

Sektor mikro, małych i średnich przedsiębiorstw – geokodowanie danych MŚP



Projekt „Usytuowanie na poziomie samorządów lokalnych instrumentów wsparcia dla MŚP, działających w oparciu o model wielopoziomowego zarządzania regionem” realizowany przez konsorcjum w składzie: Województwo Kujawsko-Pomorskie, Szkoła Główna Handlowa w Warszawie i Uniwersytet Mikołaja Kopernika w Toruniu”; finansowany z Narodowego Centrum Badań i Rozwoju w ramach Programu Strategicznego Gospostrateg – „Społeczny i Gospodarczy Rozwój Polski w warunkach globalizujących się rynków”
Nr umowy GOSPOSTRATEG1/385453/3/NCBR/2018

Spis treści

<u>OPRACOWANIE METODYKI GEOKODOWANIA MŚP W WOJEWÓDZTWIE KUJAWSKO-POMORSKIM NA PODSTAWIE DANYCH OPISOWYCH REGON OBEJMUJĄCEGO OKOŁO 190 000 MAŁYCH I ŚREDNICH PRZEDSIĘBIORSTW</u>	<u>3</u>
<u>ANALIZA DANYCH.....</u>	<u>12</u>
<u>METODYKA GEOKODOWANIA.....</u>	<u>14</u>

Opracowanie metodyki geokodowania MŚP w województwie kujawsko-pomorskim na podstawie danych opisowych REGON obejmującego około 190 000 małych i średnich przedsiębiorstw

Celem zadania było geokodowanie danych pochodzących z rejestru REGON dla małych i średnich firm mających siedziby na terenie województwa kujawsko-pomorskiego. Dane REGON zostały udostępnione Marszałkowi Województwa Kujawsko-Pomorskiego przez Urząd Statystyczny w Bydgoszczy 15.03.2019 r. na podstawie PG-II-G.052.1.2019 z 12.03.2019.

Udostępnione dane zostały dostarczone Wykonawcy w postaci arkusza kalkulacyjnego Excel (plik BDG-WOR.6010.6.2019.5_Załącznik 1) zawierającego 193092 rekordów. Udostępniona baza danych REGON zawiera następujące informacje o podmiocie:

- pełną nazwę,
- numer identyfikacyjny REGON,
- adres siedziby (o ile posiadamy),
- działalność przeważającą według PKD 2007 (sekcja i dział, o ile posiadamy).
- podstawową formę prawną,
- szczególną formę prawną,
- formę własności (o ile posiadamy),
- stan aktywności prawnej i ekonomicznej,
- daty: powstania podmiotu, rozpoczęcia działalności, zawieszenia i wznowienia działalności, wpisu do ewidencji lub rejestru, orzeczenia o ogłoszeniu upadłości, zakończenia postępowania upadłościowego, data zgonu,
- liczbę pracujących w grupach przedziałowych (A: 0-9 osób; B: 10-49 osób; C: 50-249 osób; D: 250-999 osób; E: 1000 i więcej osób).

Dane te zawierają informacje zarówno o aktywnych (podmiot prowadzący działalność), zawieszonych, jak i zlikwidowanych MŚP (podmiot z zakończoną działalnością).

Szczegółowe zestawienie i objaśnienie cech poszczególnych podmiotów zawierają poniższe zestawienia:

Podstawowe Formy Prawne (pfp)

- 1 OSOBA PRAWNA
- 2 JEDNOSTKA ORGANIZACYJNA NIEMAJĄCA OSOBOWOŚCI PRAWNEJ
- 9 OSOBA FIZYCZNA PROWADZĄCA DZIAŁALNOŚĆ GOSPODARCZĄ

Szczególne Formy Prawne (sfp)

- 019 SPÓŁKI CYWILNE PROWADZĄCE DZIAŁALNOŚĆ NA PODSTAWIE UMOWY ZAWARTEJ ZGODNIE Z KODEKSEM CYWILNYM
- 023 SPÓŁKI PRZEWDZIANE W PRZEPISACH INNYCH USTAW NIŻ KODEKS SPÓŁEK HANDLOWYCH I KODEKS CYWILNY LUB FORMY PRAWNE, DO KTÓRYCH STOSUJE SIĘ PRZEPISY O SPÓŁKACH
- 044 UCZELNIE
- 049 FUNDUSZE
- 050 KOŚCIÓŁ KATOLICKI
- 051 INNE KOŚCIOŁY I ZWIĄZKI WYZNANIOWE
- 053 EUROPEJSKIE UGRUPOWANIA WSPÓŁPRACY TERYTORIALNEJ
- 055 STOWARZYSZENIA NIEWPISANE DO KRS
- 060 ORGANIZACJE SPOŁECZNE ODDZIELNIE NIEWYMIENIONE NIEWPISANE DO KRS
- 070 PARTIE POLITYCZNE
- 076 SAMORZĄD GOSPODARCZY I ZAWODOWY NIEWPISANY DO KRS
- 080 PRZEDSTAWICIELSTWA ZAGRANICZNE

- 085 WSPÓLNOTY MIESZKANIOWE
- 099 OSOBY FIZYCZNE PROWADZĄCE DZIAŁALNOŚĆ GOSPODARCZĄ
- 114 EUROPEJSKIE ZGRUPOWANIA INTERESÓW GOSPODARCZYCH
- 115 SPÓŁKI PARTNERSKIE
- 116 SPÓŁKI AKCYJNE
- 117 SPÓŁKI Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ
- 118 SPÓŁKI JAWNE
- 120 SPÓŁKI KOMANDYTOWE
- 121 SPÓŁKI KOMANDYTOWO - AKCYJNE
- 122 SPÓŁKI EUROPEJSKIE
- 124 PRZEDSIĘBIORSTWA PAŃSTWOWE
- 126 TOWARZYSTWA UBEZPIECZEŃ WZAJEMNYCH
- 132 INSTYTUCJE GOSPODARKI BUDŻETOWEJ
- 133 ZWIĄZKI ZAWODOWE ROLNIKÓW INDYWIDUALNYCH
- 134 TOWARZYSTWA REASEKURACJI WZAJEMNEJ
- 135 GŁÓWNE ODDZIAŁY ZAGRANICZNYCH ZAKŁADÓW REASEKURACJI
- 136 GŁÓWNE ODDZIAŁY ZAGRANICZNYCH ODDZIAŁÓW UBEZPIECZEŃ
- 137 OGÓLNOKRAJOWE ZRZESZENIA MIĘDZYBRANŻOWE
- 138 OGÓLNOKRAJOWE ZWIĄZKI MIĘDZYBRANŻOWE
- 140 SPÓŁDZIELNIE
- 141 JEDNOSTKI BADAWCZO-ROZWOJOWE
- 142 SPÓŁDZIELNIE EUROPEJSKIE
- 143 ZWIĄZKI ROLNIKÓW, KÓŁEK I ORGANIZACJI ROLNICZYCH
- 145 ZWIĄZKI ROLNICZYCH ZRZESZEŃ BRANŻOWYCH
- 146 SAMODZIELNE PUBLICZNE ZAKŁADY OPIEKI ZDROWOTNEJ

- 147 CECHY RZEMIEŚLNICZE
- 148 FUNDACJE
- 152 IZBY RZEMIEŚLNICZE
- 154 ZWIĄZEK RZEMIOSŁA POLSKIEGO
- 155 STOWARZYSZENIA
- 156 ZWIĄZKI STOWARZYSZEŃ
- 157 STOWARZYSZENIA KULTURY FIZYCZNEJ
- 158 ZWIĄZKI SPORTOWE
- 159 POLSKIE ZWIĄZKI SPORTOWE
- 160 INNE ORGANIZACJE SPOŁECZNE LUB ZAWODOWE
- 161 KOLUMNY TRANSPORTU SANITARNEGO
- 162 STOWARZYSZENIA KULTURY FIZYCZNEJ O ZASIĘGU OGÓLNOKRAJOWYM
- 163 ZRZESZENIE HANDLU I USŁUG
- 164 ZRZESZENIE TRANSPORTU
- 165 INSTYTUTY BADAWCZE
- 166 OGÓLNOKRAJOWE REPREZENTACJE ZRZESZEŃ HANDLU I USŁUG
- 167 OGÓLNOKRAJOWE REPREZENTACJE ZRZESZEŃ TRANSPORTU
- 168 INNE ORGANIZACJE PODMIOTÓW GOSPODARCZYCH
- 169 IZBY GOSPODARCZE
- 171 PRZEDSIĘBIORSTWA ZAGRANICZNE
- 172 ZWIĄZKI ZAWODOWE
- 174 ZWIĄZKI PRACODAWCÓW
- 175 FEDERACJE/KONFEDERACJE ZWIĄZKÓW PRACODAWCÓW
- 177 KÓŁKA ROLNICZE
- 178 ROLNICZE ZRZESZENIA BRANŻOWE

