

## PROBLEMATYKA GEOKODOWANIA ZDARZEŃ DROGOWYCH<sup>1</sup>

### PROBLEMS OF ROAD ACCIDENTS GEOCODING

**Piotr Cichociński**

AGH Akademia Górniczo-Hutnicza w Krakowie,  
Wydział Geodezji Górniczej i Inżynierii Środowiska, Katedra Geomatyki

**Słowa kluczowe:** kolizja, OpenStreetMap, liniowy system odniesienia, pikietaż, wypadek, geokodowanie wypadków

**Keywords:** collision, OpenStreetMap, linear reference system, chainage, accident, accidents geocoding

### Wstęp

Geokodowanie jest to proces określania położenia obiektów, wyrażonego za pomocą różnego rodzaju identyfikatorów geograficznych, najczęściej adresu. Odbywa się przez porównanie odpowiednich elementów informacji adresowej z materiałem odniesienia, określającym położenie poszczególnych adresów.

Adres może być zapisywany w wielu różnych postaciach. W Polsce, zwyczajowo dla miejscowości, w których ulice mają nazwy, przyjmuje się, że adres składa się z nazwy ulicy, po której następuje numer porządkowy budynku (i ewentualnie numer lokalu). Dodatkowo podawany jest kod pocztowy i nazwa miejscowości. Taki zestaw informacji pozwala jednoznacznie zidentyfikować daną lokalizację w skali całego kraju. Innymi obiektami lub punktami charakterystycznymi, wykorzystywanymi w geokodowaniu, mogą być słupki kilometrowe i hektometrowe umieszczone przy głównych drogach lub też skrzyżowania dróg i ulic.

Zlokalizowana w przestrzeni informacja, może być podstawą do wykonywania map i przeprowadzania analiz geograficznych. Można określać relacje przestrzenne występujące pomiędzy wyznaczonymi punktami oraz badać związki pomiędzy wartościami atrybutów a położeniem obiektów. Jako przykłady wykorzystania informacji o lokalizacji uzyskanej poprzez geokodowanie, można podać analizy przestrzennego rozkładu wartości nieruchomości (Cichociński, 2010), badanie zagadnień zdrowotnych (Rushton i in., 2006), planowanie usług

---

<sup>1</sup> Praca została zrealizowana w ramach Badań Statutowych prowadzonych w roku 2014 w Katedrze Geomatyki Wydziału Geodezji Górniczej i Inżynierii Środowiska Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie.

miejskich (Brzuchowska, 2010), analizy przestępczości (Ratcliffe, 2004) lub też poszukiwanie przyczyn wypadków drogowych (Hijar i in., 2003, Li i in., 2007, Meliker i in., 2004).

Szczególnie właśnie miejsca wystąpienia zdarzeń drogowych, takich jak wypadki i kolizje, dokumentowane są za pomocą opisu geograficznego. Mimo powszechnej dostępności technologii GPS (GNSS) wciąż niewiele z nich ma położenie wyrażone za pomocą współrzędnych. W większości przypadków lokalizacja ogranicza się do wskazania najbliższego adresu, kilometrażu drogi, skrzyżowania lub nawet swobodnego opisu. Dlatego poprawny wynik geokodowania będzie miał istotne znaczenie dla wiarygodności rezultatów ewentualnych analiz (zwłaszcza przestrzennych).

Wobec dostępności odpowiedniego oprogramowania i stwierdzonej (Karimi i in., 2004) porównywalności stosowanych algorytmów geokodowania, celem pracy nie jest analiza samego procesu geokodowania, lecz z jednej strony zbadanie dostępności dla obszaru Polski odpowiednich danych referencyjnych, a z drugiej próba zwrócenia uwagi na zapis informacji adresowej, która ma podlegać geokodowaniu, ze szczególnym uwzględnieniem opisu lokalizacji zdarzeń drogowych.

## Geokodowanie

### Geokodowanie adresów

W przypadku posługiwania się informacją adresową, geokodowanie jest procesem, który składa się z trzech głównych etapów (Charif i in., 2010):

- 1) strukturyzacji i normalizacji poszukiwanych adresów, w celu zidentyfikowania poszczególnych elementów składowych wprowadzonego kompletnego adresu (numer budynku, nazwa ulicy, kod pocztowy, miasto, etc.);
- 2) znajdowania pasującego adresu w źródłowej bazie danych przez porównanie z odpowiednimi elementami adresu poszukiwanego;
- 3) właściwego geokodowania, czyli wyznaczenia współrzędnych zidentyfikowanego adresu.

System powinien znaleźć najlepiej dopasowane adresy bez interwencji użytkownika. Każdy przetworzony wiersz tabeli powinien zostać opisany atrybutami podającymi znalezione położenie oraz uzyskaną dokładność dopasowania. W tym przypadku również duże znaczenie ma jakość poszukiwanego adresu, często bowiem występują błędy literowe. Rezultat na poziomie 70% prawidłowo zlokalizowanych adresów jest często uważany za satysfakcjonujący (Paull, 2003).

Tymczasem, jakość danych uzyskanych w procesie geokodowania ma istotne znaczenie, gdyż dane te często są podstawą do podejmowania decyzji. Na jakość geokodowania, oprócz wymienionego powyżej sposobu zapisu poszukiwanego adresu, mają wpływ również cechy źródłowej bazy danych, takie jak (Karimi i in., 2004): kompletność, poprawność zapisu adresu, aktualność i dokładność położenia, a także w niektórych przypadkach technika interpolacji.

