Extrakcia sumarizovaných názorov a identifikácia autorít v konverzačnom obsahu

Spoločný seminár
”Umelej inteligencie a kognitívnej vedy“
máj 2014

Kristína Machová, TUKE

Osnova:

1. Motivácia
2. Konverzačný obsah
3. Problémy riešiteľné dolovaním
4. Analýza názorov
5. Dynamický koeficient
6. N-gramy
7. Identifikácia autorít
8. Vzťah medzi dolovaním názorov a autorít

Motivácia

- Sociálny web umožňuje a posilňuje interakcie
- Tieto interakcie sú spojené s ovplyvňovaním → rozhodovacie procesy v reálnych situáciách (kúpa drahého produktu, voľba politickej reprezentácie…)
- Rozhodovacie procesy môžu byť podporované aplikáciami dolovania názorov z konverzačného obsahu.
- Získané informácie:
  - o drahých veciach (nehnuteľnosť, dovolenková destinácia, auto…)
  - kultúrne informácie
  - Informácie spojené s bezpečnostnými aspektmi

Motivácia (2)

- Dolovanie názorov (opinion mining, sentiment cassification, sentiment analysis) dolovanie postoja jednotlivého prispievateľa (diskusie ako celku) k určitej téme.
- Téma – hodnotenie produktu, politickej situácie, udalosti, osoby, lekára, filmu, knihy, tovaru alebo pocitov autora k objektu hodnotenia.
- Dolovanie názorov je možné rozšíriť z vnímania textov na úroveň vlastností posudzovaných objektov.

Konverzačný obsah

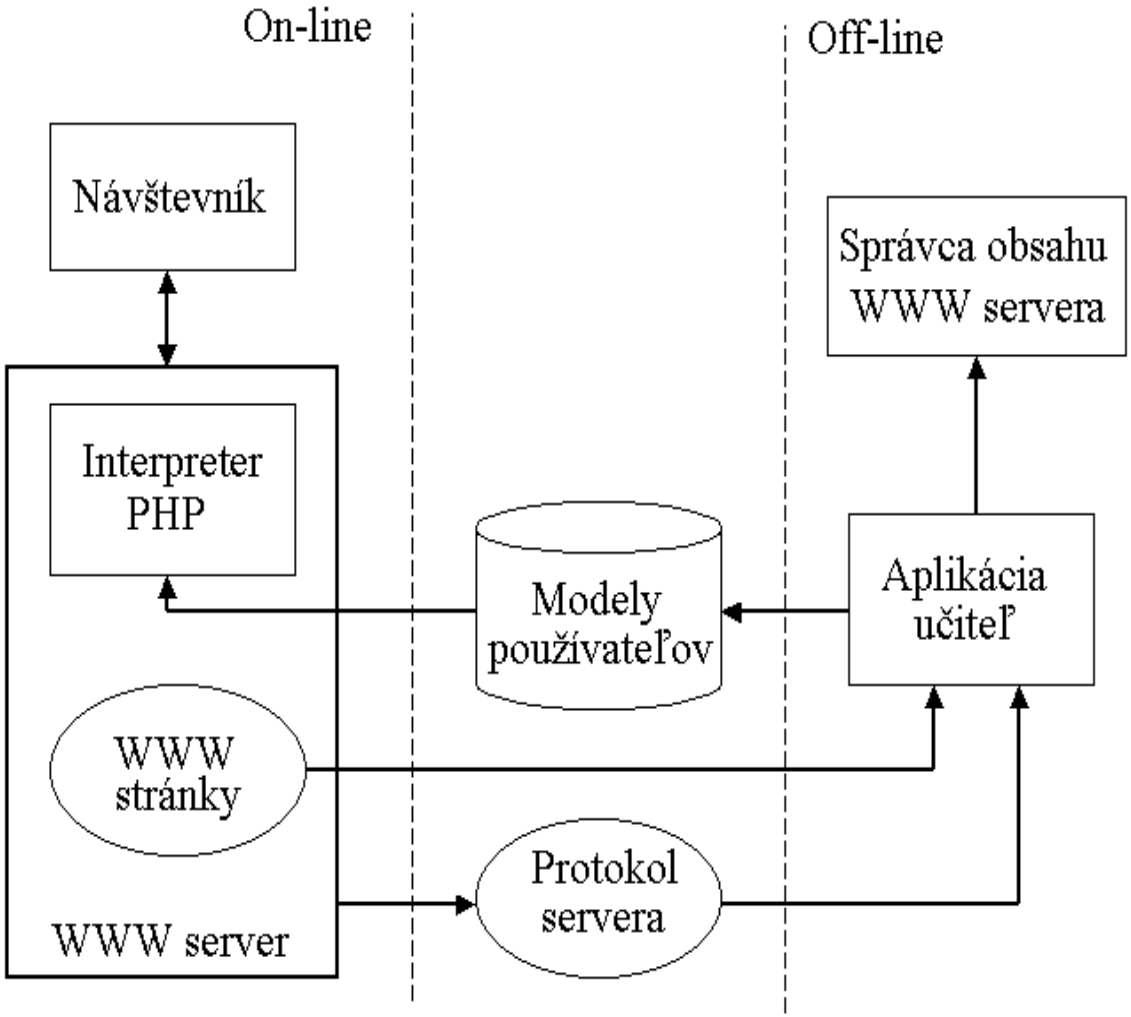
- Krátke texty (hovorené písanie, písané hovorenie – debata) k určitej téme.
- Známa téma – výsledok hodnotenia veci, produktu, osoby o ktorej sa diskutuje, resp. pocitov autora.
- Neznáma téma – modelovanie témy.
- Syntaktická odlišnosť (frekvencia typických slov, interpunkcia, slovosled, preklepy – aj úmyselné) – odráža autorovu osobnosť.
- Konverzačný obsah: sociálne siete, blog, microblog, chat, chatrooms, IRC (Internet Relay Chat), diskusné fóra, komentáre k článkom, videám a pod.

Typy dolovania z konverzácie

- Dolovanie z používania
  - doluje sa z log súborov
  - používateľ verzus linky (stránky), ktoré navštívil
  - vedie k personalizácii webu (navigácia používateľa)
- Dolovanie zo štruktúry
  - mapovanie okolia aktuálnej web stránky (navigácia používateľa)
  - dolovanie zo štruktúry konverzácie (identifikácia autorít)
- Dolovanie z obsahu konverzácie – analýza sentimentu
  - dolovanie názorov resp. klasifikácia názorov (pozitívny, negatívny)
  - detekcia emócií (strach, hnev, smútok odpor, prekvapenie, radosť - Ekman), (znechutenie, nadšenie, hanba, vina...)

