IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

I-INJECTION TOWARDS EFFECTIVE COLLABORATIVE FILTERING BY USING TKU AND TKO

SATTI RADHA [#1], K.VENKATESH [#2]

[#1] MCA Student, Master of Computer Applications,

D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.

[#2] Assistant Professor, Master of Computer Applications,

D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.

Abstract

We proposed a novel framework to address the above issues by proposing top-k high utility item set mining, where k is the desired number of HUIs to be mined. Here in our proposed thesis we developed two types of efficient algorithms one is NTKU (I.e. Novel mining of Top-K Utility item sets) and second one is NTKO (Novel mining of Top-K utility item sets in one phase) for mining the item sets automatically without the need of setting min_util value. By conducting various experiments on our proposed two algorithms with an online products data sets, its show that the performance of the proposed algorithms is best in displaying the products in a UP-Tree structure for the current thesis, where we can show all the utility items mapped in a tree manner under the root node

1. INTRODUCTION

FREQUENT item set mining (FIM) [1] is a fundamental research topic in data mining. However, the traditional FIM may discover a large amount of frequent but low-value item sets and lose the information on valuable item sets having low selling frequencies. Hence, it cannot satisfy the requirement of users who desire to discover item sets with high utilities such as high profits. To address these issues, utility emerges as an important topic in data mining and has received extensive attention in recent years. In utility mining, each item is associated with a utility (e.g. unit profit) and an occurrence count in each transaction (e.g. quantity).

The utility of an item set represents its importance, which can be measured in terms of weight, value, quantity or other information depending on the user specification. An item set is called high utility item set (HUI) if its utility is no less than a user-specified minimum utility threshold min_ util. HUI mining is essential to many applications such as streaming analysis [2], [11], [35], market analysis [13], [17], [22], mobile computing [23] and biomedicine [4]. However, efficiently mining HUIs in databases is not an easy task because the downward closure property [1], [8] used in FIM does not hold for the utility of item sets. In other words, pruning search space for HUI mining is difficult because a superset of a low utility item set can be high utility.

To tackle this problem, the concept of transaction weighted utilization (TWU) model [13] was introduced to facilitate the performance of the mining task. In this model, an item set is called high transaction-weighted utilization item set (HTWUI) if its TWU is no less than min_ util, where the TWU of an item set represents an upper bound on its utility. Therefore, a HUI must be a HTWUI and all the HUIs must be included in the complete set of HTWUIs. A classical TWU model-based algorithm consists of two phases. In the first phase, called phase I, the complete set of HTWUIs are found. In the second phase, called phase II, all HUIs are obtained by calculating the exact utilities of HTWUIs with one database scan. Although many studies have been devoted to HUI mining, it is difficult for users to choose an appropriate minimum utility threshold in practice. Depending on the threshold, the output size can be very small or very large.

Besides, the choice of the threshold greatly influences the performance of the algorithms. If the threshold is set too low, too many HUIs will be presented to the users and it is difficult for the users to comprehend the results. A large number of HUIs also causes the mining algorithms to become inefficient or even run out of memory, because the more HUIs the algorithms generate, the more resources they consume. On the contrary, if the threshold is set too high, no HUI will be found. Although many studies have been devoted to HUI mining, it is difficult for users to choose an appropriate minimum utility threshold in practice. The existing 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.In the existing system, there is no concept like mining utility item sets based on overall utility of all items.In the existing system, there is no concept like mapping all utility items on a tree manner like UP-Tree(Utility Pattern Tree) mapped based on traffic.

PROJECT SCOPE

We address all of the above challenges by proposing a novel framework for top-k high utility itemset mining, where k is the desired number of HUIs to be mined. First, two efficient algorithms named TKU (mining Top-K Utility itemsets) and TKO (mining Top-K utility itemsets in One phase) are proposed for mining the complete set of top-k HUIs in databases without the need to specify the min_util threshold.The TKU algorithm adopts a compact tree-based structure named UP-Tree to maintain the information of

transactions and utilities of itemsets. TKU inherits useful properties from the TWU model and consists of two phases.

PROJECT OVERVIEW

In this project, Empirical evaluations on different types of real and synthetic datasets show that the proposed algorithms have good scalability on large datasets and the performance of the proposed algorithms is close to the optimal case of the state-of-the art two-phase and one-phase utility mining algorithms.

2. LITERATURE SURVEY

INRODUCTION

Literature survey is the most important step in software development process. Before developing the tool, it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, ten next steps are to determine which operating system and language used for developing the tool. Once the programmers start building the tool, the programmers need lot of external support. This support obtained from senior programmers, from book or from websites. Before building the system the above consideration r taken into for developing the proposed system.

