ИЗВЕСТИЯ НА БЪЛГАРСКОТО ГЕОГРАФСКО ДРУЖЕСТВО
JOURNAL OF THE BULGARIAN GEOGRAPHICAL SOCIETY
Journal homepage: www.geography.bg/

Simple queries - Simple answers. First steps in the spatial databases

Sofia Kostadinova[1], Boris Markov[1], Nikolay Petrov[2]

[1] Department of Cartography and GIS, Faculty of Geology and Geography, University of Sofia „St. Kliment Ohridski", Sofia
[2] Department of Transport Management and Logistics, University of Transport „Todor Kableshkov", Sofia

Corresponding author: Sofia Kostadinova; Sofia University,1504 Sofia, 15 Tsar Osvoboditel Blvd;
e-mail: kostadinova.sofia@gmail.com

Key words:
spatial database, SDMS, spatial query, GIS

ABSTRACT

The increasing use of geographic information systems leads to an accumulation of information that requires its systematization and arrangement in a form suitable for multiple use. Spatial databases are a powerful tool that, in addition to storing data in an orderly manner, allows different users to retrieve information useful for decision-making and problem solving, planning, etc.

1. Introduction

The geographic information is 80% of all information. It has high applicability in a wide range of activities, so the application of Geographic Information Systems (GIS) is becoming more and more popular. Governmental structures, NGOs, business and other organizations are applying GIS for geocoding addresses, customer location, store locations, vehicle tracking, for demography, cartography, epidemiology, crime patterns, weather information, land holdings, natural resources, city planning, environmental planning, hazard detection and so on. GIS is an information system consisting of computer hardware and software, database and users, used for entering, storing, manipulating, analyzing and retrieving geographic data to solve a variety of tasks in a variety of areas (Popov and Kotzev, 2011).

GIS integrates four key components: hardware, software, data, and people. Depending on the interest in a particular application, a GIS can be considered as a data store (application of a spatial database), a tool or source of information and knowledge.

Possibly the most important component of GIS is the data. Geographic relational data consist of: spatial attributes (coordinates, geometry) and non-spatial attributes (names number of people, telephone numbers, date, etc.) (Kuba, 2001). As GIS analysis becomes more and more complicated, the need for efficient and intelligent data storage is becoming more and more important. Creating specially designed storage systems for geographic data can make the analysis of data much more efficient.The database is a collection of structured data in a certain way (Popov, 2012). The spatial database is a database, which can store and access spatial data or data that defines a geometric space. Combining database with database management system (DBMS) allows for the creation, storage, maintenance, manipulation, and retrieval of large datasets that are distributed over one or more files.

The general objective of this research is to make spatial queries for a database. For this purpose, we made a preview of the types of standard databases and spatial databases. After short comparison between them, we chose the most appropriate database and we created simple spatial queries to test it.

2. The databases

There are several advantages of storing data in data bases instead of simple data storage. It gives more flexibility in data operations and performing processing such as classification, sorting, aggregation, and calculation. Table. 1 shows some advantages of using databases instead of simple unorganized file storage (e.g. file folder).

Table. 1. Advantages of using databasesinstead of file-based storages (help.cadcorp.com/en/7.0/sis/help/databases_spatial.html)

	File-based storage	Database
File size	Usually limited to 2Gb	Unlimited
Client Memory Usage	Memory intense	Memory efficient
Multi-user Access	Only crude control, if any	Multi-user access is supported
Managing data	Difficult to manage, data may become duplicated	Central storage of data, easy to locate and backup

Security	File system security	Built-in security access to data is provided
Spatial Analysis and Attribute Searches	Carried out on the client. Therefore limited to client capability	Carried out on the Server. Servers are more powerful, therefore analysis is faster.

A DBMS is a complex software, optimized for certain kinds of purposes (e.g. answering complex queries or handling many concurrent requests), and its performance may not be adequate for certain specialized applications (Ramakrishnan, 2003).

2.1. Standard databases

Databases can be classified on different criteria. In this case, we will treat them as a standard database (or non-spatial DB) and spatial DB.

