

Krzysztof Janc

Uniwersytet Wrocławski
Instytut Geografii i Rozwoju Regionalnego
e-mail: krzysztof.janc@uwr.edu.pl

Wyszukiwarka internetowa w poznaniu przestrzeni cyfrowej

Zarys treści: Obserwujemy obecnie wzrost znaczenia przestrzeni cyfrowej – coraz więcej interakcji pomiędzy ludźmi, ich aktywności przenosi się w całości lub częściowo do tej przestrzeni. W rezultacie w nauce, w tym w geografii, pojawiły się nowe problemy oraz powstają i rozwijane są nowe sposoby badań. Uwzględniając to celem opracowania jest przybliżenie sposobów eksploracji sieci WWW w odniesieniu do wyszukiwarek internetowych. Wyszukiwarki są bowiem jednym z popularniejszych i łatwych do wykorzystania narzędzi eksploracji przestrzeni cyfrowej. Przedstawiono dwa sposoby pozyskania danych, a w konsekwencji analiz: manualne wpisywanie zapytań lub korzystanie ze skryptów oraz korzystanie ze specjalnych usług dostarczanych przez operatorów wyszukiwarki. Podjęto również rozważania dotyczące zalet i wad prezentowanego podejścia do analizy przestrzeni cyfrowej.

Słowa kluczowe: przestrzeń cyfrowa, wyszukiwarki internetowe, geografia

Wprowadzenie

Współcześnie w dużej liczbie codziennych aktywności ludzie wykorzystują urządzenia cyfrowe. Przyspieszają one wymianę informacji, ułatwiają dostęp oraz umożliwiają ich wytwarzanie. W celu zwiększenia użyteczności urządzenia cyfrowe zazwyczaj podłączone są do Internetu. Mogą być zarówno używane świadomie, jak i automatycznie rejestrować zachowania ludzi, parametry środowiska, w którym funkcjonują, stany urządzeń itp. Korzystanie z urządzeń cyfrowych tworzy przestrzeń cyfrową. Jest ona obecnie głównie efektem działań ludzi, którzy korzystają z urządzeń podłączonych do Internetu, wpisują zapytania do wyszukiwarek, odwiedzają strony internetowe, umieszczają posty na portalach społecznościowych, dokonują zakupów w Internecie, oglądają filmy itp. Tworzą w ten sposób cyfrową reprezentację świata rzeczywistego.

Ponieważ coraz więcej interakcji pomiędzy ludźmi oraz innych aktywności przenosi się w całości lub częściowo do przestrzeni cyfrowej, w nauce pojawiły

się nowe obszary badawcze. Część z nich zaczęła się krystalizować już u progu rozwoju Internetu oraz sieci WWW. Kwestie obecności przestrzeni cyfrowej w naszym życiu, przenikania świata rzeczywistego z wykreowanym przez korzystanie z urządzeń cyfrowych, można przedstawić poprzez odwołanie do światów wykreowanych w dwóch filmach (z końca XX w.) poruszających kwestie relacji człowiek–otoczenie. Pierwszym z nich jest „Matrix” (premiera w 1999 r.). Matrix to wszechobecne otoczenie człowieka, w którym żyje¹, przenikające każdą sferę działania, będące jednocześnie fikcją – cyfrowym światem. W codziennym życiu przestrzeń cyfrowa staje się dzisiaj równie wszechobecna jak filmowy Matrix. Jej poznanie, ustalenie praw nią rządzących jest kluczem do zrozumienia otoczenia. Drugim filmem, odwołującym się do funkcjonowania człowieka w „sztucznym” środowisku, jest „Truman Show” (premiera w 1998 r.). W odniesieniu do uniwersum zaprezentowanego w tym filmie, Keen (2015) zauważa, że współcześnie żyjemy w cyfrowej erze wielkiego ekshibicjonizmu. Każdy użytkownik Internetu wytwarza dane o praktycznie wszystkich aspektach swojego życia – dzieli się nimi z innymi użytkownikami. Z perspektywy badawczej to, co kiedyś było domeną świata prywatnego, staje się dostępne dla każdego. Powstaje baza danych opisująca aktywność użytkowników Internetu, która staje się domeną badań społecznych na gruncie geografii.

W nawiązaniu do powyższych odwołań do filmów problematyka przestrzeni cyfrowej w geografii może być rozpatrywana w dwóch perspektywach: przestrzeni cyfrowej jako źródła informacji o przestrzeni oraz przedmiotu badań (Janc 2019). Pierwsza perspektywa to świat Trumana Burbanka (bohatera „Truman Show”) – świat bogaty w informacje. Druga to świat bohaterów „Matrixa” – zmuszający nas do poznania/odkrycia w celu przetrwania. W jednym i drugim podejściu podstawą wszelkich rozważań powinien być odpowiedni sposób działania, który pozwala na zrozumienie samej przestrzeni cyfrowej oraz działań w niej odzwierciedlonych. Przestrzeń cyfrowa doczekała się więc dedykowanych jej sposobów analizy. W tym przypadku można wyróżnić dwie grupy sposobów prowadzenia badań. Pierwsze, to adaptacja tradycyjnych metody badań do przestrzeni cyfrowej (np. ankiety). Drugie to nowe sposoby badań powstałe specjalnie do analiz przestrzeni cyfrowej (np. webometria) (Siuda 2016).