- 179 ODDZIAŁY ZAGRANICZNYCH PRZEDSIĘBIORCÓW
- 180 SPÓŁDZIELCZE KASY OSZCZĘDNOŚCIOWO-KREDYTOWE
- 181 STOWARZYSZENIA OGRODOWE
- 182 ZWIĄZKI STOWARZYSZEŃ OGRODOWYCH
- 183 JEDNOSTKI TERENOWE STOWARZYSZEŃ POSIADAJĄCE OSOBOWOŚĆ PRAWNĄ
- 184 JEDNOSTKI ORGANIZACYJNE ZWIĄZKÓW ZAWODOWYCH POSIADAJĄCE OSOBOWOŚĆ PRAWNĄ
- 381 PRZEDSZKOLA PUBLICZNE
- 382 PRZEDSZKOLA NIEPUBLICZNE
- 383 PUBLICZNE SZKOŁY PODSTAWOWE
- 384 PUBLICZNE GIMNAZJA
- 385 PUBLICZNE SZKOŁY PONADPODSTAWOWE
- 386 PUBLICZNE SZKOŁY PONADGIMNAZJALNE
- 387 PUBLICZNE SZKOŁY ARTYSTYCZNE
- 388 NIEPUBLICZNE SZKOŁY PODSTAWOWE
- 389 NIEPUBLICZNE GIMNAZJA
- 390 NIEPUBLICZNE SZKOŁY PONADPODSTAWOWE
- 391 NIEPUBLICZNE SZKOŁY PONADGIMNAZJALNE
- 392 NIEPUBLICZNE SZKOŁY ARTYSTYCZNE
- 393 PUBLICZNE PLACÓWKI SYSTEMU OŚWIATY
- 394 NIEPUBLICZNE PLACÓWKI SYSTEMU OŚWIATY
- 395 INNE PUBLICZNE JEDNOSTKI ORGANIZACYJNE SYSTEMU OŚWIATY
- 396 INNE NIEPUBLICZNE JEDNOSTKI ORGANIZACYJNE SYSTEMU OŚWIATY
- 397 PUBLICZNE ZESPOŁY SZKÓŁ I PLACÓWEK SYSTEMU OŚWIATY
- 398 NIEPUBLICZNE ZESPOŁY SZKÓŁ I PLACÓWEK SYSTEMU OŚWIATY

- 401 ORGANY WŁADZY, ADMINISTRACJI RZĄDOWEJ
- 402 ORGANY KONTROLI PAŃSTWOWEJ I OCHRONY PRAWA
- 403 WSPÓLNOTY SAMORZĄDOWE
- 406 SĄDY I TRYBUNAŁY
- 409 SKARB PAŃSTWA
- 428 PAŃSTWOWE JEDNOSTKI ORGANIZACYJNE
- 429 GMINNE SAMORZĄDOWE JEDNOSTKI ORGANIZACYJNE
- 430 POWIATOWE SAMORZĄDOWE JEDNOSTKI ORGANIZACYJNE
- 431 WOJEWÓDZKIE SAMORZĄDOWE JEDNOSTKI ORGANIZACYJNE
- 439 INNE PAŃSTWOWE LUB SAMORZĄDOWE OSOBY PRAWNE W ROZUMIENIU ART. 9 PKT 14 USTAWY Z DNIA 27 SIERPNIĄ 2009 R. O FINANSACH PUBLICZNYCH (DZ. U. Z 2013 POZ. 885, Z PÓŹN. ZM.)
- 999 BEZ SZCZEGÓLNEJ FORMY PRAWNEJ

