iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 6 **Issue:** II **Month of publication:** February 2018

**DOI:** http://doi.org/10.22214/ijraset.2018.2083

# A Study on Large-Scale Cross-Media Retrieval of Wikipedia Images towards Visual Query and Textual Expansion

Monelli Ayyavaraiah[1]
[1]*Department of Information Technology, Mahatma Gandhi Institute of Technology, Hyderabad*

***Abstract: In this paper, we present our methodologies for the Wikipedia MM assignment at Image CLEF 2008. We initially tried different things with a content based image retrieval approach with query expansion, where the augmentation terms were naturally chosen from an information base that was semi-consequently developed from Wikipedia. Reassuringly, the exploratory results rank in any case among all submitted runs. We additionally actualized a substance based image retrieval approach with query-dependent visual concept detection. At that point cross-media retrieval was effectively done by independently applying the two meta-look instruments and afterward consolidating the results through a weighted summation of scores. Despite the fact that not presented, this approach beats our content based and substance based methodologies remarkably.***
***Keywords: textual query expansion, Image retrieval, cross-media re-ranking, query-dependent visual concept detection***.

## I. INTRODUCTION

The Wikipedia MM undertaking at Image CLEF 2008 plans to examine successful retrieval approaches in an expansive scale gathering of Wikipedia images. In the undertaking, members need to manage looking 75 themes from around 150,000 images. Pursuit over such an expansive scale image gathering offers numerous difficulties. Among them, the most glaring test is the purported semantic hole [8]. Indeed, even in the circumstance where images are related with some textual portrayals, this semantic hole is as yet present since they don't completely capture every one of the nuances of the semantics of the images.

To address the semantic hole issue, we tried different things with a few image retrieval approaches on the Wikipedia MM dataset. A retrieve motor was actualized in this investment, which comprises of four segments respectively for information pre-handling, content based image retrieval (TBIR), content-based image retrieval (CBIR), and cross-media retrieval. In TBIR, textual query expansion strategy is utilized where the augmentation terms are consequently chosen from a learning base (KB) that is semi-naturally developed from the online vast scale reference book — Wikipedia.

Reassuringly, the test results rank in any case among all submitted runs. For CBIR, visual query expansion is utilized through query-dependent visual concept detection to semantically comment on images or increase their harsh semantics gathered from related content. By examination, this approach performs superior to anything the other submitted CBIR runs. At that point cross-media retrieval is performed by independently applying the two meta-seek devices and afterward consolidating the results through a weighted summation of scores. Despite the fact that not presented, this approach beats our content based or substance based methodologies remarkably.

The rest of this paper is composed as takes after. Textual and visual query expansion approaches for two meta-seek apparatuses are depicted respectively in Section 2 and 3. At that point the cross-media re-ranking methodology is presented in Section 4. The trial results are appeared in Section 5. At last we reach a determination in Section 6.

## II. TEXTUAL QUERY EXPANSION FOR TBIR

A characteristic answer for Wikipedia MM 2008 undertaking is to utilize TBIR technique. To enable the retrieval framework to draw near to clients' real plan, query expansion methods are frequently utilized by adding terms to inquiries or changing preliminary questions. In this investment, we center around how to consequently remove the expansion terms from a KB that is semi-naturally developed from Wikipedia. Sorted out with concepts distinguished by URLs and connections amongst concepts and outside hubs, Wikipedia isn't just a Web accumulation yet in addition an online information focus which amasses every one of clients' insights. Therefore, it is normally appealing and promising that this open, and always advancing reference book can yield modest information structures that can be abused to upgrade the semantics of questions.

Recently, "Wikipedia mining" has been addressed as another research theme. WikiRelate [2] utilized connection based way length

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887*
*Volume 6 Issue II, February 2018- Available at www.ijraset.com*

for processing relatedness for given concepts; Nakayama et al. [3] proposed the PFIBF (Path Frequency – Inversed Backward connection Frequency) calculation for Web thesaurus development. Be that as it may, none of work is made on utilizing Wikipedia as the KB in data retrieval.

