

*Joanna Michalak**

SOCIAL MEDIA JAKO KOPALNIA INFORMACJI – WYBRANE OBSZARY WYKORZYSTANIA DANYCH NA PRZYKŁADZIE PORTALU TWITTER

Z a r y s t r e ś c i: Przekształcenie danych z social media jest w zainteresowaniu wielu naukowców, większość z publikacji skupia się na platformie Twitter. Twitter może być źródłem informacji dla użytkowników portalu (np. o wydarzeniach, produktach). Z drugiej strony platforma to źródło informacji dla przedsiębiorstw i badaczy o opinii i interakcjach użytkowników portalu. W artykule przeprowadzono przegląd literatury w której wykorzystane zostały dane z portalu Twitter w obszarze polityki oraz zarządzania. Autor stara się odpowiedzieć na pytanie jaką wartość wprowadza uwzględnienie danych z social media do prowadzonych analiz.

S ł o w a k l u c z o w e: social media, data mining, Twitter, Web 2.0.

K l a s y f i k a c j a J E L: C80, M31.

WSTĘP

Przekształcenie danych z social media jest w zainteresowaniu wielu naukowców, np. [Sprenger i inni, 2014, s. 178-185], [Stacey i inni, s. 1-37], [Giachanou i inni, 2016, s. 28:1-28:41], [Kaplan i inni, 2011, s. 105 -113] większość z tych publikacji skupia się na platformie Twitter. Celem niniejszego opracowania jest przegląd badań wykorzystujących dane z portalu Twitter w następujących obszarach: polityka oraz zarządzanie. Pytanie badawcze brzmi: jaką wartość wprowadza uwzględnienie danych z social media w analizach prowadzonych w wymienionych obszarach? Należy zaznaczyć, że ze względu na objętość

* Adres do korespondencji: Joanna Michalak, Uniwersytet Mikołaja Kopernika w Toruniu, Wydział Nauk Ekonomicznych i Zarządzania, ul. Gagarina 13a, 87-100 Toruń, e-mail: michalak@doktorant.umk.pl;

artykułu informacje w nim zawarte traktowane są jako sygnały do podjęcia dalszych rozważań.

1. SOCIAL MEDIA JAKO ŹRÓDŁO INFORMACJI

A. Definicja social media

Media społecznościowe to nadal nowy i dynamicznie rozwijający się obszar, dlatego wiele zjawisk go dotyczących nie posiada jeszcze powszechnie akceptowalnych definicji. Media społecznościowe zalicza się do trendu Web 2.0 – formy aktywności w sieci, która opiera się na sieciach społecznościowych oraz charakteryzuje się dwukierunkowością (użytkownicy są nie tylko odbiorcami przekazu ale również jego twórcami). Inaczej ujmując: platformy social media to wirtualna przestrzeń w której na użytkowników oddziałują różne informacje do których mogą się oni odnieść.

Uważa się, że serwisy Web 2.0 zmieniają paradygmat interakcji między właścicielami serwisu a jego użytkownikami, oddając tworzenie większości treści w ręce użytkowników. Oprócz stawiania użytkowników w roli twórców treści do specyficznych cech Web 2.0 należy zaliczyć również: (1) użycie folksonomii¹, (2) tworzenie się wokół serwisów rozbudowanych społeczności, (3) możliwość nawiązywania kontaktów, (4) wykorzystanie efektów sieciowych, (5) wykorzystanie kolektywnej inteligencji, (6) skupienie się na relacjach międzyludzkich. Sposób interakcji i prezentowania informacji zależy od punktu widzenia użytkowników, uwzględniając w tym ich wiedzę oraz doświadczenie. O'Reilly twierdzi, że idea Web 2.0 wprowadza postulat społecznego uczestnictwa w wirtualnej rzeczywistości, co przejawia się w nadawaniu kierunku rozwoju i zbiorowej odpowiedzialności za zawarte w niej treści. Treści generowane przez użytkowników określane są mianem USG (ang. User-generated Content) [Kaplan i inni, 2010, s.105-113] a umiejętne ich przetworzenie stanowi wyzwania dla szerokiego grona osób zainteresowanych. Rolę użytkowników i ich znaczenie podsumowuje tabela 1.

Tabela 1. Porównanie roli nadawcy i odbiorcy [Web 1.0 vs Web 2.0]

Cecha	Web 1.0	Web 2.0
rola nadawcy	dominująca	równa odbiorcy
rola odbiorcy	pasywna	aktywna

¹ Folksonomia, inaczej wspólne tagowanie, społeczna klasyfikacja, społeczne indeksowanie; kategoryzacja treści z wykorzystanie dowolnie dobranych słów kluczowych. Zadaniem jest uporządkowanie informacji w kategoriach. Występuję przeważnie w odniesieniu do niezhierarchizowanych wspólnot, jak np. strony www [http://www.pedagogiczna.edu.pl/warsztat/2007/4/070404.htm].

stosunek nadawca-odbiorca	nierówny: nadawca uprzywilejowany	równi: nadawca i odbiorca tak samo istotni
charakter nadawcy	zinstytucjonalizowany	prywatny

Źródło: M. Szpunar, Nowe-stare medium. Internet między tworzeniem nowych modeli komunikacyjnych, reprodukowaniem schematów komunikowania masowego, Warszawa 2012, s.56.

