Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 3, March 2016, pg.157 – 162

Secure and Dynamic Image Search for Retrieving Document

Abinaya. S[1], Dr. R. Kalpana[2]
Department of Computer Science and Engineering, IFET College of Engineering, Villupuram

Abstract- A web search engine is a software system that is designed to search for information on the World Wide Web. In the Existing search engines the accuracy of retrieving the document using the image is low. It is inefficient in the retrieval of documents. The aim of the image search is to retrieve the relevant image with respect to user query from a large image database. We first discuss the homomorphic encryption and Feature/Index Randomized technique that enable similarity comparison between features in the encrypted domain. The encrypted features along with encrypted images can protect image content privacy against untrustworthy service providers and malicious intruders. The ability to generate encrypted indexes on the user side provides an alternative for secure retrieval with reduced communication overhead. We proposed secure k-NN computation that can determine which of using to display result of correct images to the user "active re-ranking" is used. In this method the user intention is captured and used for re-ranking the images. To improve the performance of search, labelling information is collected from user and new method is proposed to actively select more informative query images through structural information. Few images are labelled by user in active re-ranking.

Keywords: K-NN algorithm, Homomorphic encryption, Feature/Index Randomized Technique.

I. Introduction

Now, mostly the commercial Web image search engines, e.g., Microsoft's Live Image Search and Google Image Search[2], are based on query by keywords scenario. That means, a user provides a keyword, e.g., rose, then the search engine returns corresponding images by processing the associated textual information, e.g., file name, surrounding text, URL, etc. Here we use the image as the input and get the relevant document as the output eg. If the user give the rose image it display the related documents about the rose in the web[1]. For the web page the encryption is not needed because all the users access the document. But in certain organisations the encrypted image is needed because unauthorised user can hack the image. So we use the encryption technique called homomorphic encryption. It is used to only encrypt the image but it cannot be decrypted by the unauthorized user because it need security key for decryption[3,4]. By using the homomorphic encryption the unauthorised used cannot know which image is given as input because they only have the public key for decryption. This technique is used to strengthen the security mechanism. To increase the accuracy of the search we use the feature/indexed randomized technique[4] it rank the image based on the accuracy of retrieving the

document related to the image. The document which is related to the image are indexed by the index randomized technique. So at the next time it cannot need to compare all the documents to retrieve the result because already the indexing is done for the document. For comparing the image we use the k-nn (K Nearest Neighbour) algorithm to retrieve the relevant document for the given images[10]. It compare all the image in the database by using the features in the image then it retrieve the relevant document for the given image from the large database. The database is large and it is efficient in retrieving the document by comparing with the text based search engine. Thus the noisy interrupts are overcome from the image search engine. It is efficient to search the document in the image search engine if we don't know anything about the image. It need only less amount of cost to design the search engine.

II. Related Work

Related work falls into a number of categories: Exploiting Image content in web search, Improve ranking, Photo based question answering.

What Can Pictures Tell Us About Web Pages?
Improving Document Search Using Images[1]:

In this paper they check the document search using the images. They use the Trec Million Query track bench mark to search the content using the image. It take the feature of the image and search the document related to the image. They compare the accuracy of the text-based search engines (UDMQ and Indri) to the different image based ranking models. We have shown that this yields a 33 percent relative improvement in accuracy over a state-of-the-art text-based retrieval baseline. All this is achieved at the small cost of a few additional hundred bytes of storage for each page.

Exploiting image content in web search [**6**]:

This paper proposes a new framework for Web search, which exploits image contents to help improve the search performance. In this framework, candidate images are retrieved at first by considering their associated text information. Then, images related to the query are identified by analyzing the density of the visual feature space. After that, an image-based rank of the Web pages is generated, which is combined with the traditional keyword-based search result to produce the final search result. It Improve the search performance. Lack of a mechanism to adaptively trade off the contributions of keyword-based and image-based ranks. But The current process for analyzing the images is not fast enough.

Improve Ranking by Using Image Information[5]:

In determining the ranking of Web pages against a given query, most (if not all) modern Web search engines consider two kinds of factors: text information (including title, URL, body text, anchor text, etc) and static ranking (e.g. PageRank). Although images have been widely used to help represent Web pages and carry valuable information, little work has been done to take advantage of them in computing the relevance score of a Web page given a query. We propose, in this paper, a framework to contain image information in ranking functions. But the online computation cost is too high.

Photo-based Question Answering[7]

In the photo based question and answering they develop a three-layer system architecture for photo-based QA that brings together recent technical achievements in question answering and image matching. The first, template-based QA layer matches a query photo to online images and extracts structured data from multimedia databases to answer questions about the photo. To simplify image matching, it exploits the question text to filter images based on categories and keywords. The second, information retrieval QA layer searches an internal repository of resolved photo-based questions to retrieve relevant answers. The third, human-computation QA layer leverages community experts to handle the most difficult cases. The QA system uses questions to identify relevant categories to ease the burden of image matching. On the other hand, even if the QA system fails to find relevant categories, image matching still achieves reasonable accuracy, which can in turn benefit the QA process. But it is too difficult to handle the first layer.

