

*Medjon HYSENAJ**

INTERACTIVE METHODOLOGY FOR SPATIAL REPLICATION AND RESEARCH

Abstract

The paper outlines geographic search technology as the future of research methodology. Due to the importance of geographical environment in the routine issues, the close connection between spatial extension and most of the phenomena running over the world, the geographic query process turns to be key factor in overcoming hazard decision-making and evaluation. We describe a two step process where derived output from spatio-textual research could be largely disseminated by spatial replication techniques. Spatial research differs in that it requires a combination of text and geographical data. The mathematical formula to calculate distances derives from the assertion that the Earth is nearly spherical (ignoring ellipsoidal effects) so great-circle distance formulas give the distance between points on the surface of the Earth correct to within 0.5% or so. Based on the search range and limit, we introduce the Cosine algorithm who offers better speed performance and data retrieval accuracy. Retrieved data exploit spatial replication technique potentiality. When it comes to spatial data we deal extensive amount of information. Web Gis Albania is an informative platform initially structured for academics purposes further enriched with modules which support the spatial data circulation.

1. INTRODUCTION

The World Wide Web nowadays has turned into an endless research environment where everybody is trying to retrieve the largest amount of data within the minimum efforts. It becomes necessary to structure the appropriate platform to handle the increasing demands of the web users.

* University of Shkoder, Department of Geography, Albania, mhysenaj@unishk.edu.al

Geographic web search engines allow users to constrain and order search results in an intuitive manner by focusing a query on a particular geographic region [1]. Geographic research brings the proper information to the proper owner, avoiding abundant data out of user's interest. Content that is tailored to a region is potentially more useful to users from that region [2].

The vast amount of geographic information concerning the real world contains rough uncategorized data. Through spatio-textual research (figure 1) we aim to create a categorized repository system object to massive dissemination. The goal is to reach a unified level of geographical information acting as a standard for most users.

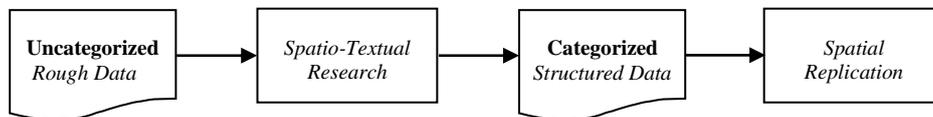


Fig. 1. Spatial-cycle data circulation [source: own study]

The terms geographic queries and spatial queries imply querying a spatially indexed database based on relationships between particular items in that database within a particular coordinate system (or compatible coordinate systems) [3]. The Web GIS Albania application (Figure 2) combines spatial with textual queries.

We exploit mathematical algorithms for spherical search coordinates mainly represented by the law of Cosine with textual queries based on the entity attributes from who derive the platform geo-database. A spatial database is a database system that offers spatial data types in its data model and query language and supports spatial data types in its implementation, providing at least spatial indexing and spatial join methods [4]. Geographic data is often subject of local or global target for multiple profile entities such as business and marketing strategy, hazard management, logistic assignment, etc.

As for large part of web research has strict implicit or explicit connection to geographic location we determine that these issues turn to be geo-sensitive. Hence the use of spatial database as the backbone repository instrument for the application could turn into a logic solution for the increasing number of geo-requests. A spatial database is considered complete from that it accomplishes thoroughly every complex request. Hence it is important enriching the geo-database with current and updated information.

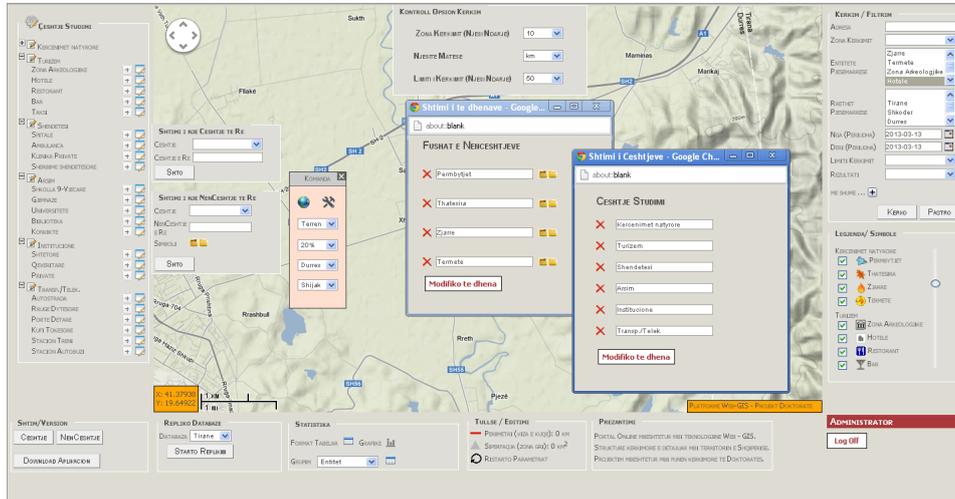


Fig. 2. WGA application, Back-office [source: own study]

2. RESEARCH METHODOLOGY

The application research process allows the user to deal with multi functionalities, covering large part of user request, without the need to perform distinct and consecutive actions. Let's analyze two distinct cases.

Case one: the request consists to retrieve a list of bars, hotels and restaurants within e distance of 10 km from a specific hotel? For management purpose the investor need to retrieve a list of these three business entities that could turn useful or maybe potential competitors fur his future investment. He isn't interested in other entities so he doesn't expect the list to contain other useless data. By so the search engine offers the exact output he needs with minimum efforts. From figure two we notice there are six entities that fulfill both requirements distance and type. Further the user needs a report containing only the bars? Although the distance remains the same, the modified query will turn already three from previous points.

Case 2: During the period 2010-2011, Shkodër district suffered from large scale flood. The population was affected from both social and economic aspect. By that time, responsible authorities were charged for uncoordinated action and slow intervention process. Basically the created situation was the product of a missing platform, able to lead to well organized instructions. The elaboration of an evacuation plan consists in finding the optimal routes (distance and security), accommodation entities (shelter for affected population) from flooded areas, etc. The current application would have solved many uncertainties in short time.

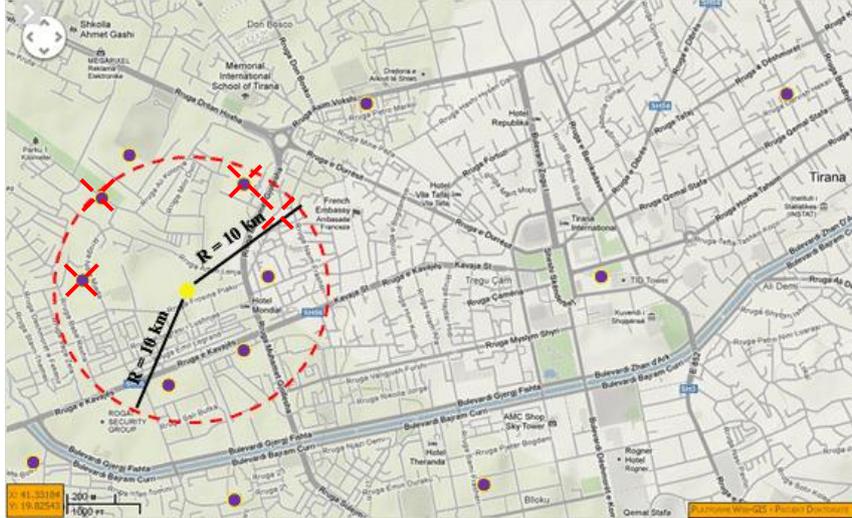


Fig. 3. Research schema

Three years ago after the first online appearance, the application handled the research process exploiting the Haversine formula (1). Principally the Haversine formula is far superior from other formulas for small distances at the range of a couple of meters. However the applications goal is to perform researches far beyond this limit.

$$dLat = (lat2-lat1) \quad (1)$$

$$dLon=(lon2-lon1)$$

$$a=\sin(dLat/2)*\sin(dLat/2)+\cos(lat1)*\cos(lat2)*\sin(dLon/2)*\sin(dLon/2)$$

$$distance = 6371 * 2 * \text{atan2}(\text{sqrt}(a), \text{sqrt}(1-a))$$

Claiming larger distances at the range of kilometers, we discovered that Vincenty formula or the law of Cosines have better performance as for accuracy matter. The Law of Cosines (2) uses spherical geometry to calculate the great circle distance for two points on the globe [6]. Although the Haversine formula is more robust to floating point errors [5], in fact, JavaScript (and most modern computers & languages) use 'IEEE 754' 64-bit floating-point numbers, which provide 15 significant figures of precision, with this precision; the simple spherical law of Cosines formula gives well-conditioned results [6]. On the other side the law of Cosines performs much faster than the Vincenty formula. So taking into account speed performance and accuracy of retrieved data, we conclude that at present the law of Cosines meets better application goal.