Uwzględniając powyższe, należy zaznaczyć, że celem opracowania jest przybliżenie sposobów eksploracji sieci WWW (części przestrzeni cyfrowej) w odniesieniu do wyszukiwarek internetowych. Środkiem osiągnięcia celu jest dyskusja nad warunkami, wadami, zaletami, ograniczeniami stosowania tych podejść w badaniu Internetu. Zakres pracy zawężono do jednego z komponentów przestrzeni cyfrowej, czyli sieci WWW. W pierwszej kolejności przedstawiono cechy przestrzeni cyfrowej, oraz ich wpływu na geografie jako naukę. W dalszej części opracowania poświęcono uwagę zagadnieniom korzyści wynikających z analizy danych pozyskanych dzięki wyszukiwarkom internetowym oraz wadom takiego podejścia do poznawania rzeczywistości społecznej użytkowników Internetu.

¹ „Chcesz wiedzieć, czym jest Matrix? Matrix jest wszędzie. On jest wszystkim, co nas otacza (...)” (Wachowski L., Wachowski A. 1999, s. 28–29 – za rok cytacji uznano rok premiery filmu, a nie powstania scenariusza).

Przestrzeń i geografia cyfrowa

Przestrzeń cyfrowa jest stworzona z danych, które reprezentują (odzworowują) funkcjonowanie ludzi w wielu kontekstach, sytuacjach, miejscach. Poznawanie przestrzeni cyfrowej umożliwia zrozumienie działań podejmowanych przez człowieka zarówno w niej, jak i w przestrzeni rzeczywistej (Boos 2017). Wynika to z faktu, że obecnie mamy do czynienia ze zhybrydyzowaną przestrzenią rzeczywistą i cyfrową (m.in. Zook, Graham 2007, Ratti, Claudel 2016). Przestrzeń cyfrowa jest „zaludniona” przez rzeczywistych użytkowników, doświadczających obydwu, połączonych ze sobą przestrzeni (Cohen 2007). Działania podejmowane w jednej z nich mają konsekwencje w drugiej. Wzajemnie się przenikają, tworząc świat, w którym odseparowanie aktywności w przestrzeni rzeczywistej i cyfrowej staje się niemożliwe.

Wszechobecność przestrzeni cyfrowej i jej silne związki z przestrzenią rzeczywistą inicjują na gruncie geografii dyskusję o kształcie dyscypliny, szczególnie w zakresie analiz społecznych w odniesieniu do przestrzeni. Obserwuje się daleko idące przemiany w geografii społecznej w kontekście wpływu cyfrowej rzeczywistości na istotę dyscypliny. Rozwija się tzw. geografia cyfrowa (*digital geography*) jako efekt zwrotu cyfrowego (*digital turn*) w geografii. Geografia cyfrowa to badania procesów i zjawisk społecznych w nowym kontekście świata w którym coraz częściej terytoria, miejsca są złączone z przestrzenią cyfrową. Cyfrowość należy rozumieć również jako zwrot ku nowemu sposobowi prowadzenia badań oraz ku zmianie przedmiotu badań. Obiekty tworzące przestrzeń oraz jej właściwości, tj. urządzenia, wytwarzane informacje (artefakty), logika, estetyka oraz relacje społeczne, podlegają ucyfrowieniu (Ash i in. 2018, Leszczynski 2018). W efekcie powstają i rozwijane są nowe sposoby badań na gruncie geografii społecznej.

Istotnym czynnikiem cyfrowego zwrotu jest wzrost znaczenia Big Data (dużych zbiorów danych), których źródłem są w głównej mierze interakcje pomiędzy użytkownikami Internetu (m.in. serwisy społecznościowe, przesyłanie danych, wykonywanie transakcji bankowych, tworzenie zawartość sieci WWW) oraz czujniki automatycznie rejestrujące stan obiektów czy działania ludzi (m.in. Mayer-Schönberger, Cukier 2017). Większość z Big Data zawiera informacje o koordynatach geograficznych, geograficzne metadane, adres czy inne odwołania do przestrzeni rzeczywistej (Leszczynski, Crampton 2016). Umożliwia to zastosowanie ich w badaniach zjawisk przestrzennych, co w konsekwencji pozwala na stworzenie niemal holistycznego obrazu społeczeństwa (Torrens 2010). Społeczeństwo funkcjonuje bowiem w środowisku „bogatym” w dane (Miller, Goodchild 2015). Dane te są cyfrową reprezentacją procesów i zjawisk zachodzących w społeczeństwie.

Samo powstawanie cyfrowej geografii nie zmienia tego, co jest istotą dyscypliny w rozumieniu jej przedmiotu badań, celu ich prowadzenia. Należy przypomnieć, że w geografii społecznej ważnymi kwestiami w badaniach są:

- zrozumienie i wyjaśnienie zachowań człowieka;
- przewidywanie zachowań człowieka i procesów na potrzeby planowania przebiegu zjawisk w przestrzeni, jej organizacji;

- rozwiązywanie problemów występujących w społeczeństwie, które mogą zostać złagodzone poprzez wiedzę o zachowaniach człowieka (Goodchild, Janelle 2004).

Wzrost udziału znaczenia przestrzeni cyfrowej powoduje wzbogacenie zakresu badań.

W naukach zajmujących się poznaniem procesów i struktur społecznych Internet otworzył nowe pola badawcze (np. społeczności internetowe) oraz dedykowane im metody badań. Na gruncie socjologii podstawowe zagadnienia związane z prowadzeniem badań, czyli skąd czerpać dane do analiz, jakie stosować metody, jak interpretować wyniki badań, przedstawia Jemielniak (2019). W odniesieniu do geografii zbliżone kwestie poruszone są w monografii Janca (2017). Powstanie tego typu opracowań jest efektem dążenia do opisanego Internetu z perspektywy nauk społecznych, gdyż stał się on czynnikiem transformacji społeczeństwa. Z drugiej strony dostrzeżono potencjał, jaki oferuje przestrzeń cyfrowa w zakresie możliwości poznania i badania zjawisk zachodzących w społeczeństwie.

