

Considerații privind utilizarea bazelor de date spațiale în domeniul economic

Conf.dr. Constanța BODEA, prof.dr. Ion LUNGU,
Catedra de Informatică Economică, A.S.E., București

Lucrarea tratează problematica datelor spațiale, în contextul organizării acestora drept colecții într-o bază de date. Sunt definite conceptele de date spațiale și baze de date spațiale. Sunt prezentate principalele tipuri de cereri care sunt formulate uzual asupra unei baze de date spațiale. Lucrarea identifică domeniile de aplicare a bazelor de date spațiale în domeniul economic și în cadrul sistemelor informatice geografice.

Cuvinte cheie: date spațiale, baze de date spațiale.

Introducere

Obiectele spațiale, constituite din puncte, linii, suprafețe, volume și obiecte de dimensiuni superioare, sunt frecvent utilizate în cadrul unor aplicații (de proiectare asistată, cartografie, sisteme informatice geografice). De cele mai multe ori, obiectele spațiale au atașată și dimensiunea temporală, devenind obiecte spațio-temporale.

Date și baze de date spațiale

Datele spațiale definesc diferite obiecte spațiale. Obiectele spațiale prezintă însă nu numai atribute spațiale - lungime, configurație, perimetru, forma suprafeței, tip volum etc. - ci și atribute nespațiale.

Datele spațiale pot fi definite pe domenii de valori discrete sau continue (de exemplu, segmente de dreaptă, regiuni etc.).

Bazele de date spațiale (BDSP) reprezintă un ansamblu de colecții de date spațiale și nespațiale, organizate în scopul facilitării gestiunii și utilizării multi-user a acestora.

Cereri dintr-o bază de date spațiale

Dintr-o BDSP se pot formula mai multe tipuri de cereri de date și anume:

- **cereri generale** (cum ar fi: afișarea datelor, regăsirea unui pattern, anticiparea unui comportament de obiect spațial), care nu necesită facilități speciale de tratare a datelor spațiale.

- **cereri spațiale**, care pot fi locale, zonale și care implică vecinătatea. Acestea din urmă sunt cel mai greu de evaluat întrucât presupun căutarea și determinarea proximității.

Interacțiunea cu o BDSP poate îmbrăca forma:

- comenzilor, atunci când utilizatorul știe precis ceea ce dorește să afle din cadrul BDSP;

- probelor, atunci când utilizatorul realizează o explorare a datelor. În acest din urmă caz sunt utilizate vizualizări și tehnici de punctare.

Interfața grafică este extrem de importantă pentru o BDSP, întrucât atât intrările cât și ieșirile trebuie să fie grafice.

Pentru accesarea BDSP s-a propus o varietate de soluții, cum ar fi, de exemplu utilizarea SQL sau QBE.

Utilizarea SQL pentru lucrul cu o BDSP presupune extinderea limbajului cu o serie de operatori spațiali sau utilizarea unor funcții și operatori definiți de utilizator și cunoscuți de sistemul de

gestiune a bazei de date spațiale doar la momentul execuției.

Să presupunem, de exemplu, că lucrăm cu SQL extins cu operatori spațiali. Declaraarea relațiilor (tabelelor) de date presupune declararea atributelor spațiale pe domenii predefinite de valori spațiale (de exemplu, LINE-SEGMENT, ARIE etc.).

Un predicat SQL este spațial dacă prezintă o condiție care implică cel puțin un atribut sau o operație spațială.

Condițiile spațiale impun aplicarea unor funcții pe atributele spațiale, precum `aria()`, `perimetrul()`, `lungimea()`, `inmerimetrul()`, `învecinătatea()` etc.

Să considerăm următoarele relații:

```
create table drumuri
  (identificator NUMBER,
   nume CHAR(30),
   tip CHAR(30),
   coordonate LINE-SEGMENT);
create table regiuni
  (:
   :
   :
   localizare REGION);
```

Să presupunem că dorim să găsim numele acelor drumuri care trec prin regiunea R1.

Cererea SQL este:

```
select nume
from drumuri, regiuni
where numereg="R1"
and intersect (localizare, coordonate);
```

Când intersecția implică mai mult de un atribut spațial (cum este cazul în cadrul exemplului de mai sus), operația ce trebuie realizată nu este o selecție spațială, ci un join spațial.

Structuri de date pentru BDSp

O modalitate uzuală de organizare a datelor spațiale este parametrizarea acestora, pentru a le reduce la un punct într-un spațiu de dimensiune superioară. Acest lucru se poate realiza relativ ușor pentru

SGBD-urile convenționale, prin adăugarea unuia sau mai multor câmpuri la tabelă.

O asemenea abordare este însă posibilă numai în cazul realizării unor regăsiri simple de date. Dacă cererile de date implică aspecte privind spațiul ocupat de obiecte (deci se presupune tratarea concomitentă a mai multor record-uri, în raport de noțiunea de proximitate), atunci se impune regăsirea record-urilor pe baza unor proprietăți spațiale ce nu sunt memorate explicit în baza de date. De exemplu, într-o bază de date cu drumurile aflate în evidența Administrației Naționale a Drumurilor nu dorim să obligăm utilizatorul să specifice explicit drumurile care se intersectează unele cu altele, sau traversează anumite regiuni. Volumul acestor informații, dacă s-ar memora explicit ar putea fi extrem de mare, iar costul preprocesării de asemenea. Din acest motiv, este preferabilă memorarea implicită a acestor informații. În acest fel poate fi tratată o gamă mai largă de cereri spațiale, fără cunoașterea anticipată a acestora.

O abordare uzuală în reprezentarea datelor spațiale este separarea acestora de cele nespațiale, întreținând un set de legături explicite între acestea. Operațiile spațiale sunt executate direct pe structurile de date spațiale, ceea ce oferă libertatea alegerii celei mai potrivite structuri spațiale. În plus, se poate utiliza un procesor spațial, special proiectat pentru a trata în mod eficient partea din cererile de date ce implică relații de proximitate (căutare spațială), în timp ce un SGBD convențional poate fi utilizat pentru tratarea părții ce implică date nespațiale.

Pentru a face posibilă tratarea cererilor bazate pe vecinătatea în spațiu a obiectelor, datele spațiale trebuie sortate/ordo-

nate. Problema deosebit de importantă în acest caz este stabilirea cheilor de sortare/ordonare.

În cazul datelor spațiale, ordonarea trebuie să se bazeze pe toate cheile spațiale, adică pe toate proprietățile spațiale ale obiectelor. Asemenea tehnici de ordonare sunt cunoscute sub numele de **indexare spațială**.

Într-o bază de date, un record poate fi interpretat drept un punct dintr-un spațiu multidimensional. Acest mod de conceptualizare a record-ului poate fi utilizat pentru a lucra cu datele spațiale, prin maparea obiectului spațial într-un punct reprezentat într-un spațiu de aceeași dimensiune sau de altă dimensiune. Problema care apare este că dimensiunea punctului poate fi prea mare. O soluție este de a aproxima obiectul spațial, pentru a realiza scăderea dimensionalității.

O altă problemă foarte importantă este păstrarea proprietăților de proximitate în cazul schimbării dimensiunii spațiului în care se face reprezentarea obiectului.

Să considerăm, de exemplu, problema reprezentării segmentelor de dreaptă în cadrul unei baze de date. Vom denumi hartă poligonală ansamblul segmentelor, indiferent dacă acestea sunt conectate sau nu. O asemenea hartă poate avea semnificația unei rețele de drumuri, de linii de înaltă tensiune etc. Utilizând punctele pentru reprezentarea segmentelor (așa numitele puncte de reprezentare), fiecare segment va fi reprezentat prin extremități. Acest lucru înseamnă că fiecare segment va fi reprezentat cu ajutorul unui tuplu format din patru elemente și anume:

- o pereche de elemente corespunzând coordonatelor X;
- o pereche de valori corespunzând coordonatelor Y.

S-a realizat astfel o mapare dintr-un spațiu bidimensional într-un spațiu de dimensiune patru, care conține punctele de reprezentare a segmentelor. Această mapare este utilă atunci când se dorește soluționarea unor cereri de date, precum: găsirea tuturor segmentelor care intersectează un anumit punct, un set de puncte sau un alt segment. Se observă că toate aceste cereri implică punctele ce aparțin segmentelor.

Maparea nu este însă potrivită pentru soluționarea cererilor care fac referire la punctele ce nu sunt parte a segmentelor, pentru că acestea nu sunt transpuse în spațiul de reprezentare (de dimensiune patru). Să presupunem că dorim să determinăm dacă două segmente sunt în apropiere unul de altul sau să găsim cel mai apropiat segment de un punct sau o dreaptă dată. Satisfacere unor asemenea cereri este dificilă, întrucât proximitatea din spațiul bidimensional în care sunt figurate segmentele nu este în mod necesar conservată în spațiul de dimensiune patru în care sunt reprezentate segmentele. Altfel spus, deși două segmente de dreaptă sunt apropiate unul de altul, distanța euclidiană între punctele lor de reprezentare poate fi foarte mare. Din acest motiv este nevoie de o altă reprezentare a segmentelor.

O posibilitate este de a lucra cu structuri bazate pe spațiul ocupat de diferite obiecte. Reprezentările bazate pe poziționarea spațială descompun spațiul în care sunt figurate datele în regiune, numite *buckete*. În general, este descompus spațiul pe care se realizează maparea. Considerăm însă, că ceea ce trebuie descompus este spațiul în care există obiectele ce trebuie reprezentate (pentru exemplul dat, descompunerea spațiului bidimensional).

Există mai multe metode de descompunere a spațiului. Cea mai utilizată are la bază noțiunea de închidere minimală. În acest caz, obiectele sunt grupate, după proximitate într-o ierarhie și sunt apoi memorate în diferite structuri ierarhice. Cele mai cunoscute structuri ierarhice folosite sunt arborii-R.

Arborii-R și variantele lor sunt proiectați pentru organizarea unei colecții de obiecte spațiale prin reprezentarea acestora prin forme rectangulare. Fiecare nod din arbore corespunde formei rectangulare minimale ce înglobează nodurile fiu. Nodurile terminale conțin în locul fiilor pointeri spre obiecte actuale din BDSp.

În BDSp, obiectele sunt reprezentate prin forma rectangulară minimală care le poate îngloba. Formele rectangulare se pot suprapune, astfel încât, un obiect poate fi spațial conținut în diferite noduri. Acest lucru înseamnă că o cerere spațială poate reclama vizitarea mai multor noduri. Regulile de bază pentru construirea arborilor-R sunt foarte apropiate de cele ale arborilor-B. Toate nodurile terminale apar pe același nivel. Fiecare intrare într-un nod terminal este un tuplu de forma (R,O) , unde R desemnează forma rectangulară minimală ce conține obiectul O. Intrarea într-un nod neterminal este un tuplu (R,P) , unde R reprezintă forma rectangulară minimală care, spațial conține formele rectangulare din nodul copil desemnat prin P. Un arbore-R de ordin (m,M) este structura în care fiecare nod, cu excepția rădăcinii, are între m și M intrări, unde $m \in [M/2]$.

Utilizarea BDSp în cadrul sistemelor informatice geografice (SIG)

Utilizarea sistemelor SIG devine în prezent o necesitate. Având în vedere

cerințele informaționale, SIG-urile au drept obiective:

- să asigure gestionarea unei mari cantități de date spațiale;
- să permită realizarea unor analize axate pe caracteristicile geografice ale datelor;
- să permită accesul facil al utilizatorilor la datele spațiale.

Realizarea acestor obiective impune ca SIG să prezinte în **arhitectură** în mod obligatoriu următoarele module:

- modulul operațiilor de bază, responsabil cu asigurarea introducerii datelor, producerii hărților simple, memorării și gestionării datelor geografice;
- module AM/FM (Automated Mapping/Facilities Management) pentru asigurarea regăsirii informațiilor descriptive referitoare la diferite imagini grafice și pentru managementul acestor asocieri;
- module pentru analize spațiale, statistici și modelare spațială, pentru prelucrarea imaginilor raster și a celor vectoriale.

Bibliografie

1. Gambosi G, Scholl M, six H-W (eds) - *Geographic DBMS*, Springer-Verlag, 1992.
2. Ivan I, Bodea C, Vasilescu A. - *Structuri de date pentru baze de date spațiale*, Sesiunea științifică: România - accelerarea tranziției, I.N.I. București, martie, 1996
3. Scholten H, Stillwell J (eds) - *Geographical Information Systems for Urban and REgional Planning*, Kluwer Academic Publishers, 1990