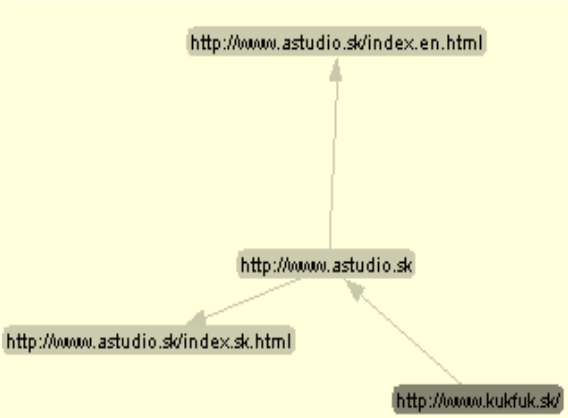
Dolovanie z používania

- použitím strojového učenia (HGS, HSG) sa učí model používateľa
- model používateľa sa použije na doporučovanie personalizovaného zoznamu nových stránok



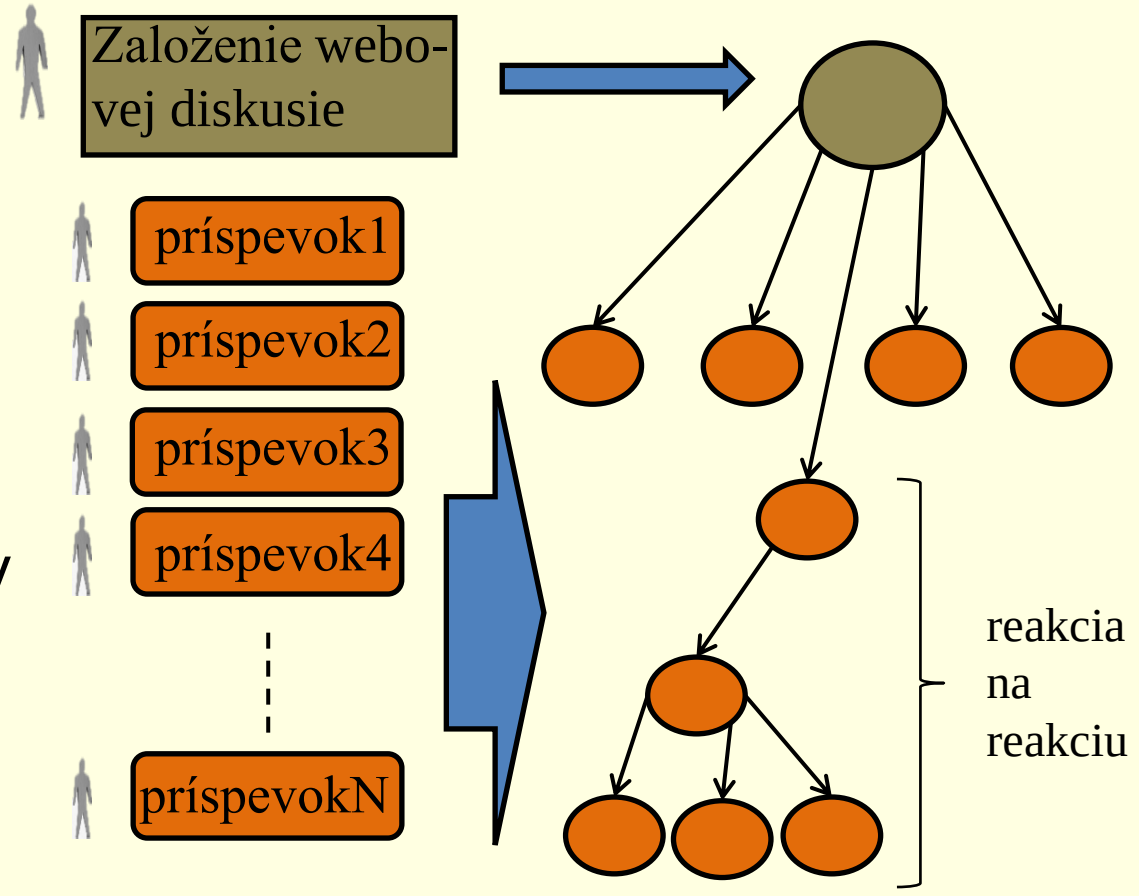
Dolovanie zo štruktúry

- Parciálne mapovanie okolia aktuálnej web stránky
- Matice susednosti, matice najkratších vzdialeností
- Rozlišujeme úrovne vnorenia (2,3,...)



Dolovanie zo štruktúry konverzácie

- Počet príspevkov daného prispievateľa
- Počet reakcií na jeho príspevky
- Počet výskytov na spodnej úrovni (uzavretá diskusia)
- a pod.



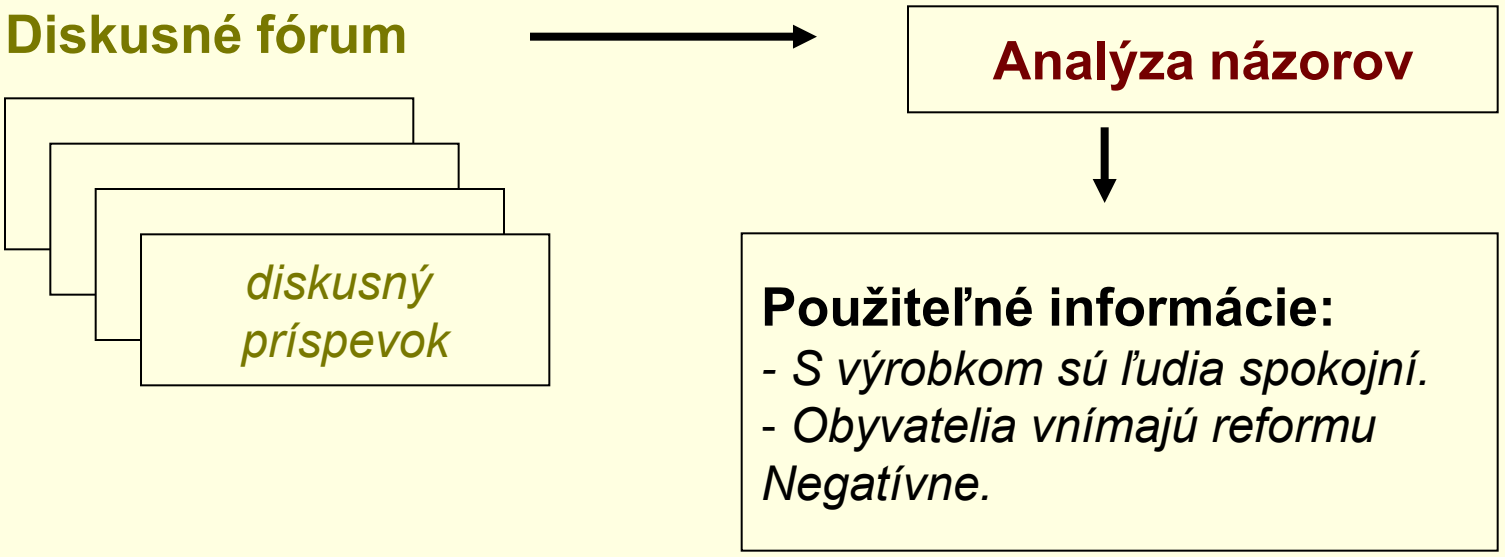
Dolovanie z obsahu konverzácie

Problémy riešiteľné dolovaním konverzačného obsahu:

