ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

Multi-Keyword Ranked Search for Spatial Network using Big Data

L.Cinthia Beatrice[1], K.Ravikumar[2]

Post Graduate Student, Dept. of CSE, Rrase College of Engineering, Chennai, Tamilnadu, India[1]

Professor, Dept. of CSE, Rrase College of Engineering, Chennai, Tamilnadu, India[2]

ABSTRACT:The amount of web-based information available has increased dramatically. Current web information gathering systems attempt to satisfy user requirements by capturing their information needs. For this purpose, profiles are created. Profiles represent the concept models possessed by users when gathering web information. Using Cloud Computing customers can remotely outsource their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. Its economic saving is inspirational to both individuals and enterprises to outsource their local complex data management system into the cloud. To provide data privacy and combat unauthorised accesses in the cloud and beyond, sensitive data may have to be encrypted by the owners before outsourcing to the commercial public cloud

KEYWORDS: Approximate String Search, Multi-Keyword Ranked Search over Encrypted cloud data, Range Query, Ranked Search, Spatial Databases.

I.INTRODUCTION

Primary objectives of utilising a spatial database is that they provide optimized storage and permit to query data that represents objects defined in a geometric space.A model is implicitly possessed by users and is generated from their background knowledge. While this model cannot be proven in laboratories, many web ontologists have observed it in user behaviour. When users read through a document, they can easily determine whether or not it is of their interest or relevance to them, a judgment that arises from their implicit concept models. If a user's concept model can be simulated, then a superior representation of user profiles can be built. To simulate user concept models, ontologies and formalization model are utilized in personalized web information gathering. Such ontologies are called ontological user profiles or personalized ontologies. To represent user profiles, many researchers have attempted to discover user background knowledge through global or local analysis. Global analysis uses existing global knowledge bases for user background knowledge representation. Commonly used knowledge bases include generic ontologies, thesauruses and online knowledge bases. The global analysis techniques produce effective Performance for user background knowledge extraction and for approximate string searches in Spatial databases

Local analysis investigates user local information or observes user behaviour in user profiles. However, because local analysis techniques rely on data mining or classification techniques for knowledge discovery, occasionally the discovered results contain noisy and uncertain information. As a result, local analysis suffers from ineffectiveness at capturing formal user knowledge. From this, we can hypothesize that user background Knowledge can be better discovered and represented if we can integrate global and local analysis within a hybrid model. The knowledge formalized in a global knowledge base will constrain the background knowledge discovery from the user local information. Such a personalized spatial model should produce a superior representation of user profiles for web information gathering. In this paper, a spatial model to evaluate this hypothesis is proposed. This model simulates users' concept models by using personalized ontologies and attempts to improve web information gathering performance by using ontological user profiles. The world knowledge and a user's local instance repository (LIR) are used in the proposed model. World knowledge is the knowledge acquired by people from experience and education an LIR is a user's personal collection of information items. The proposed spatial model is evaluated by comparison against some benchmark models through experiments using a large standard data set. The evaluation results show that the proposed spatial model is successful

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

II. RELATED WORK

In the paper 'Planar formulae and their uses' MicroRNAs are a class of small noncoding RNAs that are abnormally expressed in different cancer cells. Molecular signature of miRNAs in different malignancies suggests that these are not only actively involved in the pathogenesis of human cancer but also have a significant role in patients survival. The differential expression patterns of specific miRNAs in a specific cancer tissue type have been reported in hundreds of research articles. However limited attempt has been made to collate this multitude of information and obtain a global perspective of miRNA dysregulation in multiple cancer types.

In the paper 'Selectivity estimation in spatial databases' A prospective high-throughput gene expression study was conducted to identify transcriptional profiles predictive of a clinical response to cisplatin and fluorouracil (CF) combination chemotherapy and to identify deregulated genes associated with acquired resistance to CF. Methods: Endoscopic biopsy samples were collected from CF-treated metastatic gastric cancer (MGC) patients (pts) prior to CF (n = 123) and following the development of resistance (n = 22) at the National Cancer Centre of Korea from 2001 to 2006, and analysed using CGH and expression microarrays. I developed 2 survival risk predictors. The first predictor was constructed using genes in DNA amplicons and identified in the expression signature that correlates with survival (intrinsic resistance signature). The second predictor was based on the acquired resistance signature, which was identified by comparing matched expression array data from initially responsive patients prior to treatment with profiles obtained at progressive disease

III.PROPOSED WORK

I demonstrate the efficiency and effectiveness of proposed methods for Spatial Approximate String (SAS) queries using a comprehensive experimental evaluation. The world knowledge and a user's local instance repository (LIR) are used in the proposed model.