Obecnie istnieje wiele platform, które różnią się między sobą pod względem zastosowania, wśród najpopularniejszych wymienić należy:

- Facebook²: serwis społecznościowy w ramach, którego użytkownicy mogą tworzyć sieci i grupy, dzielić się wiadomościami, zdjęciami, linkami oraz plikami.
- LinkedIn³: międzynarodowy serwis społecznościowy specjalizujący się w kontaktach biznesowo-zawodowych.
- MySpace⁴: serwis umożliwia komunikację między internautami i nawiązywanie znajomości. Serwis oferuje m. in. tworzenie własnych galerii zdjęć oraz profili muzycznych.
- YouTube⁵: serwis umożliwiający bezpłatne umieszczanie, odtwarzanie strumieniowe, ocenianie i komentowanie filmów.
- Twitter⁶: mikroblog w ramach którego użytkownicy mogą dzielić się, ograniczonymi do 140 znaków, wiadomościami.

B. Social media jako źródło informacji

Social media mogą być rozpatrywane w dwojaki sposób:

- Jako źródło informacji dla użytkowników (głównie o: produktach, wydarzeniach bieżących, przedsiębiorstwie).
- Jako źródło informacji dla przedsiębiorstw oraz naukowców (głównie o: interakcjach i opinii użytkowników).

Niniejsza praca skupia się wyłącznie na drugim aspekcie. Podczas użytkowania platform społecznościowych, ogromna ilość tzw. *digital footprints* jest rejestrowana [https://en.oxforddictionaries.com/definition/digital_footprint], mogą one zostać następnie wykorzystywane do ekstrakcji wzorców. Dla przykładu: przeprowadzona na dużą skalę analiza bazująca na danych z Twittera i Facebook wykazała strukturę sieci społecznych milionów użytkowników [Ugander i inni, 2015, s. 1-17], [Kwak i inni, 2010, s. 591-600]. Dane z Twittera zostały wykorzystane do rozpoznawania wzorców w przepływie popularnych informa-

² facebook.com

³ linkedin.com

⁴ myspace.com

⁵ youtube.com

⁶ twitter.com

cji [Lerman i inni, 2010, s. 90-97], [Cheng i inni, 2014, s. 925-936], identyfikacji wzorców językowych na światową skalę [Mocanu i inni, 2013, s.1-9], aktywności politycznej [Conver i inni, 2011, s.89-96], [Conver i inni, 2012, s.1-19] oraz dynamiki na giełdach [Bordino i inni, 2012, s.1-17].

C. Plaster miodu social media

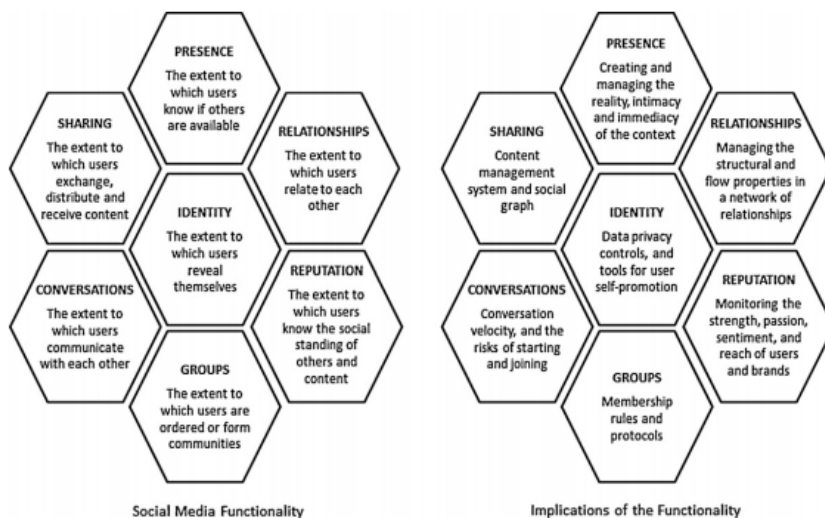
Jak wskazuje [Kietzmann i inni, 2011, s.241-251] firmy bądź osoby publiczne muszą zdecydować czy będą aktywne/bierne w social media. Bez względu na przyjętą strategię wiadomości ich dotyczące mogą pojawiać się na portalach. Przez aktywne podejście, tzn. opracowanie strategii działania w social media, rejestrowanie komunikatów oraz odpowiednią reakcją można wywierać wpływ na grono ich użytkowników. Autorzy przedstawiają tzw. plaster miodu mediów społecznościowych (ang. *honeycomb of social media*). Boki plastra miodu pozwalają na analizę zachowania użytkowników mediów społecznościowych w kontekście funkcjonalności serwisu z przełożeniem na postępowanie dla firmy. Platformy społecznościowe budowane są wokół wymienionych funkcjonalności z zachowaniem równowagi pomiędzy nimi, tzn. żaden z potarli społecznościowych nie jest zbudowany tylko i wyłącznie w oparciu o jeden obszar. Obszary te nawiązują do siebie i się przenikają, najczęściej wykorzystywane są 3-4 funkcjonalności [Kowalska, 2013, s.205-215]. W jego skład wchodzi 7 obszarów (rysunek 2):

- Tożsamość (ang. *Identity*): Dotyczy stopnia w jakim użytkownicy ujawniają swoją tożsamość w social media (tj. imię, nazwisko, płeć). Użytkownicy mogą publikować prawdziwe bądź fikcyjne informacje na swój temat. Jest to element (świadomej bądź nie) budowy wizerunku w sieci. Nieposzanowanie polityki prywatności użytkowników może generować szereg negatywnych skutków dla przedsiębiorstwa.
- Komunikacja (ang. *Conversations*): Obszar ten wskazuje na sposób komunikacji między użytkownikami portalu. Większość serwisów zbudowanych wokół tego obszaru posiada liczne narzędzia związane z konwersacją (zarówno dla dwóch użytkowników jak i grup użytkowników). Przedsiębiorstwa zaczynają dostrzegać korzyści z procesu pozyskiwania treści tych komunikatów i rozwoju narzędzi służących do ekstrakcji wzorców (np. analiza sentymentu). Poznanie opinii klientów dotyczących danego produktu i marki może przynieść wiele korzyści dla przedsiębiorstwa.
- Współdzielenie (ang. *Sharing*): obszar dotyczy dzielenia się różnymi treściami wśród użytkowników portali, stanowi to ważny sposób interakcji. Przedsiębiorstwa mogą wykorzystać ten obszar do popularyzacji własnych treści. Social media postrzegane są jako nowe kanały dystrybucji informacji o firmie, produkcie oraz działaniach firmy.

- Obecność (ang. *Presence*): Możliwość geolokalizacja wiadomości powoduje, że użytkownicy social media mogą pozyskać informację o lokalizacji swoich znajomych, natomiast przedsiębiorstwa mogą zyskać informację o mobilności użytkowników. Z drugiej strony, jak wskazuje [Kowalska, 2013, s. 205-215] obszar ten jest ściśle powiązany z następującymi obszarami: relacje oraz komunikacja. Portale umożliwiają nawiązywanie relacji i rozmów z klientami w czasie rzeczywistym. Skutkować to może bliższą relacją między tymi podmiotami a przedsiębiorstwa mogą być bardziej wiarygodne w działaniach związanych z troską o klienta.
- Relacje (ang. *Relationships*): Obszar reprezentuje zakres w jakim użytkownicy social media tworzą grupy, listy znajomych oraz nawiązują znajomości na portalach społecznościowych. Tworzenie relacji w social media znacząco różni się od tych tworzonych w rzeczywistości, obydwie strony zaangażowane (tj. przedsiębiorstwa i użytkownicy) muszą zaakceptować tę formę i ją zrozumieć. Przedsiębiorstwa mogą angażować się w badanie struktury i siły relacji. Rekomendowane jest aby identyfikować podmioty mające silny wpływ na pewną grupę użytkowników (tzw. liderów opinii).
- Reputacja (ang. *Reputation*): Obszar ten odnosi się zarówno do użytkowników jak i publikowanych przez nich treści. Reputacja jest tutaj definiowana głównie przez pryzmat zaufania do publikowanych treści. Reputacja ma istotne znaczenie dla przedsiębiorstw, które powinny angażować się w celu jej ochrony i monitorowania, w sensie śledzenia informacji pojawiających się na temat przedsiębiorstwa. Kluczowy i nadal rozwijający się obszar to budowa wskaźników, które w sposób obiektywny pozwoliłyby na monitorowanie zmian w wizerunku.
- Grupy (ang. *Groups*): Obszar dotyczący sposobu w jakim użytkownicy mogą tworzyć społeczności. Grupowanie kontaktów, tworzenie podgrup, grupy: otwarte, zamknięte bądź tajne. Przedsiębiorstwa powinny wykorzystywać tę funkcjonalność głównie w zakresie grupowania własnych kontaktów na pewne kategorie.

W wielu publikacjach powtarzane jest, że przedsiębiorstwa powinny poznać społeczny krajobraz mediów społecznościowych. Obecnie powinno być to rozszerzone na każde zainteresowane podmioty, ponieważ szeroko dostrzega się możliwości analiz nie tylko przez przedsiębiorstwa ale również polityków, inwestorów a nawet osoby związane z branżą filmową. Zrozumienie sposobu zachowań użytkowników wokół analizowanego obiektu może przynieść wiele korzyści. Wśród działań zmierzających do tego typu analiz powinno wymienić się następujące: rozpoznanie specyfiki social media, opracowanie strategii aktywności w social media, nadzór na treścią oraz sposób wykorzystania informacji, które generowane są w social media [Kietzmann i inni, 2011, s. 241-251].

Rysunek 1. Plaster miodu social media



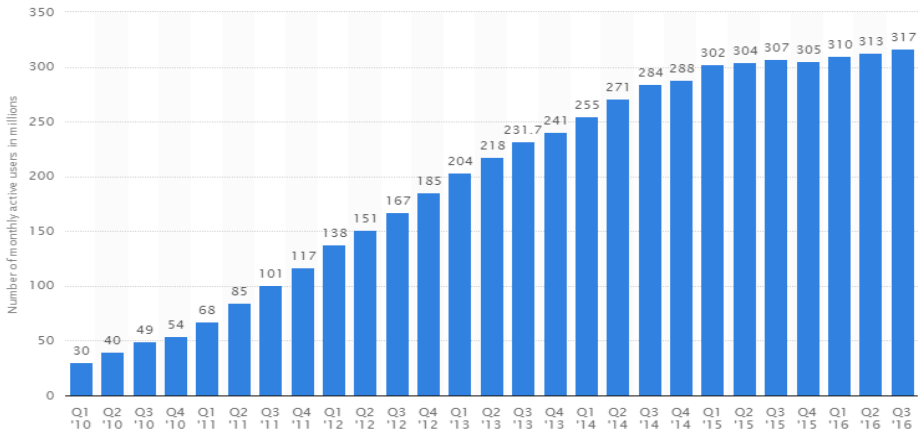
Źródło: Kietzmann i inni, 2011.

