

Internetowe źródła danych przestrzennych w kontekście precyzji informacji dostępnej dla młodzieży i studentów

BARTŁOMIEJ IWAŃCZAK

Instytut Badań Edukacyjnych, Warszawa, Polska

Proces zdobywania i przetwarzania informacji w kontekście edukacji formalnej różni się zależnie od tego, czy dotyczy uczniów lub studentów należących do pokoleń sprzed rewolucji cyfrowej, czy też z pokoleń, w których świat rzeczywisty i cyfrowy nieustannie się przenikają. Szeroki i stały dostęp do Internetu zmienił fundamentalnie sposób wyszukiwania, wyboru i przesyłania informacji (Rysunek 1) oraz przyspieszył i uprościł cały proces.



Rysunek 1. Różnice w stopniu trudności kolejnych etapów przetwarzania informacji. Źródło: opracowanie własne

Dawniej uczeń lub nauczyciel (analogicznie student i naukowiec) pozyskanie wiedzy rozpoczynali od wizyty w bibliotece w celu znalezienia odpowiedniej książki. Nawet duża biblioteka nie mogła zebrać wszystkich publikacji, dlatego w wymianie informacji najważniejsze były tzw. słabe (dalekie) powiązania sieciowe (wyjazdy do bibliotek zagranicznych, staże w ośrodkach, w których problem badawczy rozwiązano inaczej). Jeśli publikacja była już dostępna, to dotarcie do źródeł, na których opierał się dany tekst, stanowiło wyzwanie. Po wyszukaniu informacji tworzyło się zazwyczaj syntetyczną notatkę, ponieważ dosłowne przepisywanie tekstu pochłaniało zbyt wiele czasu. Po upowszechnieniu się ksero kopiowanie nawet całych publikacji stało się normą.

W świecie cyfrowym znalezienie wielu źródeł dotyczących konkretnej informacji wymaga jedynie precyzyjnego sformułowania słów kluczowych. Po wprowadzeniu ich w wyszukiwarkę pojawiają się dziesiątki tysięcy pozycji, w tym artykuły naukowe z całego świata. Więcej jednak czasu trzeba poświęcić na wyodrębnienie właściwej, poprawnej

treści. Dużym wsparciem w znalezieniu powiązanych ze sobą informacji są portale takie jak scholar.google.pl, researchgate.net, academia.edu. Łatwy dostęp do rozmaitych źródeł w Internecie wpłynął na wzrost zjawiska plagiatu. Kopiowanie treści cyfrowych można wykonywać bez zastanowienia i ograniczeń związanych z objętością tekstu. Przemodelowanie sposobu wyszukiwania informacji stało się wyzwaniem dla współczesnej edukacji.

Obecnie głównym źródłem wiedzy ucznia i studenta stała się sieć. Nabiera znaczenia popularne powiedzenie, że „jeśli czegoś nie ma w Internecie, to nie istnieje”. Informacja w nim dostępna nie jest jednak jednorodna, ma również szereg cech typowych tylko dla tego właśnie nośnika. Kluczowym aspektem jest poziom poprawności i dokładności dostępnych danych. W artykule zostanie przedstawione zróżnicowanie tego poziomu w kontekście danych środowiskowych (przestrzennych) w trzech podstawowych wymiarach: czasowym, przestrzennym i informacyjnym.

Specyfika przestrzennej informacji cyfrowej

Informacje w Internecie są wprowadzane niezależnie i przez wielu użytkowników bez scentralizowanej kontroli. Z tego powodu są to dane nieuporządkowane i w różnym stopniu mniej lub bardziej wiarygodne. Można wyróżnić cztery cechy będące konsekwencją specyfiki internetowych zasobów danych przestrzennych: 1) równoległość, 2) ograniczoność, 3) aprzestrzenność i 4) niestabilność.

Informacja w sieci jest **równoległa**, co oznacza, że te same informacje pojawiają się jednocześnie w wielu miejscach i można je pobrać z wielu niezależnych źródeł. Zwiększa to bezpieczeństwo „przechowywania” informacji, ale także powoduje jej nadmiarowość. Na przykład treści zaczerpnięte z Wikipedii prawdopodobnie powtórzą się na wielu stronach, przez co wzrasta szansa ich odnalezienia. Z drugiej strony pojedyncze wystąpienie unikalnych, cennych treści na blogu specjalisty może powodować ich pominięcie. Dane są zatem rozproszone.

Drugą cechą danych internetowych jest **ograniczoność**. Coraz częściej dostęp do informacji staje się płatny, na przykład do archiwalnych artykułów w gazetach lub artykułów naukowych. Ponadto każde dzieło (artykuł, fotografia) jest chronione prawem autorskim, a jego autor może udostępnić je korzystając z wielu rodzajów licencji.

