

Artur KRAWCZYK

Akademia Górniczo-Hutnicza, Katedra Ochrony Terenów Górniczych, Geoinformatyki i Geodezji Górniczej

PRÓBA SYSTEMATYKI ZAPISU ATRYBUTÓW I TOPOLOGII OBIEKTÓW GEOMETRYCZNYCH W SYSTEMACH INFORMACJI GEOGRAFICZNEJ

Streszczenie. Niniejszy artykuł opisuje zagadnienie zapisu danych przestrzennych w systemach informacji geograficznej. Zidentyfikowano formaty zapisu danych przestrzennych. Na podstawie analizy struktur formatów danych oraz sposobów łączenia danych atrybutowych z geometrycznymi, usystematyzowano stosowane zapisy danych. Systematyka uwzględnia możliwości formatu danych w zakresie trwałego przechowywania danych o topologii obiektów geometrycznych oraz połączenia i wykorzystywania danych atrybutowych.

Słowa kluczowe: GIS, geomatyka, formaty danych, topologia, bazy danych przestrzennych

ATTRIBUTE AND TOPOLOGY OF GEOMETRIC OBJECTS SYSTEMATICS ATTEMPT IN GEOGRAPHIC INFORMATION SYSTEMS

Summary. This article describes the issue of generalization of spatial objects in spatial databases. Identified examples of operations (the sum of surface objects, surface objects total) in which the reduction of redundant nodes strongly accelerates query. Highlighting the issue of generalization of geometric objects. It has been proven NP-difficulty of the operation the largest reduction in the number of nodes at a given accuracy

Keywords: spatial databases, object generalization, spatial queries

1. Wstęp

Baza danych przestrzennych w systemach GIS ma za zadanie odwzorować wybrane cechy realnego świata. Stopień skomplikowania świata jest wielokrotnie większy niż są w stanie odtworzyć go najbardziej nawet zaawansowane systemy informacji. Baza danych w systemie GIS z założenia stanowi zatem model, który powinien zapewnić taką reprezentację danych, aby możliwe było osiągnięcie założonego celu utworzenia systemu.

Obiektami zawartymi w bazie GIS mogą być wszystkie elementy świata rzeczywistego, np. budynki, drogi, rzeki, granice, punkty itd. Dodatkowo każdy z tych obiektów można opisać za pomocą danych niegraficznych, czyli atrybutów opisowych. Przykładem może być budynek – jego atrybutami są chociażby: określony rodzaj konstrukcji, wiek, właściciel, adres; rzeka ma swoją nazwę, przepływ, klasę czystości itd. Każda cecha przestrzenna może zostać odwzorowana za pomocą różnych modeli danych, zapisanych za pomocą różnych formatów.

2. Modele danych przestrzennych w GIS

W budowie Systemów Informacji Geograficznej dużą rolę odgrywają modele danych. Początkowo Systemy GIS opierały się na technologii macierzowej (rastrowej), w późniejszym okresie rozwoju zaczęto wykorzystywać dane wektorowe wraz z relacyjnymi bazami danych.

Rastrowy model danych składa się z rzędów i kolumn. Numery rzędu i kolumny określają współrzędne danej komórki rastra. Modele rastrowe/macierzowe (ang. *grids*) różnią się od obrazów rastrowych (ang. *images*). W przypadku macierzy zakłada się, że każda komórka jest homogeniczna, czyli jednolita pod względem opisującego ją atrybutu, co pozwala na wykonywanie dodawania i odejmowania rastrow (algebra macierzy).

Funkcjonalność modelu rastrowego rozszerza możliwość budowy referencji pliku rastra (każdej jego komórki) do relacyjnej bazy danych, wtedy wartość piksela jest tożsama z identyfikatorem kolumny w tabeli bazy danych. Do analiz można wtedy wybierać dowolną kolumnę z tabeli atrybutów rastra. Tego typu analizy umożliwia oprogramowanie GRASS.

