Normalizing Data Regions in Spatial Data Grid Using Parallel BNN with MOW

K.Balasaravanan[1], Dr. K. Duraiswamy[2]

[1]Department of Computer science and Engineering, CMS College of Engineering, Namakkal
[2]Department of Computer science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode
Tamilnadu, India
[1]kbalasaravanancse@gmail.com
[2]drkduraiswamy@yahoo.co.in

Abstract: As the promising expertise, grid computing is functional to several projects. And the administration of spatial data also features novel challenges. The appliance of grid computing commenced to the organization of spatial data highlights the exploitation and intention of spatial data grid. The notion of grid computing has infused all areas of dispersed computing, altering the mode in which dispersed systems are considered, urbanized and executed. In spatial database, if mixed regions are present in dense and sparse dataset in an arbitrary manner, then it is difficult to retrieve the items from the database. To make the search more effective, in this work, we present a scenario that illustrates for the foremost time how data grids can be functioned to facilitate the sharing of deal with data in spatial data infrastructure (SDI). Combining spatial data from dispersed diverse sources into a distinct centralized dataset entails, amid others, a significant human management endeavor. We present a reference model for aligning the sparse and dense data objects in an efficient manner based on the data object position and localization. Then the technique normalizes the data objects in the corresponding regions, processes all the queries in a parallel way and produces the results to all requested user at a short interval of time simultaneously. BNN search based on Marginal Object Weight ranking scheme used here to retrieve the items from distributed servers with high dimensional data structure. An experimental evaluation is carried over with real data sets to estimate the performance of the proposed spatial data grid formation based parallel BNN using MOW in spatial database with an existing efficient BNN search using MOW ranking scheme in spatial database.

[K.Balasaravanan, Dr. K. Duraiswam. **Normalizing Data Regions in Spatial Data Grid Using Parallel BNN with MOW.** *Academ Arena* 2019;11(6):72-79]. ISSN 1553-992X (print); ISSN 2158-771X (online). http://www.sciencepub.net/academia. 12. doi:10.7537/marsaaj110619.12.

Keywords: Grid formation, mixed regions of sparse and dense, normalization, parallel, MOW, Spatial data Infrastructure

1. Introduction

In common, spatial data mining, or knowledge detection in spatial databases is the pulling out of inherent knowledge, spatial associations and detection of remarkable uniqueness and patterns that are not openly symbolized in the databases. These methods can acts as an imperative task in accepting spatial data and in detaining essential relationships among spatial and non-spatial data. Furthermore, such exposed relationships can be utilized to current data in a brief manner and to restructure spatial databases to contain data semantics and accomplish high presentation. Spatial data mining has extensive applications in several fields, counting image database exploration, GIS systems, medical imaging and so on.

Spatial data organization is the crucial utility of Geographical Information System (GIS), in addition to a very significant guide of determining GIS software capabilities. Motivated by database tools and Internet, spatial data administration widen from desktop to WAN (Wide Area Network) and Internet, and the capability of spatial data organization also expand reasonably. With the development of human's observation and perceptive, spatial data management facade to some novel situations as pursues:

- The amounts of spatial data invention have gathered to a definite degree; sensors develop into more and more, the established data tot up.
- The scales of problems resolving develop into more and larger. It is unfeasible to resolve problems presently depend on numerous super-computerse. g. universal weather examination.

The dispensation of spatial data is flattering gradually more pertinent in abundant financial sectors and fields of examine. There models of incessant wide-area provinces consist of large amounts of spatial data. In this context modern concepts of parallel computing are needed to reach acceptable computing time for processing such data models. In context of processing spatial data the OGC provides a range of standards for a web service architecture which can handle spatial data in a specific way. With these specifications several SDIs are already realized as one kind of network, which presents structuring and dispersed processing service manacles on geospatial datasets. In systematic computing a different kind of network - grid environments - can implement computational exhaustive calculations on very huge quantity of data by employing web services. There a grid location presents a allocation of calculations and datasets on abundant rich prepared grid nodes with prospect of dispersed high-speed data transfer. The progression process of both networks - SDI and grid environment - previously expands over numerous years and fallout in abundant complicated components of its architecture. An incorporation of both networks can unite the dispersed processing by using geospatial service chains with abundant distributed rich prepared resources of a grid atmosphere.

In this work, we present a scenario that illustrates for the foremost time how data grids can be functioned to facilitate the sharing of deal with data in spatial data infrastructure (SDI). Combining spatial data from dispersed diverse sources into a distinct centralized dataset entails, amid others, a significant human management endeavor. We present a reference model for aligning the sparse and dense data objects in an optimal manner based on the data object position and localization. Then the technique normalizes the data objects in the corresponding regions, processes all the queries in a parallel way and produces the results to all requested user at a short interval of time simultaneously. BNN search based on Marginal Object Weight ranking scheme used here to retrieve the items from distributed servers with high dimensional data structure.

2. Literature Review

Spatial data management is the crucial purpose of Geographical Information System (GIS), in addition to a very significant directory of computing GIS software skills. Through these years, the introduction of grid computing knowledge presents a new method for spatial data organization.

The progression process of both networks previously expands over numerous years and fallout in several complicated components of its design. So an incorporation of SDI and grid knowledge is valuable by amalgamation of dispersed processing geospatial service cuffs with frequent dispersed affluent prepared resources of a grid situation. The paper [1] converses diverse approaches to carry mutually worlds – the SDI and the grid situation – simultaneously and "indulge" OGC submissive service cuffs as preserving the usability for spatial data communications service [4] user and receiving an optimized data and in sequence flow in the appreciative service chain. It highlights consumption and design of spatial data grid [7].

In a spatial database, indecisive assessment query processing is a considerable quality in data mining. The objects in the queries utilized to recognize the distance amongst two objects can be calculated summarily, whilst [10] distance estimation is restricted with NN search using permanent queries. For a spatial queries, [12] offered a NN search using Voronoi Neighbors by identifying the best NN object consistency. Query processing in spatial database for NN search [3] has been demonstrated in various aspects. For admitting the indecisive spatial database, querying objects [2] proposed adaptive algorithms to construct a decision whether to employ a key search or a data scrutinize for each step throughout the processing of a queries are used. For an efficient search in uncertain database, NN search using Top-k queries in a given uncertain database using query model with authentication systems [5].

An active preservation mechanism [6] to attain brilliant load balance and an algorithm for spatial series queries is offered. Both the outcomes of theoretic examination and researches explain that the administration of spatial data in the framework is proficient. In a grid situation, implementations of GIS (geographic in sequence system) should not only distress their domestic spatial distinctiveness [4], but also compact with the network resources properties. Consequently, the association and assistance can be officially confronted by the requirement of users who desire to manage diverse services, running on diverse platforms [11] in dispersed data resources [9]. A novel replication advance [8] which are adaptive, entirely decentralized, and supported on swarm intellect which is essentially a bottom-up technique.

3. Materials and Methods

Within the possibility of dealing out huge amounts of data and composite calculations the employment of grid computing is a fine preference for attaining high performance, since processing of widespread spatial data can be very computationally rigorous. At the instant there subsist spatial data infrastructures (SDI) as outline with the use of web services, which are consistent by the Open Geospatial association. The architecture of a SDI is focused on formation and dispersed processing service manacles in geospatial framework. The proposed normalizing regions of data in spatial data grid using parallel BNN with MOW is operated under two phases. The first phase describes the process of formation of grid based on the size of the grid with respect to object localization and position. The normalization of data regions takes place based on the mixed regions of spatial and dense datasets. The second phase describes the parallel BNN search with MOW ranking scheme. The architecture diagram of the proposed normalizing regions of data in spatial data grid [NRDSD] using parallel BNN with MOW is shown in figure 1.

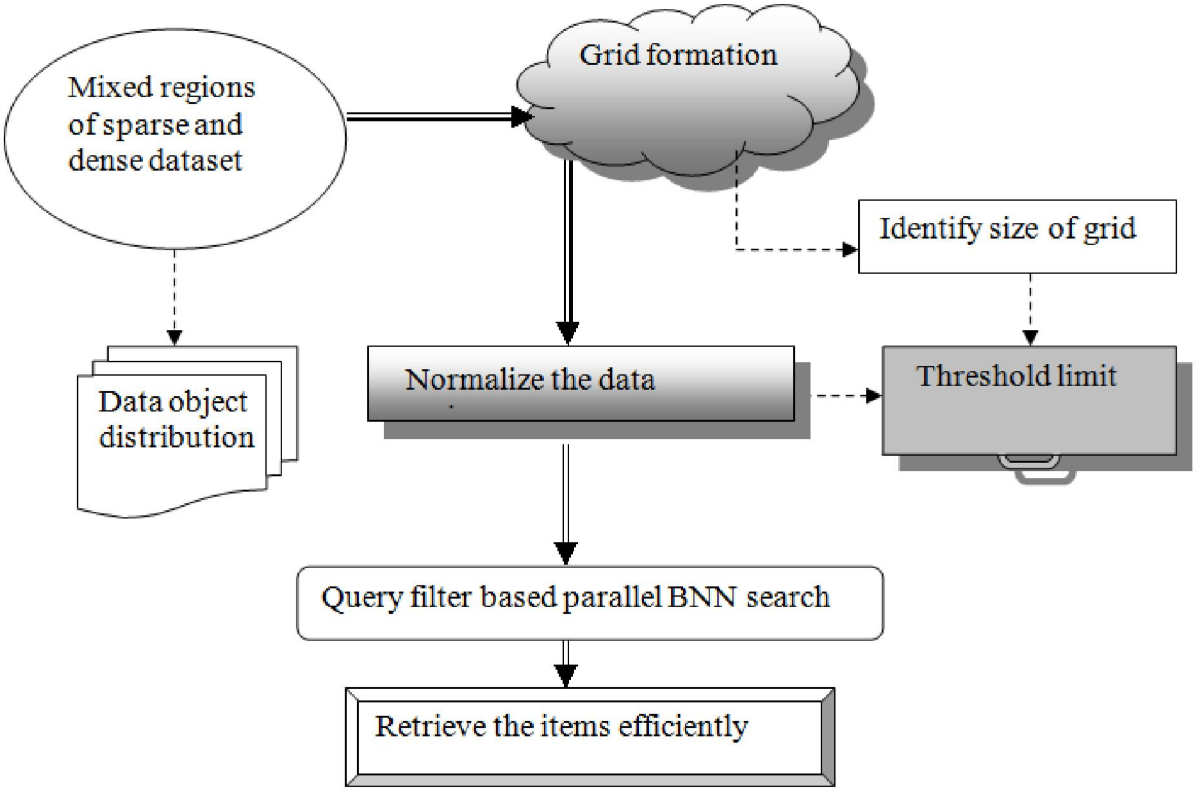

Figure 1: Architecture diagram of the proposed NRDSD

Spatial data infrastructure: The purpose of a SDI is the hosting of geographic data and attributes, adequate credentials (metadata), a way to determine, imagine and assess the data (catalogue and web planning) and some techniques to present contact to the geographic data. The construction chains energetic fastening of services by issue/identify/bind guide, where a service will be available to a tune broker. A requestor is able to discover metadata of the demanded service at the service adviser and then the requestor can connect the service at the position explained in the established metadata.

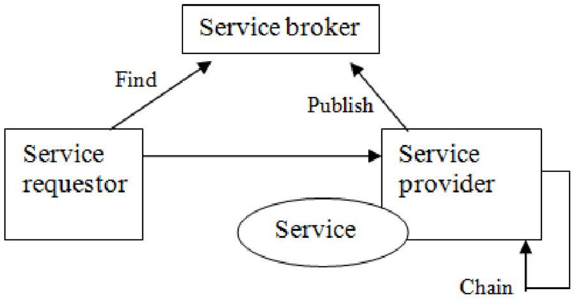

Figure 2: Publish/find/bind pattern

The architecture ropes active requisite services by concern/find/bind outline, where an examination will be available to a service adviser. A requestor can discover metadata of the demanded service at the service dealer and then the requestor is able to combine the service at the location explained in the determinacy metadata (shown in Figure 2). There a place of particular data objects compliant web services will be precise, e.g.:

- For computations on very widespread spatial data identifies a consistent boundary **Web Processing Services (WPS)** to present any kind of GIS functionality, counting admittance to calculations and/or calculation models.

- **Web Coverage Services (WCS)** are précised to illustrate and transport multidimensional exposure data.
- **Web Feature Services (WFS)** preserved data access functionality and processes on geographic features.
- **Catalogue Services (CS-W)** hold the detection and recovery of spatial data. Also a CS-W provisions and offers metadata of services.

Gridification of SDI: A grid is a method that is apprehensive with the incorporation, virtualization, and organization of services and possessions in a dispersed, diverse location that ropes compilations of users and resources (practical organizations) crosswise conventional secretarial and directorial provinces (real organizations). Spatial data grid, as its name means, will present spatial data organization and correlative services supported on spatial data. Spatial data grid supervises multi-precise and multi-scale spatial data, counting raster image, vector and DEM. These spatial data can hoard in database, file or others behaviors, consistent with condition.

In general Gridification is the shedding of offered applications and services into the construction of a grid atmosphere. Users of a SDI desire to generate the SDI service series in a verified way, but they desire to obtain the reward of using the grid construction.

Thus in center of a spatial data infrastructure three objectives of Gridification can be recognized:

1. A prospect to unite one gridified with other (probably non-spatial) grid services and vice versa.
2. An access to non-gridified OWS must be probable.
3. The Gridification of SDI service chains should be approved out in such a mode, that a exploited presentation among connected services can be attained by sufficient parallelization techniques which may deal with from a distinct service up to the complete service chain. Thus both a view on Gridification of distinct services and a sight on planning compound connected service chains to a grid environment are desired.

Normalizing data regions using grid formation: After the grid has been constructed with a mixed regions of sparse and dense datasets, the grid formation is done based on the size of the grid with respect to the data objects position and localization in the grid. Analyze the data objects in the grid and form a structure based on the threshold limit. The threshold limit is computed based on the distance between the data objects present in the data grid ie., minimum and maximum distance required to forma group.

We generate the hierarchy of data objects with their associated parameters when the data is loaded into the database. Parameters m, n, s, max, and min of bottom level data objects are calculated directly from specified data. The value of distribution could be either assigned by the user if the distribution type is known before hand or obtained by hypothesis tests such as X2-test. Parameters of higher level data objects can be easily calculated from parameters of lower level data objects. Let m, n, s, max, min, dist be parameters of current data object and m_i, n_i, s_i, min_i, and max_i be parameters of corresponding lower level data objects, respectively. Then m, n, s, max and min can be computed as follows.

Based on the threshold limit, the data objects in the grid are segregated and form a group. Through the formation of grid based on the size of the data objects, the normalization of data regions have been computed.

Query filter based Parallel BNN search with MOW: A query filter filters the queries based on its similarity. The similarity of the query is identified by the process of retrieval of same information from the spatial database system. Those similar queries are segregate out from the list and form a group. Normally, the queries are come with constraints to access the spatial database system. Based on the constraints, the similarity of the queries are identified and segregated from the list.

$$n = \sum_i n_j$$

$$m = \frac{\sum_i m_i n_i}{n}$$

$$s = \sqrt{\frac{\sum_i (s_i^2 + m_i^2) n_i}{n} - m^2}$$

$$\min = \min_i (\min_i)$$

$$\max = \max_i (\max_i)$$

Where n-number of data objects (points)
m - mean of all values of the data objects
s - standard deviation of all values of the data objects
min - the minimum distance of the data object from the current data object
max - the maximum distance of the data object from the current data object

After formation of grids with respect to the data object distributed size, the process of parallel BNN search will take place. Since the queries are segregated, the process of searching the nearest object for those queries becomes an easy approach in the spatial database system. The segregated similar queries are given as input for BNN search. BNN search also supports the queries from the distributed server. Access the spatial database present in the distributed server. Based on the data structure present in the set of queries, the BNN search process will be done.

The marginal weight of object computed for parallel BNN is derived based on the occurrence of event formed with objects adjacent to the situation / querying summit in requirements of occurrence of object being at the closest position and the objects remoteness to the querying point on the spatial database. All the adjacent objects are detached with the projected marginal weight and ranking is accomplished based on the chronological order. With the rank selected for every object for NN search, the NN object is standard using parallel BNN without any uncertainty.

4. Results

The experiments of the proposed normalizing regions of data in spatial data grid [NRDSD] using parallel BNN with MOW is evaluated with the spatial data sets time series obtained from UCI repository. The experiment is implemented in Java 1.6 SDK and core java concept with over 1200 instances of spatial dataset. We ran our experiments with various data sets obtained from UCI repository. The normalizing regions of data in spatial data grid [NRDSD] using parallel BNN with MOW efficiently designed to retrieve the best NN object from spatial database based on specifying the threshold limit. Then query filtering technique filters the queries for parallel BNN search. to identify the NN object efficiently. Parallel BNN search used MOW ranking scheme to recognize the best NN object from the set of unique objects. Based on frequency and distance of object, MOW is assigned. The MOW evaluated for the nearest objects of the reference points are points out its simplicity of the nearest neighbor using parallel BNN amongst all the accessible objects attributes in the spatial data sets. The performance of the proposed normalizing regions of data in spatial data grid [NRDSD] using parallel BNN with MOW identified the best NN object based on grid formation is measured in terms of.

- Normalization of data objects
- Search path length
- Time consumption

Table 1: No. of data regions vs. Normalization

No. of data regions	Normalization	
	Proposed NRDSD	Existing BNN-MOW
2	24	12
4	36	20
6	50	29
8	62	34
10	75	45

The above table describes the effectiveness of the normalization based on number of data regions in the grid with respect to sparse and dense data set. The effectiveness of the proposed normalizing regions of data in spatial data grid [NRDSD] is compared with an existing BNN search using MOW in spatial database.

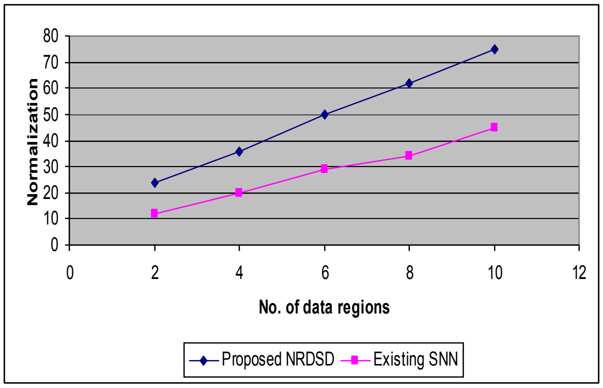

Figure 3. No. of data regions vs. Normalization

Figure 3 describes the effectiveness of normalization based on number of data regions in the grid with respect to sparse and dense data set. The dataset normalization refers to the efficiency of organizing the database in terms of collection of data objects in the data regions. In the proposed NRDSD, the formation of grid is done based on the size of the grid present with respect to threshold limit. The threshold limit is evaluated based on the minimum and maximum distance done based on the data regions. Since the data regions have been normalized based on the threshold limit, the effectiveness on BNN search for to identify uncertainty in database is identified. The effectiveness of normalization is measured in terms of how efficient the data objects are aligned and processed. Compared to an existing BNN search and MOW in spatial database, the proposed NRDSD provides better terms of normalization efficiency and produced good results and the variance is 30-40% high.

Table 2: No. of data objects vs. BNN search path length

No. of data objects	BNN search path length			
	Proposed NRDSD	Query filter	BNN	MOW
25	5	10	15	20
50	12	18	20	30
75	18	29	26	34
100	20	26	34	45
125	30	33	45	56

The above table describes the BNN search path length based on the number of data objects present in the grid. The effectiveness of the proposed normalizing regions of data in spatial data grid [NRDSD] is compared with an existing BNN search and MOW in spatial database with respect to query filtering scheme.

Figure 4. describes the BNN search path length based on number of data objects present in the grid. The objective of identifying the NN object in the spatial database with respect to the data objects should consume less interval of time and identified the search path length efficiently. The search path length is measured in terms of how much time the proposed NRDSD consumes to identify the NN object in the spatial database. The search path length using NN is considerably low in the proposed NRDSD, since it form a grid based on the size of the data objects. The size of the grid is processed to normalize the data regions. Compared to an existing BNN search and MOW in spatial database, the proposed NRDSD consumes less search path and the variance is 35-45% low.

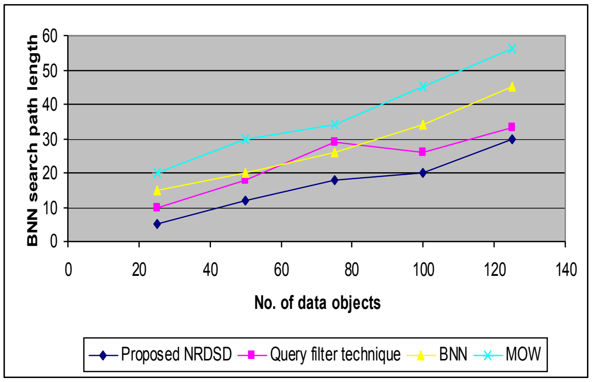

Figure 4: No. of data objects vs. BNN search path Length

Table 3: No. of data objects vs. time consumption to identify NN

No. of data objects	Time consumption to identify NN			
	Proposed NRDSD	Query filter	BNN	MOW
10	2.2	3.8	4	6
20	4.5	5.1	5.6	6.2
30	3.4	4.3	4.9	7.3
40	3.9	5.6	5.2	9.7
50	4.2	5.3	6.8	8.4

The above table describes the execution time taken for performing BNN search simultaneously for all requested queries based on the number of data objects present in the grid. The effectiveness of the proposed normalizing regions of data in spatial data grid [NRDSD] is compared with an existing BNN search and MOW in spatial database with respect to query filtering scheme.

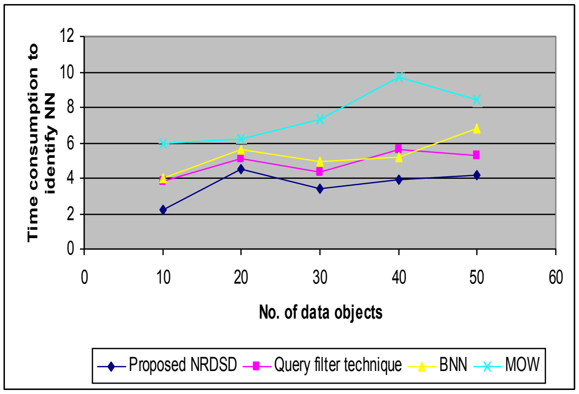

Figure 5: No. of data objects vs. time consumption to identify NN

Figure 5. depicts the execution time taken to search the NN object for all the requested queries based on number of objects present in the grid. In the proposed NRDSD, the grid formation is done based on the number of data objects with respect to size of the data object. The proposed NRDSD normalized the data regions to enhance the NN search in the spatial database. An execution time is measured in terms of seconds. Compared to an existing BNN search, MOW ranking scheme, the proposed normalizing regions of data in spatial data grid [NRDSD] in spatial database where each and every query has been processed in a short interval of time to identify the NN object and the data objects are not aligned well with the data regions and produce the outcome for all the queries in a short interval of time and variance is 44-55% low in the proposed NRDSD.

5. Discussion

In this work, we have seen how the objects in the spatial dataset are normalized based on the threshold limit. Then the query filtering process is done and the nearest neighbor object is efficiently identified based on ranking scheme using the query filter based parallel BNN search using MOW ranking scheme with an existing BNN search using MOW in spatial database written in mainstream languages such as Java. We used different sets of semantic data for comparing the results of the proposed normalizing regions of data in spatial data grid [NRDSD] using parallel BNN with MOW with an existing BNN search using MOW in spatial database. The table and graph describes the performance of the proposed normalizing regions of data in spatial data grid [NRDSD] using parallel BNN with MOW.

Finally, it is being observed that the proposed NRDSD provides a better grid formation by assigning the threshold limit. After normalizing the data objects in the grid, the spatial database is processed with the query filtering technique used to filter the queries based on the similarity and parallel BNN search has also been applied.

6. Conclusion

In this work, we efficiently performed the NN search in spatial database from distributed server by professionally introducing the proposed normalizing regions of data in spatial data grid [NRDSD] in spatial database. The data regions are normalized based on the data objects present in the spatial database. Normalizing the data regions is done with respect to threshold limit based on minimum and maximum distance of the data objects. Then query filtering technique is applied to filter the query at first based on its similarity of the constraints. After filtering process, parallel BNN search is performed with MOW ranking scheme identify the NN object and rank the objects in spatial database accurately under a limited memory. The proposed normalizing regions of data in spatial data grid supports high dimensional data structure and accessed the users' queries from distributed server. The proposed normalizing regions of data in spatial data grid eliminated the servers' data set which does not provide the result accuracy. The experimental results showed that the proposed normalizing regions of data in spatial data grid worked efficiently by improving the reliability of the NN search with less overhead and less execution time.

References

1. Andreas Krüger ET.AL., "Mapping Spatial Data Infrastructures To A Grid Environment For Optimised Processing Of Large Amounts Of Spatial Data", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B4. Beijing 2008.
2. David R. Parker, et.al., "Development of a Bayesian Framework for Determining Uncertainty in Receiver Operating Characteristic Curve Estimates", IEEE transactions on knowledge and data engineering, vol. 22, no. 1, january 2010.
3. Hu, L. ET. AL., "Spatial Query Integrity with Voronoi Neighbors", IEEE transactions on knowledge and data engineering, 2011.
4. Jianjun Cui et. Al., "Design and Implementation of WSRF-Based GIS Service in Spatial Data Grid", Eighth International Conference on Grid and Cooperative Computing, 2009. GCC '09.
5. Kyriakos Mouratidis and Man Lung Yiu " Anonymous Query Processing "ieee transactions on knowledge and data engineering, vol. 22, no. 1, january 2010.
6. Lingkui Meng et. Al., "Spatial Data Grid Framework Based on Super-Peer Network Model", ISECS International Colloquium on Computing, Communication, Control, and Management, 2008. CCCM '08.
7. Liu Hua et. al., "Large Volume Spatial Data Management Based on Grid Computing", Supported by the National Key Basic Research and Development Program of China 2005 IEEE.
8. Ping Zhang et. Al., "A replication strategy based on swarm intelligence in spatial datagrid", 18th International Conference on Geoinformatics, 2010 .
9. Qidong Yin et. Al., "Research on Spatial Data Interoperability in Spatial Data Grid Environment", Third International Joint

Conference on Computational Science and Optimization (CSO), 2010.
10. Safavi-Naini, R. et. Al., “Information Theoretic Bounds on Authentication Systems in Query Model”, IEEE transactions on Information theory, 2008, DOI: 10.1109/TIT.2008.921683
11. Xiaobo Guan et. Al., “A Design of the Architecture of Geospatial Information Collaborative Service in Spatial Data Grid Environment”, Seventh International Conference on Grid and Cooperative Computing, 2008.
12. Yang Y, S. Papadopoulos, D. Papadias, et. Al., “Authenticated Multi-step NN search”, IEEE transactions on knowledge and data engineering, 2010.

6/25/2019