RELATED WORK

1. Social contextual recommendation

Authors: M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang

Exponential growth of information generated by online social networks demands effective and scalable recommender systems to give useful results. Traditional techniques become unqualified because they ignore social relation data; existing social recommendation approaches consider social network structure, but social contextual information has not been fully considered. It is significant and challenging to fuse social contextual factors which are derived from users' motivation of social behaviors into social recommendation. In this paper, we investigate the social recommendation problem on the basis of psychology and sociology studies, which exhibit two important factors: individual preference and interpersonal influence. We first present the particular importance of these two factors in online behavior prediction. Then we propose a novel probabilistic matrix factorization method to fuse them in latent space. We further provide a scalable algorithm which can incrementally process the large scale data. We conduct experiments on both Facebook style bidirectional and Twitter style unidirectional social network data sets. The empirical results and analysis on these two large data sets demonstrate that our method significantly outperforms the existing approaches.approaches.

2. Personalized recommendation based on reviews and ratings alleviating the sparsity problem of collaborative filtering

Authors: J. Xu, X. Zheng, W. Ding

With the development of e-commerce, shopping on-line is becoming more and more popular. When we need to decide whether to purchase a product or not on line, the opinions of others become important. The convenience of new web technologies enables us to freely express our opinions and reviews for various products we have purchased which leads to a serious problem, information overloading. How to mine these review data to understand customers' preferences and make recommendations is crucial to merchants and researchers. Traditional collaborative filtering (CF) algorithm is one of the most successful recommendation system technologies. The core idea of CF algorithm is to recommend products based on other people who have similar tastes with target users. However, the ability of CF is limited by the sparsity problem, which is very common in reality. The reason derives from the fact that traditional CF method only takes users' ratings into account. In this paper, we propose a new personalized recommendation model, i.e. topic model based collaborative filtering (TMCF) utilizing users' reviews and ratings. We exploit extended LDA model to generate topic allocations for each review and then obtain each user's preference. Moreover, a new metric is designed to measure similarity between users alleviating the sparsity problem to a large extent. Finally, recommendations are made based on similar users' ratings. Experiments on seven data sets indicate better prediction accuracy than other traditional and state-of-the-art methods with substantial improvement in alleviating the sparsity problem.

3. Semantic-based location recommendation with multimodal venue semantics

Authors: X. Wang, Y. Zhao, L. Nie, Y. Gao

In recent years, we have witnessed a flourishing of location -based social networks. A well-formed representation of location knowledge is desired to cater to the need of location sensing, browsing, navigation and querying. In this paper, we aim to study the semantics of point-of-interest (POI) by exploiting the abundant heterogeneous user generated content (UGC) from different social networks. Our idea is to explore the text descriptions, photos, user check-in patterns, and venue context for location semantic similarity measurement. We argue that the venue semantics play an important role in user check-in behavior. Based on this argument, a unified POI recommendation algorithm is proposed by incorporating venue semantics as a regularizer. In addition to deriving user preference based on user-venue check-in information, we place special emphasis on location semantic similarity. Finally, we conduct a comprehensive performance evaluation of location semantic similarity and location recommendation over a real world dataset collected from Foursquare and Instagram. Experimental results show that the UGC information can well characterize the venue semantics, which help to improve the recommendation performance.

3. EXISTING SYSTEM

In the existing system there is no concept like UP tree to display the items in a tree manner.

LIMITATIONS

1. Although many studies have been devoted to HUI mining, it is difficult for users to choose an appropriate minimum utility threshold in practice.
2. The existing 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.
3. In the existing system, there is no concept like mining utility item sets based on overall utility of all items.
4. In the existing system, there is no concept like mapping all utility items on a tree manner like UP-Tree(Utility Pattern Tree) mapped based on traffic.

4. PROPOSED SYSTEM

We address all of the above challenges by proposing a novel framework for top-k high utility itemset mining, where k is the desired number of HUIs to be mined. Major contributions of this work are summarized as follows:First, two efficient algorithms named TKU (mining Top-K Utility itemsets) and TKO (mining Top-K utility itemsets in One phase) are proposed for mining the complete set of top-k HUIs in databases without the need to specify the min_util threshold.The TKU algorithm adopts a compact tree-based structure named UP-Tree to maintain the information of transactions and utilities of itemsets. TKU inherits useful properties from the TWU model and consists of two phases.

ADVANTAGES OF THE PROPOSED SYSTEM

The following are the advantages of the proposed system. They are as follows:

1. Two efficient algorithms TKU (mining Top-K Utility itemsets) and TKO (mining Top-K utility itemsets in One phase) are proposed for mining such itemsets without setting minimum utility thresholds.
2. Empirical evaluations on different types of real and synthetic datasets show that the proposed algorithms have good scalability on large datasets.
3. The performance of the proposed algorithms is close to the optimal case of the state-of-the art two-phase and one-phase utility mining algorithms.