Istnieją trzy podstawowe metody określania lokalizacji adresu. Najdokładniejsza polega na pobraniu współrzędnych obiektu punktowego. Geokodowanie na podstawie punktów jest oceniane bardziej pozytywnie (Zandbergen, 2008, Vieira i in., 2010) ze względu na doskonałą dokładność położenia. Drugą z popularnych metod jest interpolacja współrzędnych, dokonywana wzdłuż obiektu liniowego, najczęściej ulicy, przy założeniu pewnego zakresu numeracji

przyporządkowanego każdemu z takich obiektów. Metoda taka nazywana jest *street geocoding*. Z jej stosowaniem wiąże się wiele problemów, szeroko opisywanych w literaturze (Zandbergen, 2007, 2011; Zimmerman, Li, 2010). Szczególnym przypadkiem geokodowania względem ulic jest znajdowanie skrzyżowań. W takim przypadku poszukiwaną lokalizację należy opisać za pomocą dwóch nazw ulic. Natomiast, podstawą funkcjonowania trzeciej metody jest wyznaczenie w dowolny sposób punktu leżącego wewnątrz obiektu powierzchniowego lub liniowego.

### Liniowe systemy odniesienia

Adres jako forma lokalizacji zdarzeń dobrze sprawdza się na terenach zabudowanych. Jednak na obszarach, gdzie brak jest punktów adresowych, trzeba było znaleźć lub utworzyć inne punkty charakterystyczne. W ten sposób powstał liniowy system odniesienia (*linear referencing system* – LRS). Jest definiowany jako system wspomagający gromadzenie i zarządzanie informacją o zdarzeniach, które występują w sieci transportowej (Curtin i in., 2007). W tym kontekście LRS zwykle składa się z czterech elementów: sieci drogowej stanowiącej przestrzenną podstawę dla lokalizacji zdarzeń, zbioru obiektów posiadających określone położenie w przestrzeni (tworzących układ odniesienia), co najmniej jednej metody pozwalającej na lokalizację nieznanego położenia oraz zbioru zdarzeń punktowych lub liniowych, którego lokalizację wzdłuż elementów sieci trzeba określić (Curtin, 2007).

System taki upraszcza rejestrację danych, posługując się względnym położeniem wzdłuż istniejących obiektów liniowych. To znaczy, że pozycja jest definiowana za pomocą odniesienia do segmentu liniowego obiektu geograficznego oraz odległości mierzonej wzdłuż tego segmentu od danego punktu (PN-EN ISO 19116:2006). Na przykład: Droga Krajowa nr 22, kilometr 238, jednoznacznie identyfikuje położenie w przestrzeni geograficznej bez potrzeby wyrażania go w postaci współrzędnych.

Najczęściej spotykanym przykładem liniowego systemu odniesienia jest system słupków kilometrowych i hektometrowych (zwanym także słupkami pikietażowymi), ustawianych wzdłuż głównych dróg, stosowany do zarządzania infrastrukturą drogową (Smith i in., 2001). Jest on szczególnie użyteczny przy zbieraniu i wykorzystaniu danych w terenie. Położenie, wyrażone przez nazwę lub numer drogi, oznaczenie najbliższego słupka oraz odległość od tego słupka może być stosowane do jednoznacznego zlokalizowania obiektu lub zdarzenia. Słupki powinny być umieszczane co 100 metrów, wzdłuż krawędzi jezdni, od początku drogi do jej końca. W Polsce, początek i koniec drogi stanowią: granica państwa, skrzyżowanie z inną drogą publiczną, granica portu morskiego. Początkiem kilometrażu drogi jest punkt referencyjny na początku jej przebiegu, któremu zostaje nadany kilometraż początkowy równy 0 (km). Kilometraż najczęściej rośnie z północy na południe lub z zachodu na wschód. W przypadku występowania wspólnych przebiegów dróg klas technicznych A (autostrada) i S (szybkiego ruchu) z dowolnymi innymi drogami, ciągłość kilometrażu występuje na drodze o wyższej klasie technicznej, dla pozostałych przypadków ciągłość kilometrażu występuje na drodze o niższym numerze. Na pozostałych drogach występuje „nieciągłość” kilometrażu na długości wspólnego przebiegu odcinka drogi (GDDKiA, 2012).

## Lokalizacja zdarzeń drogowych

### Przygotowanie danych o lokalizacji zdarzeń drogowych

Sposób rejestrowania przez policję zdarzeń drogowych określa zarządzenie Komendanta Głównego Policji (KGP, 2006, 2012, 2013). W szczególności w rozporządzeniu zawarto wzór karty zdarzenia drogowego oraz określono sposób wypełniania takiej karty. W zakresie informacji o miejscu zdarzenia karta zawiera następujące pola: POWIAT, GMINA, MIEJSCOWOŚĆ, ULICA, NR POSESJI (dopuszczalne jest podanie zamiennie numeru innego punktu charakterystycznego, na którego wysokości zaistniało zdarzenie, na przykład numer latarni), DROGA NR, KM, HM, SKRZYŻOWANIE Z ULICĄ / DROGĄ NR, ODLEGŁOŚĆ DO SKRZYŻOWANIA (odległość w metrach od miejsca zdarzenia do najbliższego skrzyżowania, jeżeli droga nie jest wyposażona w słupki pikietażowe), KIERUNEK (nazwa miejscowości, w której kierunku mierzono odległość od miejsca zdarzenia do najbliższego skrzyżowania), WSPÓŁRZĘDNE GPS (wyrażone w stopniach, minutach, sekundach).

Dane z karty są następnie rejestrowane w komputerowym Systemie Ewidencji Wypadków i Kolizji (SEWiK). Zgromadzone informacje mają być podstawą do sporządzania okresowych zestawień o stanie bezpieczeństwa ruchu drogowego. Trudno sobie jednak wyobrazić rzetelne i kompletne analizy nieoparte prezentacją przestrzennego rozkładu zdarzeń.

