

## Posudek oponenta diplomové práce

Jméno a příjmení autora posudku: Roman Neruda

Jméno a příjmení autora práce: Peter Zvirinský

Název práce: Social networks and data mining

---

Vlastní text (sem prosím napište text posudku, délka textu posudku není omezena):

Předkládaná práce se zabývá tématem použití metod data miningu pro oblast sociálních sítí. Autor aplikuje metody bayesovské analýzy, nákupního košíku, a metody analýzy sociálních sítí na databázi Insolvenčního rejstříku ČR. V úvodu se autor věnuje popisu metodologie data miningu CRISP-DM, v kapitolách 3-4 jsou zevrubně popsána data a proces extrakce a předzpracování Insolvenčního rejstříku ČR. Vlastní analýze dat je věnována rozsáhlá pátá kapitola, kde se postupně aplikují jednotlivé metody a vyvozují se z nich konkrétní závěry o struktuře a informacích v datech insolvenčního rejstříku. Šestá kapitola práci uzavírá zhodnocením dosažených výsledků a výhledem další práce. Součástí předkládaného textu jsou i použité programy a data na příloženém CD, která jsou popsána v přílohách.

Jde o rozsáhlou práci, která na zhruba 150 stranách uvádí do problematiky data miningu, popisuje použité metody a detailně se věnuje struktuře a způsobu práce s insolvenčním rejstříkem ČR. Práce je psána relativně dobrou angličtinou, text je srozumitelný a pečlivě zpracovaný. Popsané metody jsou vyloženy jasně, stejně tak i použitý software. Autor se zvláště podrobně věnuje struktuře dat a metodám předzpracování, které zahrnovaly využití systému pro rozpoznání písma z naskenovaných dokumentů. Výsledky autorovy práce s daty jsou většinou pečlivě popsány a ilustrovány řadou grafů a tabulek s jednoduchými statistickými údaji.

Za hlavní slabinu práce považuji nedostatečně prezentovaný vlastní přínos autora. Na jednu stranu se zdá, že jde o aplikaci standardních metod známých v oblasti data miningu, včetně použitých standardních softwarových nástrojů. Na druhou stranu, výsledky z případové studie na Insolvenčním rejstříku nepřinášejí dle mého nějaké překvapivé nebo nové informace. Autor by se při obhajobě měl zaměřit na zdůraznění vlastního přínosu práce v obou těchto oblastech. Co bylo cílem práce a jak jsou tyto cíle naplněny? Specificky by mě zajímala také otázka vhodnosti využití dat analýzy sociálních sítí na oblast insolventního registru. Jaké jsou autorovy vlastní zkušenosti? Jaké výsledky analýza přinesla ve srovnání s potenciálem použití těchto metod pro sítě typu Twitter nebo Facebook?

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	ANO <input type="checkbox"/>
---	------------------------------

Seznam soutěží studentských prací, viz <http://www.mff.cuni.cz/studium/bcmgr/prvzoryace/>

Pokud jste výše zaškrtnli ANO, zdůvodněte prosím svůj návrh, případně uveďte konkrétní soutěž, pro kterou je práce vhodná (rámeček lze nechat prázdný, pokud za dostatečné zdůvodnění považujete text posudku):

V Praze dne: 19. května 2014

Podpis:\*\*

\* *nehodící se škrtněte (vymažte)*

\*\* *do SISu vkládejte formulář nepodepsaný (ve formátu PDF), podpis je potřeba doplnit až na vytištěný posudek.*