$$d = \text{acos}(\sin\phi_1 \cdot \sin \phi_2 + \cos\phi_1 \cdot \cos \phi_2 \cdot \cos \Delta\lambda) \cdot R \quad (2)$$

Below we notice the low of Cosines formula integrated into the application search code where the 's' variable according to the row of appearance stands for: latitude, longitude, latitude, search_radius and search_limit.

```
SELECT name, address, latitude, longitude, (6371 * acos(cos(radians('%s')) *
cos(radians(lat)) * cos(radians(lng) - radians('%s')) + sin( radians('%s')) *
sin( radians(lat)))) FROM geographic_table HAVING search_radius < '%s' ORDER BY
search_radius LIMIT 0, '%s'"
```

The interface (figure 4) containing the necessary tools to perform the required actions. The goal is to dismantle every issue into elementary parts so that the user can easily reach the target. The set of filters handle complex queries structured to cope various situations in a stressful and eager investigation environment, also based on the involved entities additional criteria have been added. They reflect relevant attributes derived from the responsible geo-database. Further the intention remains to keep enriching the platform with additional modules and research specifications.

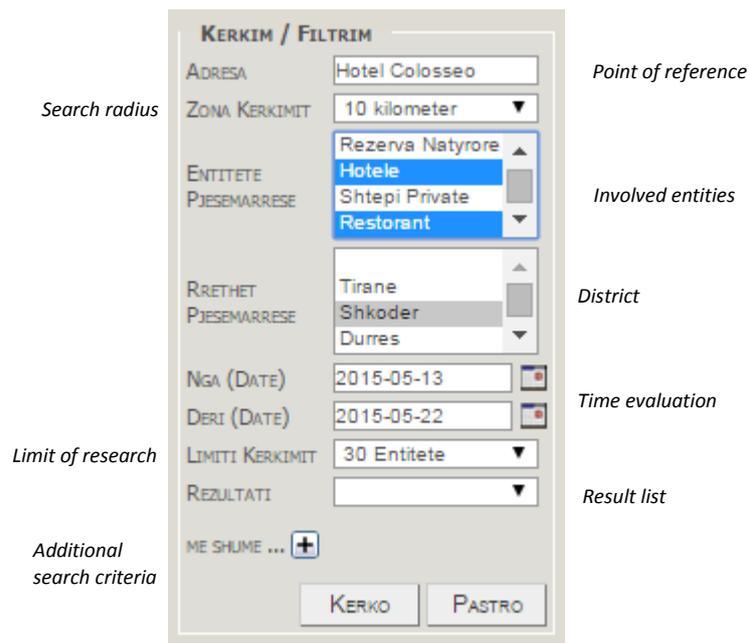


Fig. 4. Research tool [source: own study]

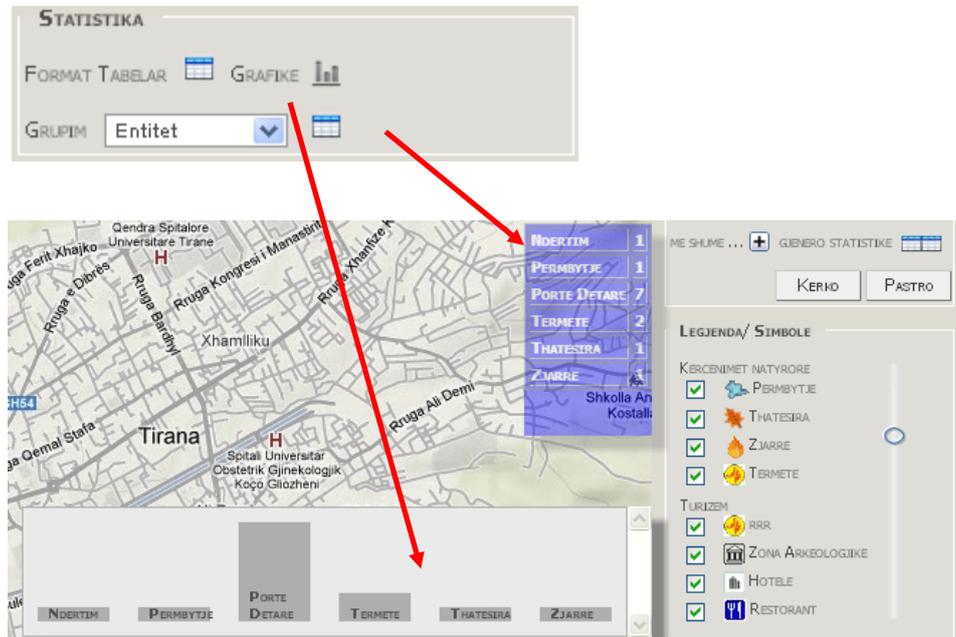


Fig. 5. Statistical tool [source: own study]

As the main goal of the researcher is to produce data in terms of reports or synopsis WGA automatically generates statistics (figure 5) based on the output. The report can be structured based on two elements entity or location. In case we scan a specific area (usually at district level) means we are interested to analyze entities for that area. In case we perform the research on multiple areas (more than one district) it is possible to group as well as by entity or location. Generated data can be analyzed on graphs or tabular formats and exported into pdf., xls. or doc. formats. The user reaches a new standard of just visualizing the current issue.

3. REPLICATION TECHNIQUE

The development of handled applications such as tablets, smart phones, laptop, for dynamic users creates optimal conditions to focus on Gis technology. Data replication stands as an important pillar for management issues for many private and government entities. Data replication is driven by the need to maintain more accurate and up-to-date spatial databases, but the same time reduce data acquisition and maintenance costs [7]. Due to the new administrative reform in the country a large number of issues need to be restructured.

The new territorial reform affects directly spatial attributes which also need to be reflected on maps. Old data must be replaced by entities. This will affect most of spatial distributed object which will need partial or full update. When it comes to geographic entities we easily presume the large quantity of data to be affected. Using traditional methods could cause process overload and inconsistent data management. Efficient data exchange is crucial to the successful future of the geospatial sector [8]. Gis web based platform for data management associated with replication techniques for data turnover and distribution produce an optimal combination to support such issue.

The spatial database contain exclusively geographical coordinates associated with relevant attributes for each object avoiding this way redundant data that could slow the application performance. It wasn't necessary to store data related to the map structure; due to internet performance has increased sharply, this allow us to exploit pre-generated maps offered by Geo suppliers which are much faster to load compared to local geo files. We exploit maps offered by Google through Google APIs by creating a second layer of markers for every accomplished research which overlays the existing map. This schema allows developing a two sided data flow between the user and the spatial database. A successful research is translated into a generated XML file containing located objects and the respective coordinates and attributes.

4. DATA FLOW SCHEMA

Nowadays creating multiple copies of information into multiple locations turns into an essential step for a healthy performance. MySQL Replication raises the bar – a whole new level in terms of performance [9]. Platforms dealing with large amount of data are closely related with such supporting techniques. Web Gis Albania primarily relies on geographical data.

Among all types of data, the necessity to keep spatial data up-to-date is imperative for their dynamic nature. We claim the term dynamic due to the rapid evolution. A restaurant may close or transfer somewhere else. Bars can change name or category. Business entities are often subject to developments. Hence information reflected geographically, associated to real world objects, evolves rapidly, as much as a week or month could be considered high value variables.

