

Kłeska cyfrowego urodzaju

Jeśli analityka Big Data miałaby swój prywatny dekalog, to jego pierwsze przykazanie brzmiałoby: „Dane należy wizualizować”. Big Data to nie tylko liczby i ciągi znaków w komórkach Excela. To żywe obrazy, interaktywne mapy i wykresy, które przemawiają do nas skuteczniej, niż dane w ich czystej formie. Tylko jak pokazać Big Data? Czy faktycznie wielkie zbiory danych to przekleństwo dla firm zajmujących się ich wizualizacją, czy może raczej klątwa urodzaju, która otwiera przed biznesem nowe drzwi do lepszego zrozumienia siebie i swojego otoczenia?

Big Data naprawdę jest „big”. Skalę wielkich zbiorów danych w sieci szacuje się dziś na poziomie ponad 8 zettabajtów (1 Zettabajt to równowartość 1 miliarda gigabajtów). Jednak w tym roku wolumen Big Data ma już przekroczyć próg 10 zettabajtów. Rokrocznie w internecie przybywa około 40 proc. nowych informacji. W sekundę w dzisiejszym Internecie przepływa tyle danych, ile liczyła cała globalna sieć jeszcze 20 lat temu. Dlatego niekiedy mówi się już nie tyle o „Big”, co Huge Data. Brytyjscy analitycy internetowi żartują nawet, że tak duża liczba danych w sieci wymaga wręcz stworzenia nowej jednostki: „hellabajta”. Ta nowa „jednostka” pomiaru danych w internecie stanowiłoby oczywiście żartobliwy skrót od „hell-a-lot-of-bytes”, co w wolnym tłumaczeniu oznacza „piekielnie dużo danych”. Jak dużo?

Według obliczeń firmy IBM każdego dnia generowanych jest ponad 2,5 mln terabajtów nowych danych, pochodzących z milionów różnych źródeł. By móc lepiej zwizualizować wielkość Big Data, należałoby przeliczyć te wartości na bardziej miarodajne i trafiające do wyobraźni wskaźniki. I tak – 1 terabajt danych to ponad 17 tys. godzin muzyki, ponad 132 tys. książek liczących powyżej 650 stron, około 310 tys. zdjęć, a także ponad 86 mln jednostronicowych dokumentów w Wordzie.

2,5 mln terabajtów nie mówi zatem jeszcze nic. Dopiero ich wizualizacja odsłania przed nami skalę Big Data, z jaką mamy obecnie do czynienia.

Nietrudno się domyślić, że fizyczne przejście wszystkich informacji generowanych jednego dnia w sieci oraz odkrycie występujących między nimi powiązań, mogłoby zająć całe lata. Cyfrowy potop, z którym mamy do czynienia każdego dnia, jest niemożliwy do okiełznania ani przez przedsiębiorstwo, ani – tym bardziej – przez pojedynczego człowieka. Dlatego firmy są dziś wyręczane przez inteligentne algorytmy oraz techniki uczenia maszynowego, które same potrafią wizualizować dane, gromadząc je z setek tysięcy różnych źródeł i tworząc z nich interaktywne mapy. To dzięki nim uzyskujemy błyskawiczny dostęp do interesujących nas zagadnień, wyklikując tylko odpowiednie opcje. Dane zaprezentowane na mapach są również zdecydowanie łatwiejsze w analizie oraz implementacji w biznesowej strategii przedsiębiorstwa. Mają przynajmniej jedną, niepodważalną przewagę nad danymi w surowej postaci: pozwalają firmie zaoszczędzić czas. Wystarczy przecież rzut oka na mapę zawierającą naniesione na nią dane, aby uzyskać pełne rozeznanie w interesującym nas temacie. Mapy danych są dziś jednym z najskuteczniejszych narzędzi stanowiących źródło insightów dla firm.

Wizualizacja danych pozwala również odkryć powiązania między poszczególnymi składowymi cyfrowego świata, rozpoznać wzorce zachowań, relacje, czy trendy, a także antycypować nadchodzące zmiany. Co więcej – dane na mapach mogą być aktualizowane w czasie rzeczywistym zmieniając np. swój kolor kształt czy rozmiar. Dzięki takim mapom biznes zawsze ma dostęp do najświeższych informacji i nie musi kolejny raz przeszukiwać raportów w poszukiwaniu zmian. Takich wiadomości nie wyczyta się w raportach. Można je tylko zobaczyć.

Biznes określa to zjawisko jako „data visualisation”. Stanowi ono jeden z kluczowych elementów strategii firm w nadchodzących latach. Umożliwia bowiem przetworzenie tzw.

„surowych danych” („raw data”), które odkoduje zapewne wąskie grono firmowych specjalistów – w dane inteligentne („smart data”), które będą zrozumiałe już dla wszystkich pracowników firmy dzięki swojej intuicyjności. Według badania przeprowadzonego przez firmę Markets and Markets, globalny rynek wizualizacji danych rośnie w tempie 10,9 proc. rok do roku. Do końca 2020 roku przekroczy już próg 2,8 mld dolarów. Szczególnie cenne są dane demograficzne oraz dane geograficzne, dotyczące np. lokalizacji. Według badania Forbes Insights 2015 ponad 60 proc. danych gromadzonych przez marketerów z całego świata stanowią właśnie dane tego typu, choć de facto zwizualizować można w zasadzie każdy element z internetowego strumienia danych.

Jednak tak rozumiana analityka Big Data nie każdemu jest w smak. David Koretz z HackerInterrupted, popełnił swego czasu ironiczny i dość szyderczy tekst wymierzony w tę technologię. Później ten materiał zapewne długo odbijał mu się czkawką. W artykule o wymownym tytule „Big Data jest głupia” (*Big Data is Dumb*)[1], Koretz pisał: „Big Data was rozczaruje. To tylko moda, która coraz bardziej zapycha ludziom gardła. Okaże się przelotnym trendem, rozczarowującym tyleż fundusze inwestycyjne, jak i dyrektorów działów informatyki. Po 20 latach wpatrywania się w dane lekarze do dziś nie są w stanie rozstrzygnąć, czy masło czy też może margaryna są lepsze dla zdrowia. Big Data będzie tylko na tyle inteligentna, na ile inteligentne są nasze zbiory danych”.

Koretz nie miał racji. Analityka Big Data to dziś biznesowy fenomen. Ten segment cyfrowej gospodarki rozwija się już w tempie sześciokrotnie szybszym niż cała branża IT. Według firmy analitycznej Gartner to właśnie wielkie zbiory danych są dziś kołem zamachowym przychodów w przedsiębiorstwach, a dzięki analityce Big Data do 2020 roku uda się zautomatyzować około 80 proc. procesów biznesowych w firmach.

Alfabet też jest jakimś zbiorem danych. A mimo to nie można uznać go za dzieło literackie. Jak to się zatem dzieje, że nie

wszystko, co wykorzystuje litery alfabetu, jest takim dziełem? To proste: genialny pisarz rozpoznaje wcześniej nieodkryte połączenia między literami, wyrazami i zdaniami, a następnie buduje z danych alfabetu całość, której wcześniej nikt przed nim nie stworzył. Nanosi dane na arkusz papieru. Dokładnie tak samo rzecz ma się z wizualizacją danych i tworzeniem z nich interaktywnych map, które w przejrzysty sposób unaoczniają kluczowe informacje ze strumienia Big Data. Dane stanowią zatem tylko pewne tworzywo, które – odpowiednio dobrane, wymodelowane i pokazane – pozwala unaocznić ich sensowność oraz zakodowaną w nich wartość dla przedsiębiorstw.

Dlatego też eksplozji danych w Sieci nie należy rozpatrywać jako biznesowego przekleństwa. To raczej „klęska” innego typu: klęska urodzaju. Problem firm nie polega dziś bowiem na deficycie danych, lecz na ich nadmiarze. Jeśli istnieje jakaś szalupa ratunkowa, która pomogłaby firmom przetrwać w tym cyfrowym potopie, to bez wątpienia jest nią właśnie „data visualisation”.

[1]

<http://forums.juniper.net/t5/Hacker-Interrupted/Big-Data-is-Dumb/ba-p/150630>