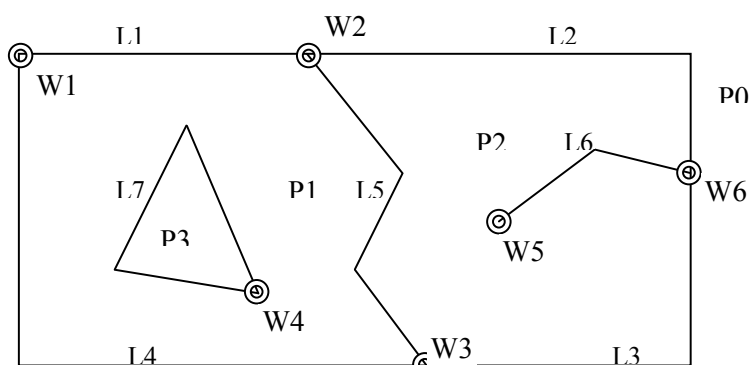
Dane rastrowe typu image są to najczęściej obrazy bardzo wysokiej rozdzielczości, gdzie rozmiar piksela jest znacznie mniejszy niż dokładność pozycjonowania (kalibracji) obrazu w układzie współrzędnych (np. wysokorozdzielcze zdjęcia satelitarne).

Model wektorowy zapewnia uzyskanie dużej dokładności w określeniu położenia każdego obiektu, który może mieć w tym przypadku postać pojedynczych punktów, linii (powstałych z połączenia punktów) lub poligonów (obszarów) tworzonych jako domknięte ciągi li-

nii. Ze względu na to, że linie i poligony są zbudowane z punktów o określonych współrzędnych (X i Y), są one podstawowymi elementami modelu wektorowego. Dane wektorowe można podzielić (w zależności od sposobu zapisu) na proste lub topologiczne modele wektorowe. Prosty model wektorowy (*spaghetti*) stanowi zbiór niepowiązanych ze sobą obiektów punktowych, liniowych i powierzchniowych. Ten model danych stosowany jest w aplikacjach CAD (Computer Aided Design). Główną zaletą tego formatu jest prostota, natomiast istotnymi wadami są: konieczność podwójnego zapisu geometrii na stykach obiektów i kłopotliwa aktualizacja oraz utrudniona analiza związków przestrzennych zachodzących między obiektami. Ważne jest to, że te elementy geometryczne mogą również posiadać referencję do bazy danych i mogą być wykonywane zapytania geoatrybutowe bez zastosowania topologii.

W wektorowym, topologicznym modelu danych, oprócz położenia obszaru definiowany jest jego związek z innymi obiektami tej samej klasy, czyli tego samego wyróżnienia. Topologia dostarcza informacji o tym, które obiekty graniczą ze sobą, które punkty tworzą boki danego poligonu, a które punkty jednocześnie należą do dowolnych dwóch poligonów. Każda linia ma określone węzły, które łączą dodatkowe punkty załamania oraz powierzchnie znajdujące się po obu jej stronach. Taki zapis ułatwia wszelkie operacje przestrzenne oraz upraszcza algorytmy obliczeniowe, wykorzystywane w badaniu związków przestrzennych między obiektami.

3. Topologia wektorowych danych przestrzennych



Rys. 1. Przykładowy, krzywoliniowy model topologiczny
Fig. 1. Example of arc-node topology model

Jednym z najbardziej znanych modeli topologicznych jest model planarny typu „arc-node”. Podstawowymi elementami tego modelu są łuk (ang. *arc*) i węzeł (ang. *node*) [1]. Łuk jest zdefiniowany jako łańcuch zaczynający się i kończący w węźle. Węzeł jest natomiast początkiem i końcem każdego łańcucha. Obiektami mogą być punkty, zbiory łańcuchów

i poligony. Model opisuje topologię obiektów, węzłów i łuków. Aby na trwale zapisać strukturę topologii, potrzebne są trzy tabele przechowujące następujące dane:

- topologia poligonów – lista poligonów łączących linie,
- topologia punktów węzłowych – lista punktów węzłowych,
- topologia linii granicznych – lista linii łączących węzły.

