

Dokumentacja projektu 2

Ewa Baranowska, Dorota Łepicka, Michał Mück, Michał Stolarczyk

12 czerwca 2016

Wstęp

Projekt ten miał na celu znalezienie dni w których użytkownicy licznie komentowali posty na fanpage'u banków. Analizie poddaliśmy kilka banków: Alior Bank S.A., Dane które otrzymaliśmy były zbierane od ... do kwietnia 2016r. Skupiliśmy się głównie na analizie rzeczowników. Posty nie zawierające rzeczowników zostały usunięte. Aby ułatwić pracę z danymi zbudowaliśmy ramkę danych zawierającą poniższe zmienne:

- id - id postu,
- parent_id - id rodzica postu, jeśli taki nie istnieje jest brak danych,
- created_at - data utworzenia postu,
- tread - numer wątku do którego należy post,
- user_name - nazwa autora postu,
- body - treść postu,
- rzeczownik - rzeczowniki które znajdują się w poście zapisane w formie podstawowej.

Każdemu rzeczownikowi w poście został przydzielony osobny wiersz. Przykład takiej ramki danych:

Aby umożliwić czytelnikowi odtworzenie wykonanych przez nas działań zbudowaliśmy pakiet o nazwie **ClockworkOranges**. Zawartość naszej biblioteki składa się z funkcji:

1) Służących do przetwarzania danych:

2)