5. SOFTWARE PROJECT MODULES

Implementation is the stage where the theoretical design is converted into programmatically manner. In this stage we will divide the application into a number of modules and then coded for deployment. We have implemented the proposed concept on Java programming language with JSE as the chosen language in order to show the performance this proposed protocol. The application is divided mainly into following 2 modules. They are as follows:

1) Admin Module

2) User Module

Now let us discuss about each and every module and sub modules which are present in this application.

5.1 Admin Module

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as

1. Viewing and Authorizing Users
2. Add Categories, Sub-Categories and Product Posts
3. View all Products with Ranks and Comments
4. View Top-K Utility Item Sets Keywords
5. View all Products in terms of Construction of UP-Tree
6. View all high Utility Item Set Mining Products
7. Find Top K Products Results in Chart

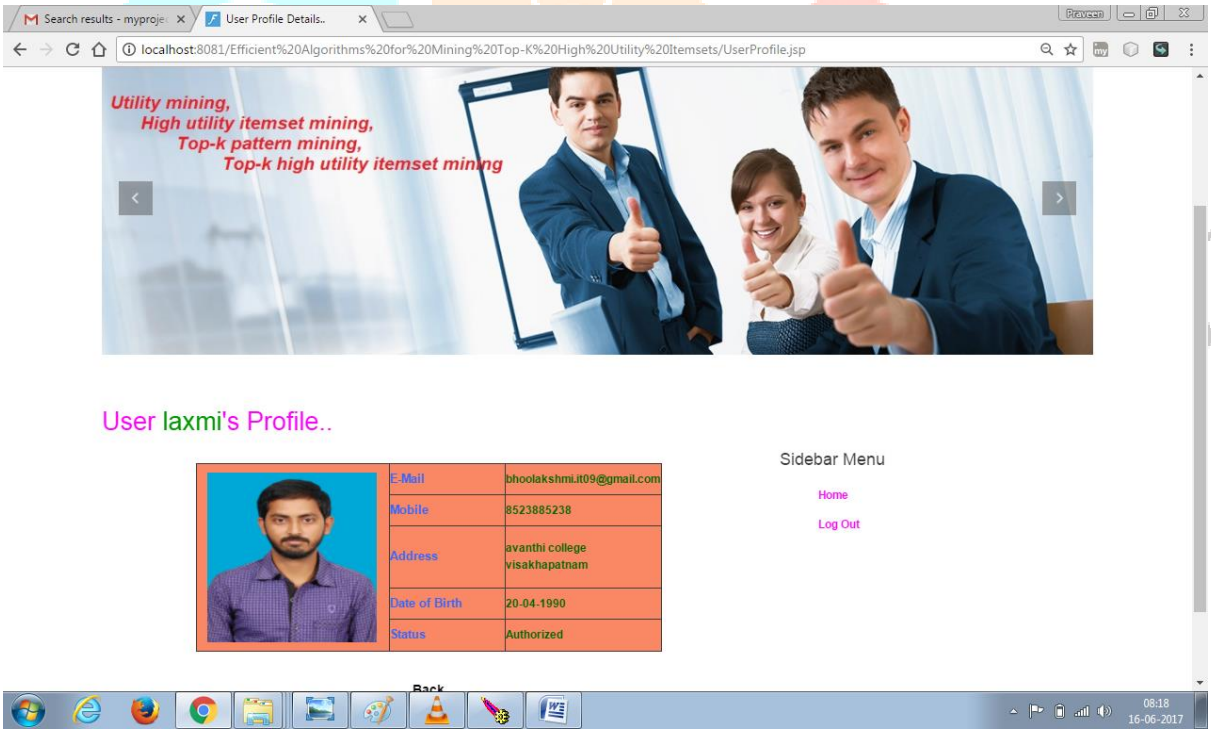
5.2 User Module

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations like

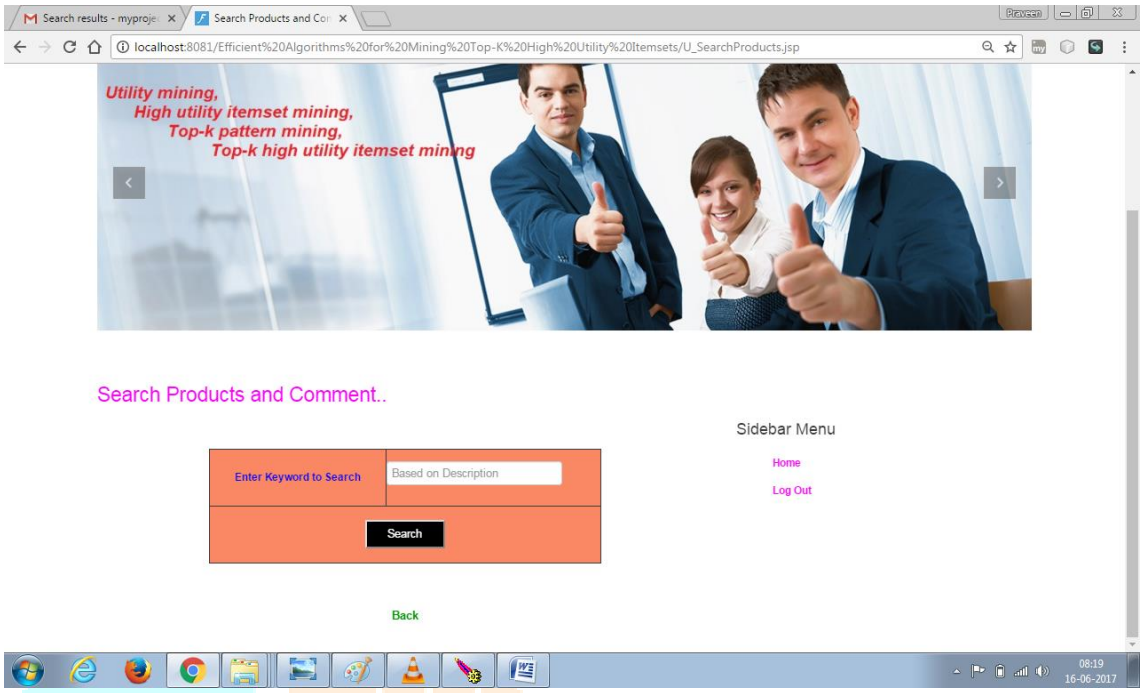
1. Viewing Profile Details
2. Search Products
3. Search and View Products in UP-Tree Format
4. Finding Top K Item Sets