Większość informacji w Internecie nie jest przechowywana w relacji **przestrzennej**, tzn. nie ma możliwości bezpośredniego zidentyfikowania miejsca, którego dotyczy. Dane o lokalizacji najczęściej mają charakter opisowy. Uniemożliwia to przeszukiwanie według kryteriów przestrzennych (np. znalezienie wszystkich artykułów prasowych o pożarach w odległości kilometra od Pałacu Kultury i Nauki). Wyszukiwarka Google pozwala szukać według słów kluczowych, fraz, czasu tworzenia, a nawet podobieństwa grafiki, nie ma jednak możliwości znajdowania informacji o obiektach powiązanych przestrzennie. Wyjątkiem są Mapy Google, jednak jest to pojedyncza baza danych należąca do giganta informatycznego, a nie mechanizm indeksowania stron.

Wreszcie dane w Internecie są **niestabilne**. Druki zwarte czy czasopisma są archiwizowane w bibliotekach według stanu na moment wydruku, natomiast informacje cyfrowe można usunąć lub zmodyfikować w dowolnym momencie. Równie łatwo je utracić, np. w wyniku nieopłacenia przez administratora strony należności za hosting na kolejny rok.

Analiza zbioru danych o ulicy Krakowskie Przedmieście w Warszawie pokazała, że około 30% źródeł z Internetu zniknęło w ciągu 12 miesięcy. Zatem część cyfrowej

wiedzy traci się bezpowrotnie, zwłaszcza tej, która pochodzi z bieżących informacji (postępy budowy, aktualne wydarzenia itp.).

Typologia

Informacja jest kluczowym elementem procesu nauczania. Szybkość dostępu, wiarygodność, czytelność i inne parametry informacji są całkowicie uzależnione od rodzaju źródła danych. W literaturze wydziela się różne źródła. W przypadku danych przestrzennych w badaniach dziedzictwa kulturowego wyróżniono: badania terenowe, pomiary GPS, informacje udostępnione przez instytucje, archiwa (źródła pisane, materiały kartograficzne i ikonograficzne), mapy (analogowe i cyfrowe), dane teledetekcyjne (cyfrowa fotogrametria i skanowanie laserowe), zdjęcia lotnicze i satelitarne (Jażdżewska, 2010). Szady (2008) dzieli źródła na kartograficzne i pisane na poziomie podstawowym, a Myga-Piątek (2005) wyróżnia badania historyczne, surowe dane historyczne, stare mapy, ikonografię, zdjęcia lotnicze (wraz ze zdjęciami satelitarnymi), własne badania terenowe (inventaryzację), nieformalizowane wywiady, formalne ankiety i oficjalną dokumentację obszaru badań.

Podziały te w sposób dość intuicyjny różnicują zbiory danych, grupując je z punktu widzenia funkcjonalności. Nie ma jednak jednolitego kryterium wydzielenia. W tym artykule podjęto próbę usystematyzowania typologii w kontekście podstawowej triady informacyjnej, czyli precyzji wartości, miejsca i czasu (co-gdzie-kiedy).

W analizie uwzględniono wspomniane wcześniej poziomy: precyzję czasową, przestrzenną i informacyjną (informacyjność) oraz wydzielono siedem głównych rodzajów źródeł informacji, wyróżnionych ze względu na ich formę i zakres przechowywania: wypowiedzi ustne, źródła pisane, zdjęcia, mapy, nagrania wideo, bazy danych i dokumentacja (Rysunek 2).

	PRECYZJA		
	CZASOWA	PRZESTRZENNA	INFORMACYJNA
DANE WERBALNE			
DANE OPISOWE			
DANE FOTOGRAFICZNE			
DANE KARTOGRAFICZNE			
DANE AUDIOWIZUALNE			
BAZY DANYCH			
DOKUMENTACJA			

Rysunek 2. Precyzja na analizowanych wymiarach w podziale na źródło danych: precyzja czasowa (+ precyzja do przedziału, ++ precyzja do roku, +++ precyzja do daty), precyzja przestrzenna (+ precyzja do obszaru, ++ precyzja do linii (ulicy), +++ precyzja do punktu), precyzja informacyjna (+precyzja do opisu, ++ precyzja do kategorii, +++ precyzja do wartości). Jasne pola oznaczają możliwość występowania kilku zaznaczonych poziomów. Źródło: opracowanie własne

Prezentowana w artykule typologia oraz przedstawione wnioski powstały na podstawie analizy internetowych źródeł danych przestrzennych o 425 obiektach w Warszawie. Badanie przeprowadzono w latach 2014–2019 i przeanalizowano w nim 9113 źródeł pochodzących z Internetu.

Dane werbalne

Pierwotnym źródłem informacji o środowisku (pomijając własną obserwację) są przekazy ustne. Niedocenienie tego nośnika wynika z faktu, że pamięć ludzka jest zawodna i nie przechowuje kompletu szczegółów, lecz ogólny obraz. Aby zapewnić ciągłość narracji, umysł uzupełnia braki na bieżąco, co powoduje zniekształcenie informacji. Charakterystyczne cechy dla tego źródła to narracja, brak konkretnych dat (respondenci wskazują raczej okresy) i konkretnych miejsc (zamiast współrzędnych częściej opisywane są obszary). Precyzja czasu jest zwykle bardziej eksponowana niż precyzja miejsca. Historia mówiona umożliwia jednak zebranie informacji lokalnej, której respondenci byli świadkami, oraz zebranie danych, które nie są na tyle znaczące, aby były elementem wiedzy zgromadzonej w innych typach źródeł. Warto zwrócić uwagę, że nie ma fałszywych źródeł werbalnych, gdyż historie przekazane ustnie są prawdziwe dla mówiącego (Fiternicka-Gorzko, 2012).

