

DESIGN OF DATA STRUCTURES FOR GIS REPRESENTING HISTORICAL EVENTS

Roman BOHOVIC, Bachelor Degree Programme (3)
Dept. of Computer Graphics, FIT, BUT
E-mail: xbohov00@stud.fit.vutbr.cz

Supervised by: Ing. Vítězslav Beran

ABSTRACT

Main purpose of this work is to design simple standards and data structures of historical and spatial information for representation in geographic information systems. The most important attribute of information is spatial representation. The time aspect is also important in this case. Other attributes are optional and can be used as supplement. This is implemented in shapefile format with some additional rules. It also deals with geographical data sources.

1 ÚVOD

Pri návrhu geografického informačného systému (ďalej GIS) je potrebné zohľadňovať účel, pre ktorý je systém vytváraný. To zahŕňa rozbor funkcií, ktoré bude systém poskytovať. Od toho sa odvíja charakter informácií, ktoré je treba získať, uchovávať a vhodne a efektívne používať. Tým je podmienený samotný návrh dátových štruktúr. Pri návrhu štandardu digitálnej reprezentácie dát pre GIS, ktorého hlavnou funkciou bude predovšetkým názorné zobrazovanie historických udalostí, vychádzam z dvoch kľúčových charakteristík – polohy a času. Pritom je potrebné rozlišovať historické dáta ako také a geografické dáta s časovým aspektom, ktorých návrhom sa zaoberám.

2 URČENIE POLOHY

Každý GIS sa snaží určitým spôsobom popísať krajinu s jej prvkami a interakciami. Úplný popis nie je nikdy možný, podstatná je úroveň abstrakcie, ktorú pri popisovaní používame. Bez ohľadu na to je potrebné vyriešiť otázku reprezentácie priestoru v systéme.

Poloha je v ľudskom ponímaní chápaná 3-rozmerne, pričom v systémoch je často generalizovaná len na dve dimenzie. Je potrebné ale jednoznačne určiť sústavu, v ktorej sa budeme pohybovať. Používajú sa sférické súradnice, geografická šírka a dĺžka, na popisanie polohy bodu na Zemskom povrchu. Nastáva ale pritom problém s výberom referenčného telesa (guľa, 2-osí elipsoid - Besselov, Krasovského, 3-osí elipsoid atď.).

V navrhovanom systéme je však jednoznačne nutné používať jediný referenčný elipsoid. Z tohto hľadiska je najvýhodnejšie použiť World Geodetic System (WGS-84), ktorý

je síce v oblasti Európy relatívne nepresný, ale globálne najpoužívanější. Dôležitým dôvodom pre jeho zavedenie je, že sa používa v systémoch GPS.

3 PRIESTOROVÉ DÁTA – RASTROVÝ A VEKTOROVÝ POPIS

V reprezentácii grafických dát v počítačoch existujú dva rozdielne prístupy: rastrový a vektorový. To platí aj o geografických objektoch (chápaných ako digitálna mapa). Prechod medzi oboma modelmi je náročný, aj keď väčšina GIS-ov vie pracovať s oboma.

Jednoznačnou nevýhodou **rastrového popisu** je objem takto uložených dát. Na druhej strane sú rastrové dáta spojité, zahŕňajú celý povrch, čo je pre grafický výstup lepšie.

Preto návrh obsahuje rastrovú výškovú mapu, ktorá pokrýva celú zemeguľu a slúži ako podklad pre výstupné zobrazenie systému. Výhodou podkladu je, že výstupná informácia nebude obsahovať žiadne „biele miesta“ a zároveň je nadmorská výška charakteristikou, ktorá sa z hľadiska ľudskej histórie v podstate nemení a je dôležitým faktorom v krajine.

Ako zdroj k vytvoreniu potrebného rastru slúži GTOPO30. Je to svetový digitálny model terénu s horizontálnym rozlíšením najmenej jednotky 30'', čo v teréne predstavuje asi 1 km. Ako referenčný je použitý pri týchto dátach systém WGS-84. Hodnota každého pixelu je rovná priemernej nadmorskej výške na povrchu v celých metroch, oceány sú reprezentované hodnotou NO DATA (-9999). Všetky dáta aj s metadátami majú neskompimovanú veľkosť 2,72 GB (1,74 GB dátová časť) a sú rozdelené na 33 častí. Kompletne a bez obmedzení sú tieto dáta poskytnuté zdarma verejnosti americkou štátnou organizáciou U.S. Geological Survey, ktorá dáta spracovala z viacerých zdrojov.

Vo vytváranom systéme je síce výškový model terénu súčasťou geografickej časti, svojou rastrovou povahou sa však líši od ostatných navrhovaných GIS dátových štruktúr.