In Wikipedia, each non-managerial page is utilized as a term/concept depicting people (e.g., Jingtao Hu), concepts (e.g., Emissions exchanging), areas (e.g., Big Ben), occasions (e.g., crumple of the World Trade Center), and classifications (e.g., microbiology). For a given term, the related terms can be effectively separated from the corresponding Wikipedia pages, and after that used to expand the query when this term is utilized as the query input. At last, the expanded query is nourished into the retrieval motor to produce the last indexed lists. In our execution, we utilize the TF-IDF worldview for content retrieval which has been generally utilized as a part of content mining and data retrieval.

As appeared in Fig. 1, three stages are utilized to develop the KB from Wikipedia:

### *A. Near Pages Selection*

We initially download and file all Wikipedia pages with TF-IDF show. Just pages with a closeness score higher than threshold is set to be 0.9 in our analyses) are picked as the related pages of the info query.

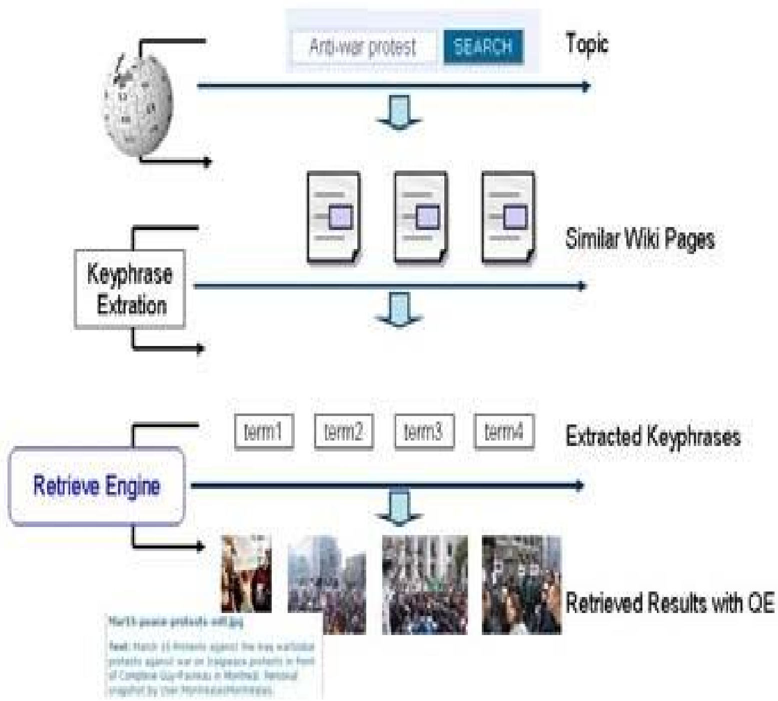



Fig. 1. Textual query expansion using the KB constructed from Wikipedia

### *B. Page Keyphrase Extraction*

In a Wikipedia page, keyphrases or watchwords quickly depict the substance of a concept. In this manner they can be utilized to improve the semantics of that concept. In our framework, we utilize an unsupervised keyphrase extraction calculation presented in our previous work [4]. By treating content in a page as a semantic system, this calculation figures a few structure factors of Small-World

### *C. Term Selection for Query Expansion*

By and by, the best positioned keyphrases can't be directly utilized for query expansion. For example, while seeking "saturn", term "moon" is extricated as the keyphrase with a high score, however "moon" may show up on numerous pages and ought to be considered more broad. To address this issue, a factual feature Inverse Backward connection Frequency (ibf) [3] is figured as:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 6 Issue II, February 2018- Available at www.ijraset.com

$$ibf \,\square\, \log\left(\frac{N}{bf(t)\,\square\square}\right), \qquad (1)$$

where bf (t) is the quantity of in reverse connections in which the connection content contains term t, N indicates the aggregate number of pages and is a parameter on the off chance that bf (t) is zero. There-fore, the last weight of a keyphrase can be processed as:

$$w_{t_k} \,\square K \,\square\, P(t_k) \,\square\, ibf(t_k). \qquad (2)$$

By then the keyphrases with their institutionalized weights are joined with the principal request to fabricate an extended inquiry to be fed into the recover engine.

