Topologikus térbeli adatstruktúrák

Elek István

Topologikus térbeli adatstruktúrák

TYPOTEX

A könyv megjelenését a Magyar Tudományos Akadémia támogatta.

© Elek István, Typotex, Budapest, 2015
Engedély nélkül semmilyen formában nem másolható!

Lektorálta: dr. Nikovits Tibor (ELTE, Információs Rendszerek Tanszék)

ISBN 978-963-279-862-2

Témakör: *térinformatika, adatbázis-technológia*

Kedves Olvasó!
Köszönjük, hogy kínálatunkból választott olvasnivalót!
Újabb kiadványainkról, akcióinkról
a *www.typotex.hu* és a *facebook.com/typotexkiado*
oldalakon értesülhet.

Kiadja a Typotex Elektronikus Kiadó Kft.
Felelős vezető: Votisky Zsuzsa
Főszerkesztő: Horváth Balázs
Műszaki szerkesztő: Csépány Gergely
A szöveget gondozta: Czene István
Borítóterv: Szalay Éva
Készült a Kódex Könyvgyártó Kft. nyomdájában
Felelős vezető: Marosi Attila

Tartalomjegyzék

Bevezetés **1**

1. A spagettimodell áttekintése **5**
   1.1. Vektoros adatok létrehozása 5
   1.2. Kapcsolatok, rétegek, feature classok 10
   1.3. Spagettileírás relációs logikával 13

2. Az OGC-topológia kezelése **15**
   2.1. Fogalommeghatározások 15
   2.2. A geometry object modell 19
   2.3. SQL-implementációk 43
   2.4. Összegzés 49

3. Ipari szoftverek megoldásai **51**
   3.1. PostGIS térbeli objektum típusai 51
   3.2. A PostGIS felépítése 52

4. Redundanciamentes adatszerkezetek **63**
   4.1. Topológialeírás relációs adatbázis-kezelőkkel 63
   4.2. Poligontopológia-leírás duális gráfokkal 81

5. A geometriai megközelítés fonákságai **91**
   5.1. A hálózatmodell 91
   5.2. Összegzés 98

6. Formátumleírások **99**
   6.1. A Well-known formátumok 99
   6.2. Mapinfo MIF/MID 100
   6.3. A DAT szabvány elemei 111

7. Gráfelmélet-összefoglaló **121**

Ábrák jegyzéke

1.1. Az ELTE IK madártávlatból 7
1.2. Egy- és kétdimenziós objektumok 8
1.3. A vektoros rendszerek elemei 9
1.4. A grafikus és a leíró adatok kapcsolata sematikusan 10
1.5. Objektumcsoportok 11
1.6. Topológiahibák 12
1.7. Spagetti és relációs modellje 14
1.8. A spagettitopológia hibái 14

2.1. Az OGC-geometria osztályai 20
2.2. A geometry osztály műveleteinek elvi felépítése 21
2.3. A geometriagyűjtemény (geometry collection) és metódusai 24
2.4. A pont és metódusai 25
2.5. A görbe és metódusai 26
2.6. Példák MultiLineStringekre 27
2.7. A MultiCurve és metódusai 27
2.8. Példák MultiLineStringekre 28
2.9. A surface és metódusai 29
2.10. Példák poligonokra 30
2.11. Példák nem megengedett poligonokra 30
2.12. A Polygon és metódusai 31
2.13. A polyhedral surface körbejárási irányokkal 31
2.14. A polyhedral surface és metódusai 32
2.15. Multisurface műveletek 33
2.16. Példák MultiPolygonokra 34
2.17. Meg nem engedett MultiPolygonok 35
2.18. Átfedő a, b poligonok 36
2.19. Példák érintő geometriai objektumokra 39
2.20. Példák *Within* geometriai relációra 41
2.21. Példák *Overlap* geometriai relációra 42

2.22. Egy adatbázisséma a feature table leíráshoz 44
2.23. Egy geometry table példa poligongeometriára 46
2.24. A *Geometry Type* sémája 47
2.25. SQL-geometriatípus hierarchiája 49

