

# Posudek oponenta diplomové práce

Jiří Kubeš: Strojové zpracování dat z insolvenčního rejstříku

Diplomová práce se zabývá zpracováním přehledu otevřených dat poskytovaných institucemi veřejné správy ČR, analýzou dat poskytovaných webovými službami insolvenčního rejstříku (ISIR) a realizací nástroje (datové pumpy) umožňujícího ukládat tato data do lokální relační databáze.

Celkově je práce provedena dosti podrobně a se snahou o důkladnost, ovšem převažují v ní – zejména v úvodní „studijní“ části – slovní kvalitativní informace a hodnocení nad technickými a kvantitativními. To je vzhledem k relativně technickému tématu na škodu, žádoucí by byl např. popis workflow ISDS formou diagramu (oddíl 2.7) nebo uvedení XSD pro datový typ „akce“ (oddíl 3.2.1). Popis insolvenčního rejstříku a procesu insolvenčního řízení (oddíl 2.4) je nepřehledný a neobsahuje dost informací na pochopení této agendy, pro práci klíčové, ať už z pohledu právního, obchodního nebo technického. V souvislosti s podrobnou hodnotící analýzou webových služeb ISIR by také bylo vhodné v úvodních kapitolách uvést standardy, vzory či doporučení týkající se návrhu rozhraní a funkcionality webových služeb.

Analýza webových služeb (WS) poskytovaných ISIR (kap. 3) je podrobná a dobře ilustruje problémy, se kterými se musí klientské aplikace vypořádat. V kapitole 4 Návrh datové pumpy bohužel chybí úvodní definice problému a popis účelu navrhovaného nástroje, ze kterých by plynuly požadavky na strukturu navrhovaného datového modelu a způsob vytváření ukládaných dat. Následující analýza paralelizovatelnosti zpracování (oddíl 4.4) a popis implementace (kap. 5) jsou vcelku dobře provedené a přehledně strukturované, v některých partiích (5.3-5.5) je ale text ale zbytečně podrobný a duplikuje srozumitelnou (a dobře komentovanou) programovou implementaci.

Za největší nedostatek práce považuji absenci hlubší analýzy kvality dat poskytovaných WS ISIR a související ověření výstupů datové pumpy z pohledu užitečnosti získaných informací. Přestože některé problémy s daty a architekturou WS jsou diskutovány na různých místech práce (oddíly 3.2-3.4, 4.2), čtenář se nedozví celkové charakteristiky dat (počty záznamů, objem dat, četnost změn resp. akcí v rámci jednotlivých řízení) ani informace o životním cyklu záznamů a jejich datové kvalitě v korpusu dostupném přes WS ISIR (míra duplicity údajů, (ne)úplnost informací o jednotlivých řízeních, atd.). Práce tedy nedává dostatečně jasnou odpověď na otázku, zda je možné WS ISIR použít pro sofistikovanější a spolehlivé analýzy dat o insolvenčních řízeních.

Vlastní implementace nástroje je v odpovídající kvalitě a rozsahu, některá návrhová rozhodnutí (připojení k databázi není definováno v samostatném konfiguračním souboru, dotazy pro ukládání dat jsou vytvářeny ve zdrojovém kódu „ručně“) ale komplikují údržbu nástroje při změnách WS ISIR. Text práce není po formální stránce příliš kvalitní, obsahuje např. množství dlouhých odstavců a pravopisných i věcných chyb. UML diagramy v příloze jsou zbytečně podrobné a dosti nečitelné.

V souhrnu konstatuji, že předložená diplomová práce má jisté nedostatky vzhledem k jejímu účelu, nicméně splňuje zadání a je využitelná jako základ pro další práci s daty insolvenčního rejstříku. Práci hodnotím stupněm **velmi dobře** a doporučuji k obhajobě.

V Plzni 31.5.2016

doc. Ing. Přemysl Brada, MSc. Ph.D.

Otázky k obhajobě

- Na str. 3 píšete „[Díky] e-governmentu došlo ... k úsporám.“ Je možné tyto úspory vyčíslit?
- Jaká notace je použita v diagramech 3.2 (str. 14) a 4.2 (str. 33)?

SOUHLASÍ  
S ORIGINÁLEM

Západočeská univerzita v Plzni  
Fakulta aplikovaných věd  
katedra informatiky a výpočetní techniky