

Posudek vedoucího diplomové práce

Název DP: **Algoritmy detekce obchodních dokumentů podle šablon**

Diplomant: **Jakub Michalko**

Obsah práce:

Předmětem diplomové práce (DP) bylo navrhnout a implementovat algoritmy pro analýzu a vyhledávání obchodních dokumentů podle šablon. Podstatnou součástí zadání byl návrh a implementace algoritmu pro podobnostní porovnávání anotované šablony a testovaného dokumentu na základě komplexní informace o strukturách rozpoznávaných v dokumentu (bloky textu, layout dokumentu). V úvodní kapitole autor stručně popisuje motivaci a specifikuje strukturu a cíl práce. V kapitole 2 jsou stručně rozebrány 3 aplikace pro automatické rozpoznávání dokumentů, z čehož jedna je výsledek softwarového projektu obhájeného na MFF, kde členem týmu byl autor DP. Ve třetí kapitole je stručně popsán model dokumentů a šablon a ve čtvrté velmi podrobně popsána implementace modelu pomocí tří komponent (parser, anotátor a vyhledávač dokumentů). Pátá kapitola obsahuje uživatelskou příručku, šestá programátorskou dokumentaci a sedmá kapitola obsahuje experimentální část. Poslední kapitola práci uzavírá.

Hodnocení:

V této verzi DP došlo k zapracování připomínek vznesených při předchozím termínu obhajoby. Mimo jiné, byly vymezeny rozdíly mezi SW projektem LANA a současnou aplikací/modelem, byl oddělen popis modelu (zavedena formalizace) od jeho implementace a byla zejména dokončena kapitola s experimenty. Práce stále trpí některými nedostatky předchozí verze – např. příliš prostoru věnováno implementačním detailům (detekce čar a předzpracování obrazu) – nicméně hlavní připomínky byly vypořádány.

Závěr:

Práci doporučuji k obhajobě.

V Praze dne 2. srpna 2016

Doc. RNDr. Tomáš Skopal, Ph.D.
vedoucí DP