Peneva (2005) gives the following brief definition for a „database" - a set of data grouped for the needs of a user. This definition also has specific properties, such as a resistance of data over time, the presence of a subject area (universe of discourse), as they represent part of the real world and databases are designed and built for a purpose. Maintenance of data is performed by Database management systems (DBMS). Examples of DBMSs are: MySQL, Microsoft SQL Server, Oracle, IBM DB2, Access, FoxPro, Visual dbase are most frequently used for personal computers.

Both a database and its DBMS conform to the principles of a particular database model. The historical development of the database models started with the hierarchical model and network model of data. Later, was developed other types of data models: Entity-relationship model, Document model, Star schema, Array model, Multivalue model and many others. Currently, the most popular model is the relational model. The model was proposed by Edgar Codd (1970) and it remains the basis of the next „post-relational" and object-relational databases. Codd (1970) proposes the introduction of multiple tables to be interconnected by relations, and demand data going through standardized query language (SQL). The relation scan be defined in the object-relational databaseby several types of relationships: „one to one", „one to many", „many to one"and „many to many" relationships.

2.2. Spatial databases

The spatial databases (SDB), or geodatabases, store information for related objects in space. They are distinguished from the standard databases by some characteristics:

• In addition to standard data types (varchar, int, timestamp, binary, xml, float and others) SDB supported geometry type. Geometric data types represent two-dimensional spatial objects.

• Besides the standard SQL, it has additional spatial operators.

The development of SQL generally passes through several stages. The first was accomplished after 1976 and formed SQL1. The second was in the early 90's - SQL2 and the third in 2003 formed SQL3.The last stage format spatial data type gives the opportunity to look at the numbers, currency, date, etc. The difference (Semantic Gap) between SQL & Spatial Query is that the old SQL is not suitable for expression oftype spatial databases and inquiries about them. Traditional SQL describes spatial data, by the use of many many tables and it is too complicated.

Here, as in the standard DB SELECT and FROM are mandatory clauses. WHERE is an optional clause, as well as ORDER BY and GROUP BY and HAVING. In WHERE clause we rely on subtypes rules that we want to return.

According Karimipour (2009) spatial operations represent different aspects of the same real world operation. Therefore, spatial operations have space-invariant properties based on which they can be described. Velicanu (2010) states that there are two important categories of spatial operations: operations that determine the spatial relations and spatial analysis operations. Spatial querying using SQL use simple SQL expressions to determine spatial relationships as distance, adjacency, containment, and others and SQL expressions to perform spatial operations as area, length, intersection, union, buffer, and others. Shekhar (1999) divides the spatial operators into three categories: applied to all geometric types (Spatial Reference, Envelope, Export, Is Simple, Boundary etc.), proof of topological connectivity (Touch, Cross, Within, Contains etc.) and analysis of Spatial Data (Distance, Buffer, Union, Sym Diff etc.).

• In addition to traditional indices, they have the Spatial index.

The spatial index is used by spatial databases (databases which store information related to objects in space) to optimizespatial queries. Spatial indexes differ depending on the data source (http://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/an-overview-of-spatial-indexes-in-the-geodatabase.htm). A spatial index is a type of extended index that allows you to index a spatial column. A spatial column is a table column that contains data of a spatial data type, such as geometry or geography (https://msdn.microsoft.com/en-us/library/bb895265.aspx).

The spatial data is managed by Spatial database management systems (SDBMS). Most standard DBMS have spatial extensions for spatial defined objects. Examples of such databases are IBM DB2 with IBM Spatial Extender, PostgreSQL with PostGIS, Oracle with Oracle Spatial and others.

3. Simple query in a test spatial database

3.1 Data characteristic

We chose PostgreSQL 9.5 because it is the world's most advanced Open Source database. We created a local spatial database with data from the pre-treated attributive information of three shapefiles. The geometry type of the objects in the shapefiles is points (settlements), linestring (rivers) and polygons (NATURA sites). The data was acquired from Civil Registration and Administrative Services, Bulgarian Ministry of Environment and Waters and National Statistical Institute. The position of the objects (Fig. 1) is located in Bulgarian part of Strumeshnitsa river watershed.

