

## I. IDENTIFIKAČNÍ ÚDAJE

<b>Název práce:</b>	<b>Algoritmy pro extrakci definic v právních předpisech</b>
<b>Jméno autora:</b>	<b>Jan Hošťálek</b>
<b>Typ práce:</b>	diplomová
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	13136
<b>Oponent práce:</b>	Karel Klíma
<b>Pracoviště oponenta práce:</b>	MFF UK, Katedra softwarového inženýrství

## II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

<b>Zadání</b>	<b>náročnější</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
<p>Zadáním práce bylo navrhnout, implementovat a otestovat algoritmy pro extrakci právních pojmů a jejich definic z českého právního řádu. Na první pohled by se mohlo zdát, že řešení lze splnit relativně jednoduše pouze vhodným “poskládáním” existujících NLP přístupů a metod, nicméně zaměření práce na právní doménu posouvá náročnost do jiné roviny. V první řadě je nutné seznámit se aspoň se základy českého právního řádu. Řešení musí nutně zahrnovat extrakci struktury a textu právních předpisů z dostupných zdrojů, které nejsou příliš kvalitní, a musí brát v potaz specifickou stylistiku z právní oblasti. Testování výsledného řešení pak vyžaduje zapojení doménových expertů, právní texty totiž vyžadují pro jejich použití větší či menší míru interpretace, které není vždy jednoznačná.</p>	

<b>Splnění zadání</b>	<b>splněno</b>
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
<p>Student vymyslel, implementoval a otestoval algoritmy v souladu se zadáním práce. Jedním z výstupů práce je webová aplikace OntoBuilder, která uživatelům umožňuje efektivně extrahovat slovníky právních termínů a definic z právních textů. Na základě testování a zpětné vazby pak student navrhl ještě alternativní řešení OntoAssistant, jehož základním principem je asistovaná extrakce informací z právních textů.</p>	

<b>Zvolený postup řešení</b>	<b>vynikající</b>
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
<p>Po metodické stránce je práce velmi kvalitně provedena. Student se seznámil s webovými standardy souvisejícími s modelováním OWL ontologií a provedl zevrubnou rešerši NLP technik pro extrahování znalostí z textu. Z existujících možností pak vybral několik řešení a provedl experimenty na české právní doméně. Výsledný návrh architektury OntoBuilder je tak přímým výsledkem korektní aplikace vědecké metody. Metodika testování byla vhodně zvolena a je v práci zevrubně popsána, včetně všech klíčových výsledků.</p>	

<b>Odborná úroveň</b>	<b>A - výborně</b>
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
<p>Student předvedl v rámci řešení práce dobré přehledové znalosti týkající se technologií sémantického webu, modelování ontologií a dobývání informací z textu, ať už klasickými metodami, nebo např. pomocí LLMs. Stejně tak se seznámil s některými aspekty českého právního řádu, které byly klíčové pro provedení práce.</p>	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>A - výborně</b>
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
<p>Formální a jazyková úroveň práce je standardní v tom nejlepší slova smyslu. Text se dobře čte a je dobře strukturován. Rozsah práce je odpovídající zadání.</p>	

**Výběr zdrojů, korektnost citací**

**A - výborně**

*Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.*

Student čerpal z řady odborných knih a článků, dále pak z W3C specifikací a v neposlední řadě z konkrétních českých právních předpisů. Zdroje jsou vhodně užité, převzaté prvky jsou označené v souladu s citačními zvyklostmi.

Jako jeden ze zdrojů právních informací, který v práci není obsažený, ale mohl by být, je projekt e-Sbírka, který obsahuje garantované a strukturované znění českých zákonů. Vzhledem k tomu, že byl projekt spuštěn relativně nedávno (v lednu 2024) však jeho absenci nehodnotím negativně.

**Další komentáře a hodnocení**

*Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.*

Student v rámci vypracování diplomové práce navrhl a implementoval webovou aplikaci pro extrakci právních termínů a definic z českého právního řádu. Zvolené technické řešení je dobré a především efektivní z pohledu časové náročnosti implementace - platforma Streamlit je skvělá na vytváření rychlých prototypů aplikací v Pythonu. Aplikace je funkční podle očekávání, umožňuje uživateli extrahovat termíny a definice z právních předpisů. Při testování jsem narazil na případy, kdy aplikace nezafungovala podle ideálního očekávání a např. nekorektně převedla jeden z termínů do infinitivu, nebo nenašla kompletní definici, avšak tyto případy a možné chyby jsou v práci popsány a v rámci možností dobře ošetřeny. Uživatelské rozhraní by mohlo být přívětivější, hodně by pomohlo například zvýraznění nalezených termínů a definic v původním textu. Vzhledem k tomu, že jde o prototyp pro ověření funkčnosti extrakčních algoritmů, které byly stěžejním výstupem práce, jsou tyto nedostatky pochopitelné.

**III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE**

*Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.*

Zadání práce bylo poměrně náročné, student k řešení použil odpovídající metodiku a výsledky práce, jak samotný text, tak i praktická implementace v podobě webové aplikace, jsou dobré.

Otázky k obhajobě: projekt e-Sbírka poskytuje strukturované znění právních předpisů ČR, kde se právní předpisy dělí na jednotlivé sekce a podsekce (části, hlavy, paragrafy, atd.). Jak by použití tohoto zdroje, na rozdíl od použití prostého textu, ovlivnilo kvalitu extrakce právních termínů v OntoBuilderu? Mohlo by použití předem strukturovaných právních předpisů vylepšit proces nebo výsledek extrakce právních termínů a definic?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 2.6.2024

Podpis: