INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.,

Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],

Index Copernicus Publishers Panel, Polandwith IC Value of 5.09 &number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 2840 Cities in 164 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

http://ijrcm.org.in/

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography: Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** ***anytime*** in ***M.S. Word format*** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website (***FOR ONLINE SUBMISSION, CLICK HERE***).

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____

THE EDITOR
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF.**

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '_____' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:

a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mentionthe following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/ Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESES

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

SCOPE FOR FURTHER RESEARCH

ACKNOWLEDGMENTS

REFERENCES

APPENDIX/ANNEXURE

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES &TABLES:** These should be simple, crystal clear, centered, separately numbered & self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS:**These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

OPTIMIZING THE DE-DUPLICATION FOLIAGE IMAGE ACCESS IN STORAGE SYSTEMS

CHITTALA RAMA SATYA
STUDENT
ST. MARTIN'S ENGINEERING COLLEGE
HYDERABAD

B.VIJAYA KUMAR
PROFESSOR
ST. MARTIN'S ENGINEERING COLLEGE
HYDERABAD

ABSTRACT

We are now in the era of the Big Data revolution where nearly every aspect of computer engineering is being driven by large-data processing and analysis. The vast volume of data is used on Personalized Searches in normal or internet applications. Valid of data is important for accessing system without De-duplication. In this paper, we proposed to optimize the de-duplicated data using Map Reduce with the help of similarity based methodology that are used for comparable system. To access the data are in data storage applications using point-to-point process. The experimental results are conducting the probability testing is applied on large data sets on ternary clusters.

KEYWORDS

Big Data, De-Duplication, Foliage Images, Optimize, Point-to-Point, Probability Testing, Ternary.

INTRODUCTION

In our daily life, increasing in the business data growth rates is high, rapidly changes in coming years. Likewise, retention and retrieval requirements for new and existing data will expand, driving still more data to disk storage. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, which is forming Big Data. It is a collection of datasets. The data sizes are constantly moving from a few terabytes to many petabytes of data in a single data set. So it becomes difficult to process using on-hand database management tools or traditional data processing applications. With this difficulty, new platforms of "big data" tools are being developed to handle various aspects of immense quantities of data such as web logs, social networks and search indexing. Using key-based word searching we can search for words anywhere in the record.

Keyword searches are useful when we don't have complete information about what we are trying to search. For example it is a good substitute for a title or author search when we have incomplete title or author information. But by using keyword searches we can't suppress the redundant data of many users. There is ever-increasing focus on improving data storage efficiencies across the information infrastructure. Data reduction is a tactic which can decrease the disk storage; network bandwidths required and optimize use of existing storage assets. Deduplication is a well-known technique for data reduction process that means eliminating the redundant data in sequential-access disk and from active-data storage pools. One unique instance of the data is retained on storage media, and redundant data is replaced with a pointer to the unique data copy. The goal of Deduplication is to reduce the overall amount of time that is required to retrieve data by letting you store more data on disk.

The proposed system incorporates large data sets i.e., Big Data by forming three ternary clusters namely, Tag cluster, User cluster and Image data cluster. Also, all the information kept into three clusters. Clusters has played a critical role for pattern recognition, image segmentation i.e., which provides similar objects to create groups in an un-labeled data. Various clustering algorithms are created to define a partitioning of a dataset. It is importance of parameters are assigned an improper value, the clustering method results in a partitioning scheme that is not optimal for the specific data set leading to wrong decisions. The problems of deciding the number of clusters better fitting a dataset as well as the evaluation of the clustering results has been subject of many research efforts

The rest of the paper is organized as follows: Section 1 deals with present surveys of related work. Section 2 offers fuzzy classifications of decision personnel signature dataset. Section 3 contains details about support vector machine. Section 4 presents the study of foliage image using Map Reduce technique. Section 5 discussed the similarity technique applied on foliage image using Map Reduce. Section 6 presents the experimental results obtained. Finally, we conclude the conclusion and future works.

1. PERSONALIZED SEARCH

In previously discussed the personalized image search framework by various authors are simultaneously considering user and query information discussed in Learn to Personalized Image Search From the Photo Sharing Websites by Jitao Sang et al., published in IEEE Transaction Multimedia, 2012 [4]. The user's preferences over images under certain query are estimated by how probable he/she assigns the query-related tags to the images. A ranking based tensor factorization model named RMTF is proposed to predict users' annotations to the images. To better represent the query-tag relationship, we build user-specific topics and map the queries as well as the users' preferences onto the learned topic spaces. In this part, three modules are follows

1.1. USER-SPECIFIC TOPIC MODELING

User-specific information is considered to distinguish the exact intentions of the user queries and to re-rank the list results. Users may have different intentions for the same query for example; searching for "jaguar" by a car fan has a completely different meaning from searching by an animal specialist.