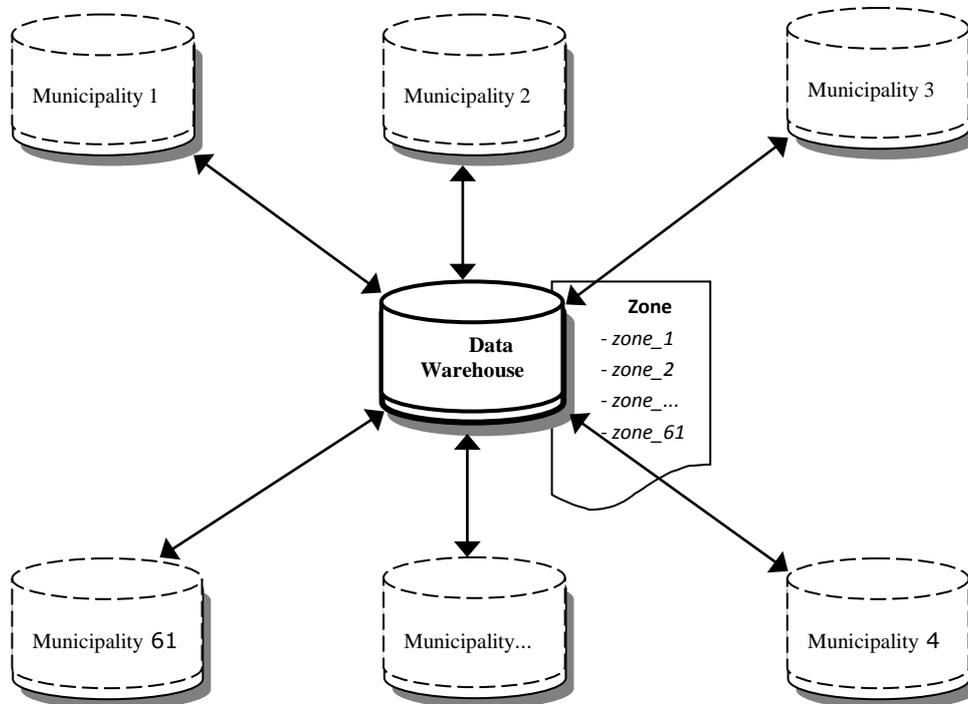


Fig. 6. Replication schema [source: own study]

The new territorial reform divides the country into 61 administrative areas called municipality. We need geographically dispersed units converging into a centralized environment. That requires 61 peripheral slave servers connected into one master server (Figure 6). Peripherals act as autonomous entity which collect, distribute and manage geographical data related to the associated area.

Each peripheral unit asks for specialists with sustainable geographical area knowledge associated with supportive skills in managing the platform system. The potentiality of WGA to generate reports has direct effect on the server performance which can lead to overload and data loss. Multiple backup and healthy recovery plans must always be on standby. This might need great efforts leading to massive consumption of physical and time sharing resources. Replication is considered a modern solution for backup and a healthy failover management.

From previous paper work [10] I described the application data flow schema (Figure 7). We use two distinct database systems. Spatial database refers to any set of data describing the semantic and spatial properties of real world phenomena (temporal properties are also possible) [11]. The so-called “spatial database” is associated the context of data which preserves. We structure spatial database responsible for all entities belonging to the analyzed area.

Data insertion occurs through a simple mechanism, who transforms objects identified in real world into coordinates which are associated with corresponding attributes. They are unique as the objects reflecting the outside reality. These coordinates constitute the groundwork from which we manage to display symbols on maps and further structure reports based on the output.