Liczba osób pracujących w grupach przedziałowych (lprac)

Liczba osób pracujących

- A 0 – 9
- B 10 – 49
- C 50 – 249
- D 250 – 999
- E 1000 i więcej

Forma Własności (fw)

- 111 WŁASNOŚĆ SKARBU PAŃSTWA
- 112 WŁASNOŚĆ PAŃSTWOWYCH OSÓB PRAWNYCH
- 113 WŁASNOŚĆ JEDNOSTEK SAMORZĄDU TERYTORIALNEGO LUB SAMORZĄDOWYCH OSÓB PRAWNYCH
- 121 WŁASNOŚĆ MIESZANA W SEKTORZE PUBLICZNYM Z PRZEWAGĄ WŁASNOŚCI SKARBU PAŃSTWA
- 122 WŁASNOŚĆ MIESZANA W SEKTORZE PUBLICZNYM Z PRZEWAGĄ WŁASNOŚCI PAŃSTWOWYCH OSÓB PRAWNYCH
- 123 WŁASNOŚĆ MIESZANA W SEKTORZE PUBLICZNYM Z PRZEWAGĄ WŁASNOŚCI JEDNOSTEK SAMORZĄDU TERYTORIALNEGO LUB SAMORZĄDOWYCH OSÓB PRAWNYCH
- 127 WŁASNOŚĆ MIESZANA W SEKTORZE PUBLICZNYM Z BRAKIEM PRZEWAGI KTÓREGOKOLWIEK RODZAJU WŁASNOŚCI PUBLICZNEJ
- 131 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z PRZEWAGĄ WŁASNOŚCI SEKTORA PUBLICZNEGO, W TYM Z PRZEWAGĄ WŁASNOŚCI SKARBU PAŃSTWA
- 132 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z PRZEWAGĄ WŁASNOŚCI SEKTORA PUBLICZNEGO, W TYM Z PRZEWAGĄ WŁASNOŚCI PAŃSTWOWYCH OSÓB PRAWNYCH
- 133 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z PRZEWAGĄ WŁASNOŚCI SEKTORA PUBLICZNEGO, W TYM Z PRZEWAGĄ WŁASNOŚCI JEDNOSTEK SAMORZĄDU TERYTORIALNEGO LUB SAMORZĄDOWYCH OSÓB PRAWNYCH
- 137 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z PRZEWAGĄ WŁASNOŚCI SEKTORA PUBLICZNEGO Z BRAKIEM PRZEWAGI KTÓREGOKOLWIEK RODZAJU WŁASNOŚCI PUBLICZNEJ
- 214 WŁASNOŚĆ KRAJOWYCH OSÓB FIZYCZNYCH
- 215 WŁASNOŚĆ PRYWATNA KRAJOWA POZOSTAŁA
- 216 WŁASNOŚĆ ZAGRANICZNA
- 224 WŁASNOŚĆ MIESZANA W SEKTORZE PRYWATNYM Z PRZEWAGĄ WŁASNOŚCI KRAJOWYCH OSÓB FIZYCZNYCH