Dzięki uprzejmości administratorów portalu internetowego [sewik.pl](http://sewik.pl) pozyskano komplet informacji o zdarzeniach drogowych z lat 2007-2012, które miały miejsce w Krakowie. Dane takie udostępnione zostały przez Komendę Główną Policji dla sieci Miasta dla Rowerów, celem upowszechniania informacji o wypadkach i kolizjach z udziałem rowerzystów. Zbiór danych miał wielkość 55 MB i postać zapisanego w języku SQL zrzutu (*dump*) z systemu zarządzania bazą danych MySQL.

Po wprowadzeniu drobnych zmian, wynikających z różnic pomiędzy dialektami języka SQL, został wczytany do bazy danych SpatiaLite. Wszystkie wymienione powyżej informacje o lokalizacji zdarzenia zapisane są w tablicy „zdarzenie”. Liczy ona 72 540 wierszy. Ze względu na różne możliwe do zastosowania metody geokodowania wybrano i zapisano w odpowiednich plikach .dbf cztery grupy zdarzeń lokalizowane odpowiednio przez atrybuty:

- ULICA (w bazie danych kolumna ULICA\_ADR) i NR POSESJI (kolumna NUMER\_DOMU),
- ULICA (ULICA\_ADR) i SKRZYŻOWANIE Z ULICĄ / DROGĄ NR (ULICA\_SKRZ)
- DROGA NR (ZSSD\_KOD) oraz KM i HM (zapisane w postaci ułamka dziesiętnego w jednej kolumnie KM\_HM),
- WSPÓŁRZĘDNE GPS (kolumny WSP\_GPS\_X i WSP\_GPS\_Y, zawierające współrzędne wyrażone w stopniach, zapisane w systemie dziesiętnym).

Za każdym razem posługiwano się ogólnym zapytaniem SQL:

```
SELECT ID, ULICA_ADR, NUMER_DOMU, ZSSD_KOD, ULICA_SKRZ,
KM_HM, ODLEGLOSC, KIERUNEK, WSP_GPS_X, WSP_GPS_Y
FROM zdarzenie
```

w każdym z przypadków formułując tylko odpowiedni warunek.

Dla wariantu opisu poprzez „tradycyjnie” pojmowany adres, składający się z nazwy ulicy i numeru adresowego warunek był następujący: `WHERE NUMER_DOMU<'A' AND NUMER_DOMU>='0'`. Już wstępne zapoznanie się z zawartością tych dwóch kolumn

pokazało, że poprawne przeprowadzenie procesu geokodowania nie będzie łatwe. Podstawowym problemem, który zaobserwowano, jest bardzo dowolny sposób wypełniania poszczególnych kolumn (tab. 1). Na przykład w 2958 przypadkach jako nazwę ulicy podano RONDO i jego nazwę. Jednak szczególnie chodzi o kolumnę NUMER\_DOMU, której zawartością jest liczba tylko

w 19 824 wierszach. W 8118 przypadkach występuje słowo BRAK. W pozostałych 16 844 znaleźć można raczej opis punktu charakterystycznego, takiego jak na przykład: MOST (225), WIADUKT (123), nazwa stacji benzynowej (ok. 700), etc.

W tej ostatniej grupie wyróżnia się określenie „na wysokości czegoś” (7514 wierszy), na przykład: innej ulicy, sklepu, stacji benzynowej, przystanku. Można też zaobserwować przypadki, że w kolumnie NUMER\_DOMU zapisana jest nazwa ulicy, natomiast nazwa centrum handlowego znajduje się w kolumnie ULICA. Takie dane nie mogą podlegać geokodowaniu bez gruntownego przetworzenia.

W przypadku zdarzeń, które miały miejsce na skrzyżowaniach zastosowano warunek: `WHERE ULICA_SKRZ<>' ' AND NUMER_DOMU = ''`. Z otrzymanych 24 898 wierszy, 117 zawierało w kolumnie ULICA\_SKRZ słowo BRAK, 1446 W RONDZIE i 51 DROGA OSIEDŁOWA bądź DROGA WEWNĘTRZNA, które usunięto z ostatecznego wyniku.

Do wybrania zdarzeń lokalizowanych z użyciem słupków kilometrowych i hektometrych posłużono się warunkiem: `WHERE KM_HM<>' '`. Z uzyskanych w wyniku 1258 wierszy przeważająca większość (1226) dotyczyła przebiegającego przez Kraków odcinka autostrady A4. 17 odwoływało się do relatywnie krótkiego fragmentu drogi S7, a 15 zawierało w kolumnie ZSSD\_KOD niezrozumiałe dla autora oznaczenie KS.

Lokalizacje zdarzeń opisane współrzędnymi wybrano przez warunek: `WHERE WSP_GPS_X <> ''`. Takich miejsc było tylko 46, co stanowi zaledwie 0,06% wszystkich zdarzeń.

