

OBJECT DATABASE SPATIAL EXTENSION

Ondřej Polách

Master Degree Programme (2), FIT BUT

E-mail: xpolac04@stud.fit.vutbr.cz

Supervised by: Marek Rychlý

E-mail: rychly@fit.vutbr.cz

Abstract: The vast majority of today's spatial database systems is based on post-relational databases. However, this work is trying to find a connection between object and spatial databases. The obtained knowledge is reflected in the implementation of an object database spatial extension.

Keywords: Exact geometry computation, SQL/MM Spatial, Object database db4o, Coordinate reference systems.

1. ÚVOD

Cílem této práce je implementovat rozšíření prostorové databáze *db4o* [1], které umožní dalším vývojářům perzistentně pracovat s geometrickými daty pomocí jednotného aplikačního rozhraní.

Důraz je kladen hlavně na použitelnost, výpočetní robustnost, přesnost, spolehlivost a výkonnost. Tyto kritéria jsou tudíž ukazateli kvality rozšíření a to v zapsaném pořadí od nejvíce důležitých. Proč právě toto pořadí? Odpověď se skrývá v pochopení daných kritérií.

Použitelnost je nejnütnější kritériem, protože bez přívětivého rozhraní pro další vývojáře, by tito jen stěží toto rozšíření používali pro řešení svých problémů. Výpočetní robustnost je úzce spjata se zbývajícími třemi. Jsou-li geometrické algoritmy postaveny na robustním jádře, jsou přesné a spolehlivé, ovšem za cenu snížení výkonu a naopak. Vše záleží na požadavcích konkrétní aplikace, zdali je ochotná akceptovat snížení výkonu pro spolehlivé výsledky, či naopak potřebuje výkon za cenu nižší přesnosti. Kompromis mezi robustností a výkonností leží někde uprostřed. Pro toto rozšíření jsem se rozhodl jít cestou maximální robustnosti za použití exaktní výpočetní geometrie (*Exact Geometry Computation* - EGC), která je v posledních dvou desítkách let hlavním směrem řešení geometrické výpočetní robustnosti. Upřednostňuji tedy přesnost a spolehlivost na úkor výkonnosti. Avšak, výzkum v oblasti EGC stále pokračuje a proto věřím, že budou objeveny stále efektivnější metody pro exaktní geometrické výpočty. Již v současné době existuje několik metod, které s úspěchem zvyšují výkonnost EGC. Patří zde hlavně použití numerických filtrů. Čtenáře, který by se chtěl dozvědět více o EGC, bych rád odkázal na zdroj [2], který mu poskytne dobrý začátek.

Rád bych také podotknul, že s největší pravděpodobností budou geometrické algoritmy implementovány pouze pro 2 dimenze (2D). Kde to bude možné, budu se snažit, aby daná část byla jednoduše rozšiřitelná o další verze algoritmů. Není-li v následujícím textu řečeno jinak, jedná se o verzi pro 2D.

2. ARCHITEKTURA

Architektura rozšíření je postavena na přístupu tvorby prostorové databáze, který preferuje oddělení geometrické knihovny a samotné databáze, přičemž jsou obě entity spojeny integrační vrstvou. Tento přístup je popsán ve zdroji [3].

3. GEOMETRICKÁ KNIHOVNA

3.1. ROBUSTNÍ JÁDRO

Jádrem většiny geometrických algoritmů je výpočet znaménka determinantu. Tento algoritmus musí být implementován robustně. Je použita technika pozdního vyhodnocování s použitím dynamického numerického filtru založeného na intervalové aritmetice [4] následovně:

1. Determinant se vyhodnotí pomocí intervalové aritmetiky.
2. Nelze-li v prvním kroku rozhodnout o znaménku, vyhodnotí se determinant exaktně za použití aritmetiky nad čísly s neomezenou přesností.

Determinant bude vyhodnocen numerickou analytickou metodou LU dekompozice ve tvaru $PA=LU$. Pro intervalovou aritmetiku bude použita knihovna *JInterval* [5]. Pro exaktní výpočet bude použit číselný typ s neomezenou přesností *BigFraction* z knihovny *Apache Commons Math*.

3.2. GEOMETRICKÉ PREDIKÁTY A OPERACE

Základním predikátem je testování orientace bodu k úsečce. Tento predikát je počítán robustně, protože používá výpočet znaménka determinantu popsaného výše. Na tomto predikátu je založen jeden z nejdůležitějších predikátů tohoto rozšíření, a sice robustní test průniku dvou úseček.

Hlavní operací je výpočet průsečíku dvou úseček. Pro dosažení robustnosti je výpočet založen na homogenních souřadnicích. Dalšími algoritmy patřící do této kategorie jsou výpočet konvexní obálky, buffer, difference apod. U těchto algoritmů je důležitá topologická konzistence. Pro její dosažení budou výpočty probíhat na planárním grafu s použitím principu překreslování, který byl poprvé popsán ve zdroji [6].

3.3. VZTAHY V PROSTORU

Vztah v prostoru dokážeme vyjádřit pomocí 5 operací, kterými jsou *dotek* (touches), *uvnitř* (within), *přes* (crosses), *přesah* (overlaps) a *disjunkce* (disjoint). Pro určení vztahů je použit model nazvaný *Dimensionally Extended Nine-Intersection Model* (DE-9IM) [7]. Ten je založen na definici hranice, vnitřku a vnějšku geometrických objektů. Pro určení vztahu se počítá dimenze průniku u devíti kombinací z hranic, vnitřků a vnějšků. Určení vztahu je opět robustní.

3.4. PROSTOROVÝ DOTAZOVACÍ JAZYK

Prostorový dotazovací jazyk rozšiřuje dotazovací jazyk o prostorovou algebru. Musí tedy podporovat prostorové datové typy a musí být schopen vyjádřit prostorové operace. Objektová databáze *db4o* obsahuje 3 typy dotazovacích jazyků. A sice, nativní dotazy, dotazy pomocí příkladu a *Simple Object Database Access* (SODA). SODA tvoří základní dotazovací jazyk a jeví se tedy jako vhodný kandidát na rozšíření o prostorové dotazy. Je nutné zakomponovat prostorové indexování a prostorovou algebru. Indexování bude postaveno na struktuře z rodiny *R-Tree*.

3.5. APLIKAČNÍ ROZHRAŇÍ

Použitelnost rozšíření velmi úzce souvisí s jeho aplikačním rozhraním. To je postaveno na standardu *SQL/MM Spatial* [7]. V současné době existuje pouze jedno prostorové rozšíření databáze *db4o*. Jedná se o *Java Topology Suite*, která používá standard *Simple Features Specification for SQL*. Zdá se tedy, že prostorové rozšíření *db4o* založené na *SQL/MM Spatial* zatím neexistuje.

3.6. SOUŘADNICOVÉ SYSTÉMY

Standard *SQL/MM Spatial* přímo nespécifikuje koncept souřadnicových referenčních systémů (SRS). Pouze obsahuje koncept jejich zakomponování. Pro použití je doporučován standard ISO 19111, který popisuje koncept SRS. Prostorové rozšíření používá knihovnu *JScience* [8], která obsahuje základní implementaci zmiňovaného konceptu z ISO 19111. Rozšíření obsahuje generické

souřadnice pro 2D a 3D, spolu s generickými 2D a 3D SRS, které mají sloužit jako základní SRS pro kartézský prostor. Díky tomu, že je použito souřadnicové rozhraní z knihovny *JScience*, bude možno použít i jiné typy souřadnic a SRS z této knihovny.

4. INTEGRACE DO DATABÁZE

Tato kapitola krátce shrnuje koncept integrace knihovny do *db4o*. Integrovaná vrstva bude nejspíše postavena na nízko-úrovňovém rozhraní *db4o* nazvaném *TypeHandlers*, které dovoluje implementovat perzistentní strategii pro vlastní typy. Dále bude nutné určitým způsobem zaintegrovat prostorový dotazovací jazyk do SODA. Některé aspekty ohledně integrace jsou zatím otevřené.

5. ZÁVĚR

Čtenář byl seznámen se základními informacemi ohledně prostorového rozšíření *db4o*. Rád bych zde zmínil náměty na budoucí rozšíření této práce. Jedná se o grafický vstup a výstup, který by měl být součástí každé prostorové databáze. Dalším rozšířením může být implementace algoritmů pro vyšší dimenze apod. Projekt na vývoj prostorového rozšíření nese název *db4o-sp* a je hostován na portále *Sourceforge*, kde jsou poskytovány informace o průběhu vývoje pomocí vestavěného blogu.

PODĚKOVÁNÍ

Tato práce byla podpořena výzkumným záměrem č. MSM 0021630528 a projektem Centra excellence IT4Innovations (CZ.1.05/1.1.00/02.0070).

REFERENCE

- [1] VERSANT CORP. *Db4objects* [online]. [cit. 2012-02-20]. Dostupné z: <http://www.db4o.com/>
- [2] LI, Chen, PION, Sylvain a YAP, Chee-Keng. Recent Progress in Exact Geometric Computation. *J. Log. Algebr. Program.* 2005, s. 27. Dostupné z: <http://www.arnetminer.org/viewpub.do?pid=984592>
- [3] GÜTING, Ralf Hartmut. An Introduction to Spatial Database Systems. *Invited Contribution to a Special Issue on Spatial Database Systems of the VLDB Journal* [online]. 1994, s. 34 [cit. 2012-02-20]. Dostupné z: http://www.cise.ufl.edu/~mschneid/Research/thesis_papers/Gue94VLDBJ.pdf
- [4] BRONNIMANN, Hervé, BURNIKEL, Christoph a PION, Sylvain. Interval Arithmetic Yields Efficient Dynamic Filters for Computational Geometry. [online]. 2001, s. 21 [cit. 2012-02-20]. Dostupné z: http://hal.inria.fr/docs/00/34/42/81/PDF/interval_journal-1.pdf
- [5] NADEZHIN, Dmitry. *Java library for interval computations: JInterval* [online]. [cit. 2012-02-20]. Dostupné z: <http://java.net/projects/jinterval>
- [6] GREENE, Daniel Hill, YAO, Frances Foong. Finite-resolution computational geometry. *SFCS '86 Proceedings of the 27th Annual Symposium on Foundations of Computer Science.* 1986, [cit. 2012-02-20]. Dostupné z: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4568205
- [7] ISO/IEC JTC 1/SC 32/WG 4. *SQL Multimedia and Application Packages Part 3: Spatial.* 2003. Dostupné z: <http://jtc1sc32.org/doc/N1101-1150/32N1107-WD13249-3--spatial.pdf>
- [8] DAUTELLE, Jean-Marie. *JScience* [online]. [cit. 2012-02-20]. Dostupné z: <http://jscience.org/>