- Analýza názorov (Pozitívny, negatívny?)
- Identifikácia autorít (Kto je autoritou v tejto diskusii?)
- Vyhľadávanie názorového spamu (Je obsah príspevku informatívny? Vykecávačky?)
- Určovanie užitočnosti názorov (Je tento názor kvalitný, autoritatívny?)
- Aspektovo orientovaná analýza sentimentu (Aká je názorová polarita v rámci jednotlivých vlastností entity?)
- Porovnávacia analýza sentimentu (Ktorý z týchto produktov je lacnejší, komfortnejší, poruchovejší?)
- Cielená reklama (Čo má obsahovať, lebo to ľudia oceňujú?)
- Detekcia emócií (Čo vyjadruje príspevok: nadšenie, znechutenie?)
- Modelovanie témy (O čom sa diskutuje?)
- Vyhľadávanie názoru (Kde sa o tom diskutuje?)
- Identifikácia autorstva (Kto je autorom príspevku? Aký typ človeka je prispievateľ?)

Analýza názorov

- Diskusné fóra – rastúce úložiská informácií: názorov, pocitov, postojov a nálad ľudí (Internet ako spôsob komunikácie).
- Na rozdiel od databáz neobsahujú štruktúrované dáta, preto vyžadujú špeciálne postupy (klasifikácia názorov).



Analýza názorov

- Uplatnenie v oblastiach s potrebou agregácie množstva názorov do jednej výslednej ucelenej informácie.
- Vývoj a predaj produktov, prieskum verejnej mienky,...
- Tieto oblasti sa skúmajú z dvoch pohľadov:
  - z pohľadu spotrebiteľa (zdroj informácií pre rozhodnutie o kúpe, webové stránky produktu, diskusia na portáloch - extrakcia sumarizovaného názoru aplikáciou KN)
  - z pohľadu výrobcu (vývoj (informácie o dodávateľoch a konkurencii) a predaj (informácie o potrebách a spokojnosti zákazníkov), marketingový prieskum – náklady (dotazníky, telefón)
  - Internetový prieskum prostredníctvom aplikácie KN (↓ náklady, ↑ rýchlosť) - rýchlosť získavania informácií o zákazníkovi je zásadná.

Metódy analýzy názorov

Podľa Taboada, dva hlavné prístupy ku analýze názorov:

- Prístup založený na klasifikácii
  - metódy strojového učenia (Naive Bayes Classifier, SVM – Support Vector Machines) vyžadujú trénovaciu množinu (anotačné nástroje, váhovacie techniky)
  - štatistické metódy (Maximal Entropy)
- Prístup založený na externom zdroji – lexikóne
  - slovníkovo založený
  - korpusovo založený

Podľa Koncza:

- Exogénne (SU, TM)
- Endogénne (externý zdroj znalostí – slovník)

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics, Vol. 37, No. 2, 267-307 (2011)

Slovníkový prístup – o čo ide?

- Pozitívny (negatívny) príspevok (diskusia): prevažujú slová (príspevky) s pozitívnou (negatívnou) polaritou
- Neutrálny príspevok:
  - Striktný prístup

    IF Pocet_pozit = Pocet_negat THEN neutralita

    vhodný pre krátke príspevky (pohltenie širším pásmom neutrality)
  - Vo všeobecnosti:

    IF |Pocet_pozit – Pocet_negat| ≤ H THEN neutralita

    vhodný pre dlhšie príspevky

    H = 0 – striktný prístup

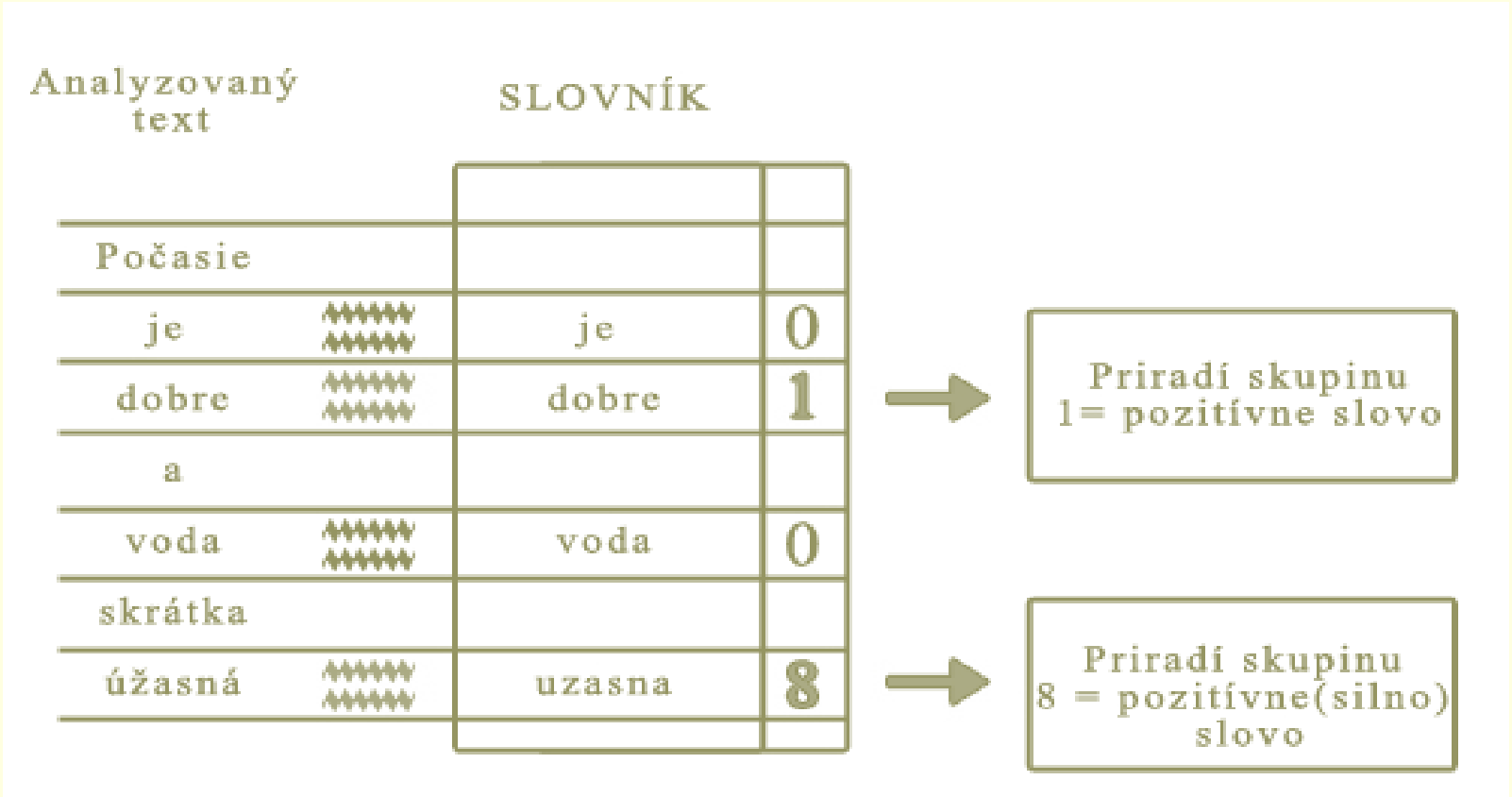
Slovníkový prístup k analýze názorov

Je potrebné získať klasifikačný slovník:

- generovaním pre danú doménu
  - nahrávanie klasifikačného slovníka z diskusie
- generovaný použitím známych lexikónov
  - Word Net
  - Word Net – Affect
  - Senti Net
  - Senti Word Net

Nahrávanie klasifikačných slovníkov

Identifikácia slov so subjektivitou a ich nahrávanie do poľa termov – slov. Každému slovu je priradená číselná hodnota (polarita, zápor, intenzita).



Nahrávanie klasifikačných slovníkov

Klasifikačný slovník:

- obsahuje slová, ktoré sú nositeľmi názoru v rámci danej domény
- prebraté z priamo z diskusie (náš prístup)
- má zabezpečiť prispôsobenie sa živej reči prispievateľov do web diskusií
- nespisovné slangové slová (coolový, dzivý,...)
- slová bez diakritiky (kvalitny, paci (sa mi))
- gramatické chyby?
- čím je slovník obsiahlejší, tým presnejšia je klasifikácia názorov

Slovníkový prístup

Ukážky klasifikačných slovníkov: Table2 - príslovky
Table1 - podstatné mená a slovesá,
Table3 – intenzifikátory

Table 1
Examples of words in the noun and verb dictionaries.

Word	SO Value
monstrosity	−5
hate (noun and verb)	−4
disgust	−3
sham	−3
fabricate	−2
delay (noun and verb)	−1
determination	1
inspire	2
inspiration	2
endear	3
relish (verb)	4
masterpiece	5

Table 2
Examples from the adverb dictionary.

Word	SO Value
excruciatingly	−5
inexcusably	−3
foolishly	−2
satisfactorily	1
purposefully	2
hilariously	4

Table 3
Percentages for some intensifiers.

Intensifier	Modifier (%)
slightly	−50
somewhat	−30
pretty	−10
really	+15
very	+25
extraordinarily	+50
(the) most	+100

Taboada, M., et al. : Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics, Vol. 37, No. 2, 267-307 (2011)

Základné problémy analýzy názorov

Nositeľmi postojov sú hlavne prídavné mená (perfektný), príslovky (katastrofálne), podstatné mená (bomba, hlúposť), slovesá (zničiť).

- Určenie subjektivity slova (nahrávanie klasifikačných slovníkov)
- Určenie orientácie, resp. polarity slova - pozitívna, negatívna a neutrálna (priemerný)
- Určenie sily polarity slova – stupnica intenzity orientácie (slovné a číselné vyjadrenie).

Základné problémy analýzy názorov je možné riešiť pomocou klasifikačných slovníkov (vyhodnocovanie zhody slov príspevku a slovníka)

Základné problémy analýzy názorov

- Určenie sily polarity slova – veľkosť podpory slova k potvrdeniu alebo vyvráteniu názoru
- Slovné a numerické stupnice (vhodnejšie pre spracovanie počítačom).

Počet stupňov	Stupnice	
2	negatívna	pozitívna
6	slabo negatívna, mierne negatívna, silno negatívna	slabo pozitívna, mierne pozitívna, silno pozitívna
10	-5, -4, -3, -2, -1	1, 2, 3, 4, 5

Problémy analýzy názorov

- Určenie sily polarity slova – stupnica so 6 hodnotami

+3 silno pozitívna	perfektný, vynikajúci, božský, úžasný
+2 mierne pozitívna	pekný, chvályhodný, kvalitný, šikovný
+1 slabo pozitívna	vhodný, dobrý, frajerský, fajn
-1 slabo negatívna	slabší, priemerný, nemastný, neslaný
-2 mierne negatívna	zlý, nefunkčný, slabý, nevyhovujúci
-3 silno negatívna	otrasný, katastrofálny, najhorší, úbohý

- Intenzifikácia – posuv polarity do vyššej/nižšej roviny
  - amplifier: prekvapujúco pekný, vysoko kvalitný
  - downtowner: o dosť slabší, nehorázne nekvalitný
- Negácia – preklopenie polarity

Intenzifikácia a negácia

- Spracovanie negácie (nie, ne...):
  - preklopenie polarity (switch negation)
  - posun polarity (shift negation) k opačnej polarite o fixnú hodnotu, napríklad „6“

  prídavné meno „a + 3“ je negované na „a − 3“ – podobné switch
  ale prídavné meno „a − 5“ je iba „a + 1“ – nepodobné switch
  „She’s not terrific (5 − 6 = − 1) but not terrible (−5 + 6 = 1) either.”
  - dynamický koeficient
- Intenzifikácia
  - zvyšuje/znižuje polaritu prostredníctvom slovníka

  really (+15) very (+25) good (3): 3x(100%+25%)x(100%+15%) = 4,3
  the most (+100) excellent (5): 5x(100%+100%) = 10
  - dynamickým koeficientom (nemusí za sebou)

Statický koeficient v negácii

Rozmanitosť vetných štruktúr v slovenčine – zápor môže byť pred ale aj za negovaným slovom aj ďalej od neho. („Povedal, že budem sklamaný, nebudem.“ 00023)

Mobil	nie	je	kvalitný
0	3	0	1
Tento	mobil	nebol	kvalitný
0	0	3	1
Tento	mobil	kvalitný	nebol
0	0	1	3

- Rovnaká polarita: 0301, 0031, 0013
  aj 3000010 „Nie je to podľa mňa kvalitný mobil“.
- Opačná polarita: 309 „Nie som najhorší“.
- Potreba prispôsobenia dĺžky kombinácie slov (dynamický koeficient)

Statický koeficient v intenzifikácii

- Slová zvyšujúce intenzitu polarity (zväčša príslovky) Uplatní sa iba v spojení s inou kategóriou stupňa polarity, napr.: 00041, 4002, (dynamický koeficient).
- Koeficient by mal zabrániť izolácii intenzifikátora (resp. záporu) od slova, ku ktorému sa vzťahujú (K=4).

