

Maria RACZYŃSKA

Uniwersytet Technologiczno-Humanistyczny im. Kazimierza Pułaskiego
w Radomiu, Polska

Big Data – szanse i zagrożenia

Wstęp

Rozwój najnowszych technologii informatycznych sprawia, że mamy do czynienia obecnie z olbrzymią ilością pojawiających się danych, dużą ich różnorodnością, dużą szybkością pojawiania się danych i ich możliwością analizy w czasie rzeczywistym oraz coraz częściej znaczącą wartością tych danych. Te 4xV, czyli Volume (ilość danych), Variety (różnorodność danych), Velocity (masowy napływ danych) oraz Value (wartość płynąca z tych danych) są czynnikami charakterystycznymi dla obszaru Big Data. Big Data to analiza różnorodnych dostępnych w ogromnych ilościach zbiorów danych, często w czasie rzeczywistym. Szacuje się, że ilość informacji wytworzona przez ludzkość do 2007 r. wynosiła ok. 295 EB (EB-eksabajt; 1 EB to $2^{60} \approx 10^{18}$, czyli trylionów bajtów; jeden bajt odpowiada jednemu znakowi pisarskiemu). Do 2007 r. informacja była przechowywana w formie analogowej, dopiero w 2007 r. sytuacja uległa całkowitej zmianie i 94% informacji było w formie cyfrowej [Błoński 2013]. Ciekawym odkryciem jest „fakt nadawania dwóch zeta bajtów informacji, czyli odpowiednio 175 gazet dziennie na osobę” [Błoński 2013]. Według szacunków badania IDC Digital Universe ilość informacji na świecie wzrasta przeszło dwukrotnie [Szerokopasmowa... 2013]. Z badań wynika, że za 10 lat przedsiębiorstwa będą zarządzać 50-krotnie większą ilością danych i 75-krotnie większą liczbą plików. Dane te odzwierciedlają dynamikę przyrostu informacji.

Zjawisko Big Data wynika z faktu, że obecnie dane coraz częściej przechowywane są w sposób cyfrowy, co ułatwia ich gromadzenie, przetwarzania i ich analizę. Kenneth Cukier – współautor wydanej niedawno książki *Big Data: A Revolution that will transform how we live, work and think* – uważa, że „niezależnie od dziedziny, możemy zbierać dziś znacznie więcej danych i nauczyć się z nich w nowy sposób korzystać”. Jak zauważają autorzy, Big Data to rewolucja, która zmieni nasz sposób życia, pracy i myśli [Schoonberger, Cukier 2013]. Najnowsze trendy w IT mają olbrzymi wpływ na gospodarkę, naukę i społeczeństwo. Zarządzanie danymi, monitorowanie aktywności społecznej – to przykład korzyści wynikających z zastosowania najnowszej technologii IT [Hurvitz 2013].

Najnowsza technologia informacyjna jest zjawiskiem globalnym. Można powiedzieć, że oddziałuje zarówno na pojedynczego człowieka, jak i na wspól-

noty i narody. W społeczeństwie informacyjnym wymusza dokonywanie wyborów, podejmowanie decyzji, często w odniesieniu nie tylko do pojedynczego człowieka, ale również do całej organizacji. Jak słusznie zauważył znany pedagog – Waldemar Furmanek – „społeczeństwo informacyjne to społeczeństwo wiedzy, to wspólnota twórców wiedzy. W tym społeczeństwie znajdzie dla siebie godne miejsce tylko ten, kto potrafi tę wiedzę wykorzystać dla wytworów, usług i utworów cyfrowych. Człowiek nieposiadający odpowiedniej wiedzy w tym społeczeństwie nie znajdzie dla siebie godnego miejsca” [Furmanek 2007: 120]. Warto zwrócić uwagę za psycholog Marią Ledzińską, iż: „efektywność funkcjonowania w świecie – zwłaszcza w warunkach nadmiaru informacji – wyznaczana jest umiejętnością dokonywania wyborów. Trafne i podjęte w odpowiednim czasie decyzje przesądzają niejednokrotnie o naszych sukcesach życiowych, choć nie wszystkie są podejmowane świadomie i dobrze przemyślane” [Ledzińska 2004: 127].