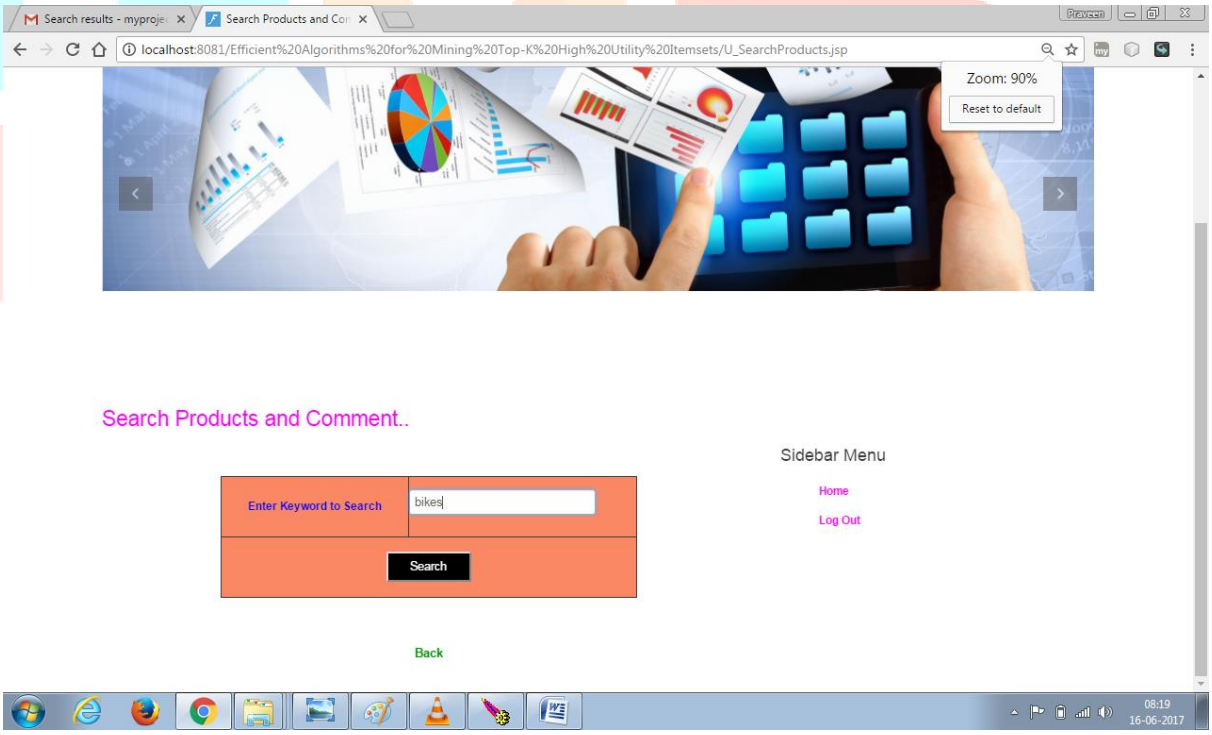
6. OUTPUT RESULTS

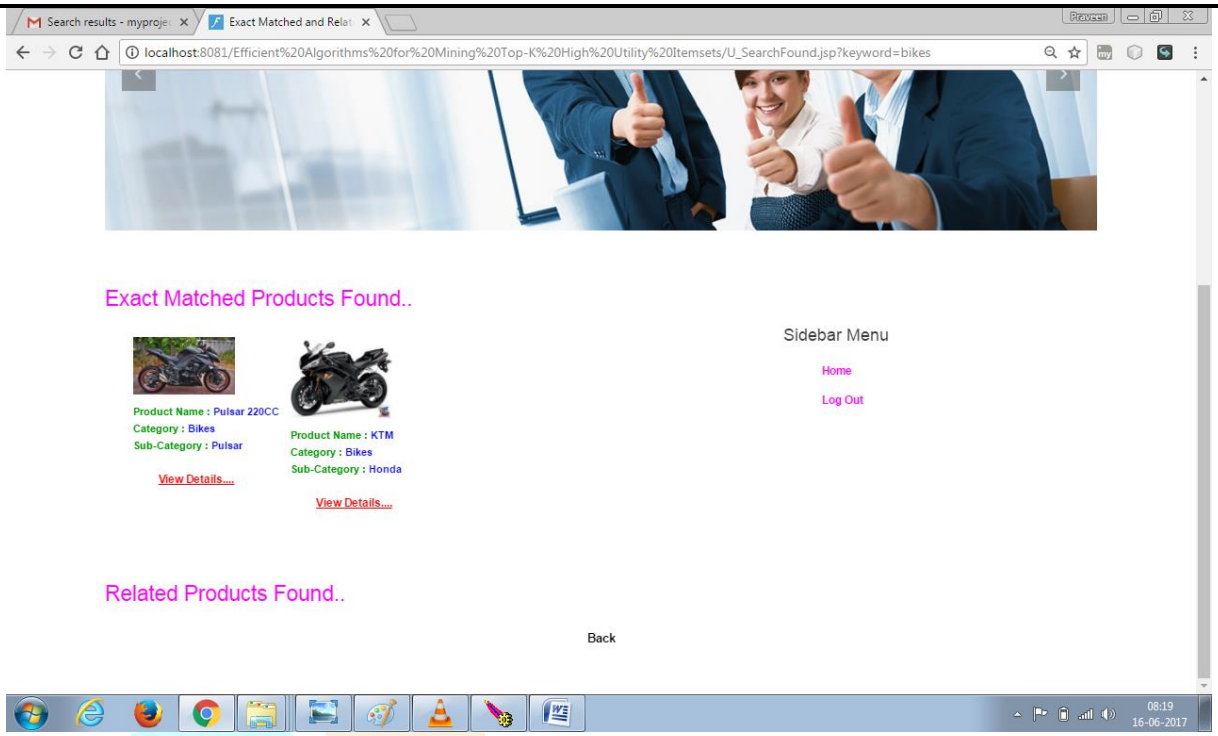
User Views His/Her profile



