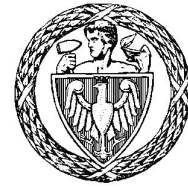


Politechnika Warszawska

W Y D Z I A Ł E L E K T R Y C Z N Y



Instytut Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych
Zakład Elektrotechniki Teoretycznej i Informatyki Stosowanej

Praca dyplomowa magisterska

na kierunku Informatyka
w specjalności Inżynieria oprogramowania

Analiza wiadomości z serwisów społecznościowych na użytek
rozpoznawania nastrojów.

Jakub Rzepliński

Numer albumu: 233608

promotor
dr inż. Marcin Kołodziej

Warszawa, 2018

Streszczenie

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Słowa kluczowe: jakieś, słowa, kluczowe, po polsku

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Keywords: some, keywords, in, english

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że przedstawiona praca dyplomowa:

- została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami,
- nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego lub stopnia naukowego w wyższej uczelni

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

.....
data

.....
podpis autora (autorów) pracy

Oświadczenie

Wyrażam zgodę / nie wyrażam zgody^{*1} na udostępnianie osobom zainteresowanym mojej pracy dyplomowej. Praca może być udostępniana w pomieszczeniach biblioteki wydziałowej. Zgoda na udostępnienie pracy dyplomowej nie oznacza wyrażenia zgody na jej kopiowanie w całości lub w części.

Brak zgody nie oznacza ograniczenia dostępu do pracy dyplomowej osób:

- reprezentujących władze Politechniki Warszawskiej,
- członków Komisji Akredytacyjnych,
- funkcjonariuszy służb państwowych i innych osób uprawnionych, na mocy odpowiednich przepisów prawnych obowiązujących na terenie Rzeczypospolitej Polskiej,

do swobodnego dostępu do materiałów chronionych międzynarodowymi przepisami o prawach autorskich. Brak zgody nie wyklucza także kontroli tekstu pracy dyplomowej w systemie antyplagiatowym.

.....
data

.....
podpis autora (autorów) pracy

^{*1} - niepotrzebne skreślić

Spis treści

1	Wstęp	1
1.1	Motywacja do pracy	1
1.1.1	Przykłady innych prac	3
1.2	Metodyka pracy	3
1.3	Plan pracy	4
2	Przegląd technologii	5
2.1	Metodyka przeglądu	5
2.2	Wyniki przeglądu	5
2.3	Kolejna sekcja	5
3	Serwis społecznościowy Twitter	7
3.1	Historia	7
3.2	Tweet	7
3.3	Architektura	7
3.4	API	8
3.4.1	Zakres działania	8
3.4.2	Rejestracja	8
3.4.3	Sposób działania	8
3.4.4	Ograniczenia	10
3.5	Dostępne narzędzia analityczne	10
3.5.1	Twitter Analytics	10
3.5.2	Hootsuite	10
3.5.3	Mentionmap	13
3.5.4	Podsumowanie	14
4	Analiza danych Big Data	17
4.1	Pojęcie Big Data	17
4.2	Charakterystyka	17
4.2.1	Objętość	17
4.2.2	Szybkość	18
4.2.3	Różnorodność i złożoność	18
4.2.4	Zmienność	19
4.2.5	Wartość	19
4.3	Paradygmat MapReduce	19
4.4	Wymagania stawiane przed systemami Big Data	20
4.5	Przykłady zastosowania systemów Big Data	20
5	Analiza sentymentu wypowiedzi	23
5.1	Sentyment wypowiedzi	23
5.2	Przetwarzanie języka naturalnego	23
5.3	Wymagania związane z oceną sentymentu	24

5.4	Zastosowanie badania wydźwięku wypowiedzi z internetu	24
-----	---	----

Rozdział 1

Wstęp

Witaj, to szablon pracy dyplomowej **częściowo** przystosowany do nowych wymagań edycyjnych PW¹. Zapoznaj się z kodem źródłowym przed przystąpieniem do pracy. Na początek wymienię kilka ważnych uwag:

- ustawienia dokumentu znajdują się w pliku `config.tex`,
- niestety ze względu na użycie płatnych fontów w stronie tytułowej aktualna wersja używa plików `.png`²,
- bibliografia używa stylu zbliżonego do stylu harwardzkiego(**ogata2010modern**), a cytowania są w większości zgodne z zaleceniami BG PW (**linh2002line**); można to zmienić w pliku konfiguracyjnym(**DUMMY:1**) na cytowania numeryczne,
- do kompilacji polecam TeXstudio 2.10.8 + MiKTeX w sekwencji `pdflatex + biber + pdflatex + pdflatex`³ (czasem warto powtórzyć 2-3 razy, by zaktualizowała się bibliografia); do podglądu zaś Sumatra PDF,
- do zarządzania bibliografią polecam oprogramowanie Zotero i opcja Better BibTeX.

Pozdrawiam, Dominik Roszkowski.

1.1 Motywacja do pracy

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed

¹Niektóre elementy z Zarządzenia Rektora nr 43/2016 są bardzo trudne do zrealizowania w \LaTeX -u.

²Jeśli uda się dostać nagłówki i rodzaj pracy w krzywych, to zostaną one podmienione.

³By dodać w TeXstudio przejdź do Opcje > Build > User commands i dodaj: `txs:///pdflatex | txs:///biber | txs:///pdflatex | txs:///pdflatex`



Rysunek 1.1: Długi podpis grafiki opisujący na przykład, co znajduje się na niej po lewej a co po prawej stronie rozciągający się na więcej niż jedną linijkę

Tablica 1.1: My caption of this table

L.p.	$\Re\{\underline{x}(m)\}$	$-\Im\{\underline{x}(m)\}$	$\underline{x}(m)$	$\frac{\underline{x}(m)}{23}$	A_m	$\varphi(m) / ^\circ$	$\varphi_m / ^\circ$
1	16.128	8.872	16.128	1.402	1.373	-146.6	-137.6
2	1.29	0.099	1.29	0.112	0.097	-175.6	-114.7
3	16.128	8.872	16.128	1.402	1.373	-146.6	-137.6
4	1.29	0.099	1.29	0.112	0.097	-175.6	-114.7
5	16.128	8.872	16.128	1.402	1.373	-146.6	-137.6
6	1.29	0.099	1.29	0.112	0.097	-175.6	-114.7
7	0.641	-0.466	0.641	0.056	0.045	133.3	-106.3

accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.



