

IDENTIFYING PERSONAL CHARACTERISTICS USING MICROBLOG

Peter Korenek

Vedúci: Ing. Marián Šimko PhD

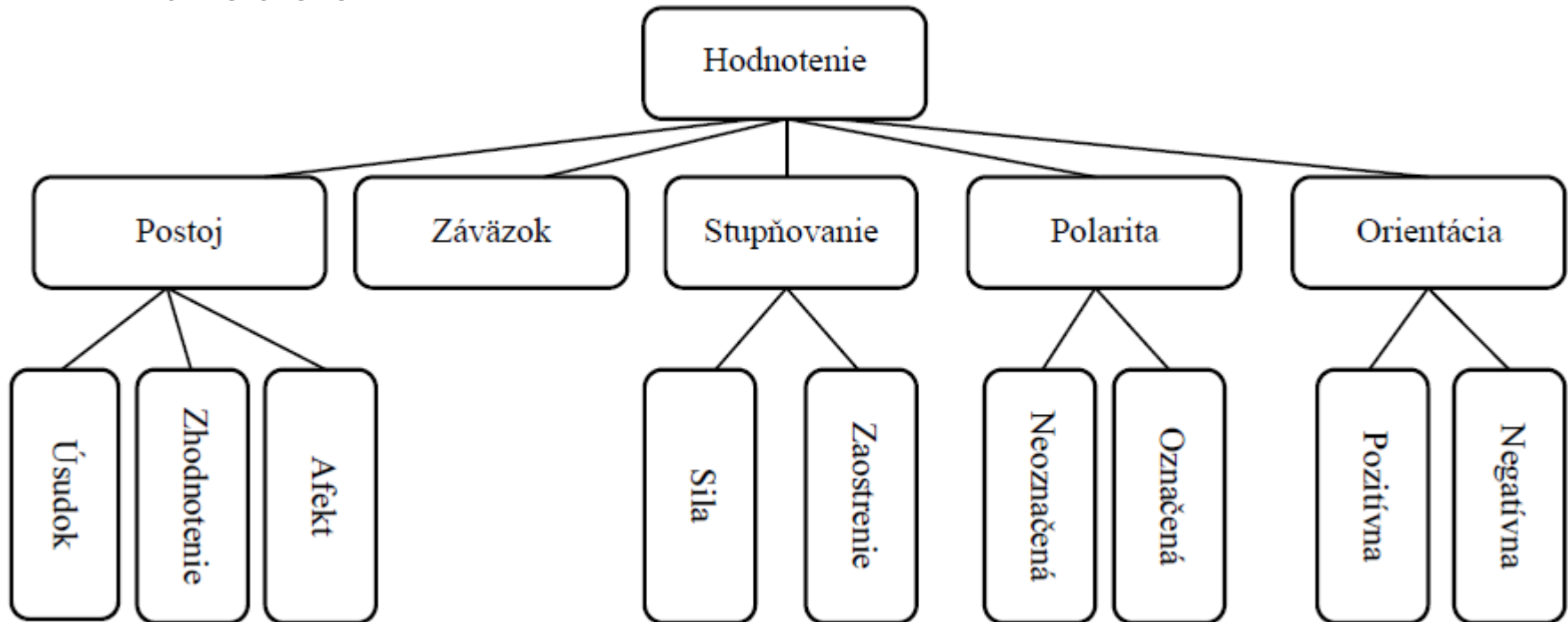
Motivácia

- Vedieť čo sa ľuďom páči a nepáči
- Vedieť aký postoj majú ľudia k istým subjektom
- Získať dôležité prieskumné názory
- Potvrdiť použiteľnosť teórie hodnotení aj pri analýze krátkych článkov ako sú mikroblogy
- V súčasnosti je bežná klasifikácia pozitívne / negatívne texty



Teória hodnotení (White, 2005)

- Emócie sú dôsledky toho ako človek ohodnotil istú situáciu



Orientácia na cieľ

- Na identifikáciu cieľov vety vytvorené jednoduché syntaktické vzory špecifické pre mikroblogy
- **I'm a current @Blackberry user, very disappointed with it! I'm moving to @Android or @Apple @iphone**
- Negatívne pre Blackberry, pozitívne pre Android a Apple – ako ohodnotiť celý microblog?
- Klasifikácia zamierená na cieľ už bola použitá (Whitelaw, 2005), ale nie v kombinácii s teóriou hodnotení na mikroblogoch

Postup

- Získanie dát
- Predspracovanie
 - ▣ Slang
 - ▣ Lematizácia
- Identifikácia cieľov mikrobloggerov
- Priradenie emócií cieľom
- Klasifikácia
- Vytvorenie stromu záujmov používateľa

