

OBRAZLOŽENJE TEME DOKTORSKE DISERTACIJE

1. PRIJEDLOG NASLOVA DISERTACIJE

„Novo generalizirano pretraživačko drvo za prostorne baze podataka“

(eng: „New generalized search tree for spatial databases“)

2. TIP ISTRAŽIVANJA

Fokus ove doktorske disertacije je na razvoju generaliziranog pretraživačkog drveta koje će obezbjediti superiorne indeksne šeme unutar prostorne baza podataka. Unutar disertacije planirano je istraživanje razvoja i implementacije pretraživačkih drveta koja se odnose na multidimenzionalne prostorne pristupne metode. Novo drvo će se temeljiti na principima generaliziranog pretraživačkog drveta GiST, predstavljenog 1995 godine od strane grupe profesora sa kaliforniskog univerziteta Berkeley, Joseph M. Hellerstein, Jeffrey F. Naughton i Avi Pfeffer [26]. Osim teorijskog razmatranja u okviru disertacije će se koristiti modeliranje, simulacija te implementacija, kako bi se analizirale performanse novokreiranog drveta.

3. STANJE U OBLASTI KOJOJ TEMA PRIPADA

Veliki broj savremenih sistema, kao što su GIS (Geografic Information Systems), CAD/CAM (Computer Aided Design/ Computer Aided Modeling) VLSI (Very Large Scale Integration), Robotika i dr., zahtjevaju manipulaciju multidimenzionalnim geometrijskim objektima, tj. prostornim podacima[4],[12],[34].

Navedeni sistemi moraju upravljati velikim količinama podatka, reda terabajta, petabajta pa čak i većeg. Nadalje, moraju podržavati prostorne upite nad djelovima baze koji će pretražiti,

dobaviti te geometrijskom analitikom dati rezultate na postavljene upite (Preklapanja, Najbliži susjed, Pripadnosti objekta, Međusobna povezanost objekata i dr.). S tim u vezi potrebna im je prostorne baza podataka koja će omogućiti efikasno procesiranje postavljenih upita[8],[12],[29].

Kako bi se vršilo efikasno procesiranje postavljenih upita nad velikom količinom prostornih podataka, koja se nalazi na sekundarnoj memoriji, sistem baze podataka zahtjeva pristupni metod (Access Metod - AM) koji će minimizirati broj jako sporih I/O operacija i dobiti podatke o objektima prema njihovoj prostornoj dispoziciji [3],[11],[14],[32].

Prvi pristupni metodi su bili jednodimenzionalni (B-drvo) i bili su osnovni načini za indeksiranje perifernih uređaja. Osim izvorne primjene, isti su naišli na visoku primjenu u klasičnim aplikacijama baziranim na alfa-numeričkim podacima.

Iako su implementacije struktura sa tradicionalnim pristupnim metodima veoma raštrkane i općeprihvaćene, iste nisu u stanju na adekvatan način indeksirati objekte koji se nalaze u multidimenzionalnom prostoru. Za indeksiranje prostornih objekata, na osnovu njihove dispozije, strukture koje koriste upoređivanje egzaktnih vrijednosti (Hash tabela i sl.) nisu od pomoći jer imaju vrlo velik opseg pretrage. Nadalje, strukture koje koriste jednodimenzionalno redanje na osnovu ključnih vrijednosti (B-drvo, ISAM i sl.) također nisu upotrebljive jer im pretraživački prostor nije multidimenzionalan [1],[8],[12],[13].

Tek u osamdesetim godinama prošlog stoljeća javlja se pojam multidimenzionalnog prostornog pristupnog metoda. Prostorni pristupni metodi su omogućili rad nad multimedijalnim, grafičkim, i vizualnim informacijama, što je stvorilo osnovu za razvoj prostorno-orijentisanih informacionih sistema. Prostorne baze podataka se razlikuju od relacionih baza podataka u tome što omogućavaju snažniji i korisničko-proširivi sistem tipova podataka, koji se može koristiti u proizvoljnim objektno-orijentisanim hijerarhijskim konceptima. S druge strane razlikuju se od objektno-orijentisanih baza podataka jer baštine mogućnosti deklarativnog pristupa, višekorisničkih operacija, transakcione izolacije i povrata na prethodno stanje, mogućnosti koje su relacione baze podataka jako dobro usavršile [9],[12].