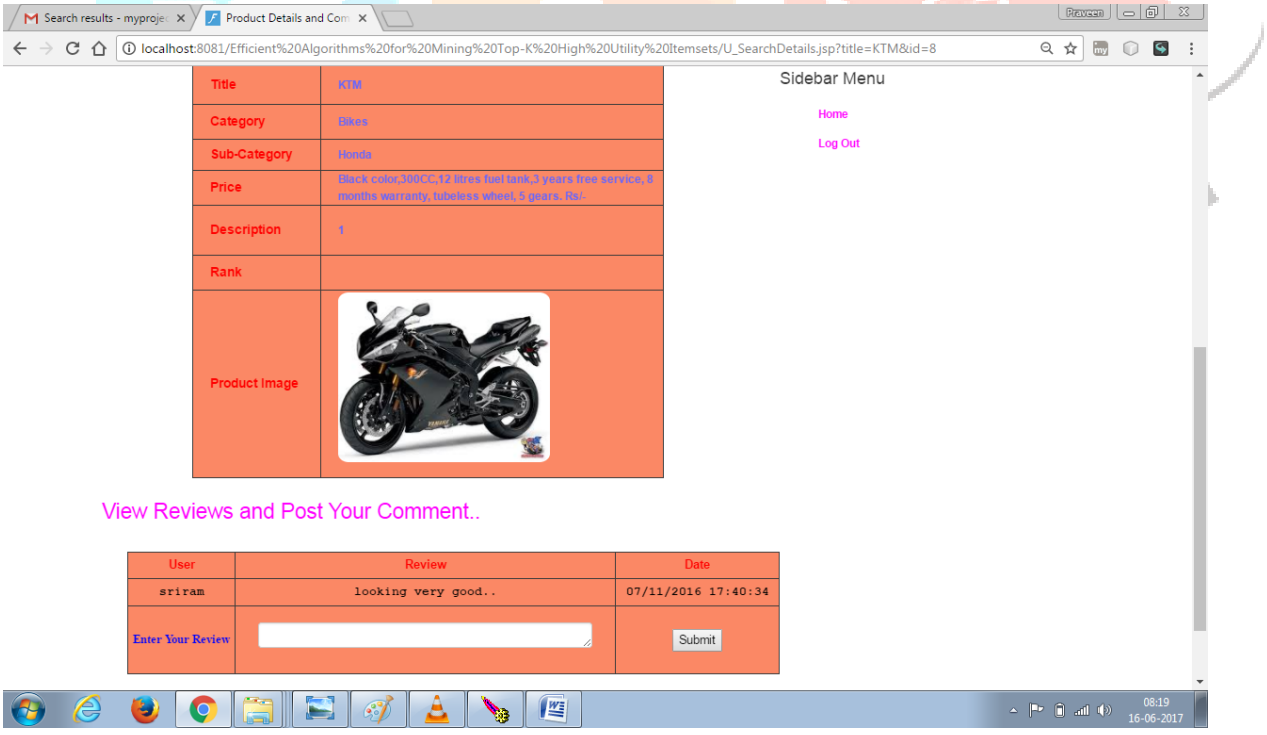
User Can search for products based on product name or category name or sub category name