Znacząca część przestrzeni cyfrowej związana jest z siecią WWW (*World Wide Web*), jest przez nią tworzona. Sieć WWW uważa się za największą bazę danych na świecie. Składa się z serwisów WWW, które połączone są ze sobą hiperłączami. Odnalezienie pośród miliardów serwisów potrzebnych użytkownikowi informacji jest możliwe poprzez korzystanie z wyszukiwarek internetowych, które pozwalają na lokalizowanie, organizowanie i rozpowszechnianie informacji. One to są jednym z elementów, którego wykorzystanie do zrozumienia sieci WWW, jest niezbędne w naukach społecznych. Głównym sposobem pozyskiwania danych na rzecz analizy sieci WWW jest eksploracja zasobów internetowych (*web mining*), w ramach której można wydzielić trzy podejścia:

- eksploracja danych strukturalnych – np. analiza hiperłączy;
- eksploracja zawartości sieci – analiza zawartości serwisów WWW;
- eksploracja użytkowania sieci – odkrywanie wzorców zachowań użytkowników (Liu 2007, Markov, Larose 2009).

Eksploracja sieci WWW dostarcza bogatego zestawu danych, zarówno ilościowych, jak i jakościowych.

Co wie o nas Google? Co my (nie)wiemy dzięki Google²?

Każda aktywność z wykorzystaniem Internetu pozostawia po sobie ślad. Jest to cyfrowy zapis tych aktywności. Określa się go często jako cyfrowy cień. Powstaje on z danych, które można uznać za publiczne (np. umieszczane w serwisach społecznościowych), oraz takich, które użytkownicy Internetu woleliby pozostawić jako prywatne (np. transakcje bankowe, dane medyczne) (Gantz, Reinsel 2011). Pozostawiane przez poszczególne osoby, społeczności cyfrowe ślady (*digital*

² Google nie jest jedyną z wyszukiwarek internetowych. Jest jednak „największą” i stała się niemal synonimem wyszukiwania informacji w sieci WWW. Stąd też w opracowaniu koncentrującym się na wyszukiwarkach internetowych najwięcej odwołań jest do wyszukiwarki Google.

traces) umożliwiają uzyskanie bogatego obrazu składającego się z wielu „warstw geografii człowieka” (González-Bailón 2013)³.

Jednym z największych „kolekcjonerów” cyfrowych śladów jest firma Google. Dzięki wyszukiwarce, indeksowaniu innych poza wyszukiwaniem dostarczanych usług (np. poczta Gmail, mapy Google), pobieraniu informacji z systemu operacyjnego Android (obsługującego smartfony), jest ona w stanie powiązać dane opisujące różne zakresy funkcjonowania użytkowników. Umożliwia to odtworzenie ich cyfrowego alter ego. W przypadku firmy Google kluczowym produktem, chronologicznie pierwszym, jest wyszukiwarka internetowa. Jest ona podstawowym narzędziem korzystania z Internetu. Obecnie również coraz częściej jest narzędziem umożliwiającym prowadzenie badań społecznych.

Siłą każdej wyszukiwarki, a w tym Google, jest to, że staje się tym bardziej zasobna w informacje i jednocześnie użyteczna, im częściej jest używana, im większą liczbę użytkowników jest w stanie zaangażować. Tym samym wyszukiwarka staje się cyfrowym odzwierciedleniem ludzi. Wpisywane zapytania, klikane odnośniki tworzą bazę danych będącą efektem wiedzy i zachowań użytkowników (Stross 2009, Keen 2015). Dlatego dane z Google mogą być podstawą analiz przemian społeczeństwa, tendencji związanych z popularnością, powstawaniem zjawisk, odzwierciedlających „puls życia” użytkowników tej wyszukiwarki. Liczba wyszukiwanych terminów w Google jest coraz większa i szacuje się obecnie, że w ciągu sekundy na świecie wpisywanych jest blisko 76 tys. zapytań (www.internetlivestats.com), co daje ponad 6,5 mld na dobę, przy udziale rynkowym 76% według stanu na czerwiec 2019 r. (www.netmarketshare.com).

Specyfika funkcjonowania wyszukiwarek umożliwia wskazanie na związek pomiędzy zachowaniami ich użytkowników, podejmowanymi na podstawie rezultatów zapytań, a samym generowaniem tych zapytań. Codziennie miliardy użytkowników podejmują decyzje odnośnie do wielu form aktywności, w tym przestrzennych (np. zakupy, miejsca do odwiedzenia, opinie o miejscach, aktywnościach ekonomicznych) w oparciu o wyniki dostarczane im przez wyszukiwarki internetowe. Wyniki te są efektem tego, jak w interakcji z wyszukiwarką (również innymi usługami internetowymi) zachowywali się wcześniej użytkownicy. Bierze się to stąd, że algorytmy dostosowują rezultaty wyszukiwania do historii wyszukań, domniemanych preferencji użytkowników.

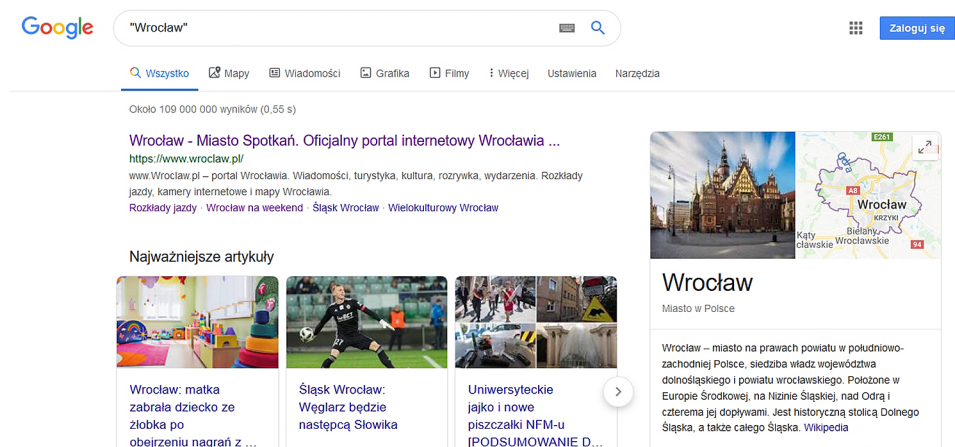
Wpisywane przez użytkowników wyszukiwarek terminy wskazują na zainteresowanie daną tematyką, zagadnieniem, również obiektami w przestrzeni. W jaki sposób możemy skorzystać z informacji zgromadzonych przez wyszukiwarki internetowe w badaniach geograficznych? Metod pozyskania danych, a w konsekwencji ich zastosowania w badaniach, jest kilka: manualne wpisywanie zapytań, korzystanie ze skryptów (automatyzacja procesu pozyskiwania danych), korzystanie ze specjalnych usług dostarczanych przez operatorów wyszukiwarki (np. Google Trends, Baidu Index). W dwóch pierwszych przypadkach można korzystać zarówno z ogólnych wyników wyszukiwania, jak i zawężonych do jednej

³ Przykłady aktywności i związanych z nimi cyfrowych śladów przedstawione są w opracowaniu Surmy (2017).

z kategorii, np. wiadomości (Janc 2015), mapy czy też „widok ulicy” (Graham, Zook 2013, Power i in. 2013), grafik, filmów.

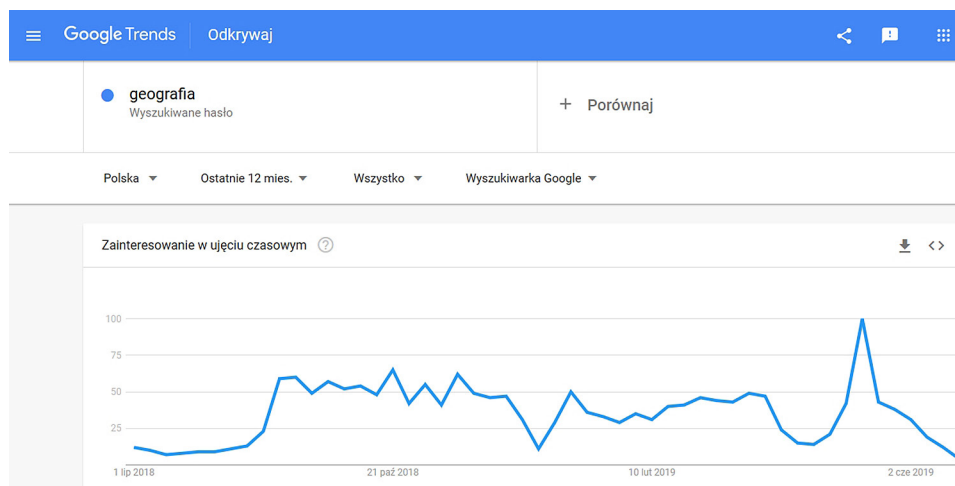
Manualne wpisywanie zapytań do wyszukiwarki czy też pozyskiwanie tych danych na drodze korzystania z narzędzi automatyzujących ten proces umożliwia uzyskanie takich samych rezultatów, czyli informacji o wolumenie zasobów odwołujących się do danego zagadnienia. Za przykład niech posłuży liczba wyników dla zapytania „Wrocław”, która wynosi około 109 mln. Rezultatem zapytania są również informacje o konkretnych stronach WWW, ich charakterystyce – w odniesieniu do miejsc, ich reprezentacji w przestrzeni cyfrowej. Jest to odzwierciedlone na standardowym ekranie wyników wyszukiwania (ryc. 1). Zestaw widocznych na nim informacji (mapa, podgląd na artykuł w Wikipedii, najważniejsze strony związane z miastem, najważniejsze najnowsze wydarzenia) to cyfrowe odzwierciedlenie miejsca.

Dane zgromadzone w wyniku interakcji użytkownik–wyszukiwarka są możliwe do eksploracji w serwisie Google Trends. Narzędzie to oferuje możliwość pozyskania zrelatywizowanych danych o zainteresowaniu użytkowników określonym terminem bądź terminami. Wyniki podawane są w zakresie 0 do 100, gdzie 100 oznacza największe zainteresowanie danym terminem w zadnym okresie (ryc. 2). Ważne jest to, że wartość ta jest obliczana nie na pełnej liczbie zapytań, ale na próbie, co może wpływać na zmiany rezultatów w zależności od dnia badania (Choi, Varian 2012). Nie można więc uzyskać danych o konkretnej liczbie wyszukiwań⁴. Doskonale jednak widoczne są okresy wzmożonego zainteresowania. W przypadku zapytania dla kilku terminów możliwe jest porównanie okresów popularności oraz relatywnej różnicy w popularności. Google Trends oferuje dużo możliwości personalizacji, w tym również ustawienie odniesień do obszaru: państwo, regiony



Ryc. 1. Wynik dla zapytania „Wrocław” w Google (standardowy ekran wyników).
Źródło: google.pl (dostęp: 26.06.2019).