III. Proposed System

The aim of the image search is to retrieve the relevant image with respect to user query from a large image database. So identifying the accurate image with user intention is the most challenging task. We propose two categories of secure retrieval schemes. This paper focuses on comparing these two major paradigms of techniques, namely, homomorphic encryption based techniques and feature/index randomization-based techniques, for confidentiality-preserving image search. Most state-of-the-art content-based image retrieval techniques utilize low-level visual features to represent and compare image content, and these visual features can potentially reveal important information about the image content. We first discuss feature protection schemes that enable similarity comparison between features in the encrypted domain. The encrypted features along with encrypted images can protect image content privacy against untrustworthy service providers and malicious intruders. The ability to generate encrypted indexes on the user side provides an alternative for secure retrieval with reduced communication overhead. We proposed secure k-NN computation that can determine which of using to display result of correct images to the user "active reranking" is used. In this method the user intention is captured and used for re-ranking the images. To improve the performance of search, labelling information is collected from user and new method is proposed to actively select more informative query images through structural information. Few images are labelled by user in active re-ranking.

K-NN Algorithm:

We first categorize the documents using KNN based machine learning approach and then return the most relevant documents. We presented a k-NN algorithm for improving the classification of large collection of documents. We extract the content and then search the index for the K-NN classifications. To improve the performance of search, information is collected from user and new method is proposed to actively select more informative query images through structural information.

KNN can be run in these steps:
1. Store the output values of the M nearest neighbors to query scenario q in vector $r=\{r^1,...r^M\}$ by repeating the following loop M times:
a. Go to the next scenario s^i in the data set, where is the current iteration within the domain {1,....,P}
b. If q is not set or $q<d(q,s^i)$:$q<-d(q,s^i)$, $t<-o^i$
c. Loop until we reach the end of the data set (i.e.i=P)
d. Store q into vector c and t into vector r
2. Calculate the arithmetic mean output across r as follows:
$\bar{r}=\frac{1}{M}\sum_{i=1}^{M} r_i$

3. Return $\bar{r}$as the output value for the query scenario q.

Homomorphic Encryption:

A homomorphic encryption scheme is a crypto system that allows computations to be performed on data without decrypting it. Users share documents with encryption format for the attackers could not retrieved the document. So users uploading the data using homomorphic encryption technique for security.

Additive Homomorphic Encryption:
A Homomorphic encryption is additive, if:
Enc $(x \oplus y) = Enc(x) \otimes Enc(y)$
l l
Enc $(\Sigma mi) = \Pi$ Enc (mi)
i=1 i=1

Multiplicative Homomorphic Encryption:
A Homomorphic encryption is multiplicative, if:

$$Enc(x \otimes y) = Enc(x) \otimes Enc(y)$$

$$Enc\left(\prod_{i=1}^{l} m_i\right) = \prod_{i=1}^{l} Enc(m_i)$$

Feature/Indexed Randomized Technique:

This technique can be used to find the high accuracy of matching content with optimal solution. Feature selection is used to identifying a content of the most relevant features in the context of images. We then choose a random neighbor of dataset and compute its accuracy. This process of searching the content is continued until no significant improvement in the accuracy can be obtained. We have found that it is indeed scalable, reliable and efficient. Randomizing feature and search indexes to enable similarity comparison between encrypted document and images.

IV. System implementation

User Interface

User Registration Module provides functionality to register viewers of the learning site in order to get access to personalized content that the site using this module provides to its users. Module can be also used to register users for custom modules that support personalization and user specific handling. For example module can be used to get awareness from the users and updating the resources of learning based on the awareness. Updating the resources means adding some additional information based on the user request if it is a valuable one.

Data Pre-processing

The user gives the input data as a images or query, first it analyze the user images or query. To achieve this, the data owner needs to build a searchable index from a collection of keywords extracted out of files, and then outsources both the encrypted index and encrypted files onto the server. This module is to analyze the query and pre-processing process takes place.

Query Analysis

After pre-processing of data, this module finds out the semantic analysis for the input images. This method finds out the meaning of the input data images. The semantic dataset is used, and it is compares with the input query, and results in the semantic data for the input data images.

Retrieving Result

It first extracts the web documents, and compares with the semantic relations. It uses the visual features for comparison and find out the relations between the web documents and the user query input data images. After calculating the relations between the web documents and the user query. The randomization-based method evaluates the available routes, which are accurate results. Routing plans are listed out and the ranking method takes place to rank the web documents as per the user query.