Tabela 1

Topologia poligonów	
poligon	linie graniczne
P0	-L4,-L3,-L2,-L1
P1	L1,L5,L4; -L7
P2	L2,L3,-L5
P3	L7

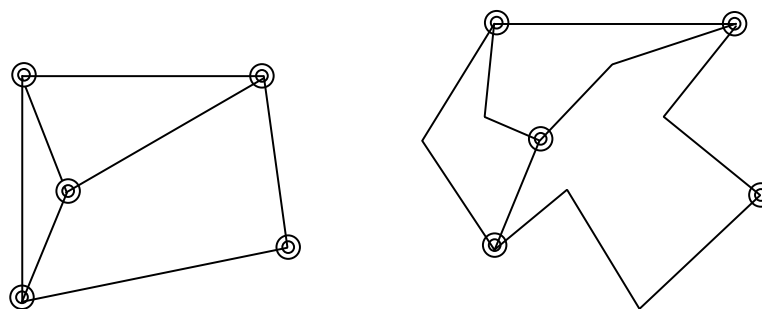
Tabela 2

Topologia punktów węzłowych	
punkty węzłowe	linie graniczne
W1	L4,-L1
W2	L1,-L2,-L5
W3	L3,L5,-L4
W4	L7,-L7
W5	L6
W6	L2,-L6,-L3

Tabela 3

Topologia linii				
linie graniczne	początkowy punkt węzłowy	końcowy punkt węzłowy	wielobok po lewej stronie	wielobok po prawej stronie
L1	W1	W2	P0	P1
L2	W2	W6	P0	P2
L3	W6	W3	P0	P2
L4	W3	W1	P0	P1
L5	W2	W3	P2	P1
L6	W6	W5	P2	P2

W modelu topologicznym rozpatruje się płaszczyznę x,y podzieloną liniami granicznymi $L1, L2, \dots, Lm$ na obszary $P0, P1, P2, \dots, Pn$. Linie graniczne nie mogą się przecinać, mają określone swoje kierunki i łączą węzły $W1, W2, \dots, Wq$. Zależnie od rodzaju linii granicznych, można mówić o prostoliniowym lub krzywoliniowym modelu topologicznym [1].



Rys. 2. a) model prostoliniowy, b) model krzywoliniowy

Fig. 2. a) line node model, b) arc node model

W pierwszym wypadku każda linia graniczna jest wektorem łączącym węzeł początkowy W_j z jego węzłem końcowym W_k . Wektor ma zdefiniowany obszar po lewej stronie Pl i po prawej Pp . W modelu krzywoliniowym linia ma dodatkowe punkty załamania, niebędące punktami węzłowymi. Topologia linii definiuje pięć oznaczeń:

L_i, W_j, W_k, Pl, Pp .

Zmiana kierunku linii L powoduje zmianę kolejności punktów węzłowych i obszarów, taką linię oznacza się jako $-L$. Jeśli według przykładowego rysunku (rys. 1) linia $L2$ biegnie od węzła $W2$ do węzła $W6$, mając po lewej obszar $P0$, a po prawej obszar $P2$, oznaczmy ją:

$L2, W2, W6, P0, P2$.

Linia $-L2$ będzie biegła od $W6$ do $W2$, mając po lewej obszar $P2$, a po prawej $P0$:

$-L2, W6, W2, P2, P0$.

Tabele 1, 2 i 3 umożliwiają zachowanie informacji o topologii, która następnie służy do wykonywania takich operacji, jak: obliczanie odległości lub powierzchni czy wyznaczanie współrzędnych punktów przecięcia linii nakładających się warstw topologicznych.

Model topologiczny wymaga, aby wszystkie linie dokładnie łączyły się w punktach węzłowych, a wieloboki były domknięte. Wadą tego modelu jest dość długi czas obliczeń potrzebnych do zidentyfikowania wszystkich punktów węzłowych.

4. Przegląd formatów zapisu modelu wektorowego danych geograficznych

Rozwój komercyjnych systemów informacji geograficznej datuje się od początków lat 70. W ówczesnych czasach były to systemy mocno scentralizowane, a pojęcie wymiany danych dotyczyło właściwie konwersji całego zasobu GIS. Pod koniec lat 90., w związku z postępem technologicznym, nastąpił dynamiczny rozwój technologii GIS. Pojawiło się wiele nowych firm oferujących oprogramowanie GIS – MapInfo, Bentley Geographics, AutodeskMap, Manifold, i wiele innych rozwiązań open source – GRASS, Jshape i innych. Sytuacja ta spowodowała powstanie dużego zamieszania w opisie formatów danych przestrzennych stosowanych przez różne firmy i organizacje.

