

# 3.5

## Część 3

Nabywcy w działaniach marketingowych organizacji

Elżbieta Wąsowicz-Zaborek\*

# Analiza treści generowanych przez użytkowników Internetu jako podstawa decyzji marketingowych w małym przedsiębiorstwie

## Streszczenie

Wobec rosnącego znaczenia Internetu nie sposób przecenić zawartych w nim danych. Szczególne wyzwanie stanowi sposób ich identyfikacji, gromadzenia oraz analizy. Rozdział ma charakter koncepcyjny. Celem opracowania jest ukazanie źródeł treści generowanych przez użytkowników Internetu istotnych z punktu widzenia podejmowania decyzji marketingowych przedsiębiorstwa oraz sposobów ich pozyskania i interpretacji. Przedstawiono narzędzia, metody i procedury badawcze, w tym z wykorzystaniem dedykowanego oprogramowania komputerowego (roboty indeksujące oraz programy komputerowego wspomaganie analizy danych jakościowych CAQDAs). Rozważania zilustrowane zostały przykładem małego przedsiębiorstwa z sektora turystycznego.

**Słowa kluczowe:** UGC, *User Generated Content*, treści generowane przez użytkowników, decyzje marketingowe

---

\* Szkoła Główna Handlowa w Warszawie, ORCID: 0000-0001-8422-4383

## Analysis of User Generated Content as the Basis for Marketing Decisions in a Small Enterprise

### Abstract

Considering the growing importance of the Internet, the utility of rich data available online cannot be overestimated. A particular challenge is how to identify, collect and analyze the data. This article is a conceptual work with the aim to show how to locate and explore on-line sources of user generated content relevant for making marketing decisions by firms. Research tools, methods and procedures, including the use of dedicated computer software (indexing robots and CAQDAs software supporting qualitative data analysis) are presented in the text. The discussed concepts are illustrated with an example of a small family business from the tourism sector.

**Keywords:** UCG, User Generated Content, marketing decisions

---

### Wprowadzenie

Internet ma ogromny potencjał do tworzenia możliwości biznesowych. Jedną z pierwszych strategicznych ram dla praktyki biznesowej w Internecie wskazał w modelu ICDT<sup>1</sup> A. Anghern, dzieląc możliwości stworzone przez sieć na cztery różne przestrzenie wirtualne (informacyjną, komunikacyjną, dystrybucyjną, transakcyjną)<sup>2</sup>. Autor tej koncepcji dostrzegał w ramach wirtualnej przestrzeni komunikacyjnej potencjał do zaangażowania odbiorców w dialog, ale chyba nie był w stanie przewidzieć prawdziwej rewolucji związanej z przejściem do drugiej fazy rozwoju Internetu, określanej jako Web 2.0, która rozpoczęła się na początku XXI wieku. Ewolucja technologii wpłynęła na sposób, w jaki ludzie się komunikują, podejmują decyzje, socjalizują, uczą się, bawią, współdziałają ze sobą, a nawet robią zakupy. Web 2.0, obok zmiany indywidualnych i grupowych zachowań ludzi, wpłynął również na struktury rynku, powodując znaczne przesunięcie dominującej pozycji rynkowej od producentów czy sprzedawców do klientów. Głównym tego powodem jest fakt, że dzisiejszy konsument internetowy ma wgląd w niespotykany wcześniej zasób informacji i wiedzy, jak również nieograniczony wybór, dostępny po kliknięciu myszką komputerową lub z poziomu mobilnej przeglądarki urządzenia przenośnego. W erze Web 2.0 tradycyjna koncepcja jednostronnej komunikacji zastę-

---

<sup>1</sup> ICDT model – nazwa to akronim angielskich słów *information, communication, distribution, transaction*.

<sup>2</sup> A. Anghern, *Designing Mature Internet Business Strategies: The ICDT Mode*, "European Management Journal" 1997, vol. 15, no. 4, s. 361–369.

powana jest indywidualizacją przekazu, ale przede wszystkim większy nacisk kładziony jest na zaangażowanie adresata. Aplikacje Web 2.0 są w dużej mierze oparte na treściach generowanych przez użytkowników. Jest to podstawowa różnica w stosunku do poprzednich aplikacji internetowych: użytkownik jako istotny uczestnik jest nowym partnerem marketingowym. Jego głos jest coraz bardziej słyszalny i stanowi istotne źródło informacji dla podejmowania decyzji zakupowych przez innych adresatów oferty.

Wobec tego Web 2.0 oferuje przedsiębiorstwom nowe wyzwania, ale także nowe możliwości w zakresie nawiązywania i utrzymywania kontaktów z uczestnikami rynku, bezpośredniego kontaktu z klientami, ale także poznawania ich potrzeb i opinii. Rozwój technologii tworzenia treści stwarza okazję do zbierania informacji od każdego, kto ma dostęp do Internetu. Informacje generowane przez użytkowników (*User Generated Content* – UGC) w coraz większym stopniu wspierają podejmowanie decyzji i analizę w odniesieniu do różnych rodzajów działalności biznesowej, zarządczej czy marketingowej.

Ilość danych generowanych przez użytkowników sieci nie jest obecnie liczona już ani w terabajtach ani nawet petabajtach, ale exabajtach<sup>3</sup> i ciągle rośnie. Ogromne zbiory danych (*big data*) możliwe do pozyskania, nie tylko pochodzące z UGC, mogą stanowić ważne źródło informacji, ale wymagają odpowiedniego przygotowania i świadomego zarządzania. Raporty opublikowane przez Eurostat wskazują, że w 2018 r. ciągle niewiele przedsiębiorstw wykorzystywało *big data*. Jedynie 12% przedsiębiorstw zatrudniających powyżej 10 osób z 28 krajów Unii Europejskiej korzystało z *big data*. W Polsce wskaźnik ten to zaledwie 8%<sup>4</sup>.

Media społecznościowe (SM), takie jak Facebook, Twitter, LinkedIn, YouTube i Instagram, są szerokimi i istotnymi źródłami danych społecznych na temat życia codziennego<sup>5</sup>. Zachowania offline przenoszone są obecnie w dużej mierze do sieci, a przez to stają się potencjalnie dostępne do analiz.

Zgodnie z danymi Eurostat *big data* z mediów społecznościowych analizowało w 2018 r. 6% przedsiębiorstw z 28 krajów w UE, co stanowi 45% tych, które w ogóle analizują *big data* z różnych źródeł. W Polsce wskaźnik ten jest jeszcze niższy. Jedynie 2% polskich firm, czyli 31% tych, które w ogóle analizują *big data*, śledzi dane z mediów społecznościowych<sup>6</sup>.

Konfrontacja świadomości potencjału, jaki drzemie w informacjach generowanych przez użytkowników (UGC), z ciągle niskim poziomem ich wykorzystania przez przed-

<sup>3</sup> Petabajt = 1 000 000 000 000 000 (biliard) bajtów; exabajt = 1 000 000 000 000 000 000 (trylion) bajtów.

<sup>4</sup> [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc\\_eb\\_bd&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_eb_bd&lang=en) (dostęp: 19.12.2019).

<sup>5</sup> A. Halavais, *Bigger Sociological Imaginations: Framing Big Social Data Theory and Methods*, "Information, Communication & Society" 2015, vol. 18, no. 5, s. 583–594; G., Bello-Orgaz, J.J., Jung, D. Camacho, *Social Big Data: Recent Achievements and New Challenges*, "Information Fusion" 2016, vol. 28, s. 45–59.

<sup>6</sup> [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc\\_eb\\_bd&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_eb_bd&lang=en) (dostęp: 19.12.2019)

siębiorców wskazuje na konieczność prowadzenia zarówno działalności badawczej, jak i edukacyjnej w zakresie pozyskiwania, przetwarzania, raportowania i wykorzystania danych generowanych przez użytkowników sieci.

Celem opracowania jest ukazanie źródeł UGC istotnych z punktu widzenia decyzji przedsiębiorstw oraz sposobów ich pozyskania i interpretacji. Dokonano także analizy możliwości ich wykorzystania w praktyce biznesu, a w szczególności w podejmowaniu decyzji marketingowych. Na początku zdefiniowano pojęcie UGC oraz dokonano przeglądu istniejących badań w tym zakresie, a następnie omówione zostały poszczególne etapy procesu badawczego, który mogą przeprowadzić przedsiębiorcy w celu podjęcia decyzji zarządczych. Omówiono kolejno potencjalne źródła danych, sposoby ich pozyskiwania oraz analizy. Dla lepszego zobrazowania tematu posłużono się przykładem przedsiębiorstwa turystycznego – firmy oferującej usługi w zakresie turystyki aktywnej wraz z zapleczem noclegowym. Studium przypadku pozwoliło na pełniejszą, a także krytyczną analizę opisywanych zagadnień. Opracowanie może stanowić zarówno wkład w dyskusję nad metodami badania zjawisk rynkowych, jak i zestaw wskazań dla praktyków biznesu, którzy na co dzień potrzebują dostępu do wiedzy w podejmowaniu decyzji.

## Treści generowane przez użytkowników Internetu (UGC) – definicja i przegląd literatury

Użytkownik jest istotnym elementem koncepcji Web 2.0, nie tylko jako odbiorca i adresat, ale przede wszystkim jako twórca treści<sup>7</sup>. Postęp technologiczny w dziedzinie komunikacji implementowany do wirtualnego świata jest tak szybki, że jednoznaczne zdefiniowanie treści generowanych przez użytkowników jest w praktyce dość trudne.

Zgodnie z definicją podaną przez OECD<sup>8</sup> „UGC to:

- 1) treści udostępniane publicznie w Internecie,
- 2) odzwierciedlające pewną dozę twórczego wysiłku oraz
- 3) tworzone poza profesjonalną rutyną i praktyką”<sup>9</sup>.

Treści generowane przez użytkowników sieci mogą dotyczyć praktycznie wszystkich sfer aktywności ludzi. Stanowią odzwierciedlenie ich postaw i opinii. Część treści generowana jest w sposób spontaniczny, jak np. opinie współdzielone na forach dyskusyjnych czy w mediach społecznościowych. Pozostałe mogą być tworzone na skutek

<sup>7</sup> T. O'Reilly, *What Is Web 2.0: Design Patterns and Business Models for the Next Generation Software*, "Communication and Strategies" 2007, vol. 65, no. 1, s. 17–37.

<sup>8</sup> Organization for Economic Co-operation and Development – Organizacja Współpracy Gospodarczej i Rozwoju.

<sup>9</sup> OECD, *Participative Web and User-Created Content. Web 2.0, WIKIS and Social Networking 2007*.

świadomych, często wymuszonych działań przedsiębiorstw i innych instytucji mających interes w angażowaniu internautów. Przykładem są tu formularze zapytań umieszczane na stronie, ankiety elektroniczne, konkursy, w których uczestnicy zachęceni są do dzielenia się tekstami lub materiałami multimedialnymi w zamian za obietnicę nagrody. Co ważne, generowane w sieci treści przyjmują różne formy, daleko wykraczające poza ramy zwykłego tekstu. Tabela 3.5.1 prezentuje zestawienie przykładowych form UGC.

**Tabela 3.5.1.**

Typy treści generowanych przez użytkowników Internetu

Typ UGC	Charakterystyka
Teksty typu <i>fan fiction</i>	Mogą przybierać formy krótkich wypowiedzi lub bardziej rozbudowanych utworów literackich pisanych prozą lub wierszem. Wypowiedzi mogą być publikowane na własnych stronach internetowych lub specjalnych dedykowanych serwisach, np. skupiających fanów danej dziedziny, którzy publikują artykuły na dany temat. Przykłady: Granice.pl, Wattpad.com, gdzie publikowane są utwory literackie
Zdjęcia	Obrazy mogą być publikowane jako samodzielne treści lub jako forma uatrakcyjnienia i wizualizacji tekstu. Publikowane obecnie w sieci zdjęcia w większości wykonane są przy użyciu smartfonów. Pozwala to na prezentację nawet w czasie rzeczywistym i znacznie podnosi aktualność informacji. Publikacja możliwa jest zarówno na własnych stronach www użytkowników, jak i w mediach społecznościowych lub w specjalnie dedykowanych serwisach. Przykładowe SM do dzielenia się zdjęciami to Instagram i Pinterest, a serwisy to Flickr czy Fotosik.pl
Filmy	Publikowane filmy mogą obejmować różne tematy i stanowić samodzielne utwory (amatorskie filmy animowane, krótko i długometrażowe), scenki komiczne, recenzje produktów, filmowe poradniki czy też relacje z podróży i wydarzeń. Prezentowane treści mogą być przygotowane z wyprzedzeniem lub też prezentowane w czasie rzeczywistym (relacja na żywo). Najpopularniejszym serwisem do dzielenia się treściami wideo jest Youtube.com
Prezentacje i animowane gify	Prezentacje multimedialne są bardziej statyczne od filmów i raczej opierają się na samych obrazach, rzadko na dźwięku. Popularnymi serwisami do współdzielenia prezentacji są np. Slideshare.com lub Prezi.com (który pozwala także na tworzenie prezentacji). Popularność zdobywają także fotostory, stanowiące połączenie kilku zdjęć umieszczanych w serwisach społecznościowych (np. relacja na Facebooku lub Instagramie)
Muzyka i dźwięk	Ten rodzaj treści jest zwykle uzupełnieniem materiałów wizualnych, ale może stanowić także przekaz sam w sobie. Możliwe jest także umieszczanie w sieci własnych utworów, np. w serwisach Bandcamp, Soundcloud, YouTube, Spotify
Wiki	Ten typ serwisu umożliwia użytkownikom współtworzenie treści i dzielenie się wiedzą poprzez dodawanie, zmianę oraz usuwanie treści już istniejących w celu ich aktualizacji. Najpopularniejszą wiki jest Wikipedia
Blogi	Nazwa pochodzi od angielskich słów <i>web</i> i <i>log</i> . Blogi to treści tworzone przez internautów w sposób regularny. Umieszczane są na specjalnych stronach, które umożliwiają także porządkowanie i przeszukiwanie treści wg słów kluczowych czy chronologicznie. Niejednokrotnie możliwe jest także komentowanie wpisów. Blogi mogą mieć różne formy i obejmować różne tematy. Częstym rozróżnieniem są blogi osobiste (autor prezentuje lub komentuje wydarzenia z własnego życia) i specjalistyczne (treści koncentrują się zwykle wokół wybranej tematyki, np. kulinarnej, modowej, podróżniczej). Przykładami popularnych serwisów blogowych są WordPress.org, Blogger, Tumblr

cd. tabeli 3.5.1

Typ UGC	Charakterystyka
Fora z pytaniami i odpowiedziami (Q&A)	Ta forma umożliwia użytkownikom zadawanie pytań i uzyskiwanie odpowiedzi. Sposoby zadawania pytań mogą przybierać różne postaci – fora, quizy, ankiety, sondy z odpowiedziami typu „tak” lub „nie”. Popularne serwisy tego rodzaju to: Yahoo! Answer, Ask.com, Quora, Zapytaj.onet.pl
Fora i grupy dyskusyjne	Mogą funkcjonować jako samodzielne serwisy lub jako część strony dostawcy. Grupy dyskusyjne zrzeszają zwykle użytkowników zainteresowanych danym zagadnieniem, np. miłośników fotografii na Fotopolis
Wiadomości społeczne	Serwisy społeczne, które wykorzystują potencjał rozproszonej sieci użytkowników do pozyskiwania najświeższych informacji o bieżących wydarzeniach, np. Kontakt24, TwojeInfo
Agregatory linków	Internauci dzielą się treściami już istniejącymi w sieci poprzez tworzenie listy ich zdaniem atrakcyjnych linków, co stanowi wartość dodaną, np. Reddit, Wykop.pl
Media społecznościowe (social media – SM)	Media społecznościowe to zbiorowe określenie opisujące serwisy, w których użytkownicy mogą tworzyć wirtualne społeczności. Różnią się one między sobą formą prezentacji treści, dostępnością, zakresem publikowanych informacji. Mnogość SM jest odpowiedzią na rosnącą liczbę internautów, którzy przenoszą do wirtualnego świata coraz to nowe dziedziny życia. Najpopularniejsze SM to Facebook, Instagram, Twitter

Źródło: opracowanie własne.

Podobnie jak wiele jest typów UGC, tak wiele jest sposobów ich analizowania. Niektóre są dość proste, inne bardziej skomplikowane i wymagają niekiedy specjalistycznej wiedzy lub narzędzi (np. analiza obrazu i filmu).

UGC stanowią przedmiot ogromnego zainteresowania naukowców. Hasło *User Generated Content* daje ponad 5,3 mln wyszukiwań w serwisie Google Scholar (dostęp: 20.12.2019). Autorzy skupiają się na różnych aspektach tematu. Badaniem motywacji tworzenia UGC przez internautów zajmowali się wraz ze współpracownikami m.in. T. Daugherty<sup>10</sup>, R. Stoeckl<sup>11</sup>, U. Schaedel<sup>12</sup>, Y. Sun<sup>13</sup>. Szeroko omówili także temat ogólnego wpływu i wykorzystania UGC w kreacji i ocenie pozycji marek np. S. Tirunillai i G.J. Tellis<sup>14</sup>, H. Shad ze współpracownikami<sup>15</sup>, A.J. Kim i K.K.P. Johnson<sup>16</sup>, X. Liu ze

<sup>10</sup> T. Daugherty, M.S. Eastin, L. Bright, *Exploring Consumer Motivations for Creating User-Generated Content*, „Journal of Interactive Advertising” 2008, vol. 8, no. 2, s. 16–25.

<sup>11</sup> R. Stoeckl, P. Rohrmeier, T. Hess, *Motivations to Produce User Generated Content: Differences Between Webloggers And Videobloggers*, BLED 2007, Proceedings 2007.

<sup>12</sup> U. Schaedel, M. Clement, *Managing the Online Crowd: Motivations for Engagement in User-Generated Content*, „Journal of Media Business Studies” 2010, vol. 7, no. 3, s. 16–36.

<sup>13</sup> Y. Sun, X. Dong, S. McIntyre, *Motivation of User-Generated Content: Social Connectedness Moderates the Effects of Monetary Rewards*, „Marketing Science” 2017, vol. 36, no. 3, s. 329–337.

<sup>14</sup> S. Tirunillai, G.J. Tellis, *Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation*, „Journal of Marketing Research” 2014, vol. 51, s. 463–479.

<sup>15</sup> H. Shad, S. Jamali, A. Aleahmad, *Online Reputation Measurement of Companies Based on User-generated Content in Online Social Networks*, „Computers in Human Behavior” 2016, no. 54, s. 94–100.

<sup>16</sup> A.J. Kim, K.K.P. Johnson, *Power of Consumers Using Social Media: Examining the Influences of Brand-related User-generated Content on Facebook*, „Computers in Human Behavior” 2016, vol. 58, s. 98–108.

współpracownikami<sup>17</sup>, S.A. Jin<sup>18</sup>, L. Leung<sup>19</sup>. W tym także obszarze analizą sentymentu, polegającą na klasyfikacji treści w zależności od wyrażonych nastrojów (pozytywny, negatywny, neutralny), zajmowali się m.in. B. Liu<sup>20</sup>, B. Pang i L. Lee<sup>21</sup>.

Z uwagi na operowanie dużymi zbiorami danych w wielu opracowaniach autorzy posługują się komputerowymi narzędziami wspomagającymi analizy, opartymi na algorytmach uczących się (*machine learning*). Większość metod analizy opiera się na identyfikacji i klasyfikacji słów lub zwrotów. Badacze wykorzystywali różne algorytmy: (1) uczenia się nadzorowanego (V. Pérez-Rosas i in.<sup>22</sup>), (2) uczenia się bez nadzoru (D. Borth i in.<sup>23</sup>), (3) uczenia się częściowo nadzorowanego (N.F. Felix i in.<sup>24</sup>) oraz (4) inne podejścia (C. Clavel i Z. Callejas<sup>25</sup>). Kilku badaczy dokonywało porównań skuteczności różnych metod, np. O. Irsoy i C. Cardie<sup>26</sup>, B. Jou i S.-F. Chang<sup>27</sup>.

Znaczna część badań odnosiła się do blogów, ich autorami byli np. L. Leung<sup>28</sup>, A. Chia<sup>29</sup>, G. Mishne i N.S. Glance<sup>30</sup>, A. Gumkowska i współpracownicy<sup>31</sup>. Stosunkowo niewiele natomiast prac badawczych poświęcono eksploracji opinii w obszarze forów

---

<sup>17</sup> X., Liu, A.C. Burns, Y. Hou, *An Investigation of Brand-Related User-Generated Content on Twitter, An Investigation of Brand-Related User-Generated Content on Twitter*, "Journal of Advertising" 2017, vol. 46, no. 2, s. 1–12.

<sup>18</sup> S.A. Jin, *The Potential of Social Media for Luxury Brand Management*, "Marketing Intelligence & Planning" 2012, vol. 30, no. 7, s. 687–699.

<sup>19</sup> L. Leung, *User-generated Content on the Internet: An Examination of Gratifications, Civic Engagement and Psychological Empowerment*, "New Media and Society" 2009, vol. 11, no. 8, s. 1327–1347.

<sup>20</sup> B. Liu, *Sentiment Analysis and Opinion Mining*, "Synthesis Lectures on Human Language Technologies" 2012, vol. 5, no. 1, s. 1–167.

<sup>21</sup> B. Pang, L. Lee, *Opinion Mining and Sentiment Analysis*, "Foundations and Trends in Information Retrieval" 2008, vol. 2, no. 1–2, s. 1–135.

<sup>22</sup> V. Pérez-Rosas, R. Mihalcea, L.P. Morency, *Multimodal Sentiment Analysis of Spanish Online Videos*, "IEEE Intell. Syst." 2013, vol. 28, no. 3, s. 38–45.

<sup>23</sup> D. Borth, R. Ji, T. Chen, T. Breuel, S.-F. Chang, *Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs*, ACM International Conference on Multimedia, 2013, s. 223–232.

<sup>24</sup> N.F. Felix, L.F. Da Silva, S. Coletta, E.R. Hruschka, *A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning*, "ACM Computing Survey" 2016, vol. 49, no. 1, s. 1–26.

<sup>25</sup> C. Clavel, Z. Callejas, *Sentiment Analysis: From Opinion Mining to Human-Agent Interaction*, "IEEE Trans. Affect. Comput." 2015, s. 74–93.

<sup>26</sup> O. Irsoy, C. Cardie, *Opinion Mining with Deep Recurrent Neural Networks*, "EMNLP" 2014, s. 720–728.

<sup>27</sup> B. Jou, S.-F. Chang, *Deep Cross Residual Learning for Multitask Visual Recognition*, ACM International Conference on Multimedia (MM), 2016.

<sup>28</sup> L. Leung, *op.cit.*

<sup>29</sup> A. Chia, *Welcome to Me-Mart: The Politics of User-Generated Content in Personal Blogs*, "American Behavioral Scientist" 2012, vol. 56, no. 4, s. 421–438.

<sup>30</sup> G., Mishne, N.S. Glance, *Predicting Movie Sales from Blogger Sentiment*, AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, 2006.

<sup>31</sup> A. Gumkowska, M. Maryl, P. Toczyński, *Blog to... blog. Blogi oczyma blogerów*, Raport z badania jakościowego zrealizowanego przez Instytut Badań Literackich PAN i Gazeta., *Tekst w sieci. Tekst. Język. Gatunki*, D. Ulicka (red.), Wydawnictwa Akademickie i Profesjonalne, Warszawa 2009, s. 285–309.

dyskusyjnych (A. Chmiel i in.<sup>32</sup>, M. Jaciow<sup>33</sup>). Wielu autorów przedmiotem badań czy- niło platformę Twitter i treści tam publikowane, np. C. Castillo i in.<sup>34</sup>, V. Lampos i in.<sup>35</sup>, A. Agarwal i in.<sup>36</sup> Prawdopodobną przyczyną jest w tym wypadku powszechna dostęp- ność publikowanych danych, bez ukrywania ich systemem haseł oraz dostępność wbu- dowanych i zewnętrznych narzędzi do monitorowania treści (w tym wielu darmowych). Dodatkowo krótkie formy wypowiedzi typowe dla Twittera są łatwiejsze do analizy.

Pomimo największej popularności Facebooka wśród serwisów społecznościowych stosunkowo niewiele można znaleźć prac badawczych na ten temat. Większość z nich dotyczy ogólnych zagadnień użytkowania tej platformy i fenomenu socjologicznego, jaki ona stanowi (A. Nadkarni i S.G. Hofmann<sup>37</sup>, C.M.K. Cheung i in.<sup>38</sup>). Rzadziej analizo- wane są same dane w niej zawarte (I.P. Cvijikj i F. Michahelles<sup>39</sup>, V.S. Jin i in.<sup>40</sup>) – głównie dlatego, że w przeciwieństwie do Twittera treści publikowane w tym serwisie dostępne są przede wszystkim dla osób zalogowanych, a w przypadku wypowiedzi prywatnych użytkowników dla wąskiego kręgu „znajomych”. W związku z tym publicznych postów do pozyskania jest zdecydowanie mniej. Nie oznacza to jednak, że platforma Facebook powinna pozostać poza kręgiem zainteresowania marketerów. Wskazane jest, aby śle- dzili oni oficjalne profile marek, za które odpowiadają, i aktywizowali uczestników sieci do publikacji użytecznych i dostępnych treści.

Podsumowując, prowadzone dotychczas badania wskazują na znaczną użyteczność prezentowanych przez użytkowników treści dla różnych zastosowań, w tym dla prak-

---

<sup>32</sup> A. Chmiel, P. Sobkowicz, J. Sienkiewicz, G. Paltoglou, K. Buckleyb, M. Thelwall, J.A. Hołyst, *Negative Emotions Boost User Activity at BBC forum*, "Physica A" 2011, vol. 390, no. 16, s. 2936–2944.

<sup>33</sup> M. Jaciow, *Fora internetowe jako źródło informacji–możliwości i ograniczenia wykorzystania w badaniach zachowań nabywczych*, „Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu” 2011, nr 236, s. 113–122.

<sup>34</sup> C. Castillo, M. Mendoza, B. Poblete, *Information Credibility on Twitter*, Proceedings of the 20<sup>th</sup> International Conference on World Wide Web (WWW'11), Association for Computing Machinery, New York 2011, 675–684.

<sup>35</sup> V. Lampos, T.D. Bie, N. Cristianini, *Flu Detector – Tracking Epidemics on Twitter*, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2010), Springer, Barcelona, s. 599–602.

<sup>36</sup> A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, *Sentiment Analysis of Twitter Data*, Proceedings of the Workshop on Language in Social Media (LSM 2011), Association for Computational Linguistics, Portland, Oregon 2011, Oregon, s. 30–38.

<sup>37</sup> A. Nadkarni, S.G. Hofmann, *Why Do People Use Facebook?*, „Personality and Individual Differences” 2012, vol. 52, no. 3, s. 243–249.

<sup>38</sup> C.M.K. Cheung, P. Chiu, M.K.O. Lee, *Computers in Human Behavior Online social networks: Why Do Students Use Facebook?*, „Computers in Human Behavior” 2011, vol. 27, no. 4, s. 1337–1343.

<sup>39</sup> I.P. Cvijikj, F. Michahelles, *Monitoring Trends on Facebook*, IEEE Ninth International Coonference on Deependable, Autoutonomic and Secure Computing 2011, s. 896–903.

<sup>40</sup> V.S. Jin, J. Phua, K.M. Lee, *Computers in Human Behavior Telling Stories about Breastfeeding through Facebook: The Impact of User-generated Content (UGC) on Pro-breastfeeding Attitudes*, „Computers in Human Behavior” 2015, vol. 46, s. 6–17.



tyki biznesu. Postęp technologiczny i pojawiające się nowe sposoby prezentacji, ale także pozyskiwania i analizy danych, pozwalają przewidywać dalszy wzrost znaczenia tego zasobu w przyszłości.

## Źródła danych w UGC dla podejmowania decyzji marketingowych w przedsiębiorstwie

Marketing jest koncepcją zarządczą prowadzącą do realizacji celów przedsiębiorstw poprzez maksymalizację zadowolenia klienta. W ramach systemu zarządzania marketingowego firmy podejmują wiele decyzji strategicznych i operacyjnych. Podstawą i punktem wyjścia są decyzje związane z wyborem rynku lub rynków docelowych, na których oferowane będą produkty lub usługi oraz z planowaniem strategii działania na tych rynkach, a także doбором narzędzi marketingowych<sup>41</sup>.

W celu minimalizacji ryzyka związanego z podjęciem potencjalnie błędnych decyzji konieczne jest pozyskanie i odpowiednie przetworzenie niezbędnych danych. Mogą one pochodzić ze źródeł pierwotnych, np. z badania rynku, sondaży, ankiet satysfakcji itp., obejmujących zarówno klientów, jak i partnerów rynkowych. Proces ich pozyskania jest często kosztochłonny i najczęściej wymusza współpracę ze specjalistycznymi firmami badawczymi, a informacje w nich zawarte szybko ulegają dezaktualizacji na skutek dynamicznych zmian rynkowych.

Uzupełnieniem i/lub alternatywą są dane pochodzące ze źródeł wtórnych. Są to zasoby danych już istniejących – czy to wewnątrz firmy, np. raporty kasowe, sprawozdania przedstawicieli i partnerów handlowych, czy na zewnątrz, takie jak publikowane raporty rynkowe, badania i opracowania naukowe i branżowe, artykuły i doniesienia medialne, a także – będące przedmiotem zainteresowania w tym rozdziale – UGC. B.N. Noone i współpracownicy potwierdzają, że przegląd UGC w mediach społecznościowych może się przyczynić do lepszego zrozumienia specyfiki konsumentów, którzy korzystają z określonych narzędzi SM<sup>42</sup>.

Pomimo wszechobecności UGC jego wykorzystanie niesie potencjalne ryzyko. Po pierwsze, należy pamiętać, że uzyskane zbiory nie są zwykle reprezentatywne dla całej badanej populacji. Nie wszyscy potencjalni uczestnicy grupy docelowej generują treści,

<sup>41</sup> Klasyczna koncepcja tzw. 4P obejmuje zestaw czterech podstawowych narzędzi – produkt, cena, promocja, dystrybucja. Pogłębiająca się konkurencja i rosnące oczekiwania konsumentów wymusiły rozbudowanie mieszanki marketingowej o kolejne instrumenty, jak np. ludzie, element materialny, procesy.

<sup>42</sup> B.N. Noone, K.A. McGuire, K.V. Rohlfs, *Social Media Meets Hotel Revenue Management: Opportunities, Issues and Unanswered Questions*, "Journal of Revenue and Pricing Management" 2011, vol. 10, no. 4, s. 293–305.

a ponadto nie do wszystkich – z uwagi na ochronę prywatności i brak odpowiednich technologii – można dotrzeć.

Po drugie, udział osób spoza samej firmy w tworzeniu danych utrudnia kontrolę nad formą ich dostarczenia. Są one często mało ustrukturyzowane i wymagają dużego zaangażowania w ich pozyskanie i przygotowanie. Ponadto ilość potencjalnych danych jest bardzo duża i konieczne jest podjęcie decyzji o ich selekcji. W przypadku dużych firm, dysponujących znacznymi budżetami na badania rynku, możliwe jest skorzystanie z odpłatnych narzędzi lub usług zewnętrznych, które dostarczą niezbędnych informacji. Mniejsze przedsiębiorstwa zmuszone są zwykle do korzystania z tańszych rozwiązań.

Często popełnianym przez decydentów błędem jest ograniczenie się do śledzenia treści generowanych przez użytkowników na stronach internetowych i profilach w mediach społecznościowych własnej firmy. Tymczasem internauci udostępniają dane w miejscach, które są im łatwiej i szybciej dostępne, jak choćby grupy dyskusyjne, społeczności, w których przebywają na co dzień, ulubione wyszukiwarki, porównywarki czy własne strony z blogami. Ponadto nie zawsze treść musi dotyczyć konkretnej firmy, ale np. samych oczekiwań i postaw wobec danej kategorii produktów, które znajdują się obecnie w ofercie lub planowane są do wdrożenia w przyszłości. Wniosek na temat potencjalnego wzrostu popytu na opony zimowe można wysnuć np. na podstawie publikacji dotyczących zakupów nowych aut oraz dyskusji na temat spodziewanego mroźnego sezonu zimowego, a nie wyłącznie na podstawie zapytań wysyłanych na profilu Facebooka producenta ogumienia.

Obecnie dostępne są także specjalne usługi i serwisy, które pozwalają na bieżące automatyczne przeszukiwanie sieci w celu wyławiania nowych treści dotyczących danej kategorii i zawierających wybrane słowa kluczowe, np. Gavagai Monitor, Brand24. Moduły do przeszukiwania sieci w celu identyfikacji miejsc zawierających interesujące dane zostały także wprowadzone do programów do analizy statystycznej, jak np. Statistica, SAS, SPSS, ale także Excel, oraz do analizy jakościowej danych – NVIVO, czy MaxQDA.

## Studium przypadku

Przedmiotem analizy jest mała firma rodzinna z siedzibą w woj. podlaskim zatrudniająca do 10 osób. Przedsiębiorstwo istnieje od 1996 r. i świadczy usługi turystyki aktywnej – organizacja spływów kajakowych, wycieczek rowerowych i innych imprez turystycznych. Firma wpisana jest do rejestru organizatorów turystyki prowadzonego przez marszałka woj. podlaskiego. Od 2013 r. firma dysponuje własnym zapleczem noclegowym w postaci głównego budynku hotelowego, bungalowu i domków letniskowych. Obiekty nie są kategoryzowane. Łączna liczba miejsc noclegowych to 60, z czego w tzw. wysokim sezonie, od początku maja do końca września, wykorzystanych jest średnio ok. 80% miejsc. W pozostałych okresach średnie obłożenie jest znacznie niższe.

Przedsiębiorstwo, aby utrzymać się na rynku, zmuszone jest do monitorowania potrzeb potencjalnych klientów w celu wprowadzania innowacji wyprzedzających ofertę konkurentów. Ponadto ważne jest przewidywanie popytu na usługi oraz ciągłe podnoszenie jakości świadczeń.

Firma prowadzi swój profil na Facebooku, obecnie śledzony przez ponad 1650 osób. Ma także konto na Instagramie, gdzie liczba obserwujących jest znacznie mniejsza (poniżej 200 osób), a także kanał na YouTube.

Ponadto utworzony został rekord na serwisie TripAdvisor dotyczący firmy. Nie zostały tam jednak zamieszczone jeszcze żadne recenzje, ale w zakładce dotyczącej miejscowości, w której firma ma swoją siedzibę, pojawiają się treści dotyczące oferowanych usług. Natomiast aktywnie komentowane jest zaplecze noclegowe. Przedsiębiorstwo pojawia się także w relacjach blogerów.

Dla właścicieli firmy istotne jest zidentyfikowanie miejsc, w których internauci poszukują informacji, ale również je zostawiają. W analizowanym przykładzie mamy do czynienia z firmą z sektora turystycznego. Jest to specyficzna branża, która w dużej mierze zaabsorbowała możliwości, jakie daje Internet. Dzisiejszy proces podejmowania decyzji o wyborze miejsca docelowego podróży czy o zakupie wybranego produktu przeniesiony został w dużej mierze do przestrzeni internetowej. Produkty turystyczne są niematerialne i stanowią przede wszystkim sumę przeżyć i emocji odbiorcy, można je określić mianem produktu-doświadczenia<sup>43</sup>. Nie jest możliwe wcześniejsze zweryfikowanie wszystkich elementów wchodzących w jego skład. W związku z tym decyzje zakupowe warunkowane są głównie wyobrażeniem o nim. Dlatego tak istotne stają się przekazy innych konsumentów. C. Költringer i A. Dickinger podkreślają, że UGC są najbogatszym i najbardziej zróżnicowanym źródłem informacji online w turystyce<sup>44</sup>.

W kontekście podróży konsumenci często rozpowszechniają informacje i rekomendacje ze względu na silne uczucia związane z doświadczaniem produktu, takie jak przyjemność, irytacja czy zawód. W niektórych przypadkach dzielenie się przyjemnymi doznaniem z podróży jest postrzegane jako część pozytywnych doświadczeń<sup>45</sup>.

Skoro Internet jest jednym z najważniejszych źródeł informacji dla klienta, a dodatkowo szczególnie cenne są rekomendacje i relacje innych osób, które wcześniej skorzystały z oferty, to warto się zastanowić, gdzie te informacje można uzyskać.

<sup>43</sup> L. Bei, E.Y.I. Chen, R. Widdows, *Consumers' Online Information Search Behavior and the Phenomenon of Search vs. Experience Products*, "Journal of Family and Economic Issues" 2004, vol. 25, no. 4, s. 449–467.

<sup>44</sup> C. Költringer, A. Dickinger, *Analyzing Destination Branding and Image from online Sources: A Web Content Mining Approach*, "Journal of Business Research" 2015, vol. 68, no. 9, s. 1836–1843.

<sup>45</sup> S.W. Litvin, R.E. Goldsmith, B. Pan, *Electronic Word-of-mouth in Hospitality and Tourism Management*, "Tourism Management" 2008, vol. 29, s. 458–468.

Zgodnie z raportem TripAdvisor 2017–2018 z badania przeprowadzonego na grupie użytkowników tego serwisu źródła informacji inspirujące wybór miejsca docelowego podróży kolejno według wskazań badanych to: TripAdvisor (37%), wyszukiwarki internetowe (10%), treści umieszczane przez użytkowników w SM (9%), przekaz ustny (WOM<sup>46</sup>) (9%), OTA<sup>47</sup> (8%), strony www z przewodnikami turystycznymi (5%), przewodniki turystyczne (5%), treści umieszczane online przez znajomych i rodzinę (3%), oficjalne rankingi hotelowe (2%) oraz inni dostawcy turystycznych treści internetowych (2%)<sup>48</sup>.

Mając świadomość, jak istotne w kierowaniu ruchu internautów są wyszukiwarki, a szczególnie Google, analizę UGC przytoczona firma powinna rozpocząć od sprawdzenia ogólnych zachowań w serwisie Google Trends. Strona prezentuje trendy wyszukań słów kluczowych. Obok weryfikacji, jak się prezentują wyszukiwania marki firmy, nazwy miejscowości i regionu, w którym prowadzona jest działalność, warto prześledzić zapytania dotyczące form turystyki oferowanych przez przedsiębiorstwo, np. spływy kajakowe, rajdy rowerowe, turystyka aktywna itp. Różnice w liczbie wyszukań w ciągu dłuższego okresu ułatwi oszacowanie przypuszczalnego rozkładu popytu w kolejnych latach. Przykładowo zapytanie „spływy kajakowe” wyraźnie częściej było wybierane w 2019 r. w miesiącach od maja do września niż w pozostałych (rysunek 3.5.1). Analiza wskazań Google Trends pozwala także na określenie, gdzie wyszukania występują, co dodatkowo dostarcza informacji na temat miejsc istotnych dla internautów, gdzie być może kumulowane jest UGC.

Do ważnych miejsc w Internecie, w których potencjalni klienci firmy poszukują informacji, należy TripAdvisor. Jest to platforma, na której internauci mogą oceniać miejsca docelowe podróży, wskazywać najważniejsze ich zdaniem atrakcje, miejsca noclegowe i gastronomiczne, a także je komentować. Liczba treści generowanych przez użytkowników jest tak duża, że serwis można uznać za jedno z najbogatszych repozytoriów UGC dla turystyki.

Kolejnym istotnym dla firmy źródłem są internetowe platformy pośredniczące (OTA). Analizowane przedsiębiorstwo współpracuje z serwisem Booking.com, na którym w dniu tworzenia niniejszego rozdziału umieszczono 170 opinii. Przy tak niewielkiej liczbie treści możliwe jest analizowanie ich na bieżąco, bez angażowania komputerowych narzędzi wspomagających. Sprawy mają się podobnie w przypadku profili na SM prowadzonych przez przedsiębiorstwo.

Bardzo istotne jest także regularne przeszukiwanie sieci za pomocą wyszukiwarek internetowych (np. Google) tak, aby wykrywać nowe zdarzenia prezentujące firmę i jej ofertę, np. na nowo tworzonych wpisach na blogach, forach i grupach dyskusyjnych. Zasadne jest także skorzystanie z usług monitoringu sieci.

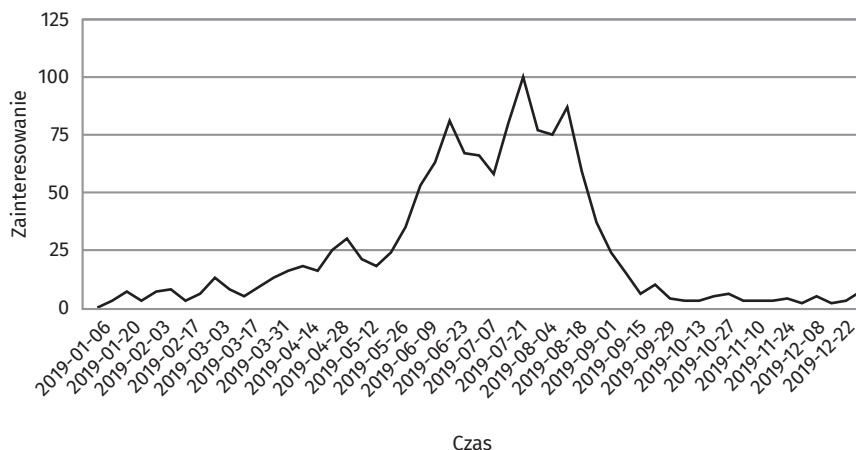
<sup>46</sup> *Word-of-Mouth.*

<sup>47</sup> Online Travel Agencies – internetowe platformy pośredniczące, jak np. Booking.com

<sup>48</sup> TripAdvisor, *TripBarometer 2017/18 Global Report*, 2018.

**Rysunek 3.5.1.**

Zainteresowanie hasłem „splywy kajakowe” w wyszukiwarkach w Polsce w 2019 r.



Liczby reprezentują poszczególne zainteresowania w wyszukiwaniu względem najwyższego punktu na wykresie. Wartość 100 oznacza najwyższą popularność hasła. Wartość 50 oznacza, że popularność hasła była dwukrotnie mniejsza. Wartość 0 wskazuje, że dla danego hasła nie ma wystarczających danych.

Źródło: Google Trends, <https://trends.google.com/trends/explore?geo=PL&q=splywy%20kajakowe>

## Pozyskiwanie danych z UGC

Prawidłowa identyfikacja źródeł UGC to pierwszy krok w procesie ich analizy. W tym celu konieczne jest przygotowanie zbiorów danych zawierających interesujące dla decydenta informacje. O ile niewielkie zbiory, zawierające kilkadziesiąt czy nawet kilkaset tekstów, nie stanowią dużego problemu, o tyle te większe są już kłopotliwe w analizie. Niewielki zasób możliwy jest do bieżącego monitorowania np. kilka razy w tygodniu, co ani nie jest czasochłonne, ani nie angażuje dodatkowych środków finansowych. W przypadku dużych zbiorów lub takich, które są bardziej rozproszone, konieczne jest już wsparcie komputerowe.

Chociaż dane z UGC są często aktualne i wartościowe dla biznesu, ich względna niedostępność wynikająca z trudności w pozyskaniu stanowi główną przeszkodę w ich wykorzystaniu w praktyce. Istnieje ciągle niewiele metod pobierania UGC wbudowanych w strony internetowe i nie ma ujednoczonych interfejsów do zbierania danych ze wszystkich stron. Problem jest mniejszy w przypadku niektórych SM, np. Twittera, gdzie zawartość jest zwykle dostępna poprzez interfejsy oprogramowania aplikacji (tzw. API), czyli interfejsy stworzone specjalnie do zautomatyzowanego realizowania zapytań o przechowywane dane.

Ze względu na brak uniwersalnego narzędzia do gromadzenia tych danych pozyskiwanie UGC z wielu źródeł jest trudne. Znaczna część danych pozostaje więc poza zasięgiem badaczy, w szczególności, jeżeli dane te znajdują się na stronach niedostępnych dla wyszukiwarek internetowych<sup>49</sup>.

Pobranie dużego zbioru danych z interesującego źródła w sposób zautomatyzowany wymaga zatem przygotowania odpowiedniego narzędzia w postaci tzw. web scrapera. Może ono być napisane za pomocą różnych języków programowania, takich jak Python lub Java. Istnieją także narzędzia, które nie wymagają umiejętności programowania, a jedynie określenia ścieżek dostępu (map strony) do interesujących informacji. Przykładem mogą być tu rozszerzenia przeglądarek internetowych, wśród których najpopularniejszy jest dodatek Googla dla przeglądarki Chrome o nazwie Webscrapler.

Wraz z rozwojem UGC i kanałów, w których są one umieszczane, oferowane są nowe narzędzia, które ułatwiają ich wykorzystanie. Są to w większości oferty komercyjne, ale ich koszt nie przekracza zwykle możliwości małego przedsiębiorstwa. Umowy zwykle oparte są na zasadzie subskrypcji lub abonamentu na usługi na dłuższy okres, co pozwala na bieżący dostęp do użytecznych danych, często przygotowanych w postaci gotowych raportów z ich analizy (jak np. analiza sentymentu).

Należy pamiętać, że dane zgromadzone w Internecie są własnością intelektualną twórcy. Prawo zapewnia właścicielowi możliwość określenia sposobu, w jaki chciałby on udostępnić dane. O ile większość UGC można swobodnie wyszukiwać, czytać i przeglądać, o tyle już ich kopiowanie, przetwarzanie, a w szczególności dalsza redystrybucja mogą podlegać już pewnym ograniczeniom. Dlatego też w przypadku samodzielnego pozyskiwania treści tak istotne jest zweryfikowanie, na jakie sposoby wykorzystania danych zezwala określone źródło.

Ogólnodostępne narzędzia monitorujące zapewniają dostęp jedynie do treści publicznych. Nie dostarczają zatem pełnych danych w ramach interesującego decydenta tematu. Mimo tego ich prostota i łatwość w użyciu sprawiają, że stają się one istotnym wsparciem dla monitorowania zachowań i opinii grup docelowych.

## Studium przypadku

Skala działalności prowadzonej przez analizowane przedsiębiorstwo nie powoduje, że w sieci pojawia się dużo UGC dotyczących samej marki firmy i odnoszących się specyficznie do jej produktów. Analiza tych danych także w zakresie oceny sentymentu nie jest więc bardzo skomplikowana i wymaga jedynie systematyczności. Jak wskazano

---

<sup>49</sup> P. Liakos, A. Ntoulas, A. Labrinidis, A. Delis, *Focused Crawling for the Hidden Web*, "World Wide Web" 2016, vol. 19, no. 4, s. 605–631.

we wcześniejszym podrozdziale, firma może jednak sięgnąć po większe zbiory danych, które ułatwią określenie potrzeb i oczekiwań grup docelowych. Podstawowym dużym zbiorem danych może być serwis TripAdvisor, a docelowo także Booking.com, ponieważ znajdują się tam tysiące opinii. Z uwagi na dużą popularność tych serwisów na rynku można znaleźć wiele narzędzi gotowych, także *open source*, które pozwolą na pobranie zestawienia danych tekstowych do dalszej analizy. W ramach usług dla klientów biznesowych same platformy też udostępniają narzędzia do pobierania danych.

## Analiza danych UGC

Analiza UGC w małych i dużych zbiorach sprowadza się głównie do odczytania faktycznego przekazu i próby jego zrozumienia. W przypadku niewielkich zbiorów analiza sprowadza się do przeczytania wszystkich dostarczonych informacji i wnioskowania na ich podstawie. Można wspomóc się prostym zestawieniem danych w formularzu Excel lub innym programie do tworzenia baz danych. Ważne jest oprócz analizowania treści także określenie jej źródła i czasu powstania, co ułatwi jej osadzenie w szerszym kontekście. Ponieważ większość uzyskanych danych ma charakter jakościowy, możliwe jest posłużenie się specjalistycznym oprogramowaniem, tzw. CAQDAs (*Computer Assisted Qualitative Data Analysis*). Programy takie jak NVIVO, MaxQDA, Atlas pozwalają na przeszukiwanie, kodowanie danych, raportowanie wniosków i zestawień w sposób bardzo intuicyjny i przystępny. Wymagają jednak trochę czasu na opanowanie funkcji oraz środków na zakup licencji.

Wraz ze wzrostem wielkości zbiorów danych ręczna weryfikacja staje się mniej praktyczna. Eksploracja dużych zbiorów danych (*data mining*) wymaga wsparcia komputerowego. *Data mining* sprowadza się głównie do odkrywania statystycznych prawidłowości i przedstawienia ich w formie reguł logicznych, drzew decyzyjnych lub sieci neuronowych<sup>50</sup>.

Pierwszym etapem jest zwykle przygotowanie danych poprzez ich wstępną ocenę i oczyszczenie. Celem jest wydobycie tych, których jakość jest wystarczająco dobra i które poddadzą się analizie. W przypadku UGC, gdzie treści tworzone są w sposób nieustrukturyzowany i często przypadkowy, jest to szczególnie istotne. Można w tym celu wykorzystać filtrowanie treści, które może być przeprowadzane ręcznie lub automatycznie przez algorytmy.

---

<sup>50</sup> A. Maheshwari, *Business Intelligence and Data Mining*, Business Expert Press, 2014. ProQuest Ebook Central, <https://ebookcentral.proquest.com/lib/sgh/detail.action?docID=1911815>

Przygotowane zbiory można poddać dalszej analizie. Istnieją dwa podstawowe rodzaje procesu eksploracji danych z wykorzystaniem uczenia maszynowego (*machine learning*): nauka nadzorowana (*supervised learning*) i nauka nienadzorowana (*unsupervised learning*). W procesie uczenia się pod nadzorem można stworzyć model decyzyjny wykorzystujący dane z przeszłości, a następnie wykorzystać ten model do przewidywania prawidłowych odpowiedzi dla przyszłych zdarzeń<sup>51</sup>. W przypadku uczenia nienadzorowanego wzorzec nie jest wcześniej dostarczany, a maszyna musi sama go wytworzyć na podstawie dostarczonych danych. Istnieją jeszcze formy pośrednie uczenia częściowo nadzorowanego oraz uczenie wzmocnione, w którym na wstępie dostarczane są ograniczenia oraz reguły dozwolonych działań<sup>52</sup>.

Obecnie dostępne rozwiązania stają się coraz doskonalsze, a wyniki dzięki nim otrzymywane coraz bardziej wiarygodne i praktyczne. Ciągle jeszcze maszyna nie jest jednak w stanie zastąpić inteligencji człowieka i część obszarów zastosowania nastrocza sporo problemów, szczególnie tam, gdzie dane mają charakter heterogeniczny i nieustrukturyzowany, tak jak na przykład w rozumieniu języka naturalnego. Podstawową trudnością jest analiza kontekstu wypowiedzi i faktycznych intencji autora. Ponadto większość narzędzi opiera się na słownikach języka angielskiego, co także ogranicza możliwość analizy treści w innych językach. Jednak i tu w ostatnich latach obserwowany jest znaczny postęp.

## Studium przypadku

W przypadku analizowanego przedsiębiorstwa, z uwagi na zakres pozyskiwanych danych oraz skalę ich wykorzystania, najlepsze wydaje się stosowanie gotowych rozwiązań do śledzenia sentymentu wobec produktów i marki. Ponieważ firma nie zatrudnia specjalistów, którzy mogą przygotować zindywidualizowane narzędzia do pozyskiwania dużych zbiorów danych o opiniach i trendach zachowań klientów, wskazane jest skorzystanie z wbudowanych możliwości analitycznych oferowanych przez dostawców portali, na których treści te są generowane.

---

<sup>51</sup> *Ibidem*, s. 50.

<sup>52</sup> Szerzej na ten temat [https://www.sas.com/pl\\_pl/news/informacje-prasowe-pl/2018/cztery-typy-uczenia-maszynowego.html](https://www.sas.com/pl_pl/news/informacje-prasowe-pl/2018/cztery-typy-uczenia-maszynowego.html)



## Podsumowanie

UGC stanowią przedmiot zainteresowania zarówno praktyków, jak i naukowców. Jedni i drudzy dostrzegają ogromny potencjał, który drzemie w bogactwie zgromadzonych w sieci informacji. Ilość UGC w Internecie jest duża, ale obecnie łatwo dostępne są tylko treści dostarczane przez API, które stanowią jedynie ułamek całkowitego zasobu UGC. Dzięki opracowaniu zautomatyzowanej metody ekstrakcji UGC w znormalizowanym formacie strony internetowe takie jak artykuły informacyjne, blogi i inne szeroko wykorzystywane media będą mogły zostać odblokowane i udostępnione jako praktyczne źródło danych dla podejmowania decyzji w biznesie.

Istotna kwestia to jakość i wiarygodność wyników otrzymanych z badań z wykorzystaniem UGC. Z całą pewnością nie są to analizy reprezentatywne. Ponadto należy się liczyć z tym, że nadawcy przekonani o anonimowości mogą zakłamywać przekaz i tworzyć byty nierzeczywiste. Z drugiej strony wielu tradycyjnych badań społecznych opartych na opiniach respondentów również dotyka ten problem.

Prowadzone intensywne prace nad przygotowaniem nowych narzędzi pozyskiwania i analizowania treści z Internetu z pewnością w niedalekiej przyszłości dostarczą rozwiązań, które jeszcze bardziej uproszczą ten proces. Potencjalne zmniejszenie koniecznych na pozyskanie interesujących informacji nakładów powinno się przyczynić do większego wykorzystania zasobów UGC także przez małe przedsiębiorstwa. Proste śledzenie opinii konsumentów i przewidywanie trendów ich zachowań z pewnością ułatwiłoby podejmowanie decyzji marketingowych na wszystkich etapach – od planowania strategicznego, po działania operacyjne.

## Bibliografia

- Agarwal A., Xie B., Vovsha I., Rambow O., Passonneau R., *Sentiment Analysis of Twitter Data*, Proceedings of the Workshop on Language in Social Media (LSM 2011), Association for Computational Linguistics, Portland, Oregon 2011, s. 30–38.
- Angehrn A., *Designing Mature Internet Business Strategies: The ICDT Model*, "European Management Journal" 1997, vol. 15, no. 4, s. 361–369.
- Bei L., Chen E.Y.I., Widdows R., *Consumers' Online Information Search Behavior and the Phenomenon of Search vs. Experience Products*, "Journal of Family and Economic Issues" 2004, vol. 25, no. 4, s. 449–467.
- Bello-Organ G., Jung J.J., Camacho D., *Social Big Data: Recent Achievements and New Challenges*, "Information Fusion" 2016, vol. 28, s. 45–59.

- Borth D., Ji R., Chen T., Breuel T., Chang S.-F., *Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs*, ACM International Conference on Multimedia, 2013, s. 223–232.
- Information Systems*, 17th European Conference on Information System.
- Castillo C., Mendoza M., Poblete B., *Information Credibility on Twitter*, 2011, s. 675–684.
- Cheung C.M.K., Chiu P., Lee M.K.O., *Online Social Networks: Why do Students Use Facebook?*, “Computers in Human Behavior” 2011, vol. 27, no. 4, s. 1337–1343.
- Chia A., *Welcome to Me-Mart: The Politics of User-Generated Content in Personal Blogs*, “American Behavioral Scientist” 2012, vol. 56, no. 4, s. 421–438.
- Chmiel A., Sobkowicz P., Sienkiewicz J., Paltoglou G., Buckley K., Thelwall M., Hołyst J.A., *Negative Emotions Boost User Activity at BBC Forum*, “Physica A” 2011, vol. 390, no. 16, s. 2936–2944.
- Clavel C., Callejas Z., *Sentiment Analysis: From Opinion Mining to Human – Agent Interaction*, “IEEE Trans. Affect. Comput.” 2015, s. 74–93.
- Cvijikj I.P., Michahelles F., *Monitoring Trends on Facebook*, IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing, 2011, s. 896–903.
- Daugherty T., Eastin M.S., Bright L., *Exploring Consumer Motivations for Creating User-Generated Content*, “Journal of Interactive Advertising” 2008, vol. 8, no. 2, s. 16–25.
- Felix N., Silva F.D.A., Coletta, L.F.S., Hruschka E.R., *A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning*, “ACM Computing Survey” 2016, vol. 49, no. 1, s. 1–26.
- Gumkowska A., Maryl M., Toczyński P., *Blog to... blog. Blogi oczyma blogerów*, Raport z badania jakościowego zrealizowanego przez Instytut Badań Literackich PAN i Gazeta.pl, *Tekst w sieci. Tekst. Język. Gatunki*, D. Ulicka (red.), Wydawnictwa Akademickie i Profesjonalne, Warszawa 2009, s. 285–309.
- Halavais A., *Bigger Sociological Imaginations: Framing Big Social Data Theory and Methods*, “Information, Communication & Society” 2015, vol. 18, no. 5, s. 583–594.
- <https://trends.google.com/trends/explore?geo=PL&q=splywy%20kajakowe>
- [https://www.sas.com/pl\\_pl/news/informacje-prasowe-pl/2018/cztery-typy-uczenia-maszynowego.html](https://www.sas.com/pl_pl/news/informacje-prasowe-pl/2018/cztery-typy-uczenia-maszynowego.html)
- Irsoy O., Cardie C., *Opinion Mining with Deep Recurrent Neural Networks*, “EMNLP” 2014, s. 720–728.
- Jaciow M., *Fora internetowe jako źródło informacji – możliwości i ograniczenia wykorzystania w badaniach zachowań nabywczych*, „Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu” 2011, nr 236, s. 113–122.
- Jin S.A., *The Potential of Social Media for Luxury Brand Management*, “Marketing Intelligence & Planning” 2012, vol. 30, no. 7, s. 687–699.
- Jin V.S., Phua J., Lee K.M., *Computers in Human Behavior Telling Stories about Breastfeeding through Facebook: The Impact of User-generated Content (UGC) on Pro-breastfeeding attitudes*, “Computers in Human Behavior” 2015, vol 46, s. 6–17.
- Jou B., Chang S.-F., *Deep Cross Residual Learning for Multitask Visual Recognition*, ACM International Conference on Multimedia (MM), 2016.
- Kim A.J., Johnson K.K.P., *Power of Consumers Using Social Media: Examining the Influences of Brand-related User-generated Content on Facebook*, “Computers in Human Behavior” 2016, vol. 58, s. 98–108.
- Költringer C., Dickinger A., *Analyzing Destination Branding and Image from Online Sources: A Web Content Mining Approach*, “Journal of Business Research” 2015, vol. 68, no. 9, s. 1836–1843.
- Lamos V., Bie T.D., Cristianini N., *Flu Detector – Tracking Epidemics on Twitter*, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2010), Springer, Barcelona 2010, s. 599–602.

- Leung L., *User-generated Content on the Internet: An Examination of Gratifications, Civic Engagement and Psychological Empowerment*, "New Media and Society" 2009, vol. 11, no. 8, s. 1327–1347.
- Liakos P., Ntoulas A., Labrinidis A., Delis A., *Focused Crawling for the Hidden Web*, "World Wide Web" 2015.
- Liu B., *Sentiment Analysis and Opinion Mining*, "Synthesis Lectures on Human Language Technologies" 2012, vol. 5, no. 1, s. 1–167.
- Liu X., Burns A.C., Hou Y., *An Investigation of Brand-Related User-Generated Content on Twitter, An Investigation of Brand-Related User-Generated Content on Twitter*, "Journal of Advertising" 2017, vol. 46, no. 2, s. 1–12.
- Litvin S.W., Goldsmith R.E., Pan B., *Electronic Word-of-mouth in Hospitality and Tourism Management*, "Tourism Management" 2008, vol. 29, s. 458–468.
- Maheshwari A., *Business Intelligence and Data Mining*, "Business Expert Press" 2014. ProQuest Ebook Central, <https://ebookcentral.proquest.com/lib/sgh/detail.action?docID=1911815>
- Mishne G., Glance N., *Predicting Movie Sales from Blogger Sentiment*, American Association for Artificial Intelligence, 2005.
- Nadkarni A., Hofmann S.G., *Why Do People Use Facebook?*, "Personality and Individual Differences" 2012, vol. 52, no. 3, s. 243–249.
- Noone B.N., McGuire K.A., Rohlf K.V., *Social Media Meets Hotel Revenue Management: Opportunities, Issues and Unanswered Questions*, "Journal of Revenue and Pricing Management" 2011, vol. 10, no. 4, s. 293–305.
- O'Reilly T., *What is Web 2.0: Design Patterns and Business Models for the Next Generation Software*, "Communication and Strategies" 2007, vol., no. 1, s. 17–37.
- OECD, *Participative Web and User-Created Content. Web 2.0, WIKIS and Social Networking* 2007.
- Pang B., Lee L., *Opinion Mining and Sentiment Analysis*, "Foundations and Trends in Information Retrieval" 2008, vol. 2, no. 1–2, s. 1–135.
- Pérez Rosas V., Mihalcea R., Morency L.P., *Multimodal Sentiment Analysis of Spanish Online Videos*, "IEEE Intell. Syst." 2013, vol. 28, no. 3, s. 38–45.
- Schaedel U., Clement M., *Managing the Online Crowd: Motivations for Engagement in User-Generated Content*, "Journal of Media Business Studies" 2010, vol. 7, no. 3, s. 16–36.
- Shad H., Jamali S., Aleahmad A., *Online Reputation Measurement of Companies Based on User-generated Content in Online Social Networks*, "Computers in Human Behavior" 2016, vol. 54, s. 94–100.
- Stoeckl R., Rohrmeier P., Hess T., *Motivations To Produce User Generated Content: Differences Between Webloggers And Videobloggers*, BLED 2007 Proceedings, 2007.
- Sun Y., Dong X., McIntyre S., *Motivation of User-Generated Content: Social Connectedness Moderates the Effects of Monetary Rewards Monetary Rewards*, "Marketing Science" 2017, vol. 36, no. 3, s. 329–337.
- Tirunillai S., Tellis G.J., *Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation*, "Journal of Marketing Research" 2014, vol. 51, s. 463–479.
- TripAdvisor, *TripBarometer 2017/18 Global Report*.