(a) A subfigure 1



(b) A subfigure 2

Rysunek 1.2: A figure with two subfigures that you can reference easily e.g. in Figure 1.2a and Figure 1.2b

1.1.1 Przykłady innych prac

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

1.2 Metodyka pracy

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipi-

scing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

1.3 Plan pracy

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Rozdział 2

Przegląd technologii

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

2.1 Metodyka przeglądu

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

2.2 Wyniki przeglądu

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

2.3 Kolejna sekcja

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc

eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Rozdział 3

Serwis społecznościowy Twitter

Serwis społecznościowy Twitter jest globalnym serwisem internetowym służącym głównie do zamieszczania wiadomości tzw. *tweet*, które użytkownicy tego serwisu mogą także czytać, komentować lub przekazywać dalej. Od kilku lat Twitter jest serwisem gdzie dochodzi do wymiany zdań na różny temat, dotyczących np. polityki, sportu, produktów, wydarzeń społecznych, a profile posiada wiele osób znanych publicznie oraz instytucji.

3.1 Historia

Serwis ten, nazywany SMS internetu, został założony w 2006 r. w Stanach Zjednoczonych przez Jacka Dorsey'a, Ev Williamsa, Noah Glassa oraz Biza Stone'a i od początku powstania sukcesywnie zwiększał swoją popularność poprzez wzrost liczby użytkowników odwiedzających jego witrynę oraz wysyłających wiadomości. W 2012 r. osiągnął ponad 100 milionów użytkowników, którzy zamieszczali łącznie ponad 340 milionów wiadomości dziennie oraz obsługiwał średnio około 1.6 miliarda wyszukujących zapytań dziennie. W 2013 r. Twitter stał się jedną z najczęściej odwiedzanych stron w całym internecie. W tym samym roku inżynierowie Twittera podali informację, że serwis ten obsługuje ok. 143 tys. wiadomości na sekundę. Na początku 2016 r. serwis ten posiadał ponad 319 milionów użytkowników aktywnych podczas każdego miesiąca. Od listopada 2013 r. akcje Twittera są obecne na nowojorskiej giełdzie.

3.2 Tweet

Tweet, czyli krótka wiadomość tekstowa, była początkowo ograniczona do 140 znaków, ale limit ten został podwojony w 2017 r. dla wszystkich języków oprócz chińskiego, japońskiego i koreańskiego. Użytkownicy mają możliwość wyróżniania wybranych przez siebie tematów przez dodanie do nich znaku '#', co czyni takie wyrażenie tagiem. Inną możliwością oferowaną przez Twittera jest odpowiadanie innym użytkownikom lub zamieszczenie referencji do nich przez dodanie znaku '@' poprzedzającego nazwę profilu innej osoby.

3.3 Architektura

Serwis społecznościowy Twitter opierał się początkowo o typową architekturę trójwarstwową składającą się z warstwy prezentacji, logiki biznesowej oraz warstwy danych. Do napisania tej aplikacji został użyty framework Ruby on Rails wykorzystujący język Ruby, a warstwa bazodanowa opierała się o technologię MySQL. Jednak wraz ze wzrostem ilości przetwarzanych danych inżynierowie Twittera podjęli decyzję w 2011 r. o zmianie technologii na język Scala, który działa na maszynie wirtualnej Javy oraz zrezygnowano z dotychczasowej architektury na

rzecz budowy rozproszonych serwisów komunikujących się między sobą. Wraz z przeprowadzonymi zmianami zanotowano ponad 10-krotne polepszenie obsługi tweetów.

3.4 API

Twitter jest platformą otwartą i udostępnia programowalny interfejs API w dwóch postaciach: Search API oraz Streaming API.

3.4.1 Zakres działania

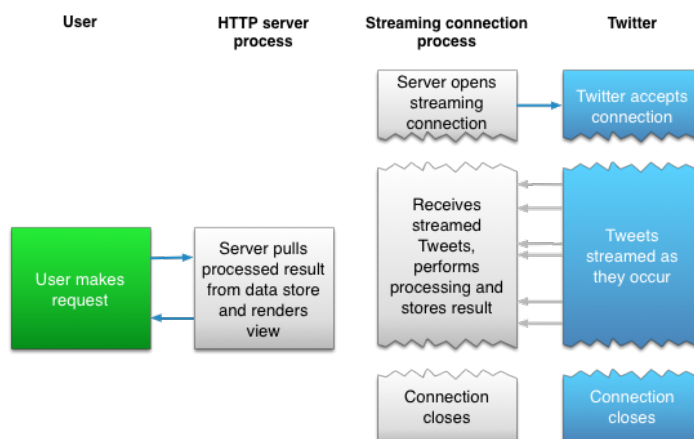
Programiści korzystający z Search API są w stanie uzyskać dostęp tylko do danych historycznych, które zostały już wcześniej zamieszczone na łamach serwisu Twitter. Natomiast w przypadku Streaming API dostajemy możliwość śledzenia strumienia danych, które są do naszego dostępu nawet już kilka sekund po zamieszczeniu w serwisie Twitter. Po podłączeniu do takiego strumienia możemy cały czas obserwować nowe wiadomości. Przyjęło się stosować nazywnictwo, że analiza Search API to analiza *back in time*, a Streaming API to śledzenie *real time*.

3.4.2 Rejestracja

Obie formy API wymagają wcześniejszej rejestracji na stronie <https://developer.twitter.com/en/apply-for-access> przeznaczonej dla deweloperów zainteresowanych wykorzystywaniem Twitter API. Po przejściu pomyślnej rejestracji dostajemy dane, które po nawiązaniu połączenia z serwisem Twitter umożliwiają mu jednoznacznie określić, że możemy uzyskać dostęp do API.

3.4.3 Sposób działania

Search API powstało z wykorzystaniem standardu REST - *Representational State Transfer*. Oba rodzaje API wykorzystują protokół HTTP: do poprawnego działania Streaming API potrzebne jest ciągłe połączenie HTTP, a w przypadku drugiego z nich każda operacja jest wykonywana przy nawiązaniu oddzielnego połączenia.



Rysunek 3.1: Schemat działania dwóch rodzajów programistycznego interfejsu API udostępnianego przez serwis społecznościowy Twitter: Search API i Streaming API.

Search API posiada ściśle określone parametry, które mogą być przesłane w żądaniu. W tabeli 3.1 zaprezentowano ich wykaz.