4.1. Rozwiązania firmy ESRI

Na przestrzeni ostatnich lat można zaobserwować ciągły rozwój formatów danych dokonywanych w firmie ESRI. Historycznie pierwsza była struktura katalogów, gdzie jeden katalog odpowiadał jednej warstwie (coverage), stosowana od v1 z 1981 roku do wersji 7. Pojawienie się produktu ArcView spowodowało powstanie nowego geoprzestrzennego formatu

danych *shape*, zapisywanego z rozszerzeniem *.shp. W wersji Arc/Info 8 pojawiła się zmodyfikowana struktura katalogów – ArcCatalog. W wersji 9 zmieniono nazwę aplikacji na ArcGIS, co powiązane było z wprowadzeniem nowego formatu danych, potocznie określanym jako geobaza lub model georelacyjny. Warto podkreślić, że obecnie oprogramowanie posiada możliwość pełnej pracy na danych przestrzennych w pięciu zupełnie różnych formatach danych. Oczywistą sprawą jest fakt, że parametry pracy się różnią, ale możliwości oferowane przez program są praktycznie identyczne.

Aplikacja ARC/INFO pierwotnie posiadała dane w postaci punktów, linii i poligonów połączonych z atrybutami. Każda warstwa (coverage) umieszczona została w pojedynczym katalogu. Tabela atrybutów dla danej warstwy zapisana jest w osobnym katalogu o stałej nazwie dla całego projektu INFO. Warstwy, w zależności od typu, mogą zawierać dane o różnych rodzajach obiektów. Łącznie zdefiniowano 23 formaty plików, które pozwalają na zapis warstw punktowych liniowych i powierzchniowych. Poniżej przedstawiono niektóre z formatów [2]:

AAT – Arc Attribute Table (atrybuty tekstowe łuku, zarządzane przez bazę INFO),

ARC – Arc coordinates and topology (odpowiednik tabeli 3 – zapis topologii),

BND – Coverage minimum and maximum coordinates,

CNT – Polygon centroid table,

LAB – Label point coordinates and topology,

PAL – Polygon topology (odpowiednik tabeli 1 – zapis danych topologicznych),

PAT – Polygon Attribute Table (atrybuty tekstowe poligonu zarządzane przez bazę INFO),

PBN – Polygon index file,

PRF – Polygon/Point cross-reference file,

TIC – Tic coordinates and IDs,

Inne to XBX, XBN, TXT, TOL, TBX, TBN, PFF, PBX, MSK, LOG, ABN, ABX, ARF.

Warstwy z tego modelu danych nie mogą być kopiowane do innych katalogów, bowiem przez bazę INFO są na stałe związane z drzewem katalogów danego projektu. Stąd też konieczne było opracowanie formatu eksportu danych – *.e00, który pozwalał na przenoszenie danych pomiędzy komputerami.

W aspekcie topologii danych istotny jest tu plik formatu *.ARC, który zdefiniowany został prawie tak dokładnie, jak model teoretyczny, zawiera on 7 kolumn i dokładnie odpowiada strukturze opisanej w teorii:

- 1. coverage#
- 2. coverage-ID
- 3. from node – od węzła
- 4. to node – do węzła
- 5. left polygon – poligon prawy

- 6. right polygon – poligon lewy
- 7. number of coordinates

Można również wskazać inne pliki, które zgodne są z modelem arc-node. Na podstawie tej analizy możemy przyjąć, że topologia obiektów geometrycznych jest trwale zapisywana w strukturze projektu. Baza danych obsługuje dane atrybutowe za pomocą własnych formatów tabel. Baza INFO obsługuje zapytania do tabel poszczególnych warstw.

ESRI Shapefile

Jednym z formatów najczęściej spotykanych podczas pracy z GIS jest format *shape*. Opracowany został przez firmę ESRI dla aplikacji ArcView. ArcInfo zostało wyposażone w narzędzie do eksportu plików *shapefile*. Ze względu na duży koszt obsługi topologii, opracowano format *shape* w ten sposób, aby nie przechował struktury danych dotyczącej topologii. Dzięki temu uzyskano bardzo dużą przenośność danych. Atrybuty tekstowe są przechowywane w formacie dBASE. Plik główny (*.shp) składa się z rekordów. Każdy rekord zawiera listę współrzędnych opisujących wierzchołki danego obiektu. Plik indeksu (*.shx) zawiera adresy rekordów pliku głównego. Plik tabeli bazy danych (*.dbf) zawiera rekordy atrybutów obiektów. Ten format wprowadzał pojęcie topologii, na żądanie. Aplikacja zapisująca kontroluje dane wektorowe pod względem topologii, ale jej nie zapisuje. Program ArcView odczytując plik, odtwarza topologię w pamięci operacyjnej. Warstwa wtedy nadaje się do wykonywania operacji topologicznych.