## III. QUERY-DEPENDENT VISUAL CONCEPT DETECTION FORCBIR

In the Wikipedia MM dataset, a few images have few or even no enlightening writings. To address this issue, query-dependent visual concept detection can be utilized to semantically comment on images or increase their harsh semantics gathered from related content. Given the pre characterized query concepts, 1-versus all visual concept indicators are prepared for every one of these concepts by utilizing the preparation images acquired by Yahoo! seek. Plainly, these preparation images can be utilized for visual query expansion to improve the CBIR assignment. As appeared in Fig 2, the preparation procedure incorporates the accompanying three stages:

### A. *Building the preparation set.*

For each query concept, top k (k=30 in our framework) images are ripped at from Yahoo! image web crawler. At that point some unrelated images with respect to the concept are physically filtered out which frames a positive preparing set. Negative images for every concept are haphazardly chosen from positive images of alternate concepts.

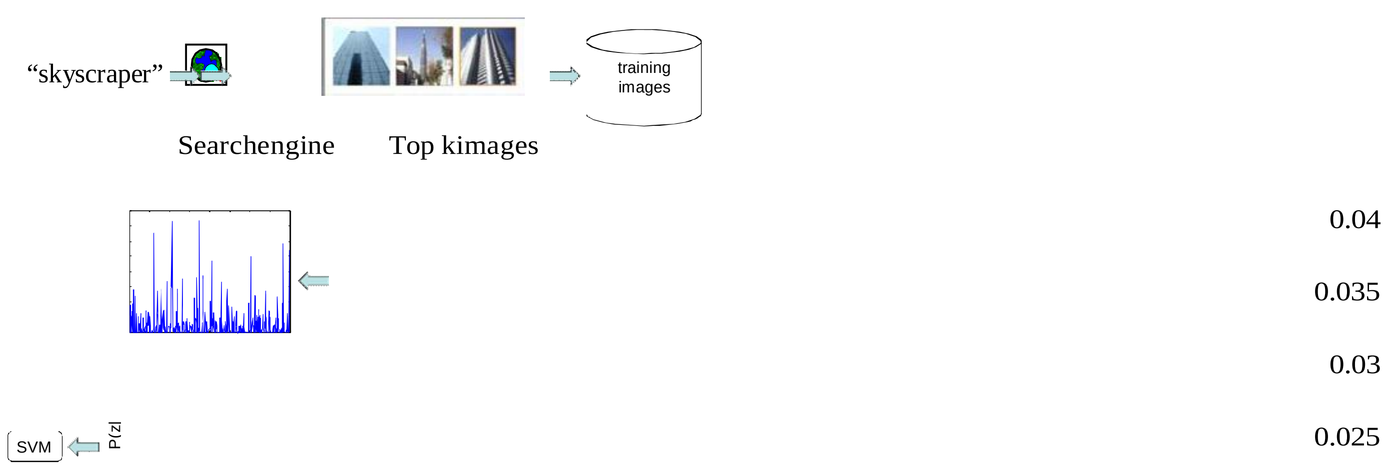
### B. *Building Bag of Words (BOW) representation*

Filter [5], Dense-SIFT [6] and Color-Dense-SIFT are removed from the preparation sets of all concepts. At that point k-implies calculation is utilized to quantize different kinds of features and create a joined visual codebook. All images are represented by an arrangement of tokens of the visual words.

### C. *Supervised preparing for every subject*

Unsupervised probabilistic inactive semantic investigation (pLSA) [7] is used to construe the idle subject appropriation of the preparation images in view of the BOW representation. At that point bolster vector machine (SVM) is utilized to prepare a one-class classifier for every concept in the inactive subject space.

Fig. 2. Query-dependent visual concept detection for CBIR



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887*
*Volume 6 Issue II, February 2018- Available at www.ijraset.com*

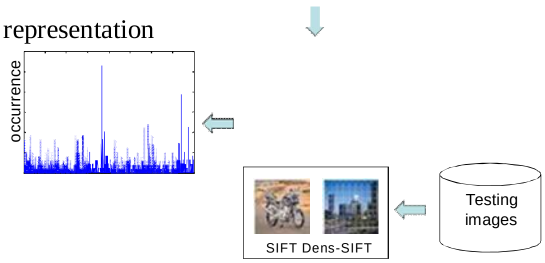
0.02

0.015

0.01

0.005

topic representation bag-of-words representation




30

25

20

15

10

5

0

SVM raining/testing

Feature extraction Given the prepared 1-versus all visual concept finders for all query subjects, we can per-frame the concept detection for each test image by initially representing it with the visual words from the prepared codebook, surmising its inactive point dispersion in view of the prepared p LSA display, lastly processing the responds of the prepared SVMs for different concepts. Thought is recognized exactly when the contrasting react is over a given edge. For CBIR, test pictures are finally situated by their re-sponds with respect to the request thought.