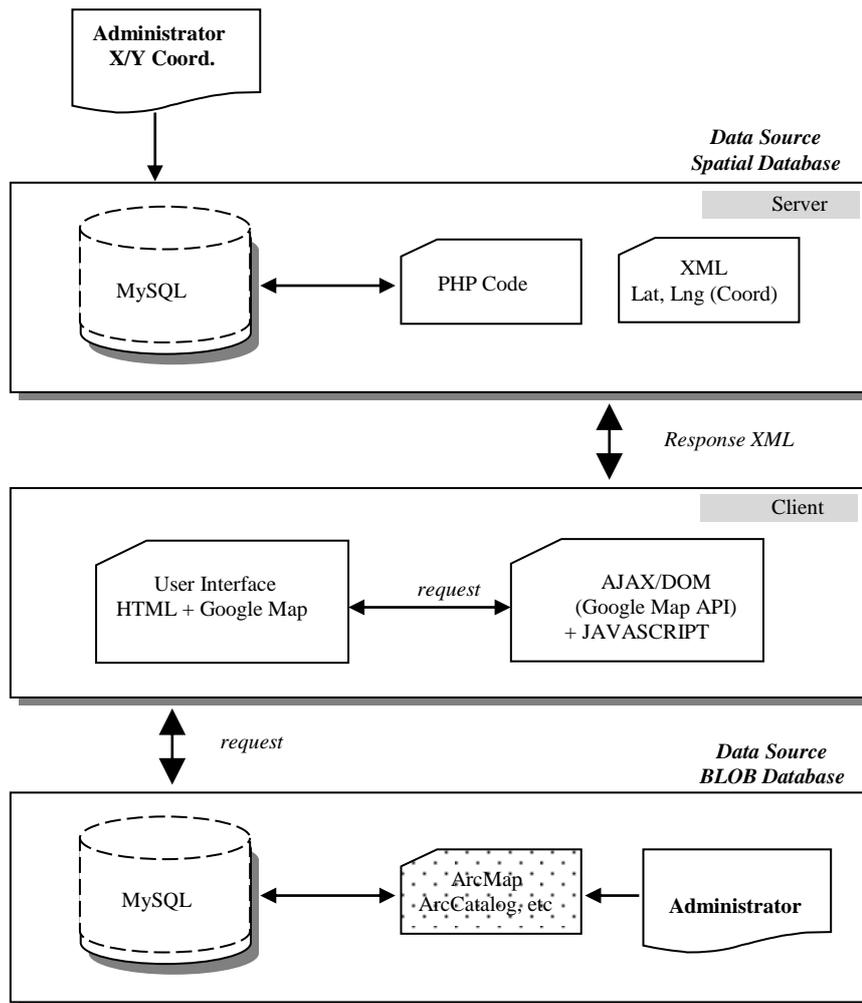


Fig. 7. Data flow schema, Web Gis Albania [source: own study]

We exploit a second database to store pre-processed dynamic maps with the help of ArcGIS package. After that, we store maps converted into images .gif or .jpeg type in the database as blob files. Blob is a binary large object that can hold a variable amount of data whose values are treated as binary strings. It is compatible when it comes to store image data [12].

Figure 8 present the spatial database schema which store any entity that can be spatially displayed and represent a point of interest for the population and decision-making authorities. Each primary entity is divided into categories in a one-to-many relationship reflecting specific issues. For each category in a one-to-many relationship, we structure databases reflecting relevant units storing coordinates and specific attributes for each object of that category. Most of information is stored in these tables.

The centralized spatial database claims the same table structure. Meanwhile peripheral databases preserve the geographical source of the derived data through the additional attribute *id_municipality*, essential to avoid data overlap. Through this column all data replicated into the Master database could be further exploited to produce general reports.

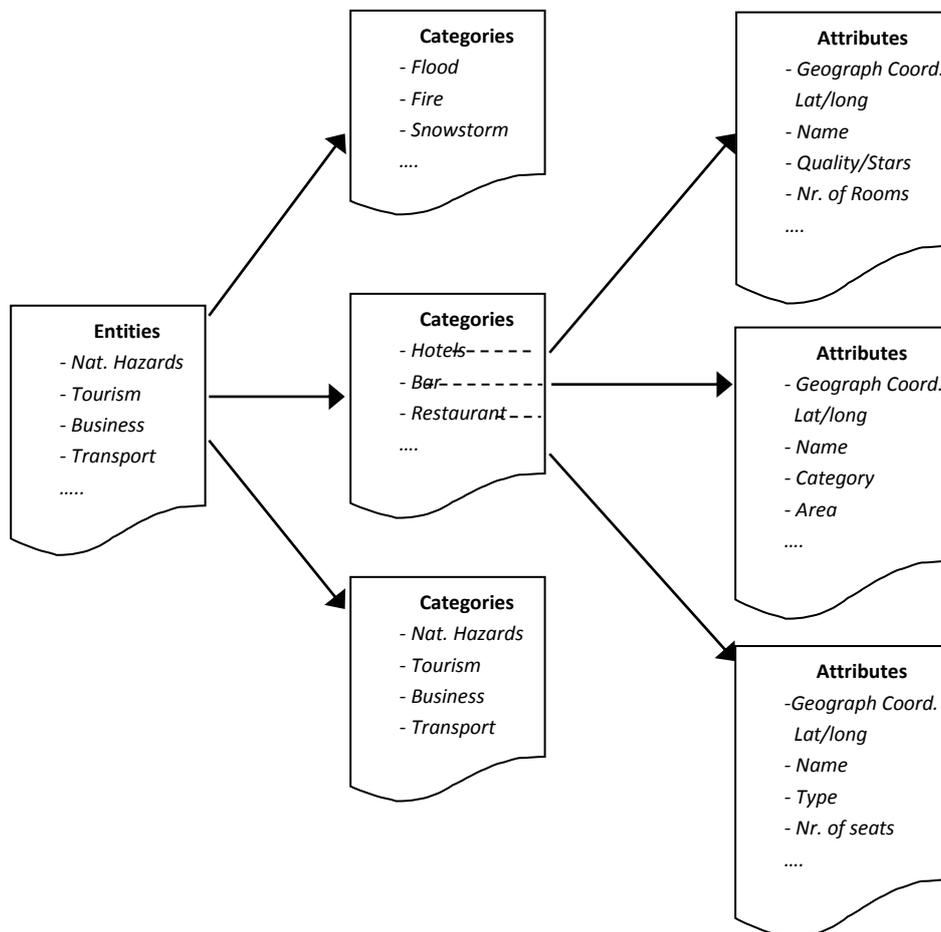


Fig. 8. Spatial database schema [source: own study]

