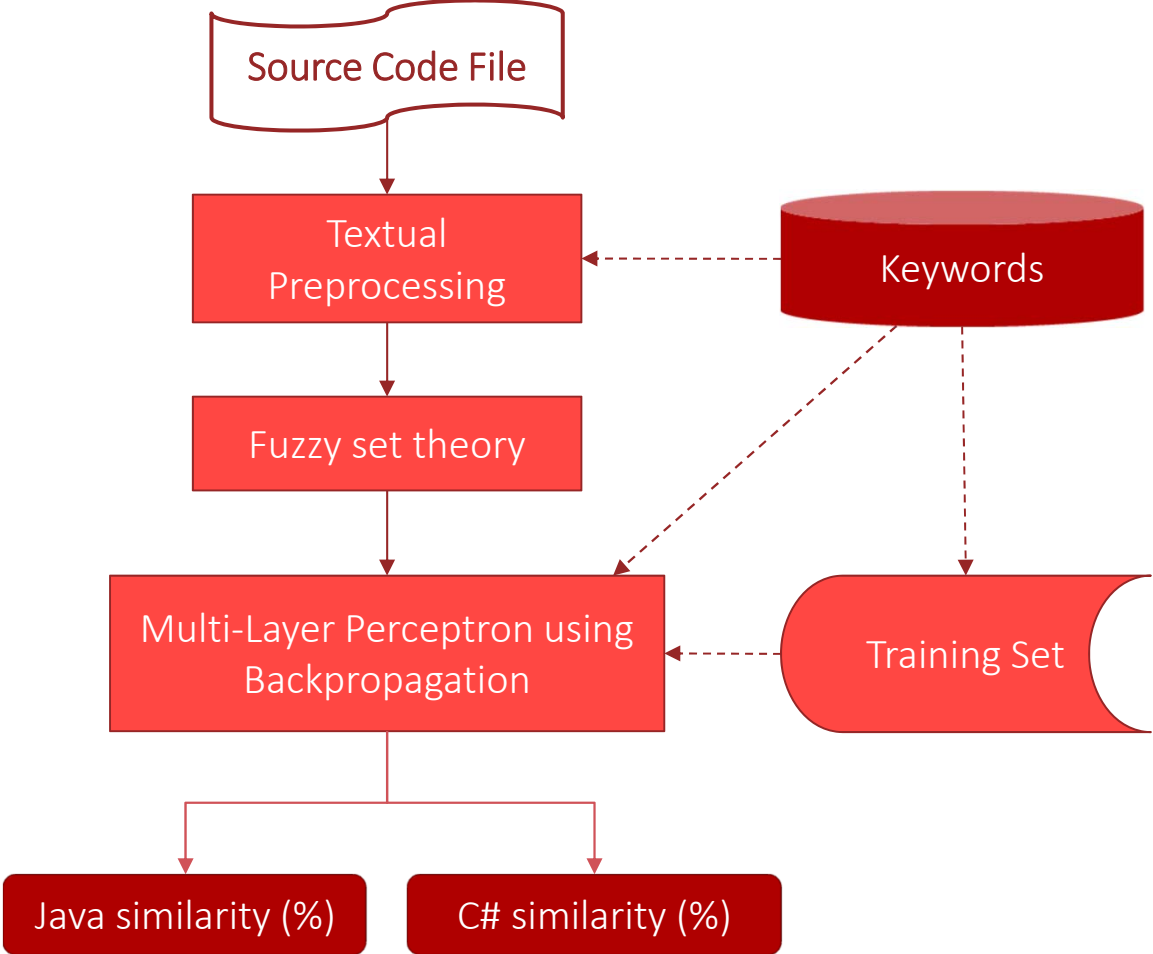


NEUROPH JAVA
FRAMEWORK

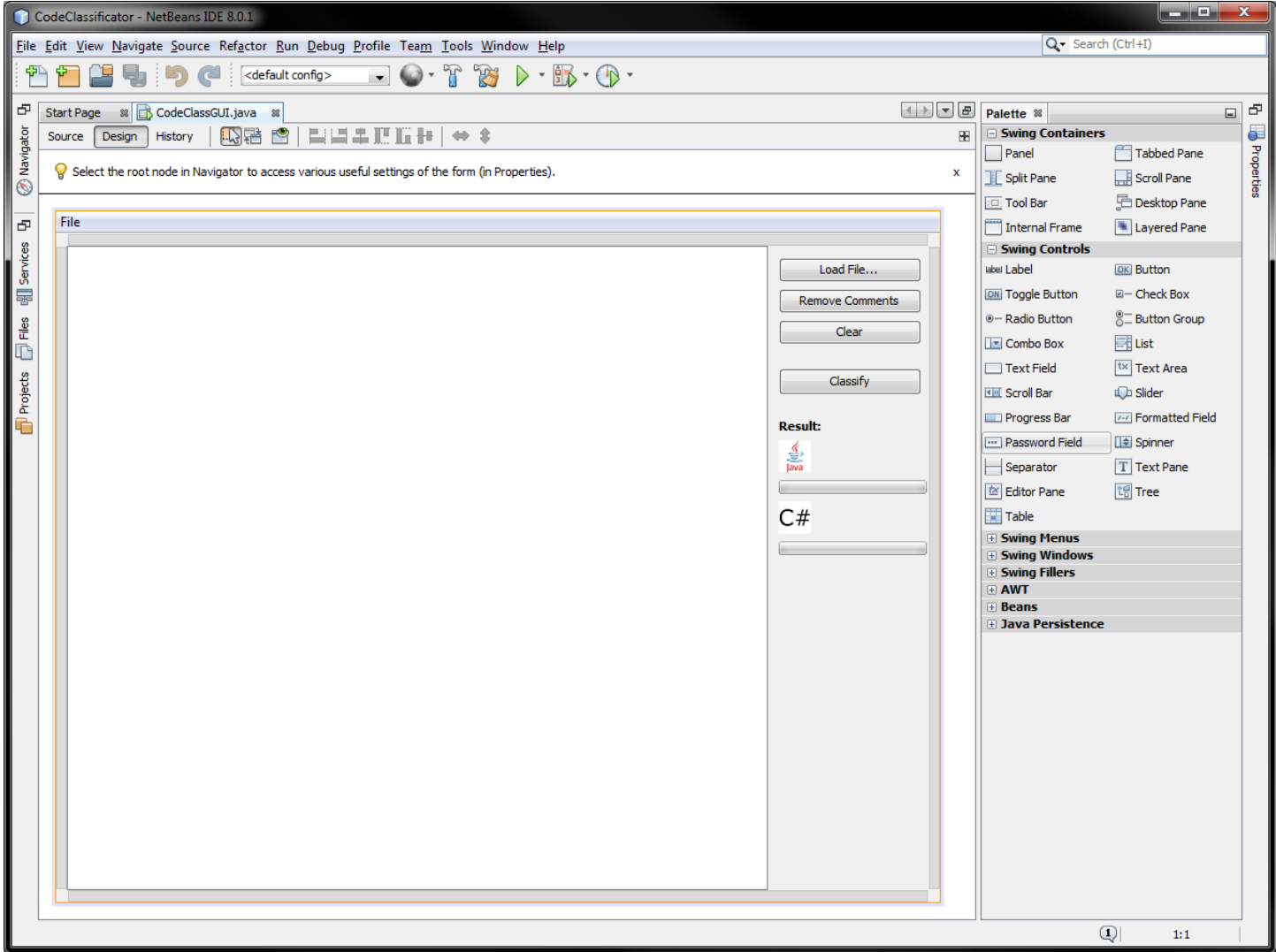


JAVA SWING GUI TOOLKIT

WORKFLOW



GUI DESIGN



EVALUATION RESULTS

#	File extension	Similarity to Java	Similarity to C#
1	.java	95.36%	5.01%
2	.java	42.26%	58.07%
3	.java	94.90%	5.48%
4	.java	78.25%	21.62%
5	.java	99.88%	0.12%
6	.cs	1.36%	98.66%
7	.cs	0.00%	100.00%
8	.cs	3.10%	96.97%
9	.cs	0.09%	99.91%
10	.cs	0.77%	99.19%

REFERENCES

- Alvares, Marcos, Tshilidzi Marwala, and Fernando Buarque de Lima Neto. "Application of computational intelligence for Source Code classification." Evolutionary Computation (CEC), 2014 IEEE Congress on. IEEE, 2014.
- CertPal blog:
<http://www.certpal.com/blogs/2010/07/java-neuroph-tutorial-the-code-classifier/>
- Neuroph Framework: <http://neuroph.sourceforge.net/>

END OF PRESENTATION

Questions?