iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** IV **Month of publication:** April 2016

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Literature Review on Scientometric in Cloud Computing

B. Ashok[1], Munisekhar Prudhvi[2]
[1]*M.Tech, Dept of CSE, SITE Engg. College, Tirupati*
[2]*Asst. prof, Dept of CSE, SITE Engg. College, Tirupati*

Abstract—The prominence and fast advancement of Cloudcomputing lately has prompted a gigantic measure of productions containing the accomplished learning of this zone of examination. Because of the interdisciplinary nature and high importance of Cloudcomputing research, it turns out to be progressively troublesome or even difficult to comprehend the general structure and advancement of this field without expository methodologies. While assessing science has a long custom in numerous fields, we distinguish an absence of a complete scientometric study in the region of Cloudcomputing. In light of a substantial bibliographic information base, this study applies scientometric intends to observationally think about the development and condition of Cloudcomputing research with a perspective from over the mists. By this, we give broad bits of knowledge into distribution designs, research effect and research profitability. Moreover, we investigate the transaction of related subtopics by dissecting catchphrase bunches. The aftereffects of this study give a superior comprehension of examples, patterns and other essential variables as a premise for coordinating exploration exercises, sharing learning and working together in the zone of Cloudcomputing research.

Index Terms—cloud computing, cloud computing research, scientometric analysis, scientometrics, keyword cluster analysis.

I. INTRODUCTION

Cloud computing means storing and accessing data and programs over the Internet instead of your computer's hard drive. The cloud is just a metaphor for the Internet. It goes back to the days of flowcharts and presentations that would represent the gigantic server-farm infrastructure of the Internet as nothing but a puffy, white cumulonimbus cloud, accepting connections and doling out information as it floats. (Griffith, 2013) “Cloud computing” is the next natural step in the evolution of on-demand information technology services and products. To a large extent, cloud computing will be based on virtualized resources. Cloud computing predecessors have been around for some time now, but the term became “popular” sometime in October 2007 when IBM and Google announced collaboration in that domain. This was followed by IBM’s announcement of the “Blue Cloud” effort. Since then, everyone is talking about “Cloud Computing” (Vouk, 2008). Cloud computing is a service where computing is provided as a commodity, much akin to electricity or cable television. Thus, cloud computing is not about a specific technology; rather it is a step in the commoditization of IT enabled by technological advances (Iyoob, Zarifoglu, & Dieker, 2013). 2 Bibliometrics represented a statistical approach to master the growing flood of scientific information and to analyse and to understand the cognitive characteristics of “big science” by measuring quantitative aspect of communication in science and by providing results to scientists & users outside the scientific community. There have been a number of techniques that have been evaluated during the passage of time to evaluate the resources. Bibliometrics is one of the techniques that have been adopted by the library professionals to explore the

impact of any field of knowledge. While bibliometric methods are most often used in the field of library and information science, bibliometrics has wide applications in other areas. In fact, many research fields use bibliometric methods to explore the impact of their field, the impact of a set of researchers, or the impact of a particular paper. It utilizes quantitative analysis and statistics to describe patterns of publication within a given field or body of literature. Researchers may use bibliometric methods of evaluation to determine the influence of a single writer, for example, or to describe the relationship between two or more writers or works, or to identify the pattern of publication and authorship, citations used for a subject etc over a period of time. First scientometric applications were developed to improve use of bibliometric databases and extend information services. Citations were considered documented use of information & have consequently applied first in the context of librarianship, scientific information and information retrieval (Glanzel, 2008).

Scientometric is the branch of science that describes the output traits in terms of organisational research structure, resource inputs

and outputs, develops benchmarks to evaluate the quality of information output. it studies characterize the discipline using the growth pattern and other attributes. These studies have potential, particularly in assessing the emerging disciplines (Ramachandran, 2012). Mapping scientific fields is quite a common operation in bibliometric studies. Mapping is a study of correlation links between the past and present research work using citation analysis. Mapping is a technique used to analyze vast literature and molded in a particular form. This particular information format can be used by researcher effectively. Different elements of a bibliographic record may used to generate a map structure. Each element reveals a specific structure, unique in a sense, but always related to the structures based on other element (Mithal, Ahmad & Singh, 2005). This study, aims to capture the overall publications of cloud computing among top 10 countries and India. It aims to study publication progress by means of Relative growth rate (RGR) and Doubling time (Dt), authorship collaboration is measured using scientometric tools such as collaborative index (CI), modified collaborative coefficient (MCC). In order to find out the quality of the research output, number of high quality paper's (NHQ) is applied.