Tablica 3.1: Parametry żądania Twitter Search API

Parametr	Wymagany/Opcjonalny	Opis	Przykład
q	wymagany	zapytanie wyszuki- jące o maksymalnej długości 500 znaków	nasa
geocode	opcjonalny	zwraca wiadomości użytkowników od- dalonych o podany promień od podanej szerokości i długo- ści geograficznej, promień może być podany w milach lub kilometrach	37.781157 -122.398720 1mi
lang	opcjonalny	ogranicza wiadomo- ści do wybranego ję- zyka spośród dostęp- nych kodów ISO 639- 1	pl
locale	opcjonalny	specyfikuje język wy- syłanego zapytania, obecnie tylko <i>ja</i> jest skuteczny	ja
result_type	opcjonalny	określa typ zwraca- nych wiadomości, obecnie dostępne są trzy wartości tego parametru: <i>recent</i> (zwracane są najnowsze wia- domości), <i>popular</i> (zwracane są naj- bardziej popularne wiadomości) <i>mixed</i> (wartość domyślna, zwracane wyniki obejmują najnowsze i najbardziej popu- larne wiadomości)	mixed
count	opcjonalny	specyfikuje ilość zwracanych wiado- mości; maksymalna wartość to 100, a domyślna to 15	100

until	opcjonalny	ten parametr odpowiada za zwracanie wiadomości, których data utworzenia jest starsza o maksymalnie tydzień niż podana; obowiązuje format YYYY-MM-DD	2015-07-19
since_id	opcjonalny	dzięki temu parametrowi zwracane są wiadomości o ID większym niż podane czyli nowsze wiadomości niż określono	12345
max_id	opcjonalny	dzięki temu parametrowi zwracane są wiadomości o ID mniejszym lub równym niż podane czyli starsze lub takie same wiadomości niż określono	54321

Streaming API nie posiada takich ograniczeń. W języku programowania Java dostępny jest pakiet *twitter4j* zawierający interfejsy *User* oraz *Status*, na które mapowane są przychodzące ze strumienia informacje. W tabeli 3.2 zamieszczam ich dokumentację.

3.4.4 Ograniczenia

Korzystając z Search API mamy możliwość wysłania 720 zapytań na godzinę, a maksymalna ilość wiadomości jaka może być zwrócona na jedno zapytanie to 100. Jeśli wykorzystalibyśmy ten limit w maksymalny sposób to daje nam to 72 000 wiadomości na godzinę. W przypadku Streaming API głównym ograniczeniem jest dostęp do ok. 1 % danych ze strumienia, a maksymalna ilość wiadomości w czasie jednej minuty to 3 000. W przypadku tego API w ciągu godziny możemy uzyskać 180 000 wiadomości na godzinę. Są to ograniczenia, które obowiązują dla rozwiązań typu *open-source*.

3.5 Dostępne narzędzia analityczne

Udostępnienie API przez serwis społecznościowy Twitter oraz rosnące znaczenie danych generowanych przez użytkowników tego serwisu spowodowało, że wiele firm oraz instytucji zaczęło przywiązywać dużą wagę do analizy opinii wyrażanych na swój temat lub na tematy pokrewne, zainteresowania pewnymi tematami oraz kształtujących się trendów. Dlatego powstały aplikacje internetowe służące do wyświetlania takich informacji i przeprowadzające wstępną analizę zebranych danych. W dalszej części pracy znajduje się omówienie najważniejszych z nich.

3.5.1 Twitter Analytics

Pierwszym z narzędzi, którym warto poświęcić uwagę jest *Twitter Analytics*. Aplikacja ta posiada trzy zakładki. Na pierwszej z nich wyświetla statystyki dotyczące profilu z serwisu Twitter, którym logujemy się do niej: naszą najbardziej popularną wiadomość, najpopularniejszą wzmiankę o naszym profilu oraz wykresy trendów m. in. liczby osób śledzących nasz profil i odwiedzin. Na kolejnej zakładce mamy możliwość tworzenia wiadomości, które zostaną wygenerowane na naszym profilu oraz dodania do nich plików graficznych, materiałów audio lub wideo. Na ostatniej stronie użytkownik ma szansę poznania informacji takich jak np. lokalizacja geograficzna i wiek osób śledzących nasz profil czyli tzw. *followers*.

3.5.2 Hootsuite

Kolejną aplikacją jest *Hootsuite*. Jest to narzędzie, które umożliwia prowadzić kampanie w kilku mediach społecznościowych na raz np. *Facebook*, *Instagram*, *LinkedIn*. Posiada wiele rozbudowanych funkcji. Co ciekawe Hootsuite pozwala na analizę nastrojów społecznych dla wybranych

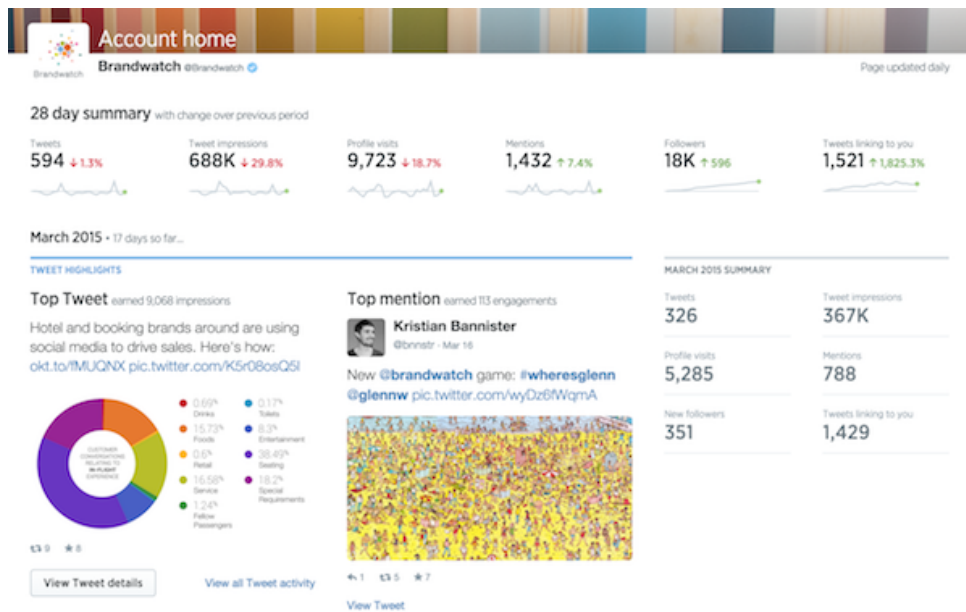
Tablica 3.2: Dokumentacja interfejsu Status

Typ zwracany	Nazwa metody	Opis
long[]	getContributors()	
java.util.Date	getCreatedAt()	zwraca datę utworzenia wiadomości
long	getCurrentUserRetweetId()	zwraca id użytkownika, którego wiadomość została podana dalej
int	getDisplayTextRangeEnd()	
int	getDisplayTextRangeStart()	
int	getFavoriteCount()	zwraca informację ile razy została polubiona wiadomość
GeoLocation	getGeoLocation()	zwraca lokalizację użytkownika zamieszczającego tą wiadomość
long	getId()	zwraca id wiadomości
java.lang.String	getInReplyToScreenName()	zwraca nazwę użytkownika, do którego kierowana jest odpowiedź
long	getInReplyToStatusId()	zwraca id wiadomości, do którego kierowana jest odpowiedź
java.lang.String	getLang()	zwraca język zamieszczonej wiadomości
Place	getPlace()	zwraca obiekt Place przypisany do tej wiadomości
Status	getQuotedStatus()	zwraca obiekt Status cytowanej wiadomości
long	getQuotedStatusId()	zwraca id cytowanej wiadomości
URLEntity	getQuotedStatusPermalink()	zwraca obiekt URLEntity reprezentujący bezpośredni odnośnik do cytowanej wiadomości
int	getRetweetCount()	zwraca ile razy wiadomość została podana dalej
Status	getRetweetedStatus()	zwraca oryginalny status, który jest podany dalej w tej wiadomości
Scopes	getScopes()	zwraca obiekt typu Scopes posiadający informację o id miejsc, do których odnosi się ta wiadomość
java.lang.String	getSource()	zwraca źródło wiadomości
java.lang.String	getText()	zwraca tekst wiadomości
User	getUser()	zwraca obiekt typu User powiązany z tą wiadomością
java.lang.String[]	getWithheldInCountries()	zwraca tablicę nazw krajów, w których wiadomość została wstrzymana

boolean	isFavorited()	zwraca informację czy wiadomość została polubiona
boolean	isPossiblySensitive()	zwraca informację czy wiadomość zawiera link do chronionych informacji
boolean	isRetweet()	zwraca informację czy tweet jest podaną dalej wiadomością
boolean	isRetweeted()	informuje czy wiadomość jest podana dalej
boolean	isRetweetedByMe()	informuje czy wiadomość jest podana dalej przez tego użytkownika
boolean	isTruncated()	informuje czy wiadomość jest skrócona (zakończona znakiem "...")

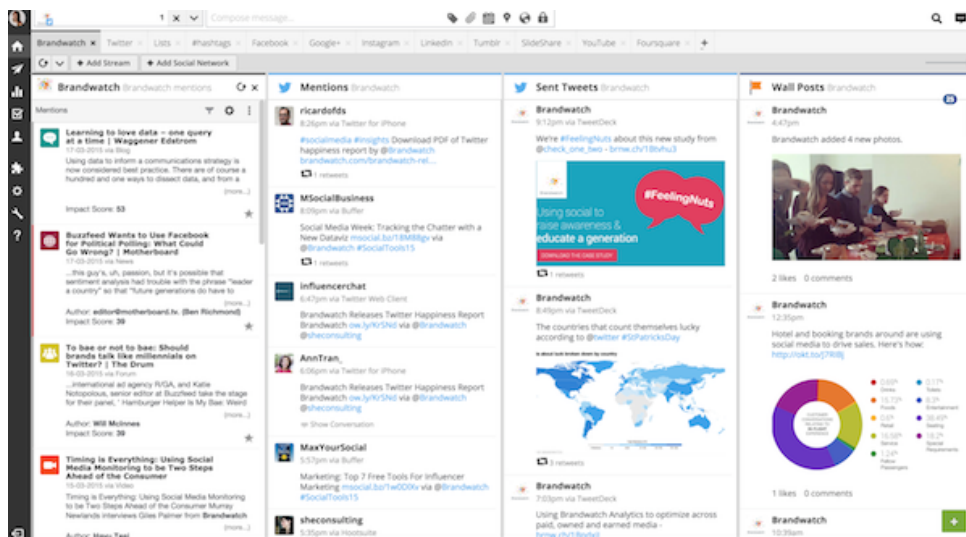
Tablica 3.3: Dokumentacja interfejsu User z pakietu twitter4j. Zamieszczone zostały tylko najważniejsze z metod.

Typ zwracany	Nazwa metody	Opis
java.util.Date	getCreatedAt()	zwraca datę utworzenia profilu użytkownika
java.lang.String	getDescription()	zwraca opis konta użytkownika
java.lang.String	getEmail()	zwraca adres e-mail powiązany z tym kontem
int	getFavouritesCount()	zwraca liczbę wiadomości, którą polubił ten użytkownik
int	getFollowersCount()	podaje ilość użytkowników śledzących profil
int	getFriendsCount()	podaje ilość śledzonych profili
long	getId()	zwraca id użytkownika
java.lang.String	getLang()	zwraca język preferowany przez użytkownika
java.lang.String	getLocation()	zwraca lokalizację użytkownika
java.lang.String	getName()	podaje nazwę użytkownika
java.lang.String	getScreenName()	zwraca nazwę konta
Status	getStatus()	zwraca obiekt typu Status reprezentujący wiadomość wysłaną przez użytkownika
int	getStatusesCount()	podaje ilość wiadomości wysłanych przez użytkownika
java.lang.String	getTimeZone()	podaje strefę czasową użytkownika
java.lang.String	getURL()	zwraca URL do profilu
boolean	isVerified()	podaje informację czy profil jest zweryfikowany



Rysunek 3.2: Narzędzie Twitter Analytics.

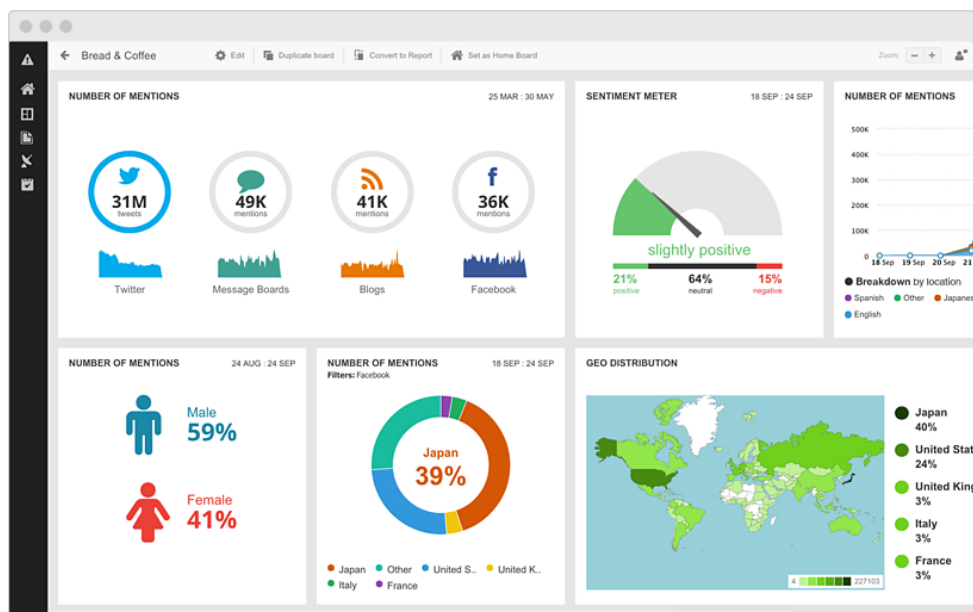
słów kluczowych. Wyświetlane informacje zawierają ogólny zarys użytkowników zamieszczających wiadomości na dany temat np. lokalizację geograficzną, podział ze względu na płeć i język oraz wykres trendu. Aplikacja ta wyświetla wszystkie wiadomości, w których użytkownicy używają wybranego słowa kluczowego. Jak podają twórcy tego narzędzia ma to główne zastosowanie jako pomoc w kampaniach marketingowych w dotarciu do użytkowników krytykujących produkt i posiadających największą ilość osób śledzących.