1.2. PERSONALIZED IMAGE SEARCH

In the research community of personalized search, evaluation is not an easy task since relevance judgment can only be evaluated by the searchers themselves. This approach is very expensive and needs large-scale real search logs. Social sharing websites provide rich resources that can be exploited for personalized search evaluation. User's social activities, such as rating, tagging and commenting, indicate the user's interest and preference in a specific document.

1.3. RANKING

Photo sharing websites differentiate from other social tagging systems by its characteristic of self-tagging: most images are only tagged by their owners. If any problem occurs to enable information propagation it calls for external resources. Here, in addition to the ternary interrelations, we also collect multiple intra-relations among users, images and tags. We assume two items with high affinities should be mapped close to each other in the learnt factor subspaces. Previous works on these lines were proposed by various authors; Jitao Sang is discussed in Learn to Personalized Image Search from the Photo Sharing Websites, IEEE Transactions on Multimedia vol.14, August 2012[4, 5, 6, and 7].

2. TERNARY CLUSTER ANALYSIS

Classification is the set of objects which are assigning to one among a set of pre-defined categories. It is a systematic approach for building classification models from an input dataset. Consider, Entire input personnel signature can be segmented into three clusters i.e., foliage (leaf) character images, users and Tagging. The cluster data is represented in the form of triplets.

Let U, I, T denote the sets of users, Images and Tags. The group of data of a cluster is denoted by C⊂U X I X T, i.e., each triplet (u, i, t) ϵ C means that user u has annotated image i with tag t. the ternary interrelations can then constitute a three dimensional tensor y ϵ $R^{|U| X|I|X|T|}$ is applied on map reduce [10].

3. SUPPORT VECTOR MACHINE

Support Vector Machine is a machine learning process. It is defined over a vector space in which the problem is to find a decision surface that "best" separates the data vectors into two classes suggested by Isabelle Moulinier et al., 1997 [4] is discussed the simplest linear form; an SVM is a hyper-plane that separates a set of positive from a set of negative with maximum margin.

The hyper-plane (dot) used on the database images and represented as linear SVMs or it can be found in a higher-dimensional space by transforming the images into a representation having more dimensions like input variables than the whole database in data images are treated as non-linear SVMs. It is used to provide simple solution by mapping the image data into a higher dimensional space and then reducing the problem to a linear.

The purpose of hyper-plane is to separates the training data by a maximal margin. All vectors lying on one side of the hyper-plane are labeled as 0, and all vectors lying on the other side are labeled as 1. The training instances that lie closest to the hyper-plane are called support vectors [1, 4 and 9].

The Linear Support Vector Machine S = **w.i** (3.1)

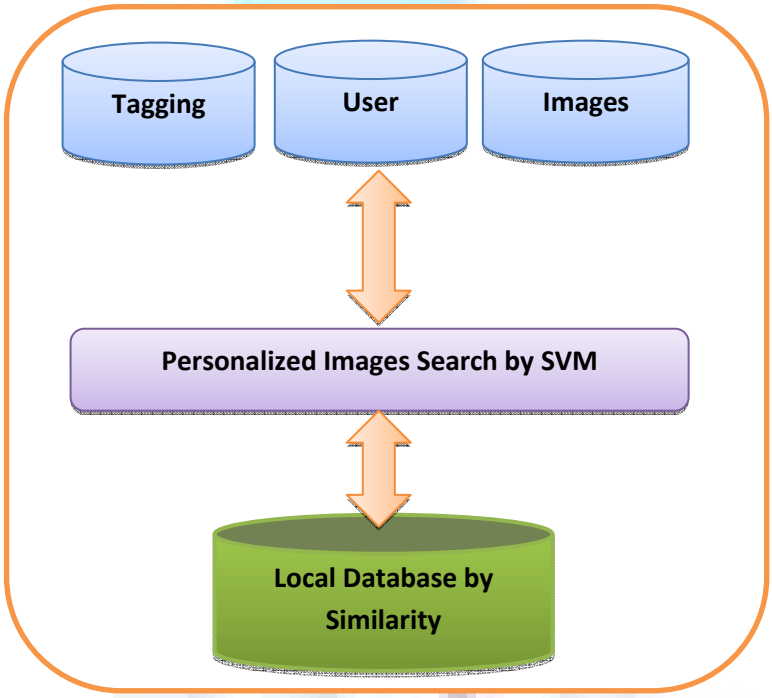
Where w is the normal vector namely called as database to the hyper-plane, and i is the input vector is known as Input Data. In the linear case, the margin is defined by the distance of the hyper-plane to the nearest of the positive and negative.

4. MAP REDUCE BASED T-NARY CLUSTER OPTIMIZATION USING SIMILARITY

The process requires validation that the target database is clean and contains the non-redundant data, from an identity perspective. The identity de-duplication process involves comparing all records of the image database among them, powered by complex mathematical algorithms that discover the some statistical study of similarity relationships within datasets.

