



Segment publika v sociálních médiích

KLÁRA VOSTŘELOVÁ

SÉMANTIKA A POČÍTAČE

9. 5. 2017 BRNO

Analýza publika

- ▶ K analýze sociálních médií stále důležitější
- ▶ Různé atributy publika – umístění autora, demografie, chování, zájmy a typy účtů
- ▶ Aplikovatelná na různé zdroje sociálních médií i na více jazyků
- ▶ V rámci programu IBM Watson Analytics

IBM Watson Analytics

- ▶ Projekt firmy IBM
- ▶ Inteligentní analýza dat s jejich automatickou vizualizací
- ▶ Pro obchodníky – měření výsledků, lepší pochopení, jak svět vnímá jejich společnost, značku a produkty nebo služby
- ▶ Komunikace s daty konverzačně – díky zjišťování vedených dat, automatizované prediktivní analýze a kognitivním schopnostem, jako je dialog v přirozených jazycích

Watson Analytics for Social Media

- ▶ Nástroj pro analýzu sociálních médií
- ▶ Založený na technologii cloudu
- ▶ Pro uživatele firemních linek z oblasti veřejných vztahů, marketingu nebo správy produktů
- ▶ Nástroj nezávislý na doméně – uživatelé seskupují témata, která jsou předmětem zájmu, např. produkty, značky, služby, politika atd.
- ▶ Na základě témat se načítá veškerý relevantní obsah:
 - ▶ Z různých zdrojů sociálních médií (Twitter, Facebook, blogy, fóra, recenze, komentáře k videím a zprávy)
 - ▶ Ve více jazycích (angličtina, arabština, francouzština, němčina, italština, portugalština a španělština)

Informace ke klasifikaci

- ▶ Jméno a příjmení autora
- ▶ Autorská biografie – krátký, volitelný popis autora, často zahrnuje umístění autora
- ▶ Obsah autora – příspěvky v sociálních médiích (Tweety, blogové příspěvky, příspěvky ve fórech, recenze), obsahují témata nastavovaná uživatelem WASM, nikoliv všechny příspěvky autora
 - ▶ Dva důvody:
 - 1) Mnoho zdrojů (fóra) nenačítá obsah založený na jménu autora
 - 2) Poskytovatelé obsahu na sociálních médiích účtují za dokumenty – nebylo by to ekonomické
- ▶ Kombinace obsahu autora s jménem a biografií autora významně zvyšuje přesnost rozdělení publika

Klasifikace publika

Umístění autora

- ▶ Město nebo oblast, kde autor žije
- ▶ Informace z biografie se porovnávají s velkými seznamy názvů měst, regionů a zemí – upřednostňují místa s vyšším počtem obyvatel
- ▶ Např. autor ze Stuttgartu je přidělen Stuttgartu v Německu, ne v USA

Chování autora

- ▶ Uživatelé **vlastní** určitý produkt nebo **využívají** určitou službu
 - ▶ Porovnání frází se slovníkem: *mám telefon XY* nebo *používám servis XY*
- ▶ **Potenciální** uživatelé – mají zájem o koupi produktu nebo služby
 - ▶ Porovnání frází se slovníkem: *hledám telefon XY* či *potřebuji telefon XY*
- ▶ **Bývalí uživatelé** – přestali používat produkt či službu
 - ▶ Porovnání frází se slovníkem: *prodal jsem svůj telefon XY*

Klasifikace publika

Demografie autora

- ▶ **Rodinný stav:** autor se oženil – hodnota je pravdivá, jinak neznámá
 - ▶ WASM slovníky a vzory shodné např. s *ženatý, manžel* nebo *na narozeniny mé ženy*
- ▶ **Rodičovský stav:** autor má děti – hodnota je pravdivá, jinak neznámá
 - ▶ Klíčová slova a vzory shodné s těmi ve slovníku – např. *otec* nebo *moje děti*
- ▶ **Pohlaví:** hodnoty – mužské, ženské a neznámé
 - ▶ Klíčová slova a vzory shodné s těmi ve slovníku a navíc jméno autora je v databázi 150 000 křestních jmen různých kultur
 - ▶ Nejednoznačnost jmen – Např. *Andrea* – identifikace pomocí jazyka

Klasifikace publika

Zájmy autora

- ▶ **První stupeň:** 8 kategorií zájmových skupin – *umění a zábava* nebo *sport*
- ▶ **Druhý stupeň:** přibližně 60 konkrétnějších kategorií – *hudba a filmy* pod *uměním a zábavou*, *míčové sporty* a *bojové umění* v rámci sportů
- ▶ **Třetí úroveň:** jednotlivé zájmy – *tenis* a *fotbal* pod *míčovými sporty*

Typ účtu

- ▶ **Organizační:** společnosti, nevládní organizace, noviny, univerzity a fanoušci – např. *Corp*, *oficiální účet Twitter*
- ▶ **Osobní:** účty jednotlivých autorů – substantiva (*režisér*, *vývojář*, *milovník*), osobní a majetková zájmena (*já*, *můj*, *náš*)

Hodnocení typů účtů

- ▶ Založeno na náhodném vzorku 50 124 Tweetů
 - ▶ Zahrnuje 43 193 odlišných autorů
 - ▶ Biografii poskytuje 36 692 autorů
- ▶ systém WASM přiřadí typ účtu 18 657 autorům
 - ▶ 16 981 osobních
 - ▶ 1 766 organizačních
- ▶ Ruční ohodnocení 500 klasifikovaných účtů
 - ▶ přesnosti 97,4 %

Zdroje

- ▶ <http://aclweb.org/anthology/E/E17/E17-3014.pdf>
- ▶ <https://www.ibm.com/analytics/watson-analytics/us-en/>

Děkuji za pozornost