

dr Krystyna Polańska<sup>1</sup>

mgr inż. Aleksander Wassilew<sup>2</sup>

Instytut Informatyki i Gospodarki Cyfrowej, Kolegium Analiz Ekonomicznych  
Szkoła Główna Handlowa w Warszawie

## **Analizy *big data* w serwisach społecznościowych**

### WPROWADZENIE

Określenie *big data* jest jednym z najbardziej popularnych pojęć funkcjonujących w dyskursie publicznym dotyczącym aktualnych technologii informacyjno-komunikacyjnych [Cuesta, 2013, s. 45–53]. Rozwój komunikacji cyfrowej i telefonii komórkowej sprawił, że dane i informacje we współczesnym świecie stały się bardziej powiązane, a użytkownik uwikłany w niezliczone sieci, nie tylko wykorzystuje dostęp do tych wszystkich informacji<sup>3</sup>, ale też sam uczestniczy w ich rozpowszechnianiu oraz w tworzeniu nowych. Duża grupa użytkowników identyfikowalnych (dzięki przynależności do różnych mikrosieci choćby na poziomie prostych danych demograficznych) powiązana ze zbiorem interesujących ich informacji generuje w Internecie nowy strumień danych, który zasila wcześniej zgromadzone ogromne zasoby nieuporządkowanych danych.

Określenie *big data* jest rozumiane jako luźno zdefiniowane pojęcie, używane do opisu dużych i skomplikowanych zbiorów danych, niewygodnych do pracy z wykorzystaniem standardowego oprogramowania statystycznego [Snijders i in., 2012, s. 1–5]. W jednej z szerzej akceptowanych definicji tego pojęcia zwraca się uwagę na fakt, że „volumen samych danych staje się też częścią rozważanego problemu” [Loukides, 2012]. Hilbert zaznacza, że w tej chwili istotne dla paradygmatu *big data* jest nie tyle coraz większa ilość samych danych, co ich analiza służąca „inteligentnemu podejmowaniu decyzji” [Hilbert, 2011, s. 4]. *Big data*

---

<sup>1</sup> Adres korespondencyjny: Instytut Informatyki i Gospodarki Cyfrowej, Kolegium Analiz Ekonomicznych, Szkoła Główna Handlowa w Warszawie, ul. Madalińskiego 6/8, pok. 239, 02-513 Warszawa, e-mail: kpolan@sgh.waw.pl, tel. 22 564 92 80.

<sup>2</sup> Adres korespondencyjny: Instytut Informatyki i Gospodarki Cyfrowej, Kolegium Analiz Ekonomicznych, Szkoła Główna Handlowa w Warszawie, ul. Madalińskiego 6/8, 02-513 Warszawa, e-mail: alex@sgh.waw.pl, tel. 22 564 92 80.

<sup>3</sup> Informacje są tu rozumiane jako zinterpretowane dane – subiektywne i znaczące dla konkretnego użytkownika, tzn. na poziomie infologicznym.

zmienia nasze postrzeganie rzeczywistości polegające na zrozumieniu co oznacza, że coś wiemy i jak jest generowana taka wiedza [Manovich, 2012, s. 1]. Najbardziej oczywistą cechą *big data* jest wielkość analizowanych danych (*volume*). Równie istotne są trzy inne cechy, a mianowicie szybkość przepływu danych (*velocity*), ich różnorodność (*variety*) [por. Genovese, Prentice, 2013, s. 45–53] i wiarygodność (*veracity*)<sup>4</sup>.

Wielkość analizowanych danych stanowi pierwsze wyzwanie dotyczące skalowalności w zarządzaniu *big data*. Dane wciąż są gromadzone, przechowywane i analizowane pod kątem potencjalnie znaczących sygnałów wzmożonej popularności określonych haseł rozproszonych w przestrzeni informacyjno-komunikacyjnej. Szybkość przepływu danych wynika z tempa napływu nowych informacji i kojarzenia z już istniejącymi, które również w sposób ciągły są aktualizowane, a także inteligentnego sortowania wg wyspecyfikowanych celów i potencjalnych zastosowań. W istocie rzeczy, szybkość dotyczy dwóch różnych typów strumieni danych:

1. strumieni nowych danych (potencjalnie generowanych w różny sposób i z różnych źródeł) stopniowo dołączanych do istniejących już dużych zbiorów danych;
2. strumieni wyników zapytań formułowanych przez rozproszonych użytkowników oraz strumieni aktualizowanych już zgromadzonych danych.