⁴ Czynnione są próby stworzenia modeli statystycznych umożliwiających oszacowanie bezwzględnej liczby wyszukań (np. Lillo, Ruggieri 2019).



Ryc. 2. Wynik dla zapytania „geografia” w Google Trends.

Źródło: trends.google.pl (dostęp: 26.06.2019).

(dla Polski – województwa). Co ciekawe, chiński odpowiednik opisywanej usługi, czyli Baidu Index, oferuje absolutną wartość wyszukiwań (Vaughan, Chen 2015). Informacje bardziej szczegółowe są też możliwe do pozyskania poprzez korzystanie z odpłatnych usług Google, a mianowicie Google Ads (dawniej Google AdWords). Jest to system reklamowy Google. Umożliwia on pozyskanie aktualnej liczby zapytań (także estymowanej) oraz zejście na niższy poziom odniesienia przestrzennego – w przypadku USA nawet na gminny (Gross, Mann 2017).

Jak podają Jun i in. (2018), obecnie trzema głównymi polami zastosowań danych z Google Trends są analizy zagadnień związanych z naukami o komunikacji, technologiami informacyjnymi, medycyną i zdrowiem, ekonomią. Jednak to właśnie geografia na polu analiz fenomenów społecznych ma bardzo duży potencjał do zwiększenia poziomu wykorzystania danych oferowanych przez Google Trends. Już obecnie można wyróżnić kilka nurtów badań, w których rzeczywistość społeczna, funkcjonowanie ludzi w przestrzeni – ich decyzje – analizowane są przy użyciu danych zgromadzonych przez wyszukiwarkę. Eksploracja Google Trends może być pomocna w analizie zainteresowania ludzi atrakcjami turystycznymi i modelowaniu napływu turystów (m.in. Dinis i in. 2017, Park i in. 2017), związkami pomiędzy religijnością a zainteresowaniem treściami pornograficznymi (Whitehead, Perry 2018), zmianami w zainteresowaniu kryptowalutą bitcoin (Yelowitz, Wilson 2015). Podobny charakter oraz pola zastosowań ma Baidu Index – odpowiednik Google Trends (m.in. Zhang, Tang 2016, Huang i in. 2017).

Wady i zalety zastosowania wyszukiwarek internetowych w badaniach geograficznych

Niezależnie od sposobu pozyskania i wykorzystania danych odnośnie do zapytań wpisywanych do wyszukiwarek należy podkreślić ich zalety z perspektywy badań w geografii. Niewątpliwie najważniejszą zaletą jest łatwość pozyskania danych, szczególnie w przypadku gdy korzysta się z gotowych usług o charakterze analitycznym (np. Google Trends). Kolejnym ważnym argumentem za stosowaniem tego typu danych jest małe opóźnienie czasowe pomiędzy zjawiskiem, wydarzeniem, które zaistniało, którego stan chcemy opisać, a dostępnością danych umożliwiających opis. Dane pochodzące z wyszukiwarek dają możliwość analiz zjawisk niemal w czasie rzeczywistym. Prowadzenie tego typu badań jest obecnie koniecznością – zwłaszcza w celach zarządzania działaniami praktycznymi związanymi z zarządzaniem procesami przestrzennymi, np. w ramach funkcjonowania *smart cities*.

Odnośnie do każdego sposobu eksploracji przestrzeni cyfrowej można podać zestaw wad, które bardzo często nie wynikają z samego sposobu pozyskiwania danych, w konsekwencji analiz, lecz z cech przestrzeni cyfrowej (tj. heterogeniczności, braku kontroli nad procesem jej tworzenia, szybkimi zmianami zawartości). Dla badań społecznych bazujących na danych pozyskanych z wyszukiwarek internetowych jest to szczególnie zauważalne i może istotnie wpływać na proces wnioskowania.

W przypadku korzystania z Google Trends należy wskazać kilka kluczowych mankamentów (m.in. Ballatore i in. 2020):

1. Brak informacji o motywacji użytkownika (celu), który wpisuje dany termin w pole wyszukiwania. Związane jest to z tzw. ambiwalentnymi terminami wyszukiwań – vide: efekt Paris Hilton (Brunn i in. 2010) czy też Ryszarda Kalisza (Janc 2012)⁵. Problem ten wynika z semantycznego obciążenia wyników badań, czyli braku pełnego zrozumienia kontekstu pytania użytkownika przez algorytm wyszukiwarki. Jednak dzięki możliwości wyboru w opcjach kategorii tematycznych, których jest blisko 30 (przykładowe to „ludzie i społeczeństwo”, „wiadomości”, „hobby i wypoczynek”) można starać się sprecyzować kontekst zapytania.
2. Obciążenie wyników badań (liczby wyników) krótkoterminowymi wydarzeniami (np. katastrofy naturalne, sprawy kryminalne, ataki terrorystyczne, megawydarzenia). Badania Janca (2015) wskazują, że duże zmiany w wolumenie liczby wyników (na podstawie Google Wiadomości) dla miast w różnych okresach związane są z wydarzeniami sportowymi, kulturalnymi, o charakterze sensacji (kryminalne). Podobne „zachowanie” się rezultatów pochodzących z wyszukiwarki (ich zmienność) występować będzie dla innych zagadnień.