D. TWITTER jako źródło informacji

Twitter, platforma założona w 2006, odnotował gwałtowny wzrost popularności w ostatnich latach, (por. rysunek 2), od tego czasu powstało wiele publikacji z obszaru m. in. socjologii, zarządzania, data mining, ekonomii traktujących platformę jako źródło informacji o interakcjach i opiniach aktywnych użytkowników. Szerokie zainteresowanie jest wynikiem m. in.:

- łatwego dostępu do danych - API Twitter można podzielić na dwie grupy: API Rest oraz API Streaming [Bonzanini, www.packtpub.com, 27.08.2016], przy nałożonych ograniczeniach (szerzej: dev.twitter.com) umożliwiając łatwy dostęp do przeszłych i publikowanych w czasie rzeczywistym wiadomości (np. przy wykorzystaniu programu Python, szerzej: [Michalak, 2016, s. 97-110],
- liczby użytkowników - szacuje się, że ponad 284 milionów użytkowników wysłała dziennie około 500 milionów wiadomości [about.twitter.com/company, 10.11.2016]. Ponieważ platforma jest zasilana w sposób ciągły przez użytkowników z całego świata, w przeciwieństwie do tradycyjnych blogów, Twitter jest dynamicznym forum, które pozwala obserwować ciągle zmieniające się wzorce w aktywności użytkowników na niespotykaną dotąd skalę [Giachanou i inni, 2016, s. 28:1-28:41].

Rysunek 2. Liczba użytkowników Twittera [Q1 2010 - Q3 2016, globalnie]



Źródło: www.statista.com/statistics/282087/number-of-monthly-active-twitter-users[28.01.2017].

Dane generowane na portalu można podzielić na dwie grupy: (1) dane o użytkowniku (statyczne) oraz (2) wiadomości wysyłane przez użytkowników (aspekt dynamiczny). Na platformie publikowane są krótkie wiadomości (ograniczenie do 140 znaków). Kopanie w tych danych skupia się przeważnie wokół następujących zadań: (1) analiza trendów (hashtags), (2) eksploracja sieci społecznościowych, (3) analiza sentymentu oraz (4) analiza ścieżek przepływu informacji. Jak wykazano w tabelach 2-3 zadania 1-4 są wykorzystywane z różnymi modyfikacjami i rozszerzeniami do odpowiedzi na zadane przez różnych autorów pytania badawcze.

Przy założeniu, że aktywność użytkowników platformy koncentruje się wokół określonych wydarzeń, odbywających się w czasie t (dla przykładu: wybory, koncerty, wydarzenia z życia publicznego, wydarzenia prywatne z życia użytkowników) należy przyjąć, że wiadomości te będą zawierały opinie oraz nastroje w konkretnym przedziale czasowym⁷ $t_{[i, i+k]}$. Wykonanie zadań 1-4 powinno prowadzić do odzwierciedlenia zachowań i nastrojów określonej grupy (zainteresowanych tematem) użytkowników w czasie $t_{[i, i+k]}$ w odniesieniu do analizowanego wydarzenia/obiektu.

Po wysłaniu zapytania Twitter API zwraca wyniki w formacie JSON (JavaScript Object Notation) [Michalak, 2016, s. 97-110]. Z analitycznego punktu widzenia *tweet* może zawierać wiele informacji [dev.twitter.com/overview/apidobj], na przykład:

- contributors (id, id_str, screen_name) – identyfikacja autora tweeta,

⁷ Oznaczenia przyjęte jak w artykule: Bollen J., Mao H., & Pepe A., *Modeling Public ...*

- coordinates – geolokalizacja,
- created_at – UTC – czas utworzenia tweeta,
- entities (hashtags, urls, user_mention),
- retweet_count – liczba dalszych podań tweeta.

Rysunek 3. Uproszczony przykład struktury wiadomości z portalu Twitter

```
{
  "created_at": "Thu May 12 19:49:29 +00002016",
  "id": "0000000000000000",
  "id_str": "0000000000000000",
  "text": "XYZ",
  "source": "XYZ",
  "truncated": false,
  "in_reply_to_status_id": null,
  "lang": "en",
  "coordinates":
    [[
      [-77.119759, 38.791645],
      [-76.909393, 38.791645],
      [-76.909393, 38.995548],
      [-77.119759, 38.995548].]]
  "retweet_count": 1585
}
```

Źródło: dev.twitter.com/overview/api/tweets [28.01.2017].

Przetworzenie danych z Twittera wymaga podejścia interdyscyplinarnego przy wykorzystaniu wiedzy m. in. z zakresu: uczenia maszynowego, teorii baz danych, przetwarzania języka naturalnego, socjologii, psychologii, ekonomii, programowania, zarządzania.

Wśród zalet analiz tego typu danych należy wyróżnić:

- dostęp do dużych próbek danych,
- szybkość przepływu danych, tzn. odpowiedź na niektóre pytania można uzyskać w ‘czasie rzeczywisty’,
- niskie koszty zbierania danych.

Wśród ograniczeń należy wymienić:

- ograniczenia technologiczne tj. brak odpowiedniego sprzętu,
- brak odpowiednich kompetencji pracowników,
- brak jasno sprecyzowanych rekomendacji, schematów postępowania, definicji i wskaźników, które mogłyby pomóc w procesie analizy danych z portali społecznościowych.

2. POLITYKA

Politycy poprzez indywidualne konta, bądź oficjalnie przypisane danej partii dzielą się newsami oraz aktualnościami, często opatrując je osobistymi przemyśleniami. [Parmelee, Bichard, 2012] w swojej książce szeroko podejmują temat wykorzystania Twittera w polityce i jego wpływu na relacje pomiędzy politykiem a wyborcą. Do kluczowych pytań jakie autorzy stawiają zaliczyć należy:

- Czy i w jakim stopniu wiadomości kierowane przez oficjalne konta polityków/partii wpływają na poglądy zwolenników i ich zachowania?
- Czy *followers* posiadają pewne cechy dzięki którym niektórzy zwolennicy są bardziej podatni niż inni na wpływ informacji? Czy wśród tych cech należy wymienić: zainteresowanie polityką, cechy demograficzne, ideologie, stopień zaufania do rządu?
- Czy jest możliwe aby, na podstawie komentarzy i interakcji określonej grupy użytkowników zbudować szczegółowy profil osób, którzy **śledzą** polityków w social media?
- Czy jest możliwe aby porównać wpływ jaki wywierają wiadomości kierowane na portalu z bardziej tradycyjnymi środkami oddziaływania na wyborców?
- Czy interakcja z osobami, które śledzą dany profil daje istotnie lepsze efekty niż tylko nadawanie informacji w ich stronę?
- Jaki wizerunek prezentuje polityk w mediach społecznościowych i jaka jest opinia osób obserwujących dane konto?
- Czy śledzimy konta tylko tych polityków, które są zgodne z naszymi przekonaniem politycznymi?
- Czy dzięki umiejętnemu sterowaniu publikowanymi wiadomościami i popularnością osób które obserwują dany profil można uzyskać tzw. efektu kuli śnieżnej (dzięki opcji: ‘podaj dalej’). Jeśli tak, czy należy przyjąć, że liczba osób, które obserwują dany profil może nie mieć tak dużego znaczenia?
- Czy dynamika popularności partii na portalu napędzana jest tylko pozytywnymi emocjami czy też wynikać może z rozczarowania bądź sarkazmu?

Tabela 2. Wybrane publikacje dotyczące wykorzystania danych z Twittera w obszarze polityka

tytuł	problem badawczy	główne wnioski
Twitter-based analysis of the dynamics of collective attention to political parties	Czy istotne informacje dotyczące wyborów mogą być uzyskane z social media? Czy liczba publikowanych wiadomości na temat partii politycznych może być traktowana jako wskaźnik uwagi poświęconej partiom przez użytkowników?	Autorzy oceniają dynamikę wolumenu wiadomości i ich relację (oraz możliwości prognostyczne) z ostatecznymi wynikami wyborów. Ostatecznie stwierdzają, że wolumen wiadomości może być dobrym wskaźnikiem sukcesu partii jeśli rozpatrywane są odpowiednie przedziały czasowe.
'News' vs 'old' leaderships the campaign of Spanish general elections 2015 on Twitter	Celem badania jest analiza aktywności użytkowników Twittera podczas kampanii w 2015 w wyborach powszechnych w Hiszpanii wokół przywódców głównych partii politycznych. Jak wiele i jakiego typu wiadomości jest publikowanych w tym czasie? Wokół jakich tematów się koncentrują?	Autorzy postrzegają Twitter jako część nowego modelu kampanii, którego cechą jest personalizacja rywalizacji wyborczej. Zwracają uwagę na wzrost świadomości bycia 'online' wśród polityków. Analizie porównawczej wśród kandydatów podlegają następujące elementy dotyczące ich aktywności na platformie: liczba wiadomości wysyłanych przez polityków, zawartość tych wiadomości, sposób w jaki politycy wysyłają wiadomości. Autorzy wskazują, że wiele czynności związanych z wyborami jest nadal uzależnionych silnie od 'starych mediów', zwłaszcza telewizji – np. wpływ na kształtowanie opinii odbiorców. Twitter został użyty jako narzędzie do jednoczenia zwolenników i działaczy partyjnych oraz jako narzędzie promocji występów kandydatów w innych mediach. Większość kandydatów nie używała Twittera do interakcji z wyborcami.
The use of Twitter for political purposes in Slovenia	Autorzy analizują ogólny charakter komunikacji przy wykorzystaniu Twittera. Stawiają hipotezę, że Twitter jest używany głównie w celu promocji. Pomimo faktu, że komunikaty wysyłane przez portal są krótkie powinny one być skuteczne.	Portale społecznościowe uczyniły arenę polityczną bardziej dynamiczną, łatwiej otwartą 'dla wszystkich' z informacjami łatwo i natychmiast dostępnymi dla wszystkich zainteresowanych. Autorzy wykazują, że Twitter jest wykorzystywany głównie do informowania a nie w celu promowania idei politycznych.