GEOBAZA ESRI

Geobaza jest trzecim z kolei formatem danych opracowanym przez ESRI. Fizycznie Geobaza to plik relacyjnej bazy danych, najczęściej Microsoft ACCESS, który przechowuje równocześnie dane atrybutowe i geometryczne w jednej relacyjnej tabeli. Koncepcja Geobazy przewiduje, że topologiczne relacje między obiektami nie są zachowywane. Istnieją natomiast warunki topologiczne, które są wprowadzane za pomocą kreatora i przechowywane w bazie danych jako tabela warunków. Podczas procesu zatwierdzania obiekty w Geobazie są sprawdzane pod względem zgodności z tymi warunkami. Podczas edycji model geometrii może mieć wydzielane obszary tzw. dirt area, w których obiekty biorące udział w tworzeniu geometrii zostały zmienione (dodane, usunięte, zaktualizowane). Jeśli geometria jakiegoś obiektu jest modyfikowana, zasięg obszaru „dirt area” jest powiększony do pokrycia się z całym obszarem zasięgu tej geometrii. Po zakończeniu modyfikacji cały ten obszar zostaje sprawdzony i zatwierdzony. Błędy budowy topologii są generowane w postaci tablicy. Program oferuje narzędzia do wyczyszczenia geometrii i taką, czystą geometrię zapisuje w tabelach w postaci BLOB. Tabele o topologii tworzy w momencie odczytu danych z bazy. Ten typ topologii można określić również jako: topologia odtwarzana przy otwarciu pliku.

4.2. Rozwiązania firm Intergraph i Bentley

Na początku lat 70. firma Intergraph opracowała wielofunkcyjny system informacji o terenie – Modular GIS Environment. W końcu lat 70. firma INTERGRAPH wykupiła firmę Bentley. Intergraph przeniósł około 400 swoich aplikacji na platformę MicroStation. Od 1994 roku firma Bentley stała się samodzielną, co skutkowało powstaniem kolejnego systemu GIS na platformie MicroStation: GeoGraphics 95, bardzo podobnego w działaniu jak MGE.

MGE i GeoGraphics

W programach GeoGraphics i MGE format danych GIS jest modelem wektorowo-relacyjnym. Dane przestrzenne są przechowywane w dualnej formie: w postaci tabel relacyjnej bazy danych oraz danych graficznych w plikach wektorowych. Oba te formaty łączone są w środowisku projektu mechanizmem opierającym się na referencjach (database link), który umożliwił kreowanie połączenia pomiędzy danym elementem grafiki a rekordami relacyjnej bazy danych. Danymi są zestawy tabel (atrybutowa baza danych np. mdb) przyłączonych przez serwer ODBC oraz zestawy map (pliki graficzne *.dgn) dotyczące tego projektu. W tym modelu zastosowano topologię na żądanie. Oznaczało to, że struktura topologiczna tworzona była z warstw wektorowych w momencie wykonywania analizy wektorowej. Dodatkową trudność w użytkowaniu tego modelu danych był problem z utrzymaniem spójności topologicznej warstw geometrycznych. Łatwość manipulacji geometrią w systemach CAD, jakim jest MicroStation, nie sprzyjała utrzymaniu topologii danych geometrycznych.

Bentley MAP

Jest to produkt, który zastąpił aplikację Microstation GeoGraphics. W tej aplikacji całkowicie zrezygnowano z zapisu tekstowych atrybutów obiektów geometrycznych w relacyjnej bazie danych. Zastąpiono ją znacznikami XML, zapisywanymi w pliku wektorowym aplikacji. Z tego powodu tradycyjne zapytania geoatrybutowe zostały zastąpione zapytaniami XPath. Topologia składowana jest w pliku wektorowym dla konkretnej, wybranej feature (warstwy) i posiada osobny zestaw narzędzi do jej zarządzania. Są trzy tabele o nazwach: Area Table, Boundary Table, Node Table, których treść zgodna jest z zawartością planarnego modelu topologii typu *arc-node*.

