

Univerzita Karlova
Matematicko-fyzikální fakulta

ZÁZNAM O PRŮBĚHU OBHAJOBY
DISERTAČNÍ PRÁCE

Název práce: Discovering the structure of natural language sentences by semi-supervised methods

Jazyk práce: Angličtina

Jméno studenta/studentky: Mgr. Rudolf Rosa

Studijní program: Informatika

Studijní obor: 4I3 Matematická lingvistika

Školitel:

doc. Ing. Zdeněk Žabokrtský, Ph.D.

Oponenti:

doc. RNDr. Aleš Horák, Ph.D.

prof. Jörg Tiedemann

Členové komise:

prof. RNDr. Jan Hajič, Dr., předseda (přítomen)

doc. RNDr. Markéta Lopatková, Ph.D. (přítomen)

Mgr. Barbora Vidová Hladká, Ph.D. (přítomen)

Ing. Alexandr Rosen, Ph.D. (nepřítomen, omluven)

Mgr. Jan Štěpánek, Ph.D. (přítomen)

doc. RNDr. Vladimír Petkevič, CSc. (přítomen)

doc. RNDr. Aleš Horák, Ph.D. (přítomen)

Datum obhajoby: 14. 6. 2018

Průběh obhajoby:

Obhajoba proběhla v angličtině ve čtvrtek 14. 6. 2018 od 10:00 hodin v budově MFF UK v místnosti S1, Malostranské nám. 25, Praha 1.

Předseda prof. RNDr. Jan Hajič, Dr. přivítal přítomné, představil členy komise a konstatoval, že komise je usnášeníschopná. Kromě komise byl přítomen i školitel, oba oponenti a řada hostů. Předseda představil studenta a jeho dosavadní studium. Konstatoval, že byly splněny všechny formální podmínky pro konání obhajoby a že nedošly žádné písemné připomínky mimo posudků oponentů. Student absolvoval zkoušku

Pokyny pro předsedy nebo místopředsedy komisi:

Práce v elektronické podobě musí být studentem vložena do SIS. Formulář vyplňte ve všech bodech v elektronické podobě. V bodě Členové komise se uvedou všichni členové komise a za jejich jména se uvede „(přítomen)“ nebo „(nepřítomen)“. Předseda nebo místopředseda komise je jejím členem. V bodě Průběh obhajoby by měly být uvedeny alespoň čtyři věty vystihující průběh obhajoby. Po vyplnění formuláře ho vytiskněte, dole formulář ještě vlastnoručně podepište a přiložte k zápisu o státní závěrečné zkoušce. Současně vložte formulář v elektronické podobě (bez vlastnoručního podpisu) do SIS.

z angličtiny, doktorandskou státní zkoušku a odevzdal disertační práci, životopis i seznam publikací. Přítomní dostali možnost nahlédnout do výtisků obhajované disertační práce.

Předseda shrnul životopis studenta, zmínil, že student studuje na MFF od bakalářského studia, strávil půl roku v Googlu na stáži a půl měsíce v Edinburghu. Jeho publikační historie je bohatá, obsahuje 31 publikací a na Google Scholar obdržel již přes 300 citací.

Školitel doc. Ing. Zdeněk Žabokrtský, Ph.D. shrnul studentovo studium a výzkum, pochválil práci na anotačním treebanku i to, že prozkoumal mnoho směrů. Dále vyzdvihl, že student je jedním ze spoluautorů Universal Dependencies.

Uchazeč dostal slovo, aby představil svou práci, která se týká objevování struktur v jazycích, které nemají vlastní treebanky a jak k této úloze lze použít jazyky, které treebanky mají. V první části přednášky srozumitelně objasnil problémy spojené s úlohou, zejména problematiku hledání jazyka, který je podobný zkoumanému a to buď, že patří do společné rodiny nebo má podobný slovosled; problematiku využití strojového překladu na jazycích s malým množstvím paralelních dat; a jak by se daly různé přeložené treebanky zkombinovat. V druhé části představil svůj přístup, který je založený na využití Kullback–Leiblerovy divergence na trigramech part-of-speech tagů. Pro překlad zdrojových treebanků se rozhodl k využití statistického strojového překladu, který nutí generovat stejně dlouhé výstupy, jako jsou vstupní věty. Nakonec přeložené treebanky kombinuje pomocí metody Maximum spanning tree. V závěru představil podrobné vyhodnocení různých experimentů. Nakonec uchazeč stručně shrnul celou svoji práci.

Předseda požádal oponenty o přednesení jejich posudků a položení otázek.

Oponent prof. Jörg Tiedemann z Univ. of Helsinki vyzdvihl kvalitu práce a oblast, které se uchazeč zajímá, pochválil úroveň angličtiny a srozumitelnost, ale současně podotkl, že by rád v práci viděl větší důraz na technické detaily systému. Hlavní přínos práce vidí ve využití KL divergence a v podrobných experimentech, které uchazeč provedl. Potvrdil, že práce je kompletní a splňuje všechny požadavky disertační práce. Dále měl několik dotazů, na které uchazeč srozumitelně odpověděl.

Jako druhý vyjádřil své kladné stanovisko k práci oponent doc. RNDr. Aleš Horák, Ph.D. Zejména pochválil stylizaci a jazykovou stránku práce, důraz na kompletnost a detailní historii cross-lingual parsingu a úplnost experimentů. Oponent měl několik připomínek a otázek, na které uchazeč uspokojivě odpověděl.

Předseda se následně zeptal, jestli mají oponenti ještě další dotazy (nebyly) a následně otevřel všeobecnou diskuzi.

Diskuzi zahájil Pavel Straňák dotazem na výsledky bez použití strojového překladu. Aleš Tamchyna se ptal na roli oracula v experimentech a jestli může uchazečův systém výsledek oracula překonat. Následovaly dva dotazy od Jana Hajiče a diskuzi uzavřela Markéta Lopatková dotazem na part-of-speech tagging cílového jazyka. Uchazeč vše zodpověděl k plné spokojenosti všech tazatelů a ukázal, že se v problematice velmi dobře orientuje.

Předseda ukončil veřejnou debatu, konstatoval, že byly provedeny všechny potřebné kroky a formální požadavky byly splněny. Následně vyzval všechny přítomné mimo členů komise, oponentů a školitele, aby opustili místnost a mohlo proběhnout hlasování.

Následovala neveřejná diskuse členů komise a hlasování.

Pokyny pro předsedy nebo místopředsedy komisí:

Práce v elektronické podobě musí být studentem vložena do SIS. Formulář vyplňte ve všech bodech v elektronické podobě. V bodě Členové komise se uvedou všichni členové komise a za jejich jména se uvede „(přítomen)“ nebo „(nepřítomen)“. Předseda nebo místopředseda komise je jejím členem. V bodě Průběh obhajoby by měly být uvedeny alespoň čtyři věty vystihující průběh obhajoby. Po vyplnění formuláře ho vytiskněte, dole formulář ještě vlastnoručně podepište a přiložte k zápisu o státní závěrečné zkoušce. Současně vložte formulář v elektronické podobě (bez vlastnoručního podpisu) do SIS.

Komise schválila udělení titulu Ph.D.

Zápis provedl 14. 6. 2018 Ing. Tom Kocmi

Počet publikací: 31

Výsledek hlasování:

Počet členů s právem hlasovacím: 6

Počet přítomných členů: 6

Odevzdáno hlasů kladných: 6

Odevzdáno hlasů neplatných: 0

Odevzdáno hlasů záporných: 0

Výsledek obhajoby: prospěl/a neprospěl/a

Předseda nebo místopředseda komise:

Pokyny pro předsedy nebo místopředsedy komisí:

Práce v elektronické podobě musí být studentem vložena do SIS. Formulář vyplňte ve všech bodech v elektronické podobě. V bodě Členové komise se uvedou všichni členové komise a za jejich jména se uvede „(přítomen)“ nebo „(nepřítomen)“. Předseda nebo místopředseda komise je jejím členem. V bodě Průběh obhajoby by měly být uvedeny alespoň čtyři věty vystihující průběh obhajoby. Po vyplnění formuláře ho vytiskněte, dole formulář ještě vlastnoručně podepište a přiložte k zápisu o státní závěrečné zkoušce. Současně vložte formulář v elektronické podobě (bez vlastnoručního podpisu) do SIS.