International Journal of Research
Available at https://pen2print.org/index.php/ijr/
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 06 Issue 1
January 2019

An Enhanced and Reliable Efficient Algorithm for Mining Authentication Utility Item Sets

Odugu swathi & P.Vijay Babu
[1] PG Scholar Department of CSE, GVR & S College Of Engineering & Technology, Guntur (D.T),Andhra Pradesh
[2]Associate Professor, Department of CSE ,GVR & S College Of Engineering & Technology, Guntur (D.T),Andhra Pradesh

Abstract:-

High utility itemsets (HUIs) mining is a developing theme in information mining, which alludes to finding all itemsets having an utility gathering a client determined least utility edge min_util. In any case, setting min_util fittingly is a troublesome issue for clients. As a rule, finding a suitable least utility limit by experimentation is a dreary procedure for clients. On the off chance that min_util is set too low, such a large number of HUIs will be created, which may cause the mining procedure to be exceptionally wasteful. Then again, if min_util is set too high, all things considered, no HUIs will be found. we propose a strategy and framework for confirming messages is given. A message validation framework creates irregular string and that will be sent to the beneficiary's portable and message sent to the beneficiary by means of mail at the opposite side the collector got the message through the mail and that will be in encoded shape and the beneficiary will decode the message with key that is sent to his versatile. The message validation framework at that point decides if the recovered message coordinates the first message. In the event that the codes coordinate, the uprightness and validness of the message are confirmed.

Key Words:- Utility mining, high utility item set, frequent item set mining, top-k pattern mining

I.INTRODUCTION

Consider a cloud-based healthcare information system that hosts outsourced personal health records (PHRs) from various healthcare providers. The PHRs are encrypted in order to comply with privacy regulations like HIPAA. In order to facilitate data use and sharing, it is highly desirable to have a searchable encryption (SE) scheme which allows the cloud service provider to search over encrypted PHRs on behalf of the authorized users (such as medical researchers or doctors) without learning information about the underlying plaintext. Note that the context we are considering supports private data sharing among multiple data providers and multiple data users. Therefore, SE schemes in the private-key setting [1], [2], [3], which assume that a single user who searches and retrieves his/her own data, are not suitable. On the other hand, private information retrieval (PIR) protocols [4], [5], [6], which allow users to retrieve a certain data-item from a database which publicly stores data without revealing the data-item to the database administrator, are also not suitable, since they require the data to be publicly available. In order to tackle the keyword search problem in the cloud-based healthcare information system scenario, we resort to public-key encryption with keyword search (PEKS) schemes, which is firstly proposed in [7]. In a PEKS scheme, a cipher text of the keywords called "PEKS cipher text" is appended to an encrypted PHR. To retrieve all the encrypted PHRs containing a keyword, say "Diabetes", a user sends a "trapdoor" associated with a look inquiry on the catchphrase "Diabetes" to the cloud benefit supplier, which chooses all the encoded PHRs containing the watchword "Diabetes" and returns them to the client while without taking in the hidden PHRs. Be that as it may, the arrangement in [7] and also other existing PEKS plans which enhance [7] just help balance inquiries [8]. Set crossing point and meta keywords1 [9], [10] can be utilized for conjunctive catchphrase seek. In any case, the methodology in light of set crossing point releases additional data to the cloud server past the aftereffects of the conjunctive inquiry, while the approach utilizing meta catchphrases require 2m meta watchwords to oblige all the conceivable conjunctive inquiries for m catch phrases. With the end goal to address the above insufficiencies in conjunctive catchphrase look, plans, for example, the ones in [11], [12] were advanced in general society key setting.

II.LITERATURE SURVEY

1) Retrieving top-k prestige-based relevant spatial web objects,

AUTHORS: X. Cao, G. Cong, and C. Jensen

The location-aware keyword query returns ranked objects that are near a query location and that have textual descriptions that match query keywords. This query occurs inherently in many types of mobile and traditional web services and applications, e.g., Yellow Pages and Maps services. Previous work considers the potential results of such a query as being independent when ranking them. However, a relevant result object with nearby objects that are also relevant to the query is likely to be preferable over a relevant object without relevant nearby objects. The paper proposes the concept of prestige-based relevance to capture both the textual relevance of an object to a query and the effects of nearby objects. Based on this, a new type of query, the Location-aware top-k Prestige-based Text retrieval (LkPT) query, is proposed that retrieves the top-k spatial web objects ranked according to both prestige-based relevance and location proximity.We propose two algorithms that compute LkPT queries. Empirical studies with real-world spatial data demonstrate that LkPT queries are more effective in retrieving web objects than a previous approach that does not consider the effects of nearby objects; and they show that the proposed algorithms are scalable and outperform a baseline approach significantly.