⁵ Przykładowo kiedy podczas kwerendy Google intencją badacza będzie pozyskanie informacji o liczbie wyników dla miasta Łodzi – rezultaty będą również obciążone występowaniem odniesień do środka lokomocji.

3. Kolejnym istotnym problemem jest fakt, że geografia wyszukiwań jest zaburzona różnymi czynnikami. Jednym z nich jest to, że funkcjonujące na rynku wyszukiwarki odznaczają się różną popularnością w poszczególnych państwach. Nie zawsze jest to związane z samymi preferencjami użytkowników (uwarunkowanymi kulturowo), lecz działaniami państwa, np. cenzura Google w Chinach.
4. Na aspekty przestrzenne wyszukiwań wpływa również fakt korzystania przez część użytkowników Internetu z rozwiązań maskujących lokalizację użytkownika. Korzystanie z VPN (*Virtual Private Network*)⁶ czy TOR (*The Onion Router*)⁷ umożliwia z jednej strony bezpieczeństwo poruszania się po sieci WWW, z drugiej skutecznie (choć nie zawsze) usuwa informacje o faktycznym związku z lokalizacją przestrzenną użytkownika, w konsekwencji podejmowanych przez niego działań.

Analizując wyniki wyszukiwań (wolumen terminów zawartych na serwisach WWW, tematyka zasobów), należy zwrócić szczególną uwagę na następujące, ściśle związane z wymienionymi powyżej, problemy:

1. Popularność wyszukiwarek (różna w różnych państwach) wpływa na to, że w zależności od kontekstu badań należy wybrać odpowiednią – przykładowo do analiz widoczności miast chińskich w przestrzeni cyfrowej wybiera się „chińskie Google”, czyli Baidu (m.in. Wang, Loo 2019); dla miast rosyjskich alternatywą może być natomiast wyszukiwarka Yandex. Nie jest to problem w przypadku badań zawężonych do jednego państwa, do jednego kręgu kulturowego (np. ten sam język). Jednak porównania międzynarodowe – analizy niektórych zjawisk w skali globu – są utrudnione, a wręcz niemożliwe.
2. Wyszukiwarki indeksują tylko (niewielki) fragment przestrzeni cyfrowej. Wynika to z tego, że jej część nie jest związana z siecią WWW. Również znaczna część sieci WWW nie podlega możliwości zbadania, gdyż jest ukryta. Zasoby te znajdują się w tzw. głębokiej sieci (*deep web*). Sieć ta, jak podają Sui i in. (2015), jest 400–500 razy większa od sieci WWW indeksowanej przez wyszukiwarki internetowe. Warto również zauważyć, że wobec trendów związanych ze wzrostem znaczenia mediów społecznościowych (np. Facebook) istotnym ograniczeniem wyszukiwarek jest brak możliwości indeksowania ich treści. Integracja mediów społecznościowych z wyszukiwarkami jest więc istotnym wyzwaniem na przyszłość (Halavais 2018).
3. Istotną wadą w przypadku pozyskiwania danych poprzez kwerendę w wyszukiwarce internetowej jest fakt, że dostarcza ona zestaw odwołań do miejsc w sieci WWW zawierających dany termin, jednocześnie jednak użytkownik nie wie, dlaczego te konkretnie zostały wybrane, z jaką precyzją podawany jest wynik (liczba „trafień”) (m.in. Ballatore i in. 2017).

⁶ VPN (Wirtualne Sieci Prywatne) – to rozwiązanie umożliwiające łączenie się użytkownika z docelowym serwerem (np. serwisem WWW) poprzez szyfrowane połączenie, wykorzystujące sieć publiczną (Internet).

⁷ TOR (jedna z sieci anonimizujących) – działa na zasadzie rozproszenia połączenia po kilku węzłach sieci zlokalizowanych na całym świecie w taki sposób, aby niemożliwe było wykrycie źródła wejścia do sieci, a w konsekwencji użytkownika.

4. Kolejnym problemem jest zagadnienie personalizacji wyników wyszukiwania – powstawanie tzw. *filter bubble* – bańki filtracyjnej (Pariser 2011). Wyniki pochodzące z wyszukiwarek internetowych są obecnie dostosowywane do poszczególnych osób – użytkowników Internetu. Na podstawie wcześniejszych wyszukiwań i innych kontekstualnych informacji (np. lokalizacja) użytkownikowi dostarczane są wyniki, które mogą daleko różnić się od wyników innego użytkownika. Również istotnie w ostatnich latach jest to, że w związku z powszechnością korzystania ze smartfonów i powstaniem tzw. mediów lokacyjnych, informacje w wyszukiwarkach dostarczane są w zależności od geograficznego kontekstu jednostki (Halavais 2012). Stąd też wykonywanie kwerendy Google obarczone jest ryzykiem niekompatybilności wyników w przypadku, gdy prowadzone jest z różnych komputerów, różnych miejsc itd. Kwestią, o której nie można zapomnieć w przypadku analiz przestrzeni cyfrowej, w tym danych możliwych do pozyskania z wyszukiwarek internetowych, jest fakt, że dotyczą one de facto tylko użytkowników Internetu. W większości państw świata kilkanaście-kilkadziesiąt procent mieszkańców nie korzysta z Internetu (Janc 2017). Co istotne, są to osoby o określonym profilu społecznym, demograficznym, ekonomicznym. Powoduje to, że nie można uogólniać wniosków wynikających z analizy przestrzeni cyfrowej na całą populację.