- 225 WŁASNOŚĆ MIESZANA W SEKTORZE PRYWATNYM Z PRZEWAGĄ WŁASNOŚCI PRYWATNEJ KRAJOWEJ POZOSTAŁEJ
- 226 WŁASNOŚĆ MIESZANA W SEKTORZE PRYWATNYM Z PRZEWAGĄ WŁASNOŚCI ZAGRANICZNEJ
- 227 WŁASNOŚĆ MIESZANA W SEKTORZE PRYWATNYM Z BRAKIEM PRZEWAGI KTÓREGOKOLWIEK RODZAJU WŁASNOŚCI PRYWATNEJ
- 234 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z PRZEWAGĄ WŁASNOŚCI SEKTORA PRYWATNEGO, W TYM Z PRZEWAGĄ WŁASNOŚCI KRAJOWYCH OSÓB FIZYCZNYCH
- 235 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z PRZEWAGĄ WŁASNOŚCI SEKTORA PRYWATNEGO, W TYM Z PRZEWAGĄ WŁASNOŚCI PRYWATNEJ KRAJOWEJ POZOSTAŁEJ
- 236 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z PRZEWAGĄ WŁASNOŚCI SEKTORA PRYWATNEGO, W TYM Z PRZEWAGĄ WŁASNOŚCI ZAGRANICZNEJ
- 237 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z PRZEWAGĄ WŁASNOŚCI SEKTORA PRYWATNEGO Z BRAKIEM PRZEWAGI KTÓREGOKOLWIEK RODZAJU WŁASNOŚCI PRYWATNEJ
- 331 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z TAKIM SAMYM UDZIAŁEM WŁASNOŚCI SEKTORA PUBLICZNEGO I PRYWATNEGO Z NAJWYŻSZYM UDZIAŁEM WŁASNOŚCI SKARBU PAŃSTWA W KAPITALE OGÓŁEM
- 332 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z TAKIM SAMYM UDZIAŁEM WŁASNOŚCI SEKTORA PUBLICZNEGO I PRYWATNEGO Z NAJWYŻSZYM UDZIAŁEM WŁASNOŚCI PAŃSTWOWYCH OSÓB PRAWNYCH W KAPITALE OGÓŁEM
- 333 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z TAKIM SAMYM UDZIAŁEM WŁASNOŚCI SEKTORA PUBLICZNEGO I PRYWATNEGO Z NAJWYŻSZYM UDZIAŁEM WŁASNOŚCI JEDNOSTEK SAMORZĄDU TERYTORIALNEGO LUB SAMORZĄDOWYCH OSÓB PRAWNYCH
- 334 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z TAKIM SAMYM UDZIAŁEM WŁASNOŚCI SEKTORA PUBLICZNEGO I PRYWATNEGO Z NAJWYŻSZYM UDZIAŁEM WŁASNOŚCI KRAJOWYCH OSÓB FIZYCZNYCH W KAPITALE OGÓŁEM
- 335 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z TAKIM SAMYM UDZIAŁEM WŁASNOŚCI SEKTORA PUBLICZNEGO I PRYWATNEGO Z NAJWYŻSZYM UDZIAŁEM WŁASNOŚCI POZOSTAŁYCH KRAJOWYCH JEDNOSTEK PRYWATNYCH W KAPITALE OGÓŁEM

336 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z TAKIM SAMYM UDZIAŁEM WŁASNOŚCI SEKTORA PUBLICZNEGO I PRYWATNEGO Z NAJWYŻSZYM UDZIAŁEM WŁASNOŚCI OSÓB ZAGRANICZNYCH W KAPITALE OGÓŁEM

338 WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z TAKIM SAMYM UDZIAŁEM WŁASNOŚCI SEKTORA PUBLICZNEGO I PRYWATNEGO Z BRAKIEM PRZEWAGI KTÓREGOKOLWIEK RODZAJU WŁASNOŚCI W KAPITALE OGÓŁEM

0 BRAK FORMY WŁASNOŚCI

Stan aktywności prawnej i ekonomicznej (ape)

01 PODMIOT PROWADZĄCY DZIAŁALNOŚĆ

02 PODMIOT JESZCZE NIE PODJĄŁ DZIAŁALNOŚCI

03 PODMIOT W STANIE LIKWIDACJI

04 PODMIOT Z ZAKOŃCZONĄ DATĄ LIKWIDACJI

05 PODMIOT W STANIE UPADŁOŚCI

06 PODMIOT Z ZAKOŃCZONYM POSTĘPOWANIEM UPADŁOŚCIOWYM

07 PODMIOT Z ZAWIESZONĄ DZIAŁALNOŚCIĄ

08 PODMIOT Z ZAKOŃCZONĄ DZIAŁALNOŚCIĄ

09 PODMIOT OŻYWIONY – BŁĘDNE INFORMACJE

10 PODMIOT NIE PODJĄŁ DZIAŁALNOŚCI

Analiza danych

Informacje zawarte w powyższych tabelach mają kluczowe znaczenie nie tylko dla procesu geokodowania (adres MŚP), ale także właściwej klasyfikacji poszczególnych instytucji. Znaczna część udostępnionych przez Urząd Statystyczny w Bydgoszczy nie dotyczy bowiem bezpośrednio aktywnych MŚP. Przykładowo agregacja danych względem atrybutu ape (stan aktywności prawnej i ekonomicznej) wskazuje na funkcjonowanie w bazie REGON blisko 20 tysięcy obiektów z zawieszoną działalnością, bądź też będących w stanie likwidacji, upadłości itp.

ape	znaczenie	liczebność
01	PODMIOT PROWADZĄCY DZIAŁALNOŚĆ	176874
07	PODMIOT Z ZAWIESZONĄ DZIAŁALNOŚCIĄ	16272
03	PODMIOT W STANIE LIKWIDACJI	574
08	PODMIOT Z ZAKOŃCZONĄ DZIAŁALNOŚCIĄ	369
02	PODMIOT JESZCZE NIE PODJĄŁ DZIAŁALNOŚCI	305
05	PODMIOT W STANIE UPADŁOŚCI	135
06	PODMIOT Z ZAKOŃCZONYM POSTĘPOWANIEM UPADŁOŚCIOWYM	17
04	PODMIOT Z ZAKOŃCZONĄ DATĄ LIKWIDACJI	17
09	PODMIOT OŻYWIONY – BŁĘDNE INFORMACJE	2

Interesująca jest także agregacja danych źródłowych zrealizowana względem atrybutu sfp (Szczególne Formy Prawne) oraz atrybutu fw (Forma Własności). Poniższe zestawienia prezentują wybrany fragment analizy (5 pierwszych rekordów zestawionych pod względem liczebności).

sfp	znaczenie	liczebność
099	OSOBY FIZYCZNE PROWADZĄCE DZIAŁALNOŚĆ GOSPODARCZĄ	142 644
019	SPÓŁKI CYWILNE PROWADZĄCE DZIAŁALNOŚĆ NA PODSTAWIE UMOWY ZAWARTEJ ZGODNIE Z KODEKSEM CYWILNYM	13 590
117	SPÓŁKI Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ	12 546
085	WSPÓLNOTY MIESZKANIOWE	7 061

155	STOWARZYSZENIA	3 525
...

fw	znaczenie	liczebność
214	WŁASNOŚĆ KRAJOWYCH OSÓB FIZYCZNYCH	168 791
215	WŁASNOŚĆ PRYWATNA KRAJOWA POZOSTAŁA	12 793
113	WŁASNOŚĆ JEDNOSTEK SAMORZĄDU TERYTORIALNEGO LUB SAMORZĄDOWYCH OSÓB PRAWNYCH	3 326
234	WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI Z PRZEWAGĄ WŁASNOŚCI SEKTORA PRYWATNEGO, W TYM Z PRZEWAGĄ WŁASNOŚCI KRAJOWYCH OSÓB FIZYCZNYCH	2 184
216	WŁASNOŚĆ ZAGRANICZNA	1 646
...

fw	znaczenie	liczebność
214	WŁASNOŚĆ KRAJOWYCH OSÓB FIZYCZNYCH	168 791
215	WŁASNOŚĆ PRYWATNA KRAJOWA POZOSTAŁA	12 793
113	WŁASNOŚĆ JEDNOSTEK SAMORZĄDU TERYTORIALNEGO LUB SAMORZĄDOWYCH OSÓB PRAWNYCH	3 326
234	WŁASNOŚĆ MIESZANA MIĘDZY SEKTORAMI (..)	2 184
216	WŁASNOŚĆ ZAGRANICZNA	1 646
...

Analizy te pokazują, iż udostępnione dane zawierają zarówno informacje o MŚP (istniejących, zawieszonych i zlikwidowanych), jak również o własności Skarbu Państwa, wyższych uczelniach, sądach, szkołach, przychodniach, szpitalach itp.

Metodyka geokodowania

Wobec złożoności informacji zawartych w udostępnionej bazie REGON Wykonawca podjął starania opracowania bazy danych przestrzennych, która zawierając dane lokalizowane przestrzennie, umożliwi zarazem analizę wszystkich zawartych w tabeli źródłowej informacji opisowych dotyczących formy własności, stanu aktywności, liczby pracowników itp.

Istotą geokodowania jest proces integracji danych o charakterze opisowym z informacją lokalizowaną przestrzennie realizowany za pomocą odniesienia adresu do współrzędnych o charakterze georeferencyjnym. W celu wykonania procesu geokodowania niezbędne było pozyskanie niezależnych od udostępnionej bazy REGON danych geoprzestrzennych. Istnieje wiele opracowań – zarówno o charakterze urzędowym, jak i społecznościowym i komercyjnym, które umożliwiają realizację procesu geokodowania.

Do danych georeferencyjnych o charakterze urzędowym należy zasób PRG (Państwowy Rejestr Granic). Dane te można pobrać bezpłatnie ze strony Centralnego Ośrodka Dokumentacji Geodezyjnej i Kartograficznej - <http://www.gugik.gov.pl/pzgif/dane-bez-oplat/dane-z-panstwowego-rejestru-granic-i-powierzchni-jednostek-podzialow-terytorialnych-kraju-prg>.

Do danych georeferencyjnych o charakterze społecznościowym należy OSM (Open Street Map), zaś komercyjnym – np. serwis Google Maps udostępniający API służące do realizacji procesu geokodowania.

Należy podkreślić, iż uzyskane wyniki, tj. lokalizacja przestrzenna poszczególnych instytucji powstała na skutek procesu geokodowania jest istotnie zależna od dwóch czynników:

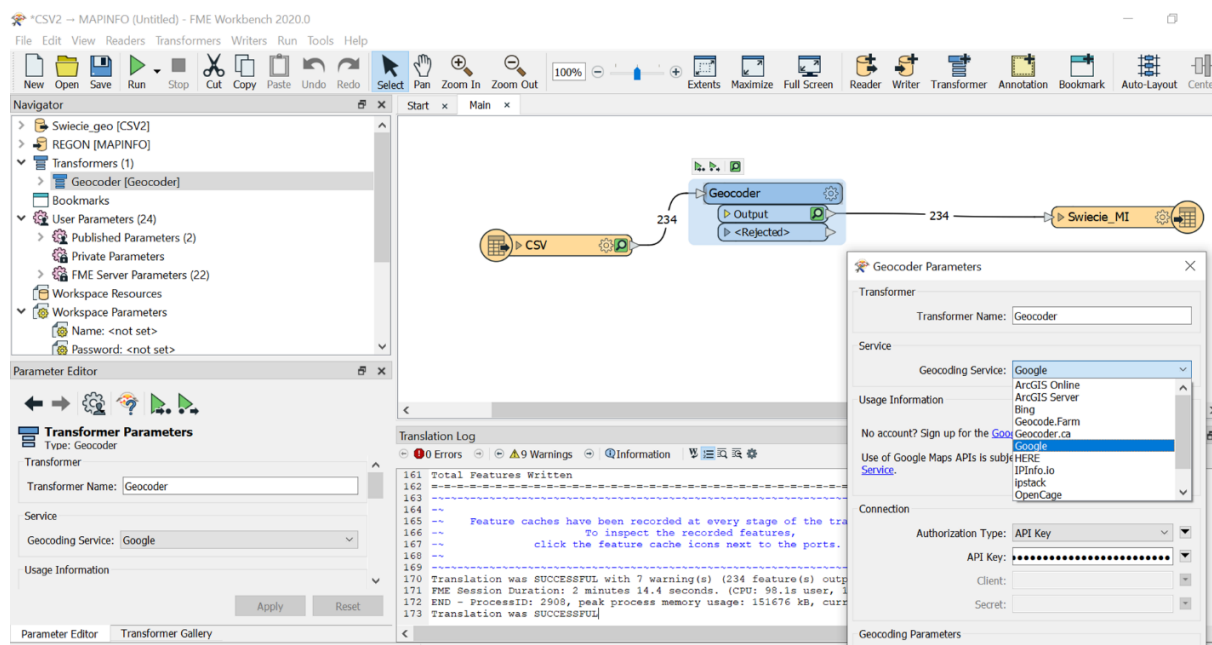
- jakości, aktualności i wiarygodności wykorzystanego serwisu geolokalizacyjnego
- poprawności zapisu poszczególnych adresów zawartych w udostępnionej bazie danych REGON.