<p>Campaigning on Twitter: Towards the 'Personal Style' Campaign to Activate the Political Engagement During the 2011 Spanish General Elections</p>	<p>Celem artykułu jest ocean w jakim zakresie kandydaci na prezydenta w 2011 dwóch wiodących partii w Hiszpanii wdrożyli nowy model zakładający personalizację kampanii.</p>	<p>Każdy z polityków wykorzystał inną strategię w swojej kampanii. Jeden z kandydatów podkreślał swój własny indywidualny styl i unikał skojarzeń z poprzednikiem. Drugi kandydat wprowadził świadomą kampanię instytucjonalną w której trzymał się oficjalnych stanowisk partyjnych. Autorzy porównywali kandydatów pod względem interakcji z wyborcami, w przypadku braku komunikacji i zamieszczania informacji tylko o swoich działaniach udział kandydata nie był przyczyną do zaangażowania wyborców w dyskusję na wybrane (ważne z ich punktu widzenia) tematy. Takie publikowanie zasobów nie zachęcało wyborców do otwartej i częstej dyskusji. W takim przypadku Twitter nie jest postrzegany jako narzędzie do personalizacji kampanii ale jako kolejny kanał informowania o sprawach już znanych użytkownikom. Jak wykazali autorzy pomimo odmiennej strategii zastosowanej przez kandydatów ich działania w social media były podobne w sensie braku skupienia się na angażowaniu widowni do dyskusji.</p>
<p>Predicting Elections with Twitter: what 140 characters reveal about political sentiment</p>	<p>Badanie wykorzystuje przeprowadzone w Niemczech wybory w celu odpowiedzi na pytanie czy Twitter jest użyty jako forum do politycznej dyskusji oraz czy te wypowiedzi odzwierciedlają ogólne polityczne nastroje.</p>	<p>Analiza nastrojów politycznych w wiadomościach wykazała, że Twitter w sposób wiarygodny odzwierciedla krajobraz polityki offline. Tym samym należy stwierdzić, że wiadomości z platformy mogą stanowić ważny wskaźnik nastrojów politycznych, który należy uwzględnić w rozważaniach i dalszych badaniach.</p>

Źródło: opracowanie własne na podstawie: [López-García i inni, 2016], [Eom i inni, 2015], [Godnov i inni, 2014], [Zamora Medina i inni, 2014], [Tumasjan i inni].

Jakie informacje mogą być uzyskane w danym obszarze na podstawie analizy wzorców z portali społecznościowych?

- ogólny sentyment osób wypowiadających się na temat sytuacji politycznej,
- czy strategia działania w social media przyjęta podczas kluczowych politycznych wydarzeń skutecznie angażuje odbiorców.

Preferowane techniki⁸ analizy danych: text mining, analiza sentymentu, wizualizacja sieci społecznościowych oraz analiza informacji o aktywności użytkowników (tutaj polityków).

⁸ Część ta dotyczy tylko ekstrakcji wzorców (data mining).

3. ZARZĄDZANIE PRZEDSIĘBIORSTWEM

Social media postrzegane są: jako stosunkowo niedroga dla organizacji forma wdrożenia marketingowych kampanii. Z drugiej strony powinny być docenione jako źródło informacji o opinii użytkowników odnośnie marki bądź produktów nie tylko przedsiębiorstwa ale również konkurencji. Wskazuje się, że z uwagi na dynamiczny rozwój mediów społecznościowych pomijanie informacji o przedsiębiorstwie i jego otoczeniu wraz z wyrażonym kontekstem emocjonalnym może mieć negatywne skutki. Klienci publikują informacje na portalach o tym co robią, co lubią, co czują i skąd pochodzą. Identyfikowanie wzorców w tych danych ma liczne zastosowania, dla przykładu:

- zarządzanie relacjami z klientem,
- analiza popularności produktów, wydarzeń etc.,
- gromadzenie informacji zwrotnych dotyczących problemów w nowo wydanych produktach,
- monitorowanie tendencji w nastrojach dotyczących analizowanego obiektu wraz z identyfikacją w nagłych skokach bądź spadkach zadowolenia,
- identyfikacja tzw. liderów opinii wraz z analizą ich sentymentu i zasięgu postów,

Wszystkie wymienione elementy mają na celu wspomaganie podejmowanych decyzji bądź stanowią przyczynę do pogłębienia badań. Od momentu wzrostu popularności portali społecznościowych powstało wiele publikacji wykorzystujących dane stamtąd w celu optymalizacji decyzji. Większość z tych publikacji koncentruje się na ekstrakcji wzorców dotyczących wizerunku analizowanego obiektu oraz wpływu komunikacji z otoczeniem i komunikatów wysyłanych przez przedsiębiorstwo na osoby zainteresowane. Niemożliwe wydaje się przedstawienie większości analiz jednak tabela 3 zawiera artykuły, które reprezentują następujące analizy (1) marketing, (2) komunikacja i zaangażowanie oraz (3) wykorzystanie danych z mediów społecznościowych jako praktycznych przykładów w procesie nauczania.

Preferowane techniki analizy danych: analiza sentymentu, wizualizacja sieci społecznościowych oraz identyfikacja liderów opinii.

Tabela 3. Wybrane publikacje dotyczące wykorzystania danych z Twittera w obszarze zarządzanie przedsiębiorstwem

tytuł	problem badawczy	główne wnioski
Twitter: Power: Tweets as Electronic Word of Mouth	Jakie są ogólne trendy w eWOM microblogging? Jakie są cechy zarządzania marką na mikroblogu? Jakie są wzorce komunikacji mikroblogowej pomiędzy firmami a klientami?	W badaniu autorzy analizowali sposoby wykorzystania eWOM branding. Należy stwierdzić, że coraz częściej decyzje o zakupie klient podejmuje pod wpływem informacji zawartych w internecie. Konsumenci coraz częściej obdarzają platformy społecznościowe zaufaniem i cenią wyrażone na nich opinie. Ze względu na łatwość monitorowania nastrojów dotyczących marki i powszechność Twittera należy sądzić, że może on stać się kluczowy w obronie przez przedsiębiorstwo strategii. Autorzy wskazują na możliwości jakie analityka danych w czasie rzeczywistym stwarza dla budowania relacji z klientem.
How the larger Corporation engage with stakeholders through Twitter	Autorzy analizują w jaki sposób niektóre korporacje korzystają z Twittera w celu zaangażowanie swoich interesariuszy.	Autorzy analizowali zawartość 5,352 tweetów pod względem częstotliwości pojawiania się, followers i following, zachowania oraz retweety. Badania wykazały, że korporacje nieefektywnie wykorzystują potencjał social media w celu wzajemnych budowania relacji z interesariuszami.
Exploring the use of Twitter opinion mining (TOM) in marketing courses	W artykule omówiono podstawowe techniki analizy danych z social media mining i zwrócono uwagę na możliwość wykorzystania tego pozyskanych danych w nauczaniu niektórych przedmiotów	Autorka wskazuje na podstawowe techniki związane z mierzaniem, modelowaniem i kopaniem w dużych zbiorach danych z mediów społecznościowych. Dane te mogą być wykorzystywane jako case study dla zobrazowania treści przekazywanych studentom podczas zajęć, w następujących przedmiotach: marketing, badania marketingowe, marketing mediów społecznościowych.