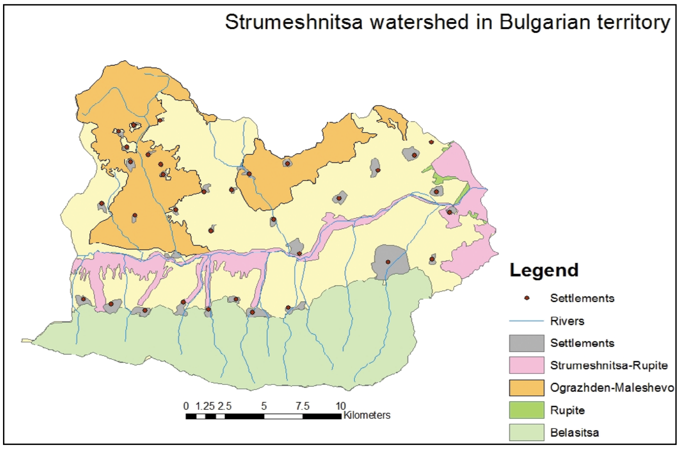

Fig. 1. Location of the objects

The territory is located in southwestern part of the country. Strumeshnitsa river is the largest tributary of Struma river and drains area of 1900 km², but only 441 km² of them are in Bulgaria. The study area falls within three SCI protected areas: Ograzhden-Maleshevo (BG0000224), Rupite-Strumeshnitsa (BG0001023) and Belasitsa (BG0000167), and one SPA protected area: Rupite (BG0002098). The whole territory of the watershed is located in Petrich municipality. It comprises 34 settlements.

When the shapefile is imported an identical attribute table is automatically created in the database. Automatically generated queries look like this:

gid integer NOT NULL DEFAULT nextval ('natura_sites_gid_seq' :: regclass),
siteCode character varying (10)
area double precision,
nameEn character varying (254)
types character varying (254)
nutsLsecond character varying (254)
regionName character varying (254)
bioRegion character varying (254)
bioreg_percent double precision,
geom geometry (MultiPolygon),
CONSTRAINT natura_sites_pkey PRIMARY KEY (gid)

The other files are imported similarly. In all tables appears new column "geom", which has a geometry type.

3.2. Examples spatial SQL

• Simple query #1 - What is the area of Natura2000 sites?

SELECT SUM(area) FROM natura_sites;

In this case, we use SUM() function to calculate the sum of all features in the naturaSites table. PostgreSQL SUM() function is used to find out the sum of a field in various records, and also returns the sum of values for the list of selected columns. It takes as a parameter a cologne name with Numeric data types.

Data Output | Explain | Me

	sum double precision
1	248.81281

Fig. 2. The results of Simple query #1

• Simple query #2 - Which are the settlements that have the same zip code?

SELECT name_en, post_code
FROM settlements_points
WHERE post_code in
(SELECT post_code FROM settlements_points
GROUP BY post_code
HAVING COUNT(*)>1);

Output pane

Data Output | Explain | Messages | History

	name_en character varying(25)	post_code double precision
1	Gega	2882
2	Kladentsi	2895
3	Zoychene	2882
4	Gorchevo	2882
5	Kukurahtsevo	2882
6	Vishlene	2895
7	Drenovitsa	2895

Fig. 3. The results of Simple query #2

This simple query shows points with the same postcode (Fig. 3). It takes as a parameter a cologne name with any data types. We use Count() function, which returns the number of rows that match specified criteria. It also returns the number of values (NULL values will not be counted) of the specified column. The result shows that there are seven settlements that have the same zip code. Four of them using 2882 as zip code and 3 of them using 2895.

• Simple query # 3 - Which rivers originate from Ograzhden?

SELECT r.name_en as name, r.origin
FROM rivers_strbasin as r
WHERE r.origin = 'Ograzhden';

This is a simple query for a river with a given origin. Using the basic CRUD operations, for query context. With this selection are chosen all rivers that have „Ograzhden" in column „origin". Fig. 4 shows that there are three rivers that originate from Ograzgden.