Początki archiwów źródeł danych werbalnych, zarejestrowanych w różnych formach, sięgają lat 60. XIX w. (prace Huberta H. Bancrofta, badania Williama T. Coucha). Pierwsze metodologiczne uwarunkowania powstały w latach 40. XX w. m.in. w Oral History Research Office (Fiternicka-Gorzko, 2012). W Polsce dopiero na początku lat 80. XX w. uznano wywiady za rodzaj dokumentów archiwalnych, odróżniając je od potocznej historii ludowej (Lewandowska, 2011). Rozwój tej dyscypliny

przyśpieszył dzięki upowszechnieniu w latach 60. XX w. przenośnych magnetofonów kasetowych (Fiternicka-Gorzko, 2012), co umożliwiło rejestrowanie wywiadów i ich późniejszą transkrypcję. Kolejne dwa kamienie milowe w rozwoju tego obszaru to 1) zapis cyfrowy, umożliwiający indeksowanie i udostępnianie, oraz 2) automatyczne rozpoznawanie mowy i zapisywanie jej w postaci tekstowej. Dzięki temu współczesne wywiady gromadzone w postaci zbiorów historii mówionej są ustrukturalizowane oraz często mają wyższe standardy naukowej rzetelności niż dawne pisane teksty historyczne.

Przykładem baz danych werbalnych są utworzone w 1955 r. Archiwum Dokumentacji Mechanicznej (przekształcone w 2008 r. w Narodowe Archiwum Cyfrowe) i Narodowe Centrum Kultury, które realizuje program Świadkowie Historii polegający na zbieraniu świadectw zapomnianych bohaterów. Dobrym przykładem są też Archiwa Historii Mówionej prowadzone przez Dom Spotkań z Historią (Lewandowska, 2011) oraz Muzeum Powstania Warszawskiego. Większość z tych źródeł jest dostępna w Internecie w formie nagrań lub transkrypcji. Lokalne projekty tworzą też nauczyciele w ramach działań edukacyjnych (Ziętał, 2014).

Dane opisowe

Źródła pisane uważa się za podstawowe źródła informacji o środowisku. Najczęściej opisywane są jedynie procesy, zdarzają się jednak dokładne informacje o czasie i nieco rzadziej o lokalizacji. Rzetelność informacyjna tych danych jest podobna do źródeł mówionych, ale mniejsza niż w przypadku map, zdjęć i filmów. Dane opisowe najczęściej wymagają przetworzenia i interpretacji przez człowieka, chociaż z pomocą coraz częściej przychodzi tzw. uczenie maszynowe (ang. *Machine Learning*). Zaawansowane są też obecnie komputerowe mechanizmy przeszukiwania treści tekstowej. Źródła opisowe dość często bywają jedynymi dostępnymi zasobami o danych przestrzennych sprzed XIX w.

Ze względu na poziom rzetelności informacji można wyróżnić cztery rodzaje danych opisowych dostępnych w Internecie. Największą wartość mają książki, monografie i artykuły naukowe udostępniane w postaci pełnotekstowej (najczęściej dostępne w formatach pdf lub djvu). Te zasoby są równorzędne utworom dostępnym w formie papierowej, ale umożliwiają szybsze wyszukiwania informacji. Drugi poziom stanowią encyklopedie internetowe, z których najpopularniejsza jest Wikipedia. Mają dwie ważne cechy: 1) Nie traktuje się ich jako pełnoprawnych źródeł, gdyż każdy może zaproponować zmianę treści, dodając zarówno prawdziwą, jak i fałszywą informację. Jednakże o wiarygodność dba tutaj znacznie szerszy zespół niż w przypadku prac naukowych. Jest to istotne, ponieważ przeprowadzono wiele eksperymentów, w których udawało się opublikować (jako prawdziwe) fikcyjne badania naukowe lub badania prowadzone przez fikcyjnych naukowców czy też artykuły generowane maszynowo. 2) Taka encyklopedia nie ma stałej wersji informacji. Artykuły są nieustannie modyfikowane, zarówno poprzez zwiększenie, jak i przez zmniejszenie poziomu informacyjności. Na trzecim poziomie znajdują się teksty popularnonaukowe – zarówno opublikowane artykuły, jak również strony internetowe urzędów, fundacji czy pasjonatów z danego obszaru tematycznego. Największym zagrożeniem jest niewielka trwałość czasowa tej informacji – nie ma wielu

witryn, które istnieją dłużej niż kilkanaście lat. Modernizacja strony w najgorszym przypadku usuwa stare wpisy, a w najlepszym powoduje przebudowanie zasobów, co w efekcie uniemożliwia odnalezienia poszukiwanego artykułu. Czwarty poziom to wypowiedzi internautów na blogach, forach czy w portalach społecznościowych. Dynamika jest tutaj jeszcze większa, a popularność wśród młodzieży zyskują takie narzędzia jak Snapchat, InstaStories czy relacje na Facebooku, gdzie z założenia informacja nie podlega zachowaniu.

