

## ***Big data* w perspektywie memetycznej**

### **Streszczenie**

W artykule postawione zostaje pytanie o adekwatność opisów i wyjaśnień memetycznych do fenomenu *big data*. Takie ujęcie rzuca nowe światło zarówno na zjawisko *big data*, jak i na memetykę pretendującą do roli adekwatnej i zupełnej teorii kultury. Artykuł składa się z trzech części. W pierwszej zostaje przedstawione, czym jest *big data* i jakie niesie za sobą konsekwencje. Druga zawiera dociekania, na ile memetyczne opisy i wyjaśnienia kultury stosują się do *big data*. Podsumowanie stanowi odpowiedź na postawione w tekście pytania.

### **Słowa-klucze**

big data, memetyka, mem, kultura, teoria kultury

## **Big data from the memetic point of view**

### **Summary**

The article puts a question about the adequacy of memetic descriptions and explanations to the phenomenon of big data. This view throws light on big data and the memetics as an adequate and complete theory of culture. The article consists of three parts. In the first part I present the phenomenon of big data and its consequences. The second is discussion about the applications of memetic descriptions and explanations to big data. The summary contains the answers to posed questions.

### **Keywords**

big data, memetics, meme, culture, theory of culture

---

<sup>1</sup>Ze względu na występowanie w literaturze polskojęzycznej zwrotu angielskiego i brak kanonicznych tłumaczeń na polski zdecydowałem się na pozostawienie zwrotu *big data* w oryginalnej wersji.

*Big data* to fenomen wnikliwie ostatnio analizowany i komentowany zarówno w publikacjach specjalistycznych, jak i w tych przeznaczonych dla szerszego grona odbiorców. Najwięcej miejsca poświęca się zastosowaniom *big data*, zwracając szczególną uwagę na ich potencjał biznesowy. Dyskutuje się kwestie prawne i etyczne związane z gromadzeniem i przetwarzaniem danych, a także z przewidywaniem na ich podstawie ludzkich zachowań. Spory związane z *big data* sięgają jednak też na grunt metodologii i filozofii nauki.

Niniejszy artykuł nie będzie stanowił omówienia wszystkich aspektów i kwestii związanych z tym zjawiskiem, lecz jedynie tych, które są niezbędne do realizacji stawianego celu, którym jest odpowiedź na pytanie o stosowalność opisów i wyjaśnień memetycznych do *big data*. Zbadam, czy memetyka jest w stanie rzucić nowe światło na omawiany fenomen, oraz czy *big data* zmusza do uzupełnienia czy nawet rewizji niektórych założeń i twierdzeń memetyki.

Artykuł składa się z trzech części. W pierwszej przybliżę, czym jest *big data* i jakie niesie za sobą istotne konsekwencje. W drugiej zbadam, na ile memetyczne opisy i wyjaśnienia kultury stosują się do *big data*. Na koniec zastanowię się, czy powinniśmy zmodyfikować memetykę i jeśli tak, to w jakich aspektach i w jakim stopniu.

## Czym jest *big data*?

*Big data* wiąże się z błyskawicznym przetwarzaniem ogromnych ilości danych, zbieranych często w różnych sytuacjach i za pomocą odmiennych narzędzi, zapisanych na różnych nośnikach i w różnych formatach. Dane te analizowane są w celu wychwycenia korelacji, wzorców i trendów, pomocnych m.in. w prognozowaniu zjawisk i tendencji (np. rozprzestrzeniania się epidemii – Google Flu, czy inflacji – PriceStats), ale też ludzkich zachowań, preferencji i nawyków (np. zachowań konsumenckich czy preferencji czytelniczych – Amazon).

Viktor Mayer-Schönberger i Kenneth Cukier zwracają uwagę na wieloznaczność wyrażenia *big data*<sup>2</sup>. Sami proponują niezbyt precyzyjną definicję:

Obecnie myśli się o zagadnieniu *big data* – i takie podejście jest prezentowane w niniejszej książce – następująco: uważa się, że termin ten obejmuje to, co może być zrealizowane w dużej skali, a nie może być wykonane w małej, w celu zyskania nowej wiedzy lub stworzenia nowej wartości w sposób, który zmieni rynki, organizacje, relacje między rządami a obywatelami itp.<sup>3</sup>

Takie dookreślenie jest – według mnie – za szerokie, ponieważ działania realizowalne tylko w dużej skali w celu zyskania nowej wiedzy lub stworzenia nowej wartości wcale nie muszą wiązać się z analizą wielkiej ilości danych, a więc z tym, co standardowo rozumie się pod pojęciem *big data*. Przykładem tego mogą być protesty społeczne zmierzające do zmiany władzy (jak chociażby na Ukrainie czy podczas tzw. Arabskiej Wiosny). Mamy tu do czynienia z działaniami, które mogą osiągnąć swój cel wyłącznie przy dużej skali protestów, przyczyniają się one do stworzenia nowej wartości i zmieniają relacje między rządem a obywatelami, natomiast nie wiążą się z przetwarzaniem wielkiej ilości danych w celu wydobycia z nich korelacji i wzorców. Nie powinno się ich zatem zaliczać do *big data*, choć zdają się spełniać definicję Mayer-Schönbergera i Cukiera.

<sup>2</sup> V. Mayer-Schönberger, K. Cukier, *Big data. Rewolucja, która zmieni nasze myślenie, pracę i życie*, przeł. M. Glatki, Wydawnictwo MT Biznes, Warszawa 2014, s. 19.

<sup>3</sup> Tamże, s. 20.

Trafniejszą definicję przedstawia ekspert z firmy Gartner<sup>4</sup> – Doug Laney. Charakteryzuje on *big data* poprzez trzy V – ogromną wielkość [*volume*] zbiorów danych, dużą prędkość [*velocity*] ich przetwarzania (ze szczególnym naciskiem na przetwarzanie w czasie rzeczywistym) oraz wielką różnorodność [*variety*] danych, zapisanych na różnych nośnikach, w różnych formatach<sup>5</sup>. Doskonałym przykładem działania *big data* jest personalizacja reklam na podstawie danych zebranych z portali społecznościowych, odwiedzanych przez użytkownika stron, poleceń, pozostawionych wpisów i komentarzy. Jest to pokaźny zbiór danych, przechowywanych w różnych formatach i na różnych serwerach, poddanych błyskawicznej analizie w celu wydobycia preferencji, zainteresowań, pragnień, potrzeb, nawyków, a także statusu konsumenta.

Mayer-Schönberger i Cukier dostrzegają rewolucyjny potencjał tkwiący w *big data*:

Pojawienie się zjawiska *big data* stanowi ważny krok w ludzkim dążeniu do skwantyfikowania i zrozumienia świata. Większość fenomenów, które nigdy wcześniej nie mogły być zmierzone, gromadzone, przeanalizowane i udostępnione, staje się przeliczalnymi danymi. Korzystanie z dużych ilości danych zamiast małych porcji i akceptowanie niedokładnych informacji otwierają drzwi do nowych poziomów poznania. Sprawia to, że społeczeństwo zrezygnuje z preferowanego wcześniej poszukiwania przyczynowości na rzecz korzyści płynących z korelacji.<sup>6</sup>

Autorzy wskazują na trzy zmiany, jakie wiążą się z rozwojem *big data*<sup>7</sup>. Pierwsza dotyczy możliwości efektywnego analizowania ogromnych zbiorów danych, bez konieczności ograniczania się do mniejszych zbiorów (np. do stosowania metody próby losowej). Dzięki temu jesteśmy w stanie wydobyć interesujące korelacje niemożliwe do uzyskania dawniej stosowanymi metodami.

Druga zmiana związana jest z tym, że przy korzystaniu z odpowiednio dużych zbiorów danych nie musimy, jak przy wcześniejszych metodach (np. losowej próby), martwić się o dokładność i uporządkowanie danych. Jest bowiem tak, że występujące gdzieś niedokładności są pomijalne w obliczu ogromnej ilości danych, nie wpływają na uzyskany wynik, a także nie przeszkadzają w wykrywaniu korelacji i trendów<sup>8</sup>. Omawiane niedokładności mogą mieć różne pochodzenie; Mayer-Schönberger i Cukier wymieniają następujące<sup>9</sup>:

- wzrost liczby błędów postępujący wraz ze wzrostem ilości danych;
- łączenie niepasujących do siebie informacji;
- niekonsekwentne, odmienne sposoby formatowania danych;
- niedokładności biorące się z przetwarzania danych.

Trzeba zaznaczyć, że nieuporządkowanie nie jest konstytutywną cechą *big data*, lecz bierze się z niedoskonałości naszych narzędzi używanych do mierzenia, zapisywania i analizy danych – istnieje zatem szansa, że wraz z rozwojem technologii nieuporządkowanie danych

<sup>4</sup> Dawniej META Group.

<sup>5</sup> <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (stan na 30 kwietnia 2015 r.). W niektórych publikacjach dodaje się kolejne dwa V: *veracity*, czyli wiarygodność, prawdziwość danych, oraz *value*, czyli ich wartość dla użytkownika. J. Dygaszewicz, *Big data w statystyce publicznej*, w: *Internet. Publiczne bazy danych i big data*, red. G. Szpor, Wydawnictwo C.H. Beck, Warszawa 2014, s. 51.

<sup>6</sup> V. Mayer-Schönberger, K. Cukier, *Big data...*, s. 34.

<sup>7</sup> Tamże, s. 28–30, 35–36.

<sup>8</sup> Tamże, s. 61–62.

<sup>9</sup> Por. tamże, s. 53.

zostanie zmniejszone, a może nawet całkowicie wyeliminowane<sup>10</sup>. Nie oznacza to jednak, iż do tego czasu powinniśmy poprzestać na analizowaniu niewielkich zbiorów zawierających dokładne dane<sup>11</sup>. Zdaniem Mayer-Schönbergera i Cukiera, powinniśmy wyjść poza nawykowe ograniczenia i zacząć wykorzystywać potęgę *big data*:

Spółeczeństwo przystało na dwa niejawne kompromisy, które tak wrosły w naszą naturę, że traktujemy je jako naturalną kolej rzeczy. Po pierwsze, zakładamy, że nie możemy wykorzystywać dużo większej liczby danych, więc tego nie robimy. Jednak ograniczenie to staje się coraz mniej istotne, ponieważ powoli uświadamiamy sobie, jak wiele możemy zyskać, używając wielkości zbliżających się do  $N = \text{całość}$ . Drugi kompromis dotyczy jakości informacji. W epoce małych ilości danych położenie nacisku na dokładność było racjonalnym zachowaniem, bo kiedy zbiera się niewielką liczbę informacji, muszą być one jak najdokładniejsze. W wielu przypadkach reguła ta dalej ma zastosowanie, ale w innych rygorystyczna precyzja jest mniej ważna niż błyskawiczne otrzymanie wyniku z zakresu jakiegoś zagadnienia lub szybkie rozpoznanie pojawiających się trendów<sup>12</sup>.

I wreszcie trzecia zmiana wiąże się z przeniesieniem nacisku z szukania związków przyczynowo-skutkowych na wyszukiwanie korelacji. Choć skłonność do odnajdywania relacji kauzalnych jest dla człowieka naturalna, to powinna być powściągnięta z kilku powodów<sup>13</sup>. Po pierwsze, wspomniana skłonność często prowadzi na manowce, czego najlepszym przykładem jest powszechność błędu *post hoc ergo propter hoc*<sup>14</sup>. Po drugie, twierdzenia o istnieniu związków przyczynowych są o wiele trudniejsze do uzasadnienia niż twierdzenia o istnieniu korelacji. Po trzecie, skoncentrowanie się na badaniu korelacji może sporo wnieść do poszukiwań relacji przyczynowych. I wreszcie po czwarte, zbyt ni nacisk na przyczynowość nie zawsze jest pożądany z tego względu, że zaciemnia obraz sytuacji. Autorzy ilustrują to w następujący sposób: analiza korelacji pokazała, że pomarańczowe samochody psuły się o połowę rzadziej od średniej, jednakże mimo powstania wielu hipotez przyczynowych tłumaczących tę zależność nie udało się ustalić, która jest adekwatna, co wszak nie przekreśliło korzyści biorących się z poznania owej korelacji<sup>15</sup>. I choć analiza korelacji nie zastąpi całkowicie szukania związków przyczynowo-skutkowych<sup>16</sup>, to właśnie na nią zostanie w przyszłości położony nacisk: „Nie

<sup>10</sup> Por. tamże, s. 62.

<sup>11</sup> Tym bardziej, że ideał polegający na dysponowaniu wszystkimi możliwymi danymi o największym możliwym stopniu dokładności jest nieziszczalny. Jak bowiem realistycznie zauważają Mayer-Schönberger i Cukier: „Ilości danych, które będziemy w stanie zgromadzić i przetworzyć zawsze będą niewielką częścią wszystkich informacji istniejących we wszechświecie [...] Ponieważ nigdy nie będziemy mieli dokładnych informacji, nasze przewidywania zawsze będą obarczone błędem. Nie oznacza to, że będą złe, będą po prostu niekompletne. Nie unieważnia to jednak wiedzy, którą poznajemy dzięki *big data*”. Tamże, s. 258.

<sup>12</sup> Tamże, s. 71.

<sup>13</sup> Tamże, s. 90–96.

<sup>14</sup> Inaczej błąd fałszywej przyczyny. Polega na wyciągnięciu błędnego wniosku o powiązaniu przyczynowym dwóch zjawisk z faktu, że jedno zjawisko wystąpiło czy występuje po drugim.

<sup>15</sup> Tamże, s. 94–95.

<sup>16</sup> Mayer-Schönberger i Cukier są mniej radykalni niż inni piewcy *big data*. Chris Anderson stwierdził, że tradycyjny model odkrycia naukowego, sprowadzający się do sformułowania hipotezy i jej empirycznego przetestowania, zostanie zastąpiony przez analizę danych, która sama nie będzie potrzebowała teorii. Ch. Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, „Wired” 2008. Mayer-Schönberger i Cukier słusznie wskazują na absurdalność takiego ujęcia, ponieważ sama analiza *big data* opiera się na teoriach matematycznych, statystycznych i informatycznych, a ram pojęciowych i teorii potrzebujemy już na etapie selekcji danych, nie mówiąc o ich analizie i interpretacji wyników. V. Mayer-Schönberger, K. Cukier, *Big data...*, s. 100–101.

zrezygnowaliśmy z szukania przyczynowości, ale spadła ona z piedestału głównego źródła poznania. *Big data* jest turbodoładowaniem dla analizy nieszukającej związków przyczynowych, które tak często zastępuje badanie przyczynowości<sup>17</sup>.

Kolejnym interesującym fenomenem związanym z *big data* jest tendencja do gromadzenia danych, co autorzy określają mianem danetyzacji.

Danetyzacja oznacza zbieranie informacji o wszystkim – wliczając w to kwestie, o których nigdy nie myśleliśmy jak o źródłach danych, takie jak miejsce przebywania konkretnej osoby, wibracje silnika czy naprężenia występujące w moście i przetworzenie ich w określony format w celu skwantyfikowania.<sup>18</sup>

Autorzy przedstawiają później zresztą nieco inne rozumienie tego pojęcia, traktując danetyzację jako przedstawienie czegoś w skwantyfikowanej formie, czyli takiej, którą można zestawić w tabelach i przeanalizować<sup>19</sup>. Trzeba przy tym pamiętać, że danetyzacja nie rozpoczęła się wraz z cyfryzacją, lecz jest od niej wcześniejsza, choć trzeba przyznać, że cyfryzacja bardzo mocno napędziła danetyzację<sup>20</sup>. Wystarczy przyjrzeć się szacunkom Martina Hilberta<sup>21</sup>. Jego zdaniem, w 2007 r. zgromadzono 300 eksabajtów<sup>22</sup> danych, z tego jedynie 7% było zapisane w formie analogowej. Szacunki na 2013 r. wskazują już na 1200 eksabajtów, z czego jedynie 2% znajdowało się w formie analogowej<sup>23</sup>.

Danetyzacja postępuje dlatego, że wszelkie dane stały się wartościowe. Ich wartość nie tkwi jedynie w obecnym czy nawet planowym wykorzystaniu, ale jest funkcją wszelkich zastosowań przeszłych, obecnych i przyszłych<sup>24</sup>. Dane, które dzisiaj wydają się bezwartościowe, jutro mogą znaleźć zastosowanie i odegrać kluczową rolę. Z tego względu coraz bardziej widoczna jest tendencja do gromadzenia danych „na zapas”, do późniejszego wykorzystania lub udostępnienia (najczęściej odpłatnego) podmiotom, które będą umiały je zagospodarować. Ma to tym większy sens, że koszty gromadzenia i przechowywania danych sukcesywnie się zmniejszają.

Nasza epoka – piszą Mayer-Schönberger i Cukier – różni się tym, że wiele wewnętrznych ograniczeń związanych z gromadzeniem danych przestało istnieć. Technologia osiągnęła punkt, w którym duże ilości informacji mogą być tanio wychwycone i zapisane. Dane często mogą być zbierane pasywnie<sup>25</sup>, bez dużego wysiłku, ci zaś, o których informacje są gromadzone, często nie mają świadomości, że są podmiotem takich działań. Ponieważ

---

<sup>17</sup>Tamże, s. 96.

<sup>18</sup>Tamże, s. 31.

<sup>19</sup>Tamże, s. 109.

<sup>20</sup>Tamże, s. 103–118.

<sup>21</sup>M. Hilbert, P. Lopez, *The World's Technological Capacity to Store, Communicate, and Compute Information*, "Science" 2011, nr 4, s. 60–65; M. Hilbert, P. Lopez, *How to Measure the World's Technological Capacity to Communicate, Store and Compute Information*, "International Journal of Communication" 2012, s. 1042–1045.

<sup>22</sup>1 eksabajt = 1 miliard gigabajtów.

<sup>23</sup>V. Mayer-Schönberger, K. Cukier, *Big data...*, s. 24.

<sup>24</sup>Tamże, s. 135–141.

<sup>25</sup>Dotyczy to szczególnie tzw. danych resztkowych, czyli śladów powstałych w wyniku ludzkiej aktywności. W świecie realnym takimi danymi są np. informacje o lokalizacji człowieka, zmianach jego parametrów fizjologicznych (monitorowane przez urządzenia wyposażone w GPS i odpowiednie czujniki), w Internecie są to odwiedzane strony, miejsca kliknięć i ruchy myszki itp. Zbieranie takich danych odbywa się automatycznie przez przeznaczone do tego urządzenia i programy. Tamże, s. 149–154.

koszt przechowywania informacji spadł bardzo mocno, łatwiej usprawiedliwić [to] zachowanie, niż pozbyć się danych. Wszystko to sprawia, że dużo więcej danych niż kiedykolwiek wcześniej jest dostępniejszych i tańszych<sup>26</sup>.

Danetyzacja przyczyniła się zresztą do powstania nowych dyscyplin badawczych. Świetnym tego przykładem jest kulturomika, będąca odpowiedzią na prace Google'a nad zdanetyzowaniem każdej dostępnej książki, badająca ludzkie zachowania i kulturowe trendy przy pomocy statystycznej analizy tekstów<sup>27</sup>.

## **Big data a memetyczne opisy i wyjaśnienia kultury**

Po ogólnym przedstawieniu fenomenu *big data* warto zbadać, na ile opisy i wyjaśnienia w kategoriach memetycznych dają się zastosować do omawianego zjawiska. W dalszej kolejności, o ile memetyka nie poradzi sobie z tym zadaniem, trzeba będzie rozstrzygnąć, czy memetyka jako teoria wymaga odpowiedniej korekty. Innym możliwym rozwiązaniem jest uznanie, że *big data*, a przynajmniej niektóre jego elementy czy aspekty, nie należą do przedmiotu zainteresowań memetyki. W obu wypadkach istnieje spora szansa, że dowiemy się czegoś nowego zarówno o *big data*, jak i o samej memetyce.

Na wstępie wypada wyraźnie zaznaczyć, że wśród memetyków można zaobserwować różnice związane z rozumieniem memów i ich roli w kulturze. Z tego względu niezwykle trudno jest wyznaczyć memetyczny „standard” w opisie i wyjaśnianiu kultury. Mimo tego postaram się taki „standard” zrekonstruować, posiłkując się twierdzeniami przedstawicieli memetyki.

Twierdzeniem łączącym memetyków jest uznanie memu za podstawową jednostkę przekazu kulturowego<sup>28</sup>, a zarazem za pewną porcję informacji<sup>29</sup> – analogicznie do genu, który stanowi podstawową jednostkę przekazu informacji biologicznej. W wypadku memu najczęściej jest to informacja analogowa, czasami podkreśla się jej semantyczny charakter<sup>30</sup>. Zdaniem prekursora memetyki, memy – podobnie zresztą jak geny – należą do replikatorów, czyli cechują się długowiecznością (jako memy-typy, a niekoniecznie jako poszczególne egzemplarze), płodnością, czyli łatwością rozprzestrzeniania się, a także względną wiernością kopiowania<sup>31</sup>. Jako replikatory są „samolubne”, czyli „dążą” do przetrwania przez wytworzenie jak największej liczby swoich kopii<sup>32</sup>. Nie jest to łatwe, gdyż zasoby ich pierwotnego środowiska – ludzkich

<sup>26</sup> Tamże, s. 136.

<sup>27</sup> Tamże, s. 116.

<sup>28</sup> R. Dawkins, *Samolubny gen*, przeł. M. Skoneczny, Prószyński i S-ka, Warszawa 2007, s. 244.

<sup>29</sup> Tamże, s. 176. R. Dawkins, *Fenotyp rozszerzony*, przeł. J. Gliwicz, Prószyński i S-ka, Warszawa 2003, s. 146. R. Brodie, *Wirus umysłu*, przeł. P. Turski, TeTa Publishing, Łódź 1997, s. 29. H. Bjarneskans, B. Grönnevik, A. Sandberg, *Cykl życiowy memów*, w: *Infosfera. Memetyczne koncepcje kultury i komunikacji*, red. D. Wężowicz-Ziółkowska, WSZOP, Katowice 2009, s. 67. F. Heylighen, *Memetyka*, w: *Infosfera...*, Katowice 2009, s. 27.

<sup>30</sup> Czynią to Nowak i Borkowski: W. Borkowski, A. Nowak, *Wpływ społeczny – alternatywny model rozprzestrzeniania się memów*, w: *Infosfera...*, Katowice 2009, s. 176.

<sup>31</sup> R. Dawkins, *Samolubny gen...*, s. 246–247.

<sup>32</sup> Cudzysłów nie jest tu przypadkowy, mamy tu bowiem do czynienia z metaforami, będącymi użytecznymi skrótami. Przykłady rozwinięć czy parafraz tego typu skrótów: S. Blackmore, *Maszyna memowa*, przeł. N. Radomski, Dom Wydawniczy Rebis, Poznań 2002, s. 28–29. H. Bjarneskans, B. Grönnevik, A. Sandberg, *Cykl życiowy memów...*, s. 69–70.

i zwierzęcych mózgów (*resp.* umysłów)<sup>33</sup> – są ograniczone, a przez to memy zmuszone są do konkurowania. Konkurencja zmusza je do tworzenia większych całości – mempleksów – czyli zbiorów „mutacyjnie wspomagających się, stowarzyszonych memów, koewoluujących w relacji symbiotycznej”<sup>34</sup>. U podstaw konkurowania memów leży to, że wywierają one wpływ na zachowania swoich nosicieli<sup>35</sup>. Ze względu na ten wpływ – nazywany femotypowym lub socjotypowym<sup>36</sup> – memy i mempleksy stają się mniej lub bardziej atrakcyjne dla nosicieli, z czym powiązana jest liczba ich kopii na pierwotnych nośnikach, jakimi są mózgi czy też umysły, a także na wtórnych nośnikach, takich jak czasopisma, książki, dokumenty, dzieła sztuki, nośniki cyfrowe (dyski, twarde dyski, pendrive’y itp.).

Warto przy tym podkreślić, że między memami a powodowanymi przez nie zachowaniami nie zachodzi relacja jedno-jednoznaczna. Dany mem może wywoływać odmienne zachowania u różnych nosicieli, a nawet u tego samego nosiciela w różnych momentach czy sytuacjach, z kolei dane zachowanie może być efektem działania różnych memów. To wskazuje na niejednoznaczne skutki informacji memetycznej i ich zależność od wielu czynników: od obecności innych memów, poprzez cechy danego nosiciela, aż po właściwości sytuacji, w jakiej nosiciel się znajduje.

Kolejne powszechnie uznane w memetyce twierdzenie głosi, że memy przekazywane są na drodze niegenetycznej. Istnieją wyraźne różnice w stanowiskach poszczególnych memetyków co do charakteru tego przekazu. Richard Dawkins i Susan Blackmore podkreślają, że memy rozprzestrzeniają się przez naśladownictwo<sup>37</sup>. Jest to ujęcie dość wąskie. Inni memetycy uznają, że transmisja memów przebiega na drodze szeroko pojętego uczenia się lub też – jeszcze szerzej – w dowolny, niegenetyczny sposób (tak twierdzą np. Nowak i Borkowski<sup>38</sup>).

Przejdźmy do charakterystyki *big data* za pomocą kategorii memetycznych, koncentrując się przy tym na porównaniu memów i danych. W przypadku *big data* mamy do czynienia z danymi, które – w myśl hierarchii DIKW<sup>39</sup> – należy odróżniać od informacji. Według Jona Michaela Dunna, tylko zarejestrowane i zinterpretowane przez kogoś dane stanowią informację<sup>40</sup>. Peter Drucker uważa z kolei, że informacje to „dane przedstawione w jakimś

---

<sup>33</sup> Memetycy różnią się szczegółowością, z jaką charakteryzują pierwotne nośniki czy też pierwotne środowisko memów. Niektórzy wskazują na struktury mózgowe (R. Dawkins, *Fenotyp rozszerzony*, s. 146. W. Borkowski, A. Nowak, *Wpływ społeczny...*, s. 176; S. Blackmore, *Maszyna memowa...*, s. 79), inni na umysły (R. Brodie, *Wirus umysłu...*, s. 29), jeszcze inni na ich składniki np. pamięć semantyczną i odpowiadające jej korelaty neuronalne (E.O. Wilson, *Konsiliencja*, przeł. J. Mikos, Zysk i S-ka, Poznań 2002, s. 206). Z kolei wtórnymi nośnikami mogą być fale akustyczne (mowa, melodie, dźwięki itp.), ruchy ciała, czasopisma, książki, dzieła sztuki, urządzenia magazynujące informacje w rozmaitych formatach itp.

<sup>34</sup> *Infosfera...*, s. 237.

<sup>35</sup> R. Brodie, *Wirus umysłu*, s. 29; H. Bjarneskans, B. Grønnevik, A. Sandberg, *Cykl życiowy memów...*, s. 67.

<sup>36</sup> *Infosfera...*, s. 229–243.

<sup>37</sup> R. Dawkins, *Samolubny gen...*, s. 244; S. Blackmore, *Maszyna memowa...*, s. 79.

<sup>38</sup> W. Borkowski, A. Nowak, *Wpływ społeczny...*, s. 176.

<sup>39</sup> Hierarchia DIKW wzięła nazwę od pierwszych liter słów: dane [*Data*], informacje [*Information*], wiedza [*Knowledge*] i mądrość [*Wisdom*]. Mimo tego, że nie ma zgody co do definicji poszczególnych elementów hierarchii DIKW, wielu jej zwolenników przyjmuje istnienie następujących zależności między nimi:

- nieprawdą jest, że każda dana lub zbiór danych stanowi informację
- nie jest tak, że każda informacja stanowi wiedzę;
- również nieprawdą jest, że każdy zasób wiedzy stanowi mądrość.

<sup>40</sup> J.M. Dunn, *Information in Computer Science*, w: *Philosophy of Information*, red. P. Adriaans, J. van Benthem, Elsevier, Amsterdam, Oxford 2008, s. 582.

kontekście i w związku z jakimś celem"<sup>41</sup>. Natomiast Dariusz Jemielniak pisze: „[...] wiedza to zorganizowany zbiór informacji wraz z regułami ich interpretowania. Przez informacje rozumiemy uporządkowane dane, natomiast dane traktować będziemy jako pozbawione kontekstu ciągi znaków”<sup>42</sup>. Na tej samej stronie autor ujmuje relacje między elementami DIKW za pomocą następującego schematu: znaki + składnia → dane + kontekst → informacje + struktura → wiedza. Wszystkie te ujęcia podkreślają fakt, iż nie każda dana, a nawet zbiór danych stanowi informację. Z tego względu memy – jako jednostki informacji – mogą nie obejmować przynajmniej niektórych zbiorów danych wchodzących w skład *big data*. Dodatkowo trzeba podkreślić, że informacja memetyczna ma charakter w większości analogowy i semantyczny. W opozycji do tego dane wykorzystywane w analizie *big data* mają charakter cyfrowy i są przetwarzane na poziomie syntaktycznym<sup>43</sup>.

Następne różnice są równie uderzające. W wypadku informacji memetycznej mamy do czynienia z przekazem od mózgu (*resp.* umysłu) do mózgu (*resp.* umysłu), najczęściej przy wykorzystaniu wtórnych (nieneurologicznych) nośników memów, z kolei transmisja danych odbywa się od maszyny do maszyny bez pośrednictwa ludzkiego mózgu. Ze względu na swój ogrom dane wykorzystywane w analizie *big data* nie trafiają do mózgu użytkownika, choć zdarza się, że same wyniki analizy już tak. Przyjrzyjmy się dwóm przykładom wykorzystania *big data*. W pierwszym klient banku składa elektronicznie wniosek o przyznanie kredytu. Komputerowy algorytm, posiłkując się danymi z wielu źródeł, szacuje ryzyko kredytowe, decyduje o przyznaniu kredytu i jego kosztach. Na żadnym etapie przetwarzania dane nie trafiają do ludzkiego mózgu. Ani twórca algorytmu, ani jego użytkownik nie zapoznają się z przetwarzanymi danymi, lecz ewentualnie z wynikami analizy. W kolejnym przypadku nawet wyniki nie trafiają do ludzkiego mózgu. Wyobraźmy sobie inteligentny system ogrzewania domu, który na podstawie analizy zachowań lokatorów, przeprowadzanych pomiarów (temperatury w poszczególnych pomieszczeniach i parametrów fizjologicznych mieszkańców), czy też ich aktywności w Internecie, odpowiednio dostosowuje temperaturę. Wyniki analiz w postaci zmian temperatury najczęściej nie są rejestrowane przez lokatorów, nie mówiąc już o tym, że dane wykorzystywane przez system nie trafiają do niczyjego mózgu.

Zastąpienie przekazu informacji od mózgu do mózgu przekazem od maszyny do maszyny skutkuje zmianą samego sposobu transmisji. Memy przekazywane są na drodze niegenetycznej, co konkretyzuje się, wskazując na uczenie się (szerzej) lub naśladownictwo (węziej) jako

---

<sup>41</sup> P.F. Drucker, *Nadchodzi nowa organizacja*, w: *Zarządzanie wiedzą*, red. P.F. Drucker, przeł. M. Witkowska, Helion, Gliwice 2006, s. 11.

<sup>42</sup> D. Jemielniak, *Rozdział 1*, w: *Zarządzanie wiedzą*, red. D. Jemielniak, A.K. Koźmiński, wyd. 2, Oficyna A Wolters Kluwer Business, Warszawa 2012, s. 24.

<sup>43</sup> Należy tu doprecyzować, co rozumie się pod określeniami syntaktyczny i semantyczny. Wielu informatyków posługuje się słabszym pojęciem semantyczności, stąd nie jest dla nich problemem uznanie, że przynajmniej niektóre komputerowe analizy danych mają charakter semantyczny. W niniejszym tekście posługujemy się mocniejszym, filozoficznym rozumieniem semantyczności, które wiąże się z intencjonalnością, czyli z własnością umysłu, dzięki której skierowany jest on na przedmioty i stany rzeczy niezależnie istniejącego świata, oraz dzięki której stany umysłu dotyczą tychże przedmiotów i stanów rzeczy. Por. J.R. Searle, *Umysł. Krótkie wprowadzenie*, przeł. J. Karłowski, Rebis, Poznań 2010, s. 176. Zdaniem Searle'a, istnieją silne racje za tym, że komputer nie dysponuje intencjonalnością i działa wyłącznie na poziomie syntaktycznym, a nie semantycznym jak umysł ludzki. Por. J.R. Searle, *Umysł, mózg i nauka*, przeł. J. Bobryk, PWN, Warszawa 1995. Komputery operując na symbolach i nie odnosząc ich do niczego poza nimi, same nie są intencjonalne, gdyż intencjonalność zapewnia z jednej strony programista, a z drugiej użytkownik programu.



sposób przekazu. W wypadku transmisji danych między maszynami trudno mówić, że maszyny naśladują inne maszyny, czy też że jedne się od drugich uczą<sup>44</sup>. O przekazie danych daje się powiedzieć tylko tyle, że ma on charakter niegenetyczny, jednak jest to zbyt szeroka kategoria, by doszukiwać się istotnych podobieństw w sposobie transmisji memów i danych.

Kolejna różnica dotyczy zasobów środowiska. Pierwotnym środowiskiem memów są mózgi (*resp.* umysły), a ponieważ zasoby tego środowiska (struktury neuronalne) są ograniczone, memy zmuszone są do konkurencji o te zasoby. Inaczej sprawa ma się z danymi – tendencja do gromadzenia wszystkich danych „na zapas”, a także zmniejszenie kosztów ich pozyskiwania i przechowywania sprawiają, że konkurencja jest wyraźnie ograniczona. Jak już wspomniałem, wartość danych w kontekście *big data* nie ogranicza się wyłącznie do planowanych zastosowań, lecz jest funkcją wszystkich przeszłych, obecnych i przyszłych użyć. Ponieważ nie możemy przewidzieć wszystkich przyszłych zastosowań danych, nie jesteśmy w stanie oszacować ich wartości. To powoduje, że wszystkie dane są potencjalnie wartościowe i warte gromadzenia. Dane nie muszą zatem rywalizować o miejsce w magazynach, gdyż archiwizuje się je bez wyjątku z myślą o potencjalnych zastosowaniach.

Rywalizacja memów opiera się na różnicach w ich „atrakcyjności” dla nosicieli, zależnych od powodowanych przez nie efektów fenotypowych czy socjotypowych. Memy i mempleksy skłaniają swoich nosicieli do określonych zachowań, które mogą być mniej lub bardziej korzystne w danym środowisku, w odniesieniu do pewnych sytuacji. Posiadane przez nie przekonania i postawy, mające w sobie komponent memetyczny, bezpośrednio motywują mnie do podjęcia pewnych działań, do zaniechania innych, czy wreszcie do działania w określony sposób. W przypadku danych ich wpływ jest o wiele bardziej pośredni na człowieka. Ilustracją tego niech będzie komputerowy system sterowania pojazdem. System ten zbiera dane z różnych czujników i na tej podstawie wyznacza kierunek i prędkość jazdy. Dane wpływają więc bezpośrednio na zachowanie pojazdu, a pośrednio na pasażerów i innych użytkowników drogi. Ważną różnicę daje się też dostrzec w odniesieniu do kwestii jednoznaczności skutków informacji. Wspomnieliśmy, że relacja między memami a powodowanymi przez nie zachowaniami jest wielo-wieloznaczna. Z kolei w odniesieniu do *big data* możemy uznać, że skutki danych – tzn. wyniki ich analizy – są jednoznacznie wyznaczone przez zastosowane algorytmy<sup>45</sup>.

Wszystkie wymienione różnice sprawiają, że „standardowy” sposób memetycznego opisu kultury słabo przystaje do zjawiska *big data*. Podsumowaniem owych różnic jest poniższa tabela.

---

<sup>44</sup>Co prawda, w informatyce mówi się o uczeniu sieci neuronowych. Sieci neuronowe to programowe lub sprzętowe symulatory modeli matematycznych, opisujących działanie sieci neuronowych, składające się z wielu wzajemnie połączonych (połączeniom przypisane są wagi) sztucznych neuronów, ułożonych w warstwy (wejściową, obliczeniową, wyjściową). Jedną z wielu istotnych cech sieci jest zdolność do uczenia, przy czym uczenie się sieci to „wymuszenie na niej określonego reagowania na zadane sygnały wejściowe” J. Żurada, M. Barski, W. Jędruch, *Sztuczne sieci neuronowe*, PWN, Warszawa 1996, s. 57. Wyróżnia się uczenie z nadzorem (z nauczycielem), podczas którego nauczyciel prezentuje właściwą odpowiedź na dane wejściowe, dzięki czemu sieć koryguje wagi połączeń synaptycznych, oraz uczenie bez nadzoru (bez nauczyciela), gdzie sieć nie otrzymuje właściwych odpowiedzi, lecz musi sama do nich dojść. Por. K. Ficoń, *Sztuczna inteligencja nie tylko dla humanistów*, BEL Studio, Warszawa 2013, s. 134–172; R.A. Kosiński, *Sztuczne sieci neuronowe*, WNT, Warszawa 2014. Tak rozumiane uczenie znacznie różni się od sposobu, w jaki przekazywane są memy.

<sup>45</sup>Rezultaty działania programów komputerowych są wyznaczone przez zastosowane algorytmy i dane wejściowe. O niektórych algorytmach – np. genetycznych – mówimy, że są randomizowane, ponieważ dobierają dane wejściowe w sposób losowy (w wypadku algorytmów genetycznych najczęściej losowo tworzy się populację początkową). K. Ficoń, *Sztuczna inteligencja...*, s. 176–226.

Kryterium porównania	Memy w ujęciu memetyki	<i>Big data</i>
<b>Elementy łańcuch przekazu</b>	Przekaz od mózgu do mózgu	Przekaz od urzędnika do urzędnika
<b>Charakterystyka informacyjna</b>	Informacja głównie analogowa, aspekt semantyczny	Dane w postaci cyfrowej, aspekt syntaktyczny
<b>Rodzaj przekazu</b>	Przekaz na drodze niegenetycznej /uczenia się naśladownictwa	Przekaz nie związany z uczeniem się (raczej) czy naśladownictwem
<b>Środowisko</b>	Ograniczone zasoby środowiska (rywalizacja o reprezentację w strukturach neuronalnych mózgu)	Prawie nieograniczone (w pewnym zakresie) możliwości przechowywania danych
<b>Efekt behawioralny</b>	Wpływ na ludzkie zachowania	Ewentualny, pośredni wpływ na ludzkie zachowania
<b>Jednoznaczność efektu behawioralnego</b>	Niejednoznaczne skutki informacji	Skutki informacji wyznaczone przez algorytmy
<b>Aspekt kulturowy</b>	Memy jako nośniki przekazu kulturowego	Część danych nie kwalifikuje się jako elementy przekazu kulturowego

Tabela 1: Porównanie memów i *big data* (opracowanie własne)

Zostaje nam analiza konsekwencji nieadekwatności narzędzi memetycznych do opisu i wyjaśnienia fenomenu *big data*.

## Podsumowanie

Skontrastowanie memetycznego obrazu kultury z fenomenem *big data* pokazało, jak wiele różnic dzieli memy od przechowywanych i przetwarzanych danych. Stawia to pytania o zupełność i adekwatność memetyki jako ogólnej teorii kultury, a także o miejsce *big data* w kulturze. Bez wątpienia, bardzo wiele przechowywanych na cyfrowych nośnikach danych, a nawet ich zbiorów nie kwalifikuje się do nazwania ich memami<sup>46</sup>. Co prowadzi do konkluzji, że memetyka jest niekompletna lub że dane te nie wchodzą w skład kultury, a przez to nie stanowią przedmiotu zainteresowania memetyki. W przypadku pierwszej odpowiedzi możemy zgodzić się na niezupełność memetyki jako ogólnej teorii kultury, bądź starać się poprawić memetykę. Taka poprawa może przybrać postać rozszerzenia pojęcia memu lub uzupełnienia memetyki o nowy replikator. To ostatnie czyni Blackmore, wprowadzając trzeci replikator – technologiczny mem, czyli tem<sup>47</sup> – który mógłby spróbować objąć swym zasięgiem *big data*. Jest to, moim zdaniem,

<sup>46</sup>W sprawie definicji memu: A. Stępnik, *Memetyczny zwrot w badaniach nad kulturą. O definicji memu w kontekście roli memetyki jako ogólnej teorii kultury* (w druku).

<sup>47</sup>Blackmore najpierw używała słowa *teme* (liczba mnoga: *temes*), by zastąpić je łatwiej wymawialnym *treme*

o wiele lepsze rozwiązanie niż rozszerzenie zakresu słowa „mem”, które i bez tego typu operacji budzi spore wątpliwości.

Bardziej właściwe wydaje mi się wyraźne rozgraniczenie przedmiotu memetyki, jakim jest kultura, od przynajmniej niektórych zbiorów danych, mając w pamięci, iż nie każdy zbiór danych stanowi informację, a do tego nie każda informacja ma charakter kulturowy. Przy takim ujęciu fakt, że niektóre zbiory danych nie dają się wyrazić w kategoriach memetycznych nie musi przekreślać pretensji memetyki do ujmowania całości kultury. Choć nie uważam tych pretensji za możliwe do realizacji. Ale zupełnie z innych względów niż z powodu *big data*.

## Bibliografia

- Anderson Chris, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, "Wired" 2008.
- Biedrzycki Mariusz, *Genetyka kultury*, Prószyński i S-ka, Warszawa 1998.
- Bjarneskans Henrik, Grønnevik Bjarne, Sandberg Anders, *Cykl życiowy memów*, w: *Infosfera. Memetyczne koncepcje kultury i komunikacji*, red. D. Wężowicz-Ziołkowska, WSZOP, Katowice 2009, s. 65–89.
- Blackmore Susan, *Maszyna memowa*, przeł. Norbert Radomski, Dom Wydawniczy Rebis, Poznań 2002.
- Borkowski Wojciech, Nowak Andrzej, *Wpływ społeczny – alternatywny model rozprzestrzeniania się memów*, w: *Infosfera. Memetyczne koncepcje kultury i komunikacji*, red. D. Wężowicz-Ziołkowska, WSZOP, Katowice 2009, s. 173–212.
- Brodie Richard, *Wirus umysłu*, przeł. Przemysław Turski, TeTa Publishing, Łódź 1997.
- Dawkins Richard, *Fenotyp rozszerzony*, przeł. Joanna Gliwicz, Prószyński i S-ka, Warszawa 2003.
- Dawkins Richard, *Samolubny gen*, przeł. Michał Skoneczny, Prószyński i S-ka, Warszawa 2007.
- Drucker Phill F, *Nadchodzi nowa organizacja*, w: *Zarządzanie wiedzą*, red. P. F. Drucker, przeł. M. Witkowska, Helion, Gliwice 2006, s. 7–27.
- Dunn John M., *Information in Computer Science*, w: *Philosophy of Information*, red. P. Adriaans, J. van Benthem, Elsevier, Amsterdam, Oxford 2008, s. 581–608.
- Dygaszewicz Jan, *Big data w statystyce publicznej*, w: *Internet. Publiczne bazy danych i big data*, red. G. Szpor, Wydawnictwo C.H. Beck, Warszawa 2014, s. 49–62.
- Ficoń Krzysztof, *Sztuczna inteligencja nie tylko dla humanistów*, BEL Studio, Warszawa 2013.
- Heylighen Francis, *Memetyka*, w: *Infosfera. Memetyczne koncepcje kultury i komunikacji*, red. D. Wężowicz-Ziołkowska, WSZOP, Katowice 2009, s. 27–31.
- Hilbert Martin, Lopez Priscila, *The World's Technological Capacity to Store, Communicate, and Compute Information*, "Science" 2011, nr 4, s. 60–65.
- Hilbert Martin, Lopez Priscila, *How to Measure the World's Technological Capacity to Communicate, Store and Compute Information*, "International Journal of Communication" 2012, s. 1042–1045.
- Infosfera. Memetyczne koncepcje kultury i komunikacji*, red. D. Wężowicz-Ziołkowska, WSZOP w Katowicach, Katowice 2009.
- Jemieliński Dariusz, *Rozdział 1*, w: *Zarządzanie wiedzą*, red. D. Jemieliński, A.K. Koźmiński, wyd. 2, Oficyna A Wolters Kluwer Business, Warszawa 2012, s. 21–42.
- Kosiński Robert A., *Sztuczne sieci neuronowe*, WNT, Warszawa 2014.
- Mayer-Schönberger Viktor, Cukier Kenneth, *Big data. Rewolucja, która zmieni nasze myślenie, pracę i życie*, przeł. M. Glatki, Wydawnictwo MT Biznes, Warszawa 2014.

- Laney Doug, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, META Group 2001  
[<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (stan na 30 kwietnia 2015 r.)]
- Searle John R., *Umysł, mózg i nauka*, przeł. J. Bobryk, PWN, Warszawa 1995.
- Searle John R., *Umysł. Krótkie wprowadzenie*, przeł. J. Karłowski, Rebis, Poznań 2010.
- Stępnik Andrzej, *Memetyczny zwrot w badaniach nad kulturą. O definicji memu w kontekście roli memetyki jako ogólnej teorii kultury* (w druku).
- Wilson Edward O., *Konsiliencja*, przeł. J. Mikos, Zysk i S-ka, Poznań 2002.
- Żurada Jacek, Barski Mariusz, Jędruch Wojciech, *Sztuczne sieci neuronowe*, PWN, Warszawa 1996.

### Nota o autorze

Andrzej Stępnik – urodzony w 1979 r. w Warszawie, doktor filozofii, absolwent Instytutu Filozofii Uniwersytetu Warszawskiego. Autor ponad 50 publikacji naukowych, w tym książki *Pragmatyzm Williama Jamesa: ujęcie systemowo-krytyczne*. Zajmuje się epistemologią, filozofią umysłu, filozofią nauki i metodologią ogólną, filozofią religii i aksjologią, interesują go zagadnienia z pogranicza filozofii i nauk szczegółowych.

### About the author

Andrzej Stępnik – born in 1979 in Warsaw, Ph.D. in philosophy, University of Warsaw's Institute of Philosophy graduate. Author of over 50 publications, including a book *Pragmatyzm Williama Jamesa: ujęcie systemowo-krytyczne*. Interested in epistemology, philosophy of mind, philosophy of science, general methodology, philosophy of religion, axiology. Also: topics on the borders between philosophy and specific science.