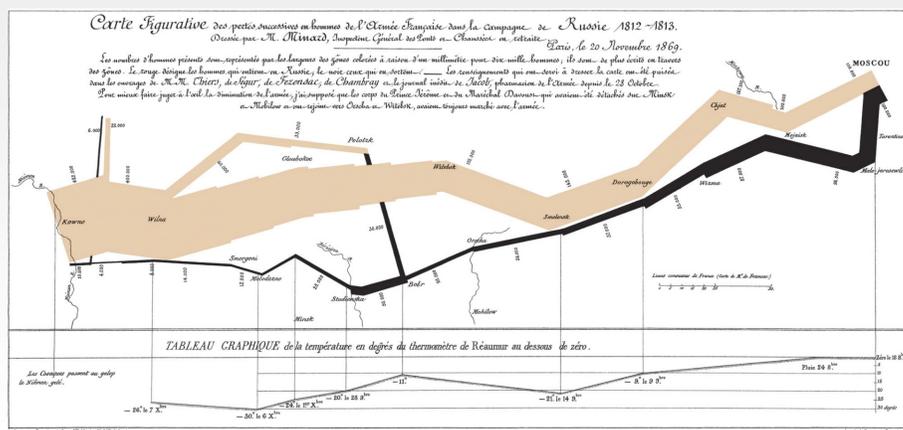


Visualization of Plagiarism

Adrian Garcia
Supervisor Prof. A. Jung
School of Computer Science University of Birmingham
msc38adg@cs.bham.ac.uk

Introduction

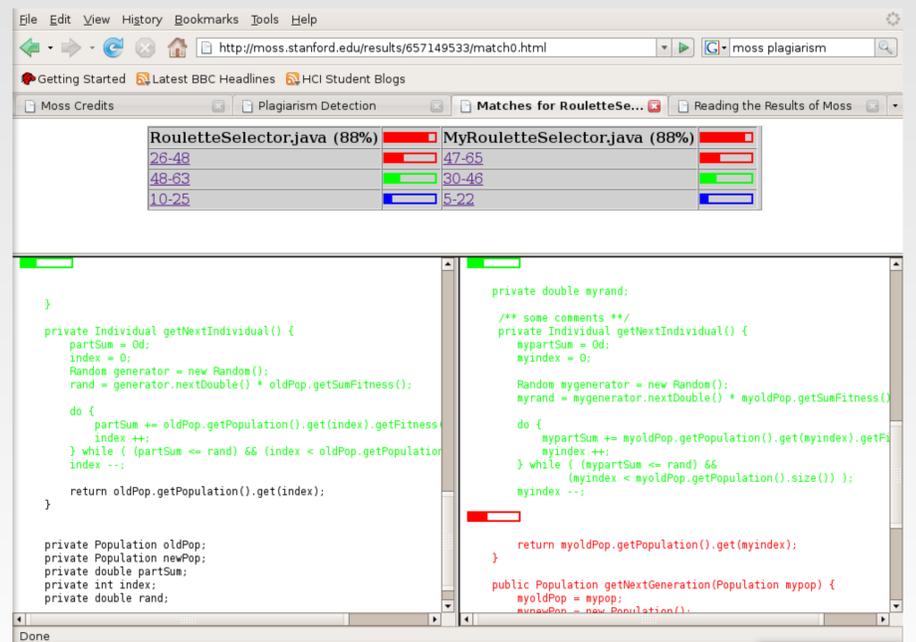
Plagiarism is passing off the work or ideas of others as though it were one's own and is often reported to be a growing problem in academia. Visualization can be seen in terms of graphical representations of concepts or information that make insights into complexity easier. For example in Minard's 1861 map of Napoleon's march and retreat on Moscow the death rates, temperature and geographic locations are all represented in a way that is more understandable than representing the same information in a tabular way.



When visualization techniques are used with computing then interaction can be introduced. This allows better exploration of the domain in question and possibility of alternative presentations.

