



# A Framework Outline for Word Catalog Based Model in Cloud

Mr.C.Sivakumar<sup>1</sup>, Mr.M.Arulprakash M.Tech<sup>2</sup>

M.E II year, Department of CSE, Sri Subramanya College of Engineering and Technology, Palani, Dindigul, Tamilnadu, India-624 615<sup>1</sup>

Assistant Professor, Department of CSE, Sri Subramanya College of Engineering and Technology, Palani, Dindigul, Tamilnadu, India-624 615<sup>2</sup>

**ABSTRACT:** Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy. We add a ranking procedure that turns a pairwise preference relation into a complete preorder of the option The statistical measure approach, i.e., relevance score, from information retrieval to build a secure searchable index, and develop a one-to-many order-preserving mapping technique to properly protect those sensitive score information. The resulting design is able to facilitate efficient server-side ranking without losing keyword privacy. Thorough analysis shows that our proposed solution enjoys “as strong-as-possible” security guarantee compared to previous searchable encryption schemes, while correctly realizing the goal of ranked keyword search. We introduce a general preference based racing (PBR) algorithm that provides the basic statistical information’s which is needed to solve the selection problem, notably estimates of the  $y_{i,j}$  and corresponding confidence intervals. It contains a subroutine that implements sampling strategies for the different ranking models based on user preferences. A Manipulation mining system will take care of calculating the users interest and provide an optimistic approach to fetch the data based on user’s interest.

**KEYWORDS:** Upload Document, Keyword Search, Search by Date, Search by Doc name

## I. INTRODUCTION

Dominance analysis is important in many multicriteria decision-making applications. we propose a dynamic programming approach which finds an optimal solution when there are two attributes to be considered. But, we show that this problem is NP-hard when there are more than two attributes to be considered. Thus, we propose two greedy algorithms for this problem. One greedy algorithm has a theoretical guarantee on the profit returned while the other greedy algorithm performs well empirically.

In some cases, data sets are dynamic and change from time to time. In this paper, we also study how to find top-k profitable products when data sets change. For example, some new products are launched in the existing market while some products which were present in the existing market become unavailable. Besides, the prices of existing products in the market may change due to various reasons, such as inflation and cost increase.

The second instance is called finding top-k popular products. In some cases, if we know how many customers are interested in some potential products, we can better find potential products. One well-known application which allows customers to provide their preferences. Similarly, a naive way is to enumerate all possible subsets of size k from Q, calculate the total number of customers interested in some packages in this subset, and choose the subset with the greatest

## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

### Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014

number of customers. But, it is not scalable. We show that this problem is NP-hard. But, interestingly, we propose a 0.63-approximate algorithm which runs in polynomial time.

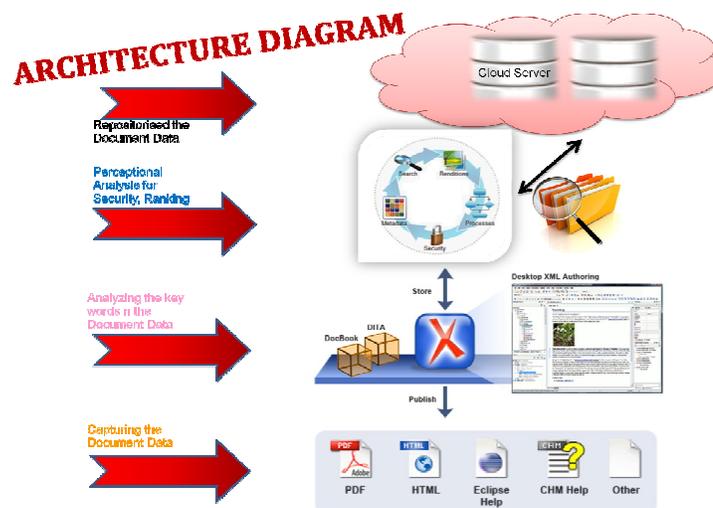
## II. EXISTING SYSTEM

To find top k profitable products: A naïve way for this instance problem is to enumerate all possible subsets of size k from the available set and then calculate the sum of the profits of each possible subset and finally choose the subset with greatest sum. To find top k popular products: A naïve way for this instance problem is similar to that of the first instance. First find all possible subsets of size k from the available set and then choose the subset with greatest number of customers.

### Disadvantages:

- The method used for finding top k profitable products (first problem instance) is not scalable because there are an exponential number of all possible subsets.
- The method used for finding top k popular products (second problem instance) is also not scalable because this also involves finding all possible subsets which is exponential.

