

Posudek diplomové práce

Název diplomové práce: *Natural Language Processing of Textual Use Cases*

Autor diplomové práce: Jaroslav Dražan

Vedoucí diplomové práce: RNDr. Vladimír Mencl, Ph.D.

Cílem této diplomové práce bylo navázat na mé předchozí výsledky v použití lingvistických nástrojů na odvozování specifikací chování softwarových systémů z textových *use cases* („případů užití“). Předem daný úkol byl zlepšit přesnost odvozování, zejména správným výběrem z více *parse trees* („derivačních stromů“), které mohou být pro danou větu získané použitím několika parserů. Další, obecněji daný úkol byl hledat další možnosti vylepšení mého algoritmu. Součástí zadání bylo i nadále využívat omezení, která platí o struktuře vět používaných v *use cases*, a místo obecně těžkého zpracování přirozeného jazyka řešit pouze speciální případ *use cases*.

Student se práce chopil velmi samostatně a dokázal si najít i vlastní zdroje testovacích dat. Při jejich zpracování postupně nacházel možnosti vylepšení mého původního algoritmu; na získaných datech dosahuje vylepšený algoritmus velmi úspěšných výsledků.

Hlavním přínosem této práce je rozšíření konverze i na složitější věty, které neodpovídají vstupním požadavkům mého původního algoritmu, ale které jsou v praxi psaní *use cases* běžně používané – a zde se studentovi podařilo vhodným způsobem vymežit takovou třídu vět, které ještě lze popsat jednoduchými pravidly, a úspěšně se vyhnul slepé uličce vytváření mnoha jednoúčelových pravidel.

V rámci tohoto širšího schematu student přepracoval vstupní předpoklady pro algoritmus a současně i vytvořil složitější systém jeho výstupů, který reflektuje komplexnější informaci, kterou je nyní možné z jednoho *use case* kroku získat.

Druhým hlavním výsledkem je vytvoření metriky, která umožňuje vybírat z více různých *parse stromů* a tím úspěšně zpracovat větu i v případě, že jeden z parserů selže (a tedy je pomocí této metriky pro tento speciální případ možné detekovat selhání parseru). Aby mohl používat i výstupy z jiných parserů, musel student přepracovat můj algoritmus a odstranit závislost na *headword*, které poskytuje pouze Collinsův parser. Zde pro některé případy musel vyřešit i úlohu „výběru nejvýznamnějšího slova“, jejíž řešení bylo v mé původní práci delegováno na Collinsův parser.

Za mírnou slabinu práce lze považovat přesnost vyjadřování, zejména v kapitole 3. Vše podstatné pro popis algoritmu v ní řečené je, čtení textu je zde ale mírně obtížnější a k některým aspektům navrhovaného systému se bohužel nevyjadřuje. Na následující otázky jsem v textu nenašel uspokojivou odpověď:

1. Je *action set* skutečně pouze (neuspořádaná) množina, nebo se bude zachovávat pořadí akcí? Pokud ano (jak doufám), pak by bylo vhodnější nazývat spojení jednotlivých *action sets* *concatenation* and ne *union*.
2. Popis vyhledávání „S“ uzlů se nevyjadřuje, zda bude zpracovávat i „S“ uzel (různý od kořenového), do kterého je vložený jiný „S“ uzel – ani nevyslovuje předpoklad, že taková


situace v korektním parse stromu nenastane.

3. Sekce 3.4.6 na straně 27 se ani nevyjadřuje, co bude výsledný action set v případě, že věta obsahuje více *indirect objects*. Odpověď je vidět teprve až z příkladu na straně 42.

I přes tyto drobné nedostatky ale práci celkově hodnotím jako velmi úspěšnou a přínosnou. Práce je dobře strukturována, po popisu hlavních tezí práce je student předvádí na případové studii a porovnává podle dosažených výsledků s původním algoritmem; ve svém řešení i vysvětluje motivace pro zvolený postup. V kapitolách 6 a 7 poté rozebírá případné alternativní postupy i možná pokračování jeho práce.

Student práci zvládl vypracovat v angličtině na velmi dobré jazykové úrovni. Dosažené nové výsledky považuji za publikovatelné a i přes uvedené drobné nedostatky práci navrhuji uznat jako diplomovou.

V Macau, 31. ledna 2006.



RNDr. Vladimír Mencl, Ph.D.