Czy wymienione (liczne) wady powinny nas powstrzymać od korzystania z przytoczonych źródeł danych, sposobów analiz? Na pewno nie. Pomijanie tego typu sposobów analizy społecznych aspektów funkcjonowania ludzi skazuje badaczy na zawężenie możliwości opisanie rzeczywistości. Jednak nie należy traktować przestrzeni cyfrowej i danych z niej pochodzących jako jedynej – absolutnego – punktu odniesienia. Jak zauważają Ballatore i in. (2020), właściwe interpretowanie wyników z tego typu badań wymaga wykorzystania lokalnej wiedzy eksperckiej. Jest to zresztą kwestią niezmienną w badaniach geograficznych.

Podsumowanie

Przedstawiony sposób eksploracji przestrzeni cyfrowej jest jednym z wielu i dotyczy tylko wąskiego jej zakresu. Jest jednak bardzo popularny i umożliwia analizę szerokiego spektrum zagadnień. Obecnie korzystanie z wyszukiwarek jest naturalnym – pierwszym, intuicyjnym – sposobem pozyskania informacji. Również na gruncie nauki, w tym zajmującej się społecznymi aspektami funkcjonowania ludzi w przestrzeni, badacze coraz częściej sięgają do tych prostych sposobów, żeby wskazać istotność danego zagadnienia, odwołać się do zasobów informacji o zjawiskach, wydarzeniach.

Świat wyszukiwarek internetowych jest bogaty w informacje. Pozwala, dzięki ich powiązaniu z innymi zasobami, na w miarę precyzyjne opisanie wielu społecznych procesów. Metody oparte na eksploracji danych pochodzących z wyszukiwarek internetowych mają wady, m.in.: brak pełnego „rozumienia” na linii użytkownik–algorytm wyszukiwujący; wpływ regulacji państwowych oraz uwarunkowań kulturowych na korzystanie z określonych wyszukiwarek. Poza wymienionymi

w opracowaniu problemami należy jeszcze poruszyć kwestię tego, czy nasze poznanie nie jest w nadmiernym zakresie limitowane przez dostarczenieli danych. Czy tak jak Truman Burbank nie poznajemy tylko tego świata, który korporacja chce, byśmy poznali? Czy więc korzystając ze sposobów umożliwiających eksploatację przestrzeni cyfrowej, wbrew nadziejom, nie oddaliśmy się od możliwości poznania rzeczywistości społecznej, procesów przebiegających w przestrzeni? Stąd też nie należy bezkrytycznie podchodzić do danych pozyskanych z przestrzeni cyfrowej.

Trzeba również finalnie zauważyć, że przestrzeń cyfrowa, podobnie jak Matrix, powstała jako twór sztuczny, odizolowany od realiów funkcjonowania większości ludzi. Obecnie jednak, jako tworzona przez miliardy użytkowników i urzędów, stała się częścią codziennego życia, osnową większości procesów społecznych. Dlatego poznanie jej ma nie tylko wymiar poznawczy, ale jak najbardziej praktyczny. Rodzi to konieczność opracowywania, rozwijania dedykowanych metod analizy przestrzeni cyfrowej.

Literatura

- Ash J., Kitchin R., Leszczynski A. 2018. Digital turn, digital geographies? *Progress in Human Geography*, 42: 25–43.
- Ballatore A., Graham M., Sen S. 2017. Digital hegemonies: the localness of search engine results. *Annals of the American Association of Geographers*, 107: 1194–1215.
- Ballatore A., Scheider S., Spierings B. 2020. Tracing tourism geographies with Google Trends: a Dutch case study. [W:] P. Kyriakidis, D. Hadjimitsis, D. Skarlatos, A. Mansourian (red.), *Geospatial Technologies for Local and Regional Development. AGILE 2019. Lecture Notes in Geoinformation and Cartography*. Springer, Cham, s. 145–163.
- Boos T. 2017. *Inhabiting Cyberspace and Emerging Cyberplaces. The Case of Siena, Italy*. Palgrave Macmillan, Cham.
- Brunn S.D., Devriendt L., Boulton A., Derudder A., Witlox F. 2010. Networks of European cities in worlds of global economic and environmental change. *Fennia*, 188: 37–49.
- Choi H., Varian H. 2012. Predicting the present with Google Trends. *Economic Record*, 88: 2–9.
- Cohen J.E. 2007. Cyberspace as/and Space. *Columbia Law Review*, 107: 210–256.
- Dinis G., Costa C., Pacheco O., 2017. Similarities and correlation between resident tourist overnights and Google Trends information in Portugal and its tourism regions. *Tourism & Management Studies*, 13: 15–22.
- Gantz J., Reinsel D. 2011. Extracting value from chaos. *IDC iView*, 1142: 1–12.
- González-Bailón S. 2013. Big data and the fabric of human geography. *Dialogues in Human Geography*, 3: 292–296.
- Goodchild M.F., Janelle D.G. 2004. Thinking spatially in the social sciences. [W:] M.F. Goodchild, D.G. Janelle (red.), *Spatially integrated social science*. Oxford University Press, Oksford, s. 3–22.
- Graham M., Zook M. 2013. Augmented realities and uneven geographies: exploring the geolinguistic contours of the web. *Environment and Planning, A*, 45: 77–99.
- Gross N., Mann M. 2017. Is there a “Ferguson Effect?” Google searches, concern about police violence, and crime in US cities, 2014–2016. *Socius*, 3: 1–16.
- Halavais A. 2012. *Wyszukiwarki internetowe a społeczeństwo*. Wydawnictwo Naukowe PWN, Warszawa.
- Halavais A. 2018. How Search Shaped and Was Shaped by the Web. [W:] N. Brügger, I. Milligan (red.), *The SAGE Handbook of Web History*. SAGE, Londyn, s. 242–255.
- Huang X., Zhang L., Ding Y. 2017. The Baidu Index: Uses in predicting tourism flows – A case study of the Forbidden City. *Tourism Management*, 58: 301–306.