Rysunek 3.3: Narzędzie Hootsuite - główny pulpit.

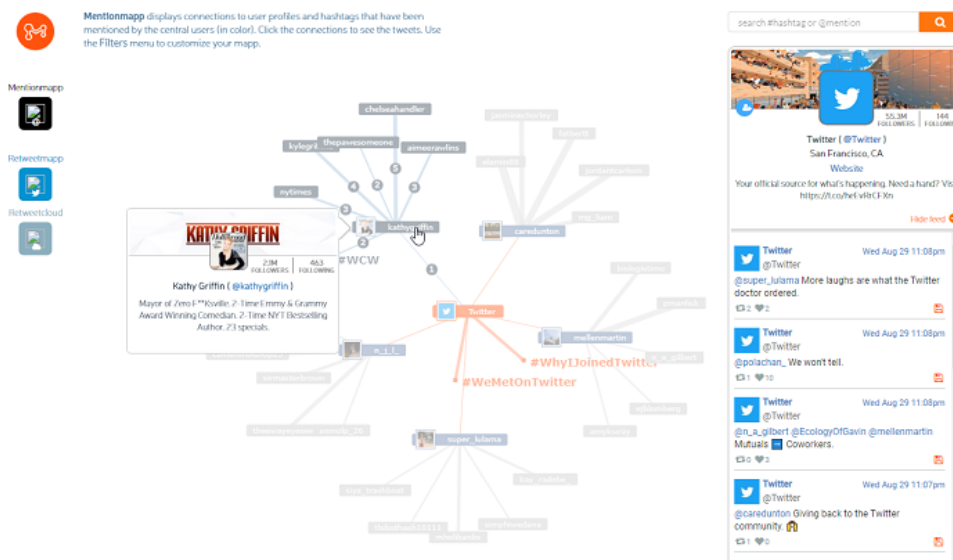
3.5.3 Mentionmap

Trzecim narzędziem wartym omówienia jest *Mentionmap*. Jest to aplikacja, która wyróżnia się spośród innych tym, że rysuje wykres powiązań pomiędzy użytkownikami wchodzącymi w interakcje z danym profilem, a także z ich profilami. Posiada także podstronę umożliwiającą badanie nastrojów społecznych osób zamieszczających wiadomości z konkretnym słowem klu-



Rysunek 3.4: Narzędzie Hootsuite - badanie nastrojów społecznych.

czowym. Nastroje te są przedstawione w postaci chmury nazw kont użytkowników, gdzie kolor nazwy zależy od nastroju prezentowanego przez użytkownika pod kątem słowa kluczowego.



Rysunek 3.5: Narzędzie Mentionmap - wykres powiązań.

3.5.4 Podsumowanie

Podsumowując warto zauważyć, że nie ma obecnie na rynku aplikacji, która umożliwiałaby śledzenie występowania dowolnego słowa w wiadomościach zamieszczanych w czasie rzeczywistym w serwisie Twitter, rysowałaby wykres zależności pomiędzy użytkownikami tego serwisu, analizowałaby nastroje społeczne użytkowników z wyświetleniem informacji o nastroju wyrażanym w poszczególnych wiadomościach oraz pozwalałaby na analizowanie danych historycznych i zamieszczanych w czasie rzeczywistym. Taka sytuacja pozwala na utworzenie nowej aplikacji, która udostępniłaby te funkcjonalności.



Rysunek 3.6: Narzędzie Mentionmap - badanie nastrojów społecznych.

Rozdział 4

Analiza danych Big Data

Termin *Big Data* odnosi się do dużych, zmiennych i różnorodnych zbiorów danych, których przetwarzanie i analiza jest pracochłonne, ale może prowadzić do ciekawych wniosków oraz pozyskania nowej wiedzy. Zbieranie oraz przechowywanie dużej ilości danych do analizy było praktykowane od bardzo dawna, jednak dokładniejsza koncepcja Big Data została poznana w 2001 roku kiedy to analityk Doug Laney zaprezentował znaną dzisiaj definicję 3V: *volume*, *velocity*, *variety* czyli: ilość, szybkość, złożoność, a później dodano jeszcze czwarty atrybut *veracity* czyli: wiarygodność.

4.1 Pojęcie Big Data

Wraz ze wzrostem zainteresowania Big Data podjęto próby dokładniejszego opisanie tego terminu. Obecnie definiując to pojęcie trzeba odnieść się do nowych rozwiązań technologicznych dotyczących wielkich wolumenów danych o innym charakterze ilościowym oraz jakościowym niż dotychczas.

Jedna z pierwszych definicji Big Data została zaprezentowana przez M. Cox i D. Ellsworth w 1997 r. jako duża ilość danych, którą należy zwiększać, aby wydobyć wartości informacyjne. Inna i najbardziej popularna, została przedstawiona w 2001 r. przez pracującego dla firmy analityczno-doradczej wspomnianego analityka D. Laney, opiera się o trzy atrybuty: ilość, szybkość i złożoność. W 2012 r. ta sama firma dodała do swojej definicji kolejne dwa atrybuty: zmienność i złożoność. Autorzy innej publikacji *Big Data: Issues, Challenges, Tools and Good Practices* z 2013 r. definiują pojęcie Big Data jako wymagające stosowania nowych technologii i architektur z powodu potrzeby ekstrakcji wartości płynącej z tych danych.