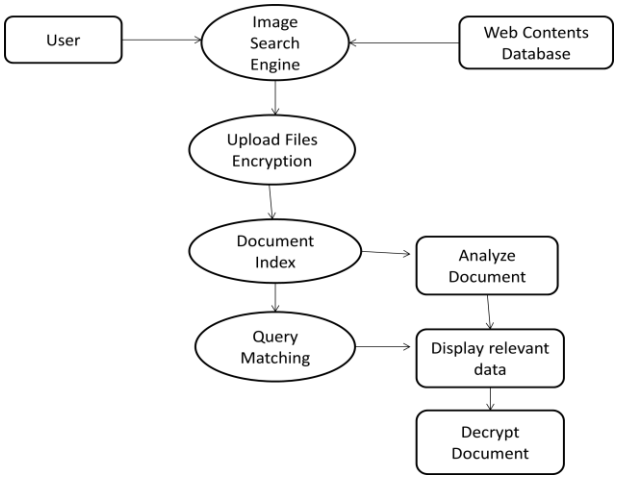
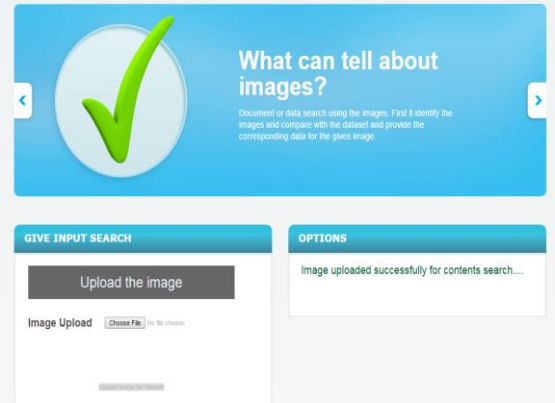


Diagram: Implementation

V. Experimental result

The image is first uploaded in the database. The documents related to the image are also stored in the database. Once the user uploaded the image it is encrypted and then search the content related to the image by using the K-nn algorithm. Now the indexing is done and it search the relevant document. Once it encrypted it cannot be decrypted unless the user have the security key. Because it cannot be decrypted using the private key.



It then display the relevant document to the image which is given as input.

VI. Conclusion

In the image domain, the idea is to start from an image query complemented by some other (possibly not relevant) images that contextualize the query by associating the right semantic concept. Even if the approach is simple, it is indeed an effective query model, more flexible with respect to the usual Query. It have advantage of being highly efficient makes them good candidates for practical web applications that have less stringent requirement on security but demand high efficiency and least user involvement. Index randomization techniques have the advantage of being highly efficient and requiring minimum user-involvement. All this is achieved at the small cost of a few additional hundred bytes of storage for each page. While in this work we have focused on a reranking strategy, we believe that our framework is sufficiently efficient to support in the near future the application of a single joint search model over text and images in the Web collection.

References

1. Sergio Rodriguez-Vaamonde, Lorenzo Torresani, Member, IEEE, and Andrew W. Fitzgibbon, Senior Member, IEEE What Can Pictures Tell Us About Web Pages? Improving Document Search Using Images in vol.37, no.6, June 2015.
2. Kalpana.R and Elavarasi.K, "Intelligent transport system for human detection with an efficient HOG Extraction method" Internationsal journal of innovative research in Computer Science and Technology, ISSN:2347-5552, Volume: 3, ISSUE-3, MAY-2015.
3. Raunak Joshi1, Bharat Gutal2, Rajkumar Ghode3, Manoj Suryawanshi4, Prof U.H. Wanaskar5, Data Mining Using Secure Homomorphic Encryption in Vol. 4, Issue 10, October 2015.
4. Wenjun lu1, avinash l. Varna2, (member, ieee), and min wu3, (fellow, ieee),Confidentiality-Preserving Image Search: A Comparative Study Between Homomorphic Encryption and Distance-Preserving Randomization in Mar 2014.
5. Q. Yu, S. Shi, Z. Li, J.-R. Wen, and W.-Y. Ma, "Improve ranking by using image information," inProc. Adv. Inf. retrieval, 29th Eur. Conf. IR Res., Rome, Italy, 2007, pp. 645–652.
6. Z.-H. Zhou and H.-B. Dai, "Exploiting image contents in web search." inProc. Int. J. Conf. Artif. Intell., 2007, pp. 2922–2927.
7. T. Yeh, J. J. Lee, and T. Darrell, "Photo-based question answering," inProc. 16th ACM Int. Conf. Multimedia, ser. MM '08, New York, NY, USA: ACM, 2008, pp. 389–398.
8. J. Krapac, M. Allan, J. J. Verbeek, and F. Jurie, "Improving web image search results using query-relative classifiers," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2010, pp. 1094–1101.
9. Maha TEBAA, Saïd EL HAJJI, Abdellatif EL GHAZI, Homomorphic Encryption Applied to the Cloud Computing Security in Vol I July 2012
10. N. Bhatia et al, Survey of Nearest Neighbor Techniques. International Journal of Computer Science and Information Security, Vol. 8, No. 2, 2010.