5. CONCLUSIONS

All aspects of human activity are rooted in geographic space in some aspect [13]. As a result references to geographical locations are prevalent everywhere. The paper goal is to present a spatial cycle data circulation based on a spatio-textual research transforming rough data into structured information further disseminated by spatial replication techniques. The spherical algorithm such as the law of Cosin associated with additional text queries produced interesting results. We build a GIS based platform, initially for academics research, offering a clear perspective for a spatial search engine. On the basis of well performed query stands a fully developed spatial database. That's why the platform needs frequent "attention".

The new territorial reform is an incentive to reflect geographical evolution. Good territory administration creates accurate, accessible, interoperable, timely, secure, and complete information about land in an affordable and efficient way that promotes confidence between the public, its commercial enterprises, and government [14]. Wherever geographic evolution is present there is always necessity to reflect spatial changes. Administrative units include vast entities which by them undergo deep changes not only geographically, reflecting their spatial location evolution; also we face consistent changes in their associative attributes.

The paper addresses problems associated with spatial data replication within a data-centralized mechanism and geographically distributed environment. The scope is to settle a healthy multiple data circulation among resources whose location are spatially extended. We concentrate on object location and their geographic coordinates followed by their associative attributes. The maps act as linen where results from the spatio-textual research overlay on it. The data flow allows us to produce in short time complex data analyze and explicit reports based on decision-makers requirements.

The replication technique will allow us to act as a connection bridge for all units into a centralized repository from which initiate most of the consentaneity for the geographic data flow. Through spatial replication we aim to reduce time extension of processes, increase availability and failure management, create a reliable standard from which people can easily retrieve informative data. WGA act as a source of communication also as a tool for most academics projects. The goal is to prompt Gis supporters to rely on spatial technology and to further enrich their skills on spatial technology. Location information integrated into our digital information environment adds a previously missing conceptual dimension to our formulation of problems and questions [15]. The spatial component will turn into an underlying stone for many interactive processes such as urban planning, environmental issues, research methodology, business operations, etc. It is necessary to enlarge and upgrade the application. In the future perspective further enrichment of the spatial database will turn a key issue.

REFERENCES

- [1] YEN-YU CH., TORSTEN S., ALEXANDER M.: *Efficient Query Processing in Geographic Web Search Engines*. 2006.
- [2] KEVIN S.: *Geospatial Mapping and Navigation of the Web*. IBM Almaden Research Center, U.S.A, 2001.
- [3] RAY L.: *Geographic Information Retrieval and Spatial Browsing*. School of Library and Information Studies, University of California, Berkeley, 2000.
- [4] RALF G.: *An Introduction to Spatial Database Systems*. Praktische Informatik IV, FernUniversität Hagen, D-58084 Hagen, Germany, 1994.
- [5] Formulas to Calculate Geo Proximity, www.stackoverflow.com.
- [6] CHRIS V.: *Calculate distance, bearing and more between Latitude/Longitude points*. www.movable-type.co.uk/scripts/latlong.html
- [7] MARK S., Katherine I.: *Data Replication and Data Sharing – Integrating Heterogeneous Spatial Databases*.
- [8] SERRYN E., FRANCISCO E., IAN W.: *Hierarchical spatial reasoning theory and gis technology applied to the automated delineation of administrative boundaries*. 2000.
- [9] LUIS S., LARS T.: *MySQL Replication High Performance: Multi-Threaded Slaves and Group Commit*, 2012.
- [10] MEDJON H.: *Web GIS Albania Platform, an Informative Technology for the Albanian Territory*, Informatica 2012.
- [11] BEDARD Y.: *Principles of spatial database analysis and design*. 2005.
- [12] ORACLE, MySQL reference manual, 2015.
- [13] CHRISTOPHER B., ALIA I., DAVID F., GAIHUA F., SUBODH V.: *The SPIRIT Spatial Search Engine: Architecture, Ontologies and Spatial Indexing*. School of Computing, Cardiff University, Cardiff, Wales, CF24 3XF, UK, 2004.
- [14] ROBIN M., VICTORIA S.: *Ict for land administration and management*. 2011.
- [15] The Spatial Web, An Open GIS Consortium (OGC) White Paper, www.opengis.com, 2004.