

# Posudek oponenta diplomové práce

Autor/autorka práce: **David Steinberger**

Název práce: **Metody statistické sémantické analýzy**

## Obsah práce

Rozsah práce je v očekávaném rozmezí. V textu práce autor dostatečně popisuje existující řešení, specifikum statistické sémantické analýzy češtiny, použitá řešení a srovnání jejich výsledků. Text je přehledný a jeho struktura odpovídá očekávané struktuře diplomové práce. Použité rovnice jsou správné s výjimkou (2.8), kde namísto kritéria má být jeho gradient.

## Kvalita řešení a dosažených výsledků

Práce používá v současnosti nejúspěšnější metody, přičemž tyto metody jsou vhodným způsobem adaptovány pro použití na český jazyk. O funkčnosti návrhu svědčí prezentované výsledky, a to především zvýšení syntaktické přesnosti. Analýza dosažených výsledků není rozsáhlá, což je ovšem při použití neuronových sítí, které podobnou analýzu znemožňují, pochopitelné. Vlastnosti použitých programů, vlastnosti formátů dat i parametry datových sad jsou dostatečně dokumentovány. Dokumentace jednotlivých experimentů již tak propracována není a opakovat experimenty je možné jen s jistým úsilím.

## Formální úroveň

Typografická stránka práce je velmi dobrá. Stylistická stránka je na diplomovou práci nadprůměrná. Práce je dostatečně vybavena seznamem zkratk a přílohami s detaily použitých dat, implementace a výsledků.

## Práce s literaturou

Autor ke všem popisům existujících řešení připojuje relevantní citace aktuálních prací. I ostatní citace jsou relevantní. Práce neobsahuje citaci, kterou by bylo možné považovat za nevhodnou.

## Splnění zadání

Zadání bylo splněno bez výhrad.

## Doplňující informace k práci

K zvolenému způsobu řešení je třeba uvést, že pracnost vytváření databáze pro sémantickou analýzu nedovoluje vytvořit databázi, ze které by bylo možné vyčlenit trénovací množinu pro systémy strojového učení. Systémy sématické analýzy tedy vznikají z jiných dat jako víceméně vedlejší efekt, což značně ztěžuje návrh těchto metod. Dále, je třeba také zdůraznit, že jedním z cílů práce, který byl bezesbtyku splněn, byl návrh a zajištění tvorby české databáze. Tento výsledek není důležitý jen pro tuto práci, ale především dovoluje snadno v práci pokračovat.

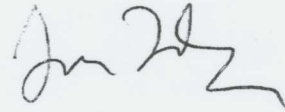
Dotazy k práci

- 1) Lze nějakým způsobem např. pomocí údajů v tabulkách 4.4 a 4.5 odhadnout, jaké sémantické přesnosti by mohl dosáhnout člověk nebo jakou maximální sémantickou přesnost je možné očekávat?
- 2) Jakým způsobem byly stanoveny počty neuronů v těch vrstvách, ve kterých je počet neuronů možné volit? Byly použity v literatuře běžné počty, nebo byly počty voleny experimentálně?
- 3) Byly kromě funkce softmax použity i jiné funkce než sigmoidální (logistické) funkce?

Navrhuji hodnocení známkou **v ý b o r n ě** a práci doporučuji k obhajobě.

V Plzni 12.8.2016

Ing. Jan Zelinka, Ph.D.



**SOUHLASÍ  
S ORIGINALÉM**



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

①