Cechą cyfrowej informacji opisowej jest brak możliwości odróżnienia oryginału od kopii umieszczonej na innej stronie. Stosunkowo łatwe powielanie powoduje, że strony w Internecie można podzielić na te, które podają nowe, unikalne treści oraz te, które powtarzają kilka najbardziej rozpowszechnionych na dany temat. Google przy ok. 10 tys. wyników wyszukiwania określonej informacji przestrzennej wyświetla zazwyczaj od 20 do 30 stron z unikalną treścią spełniającą wymagania zapytania.

Dane fotograficzne

W tym akapicie zdjęcie jest rozumiane jako graficzny (obrazowy) nośnik informacji przestrzennej. Jest nim zarówno oryginalna fotografia wykonana aparatem (cyfrowym lub analogowym), jak i fotografia innej fotografii. Częściowo za zdjęcie może być uznany szkic czy weduta. W przyjętym rozumieniu fotografia strony książki, która zawiera tekst, nie jest zdjęciem.

Charakterystyczne dla zdjęć jest to, że są wykonywane w określonym punkcie oraz w jednym czasie, natomiast parametry te zazwyczaj nie są na nich oznaczone. Przetworzenie informacji ze zdjęcia wymaga analizy a następnie interpretacji przez człowieka. Dostępne są już mechanizmy automatycznego rozpoznawania elementów na zdjęciach (np. Tensorflow).

Kwestia identyfikacji zdjęcia jest kluczowa dla informacyjności. Znacznie łatwiej odnaleźć kolejny tekst wiedząc, że poszukiwany jest Pałac Pacy, niż odnaleźć obiekt, mając tylko zdjęcie Pałacu Pacy bez wcześniejszej wiedzy, co się na nim znajduje. Jeżeli chodzi o aspekt czasu, najczęściej można określić jego przedział podczas wykonywania zdjęcia, a następnie zawężyć go (zarówno po widocznych na zdjęciu obiektach, np. samochodach czy ubraniach, jak i po specyficznej tonacji barwnej zdjęcia). Przy określaniu lokalizacji problemem jest rozstrzygnięcie, czy zdjęcie przypisać do miejsca jego wykonania (którego nie widać), czy do któregoś z obiektów widocznych na nim w bliższej lub dalszej perspektywie. Dobrą praktyką jest schemat opisanego fotografii stosowany przez Google StreetView, w którym widok określają kolejne wartości liczbowe: długość geograficzna (-180,180), szerokość geograficzna (-90,90), azymut względem północy (0,360), kąt nachylenia zdjęcia w pionie (0,90), zoom (~8-800). Dodatkowo w przypadku określenia konkretnej informacji na zdjęciu można dołączyć jej położenie w bieżącym kadrze (x, y). Zaletą zdjęć jest wierne przedstawienie zastanej sytuacji, dzięki czemu poziom szczegółowości już rozpoznanego zdjęcia zdecydowanie przewyższa źródła opisowe, gdyż pozwala na zarejestrowanie informacji, które mogły nie być istotne z punktu widzenia autora informacji opisowej albo werbalnej.

Szczególnie cenne i warte archiwizowania są zdjęcia z czasów, gdy wykonywanie fotografii było kosztowne i dostępne nielicznym. Obecnie wykonuje się tysiące

zdjęć przedstawiających jedną informację w tym samym czasie, co prowadzi do nadmiarowości danych. W przypadku poszukiwania tego rodzaju źródeł należy zawsze zastanowić się nad kryterium wyboru, takiego jak kryterium unikatowości (szukamy konkretnego wydarzenia czy detalu) lub kryterium kroków czasowych (szukamy miejsca w podobnym ujęciu, np. co 5 lat).

Bazy współczesnych zdjęć można podzielić na dwa poziomy zorganizowania. W sposób mało uporządkowany są umieszczane na blogach albo portalach społecznościowych. Najczęściej możliwe jest określenie obiektu na zdjęciu, które jest opisane. Jeżeli nie zostało ono przekształcone w programie graficznym, w metadanych dostępna jest data wykonania lub rzadziej współrzędne geograficzne zapisane podczas geotagowania. Nie ma jednak możliwości szukania zdjęć ze względu na te kryteria. Drugim poziomem są zdjęcia już pogrupowane ze względu na atrybut czasu lub lokalizacji. Zdjęcia mogą być umieszczone jako pojedyncze fotografie, galerie, panoramy 360, galerie panoram czy zdjęcia powiązane z interaktywną mapą (Pokojski i Pokojska, 2012). Przykładami portali indeksujących przestrzennie fotografie są: Fotopolska, której zasoby przekroczyły w 2017 r. milion zdjęć, projekty Panoramio (zakończony w 2017 r.) i Mapio. W przypadku zdjęć historycznych przykładem może być uporządkowana baza Referatu Gabarytów m. st. Warszawy.