## IV. EXPERIMENTS

This area depicts our analyses for the Wikipedia MM undertaking. Note that a portion of the test results reported here were not submitted before the due date. The analyses are assessed by MAP (Mean Average Precision), P@N (preci-sion of best N images), and R-precision. The ground-truth results are given in the assessment period of the Wikipedia MM undertaking.

### *A. Experiments with TBIR*

The primary arrangement of investigations is to assess the execution of TBIR approach with different query expansion strategies.

1) *Query expansion by utilizing the naturally developed KB.* Different strategies are utilized to naturally build KB from Wikipedia for query expansion, by utilizing different content sources (e.g., titles, connections or full text of Wikipedia articles) and different term determination calculations (e.g., TFIDF-based, Small-World (SW)- based, SWIBF-based). Therefore, four programmed query expansion strategies were assessed in our trials, respectively meant by QE-Title-TFIDF, QE-Link-TFIDF, QE-Fulltext-SW, and QE-Fulltext-SWIBF. We additionally utilize NO-QE to indicate TBIR without query expansion. In all trials, just best 20 terms are utilized. Shockingly, all these programmed query expansion techniques cannot essentially enhance

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 6 Issue II, February 2018- Available at www.ijraset.com

the TBIR execution, compared with NO-QE (See Table 1). Hence we ought to consider how to enhance the nature of the developed KB.

2) *Query expansion by utilizing the semi-consequently developed KB.* After the KB was consequently built from Wikipedia, we at that point played out some manual confirmations. Here we utilize QE-Full text-Semi to indicate this query expansion strategy. Note that for this situation, the query expansion strategy still consequently chooses terms from the KB to semantically grow a given query term. From Table 1, we can see that this QE-Fulltext-Semi strategy performs much superior to every single other model.

Table 1. The experimental results of different textual query expansion methods

| Run ID | QE | Modality | MAP | P@5 | P@10 | R-Prec |
|---|---|---|---|---|---|---|
| NO-QE | without | TXT | 0.2565 | 0.4427 | 0.3747 | 0.2929 |
| QE-Title-TFIDF | with | TXT | 0.2566 | 0.4187 | 0.3627 | 0.2967 |
| QE-Link-TFIDF | with | TXT | 0.2271 | 0.376 | 0.3147 | 0.2533 |
| QE-Fulltext-SW | with | TXT | 0.2365 | 0.3733 | 0.336 | 0.2618 |
| QE- Fulltext-SWIBF | with | TXT | 0.2609 | 0.44 | 0.3693 | 0.2859 |
| QE- Fulltext-SEMI | with | TXT | 0.3444 | 0.5733 | 0.476 | 0.3794 |

### B. *Experiments with CBIR*

Compared with TBIR, our CBIR acquired an equivalent precision in the best positioned images (P@5=0.5307 and P@10= 0.4507 of CBIR versus P@5=0. 5733 and P@10=0.476 of TBIR), yet much lower MAP (0.1928 of CBIR versus 0.3444 of TBIR) and R-Prec (0.2295 of CBIR versus 0.3794 of TBIR). Albeit visual substance vagueness reduces the general execution (MAP) by returning images with comparative low-level features, the test results demonstrate that taking in visual models from Web images (e.g., from Yahoo! look) do rank the substance relevant images higher. It additionally ought to be noticed that, our CBIR approach performs best among all submitted CBIR keeps running in Wikipedia MM 2008 assignment.

Table 2. The experimental results of CBIR

| Run ID | QE | Modality | MAP | P@5 | P@10 | R-Prec |
|---|---|---|---|---|---|---|
| *CBIR run1* | with | IMG | 0.1912 | 0.5333 | 0.4427 | 0.2929 |
| *CBIR run2* | with | IMG | 0.1928 | 0.5307 | 0.4507 | 0.2295 |

