

Ocena rozdzielczości pojęciowej zbioru danych

Determining the conceptual resolution of a data set

Bartłomiej Bielawski

Politechnika Warszawska, Wydział Geodezji i Kartografii

Słowa kluczowe: schemat aplikacyjny, UML, IFPUG, cecha atomowa
Keywords: application schema, UML, IFPUG, atomic property

Wprowadzenie

Rozdzielczość danych przestrzennych dostarcza podstawowej wiedzy o gęstości danych przestrzennych w zbiorze danych (Baranowski i in., 2008). Można wyróżnić kilka aspektów rozdzielczości, na przykład: geometryczną, relacji przestrzennych lub pojęciową. Rozdzielczość zbioru danych może być określana w sposób względny jako stosunek rozdzielczości badanego modelu danych do modelu referencyjnego (np. BDOT10k – Baza Danych Obiektów Topograficznych o szczegółowości odpowiadającej mapie topograficznej 1:10 000 oraz MGCP – *Multinational Geospatial Co-Production Program* – baza danych przestrzennych w specyfikacji *National Geospatial-Intelligence Agency*) lub bezwzględny – charakteryzujący tylko opisywany zbiór danych bez odniesienia się do zbiorów danych o znanej zawartości informacyjnej.

Aktualnie stosowane wskaźniki dotyczące oceny zawartości informacyjnej zbiorów danych przestrzennych są zdaniem autora niewystarczające. Co prawda istnieją systemy oceny złożoności schematów aplikacyjnych lub wielkości systemów informatycznych, lecz żadne z nich nie pozwalają na charakterystykę modelu danych uwzględniających zrozumienie informacji zawartych w modelu. Uwzględnienie aspektów aksjomatycznej teorii poznania pozwala uzupełnić aspekt rozumienia informacji na najwyższym poziomie uogólnienia – identyfikację najmniejszych porcji informacji w schemacie aplikacyjnym.

Analiza literatury przedmiotu w zakresie oceny rozmiaru systemów informatycznych, oceny złożoności schematów aplikacyjnych oraz teorii poznania, pozwala na identyfikację elementów istotnych dla przedmiotu badań. Zidentyfikowane elementy pozwalają na skonstruowanie nowego terminu rozdzielczości pojęciowej oraz zaproponowanie stosowania go jako komponentu charakteryzującego gęstość informacji modelu danych.

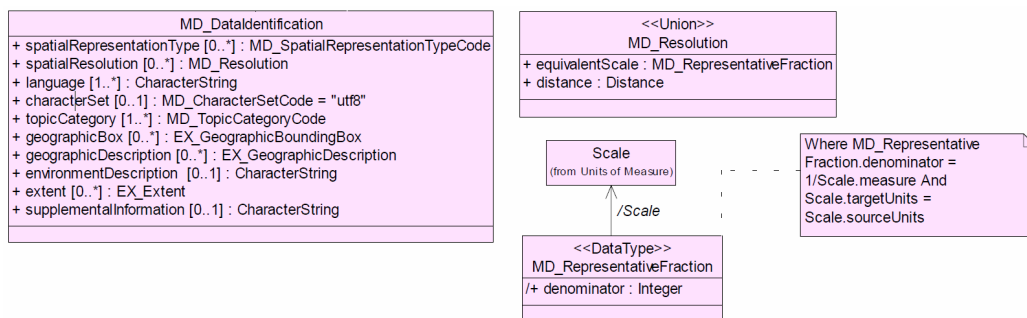
W artykule zostanie omówiona autorska koncepcja dotycząca określania bezwzględnej rozdzielczości pojęciowej zbiorów danych.

Rozdzielczość w geodezji i kartografii

Pojęcie rozdzielczości znane jest w geodezji i kartografii w odniesieniu do danych rastrowych, rozumianych jako macierz elementów w rekordach, w których przechowywane są istotne dane dotyczące zdefiniowanych atrybutów. Wartości atrybutów mogą być wyrażone w systemie binarnym, stało- lub zmiennoprzecinkowym (Felncenloben, 2011). Cechą charakterystyczną danych rastrowych jest rozdzielczość, rozumiana jako liczba pikseli w poziomie i pionie. Obrazy rastrowe mogą służyć do bezpośredniego zobrazowania rzeczywistości przez wykonanie zdjęć lotniczych lub satelitarnych, gdzie każdemu pikselowi przyporządkowany jest odpowiedni fragment terenu. Cechą charakterystyczną takich zobrazowań jest wycinek terenu przestawiany jednym pikselem rastra. Przykładowo satelita LANDSAT TM dostarcza sceny satelitarne o rozmiarze piksela odpowiadającego 30×30 metrów w terenie, a satelita SPOT wykonuje zobrazowania o wielkości piksela 20×20 metrów (Izdebski, 2008). Ten rodzaj rozdzielczości nazwany bywa rozdzielczością terenową (Będkowski, 2008).