**Tabela 1.** Fragment tablicy „zdarzenie”

ULICA_ADR	NUMER_DOMU
RONDO POLSAD	
ZAKOPIAŃSKA	WIADUKT-WYS. WĘZEL OPATKOWICKI
WIELICKA	STATOIL
DASZYŃSKIEGO	WYJAZD Z GALERII KAZIMIERZ
KONOPNICKIEJ	WYS.MANGHA
GALERIA KRAKOWSKA	PAWIA 5 - PARKING

### Przygotowanie danych referencyjnych

Najistotniejszym źródłem otwartych danych przestrzennych w Polsce (Michalak, 2007), szczególnie dotyczących sieci drogowej, jest OpenStreetMap (OSM). Model danych OSM różni się od klasycznego modelu danych, stosowanego w systemach informacji geograficznej. Opiera się na trzech podstawowych elementach: punktach, zwanych węzłami (*node*), liniach (*way*) i relacjach, które opisują przestrzenne cechy obiektów. Charakterystyki opisowe są zapisywane w postaci atrybutów przyjmujących postać par klucz-wartość (*key-value*) (Zhou i in., 2013). W szczególności informacja adresowa zapisywana jest w postaci dwóch atrybutów: *addr:street* i *addr:housenumber*. Przewagą OSM nad większością komercyjnie dostępnych cyfrowych zbiorów danych sieci drogowych oraz popularnymi portalami internetowymi udostępniającymi mapy (takimi jak na przykład Google Maps) jest to, że rejestruje położenie słupków kilometrowych.

Mimo, że możliwe jest pobranie wybranych fragmentów OpenStreetMap (OSM) w postaci plików *shape* (Esri, 1998), to jednak, aby mieć pełny dostęp do wszystkich danych zapisanych w bazie OSM, zdecydowano się na pobranie ich w formacie OSM XML. Skorzystano z portalu [metro.teczno.com](http://metro.teczno.com), oferującego tak zwane *Metro Extracts*, czyli fragmenty bazy OSM, obejmujące większe miasta świata. W tej grupie znajduje się również Kraków. Wybrany format danych idealnie odpowiada strukturze bazy danych OSM, lecz ponieważ struktura ta jest różna od tradycyjnie stosowanego w systemach informacji geograficznej georelacyjnego modelu danych, nie było łatwo załadować pobrane dane do oprogramowania GIS. Ostatecznie zdecydowano się skorzystać z systemu zarządzania bazą danych przestrzennych SpatiaLite, który umożliwia przetworzenie danych OSM na następujące trzy sposoby (Furieri, 2012):

1. Elementy geometryczne zostają przetworzone do typowej dla systemów informacji geograficznej postaci punktów, linii i poligonów. Obiekty jednego rodzaju zostają następnie pogrupowane w warstwy tematyczne (tabele bazy danych) na podstawie posiadanych atrybutów. Wynikowa baza danych może być od razu użyta w oprogramowaniu GIS.
2. Zostaje utworzona poprawna sieć drogowa, składająca się z węzłów i elementów liniowych. Wynikowy graf przeznaczony jest do wykorzystania w module SpatiaLite, służącym do wykonywania analiz sieciowych.
3. Powstaje struktura bazy danych wiernie odwzorowująca obiekty OSM i ich atrybuty oraz związki występujące pomiędzy nimi, w pełni zachowująca oryginalny model danych OSM, unikająca wprowadzania jakichkolwiek ograniczeń w danych. Taka baza danych może być stosowana w szczególnych przypadkach, wymagających dostępu do absolutnie wszystkich danych, zwłaszcza atrybutowych, zapisywanych w OpenStreetMap.

Dzięki sformułowaniu zapytań w języku SQL było możliwe wydobycie z wygenerowanych zgodnie z powyższym opisem baz danych odpowiednich podzbiorów, które następnie wyeksportowano do plików *shape* (z wyjątkiem ulic, pochodzących bezpośrednio z tabeli przeznaczonej do analiz sieciowych).

Jako punkty adresowe wybrano wszystkie węzły (*node*) opisane jednocześnie atrybutami *addr:street* i *addr:housenumber*:

```
SELECT n.node_id as ID, nt1.v as numer, nt2.v as ulica, n.geometry
FROM osm_node_tags as nt1, osm_node_tags as nt2, osm_nodes as n
WHERE (nt1.k="addr:housenumber" and nt2.k="addr:street")
and nt1.node_id=n.node_id and nt2.node_id=n.node_id
```