Zródło: opracowanie własne na podstawie: [Jansen i inni, 2009], [Ivana Mamic, 2013],[Luna-Nevarez, 2016], [Li, 2013].

PODSUMOWANIE

Jak wskazano w opracowaniu, analiza danych z portali społecznościowych to obiecujący obszar. Dane z social media wykorzystywane są w celu identyfikacji ważnych dla osób decyzyjnych zależności na bardzo dużą skalę. Badania prowadzone są w coraz większej liczbie obszarów, nie tylko polityka i zarządzanie, ale również socjologia, psychologia, rynki finansowe. Ze względu na problemy wynikające ze specyficznego charakteru mediów społecznościowych ważne jest aby zrozumieć potrzeby osób decyzyjnych oraz zrozumieć dane, które można pozyskać. Przez zrozumienie potrzeb należy rozumieć odpowiedź na następu-

jące przykładowe pytania: Jakie informacje chce pozyskać (opinia klientów, zachowania wyborców, ścieżka przepływu informacji)? Jakie są pytania badawcze? Czy będzie wymagało to użycia technik i narzędzi dotyczących analizy sentymentu/ sieci społecznościowych? Czy informacje będą głównym źródłem wiedzy na wybrany temat, czy może ich zadaniem jest uzupełnienie innych badań? Innymi słowy następuje przełożenie potrzeb na problemy eksploracji danych. Przez zrozumienie danych należy tutaj rozumieć uważne podejście do danych, precyzję pytań badawczych i celów. Ze względu na ograniczenia nałożone przez API Twitter należy pamiętać o ograniczonym czasowo dostępie do wiadomości.

Jaką wartość stanowią tego typu analizy? Mogą dostarczać informacji o opinii dużej liczby użytkowników oraz stanowić podstawę do oceny skuteczności wybranych strategii.

LITERATURA

- Bollen J., Mao H., & Pepe A., *Modeling Public Mood and Emotion: Twitter Sentiment and Socio – economic phenomena* ., Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, online: www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2826/3237.
- Bonzanini, M., *Mastering Social Media Mining with Python*, dostęp online: <https://www.packtpub.com/> [27.08.2016].
- Bordino I, Battiston S, Caldarelli G, Cristelli M, Ukkonen A, (2012), Weber I. *Web search queries can predict stock market volumes*. PLoS ONE. 2012;7(7): e40014 doi: 10.1371/journal.pone.0040014.
- Cheng J, Adamic L, Dow PA, Kleinberg JM, Leskovec J., (2014), *Can cascades be predicted?* Proceeding of the 23rd International World Wide Web (WWW). 925–936.
- Conover MD, Gonçalves B, Flammini A, Menczer F. Partisan (2012), *Asymmetries in Online Political Activity*. EPJ Data Science;1: 6 doi: 10.1140/epjds6
- Conover MD, Ratkiewicz J, Francisco M, Gonçalves B, Flammini A, Menczer F., (2011), *Political Polarization on Twitter*. Proceeding of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM), 89–96.
- De Milto, L. (2015). *USING SOCIAL MEDIA FOR MARKETING: HARNESSING THE POWER OF TWITTER, LINKEDIN, AND BEYOND*. AMWA Journal: American Medical Writers Association Journal, 30(4), 167-168.
- Eom, Y., Puliga, M., Smailović, J., Mozetič, I., & Caldarelli, G. (2015). *Twitter-Based Analysis of the Dynamics of Collective Attention to Political Parties*. Plos One, 10(7), e0131184. doi:10.1371/journal.pone.0131184.
- Giachanou, A., Crestani, F., (2016) *Like it or not: A survey of Twitter sentiment analysis methods*, ACM Comput. Surv. 49, 2, Article 28.
- Godnov, U., & Redek, T. (2014). *The use of Twitter for political purposes in Slovenia*. Romanian Journal Of Political Science, 14(1), 4-34.
- Ivana Mamic, L., & Arroyo Almaraz, I. (2013). *How the larger corporations engage with stakeholders through Twitter*. International Journal Of Market Research, 55(6), 851-872. doi:10.2501/IJMR-2013-070
- Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). *Twitter power: Tweets as electronic word of mouth*. Journal Of The American Society For Information Science And Technology, 60(11), 2169-2188.