Człowiek jest twórcą najnowszej analizy danych. Jak wykorzystuje najnowsze wytwory cyfrowe? Czy potrafi z najnowszej technologii godnie korzystać? Czy w zalewie informacyjnym człowiek dysponuje własnym aksjologicznym systemem filtrującym?

Big Data zmienia sposób myślenia o biznesie, zdrowiu, polityce, edukacji. Zmienia się również nasze życie – stajemy przed nieuchronnym końcem naszej prywatności. Nurtujące stają się pytania:

- Co jest wynikiem? Do czego ten wynik ma doprowadzić? Jak wielkie mają być te wytwory i usługi cyfrowe?
- Z czego ten wynik ma być uzyskany? Czy wszystkie informacje uzyskane z przetworzonych danych są zgodne z zapotrzebowaniami odbiorcy? W jaki sposób są one zdobyte? Czy zachowane zostały normy etyczne?
- Co dalej? Czy i w jaki sposób przechowywać przetworzone dane? Jak je zabezpieczyć?
- Czy istnieje granica ilości danych i przetworzonych zeń informacji?

„Człowiek skazany jest na informacje. Oznacza to nie tylko zależność jakości jego życia od informacji, i od tego, jakie są to informacje, w jakiej ilości występują, a także na ile są one obecne w doświadczaniu świata przez człowieka. Znaczą to także, że człowiek wartościuje, selekcjonuje informacje oraz je wytwarza, ciągle dążąc, w sposób świadomy, do przekraczania zadanych granic [Furmanek 2007: 185].

Jesteśmy świadkami radykalnego wzrostu znaczenia informacji i jej przetwarzania. Rozwój najnowszej technologii znacznie zmienia nie tylko skalę dostępu do informacji, ale również stwarza nowe możliwości analizy informacji. „Analiza informacji leży u podstaw niemal każdej codziennej czynności, podejmowanej decyzji, zaplanowanych działań; to sposób i narzędzie, które pozwala odnaleźć się w złożonej i wymagającej rzeczywistości” [Liedel, Piasecka, Aleksandrowicz 2012].

1. Big Data – nowe możliwości analizy danych

Wiele firm stoi obecnie przed problemem olbrzymiej ilości danych i trudności znalezienia najbardziej wartościowej informacji. Co zrobić z olbrzymią ilością danych, które napływają do firmy? Czy przechowywać wszystkie dane w firmie? Jeśli – tak – to w jakiej formie? Czy analizować wszystkie dane? Jak dowiedzieć się, które dane są najbardziej wartościowe? Analitycy biznesowi, menedżerowie w przedsiębiorstwach, architekci, deweloperzy, lekarze, prawnicy szukają metod i narzędzi, które pomogą przekształcić ogromną ilość danych w wartość danych. Biorąc pod uwagę olbrzymią ilość pojawiających się danych, analityka Big Data staje się wysoko ocenianą technologią. Z wypowiedzi jednego z przedstawicieli amerykańskiej firmy Gartner specjalizującej się w zagadnieniach strategicznego wykorzystania technologii informatycznych wynika, iż „Ilość rozproszonych danych ma wzrosnąć o 80% w ciągu kolejnych 5 lat [...]. Analityka *business intelligence* została najwyższej ocenioną technologią na rok 2012; szefowie informatyczni przedsiębiorstw dzięki łączeniu analityki z innymi rozwiązaniami uzyskują nowe możliwości” [Gartner... 2012]. Na przestrzeni ostatnich kilku lat trwają prace nad tworzeniem nowych narzędzi, które stworzą nowe możliwości analizy danych. Podstawową rolę w analizie tych danych odgrywają przejrzyste i użyteczne informacje zbierane z wielu źródeł danych i magazynowane pod postacią cyfrową.

Zdobywanie i wykorzystywanie wiedzy przez kadre kierowniczą przedsiębiorstw jest obecnie zaliczane do najbardziej konkurencyjnych czynników w gospodarce. Decydujące znaczenie dla praktyki gospodarczej ma obecnie Gospodarka Oparta na Wiedzy (GOW). Według definicji OECD jest to gospodarka oparta wprost na tworzeniu, traktowanym jako produkcja, oraz dalszym przekazywaniu, czyli dystrybucji oraz praktycznym wykorzystywaniu wiedzy i informacji. System informacyjno-komunikacyjny tworzy kluczową infrastrukturę Gospodarki Opartej na Wiedzy [Grudzewski, Hejduk 2006: 16]. Efektywność tego systemu zależy między innymi od umiejętności wykorzystania przez pracowników firmy możliwości systemu informacyjno-komunikacyjnego. Zatem nie tylko dostępność infrastruktury informacyjno-komunikacyjnej, w tym narzędzi systemu Big Data, ale nade wszystko umiejętność wykorzystania jej możliwości, umiejętność wyciągania prawidłowych wniosków, umiejętność przystosowania się do zmiennych warunków otoczenia.

Choć zjawisko Big Data jest stosunkowo nowym zjawiskiem, to zachodzące zmiany w nauce i technice, technologii oraz informatyce i telekomunikacji sprawiają, iż na rynku pojawia się wiele narzędzi umożliwiających zbieranie, gromadzenie, przetwarzanie i analizę dużej ilości danych. Są cennymi narzędziami dla szukających dróg wykonania zadania, analizowania zagrożenia czy ocenie posiadanych np. zasobów. Informacja jest jednym z najcenniejszych dóbr, obecnie oraz często w przeszłości. Cechą charakterystyczną dzisiejszej

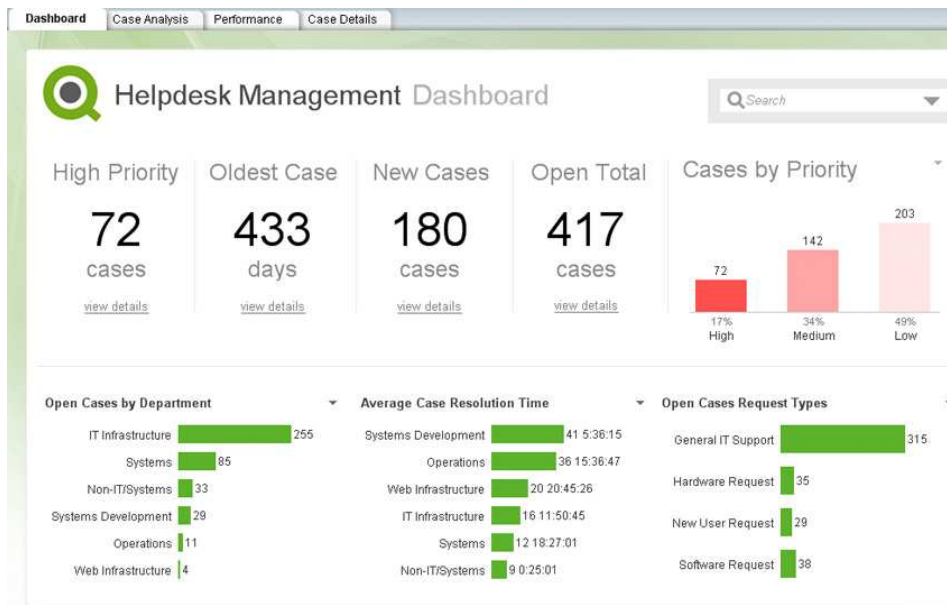
informacji jest jej cyfrowy format – łatwość gromadzenia, poszukiwania i analizy. Wiele instytucji i osób dzięki najnowszej technologii może obecnie korzystać z pełnej informacji. Jak słusznie zauważył prof. Włodzimierz Gogołek podczas jednej z konferencji „Informatyka w dobie XXI wieku”: „nigdy wcześniej tak wiele informacji o przeciętnym obywatelu nie było tak łatwo dostępne tak wielu osobom. W konsekwencji świat staje się globalną wioską, w której prywatność może okazać się anomalią” [Gogołek 2011: 235]. Informacja i informatyka, w najbardziej szerokim tego słowa znaczeniu, stały się obecnie nieodłączną częścią życia każdego społeczeństwa i prawie każdego człowieka. Nie jest istotne, czy dotyczy to jego bezpośrednio, czy też tylko pośrednio. Zarówno społeczeństwo, jak i człowiek jako jednostka musi w tym uczestniczyć. Zjawisko przymusu nowoczesności uznać należy za rys obecnych czasów [Furmanek 2011: 47].

Jednym z narzędzi, które może przekształcić ogromną ilość danych w wartość biznesową, jest aplikacja QlikView. QlikView to narzędzie *business intelligence* (BI), które pozwala na zbudowanie całościowego systemu BI w przedsiębiorstwie. To przykład oprogramowania, które łączy dane ze wszystkich dostępnych źródeł oraz dostarcza użytkownikom błyskawicznych odpowiedzi na zadawane pytania. Dzięki zastosowanej technologii AQL (*Associative Query Logic*), która w sposób automatyczny szuka powiązań w wewnętrznej bazie danych QlikView, istnieje nieskończenie wiele możliwości zadawania pytań ad hoc. Dzięki tej aplikacji można szybko przygotowywać raporty. Program ten pozwala w czasie rzeczywistym tworzyć analizy i wychwytywać np. jednocześnie zmiany na rynku i w kanale sprzedaży, dzięki czemu pozwala zaoszczędzić czas i ułatwia wnioskowanie. Na rys. 1 przedstawiony został przykład analizy danych przeprowadzonej za pomocą narzędzia QlikView.

Główne zalety i korzyści QlikView to:

- „użycie modelu danych ładowanego do pamięci;
- możliwość natychmiastowej (*in memory*) manipulacji ogromną ilością danych;
- nie wymaga dużych nakładów finansowych na sprzęt;
- automatyczna integracja danych i atrakcyjne dla klientów analityczne środowisko graficzne;
- bardzo efektywne i rozbudowane możliwości wizualizacji danych;
- łatwość użycia – użytkownik końcowy praktycznie nie wymaga treningów;
- wysoce skalowalne – niemal natychmiastowy czas odpowiedzi, nawet w przypadku olbrzymich wolumenów danych;
- szybka implementacja [...];
- elastyczność – umożliwia nieograniczone definiowanie wymiarów i miar, które mogą być modyfikowane w przeciągu kilku sekund;
- integracja – jest rozwiązaniem kompletnym, które na pojedynczej platformie umożliwia przeprowadzanie skomplikowanych analiz, raportów;
- niskie koszty – skrócenie okresu zwrotu z inwestycji, będące następstwem krótkiego okresu wdrażania;

- produkt wolny od ryzyka – dostępność w pełni funkcjonalnej wersji programu na okres próbny” [Data Warehouse 2013].



Rys. 1. QlikView – przykład analizy danych

Kolejnymi przykładami narzędzi Big Data są:

- Hadoop – platforma pamięci masowej, platforma do przechowywania i zarządzania dużych zbiorów danych; możliwość analizy dużych danych strukturalnych i niestructuralnych.
- MapReduce – system do tworzenia aplikacji działających jednocześnie na wielu tysiącach komputerów.

2. Przykład innowacyjnego wykorzystania danych Big Data

Jednym z przykładów wykorzystania Big Data jest wykorzystanie danych do poprawy usług publicznych, zwalczania przestępczości. Autorzy Raportu *A Revolution that will transform...* przedstawiają przykłady wykorzystania Big Data w Nowym Yorku. W mieście tym np. dokonano analizy danych z ponad 20 aptek, celem zwalczania procederu ubiegania się o nielegalną refundację leków. Okazało się po analizie danych oraz po przeprowadzeniu tradycyjnych audytów, że w 18 z badanych aptek rzeczywiście doszło do nadużyć: wartość wniosków pewnego lekarstwa znacznie odbiegała od normy [Hurvitz 2013]. Wykorzystanie Big Data pomogło zwalczyć proceder nielegalnego ubiegania się o zwrot refundacji za leki.

Kolejnym przykładem wykorzystania Big Data jest zapobieganie pożarom. Jak wiadomo, pożary wybuchają najczęściej w domach nielegalnie zasiedlanych, gdzie mieszka zbyt dużo osób, często lokalach nielegalnie rozbudowywanych. Analitycy z Nowego Yorku stworzyli bazę 900 tys. budynków w tym mieście, dołączyli dane z innych urzędów (podatki, wskaźniki przestępczości, liczba wezwań karetek pogotowia, straży pożarnej, anomalii wykorzystania prądu itp.). Dane te porównali z informacjami o pożarach z pięciu ostatnich lat. Przeprowadzona analiza danych wskazała na korelacje, które zwiększały ryzyko pożaru. Dodatkowo dzięki przeprowadzonej analizie ujawniony został czynnik, który wymagał pilnej interwencji. Czynnik ten dotyczył przeludnienia budynków.

Jednym z najbardziej znanych narzędzi wykorzystania Big Data jest narzędzie stworzone przez Google. Jest to narzędzie do szacowania analizy zachorowalności na gripę w czasie rzeczywistym. Analitycy wyodrębnili dane z lat 2003–2008 dotyczące zapytań w wyszukiwarce na temat grypy. Analizując dane dotyczące częstotliwości zapytań w wyszukiwarce na temat grypy, można było wysnuć wnioski dotyczące nasilania się wirusa grypy zarówno w kraju, jak i w różnych częściach świata. Analizując te dane, można było przeciwdziałać nasilaniu się aktywności tego wirusa.

Na podstawie zachowań klientów na stronach WWW oraz np. poszukiwanych przez niego towarów czy usług można zbudować profil klienta oraz dzięki narzędziom Big Data można dokonać automatyzacji marketingu. Poprzez np. monitorowanie poczty e-mail (nie tylko w formie newsletterów, ale również poczty wysyłanej z indywidualnych kont pracowników), które dostarczy informacji o otwartych mailach, klikniętych e-mailach, ściągniętych załącznikach, można zrealizować w firmie skuteczne działania sprzedażowe.

Trwają obecnie prace nad wykorzystaniem Big Data w Alior Banku. Bank obecnie określa możliwości wykorzystania innowacyjnych sposobów przetwarzania danych (np. danych pozyskiwanych z serwisów społecznościowych), by w przyszłości – jak wynika z wypowiedzi wiceprezesa Zarządu Alior Banku – zaproponować klientom lepsze rozwiązania, uprościć procesy oceny zdolności kredytowej czy spersonalizować produkty bankowe. Działania te z jednej strony mogą przyczynić się do lepszego funkcjonowania Alior Banku, z drugiej strony stać się mogą ostrzeżeniem dla użytkowników portali społecznościowych.

3. Big Data – czy to koniec prywatności?

Big Data to bezsprzecznie olbrzymie możliwości dla wielu firm. Warto jednak spojrzeć na to nowe zjawisko z perspektywy każdego z nas – użytkowników najnowszej technologii. Mając konto w banku i korzystając z karty kredytowej, jesteśmy niemal na każdym kroku „kontrolowani”: jakie artykuły kupujemy? Kiedy i gdzie? Z jaką częstotliwością? Banki i operatorzy kart płatniczych gromadzą na nasz temat olbrzymie bazy danych. Mając np. swój profil na jednym

z portali społecznościowych: Facebooku (popularny serwis społecznościowy ma ponad 900 milionów aktywnych użytkowników, w tym w Polsce 9,95 miliona) [Komputer Świat 2013] czy Twitterze i umieszczając tam swoje wpisy – pozostawiamy dane dla analityków, którzy dokonują różnego typu analiz: czym się interesujemy, z kim utrzymujemy kontakty, czy jesteśmy np. wiarygodnymi współpracownikami do biznesu, czy można nam udzielić kredytu, czy często chorujemy itp. W zależności od ilości i tematyki umieszczanych przez nas wpisów na tych portalach – baza o nas może być wykorzystana przez analityków różnych firm. Zbieranie i przetwarzanie danych budzi obawy wielu z nas. Zjawisko Big Data budzi wiele obaw związanych z polityką prywatności osób i firm, a także z bezpieczeństwem danych. Obok najnowszej generacji technologii informatycznych, sprzętu komputerowego, wysokospecjalistycznych aplikacji komputerowych, za pomocą których możliwe będzie zbieranie i przetwarzanie dużych ilości danych, a także ich analizowanie, niezbędny jest człowiek, który potrafi skorzystać nie tylko z najnowszej technologii, potrafi skorzystać z wiedzy, ale również myśli.

Wnioski

Otoczające nas zewsząd komputery, sieć Internetu, dane, liczby, fakty, zachodzące zmiany w nauce i technice, technologii oraz informatyce i telekomunikacji sprawiają, że zdobywanie i gromadzenie olbrzymich ilości danych oraz ich przetwarzanie nie stanowi dzisiaj problemu. Dane w erze Internetu nie stanowią już dzisiaj prawie dla nikogo problemu. Ważne staje się natomiast wydobywanie znaczącej wartości z informacji pozyskanych z danych. Ważny jest również atrybut informacji, jakim jest relewancja, czyli ważność nadana informacji zależna od odbiorcy. Ważne staje się również wyciąganie odpowiednich wniosków i uzyskana wiedza (na podstawie uporządkowanych informacji) oraz w konsekwencji użyta mądrość w praktyce. Nieodzowny staje się człowiek.

Big Data – nowe zjawisko analizy dużej ilości danych, z jakim mamy do czynienia od kilku lat ma już zarówno swoich wielu zwolenników, jak również przeciwników.

Czy możemy zatem mówić o jakiejś granicy dla innowacji? W dobie olbrzymiej konkurencji, wzrostu wydajności warto zastanowić się nad stale rosnącą ilością danych, zwiększeniem objętości i szczegółowości informacji uzyskiwanych z tych danych, informacji, które coraz częściej stają się kartą przetargową dla firm. Wykorzystanie dużych ilości danych stanowi niewątpliwie olbrzymi potencjał w działalności różnych sektorów. Mając na uwadze ten potencjał, warto jednak zastanowić się nad skutkami dużych danych. Zasady dotyczące prywatności, bezpieczeństwa, architektury danych, zarządzania dostępnością, przechowywania informacji, monitorowania sieci, własności intelektualnej, nadmiaru/chaosu danych – to kwestie, które wymagają badań w cyfrowym świecie danych. Zjawisko Big Data jest niewątpliwie olbrzymim sukcesem w dobie cyfryzacji danych, ale jed-

nocześnie budzi wiele obaw związanych z polityką prywatności pojedynczego człowieka i firm, a także z bezpieczeństwem danych. Budzi również wiele obaw w sferze moralności i etyki, która w dobie postępu, rozwoju cywilizacji współczesnej pozostaje coraz częściej w tyle.

Powraca uporczywe pytanie: **czy człowiek – w kontekście tego postępu – tworzący niezliczoną ilość danych, potrafi świadomie, odpowiedzialnie korzystać z tych danych?** Czy to, co przewidział już w latach siedemdziesiątych znany futurolog Alvin Toffler, iż szybko rosnąca ilość informacji wkrótce stanie się problemem dla ludzkości [Toffler 1970], jest już faktem? Czy istnieje granica ilości danych i przetworzonych zeń informacji? Big Data... i co dalej?

Literatura

- Furmanek W. (2007), *Jutro edukacji technicznej*, Rzeszów.
- Furmanek W. (2011), *Paradygmaty nauki i techniki współczesnej* [w:] *Informatyka w dobie XXI wieku. Nauka, Technika, Edukacja a nowoczesne technologie informatyczne*, red. A. Jastriebow, B. Kuźmińska-Sołśnia, M. Raczyńska, Radom.
- Gogołek W. (2011), *Rafinacja informacji sieciowej* [w:] *Informatyka w dobie XXI wieku. Nauka, Technika, Edukacja a nowoczesne technologie informatyczne*, red. A. Jastriebow, B. Kuźmińska-Sołśnia, M. Raczyńska, Radom.
- Grudzewski W.M., Hejduk I.K. (2006), *Wybrane problemy budowy systemu Gospodarki Opartej na Wiedzy GOW w Polsce* [w:] *Wybrane aspekty zarządzania wiedzą w przedsiębiorstwach Unii Europejskiej*, red. T. Krupa, Opole.
- Hurvitz J., Nugent A., Halper F., Kaufman M. (2013), *Big Data for Dummies*.
- Liedel K., Piasecka P., Aleksandrowicz T.R. (2012), *Analiza informacji. Teoria i praktyka*, Warszawa.
- Schoonberger V.M., Cukier K. (2013), *A Revolution that will transform how we live, work and think*, Boston, Nowy York.
- Toffler A. (1970), *Future Shock*, Bantam Book.

Netografia [2.05.2013]:

- Błoński M., *Ile informacji zgromadziła ludzkość* [w:] <http://kopalniawiedzy.pl/informacja-Martin-Hilbert-University-of-Southern-California,12495>
- Data Warehouse, info: <http://datawarehouse4u.info/Qlikview.html>
- Gartner Executive Programs' Worldwide Survey of More Than 2,300 CIOs Shows Flat IT Budgets in 2012, but IT Organizations Must Deliver on Multiple Priorities [w:] <http://www.microsoft.com/enterprise/pl/pl/trends/big-data.aspx>
- Komputer Świat, *Ilu Polaków jest na Facebooku*, www.komputerswiat.pl
- QlikView: <http://www.qlikview.com/pl/explore/products/overview>
- Szerokopasmowa: <http://szerokopasmowapolska.pl/co-dwa-lata-ilosc-danych-na-swiecie-wzrasta-przeszlo-dwukrotnie/>

Streszczenie

W artykule przedstawione zostały różne aspekty zjawiska Big Data. Ukazany został wpływ technologii informacyjnej na analizę różnorodnych dostępnych w ogromnych ilościach zbiorów danych, często w czasie rzeczywistym. Przedstawione zostały przykłady wykorzystania Big Data. Zwrócona została także uwaga na kwestie, które wymagają badań w cyfrowym świecie danych.

Słowa kluczowe: duże ilości danych, zbiory danych, uporządkowanie danych, przetwarzanie w chmurze, prywatność człowieka.

Big Data – chances and threats**Abstract**

The article presents the different aspects of phenomenon of Big Data. It shows the effect of information technology on the analysis of various available in huge amounts of data sets in real time. Presented are examples of Big Data. The return was a note on the issues that require research data in the digital world.

Key words: big data, sets of data, cloud computing, privacy, human.