

Posudek diplomové práce

Autor diplomové práce: Martin Kubelka

Název diplomové práce: Datamining sociálních sítí

Vedoucí diplomové práce: Josef Šlerka

Oponent: Veronika Trachtová

Celkové hodnocení: velmi dobře, práci doporučuji k obhajobě

Martin Kubelka si ve své diplomové práci klade za cíl prozkoumat možnosti analýzy obsahu sociálních sítí, a to především z pohledu Social Network Analysis (SNA) a analýzy sentimentu. Práce přináší přehled metod, jejichž prostřednictvím lze analyzovat obrovský objem dat, který je dostupný díky expanzi sociálních sítí a s ní souvisejícímu nárůstu informací tvořených uživateli (tzv. Big data). Autor mj. shrnuje i podmínky a smysl zpracování těchto dat (nárůst výpočetní kapacity a zároveň schopnost zpracovat data v únosném časovém horizontu).

Na práci lze ocenit především představení praktických výsledků datamingu v sociálních sítích, které autor provedl. Ať už se jednalo o ukázky určování vlivu jednotlivců v rámci sítě pomocí NodeXL, word cloudy, či výpočet korelací, u kterých se ukázal vztah mezi „buzzem“ na sociálních sítích a návštěvnosti kin (a tržeb) vybraných filmů. V některých případech mohl Martin Kubelka detailněji popsat výsledky, neboť pro všechny čtenáře není spojitost mezi konkrétními uživateli tak jasná.

V teoretických částech práce se Martin Kubelka v několika případech nevyhnul příliš stručným informacím, které by mohl popsat detailněji (např. definice méně známých pojmů jako acyklické směrované sítě či RDF technologie). Podobně v některých pasážích jeho odkazování na obecné zdroje jako Wikipedia přivádí čtenáře k otázce, zda tento přístup nevede k příliš nekritickému zhodnocení daných metod. Práce splňuje náležitosti co se týče rozsahu i formy zpracování. Drobné výtky lze mít vůči několika hovorovým obrátům, které se v práci objevily.

Práce Martina Kubelky reflektuje změnu přístupu ke zpracování dat a z jejího textu je zřejmé, že autor při provádění praktických výzkumů došel k samostatným závěrům, které se dobře uplatní zejména při social media monitoringu v českém prostředí (např. problémy s automatickým určováním sentimentu, význam analýzy příspěvků nejen ze sociálních sítí, ale rovněž z diskusních fór, jejichž struktura je klíčová při určování kontextu a tedy významu příspěvku). Podobně zajímavé jsou i výsledky výzkumu, v němž autor zaznamenal, že hodnocení sentimentu různými respondenty vede k poměrně nejednoznačným výsledkům.

V rámci obhajoby práce, by mě zajímalo, jak Martin Kubelka vnímá etický rámec dataminingu, a rovněž, jak hodnotí možnosti predikce na celou populaci, když na některých sítích zatím převažují lidé ze skupiny technologických early-adopters.