

PREPÁJANIE MULTIMEDIÁLNYCH METADÁT S VYUŽITÍM MIKROBLOGOVEJ SIETE

Peter Gašpar ■ Vedúci: Jakub Šimko

NÁŠ CIEĽ

Nájsť **prepojenie** medzi TV programom a mikrobloggermi. Toto prepojenie nám umožní extrahovať ďalšie **metadáta z mikrobloggerov**.

1

PREDSPRACOVANIE TV PROGRAMU

Z názvu TV relácie získame pomenované entity, hashtagy a prefixový index. **PREFIXOVÝ INDEX** pozostáva z prvých N začiatkových písmen názvu TV relácie

názov TV relácie
Teória veľkého tresku

TELEVÍZNE METADÁTÁ

pomenované entity	Teória veľkého tresku
hashtagy	#teoriavelkehotresku
prefixový index	Teória veľkeho tresku

2

EXTRAKCIA ČŔT Z PRÍSPEVKU

Z každého príspevku extrahujeme nasledujúce črty: **pomenované entity, hashtagy, externé prepojenia**.



Pomenované entity

Slová začínajúce na znaky z prefixového indexu.



Hashtagy

Slová začínajúce na "#".



Hashtagy z URL prepojenia

Ak príspevok obsahuje URL, rozdelíme ju na časti oddelené "/".

PRÍKLAD PRÍSPEVKU

Sledujte novú sériu **Teórie veľkého tresku** dnes na JOJ
#tvt <http://teoriavelkehotresku.sk>

3

NÁJDENIE MAPOVANIA

Črty extrahované v kroku 2 vyhladávame v televíznych metadátach z kroku 1.

PRÍSPEVOK

extrahované
pomenované entity

N-GRAMOVÁ PODOB.

TV RELÁCIA

generované
pomenované entity

extrahované
hashtagy

LEVENSHTEINOVA VZD.

generované
hashtagy

extrahované
hashtagy z prepojenia

N-GRAMOVÁ PODOB.

generované
hashtagy

Celková podobnosť medzi príspevkom a TV reláciou je vypočítaná ako priemer všetkých podobností.

Výsledkom mapovania pre príspevok je zoznam TV relácií s príslušnými podobnosťami. TV relácie, ktoré majú podobnosť nižšiu ako **hranica** sú zo zoznamu vyradené.



Výsledky mapovania používame v ďalšej fáze nášho riešenia - **ko-učenie**. Túto fázu ešte aktívne vyvíjame.

OVERENIE

Pre overenie nášho riešenia sme anotovali **1300 príspevkov** odoslaných Facebook profilmi 2 slovenských a 8 anglických televíznych staníc.

Postupné pridávanie črt príspevku:

PRESNOSŤ
85%



PRESNOSŤ
91%



PRESNOSŤ
93%



Nami navrhnuté vylepšenia
taktiež zvýšili presnosť:

ZLEPŠENIE
+7%
**PREFIXOVÝ
INDEX**

ZLEPŠENIE
+35%
HRANICA
(> 0.70)

ZLEPŠENIE
+12%
KO-UČENIE

Uvedené hodnoty uvádzajú zmenu
presnosti voči nastaveniu bez vylepšení.