

DOTAZOVACÍ JAZYK PRO GEOGRAFICKOU ANALÝZU

QUERY LANGUAGE FOR GEOGRAPHICAL ANALYSIS

Tomáš Vaníček

Anotace:

V práci je definován pojem geografický výraz. Pomocí tohoto pojmu je definován dotazovací jazyk GSQL sloužící k jednoduchému a jednoznačnému zadávání úloh klasické geografické (prostorové) analýzy. Dále je zaveden pojem geografický analytický stroj a jsou naznačeny úlohy, které by měl takový program zvládat.

Summary:

In this paper concept geographical expression is defined. More there is defined query language GSQL useful for simply inputting of classical geographical (space) analysis problems. Concept of Geographical Analytical Machine is given and a class of problems to be solve by such program is described.

Klíčová slova:

Geografické informační systémy, Geografická analýza, Dotazovací jazyky

Key words:

Geographical Information Systems, Geographical Analysis, Query Languages

V této práci je definován pojem geografický výraz a jeho pomocí je definován dotazovací jazyk GSQL (Geographical Structured Query Language) sloužící pro zjednodušené zadávání klasických úloh geografické (prostorové) analýzy. Dále je zaveden pojem geografický analytický stroj a jsou naznačeny úlohy, které by takový program měl zvládat.

Geografická analýza

Geografická analýza je hlavní a nejdůležitější činností geografických informačních systémů. Možností provádět geografickou analýzu se GISy liší od "obyčejných" informačních systémů a "obyčejných" kreslících programů. Z toho důvodu jsou i datové formáty GISů stejně tak jako všechny činnosti vedoucí k získávání, údržbě a prezentaci dat koncipovány tak, aby co nejvíce vyhovovaly potřebám geografické analýzy. Geografická analýza nám pomůže odpovědět na otázky, pro jejichž vyhodnocení je nutné znát prostorové souvislosti mezi geografickými daty. Typické otázky, na které lze odpovědět pomocí geografické analýzy jsou například:

- Najdi mi nejbližší okresní město k vesnici jménem XY.
- Najdi v okolí 2 km Lipenské přehrady plochu o velikosti alespoň 10 ha vzdálenou alespoň 500 m od silnice a maximálně 800m od vodního toku.
- Zjisti, které oblasti státu nejsou pokryty televizním signálem na základě znalostí polohy a dosahu jednotlivých vysílačů.
- Přiřaď ke každé škole ve městě počet dětí, které bydlí v okruhu 10 minut cesty pěšky od školy.

Základním prostředkem pro provádění geografické analýzy jsou operace nad mapovými vrstvami. Vstupem do těchto operací je vždy jedna nebo více mapových vrstev, jejich

vstupem pak jedna výsledná mapová vrstva. Každý GIS obsahuje jistou množinu programů pro provádění těchto operací. Mezi základní prostorové operace patří následující:

- Vytváření obalových zón (buffer).

Vstupem je vrstva A, která může být bodová, čárová i polygonální, a vzdálenost k zadaná obvykle číselnou hodnotou. Výstupem je polygonální vrstva B. Polygony vrstvy B obsahují ty body, pro které lze nalézt alespoň jeden bod uvnitř polygonu (na lomené čáře) vrstvy A ve vzdálenosti menší nebo rovné k.

- Sjednocení (union).

Vstupem jsou polygonální vrstvy A a B, výstupem polygonální vrstva C. Vrstva C obsahuje oblast, která je sjednocení oblastí popsaných ve vrstvách A a B. Přitom hranice polygonů ve vrstvě C vzniknou tam, kde byly hranice polygonů buď ve vrstvě A nebo ve vrstvě B. Jedná se tedy vlastně o „položení dvou mapových vrstev přes sebe“.

- Ztotožnění (identity).

Vstupem je polygonální vrstva A a vrstva B libovolného typu. Výstupem je polygonální vrstva C. Na rozdíl od předcházejícího případu obsahuje vrstva C oblast totožnou s oblastí popsanou vrstvou A. Uvnitř této oblasti jsou opět všechny prvky vrstvy A i vrstvy B. Všimněte si, že na rozdíl od operace sjednocení není operace ztotožnění komutativní. Záleží při ní na pořadí operandů.

- Průnik (intersect).

Vstupem jsou polygonální vrstvy A a B, výstupem polygonální vrstva C. Vrstva C bude obsahovat oblast, která je průnikem oblasti popsané vrstvou A a oblasti popsané vrstvou B. Uvnitř této oblasti budou opět obsaženy všechny prvky vrstvy A i všechny prvky vrstvy B.