Oprócz rozdzielczości przestrzennej (rozdzielczości terenowej) cechą sensorów fotogrametrycznych i teledetekcyjnych są: **rozdzielczość radiometryczna** – wiążąca się z dyskretyzacją wielkości wykrywanego promieniowania i wyrażaną w bitach, **rozdzielczość spektralna** – która jest zdolnością systemu sensorowego do rozróżniania promieniowania elektromagnetycznego różnych częstotliwości, **rozdzielczość czasowa** – która określa odstęp czasu, podawany w dniach, między dwoma kolejnymi pokryciami danego fragmentu powierzchni ziemskiej przez satelitarny system sensorowy (Gaździcki, 2016). Wszystkie rodzaje omawianych rozdzielczości są dobrze opisane w literaturze i szeroko stosowane w praktyce produkcyjnej.

Termin „rozdzielczość” istnieje również w kontekście danych przestrzennych. Norma PN-EN-ISO 19115 (Informacja geograficzna – Metadane) przewiduje możliwość zapisu rozdzielczości danych w elemencie *MD_DataIdentification*.



Rysunek 1. Klasy definiujące rozdzielczość zbioru danych
(źródło: PN-EN-ISO 19115, wyciąg z diagramu A.2 – Identification information)

Jak pokazano na diagramie (rys. 1), atrybut *spatialResolution* w klasie *MD_DataIdentification* jest definiowany jako wskaźnik dostarczający ogólnej wiedzy na temat gęstości danych przestrzennych w zbiorze danych¹ (ISO 19115). Atrybut ten jest typu *MD_Resolution*,

¹ factor which provides a general understanding of the density of spatial data in the dataset

zdefiniowany jako poziom szczegółowości wyrażony w postaci wskaźnika skali albo odległości terenowej². Typ ten jest typem złożonym i definiowany jest przez dwa atrybuty: *equivalentScale* o typie *MD_RepresentativeFraction* oraz *distance* typu *Distance*. Należy zwrócić uwagę, że klasa *MD_Resolution* ma stereotyp *Union*, co przekłada się na alternatywne zastosowanie jednego z dwóch atrybutów na etapie implementacji fizycznej schematu pojęciowego (Chojka, Parzyński, 2013).

Atrybut *equivalentScale* w klasie *MD_Resolution* jest typu *MD_RepresentativeFraction* – co jawnie przekłada się na mianownik skali wyrażony za pomocą liczby całkowitej: atrybut *denominator* w klasie *MD_RepresentativeFraction*.

W praktyce oznacza to, że domeną elementu *MD_Resolution* są wartości rozdzielczości zbioru danych o reprezentacji wektorowej za pomocą **mianownika skali** albo wykorzystując rozdzielczość terenową obrazu rastrowego wyrażonego w **jednostkach odległości** (wielkości piksela).

Mianownik skali nie jest cechą charakterystyczną zbioru danych. Skala mapy to stosunek odległości na mapie do odpowiadającej jej odległości rzeczywistej zredukowanej na powierzchnię odniesienia. Stosowanie tej wartości jest uwarunkowane historycznie, ponieważ wiele zbiorów danych tworzonych w latach 80. i 90. XX wieku pochodziło z wektoryzacji map papierowych. W ocenie autora, wartość ta nie powinna być stosowana do oceny rozdzielczości zbiorów danych, ponieważ parametr ten jest cechą mapy w postaci analogowej, a nie cechą zbioru danych.

Elementy aksjomatycznej teorii poznania

Zgodnie ze słownikiem etymologicznym angielskie słowo *resolution* (pol. rozdzielczość) wywodzi się z języka starofrancuskiego (st. fr. *resolution*) lub łaciny (łac. *resolutionem*) i oznacza proces redukcji obiektów do prostszych form (ang. *process of reducing things into simpler forms*) (Harper, 2016).

Aksjomatyczna teoria poznania (Kossecki, 2005) opiera się na pojęciach pierwotnych, których nie definiujemy, ale ich znaczenie przyjmujemy za oczywiste. W duchu tej teorii Kossecki wyróżnia trzy niedefiniowalne pojęcia pierwotne: **Obiekt elementarny** – którego nie dzieli się na mniejsze części; **Relacje pierwotne** – powiązania między obiektami elementarnymi (przyp. autora) oraz **Zbiór** – zawierający obiekty lub relacje. Obiekt elementarny oraz relacje elementarne są jednoznaczne z ostatnim etapem podziału na części. Pojęcia te – z punktu widzenia oceny rozdzielczości pojęciowej – są **cechami atomowymi**.

W przypadku danych rastrowych, takim najmniejszym elementem (obiektem elementarnym w teorii Kosseckiego) jest jeden piksel, który jest najprostszym elementem budującym zobrazowanie rastrowe. W fizyce cząstek elementarnych jako obiekty elementarne traktuje się właśnie te cząstki, w demografii jako obiekty elementarne traktujemy ludzi (Kossecki, 2005).

Zadanie jakie stoi przed autorem to identyfikacja najmniejszej, atomowej części schematu pojęciowego. Po zidentyfikowaniu najmniejszych porcji informacji będzie możliwe wyznaczenie wskaźnika opisującego w sposób obiektywny każdy zbiór danych.

² level of detail expressed as a scale factor or a ground distance

Ocena złożoności schematów aplikacyjnych UML

Istnieje wiele sposobów wyliczania złożoności i wielkości diagramów klas UML.

Interesujące zestawienie metryk złożoności schematu aplikacyjnego, w postaci diagramu klas UML, przedstawiła Chojka (2014) w artykule pt. „Złożoność schematów aplikacyjnych UML i GML”. Autorka wyróżniła dwie grupy metryk: metryki wielkości oraz metryki złożoności (tab. 1). Z wykorzystaniem zidentyfikowanych metryk dokonała oceny złożoności wybranych krajowych schematów aplikacyjnych UML w oparciu o wybrane metryki. Wyniki oceny złożoności dla wybranych schematów aplikacyjnych zestawiała w tabeli 2.

Tabela 1. Metryki wielkości i złożoności (opracowanie własne na podstawie: Chojka, 2014; Genero, Piattini, Calero, 2005; Vargas, Nugroho, Chaudron, Visser, 2012)

Metryki wielkości	Metryki złożoności
NC – liczba klas NA – liczba atrybutów NM – liczba metod	NAssoc – całkowita liczba asocjacji NAgg – całkowita liczba agregacji NDep – całkowita liczba zależności NGen – całkowita liczba związków generalizacji NGenH – całkowita liczba hierarchii dziedziczenia między klasami MaxDIT – ścieżka od danej klasy do "korzenia" w ramach hierarchii dziedziczenia MaxHAgg – ścieżka od danej klasy do liścia w ramach hierarchii agregacji AscNoRole – liczba powiązań bez nazwanych ról LoneClass – liczba klas, które nie są w żaden sposób powiązane z innymi klasami

Tabela 2. Złożoność wybranych schematów aplikacyjnych IIP (źródło: Chojka, 2014)

	NC	NA	NAssoc	NGen	LoneClass
EGIB	71	699	78	30	38
RCIWN	20	158	10	6	13
PRG	10	71	4	5	3
EMUIA	15	83	9	1	10
BDOT	60	244	3	27	32
GESUT	36	182	4	17	17
MZ	1	0	7	0	0
SytWys	12	45	17	0	6
Osnowa	29	180	9	11	10

Na podstawie analizy zestawienia (tab. 2), można uzyskać informacje o stopniu złożoności/komplikacji diagramu klas. Wartości uzyskiwane są w sposób automatyczny poprzez zliczenie określonych cech (metryk złożoności) diagramu klas. Metoda ta pozwala na relatywnie szybkie i precyzyjne obliczenie stopnia skomplikowania diagramu klas.

Przetestowane metryki złożoności nie oddają w pełni charakteru schematów aplikacyjnych (Chojka, 2014). Tego typu ocena nie uwzględnia zawartości informacyjnej modelowanej dziedziny, a jedynie wskazuje na złożoność diagramu klas, bez próby analizy zawartości informacyjnej. W szczególnych przypadkach stopień złożoności diagramu może być zbieżny z zawartością informacyjną, lecz jak zauważa autorka nadmierna złożoność diagramu może być efektem nieefektywnego zaprojektowania schematu aplikacyjnego (Chojka, 2014).

Ocena złożoności systemu informatycznego według ISO

Budowa każdego systemu informatycznego wymaga oceny nakładu pracy (kosztów), jaki wykonawca będzie musiał ponieść. W przypadku konstruowania dużych systemów informatycznych prawidłowa ocena wymiaru prac jest kluczowa. Pionierem w zakresie badania złożoności systemów informatycznych był Albrecht, który w późnych latach 70. XX wieku, zbudował podwaliny pod metodykę szacowania i pomiaru rozmiarów systemów informatycznych (Dale, 2009). Aktualnie istnieje duża liczba metod pomiaru oprogramowania bazującego na założeniach Albrechta. Wiele z nich przyjęte zostało jako normy ISO. Przykładem takich standardów są metody: COSMIC-FPP przyjęta jako norma ISO/IEC 19761:2001 (ISO 19761) czy IFPUG – aktualnie norma ISO/IEC 20926:2009 (ISO 20926).

Wspomniane wyżej metody pozwalają na estymację i mierzenie tzw. rozmiaru funkcjonalnego systemów informatycznych. Obie metody opierają się na empirycznych modelach parametrycznych, które buduje się na podstawie danych pochodzących z wielu przedsięwzięć wcześniej zrealizowanych (Czarnocka-Chrobot, 2004). Podstawą obu metod jest wyznaczanie tzw. umownych punktów funkcyjnych. Abstrahując od szczegółów obie metody wykorzystywane są do pomiaru funkcjonalności, a nie złożoności architektury systemu (Czarnocka-Chrobot, 2004) – w tym oceny złożoności schematu pojęciowego.

W wyniku przeprowadzonej analizy metod pomiaru oprogramowania, zidentyfikowano aspekt mający zastosowanie do pomiaru rozmiaru modelu pojęciowego. W ramach metody IFPUG (ISO 20926) wyróżniono dwie grupy mierzonych wskaźników. Jedną z nich dotyczy **zasobów danych** (ang. *data storage*) drugą **transakcji** (ang. *transactions*), które skupiają się na wskaźnikach związanych z przepływem procesów. W ramach grupy pomiaru zasobów danych wyróżniono dwa wskaźniki:

- wewnętrzny wskaźnik logiczny ILF (*Internal Logical File*³),
- zewnętrzny wskaźnik logiczny EIF (*External Interface File*).

Najważniejszą intencją wyróżnienia wskaźnika ILF była identyfikacja danych wykorzystywanych przez jeden lub więcej elementarnych procesów wewnątrz aplikacji (IFPUG, 2000), czyli identyfikacja najmniejszej porcji informacji, która może podlegać zliczaniu. Zewnętrzny wskaźnik logiczny (EIF) pozwala na identyfikację dokładnie tych samych elementów co ILF, lecz zarządzanych na zewnątrz systemu (IFPUG 2000, ISO 20926). Można uznać, że ta cecha odpowiada zawartości informacyjnej odpowiadającej wiązaniom pomiędzy różnymi schematami pojęciowymi, przykładowo. BDOT10k i EGIB.

Zgodnie z metodologią IFPUG wyznaczenie wskaźników ILF i EIF odbywa się z wykorzystaniem dwóch rodzajów reguł dotyczących atrybutów DET (ang. *Data Element Type*) oraz grup atrybutów RETs (ang. *Record Element Type*) (IFPUG 2000; ISO 20926). Prawidłowa identyfikacja tych dwóch elementów w schematach aplikacyjnych jest kluczowa do właściwej oceny rozdzielczości pojęciowej zbioru danych. **DET** jest unikalnym, niepowtarzalnym elementem typu danych identyfikowalnym przez użytkownika⁴. **RET** to wskaźnik mówiący o zidentyfikowanej grupie elementów danych w ramach ILF lub EIF. W ramach

³ Metoda IFPUG stosuje pojęcie „pliku” (ang. *file*) przy czym zastrzega, że nie chodzi o plik w tradycyjnym ujęciu informatyki, tylko o powiązaną logicznie grupę danych (IFPUG, 2000) w sensie atrybutu lub klasy obiektów. Nie chodzi tutaj o fizyczną implementację modelu na poziomie PSM (Platform Specific Model w podejściu Model Driven Architecture), lecz o wydzielenia charakterystyczne dla modelu pojęciowego na poziomie PIM (Platform Independent Model w MDA).

⁴ A *data element type* is a unique user recognizable, non-repeated field (IFPUG, 2000).

RET wyróżnia się dwa typy wskaźników: obligatoryjne i fakultatywne. Szczegółowe zasady identyfikacji zostały przedstawione w podręczniku „Function Point Counting Practices” (IFPUG, 2000).

Koncepcja oceny rozdzielczości pojęciowej

Podstawą oceny rozdzielczości pojęciowej zbiorów danych jest powiązanie wiedzy z trzech zakresów:

- 1) aksjomatycznej teorii poznania, jako precyzyjnego narzędzia w wyróżnieniu obiektów pierwotnych i relacji pierwotnych schematu aplikacyjnego,
- 2) modelowania pojęciowego, bazującego na normach z zakresu informacji geograficznej ISO 19100.
- 3) wymiarowania oprogramowania metodą IFPUG.

Ocenie podlega wyłącznie schemat aplikacyjny, utworzony na podstawie norm i dokumentów normatywnych z serii ISO 19100. Praktyczne wykorzystanie zasad modelowania danych opisanych w tych dokumentach gwarantuje wiarygodność oceny zbioru danych oraz rzeczywistą możliwość porównania rozdzielczości pojęciowej dwóch zbiorów danych. Podejście takie jest krytyczne z punktu widzenia powtarzalności pomiarów dla różnych schematów pojęciowych.

Jednym z najistotniejszych elementów koncepcji jest zidentyfikowanie **cechy atomowej** – czyli najmniejszej niepodzielnej części informacji zawartej w schemacie aplikacyjnym. Taką cechą może być klasa obiektów, atrybut, typ atrybutu, wartość atrybutu, czyli **obiekt elementarny** z aksjomatycznej teorii poznania lub rola asocjacyjna czyli **relacja pierwotna** w tej teorii (Kossecki, 2005).

