

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy v Praze

Autor práce Erik Lux
Název práce Feature selection for text classification with Naive Bayes
Rok odevzdání 2012
Studijní program Informatika **Studijní obor** Obecná informatika

Autor posudku Mgr. Zuzana Reitermanová **Role** Vedoucí
Pracoviště KTIML MFF UK

Prosím vyplňte hodnocení křížkem u každého kritéria. Hodnocení *OK* označuje práci, která kritérium vhodným způsobem splňuje. Hodnocení *lepší* a *horší* označují splnění nad a pod rámec obvyklý pro bakalářskou práci, hodnocení *nevyhovuje* označuje práci, která by neměla být obhájena. Hodnocení v případě potřeby doplňte komentářem. Komentář prosím doplňte všude, kde je hodnocení jiné než *OK*.

| K celé práci | lepší | OK | horší | nevyhovuje |
|---|--------------------------|-------------------------------------|-------------------------------------|--------------------------|
| Obtížnost zadání | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Splnění zadání | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Rozsah práce ... <i>textová i implementační část, zohlednění náročnosti</i> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| Komentář Předložená práce shrnuje existující techniky pro klasifikaci textu a relevantní metody pro výběr příznaků. Následně student implementoval zvolené algoritmy (Naive Bayes a několik různých metod pro výběr příznaků a předzpracování dokumentů). Chování těchto metod student otestoval a porovnal na dvou referenčních datasetech. Za hlavní nedostatky práce považuji slabší jazykovou úroveň textu a příliš stručnou řešeršíní část. Naopak oceňuji poměrně zdařilý návrh a implementaci aplikace. Student vytvořil modulární a snadno rozšiřitelnou knihovnu, která může být snadno začleněna do různých projektů a použita na různé typy textu. Nad knihovnou vytvořil student jednoduché rozšíření existujícího emailového klienta. Přes nedostatky v textové části práce považuji práci za poměrně zdařilou především co se týká návrhu, implementace a provedených experimentů. Mimoto, vytvořená knihovna může sloužit jako základ pro další výzkum v oblasti klasifikace textu. Doporučuji proto, aby práce byla uznaná jako bakalářská práce. | | | | |

| Textová část práce | lepší | OK | horší | nevyhovuje |
|---|--------------------------|-------------------------------------|-------------------------------------|--------------------------|
| Formální úprava ... <i>jazyková úroveň, typografická úroveň, citace</i> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| Struktura textu ... <i>kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| Analýza | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| Vývojová dokumentace | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Uživatelská dokumentace | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Komentář | | | | |
| <p>Práce má pěknou strukturu a odpovídající typografickou úroveň (až na občasné chyby v odkazech v rámci textu, např. str. 24). Text je psán anglicky a jeho jazyková úroveň je slabší, text se poměrně obtížně čte.</p> <p>Rešeršní část práce je přehledná, zmiňuje snad všechny důležité relevantní techniky a aplikace, nicméně je dost stručná. K lepší čitelnosti by prospělo lepší propojení jednotlivých kapitol a průběžně nastíněná motivace.</p> <p>Student provedl řadu experimentů, jejich výsledky přehledně vizualizoval pomocí tabulek a grafů, textový popis a zhodnocení výsledků jsou ovšem opět velmi stručné.</p> <p>Vývojová a uživatelská dokumentace jsou odpovídající.</p> | | | | |

| Implementační část práce | lepší | OK | horší | nevyhovuje |
|--|--------------------------|-------------------------------------|--------------------------|--------------------------|
| Kvalita návrhu ... <i>architektura, struktury a algoritmy, použité technologie</i> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Kvalita zpracování ... <i>jmenné konvence, formátování, komentáře, testování</i> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Stabilita implementace | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Komentář | | | | |
| <p>Knihovna poskytuje přehledné rozhraní, je snadno rozšiřitelná.</p> <p>Co se týká zpracování, kód zcela neodpovídá jmenným konvencím Javy, nicméně jména jsou jednotná, kód je dobře okomentovaný.</p> | | | | |

Celkové hodnocení Velmi dobře - Dobře
Práci navrhuji na zvláštní ocenění Ne

Datum 5. června 2012

Podpis