1) World knowledge is commonsense knowledge acquired by people from experience and education
2) An LIR is a user's personal collection of information items. From a world knowledge base, we construct personalized ontologies by adopting user feedback on interesting knowledge. The users' LIRs are then used to discover background knowledge and to populate the personalized ontologies.

Spatial mining discovers interesting and on-topic knowledge from the concepts, semantic relations, and instances in a spatial. In this paper, a 2D spatial mining method is introduced: Specificity and Exhaustivity. Specificity (denoted spe) describes a subject's focus on a given topic. Exhaustivity restricts a subject's semantic space dealing with the topic. This method aims to investigate the subjects and the strength of their associations in a spatial.

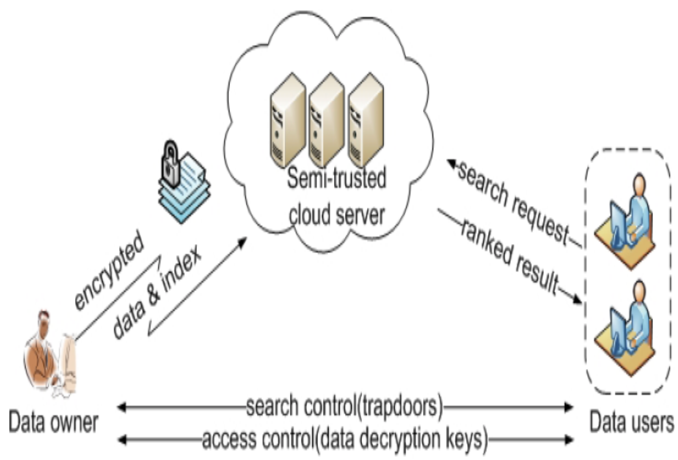

Fig 1. Architecture of the Mining Model

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

These concepts are discussed in the following sections. Fig 1 shows the architecture of the Mining Model.

DATA USER: This module includes the user registration login details. An optional Data User, who has expertise and capabilities that users may not have, is trusted to assess and expose risk of cloud storage services on behalf of the users upon request.

The Authentication Server (AS) functions as any AS would with a few additional behaviours added to the typical client-authentication protocol. The first addition is the sending of the client authentication information to the masquerading router. The AS in this model also functions as a ticketing authority, controlling permissions on the application network. The other optional function that should be supported by the AS is the updating of client lists, causing a reduction in authentication time or even the removal of the client as a valid client depending upon the request. One of the key issues is to effectively detect any unauthorized data modification and corruption, possibly due to server compromise and/or random Byzantine failures. Besides, in the distributed case when such inconsistencies are successfully detected, to find which server the data error lies in is also of great significance.

DATA OWNER: This module helps the owner to register those details and also include login details. Admin Initializing the cloud server to check the network connection from cloud server and user. The connections are successfully, and then the processes are executed. A Cloud Service Provider, who has significant resources and expertise in building and managing distributed cloud storage servers, owns and operates live Cloud Computing systems.

Cloud data storage, a user stores his data through a CSP into a set of cloud servers, which are running in a simultaneous, the user interacts with the cloud servers via CSP to access or retrieve his data. In some cases, the user may need to perform block level operations on his data. Users should be equipped with security means so that they can make continuous correctness assurance of their stored data even without the existence of local copies.

In case those users do not necessarily have the time, feasibility or resources to monitor their data, they can delegate the tasks to an optional trusted TPA of their respective choices. In our model, we assume that the point-to-point communication channels between each cloud server and the user is authenticated and reliable, which can be achieved in practice with little overhead.

FILE UPLOAD: Cloud data storage, an Admin can stores his data through a CSP into a set of cloud servers, which are running in a simultaneous, the user interacts with the cloud servers via CSP to access or retrieve his data. The Admin encrypts the data's, and then stored the cloud server. Cloud server is considered as "honest-but-curious" in our model, which is consistent with the most related works on searchable encryption. The cloud server collects the some different encrypted documents. In this process encrypted data to store with cloud server. Search result should be ranked by the cloud server according to ranking criteria.