Ważną inicjatywą z zakresu gromadzenia informacji w postaci fotografii jest projekt Google StreetView. Jest to seria panoram wykonanych w określonym punkcie czasowym, w regularnych odstępach i co kilka metrów. Jej istotą jest dokumentacja miejsc, które zazwyczaj nie byłyby obiektem zainteresowania fotografa, precyzyjnie wskazana lokalizacja, wiele serii czasowych pokazujących zmiany przestrzenne oraz panoramiczny charakter zdjęć. Usługa StreetView jest dostępna w serwisie Mapy Google od 2007 r., a pierwsze zdjęcia Polski pojawiły się w 2012 r. (Pokojski i Pokojska, 2012).

Jeżeli chodzi o mechanizmy przeszukiwania zdjęć przydatne uczniom i studentom, to na pierwszym miejscu jest wyszukiwarka grafiki Google. Pozwala ona nie tylko wyszukiwać zdjęcia na podstawie słów kluczowych, ale też wyszukiwać strony internetowe, na których umieszczono posiadane przez nas zdjęcie (mechanizm wyszukiwania obrazem).

Dane kartograficzne

Mapa w tej klasyfikacji jest rozumiana jako odwzorowanie obiektów na powierzchni ziemi widziane z góry. Dlatego też w przyjętej klasyfikacji do map, a nie zdjęć, zaliczane są poziome fotografie satelitarne i lotnicze. Mapa podobnie jak zdjęcie jest wykonana w konkretnym czasie. Nie wykonuje się jej jednak co kilka dni lub tygodni. Współcześnie aktualizuje się treści zazwyczaj raz do roku. Dawniej aktualizacja ta z powodów logistycznych trwała kilka lat, a różne fragmenty mapy przedstawiały inne przedziały czasowe. W przypadku ogólnodostępnych zdjęć satelitarnych przyjmuje się podobną rozdzielczość czasową, natomiast zdjęcia specjalistyczne, np. w projekcie Sentinel, wykonywane są nawet co kilka dni.

Mapa przedstawia obszar, a nie punkt, pozwala jednak na odczytanie, czy w danym punkcie coś się znajdowało. Kluczowa w korzystaniu z dawnej mapy, jako źródła informacji, jest świadomość, że chociaż wiadomo, iż coś się w danym miejscu

znajdowało, to (o ile wówczas nie był to obiekt charakterystyczny lub zabytkowy) brakuje dodatkowych informacji, a tym samym punktu zaczepienia do dalszego wyszukiwania informacji o tym obiekcie.

Wyjątkiem są portale indeksujące informację w postaci map cyfrowych, takie jak Wikimapia, Targeo czy Geoportal.gov.pl, na których przy obiekcie wyświetlają się jego atrybuty, na przykład nazwa, funkcja albo adres. Dane są wówczas zapisane w sposób typowy dla Systemów Informacji Geograficznej (GIS), w których atrybut (np. okres) jest podrzędny względem lokalizacji. Standardem są trzy rodzaje obiektów: punkty, linie i poligony. Najwygodniejsza jest reprezentacja w postaci punktów, jednakże nie wszystkie informacje można tak przedstawić.

Buczek wskazał pięć kryteriów oceny map archiwalnych pod względem przydatności jako źródła informacji: potencjalną wiarygodność (okres i okoliczności powstania oraz osoba autora), techniczną poprawność (precyzja i standaryzacja legendy i symbolizacji), matematyczną poprawność (odwzorowanie oraz sposób pomiaru odległości), dokładność i poprawność informacji opisowej (rzetelność) oraz czas przygotowywania map (aktualność danych) (za: Panecki, 2015). W przypadku map historycznych rzetelność jest powiązana z czasem ich powstania. Im starsze są dane, tym precyzja (czasowa i przestrzenna) jest mniejsza w stosunku do dzisiejszych standardów (Szady, 2008).