W złożonym globalnym społeczeństwie, gdzie rzeczy znaczące mogą zdarzać się wszędzie i w dowolnej chwili, możliwość natychmiastowego nagrywania, przekazywania i komunikowania swoich obserwacji przyczynia się do zmiany stosunku do tych zdarzeń, miejsc i okoliczności. Szybkość oznacza więc także aktualność, świeżość publikowanych informacji i sprawia wrażenie bezpośredniego uczestnictwa w aktualnych wydarzeniach.

Różnorodność danych odnosi się do stopnia różnicowania struktury w *big data* [Halfon, 2014]. Zasadność brania pod uwagę tego wymiaru związana jest z faktem, że *big data* może integrować wiele źródeł. W literaturze przedmiotu wskazuje się na trzy podstawowe grupy danych wyodrębnione wg kryterium ich powstawania [Drapp, Heine, 2014, s. 9]:

1. dane generowane przez maszyny (np. dane z logów systemowych, lokalizacja GPS, wiele rodzajów czujników logowania na określonych stronach www, statystyki klikalności, skanowanie kodów kreskowych lub RFID; spodziewany jest boom tego typu danych – wraz z upowszechnieniem się tzw. Internetu rzeczy (IoT));
2. dane generowane przez użytkowników (np. z sieci społecznościowych, e-maili, formularzy elektronicznych, publikowanych zdjęć, filmów, tekstów),
3. dane transakcyjne (np. z firmowych systemów CRM, ewidencji transakcji).

---

<sup>4</sup> McNully podaje dodatkowe trzy „v” – zmienność (*variability*), wartość (*value*) oraz wizualizacja (*visualisation*). Por. [McNully, 2014].

Każde źródło danych opisywane jest przez swoją własną semantykę, co powoduje powstawanie różnych struktur danych i rodzi trudność zintegrowania danych w jednym modelu logicznym.

*Big data* to zbiór danych, które w przeciwieństwie do uporządkowanych baz danych, są nieustrukturyzowane, niezorganizowane, a wręcz niewygodne do zastosowania w prostych aparatach analitycznych. Jednakże istnieje ogromna liczba sygnałów w szumie informacyjnym, czekających na odkrycie [McAfee, Brynjolfsson, 2012, s. 61–67]. Z dzisiejszego szumu jutro może wyłonić się wiadomość [Hoffmann, 2013, s. 31]. Szумы, brakujące lub niepoprawnie zapisane dane wpływają na wiarygodność wniosków wynikających z analiz dużych zbiorów danych. Jednocześnie obserwowana jest stała tendencja spadku kosztów przechowywania i przetwarzania danych (np. pamięci stałej i obliczeniowej, przepustowości łącza) [por. Wieczorkowski, 2014, s. 401; Wieczorkowski, Polak, 2014, s. 184]. Dzięki temu opłacalna stała się intensywne eksploracja danych, którą jeszcze niedawno uważano za zbyt kosztochłonną.

## SIECI SPOŁECZNOŚCIOWE JAKO GENERATOR DANYCH

Zrozumienie sieci i zasad ich tworzenia jest jednym z kluczowych obecnie tematów w badaniach nad złożonością procesów społecznych. Przy nadmiernej złożoności nie można rozróżnić struktury od procesu, czyli sposobu działania. Złożone systemy połączone są pętlami pozytywnego i negatywnego sprzężenia zwrotnego. Oddziałują w sposób dysypatywny<sup>5</sup> ze środowiskiem, na globalnym poziomie funkcjonują w warunkach dalekich od stanu równowagi, praktycznie na granicy chaosu, po części dlatego, że każdy element odpowiada jedynie na „lokalne” źródła informacji. Małe sieci są uważane za odporne na przypadkowe uszkodzenia, ale są jednocześnie podatne na skoordynowane, zamierzone ataki [Barabasi, 2003, s. 1–5].

Usieciowienie użytkowników to wg Lee Raine’go i Barry’ego Wellmana [Raine, Wellman, 2012, s. 1–5] szczególna cecha tworzących się w Internecie dużych zbiorów danych. John Urry [Urry, 2009, s. 171] uważa, że sieć „kanalizuje bezładną energię złożoności”. Według niego sieć jako jedyna forma organizacji może pomieścić w sobie „rzeczywistą różnorodność działającą jako całość”.

Serwisy społecznościowe, choć kojarzą się głównie jako narzędzie utrzymania dobrych relacji ze znajomymi, są przy tym przedsięwzięciami biznesowymi. Podstawowy model biznesowy zakłada udostępnienie społeczności platformy użytkowej w zamian za dostęp do spersonalizowanych strumieni informacji współtworzonych i współdzielonych przez społeczność.

---

<sup>5</sup> Rozproszony: energia uporządkowanego procesu zmniejsza się („rozprasza się”), przechodząc w energię procesu nieuporządkowanego.

