

Hodnocení vedoucího bakalářské práce

Autor/autorka práce: **Petr Hlaváč**

Název práce: **Klasifikace textových dokumentů pomocí neuronových sítí**

Práce se zabývá automatickou klasifikací dokumentů. Hlavní důraz je kladen na porovnání různých způsobů předzpracování textových dat a posouzení jejich vlivu na výsledek klasifikace. Pro klasifikaci je využito několik typů neuronových sítí. Metody jsou testovány na standardních českých a anglických datech (korpusy Czech Text Document Corpus a Reuters RCV1-v2).

V práci je nejprve definována klasifikační úloha a jsou popsány způsoby předzpracování, které budou využity. Teoretická část se dále věnuje popisu datových sad a obecnému popisu neuronových sítí, které se pro klasifikaci textových dat obvykle používají. Praktická část nejprve uvádí konkrétní architektury neuronových sítí, které jsou v práci využity. Rovněž je zde popsána implementace sítí pomocí knihoven Tensorflow a Keras a postup pro spouštění výpočtů na Metacentru. Metacentrum bylo využito z důvodu vysoké výpočetní náročnosti úloh.

Kapitola dosažené výsledky přináší rozsáhlé srovnání výsledků při použití stemmingu, lematizace a sofistikovanější tokenizace v kombinaci s různými neuronovými sítěmi.

Důležitou částí práce byla příprava dat, kdy bylo nutné zejména anglický korpus upravit a doplnit některé informace tak, aby odpovídal datům použitým v literatuře. Pomocí několika nástrojů pak byla data předzpracována a připravena v několika verzích pro další testování. Bylo prokázáno, že díky předzpracování je možné dosáhnout lepších výsledků, než bylo dosud publikováno v literatuře.

Práce je přehledně členěna. Text práce je kvalitní s minimálním množstvím překlepů .

Přístup k řešení práce byl aktivní a práce byla průběžně konzultována s vedoucím.

Zadání bylo splněno bez výhrad

Navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.

V Plzni 8.8.2018

Ing. Ladislav Lenc, Ph.D.