State of the Art

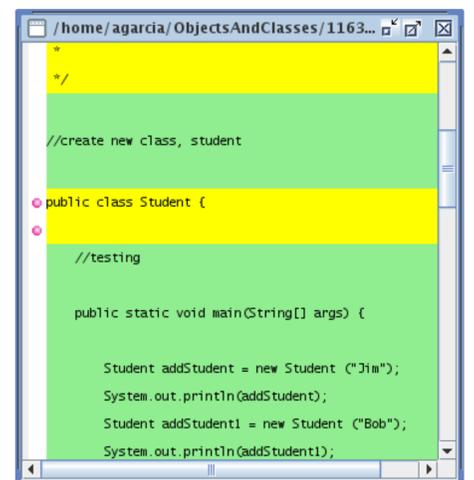
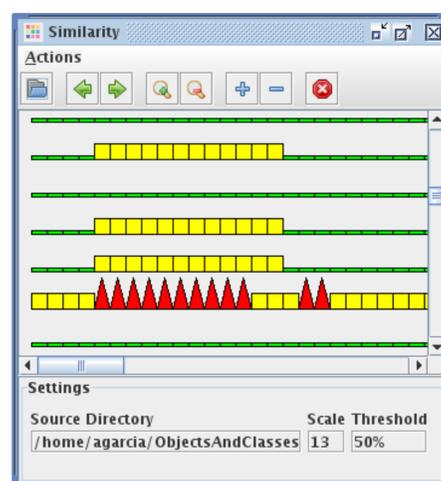
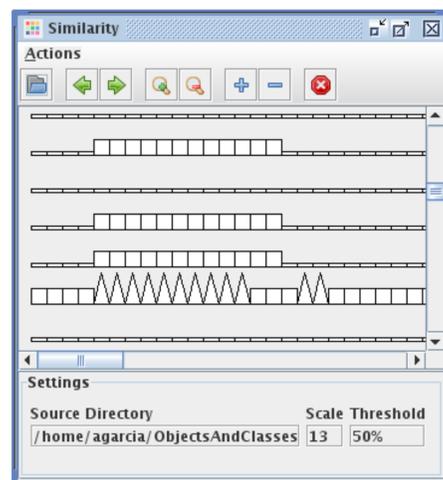
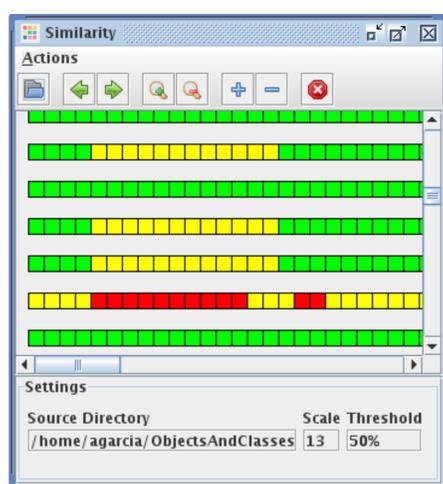
MOSS, a plagiarism detection system.



Current plagiarism detectors are effective in finding similarity between plagiarised documents and their sources. However the way that they present their results is often static (non interactive) and is limited in the visual techniques used. For example MOSS uses colour and bar charts to indicate similarity.

An Alternative Visualization System

As part of this project a system was developed to provide a framework for alternative visualizations of the relationships between plagiarized text and its source. A model was developed based on diff output that inferred similarity for individual lines of text. The system allows user interaction to control the views based on the currently selected line within a file. The similarity relationships can then be explored by following graphical markers to corresponding groupings of similar lines in other files.



Discussion

The system is designed to allow alternative views to be added and there are plans to develop visualizations based on Gestalt principles such as proximity, symmetry and chronology. Integration with other detection systems or applying the visualizations to alternative domains such as DNA sequencing are also considered to be further directions for the project.