Data Output | Explain | Messages | History

	name character varying(50)	origin character varying(254)
1	Ribnishka reka	Ograzhden
2	Krushka reka	Ograzhden
3	Gradeshnitsa	Ograzhden

Fig. 4. The results of Simple query #3

• Spatial query #4 - Which settlements lies within the Natura sites?

SELECT a.name_en, b.name_en
FROM natura_sites as a INNER JOIN set_point_1 as b ON
(a.geom && b.geom AND st_intersects(a.geom, b.geom));

Here we use spatial operator ST_Intersects, which returns TRUE if the Geometries / Geography spatially intersect in 2D and FALSE if they do not (they are Disjoint). For geography - tolerance is 0.00001 meters (so any points that close are considered to intersect). (http://postgis.org/docs/ST_Intersects.html)

Here, we can easily define if a given point is located into a polygon (Fig. 5). With that query, we can find a discrepancy between the data if Natura sites fall outside settlements and their lands.

Data Output | Explain | Messages | History

	gid integer	name_en character varying(254)	gid integer	name_en character varying(25)
1	2	Ograzhden - Maleshevo	5	Borovichene
2	2	Ograzhden - Maleshevo	11	Zoychene
3	2	Ograzhden - Maleshevo	12	Gorchevo
4	2	Ograzhden - Maleshevo	22	Dolna Ribnitsa
5	2	Ograzhden - Maleshevo	25	Churicheni

Fig. 5. Results of Spatial query #4

• Spatial query #5 - Which is the nearest river to a settlement at a distance of 2 km (2000 m)?

```
SELECT DISTINCT ON (s.gid) s.gid, s.name_en, s.geom, river.name_en
    FROM settlements_points s
    LEFT JOIN rivers_strbasin river ON ST_DWithin(s.geom, river.geom, 2000)
    WHERE river.name_en IS NOT NULL
    ORDER BY s.gid, ST_Distance(s.geom, river.geom);
```

In this query, we use two spatial operators - ST_Distance and ST_DWithin. ST_Distance is a spatial operator for geometry type and returns the 2D Cartesian distance between two geometries in projected units (based on the spatial reference). For geography type defaults returning the minimum geodesic distance between two geographies in meters. ST_DWithin - Returns true if the geometries are within the specified distance of one another. For geometry, units are in those of spatial reference and for geography, units are in meters and measurement is defaulted to use_spheroid = true (measure around spheroid), for a faster check, use_spheroid = false to measure along sphere (http://postgis.net/docs/ST_Distance.html;http://postgis.net/docs/ST_DWithin.html).

Data Output | Explain | Messages | History

	gid integer	name_en character varying(25)	geom geometry(Point)	name_en character varying(50)
1	3	Yavornitsa	010100000053EF33200DB70441AAA47259A47F5141	Yavornishka reka
2	11	Zoychene	01010000003E4C0A3F346F0441E005E8AD62895141	Gradeshnitsa
3	14	Dolna Krushitsa	010100000045CCA32CE011044120F08A5C2D865141	Krushka reka
4	18	Belasitsa	01010000005CC8C70DE4890541B1585D2B507F5141	Luda reka
5	19	Kamena	010100000078AED0B5E8E8044169F08767297F5141	Kameshnitsa
6	21	Mendovo	01010000009F3D00286F6D0541EF2010741A855141	Ribnishka reka
7	22	Dolna Ribnitsa	010100000021D15F8543B054122EA359E1D885141	Ribnishka reka
8	27	Parvomay	0101000000DD186220A4A00541AB706ADDD9825141	Strumeshnitsa
9	29	Petrich	0101000000A728E1BB45530641E450AD6553825141	Petrichka
10	31	Mitino	0101000000E2337A86CFCF0641D1FBC8CD9A855141	Strumeshnitsa

Fig. 6. Results of Spatial query #5

In this case, we look for the nearest river to some point, by given distance in meters. The meters are the third parameter of ST_DWithin () function. The result (Fig. 6) shows, that only ten settlements have a nearby passing river at 2 km distance.

4. Discussion and conclusion

Not every collection of data in one place can be called a database. Database use rules, structures, and connectivity of tables. The spatial database is an even better option for stored information, especially if it is oriented towards a representation of models of the real world. And even more, when the spatial database contains the required data, the user can use the geographic information about many solutions, such as: – measuring an area (Example 1), distance (Example 5), finding a location of objects (Example 3) or many, many others.