Serwis społecznościowy oferuje użytkownikowi możliwość budowania powiązań z innymi użytkownikami w ramach sieci w celu utrzymywania bieżącego kontaktu, komunikowania się w trybie każdy z każdym i każdy z grupą, współdzielenia i współtworzenia treści, ale także pozyskiwania nowych informacji z profili firmowych (tzw. *fan pages*). W początkach mediów społecznościowych zakładano profile firmowe, które służyły do publikowania informacji, zdjęć, filmów. Czynniono to głównie po to, by poprawić wizerunek firmy lub marki. Z czasem jednak walor sieciowej komunikacji z klientem okazał się przydatny także do doskonalenia oferty produktowej przy współpracy ze społecznością, do angażowania klientów w proces kształtowania oraz testowania nowych produktów i usług. Łatwość dotarcia do nowych klientów, szczególnie młodych, dla których jest to naturalny i popularny kanał komunikowania, okazała się czynnikiem pobudzającym pomysłowość zarówno firm odkrywających nową przestrzeń dla swojej dotychczasowej działalności, jak i nowych przedsięwzięć biznesowych, które swój model działania od razu zbudowały na fundamencie opartym na relacjach społecznościowych z potencjalnymi klientami. Dla firm poszukujących nowych obszarów wykorzystania mediów społecznościowych było to najczęściej prowadzenie szkoleń firmowych, możliwość pracy grupowej nad projektem, informowanie współpracowników o planach i działaniach firmy, a także pozyskiwanie nowych ofert i weryfikowanie informacji przy rekrutacji pracowników.

Kreatywność przy tworzeniu nowych serwisów wynika zwykle z obserwacji zarówno użytkowników (co warto zaoferować, by pozyskać dużą grupę użytkowników skłonnych do regularnego odwiedzania serwisu i przyciągania do niego znajomych), jak i potrzeb biznesu, który uzna zgromadzoną wokół serwisu społeczność za interesującą go grupę docelową. W Polsce najlepiej jednak funkcjonują pomysły, które wcześniej sprawdziły się już w Stanach Zjednoczonych. Z tej przyczyny wiele nowatorskich pomysłów nie znajduje zainteresowania wśród inwestorów, którzy starają się ograniczyć ryzyko niepowodzenia i decydują się głównie na inwestycje sprawdzone na innych rynkach.

Sukces rynkowy odnoszą tylko te przedsięwzięcia przetestowane wcześniej w innych krajach, które jednocześnie uwzględniają specyfikę polskiego użytkownika i lokalne memy<sup>6</sup> kulturowe. Memy to określenia mające znaczenie rozpoznawalne przez używające je grupy. Przykładami memów mogą być moda, melodie, idee, język (zwroty językowe), zachowania, wzory, trendy w sztuce itp. Memy rozprzestrzeniają się dryfując między umysłami „w procesie szeroko rozumianego naśladownictwa” [Dawkins, 1996, s. 267]. Jako replikator „gwarantujący” przetrwanie, mem powinien posiadać takie cechy jak: długowieczność, płodność i wierność kopiowania [Dawkins, 1996, s. 269]. Śladami transferu idei oraz wskaźnikiem jej płodności w środowisku naukowym może być np. liczba cyto-

---

<sup>6</sup> Pojęcie *mem* oznaczające replikator w sferze kulturowej (w analogii do genu w sferze biologicznej) wprowadził Richard Dawkins. Pomocne jest ono przy wyjaśnieniu transmisji kulturowej podczas ewolucji określonej społeczności (por. [Dawkins, 1996, roz. 11]).

wań, odnośników, luźnych wzmianek, ale także nieświadome zapożyczenia. Mogą nimi być zarówno polubienia, jak i udostępniania na portalach społecznościowych, a także inspirowanie i powielanie różnego rodzaju zachowań, współdzielenie i współtworzenie treści z innymi użytkownikami itp. Samo słowo mem okazuje się też dobrym przykładem memu [Dawkins, 1996, s. 434].