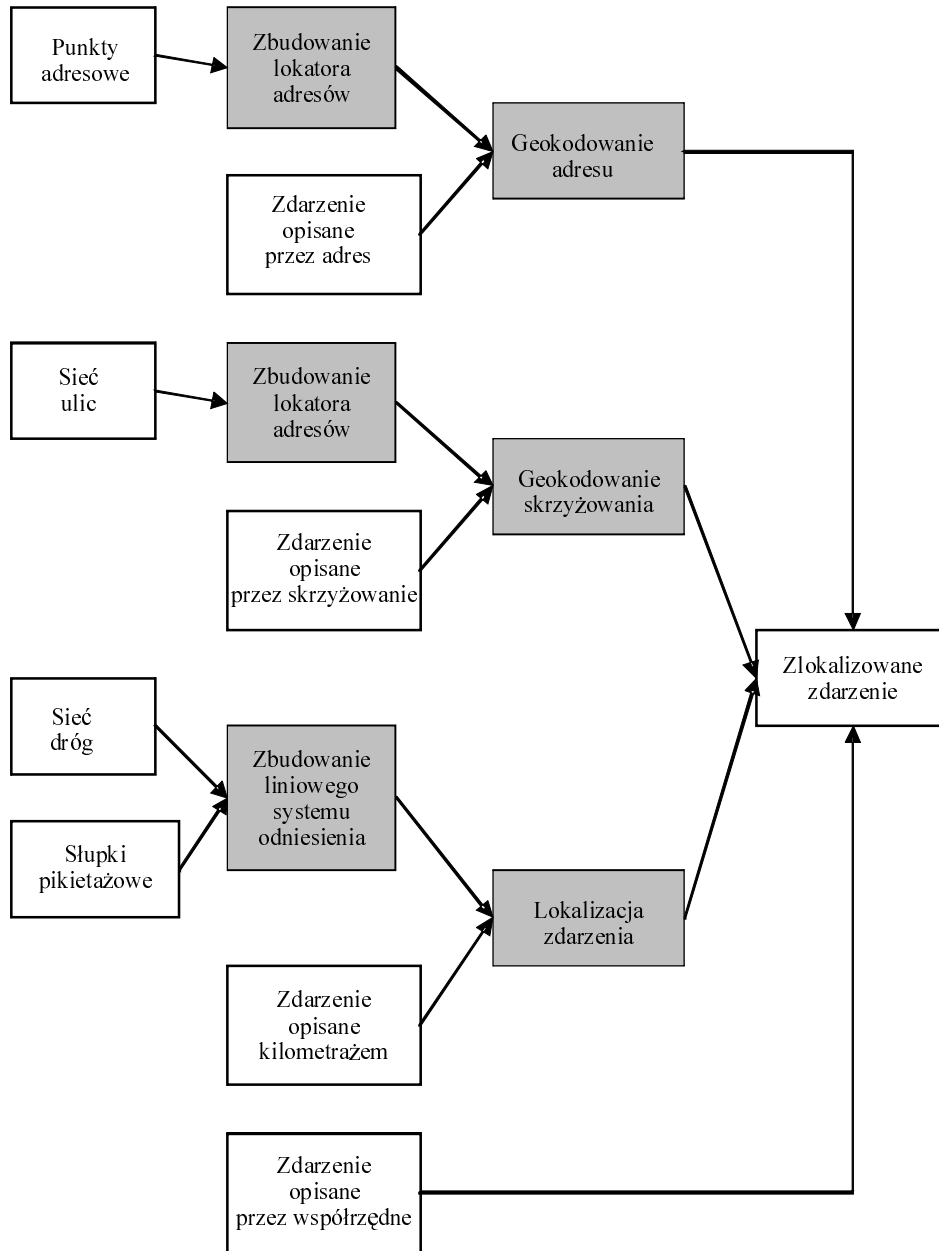
Słupki pikietażowe, to obiekty punktowe związane z drogami, opisane atrybutem *ref*, podającym numer drogi oraz atrybutem *distance*, określającym odległość słupka od początku drogi:

```
SELECT p.id, p.sub_type, p.geometry, r.v as ref_tag, d.v as dist_tag
FROM pt_highway as p, osm_node_tags as r, osm_node_tags as d
WHERE p.sub_type="milestone" and p.id=r.node_id and r.k="ref"
and p.id=d.node_id and d.k="distance"
```

Osie dróg to obiekty liniowe opisane atrybutem *ref*, zawierającym numer drogi:

```
SELECT *
FROM ln_highway, osm_way_tags
WHERE ln_highway.id=osm_way_tags.way_id and osm_way_tags.k="ref"
```

## Przeprowadzenie geokodowania



**Rysunek.** Schemat prezentujący zbiory danych oraz działania prowadzące do określenia lokalizacji zdarzeń drogowych opisanych czterema różnymi sposobami

Na rysunku zaprezentowano użyte w badaniach zbiory danych oraz zrealizowane działania prowadzące do określenia lokalizacji zdarzeń drogowych, opisanych czterema różnymi sposobami.

Geokodowanie 19 824 zdarzeń, lokalizowanych przez adres, przeprowadzono w programie ArcGIS firmy Esri, który posiada odpowiednie narzędzie do tego celu (Esri, 2011). Proces geokodowania polegał w tym przypadku na zbudowaniu najpierw tak zwanego lokatora adresów, który określa jakie role pełnią poszczególne pola (kolumny) tabeli atrybutów źródłowego zbioru danych, to znaczy w jakich polach zapisane są numery, nazwy ulic, kody pocztowe, etc. Pierwszy lokator adresów, który zbudowano, korzystał z dwóch informacji, a mianowicie nazwy ulicy i numeru adresowego. Opcje dopasowania pozostawiono na domyślnych, sugerowanych przez program poziomach, to jest: wrażliwość pisowni 80%, minimalny wynik kandydatów 10%, minimalny wynik dopasowania 85%. Uzyskano następujące wyniki: dopasowane 7715 (39%), powiązane 929 (5%), niedopasowane 11 180 (56%), które należy uznać za absolutnie niesatysfakcjonujące (w przypadku wyniku „powiązane”, należy rozumieć, że znaleziono więcej niż jeden pasujący adres i wymagana jest interwencja operatora). Jako przyczyny zidentyfikowano dwa problemy: małą aktualność posiadanej referencyjnej bazy adresowej (dane zostały pobrane ponad pół roku przed przeprowadzeniem badań, a w tym właśnie czasie dosyć intensywnie część adresowa OSM była rozbudowywana) oraz niejednorodność sposobu zapisu nazw ulic, pochodzących od osób (w opisach zdarzeń drogowych z reguły występowało samo nazwisko, natomiast w bazie danych OpenStreetMap najczęściej było zapisane imię i nazwisko). Rozwiązanie pierwszego z tych problemów wymagałoby zaktualizowania posiadanej bazy adresowej, natomiast problem drugi spróbowano prowizorycznie rozwiązać, wybierając z nazw ulic ostatni wyraz. Było to łatwe do zrealizowania przez zastosowanie w programie ArcGIS narzędzia *Kalkulator pól* i użycie do wyznaczenia wartości nowej kolumny zapisanego w języku programowania Python wyrażenia `!ULICA_ADR!.split() [-1]`. Poprawiało to sprawę imion i nazwisk, lecz jednocześnie psuło wielocłonowe nazwy ulic. Przeprowadzone powtórne geokodowanie dało następujące wyniki: dopasowane 10 701 (54%), powiązane 1196 (6%), niedopasowane 7927 (40%).