- Přepsání (update).

Vstupem jsou polygonální vrstvy A a B, výstupem polygonální vrstva C. Vrstva C obsahuje ty prvky vrstvy A, které neleží uvnitř žádného polygonu vrstvy B a všechny prvky vrstvy B. Jedná se vlastně o přepsání informací ve vrstvě A informacemi ve vrstvě B v těch oblastech, které jsou vrstvou B popsány.

- Vymazání (erase).

Vstupem jsou polygonální vrstvy A a B, výstupem polygonální vrstva C. Vrstva C obsahuje pouze ty prvky vrstvy A, které neleží uvnitř žádného polygonu vrstvy B. Jedná se o vymazání oblasti popsané vrstvou B z vrstvy A. Vymazaná oblast nebude ve výsledné vrstvě C popsána.

- Ořezání (clip).

Vstupem jsou polygonální vrstvy A a B, výstupem polygonální vrstva C. Vrstva C obsahuje ty prvky vrstvy A, které leží uvnitř nějakého polygonu vrstvy B. Ořezání je tedy v jistém smyslu doplňkovou operací k vymazání. Zužuje platnost informací obsažených ve vrstvě A pouze na oblast popsanou vrstvou B.

Kromě uvedených prostorových operací přijdou při geografické analýze ke slovu i dobře známé operace pro práci s relačními tabulkami. Tyto operace mohou využívat i hodnoty obsažené v atributech pro popisná data. Jedná se především o dvě základní operace:

- Filtrování tabulky.

Vstupem je jedna relační tabulka A (v našem případě tedy jedna mapová vrstva) a jeden logický výraz L, výstupem opět jedna relační tabulka B se stejnými atributy jako měla tabulka A. Tabulka B obsahuje pouze ty řádky tabulky A, pro které byl splněn výraz L. V praxi to znamená, že v mapové vrstvě B budou obsaženy pouze některé z prvků mapové vrstvy A. O jejich zařazení rozhodne splnění podmínky L.

- Faktorizace tabulky.

V nejjednodušším případě je vstupem relační tabulka A a jeden její atribut M. Výstupem je relační tabulka B obsahující pouze atribut M. Řádky tabulky B tvoří ty hodnoty atributu M, které se vyskytly v tabulce A. Při použití v geografické analýze se faktorizace tabulky projeví tak, že prvky, které mají stejnou hodnotu jednoho atributu se spojí do jednoho prvku. Takto můžeme například z polygonální vrstvy okresů, která jako popisný atribut ke každému okresu obsahuje název kraje, vytvořit polygonální vrstvu krajů.

Geografický výraz

Definici dotazovacího jazyka GSQL zahájíme definicí pojmu geografický výraz. Jedná se o jistou obdobu aritmetického výrazu, kde tvoříme výraz z čísel a proměnných pomocí aritmetických operátorů. V případě geografického výrazu tvoříme výraz z mapových vrstev pomocí operátorů geografických (prostorových). Po sémantické stránce má každý geografický výraz svou hodnotu, což je plošná mapová vrstva.

Definice geografického výrazu zní následovně:

Je-li XYZ plošná mapová vrstva, je XYZ též (atomický) geografický výraz

2. Je-li XYZ bodová, liniová, nebo plošná mapová vrstva a ε

reálné číslo, je

$buffer(XYZ, \varepsilon)$ geografický výraz. Jeho sémantická hodnota odpovídá plošné mapové vrstvě získané operací obalové zóny (buffer) o poloměru ε z mapové vrstvy XYZ.

3. Jsou-li XYZ a ABC dva geografické výrazy, je $union(XYZ, ABC)$ též geografický výraz. Jeho sémantickou hodnotou je mapová vrstva vzniklá operací sjednocení z mapových vrstev odpovídajících výrazům ABC a XYZ.

4. Jsou-li XYZ a ABC dva geografické výrazy, je $identity(XYZ, ABC)$ též geografický výraz. Jeho sémantickou hodnotou je mapová vrstva vzniklá operací identity z mapových vrstev odpovídajících výrazům ABC a XYZ. Samozřejmě i $identity(ABC, XYZ)$ je geografický výraz a vzhledem k nekomutativnosti operace identity je jeho sémantická hodnota odlišná od sémantické hodnoty výrazu $identity(XYZ, ABC)$.

5. Jsou-li XYZ a ABC dva geografické výrazy, je $erase(XYZ, ABC)$ též geografický výraz. Jeho sémantickou hodnotou je mapová vrstva vzniklá operací vymazání mapové vrstvy dané výrazem XYZ z mapové vrstvy dané výrazem ABC. Samozřejmě i $erase(ABC, XYZ)$ je geografický výraz a vzhledem k nekomutativnosti operace vymazání je jeho sémantická hodnota odlišná od sémantické hodnoty výrazu $erase(XYZ, ABC)$.