In some cases, the user may need to perform block level operations on his data. Users should be equipped with security means so that they can make continuous correctness assurance of their stored data even without the existence of local copies. In case those users do not necessarily have the time, feasibility or resources to monitor their data, they can delegate the tasks to an optional trusted TPA of their respective choices. In our model, we assume that the point-to-point communication channels between each cloud server and the user is authenticated and reliable, which can be achieved in practice with little overhead.

RANK SEARCH:A ranked searchable encryption scheme (RSSE) consists of four algorithms like KeyGen, BuildIndex, TrapdoorGen, SearchIndex. The RSSE Framework can be categorized into the following two phases, Setup and Retrieval. The Setup phase can be done by the Admin and the Retrieval can be done by the User. While uploading the Files in the cloud server RSSE Framework generate unique Index for each and every files. On the one hand, to meet the effective data retrieval need, large amount of documents demand cloud server to perform result relevance ranking, instead of returning undifferentiated result.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data. And it automatically generates Rank for the Files based on the same criteria done by the user side. All these are implemented with the help of OPSE (Order Preserving Symmetric Encryption) Mapping. It is used for the Clustering of all the Cloud Files. And also helps to retrieve the files efficiently.Few of the Big Data uses have been listed below:
Classification:This means getting to know your data. If you can categorize, classify, and/or codify your data, you can place it into chunks that are manageable by a human. Rather than dealing with 3.5 million merchants at a credit card company, if we could classify them into 100 or 150 different classifications that were virtually dead on for each merchant, a few employees could manage the relationships rather than needing a sales and service force to deal with each customer individually. Likewise, at a university, if an alumni group treats its donors according to their classifications, part-time students might be the representatives who work with minor donors and full-time professionals might receive incoming calls from the donors whose names appear on buildings on campus.

Estimation:This process is useful in just about every facet of business. From finance to marketing to Sales, the better you can estimate your expenses, product mix optimization, or potential customer value, the better off you will be. This and the next use are fairly self-evident if you have ever spent a day at a business.

Prediction: Forecasting, like estimation, is ubiquitous in business. Accurate prediction can reduce Inventory levels (costs), optimize sales, etc...

Affinity Grouping/Market Basket Analysis:This is a use that marketing loves. Product placement within a store can be set up based on sales maximization when you know what people buy together. There are several schools of thought on how to do it. For example, you know people buy paint and paint brushes together. One, do you make a sale on paint then jack up the prices on brushes, two do you put the paint in aisle 1 and the brushes in aisle 7 hoping that people walking from one to the other will see something else they will need, three do you set cheap stuff on the end of the aisle for everyone to see hoping they will buy it on impulse knowing they will need something else with that impulse buy (chips and dip, charcoal briquettes and lighter fluid, etc.). As you can see, knowing what people buy together has serious benefits for the retail world.

Clustering/Target Marketing: Target marketing saves millions of dollars in wasted coupons, promotions. If you send your promo to only the most likely to accept the offer, use the coupon, or buy your product, you will be much better served. If you sell acne medication, sending coupons to people over sixty is usually a waste of your marketing dollars. If, however, you can cluster your customers and know which households have a 75% chance of having a teenager, you are pushing your marketing on a group most likely to buy your product.

IV.EXPERIMENT RESULTS

Data set: I demonstrate the efficiency and effectiveness of our proposed methods for SAS queries using a comprehensive experimental evaluation. For ESAS queries, our experimental evaluation covers both synthetic and real data sets. For RSAS queries, our evaluation is based on two large, real road network datasets. In both cases, my methods have significantly outperformed the respective baseline methods.

Database Fetch & Mining: User can login to the Ontology Learning Environment in three modes: Data User (or) Data Owner (or) Admin. Admin will have the privilege to View Users, Upload Files to the Cloud, View Personalized Files. Guest users can search for a topic; however they are not authorized to read through the retrieved file content. Authorized Data Users can enter a topic and read through the fetched results. Based on the file content these users can rate the topic as relevant or irrelevant.

Consider the following example to differentiate between Positive and Negative subjects. A student enters a room full of chocolates, and I question this student to tell me the number of White Papers available in the room. Positive Subject - Number of white papers. Negative Subject - Number of colour papers. It is deemed that the Negative Subject will cover ambiguous interpretation of the question in place.