Do geokodowania 23 284 zdarzeń, lokalizowanych względem skrzyżowań, ponownie posłużono się oprogramowaniem ArcGIS, lecz tym razem zbudowano lokator adresów dla geokodowania względem ulic. W tym przypadku, w bazie referencyjnej już od pierwszego podejścia wykorzystano nazwy ulic skrócone do ostatniego członu. Ponownie okazało się, że najlepsze rezultaty dało pozostawienie opcji dopasowania na domyślnych poziomach. Uzyskano następujące wyniki: dopasowane 9019 (39%), powiązane 5518 (24%) i niedopasowane 8747 (38%).

Do określenia miejsc wystąpienia 1243 zdarzeń, lokalizowanych przez odniesienie do kilometrażu dróg, posłużono się wtyczką LRS działającą w środowisku programu QGIS (QGIS Development Team, 2014). Narzędzie to pozwala na wykonanie dwóch zasadniczych zadań: zbudowania liniowego systemu odniesienia (włączając w to jego kalibrację), a następnie lokalizację obiektów i zjawisk opisanych odpowiednimi miarami.

Pierwszym krokiem było zbudowanie tras, odpowiadających poszczególnym drogom. Każda droga w OpenStreetMap składa się z wielu krótkich odcinków, które trzeba było połączyć w części tras. W tym celu zostały powiązane sąsiadujące ze sobą elementy geometryczne, opisane tym samym numerem drogi. Dodatkowo, punkty reprezentujące słupki pikietażowe wraz z ich miarami użyto do kalibracji tak utworzonych tras. Na tym etapie konieczne było poprawienie opisów, zawierających numery kilku przebiegających jednocze-



śnie na niektórych odcinkach dróg w taki sposób, żeby pozostał tylko jeden numer najważniejszej w danym miejscu drogi.

Po ukończeniu kalibracji można było przejrzeć listę wykrytych błędów. Najpoważniejsze były związane z faktem, że część dróg była dwujezdniowa. Wtyczka LRS nie dopuszcza równoległych linii opisanych tym samym numerem drogi. Rozwiązaniem zgodnym z zarządzeniem nr 18 w sprawie zasad ustalania i prowadzenia kilometrażu dróg krajowych (GDDKiA, 2012), które stwierdza, że kilometraż wyznaczany jest w odniesieniu do osi drogi głównej zarówno dla dróg jednojezdniowych, jak i wielojezdniowych, byłoby pracochłonne zastąpienie dwóch jezdni jedną linią. Prostszy rozwiązaniem okazało się usunięcie ze zbioru danych linii, reprezentujących lewą jezdnię. Podstawę do takiego postępowania dało wcześniejsze, nieobowiązujące już zarządzenia nr 14 w sprawie zasad ustalania i prowadzenia kilometrażu dróg krajowych (GDDKiA, 2003) stwierdzające, że pomiar kilometrażu prowadzony jest dla dróg wielojezdniowych w osi jezdni znajdującej się po prawej stronie drogi, zgodnie z rosnącym kilometrażem. Geokodowanie przeprowadzone po wprowadzeniu tych poprawek pozwoliło bez najmniejszego problemu określić położenie wszystkich zdarzeń.

Oczywiście, lokalizacje zdarzeń opisane współrzędnymi nie wymagały już żadnych dodatkowych działań. W tabeli 2 zestawiono parametry statystyczne dokumentujące skuteczność zastosowanych metod geokodowania zdarzeń drogowych.

**Tabela 2.** Zestawienie parametrów statystycznych ilustrujących skuteczność poszczególnych metod geokodowania

Metoda geokodowania	Lokalizacje		
	dopasowane	powiązane	niedopasowane
	[%]		
Geokodowanie adresów	54	6	40
Geokodowanie skrzyżowań	39	24	38
Liniowy system odniesienia	100	0	0
Lokalizacja przez współrzędne	100	0	0

## Wnioski

Przeprowadzone badania pokazały, że obowiązujący obecnie sposób dokumentowania lokalizacji zdarzeń drogowych jest zupełnie nieodpowiedni do automatycznego przeprowadzania geokodowania miejsc wypadków i kolizji. Podstawowym problemem jest duża dowolność w sposobie zapisywania poszczególnych informacji określających lokalizację zdarzenia.

Z jednej strony należałoby więc dążyć do tego, żeby dane wprowadzane były do bazy w sposób bardziej usystematyzowany (Lewandowicz, Packa, 2011). W szczególności wskazane byłoby zobligowanie policjantów wypełniających karty zdarzenia do przestrzegania wszystkich zapisów zarządzenia Komendanta Głównego Policji (2006, 2012, 2013). Chodzi zwłaszcza o zawartość pola NUMER\_DOMU, które ewidentnie powinno mieć zawartość wyłącznie liczbową, chociaż podawanie sugerowanego w rozporządzeniu numeru latarni na pewno nie wpłynie na poprawę jakości procesu geokodowania. Problem leży też, przynajmniej częściowo, po stronie bazy danych OpenStreetMap, w której nazwy ulic wywodzące się od osób, najczęściej są zapisywane w postaci imienia i nazwiska, gdy tymczasem w kartach zdarzeń drogowego stosowano zapis skrócony do samego nazwiska.

Dla pokonania dwóch powyższych przeszkód można by posłużyć się metodą zaproponowaną w pracy Levine'a i Kim'a (1998). Sformułowano tam i zapisano reguły, powodujące konwersję opisów lokalizacji, pochodzących z raportów policyjnych, na postać zgodną z regułami obowiązującymi w materiale źródłowym. Jest to jednak działanie pracochłonne, gdyż w cytowanym przypadku Honolulu reguł takich było ponad tysiąc dwieście. Jednak pewną zaletą takiego rozwiązania jest to, że wykonane byłyby tylko jednokrotnie i z dużym prawdopodobieństwem mogłyby być zastosowane do kolejnych, nowych danych dodawanych do bazy zdarzeń. W podobny sposób można by także rozważyć zdefiniowanie reguł, które odwzorowywałyby punkty charakterystyczne, takie jak stacje benzynowe lub centra handlowe od razu na odpowiednie współrzędne.

Oczywiście w każdym przypadku istotna jest aktualność bazy danych referencyjnych, która ma szczególne znaczenie, gdy baza ta jak na przykład OSM jeszcze dynamicznie się rozwija.

Wobec skali problemu, z jednoznacznością identyfikacją miejsca zdarzenia wydaje się jednak, że najlepszym rozwiązaniem byłoby zastąpienie dotychczasowej papierowej karty zdarzenia drogowego formularzem elektronicznym, w którym osoba rejestrująca zdarzenie mogłaby jednoznacznie wskazywać na podkładzie mapowym punkt, w którym zdarzenie miało miejsce. Dodatkowo, umożliwiłoby to rezygnację z konieczności podawania pewnych dodatkowych cech zdarzenia, takich jak: informacja o obszarze zabudowanym lub rodzaj drogi, które mogłyby być automatycznie wyznaczane metodami relatywnie prostych analiz geograficznych, na podstawie bazy danych przestrzennych. Ponadto, możliwe byłoby wtedy zastosowanie słowników dla pewnych powtarzających się zapisów, co pozwoliłoby uniknąć błędów literowych związanych z wpisywaniem odręcznym.

### Literatura

- Brzuchowska J., 2010: Nowy wymiar planowania systemów usług miejskich dzięki geokodowaniu. *Roczniki Geomatyki* t. 8, z. 1(37): 17–21.
- Charif O., Omrani H., Klein O., Schneider M., Trigano P., 2010: A method and a tool for geocoding and record linkage. *CEPS/INSTEAD Working Paper Series* 2010-17.
- Cichociński P., 2010: Zastosowanie systemów informacji geograficznej do analizy rynku nieruchomości. III Krajowa Konferencja Naukowa „Technologie Przetwarzania Danych”. Materiały konferencyjne. Poznań, 21-23 czerwca 2010 r. WNT, Warszawa.
- Curtin K.M., Nicoara G., Arifin R.R., 2007: A comprehensive process for linear referencing. *URISA Journal*, 19(2): 41–50.
- Curtin K.M., 2007: Network analysis in geographic information science: Review, assessment, and projections. *Cartography and Geographic Information Science*, 34(2): 103–111.
- Esri, 1998: ESRI Shapefile Technical Description, An ESRI White Paper. Environmental Systems Research Institute, Redlands.
- Esri, 2011: ArcGIS Desktop 10 Help.
- Furieri A., 2012: OSM Tools. <https://www.gaia-gis.it/fossil/spatialite-tools/wiki?name=OSM+tools>
- GDDKiA (Generalny Dyrektor Dróg Krajowych i Autostrad), 2003: Zarządzenie nr 14 z 27 października 2003 roku w sprawie zasad ustalania i prowadzenia kilometrażu dróg krajowych.
- GDDKiA (Generalny Dyrektor Dróg Krajowych i Autostrad), 2012: Zarządzenie nr 18 z 1 czerwca 2012 roku w sprawie zasad ustalania i prowadzenia kilometrażu dróg krajowych.
- Hijar M., Trostle J., Bronfman M., 2003: Pedestrian injuries in Mexico: a multi-method approach. *Social science & medicine*, 57(11): 2149–2159.
- Karimi H.A., Durcik M., Rasdorf W., 2004: Evaluation of uncertainties associated with geocoding techniques. *Computer-Aided Civil and Infrastructure Engineering* 19: 170–185.
- KGP (Komendant Główny Policji), 2006: Zarządzenie nr 635 z 30 czerwca 2006 roku w sprawie metod i form prowadzenia przez policję statystyki zdarzeń drogowych, Dz.Urz. KGP nr 11, poz. 67.