Autor proponuje uznanie za cechy atomowe następujących elementów schematu pojęciowego, zapisanego w postaci diagramu klas UML⁵. Cechy atomowe są wyliczane na poziomie fizycznej klasy FC_f obiektów o stereotypie *FeatureType*, uwzględniając:

- wszystkie abstrakcyjne klasy obiektów o stereotypie *FeatureType* FC_a , z których dziedziczy klasa fizyczna,
- atrybuty o typie prostym (*simple type, basic data types*) fizycznej klasy obiektów A , również odziedziczone po nadklasach,
- atrybut o typie złożonym podlega dekompozycji do typów prostych, z wyłączeniem typów: geometrycznych, daty/godziny, identyfikatora; typy proste, powstałe w wyniku dekompozycji typów złożonych są uwzględniane dokładnie tak jak typy proste,
- typy złożone: geometria, data/godzina, identyfikator są z punktu modelowania danych przestrzennych cechami atomowymi i nie podlegają dekompozycji i uwzględniane są tak jak typy proste,
- atrybut o typie innym niż typ wyliczeniowy (*enumeration*); atrybut o typie wyliczeniowym przyjmuje wartość równą wszystkim wystąpieniom wartości typu wyliczeniowego,
- wartości typu wyliczeniowego (*enumeration values*) E_v , dla każdej klasy obiektów korzystającej z typu wyliczeniowego; o ile jest to możliwe nie uwzględnia się wartości dopełniających np. „inny”, „pozostałe” lub nieokreślonych „niezidentyfikowany”.
- role asocjacyjne (*role name*) R_n .

⁵ Nazwy elementów w języku angielskim pochodzą ze specyfikacji technicznej ISO/TS 19103 (ISO 2005).

W prezentowanym ujęciu przy określaniu cech atomowych nie są uwzględniane:

- klasy obiektów ze stereotypami innymi niż *FeatureType* lub *Enumeration*,
- atrybuty klas abstrakcyjnych; atrybuty tych klas uwzględnia się na poziomie klasy fizycznej, jako odziedziczone atrybuty po wszystkich klasach nadrzędnych,
- atrybuty oznaczone stereotypem *voidable* są traktowane jak atrybuty typów prostych lub złożonych; oznaczenie atrybutu tym stereotypem nie wpływa na wyróżnienie cechy atomowej,
- metadane na wszystkich poziomach szczegółowości,
- atrybuty redundantne lub możliwe do wyliczenia na podstawie innych atrybutów lub cech geometrycznych.

Rozdzielczość bezwzględna R jest sumą cech atomowych (CA) fizycznych klas obiektów składających się na schemat pojęciowy:

$$R = CA(FC_p) + CA(FC_a) + CA(A) + CA(E_v) + (R_n)$$

W celu poprawienia czytelności wskaźnika i podkreślenia dużej dynamiki wzrostu rozdzielczości przy małych liczbach cech atomowych, proponuje się stosować logarytm z rozdzielczości:

$$\text{Rozdzielczość Pojęciowa} = \text{Log}_2 R$$

Propozycja zastosowania skali logarytmicznej wynika przede wszystkim z charakteru wskaźnika. Małe zmiany w zawartości informacyjnej przy małych rozdzielczościach mają duży wpływ na zawartość informacyjną. Małe zmiany zawartości informacyjnej przy średnich rozdzielczościach mają mały wpływ na ogólną ocenę zawartości informacyjnej. Zastosowanie skali logarytmicznej pozwala na uwypuklenie małych zmian przy małych rozdzielczościach, natomiast przy dużych rozdzielczościach – te same zmiany mają charakter marginalny.

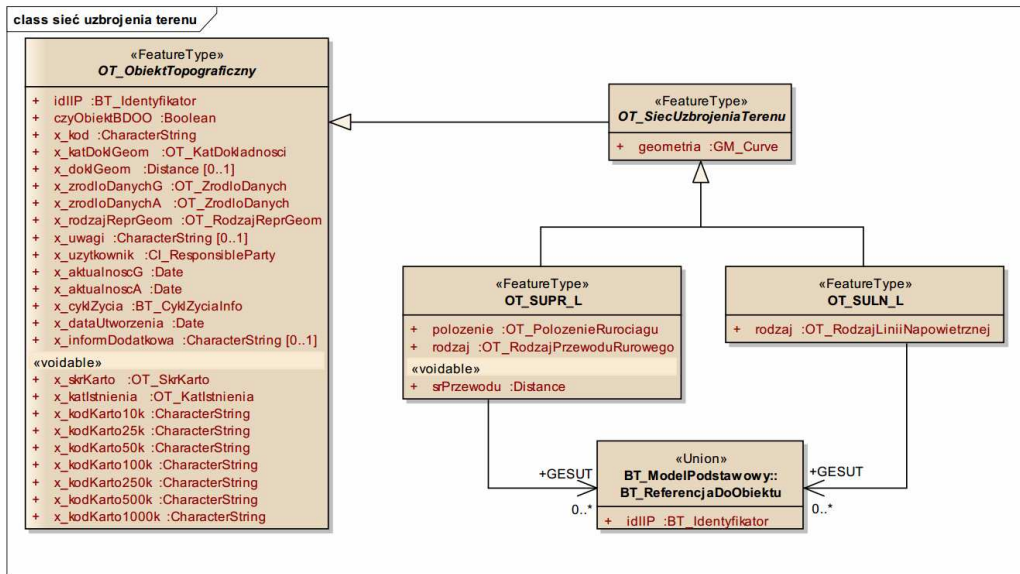
Przykład (tab. 3) dotyczy wyliczania rozdzielczości pojęciowej dla fragmentu schematu aplikacyjnego BDOT10k dla kategorii klas obiektów sieci uzbrojenia (MSWiA, 2011). Na rysunku 2 przedstawiono fragment schematu aplikacyjnego BDOT10k. Zgodnie z notacją UML, na diagramie klas, klasy fizyczne mają nazwy pisane pismem prostym. Pismem pochylonym oznaczane są klasy abstrakcyjne. Klasy fizyczne, dla których wyliczane będą cechy atomowe to *OT_SUPR_L* (przewód rurowy) oraz *OT_SULN_L* (linia napowietrzna).