GEOMEDIA firmy INTERGRAPH

Aplikacja Geomedia zastąpiła system MGE. W tej aplikacji dane graficzne i atrybutowe są umieszczone w tej samej tabeli bazy danych. Jedna tabela przechowuje jedną warstwę, a każdy pojedynczy obiekt tej warstwy jest umieszczony w osobnym rekordzie. W przypadku GeoMedia formatem zapisu jest plik .mdb Accessa, w którym oprócz atrybutów dodana jest

kolumna *geometry*, gdzie w postaci obiektu typu BLOB zakodowana jest informacja o geometrii pojedynczego obiektu. Ten format w zakresie koncepcji jest praktycznie identyczny z Geobazą ESRI, bazy różnią się konstrukcją geometrii w obiektach BLOB oraz najprawdopodobniej sposobami indeksowania.

4.3. Rozwiązania stosowane w bazach danych na przykładzie bazy Oracle

Od wielu lat firma Oracle posiada zaawansowane możliwości wspomagające aplikacje GIS w przetwarzaniu danych przestrzennych. Moduł *Oracle Spatial* rozszerza możliwości bazy danych *Oracle*. Jest to narzędzie do przechowywania danych przestrzennych oraz obsługuje ono wykonywanie analiz, takich jak: tworzenie buforów, przecięć przestrzennych, obliczanie odległości i powierzchni [3]. Rozszerzenie *Spatial* zapisuje dane przestrzenne bezpośrednio do tabel bazy danych, dla których tworzona jest w tym celu kolumna SDO_GEOMETRY, w której zapisywane są dane formatu SDO_Geometry. Typ danych SDO_Geometry nie przechowuje danych o topologii danej geometrii.

Dopiero od wersji Oracle Spatial 10g wprowadzono nowy typ danych SDO_TOPO_GEOMETRY. Pozwala on na trwały zapis topologii warstwy geometrycznej do bazy danych Oracle. W tym wypadku również zdecydowano się na odwzorowanie relacji topologicznych za pomocą modelu planarnego typu „arc-node”. Z tego powodu do zapisu struktury topologii zdefiniowano trzy tabele przeznaczone do obsługi składowych elementów geometrii [6]:

<topology_name>_NODE\$ – dane o węzłach (odpowiednik tabeli 2),

<topology_name>_EDGE\$ – dane o współrzędnych i kierunku linii danej geometrii (odpowiednik tabeli 3),

<topology_name>_FACE\$ – dane o liniach tworzących poligony (odpowiednik tabeli 1),

W celu porównania z tabelą pliku ARC szczegółowo przedstawiono kolumny tabeli <topology_name>_EDGE\$:

EDGE_ID – identyfikator linii,

START_NODE – początkowy węzeł linii,

END_NODE – końcowy węzeł linii,

LEFT_FACE_ID – identyfikator poligonu po lewej stronie,

RIGHT_FACE_ID – identyfikator poligonu po prawej stronie,

pozostałe kolumny to NEXT_RIGHT_EDGE_ID, PREV_RIGHT_EDGE_ID, NEXT_LEFT_EDGE_ID, PREV_LEFT_EDGE_ID,

GEOMETRY to linia typu SDO_GEOMETRY, która ma początek i koniec zgodny z węzłami zadeklarowanymi w tej tabeli.

Jak wynika z porównania, tabele zawierają bardzo podobne typy kolumn.

Główną przyczyną pojawienia się trwałego zapisu topologii dla geometrii okazała się niska wydajność realizacji zapytań topologicznych opartych na topologii tworzonej na żądanie. Topologia w tej wersji bazy danych może już być zapisywana na stałe i wykorzystana do realizacji efektywnych, bardzo szybkich zapytań geoprzestrzennych.

Drugim formatem danych przestrzennych, które mogą być składowane w bazie danych Oracle, jest wykorzystanie obiektów BLOB na podobnych zasadach jak realizowane jest to przez oprogramowania ArcGIS i Geomedia.