Vektorový popis objektov (bod, úsečka, polygón) v GIS-och má niekoľko výhod. Predovšetkým je presnejší (čo v tomto návrhu nie je určujúca vlastnosť) a potrebné miesto na jeho uloženie je rádovo menšie ako pri rastrovom popise. Zároveň niektoré implementácie umožňujú ukladať objekty so zachovaním topológie, čo umožňuje náročnejšiu a efektívnejšiu analýzu dát v systéme. Cenou za to je však podstatne zložitejší návrh a interpretácia.

Preto sa prikláňam k jednoduchému popisu, ktorý je menej náročný, prehľadnejší a podstatne rýchlejší sa s ním pracuje. Vzhľadom k tomu, že zložitejšia analýza sa pri zobrazovaní historických udalostí nepredpokladá, nie je to veľkým obmedzením.

Navrhnuté vektorové priestorové primitívi pre popis tvaru GIS objektov, vychádzajú zo štandardu vytvoreného firmou ESRI, ktorá navrhla pre svoje produkty shapefile formát. Dôvod pre použitie tohto formátu je existencia voľne dostupných dát (napríklad hranice štátov, riek, jazier atď.) v tomto formáte ako aj jeho jednoduchosť. Formát pozostáva z 3 súborov, ktorých meno musí byť rovnaké, líšia sa iba koncovkou:

- Hlavný – name.shp – obsahuje premenlivý počet záznamov (rovnakého typu) s popisom tvaru objektov pomocou ľubovoľného počtu bodov [lat/lon], schéma súboru je v Tab. 1.
- Atribútový – name.dbf – obsahuje položky pre každý záznam (vzťah 1:1) z hlavného súboru, pričom položka pozostáva z ľubovoľného počtu vlastností - atribútov.
- Indexový – name.shx – obsahuje offset príslušného záznamu z hlavného súboru.

Hlavička súboru [100B] - kód (9994) a dĺžka súboru, typ popisovaných tvarov (1 bod, 3 čiara, 5 polygón), hranice oblasti			
Číslo záznamu (identifikačné)	Dĺžka záznamu	Typ tvaru	Popis (závisí od tvaru objektu)
....
Číslo záznamu (identifikačné)	Dĺžka záznamu	Typ tvaru	Popis (závisí od tvaru objektu)

Tab. 1: Schéma stavby hlavného súboru shapefile formátu (podľa [2])

Popis objektov sa líši podľa potreby jednotlivých tvarov (vždy však pozostáva z pospájaných bodov), obsahuje aj súradnice obdĺžnika, v ktorom sa objekt nachádza, počet bodov, z ktorých sa skladá a v prípade polygónu aj počet častí. Oblasti totiž môžu byť aj vnorené a zložené z viacerých častí, čo je veľmi užitočné (napríklad ostrovný štát je uložený ako jediný záznam, aj keď pozostáva z viacerých nespojitých území).

4 UCHOVANIE ČASU

Existujúce GIS-y väčšinou nepočítajú s časovou dimenziou. Tým sa odlišujú klasické geografické dáta od dát používaných v systéme pre zobrazovanie historických udalostí. Preto je potrebné počítať s ďalším rozmerom. Vzhľadom k tomu, že geografické dáta majú pomerne stály charakter, je možné čas riešiť ako vlastnosť každého objektu. Preto v návrhu objektov zavádzam dva atribúty s časovou platnosť geografického objektu OD a DO.

To umožňuje modelovanie historických udalostí spolu s GIS dátami. Pritom nie je potrebné, aby boli zadané časové údaje vždy. Hodnota NO DATE (00.00.0000) znamená neobmedzenú platnosť objektu jedným, alebo oboma smermi (do minulosti aj budúcnosti). Ak nie je podstatný (alebo známi) presný dátum, je možné zadať nultý deň a mesiac, to neplatí pre rok. Udalosti pred našim letopočtom sa jednoducho udávajú ako záporné čísla pre roky.

5 ZÁVER

Dátové štruktúry sú navrhnuté tak, aby bolo možné použiť ich pre akékoľvek priestorové informácie, nielen geografického charakteru. Zároveň použité rozšírenie o časový atribút umožňuje širšie použitie v systéme zobrazujúcom historické udalosti. Vytvorený rámec však nekladie obmedzenia, ktoré by ho predurčovali len na zobrazovanie. Naopak je možné ho ďalej využívať a prípadne rozširovať (napríklad analýza dát). Pri návrhu spolupracujem s ďalšími študentmi zapojenými do projektu Globe of History, ktorý má za cieľ kompletný návrh a realizáciu systému, ktorý umožní zobrazovanie historických udalostí.

LITERATURA

- [1] ESRI : ESRI Shapefile Technical Description, An ESRI White Paper – July 1998, USA
- [2] Kolář, J.: Geografické informační systémy 10, Praha, Vydavatelství ČVUT 1997, ISBN 80-01-01698-6
- [3] <http://edcdaac.usgs.gov/main.asp>