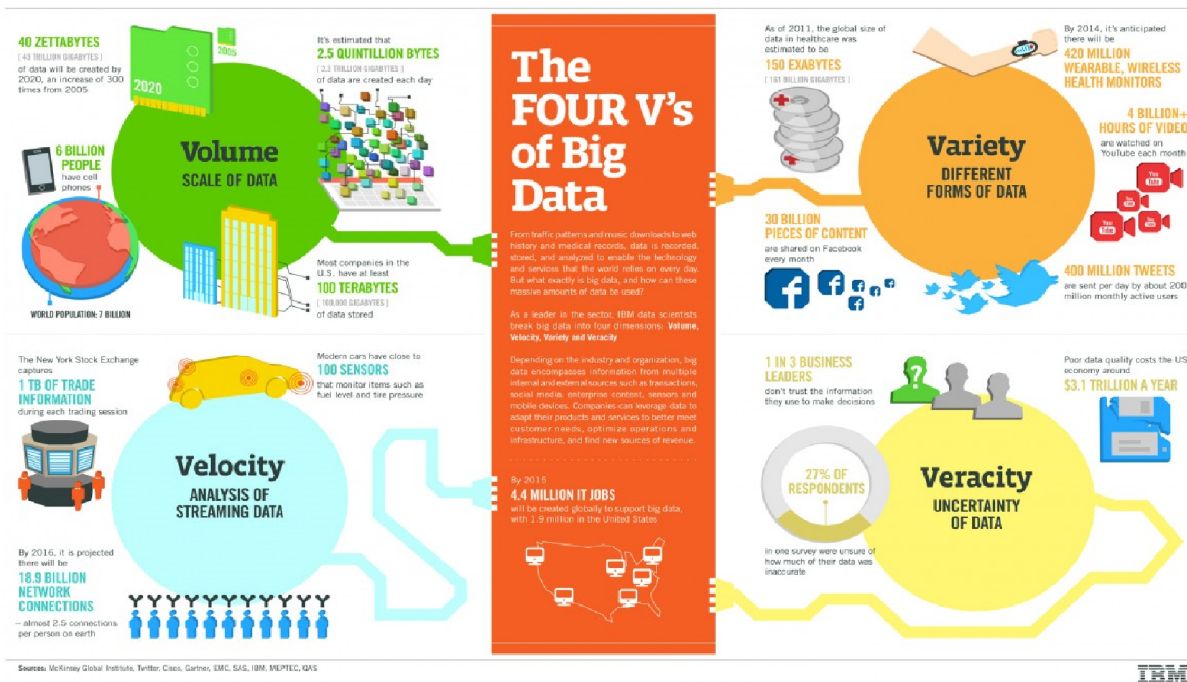
Podsumowując, określenie Big Data to pojęcie odnoszące się do zbiorów danych, które jednocześnie charakteryzują się dużą objętością, różnorodnością, strumieniowym napływem w czasie rzeczywistym, zmiennością, złożonością oraz wymagają stosowania innowacyjnych technologii i narzędzi, aby możliwe było wydobycie z nich wartościowych informacji.

4.2 Charakterystyka

Termin Big Data charakteryzują atrybuty: objętość, szybkość, różnorodność, zmienność, złożoność i wartość. W dalszej części tej pracy dyplomowej przedstawiono omówienie każdego z tych atrybutów.

4.2.1 Objętość

Atrybut ten odnosi się do dużej ilości danych, które wymagają nowych technologii. Rozmiar danych zależy od dziedziny i może wynosić od terabajtów lub petabajtów w zagadnieniach ta-



Rysunek 4.1: Definicja Big Data w ujęciu 4V.

kich jak np. analiza zderzeń cząstek elementarnych w fizyce do megabajtów lub gigabajtów np. w telekomunikacji przy analizie połączeń wykonywanych przez abonentów. Najnowsze badania prognozują, że ilość danych wzrośnie do 2020 r. o 40% zeta bajtów, co będzie skutkować 50-krotnym wzrostem od początku 2010 r..

4.2.2 Szybkość

Duża szybkość napływających danych charakteryzuje się strumieniowym napływem wymagającym analizy w czasie rzeczywistym. Z powodu ograniczeń przepustowości sieci dane takie należy pobierać porcjami i filtrować pod kątem przydatności oraz wartości informacji jaką ze sobą niosą.

4.2.3 Różnorodność i złożoność

Kolejne atrybuty czyli różnorodność i złożoność są ze sobą powiązane i odnoszą się do dużego zróżnicowania źródeł pochodzenia danych i ich formatu. Informacje mogą przychodzić w postaci ustrukturyzowanej jak i m.in. niestrukturalnych dokumentów tekstowych, wiadomości e-mail, materiałów audio i video, transakcji finansowych. Natomiast źródłami ich pochodzenia mogą być wszystkie systemy, które generują i gromadzą dane czyli np.:

- wewnętrzne systemy organizacji takie jak systemy księgowe, kadrowe i transakcyjne,
- źródła zewnętrzne - dane ze stron internetowych (także te nieindeksowane przez wyszukiwarki - pochodzące z głębokiego internetu), blogów, wiadomości tweet i forów internetowych,
- sklepy, usługodawcy i instytucje finansowe,
- instytucje zdrowia.

4.2.4 Zmienność

Dane typu Big Data podlegają okresowym trendom i wahaniom. Napływają ze zmiennym w czasie natężeniem. Mają na to wpływ wydarzenia z wielu dziedzin naszego życia - od pór roku do wydarzeń z dziedziny polityki czy sportu. Przykładem jest np. zainteresowanie inną tematyką w czasie Bożego Narodzenia niż w czasie lata lub bardziej dynamiczne zmiany na rynkach finansowych.

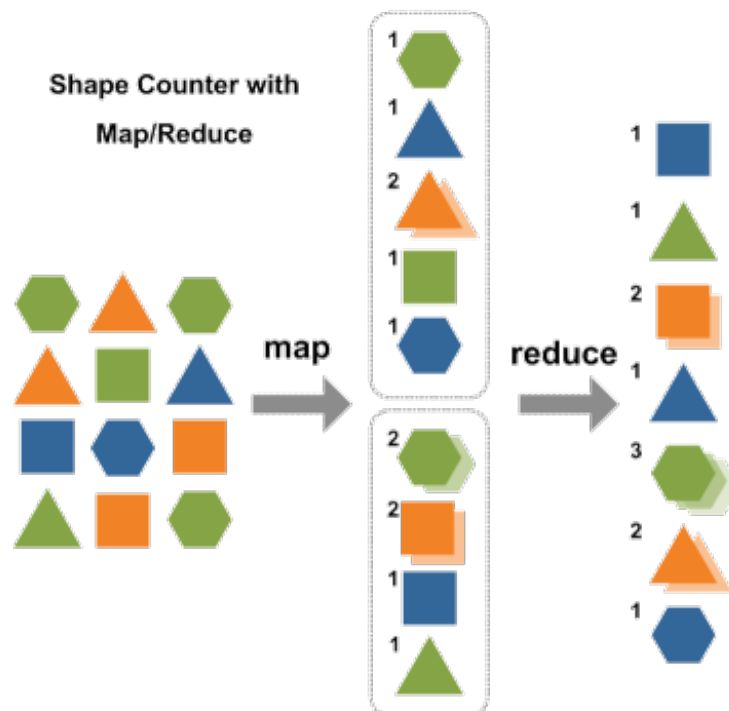
4.2.5 Wartość

Wartość danych napływających z systemów jest trudna do analizy, ponieważ struktura tych danych charakteryzuje się dużą złożonością, a korzyść płynąca z analizy jest ukryta. Obecnie istnieją już systemy, które potrafią radzić sobie z dużym wolumenem informacji, ale to jak te informacje zostaną wykorzystane zależy już od człowieka.

4.3 Paradygmat MapReduce

Większość systemów przetwarzających duże ilości danych opiera swoje działanie o paradygmat *MapReduce*. MapReduce polega na dzieleniu zbioru danych wejściowych na mniejsze niezależne od siebie podzbiory, które są następnie przetwarzane przez równoległe zadania typu *map*. Wynik wyjścia z zadań typu *map* jest sortowany, a następnie podawany na wejście zadań typu *reduce*. Zastosowanie tego paradygmatu skutkuje zwiększeniem wydajności dzięki przetwarzaniu równoległemu niezależnych bloków danych co może zostać wykorzystane przez zastosowanie wielu procesorów.

MapReduce przetwarza wyłącznie pary *<klucz, wartość>*. Dlatego większość operacji polega na znalezieniu elementów spełniających pewne kryteria, umieszczenie ich w zbiorze oraz zwiększenie licznika sygnalizującego częstotliwość występowania elementu tak jak pokazano na ilustracji 4.2..

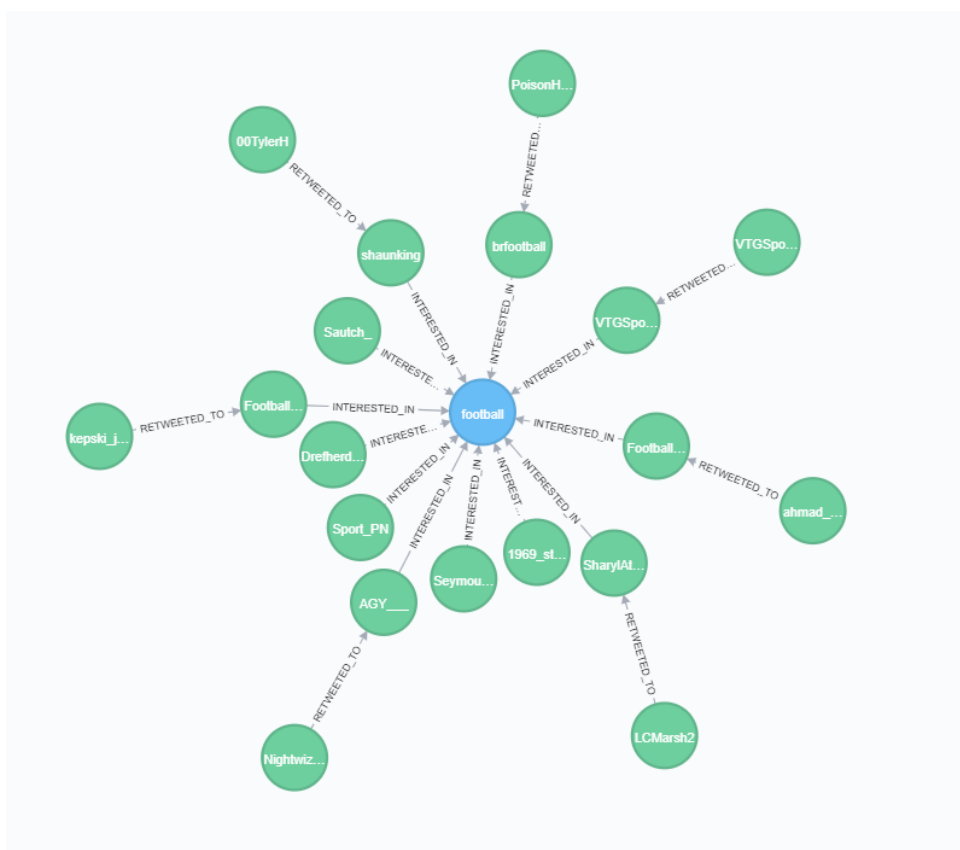


Rysunek 4.2: Przykład zastosowania paradygmatu *MapReduce*.

4.4 Wymagania stawiane przed systemami Big Data

Przetwarzanie danych napływających w sposób strumieniowy wiąże się z dużym wyzwaniem związanym z koniecznością niezwłocznej obsługi danych pojawiających się w dużej ilości i z dużą prędkością. Są to wymagania stawiane przede wszystkim przed platformami zarządzania danymi - ich przetwarzaniem oraz przechowywaniem. Jednymi z najbardziej znanych obecnie rozwiązań tego typu są *Apache Hadoop* oraz *Apache Spark*, które zostaną omówione w dalszej części pracy przy okazji sformułowania wymagań stawianych przed aplikacją zbudowaną na potrzeby tej pracy dyplomowej i dostępnych rozwiązań.

Innym napotkanym problemem jest duża złożoność danych, szczególnie danych niestrukturalnych, która stawia wyzwania związane z efektywnym przechowywaniem informacji oraz przeszukiwaniem bazy danych. Szytwno zdefiniowane relacyjne bazy danych nie odpowiadają wymaganiom stawianym przed systemami Big Data. Bazami, które znalazły swoje zastosowanie w takich systemach są bazy *NoSQL*. Dzięki ich zastosowaniu można uzyskać korzyści związane z poprawą zrozumienia danych, możliwość gromadzenia danych niestrukturalnych, skalowalność oraz elastyczność bazy danych. Przykładem baz danych typu NoSQL są *Cassandra* oraz *Neo4j*, które tak jak i *Apache Hadoop* oraz *Apache Spark* zostaną omówione w dalszej części pracy.



Rysunek 4.3: Przykład zastosowania grafowej bazy danych Neo4j.

4.5 Przykłady zastosowania systemów Big Data

Systemy Big Data mogą mieć bardzo szerokie zastosowanie. Poniżej podano listę istniejących już zastosowań, są to np.:

- wykorzystanie danych pogodowych generowanych przez Interdyscyplinarne Centrum Mo-

delowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego ICM do m.in. prognozowania ilości energii produkowanej przez elektrownie wiatrowe i słoneczne, prognozowania zagęszczenia ruchu, pomoc sieciom handlowym w prognozowaniu jakie produkty będą się lepiej sprzedawały i powinny być w promocyjnej cenie,

- wykorzystanie zegarków, należących do układów typu wearables, wraz z smartfonami oraz systemów Big Data do analizy zdrowia pacjenta np. przy leczeniu choroby Parkinsona, ale też w innych chorobach gdzie ważne jest codzienne monitorowanie,



Rysunek 4.4: Przykład zastosowania układów wearables i systemu Big Data do analizy stanu zdrowia pacjentów.