Ze względu na sposób przechowywania map w cyfrowej bazie danych można wyróżnić trzy warianty. Pierwszym wariantem, najczęściej spotykanym, jest zeskanowanie mapy papierowej i udostępnienie skanu w formie rastrowej (obrazu) jako jednego z plików do pobrania. Przykładem takiego zbioru danych jest Mapster. W jego zasobach jest ok. 15 tys. skanów archiwalnych map (Ziętał, 2014). Bogatą kolekcję map z lat 1919–1939 znajdujących się w archiwum Wojskowego Instytutu Geograficznego udostępnia portal Mapywig.org (Jaskulski, Nalej, 2015). Dużo większe możliwości analityczne daje drugi wariant, czyli nadanie mapie georeferencji. Umożliwia to porównywanie map nałożonych na siebie albo odniesienie informacji na mapie do obecnego zagospodarowania tego terenu. Najczęściej takie mapy mają postać kafelków, czyli fragmentów skanu wczytywanych na bieżąco w zależności od położenia i przybliżenia, co znacznie przyspiesza pracę serwera. Przykładami są portal stareplanymiast.pl oraz baza historycznych map i planów Warszawy dostępnych na stronie mapa.um.warszawa.pl. Trzeci wariant to dokonanie digitalizacji, czyli przetworzenie treści mapy na formę wektorową (przerysowanie obiektów na mapie jako niezależnych). Są one obiektami bazy danych, mającymi swoje atrybuty (np. opis, rok powstania, wysokość). Wariant ten jest najbardziej optymalny ze względu na niewielki zbiór danych i możliwość przeszukiwania. Traci się jednak wtedy oryginalny plik mapy, dlatego najbardziej efektywnym sposobem zapisu jest połączenie wariantu drugiego (jako warstwy źródłowej) i trzeciego (jako warstwy informacyjnej). Zakłada się, że digitalizacja informacji wcześniejszych niż mapy XIX-wieczne jest niepoprawna ze względu na brak ówczesnych pomiarów geodezyjnych (Szady, 2008).

Komputerowe możliwości analizy map historycznych pojawiły się dopiero w latach 90. XX w. i nadal nieliczne są opracowania dotyczące tego zagadnienia (Knowle, 2007; Gregory, 2007; Rumsey, 2004). Stosunkowo częściej powstają cyfrowe repozytoria map nastawione na praktyczne udostępnianie informacji przestrzennej. Wnęk (2010) wymienia przykłady takie jak Great Britain Historical Geographical Information

System (GBHGIS), Belgisch Historisch GIS (BHGIS), HGIS Germany (HIGISG), National Historical Geographic Information System (NHGIS). Różnią się one sposobem działania: od udostępniania skanów map (GBHGIS) do udostępniania danych statystycznych dotyczących obszarów (HIGISG, też: Bank Danych Lokalnych GUS).

Dane audiowizualne

Już pierwsze rejestrujące przestrzeń nagrania z XX w. są cennym źródłem informacji (np. odcinki Kroniki Filmowej), jednak dopiero w XXI w. to źródło danych stało się powszechne. Nagrania, w przeciwieństwie do zdjęć, mają charakter krótkiego odcinka w czasie oraz najczęściej krótkiego odcinka w przestrzeni (przejazd kamery), chociaż zdarzają się publikowane nagrania z jednego punktu w formie poklatkowej. W przypadku filmów ciągłość liniowa i przestrzenna podczas montażu może ulec przerywaniu. Dlatego film wymaga udziału eksperta i interpretacji nagrania.

W ostatnich latach nowym źródłem danych o przestrzeni są nagrania wykonane z użyciem dronów. Takie filmy tworzy się najczęściej na potrzeby promocyjne inwestycji, pozostałe to zazwyczaj nagrania amatorskie. Rejestracje z dronów przewyższają rozdzielczością przestrzenną zdjęcia lotnicze i satelitarne, ale obejmują tylko niewielki obszar, mają zakodowaną trasę przelotu, zatem ich charakter jest liniowy.

Materiały audiowizualne przechowywane są jak inne multimedia (dźwięk, zdjęcia, pomiary sensoryczne) najczęściej w postaci załącznika (atrybutu) powiązanego z danym miejscem. Nagrania są zbiorem pojedynczych zdjęć wykonanych w czasie rzeczywistym. Największym publicznie dostępnym ich źródłem, w tym przedstawiających środowisko, jest portal Youtube.

Bazy danych

Najczęściej spotykane bazy danych przestrzennych to bazy statystyczne. Ich struktura zwykle obejmuje wartość w danym czasie i na danym obszarze (często uśrednione). Takim źródłem jest Bank Danych Lokalnych. Rzadziej spotyka się bazy danych z mapami wektorowymi (np. Państwowy Rejestr Granic czy Topograficzna Baza Danych GUGiK), Niekiedy takie bazy danych zawierają informację o roku powstania każdego obiektu (np. projekt Code Waag).

Baza danych jest tu rozumiana również jako projekt, w którym informacja została uporządkowana i umożliwia przeszukiwanie przestrzenno-czasowe. Przykładem może być wspomniana Wikimapia albo w przypadku Warszawy projekt Warszawa1939.pl. W takich bazach możliwy do przeglądania jest finalny produkt, natomiast nie ma bezpośredniego dostępu do danych surowych. Przeciwnie są zbiory BigData uwzględniające geolokację i pochodzące najczęściej z mediów społecznościowych (Panoramio, Tweeter, Facebook). Ich przeglądanie jest stosunkowo trudne (wyjątkiem są narzędzia jak Trends Google), ale zawierają surowe dane i na ich podstawie można identyfikować różnego rodzaju zależności. Przykładem jest projekt Smellymaps identyfikujący zapachy w przestrzeni na podstawie analizy wypowiedzi użytkowników o tym miejscu w portalach społecznościowych lub wykrywanie epidemii poprzez wzrost liczby zapytań o chorobę w wyszukiwarkach internetowych.