4.1. Poligonokat felépítő kontúrok rendszere 65
4.2. A relációs táblák kapcsolatrendszere 66
4.3. A parti települések lekérdezésének eredménye 67
4.4. Egy földtani térképrészlet 68
4.5. Az A,B,C,D poligonok és node-jaik 69
4.6. Az A,B,C,D poligonok reprezentációja relációs táblákkal . . 70
4.7. Az A,B,C,D poligonok grafikai megjelenése 71
4.8. Az A,B,C,D poligonok leírása MIF/MID fájlokkal 72
4.9. Az A,B,C,D poligonok leírása topologikus adatszerkezettel . 75
4.10. Budapesti kerületek poligonjai 76
4.11. Magyarország megyéi a Dunától keletre 77
4.12. Járások poligonjai a Dunától keletre 77
4.13. Egy város egy részének telkei 78
4.14. Kereszteződő vonalak 78
4.15. Kereszteződő vonalak közös pontjai 79
4.16. Vonalak redundanciamentes tárolása 80
4.17. Példa síkgráfra 83
4.18. Síkgráf és duálisa 84
4.19. Befoglalási fa 85
4.20. Síkgráf létrehozása 87
4.21. Síkgráf megváltoztatása 89

5.1. Hálózatok 92
5.2. Node-ok és szelvénybeosztások 93
5.3. Lineáris referencia 94
5.4. Vonalláncok és szelvényezés 95
5.5. Dinamikus szegmentálás 96
5.6. Szegmentálás több attribútumadat szerint 97

7.1. A königsbergi hidak sematikusan ábrázolva 121
7.2. A königsbergi hidak gráfja 122
7.3. A G gráf csúcsai és élei 123
7.4. Példa irányított gráfra 123

7.5. Példa hurkot is tartalmazó gráfra 123
7.6. Példa többszörös élt is tartalmazó gráfra 124
7.7. Gráf szomszédsági mátrixa 125
7.8. Gráfok és komplemenseik 127
7.9. Példák izomorf gráfokra [8] 127
7.10. Példa teljes gráfokra [8] 128
7.11. Út, vonal, kör, séta 129
7.12. Vasúti példa [2] 130
7.13. Egy négypontos gráf és szomszédsági mátrixa 133
7.14. Egy egyszerű, öt vertexű gráf, súlyozott élekkel 135
7.15. Nem összefüggő gráfok 136
7.16. Fák és erdők [8] 136
7.17. Bináris fa 137
7.18. Egy gráf és duálisa 138

Táblázatok jegyzéke

2.1. Rövidítések 16
2.2. Jelölések 17
2.3. Az intersection mátrix a DE-9IM modell szerint 36
2.4. Az intersection mátrix a,b átfedő poligonokra 36
2.5. Magyarázat a feature table-t leíró adatbázissémához 44

5.1. Az n_1,n_2,n_3,n_4 node-ok georeferencia táblázata 95
5.2. Az n_1,n_2,n_3,n_4 node-ok a lineáris referenciatáblázata 95
5.3. A **K** jelű szegmens meghatározása a lineáris referencia szerint 96

6.1. Geometriareprezentációk WKT-ben 101
6.2. Matematikai transzformációk leírása WKT-ben 102
6.3. Koordináta-rendszer leírása WKT-ben 103

7.1. A K_i-ből L_j-be közlekedő vonatok táblázata 131
7.2. A M_i-ből L_j-be közlekedő vonatok táblázata 131
7.3. A K_i-ből M_j-be közlekedő vonatok táblázata 132

Bevezetés

A topológia jól ismert fogalom a térinformatikával foglalkozók körében. Minden piacvezető szoftver tartalmaz funkciókat a topológia kezelésére, létrehozására, javítására. A térbeli lekérdezések miatt a topologikus adatok megléte alapvető adatminőségi követelmény.

A térképi adatok létrehozásakor, importjakor kell felépülnie a megfelelő topológiának, amely eléggé gyakran nem teremthető meg automatikusan. Raszteres adatok feletti fél automatikus (kézi) vektorizáláskor a topológia biztosítása alapvetően a vektorizálást végző személy szakértelmétől függ. Továbbá az export/import műveletekből származó adatok topológiája erősen függ az eredeti adatforrás minőségétől. Amennyiben az adatforrás strukturálatlanul tartalmazza a vektoros adatokat (spagettimodell), és topológiai hibákat tartalmaz, akkor az import során keletkezett adatok sem lesznek topologikus minőségűek.

Az importált adatok topológiájának létrehozása külön beavatkozást igényel, amennyiben az eredeti adathalmaz tartalmazott topológiai hibákat. A piacvezető GIS szoftverek funkcionalitása számos eszközt biztosít erre a célra. Nemcsak a drága kereskedelmi szoftverek, hanem az ingyenesen hozzáférhető nyílt forráskódú rendszerek is fel vannak készülve a topológiai hibák javítására, a topologikus minőségű adatrendszer megteremtésére.