Rozdzielczość pojęciowa dla kategorii klas obiektów sieci uzbrojenia terenu wynosi:

$$\text{Log}_2 30 = 4,90$$

Analizie podlega cały model, a nie wybrane diagramy, które mogą nie przedstawiać wszystkich zależności. W omawianym przykładzie analizie podlega komplet informacji przedstawiony na schemacie pojęciowym. Ze względów redakcyjnych nie wszystkie informacje pokazywane są na jednym diagramie.

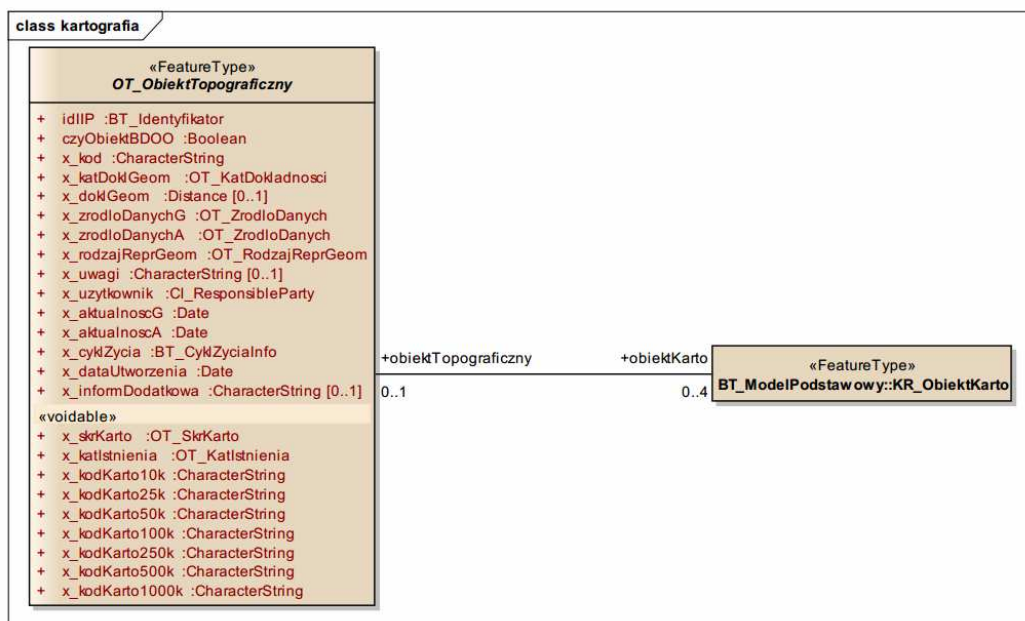
Analiza modelu BDOT10k wykazała, że istnieje powiązanie klasy *OT_obiektTopograficzny* z klasą *KR_ObiektKarto* ze schematu *BT_ModelPodstawowy* (rys. 3). Jednak nie uwzględniono tej relacji, ponieważ reprezentacja kartograficzna obiektu jest rodzajem metadanych i nie jest uwzględniana przy określaniu rozdzielczości pojęciowej.



Rysunek 2. Diagram klas dla kategorii sieć uzbrojenia terenu (źródło: MSWiA 2011)

Tabela 3. Przykład identyfikacji cech atomowych na przykładzie diagramu klas kategorii sieć uzbrojenia terenu (źródło: opracowanie własne)

	T_SUPR_L [Ca]	OT_SULN_L [Ca]	Uwagi
Klasa obiektów	1	1	Fakt wydzielenia klasy fizycznej o stereotypie <i>FeatureType</i> (\mathbf{FC}_p) 1 CA
Atrybuty o typie innym niż wyliczeniowy	1	0	Typ <i>distance</i> wyraża się liczbą naturalną lub całkowitą (\mathbf{A})
Dekompozycja typu wyliczeniowego <i>OT_PolozenieRurociagu</i>	4	nie dotyczy	4 wartości o typie prostym (\mathbf{E}_v): nadziemnyNaPodporach, nadziemnyZawieszony, naziemny, podziemny
Dekompozycja typu wyliczeniowego <i>OT_RodzajPrzewoduRurowego</i>	6	nie dotyczy	6 wartości o typie prostym (\mathbf{E}_v): benzynowy, ciepłowniczy, gazowy, kanalizacyjny, naftowy, wodociągowy
Dekompozycja typu wyliczeniowego <i>OT_RodzajLiniiNapowietrznej</i>	nie dotyczy	5	5 wartości o typie prostym (\mathbf{E}_v): najwyzszeNapiecie, wysokieNapiecie, srednieNapiecie, niskeNapiecie, telekomunikacyjna
Liczba odziedziczonych cech atomowych po klasie <i>OT_SiecUzbrojeniaTerenu</i>	2	2	Fakty wydzielenia klasy nadrzędnej 1 CA (\mathbf{FC}_p), atrybuty geometria 1 CA (\mathbf{A})
Liczba odziedziczonych cech atomowych po klasie <i>OT_ObiektTopograficzny</i>	3	3	Nie uwzględnia się metadanych - atrybutów z przedrostkiem <i>_x</i> . Fakt wydzielenia klasy nadrzędnej 1 CA (\mathbf{FC}_p), 2 atrybuty (\mathbf{A})
Liczba ról asocjacyjnych własnych	1	1	Powiązanie klasy o stereotypie <i>Union</i> (\mathbf{R}_n)
SUMA	18+12=30		



Rysunek 3. Diagram – powiązanie obiektu topograficznego z klasą KR_ObjektKarto
(źródło: MSWiA, 2011)

Zastosowania

Opracowana metoda może posłużyć przede wszystkim do ogólnej oceny zawartości informacyjnej modelu danych, tj. do oceny ilości informacji jaką może przenosić zbiór danych.

Metoda może znaleźć zastosowanie w metadanych i stanowić jeden z elementów mówiących o rozdzielczości danych przestrzennych. Przy aktualnej koncepcji konstrukcji metadanych już teraz istnieje możliwość uzupełnienia metadanych zgodnych z ISO elementem opisującym wartość rozdzielczości pojęciowej.

Kolejnym potencjalnym zastosowaniem oceny rozdzielczości jest wykorzystanie wiedzy o rozdzielczości przez usługi działające w ramach infrastruktury informacji przestrzennej. Może to być istotny element w podjęciu decyzji o możliwości integracji kilku usług oferujących różną zawartość informacyjną danych.

Wnioski

Opracowana koncepcja stanowi propozycję oceny rozdzielczości pojęciowej zbioru danych – jednego z elementów charakteryzujących zbiór danych. Przedstawiony sposób wyliczania rozdzielczości dostarcza wiedzy na temat pojemności informacyjnej modeli danych, a także umożliwia porównywanie ładunku informacyjnego określonego w różnych zbiorach danych przestrzennych. Proponowana metoda pozwala na ocenę modelu danych pod względem liczby najmniejszych jednostek informacji (cech atomowych) w schemacie aplikacyjnym. Obliczając wartości rozdzielczości pojęciowej dla dwóch zbiorów, a następnie porów-

nując uzyskane wartości, otrzymuje się ogólną informację na temat zawartości informacyjnej modelu danych. Należy zauważyć, że aspekt rozdzielczości pojęciowej jest jednym z kilku aspektów rozdzielczości zbioru danych. Ocena rozdzielczości geometrycznej oraz rozdzielczości relacji wraz z oceną rozdzielczości pojęciowej w pełni opisuje rozdzielczość zbioru danych.

Przedstawiony sposób oceny rozdzielczości pojęciowej, wraz z charakterystyką rozdzielczości geometrii i relacji, zapewnia znacznie lepszą charakterystykę poziomu szczegółowości zbioru danych niż mianownik skali stosowany aktualnie do określania gęstości danych w metadanych zgodnie z ISO 19115. Proponowana metoda wyliczania uwzględni wiele istotnych cech informacyjnych schematu aplikacyjnego, takich jak: klasy obiektów, atrybuty, typy (w tym typy wyliczeniowe). Autor nie wyklucza, że w wyniku praktycznych doświadczeń zostanie zmodyfikowany sposób wyliczania rozdzielczości pojęciowej – w szczególności jako cechy atomowe zostaną uwzględnione dodatkowe składowe schematu aplikacyjnego.