Historyczne bazy danych przestrzennych to najczęściej ucyfrowione źródła spisowe, rejestry nieruchomości, taryfy miejskie, a także wszelkiego rodzaju zestawienia, w których wartości odnoszą się do przestrzeni. Szady (2008) wskazuje, że przed rozbiorami największe znaczenie miały rejestry kościelne (księgi uposażeń i retaksacji, spisy świętopietrza). Informacja o lokalizacji dostępna jest tu w postaci pisanej (np. nazwa ulicy) i dopiero poprzez tę cechę następuje powiązanie z miejscem. Z rejestrów współczesnych należy wskazać Elektroniczne Księgi Wieczyste oraz komercyjne portale z bazami inwestycji budowlanych. W przypadku nowych inwestycji w Warszawie są to m.in. mapymieszkaniaowe.pl.

Osobnym rodzajem baz danych są strony internetowe archiwizujące inne strony internetowe. Jeśli interesująca nas strona już nie istnieje albo została przebudowana, to z dużym prawdopodobieństwem możemy przejrzeć jej zawartość tekstową m.in. dzięki projektowi Webarchive.org. Własną archiwizację wykonuje również wyszukiwarka Google.

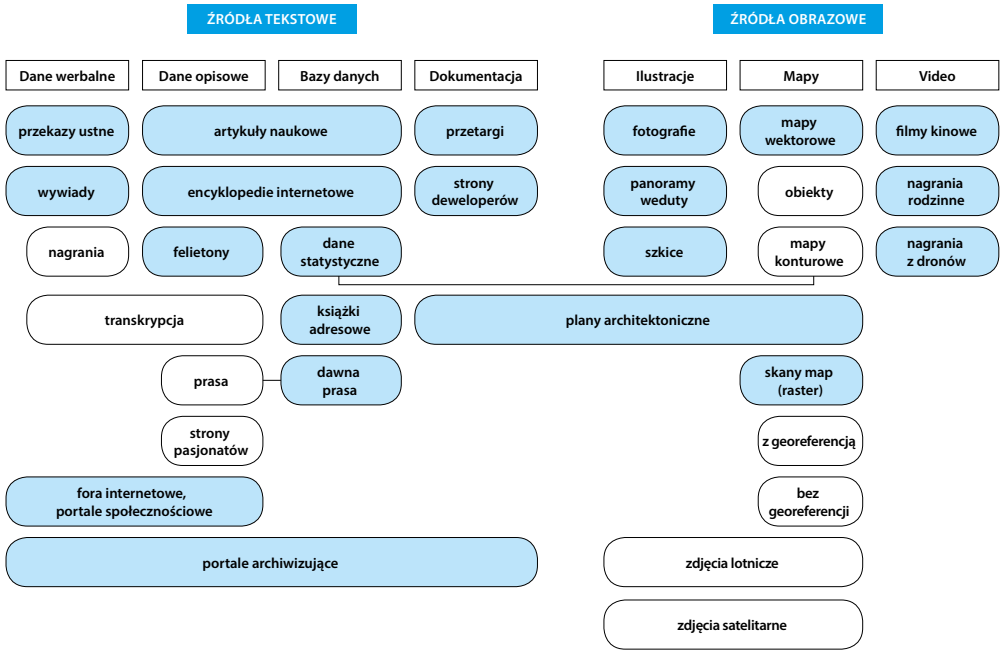
Dokumentacja

Jednym z najbardziej niedocenianych źródeł cyfrowych o wysokim potencjale informacyjnym jest dokumentacja projektowa, udostępniana jako załącznik do przetargów remontowo-budowlanych w Biuletynach Informacji Publicznej. Dotyczy to zwłaszcza obiektów i miejsc, które są budowane lub remontowane. Zasoby te, często unikalne, zawierają przekroje, rzuty, zdjęcia i opracowania na temat tego miejsca. Ponieważ są to części zamówienia publicznego, dostęp do nich jest jawny, jednak ograniczony w czasie, gdyż często treści są usuwane po zakończeniu prac, w kolejnym roku sprawozdawczym lub w momencie uruchomienia nowej strony internetowej instytucji. Zasoby te rzadko też pojawiają się w wynikach wyszukiwania.

Drugim rodzajem danych w postaci dokumentacji są informacje deweloperskie, publikowane na stronach powstających nieruchomości i w okresie ich sprzedaży. Są to często rozkłady mieszkań, informacje o standardach wykończenia, wizualizacje czy zdjęcia z budowy. Strony te są zazwyczaj usuwane po oddaniu inwestycji do użytkowania. Tylko wybrane informacje pozostają na portalach ogłoszeń o nieruchomościach lub na stronach internetowych agregujących zasoby, np. Urbanity.pl albo Remobile.pl. W obu przykładach informacja ma raczej charakter opisowy i jest uzupełniona o załączniki graficzne. Dane o lokalizacji oraz czasie jest wówczas wnioskowana pośrednio na podstawie daty opublikowania każdej z nich.