Kako bi podržali prostorno-orijentisane informacionih sistema, prostorni pristupni metodi moraju omogućiti prostorno indeksiranje te tehnike klasterizacije. Bez prostornog indeksa, svaki objekat unutar baze podataka bi se morao ispitati kako bi se ustanovilo da li isti zadovoljava kriterije postavljenog prostornog upita. Kako je set prostornih podataka veoma velik i nalazi se na sekundarnoj memoriji (u stranicama na disku), pristupni metod bez prostornog indeksa jednostavno nije primjenjiv. S tim u vezi prostorni indeks koji omogućava efikasan pronalazak adrese traženog objekta na sekundarnoj memoriji predstavlja nezaobilazan dio prostorne baze podataka[4],[10],[14].

Tehnike klasterizacije se koriste kako bi se izvršilo grupisanje objekata koji se obično pretražuju skupa. Ova tehnika vrši skladištenje objekata, koji su u stvarnom životu jedni blizu drugih, unutar istih ili susjednih stranica na disku. Na taj način klasterizacija u značajnoj mjeri doprinosi efikasnom pristupnom metodu, jer dobavljanje međusobno bliskih stranica sa diska rezultuje bržim I/O operacijama i većim brojem pogodaka u traženim stranicama RAM-a [4],[10],[14].

Na samim počecima, broj multidimenzionalnih pristupnih metoda je bio neznatan, dok je podrška za nove pristupne metode unutar prostornih baza podataka je bila jako loša, (gotovo da je i nije bilo). Veliki broj istraživačkih zajednica su bile dosta nezainteresovane za navedeni problem i umjesto da su se zalagale za proširivost baza podataka sa korisnički-definiranim pristupnim metodama, (kao što su radile Exodus, Postgres i Starburst), iste su se fokusirale na razvoj novih pretraživačkih drveća u domeni-svojstvenim izolovanim informacionim sistemima[1]. Naime, prema [8] u to vrijeme je postojalo preko 50 alternativnih indeksnih struktura i neka od njih su vrhunska u svojoj domeni, međutim cijena njihove implementacije, krenuvši od nultog stanja je bila preskupa. S tim u vezi brojne indeksne strukture nisu doživjele implementaciju unutar sistema, o čemu dovoljno govori činjenica da se samo par prostornih pretraživačkih drveća našlo u široj primjeni indeksiranja multidimenzionalnih prostora (Oracle – R i Quadtree drvo; IBM- Spatial Grid). Osnovni razlog zbog nedostupnosti većeg broja pristupnih metoda je taj što jezgra ovih DBMS-ova tada nisu bila dovoljno fleksibilna i proširiva za nove pristupne metode [1],[12],[13].

Tek od prošle decenije baze podataka kao što su PostgreSQL, Oracle 8 i IBM DB2, su počele da obezbjeđuju korisnički-proširiv sistem tipova podataka i indeksnih šema pomoću arhitektura unutar koje se mogu povezati sa ekstenzijama baza podataka iz specijalističkih

domena napravljenih od trećih strana. Unutar ovog pristupa server baze podataka obezbeđuje standardne DBMS funkcionalnosti, kao što su upravljanje skladištenjem te optimizacija i izvršavanje upita neovisno od sistema tipova podataka. Domen-specifične funkcije pretrage i funkcije računalne geometrije, razvijene od treće strane, se nalaze u formi funkcionalnih biblioteka te se sa serverom baze podataka dinamički linkuju. Kada se izvršava upit nad korisnički-definisanim tipovima podataka, DBMS poziva funkcije koje su dostavljene u gore navedenim funkcionalnim bibliotekama.

Trenutno tehnički najsavremenije dostupna baza podataka Oracle 11g veoma uspješno obezbeđuje proširivost za nove pristupne metode pružajući SQL-bazirani okvir za kreiranje novih domensko-specifičnih indeksnih šema zajedno sa setom odgovarajućih operatora. Implementacija indeksne šeme je data kao set Oracle Data Cartridge Interface (ODCIIndex) rutina za definisanja indeksa, njegovo održavanje te operacije indeksne pretrage. Indeks kreiran pomoću nove indeksne šeme se ponaša kao i indeksi već inkorporirani u bazu podataka.

S druge strane PostgreSQL ide korak dalje i obezbeđuje dva generalizirana pretraživačka drveća koja daju još veći nivo apstakcije pri kreiranju specifičnih pristupnih metoda. Navedena arhitektura omogućava korisniku kreiranje vlastitog pristupnog metoda bez poznavanja veoma zahtjevnih implementacionih problema kao što su algoritmi za konkurentnost, povratak u prvobitno stanje (recovery) te transakciona izolacija.

Pomoću kreiranja generaliziranih pretraživačkih drveća, koje bi trebala svaka baza podataka da posjeduje, ekspertima iz domene bi se omogućio okvir za jednostavniju primjenu domeni-svojstvenih pristupnih metoda. Ova disertacije bi trebala da rezultira novim generaliziranim pretraživačkim drvetom koje će da doprinose u popularizaciji proširenja za nove pristupne metode unutar prostornih baza podataka.

4. OSNOVNI CILJEVI I PLAN ISTAŽIVANJA

Osnovni cilj istraživanja je dobiti balansirano pretraživačko drvo koje će omogućiti implementacije novih prostornih pristupnih metoda unutar istog. Navedeno drvo se mora moći inkorporirati unutar standardnih dostupnih DBMS-ova kako bi se istim omogućilo uvođenje prostorne dimenzije.

Da bi se postigao navedeni cilj potrebno je detaljno proučavanje savremenih aplikacionih sistema koji koriste multidimenzionalne objekte (GIS, CAD i dr.) u cilju analize zahtjeva koji se postavljaju nad prostornim bazama podataka. Osim navedenog planirano je i proučavanje dostupnih multidimenzionalnih pretraživačkih drveta te algoritama računalne geometrije (computational geometry) koje savremene prostornim baze podataka moraju podržavati.

Konačno, definisati će se i implementirati, novi set metoda i funkcija koje će od drveta napraviti solidan okvir za implementaciju brojnih prostornih pristupnih metoda potrebnih za primjenu u mnogim sferama računarski podržanog vođenja i upravljanja.

5. METODOLOGIJA ISTRAŽIVANJA

U okviru predložene doktorske disertacije trebao bi se izvršiti:

- ✓ Pregled i analiza modeliranja i reprezentacije prostornih objekata;
- ✓ Pregled i analiza algoritama računalne geometrije;
- ✓ Pregled i analiza nivoa apstracije upitnog jezika (query language);
- ✓ Detaljno proučavanje prostornih pristupnih metoda;
- ✓ Pregled i analiza algoritama za konkurentnost, povratak u prvobitno stanje (recovery) te transakcionu izolaciju;
- ✓ Definicija i implementacija novog seta metoda i funkcija koje će omogućiti okvir za implementaciju brojnih prostornih pristupnih metoda unutar novog generaliziranog pretraživačkog drveta;
- ✓ Izvući zaključke i dati smjernica za dalje istraživanje;

Metodologija istraživanja zasnivat će se na razvoju i praktičnoj primjeni novog generaliziranog pretraživačkog drveta. Za izradu drveta koristit će se C, C++, VB programirski alati, zatim već razvijena okruženja i biblioteke za testiranje indeksnih struktura i pristupnih metoda kao što je “SpatialIndex Library“ (Marios Hadjieleftheriou) , te baze podataka Oracle, PostgreSQL i dr..

Nakon implementacije izvršit će se poređenje performansi novih prostornih pristupnih metoda kreiranih unutar drveta u odnosu na iste implementirane kao samostalne indeksne strukture, te

analiza kompleksnosti implementacije novih prostornih pristupnih metoda unutar drveta u odnosu na implementaciju istih krenuvši od nultog stanja, kako bi se na osnovu toga mogla napraviti procjena rada na razvoju generaliziranog pretraživačkog drveta.

6. OČEKIVANI IZVORNI NAUČNI DOPRINOS DISERTACIJE

Osnovni doprinos ove disertacije je razvoj generaliziranog pretraživačkog drveta koje će omogućiti implementacije novih prostornih pristupnih metoda unutar istog. Isti će obezbjediti implementaciju prostorne baze podataka koja koristi ovo pretraživačko drvo u brojnim specijalističkim domenama.

Osim navedenog disertacija će dati odgovore na veoma značajna pitanja, a neka od njih su:

- ✓ U kojoj mjeri je pojednostavljena implementacija novih prostornih pristupnih metoda unutar drveta u odnosu na implementaciju istih krenuvši od nultog stanja;
- ✓ Da li su prostorni pristupni metodi unutar ovog drveta sa performansama lošiji u odnosu na implementaciju istih kada se krene od nultog stanja, i koliko;
- ✓ U kom bi se pravcu trebala odvijati sljedeća istraživanja s aspekta pretraživačkih struktura unutar prostorne baza podataka;

7. POLAZNA LITERATURA

- [1] **PostgreSQL 9.1.1 Documentation**, The PostgreSQL Global Development Group, 2011;
- [2] Manfred M. Fischer • Jinfeng Wang, **Spatial Data Analysis Models, Methods and Techniques**, 2011 Springer;
- [3] Jinbao Wang, Sai Wu, Hong Gao, Jianzhong Li, Beng Chin Ooi, **Indexing Multi-dimensional Data in a Cloud System**, *SIGMOD'10*, June 6–11, 2010, Indianapolis, Indiana, USA. Copyright 2010 ACM;
- [4] Luc Anselin Sergio J. Rey, **Perspectives on Spatial Data Analysis**, 2010 Springer;

- [5] Sisi Zlatanova and Arta Dilo, **A data model for operational and situational information in emergency response: the Dutch case**, In: Proceedings of the Gi4DM Conference - Geomatics for Disaster Management, February 2010, Torino, 4 p.
- [6] Anda VELICANU, Stefan OLARU, **Optimizing Spatial Databases**, Informatica Economica vol. 14, no. 2/2010;
- [7] A. Abouzeid, K. Bajda-Pawlikowski, D. J. Abadi, A. Rasin, and A. Silberschatz. Hadoopdb: **An architectural hybrid of mapreduce and dbms technologies for analytical workloads**. *PVLDB*, 2(1):922–933, 2009;
- [8] Krassimir Markov, Krassimira Ivanova, Ilia Mitov, Stefan Karastanev, **Advance of the access methods**, International Journal "Information Technologies and Knowledge" Vol.2 / 2008;
- [9] Sisi Zlatanova, **Freeform Data Types in Spatial Database**, In: GIS Development, Volume 12, 4, 2008 pp. 50-54;
- [10] Dilo, A. and Zlatanova, **Spatiotemporal data modeling for disaster management in the Netherlands**, in: Information Systems for Crisis Response and Management, Harbin, 2008, China, pp. 517–528;
- [11] Chuck Murray and others, **Oracle Spatial, Oracle Spatial Developer's Guide, 11g Release 1** (11.1), 2007;
- [12] ALBERT K.W. YEUNG and G. Brent Hall, **Spatial Database Systems: Design, Implementation and Project Management**, 2007 Springer;
- [13] Ravi Kothuri, Albert Godfrind, Euro Beinat, **Pro Oracle Spatial for Oracle Database 11g**, 2007, Apress;
- [14] Mei Li, Wang-Chien Lee and Anand Sivasubramaniam, **DPTree: A Balanced Tree Based Indexing Framework for Peer-to-Peer Systems**, 2006 IEEE;
- [15] Michael Zeiler, **Modeling Our World: The ESRI Guide to Geodatabase Design, and Designing Geodatabases: Case Studies in GIS Data Modeling**, 2005, PMA, The Independent Book Publishers Association;
- [16] Peter van Oosterom, **Spatial Access Methods**, In: P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind (Eds.); **Geographical Information Systems: Principles, Techniques, Management and Applications** (second, abridged edition), John Wiley, 2005, pp. 175-190;

- [17] Yannis Manolopoulos, Apostolos N. Papadopoulos and Michael Gr. Vassilakopoulos, **Spatial databases : technologies, techniques and trends** , 2005, Idea Group Inc.;
- [18] Robert Laurini , Sylvie Servigne, Guillaume Noël, **Soft real-time GIS for disaster monitoring** , International Symposium on Geo-information for Disaster Management 2005.;
- [19] Sylvie Servigne, Guillaume Noël, **Real-time and spatiotemporal data indexing structure**, Proceedings of the 7th AGILE Conference, Heraklion, 2004, 261-268
- [20] Sisi Zlatanova, Daniel Holweg and Volker Coors, **Geometrical and Topological Models for Real-time GIS**, In: Elfriede M. Fendel and Massimo Rumor (Eds.); Proceedings of the 24th Urban Data Management Symposium, Chioggia, October 2004, pp. 3.33-3.46;
- [21] Malinowski, E. and Zimányi, E. (2004), **Representing Spatiality in a Conceptual Multidimensional Model**, *Proceedings*, 12th ACM International Symposium on Advances in Geographical Information, Washington, DC, (2004);
- [22] Gewin, Virginia (2004), **Mapping opportunities**, *Nature*, 427, 376-377.
- [23] H.J.G.L. Aalders, **History in Spatial Databases**, 7th South East Asian Survey Congress, Hongkong, November 2003, 8 p.;
- [24] Shashi Shekhar and Sanjay Chawla, **Spatial Databases: A Tour**, Prentice Hall, 2003
- [25] Philippe Rigaux, Michel Scholl and Agnes Voisard, **Spatial Databases - With Application to GIS** . Morgan-Kaufman Publishers. 2002;
- [26] Joseph M. Hellerstein, Jeffrey F. Naughton and Avi Pfeffer, **Generalized Search Trees for Database Systems**, Proceedings of the 21st VLDB Conference Zurich, Switzerland, 1995;
- [27] Hanan Samet, **The Design and Analysis of Spatial Data Structures**, Addison-Westey Publication Company, Inc, 1994;
- [28] Brinkhoff T., Kriegel H.-P., Schneider R. **Scene Organization: A Technique for Global Clustering in Spatial Database Systems**, Proceedings of the 20th VLDB Conference Santiago, Chile, 1994;
- [29] Bernhard Seeger, Hans-Peter Kriegel, **Efficient Spatial Query Processing in Geographic Database Systems**, IEEE DATA ENGINEERING BULLETIN 1993

- [30] Hans-Peter Kriegel, Thomas Brinkhoff, Ralf Schneider, **The Combination of Spatial Access Methods and Computational Geometry in Geographic Database Systems**, Lecture Notes in Computer Science, Val. 525, Springer, 1991, pp. 5-22
- [31] Schneider R., Kriegel H.-P. **The TR*-tree: A New Representation of Polygonal Objects Supporting Spatial Queries and Operation**, Lecture Notes in Computer Science, Vol. 553, Springer, 1991, pp. 249-264.;
- [32] Norbert Beckmann, Hans-Peter Kriegel, Ralf Schneider, Bernhard Seeger: **The R*-Tree: An Efficient and Robust Access Method for Points and Rectangles**. Proc. ACM SIGMOD Int. Conf. on Management of Data, Atlantic City, N.J., 1990, pp. 322-331;
- [33] Abel, D. J., and Mark, D. M., **A Comparative Analysis of Some Two-Dimensional Orderings**, International Journal of Geographical Information Systems, 1990, 4 (1), 21-31;
- [34] Bernhard Seeger, Hans-Peter Kriegel, **The Buddy-Tree: An Efficient and Robust Access Method for Spatial Data Base Systems**, Proceedings of the 16th VLDB Conference Brisbane, Australia 1990;
- [35] Peter van Oosterom, P., and van den Bos, J. **An Object-Oriented Approach to the Design of Geographic Information Systems**, Computers & Graphics, 1989, 13 (4), 409-418.;
- [36] Bernhard Seeger, Hans-Peter Kriegel, **Techniques for Design and Implementation of Efficient Spatial Access Methods**, Proceedings of the 14th VLDB Conference Los Angeles, California 1988;
- [37] T. Sellis, N. Roussopoulos, and C. Faloutsos, **The R+-Tree: A dynamic index for multi-dimensional objects**, In VLDB, 1987;
- [38] Bentley, J. L., **Multidimensional Binary Search Trees Used for Associative Searching**, Communications of the ACM, 1975, 18 (9), 509-517.;

Sarajevo 30.09.2011.godine

Kandidat

Mr. Zoran Cico dipl.el.ing.