A könyvben bemutatjuk a különböző vektoros adatmodelleket, a spagettimodellt, és a topológia megőrzésére alkalmas más modelleket is. Különösen sokat fogunk foglalkozni a poligonok topológiájával, amely a legösszetettebb ebben a vonatkozásban. A spagettimodellben is lehetséges topologikus minőségű adatrendszerek létrehozása, de az adatbázis szerkezetileg lehetővé teszi topológiai hibák létrejöttét. A spagettimodell nemcsak ezért korszerűtlen, hanem azért is, mert redundánsan tárolja az objektumok node-jait. A redundáns tárolás további hibák forrása lehet, ezért fontos olyan adatszerkezetek kidolgozása, amely megszünteti a redundanciát, és garantálja az egyszer már létrejött topológiai megőrzését. Ilyen adatszerkezeteket is ismertetni fogunk.

Be fogjuk mutatni a piacvezető adatbázis-kezelők által a térbeliség kezelésére kialakított fogalmakat és funkcionalitásokat. Kiválasztottunk egy meghatározót a neves adatbázis-kezelők közül, amely végül a Postgres/PostGIS lett. Látni fogjuk, hogy térbeli függvénykönyvtár a spagettimodellre épül. Viszonylag részletesen, de nem minden részletre kitérve, ismertetni fogjuk az OGC elvi megalapozását, és annak egy SQL-implementációját.

Részletesen ismertetünk egy redundanciamentes, a topológiát szerkezetéből következően megőrző adatstruktúrát. Ha korszerűbb, megbízhatóbb is a topológiát megőrző adatszerkezet, a spagettire kidolgozott függvénykönyvtár nem alkalmazható rá. Ez az adatszerkezet tehát nem OGC kompatibilis, de előnyös tulajdonságai miatt a tárgyalását indokoltnak találtuk. Ha bebizonyosodik, hogy a topologikus adatszerkezet a spagettinél jobb tulajdonságokkal rendelkezik, és performancia szempontból is megfelelő, akkor implementálni kell a teljes térbeli függvényhalmazt.

A könyv utolsó fejezetében részletesebben tárgyaljuk a vonalas objektumok mentén való helymeghatározást, noha nem csak a topológia szempontjából. Vonalas objektumokra is felvázolunk egy topológia megőrző adatszerkezetet, amely szintén nem OGC kompatibilis, de a redundanciamentes tárolás miatt úgy gondoltuk, hogy érdekes lehet. A vonalas objektumok menti helymeghatározás megtárgyalása során megmutatjuk, hogy a tisztán geometriai elemeken nyugvó leírás számos esetben nem kielégítő a valóság megfelelő leírására.

Úgy gondoljuk, hogy mindkét redundanciamentes topológiamegőrző adatstruktúra érdekes lehet bizonyos állami alapadatok (pl. kataszteri térképek, topográfiai térképek) tárolása szempontjából. Igaz, ma még egyetlen GIS szoftver sem követ ehhez hasonló struktúrát, hiszen vagy az OGC szabvány szerint működik, vagy saját implementációt használ, de a redundanciamentes struktúrából bármikor exportálható spagettitopológiájú adat (pl. ESRI shape, vagy Mapinfo MIF), és így bármely szoftverben előállítható az annak megfelelő munkaformátum. Egyébként bármely standard formátumú adat importjakor át kell esnünk ezen a procedurán, függetlenül attól, hogy OGC szabványt követ vagy sem.

Mintegy mellékesen bemutatunk néhány önkényesen kiválasztott, ismert adatformátumot, amely részben munkaformátum is (WKT), részben pedig csak adatcsere céljára szolgál (MIF/MID, DAT adatcsere formátum).

A könyv készítésekor azt az elvet követtük, hogy az angol terminológiát megtartottuk, mivel számos szónak jelenleg nincs még magyar fordítása,

és ahol lehetett, a szavak, fogalmak magyar megfelelőjét zárójelben feltüntettük az angol szavak mellett. Ettől persze kicsit vegyes lett a kép, de arra semmiképpen nem akartunk vállalkozni, hogy kísérletezzünk a fogalmak magyar fordításával. Remélhetőleg idővel kialakulnak a szakma által elfogadott, meghonosodott magyar szavak is.

2014. június, Budapest Dr. Elek István

1

A spagettimodell áttekintése

A vektoros adatmodell az ábrázolni kívánt objektumok általunk jellemzőnek, fontosnak tartott pontjainak helyvektorait tárolja, ami kétdimenziós esetben x, y koordinátákat, háromdimenziós esetben x, y, z koordinátákat jelent. Azt is tárolnunk, tudnunk kell, hogy a megadott koordináták milyen koordináta-rendszerben értelmezettek. Ez a tény feltételezi, hogy az a szoftver, amely tárolja a pontok koordinátáit, kezelni tudja a koordináta-rendszereket.

