

# OPONENTSKÝ POSUDEK DIZERTAČNÍ PRÁCE

**Jméno:** Anna Janíková

**Název:** Aplikace data miningu v oboru informačních studií: Analýza chování uživatelů knihovny

**Školitel:** prof. RNDr. Jiří Ivánek, CSc.

**Oponentka:** PhDr. Ladislava Zbiejczuk Suchá, PhD.

Dizertační práci s názvem „Aplikace data miningu v oboru informačních studií: Analýza chování uživatelů knihovny“ jsem přijala k oponování velmi ráda. Osobně vítám, že se využití data miningu (DM) pro knihovny v českém prostředí věnuje někdo takto systematicky, představení možností DM totiž v lokální literatuře citelně chybí, stejně jako analýzy s lokálním kontextem.

Text má odpovídající odbornou úroveň, je psán kultivovaným akademickým jazykem a autorka v něm od prvních stránek dokazuje svůj velký ponor i zájem o problematiku. Struktura dizertační práce, ve které nejprve autorka představuje fenomén data miningu v kontextu oboru, poté oblasti jeho využití, načež přechází k samotné analýze možností data miningu v konkrétním knihovním systému (přičemž se odkazuje na metodologie a témata studií představených v předchozích kapitolách) je logická a potenciální čtenář ji ocení. Otázkou je, nakolik je takový postup, kdy autorka při stanovování cíle své analýzy dat vychází především z literatury a nikoliv ze samotných dat a potřeb konkrétní knihovny, typický pro data mining, respektive užitečný v kontextu této konkrétní studie (nabízí se otázky, zda by systém Clavius a kontext konkrétní knihovny neumožňoval formulovat ještě zcela jiné, vlastní cíle zkoumání<sup>1</sup>), nicméně pro představení možností DM slouží autorce dobře a důvody takového postupu jsou vysvětleny poměrně přesvědčivě v kap. 3.5. Teoretická část práce, ve které autorka představuje množství výzkumů využívajících data mining ve fyzických i akademických knihovnách, je obdivuhodná rozsahem i pečlivou systematickostí zpracování. Autorce se navíc daří poměrně náročné téma představit srozumitelnou a čtivou formou, která zároveň nic neztrácí na své odbornosti.

Jedním z klíčových výstupů je analýza dostupnosti dat, průkaznosti a náročnosti zpracování, ke které mám několik otázek pro diskuzi: Je relevantní stanovovat náročnost provedení ve chvíli, kdy je konstatováno, že data pro takovou studii nejsou k dispozici (viz řádky Blečić + Šoskova, Leonard, Goswami a další)? Bylo by možné více specifikovat kritéria rozdělení náročnosti provedení na jednotlivé stupně – nízká, střední, vysoká / vysoká (potřeba spec. znalostí)?

Protože ne nepodstatná část výzkumných námětů byla označena za neproveditelnou z důvodu, že data z Clavia a z univerzitního informačního systému STAG nejsou provázaná, chybí mi v této části také obsáhlejší diskuze nad tím, jak velkou chybovost (respektive přínosy či rizika) by přinesly různé alternativní způsoby provázání dat z knihovního systému s daty o uživatelích (například propojení přes jiné identifikátory uživatelů či propojení s daty z dotazníku realizovaného ad hoc).

Samotná praktická práce s daty tvoří zhruba třetinu obsahu práce. V úvodní části, kdy autorka využívá většinou možností deskriptivní statistické analýzy, výsledky v některých případech nepřinášejí zcela překvapivé informace (například vyšší poměr prezenčních studentů mezi výpůjčkami a přihlašování se dal pravděpodobně předvídat na základě celkového počtu studentů v jednotlivých sledovaných kategoriích, podobně analytika stálých uživatelů je samozřejmě zkreslena tím, jak dlouho studenti/zaměstnanci na univerzitě působí), jindy však budou pro instituci výsledky zcela jisté

---

<sup>1</sup> V souladu s tvrzením převzatým autorkou, že „ve věku big data se dokonce ukazuje produktivnějším stanovení vlastní metodiky v závislosti na zpracovávaném úkolu (Piatetsky 2014).“

zajímavé, neb nemusí být z běžného provozu zřejmé (např. fakt, že po dvou letech se promění 80 % uživatelů knihovny). Určitě nejzajímavější zjištění poté přináší analýza uživatelského chování (analýza používání po měsících, dnech, hodinách a další následné analýzy). Tuto část vnímám jako jednu z nejpřínosnějších částí práce. V závěru analytické kapitoly poté autorka doplňuje svou analýzu srovnáním s výsledky předešlých zahraničních studií, což je také určitě vhodné.

Dílčí výhrady k analytické části:

- v analytické části jsou v textu uvedeny chybně čísla tabulek (s. 67),
- u obrázku 3.19 není jasné, proč jsou hodnoty na ose Y zobrazeny v procentech (vysvětlení jsem posléze našla u obrázku 3.22 – pro kontext by ale bylo vhodné doplnit i absolutní hodnoty alespoň u maximálních hodnot sledovaných proměnných),
- autorka uvádí výsledky svých analýz, ale nevyšimá si jejich statistické významnosti.

Co mi v textu ztelně chybí, je závěrečná reflexe celého postupu i doporučení pro další uchopení tématu. Proto navrhuji, aby diskuze u obhajoby byla vedena i tímto směrem. (Jaké náměty pro další výzkum autorka identifikovala? Jaké dopady na institucionální úrovni mělo či může mít zpracovávání dat z knihovních systémů?)

#### **Celkové hodnocení:**

Práci navzdory výše uvedeným dílčím výtkám hodnotím jako velice přínosnou, protože poctivým způsobem zpracovává téma, které je v českém prostředí v akademické diskuzi i v praxi velmi opomíjeno. V tomto ohledu se jedná o velký přínos pro obor.

Práce splňuje všechny požadavky standardně kladené na dizertační práci, doporučuji ji proto bez váhání k obhajobě a navrhuji klasifikaci „prospěla“.



PhDr. Ladislava Zbiejczuk Suchá, PhD.

V Brně dne 6. 6. 2016