

Posudek tezí

Jméno a příjmení autora posudku: Mgr. Barbora Vidová Hladká, Ph.D.

Jméno a příjmení autora práce: Mgr. Rudolf Rosa

Název práce: Parsing Natural Language Sentences by Semi-supervised Methods

Teze doktorské práce Rudolfa Rosy se zaměřují na vybrané aspekty automatické syntaktické analýzy, tzv. parsingu. Konkrétně je diskutován a experimentálně testován delexikalizovaný závislostní parsing napříč jazyky s využitím monolingválních anotovaných korpusů. Struktura textu kopíruje zpracování vybraných aspektů: pro každý z nich je systematicky uveden komentovaný přehled relevantní literatury a scénář experimentů s popisem teoretických východisek, použitých dat a diskutovaných výsledků. Souvislost mezi jednotlivými aspekty je jasně popsána, stejně tak i plány na další experimenty jsou srozumitelné. Z tezí bohužel nevyplývá, které dílčí téma již bylo publikováno.

Není pochyb, že uvedená dílčí témata zapadají do tématu Rudolfova doktorského studia. Již ale není zřejmé, jak do tématu zapadají. V tomto smyslu postrádám v textu navrhovanou osnovu zpracování tématu. Samozřejmě, že dílčí body osnovy se mohou měnit, přesouvat atd. podle průběžných výsledků, ale směr odkud kam se jde by měl být *víceméně* známý. Tento nedostatek může být způsoben technickým nastavením pro text tezí, čímž myslím použití šablony, která se používá při publikaci konferenčních článků.

Chápu teze jako základ pro text doktorské práce, proto mám v tomto směru dvě obecná doporučení: uvádět více ilustračních příkladů a přehodnotit nastavení toho, co uvádět v poznámkách pod čarou a co začlenit přímo do textu (motivační úvahy, např. pozn. 3, mají své místo přímo v textu - viz *i cesta je cíl*).

Dotazy (uvádím je v pořadí, v jakém vznikaly při čtení tezí):

1. Jak ve zpracování Vašeho tématu chápete *částečně řízené metody*?
2. Zaměřujete se výhradně na závislostní parsing. Proč?
3. Pracujete s pražským a standfordským anotačním stylem. Bylo by dobré doplnit popis jejich podobností a rozdílů ilustračními příklady.
4. V předposledním odstavci sekce 2 je zmíněn *hyperparameter* – čtenář nemusí být s tímto termínem seznámen. Proto zvažte, jestli je na daném místě jeho užití nutné, nebo ho popište.
5. V parsingu se zaměřujete pouze na hrany stromu. Uvažujete i o zapojení analytických funkcí?
6. V experimentech používáte výhradně MST parser. Co očekáváte od použití jiných parserů?
7. V sekci 3.1 je uveden základní přehled příznaků, se kterými se pracujete. Jaké je jejich množství v datech, se kterými experimentujete?
8. Obr. 1 není pravděpodobně úplný.
9. Zabýval jste se při kombinaci parserů studiem jejich komplementarity, tj. jak různé chyby dělají?
10. Při analýze vlivu anotačního stylu na úspěšnost parsingu se zaměřujete na *adpositions*. Neuvádím český ekvivalent, protože v textu není tento jev vysvětlen a ani z vlastní řešerše literatury mi není význam tohoto jevu zřejmý. Nicméně objevuje se v kontextu popisu standfordského anotačního stylu.
11. Ve druhém odstavci úvodu sekce 4 je zmíněn *Prague tectogrammatical tree* – čtenář nemusí znát tectogramatické stromy (viz dotaz 4.)
12. Obsah sekce 4.2 by bylo dobré ilustrovat příkladem.
13. Při analýze výsledků se odvoláváte na podobnosti stromů v anotačních stylech. Toto by bylo dobré ilustrovat konkrétními příklady (i v návaznosti na dotaz č. 3)
14. Myslíte si, že na výsledky parsingu má vliv i původní anotační schéma korpusu daného jazyka, tj. schéma, které bylo harmonizováno.
15. V čem jsou harmonizované korpusy rozdílné?
16. Ovlivňují různé žánry v harmonizovaných korpusech úspěšnost parsingu?
17. Pro KL_{upos3} by bylo zajímavé vyzkoušet tzv. skip n-gramy (z posloupnosti slov/značek/... se vybírají n-gramy, které netvoří souvislý úsek posloupnosti)
18. Provedl jste nějakou analýzu nad vahami příznaků v natrénovaných modelech?

Na základě předložených tezí konstatuji, že se Rudolf Rosa ve svém doktorském studiu věnuje vysoce aktuálnímu tématu z oblasti počítačového zpracování jazyka a že je schopen tvůrčím způsobem řešit a prezentovat vědecká témata.

V Praze dne: 7. června 2015

Podpis: