

RSS čtečka s inteligentní filtrací

Projekt z umělé inteligence

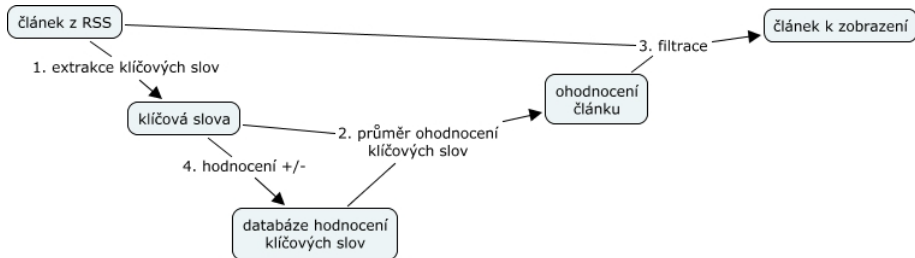
Jiří Procházka

FI MUNI



26. dubna 2013

Aktuální stav implementace

- Více-uživatelský systém pro perzistenci dat ✓
- Extrakce klíčových slov ✓
- Průměr ohodnocení klíčových slov ✓
- Hodnocení klíčových slov ✓
- Server interface ✓
- Klient interface ✓
- UI, filtrace ✓



topnews (+)

  United States, Russia agree to try to revive Syria plan

LF [filtered] (show)

 topnews (+)

  Immigration bill could fix some Boston security issues: official

LF keywords: immigration bill(+/-), Boston Marathon bombings(+/-), Boston security issues(+/-), Obama administration official(+/-), WASHINGTON Reuters(+/-), border control weaknesses(+/-), [all](+/-)

WASHINGTON (Reuters) - A top Obama administration official said on Tuesday that an immigration bill starting to move through Congress would fix some border control weaknesses that may have contributed to last week's Boston Marathon bombings.

 Email this  Digg This!

 topnews (+)

  Boston bombing suspect condition upgraded to "fair" from "serious"

LF keywords: Boston(+/-), condition(+/-), April 23(+/-), Beth Israel Deaconess Medical Center(+/-), Boston Marathon bombing attack(+/-), Boston bombing(+/-), [all](+/-)

BOSTON, April 23 - The suspect in last week's Boston Marathon bombing attack is now in "fair" condition at Beth Israel Deaconess Medical Center, an upgrade from his prior condition of "serious," the U.S. Attorney's Office in Boston said on Tuesday

Průměr ohodnocení klíčových slov

Používám depth limited search. Váha ohodnocení hyperným klesá se vzdáleností od klíčového slova v hierarchii podle funkce $f(x) = 1/x$.

Problém: přidané stejně hodnocené hypernum *snižuje* ohodnocení celého textu → použití koeficientů na zmírnění tohoto účinku:

$$\text{depthRating}(\text{rating}, \text{depth}, Q) = \text{rating} \cdot (\text{depth} + 1)^{-Q}$$

Pro výpočet dělitele v průměru se použije tento vzorec s rating 1.0 a druhým hloubkovým koeficientem:

| pole (slovo,rating,hloubka) | Q=1.0, normalní průměr | Q=1.0,Qd=1.0 | Q=1.0,Qd=0.5 |
|-----------------------------------|------------------------|---------------------|--------------|
| [('rudá',1.0,0)] | 1.0 | 1.0 | 1.0 |
| [('rudá',1.0,0),('barva',1.0,2)] | 0.6667 | 1.0 | 0.8453 |
| [('rudá',1.0,0),('barva',0.5,2)] | 0.5833 | 0.875 | 0.7396 |
| [('rudá',-1.0,0),('barva',1.0,2)] | -0.3333 | -0.5 | -0.4226 |

Přínos projektu vzhledem k cizím řešením

Projekt umožňuje filtraci na základě extrahovaných klíčových slov řízenou samotným uživatelem, což známé existující řešení nenabízí, čímž umožňuje aby individuální preference uživatelů byly zohledněny.

Pozitivní je, že lze projekt kombinovat s ruční kategorizací zpravodajských serverů či agregátorů odebráním RSS/Atom feedu žádaných kategorií, pokud to nabízí.

V případě Tiny Tiny RSS projekt lze kombinovat i se zabudovaným tagováním a filtrováním pomocí regulérních výrazů.

Způsob vyhodnocení a testování

Uživatelské testování, zpětná vazba formou dotazníků.

Dotazník:

- Počet zkoušených feedů
- Počet čtených (abstraktů) článků (zhruba)
- Počet provedených hodnocení klíčových slov (zhruba)
- Počet provedených hodnocení [all] (zhruba)
- Počet odfiltrovaných článků (zhruba)
- Hodnocení kvality extrahovaných klíčových slov (1-10)
- Celkové hodnocení konceptu projektu a jeho provedení (1-10)
- Poznámky, návrhy k zlepšení

Konec prezentace

Děkuji za pozornost