Minden, a pontnál bonyolultabb objektum, a helyvektorosan tárolt pontok sokaságából épül fel valamely összekötési szabály alapján (amely legegyszerűbb esetben lehet az összetartozó pontok beviteli sorrendje). A matematikai összefoglaló gráfelméleti fejezetével összhangban, a vektoros adatmodellben az ábrázolni kívánt objektumokat gráfokkal írjuk le. A jellegzetes pontok a gráf pontjai (vertexei, node-jai, magyarul csúcsai, csomópontjai), amelyek alkotják a vonalakat és a poligonokat.

1.1. Vektoros adatok létrehozása

Kérdés, hogy ki dönti azt el, hogy melyik pont jellegzetes illetve fontos, elvégre éppen ezeket akarjuk tárolni. A kérdés megválaszolásához vizsgáljuk meg a vektoros adatbevitel módszereit. Ha űrfelvételekről vagy légifotókról vektorizálunk, előbb értelmeznünk kell, hogy mit látunk. Ha felismertük a folyókat, utakat, házakat, akkor azok jellegzetes pontjai (ház sarkai, folyók, utak jellemző pontjai, úgy mint torkolat, útkereszteződés, stb.) egyszerűen megállapíthatók. A folyamat kiemelten fontos momentuma a felismerés, az emberi szemlélő, interpretátor tudása, tapasztalata, háttér ismeretei. Enélkül a vektorizálás lehetetlen. Az automatikus vektorizálás eddig megoldatlan mivolta éppen azt bizonyítja, hogy a vektorizálás igen magas szintű intelligenciát igényel, mint amilyen az interpretátor tapasztalata, térképészeti és légi-

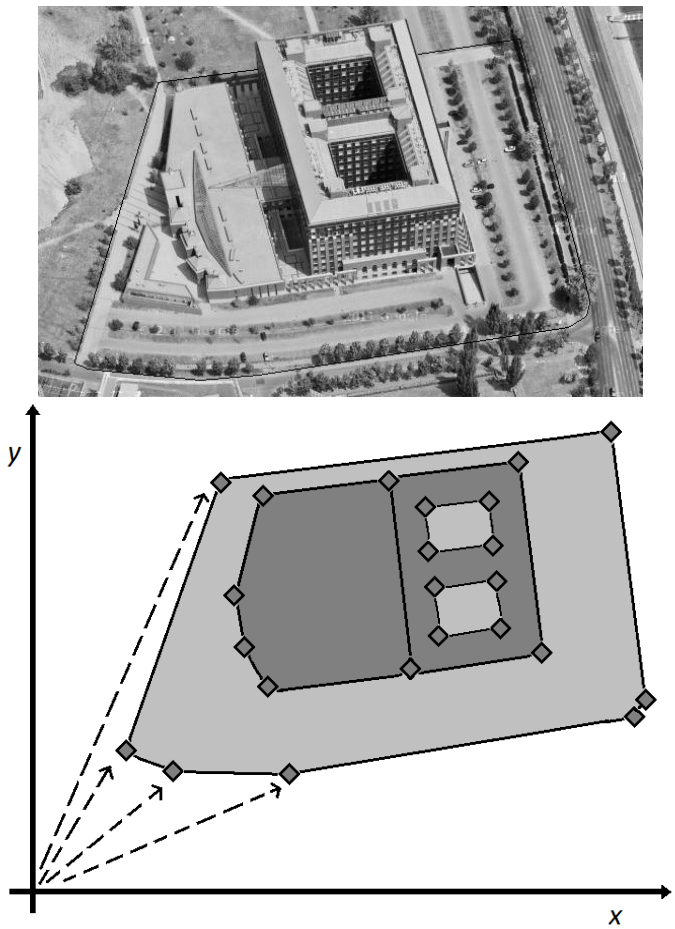
fotó értelmezési képességei. Ezek a legtöbb esetben fejlett képfeldolgozási, képfelismerési képességeket jelentenek, amelyekre mind ez idáig számítógépeket nem sikerült megtanítani. Az eddig legsikeresebb térképvektorizáló programok félautomatikus módon képesek csak működni, emberi szakértő folyamatos jelenléte mellett, aki olyan esetekben, amikor a program értelmezhetetlen grafikai szituációt észlel, beavatkozik.

