

# Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Vladan Glončák

**Název práce** Using Syntactic Features for Opinion Target Identification

**Rok odevzdání** 2019

**Studijní program** Informatika **Studijní obor** Matematická lingvistika

**Autor posudku** Jan Hajič jr. **Role** vedoucí

**Pracoviště** Ústav formální a aplikované lingvistiky

## Text posudku:

Cílem předložené práce je prozkoumat vliv syntaktických rysů (features) na identifikaci hodnocených entit v evaluativních textech. Jedná se tak o standardní experimentální diplomovou práci na poli zpracování přirozeného jazyka. Zjištěním práce práce je, že tyto rysy zlepšení nepřináší. Řešitel trénuje řadu různých sekvenčních modelů, jak tradiční Conditional Random Fields, tak aktuálně nejlepších LSTM, a porovnává výsledky při zahrnutí syntaktických informace s výsledky bez ní stejně jako tato úloha byla vyhodnocována na soutěži SemEval. Syntaktické rysy extrahuje state-of-the-art systémem UDPipe, omezení malých dat částečně obchází používáním předtrénovaných slovních reprezentací Word2vec z velkých korpusů. Přidává i další externí zdroje jako slovníky evaluativních výrazů. Pokusy provádí na anglických i českých textech, a provádí také vyhodnocení o něco více „optimistickým” způsobem, kde se však vliv syntaktických rysů také neukázal.

Skutečnost, že práce přinesla v podstatě negativní výsledky (žádný konzistentní zjitelný vliv syntaktických rysů), nepovažují v principu za nedostatek diplomové práce. Řešitel pracoval metodologicky správně, bohužel s velmi omezenými (a zároveň jedinými dostupnými) daty a nedostatkem zdrojů na vytváření rozsáhlejších dat, a jednalo se o otevřenou výzkumnou otázku, kdy skutečně nebylo zřejmé, zda syntaktické rysy nějaký (kladný) vliv mít mohou. Provedené pokusy nenaznačují, že by se při dalších pokusech na dostupných datasetech ještě nějaký významný vliv objevil.

Nejsilnější stránkou diplomové práce je samotný text. Nejen že je psaný prakticky bezchybnou angličtinou, zároveň je velmi precizní a úsporný, s vysokou a rovnoměrnou informační hustotou. Všechny pokusy jsou popsány replikovatelně: část pokusné konfigurace je pouze v konfiguračních souborech v elektronické příloze, nicméně z textu práce je na ně v příslušném místě odkazováno (sekce 6.3). Nejslabší stránka práce je pravděpodobně analýza výsledků, nicméně to je dáno skutečností, že prakticky není co analyzovat. Detailní výčet změn, které syntaktické rysy přinesly,

v přílohách práce ukazuje, že diplomant analýzu nezanedbal. Relativně malý rozsah je způsoben vysokou informační hustotou textu a tím, že nejsou zbytečně obsírně popisovány standardní modely strojového učení. I přes zdánlivě neuspokojivý výsledek vlastně nevidím nic zásadního, co by se bývalo práci dalo vytknout.

Po formální stránce je práce taktéž v pořádku. Citace a bibliografické údaje jsou použity v souladu s oborovými zvyklostmi, bibliografie je poměrně bohatá.

Dotazy k obhajobě nemám.

**Práci doporučuji k obhajobě.**

**Práci nenavrhuji na zvláštní ocenění.**

V Praze dne 26. 8. 2019

Podpis: