

Wikipedia jako narzędzie badawcze

TOMASZ RABURSKI

Wprowadzenie

Wikipedia jest jednym z najbardziej rozpoznawalnych miejsc w sieci i sztandarowym projektem społeczeństwa informacyjnego. Jest ona obecnie autonomicznym i dynamicznie rozwijającym się obszarem badawczym, nie tylko wśród osób zainteresowanych społeczeństwem sieciowym czy mediami, ale również wśród badaczy zarządzania czy różnych subdyscyplin socjologii. Na jej temat wydano już ponad tysiąc prac badawczych¹, a kilka tysięcy ma Wikipedię w tytule². Najbardziej aktualne całościowe podsumowanie badań i metod badawczych dotyczących Wikipedii można znaleźć w niepublikowanym jeszcze artykule Finna Årupa Nielsena *Wikipedia Research and Tools: Review and Comments*³.

¹ F. Å. Nielsen, *Wikipedia research and tools: Review and comments*, [online:] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2129874 [dostęp 01.11.2015].

² Wikipedia – Signpost, [online:] https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2011-04-11/Recent_research [dostęp 01.11.2015].

³ F. Å. Nielsen, *Wikipedia research...*, dz. cyt.

Wikipedyści gromadzą informacje na temat badań społecznych dotyczących tego projektu. Najlepszym źródłem jest miesięcznik „Wikimedia Research Newsletter”, przygotowywany przez Wikimedia Research Committee i redakcję newslettera „Singpost”, a publikowany na stronie Wikimedia Foundation⁴. Odnotowuje on najważniejsze (przede wszystkim anglojęzyczne) prace badawcze na temat Wikipedii i podsumowuje ich wyniki. Więcej materiałów (w postaci danych bibliograficznych) zebrano na stronie Academic Studies of Wikipedia⁵. Obok newslettera na stronie Wikimedia Foundation można też znaleźć odnośniki do repozytoriów danych, projektów badawczych wspieranych przez fundację czy dane badaczy⁶. Istnieje również kanał na Twitterze informujący o aktualnych badaniach⁷. Niezależnym od fundacji portalem zbierającym informacje o badaniach na temat projektów i oprogramowania wiki jest Wikipapers⁸.

Dane bibliograficzne polskich badań nad Wikipedią znaleźć można na stronie „Wikipedia w publikacjach

⁴ Wikimedia Meta-Wiki – Newsletter, [online:] <https://meta.wikimedia.org/wiki/Research:Newsletter> [dostęp 01.11.2015].

⁵ Wikimedia Meta-Wiki – Index, [online:] <https://meta.wikimedia.org/wiki/Research:Index> [dostęp 01.11.2015].

⁶ Tamże.

⁷ WikiResearch na Twitterze, użytkownik, [online:] <https://twitter.com/wikiresearch> [dostęp 01.11.2015].

⁸ Wikipapers – Main Page, [online:] http://wikipapers.referata.com/wiki/Main_Page [dostęp 01.11.2015].

naukowych”⁹. Z kolei niemieckojęzyczne prace zebrane są na portalu Wikipedia:Wikipedistik¹⁰. W przygotowaniu tego tekstu obficie korzystałem z wymienionych źródeł. Trzeba jednak pamiętać, że są one dalece niewyczerpujące.

W tym artykule chciałbym zająć się kwestią wykorzystania Wikipedii jako narzędzia badawczego czy wsparcia w badaniach. Podejście takie może być bardzo owocne, zarówno z powodu centralności Wikipedii w sieci i jej rosnącej roli w życiu codziennym milionów osób na całym świecie, ale jak i tego, że jest to masowy projekt oparty na idei wolnych licencji¹¹. Tym samym jest to źródło tanie oraz w wysokim stopniu przejrzyste i łatwe do wykorzystania przez badaczy.

Najpierw omówię podstawowe kwestie związane ze strukturą Wikipedii i sposobem jej funkcjonowania (osoby znające ten projekt mogą tę część pominąć). Następnie zaś przejdę do zestawów narzędzi i danych dostępnych w otoczeniu Wikipedii, a także opiszę typy badań, w których mogą one zostać wykorzystane. Dobierając przykłady badań mające ilustrować ten artykuł, kierowałem się przede wszystkim ich różnorodnością czy pomysłowością.

⁹ Wikipedia – Wikipedia w publikacjach naukowych, [online:] https://pl.wikipedia.org/wiki/Wikipedia:Wikipedia_w_publicacjach_naukowych [dostęp 01.11.2015].

¹⁰ Wikipedia – Wikipedistik, [online:] <https://de.wikipedia.org/wiki/Wikipedia:Wikipedistik> [dostęp 01.11.2015].

¹¹ Na dzień 1.11.2015 aktualną wersją licencji jest CC BY SA 3.0.

Wybór ten jest z oczywistych względów niereprezentacyjny. Badacze, którzy byliby zainteresowani aktualnym stanem badań na temat Wikipedii, będą musieli sięgnąć do wskazanych źródeł. Obok samej Wikipedii (i jej projektów siostrzanych) wiele kwestii omówionych w tym artykule odnosić się może także do innych portali opartych na oprogramowaniu MediaWiki, w szczególności do multiwersum Wikii¹², niektórych stron czy portali prywatnych.

Wikipedia nie jest projektem jednolitym, lecz skomplikowanym konglomeratem zgromadzonych i uporządkowanych informacji, oprogramowania i ludzkich społeczności. To, co najbardziej widoczne dla użytkowników, to przestrzeń główna (przestrzeń artykułów lub *main*). Równie ważne są jednak pozostałe przestrzenie Wikipedii, takie jak m.in. przestrzeń dyskusji haseł, przestrzeń użytkowników (edytorów, wikipedystów), przestrzeń pomocy czy przestrzeń Wikipedii (w której zamieszczone są m.in. strony techniczne czy redakcyjne)¹³. W przestrzeniach tych toczą się dyskusje nad kształtem haseł, wspólnymi działaniami, ustalaniem standardów. Każda z przestrzeni oferuje inny rodzaj danych. Na przykład

¹² Fandom, [online:] <http://www.wikia.com/Wikia> [dostęp 01.11.2015].

¹³ Pod względem technicznym przestrzenie poza *mainem* wyróżniane są przez dodanie przed nazwą artykułu nazwy przestrzeni i dwukropka (np. *Pomoc:Pierwsze_kroki*, *Wikipedia:Kawiarenka*, *Wikipedysta:Tomasz_Raburski*).

komunikacja między użytkownikami przebiega w przestrzeniach dyskusji, użytkowników oraz w przestrzeni Wikipedii. Informacje na temat funkcjonowania Wikipedii, dostęp do zagregowanych danych i narzędzi badawczych znajduje się poza przestrzenią główną. Całość oparta jest na oprogramowaniu MediaWiki, które również ma charakter otwarty i transparentny (licencja GNU GPL)¹⁴. Wikipedia stara się utrzymywać ścisłą łączność z zewnętrznymi projektami opartymi na wolnych licencjach, takimi jak ruch wolnych licencji, Open Street Map czy Wolne Lektury.

W ścisłym związku z Wikipedią rozwijane są jej projekty siostrzane: m.in. Commons, Wikisłownik, Wikiźródła, Wikicytaty, Wikipodróżę, Wikibooks i Wikidane¹⁵. Projekty te są ściśle powiązane ze sobą i korzystają wzajemnie ze swoich danych. Wszystkie zawierają materiały na wolnych licencjach. Duża część edytorów funkcjonuje jednocześnie w kilku projektach. Najbardziej rozbudowanym (obok samej Wikipedii) jest Commons, będące repozytorium materiałów audiowizualnych. Wikiźródła nastawione są na cyfryzację i udostępnianie utworów tekstowych. Wikibooks zawierają podręczniki (w dużej mierze oparte na materiałach z Wikipedii i Commons), a Wikipodróżę – przewodniki turystyczne. Jeden z młodszych

¹⁴ MediaWiki, [online:] <https://www.mediawiki.org/wiki/MediaWiki> [dostęp 01.11.2015].

¹⁵ Wikimedia Foundation – Our Projects, [online:] https://wikimediafoundation.org/wiki/Our_projects [dostęp 01.11.2015].

projektów siostrzanych, Wikidane, zawierają uporządkowane metadane, wykorzystywane w poszczególnych artykułach. Ze względu na duże znaczenie w badaniach projekt ten zostanie opisany samodzielnie w kolejnej sekcji. Poszczególne projekty spaja Meta-Wiki: przestrzeń działań dotyczących wielu projektów i społeczności (np. rozwoju oprogramowania, badań itp.).

Wikipedia (i projekty siostrzane) rozwijane są w wielu wersjach językowych (obecnie funkcjonuje 291 Wikipedi). Rozwojem poszczególnych wersji językowych zajmują się autonomiczne społeczności, które w większości przypadków samodzielnie piszą hasła (tłumaczenia są raczej mniejszością haseł). Poszczególne wersje językowe różnią się nie tylko zawartością merytoryczną, ale też ustalonymi zasadami edycji i standardami czy sposobem funkcjonowania społeczności¹⁶.

Wikipedię tworzą społeczności wolontariuszy. Obok tych społeczności działa Fundacja Wikimedia odpowiedzialna za utrzymanie serwerów, rozwój oprogramowania, obsługę prawną i wspomaganie społeczności (przez różne programy celowe)¹⁷. W niektórych krajach istnieją lokalne chaptery fundacji, na przykład w Polsce stowarzyszenie

¹⁶ W dalszej części artykułu stosować będę powszechną przy porównaniach konwencję skrótowego odwoływania się do poszczególnych wersji językowych: wersja anglojęzyczna to „enwiki” (en.wikipedia.org), niemieckojęzyczna do „dewiki” (de.wikipedia.org), polskojęzyczna to „plwiki”.

¹⁷ Wikimedia Foundation – Home, [online:] <https://wikimedia-foundation.org/wiki/Home> [dostęp 01.11.2015].

Wikimedia Polska¹⁸. Sama Wikipedia nie ma podmiotowości prawnej i nie zatrudnia żadnego pracownika. Pracownicy (informatycy, specjaliści od zarządzania i public relations, prawnicy) zatrudniani są przez fundację i niektóre chaptery. Wikimedia Foundation ma swój dział badań, na którego czele stoi Dario Taraborelli¹⁹.

Chociaż Wikipedię można edytować w sposób zupełnie anonimowy i dorywczy, większość edycji dokonywana jest przez zalogowanych użytkowników, tworzących luźną społeczność edytorów. Transparentność projektu ma swoje (nieliczne) ograniczenia, są nimi dostęp do treści skasowanych i ukrytych. Dostęp do tych treści mają administratorzy (w polskiej Wikipedii jest ich obecnie 118²⁰) i nie ma większych przeciwwskazań, by je udostępniali do celów badawczych. Wyjątkiem są tutaj ujawnione w edycjach dane osobowe oraz treści naruszające prawa autorskie, które są ukrywane na wyższym poziomie zabezpieczeń.

Duże znaczenie dla badaczy ma fakt, że właściwie wszystko, co znajduje się w przestrzeni Wikipedii i jej

¹⁸ Wikimedia Polska, [online:] https://pl.wikimedia.org/wiki/Strona_g%C5%82%C3%B3wna [dostęp 01.11.2015].

¹⁹ Wikimedia Foundation – Dario Taraborelli, użytkownik, [online:] <https://wikimediafoundation.org/wiki/User:Dtaraborelli> [dostęp 01.11.2015].

²⁰ Administratorzy w Wikipedii mają znacznie mniejsze uprawnienia niż w wielu innych projektach internetowych. Ich dodatkowe funkcje ograniczają się właściwie do kasowania i zabezpieczania haseł oraz blokowania użytkowników. Podobnie jak inni edytorzy, nie są oni opłacani.

projektów siostrzanych, dostępne jest na wolnych licencjach. Artykuły (a także treści w innych przestrzeniach Wikipedii i samo oprogramowanie) udostępniane są na licencji CC BY SA 3.0. Można więc z nich korzystać w dowolny sposób, jeśli poda się licencję i autorów danego utworu²¹. Niektóre treści Wikipedii (a większość w Wikiźródłach i duża część w Commons) znajdują się w domenie publicznej, jako że majątkowe prawa autorskie już do nich wygasły (stare ryciny, zdjęcia, utwory literackie). W Wikipedii dopuszcza się też stosowanie prawa cytatu (a więc zamieszczania krótkich fragmentów dzieł objętych pełnymi majątkowymi prawami autorskimi). Wolne licencje zapewniają pewną wspólną podstawę prawną, nie gwarantują jednak zupełnej odporności na typową dla internetu konkurencyjność różnych instytucji i jurysdykcji prawnych oraz niepewność co do obowiązującego prawa. Konkuruje ze sobą regulacje prawne właściwe dla miejsca, w którym umieszczone są serwery z danymi (główne serwery Wikipedii znajdują się w Ashburn w stanie Wirginia), prawo lokalne właściwe dla danego edytora oraz prawo właściwe ze względu na miejsce wystąpienia zdarzenia mającego skutki prawne (np. naruszenia dóbr

²¹ Z punktu widzenia praw autorskich niewłaściwe jest więc oznaczanie tych treści przez podanie „Źródło: Wikipedia”. Autorzy wyliczeni są w historii edycji artykułu. W przypadku wielu autorów możliwe jest podanie linku zamiast ich listy. Zob. też: Wikipedia – Pomoc:Kopiowanie poza Wikipedię, [online:] https://pl.wikipedia.org/wiki/Pomoc:Kopiowanie_poza_Wikipedi%C4%99 [dostęp 01.11.2015].

osobistych). Jeśli więc polski edytor naruszy dobre imię Francuza, możliwe jest rozpatrywanie danego sporu na bazie prawa polskiego, francuskiego czy stanu Wirginia. Szczególną uwagę trzeba zwrócić na instytucje prawa autorskiego występujące jedynie w niektórych systemach prawnych, a tym samym mocno ograniczające możliwość wykorzystywania niektórych treści²².

Dane i narzędzia

W tej części artykułu scharakteryzuję krótko rodzaje danych, które można w łatwy sposób pozyskać w przestrzeni projektów Wikimedia, jak i podstawowe narzędzia, wykorzystywane w głównej mierze przez edytorów czy administratorów, mogące mieć jednak znaczenie badawcze.

Treść Wikipedii (zarówno *main* jak i inne przestrzenie) kodowana jest w postaci zwykłego tekstu zawierającego tzw. wikikod (*MediaWiki markup*). Wprowadzony dwa lata temu edytor wizualny WYSIWYG (*Visual Editor*) ma charakter pomocniczy, a wykonywane w nim edycje również przekształcane są w zwykły tekst. Kodowanie

²² Na przykład nie we wszystkich systemach prawnych obowiązuje wolność panoramy, umożliwiającą udostępnianie zdjęć wolnostojących obiektów objętych prawami autorskimi (rzeźby czy budynki). W enwiki (w przeciwieństwie do dewiki i plwiki) dopuszczalne jest stosowanie materiałów objętych majątkowymi prawami autorskimi na podstawie instytucji *fair use* (co pozwala na zamieszczanie okładek książek i płyt czy logo przedsiębiorstw).

takie zapewnia przejrzystość danych, ułatwia ich agregację i przetwarzanie, a także zmniejsza ilość potrzebnego miejsca. Jest też dużym ułatwieniem dla potencjalnych badaczy.

Każda edycja w dowolnym artykule (również ta wycofana czy skasowana) jest zapisywana, oznaczana czasem, datą i autorstwem (co jest wymogiem licencyjnym). Dzięki stronom historii edycji hasła możliwe jest podejrzenie wszystkich tych edycji i porównanie różnych wersji artykułu. W efekcie możliwy jest dostęp nie tylko do aktualnej jego wersji, lecz do wszystkich poprzednich wersji wszystkich artykułów i porównania wersji projektów w różnych okresach. Pozwala to na śledzenie długotrwałych procesów społecznych, jak i na badania porównawcze.

W roku 2012 uruchomiony został projekt Wikidata (Wikidane), mający być bazą danych, które mogą być współdzielone między różnymi projektami²³. Zmiana informacji na poziomie Wikidata powoduje, że jest ona zauważalna we wszystkich artykułach, które się do niej odwołują²⁴. Obecnie Wikidane obejmują przede wszystkim informacje dotyczące obiektów geograficznych (na przykład dane geolokalizacyjne, nazwy w różnych językach,

²³ Wikidata, [online:] <https://www.wikidata.org/> [dostęp 01.11.2015].

²⁴ Wcześniej różne wersje językowe artykułów połączone były wzajemnymi linkami (interwiki), a zmiany w jednym artykule (np. przy dacie śmierci) nie pociągały za sobą zmiany w innych.

liczba mieszkańców, rok powstania), biografii (lata życia, miejsce urodzenia i śmierci, zawód), taksonów biologicznych i tym podobnych. Informacje z Wikidanych są też szeroko wykorzystywane poza projektami Wikimedia, na przykład są one jednym z podstawowych źródeł, na których opiera się googłowski system Knowledge Graph, wspierający i wzbogacający wyszukiwanie²⁵.

Obok treści innym rodzajem danych są sami edytorzy. Każda edycja jest przypisana do pewnego konta. Anonimowe edycje przypisane są pseudokontu o nazwie IP, spod którego dokonano edycji (wymagają tego kwestie licencyjne). Każdy użytkownik może podejrzeć, jakie edycje wykonano z danego konta. Wielu zalogowanych użytkowników ujawnia pewne dane na swój temat (płeć, wiek, znajomość języków, zainteresowania), a niektórzy ujawniają swoją rzeczywistą tożsamość. Podstawowe informacje metryczkowe są też zbierane przy zakładaniu konta (choć są nieobowiązkowe). Ich wiarygodność jest jednak dość niska. Bardziej pewne dane mogą dotyczyć użytkowników aktywnych w społeczności (np. pojawiających się na spotkaniach) lub członków stowarzyszeń Wikimedia (gdzie wymagane są prawdziwe dane osobowe). Dotyczą one jednak drobnego (choć bardzo aktywnego) wycinka społeczności. Trzeba też ostrożnie podchodzić do

²⁵ N. Shapiro, *Wikidata Meets the Google Knowledge Graph*, [online:] <http://www.searchenginejournal.com/wikidata-meets-google-knowledge-graph/130459/> [dostęp 01.11.2015].

utożsamiania kont edytorów z osobami fizycznymi. Zdąrza się, że z jednego konta edytuje kilka osób. Ponieważ dla anonimowych edycji tworzona jest historia edycji pod numerem danego IP, w wypadku edycji spod zmiennego IP historia edycji może zawierać edycje różnych osób. Edytowanie spod wielu kont (*pacynkowanie*) nie jest zakazane, o ile nie omija to innych zasad.

Istotą oprogramowania wiki są jednak połączenia, a analiza ich struktury jest jednym z zasadniczych kierunków badań. Artykuły łączone są na kilka sposobów i każdy z nich jest łatwy do prześledzenia w badaniach. Podstawowym sposobem jest linkowanie do innych stron w ramach tego samego projektu. W bocznym panelu znajduje się funkcja „Linkujące”, dająca listę haseł, w których wzmiankowane jest dane hasło. Różne wersje językowe artykułów połączone są (jak już wspominałem) przez Wikidane, a ich lista również jest widoczna w pasku bocznym. Możliwe jest też zamieszczanie linków do projektów siostrzanych (np. link do Commons, Wiki-cytatów i Wikiźródeł w przypadku pisarzy) lub innych artykułów w innej wersji językowej (rzadko stosowane), a także linków zewnętrznych (do stron poza projektami Wikimedia). Specyficzną formą odsyłaczy zewnętrznych niebędących linkami są dane bibliograficzne (źródła w artykułach). W dużych i średnich Wikipediach (tysiące aktywnych edytorów, powyżej 500 000 haseł) standardem jest podawanie przypisów w formie szablonu, w którym wyszczególnione są takie pola jak autorzy, tytuł, miejsce

publikacji czy ISBN. Dane te można następnie zaimportować programowo, co pozwala na przykład na określenie, jak często pewne źródła są używane. Trzeba jednak pamiętać, że edytorzy nie zawsze wprowadzają przypisy w postaci szablonu lub wprowadzają informacje niepełne (np. nie podają numeru ISBN), co znacznie utrudnia obróbkę danych lub generuje luki w korpusie. W przyszłości prawdopodobne jest przeniesienie danych o źródłach do Wikidanych, co znacznie ułatwi ich agregację i analizę.

Szczególnym sposobem porządkowania treści jest łączenie ich w kategorie. Każdy artykuł (również poza przestrzenią główną) należy do przynajmniej jednej kategorii. System kategorii ma charakter hierarchiczny, jednak kategorie są kształtowane i przypisywane społecznościowo (co różni je od tradycyjnych taksonomii). W konsekwencji system kategorii jest niespójny, nieuporządkowany, występują w nim nieścisłości i zapętlenia. Mimo wielu niedoskonałości jest alternatywą wobec wyszukiwarki pełnotekstowej i wygodnym sposobem na filtrowanie treści. Na systemie kategorii opiera się wiele narzędzi i badań szczegółowych, które przeszukują treści w obrębie pewnego obszaru Wikipedii (np. biografie prawników czy czeskie utwory literackie).

Połączenia między artykułami pozwalają w dość łatwy sposób tworzyć komputerowo wspomagane modele sieci społecznych. Pewnym wyjątkiem jest odtwarzanie struktury dyskusji między użytkownikami. Dyskusje na

stronach dyskusji haseł, w przestrzeni Wikipedia, Pomoc czy Meta prowadzone są w jednym miejscu, w postaci kolejnych, podpisanych wpisów. Dają się one łatwo odtworzyć na podstawie historii edycji. Specyfiką Wikipedii jest natomiast organizacja dyskusji między użytkownikami (w przestrzeni Dyskusja Wikipedysty). Wpisy nie są umieszczane jeden pod drugim, lecz każdy Wikipedysta wpisuje swoje uwagi na stronie interlokutora. Dyskusja toczy się więc w dwóch miejscach i jej odtworzenie wymaga większego wysiłku ze strony badacza.

Pełne bazy danych Wikipedii (wraz z archiwami) są dostępne dla wszystkich zainteresowanych²⁶. Badacz może sięgnąć po różnego rodzaju narzędzia pomagające w gromadzeniu danych surowych. Należą do nich między innymi Wikipedia Miner Toolkit²⁷, Wikipedia Extractor²⁸ czy Wiki Slurp²⁹. Oprogramowanie MediaWiki ma też własne

²⁶ Wikimedia – Data dumps, [online:] https://meta.wikimedia.org/wiki/Data_dumps [dostęp 01.11.2015]. Informacje o sposobach importu: Wikimedia – Data dumps/Tools for importing, [online:] https://meta.wikimedia.org/wiki/Data_dumps/Tools_for_importing [dostęp 01.11.2015]. Por. również portal Referata, zawierający inne bazy danych oparte na technologii wiki: Referata – Main Page, [online:] http://www.referata.com/wiki/Main_Page [dostęp 01.11.2015].

²⁷ Sourceforge – Wikipedia Miner, [online:] <http://sourceforge.net/projects/wikipedia-miner/> [dostęp 01.11.2015].

²⁸ Medialab – Wikipedia Extractor, [online:] http://medialab.di.unipi.it/wiki/index.php/Wikipedia_Extractor [dostęp 01.11.2015].

²⁹ WikiSlurp, [online:] <http://c2.com/cgi/wiki/wiki?WikiSlurp> [dostęp 01.11.2015].

API, pozwalające na łatwą komunikację z zewnętrznymi aplikacjami (np. pobieranie danych surowych)³⁰.

Obok danych surowych w wielu miejscach można znaleźć dane zagregowane. Poprzez historię edycji każdego z artykułów można dotrzeć do informacji na temat tego, ile razy, kiedy i przez kogo był edytowany i ile osób go obserwuje³¹. Są też dostępne podstawowe dane dotyczące struktury edycji wikipedystów (lista utworzonych artykułów, liczba edycji w poszczególnych przestrzeniach). Bardziej zaawansowane statystyki (opisujące m.in. najczęściej edytowane artykuły czy zawierające wykresy aktywności edycyjnej) wymagają włączenia tej funkcji przez edytora, a w konsekwencji są dostępne tylko w odniesieniu do bardziej zaangażowanych wikipedystów.

Gromadzone są również podstawowe statystyki globalne dotyczące poszczególnych projektów (szczególnie dynamika edycji, liczby edytorów, odwiedzalności stron)³². Narzędzia te rozwijane są przez Fundację Wikimedia, niezależnych edytorów oraz przez różne grupy niezależne³³.

³⁰ MediaWiki – API:Main Page, [online:] https://www.mediawiki.org/wiki/API:Main_page [dostęp 01.11.2015].

³¹ Zakładki: „Lista autorów” i „Statystyki” (Aka-Online, [online:] <http://vs.aka-online.de/cgi-bin/wppagehiststat.pl> [dostęp 01.11.2015]) oraz „Liczba obserwujących”.

³² Wikimedia Statistics, [online:] <http://stats.wikimedia.org/> [dostęp 01.11.2015].

³³ Zob. np. DBpedia, [online:] <http://wiki.dbpedia.org/> [dostęp 01.11.2015].

Ważnym, choć dość mało precyzyjnym, narzędziem jest licznik odwiedzalności strony (Article Traffic Statistics), zliczający liczbę odwiedzin strony w danym odstępie czasowym³⁴. Na jego podstawie tworzone są statystyki popularności haseł.

Choć dane są łatwo dostępne, to ich wykorzystanie wymaga często pracochłonnej obróbki. Dominacja badań ilościowych czy związanych z analizą sieciową jest wynikiem tego, że łatwo je oprzeć na programach komputerowych. W niektórych przypadkach dostępne są dane opracowane już przez innych badaczy. Spis niektórych zebranych korpusów danych dostępnych do analizy można znaleźć na stronie Wikipapers³⁵, na przykład korpus danych dotyczących dyskusji użytkowników w Simple English Wikipedia³⁶. Jest to edycja językowa oparta na uproszczonym korpusie języka angielskiego, przeznaczona dla obcokrajowców i młodych użytkowników. Edycję tę znacznie częściej edytują osoby, dla których język angielski nie jest językiem ojczystym (brak jest jednak dokładnych danych demograficznych). Korpus dyskusji użytkowników był podstawą lingwistycznej

³⁴ *Wikipedia article traffic statistics*, [online:] <http://stats.grok.se/> [dostęp 01.11.2015].

³⁵ Wikipapers – List of datasets, [online:] http://wikipapers.referata.com/wiki/List_of_datasets [dostęp 01.11.2015].

³⁶ *Wikipedia Discussion Corpora*, [online:] www.ukp.tu-darmstadt.de/data/discourse-analysis/wikipedia-discussion-corpora/ [dostęp 01.11.2015].

analizy procesów komunikacyjnych opartych na teorii aktów mowy³⁷.

Większość dostępnych narzędzi została zaprojektowana z myślą o pomocy edytorom i administratorom Wikipedii³⁸. Głównymi narzędziami są tutaj wyszukiwarki. Standardowa wyszukiwarka pełnotekstowa jest bardzo niewygodna³⁹. Bardziej zaawansowanym narzędziem jest Catscan, który łączy wyszukiwanie pełnotekstowe z wyszukiwaniem po drzewie kategorii⁴⁰. Można określić głębokość wyszukiwania oraz takie kryteria, jak wyszukiwana fraza, występujący element kluczowy (taki jak przypis czy grafika), czas powstania i tym podobne. Z kolei Wiki-blame pozwala odnaleźć autora konkretnej frazy i czas jej wprowadzenia⁴¹.

W dalszej części omawiane będą badania, w których Wikipedia może służyć jako narzędzie badawcze lub

³⁷ O. Ferschke, I. Gurevych, Y. Chebotar, *Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages*, w: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Stroudsburg: EACL 2012.

³⁸ Szersza (choć nadal niepełna) lista znajduje się pod adresem: [online] <https://en.wikipedia.org/wiki/Wikipedia:Tools> [dostęp 01.11.2015].

³⁹ Zaawansowane funkcje nie są dostępne w interfejsie, lecz wymagają stosowania składni. Informacje na ten temat znajdują się pod adresem: MediaWiki – Help:CirrusSearch/pl, [online:] <https://www.mediawiki.org/wiki/Help:CirrusSearch/pl> [dostęp 01.11.2015].

⁴⁰ CatScan V3.0, [online:] <https://tools.wmflabs.org/catscan2/catscan2.php> [dostęp 01.11.2015].

⁴¹ WikiBlame, [online:] <http://wikipedia.ramselehof.de/wiki-blame.php> [dostęp 01.11.2015].

źródło danych. Opierać się tu będą na analizie przeprowadzonych już badań. Zainteresowany badacz nie powinien się jednak do nich ograniczać, jako że ilość oferowanych danych i sposobów ich pozyskiwania jest tu przeogromna. Wśród badań zdecydowanie dominują metody ilościowe, zwłaszcza wspomaganą komputerowo analiza statystyczna treści czy historii edycji oraz analiza sieciowa. Rzadsza jest jakościowa analiza treści i analiza dyskursu, chociaż materiał do tych rodzajów badań jest szczególnie bogaty i łatwy w obróbce. Badania jakościowe są nieliczne, często powierzchowne lub skupione na pewnym wąskim wycinku.

Wikipedia będzie miała z pewnością rosnący wpływ na metodologię badań społecznych, nie tylko tych dotyczących jej samej czy szerzej społeczności internetowych. Obok funkcji edukacyjnych (młode pokolenie badaczy z pewnością korzystało z zawartych w niej treści) czy omawianych tutaj funkcji instrumentalnych, konsekwencje rozwoju Wikipedii mogą być nieraz trudne do przewidzenia. Rzeczą wartą odnotowania jest też dość specyficzny (choć prawdopodobnie nie jedyny) przypadek testu Rorschacha. Znaczna rozbudowa tego artykułu na angielskojęzycznej Wikipedii (2009)⁴² wraz z opisem stosowania go w diagnostyce spowodowała protesty ze strony psychologów. Wskazywali oni, że umożliwia to łatwą

⁴² Wikipedia, *Rorschach test*, hasło encyklopedyczne, [online:] https://en.wikipedia.org/wiki/Rorschach_test [dostęp 01.11.2015].

manipulację tymi testami, co powoduje utratę przez test wartości diagnostycznej⁴³. Duży zasób otwartych i dynamicznych danych sprawia, że Wikipedia jest też dobrym poligonem do testowania nowych metod ilościowej analizy danych, algorytmów obliczeniowych, programów raportujących czy analizy statystycznej⁴⁴.

Studium przypadku

Najczęstszym sposobem wykorzystania Wikipedii w badaniach jest traktowanie jej jako studium przypadku i to w różnych postaciach tej techniki badawczej⁴⁵. Najczęściej Wikipedia przyjmowana jest jako samodzielny obiekt analizy (często niestety powierzchownej) będący jedynie ilustracją pewnych ogólniejszych rozważań. Czasami służy również testowaniu teorii, jak na przykład w badaniach Olofa Sundina przyjmującego perspektywę teorii aktora-sieci i społecznych studiów nad nauką (STS) do

⁴³ Podsumowanie dyskusji wraz z odnośnikami do artykułów: *Psychologists gauge impact of Wikipedia's Rorschach test coverage*, w: Wikipedia, *Rorschach test*, hasło encyklopedyczne, [online:] https://meta.wikimedia.org/wiki/Research:Newsletter/2011/December#Psychologists_gauge_impact_of_Wikipedia.27s_Rorschach_test_coverage [dostęp 01.11.2015].

⁴⁴ Np. DBpedia, dz. cyt. Por. też: M. Thij, Y. Volkovich, D. Laniado, A. Kaltenbrunner, *Modeling and predicting page-view dynamics on Wikipedia*, [online:] <http://arxiv.org/abs/1212.5943v2> [dostęp 01.11.2015].

⁴⁵ J. Gerring, *Case Study Research. Principles and Practices*, Cambridge: Cambridge University Press 2007.

analizy wykorzystywania odnośników bibliograficznych i konstrukcji tożsamości aktorów⁴⁶. Najczęściej jednak Wikipedia jest po prostu opisywana w kategoriach danej teorii, co można potraktować jako testowanie jej mocy deskryptywnej. Tak robi Justyna Hofmokl w książce *Internet jako nowe dobro wspólne*, gdzie Wikipedię opisuje się w kategoriach pojęciowych nowego instytucjonalizmu⁴⁷.

Dobrze zrobione studium przypadku jest jednak rzadkością. Grzeszą one bądź nadmierną powierzchownością, bądź też brakiem wskazania ogólniejszego, pozaopisowego znaczenia. Dobrym przykładem licznych studiów przypadku, osadzonych w szerszym projekcie badawczym, jest monograficzna netnografia Wikipedii Dariusza Jemielnika, który obok pogłębionej analizy projektu wskazuje jego znaczenie dla rozwoju nowych form komunikacji i organizacji czy funkcjonowania struktur wiedzy i władzy⁴⁸.

Właśnie tego rodzaju badania nad nowymi formami organizacji czy komunikacji są szczególnie dobrze rozwinięte i mogą mieć duże znaczenie poznawcze. Internet jest wielkim laboratorium, w którym powstają nowe

⁴⁶ O. Sundin, *Janitors of Knowledge: Constructing Knowledge in the Everyday Life of Wikipedia Editors*, „Journal of Documentation” 2011, [online:] <http://www.emeraldinsight.com/doi/abs/10.1108/00220411111164709?journalCode=jd> [dostęp 01.11.2015].

⁴⁷ J. Hofmokl, *Internet jako nowe dobro wspólne*, Warszawa: Wydawnictwa Akademickie i Profesjonalne 2009, s. 166–177.

⁴⁸ D. Jemielnik, *Życie wirtualne dzikich. Netnografia Wikipedii – największego projektu współtworzonego przez ludzi*, Warszawa: Poltext 2013.

formy współpracy i interakcji. Różne rodzaje komunikacji, struktury władzy, dysfunkcje czy role organizacyjne są porównywane przez badaczy zarządzania i teorii organizacji. Wikipedia, choć analizowana indywidualnie, jest wtedy traktowana jako ilustracja pewnych procesów organizacyjnych czy zapowiedź nadchodzących zmian. Jest bowiem traktowana jako sztandarowy przykład społeczności praktyki (*communities of practice*), społeczności pracowników wiedzy (*knowledge workers*) czy społeczności skupionych na zarządzaniu wiedzą (*knowledge management*). Jest typowym przykładem otwartych projektów współpracy (*open collaboration project*), projektów z treściami tworzonymi przez uczestników (*user collaboration projects*) czy szerzej web 2.0.

Badań takich przeprowadzono już bardzo wiele. W Polsce oprócz Dariusza Jemielniaka tematyką nowych struktur władzy i ról społecznych zajmował się też Sebastian Skolik⁴⁹. Analiza Wikipedii służyła między innymi do stworzenia ogólniejszych koncepcji związanych z systemami budowania reputacji w społecznościach *open collaboration*⁵⁰ czy skuteczności przywództwa rozproszonego⁵¹. W wielu

⁴⁹ S. Skolik, *Typologia nieformalnych ról społecznych w Wikipedii*, „Zeszyty Naukowe Politechniki Częstochowskiej” 2011, s. 92–100.

⁵⁰ B. Adler, L. De Alfaro, A. Kulshreshtha, I. Pye, *Reputation systems for open collaboration*, „Communications of the ACM” 2011, nr 54 (8), s. 81.

⁵¹ *Effectiveness of shared leadership in online communities*, w: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, New York: ACM Press 2012.

badaniach widziano w niej zapowiedź kształtujących się form organizacyjnych. Na przykład Jemielniak i Raburski wskazywali, że może być ona traktowana jako nowa forma współpracy w społeczeństwach płynnej nowoczesności, wymykająca się tradycyjnej dychotomii współpraca – konflikt⁵². To konflikt jest siłą napędową i motywacyjną w nowych formach organizacji. Nowe technologie pozwalają ponadto na współpracę bez bezpośrednich interakcji, rozciągniętych w czasie, zatomizowanych działań.

Innym nurtem badań, który może skorzystać z analizy Wikipedii, jest badanie interakcji społecznych. Ogromna ilość szczegółowych danych, a także liczne narzędzia do śledzenia interakcji sprawiają, że jest ona bardzo dobrym materiałem źródłowym zarówno dla analiz ilościowych, jak i jakościowych (nawet tych bardzo wymagających, jak etnometodologia, analiza konwersacji czy teoria ugruntowana). Jedynie nieliczne kanały komunikacyjne pozostają poza zasięgiem badacza (spotkania wikipedystów, zamknięte listy administracyjne, prywatne maile, kanał IRC). Do dyspozycji badaczy znajduje się wiele narzędzi, które mogą ułatwiać analizę. Obok wymienionych już wcześniej warto zwrócić uwagę na Editor Interaction Analyser⁵³, pokazujący listę artykułów

⁵² D. Jemielniak, T. Raburski, *Liquid Collaboration*, w: *Liquid Organization: Zygmunt Bauman and Organization Theory*, red. J. Kociatkiewicz, M. Kostera, London – New York: Routledge 2014, s. 85–103.

⁵³ Editor Interaction Analyser, [online:] <http://tools.wmflabs.org/sigma/editorinteract.py> [dostęp 01.11.2015].

koedytowanych przez wskazanych użytkowników, czy program whoVis⁵⁴, który wizualizuje w postaci sieci społecznej konflikty toczące się wokół artykułów (wycofywanie edycji, przeciągające się spory o kształt haseł i podawane w nich fakty).

Wikipedia jako reprezentacja rzeczywistości społecznej

Wikipedię można potraktować też szerzej jako pewien element rzeczywistości społecznej, i to bardzo istotny. Jest ona jednym z centralnych węzłów internetu, odwiedzanym właściwie przez wszystkich internautów i włączony w infrastrukturę wielu innych centralnych stron (np. Google)⁵⁵. Również rzeczywistość pozainternetowa znajduje swoje odzwierciedlenie w Wikipedii. Życie *online* tworzy struktury, które w pewien sposób odwzorowują czy imitują struktury świata realnego. Tym samym analiza struktur wirtualnych może nam wiele powiedzieć o świecie pozainternetowym.

⁵⁴ *Whovisual: interactive visualizations to explore wiki article development*, [online:] people.aifb.kit.edu/ffl/whovisual/# [dostęp 01.11.2015]. Narzędzie w wersji alfa, obecnie przydatne jedynie do analizy szczególnie spornych tematów, wokół których narosło wiele konfliktów.

⁵⁵ Wikipedia znajduje się w czołówce wielu rankingów najważniejszych portali internetowych, np. *The Moz Top 500*, [online:] <https://moz.com/top500> [dostęp 01.11.2015] – miejsce ósme; *The top 500 sites on the web*, [online:] <http://www.alexa.com/top-sites> [dostęp 01.11.2015] – miejsce siódme.

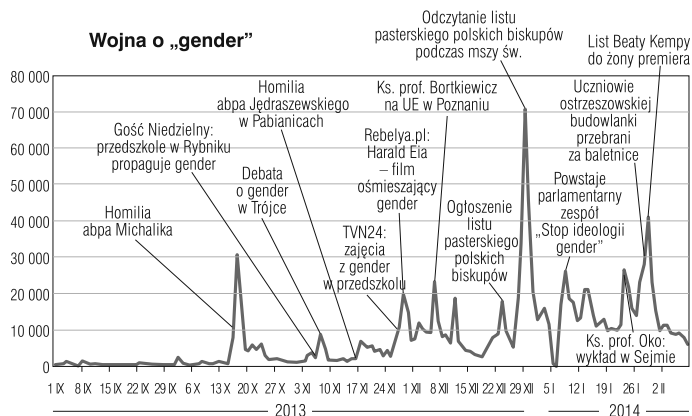
Przykładem odzwierciedlonych struktur społecznych mogą być gęstość zaludnienia, penetracja internetu, centralność pewnych wątków kulturowych czy charakterystyczne cechy różnych dyskursów. Wikipedia może być traktowana jako wskaźnik przemian w ludzkiej aktywności, przede wszystkim zdobywania i użytkowania wiedzy, ale także korzystania z kultury, podróżowania czy sposobów spędzania czasu wolnego.

Warto też wspomnieć, że może być ona narzędziem używanym poza naukami społecznymi, na przykład odwiedzalność stron związanych z chorobami (choćby z grypą) może być dobrym wskaźnikiem rozpoczynającej się epidemii⁵⁶. W podobny sposób można przy użyciu wspomnianego Article Traffic Statistics analizować pobudzenie opinii publicznej w odpowiedzi na pewne realne wydarzenia (por. rycina 3.1.). Taha Yasseri i Jonathan Bright twierdzą nawet, że zmiany w odwiedzalności pewnych stron na Wikipedii mogą być dobrym predyktorem wyników wyborów⁵⁷.

W ostatnich latach wraz z rozwojem mocy obliczeniowej komputerów oraz technik analizy Big Data, a także dzięki temu, że artykuły czy edycje mają swoje oznaczenie

⁵⁶ M. R. Laurent, T. J. Vickers, *Seeking health information online: does Wikipedia matter?*, „Journal of the American Medical Informatics Association: JAMIA” 2009, 16 (4), s. 471–479.

⁵⁷ T. Yasseri, J. Bright, *Predicting elections from online information flows: towards theoretically informed models*, [online:] <http://arxiv.org/abs/1505.01818> [dostęp 01.11.2015].



RYCINA 3.1.

Wykres pokazujący „wojnę o gender” w polskiej przestrzeni publicznej, odzwierciedloną w odstonach artykułu „gender” w polskiej Wikipedii (na podstawie Article Traffic Statistics)

Źródło: opracowanie własne na potrzeby wykładu „Edycyjne wojny religijne i konfliktowe wizje rzeczywistości w Wikipedii” (Pracownia Pytań Granicznych UAM, 29 stycznia 2015 roku).

czasowe i miejscowe (geotagi), zapisane w łatwo przetwarzalnej postaci na Wikidata, możliwe stają się skomplikowane analizy dużych korpusów danych czy ich wizualizacja⁵⁸. Różnorodność badań wykorzystujących dane przestrzenno-czasowe jest ogromna.

Najbardziej rozpowszechnione są analizy i wizualizacje, które (roboczo) nazwę *coverage*. *Coverage* określić

⁵⁸ Przykładowe wizualizacje (często interaktywne) znaleźć można pod adresem: Wikipedia Visualizations, [online:] <http://infodisiac.com/Wikimedia/Visualizations/> [dostęp 01.11.2015].

można jako stopień, w jakim byty realne (np. urzędy gminy, parlamentarzyści) mają odpowiadające im reprezentacje wirtualne (strony internetowe). Może on być uznany za jedną z miar informatyzacji społeczeństwa lub też rozwoju cyberprzestrzeni. Wikipedia gromadzi wiedzę z bardzo różnorodnych obszarów. Tworzenie i rozwój artykułów w różnych obszarach jest nierównomierny. Analiza *coverage* poszczególnych obszarów (szczególnie w ujęciu dynamicznym) pozwala wyciągnąć ważne wnioski co do kondycji społeczeństwa. Takich projektów jest obecnie bardzo dużo, a coraz doskonalsze narzędzia wizualizacji i analizy Big Data pozwalają sądzić, że w najbliższych latach będzie to nadal rozwijająca się gałąź badań.

Popularnym nurtem analizy typu *coverage* jest rzutowanie pewnych struktur wirtualnych na mapę (np. dzięki współpracy z Open Street Map), niekiedy również z uwzględnieniem czynnika czasowego. Wielokrotnie realizowane były badania nad rozkładem geotagowanych artykułów⁵⁹. Badania te pokazują zarówno ogromne możliwości tego typu analiz, jak i konieczność ostrożnego wyciągania wniosków. Przykładowo Wikipedia zawiera dziesiątki tysięcy artykułów na temat zabytków. Jest to jednak lista daleka od kompletności. Pewne obszary są w niej nadreprezentowane, a inne prawie nieobecne.

⁵⁹ Liczne przykłady na stronie Oxford Internet Institute: Information Geographies, [online:] <http://geography.oii.ox.ac.uk> [dostęp 01.11.2015]; Mapping Wikipedia, [online:] http://wikiproject.oii.ox.ac.uk/mapping_wikipedia/ [dostęp 01.11.2015].

Mapa ich rozłożenia zrekonstruowana na podstawie Wikipedii nie odzwierciedla wiernie geografii realnej, lecz jest odbiciem jej społecznego wymiaru – aktywności edytorów⁶⁰. Nadreprezentowane są szeroko znane atrakcje turystyczne, duże miasta, miejsca, do których istnieje łatwy dostęp. Niedoreprezentowane są miejsca o niskiej populacji, oddalone od linii komunikacyjnych, słabo opisane w źródłach turystycznych⁶¹. Każda wersja językowa w oczywisty sposób nadreprezentuje również treści związane z jej obszarem kulturowym. Stąd frwiki posiada znacząco większy *coverage* geograficznych artykułów dotyczących krajów frankofońskich. Różnice *coverage* dobrze obrazują również wykluczenie cyfrowe i kulturowe pewnych obszarów geograficznych.

⁶⁰ M. Graham, *Wiki Space: Palimpsests and the Politics of Exclusion*, w: *Critical Point of View. A Wikipedia Reader*, red. G. Lovink, N. Tkacz, Amsterdam: Institute of Network Cultures 2011, s. 272.

⁶¹ Nie jest tak, jeśli geotagowanie haseł nie przebiega w sposób „oddolny” i niezorganizowany, lecz zostaje przeprowadzone na przykład przez podmioty profesjonalne, służby rządowe czy na podstawie pełnych baz danych obiektów geograficznych. Takie kompletne bazy, jeśli zostają załadowane na Wikidata, powodują, że pewne obszary stają się „anomaliami” w wizualizacji. Przykładem może być „anomalia nepalska” w plwiki, gdzie wszystkie nepalskie miejscowości mają swoje geotagowane hasło, a na wizualizacjach geotagowanych artykułów polskiej Wikipedii, Nepal jest obok Polski najlepiej opracowanym obszarem świata. Jest to efektem pracy bota zaprogramowanego przez Marka Stelmasika, bazującego na danych rządu nepalskiego. Tego typu anomalie są jak na razie rzadkie z powodu niedostępności oficjalnych baz danych miejscowości lub trudności w ich przetwarzaniu.

Przykładowym narzędziem do samodzielnej analizy może być ViziData⁶² prezentujące na mapie świata dane dotyczące biogramów (daty i miejsca narodzin i śmierci) oraz geotagowanych miejsc i przedmiotów. Wartościowym elementem jest możliwość skalowania mapy oraz definiowania przedziału czasowego. Dzięki temu możliwe jest szczegółowe badanie *coverage* w różnych okresach (np. jakie jest nasycenie biogramami różnych epok historycznych)⁶³. Istnieją również narzędzia pozwalające wyciągać dane czasowo-przestrzenne z artykułów (np. biograficznych)⁶⁴. Bardzo wygodnym i ciągle rozbudowywanym narzędziem porównawczym jest wirtualna mapa na Terra Incognita⁶⁵. Z kolei Locations on Wikidata⁶⁶ pozwala sprawdzić położenie na wirtualnej mapie geotagowanych obiektów, na których temat istnieją napisane artykuły.

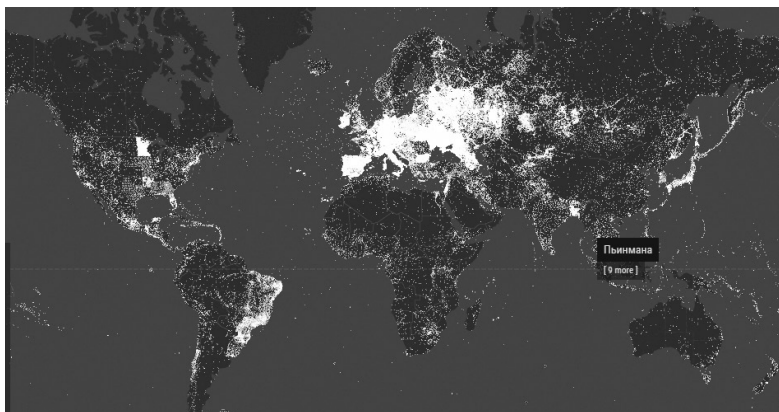
⁶² ViziData, [online:] wwwpub.zih.tu-dresden.de/~s5219191/vizidata/#d=0&m=humans&l=en&f=1&e=1700,2015&g=1&h=1.2&o=1&p=3&x=-1.0546875&y=0.3515602939922709&z=2 [dostęp 1.11.2015].

⁶³ *Time-series of the Distribution of Biographies on Wikipedia over the Last Five Centuries*, [online:] <http://geography.oii.ox.ac.uk/?page=time-series-of-the-distribution-of-biographies-on-wikipedia-over-the-last-five-centuries> [dostęp 01.11.2015].

⁶⁴ E. G. Morton-Owens, *A tool for extracting and indexing spatio-temporal information from biographical articles in Wikipedia*, [online:] https://cs.nyu.edu/web/Research/MsTheses/owens_emily.pdf [dostęp 01.11.2015].

⁶⁵ Terra Incognita, [online:] <http://tracemedia.co.uk/terra/map/> [dostęp 01.11.2015].

⁶⁶ Wikidata locations, [online:] <https://tools.wmflabs.org/wikidata-todo/around.html> [dostęp 01.11.2015].



RYCINA 3.2.

Mapy wygenerowane w projekcie Terra Incognita, porównujące *coverage* w polsko- i rosyjskojęzycznej Wikipedii. Mimo dużego pomniejszenia widoczne są różnice w zainteresowaniu obszarami geograficznymi i pewne anomalie (Nepal w plwiki, stan Minnesota w ruwiki).

Źródło: <http://tracemedia.co.uk/terra/map/> [dostęp 01.11.2015].

Coverage nie dotyczy jednak wyłącznie analizy danych geograficznych. Oczywiście kwestią jest nadreprezentacja artykułów dotyczących współczesnych wydarzeń nad wydarzeniami z poprzednich epok (co wynika zarówno z dostępności źródeł, jak i prezytyzmu opinii publicznej). Rzutując artykuły dotyczące wydarzeń na oś czasu, można badać, które okresy cieszą się szczególnym zainteresowaniem. Inne badania typu *coverage* dotyczące kwestii genderowych (i dyskryminacji dyskursywnej) omówione zostaną w dalszej części artykułu.

Badania nad *coverage* nie są jedynymi formami reprezentacji świata zewnętrznego. W większych Wikipediach, mających szeroką bazę artykułów, wizualizacja połączeń między artykułami (metodologią sieci społecznych) pozwala stwierdzić centralność danego obiektu (biografii, miejsca, idei) w szerszej sieci kulturowej. W ten sposób za pomocą wizualizacji i analizy powiązań pomiędzy artystami na przestrzeni wieków badano na przykład pole historii sztuki⁶⁷. Innym przykładem zastosowania analizy sieci społecznych jest projekt, w którego wyniku powstał model sieci społecznej miejscowości partnerskich na świecie, będący obecnie najbardziej kompletną mapą tego typu⁶⁸.

⁶⁷ D. Goldfarb, D. Merkl, M. Arends, J. Froschauer, *Art History on Wikipedia, a Macroscopic Observation*, „WebSci” 2012.

⁶⁸ A. Kaltenbrunner, P. Aragón, D. Laniado, Y. Volkovich, *Not all paths lead to Rome: Analysing the network of sister cities*, [online:] <http://arxiv.org/abs/1301.6900> [dostęp 01.11.2015].

Obok badań *coverage* czy analizy sieci społecznych, równie istotne są badania dyskursu. Istnieje wiele powodów, dla których Wikipedia może być traktowana jako uprzywilejowany obszar dyskursu sieciowego, czy (szerzej) publicznego. Głównymi są oczywiście jej popularność, centralność w strukturze internetu czy różnorodność edytorów. Jest ona podstawowym źródłem odniesienia dla ogromnej rzeszy społeczeństwa. Nawet dla osób, dla których nie jest źródłem wiedzy, jest pewnym drogowskazem wiedzy czy opinii powszechnych. W ten właśnie sposób zaczyna być traktowana między innymi przez polskie sądownictwo, uznające wikipedyczne artykuły jako „wiedzę powszechnie znaną” (w języku prawniczym „notoryjnie znane fakty”) czy słownik potocznego znaczenia pojęć⁶⁹.

Dyskurs odzwierciedlony w Wikipedii nie jest jednak, o czym trzeba pamiętać, dokładnym odwzorowaniem innych dyskursów. Jest on specyficzny, dlatego musi być traktowany ostrożnie: jako jedno z obliczy dyskursu publicznego lub wskaźnik pewnych zjawisk dyskursywnych. Jednym z czynników bardzo mocno wpływających na jego specyfikę jest forma. Przestrzeń główna Wikipedii ma być pisana jako encyklopedia. Ten specyficzny gatunek literacki bardzo mocno strukturyzuje dynamikę społecznych interakcji poza *mainem* (mają one być służebne wobec

⁶⁹ K. Dąbrowski, S. Grabarczuk, 223 *orzeczenia z Wikipedią. Wykorzystanie internetowej encyklopedii Wikipedia w uzasadnieniach wyroków sądów polskich*, [online:] <https://www.depot.ceon.pl/handle/123456789/2232> [dostęp 01.11.2015].

tworzenia *mainu*), a także prezentowaną w nim wiedzę (filtrowanie, porządkowanie i organizacja treści). Wymienione cechy gatunku literackiego, jakim jest encyklopedia internetowa, powinny być brane pod uwagę przez analityków dyskursu⁷⁰.

Dyskurs wikipedyczny jest ściśle zespolony z innymi dyskursami publicznymi. Centralność Wikipedii powoduje, że właściwie wszystkie publiczne dyskursy chcą znaleźć w nim odzwierciedlenie. Wiedza prezentowana w Wikipedii, pretendując do neutralności, uniwersalności i (w długiej perspektywie) zupełności, preferując scjentyzm czy źródła oficjalne, mocno wspiera dominujące w społeczeństwie struktury wiedzy i władzy. Ponieważ Wikipedia nie opiera się na własnych badaniach i opiniach, lecz ma być jedynie odzwierciedleniem wiedzy już opublikowanej w źródłach, można ją traktować jako odbicie praktyk dyskursywnych społeczeństwa jako takiego. Warty uwagi przykładami wykorzystania treści zawartych w Wikipedii jako materiału do jakościowej analizy dyskursu jest niepublikowana (choć dostępna w internecie) praca magisterska Dušana Miletica *Europe According to English Wikipedia: Open-sourcing the Discourse on Europe*⁷¹. Praca, broniona na

⁷⁰ A. Tereszkiwicz, *Genre Analysis of Online Encyclopedias. The Case of Wikipedia*, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego 2010.

⁷¹ D. Miletic, *Europe According to English Wikipedia. Open-sourcing the Discourse on Europe*, Jagiellonian University – Groningen University 2012 [praca magisterska].

Uniwersytecie Jagiellońskim i Uniwersytecie Groningen, stosuje metodologię Foucaultowską do analizy dyskursu na temat Europy. Autor wskazuje na rozmyte granice pomiędzy różnymi koncepcjami Europy: kulturową, geograficzną, demograficzną, jak również między Europą a Unią Europejską, oraz analizuje wizualne reprezentacje europejskości.

Badania porównawcze

Bardzo szerokie możliwości otwiera Wikipedia przy wszelkiego rodzaju badaniach porównawczych. Już same porównania w ramach jednego projektu mogą być niezwykle interesujące. Dzięki temu, że Wikipedia jest zbiorem połączonych projektów osadzonych w różnych społecznościach i kulturach, łatwo przeprowadzić za jej pomocą badania porównawcze między poszczególnymi społecznościami. Takie szerokie badania międzykulturowe są zwykle bardzo kosztowne i pracochłonne. Koszty te w przypadku Wikipedii są znacznie mniejsze, jako że korzystamy z danych zastanych, powstałych w sposób spontaniczny i łatwych do opracowania.

Najbardziej rozpowszechnione są badania porównawcze treści czy konstrukcji analogicznych artykułów, porównania *coverage* pewnych obszarów czy różnic w interakcjach. Ponieważ różne wersje Wikipedii bazują na odmiennej bazie społecznej, językowej, kulturowej czy bibliograficznej (odmienny korpus źródeł), zakłada się, że

różnice w świecie wirtualnym są odzwierciedleniem różnic w świecie realnym. Trzeba przy tym jednak pamiętać o kilku zastrzeżeniach. Po pierwsze, różnice w konstrukcji artykułów nie muszą być odzwierciedleniem różnic kulturowych, lecz na przykład nieprzeciętną aktywnością grupy wyspecjalizowanych edytorów. Po drugie, bardzo ostrożnie trzeba dobierać materiał porównawczy. Istnieją ogromne różnice pomiędzy największymi projektami (Wikipedią angielsko-, niemiecko- czy francuskojęzyczną), projektami średniej wielkości (np. Wikipedię polsko-, rosyjsko- czy holenderskojęzyczne), a projektami małymi (łotewsko-, śląsko-, baskijskojęzyczne)⁷². O pewnej reprezentacji szerszej populacji można mówić tylko w przypadku dużych i średnich projektów. Ponadto różnice w rozwoju pewnych artykułów (czy w *coverage*) wynikać mogą z różnic demograficznych (liczba edytorów i osób mówiących danym językiem), a nie z różnic kulturowych. Po trzecie, porównywane populacje są niezbyt ostro zdefiniowane i odpowiadają raczej użytkownikom danego języka niż etnosom czy kulturom. W rozbudowie wersji anglojęzycznej uczestniczą przedstawiciele wielu kultur, dla których często angielski nie jest językiem ojczystym.

⁷² Wielkość nie oznacza przy tym liczby artykułów, lecz rozmiar aktywnej społeczności (co implikuje możliwość specjalizacji, liczbę ekspertów czy aktywność pozaedycyjną społeczności). Wikipedię w języku szwedzkim czy waray liczą ponad milion artykułów, w większości składają się jednak na nie artykuły tworzone przez boty opierające się na bazach danych (np. gatunków, planetoid, wiosek), a społeczności są niewielkie.

Mimo tych zastrzeżeń Wikipedia jest bardzo przydatna do śledzenia różnic międzykulturowych, a są one zauważalne nawet w drobiazgach. Na przykład Wikipedie tworzone przez społeczności w dużej mierze muzułmańskie używają odmiennego stylu, gdy piszą o kwestiach religijnych, na przykład tam, gdzie występuje imię Mahometa, towarzyszy mu zawsze zwyczajowa i stylizowana formuła „pokój niech będzie z nim”⁷³.

Dobrym przykładem są badania, w których analiza porównawcza projektów Wikimedia łączona jest z wynikami innych badań. W ten sposób badano za pomocą analizy sieciowej wzorce komunikacyjne występujące w różnych wersjach językowych⁷⁴. Wzorce komunikacji w społecznościach angielskojęzycznej i fińskojęzycznej okazały się bardzo egalitarne, natomiast w społeczności japońskojęzycznej bardzo hierarchiczne (co zgadza się z innymi badaniami porównującymi organizacje społeczne w tych kulturach). Społeczności niemiecko- i koreańskojęzyczne znajdowały się pomiędzy tymi biegunami.

Szczególnie rozwinięte są porównawcze badania nad biografiami. Ewa S. Callahan i Susan C. Herring za

⁷³ Por. Wikipedia, Кыргыз, hasło encyklopedyczne, [online:] <https://kk.wikipedia.org/wiki/%D2%9A%D2%B1%D1%80%D0%B0%D0%BD> [dostęp 01.11.2015].

⁷⁴ K. Nemoto, P. A. Gloor, *Analyzing Cultural Differences in Collaborative Innovation Networks by Analyzing Editing Behavior in Different-Language Wikipedias*, „Procedia – Social and Behavioral Sciences” [online:] http://www.ickn.org/documents/COINS2010_Nemoto_Gloor.pdf [dostęp 01.11.2015].

pomocą jakościowej i ilościowej analizy treści badali różnice kulturowe w opisie słynnych osób w anglo- i polskojęzycznej Wikipedii⁷⁵. Analiza strukturalna próbki artykułów biograficznych, biorąc pod uwagę różnice w potencjałach obu społeczności, ujawniła przede wszystkim skupienie się społeczności na przedstawicielach własnego kręgu kulturowego (choć jak wskazano, anglojęzyczne wpisy o Polakach prawdopodobnie były w dużej mierze pisane przez Polaków). Oprócz dość oczywistych dużych różnic w długości i szczegółowości biografii w polskojęzycznej Wikipedii widać większe skupienie na suchych danych biograficznych, w szczególności związanych z karierą zawodową i rodziną, znacznie rzadsze wspomnianie o kontrowersjach czy małą uwagę przypisywaną do kwestii narodowych (pochodzenie etniczne).

Analiza sieciowa była stosowana do badania relatywnej ważności postaci w danym języku⁷⁶. Korzystając z różnych miar centralności artykułów (PageRank, 2DRank, CheiRank), badacze starali się porównać centralność

⁷⁵ E. S. Callahan, S. C. Herring, *Cultural Bias in Wikipedia Content on Famous Persons*, „Journal of the American Society for Information Science and Technology” 2011.

⁷⁶ Y. H. Eom, D. L. Shepelyansky, *Highlighting Entanglement of Cultures via Ranking of Multilingual Wikipedia Articles*, „PLOS ONE” 2013, nr 8 (10); P. Aragón, A. Kaltenbrunner, D. Laniado, Y. Volkovich, *Biographical Social Networks on Wikipedia – A cross-cultural study of links that made history*, Barcelona: Barcelona Media Foundation 2012.

postaci w danej sieci językowej. Badania te wskazują, że większość biogramów istotnych postaci jest związana z lokalnym (dla danego języka) kręgiem kulturowym. Nieliczne są natomiast biogramy ważne transjęzykowo, będące pomostami pomiędzy różnymi kręgami kulturowymi. Szczegółowa analiza pozwoliła stworzyć listę takich kluczowych postaci, wskazać ich wzajemne relacje, rozmieszczenie geograficzne i chronologiczne, strukturę płciową i tematyczną⁷⁷.

Bardzo interesujące są też badania nad uniwersalnością wiedzy. Brent Hecht i Darren Gergle podważyli tezę, że istotna encyklopedycznie wiedza jest w miarę jednolita w różnych kulturach⁷⁸. Porównując zawartość dwadzieścia pięć wersji językowych, stwierdzili, że drobny ułamek ich zawartości składa się na wspólną wiedzę. Siedemdziesiąt cztery procent artykułów zawartych było jedynie w jednej wersji językowej, a dziewięćdziesiąt pięć procent w mniej niż siedmiu wersjach. Ta duża lokalność wiedzy może w dużej mierze wynikać z odmiennych form organizacji treści. Należy też brać pod uwagę to, że różne wersje językowe rozwijają się w odmiennym tempie, nie

⁷⁷ Y.-H. Eom, P. Aragón, D. Laniado, D. Kaltenbrunner, S. Vigna, D. L. Shepelyansky, *Interactions of cultures and top people of Wikipedia from ranking of 24 language editions*, „PLoS ONE” 2015, nr 10 (3).

⁷⁸ B. Hecht, D. Gergle, *The Tower of Babel meets Web 2.0: User-generated content and its applications in a multilingual context*, w: *ACM Proceedings of Conference on Human Factors in Computing Systems (CHI 2010)*, New York: ACM Press 2010.

można więc z ich aktualnego stanu wnioskować o rzeczywistej strukturze wiedzy w danej kulturze. W 2012 roku (a więc już po badaniach Hechta i Gergela) powstał projekt siostrzany Wikidata, którego celem jest integracja poszczególnych wersji językowych, co będzie sprzyjało konwergencji struktur wiedzy.

Narzędzia do systematycznych porównań międzyjęzykowych są na razie nieliczne i wąsko wyspecjalizowane. Stosunkowo prostym narzędziem jest Wikirank⁷⁹, porównujący wersje artykułu w językach białoruskim, niemieckim, angielskim, polskim, rosyjskim i ukraińskim pod względem ich jakości (definiowanej jako ich długość, liczba i długość przypisów, liczba nagłówków oraz liczba ilustracji).

Od 2012 roku zespół CollabLab rozwija projekt Omnipedia, który pozwolić ma na porównawczą analizę treści artykułów w dwudziestu pięciu wersjach językowych⁸⁰. Narzędzie jest nadal w fazie tworzenia. Ma ono pozwalać na porównanie różnorodności treści i sposobów jej prezentacji i powiązania w różnych Wikipediach. Dotychczas opublikowano kilka artykułów na podstawie jego wykorzystania (m.in. analiza haseł „teoria spiskowa”, „Siemens”, „Northwestern University”). Badania te pokazują wyraźnie znacznie większą różnorodność struktur

⁷⁹ Wikirank, [online:] <http://wikirank.net/> [dostęp 01.11.2015].

⁸⁰ [Online:] <http://omnipedia.northwestern.edu/> [dostęp 01.11.2015].

wiedzy niż dotychczas przypuszczano, podważając tezy o globalnym konsensie wiedzy⁸¹. Projekt ten jest wart uwagi również z tego względu, że rozwija metodologię ilościowego porównywania struktur wiedzy, która może być wykorzystywana również poza Wikipedią. Narzędzie nie jest jeszcze dostępne publicznie, choć taki jest ostateczny cel projektu. Badacze zachęcają jednak do kontaktu i współpracy inne zespoły badawcze.

Inne rodzaje badań

Wikipedia jest doskonałym środowiskiem do przeprowadzania badań uczestniczących, na przykład eksperymentu, badań interwencyjnych⁸² czy prowadzonych w paradygmacie teoriokrytycznym⁸³.

Ogromna liczba i różnorodność uczestników projektu, łatwość dostępu do danych wydają się zachęcać do

⁸¹ P. Bao, B. Hecht, S. Carton, M. Quaderi, M. Horn, D. Gergle, *Omnipedia: Bridging the Wikipedia Language Gap*, [online:] <http://collablab.northwestern.edu/pubs/CHI2012-BaoEtAl-Omnipedia.pdf> [dostęp 01.11.2015].

⁸² S. Kemmis, R. McTaggart, *Uczestniczące badania interwencyjne. Działanie komunikacyjne i sfera publiczna*, w: *Metody badań jakościowych*, t. 1, red. N. K. Denzin, Y. S. Lincoln, przeł. K. Podemski, Warszawa: Wydawnictwo Naukowe PWN 2009, s. 775–831.

⁸³ J. L. Kincheloe, P. McLaren, *Teoria krytyczna i badania jakościowe. Rewizja*, w: *Metody badań jakościowych*, t. 1, red. N. K. Denzin, Y. S. Lincoln, przeł. K. Podemski, Warszawa: Wydawnictwo Naukowe PWN 2009, s. 431–485.

eksperymentów internetowych⁸⁴. Technika ta napotyka jednak na dość istotne ograniczenia etyczne. W niektórych projektach przeprowadzono eksperymenty mające postać ukrytej obserwacji uczestniczącej. Testowano między innymi sposoby rozwiązywania konfliktów (poprzez wywoływanie tych konfliktów), wykrywania wcześniej wprowadzonych przez „badacza” nieprawdziwych informacji (*hoaxów*) czy traktowania nowicjuszy przez społeczność (przez tworzenie fikcyjnych kont). Po wykryciu takich „eksperymentów” spotkały się one z bardzo negatywną reakcją społeczności, budząc atmosferę nieufności. Wiele z tych eksperymentów było przeprowadzanych przez osoby niekompetentne, nie miały należytego uzasadnienia badawczego, a tym samym były niezgodne z etyką badań.

Odnotować też warto wykorzystanie Wikipedii jako narzędzia w eksperymentach edukacyjnych. Badano na przykład, w jaki sposób włączenie w proces edukacji edytowania Wikipedii wpływa na aktywność uczniów i przyswajanie wiedzy⁸⁵. Przykładem ewolucji w podejściu do Wikipedii jako wskaźnika (w tym wypadku w badaniach kompetencji komunikacyjnych młodzieży) są dwie edycje

⁸⁴ P. Siuda, *Eksperyment w internecie – nowa metoda badań w naukach społecznych*, „Studia Medioznawcze” 2009, t. 3, nr 38, s. 152–168.

⁸⁵ P. Mareca, V.A. Bosch, *Editing the Wikipedia: Its role in science education*, w: *6th Iberian Conference on Information Systems and Technologies (CISTI)*, Madrid: ETSI de Telecomun 2011.

raportu *Dzieci sieci* (z 2012 i 2013)⁸⁶. W pierwszej edycji sięgnięcie po Wikipedię przez uczniów świadczyć miało o pewnych podstawowych (i to niezbyt wysokich) kompetencjach w wyszukiwaniu wiedzy. W drugiej edycji badacze traktowali Wikipedię w sposób bardziej zróżnicowany. Była ona uznawana za ważną część infosfery i samo korzystanie z niej nie było istotnym wskaźnikiem. Ocena kompetencji polegała na umiejętności wyszukania, oceny wiarygodności i wykorzystania informacji, niezależnie od ich pochodzenia.

W badaniach interwencyjnych obiekt badawczy jest jednocześnie celem reformistycznego przekształcenia rzeczywistości społecznej. W Wikipedii tego typu badania były przeprowadzane w kontekście problemu *gender gap*, czyli dysproporcji genderowych występujących na wielu poziomach (liczba edytorów, liczba artykułów związanych z daną płcią, ich jakość, ustrukturyzowanie i miejsce w sieci)⁸⁷. Problem ten był wielokrotnie podnoszony w społeczności Wikipedystów, jak i przez Fundację Wikimedia, stał się też przedmiotem wielu studiów oraz projektów zaradczych. Jednym z przykładów takiego

⁸⁶ Raporty: *Dzieci sieci. Kompetencje komunikacyjne najmłodszych*, red. P. Siuda, G. D. Stunża, Gdańsk: Instytut Kultury Miejskiej 2012; P. Siuda, G. D. Stunża, A. J. Dąbrowska, M. Klimowicz, E. Kulczycki, R. Piotrowska, E. Rozkosz, M. Sieńko, K. Stachura, *Dzieci sieci 2.0. Kompetencje komunikacyjne najmłodszych*, Gdańsk: Instytut Kultury Miejskiej 2013.

⁸⁷ Wikimedia, *Gender gap*, hasło słownikowe, [online:] https://meta.wikimedia.org/wiki/Gender_gap [dostęp 01.11.2015].

projektu był Passing On⁸⁸, zbierający nekrologi znanych osób, które ukazały się w New York Timesie, dokonujący ich analizy ze względu na proporcje płci (zarówno w NYT, jak i następnie ich odbicia w Wikipedii), oraz oferujący narzędzia do uzupełnienia brakujących biogramów znanych kobiet w Wikipedii.

Konkluzje

Zaprezentowane tu narzędzia i badania są jedynie drobnym fragmentem dynamicznie rozwijającego się pola badawczego. Coraz częściej Wikipedia nie jest traktowana jako jedno z wielu miejsc w przestrzeni wirtualnej, lecz jako miejsce w pewien sposób centralne, w którym schodzą się różnorodne wątki interakcji społecznych, struktur znaczeń czy wzorców zachowań. Ogromna liczba danych, ich dostępność i przejrzystość znacząco ułatwiają wykorzystanie Wikipedii w badaniach. Można więc się spodziewać jej rosnącego znaczenia dla metodologii badań społecznych w przyszłości.

BIBLIOGRAFIA

Adler B., De Alfaro L., Kulshreshtha A., Pye I., *Reputation systems for open collaboration*, „Communications of the ACM” 2011, nr 54 (8).

⁸⁸ Passington, [online:] <http://passington.natematias.com/> [dostęp 01.11.2015].

- Aragón P., Kaltenbrunner A., Laniado D., Volkovich Y., *Biographical Social Networks on Wikipedia – A cross-cultural study of links that made history*, Barcelona: Barcelona Media Foundation 2012.
- Bao P., Hecht B., Carton S., Quaderi M., Horn M., Gergle D., *Omnipedia: Bridging the Wikipedia Language Gap*, [online:] <http://collablab.northwestern.edu/pubs/CHI2012-BaoEtAl-Omnipedia.pdf> [dostęp 01.11.2015].
- Callahan E. S., Herring S. C., *Cultural Bias in Wikipedia Content on Famous Persons*, „Journal of the American Society for Information Science and Technology” 2011, Vol. 62, Is. 10.
- CatScan V3.0, [online:] <https://tools.wmflabs.org/catscan2/catscan2.php> [dostęp 01.11.2015].
- Dąbrowski K., Grabarczuk S., *223 orzeczenia z Wikipedią. Wykorzystanie internetowej encyklopedii Wikipedia w uzasadnieniach wyroków sądów polskich*, [online:] <https://www.depot.ceon.pl/handle/123456789/2232> [dostęp 01.11.2015].
- DBpedia, [online:] <http://wiki.dbpedia.org/> [dostęp 01.11.2015].
- Dzieci sieci. Kompetencje komunikacyjne najmłodszych*, red. P. Siuda, G. D. Stunża, Gdańsk: Instytut Kultury Miejskiej 2012.
- Editor Interaction Analyser, [online:] <http://tools.wmflabs.org/sigma/editorinteract.py> [dostęp 01.11.2015].
- Eom Y. H., Shepelyansky D. L., *Highlighting Entanglement of Cultures via Ranking of Multilingual Wikipedia Articles*, „PLOS ONE” 2013, No. 8 (10).
- Eom Y.-H., Aragón P., Laniado D., Kaltenbrunner D., Vigna S., Shepelyansky D. L., *Interactions of cultures and top people of Wikipedia from ranking of 24 language editions*, „PLOS ONE” 2015, nr 10 (3).
- Fandom, [online:] <http://www.wikia.com/Wikia> [dostęp 01.11.2015].

- Ferschke O., Gurevych I., Chebotar Y., *Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages*, w: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Stroudsburg: EACL 2012.
- Gerring J., *Case Study Research. Principles and Practices*, Cambridge: Cambridge University Press 2007.
- Goldfarb D., Merkl D., Arends M., Froschauer J., *Art History on Wikipedia, a Macroscopic Observation*, „WebSci” 2012.
- Graham M., *Wiki Space: Palimpsests and the Politics of Exclusion*, w: *Critical Point of View. A Wikipedia Reader*, red. G. Lovink, N. Tkacz, Amsterdam: Institute of Network Cultures 2011.
- Hecht B., Gergle D., *The Tower of Babel meets Web 2.0: User-generated content and its applications in a multilingual context*, w: *ACM Proceedings of Conference on Human Factors in Computing Systems (CHI 2010)*, New York: ACM Press 2010.
- Hofmokr J., *Internet jako nowe dobro wspólne*, Warszawa: Wydawnictwa Akademickie i Profesjonalne 2009.
- Information Geographies, [online:] <http://geography.oii.ox.ac.uk> [dostęp 01.11.2015].
- Jankowski S. J., *Wikipedia and Encyclopaedism: A genre analysis of epistemological values*, [online:] <https://www.ruor.uottawa.ca/handle/10393/24160> [dostęp 01.11.2015].
- Jemielniak D., Raburski T., *Liquid Collaboration*, w: *Liquid Organization: Zygmunt Bauman and Organization Theory*, red. J. Kociatkiewicz, M. Kostera, London – New York: Routledge 2014.
- Jemielniak D., *Życie wirtualne dzikich. Netnografia Wikipedii – największego projektu współtworzonego przez ludzi*, Warszawa: Poltext 2013.
- Kaltenbrunner A., Aragón P., Laniado D., Volkovich Y., *Not all paths lead to Rome: Analysing the network of sister cities*, [online:] <http://arxiv.org/abs/1301.6900> [dostęp 01.11.2015].

- Kemmis S., McTaggart R., *Uczestniczące badania interwencyjne. Działanie komunikacyjne i sfera publiczna*, w: *Metody badań jakościowych*, t. 1, red. N. K. Denzin, Y. S. Lincoln, przeł. K. Podemski, Warszawa: Wydawnictwo Naukowe PWN 2009.
- Kincheloe J. L., McLaren P., *Teoria krytyczna i badania jakościowe. Rewizja*, w: *Metody badań jakościowych*, t. 1, red. N. K. Denzin, Y. S. Lincoln, przeł. K. Podemski, Warszawa: Wydawnictwo Naukowe PWN 2009.
- Laurent M. R., Vickers T. J., *Seeking health information online: does Wikipedia matter?*, „Journal of the American Medical Informatics Association: JAMIA” 2009, 16 (4).
- Mapping Wikipedia, [online:] http://wikiproject.oii.ox.ac.uk/mapping_wikipedia/ [dostęp 01.11.2015].
- Mareca P., Bosch V. A., *Editing the Wikipedia: Its role in science education*, w: *6th Iberian Conference on Information Systems and Technologies (CISTI)*, Madrid: ETSI de Telecomun 2011.
- Medialab – Wikipedia Extractor, [online:] http://medialab.di.unipi.it/wiki/index.php/Wikipedia_Extractor [dostęp 01.11.2015].
- MediaWiki – API:Main Page, [online:] https://www.mediawiki.org/wiki/API:Main_page [dostęp 01.11.2015].
- MediaWiki – Help:CirrusSearch/pl, [online:] <https://www.mediawiki.org/wiki/Help:CirrusSearch/pl> [dostęp 01.11.2015].
- MediaWiki – WikiMedia, [online:] <https://www.mediawiki.org/wiki/MediaWiki> [dostęp 01.11.2015].
- Miletic D., *Europe According to English Wikipedia. Open-sourcing the Discourse on Europe*, Jagiellonian University – Groningen University 2012 [praca magisterska].
- Morton-Owens E. G., , *A tool for extracting and indexing spatio-temporal information from biographical articles in Wikipedia*, [online:] https://cs.nyu.edu/web/Research/MsTheses/owens_emily.pdf [dostęp 01.11.2015].

- Nemoto K., Gloor P.A., *Analyzing Cultural Differences in Collaborative Innovation Networks by Analyzing Editing Behavior in Different-Language Wikipedias*, „Procedia – Social and Behavioral Sciences” [online:] http://www.ickn.org/documents/COINS2010_Nemoto_Gloor.pdf [dostęp 01.11.2015].
- Nielsen F. Å., *Wikipedia research and tools: Review and comments*, [online:] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2129874 [dostęp 01.11.2015].
- Passingon, [online:] <http://passingon.natematias.com/> [dostęp 01.11.2015].
- Psychologists gauge impact of Wikipedia’s Rorschach test coverage*, w: Wikipedia, *Rorschach test*, hasło encyklopedyczne, [online:] https://meta.wikimedia.org/wiki/Research:Newsletter/2011/December#Psychologists_gauge_impact_of_Wikipedia.27s_Rorschach_test_coverage [dostęp 01.11.2015].
- Referata – Main Page, [online:] http://www.referata.com/wiki/Main_Page [dostęp 01.11.2015].
- Shapiro N., *Wikidata Meets the Google Knowledge Graph*, [online:] <http://www.searchenginejournal.com/wikidata-meets-google-knowledge-graph/130459/> [dostęp 01.11.2015].
- Siuda P., *Eksperyment w internecie – nowa metoda badań w naukach społecznych*, „Studia Medioznawcze” 2009, t. 3, nr 38.
- Siuda P., Stunża G. D., Dąbrowska A. J., Klimowicz M., Kulczycki E., Piotrowska R., Rozkosz E., Sieńko M., Stachura K., *Dzieci sieci 2.0. Kompetencje komunikacyjne najmłodszych*, Gdańsk: Instytut Kultury Miejskiej 2013.
- Skolik S., *Typologia nieformalnych ról społecznych w Wikipedii*, „Zeszyty Naukowe Politechniki Częstochowskiej” 2011.
- Sourceforge – *Wikipedia Miner*, [online:] <http://sourceforge.net/projects/wikipedia-miner/> [dostęp 01.11.2015].

- Sundin O., *Janitors of Knowledge: Constructing Knowledge in the Everyday Life of Wikipedia Editors*, „Journal of Documentation” 2011, [online:] <http://www.emeraldinsight.com/doi/abs/10.1108/00220411111164709?journalCode=jd> [dostęp 01.11.2015].
- Tereszkiewicz A., *Genre Analysis of Online Encyclopedias. The Case of Wikipedia*, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego 2010.
- Terra Incognita, [online:] <http://tracemedia.co.uk/terra/map/> [dostęp 01.11.2015].
- The Moz Top 500, [online:] <https://moz.com/top500> [dostęp 01.11.2015].
- Thij M., Volkovich Y., Laniado D., Kaltenbrunner A., *Modeling and predicting page-view dynamics on Wikipedia*, [online:] <http://arxiv.org/abs/1212.5943v2> [dostęp 01.11.2015].
- Time-series of the Distribution of Biographies on Wikipedia over the Last Five Centuries*, [online:] <http://geography.oii.ox.ac.uk/?page=time-series-of-the-distribution-of-biographies-on-wikipedia-over-the-last-five-centuries> [dostęp 01.11.2015].
- ViziData, [online:] wwwpub.zih.tu-dresden.de/~s5219191/vizidata/#d=0&m=humans&l=en&f=1&e=1700,2015&g=1&h=1.2&o=1&p=3&x=-1.0546875&y=0.3515602939922709&z=2 [dostęp 1.11.2015].
- Whovisual: interactive visualizations to explore wiki article development, [online:] people.aifb.kit.edu/ffl//whovisual/# [dostęp 01.11.2015].
- WikiBlame, [online:] <http://wikipedia.ramselehof.de/wiki-blame.php> [dostęp 01.11.2015].
- Wikidata locations, [online:] <https://tools.wmflabs.org/wikidata-todo/around.html> [dostęp 01.11.2015].

- Wikidata, [online:] <https://www.wikidata.org/> [dostęp 01.11.2015].
- Wikimedia – Data dumps, [online:] https://meta.wikimedia.org/wiki/Data_dumps [dostęp 01.11.2015].
- Wikimedia – Data dumps/Tools for importing, [online:] https://meta.wikimedia.org/wiki/Data_dumps/Tools_for_importing [dostęp 01.11.2015].
- Wikimedia Foundation – Dario Taraborelli, użytkownik, [online:] <https://wikimediafoundation.org/wiki/User:Dtaraborelli> [dostęp 01.11.2015].
- Wikimedia Foundation – Home, [online:] <https://wikimediafoundation.org/wiki/Home> [dostęp 01.11.2015].
- Wikimedia Meta-Wiki – Newsletter, [online:] <https://meta.wikimedia.org/wiki/Research:Newsletter> [dostęp 01.11.2015].
- Wikimedia Polska, [online:] https://pl.wikimedia.org/wiki/Strona_g%C5%82%C3%B3wna [dostęp 01.11.2015].
- Wikimedia Statistics, [online:] <http://stats.wikimedia.org/> [dostęp 01.11.2015].
- Wikimedia, *Gender gap*, hasło słownikowe, [online:] https://meta.wikimedia.org/wiki/Gender_gap [dostęp 01.11.2015].
- Wikipapers – List of datasets, [online:] http://wikipapers.referata.com/wiki/List_of_datasets [dostęp 01.11.2015].
- Wikipapers – Main Page, [online:] http://wikipapers.referata.com/wiki/Main_Page [dostęp 01.11.2015].
- Wikipedia – Pomoc:Kopiowanie poza Wikipedię, [online:] https://pl.wikipedia.org/wiki/Pomoc:Kopiowanie_poza_Wikipedi%C4%99 [dostęp 01.11.2015].
- Wikipedia – Signpost, [online:] https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2011-04-11/Recent_research [dostęp 01.11.2015].

- Wikipedia – Wikipedia w publikacjach naukowych, [online:] https://pl.wikipedia.org/wiki/Wikipedia:Wikipedia_w_publicacjach_naukowych [dostęp 01.11.2015].
- Wikipedia – Wikipedistik, [online:] <https://de.wikipedia.org/wiki/Wikipedia:Wikipedistik> [dostęp 01.11.2015].
- Wikipedia article traffic statistics*, [online:] <http://stats.grok.se/> [dostęp 01.11.2015].
- Wikipedia Discussion Corpora*, [online:] www.ukp.tu-darmstadt.de/data/discourse-analysis/wikipedia-discussion-corpora/ [dostęp 01.11.2015].
- Wikipedia Foundation – Our Projects, [online:] https://wikimediafoundation.org/wiki/Our_projects [dostęp 01.11.2015].
- Wikipedia Meta-Wiki – Index, [online:] <https://meta.wikimedia.org/wiki/Research:Index> [dostęp 01.11.2015].
- Wikipedia Visualizations, [online:] <http://infodisiac.com/Wikimedia/Visualizations/> [dostęp 01.11.2015].
- Wikipedia, *Rorschach test*, hasło encyklopedyczne, [online:] https://en.wikipedia.org/wiki/Rorschach_test [dostęp 01.11.2015].
- Wikipedia, Кыргызан, hasło encyklopedyczne, [online:] <https://kk.wikipedia.org/wiki/%D2%9A%D2%B1%D1%80%D0%B0%D0%BD> [dostęp 01.11.2015].
- Wikirank, [online:] <http://wikirank.net/> [dostęp 01.11.2015].
- WikiResearch na Twitterze, użytkownik, [online:] <https://twitter.com/wikiresearch> [dostęp 01.11.2015].
- WikiSlurp, [online:] <http://c2.com/cgi/wiki/wiki?WikiSlurp> [dostęp 01.11.2015].
- Yasseri T., Bright J., *Predicting elections from online information flows: towards theoretically informed models*, [online:] <http://arxiv.org/abs/1505.01818> [dostęp 01.11.2015].

Zhu H., Kraut R., Kittur A., *Effectiveness of shared leadership in online communities*, w: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work – CSCW '12*, New York: ACM Press 2012.