2) Efficient retrieval of the top-k most relevant spatial web objects

AUTHORS: G. Cong, C. Jensen, and D. Wu

The conventional Internet is acquiring a geo-spatial dimension. Web documents are being geo-tagged, and geo-referenced objects such as points of interest are being associated with descriptive text documents. The resulting fusion of geo-location and documents enables a new kind of top-k query that takes into account both location proximity and text relevancy. To our knowledge, only naive techniques exist that are capable of computing a general web information retrieval query while also taking location into account. This paper proposes a new indexing framework for location-aware top-k text retrieval. The framework leverages the inverted file for text retrieval and the R-tree for spatial proximity querying. Several indexing approaches are explored within the framework. The framework encompasses algorithms that utilize the proposed indexes for computing the top-k query, thus taking into account both text relevancy and location proximity to prune the search space. Results of empirical studies with an implementation of the framework demonstrate that the paper's proposal offers scalability and is capable of excellent performance.

3) Location-aware type ahead search on spatial databases: Semantics and efficiency

AUTHORS: S. B. Roy and K. Chakrabarti
Users often search spatial databases like yellow page data using keywords to find businesses near their current loca- tion. Such searches are increasingly being performed from mobile devices. Typing the entire query is cumbersome and prone to errors, especially from mobile phones. We address this problem by introducing type-ahead search functional- ity on spatial databases. Like keyword search on spatial data, type-ahead search needs to be location-aware, i.e., with every letter being typed, it needs to return spatial ob- jects whose names (or descriptions) are valid completions of the query string typed so far, and which rank highest in terms of proximity to the user's location and other static scores. Existing solutions for type-ahead search cannot be used directly as they are not location-aware. We show that a straight-forward combination of existing techniques for per- forming type-ahead search with those for performing prox- imity search perform poorly. We propose a formal model for query processing cost and develop novel techniques that optimize that cost. Our empirical evaluations on real and synthetic datasets demonstrate the effectiveness of our tech- niques. To the best of our knowledge, this is the first work on location-aware type-ahead search.

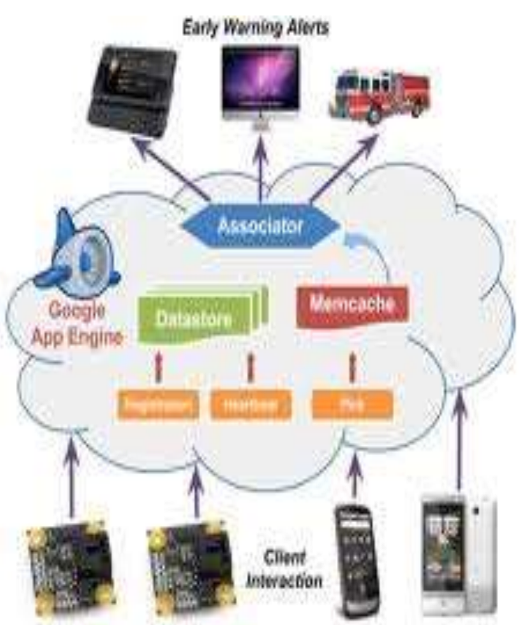
4) Processing and optimization of multiway spatial joins using r-trees
AUTHORS: D. Papadias, N. Mamoulis, and Y. Theodoridis
One of the most important types of query processing in spatial databases and geographic information systems is the spatial join, an operation that selects, from two relations, all object pairs satisfying some spatial predicate. A multiway join combines data originated from more than two relations. Although several techniques have been proposed for pairwise spatial joins, only limited work has focused on multiway spatial join processing. This paper solves multiway spatial joins by applying systematic search algorithms that exploit R-trees to efficiently guide search, without building temporary indexes or materializing intermediate results. In addition to general methodologies, we propose cost models and an optimization algorithm, and evaluate them through extensive experimentation.

5) Fast algorithms for mining association rules in large databases

International Journal of Research
Available at https://pen2print.org/index.php/ijr/
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 06 Issue 1
January 2019

AUTHORS: R. Agrawal and R. Srikant
We consider the problem of discovering association rules between items in a large database of sales transactions. We present two new algorithms for solving this problem that are fundamentally di erent from the known algorithms. Experiments with synthetic as well as real-life data show that these algorithms outperform the known algorithms by factors ranging from three for small problems to more than an order of magnitude for large problems. We also show how the best features of the two proposed algorithms can be combined into a hybrid algorithm, called AprioriHybrid. Scale-up experiments show that AprioriHybrid scales linearly with the number of transactions. AprioriHybrid also has excellent scale-up properties with respect to the transaction size and the number of items in the database.

