



AGH

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

WYDZIAŁ INFORMATYKI, ELEKTRONIKI I TELEKOMUNIKACJI

KATEDRA TELEKOMUNIKACJI

Praca magisterska

*Klasteryzacja tematów aktywności użytkowników w sieci
społecznościowej*

Clustering topics of user activity in social networking service

Autor:

Kacper Żuk

Kierunek studiów:

Teleinformatyka

Opiekun pracy:

dr inż. Jarosław Bułat

Kraków, 2018

Uprzedzony o odpowiedzialności karnej na podstawie art. 115 ust. 1 i 2 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.): „Kto przywłaszcza sobie autorstwo albo wprowadza w błąd co do autorstwa całości lub części cudzego utworu albo artystycznego wykonania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3. Tej samej karze podlega, kto rozpowszechnia bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystycznego wykonania albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, wideogram lub nadanie.”, a także uprzedzony o odpowiedzialności dyscyplinarnej na podstawie art. 211 ust. 1 ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (t.j. Dz. U. z 2012 r. poz. 572, z późn. zm.): „Za naruszenie przepisów obowiązujących w uczelni oraz za czyny uchybiające godności studenta student ponosi odpowiedzialność dyscyplinarną przed komisją dyscyplinarną albo przed sądem koleżeńskim samorządu studenckiego, zwanym dalej «sądem koleżeńskim».”, oświadczam, że niniejszą pracę dyplomową wykonałem(-am) osobiście i samodzielnie i że nie korzystałem(-am) ze źródeł innych niż wymienione w pracy.

Spis treści

| | |
|--|----|
| 1. Wstęp | 5 |
| 2. Wprowadzenie | 7 |
| 3. Pozyskiwanie i analiza danych źródłowych | 9 |
| 3.1. Charakterystyka | 9 |
| 3.2. Pozyskanie i przetrzymywanie | 9 |
| 4. Klasteryzacja tematów aktywności | 11 |
| 4.1. Graf relacji hashtagów | 11 |
| 4.2. Metryka powiązania hashtagów | 11 |
| 4.3. Klasteryzacja | 11 |
| 5. Optymalizacja wydajności | 13 |
| 5.1. Przetwarzanie strumieniowe | 13 |
| 5.2. Przetwarzanie równoległe | 13 |
| 6. Analiza wyników | 15 |
| 6.1. Wizualizacja grafu relacji | 15 |
| 6.2. Prezentacja praktycznych zastosowań | 15 |
| 7. Podsumowanie | 17 |

1 Wstep

2 Wprowadzenie

Dlaczego Twitter, co jeszcze jest innego, co inni w tym temacie porobili.

3 Pozyskiwanie i analiza danych źródłowych

3.1. Charakterystyka

3.2. Pozyskanie i przetrzymywanie

4 Klasteryzacja tematów aktywności

4.1. Graf relacji hashtagów

4.2. Metryka powiązania hashtagów

4.3. Klasteryzacja

5 Optymalizacja wydajności

5.1. Przetwarzanie strumieniowe

5.2. Przetwarzanie równoległe

6 Analiza wyników

6.1. Wizualizacja grafu relacji

6.2. Prezentacja praktycznych zastosowań

Rekomendacja treści w oparciu o aktywność

Miara podobieństwa użytkowników

7 Podsumowanie