6. Jsou-li XYZ a ABC dva geografické výrazy, je $update(XYZ, ABC)$ též geografický výraz. Jeho sémantickou hodnotou je mapová vrstva vzniklá operací přepsání mapové vrstvy odpovídající výrazu ABC mapovou vrstvou odpovídající výrazu XYZ. Samozřejmě i $update(ABC, XYZ)$ je geografický výraz a vzhledem k nekomutativnosti operace přepsání je jeho sémantická hodnota odlišná od sémantické hodnoty výrazu $update(XYZ, ABC)$.

7. Jsou-li XYZ a ABC dva geografické výrazy, je $clip(XYZ, ABC)$ též geografický výraz. Jeho sémantickou hodnotou je mapová vrstva vzniklá operací ořezání mapové vrstvy odpovídající výrazu ABC mapovou vrstvou odpovídající výrazu XYZ. Samozřejmě i

clip(ABC,XYZ) je geografický výraz a vzhledem k nekomutativnosti operace přepsání je jeho sémantické hodnoty odlišná od sémantické hodnoty výrazu *clip(XYZ,ABC)*.

8. Je-li *XYZ* geografický výraz, jsou:

*select * from XYZ where <logická podmínka>*

select <seznam atributů> from XYZ

select <seznam atributů> from XYZ where <logická podmínka>

též geografické výrazy. Jejich sémantické hodnoty vzniknou z mapové vrstvy odpovídající výrazu *XYZ* databázovými operacemi filtrace, faktorizace, popřípadě oběma operacemi najednou. Zde je použití výrazů obdobné jako v běžně používaném databázovém dotazovacím jazyce SQL.

V seznamu atributů i při tvorbě logických výrazů je možné používat standardní atributy přiřazené plošným mapovým vrstvám, tedy především:

- *INSIDE*: plocha leží uvnitř vrstvy. Typicky se použije při závěrečné faktorizaci mapové vrstvy, pomocí které se spojí jednotlivé rozdrobené a sousedící plošky vzniklé geografickou analýzou.
- *AREA*: Plocha (výměra) dané plochy plošné mapové vrstvy.
- *PERIMETR*: Poloměr dané plochy. Tedy nejmenší vzdálenost vnitřního bodu plochy od hranic plochy pro ten bod, kdy je tato vzdálenost maximální.
- *CIRCUMFERENCE*: Obvod plochy. Délka křivky (obvykle aproximované lomenou čarou), tvořící hranici plochy.

Jazyk GSQL

Závěrem zbývá definovat příkazy jazyka GSQL. Pro vyhodnocování operací geografické analýzy lze vystačit zatím s jediným příkazem *select*, jehož syntaxe i sémantika je podobná jako u bodu 9) definice geografického výrazu, tedy:

```
select * from XYZ where <logická podmínka>;
select <seznam atributů> from XYZ;
select <seznam atributů> from XYZ where <logická podmínka>;
```

popřípadě i

```
select * from XYZ;
```

Každý příkaz jazyka GSQL je zakončen středníkem. V posledním případě se s mapovou vrstvou odpovídající výrazu *XYZ* nedělají na závěr žádné databázové operace.

Příklad

Použití dotazovacího jazyka GSQL lze osvětlit na klasickém školním příkladě. Hledáme místo pro stavbu rekreačního střediska, které má dodržovat následující prostorové podmínky:

- Má být alespoň 1km od nejbližší silnice 1. třídy.
- Má být alespoň 1km a maximálně 3km od nejbližší vesnice.
- Má být maximálně 0.5 km od nejbližší vodní plochy.
- Plocha splňující tyto vlastnosti má mít výměru alespoň 100 ha.

K dispozici máme mapové vrstvy SILNICE (liniová, jedním atributem je TRIDA silnice), VODNI_PLOCHY (plošná) a SIDLA (plošná).