Architecture:-



III.Related Work:-

This subsection introduces related works about top-*k* high utility itemset mining, including *high utility itemset mining*, *top-k frequent pattern mining* and *top-k high utility itemset mining*.
High Utility Itemset Mining
In recent years, high utility itemset mining has received lots of attention and many efficient algorithms have been proposed, such as *Two-Phase* [13], *IHUP* [2], *IIDS* [17], *UP-Growth* [25], *d2HUP* [15] and *HUI-Miner* [14]. These algo-rithms can be generally categorized into two types: *two-phase* and *one-phase* algorithms. The main characteristic of two-phase algorithms is that they consist of two phases. In the first phase, they generate a set of candidates that are potential high utility itemsets. In the second phase, they calculate the exact utility of each candidate found in the first phase to identify high utility itemsets. Two-Phase, IHUP, IIDS and UP-Growth are two-phase based algo-rithms. UP-Growth is one of the state-of-the-art two-phase algorithms, which incorporates four effective strategies *DGU*, *DGN*, *DLU* and *DLN* for pruning candidates in the first phase. One the contrary, the main characteristic of one-phase algorithms is that they discover high utility itemsets using only one phase and produce no candidates. *d2HUP* and *HUI-Miner* are one-phase algorithms. *d2HUP* transforms a horizontal database into a tree-based structure called *CAUL* [15] and adopts a pattern-growth strategy to direct-ly discover high utility itemsets in databases. HUI-Miner considers a database of vertical format and transforms it into *utility-lists* [14]. The utility-list structure used in HUI-Miner allows directly computing the utility of generated itemsets in main memory without scanning the

original database. Although the above studies may perform well in some applications, they are not developed for top-*k* high utility itemset mining and still suffer from the subtle problem of setting appropriate thresholds.

Top-*k* Pattern Mining

Many studies have been proposed to mine different kinds of top-*k* patterns, such as *top-k frequent itemsets* [3, 19, 20], *top-k frequent closed itemsets* [3, 28], *top-k closed sequen-tial patterns* [24], *top-k association rules* [6], *top-k sequential rules* [5], *top-k correlation patterns* [31, 32, 33] and *top-k cosine similarity interesting pairs* [38]. What distinguishes each top-*k* pattern mining algorithm is the type of patterns dis-covered, as well as the data structures and search strate-gies that are employed. For example, some algorithms [5, 6] use a rule expansion strategy for finding patterns, while others rely on a pattern-growth search using structures such as FP-Tree [19, 20, 28]. The choice of data structures and search strategy affect the efficiency of a top-*k* pattern mining algorithm in terms of both memory and execution time. However, the above algorithms discover top-*k* pat-terns according to traditional measures instead of the utili-ty measure. As a consequence, they may miss patterns yielding high utility.

IV.CONCLUSION

In this paper, we have considered the issue of best k high utility itemsets mining, where k is the coveted number of high utility itemsets to be mined. Two proficient algorithms TKU (mining Top-K Utility itemsets) and TKO (min-ing Top-K utility itemsets in One stage) are proposed for mining such itemsets without setting least utility edges. TKU is the initial two-stage calculation for min-ing top-k high utility itemsets, which fuses five systems PE, NU, MD, MC and SE to viably raise the outskirt least utility edges and further prune the pursuit space. Then again, TKO is the first stage calculation created for best k HUI mining, which incorporates the novel methodologies RUC, RUZ and EPB to significantly enhance its execution. Exact assessments on various sorts of genuine and manufactured datasets demonstrate that the proposed calculations have great adaptability on expansive datasets and the execution of the proposed calculations is near the ideal instance of the best in class two-stage and one-stage utility mining calculations [14, 25]. In spite of the fact that we have proposed another structure for best k HUI mining, it has not yet been consolidated with other utility mining assignments to find diverse kinds of best k high utility examples, for example, top-k high utility scenes, top-k closed+ high utility itemsets, top-k high utility web get to pat-terns and best k portable high utility consecutive examples. These leave wide spaces for investigation as future work.

V.BIBLIOGRAPHY

References Made From:-

[1] R. Agrawal and R. Srikant, ―Fast Algorithms for Mining Association Rules,‖ *n Proc. of Int'l Conf. on Very Large Data Bases*, pp. 487-499, 1994.

[2] C. Ahmed, S. Tanbeer, B. Jeong and Y. Lee, ―Efficient Tree Structures for High-utility Pattern Mining in Incremental Databases,‖ *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21(12), pp. 1708-1721, 2009.

[3] K. Chuang, J. Huang and M. Chen, ―Mining Top-K Frequent Patterns in the Pres-ence of the Memory Constraint,‖ *The VLDB Journal*, Vol. 17, pp. 1321-1344, 2008.
[4] R. Chan, Q. Yang and Y. Shen, ―Mining High-utility Itemsets,‖ *in Proc. of IEEE Int'l Conf. on Data Mining*, pp. 19-26, 2003.
[5] P. Fournier-Viger, V. S Tseng, ―Mining Top-K Sequential Rules,‖ *in Proc. of Int'l Conf. on Advanced Data Mining and Applications*, pp. 180-194, 2011.
[6] P. Fournier-Viger, C. Wu, V. S. Tseng, ―Mining Top-K Association Rules,‖ *in Proc. of Int'l Conf. on Canadian conference on Advances in Artificial Intelligence*, pp. 61–73, 2012.
[7] P. Fournier-Viger, C. Wu, V. S. Tseng, ―Novel Concise Representations of High Utility Itemsets Using Generator Patterns," *in Proc. of Int'l. Conf. on Advanced Data Mining and Applications and Lecture Notes in Computer Science*, Vol. 8933, pp. 30-43, 2014.
[8] J. Han, J. Pei and Y. Yin, ―Mining Frequent Patterns without Candidate Genera-tion,‖ *in Proc. of ACM SIGMOD Int'l Conf. on Management of Data*, pp. 1-12, 2000.
[9] J. Han, J. Wang, Y. Lu and P. Tzvetkov, ―Mining Top-K Frequent Closed Patterns without Minimum Support,‖ *in Proc. of IEEE Int'l Conf. on Data Mining*, pp. 211-218, 2002.
[10] S. Krishnamoorthy, ―Pruning Strategies for Mining High Utility Itemsets,‖ *Expert Systems with Applications*, pp. Vol. 42(5), pp. 2371-2381, 2015.
[11] C. Lin, T. Hong, G. Lan, J. Wong and W. Lin, ―Efficient Updating of Discovered High-utility Itemsets for Transaction Deletion in Dynamic Databases,‖ *Advanced Engineering Informatics*, Vol. 29(1), pp. 16-27, 2015.
[12] G. Lan, T. Hong, V. S. Tseng and S. Wang, ―Applying the Maximum Utility Meas-ure in High Utility Sequential Pattern Mining,‖ *Expert Systems with Applications*, Vol. 41(11), pp. 5071-5081, 2014.
[13] Y. Liu, W. Liao, and A. Choudhary, ―A Fast High Utility Itemsets Mining Algo-rithm,‖ *in Proc. of the Utility-Based Data Mining Workshop*, pp. 90-99, 2005.
[14] M. Liu and J. Qu, ―Mining High Utility Itemsets without Candidate Generation,‖ *in Proc. of ACM Int'l Conf. on Information and Knowledge Management*, pp. 55-64, 2012.
[15] J. Liu, K. Wang and B. Fung, ―Direct Discovery of High Utility Itemsets without Candidate Generation,‖ *in Proc. of IEEE Int'l Conf. on Data Mining*, pp. 984-989 , 2012.
[16] Y. Lin, C. Wu and V. S. Tseng, ―Mining High Utility Itemsets in Big Data,‖ *in Proc. of Int'l Conf. on Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 649-661, 2015.
[17] Y. Li, J. Yeh and C. Chang, ―Isolated Items Discarding Strategy for Discovering High-utility Itemsets,‖ *Data & Knowledge Engineering*, Vol. 64(1), pp. 198-217, 2008.
[18] J. Pisharath, Y. Liu, B. Ozisikyilmaz, R. Narayanan, W. K. Liao, A. Choudhary and G. Memik, NU-MineBench version 2.0 dataset and technical report, http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html
[19] G. Pyun and U. Yun, ―Mining Top-K Frequent Patterns with Combination Reduc-ing Techniques, ―*Applied Intelligence*, Vol. 41(1), pp. 76-98, 2014.
[20] T. Quang, S. Oyanagi, and K. Yamazaki, ― ExMiner: An Efficient Algorithm for Mining Top-K Frequent

Patterns,‖ *in Proc. of Int'l Conf. on Advanced Data Mining and Applications*, pp. 436 – 447, 2006.