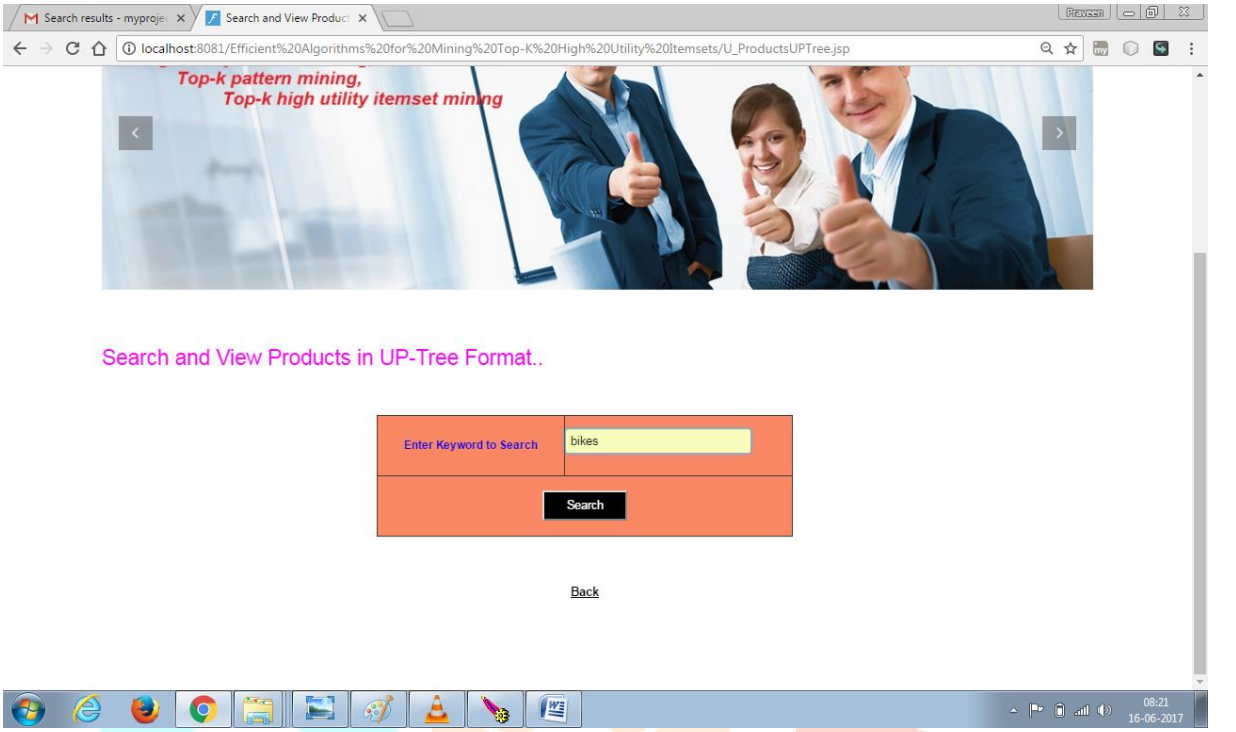


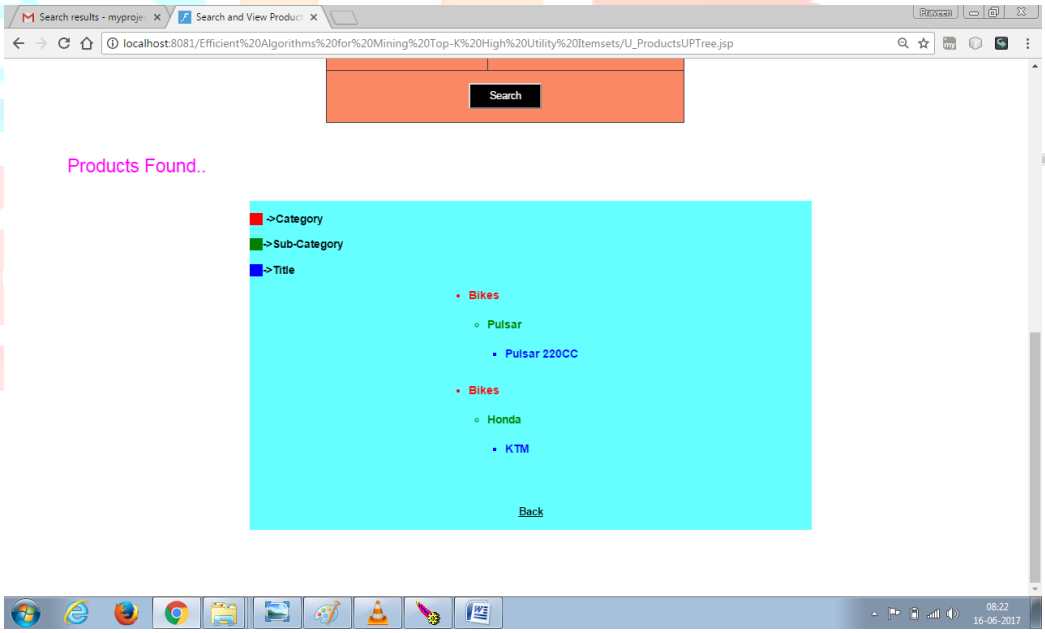
User Click on View Details button to view the products