### C. *Experiments with Cross-Media Retrieval*

In the last arrangement of investigations, cross-media retrieval approach is utilized to accomplish better execution by consolidating content based and substance based retrieval results. In the tests, we set M2 littler than M1. This infers only the best situated pictures returned by CBIR are consolidated into the re-situating stage since the lower-situated pictures may have altogether higher probabilities to be disturbances. Table 3 demonstrates the exploratory results, where ReRank-Text-Visual-N means the mix of CBIR and TBIR without query expansion, and ReRank-Semi-Visual-N signifies the mix of CBIR and TBIR with self-loader query expansion, and N indicates the corresponding parameter in Eq. (5).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887*
*Volume 6 Issue II, February 2018- Available at www.ijraset.com*

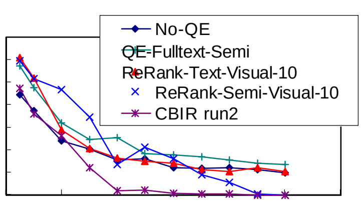
Table 3.Some experimental results of cross-media retrieval

| Run ID | QE | Modality | MAP | P@5 | P@10 | R-Prec |
|---|---|---|---|---|---|---|
| NO-QE | without | TXT | 0.2565 | 0.4427 | 0.3747 | 0.2929 |
| CBIR run2 | without | IMG | 0.1928 | 0.5307 | 0.4507 | 0.2295 |
| ReRank-Text-Visual-10 | without | TXTIMG | 0.3099 | 0.608 | 0.5213 | 0.3387 |
| ReRank-Text-Visual-20 | without | TXTIMG | 0.3035 | 0.6027 | 0.512 | 0.3420 |
| ReRank-Text-Visual-40 | without | TXTIMG | 0.2972 | 0.584 | 0.4893 | 0.3393 |
| ReRank-Text-Visual-60 | without | TXTIMG | 0.2928 | 0.5547 | 0.4733 | 0.3366 |
| ReRank-Text-Visual-80 | without | TXTIMG | 0.2910 | 0.5387 | 0.4693 | 0.3349 |
| QE- Fulltext-SEMI | with | TXT | 0.3444 | 0.5733 | 0.476 | 0.3794 |
| CBIR run2 | with | IMG | 0.1928 | 0.5307 | 0.4507 | 0.2295 |
| ReRank-Semi-Visual-10 | with | TXTIMG | 0.3584 | 0.6293 | 0.5147 | 0.3993 |
| ReRank-Semi-Visual-20 | with | TXTIMG | 0.3568 | 0.6187 | 0.5147 | 0.3974 |
| ReRank-Semi-Visual-40 | with | TXTIMG | 0.3519 | 0.5867 | 0.5013 | 0.3988 |
| ReRank-Semi-Visual-60 | with | TXTIMG | 0.3487 | 0.568 | 0.492 | 0.3988 |
| ReRank-Semi-Visual-80 | with | TXTIMG | 0.3483 | 0.5653 | 0.4907 | 0.3988 |

0.7
0.6
0.5
0.4
0.3
0.2
0.1

Precision

Precision of Top N Results




0 2040 60 80 100120

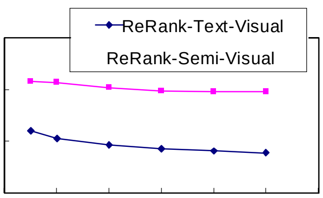
0.4

MAP

0.35

.3
.25

MAP of Re-Ranking




0 20 40 60 80 100120

N

i. (a) (b)

Fig. 3. Performance of cross-media retrieval: (a) P@N and (b) MAP results with different values of *N* in Eq. (5)

From Table 3 and Fig. 3, it's interesting to find that when N increases, the preliminary result of every framework is more prone to be similarly treated and the general execution decreases. For the blend of CBIR and content based retrieval without query expansion, the normal change of the considerable number of questions in ReRank-Text-Visual-10 is around 5.34% over the single content based retrieval approach (25.65% of MAP). While for the blend of CBIR and content based retrieval with self-loader query expansion, the normal change for every one of the inquiries in ReRank-Semi-Visual-10 is around 1.4% over the single content based retrieval approach (34.44% of MAP). We additionally watched that the cross-media retrieval results have significantly higher precision of best positioned images than both content based retrieval or CBIR results. By and large talking, content based retrieval can return more relevant images via looking catchphrases with image portrayals, while CBIR can get high precision of best positioned images however an excessive number of commotions in bring down positioned images. Accordingly consolidating CBIR with content based retrieval can help increase the precision of best positioned images. Taking everything into account, the cross-media retrieval approach performs remarkably well. This demonstrates cross-media combination is unquestionably a promising direction to explore compelling retrieval approaches with regards to an extensive scale and heterogeneous gathering of images.

## V. CONCLUSION AND FUTUREWORK

This paper reported our methodologies for the Wikipedia MM errand at Image CLEF 2008. We tried different things with TBIR, CBIR and cross-media image retrieval approaches with query expansion. Reassuringly, the test results of our TBIR approach rank in any case among all submitted runs. In spite of not presented, the cross-media approach performs much superior to the single TBIR or CBIR approaches. Additionally analyses will be finished by streamlining the KB development procedure and considering better cross-media re-ranking methodologies.

## REFERENCES

[1] Tian, Y.H., Huang, T.J., Gao, W.: Exploiting multi-context nalysis in semantic image clas- sification. J. Zhejiang Univ. SCI. 6A(11), 1268–1283(2005)
[2] Strube, M., Ponzetto, S.: WikiRelate! Computing semantic relatedness using Wikipedia. In: Proc. of National Conference on Artificial Intelligence (AAAI 2006), Boston, Mass, pp. 1419–1424(2006)
[3] Nakayama, K., Hara, T., Nishio, S.: A thesaurus construction method from large scale web dictionaries. In: Proc. of IEEE International Conference on Advanced Information Networking and Applications (AINA 2007), pp. 932–939(2007)
[4] Huang, C., Tian, Y.H., Zhou, Z., Ling, C.X., Huang, T.J.: Keyphrase extraction using Se- mantic Networks Structure Analysis. In: Proc. of the sixth IEEE

Int'l. Conf. on Data Mining (ICDM 2006), pp. 275–284. IEEE press, Hong Kong(2006)

[5] Lowe, D.: Object recognition from local scale-invariant feature. In: Proc. Int'l Conf. Computer Vision (ICCV 1999), pp. 1150–1157(1999)

[6] Lazebnik, S., Schmid, C., Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In: Proceedings of the IEEE CVPR 2006, pp. 2169–2178(2006)

[7] Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. Machine Learning 41, 177–196(2001)

[8] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach Intell. 22(12), 1349– 1380(2000)

[9] Shoban Babu Sriramoju, "Analysis and Comparison of Anonymous Techniques for Privacy Preserving in Big Data" in "International Journal of Advanced Research in Computer and Communication Engineering",Vol 6, Issue 12, December 2017, DOI 10.17148/IJARCCE.2017.61212[ ISSN(online) : 2278-1021, ISSN(print) : 2319-5940 ]

[10] Namavaram Vijay, S Ajay Babu, "Heat Exposure of Big Data Analytics in a Workflow Framework" in "International Journal of Science and Research", Volume 6, Issue 11, November 2017, 1578 - 1585, #ijsrnet

[11] Shoban Babu Sriramoju, " Review on Big Data and Mining Algorithm" in "International Journal for Research in Applied Science and Engineering Technology", Volume-5, Issue-XI, November 2017, 1238-1243 [ ISSN : 2321-9653], www.ijraset.com

[12] Shoban Babu Sriramoju, "OPPORTUNITIES AND SECURITY IMPLICATIONS OF BIG DATA MINING" in "International Journal of Research in Science and Engineering",Vol 3, Issue 6, Nov-Dec 2017 [ ISSN : 2394-8299 ].

[13] Shoban Babu Sriramoju, "Heat Diffusion Based Search for Experts on World Wide Web" in "International Journal of Science and Research", https://www.ijsr.net/archive/v6i11/v6i11.php, Volume 6, Issue 11, November 2017, 632 - 635, #ijsrnet

[14] Dr. Shoban Babu Sriramoju, Prof. Mangesh Ingle, Prof. Ashish Mahalle "Trust and Iterative Filtering Approaches for Secure Data Collection in Wireless Sensor Networks" in "International Journal of Research in Science and Engineering" Vol 3, Issue 4, July-August 2017 [ ISSN : 2394-8299 ].

[15] Ajay Babu, Namavaram Vijay, Ramesh Gadde, "SKETCHING-BASED HIGH-PERFORMANCE BIG DATA PROCESSING ACCELERATOR" in "International Journal of Research in Science and Engineering",Vol-3, Issue-6, Nov-Dec 2017, 92-99 [ ISSN : 2394-8299 ].