When we send a request from local database to cloud database the requirement of basic algorithm for any recognition can be solved. The interface of classifier with the help of SVM classifier is followed by two classes namely called cloud database and local database by using map reduce classification problem; the goal is to find the image object and store to local database by a function which is induced from the mention figure below.

FIG: 4.1: MAP REDUCE BASED PERSONALIZED IMAGE SEARCHES USING SIMILARITY



MapReduce programming model is proposed. MapReduce is a functional programing model that implements for processing and generating large data sets as in[2][3]. It is well-developed technology for distributing and parallelized environments and also working on large data sets. Initially, the user makes a request to the cloud data storage about the triplets. Then the image data produces a set of intermediate key value pairs (images) with the help of map functions written by third party users. The existing cloud data storage is used to calculate the mean value for each image i.e., intermediate key value with the help of image normalization process and feature extraction values. The cloud database consisting of intermediate key values associated with the same intermediate key and passes them to the Reduce function with the help of MapReduce library by SVM classifier. Finally, the user of the reduce function is accepts an intermediate key and a set of values for that key for the database. We compare the mean value with the target image of database with the help of similarity approach to calculate the distances i.e., Euclidean distance.

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{n}(x-y)^2} \quad (4.1)$$

5. EXPERIMENTAL RESULTS

The experimental Results are produced on personnel signature image cluster datasets uses one of the approach test of Hypothesis (the estimating the value of a parameter we need to decide whether to accept or reject a statement about the parameter.), which is to generalize the two-sample software reliability testing if the data belong to more than two groups. The product reliability test is conducted on personnel signatures data sets to test for the equality of the means of two or more normal data set and also uses variances.

$$\textbf{Pobability} = \frac{\textbf{Number of cases when we find failure}}{\textbf{Total number of cases under consideration}}$$

By using this formula, to calculate the failure probability with the help of testing from input images. To find reliability of software, we need to find the respect is done or not from given input datasets [11].

In the above case tagging, user and images are ternary relations data sets. The three equivalents sample data clusters using searching and comparing purposes. To conduct the test whether check the three equal relations data or not and taking the data are equal or not.

Let H_0 is the sample mean are not equal.
Let H_1 is the sample mean are equal
If $H_0 = H_1$ then, Display both are Equal mean. So test is accepted.
Otherwise both are Not Equal, so test is rejected.

6. CONCLUSION AND FUTURE WORKS

De-duplication is one of the best concepts in Data Storage and it is a kind of Cloud Computing application. In this paper, we describe the cloud storage as image data storage which shows that the proposed method can provide the reliable data, fault tolerant and can improve efficiency and scalability. This approach has several applications such as Intrusion Detection and new technologies like Biometric and Video survival lance Systems, etc. Furthermore, the magnitude of the recognition level of an image which will follow some probability distribution using de-duplication.

REFERENCES

1. A.Thousoo, J.S.Sarma, N.Jain, Z.Shao, P. Chakka, S.Anthony, H.Liu, P.Wychoff, and R.Murthy. Hive- a warehousing solution over a map-reduce framework. In VLDB, 2009.
2. A.Thousoo, R.Murthy, J.S.Sarma, Z.Shao, N.Jain, P.Chakka, S.Anthony, H.Liu and N.Zhang. Hive-a petabyte scale data warehousing using hadoop. In ICDE, 2010.
3. Ali Meligy and Manar Al-Khatib, **A** Grid-Based Distributed SVM Data Mining Algorithm, European Journal of Scientific Research, ISSN 1450-216X Vol.27 No.3 (2009), pp.313-321.
4. Isabelle Moulinier Feature Selection: A Useful Preprocessing Step. In Proceedings of the 19th Annual BCS-IRSG Colloquium on IR Research, P.P. 140-158, 1997.
5. J. Teevan and S T Dumais, and D. J. Liebling published in "To personalize or not to personalize: Modeling Queries with variation in user intent, in Proc.SIGIR, 2008, PP 163-170.
6. Jitao Sang and Changsheng Xu and Dongyuan Lu., Learn to Personalized Image Search from Photo Sharing Websites published in IEEE Transactions Multimedia Vol.14, August 2012.
7. Learman K and A. Plangprasophchok, and C. Wong. Published in personalizing image search results on flickr, J.CoRR, Vol abs/0704.1676.2007.
8. Naeem Naik, and L. M. R. J. Lobo, and Riyaz Jamadar. Published in a Novel Personalized Image Search Framework, International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 10, December- 2012, ISSN: 2278-0181.
9. Wiki-pedia.com/software reliability testing.
10. wiki-pedia/Cluster Analysis.
11. wiki-pedia/Support Vector Machine.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Commerce, IT and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail i.e. **infoijrcm@gmail.com** for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals