



FAKULTA
APLIKOVANÝCH VĚD
ZÁPADOČESKÉ
UNIVERZITY
V PLZNI

Ing. Ondřej Pražák
Katedra informatiky a výpočetní techniky
Fakulta aplikovaných věd
Technická 8, CZ-30614 Plzeň
Tel.: (+420) 377 63 2151
E-mail: ondfa@ntis.zcu.cz

Posudek oponenta diplomové práce

Bc. Michal Tušl
Vícejazyčná sémantická podobnost textů

Michal Tušl se ve své diplomové práci zabývá metodami učení bez učitele pro určování sémantické podobnosti textů a metodami lineární transformace pro přenos sémantiky mezi jazyky. Na diplomovou práci se jedná o velice náročné téma.

Práce se velice dobře čte, je psaná velmi věcně a stručně, ale neopomíjí žádné důležité informace. Práce má dobrou logickou strukturu, je doplněna velkým množstvím užitečných a názorných obrázků. Z textu je patrné, že autor výborně porozuměl dané problematice. Po obsahové stránce práci vůbec nic nechybí a jsem s ní nadmíru spokojen. U popisu algoritmů pro vícejazyčnost velice oceňuji přehledná schématická znázornění, jak algoritmy fungují. V textu jsem našel pár faktických chyb (například špatné dimenze matic u singulárního rozkladu) a nepřesností (definice one-hot vektoru), ale jejich množství je malé. V práci se vyskytuje pouze malé množství chyb a překlepů. Jejich formální úroveň je výborná. Grafickému zpracování dokumentu nelze téměř nic vytknout (snad jen pár stránek s příliš krátkým obsahem).

Diplomant ve své práci cituje odpovídající množství relevantních zdrojů. Jeho práci s literaturou nelze nic vytknout.

Implementovaný software napsaný v jazyce Java je poměrně rozsáhlý. Oceňuji, že si autor značnou část metod implementoval sám, což dokazuje porozumění problematice. Návrh aplikace není ideální, aplikace obsahuje poměrně velké množství duplicitního kódu. Zdrojové kódy obsahují poměrně málo komentářů, ale jsou v celku srozumitelné.

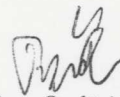
Práce představuje nové metody a z vědeckého pohledu je bezesporu přínosem. Metodice experimentů nelze nic vytknout. Výsledky vypadají slibně a v postupu jejich měření nevidím žádný problém, takže je považuji za správné. Jedinou výtku mám k předzpracování textu, pro rekurentní neuronovou síť (Skip-thoughts). Odstraněním interpunkce, stopwords a dalších sebereme modelu kontextovou informaci, ze které by se mohl něco naučit. Tento problém by mohl být jedním z důvodů, proč Skip-thoughts funguje výrazně hůře, než ostatní metody.

Práce splňuje zadání ve všech bodech. Přes výše zmíněné výtky práci považuji za vynikající. Proto práci **doporučuji** k obhajobě a hodnotím klasifikačním stupněm

„výborně“.

Doplňující otázky:

1. Proč transformace Paragraph2Vec modelu funguje podstatně hůře, než ostatní metody?


Ing. Ondřej Pražák
(oponent DP)

V Plzni 3. června 2019

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
katedra informatiky a výpočetní techniky
①



**SOUHLASÍ
S ORIGINÁLEM**