Podstawowym problemem na obecnym etapie rozwoju mediów społecznościowych jest spójny projekt komplementarnych sposobów komunikowania się z klientem poprzez media społecznościowe uwzględniający specyfikę branży, charakter oferowanych produktów i prawidłowo określone cechy konsumenta, któremu dedykowany jest konkretny produkt. Wiadomo, że sieci społecznościowe są szczególnie pomocne przy poznawaniu nowych ludzi i pozostawaniu w stałym kontakcie z już zaprzyjaźnionymi. Potwierdzają to zarówno badania Graeme Maggie Hutton i Maggie Fosdick [Hutton, Fosdick, 2011, s. 566] z lat 2006–2010, jak i Krystyny Polańskiej [Polańska, 2013a, s. 40] z 2012 i 2013 roku. W tych ostatnich badaniach tylko 6,5% respondentów nie było zainteresowanych żadnym fanpage'em firmowym; głównymi motywami decyzji o dołączeniu do fanpage'a firmowego na Facebooku były dla badanych (zarówno wówczas, jak i w badaniach przeprowadzonych rok wcześniej) studentów SGH: „poczucie wspólnoty z lubiącymi to samo”, „dostęp do najświeższych informacji”, „konkursy”, „informacje o planowanych imprezach”, „poczucie wpływu na coś” i „możliwość wypowiedzi” [Polańska, 2013b].

Zazwyczaj tylko niewielka część konsumentów aktywnie uczestniczy w przedsięwzięciach społeczno-medialnych. Z tego powodu w kampaniach reklamowych uzależnionych od udziału użytkowników najistotniejsze jest zaangażowanie użytkowników zorientowanych na relacje z innymi [Yeo, 2012, s. 307], bo to oni angażują grono swoich przyjaciół do zainteresowania się produktem lub firmą oraz do współudziału w akcjach promocyjnych. Te szczególnie aktywne osoby generują zwykle opinie, które prowokują kolejne komentarze. Udział w serwisach społecznościowych pozwala firmom nie tylko kontrolować kto jest odbiorcą przekazu, ale też uzyskiwać zwrotne komunikaty ułatwiające dalsze działania, konstruowanie nowej oferty produktowej i jej personalizowanie.

### POŻYTKI PŁYNĄCE Z *BIG DATA*

Powszechny jest pogląd, że połączenie wysiłków *big data* z teorią nauk społecznych jest przydatne do przewidywania kryzysów społecznych i gospodarczych [Helbing, Baliotti, 2011, s. 3–68; Conte i in., 2011, s. 412–413]. Sygnały o mobilizacji społeczności wokół określonej idei lub zdarzenia mogą ostrzegać przed budzącymi się ruchami społecznymi. Mogą też być przydatne do prognozowania zachowań społecznych wobec konkretnych zjawisk (np. rodzących się ruchów społecznych, rozprzestrzeniania się epidemii). Prosta ekstrapolacja prze-

biegu współczesnych zdarzeń na podstawie ich opisu z przeszłości nie znajduje odzwierciedlenia w rzeczywistości. Jednostki i całe społeczeństwa zmieniają się nieustannie, mają szeroki dostęp do informacji, mogą obserwować zachowania innych społeczeństw lub wybranych grup, dlatego reakcje społeczne łatwiej przewidzieć na podstawie bieżących analiz semantycznych niż na podstawie doświadczeń historycznych. Odpowiednio zastosowane algorytmy wobec dużych zbiorów różnorodnych danych mogą wskazywać na tendencje w myśleniu, rozumieniu określonych zagadnień, polaryzację poglądów politycznych, gospodarczych, społecznych i religijnych. Dostęp do dużych zbiorów danych pochodzących z serwisów społecznościowych umożliwia dodatkowo różnicowanie grup społecznych wg cech demograficznych, światopoglądowych, zawodowych itp. Staje się wówczas możliwa obserwacja wrażliwości i reakcji społeczności (lub tylko wybranych grup) na nowe bodźce płynące z otoczenia. Ułatwione jest też dotarcie z określonym przekazem do zdefiniowanej grupy odbiorców.

#### WNISKOWANIE NA PODSTAWIE WSKAŹNIKÓW Z SERWISÓW SPOŁECZNOŚCIOWYCH

Dane pozyskiwane z sieci społecznościowych są najczęściej ekstraktowane przez wyspecjalizowane agencje usługowe. Tylko duże firmy posiadają odpowiednie środki, by samodzielnie formułować odpowiednie algorytmy i składować dane oraz wyniki analiz. Tym bardziej, że dane są łączone z pozostałymi strumieniami danych. Jednakże już z poziomu zalogowanego użytkownika sieci możliwe jest porównywanie alternatywnych zabiegów prowadzących do określonego celu w wybranych serwisach społecznościowych, np. kampanii wyborczych dwóch kandydatów na jeden urząd, kampanii promocyjnych produktów bankowych głównych konkurentów na rynku itp. Analizę danych w serwisach społecznościowych można podzielić na trzy grupy:

1. dane obserwowalne – widoczne z poziomu zalogowanego użytkownika:
  - a) zadeklarowane zainteresowania;
  - b) polubienia (postów, komentarzy, stron)<sup>7</sup> [por. Hilbert, 2013, s. 30];
  - c) zamieszczane posty – własne lub „zewnętrzne” teksty, linki, zdjęcia, filmy;
  - d) aktywność przejawiana w reakcjach na komentarze, wpisy na tablicy oraz udostępnienia. Do tej grupy danych można dołączyć znajomych – ich liczbę i stratyfikację demograficzną – graf umożliwiający pośrednie wnioskowanie na temat użytkownika – uwzględniać częstotliwość i kierunki podążania (np. kto za kim udostępnia określone treści, kto komu i co odpisuje) oraz kto kogo obserwuje, można utworzyć dokładniejszy socjogram powiązań;

---

<sup>7</sup> Przy polubieniach pojawia się problem określenia co dokładnie zyskało aprobatę – tekst, obrazek, idea, czy wzgląd na nadawcę komunikatu. Innymi słowy – dlaczego ludzie postępują tak, jak postępują i co oznacza to konkretne posunięcie. Por. [Hilbert, 2013, s. 30].

2. analiza semantyczna stron, postów, wiadomości do określania chmur znaczeniowych;
3. grupowanie użytkowników wokół wspólnych tematów;
4. analiza aktywności użytkowników – polegająca na zastosowaniu bardziej zaawansowanych narzędzi programowych (np. analiza aktywnego czasu poświęconego danej stronie internetowej, analiza przeczytanych postów oraz linków, za którymi użytkownik podążał, bo wzbudziły jego zainteresowanie, czasu zalogowania do serwisu – jego częstotliwości i długości, pory dnia, miesiąca itp.).

Powyższe analizy mogą mieć zastosowanie przy śledzeniu rozprzestrzeniania się memów – określania ich „płodności”, „żywności” oraz ewolucji sposobów ich rozumienia w trakcie „wędrówki” w sieci. Istotne byłoby uchwycenie tendencji społecznych w początkowej fazie ich funkcjonowania w przestrzeni publicznej – narodziny nowych idei, pomysłów, ruchów.

Interesujące są też porównania równoległych strumieni informacji dotyczących tych samych zjawisk. Na przykład kampania wyborcza dwóch polityków konkurujących do objęcia jednego urzędu ze znanym wynikiem końcowym (wiemy który z nich wygrał) lub nie i analiza komparatywna ich przekazu informacyjnego do wyborców realizowanego za pomocą dwóch wybranych serwisów społecznościowych. W celu uproszczenia można wziąć pod uwagę tylko dwa najpopularniejsze serwisy społecznościowe, takie jak: Facebook i Twitter (tabela 1 zawiera zestawienie typowych form aktywności w obu serwisach), a następnie spróbować dokonać analizy porównawczej zastosowanych technik. Zestawienie ich z osiągniętym rezultatem walki konkurencyjnej może posłużyć do określenia skuteczności zastosowanych metod i doboru treści.

**Tabela 1. Twitter vs. Facebook – specyfika aktywności w serwisie**

Aktywność w serwisie	Twitter	Facebook
komentarze do bieżących wydarzeń, myśli dręczące autora zamieszczane na tablicy w profilu użytkownika	wpis/tweet	post/comment
akceptacja wyrażonych sądów	retweet	polubienie/like
podzielanie wyrażonych sądów tak dalece, że są rozpowszechniane dalej	replie	udostępnienie/share
obserwator, który jest zainteresowany osobą, marką lub produktem na tyle, by śledzić komentarze zamieszczane na fanpage`u	obserwator/follower	fan
profil osoby, marki lub produktu założony w celach marketingowych	profil firmowy	fanpage

Źródło: opracowanie własne.

Ekspansja Facebooka opiera się na dużych danych, bo jest on jednocześnie producentem i konsumentem dużych danych [Montcheuil, 2014]. W celu wyszukania i dopasowania odpowiedniej osoby do kontaktu, algorytmy analizują cha-

rakter i preferencje każdego uczestnika – na podstawie polubienia i udostępnienia konkretnych treści, otwierania proponowanych linków, wysłanych wiadomości itp. Z profilu użytkownika – o ile zostały upublicznione – można uzyskać dane socjo-demograficzne takie jak: wiek, płeć, znajomość języków obcych, odwiedzone miejsca (geolokacja), zainteresowania, hobby, a następnie połączyć je z innymi strumieniami danych [Zembik, 2014, s. 12].

Podobnie ma się rzecz z Twitterem, który jest zaliczany do mikroblogów (umożliwia publikowanie krótkich komentarzy<sup>8</sup>, zdjęć, filmów i linków docierających do grupy zainteresowanych użytkowników).

