Review of Bachelor's Thesis

Student: Pružina Tomáš
Title: Source Code Authorship Attribution (id 17751)
Reviewer: Dytrych Jaroslav, Ing., UPGM FIT VUT

1.   Assignment complexity   **average assignment**
The assignment could be considered as more difficult, but I think it is rather average in the student's conception.

2.   Completeness of assignment requirements   **assignment fulfilled**

3.   Length of technical report   **in usual extent**
The technical report has 28 standard pages and approximately 4.5 pages of images. Content is, in my view, adequate, although development and experiments could be described in more detail.

4.   Presentation level of technical report   **75 p. (C)**
The technical report has logical structure and the individual chapters follow up each other. However, division to the subsections could be made better so that each subsection contains some text and not just subsections (2 headings below each other). The third level of numbered headings is, in my opinion, unnecessary. I have an impression that the seventh chapter is unfinished. Description of experiments, in my opinion, should be more detailed.

5.   Formal aspects of technical report   **76 p. (C)**
Because the work is written in English, I am unable to fully assess the language quality of the work. From my point of view, the number of errors is acceptable, despite that there are typos and there are some missing spaces. Typographically, the work could have been processed better so that indefinite articles were not present at the ends of lines etc.

6.   Literature usage   **95 p. (A)**
The choice of literature is, in my opinion, appropriate and covers the solved problematic. The most of the literature is compound from scientific articles.

7.   Implementation results   **74 p. (C)**
The realisation output is fully functional and corresponds to the description in the technical report.
As a weakness, I find it more difficult to run and use it (individual scripts do not offer a help and output is not user-friendly).

8.   Utilizability of results
If we have enough source files to train the model, the created program can be used to determine the authorship of source codes. In the future, a system for detecting plagiarism can be built on it.

9.   Questions for defence
-

10.   Total assessment   **76 p. good (C)**
Average assignment was fulfilled in an average manner. Work with literature was above average. The technical report is clear and easy to understand, but the description of the experiments gives an incomplete impression. The realisation output is fully functional but not user-friendly. I propose an assesment by grade C.

In Brno 1. June 2017

....
signature