



Prague, October 3, 2008

The Board of Doctoral Study
Faculty of Mathematics and Physics

Report on the PhD Thesis “Syllable-based Compression

by Jan Lánský

This thesis summarizes works covering the author's study of syllable-based compression. The research described in the thesis continues the results achieved in his master thesis.

In Chapter 1 the author shortly specifies the topics of his research, i.e. a compression of: small text files, large text files, XML data, and texts based on a large alphabet. Chapters 5-8 are structured in the same way. In his thesis the author focuses on the lossless compression. He introduces a reader to the topic in Chapter 2 in a relatively short overview of representative methods of this area. The methods mentioned are today used in many software packages of compression programs working on characters or words of texts. But what is less usual is to use the methods on syllables. In Chapter 3 the author presents shortly related works. Their list is short since the topic was investigated only rarely in the past.

In Chapter 4 there is an approach to model a text by syllables considering vowels and consonants of the languages considered. In the rest of thesis the author describes details of syllable-based approach applied on situations stated in Chapter 1. Experiments accompanying all these results use large datasets, which in fact increases their plausibility. In Conclusions the cases for which syllable compression is the appropriate choice are summarized.

The author uses partly a formal approach in his thesis; however his main research method is an experiment, which is relevant in this case. The thesis contains a lot of interesting results and a deep analysis of the performed experiments. It is worthy of mention that Jan Lánský was given the Bolzano award in 2007 for a collection of works about syllable compression. Also his work with students has been exemplary during his PhD study. He involved them to the research both on the level of master theses and software projects. Some of them became co-authors of his papers.

The results presented in the thesis have been published in proceedings papers of representative international conferences, like e.g. IEEE Data Compression Conference, SOFSEM and ICADIWT. These papers are also cited internationally in a number of scientific works. The DBLP Bibliography Server records 11 items of which Lánský is a co-author. Consequently, I recommend that the candidate be awarded the Doctor degree.

Prof. Jan
supervis