- KGP (Komendant Główny Policji), 2012: Zarządzenie nr 123 z 31 maja 2012 roku zmieniające zarządzenie w sprawie metod i form prowadzenia przez Policję statystyki zdarzeń drogowych Dz.Urz. KGP 2012.28.
- KGP (Komendant Główny Policji), 2013: Obwieszczenie z 22 sierpnia 2013 roku w sprawie ogłoszenia jednolitego tekstu zarządzenia Komendanta Głównego Policji w sprawie metod i form prowadzenia przez Policję statystyki zdarzeń drogowych Dz.Urz. KGP 2013.75.
- Levine N., Kim K.E., 1998: The location of motor vehicle crashes in Honolulu: A methodology for geocoding Intersections. *Computers, Environment, and Urban Systems* 22(6): 557–576.
- Lewandowicz E., Packa A., 2011: Interoperacyjność danych w praktyce: problemy harmonizacji nazw jednostek administracyjnych w rejestrach publicznych. *Roczniki Geomatyki* t. 9, z. 3(47): 95–104, PTIP Warszawa.
- Li L., Zhu L., Sui D.Z., 2007: A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes. *Journal of Transport Geography* 15(4): 274–285.
- Meliker J.R., Maio R.F., Zimmerman M.A., Kim H. M., Smith S. C., Wilson M. L., 2004: Spatial analysis of alcohol-related motor vehicle crash injuries in southeastern Michigan. *Accident Analysis & Prevention* 36(6), 1129–1135.
- Michalak J., 2007: Otwarte oprogramowanie i otwarte dane w geomatyce. *Roczniki Geomatyki* t. 5, z. 2: 11–20.
- Paul D.L., 2003: A geocoded National Address File for Australia: The G-NAF What, Why, Who and When? PSMA Australia Limited, Griffith, ACT, Australia. <http://www.g-naf.com.au/>
- PN-EN ISO 19116:2006: Informacja geograficzna – Usługi wyznaczania położenia.
- QGIS Development Team, 2014: QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>
- Ratcliffe J.H., 2004: Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science*, 18(1): 61–72.
- Rushton G., Armstrong M.P., Gitler J., Green B.R., Pavlik C., West M.M., Zimmerman D.L., 2006: Geocoding in cancer research : a review. *American Journal of Preventive Medicine* 30(2): 16–24.
- Smith R.C., Harkey D.L., Harris B., 2001: Implementation of GIS-based highway safety analyses: bridging the gap. Publication no. FHWA-RD-01-039. Federal Highway Administration.
- Vieira V.M., Howard G.J., Gallagher L.G., Fletcher T., 2010: Geocoding rural addresses in a community contaminated by PFOA: a comparison of methods. *Environmental Health*: 9–18.
- Zandbergen P.A., 2007: Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. *BMC Public Health*: 7–37.
- Zandbergen P.A., 2008: A comparison of address point, parcel and street geocoding techniques. *Computers, Environment and Urban Systems*, 32: 214–232.
- Zandbergen P.A., 2011: Influence of street reference data on geocoding quality. *Geocarto International* Vol. 26, No. 1, February 2011: 35–47.
- Zhou X., Jiang Y., Zhou K., Zeng L., 2013: A dynamic integration method for borderland database using OSM data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* Vol. XL-4/W3: 141–147.
- Zimmerman D.L., Li J., 2010: The effects of local street network characteristics on the positional accuracy of automated geocoding for geographic health studies. *International Journal of Health Geographics*: 9–10.

### Streszczenie

Geokodowanie jest to proces określania położenia obiektów wyrażonego za pomocą różnego rodzaju identyfikatorów geograficznych, najczęściej adresu. Odbyna się poprzez porównanie odpowiednich elementów informacji adresowej z pewnym materiałem odniesienia. Podstawą skutecznego przeprowadzenia geokodowania jest z jednej strony posiadanie odpowiedniego materiału źródłowego, który określi położenie adresu, natomiast z drugiej precyzyjne opisanie poszukiwanej lokalizacji. Szczególnymi obiektami, których położenie trzeba określać tą metodą, są miejsca zdarzeń drogowych, takich jak wypadki i kolizje. Niewiele z nich ma bowiem położenie wyrażone za pomocą współrzędnych. W większości przypadków lokalizacja ogranicza się do wskazania najbliższego adresu, kilometrażu drogi, skrzyżowania lub nawet swobodnego opisu.

Najłatwiej dostępnym w Polsce źródłem informacji adresowej jest OpenStreetMap (OSM) – baza danych, która powstaje na podstawie danych gromadzonych przez wolontariuszy. Jest ona udostępniana nieodpłatnie i dla dowolnego zastosowania. Jednym z jej elementów jest informacja adresowa. Ponadto zaletą OSM jest także to, że w przeciwieństwie do większości popularnych portali mapowych rejestruje położenie słupków pikietażowych, umieszczanych wzdłuż dróg.

Wobec dostępności odpowiedniego oprogramowania i stwierdzonej w literaturze porównywalności stosowanych algorytmów celem pracy nie jest analiza samego procesu geokodowania, lecz z jednej strony zbadanie dostępności dla obszaru Polski odpowiednich danych referencyjnych, a z drugiej próba zwrócenia uwagi na sposób zapisu informacji adresowej, która ma podlegać geokodowaniu, ze szczególnym uwzględnieniem opisu lokalizacji zdarzeń drogowych. Zbadano, czy wobec zupełnie niesformalizowanego opisu lokalizacji geokodowanie w ogóle będzie możliwe do przeprowadzenia. Zaproponowano metody geokodowania dla występujących w praktyce różnych form opisu miejsca zdarzenia oraz zasugerowano sposoby rozwiązania problemów pojawiających się podczas geokodowania.

### **Abstract**

Geocoding is the process of determining the position of features, expressed by different kinds of geographic identifiers, usually addresses. This is done by comparing the relevant elements of address information with reference material. The basis for effective implementation of geocoding is, on the one hand, right source material, which determines the position of the address, and, on the other hand, precise description of the required location. Particular features, whose positions need to be specified using this method are traffic incidents such as accidents and collisions. Few of them have in fact the position expressed by coordinates. In most cases the location is limited to indication of the nearest address, road chainage, intersection, or even free-form description.

The most easily accessible source of address information in Poland is OpenStreetMap (OSM) – a database that is created on the basis of data collected by volunteers. It is provided free of charge and for any application. One of its elements is address information. In addition, OSM has also the advantage that, unlike most popular mapping portals, it records kilometer posts, placed along roads.

Considering availability of appropriate software and comparability of used algorithms found in the literature, the objective of this study was not to analyze the geocoding process itself, but on the one hand to investigate the availability of relevant reference data for the area of Poland, and on the other hand an attempt to draw attention to the method of recording address information, which is subject to geocoding, with particular emphasis on the location of traffic incidents. It was examined whether because of completely informal description of the location, geocoding in general will be possible to carry out. Methods of geocoding of different forms of description of incident location existing in practice were proposed, and ways to solve problems that arise during geocoding were suggested.

dr hab. inż. Piotr Cichociński  
Piotr.Cichocinski@agh.edu.pl