- dzięki zbieraniu i analizie danych takich jak np. temperatura, czas otwarcia drzwi, ilość kursów z wind oraz specyfikacji sprzętu przewidywany jest czas kiedy należy wykonać naprawę, a służby serwisowe wysyłane są tylko wtedy i tylko tam gdzie są rzeczywiście potrzebne,
- gromadzenie danych ze stron internetowych wraz z wyrażanymi przez internautów opiniami i analiza ich za pomocą systemów wykorzystujących sztuczną inteligencję służy firmom, ale także partiom politycznym, do monitorowania opinii na temat produktów oraz nastrojów społecznych.

Rozdział 5

Analiza sentymentu wypowiedzi

Zagadnieniem dobrze ilustrującym postęp technologiczny jest prędkość rozprzestrzeniania się informacji. W starożytności informacja była przekazywana zazwyczaj przez posłańców, od których losów zależało czy i kiedy informacja zostanie dostarczona do nadawcy. Tak było w przypadku posłańca Filippidesa, który według legendy podążył po wygranej bitwie pod Maratonem do Aten, aby uprzedzić Greków przed zbliżającym się atakiem Persów. W dzisiejszych czasach coraz częściej informacja jest dostępna na ekranach urządzeń mobilnych takich jak telefon komórkowy w czasie rzeczywistym po wystąpieniu jakiegoś zdarzenia. Dlatego bardzo ważne stało się monitorowanie opinii i nastrojów społecznych. Dostępne obecnie narzędzia pozwalają klasyfikować wydźwięk tekstu jako pozytywny, negatywny albo neutralny.

5.1 Sentyment wypowiedzi

Jako istoty ludzkie posiadamy zdolność do rozpoznawania cudzych emocji po treści wypowiedzi i towarzyszących jej czynników pozawerbalnych, której wydźwięk określa się jako stosunek lub postawa pozostająca w korelacji do pewnego zdarzenia lub sytuacji. Treści publikowane w internecie w formie tekstu niosą ze sobą tylko reprezentację tekstową, dlatego odczytanie nastroju wyrażanego w taki sposób jest trudnym zadaniem. Tym bardziej złożonym procesem wymagającym dokładnego przeanalizowania każdego słowa jest zadanie odczytania wydźwięku wypowiedzi przez napisany system.

Dziedziną zajmującą się analizą wypowiedzi jest przetwarzanie języka naturalnego (ang. *NLP - natural language processing*), gdzie językiem naturalnym określa się język stosowany przez ludzi do komunikacji interpersonalnej. Zadaniem systemów rozumiejących język naturalnych jest przekształcenie go na formę bardziej przyjazną dla komputerów. Natomiast podzbiór tych systemów służący do określania sentymentu zwraca ogólną informację o wydźwięku w ustalonej skali. Przykładowo analiza sentymentu w uważanej za punkt odniesienia dla takich algorytmów bibliotece CoreNLP, napisanej przez pracowników i studentów amerykańskiego Uniwersytetu Stanforda, zwraca informację o sentymencie w pięciostopniowej skali: bardzo negatywny, negatywny, neutralny, pozytywny, bardzo pozytywny. Jest to ogólna informacja, ale nawet taka w postaci statystyk jest w stanie posłużyć jako cenne źródło informacji.

5.2 Przetwarzanie języka naturalnego

Analiza nastrojów społecznych wypowiedzi jest możliwa po poddaniu tekstu zamianie na reprezentację, którą mogą posługiwać się systemy informatyczne. Poniżej przedstawiono kilka takich reprezentacji:



Rysunek 5.1: Tweet firmy *Google* zamieszczony z okazji Dnia Dziękczynienia jako przykład tekstu o pozytywnym wydźwięku.

5.3 Wymagania związane z oceną sentymentu

Algorytmy analizujące sentyment muszą poradzić sobie z następującymi wymaganiami:

- złożoność języka naturalnego,
- niejednoznaczność wypowiedzi np. na przykładzie wieloznaczności słowa zamek: *akcja Zemsty Fredry dzieje się na zamku w Odrzykoniu* lub *zamek w moich drzwiach nie chce się otworzyć* lub syntaktyczna niejednoznaczność gdy w jednej części zdania znajduje się pozytywny wydźwięk, a w kolejnej negatywny np. *pogoda była okropna, ale obiad bardzo nam smakował*,
- specyfika stosowanego języka np. w internecie z powodu niejednokrotnie nie zachowanych zasad gramatycznych lub wiadomości nie niosących ze sobą żadnej treści,
- neologizmy np. retweet,
- idiomy np. "urwanie głowy"
- problemów z rozpoznawaniem nazw np. *byliśmy wczoraj na K2* - czy chodzi o film czy o szczyt górski,
- problem wyrwania wypowiedzi z kontekstu,
- tekst o charakterze sarkastycznym.

5.4 Zastosowanie badania wydźwięku wypowiedzi z internetu

Jak wynika z badań przeprowadzonych na grupie ponad 2000 dorosłych Amerykanów 81% internautów przynajmniej raz szukało opinii o produkcie w internecie, a 20% Amerykanów robi to na co dzień. 80% użytkowników internetu spośród wspomnianej grupy badawczej przyznało, że opinie przeczytane w internecie miały znaczny wpływ na dokonane przez nich zakupy. Wyniki

tych badań pokazują jak bardzo wydźwięk informacji, opinii, recenzji oraz poglądów w internecie ma wpływ na zachowania ludzi. Najczęściej spotykane zastosowania analizy sentymentu wypowiedzi zamieszczanych w internecie to:

- analizowanie opinii wyrażanych na temat produktów np. krótko po premierze nowego produktu firmy chcą wiedzieć jak są one odbierane tak aby móc w porę zaplanować wprowadzenie poprawek lub aby prognozować wyniki sprzedaży oraz cenę akcji, firmy próbują także docierać do osób krytykujących aby przekonać ich do zmiany zdania lub żeby krytyków konkurencji zachęcić do zakupu swoich produktów,
- badanie opinii na temat firm np. jako pracodawców lub ich odbiór na rynku,
- analiza opinii na temat ugrupowań politycznych i polityków np. dotyczących obecnej sytuacji lub przyszłych decyzji oraz nastawienia badanych grup społecznych,
- badanie nastrojów społecznych podczas wydarzeń sportowych może posłużyć np. do uzyskania wiedzy jak odbierane są decyzje właścicieli klubów piłkarskich.