W podsumowaniu niniejszego rozdziału, z punktu widzenia obsługi topologii, możemy podzielić oprogramowanie GIS na trzy rodzaje:

1. „Persistent Topology” – trwałe (statyczny) zapis danych topologicznych. Miejsce zapisu tabel opisujących topologię może być różne: każda tabela to osobny plik binarny, tabele mogą zostać zapisane bezpośrednio do wektorowego pliku binarnego albo w relacyjnej bazie danych jako osobne tabele. Granice stykających się obszarów nie dublują się.
2. „On Load Topology” – (topologia dynamiczna) – obiekty przechowywane są osobno i podczas otwarcia programu lub załadowania warstwy odtwarzana jest topologia na czas pracy aplikacji GIS. W trakcie pracy aplikacja kontroluje topologię i najczęściej uniemożliwia zapisanie zmian w geometrii, które naruszałoby zasady topologii. W większości wektorowych formatów danych granice stykających się obiektów dublują się.
3. „On Demand Topology” – topologia tworzona jest tylko i wyłącznie w momencie, w którym uruchomimy narzędzie wymagające do swego działania warstw topologicznych, na przykład przy analizie nakładania warstw. Również w tym wypadku w większości wektorowych formatów danych granice stykających się obiektów dublują się.

5. Propozycja systematyki formatów zapisu danych w GIS

Na podstawie analizy informacji gromadzonych w systemach GIS przyjęto następujące założenia podziału formatów bazy danych GIS:

- a) obsługę zapisu topologii danych przestrzennych (Statyczna, Dynamiczna, Na Żądanie),
- b) format zapisu danych geometrycznych (plik binarny, BLOB, pole tekstowe w bazie),
- c) typ bazy danych systemu GIS (baza relacyjna, pojedynczy plik bazy relacyjnej, znaczniki XML).

Mając na uwadze powyższe parametry, zaproponowano następującą systematykę formatów danych GIS:

„Format katalogowo-plikowy” – warstwy geometryczne są umieszczane w pojedynczych plikach binarnych, tabele atrybutów każdej warstwy są również w pojedynczych plikach baz danych. W tych wypadkach istotna jest organizacja katalogów.

Dla otwarcia projektu należy wskazać katalog główny. Model ten używany jest przez stare wersje Arc/Info, system GRASS. Struktury katalogów tych programów różnią się. Katalog w systemie Grass pozwala na przechowywanie wielu warstw. W Arc/Info każdy katalog reprezentuje osobną warstwę, tabela atrybutów danej warstwy znajduje się w osobnym katalogu INFO. W systemie GRASS istnieje podział na lokalizacje, każda lokalizacja ma swój układ odniesienia i położenie, nazwy podkatalogów definiują rodzaj przechowywanych danych, natomiast dane jednej warstwy są przechowywane w kilku katalogach, w plikach o tej samej nazwie.

„Format wektorowo-relacyjny” – jeden plik geometrii zawiera wiele warstw, jedna baza zawiera wiele tabel. Przykładem mogą tu być MGE, AutoCad Map, i Geographics, które obsługują połączenia obiektów geometrycznych do atrybutów zawartych w relacyjnej bazie danych. Takie rozwiązanie powstało w wyniku rozszerzenia możliwości systemu CAD przez powiązanie obiektów z relacyjną bazą danych.

„Format wektorowo-znacznikowy” – jeden plik geometrii zawiera wiele warstw, jedna feature (warstwa) zawiera jedną tabelę w formacie XML. Przykładem może tu być Bentley MAP, który obsługuje model XFM (XML Feature Modeling). Obiekt geometryczny posiada atrybuty xml. Geoatrybutowe zapytania do obiektów geometrycznych realizowane są przez X-path. Świetnie funkcjonuje w środowisku CAD, może też zapisywać topologię do pliku wektorowego. Istnieje opcja tego formatu związana ze stałym przechowywaniem topologii.

„Format znacznikowy” – podobnie do formatu katalogowo-plikowego, warstwy umieszczane są w plikach stworzonych zgodnie ze standardem XML, a geometria zdefiniowana jest przez określony standard. W przypadku GIS mamy gotowe standardy OGC [4], takie jak GML 3.1 a w oraz identyczny format standard ISO – 19136 „Geography Markup Language” [5]. Pozostałe formaty znacznikowe (VML, SVG czy VRML), obsługujące grafikę wektorową, nie obsługują geograficznych układów współrzędnych. W tym formacie danych można również zapisać topologię, która zdefiniowana jest w schemacie GML o nazwie topology.xsd standardu GML.