Wnioskowanie porównawcze można prowadzić na podstawie:

1. liczby polubień – profilu konkurentów w obu serwisach, zamieszczanych tam komentarzy i materiałów audiowizualnych (analiza ilościowa i semantyczna);
2. współczynnika zaangażowania, czyli wskaźnika wyrażonego w procentach, który jest ilorazem sumy interakcji (komentarze plus oznaczenia „Lubię to!”) i liczby odsłon badanego wpisu, pomnożonym przez 100<sup>9</sup>.
3. chmur znaczeniowych opracowanych w oparciu o frekwencję używanych słów (analiza semantyczna) w publikowanych komunikatach.

Wnioskowanie porównawcze prowadzone jest na podstawie zauważalnie dużych przewag w liczbie konfrontowanych przejawów aktywności społeczności spowodowanych przez działania administratorów kont stworzonych w określonym celu biznesowym lub politycznym. Zestawienie konkretnych działań i reakcji społeczności na nie wskazuje na oczekiwania społeczności i dopasowanie do nich porównywanych konkurentów.

## ZAKOŃCZENIE

Wielkość wszystkich informacji przechowywanych na świecie w 2013 r. szacowano na 1200 eksabajtów, z czego mniej niż 2% stanowiły informacje analogowe (tj. książki drukowane, odbitki fotograficzne i inne dokumenty papierowe) [Mayer-Schönberger, Cukier, 2013, s. 9]. W 2007 r. w światowych repozytoriach było 4 razy mniej dostępnych informacji (300 eksabajtów), w tym 7% stanowiły informacje analogowe [Mayer-Schönberger, Cukier, 2013, s. 8]. Jednocześnie dzięki rozwojowi technologii telekomunikacyjnych w niespotykanym wcześniej tempie wzrastają możliwości zarówno przechowywania, jak i analizowania tak gigantycznych zbiorów danych. Wskazują one dokładniej i szybciej na powstawanie nowych zjawisk niż dotychczasowe metody oparte na analizach danych, pochodzących z badań wycinków społeczeństwa (prób badawczych) i mających w momencie analizy już charakter historyczny. Ponadto należy pamiętać, że poza

<sup>8</sup> Do 140 znaków.

<sup>9</sup> Współczynnik zaangażowania fanów (Engagement Rate) został opracowany przez firmę Socialbakers.com.



obszarem obserwacji badawczych *big data* pozostają wykluczeni cyfrowo, którzy są niewidocznymi w sieci, choć nadal stanowią część społeczeństwa. Użytkownik ma także prawo do bycia zapomnianym w Internecie<sup>10</sup>, choć tego typu ubytek informacji jest nieistotny w porównaniu z kreacją nowych danych w tym samym czasie. W przestrzeni wirtualnej powszechny jest problem zaufania. Jest on szczególnie dostrzegalny w sieciach społecznościowych, gdzie brak zaufania dotyczy przede wszystkim:

- samego serwisu społecznościowego i stosowanej w nim polityki prywatności oraz używanych zabezpieczeń;
- proponowanych linków – zarówno przez społeczność, jak i linków sponsorowanych;
- wiarygodności i rzeczywistej istotności publikowanych informacji.

Omówione analizy dużych zbiorów danych, jak też zaproponowana komparatyka równoległych strumieni informacji dotyczących tych samych zjawisk są zwykle realizowane przez wyspecjalizowane firmy mające dostęp do danych w serwisach społecznościowych.

Filozofia *big data* oznacza nie tylko uzyskiwanie informacji przetworzonych z dużych zbiorów danych, ale przede wszystkim poprawną ich interpretację. W przestrzeni wirtualnej pojawiają się czasem krótkotrwałe fale memów, które mogą zwiastować nowe zjawiska, ale mogą też szybko zanikać. Dlatego tak istotne są obserwacje społeczności, ich reakcji na nowe idee, w tym przede wszystkim wyczuwanie narastania entuzjazmu lub znużenia nimi.

## BIBLIOGRAFIA

- Barabasi A., 2003, *Linked: How everything is connected to everything else and what it means*, Plume, New York [za:] Ch. Snijders, U. Matzat, U-D. Reips, *Big Data: Big Gaps of Knowledge in the Field of Internet Science*, „International Journal of Internet Science” 2012, Vol. 7(1), [http://iscience.deusto.es/wp-content/uploads/2012/08/ijis7\\_1\\_editorial.pdf](http://iscience.deusto.es/wp-content/uploads/2012/08/ijis7_1_editorial.pdf) (dostęp: 12.08.2014 r.).
- Conte R., Gilbert N., Bonelli G., Helbing D., 2011, *Futur ICT and social sciences: Big Data, big thinking*, „Zeitschrift für Soziologie”, nr 40.
- Cuesta C.E., Martinez-Prieto M.A., Fernandez J.D., 2013, *Towards an Architecture for Managing Big Semantic Data in Real-Time [w:] Software Architecture. Springer Berlin Heidelberg*, <http://dataweb.infor.uva.es/wp-content/uploads/2013/04/ecs2013.pdf> (dostęp: 12.08.2014).