In addition to decision-making from the spatial relationships between the objects, we can remotely get information of existing sites. In Example 4 we make a request if there are settlements that fall within Natura sites, but we know that densely built-up areas can't fall within Natura sites. Thus, we understand that actually these settlements are too small and they have only several buildings. We may ask why these places are with so small area and population? We can find the answer in spatial relations between objects.

We can make a request to the database for distance to the biggest settlement in the watershed - Petrich, or is there a bigger settlement, or how far is the main road, or the density of the transport network, and many other requests for data that carry information about the economic benefits that define a location as prosperous.

The modern use of the spatial database is oriented of prediction on various events based on the relationships of the objects e.g. „What-IF" scenarios. If the answer should be correct, the question it has to be relevant. Spatial queries What-IF scenarios is proof that geographic information systems combined with the spatial database will increase their use in more aspects of life.

References:

• Codd, E. 1970. A Relational Model of Data for Large Shared Data Banks. Communications of the ACM. 13 (6): 377–387.

• Frank, A. 1992. Spatial concepts, geometric data models and data structures. Computer & Geosciences. Vol. 18, No. 4, pp. 409-417

• Velicanu A., S. Olaru. 2010. Optimizing spatial databases, Revista de Informatica Economica, Vol. 14, Nr. 2 / 2010, pag. 61-71, ISSN 1453-1305.

• Velicanu, A. 2010. Spatial Operations. Database Systems Journal vol. 1, no. 1/2010. Bucharest, Romania

• Karimipour, F., M. Delava, A. Frank. 2009. "An Algebraic Approach to Extend Spatial Operations to Moving Objects", World Applied Sciences Journal 6 (10): 1377-1383, 2009, ISSN 1818-4952.

• Kuba, P., 2001. Data Structures for Spatial Data Mining. FI MU Report Series. FIMU-RS-2001-05. Faculty of Informatics Masaryk University.

• Peneva, Y. 2005. Databases. Regaliya 6. ISBN-954-745-088-3. Sofia (in Bulgarian)

• Popov, A.. A. Kotzev.2011. GIS and Internet. Fondation "LOPS", Sofia.

• Popov, A. 2012. Geographic information systems. Fundamentals of geo-information modeling. Anubis. Foundation "LOPS", SU "St. Kliment Ohridski", Sofia. (in Bulgarian)

• Ramakrishnan, R., J. Gehrke. 2003. Database management systems. ISBN 0-07-246563-8-ISBN 0-07-115110-9 (ISE). 3rdInternationalEdition. US.

• Shekhar, Sh., Sanjay Chawla, Sivakumar Ravada. 1999. "Spatial databases-accomplishments and research needs". In: IEEE transactions on knowledge and data engineering, V.1, p. 45-55.

• http://help.cadcorp.com/en/7.0/sis/help/databases_spatial.html (Accessed: 25.09.2016)

• http://www.postgresql.org/(Accessed: 19.09.2016)
• http://www.cubrid.org/blog/dev-platformhttp://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7556/20-minutes-to-understanding-spatial-database/(Accessed: 25.09.2016)
• http://www.w3resource.com/PostgreSQL/postgresql-sum-function.php(Accessed: 27.09.2016)
• https://www.tutorialspoint.com/postgresql/postgresql_data_types.htm(Accessed: 27.09.2016)
• http://www.w3schools.com/sql/sql_func_count.asp, (Accessed: 27.09.2016)
• http://postgis.org/docs/ST_Intersects.html(Accessed: 21.09.2016)
• http://postgis.net/docs/ST_Distance.html(Accessed: 21.09.2016)
• http://postgis.net/docs/ST_DWithin.html(Accessed: 21.09.2016)
• http://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/an-overview-of-spatial-indexes-in-the-geodatabase.htm(Accessed: 22.09.2016)
• https://msdn.microsoft.com/en-us/library/bb895265.aspx(Accessed: 20.09.2016)

DOI 10.3897/jbgs.2018.39.16