[16] Dr. Shoban Babu, Prof. Mangesh Ingle, Prof. Ashish Mahalle, "HLA Based solution for Packet Loss Detection in Mobile Ad Hoc Networks" in "International Journal of Research in Science and Engineering" Vol 3, Issue 4,July-August 2017 [ ISSN : 2394-8299 ]

[17] Shoban Babu Sriramoju, "A Framework for Keyword Based Query and Response System for Web Based Expert Search" in "International Journal of Science and Research" Index Copernicus Value(2015):78.96 [ ISSN : 2319-7064 ].

[18] Mounika Reddy, Avula Deepak, Ekkati Kalyani Dharavath, Kranthi Gande, Shoban Sriramoju, "Risk-Aware Response Answer for Mitigating Painter Routing Attacks" in "International Journal of Information Technology and Management" Vol VI, Issue I, Feb 2014 [ ISSN : 2249-4510 ]

[19] Mounica Doosetty, Keerthi Kodakandla, Ashok R, Shoban Babu Sriramoju, "Extensive Secure Cloud Storage System Supporting Privacy-Preserving Public Auditing" in "International Journal of Information Technology and Management" Vol VI, Issue I, Feb 2012 [ ISSN : 2249-4510 ]

[20] Shoban Babu Sriramoju, "Multi View Point Measure for Achieving Highest Intra-Cluster Similarity" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2,Issue 3,March 2014 [ ISSN(online) : 2320-9801, ISSN(print) : 2320-9798 ]

[21] Shoban Babu Sriramoju, Azmera Chandu Naik, N.Samba Siva Rao, "Predicting The Misusability Of Data From Malicious Insiders" in "International Journal of Computer Engineering and Applications" Vol V,Issue II,Febrauary 2014 [ ISSN : 2321-3469 ]

[22] Ajay Babu Sriramoju, Dr. S. Shoban Babu, "Analysis on Image Compression Using Bit-Plane Separation Method" in "International Journal of Information Technology and Management", Vol VII, Issue X, November 2014 [ ISSN : 2249-4510 ]

[23] Shoban Babu Sriramoju, "Mining Big Sources Using Efficient Data Mining Algorithms" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2, Issue 1,January 2014 [ ISSN(online) : 2320-9801, ISSN(print) : 2320-9798 ]

[24] Ajay Babu Sriramoju, Dr. S. Shoban Babu, "Study of Multiplexing Space and Focal Surfaces and Automultiscopic Displays for Image Processing" in "International Journal of Information Technology and Management"Vol V, Issue I, August 2013 [ ISSN : 2249-4510 ]

[25] Shoban Babu Sriramoju, Dr. Atul Kumar, "An Analysis around the study of Distributed Data Mining Method in the Grid Environment : Technique, Algorithms and Services" in "Journal of Advances in Science and Technology" Vol-IV, Issue No-VII, November 2012 [ ISSN : 2230-9659 ]

[26] Namavaram Vijay, Ajay Babu, Ramesh Gadde, "Two Layered Privacy Architecture for Big Data Framework" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 5, Issue 10, October 2017 [ ISSN(online) : 2320-9801, ISSN(print) : 2320-9798 ]

[27] Shoban Babu Sriramoju, Dr. Atul Kumar, "An Analysis on Effective, Precise and Privacy Preserving Data Mining Association Rules with PartitioningDistributed Databases" in "International Journal of Information Technology and management" Vol-III, Issue-I, August 2012 [ ISSN : 2249-4510

[28] Shoban Babu Sriramoju, Dr. Atul Kumar, "A Competent Strategy Regarding Relationship of Rule Mining on Distributed Database Algorithm" in "Journal of Advances in Science and Technology" Vol-II, Issue No-II, November 2011 [ ISSN : 2230-9659

[29] Ramesh Gadde, Namavaram Vijay, "A SURVEY ON EVOLUTION OF BIG DATA WITH HADOOP" in "International Journal of Research in Science and Engineering",Vol-3, Issue-6, Nov-Dec 2017, 92-99 [ ISSN : 2394-8299 ].

[30] Shoban Babu Sriramoju, Dr. Atul Kumar, "Allocated Greater Order Organization of Rule Mining utilizing Information Produced Through Textual facts" in "International Journal of Information Technology and management" Vol-I, Issue-I, August 2011 [ ISSN : 2249-4510 ]