<sup>10</sup> Wniosek o usunięciu z wyników wyszukiwania w przeglądarce Google obsługującej ok. 90% ruchu w Europie jest dostępny na podstawie orzeczenia Trybunału Sprawiedliwości Unii Europejskiej stwierdzającym, że „określone osoby mogą zwrócić się do właścicieli wyszukiwarek z prośbą o usunięcie konkretnych wyników wyszukiwania haseł zawierających ich imię i nazwisko, jeśli prawo do prywatności tych osób ma priorytet nad prawem dostępu do tych wyników”. [https://support.google.com/legal/contact/lr\\_eudpa?product=websearch&hl=pl](https://support.google.com/legal/contact/lr_eudpa?product=websearch&hl=pl) (dostęp: 2.09.2014 r.).

- Dawkins R., 1996, *Samolubny gen*, Prószyński i S-ka, Warszawa.
- Drapp T.F., Heine V., *Big Data. The untamed force*, Deutsche Bank Research, Frankfurt am Main 2014 (May 5), [http://www.dbresearch.com/PROD/DBR\\_INTERNET\\_EN-PROD/PROD000000000334340/Big+data+-+the+untamed+force.pdf](http://www.dbresearch.com/PROD/DBR_INTERNET_EN-PROD/PROD000000000334340/Big+data+-+the+untamed+force.pdf) (dostęp: 10.08.2014 r.).
- Genovese Y., Prentice S., 2011, *Pattern-Based Strategy: Getting Value from Big Data*, Gartner Special Report, June [za:] C.E. Cuesta, M.A. Martinez-Prieto, J.D. Fernandez, *Towards an Architecture for Managing Big Semantic Data in Real-Time* [w:] *Software Architecture*. Springer, Berlin Heidelberg 2013, <http://dataweb.infor.uva.es/wp-content/uploads/2013/04/ecsa2013.pdf> (dostęp: 12.08.2014 r.).
- Halfon A., *Handling Big Data Variety*, <http://www.finextra.com/community/fullblog.aspx?blogid=6129> (dostęp: 1.08.2014 r.).
- Helbing D., Baliatti S., 2011, *From social data mining to forecasting socio-economic crises*, "European Physical Journal", Special Topics, Vol. 195, <http://dx.doi.org/10.1140/epjst/e2011-01401-8>.
- Hilbert M., 2013, *Big Data for Development: From Information to Knowledge Societies*, January, <http://www.geogeo.com/wp-content/uploads/2013/03/Big-data-for-development.pdf> (dostęp: 22.08.2014).
- Hoffmann J., Voss A., 2013, *Big data und seine Bedeutung für das Wissensmanagement*, "Wissenmanagement", Fraunhofer IAIS, nr. 5, [http://www.iais.fraunhofer.de/fileadmin/user\\_upload/Abteilungen/KD/uploads\\_BDA/2013\\_05\\_009\\_Wissensmanagement\\_BigData.pdf](http://www.iais.fraunhofer.de/fileadmin/user_upload/Abteilungen/KD/uploads_BDA/2013_05_009_Wissensmanagement_BigData.pdf) (dostęp: 12.08.2014).
- Hutton G., Fosdick M., *Consumer Relationships with Brands Evolve in the Digital Space*, "Journal of Advertising Research", Vol. 51(4), <http://dx.doi.org/10.2501/JAR-51-4-564-570>.
- Loukides M., 2012, *Data Science and Data Tools* [w:] *Big Data Now*, O'Reilly, Sebastopol CA, <http://www.safaribooksonline.com/library/view/big-data-now/9781449316297/ch01.html> (dostęp: 12.08.2014).
- Manovich L., 2012, *How to Follow Software Users*, <http://manovich.net/index.php/projects/how-to-follow-software-users> (dostęp: 24.8.2014 r.).
- Mayer-Schönberger V., Cukier K., 2013, *Big Data. A revolution That Will Transform How We Live, Work and Think*, John Murray Publ., London.
- McAfee A., Brynjolfsson E., 2012, *Big Data: The management revolution*, "Harvard Business Review", Vol. 90(10), [http://www.buyukverienstitusu.com/s/1870/i/Big\\_Data\\_2.pdf](http://www.buyukverienstitusu.com/s/1870/i/Big_Data_2.pdf) (dostęp: 12.08.2014 r.).
- McNully E., *Undarstanding Big Data*, <http://dataconomy.com/seven-vs-big-data> (dostęp: 22.08.2014 r.).
- Montcheuil de Y., *Facebook: A Decade of Big Data*, "WIRED" 06.03.2014, <http://www.wired.com/2014/03/facebook-decade-big-data> (dostęp: 2.08.2014 r.).
- Polañska K., 2013a, *Sieci społecznościowe. Wybrane zagadnienia ekonomiczno-społeczne*. Oficyna Wydawnicza SGH, Warszawa.
- Polañska K., 2013b, *Studenci SGH wobec Facebooka* (wyniki badań – materiał niepublikowany).

- Raine L., Wellman L.B., 2012, *Networked. The new social operating system*, MIT Press, Cambridge [za:] Ch. Snijders, U. Matzat, U-D. Reips, *Big Data: Big Gaps of Knowledge in the Field of Internet Science*, “International Journal of Internet Science” 2012, Vol. 7(1), [http://iscience.deusto.es/wp-content/uploads/2012/08/ijis7\\_1\\_editorial.pdf](http://iscience.deusto.es/wp-content/uploads/2012/08/ijis7_1_editorial.pdf) (dostęp: 12.08.2014 r.).
- Snijders Ch., Matzat U., Reips U., 2012, *Big Data: Big Gaps of Knowledge in the Field of Internet Science*, “International Journal of Internet Science”, Vol. 7(1), [http://iscience.deusto.es/wp-content/uploads/2012/08/ijis7\\_1\\_editorial.pdf](http://iscience.deusto.es/wp-content/uploads/2012/08/ijis7_1_editorial.pdf) (dostęp: 12.08.2014 r.).
- Urry J., 2009, *Socjologia mobilności*, PWN, Warszawa.
- Wieczorkowski J., 2014a, *Big data – aspekt technologiczny i ekonomiczny vs. aspekt społeczny* [w:] *Ekonomiczno-społeczne i techniczne wartości w gospodarce opartej na wiedzy*, t. 2, red. J. Buko, „Zeszyty Naukowe” nr 113, „Ekonomiczne Problemy Usług”, Uniwersytet Szczeciński, Szczecin.
- Wieczorkowski J., Polak P., 2014, *Big data: Three-aspect approach*, “Online Journal of Applied Knowledge Management”, Vol. 2.
- Yeo T.E.D., 2012, *Social-Media Early Adopters Don't Count: How to Seed Participation in Interactive Campaigns by Psychological Profiling of Digital Consumers*, “Journal of Advertising Research”, Vol. 52(3), <http://dx.doi.org/10.2501/JAR-52-3-297-308>.
- Zembik M., 2014, *Social media as a source of knowledge for customers and enterprises*, „Online Journal of Applied Knowledge Management” (A Publication of the International Institute for Applied Knowledge Management), Vol. 2(2), [http://www.iiakm.org/ojakm/articles/2014/volume2\\_2/OJAKM\\_Volume2\\_2pp132-148.pdf](http://www.iiakm.org/ojakm/articles/2014/volume2_2/OJAKM_Volume2_2pp132-148.pdf) (dostęp: 4.08.2014 r.).

### Streszczenie

Analizy *big data* otworzyły nowy rozdział w podejściu do wnioskowania na podstawie danych pozyskiwanych z Internetu. Przede wszystkim dostępne dane mają już nie tylko historyczny charakter i nie dotyczą tylko prób z badanych populacji, ale zyskały walor aktualności i masowości dzięki ekstrakcji danych z Internetu.

Nadal jednak poza obszarem obserwacji badawczych pozostali wykluczeni cyfrowo, którzy są niewidoczni w sieci choć pozostają częścią społeczeństwa.

Celem artykułu jest wskazanie możliwości wykorzystania analiz dużych zbiorów danych pozyskiwanych z Internetu oraz określenie wskaźników pochodzących z sieci społecznościowych, które mogą służyć do wnioskowania w opisie zjawisk społecznych, politycznych i gospodarczych.

*Słowa kluczowe:* big data, sieci społecznościowe, polityka prywatności

### **Big data Analysis in Social Network**

#### Summary

Analysis of *big data* opened a new chapter in the approach to inference on the basis of data obtained from the Internet. First of all available data no longer has just a historical nature and does not apply only to samples of the studied population, but have gained the values of current affairs and mass scale through extraction of data from the Internet.

Still, beyond the research observations are the digitally excluded who are invisible in the network although they remain a part of society.

The aim of this article is to show the possibility of analysing *big data* obtained from the Internet to identify indicators from social networks, which can be used for inference in the description of social, political and economic phenomena.

*Keywords:* *big data*, social networks, political of privacy

JEL: C55, L86, C80