



MATEMATICKO-FYZIKÁLNÍ
FAKULTA
Univerzita Karlova

Diploma thesis review

Review of opponent

Author: Daria Bilan
Thesis title: Distinguishing Pairs of Words Using Finite Automata
Advisor: prof. Mgr. Michal Koucký, Ph.D.
Opponent: doc. Mgr. Robert Šámal, Ph.D.
E-mail: samal@iuuk.mff.cuni.cz
Year: 2023

Given two words, each with at most n letters, how large finite automaton is needed to distinguish them? This question was posed by Goralčík and Koubek in 1986 and it received some attention since then. Despite that, the known bounds are far apart and this, together with easy to state formulation of the problem is a recipe for an interesting diploma thesis.

Daria Bilan started her thesis by presenting main known results about the problem in Chapters 1–3. This introduction is well organized and easy to read.

In Chapter 4 the author presents the concept of discerning sets. It was a basis of a failed attempt to solve the problem under study that was published as a technical result in the 80s (and this attempt is clearly explained as well). The curious properties of these sets are studied in some details, also using computer experiments. I found the results interesting and worthwhile. Still, it feels a bit unfortunate to present Lemma 88, that quenches this approach to finding discerning automata, only at the end of the chapter.

Chapter 5 deals with random automata, in particular so-called random permutation automata. The author presents some known results and adds many promising experiments.

The thesis is well-written, I found it easy to read. The presented results are correct, with one exception given below.

- In Section 4.1.2, the author claims $\lim_{n \rightarrow \infty} \frac{\text{lcm}(1, \dots, n)}{e^n} = 1$. I don't think this is true, the sequence even appears unbounded. The reference given claims only

$$\lim_{n \rightarrow \infty} \frac{\log \text{lcm}(1, \dots, n)}{n} = 1.$$

On the other hand, it certainly is true that $\text{lcm}(1, \dots, n) \leq e^{2n}$, so the argument does not break down.

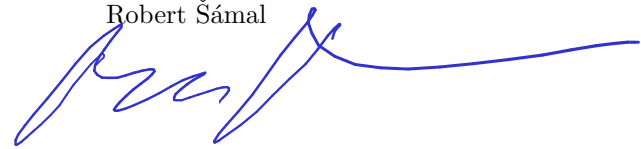
- In proof of Lemma 31 – it is claimed to be easy to construct an automaton accepting words ending in a given string. It would be better to explain how to do this.
- Figure 2.5 – a 0-arrow out of 110 is missing. The 1-arrow out of 11 is missing/wrong.
- Lemma 32, in the definition of N , k should be d .
- Beginning of Chapter 4, page 31: “which solution provides a valid upper bound” is Czenlish.
- Lemma 62: “W.L.O.G.” should be written in full.
- Lemma 63: sets of even **length** – “size” would be better word.

Compared to other diploma theses I have read, this is a very modest amount of comments. Moreover, they do not impede reading the big story of the thesis. Altogether I consider this to be a worthwhile contribution that will perhaps revive interest in this problem and lead to its resolution.

Recommendation: In accordance with the above, I recommend to accept the thesis as a diploma thesis and to classify it by the best grade.

In Liberec on Sep 1, 2023

Robert Šámal

A handwritten signature in blue ink, consisting of several loops and a long horizontal stroke extending to the right.