- Kaplan A., Haenlein M., (2010), *Users od the World, Unite! The Challenges and Opportunities of Social media*. Business Horizons, nr 53, s. 61.
- Kaplan, A., Haenlein, M., (2011). *The early bird catches the news: Nine things you should know about micro-blogging*. Business Horizons 54, 105-113.
- Kietzmann, J.H., Hermkens K., McCarthy Silvestre, B. S., (2011), *The Honeycomb of Social Media I.P. Social media? Get serious! Understanding the functional building blocks of social media*. Business Horizons 54, 241-251.
- Kowalska M., (2013), *Funkcjonowanie mediów społecznościowych jako siedem boków plastra miodu*, AUNC Zarządzanie UMK, Zeszyt 413, DOI: http://dx.doi.org/10.12775/AUNC_ZARZ.2013.017 ISSN
- Kuczamer – Kłopotowska S., (2016), *Rola Mediów społecznościowych w komunikacji pokolenia Y Handel Wewnętrzny*. 3(362:2016-227), dostęp online: http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-4e898ef6-f50f-4381-aada-a286d1d383fd/c/HW_3_2016.216-227.pdf
- Kwak, H, Lee, C, Park, H, Moon, S., (2010), *What is twitter, a social network or a news media?* Proceeding of the 19th International World Wide Web (WWW). 2010; 591–600.
- Lerman K, Ghosh R., (2010), *Information contagion: An empirical study of the spread of news on Digg and Twitter social networks*. Proceedings of the 4th AAAI International Conference on Weblogs and Social Media (ICWSM), 90–97.
- Li, T., Berens, G., & de Maertelaere, M. (2013). *Corporate Twitter Channels: The Impact of Engagement and Informedness on Corporate Reputation*. International Journal Of Electronic Commerce, 18(2), 97-126.
- López-García G., (2016), *'New' vs 'old' leaderships: the campaign of Spanish general elections 2015 on Twitter*. Communication & Society [serial online]. July 2016; 29(3): 149-168. Available from: Academic Search Complete, Ipswich, MA. Accessed March 26, 2017.
- Luna-Nevarez, C. (2016). *EXPLORING THE USE OF TWITTER OPINION MINING (TOM) IN MARKETING COURSES*. Proceedings Of The Marketing Management Association, 110-111.
- McLean L, D Gorman., (2012), *Media i społeczeństwo*. Wprowadzenie Historyczne, Kraków, s. 278;
- Michalak J., (2016), *Detecting sentiment in Twitter data – supervised machine learning approach for Twitter Sentiment Analysis in Python*, Torun Business Review, v 15, n. 4, s. 97-110.
- Mocanu D, Baronchelli A, Perra N, Goncalves B, Zhang Q, Vespignani A., (2013), *The Twitter of Babel: Mapping World Languages through Microblogging platforms*. PLoS ONE. 2013;8(4): e61981 doi: 10.1371/journal.pone.0061981.
- Parmelee J., & Bichard S., (2012), *Politics And The Twitter Revolution : How Tweets Influence The Relationship Between Political Leaders And The Public* [e-book]. Lanham, Md: Lexington Books; Available from: eBook Academic Collection (EBSCOhost), Ipswich, MA. Accessed March 5, 2017.
- Sprenger O. T., Sandner G. P., Tumasjan A., Welpe M.I *Predicting Elections with Twitter: what 140 characters reveal about political sentiment*. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, dostęp online: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1441/1852>.
- Sprenger O. T., Sandner G. P., Tumasjan A., Welpe M.I., (2014), *News or Noise? Using Twitter to Identify and Understand Company- specific News Flow.*, Journal of Business Finance & Accounting, 41(7) & (8), 791–830, September/October 2014, 0306-686X doi: 10.1111/jbfa.12086
- Stacey C. E., Pauwels K., Lackman A., *Beyond Likes and Tweets: How conversation content drives store and site traffic*, Working copy, dostęp online: <pages.stern.nyu.edu/~atakos/centerresearch/likesandtweets.pdf>

- Szpunar M., (2012), *Nowe-stare medium. Internet między tworzeniem nowych modeli komunikacyjnych, reprodukowaniem schematów komunikowania masowego*. Warszawa, s.56.
- Ugander J, Karrer B, Backstrom L, Marlow C. *The anatomy of the Facebook social graph*. Preprint. Available: arXiv: 1111.4503v1. Accessed 06 June 2015.
- Westerman, D., Spence, P. R. & Van Der Heide, B. (2014), *Social Media as Information Source: Recency of Updates and Credibility of Information*. *J Comput-Mediat Comm*, 19: 171–183. doi:10.1111/jcc4.12041.
- ZAMORA MEDINA, R., & ZURUTUZA MUÑOZ, C. (2014). *Campaigning on Twitter: Towards the “Personal Style” Campaign to Activate the Political Engagement During the 2011 Spanish General Elections*. *Comunicación Y Sociedad*, 27(1), 83-106.
- about.twitter.com/company
dev.twitter.com/overview/apidobj
facebook.com
<http://www.pedagogiczna.edu.pl/warsztat/2007/4/070404.htm>
https://en.oxforddictionaries.com/definition/digital_footprint
linkedin.com
myspace.com
twitter.com
www.statista.com/statistics/282087/number-of-monthly-active-twitter-users
www.twitter.com
youtube.com

SOCIAL MEDIA AS INFORMATION SOURCE – SELECTED EXAMPLES OF MINING DATA FROM TWITTER

Abstract: Many researches are interested in the analysis of data from social media and most of them focuses on Twitter. Twitter can be a source of information for managers and researchers about the opinions and interactions of users. The aim of this paper is a general overview of selected works in which data from Twitter were used. The author focuses on two areas: management and policy.

Keywords: social media, data mining, Twitter, Web 2.0.