[21] H. Ryang and U. Yun, ―Top-K High Utility Pattern Mining with Effective Thre-shold Raising Strategies,‖ Knowledge-Based Systems, Vol. 76, pp. 109-126, 2015.

[22] H. Ryang, U Yun and K. Ryu, ―Discovering High Utility Itemsets with Multiple Minimum Supports,‖ *Intelligent Data Analysis*, Vol. 18(6), pp. 1027-1047, 2014.

[23] B. Shie, H. Hsiao, V. S. Tseng and P. S. Yu, ―Mining High Utility Mobile Sequential Patterns in Mobile Commerce Environments,‖ *in Proc. of Int'l. Conf. on Database Sys-tems for Advanced Applications and Lecture Notes in Computer Science,* Vol. 6587, pp. 224-238, 2011.

[24] P. Tzvetkov, X. Yan and J. Han, ―TSP: Mining Top-K Closed Sequential Patterns,‖ *Knowledge and Information System*, Vol. 7(4), pp. 438-457, 2005.

[25] V. S. Tseng, C. Wu, B. Shie, and P. S. Yu, ―UP-Growth: An Efficient Algorithm for High Utility Itemset Mining,‖ *in Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 253–262, 2010.

[26] V. S. Tseng, C. Wu, P. Fournier-Viger, P. S. Yu, ―Efficient Algorithms for Mining the Concise and Lossless Representation of High Utility Itemsets,‖ *IEEE Transactions on Knowledge and Data Engineering*, Vol. 27(3), pp. 726-739, 2015.

[27] C. Wu, P. Fournier-Viger, P. S. Yu, and V. S. Tseng, ―Efficient Mining of a Concise and Lossless Representation of High Utility Itemsets,‖ *in Proc. of IEEE Int'l Conf. on Data Mining*, pp. 824-833, 2011.

[28] J. Wang and J. Han, ―TFP: An Efficient Algorithm for Mining Top-K Frequent Closed Itemsets,‖ *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17(5), pp. 652-664, 2005.

[29] C. Wu, Y. Lin, P. S. Yu and V. S. Tseng, ―Mining High Utility Episodes in Complex Event Sequences,‖ *in Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pp.536-544, 2013.

[30] C. Wu, B. Shie, V. S. Tseng and P. S. Yu, ―Mining Top-K High Utility Itemsets,‖ *in Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 78–86, 2012.

[31] H. Xiong, M. Brodie and S. Ma, ―TOP-COP: Mining TOP-K Strongly Correlated Pairs in Large Databases,‖ *in Proc. of IEEE Int'l Conf. on Data Mining*, pp. 1162-1166, 2006.

[32] H. Xiong, P. Tan, V. Kumar, ―Mining Strong Affinity Association Patterns in Data Sets with Skewed Support Distribution,‖ *in Proc. of IEEE Int'l Conf. on Data Mining*, pp. 387-394, 2003.

[33] H. Xiong, P. Tan, V. Kumar, ―Hyperclique Pattern Discovery,‖ *Data Mining and Knowledge Discovery*, Vol. 13(2), pp. 219-242, 2006.

[34] U. Yun, J. Kim, ―A Fast Perturbation Algorithm using Tree Structure for Privacy Preserving Utility Mining,‖ *Expert Systems with Applications*, Vol. 42(3), pp. 1149-1165, 2015.

[35] U. Yun, H. Ryang, ―Incremental High Utility Pattern Mining with Static and Dy-namic Databases,‖ *Applied Intelligence*, Vol. 42(2), pp. 323-352, 2015.

[36] J. Yin, Z. Zheng, L. Cao, Y. Song and W. Wei, ―Mining Top-K High Utility Sequen-tial Patterns,‖ *in Proc. of IEEE Int'l Conf. on Data Mining*, pp. 1259-1264, 2013.

[37] M. Zihayat and A. An, ―Mining Top-K High Utility Itemsets over Data Streams,‖

Information Sciences, Vol. 285 (20), pp. 138–161, 2014.
[38] S. Zhu, J. Wu, H. Xiong and G. Xia, ―Scaling Up Top-K Cosine Similarity Search,‖ Data & Knowledge Engineering, Vol. 70(1), pp. 60–83, 2011.
[39] Frequent Itemset Mining Implementations Repository, http://fimi.cs.helsinki.fi/
[40] FoodMart2000, Microsoft Developer Network (MSDN), https://technet.microsoft.com/en-us/library/aa217032(v=sql.80).aspx