Klasický postup geografické analýzy bude následující:

1. Vytvoříme operací filtrace vrstvu SILNICE1 obsahující pouze silnice 1. třídy.
2. Vytvoříme vrstvu N1 jako obalovou zónu vrstvy SILNICE1 o poloměru 1km.
3. Vytvoříme vrstvu SIDLA1 jako obalovou zónu vrstvy SIDLA o poloměru 1km.
4. Vytvoříme vrstvu SIDLA3 jako obalovou zónu vrstvy SIDLA o poloměru 3km.
5. Vytvoříme vrstvu A3 jako obalovou zónu vrstvy VODNI_PLOCHY o poloměru 0.5km.
6. Vymažeme vrstvu SIDLA1 z vrstvy SIDLA3 a vznikne vrstva A2.
7. Vytvoříme průnik vrstev A2 a A3, vznikne vrstva A23.
8. Z vrstvy A23 vymažeme vrstvu N1, vznikne vrstva A123.
9. U vrstvy A123 provedeme faktorizaci podle atributu INSIDE a poté filtraci podle logické podmínky $AREA \geq 1$ ($1\text{km}^2 = 100\text{ha}$).

Pomocí jazyka GSQL lze uvedený postup zapsat následovně:

```
select INSIDE from erase(buffer(select * from SILNICE where TRIDA=1,1),
                               intersect(erase(buffer(SIDLA,1),buffer(SIDLA,3)),
                                           buffer(VODNI_PLOCHY,0.5)
                               )
                               )
where AREA >= 1;
```

Jak je vidět z příkladu, uvedený postup pomocí jednoho příkazu jazyka GSQL je syntakticky jednoznačnější a kratší, než postup zapsaný pomocí algoritmu (posloupnosti operací). Navíc odpadá nutnost pojmenovávání velkého počtu mapových vrstev vzniklých jako mezivýsledky geografické analýzy.

Možnosti použití a výhledy do budoucna

Kde je místo pro použití dotazovacího jazyka GSQL. Klasickým využitím by mohly být geograficko analytické stroje, jakási obdoba databázových strojů, fungující na architektuře zákazník - sluha (client - server). V dnešní době jsou moduly pro geografickou analýzu sice součástí každého slušného GISu.

Ovládání geografické analýzy je však řešeno poměrně nekonceptně a to buď interaktivním způsobem, tedy zadáváním jednotlivých operací v dialogových panelech, nebo přímo psaním příkazů pro geografickou analýzu v příkazovém řádku, či v dávkovém příkazovém souboru.

Při použití dotazovacího jazyka by postup při geografické analýze vypadal následovně. Na straně zákazníka (klienta) by existoval uživatelsky příjemný program, pomocí kterého by (pravděpodobně pomocí dialogových panelů) bylo možné vytvořit příslušný geografický výraz. Pro zkušenější programátory analytiky by zůstala zachována i možnost zapsat geografický výraz přímo v textovém režimu. Na straně služebníka (serveru) by pak byl geografický výraz zpracován geograficky analytickým strojem a zákazníkovi by byla zpět předána jeho hodnota, tedy plošná mapová vrstva v domluveném formátu. Geografický analytický stroj by přitom mohl pro potřeby analýzy umět zpracovat mapové vrstvy uložené v různých formátech, popřípadě i digitalizované v různých souřadných systémech.

Jednoduchým příkladem programů, které se blíží zde definovanému pojmu geografický analytický stroj může být modul OVERLAY programů ARCINFO a ARCGIS, nebo v ještě jednodušší a průzračnější podobě program MAPOS [plavec] vytvořený studenty a pracovníky katedry inženýrské informatiky FSV ČVUT.

Definici jazyka GSQL by bylo možné do budoucna rozšířit i o obdobu dalších příkazů databázového dotazovacího jazyka SQL, tedy o příkazy sloužící k tvorbě a údržbě databází. Speciálně lze uvažovat o definici obdoby příkazů CREATE TABLE (založení nové mapové vrstvy), UPDATE (aktualizace dat, ovšem zřejmě jen popisných) a DELETE (vymazání objektů z mapové vrstvy).

Zajímavým počinem by též bylo vytvoření jednoduchého překladače či interpretu z jazyka GSQL do příkazového jazyka nějakého prostředku pro geografickou analýzu (OVERLAY, MAPOS). Tím by byla dokončena myšlenka realizace geografického analytického stroje. Též tvorba uživatelsky přátelského klientského programu může být zajímavou programátorskou myšlenkou.

Literatura

[PGPSP]

Mayer, P., Vaníček, T.: Počítačová grafika pro stavební praxi. Skripta ČVUT Praha, 1995.

[plavec]

Plavec, J.: Množinové operace v dvourozměrné geometrii a jejich využití v geografických informačních systémech}, Diplomová práce, Stavební fakulta ČVUT, 2000.

Adresa autora:

RNDr. Tomáš Vaníček, PhD., katedra inženýrské informatiky, Stavební fakulta ČVUT

Thákurova 7, Praha 6

email: vanicek@fsv.cvut.cz