Speciális esetekben persze lehetséges automatikus vektorizálás, mint például csak szintvonalakat tartalmazó raszteres térkép, de ha komplett térképek raszteres állományai képezik a vektorizálás tárgyát, a programok képtelenek kezelni a túlzottan bonyolult grafikai szituációt. Gondoljunk csak egy topográfiai térképre, amelyet színnel kitöltött felületek, eltérő vastagságú, mintázatú és színű vonalak, szintén sokféle pont, szimbólum, és végül számos felirat alkot. A különböző tematikák egy papírlapon való ábrázolása gyakran eltérő stílusú vonalak együttfutását eredményezi. A térkép olvashatósága miatt ilyenkor különböző ábrázolási konvenciók nehezítik a gépi felismerést. A térképet kevéssé ismerő ember számára is nehezen érthetőek ezek a helyzetek.

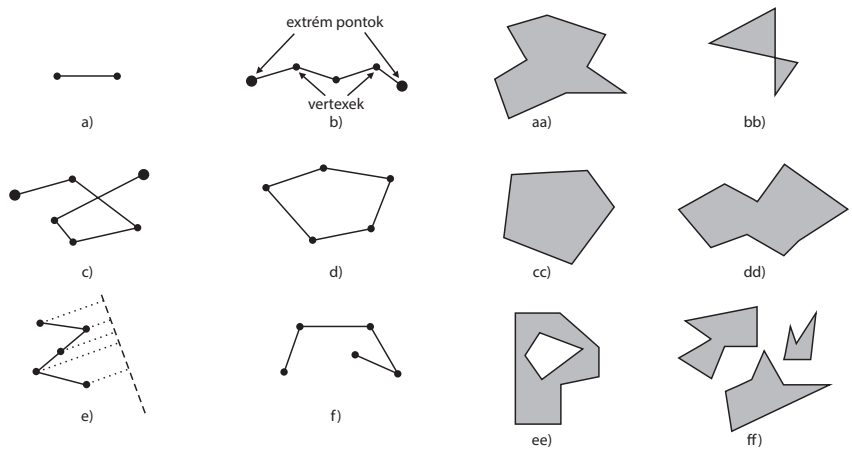
További vektorizálási szempont, hogy milyen célra készül a vektortérkép. Az előre definiált célokból következtetni lehet arra, hogy mely objektumot tekintsük fontosnak, és melyet csak háttérnek, vagy éppen elhanyagolhatónak. Példaként vegyük azt az esetet, amikor közigazgatási célból készítünk vektoros térképet szkennelt papírtérkép alapanyagból. Ilyenkor minden közigazgatási vonatkozású adat fontos (településhatár, megyehatár, stb.), ellenben a folyók pontos ábrázolást nem igényelnek, noha jelenlétük a tájékozódás szempontjából szükséges. A patakok elhanyagolhatók, nem is kell, hogy rákerüljenek a digitális térképre (még akkor sem, ha rajta vannak a papír alapanyagon).

Amikor papírtérkép a vektorizálás nyersanyaga, akkor is igaz, hogy valamikor valakinek értelmeznie kellett a képet, valakinek valamilyen előismeretei alapján ki kellett tűzni mérési pontokat. Összefoglalva tehát elmondhatjuk, hogy a vektoros adatmodellt követő digitális térképekben a helyvektorokon és az összekötési szabályon kívül nagy mennyiségű emberi tudás és szakértelem is benne van.

A 1.1. ábrán látható egy idealizált kép a valóság egy szeletéről, amelyet vektorosan írunk le. A valóságról alkotott modellünkhöz geometriai elemeket használunk, úgymint a nulla dimenziós pont, és az egydimenziós vonal.



1.1. ábra. A valóság egy darabja, az ELTE IK madártávlatból (fenti ábra), és ugyanez a vektoros adatmodellel leírva (alsó ábra). A helyvektoros leírás – mint látható – elhelyezte az ábrázolni kívánt objektumok felszínre merőleges vetületeinek jellemző pontjait egy koordináta-rendszerben. Az erősen kiemelt sarokpontok helyvektorait tároljuk, valamint a pontok összekötési szabályát



1.2. ábra. Példák egydimenziós objektumokra (bal oldali ábra): a) vonal szegmens (szakasz); b) vonallánc (polyline) extrém pontokkal (végpontokkal) és vertexekkel; c) nem kereszteződő vonallánc; d) zárt polyline; e) monoton polyline; f) nem monoton polyline.
Példák kétdimenziós objektumokra (jobb oldali ábra): aa) egyszerű poligon; bb) nem egyszerű poligon (ilyen poligon csak hibaként fordul elő a geoinformatikában); cc) konvex poligon; dd) monoton poligon; ee) poligon szigetstruktúra (lyukas poligon); ff) diszjunkt poligonokból álló régió