## III. ARCHITECTURAL DIAGRAM



## IV. PROPOSED SYSTEM

To find top k profitable products: A dynamic programming approach is proposed, which finds an optimal solution when there are two attributes to be considered. Here we are utilizing the option of find Optimal Incremental Property Algorithm. In which, we are trying to validate/identify the quasi dominance of the products and apart from that our system will recognize the skyline checks on the available data. Based on an optimized check among the two technical jargons, the profitable products will be identified. To find top k popular products: The adaptive pulling strategy related to the products prioritizes access among the two relations based on the observed data. The main idea behind this approach is to read the tuples from a relation only if there is possible evidence regarding the new tuples which will help and satisfy the termination condition. Intuitively, this prioritization process helps the algorithm terminate faster and sooner, thus improving its



## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

### Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014

performance. Obviously, the popular products based on skyline processing are done. Performance evaluation on various factors is one of the major breakthroughs in our new system.

#### Advantages:

- Finding top-k preferable products can help the effort of companies to find a subset of products together with their corresponding profitable prices, which is not addressed by existing methods.
- For the first problem of finding top-k profitable products, a dynamic programming approach which can find an optimal solution when there are two attributes to be considered is proposed.
- Finding top k popular products (second problem instance) is NP-hard and therefore a 0.63-approximate algorithm is used.
- An incremental approach is used to handle **dynamic datasets** that change over time.
  - Deletion of an existing package
  - Insertion of a new package
  - Modifying the attribute values of an existing package

## V. TECHNIC USED TO IMPLEMENT

SHA-1 Algorithm:

- SHA-1 produces a 160-bit message digest based on principles in the design of the MD4 and MD5 message digest algorithms, but has a more conservative design.
- SHA-1 differs from SHA-0 only by a single bitwise rotation in the message schedule of its compression function.

## VI. MODULE DESCRIPTION

- DOCUMENT ACCESSOR SPECIFICATION MODULE
- DOCUMENT UPLOAD MODULE
- DOCUMENT PARSE MODULE
- DOCUMENT SEARCH MODULE
- KEY APPLY MODULE
- DOCUMENT SUMMARIZATION

### DOCUMENT ACCESSOR SPECIFICATION MODULE

- An initial entry for the system to access the application. Obviously, the data related to the user will be amended through a new user form.
- In this module, we are using SHA 1 algorithm to store the user credentials.
- Obviously, the data will be validated and the valid user will be permitted to enter into the system.

### DOCUMENT UPLOAD MODULE

- Uploading is the transmission of a file from one computer system to another, usually larger computer system. From a network user's point-of-view, to upload a file is to send it to another computer that is set up to receive it.



## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

### Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014

#### DOCUMENT PARSE MODULE

- In this module, the uploaded document in the previous module is parsed.
- Document parsers can access the content type that is assigned to a document and store the content type in the document itself.
- In addition, document parsers can update the content type definition that is stored in a document so that it matches the version of the content type definition that is used by a list or document library.
- The privilege to each users has been provided like sharable/viewable/editable so as to use the document

#### DOCUMENT SEARCH MODULE

- Document search is a searching strategy of document present in the cloud.
- In this search strategy, input in the form of keyword has been given. This include searching under file name/file type/file size.

#### KEY APPLY MODULE

- Once after the successful search by the user by particular file type / file size / file name, the required document can be requested to the owner of the document.
- Data owner issues a cryptographic key so as to access privilege to view the document request.

#### DOCUMENT SUMMARIZATION

- Document summarization is an automatic procedure aimed at extraction of information from multiple texts written about the same topic.
- Resulting summary report allows individual users, such as professional information consumers, to quickly familiarize themselves with information contained in a large cluster of documents

## VII. CONCLUSION

### SUMMARY:

We identify and tackle the problem of finding top-k preferable products, which has not been studied before. We study two instances of preferable products, namely profitable products and popular products. We propose methods to find top-k profitable products and topk popular products efficiently. An extensive performance study using both synthetic and real data sets is reported to verify its effectiveness and efficiency

### FUTURE ENHANCEMENT:

As future work, we will study other instances of the problem of finding top-k preferable products by setting the utility function to other meaningful objective functions. One promising utility function is the function which returns the sum of the unit profits of the selected products multiplied by the number of customers interested in these products.

## REFERENCES

- [1] X. Wan and J. Xiao, "Exploiting Neighborhood Knowledge for Single Document Summarization and Keyphrase Extraction," ACM Trans. Information Systems, vol. 28, pp. 8:1-8:34, <http://doi.acm.org/10.1145/1740592.1740596>, June 2010.
- [2] K.S. Jones, "Automatic Summarising: Factors and Directions," Advances in Automatic Text Summarization, pp. 1-12, MIT Press, 1998.



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.2, Special Issue 1, March 2014

**Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)**

**Organized by**

**Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014**

[3] L.L. Bando, F. Scholer, and A. Turpin, "Constructing Query- Biased Summaries: A Comparison of Human and System Generated Snippets," Proc. Third Symp. Information Interaction in Context, pp.195-204, <http://doi.acm.org/10.1145/1840784>. 1840813, 2010.

[4] X. Wan, "Towards a Unified Approach to Simultaneous Single-Document and Multi-Document Summarizations," Proc. 23rd Int'l Conf. Computational Linguistics, pp. 1137-1145, <http://portal.acm.org/citation.cfm?id=1873781.1873909>, 2010.