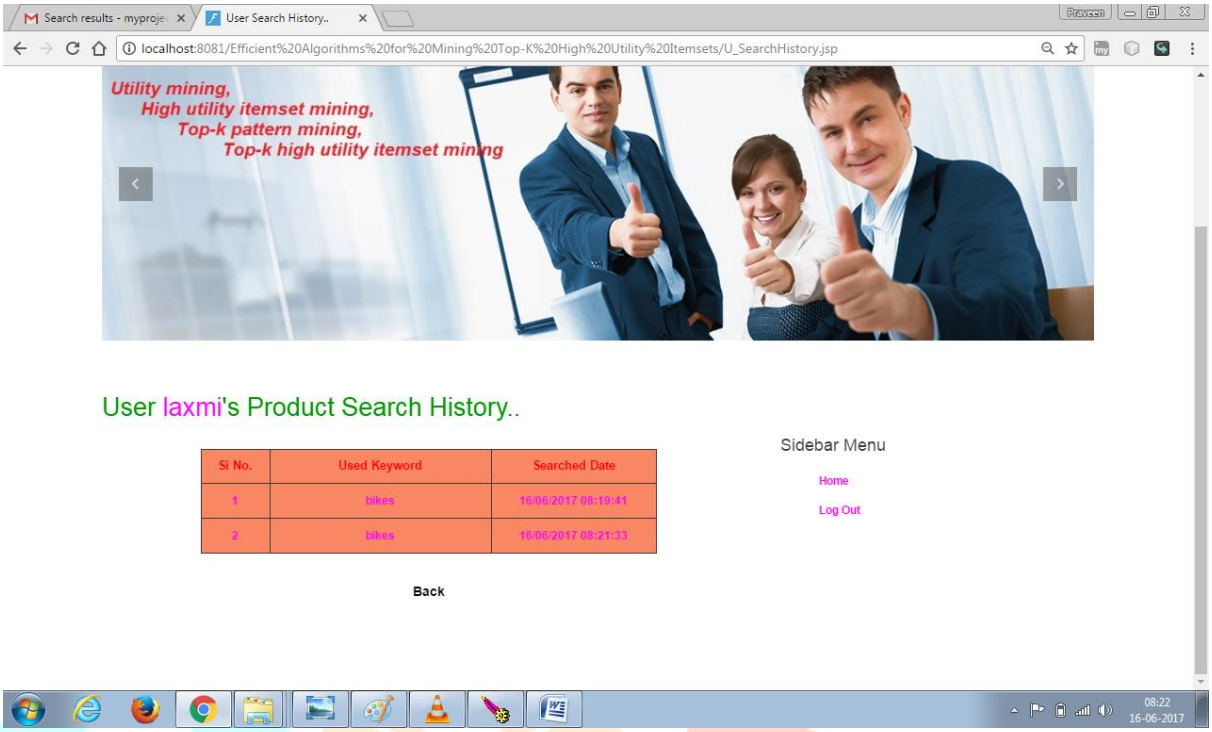
User can view top K Priority wise items

User Can view the UP Tree Based on category Wise (Extension for the current paper)





User can view his/Her Search History



7. CONCLUSION

In this project, we have studied the problem of top-k high utility item sets mining, where k is the desired number of high utility item sets to be mined. Two efficient algorithms TKU (mining Top-K Utility item sets) and TKO (mining Top-K utility item sets in One phase) are proposed for mining such item sets without setting minimum utility thresholds. TKU is the first two-phase algorithm for mining top-k high utility item sets. In our research work we have worked on just user reviews. In future, we want to extract the UP tree with automatic assignment of categories and sub categories.

8. REFERENCES

[1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in Proc. Int. Conf. Very Large Data Bases, 1994, pp. 487– 499.

[2] C. Ahmed, S. Tanbeer, B. Jeong, and Y. Lee, "Efficient tree structures for high-utility pattern mining in incremental databases," IEEE Trans. Knowl. Data Eng., vol. 21, no. 12, pp. 1708–1721, Dec. 2009.

[3] K. Chuang, J. Huang, and M. Chen, "Mining top-k frequent patterns in the presence of the memory constraint," VLDB J., vol. 17, pp. 1321–1344, 2008.

[4] R. Chan, Q. Yang, and Y. Shen, "Mining high-utility itemsets," in Proc. IEEE Int. Conf. Data Mining, 2003, pp. 19–26.

[5] P. Fournier-Viger and V. S. Tseng, "Mining top-k sequential rules," in Proc. Int. Conf. Adv. Data Mining Appl., 2011, pp. 180–194.