„Format geobazy” – jedną z coraz bardziej popularnych metod jest łączny zapis danych graficznych i atrybutowych bezpośrednio do jednej, tej samej tabeli bazy danych. Stosowane są generalnie dwa rodzaje rozwiązań: grafika w typie danych BLOB lub w postaci zakodowanego pola (macierzy) SDO_Geometry, lub stosowanych w rozwiązaniach open source formatach WKT (Well Known Text) i WKB (Well Known Binary). I tak firmy Intergraph w GeoMedia oraz ESRI w nowym produkcie ArcGIS umieszczają dane graficzne i atrybutowe w tej samej tabeli bazy danych (najczęściej jest to ACCESS). Jedna tabela przechowuje jedną warstwę, a każdy pojedynczy obiekt tej warstwy jest umieszczony w osobnym rekordzie zawierającym opis tego obiektu.

„Format topologicznej geobazy” – jedną z absolutnie najnowszych metod zapisu danych geometrycznych jest format zapisu tabel topologii bezpośrednio do bazy danych. Możliwe jest to za pomocą nowego typu danych SDO_TOPO_GEOMETRY. Cztery tabele zapisujące dane o topologii poligonów zostały odwzorowane w systemowych tabelach bazy danych *Oracle*. Tabele te wprowadzono do bazy *Oracle* w wersji 10g [6].

6. Wnioski

W ramach niniejszego artykułu zidentyfikowano trzy parametry stosunkowo precyzyjnie charakteryzujące formaty danych w Systemach Informacji Przestrzennej, są to: zapis geometrii danych wektorowych, wykorzystanie danych o topologii warstwy wektorowej oraz miejsce zapisu danych atrybutowych. Na tej podstawie usystematyzowano formaty danych przestrzennych stosowanych w Systemach GIS. Można je podzielić na następujące formaty: topologiczna geobaza, geobaza, znacznikowy, wektorowo-relacyjny, wektorowo-znacznikowy i katalogowo-plikowy. Oczywiście zaproponowana przez autora oryginalna systematyka stanowi dopiero wstęp do dyskusji na ten temat i powinna zostać w przyszłości doprecyzowana. Jednak z tego pobieżnego przeglądu widać, jak dużym problemem jest osiągnięcie optymalnej metody składowania i efektywnego operowania na danych przestrzennych. Zaprezentowane rozwiązania firm komercyjnych zmieniają się w czasie oraz widać jak w ostatnich latach firmy zintensyfikowały poszukiwania optymalnego formatu dla zarządzania danymi przestrzennymi. Opracowana systematyka pozwoli na lepszy i bardziej precyzyjny opis formatów danych stosowanych do realizacji projektów GIS oraz pewnych uwarunkowań z nimi związanych, z istnienia których zwykły użytkownik GIS często nie zdaje sobie sprawy.

BIBLIOGRAFIA

1. Gaździcki J.: Systemy Informacji Przestrzennej. PPWK, Warszawa 1990.
2. Support ESRI,
<http://support.esri.com/index.cfm?fa=knowledgebase.techarticles.articleShow&d=12368>.
3. Kothuri R., Godfrind A., Beinat E.: Pro Oracle Spatial for Oracle Database 11g. Apress, USA 2007.
4. OGC – The Open Geospatial Consortium, <http://www.opengeospatial.org/>.
5. ISO/IEC 19136, Information technology – Geography Markup Language.
6. Oracle Spatial Topology and Network Data Models Developer's Guide 11g Release 2 (11.2) E11831-0.

Recenzenci: Dr inż. Piotr Bajerski
Dr hab. inż. Marcin Gorawski

Wpłynęło do Redakcji 31 stycznia 2011 r.

Abstract

This article describes the issue of storing spatial data in geographic information systems. The method of storage of spatial data was identified. Based on the analysis of structures, data formats, and ways of combining textual data with geometrical data records used. Systematic data format takes into account the possibility of permanent storage of data on the topology of geometric objects and to connect to and use of textual data. At the end of article the attempts to develop parameters for classifying GIS data formats was made.

Adres

Artur KRAWCZYK: Akademia Górniczo-Hutnicza, Katedra Ochrony Terenów Górniczych,
Geoinformatyki i Geodezji Górniczej, al. Mickiewicza 30, 30-059 Kraków, Polska,
artkraw@agh.edu.pl.