Ten	mobil	je	totálne	kvalitný
0	0	0	4	1
neutrálne	neutrálne	neutrálne	+ intenzita	mierne pozitívne

Dosť	ma	to	hnevá
4	0	0	2
+ intenzita	neutrálne	neutrálne	mierne negatívne

Typovanie kombinácií slov

Každá z kombinácií reprezentuje práve jednu interpretáciu a je jej priradená práve jedna hodnota polarity.

Interpre -tácia	SP + I	SP MP + I	MP	MN	SN MN + I	SN + I
K = 2	48	80, 41	10, 32, 23	20, 31, 13	90, 42	49
K = 3	480,408	800, 410, 401	100, 320, 230, 302, 203	200, 310, 130, 301, 103	900, 420, 402	490, 409
K = 4	4800, 4080, 4008	8000, 4100, 4010, 4001	1000, 3200,2300, 3020,2030, 3002,2003	2000, 3100,1300, 3010,1030, 3001,1003	9000, 4200, 4020, 4002	4900, 4090, 4009
polarita	**3**	**2**	**1**	**-1**	**-2**	**-3**

Statický koeficient

KLAN – systém KLAsifikácie Názorov

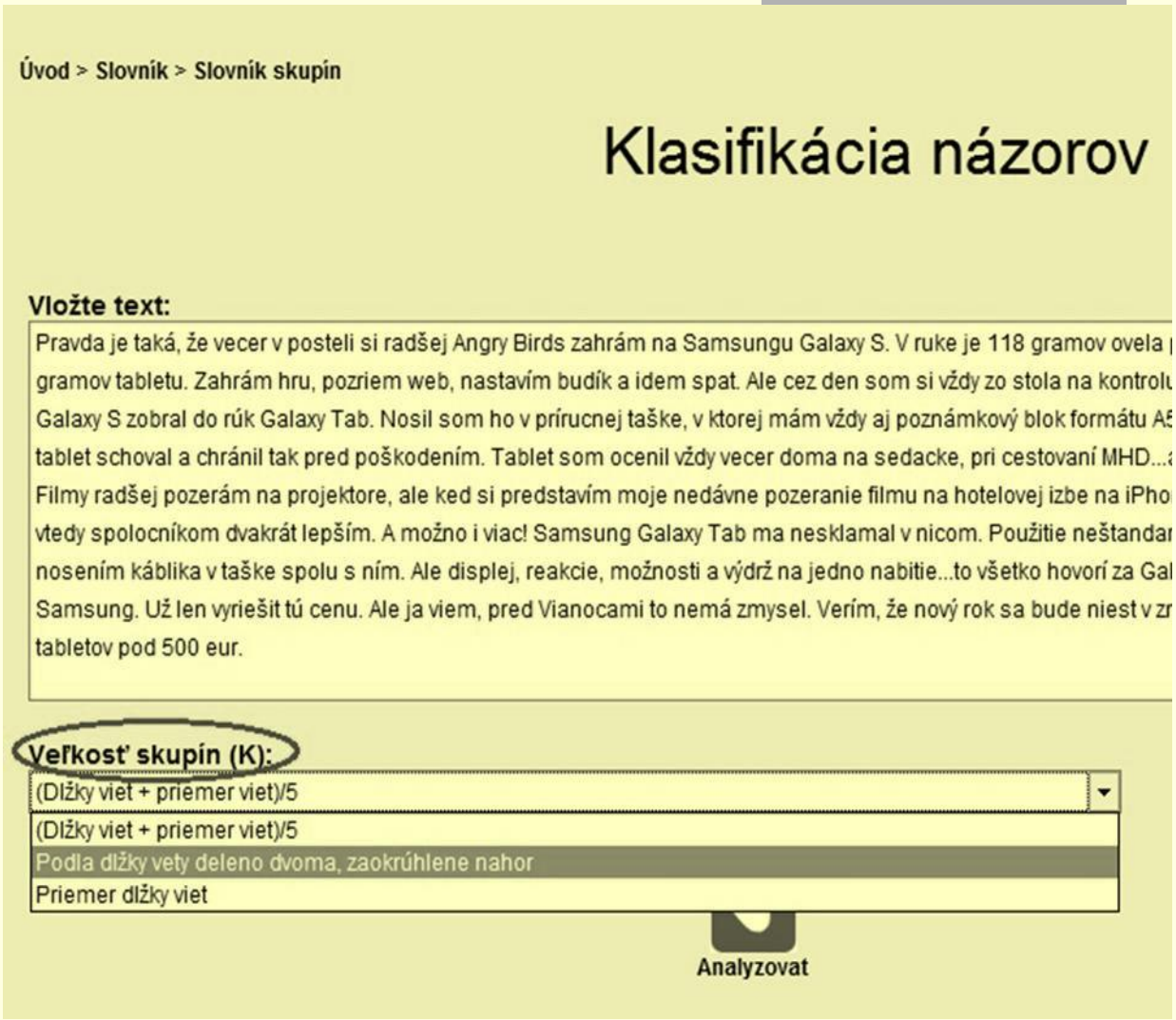
- Rozhranie „Guest“ môže klasifikovať zvolený text a nastavovať statický koeficient K.
- Rozhranie „Admin“ môže nahrávať a editovať klasifikačný slovník.

Dynamický koeficient

- Priemerná dĺžka vety
  - početnosť slov každej lexikálnej jednotky analyzovaného textu
  - aritmetický priemer
  - dynamický koeficient je rovnaký pre všetky vety
- Polovica dĺžky vety
  - početnosť slov lexikálnej jednotky delený dvoma so zaokrúhlením na hor
  - dynamický koeficient sa nastavuje zvlášť pre každú vetu analyzovaného textu
- Hybridný prístup
  - (dĺžka lexikálnej jednotky + priemerná hodnota všetkých viet) delené piatimi

Dynamický koeficient

- Priemerná dĺžka
- Polovica dĺžky
- Hybridný prístup



Použitie n - gramov

Dynamický koeficient rozdelí text do lexikálnych jednotiek, ktoré sa neprekrývajú. Môže dôjsť k izolácii negácie alebo intenzifikátora od vzťahovaného slova (neuspokojivé riešenie).