[6] P. Fournier-Viger, C.Wu, and V. S. Tseng, "Mining top-k association rules," in Proc. Int. Conf. Can. Conf. Adv. Artif. Intell., 2012, pp. 61–73.

[7] P. Fournier-Viger, C. Wu, and V. S. Tseng, "Novel concise representations of high utility itemsets using generator patterns," in Proc. Int. Conf. Adv. Data Mining Appl. Lecture Notes Comput. Sci., 2014, vol. 8933, pp. 30–43.

[8] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in Proc. ACM SIGMOD Int. Conf. Manag. Data, 2000, pp. 1–12.

[9] J. Han, J. Wang, Y. Lu, and P. Tzvetkov, "Mining top-k frequent closed patterns without minimum support," in Proc. IEEE Int. Conf. Data Mining, 2002, pp. 211–218.

[10] S. Krishnamoorthy, "Pruning strategies for mining high utility itemsets," Expert Syst. Appl., vol. 42, no. 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," Adv. Eng. Informat., vol. 29, no. 1, pp. 16–27, 2015.

[12] G. Lan, T. Hong, V. S. Tseng, and S. Wang, "Applying the maximum utility measure in high utility sequential pattern mining," Expert Syst. Appl., vol. 41, no. 11, pp. 5071–5081, 2014.

[13] Y. Liu, W. Liao, and A. Choudhary, "A fast high utility item sets mining algorithm," in Proc. Utility-Based Data Mining Workshop, 2005, pp. 90–99.

[14] M. Liu and J. Qu, "Mining high utility item sets without candidate generation," in Proc. ACM Int. Conf. Inf. Knowl. Manag., 2012, pp. 55–64.

[15] J. Liu, K. Wang, and B. Fung, "Direct discovery of high utility item sets without candidate generation," in Proc. IEEE Int. Conf. Data Mining, 2012, pp. 984–989.

[16] Y. Lin, C. Wu, and V. S. Tseng, "Mining high utility itemsets in big data," in Proc. Int. Conf. Pacific-Asia Conf. Knowl. Discovery Data Mining, 2015, pp. 649–661.