- Janc K. 2012. Miasto w sieci – obecność i powiązania Wrocławia i innych dużych miast w cyberprzestrzeni. *Przegląd Geograficzny*, 84: 509–528.
- Janc K. 2015. Visibility and Connections among Cities in Digital Space. *Journal of Urban Technology*, 22: 3–21.
- Janc K. 2017. Geografia internetu. *Rozprawy Naukowe Instytutu Geografii i Rozwoju Regionalnego*, 41. IGRR UW, Wrocław.
- Janc K. 2019. Przestrzeń cyfrowa i internet jako przedmiot zainteresowań w badaniach geograficznych. *Przegląd Geograficzny*, 91: 21–37.
- Jemielniak D. 2019. Socjologia internetu. WN Scholar, Warszawa.
- Jun S.P., Yoo H.S., Choi S. 2018. Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130: 69–87.
- Keen A. 2015. *The Internet is not the answer*. Atlantic Monthly Press, Nowy Jork.
- Leszczynski A., Crampton J. 2016. Introduction: Spatial big data and everyday life. *Big Data & Society*, 3: 1–6.
- Leszczynski A. 2018. Digital methods I: Wicked tensions. *Progress in Human Geography*, 42: 473–481.
- Lillo F., Ruggieri S. 2019. Estimating the Total Volume of Queries to Google. *The World Wide Web Conference ACM*, s. 1051–1060.
- Liu B. 2007. *Web DataMining. Exploring Hyperlinks, Contents, and Usage Data*. Springer, Berlin.
- Markov Z., Larose D.T. 2009. *Eksploracja zasobów internetowych*. WN PWN, Warszawa.
- Mayer-Schönberger V., Cukier K. 2017. *Big data. Rewolucja, która zmieni nasze myślenie, pracę i życie*. MT Biznes, Warszawa.
- Miller H.J., Goodchild M.F. 2015. Data-driven geography. *GeoJournal*, 80: 449–461.
- Pariser E. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin, Londyn.
- Park S., Lee J., Song W. 2017. Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data. *Journal of Travel & Tourism Marketing*, 34: 357–368.
- Power M.J., Neville P., Devereux E., Haynes A., Barnes C. 2013. 'Why bother seeing the world for real?': Google Street View and the representation of a stigmatised neighbourhood. *New Media & Society*, 15: 1022–1040.
- Ratti C., Claudel M. 2016. *The City of Tomorrow: Sensors, Networks, Hackers, and the Future of Urban Life*. Yale University Press, New Haven.
- Siuda P. 2016. Przelamywanie metodologicznych podziałów w rozważaniach o internetowych metodach badawczych. [W:] P. Siuda (red.), *Metody badań on line*. Wydawnictwo Naukowe Katedra, Gdańsk, s. 7–25.
- Stross R. 2009. *Planeta Google. Cel: skatalogować wszystkie informacje świata*. Wydawnictwo Studio Emka, Warszawa.
- Sui D., Caverlee J., Rudesill D.S. 2015. *The Deep Web and the Darknet: A Look Inside the Internet's Massive Black Box*. Woodrow Wilson International Center for Scholars, Waszyngton.
- Surma J. 2017. *Cyfryzacja życia w erze Big Data: człowiek, biznes, państwo*. Wydawnictwo Naukowe PWN, Warszawa.
- Torrens P.M. 2010. Geography and computational social science. *GeoJournal*, 75: 133–148.
- Wachowski L., Wachowski A. 1999. *The Matrix* (www.dailyscript.com/scripts/the_matrix.pdf).
- Wang B., Loo B.P. 2019. The hierarchy of cities in Internet news media and Internet search: Some insights from China. *Cities*, 84: 121–133.
- Whitehead A.L., Perry S.L. 2018. Unbuckling the Bible belt: A state-level analysis of religious factors and Google searches for porn. *The Journal of Sex Research*, 55: 273–283.
- Vaughan L., Chen Y. 2015. Data mining from web search queries: A comparison of google trends and baidu index. *Journal of the Association for Information Science and Technology*, 66: 13–22.
- Yelowitz A., Wilson M. 2015. Characteristics of Bitcoin users: an analysis of Google search data. *Applied Economics Letters*, 22: 1030–1036.
- Zhang Z., Tang W. 2016. Analysis of spatial patterns of public attention on housing prices in Chinese cities: A web search engine approach. *Applied Geography*, 70: 68–81.
- Zook M.A., Graham M. 2007. Mapping DigiPlace: geocoded Internet data and the representation of place. *Environment and Planning B: Planning and Design*, 34: 466–482.

Internet search engines in the digital space study

Abstract: Today we notice the growing importance of digital space – more and more human interactions, their activity are moving, partly or entirely, to this space. As a result, new problems have occurred in science, including geography, and new ways of research are being developed. The aim of paper is to show methods of World Wide Web exploration in relation to the Internet search engines. Internet search engines are one of the most popular and easy to use tool for digital space exploration. The paper shows two methods of data acquisition and consequently analyses: manual entry of queries and use of services provided by search engine operators. We also discussed the pros and cons of the presented approach to digital space analysis.

Key words: digital space, Internet search engines, geography