- ❑ Používali sme 4-gramy (riešenie problému izolácie)
- ❑ Cyklický posuv o jedno slovo

„Naozaj je to pekné a na viac aj veľmi praktické.“

4-gramy:

„naozaj je to pekné“	*P = 1 x (1+0,5) = 1,5*
„je to pekné a“	*P = 1*
„to pekné a na“	*P = 1*
„pekné a na viac“	*P = 1*
„a na viac aj“	*P = 0*
„na viac aj veľmi“	*P = 0 x 1 = 0*
„viac aj veľmi praktické.“	*P = 1 x (1+1) = 2*

Použitie n - gramov

Dva slovníky

- 1.slovník – 1.suma
  riešenie základných problémov
  (skladanie jednoduchých polarít)
  prídavné, podstatné mená, slovesá a emotikony
- 2.slovník – násobenie 2. sumou
  negácia a intenzifikácia (posuvy polarity)
  príslovky a negácie

$$P = \sum v(w_i^1)[1+\sum v(w_j^2)]$$

Použitie n - gramov

- Ukážky slovníkov používaných v aplikácii analýzy názorov použitím 4-gramov

Stupeň polarity	Slová a emotikony
3	:D, boží, špičkový
2	:), super, vynikajúci
1	pekný, funkčný, praktický
-1	nepríjemný, slabý
-2	:(, otrasný, chatrný
-3	:((, mizerný, katastrofálny

Pozitívny	Negatívny
:)	:(
:))	:((
:)))	:(((
:-)	:-(
=)	=(
:D	
=D	

Stupeň polarity	Intenzifikátory a negátory
1	veľmi, dokonale, výnimočne
0.5	vhodne, naozaj, fakticky
-0.5	málo, príliš, zbytočne
-2	negácie: nie,nie je, ne, nebol …

Použitie n - gramov

Príklady výpočtu polarity

- Jednoduché polarity

„Ako samotná myška je pekná, ale spracovanie je mizerné a celkovo je nepodarená.“
pekná(+1) + mizerné(-3) + nepodarená(-1)
P = 1 + (-3) + (-1) = -3

- Negácia

„Nie je to dobré riešenie.“
násobené: Nie(-2), pripočítané: dobré(+1)
*P = 1 * (1 + (-2)) = 1* (-1) = -1*

- Intenzifikácia

„Celkovo je spracovanie veľmi slušné.“
násobené: veľmi(+1), pripočítané: slušné(+1)
*P = 1 * (1 + 1) = 1 * 2 = 2*

Testy implementácií

Statický koeficient

http://www.mobilmania.sk (diskusné vlákno recenzií k mobilu LGKU990)

Dynamický koeficient

http://recenzie.sme.sk

N-gramy 1

http://www.mojandroid.sk (diskusné vlákno k mobilom HTC One X a HCT One S)

http://www.pocitace.sme.sk (diskusné vlákno k produktom Asus Transformer Prime TF201 and Asus Transformer Pad TF300T)

N-gramy 2

http://tech.sme.sk (recenzie telefónu Samsung Galaxy S4)

http://www.mojandroid.sk (recenzie telefónov HTC ONE a Samsung Galaxy S4)

Version	Positive	Negative	Average precision
Static coefficient	0.86	0.69	0.78
Dynamic coefficient 1	0.76	0.84	0.80
Dynamic coefficient 2	0.80	0.88	0.84
Hybrid	0.80	0.84	0.82
N-grams 1	0.83	0.57	0.70
N-grams 2	0.76	0.42	0.59

Diskusia k analýze názorov

- Odhaľovanie skrytej irónie „Táto úžasná kniha mi pomohla znížiť IQ o päť bodov!“ (čierna ovca) a dvojzmyslov „Ovládanie je jednoduché pre jednoduchých ľudí.“
- Názor vyjadrený nepriamo (text obsahuje iba neutrálne slová): „Už štyri roky dovolenkujeme v tejto destinácii.“

Ďalšie problémy znižujúce úspešnosť klasifikácie názorov

- Slovo s kladnou (zápornou) orientáciou nesie opačný postoj (zápor posunutý do inej lexikálnej jednotky): „Mám rád dobré knihy. Tak nič, možno nabudúce.“
- Prídavné mená a príslovky majú opačnú orientáciu ako sa predpokladalo:
  „Tento výrobok je obdivuhodne zbytočný.“

Identifikácia autorít

Autorita spravidla overená (reálne situácie, sociálny web?)
Typy autorít:

- Neformálna, prirodzená
  - schopnosti, primerané sebavedomie, osobný profil, sociálne aktivity, ...
  - posilňovaná rešpektom vedených ľudí
  - čestnosť, statočnosť, rozhodnosť, predvídateľnosť – odhad
- Formálna
  - pozícia, titul, funkcia v organizácii
  - status podlieha zmene
  - vyžadovaná poslušnosť, podriadenosť

- Formálna a prirodzená autorita môžu byť totožné.
- Formálna autorita sa môže meniť na prirodzenú a vice versa.

Identifikácia autorít webu

Typy autorít:

- Priateľ
  - veľké množstvo priateľov v rámci sociálneho webu
  - autorita podporovaná vzťahmi
- Šíriteľ vplyvu (influencer)
  - často citovaný (odvolávaju sa na jeho autoritu)
  - zaujme iných (prekvapí, ohromí ...)
  - autorita podporovaná názormi, vedomosťami o objekte diskusie

- Dolovanie zo štruktúry
- Dolovanie z obsahu

Identifikácia autorít webu (2)

Prístupy:

- Autority vo vede
  - vedecké články na osobných stránkach, v profiloch
  - digitálne knižnice ...
- Autority vo webových diskusiách
  - diskusie k produktom, recenzie filmov, kníh
  - sociálne siete ...

Autority vo vede

- ACM Digital Library
- IEEE Database
- Definícia vedeckej oblasti (kľúčová fráza)
- Sústredenie na referencie
  - variabilita citačných štandardov – netriviálna identifikácia autorov
  - prvý autor – najväčší podiel (?)
- Celkový počet citácií v danej oblasti
- Vizualizácia prostredníctvom TagClouds

Autority vo vede

Nájdené autority v oblasti „opinion analysis“ (ACM)

Poradie	Slovo	Počet celkových výskytov	Počet dokumentov s výskytom
1	Wiebe	34	9
2	Lee	22	17
3	Pang	19	15
4	Liu	19	9
5	Chen	18	7
6	Cardie	17	8
7	Wilson	17	6
8	Janyce	16	4
9	Zhang	13	8
10	Yu	12	10

Autority vo vede

Nájdené autority v oblasti „opinion analysis“ (IEEEE)