Experiment – klasifikácia emócií

- Emocionálna klasifikácia mikroblov orientovaná na ciele použitím teórie hodnotení
- Použitý existujúci dataset s 2500 mikroblovami a priradenými emóciami (orientácia)
- Nutná oprava asi tretiny anotácií
- 10-cross validation
- Vykonané predspracovanie a klasifikácia
- Použité atribúty:
 - ▣ Appraisal, Emotikony, interpunkcia, URL, RT, Mention
- SVM klasifikátor

Výsledky experimentu 1

- Celková úspešnosť 88%

Identifikovaná trieda

	Positive	Negative	Neutral	Total
Positive	210	17	25	252
Negative	214	20	11	245
Neutral	150	115	1688	1753
Total	374	152	1724	2250

Predpokladaná
trieda

Zhodnotenie experimentu 1

- Výsledky porovnateľné s inými prácami
- Otestované pre datasey s rôznymi témami (nielen pre jednu tému – filmové recenzie a pod)
- Teória hodnotení zvýšila presnosť o 5-10%
- Cieľený sentiment zvýšil presnosť o 5 %

Experiment 2

- Sumarizácia postojov blogera k subjektom mikroblogov
- Manuálne označené ciele v mikroblogoch
- Priradenie orientácie vzhľadom k cieľom
- Experiment zatiaľ na 10 používateľoch po 100 mikroblogoch

Výsledky E2 - Identifikácia cieľov

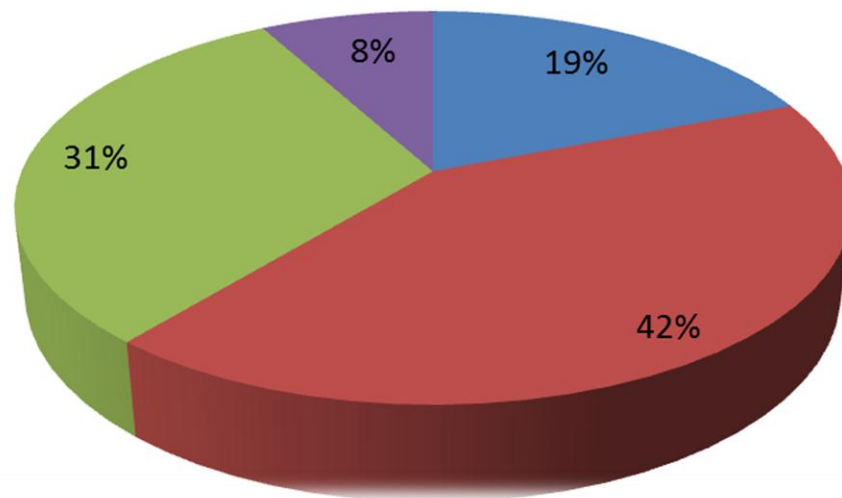
Používateľ	Počet microblogov	Počet cieľov	Počet správne identifikovaných cieľov	%
Jozef Chutka	180	200	180	90
Larissa Sutherland	150	168	155	92
Jeremiah Griffin	100	112	79	70
Selena Gomez	160	167	150	89
Spolu	590	647	564	85

Výsledky E2 – sumarizácia emócií

□ Autor - Selena Gomez

Love You Like a Love Song

- Zhodnotenie - kvalita
- Postoj - šťastie - radosť
- Postoj - spokojnosť - záujem
- Postoj - spokojnosť - potešenie



Zhodnotenie E2

- Použitie hybridných vzorov (syntax + črty mikroblogo) výrazne zvýšilo presnosť identifikácie cieľov
- Metóda poskytuje spôsob ako zhodnotiť postoje autora k cieľom – začlenenie do kategórii podľa teórie hodnotení
- Ťažšie overiť

Nedostatky

- Výskyt sarkazmu
- Myšlienky myslené v „“
- Veľkosť slovníka je dosť podstatná
- Neošetroval som preklepy

Ďalší postup

- Dokončiť experiment 2 na viacerých používateľoch
- Rozšíriť slovník

Zhodnotenie

- Využitá teória hodnotení pri klasifikácii emócií
- Identifikácia cieľov pomocou syntaktických vzorov mikroblogov
- Uskutočnené 2 experimenty
 - ▣ Prvý úspešne dokázal vhodnosť teórie aj pre mikroblogy
 - ▣ Druhý experiment ukázal použiteľnosť metódy pre sumarizáciu emócií k subjektom mikroblogov