Lokalizacja przestrzenna poszczególnych punktów adresowych może różnić się w zależności od wykorzystywanego źródła danych przestrzennych. W przypadku wykorzystania serwisów urzędowych wiarygodność punktów adresowych jest z reguły wyższa niż dla danych geolokalizacyjnych udostępnianych np. w Open Street Map, jednakże aktualność OSM z reguły jest wyższa niż danych urzędowych. Serwis Google Maps umożliwia geokodowanie największej liczby rekordów, jednakże jego wykorzystanie ma charakter komercyjny, zaś lokalizacja przestrzenna obiektów jest mniej precyzyjna niż w przypadku danych PRG.

Kluczowe znaczenie dla poprawności procesu geokodowania (lub wręcz możliwości jego realizacji) ma poprawność i kompletność zapisu informacji adresowej w bazie danych REGON. Dane o charakterze niepełnym (np. brak informacji o miejscowości), błędna nazwa ulicy czy też sposób zapisu numeru adresowego prowadzą do powstawania problemów z uzyskaniem georeferencji. Przykładowo: zapisy Ul. K. Wielkiego 13A, Kazimierza Wielkiego 13 A, Kazimierza Wlk.13a odwołują się do tej samej lokalizacji, jednak przyjęty sposób notacji uniemożliwia, lub znacznie utrudnia, automatyzację procesu geokodowania.

Do geokodowania wykorzystano szereg narzędzi geoinformatycznych, np. pakietów GIS: ArcGIS, QGIS, MapInfo oraz narzędzi dedykowanych transformacji informacji przestrzennej – FME.

Fragment struktury procesu geokodowania z wykorzystaniem narzędzia FME oraz API serwisu Google Maps przedstawia poniższy diagram.



Wobec występowania w udostępnionej bazie danych znacznej liczby błędów adresowych obejmującej około 15% danych źródłowych, niezbędne było zarówno ręczne modyfikowanie poszczególnych rekordów, jak i iteracyjne geokodowanie informacji.

Proces ten miał charakter czteroetapowy:

- bezpośrednie geokodowanie udostępnionych danych z wykorzystaniem punktów adresowych Państwowego Rejestru Granic,
- analiza błędów występujących w rekordach, które nie zostały zgeokodowane w sposób automatyczny realizowane poprzez autorskie skrypty opracowane przez Wykonawcę,
- powtórne geokodowanie pozostałych obiektów z wykorzystaniem PRG,

- geokodowanie pozostałych obiektów z wykorzystaniem serwisów OSM i Google Maps.

Efektem finalnym procesu geokodowania jest przygotowanie dwóch plików:

- BDG_GEO (w formacie Excel) zawierającego dane źródłowe poszerzone o kolumny DLUG (długość geodezyjna), SZER (szerokość geodezyjna), GEOKOD (metadane)
- wynik_geo_point (w formacie SHP) – warstwa geoinformacyjna systemu GIS zawierająca zestaw danych źródłowych.

Zgeokodowane dane zawierają metadane charakteryzujące dokładność geometryczną lokalizacji przestrzennej:

- 1 – geokodowanie z wykorzystaniem Państwowego Rejestru Granic (dokładność rzędu kilku m)
- 2 – geokodowanie z wykorzystaniem API Google Maps (dokładność rzędu kilkunastu metrów)
- 3 – geokodowanie danych o niepełnej lub nieprawidłowej informacji dotyczącej adresu (np. błędny lub nieistniejący numer budynku) – realizowane z wykorzystaniem PRG, OSM, Google Maps (dokładność rzędu kilkudziesięciu metrów)
- 4 – geokodowanie danych o niepełnej lub nieprawidłowej informacji dotyczącej adresu (np. ograniczonej do nazwy miejscowości) – realizowane z wykorzystaniem PRG, OSM, Google Maps (dokładność rzędu kilkuset metrów).

Pod względem ilościowym dominują dane o wysokiej lub bardzo wysokiej precyzji lokalizacji przestrzennej, stanowiące niemal 90% wszystkich zgeokodowanych obiektów MŚP.

dokładność geokodowania	liczebność	%
1	157 231	82
2	13 724	7
3	17 754	9
4	4 383	2