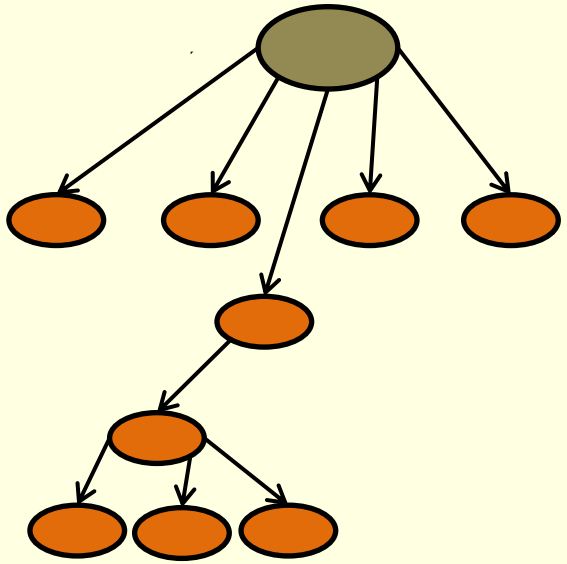
Poradie	***Slovo***	***Počet celkových výskytov***	***Počet dokumentov s výskytom***
1	*Liu*	*26*	*11*
2	*Lee*	*25*	*13*
3	*Chen*	*20*	*12*
4	*Pang*	*18*	*11*
5	*Zhang*	*18*	*8*
6	*Li*	*16*	*9*
7	*Kim*	*16*	*9*
8	*Hu*	*15*	*8*
9	*Hovy*	*14*	*7*
10	*Xu*	*12*	*7*

Autority vo vede

Výsledky vyhľadávania autorít v oblasti „opinion analysis“ vizualizované v TagCloud

Autority vo webových diskusiách

- Každý používateľ SW:
  - založenie diskusie
  - prispievanie do diskusie
- Nie každý je autoritou - rozpoznať rady odborníka
- Štruktúra diskusie – acyklický strom



Autority vo webových diskusiách (2)

Dôvody prispievania do diskusie

- Hľadanie odpovedí
  - rozhodovanie, informované rady od múdrejších, očakávanie pravdivých informácií
  - nie sú autority, je ich najviac, jadro fóra
- Príležitosť prezentovať sa, svoju dôležitosť
  - nepravdivé informácie, vyvolávanie konfliktov, degradovanie diskusie
  - problematickí provokatéri, vylúčenie, riadenie diskusie
  - nie sú autority, nie je ich veľa
- Príležitosť vyjadriť vedomosti
  - uistenie sa o správnosti nápadov, revidovanie názorov
  - pravdivé informácie, seriózny prístup, prispievajú iba keď sa cítia orientovaní
  - sú to autority, je ich málo

Vyvinuli sme prístup k odhadu autorít

Dolovanie autorít

Vstupné (predspracované) dáta obsahujú:

- meno prispievateľa
- polarita príspevku
- dĺžka príspevku
- príspevky - reakcie
- pozícia príspevku v strome – štruktúra diskusie

Tieto dáta vstupujú do procesu odhadu autority

Autorita nie je vzťahovaná k príspevkom, ale k prispievateľom (integrácia všetkých informácií o prispievateľovi).

Dolovanie autorít

V procese odhadu autority sa vytvára
zostupne usporiadaný rebríček
indikujúci prispievateľov:

- Prezentujúcich znalosť problematiky
- vyvolávajúcich mnoho reakcií
- inicializujúcich najčastejšie prechod na novú tému

Prístup k odhadu autorít

Primárne vplyvy:

- počet príspevkov prispievateľa (PP)
- počet reakcií na príspevky prispievateľa (PR)
- počet výskytov na koncovej úrovni stromu (PKU)

Sekundárne vplyvy:

- zhoda polarity (ZP)
- pozície v strome (počet úrovní - PU)
- počet termov (PT)

$$OA = 4PP^2 + 2PR^3 + 4PKU + ZP + PU + PT$$

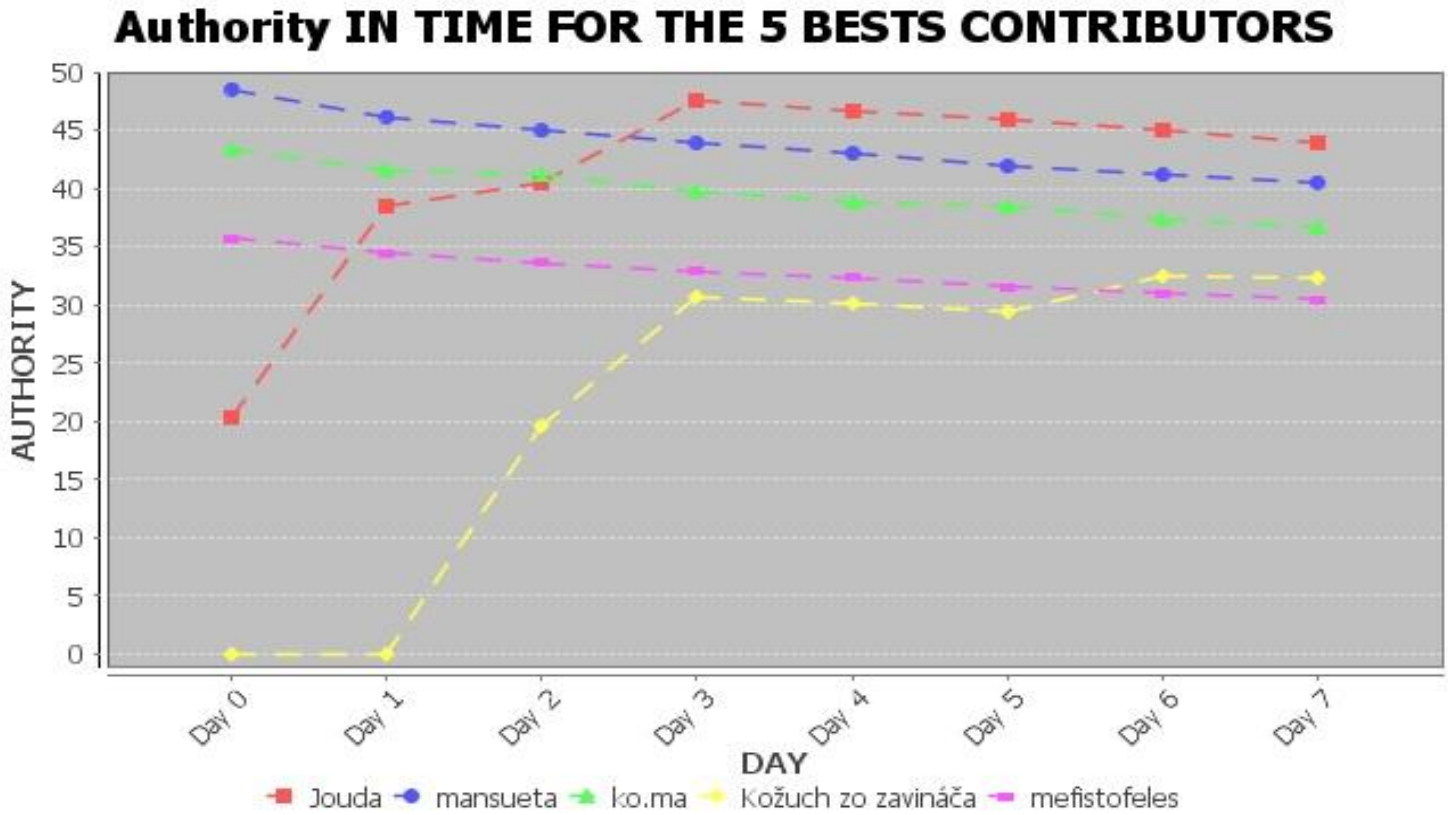
Prístup k odhadu autorít

Testovanie výsledkov navrhnutého prístupu:

Téma diskusie	Presnosť
Autorita a počet “likes”	0.94
Slovenskí politici	0.96
Bomby, letecké útoky a sirény	0.93

Dynamická zmena autority

Sledovanie dynamickej zmeny autority pre päť najvýraznejších prispievateľov:



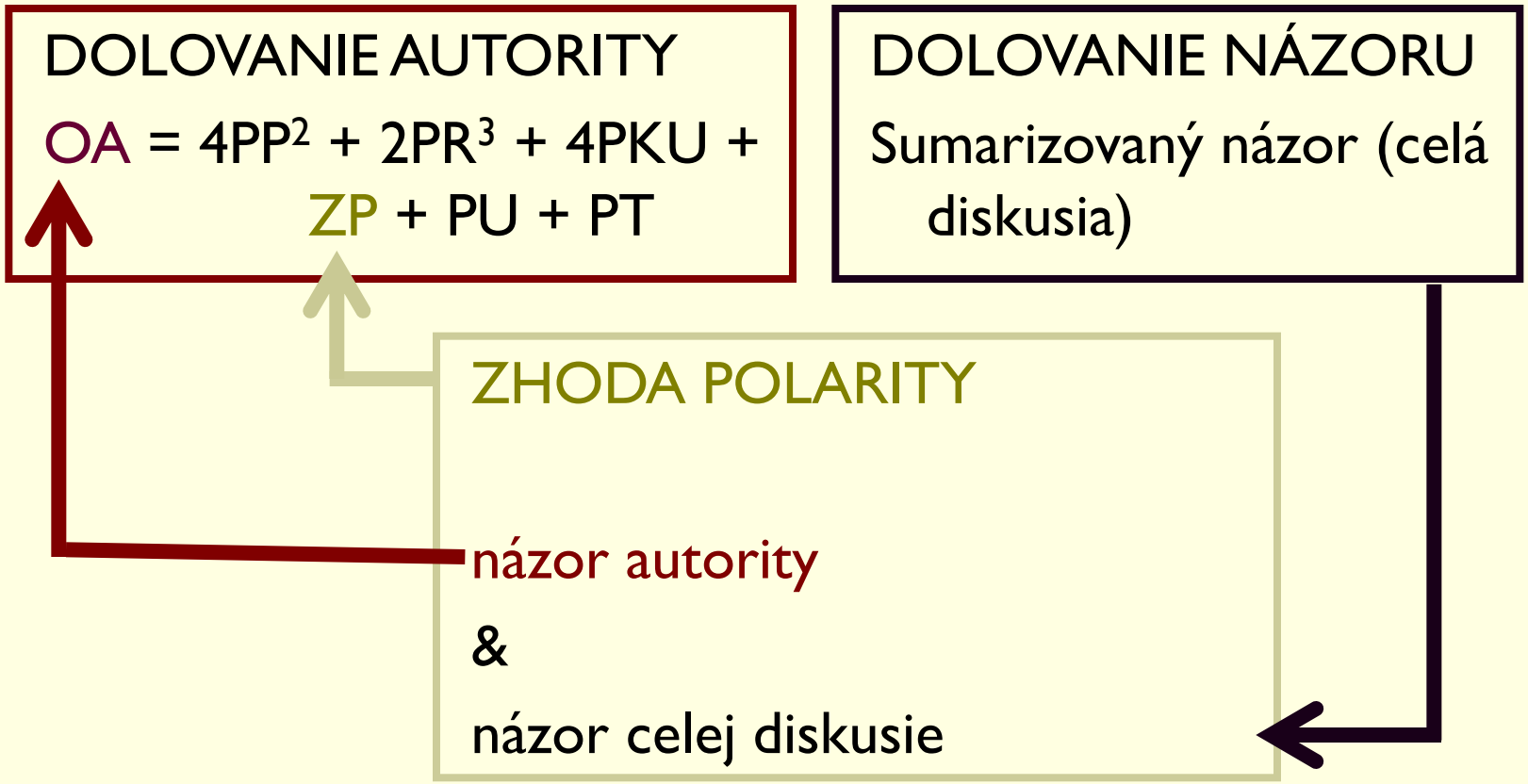
Diskusia k odhadu autorít

Implementácia metódy odhadu autorít:

- Bola testovaná s veľmi dobrými výsledkami na:
  - doméne z reálneho života
  - doméne z technickej oblasti
- Kombinuje dolovanie zo štruktúry s dolovaním z obsahu
- Dá sa použiť na vylepšenie klasifikácie názorov
  - každý príspevok má rovnakú váhu
  - každý príspevok sa svojou pozitivitou/negativitou podieľa na sumarizovanom názore s určitou váhou – vyčíslená autorita
- Nulté kolo pohovoru (organizácia založí profesionálnu diskusiu)

Vzťah medzi dolovaním názorov a autorít

1. **Použite vydolovaného názoru v dolovaní autority**



2. **Použitie informácie o autoritách v dolovaní názoru**

Použitie informácie o autoritách v dolovaní názoru

DOLOVANIE NÁZORU

Klasická cesta:

Sumár názorov diskusie

Nový prístup:

prispievateľ1→názor11

prispievateľ2→názor21

prispievateľ1→názor12

etc.

p1 (autorita w1)→n11

p2 (autorita w2)→n21

p1 (autorita w1)→n12

etc.

ODHAD AUTHORITY

OA = $4PP^2 + 2PR^3 + 4PKU + ZP + PU + PT$

prispievateľ1 → $AC_1 = w_1$

etc.

prispievateľN → $AC_N = w_N$

Váhovaný názor celej diskusie

$VN = \sum_k [w_k \sum_j n_{kj}]$

$w_k \in <0,1>$; $n_{kl} \in \{-1,+1\}$ (resp. <-5,+5>)

Ďakujem za pozornosť

kristina.machova@tuke.sk