Literatura

- Baranowski M., Gotlib D., Soczewski P., 2008: Polski krajowy profil metadanych w zakresie geoinformacji. GUGIK.
http://arch.gugik.gov.pl/__data/assets/pdf_file/0005/26438/1000_krajowyprofilmetadanych_v1_5.pdf
- Będkowski K., 2008: Fotogrametryczny pomiar wysokości drzew na obrazach z kamery cyfrowej DMC. *Roczniki Geomatyki* t. 6, z. 8: 41-48, PTIP, Warszawa.
- Chojka A., Parzyński Z., 2013: Infrastruktura Informacji Przestrzennej w UML. Geodeta, ISBN 978-83-935447-1-4.
- Chojka A., 2014: Złożoność schematów aplikacyjnych UML i GML. *Roczniki Geomatyki* t.12, z. 2(64): 197-204, PTIP, Warszawa.
- Czarnecka-Chrobot B., 2004: Pomiar rozmiaru funkcjonalnego systemu informatycznego – sposób na „CHAOS PERMANENS”? Współczesne kierunki rozwoju informatyki, PTI.
- Dale S., 2009: System Cost Engineering. Program Affordability Management and Cost Control, Gower Publishing Limited, Burlington, USA.
- Drafting Team Metadata, 2013: INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (MD_IR_and_ISO_20131029.doc).
- Felcenloben D., 2011: Geoinformacja – wprowadzenie do systemów organizacji danych i wiedzy. Wydawnictwo Gall, ISBN 978-83-60968-11-6.
- Gaździcki J. (red.): Internetowy leksykon geomatyczny. Dostęp 4.09.2016 r. www.ptip.org.pl
- Grudzień M., Soczewski P., 2008: Wzorcowe metadane dla standardowych danych przestrzennych stworzonych przez służbę geodezyjną i kartograficzną w Polsce – poziom krajowy. GUGiK.
http://arch.gugik.gov.pl/__data/assets/file/0016/26440/1002_codgik.zip
- Genero M., Piattini M., Calero C., 2005: A Survey of Metrics for UML Class Diagrams. *Journal of Object Technology* vol. 4, no. 9: 59-92, ETH Zurich.
- Harper D., 2016: Online Etymology Dictionary. Dostęp 20.07.2016.
<http://www.etymonline.com/index.php?term=resolution>
- IFPUG (International Function Point Users Group), 2000: Function Point Counting Practices Manual. Release 4.1.1, Princeton Junction, IBSN 0-963-1742-7-4.
- ISO, 2011: ISO/IEC 19761:2011 Software engineering – COSMIC: a functional size measurement method.
- ISO, 2009: ISO/IEC 20926:2009 Software and systems engineering – Software measurement – IFPUG functional size measurement method 2009.
- ISO, 2005: ISO/TS 19103 Geographic information – Conceptual schema language.
- ISO, 2015: ISO 19115 Geographic information – Metadata Part 1: Fundamentals.
- Izdebski W., 2008: Wykłady z przedmiotu SIT.
- Kossecki J., 2005: Metacybernetyka. Monografia, wyd. bd. Kielce-Warszawa.

- Orych A., Walczykowski P., 2010: Wyznaczanie terenowej zdolności rozdzielczej sensorów cyfrowych w oparciu o cele kalibracyjne. *Archiwum Fotogrametrii, Kartografii i Teledetekcji* vol. 21, ISBN 978-83-61576-13-6.
- Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 17 listopada 2011 r. w sprawie bazy danych obiektów topograficznych oraz bazy danych obiektów ogólnogeograficznych, a także standardowych opracowań kartograficznych. Dz.U. 2011 nr 279 poz. 1642.
- Vargas R.T., Nugroho A., Chaudron M., Visser J., 2012: The Use of UML Class Diagrams and Its Effect on Code Change-proneness. Proceedings of the Experiences and Empirical Studies in Software Modelling Workshop 2012, Innsbruck, Austria.

Streszczenie

Aktualny sposób określania rozdzielczości zbiorów danych w postaci mianownika skali mapy wydaje się być niewystarczający. Zdaniem autora można zaproponować sposób opisu rozdzielczości zbiorów danych, bazujący na standardach ISO z zakresu informacji geograficznej (seria ISO 19100) oraz ocenie złożoności systemów informatycznych (ISO 19761; ISO 20926), a także aksjomatycznej teorii poznania. Proponowana metoda obejmuje tylko jeden aspekt rozdzielczości zbiorów danych – rozdzielczość pojęciową. Dotyczy ona zbiorów danych zdefiniowanych na poziomie Platform Independent Model (PIM) w podejściu Model Driven Approach (MDA).

Najważniejszym elementem koncepcji oceny rozdzielczości pojęciowej jest identyfikacja najmniejszych porcji informacji (cech atomowych) zakodowanych w schemacie pojęciowym wyrażonym w postaci diagramów klas UML. Autor proponuje zestaw reguł służących do identyfikacji cech atomowych i na ich podstawie obliczenie rozdzielczości pojęciowej schematu aplikacyjnego.

Proponowany sposób wyliczania rozdzielczości znajduje zastosowanie wyłącznie dla specyfikacji zbiorów danych utworzonych na podstawie norm i dokumentów normatywnych ISO serii 19100 dotyczących informacji geograficznej.

Abstract

Actually the way of determining a data set resolution by an equivalent scale seems to be insufficient. In the author's opinion, there is a way to describe a data set resolution based upon the concepts of the ISO 19100 series of Geographic Information Quality Standards, the ISO 19761 COSMIC Measurement Standard and the software and system engineering (ISO 20926), as well as the base concepts of the axiomatic theory of knowledge. The proposal allows to calculate only one aspect of data density – the conceptual resolution. In practice it concerns datasets defined on the PIM level (Platform Independent Model) according to the Model Driven Approach – MDA).

The most important element of the proposed method is to identify the smallest information portion (atomic property) coded in the conceptual schema expressed by the UML class diagram. The key of identifying is to understand information given by the application schema. In this paper the author proposes a set of rules to identify the atomic property in the application schema. Based upon identified atomic properties the author proposes the way to calculate the conceptual resolution of the application schema.

mgr inż. Bartłomiej Bielawski
bartlomiej.bielawski@hexagon.si.com