To meet the effective data retrieval need, the large amount of documents demand the cloud server to perform result relevance ranking, instead of returning undifferentiated results. Such ranked search system enables data users to find the most relevant information quickly, rather than burdensomely sorting through every match in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the "pay-as-you use" cloud paradigm.

For privacy protection, such ranking operation, however, should not leak any keyword related information. On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse results. As a common practice indicated by today's web search engines, data users may tend to provide a set of keywords instead of only one as the indicator of their search interest to retrieve the most relevant data. And each keyword in the search request is able to help narrow down the search result further. "Coordinate matching", i.e., as many matches as possible, is an efficient similarity measure among such multi-keyword semantics to refine the result relevance, and has been widely used in the plaintext information retrieval community. However, how to apply it in the encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirements like the data privacy, the index privacy, the keyword privacy, and many others.

Data retrieval will happen through Big Data and authentication services will be handled by Azure cloud services. Usage of Big Data over conventional database helps in statistical analysis of huge data.

V.PROBLEM FORMULATION

Considering a cloud data hosting service, this involves three different entities: the data owner, the data user, and the cloud server. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C. To enable the searching capability over C for effective data utilization, the data owner, before outsourcing, will first build an encrypted searchable index I from F, and then outsource both the index I and the encrypted document collection C to the cloud server. To search the document collection for t given keywords, an authorized user acquires a corresponding trapdoor T through search control mechanisms, e.g., broadcast encryption. Upon receiving T from a data user, the cloud server is responsible to search the index I and return the corresponding set of encrypted documents. To improve the document retrieval accuracy, the search result should be ranked by the cloud server according to some ranking criteria (e.g., coordinate matching, as will be introduced shortly). Moreover, to reduce the communication cost, the data user may send an optional number k along with the trapdoor T so that the cloud server only sends back top-k documents that are most relevant to the search query. Finally, the access control mechanism is employed to manage decryption capabilities given to users.

VI.CONCLUSION

I perform a ranked search of multi-keyword over encrypted cloud data. Among several multi-keyword semantics, I choose the efficient similarity measure of "coordinate matching", i.e., as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use "inner product similarity" to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, I propose a basic idea ofMulti-Keyword Ranked Search over Encrypted cloud data (MRSE)using secure inner product computation. Finally the Graphical user interface is designed for user interaction where the user can fill the query in the form of keywords and find the desired result in integrated form.

REFERENCES

[1] C. Li, J. Lu, and Y. Lu. "Efficient merging and filtering algorithms for approximate string searches." In ICDE, pages 257–266, 2008.
[2] D. Lichtenstein. "Planar formulae and their uses." SIAM J. Comput., pages 329–343, 1982.
[3] K. Yi, X. Lian, F. Li, and L. Chen. "The world in a nutshell: Concise range queries." TKDE, 23:139–154, 2011.
[4] S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS,January 2010, LNCS. Springer, Heidelberg.
[5] A. Singhal, "Modern information retrieval: A brief overview," IEEEData Engineering Bulletin, vol. 24, no. 4, pp. 35–43, 2001.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

[6] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressingand indexing documents and images," Morgan Kaufmann Publishing,San Francisco, May 1999.
[7] S. Acharya, V. Poosala, and S. Ramaswamy. "Selectivity estimation in spatial databases." In SIGMOD, pages 13–24, 1999
[8] S. Alsubaiee, A. Behm, and C. Li. "Supporting location-based approximate-keyword queries." In GIS, pages 61–70, 2010.
[9] Arasu, S. Chaudhuri, K. Ganjam, and R. Kaushik. "Incorporating string transformations in record matching." In SIGMOD, pages 1231– 1234, 2008.
[10] Arasu, V. Ganti, and R. Kaushik. "Efficient exact set-similarity joins." In VLDB, pages 918–929, 2006

BIOGRAPHY

Cinthia Beatrice. L is a Post Graduate student in the Department of Computer Science and Engineering, Rrase College of Engineering, Anna University. She received Bachelor of Engineering degree (B.E) in 2010from St. Joseph's College of Engineering, Sholinganalur, Chennai, India. Her research interests are Networking,Cloud Computing, Information Security, Big Data, Database Management Systems, etc..