Wnioski

W klasyfikacji źródeł danych uwzględniającej potencjał informacyjny, dokładność lokalizacji i datowania wyróżniono siedem rodzajów danych, a w nich szereg podkategorii. Efektem przedstawionej w artykule analizy jest poniższa typologia (Rysunek 3).



Rysunek 3. Typologia cyfrowych źródeł danych o środowisku. Źródło: opracowanie własne

W analizowanych źródłach pojawiły się w pewne regularności. Najbardziej rozpowszechniona jest informacja tekstowa. Informacja werbalna jest najczęściej unikalna (choć może być zafałszowana) i wymaga znacznego przetworzenia w celu wydobycia aspektu przestrzennego. Informacja graficzna służy najczęściej do potwierdzenia informacji opisowej i umożliwia identyfikację detali. Mapy pozwalają na określenie lokalizacji, jednak należy wcześniej rozpoznać obiekt, gdyż same są słabym źródłem referencyjnym. Najbardziej szczegółowym i zapewne najmniej docenianym źródłem informacji jest dokumentacja projektowa dostępna ze względu na wymogi ustawowe dotyczącego zamówień publicznych. Okazało się również, że powszechnie dostępne źródła papierowe wcale nie muszą być bogatsze w treść informacyjną (z wyjątkiem opracowań monograficznych). Ale im starsze źródło, tym mniejsza rzetelność informacji. Pisane dawniej książki opierały się na przekazie ustnym, a dawne mapy nie miały precyzji geodezyjnej stosowanej obecnie.

Drugi wniosek z analizy wskazuje na fakt, że powszechny dostęp cyfrowy do informacji oszczędza czas i zwiększa możliwości zdobycia wiedzy przez ucznia. Swobodny dostęp do źródeł danych zmienił mechanizm przepływu informacji. Wiedza dostępna w sieci Internet dociera do każdego odbiorcy na świecie w tym samym czasie, a informacje mają znacznie wyższy potencjał niż te, dostępne tylko w formie papierowej, ponieważ koszty i czas poświęcony na ich uzyskanie są znacznie mniejsze. Zaletą jest też nieograniczona liczba „egzemplarzy” i możliwość bieżącej aktualizacji. Równocześnie odpowiedzialność za zgodne z faktami informacje jest mniejsza niż w innych źródłach. Zdarza się, że częściej cytowane są szeroko rozpowszechnione publikacje bezwartościowe (efekt kuli śnieżnej), niż wysoko cenione

publikacje specjalistyczne. W korzystaniu ze źródeł internetowych niezbędne są umiejętności wyszukiwania informacji, ich weryfikacji, a także znajomość specyfiki odmiennych źródeł danych. Należy też pamiętać, że mimo postępującej digitalizacji wiele treści nadal istnieje tylko w formie analogowej.

Bibliografia

- Fiternicka-Gorzko M. (2012). Historia mówiona: od metody historycznej do interdyscyplinarnego podejścia badawczego. *Opuscula Sociologica*. Vol 2. p. 5–25.
- Gregory I. N., Healey R. G. (2007). Historical GIS: structuring, mapping and analysing geographies of the past. *Progress in Human Geography*. Vol. 31(5). p. 638–653.
- Jaskulski M., Nalej M. (2015). Preparing historical maps for presentation in a geoportal. *Acta Universitatis Lodzianis. Folia Geographica Socio-Oeconomica*. Vol 22. p. 141–159.
- Lewandowska I. (2011). Oral history we współczesnej Polsce – badania, projekty, stowarzyszenia. *Wrocławski Rocznik Historii Mówionej*. Vol. 1. P. 81–103.
- Myga-Piątek U. (2005) Historia, metody i źródła badań krajobrazu kulturowego. [w:] A. Szponar, S. Horska-Schwarz (red.): *Przemiany środowiska przyrodniczego polski a jego funkcjonowanie, Problemy Ekologii Krajobrazu*, vol. 17, p. 71–77.
- Panecki T. (2015). The evaluation of archival maps in geohistorical research. *Miscellanea Geographica – Regional Studies On Development*. Vol 19(4). p. 72–77.
- Pokojski W., Pokojńska P. (2012). Wirtualne podróże w serwisach internetowych i aplikacjach webgis. *Turystryka i Rekreacja*. Vol. 9. p. 31–37.
- Szady B. (2008). Zastosowanie systemów informacji geograficznej w geografii historycznej, „Polski Przegląd Kartograficzny” 40, nr 3, p. 279–283.
- Wnęk K. (2010). Systemy GIS w badaniach historycznych. *Prace Historyczne*. Vol. 1 p. 153–171.
- Ziętał, K. (2014). KARTA i archiwa społeczne. *Biuletyn EBIB [on-line]* 2014, nr 6 (151), *Archiwa prywatne i społeczne jako składnik dziedzictwa narodowego*. <http://open.ebib.pl/ojs/index.php/ebib/article/view/271.ISSN1507-7187>