II. LITERATURE REVIEW

Authors have defined cloud computing on the basis of its use, capability and latest trends evolved in the field. According to (Plummer, et al., 2009; Buyya et al., 2009) it is a style of computing where massively scalable IT-related capabilities are provided as a service across the Internet to multiple external customers. Staten, 2008 found it as a pool of abstracted, highly scalable, and managed infrastructure capable of hosting end-customer applications and billed by consumption. Armbrust, et al., 2009 regard it as the ability to pay for use of computing resources on a short-term basis as needed. While as (Vouk, 2008) visioned it as a technology, embraces cyber-infrastructure, and builds on virtualisation, Cloudcomputing, grid computing, utility computing, networking, and Web and software services. Various studies focus on the evaluation and optimisation of the performance of the clouds. It includes studies that attempt to quantify and compare performance across different clouds (Iosup et al., 2011), to enhance workflow scheduling and load balancing (Byun, Kee, Kim, and Maeng, 2011; Kong, Lin, Jiang, Yan, and Chu, 2011), to improve dynamic resource allocation (Streitberger and Eymann, 2009; Warneke and Kao, 2011), to enable automatic bottleneck detection (Iqbal, Dailey, Carrera, and Janecek, 2011), to estimate performance of cloud network with nodes failure (Lin and Chang, 2011), and to improve interoperability across different clouds. (Beloglazov, Abawajy, and Buyya, 2011; Berl, Gelenbe, di Girolamo, Giuliani, de Meer, Dang, et al., 2010; Dougherty, White, and Schmidt, 2011; Katz, 2009) concentrated on energy efficiency, power conservation, and environmental considerations in the design of data centres. Cloud security has been a common concern for the public (Bellovin, 2011). Some articles in this subcategory look at general security mechanisms such as restrictions and audits (Spring, 2011a; Wang, Wang, Ren, Lou, and Li, 2011), multi-tenancy authorisation (Calero, Edwards, Kirschnick, Wilcock, and Wray, 2010), third-party assurance (Zissis and Lekkas, 2010), and cloud-based security services (Li, Li, Wo, Hu,Huai, Liu, et al., 2011). Other articles addressing specific cloud related security issues fall into two categories: data security and network security. The data security category includes papers looking at data encryption (Anthes, 2010), data colouring, and software watermarking for multi-way authentications (Hwang and Li, 2010), and a datapartitioning scheme for implicit security (Parakh and Kak, 2009).

Bibliometric and Scientometric Analysis of Cloud Computing There is a lack of a comprehensive scientometric study in the area of cloud computing. However following study encompass the field in terms of bibliometric and scientometric studies. Sriram & Hosseini (2010) present a review of the work published by the academic community in the field of cloud computing using ACM Digital Library, IEEE Xplore, SpringerLink, ScienceDirect and Google Scholar. This study provides an overview of the swiftly developing advances in the technical foundations of cloud computing and their research efforts. Ahmed, Chowdhury, Ahmed, & Rafee (2012) presented an overview of cloud computing and focused on the state-of-the-art research and future issues to be handled by the research community. Mirzaei (2008) presents a brief survey based of readings on "cloud" computing and tries to address related research topics, challenges ahead and possible applications. Haag & Eckhardt (2014) applied scientometric research approach and undertakes a categorized literature analysis to provide a comprehensive and systemic overview of the current status of research on cloud services and their adoption by organizations. They review 52 journals and proceedings of the information systems field to identify systematically categorize 36 articles on the topic. The content-based analysis shows that the scarce theoretical and empirical work on organizational cloud service adoption has developed and explored factors that directly or indirectly drive organizations to adopt or inhibit them from adopting cloud services from different perspectives and dimensions. Study by (Thirumagal, A. , Sethukumari, Niruba S ,2013) analyse the quantum literature output in the area of cloud computing from 2008-2012 from the Web of Knowledge. The downloaded 2207 data were analysed with the Bibexcel Tool to identify and analyse the rate and growth of scholarly publication, analyse the authorship pattern and to examine the publication type of research, application of

Lotka's Law, Creating Label View, Cluster View and finding the Citation map. Similar type of study carried out by Sivakumaren, K.S., Swaminathan, S., Karthikeyan G. 2012 found that 510 records related to Cloud Computing in "Web of Science' appeared during the periods 2001-2010. It is found that the author "Buyya.R" and the country "USA" have produced the majority of records. An extensive insights into publication patterns, research impact and research productivity was undertaken by (Heilig, L, 2014.) Furthermore, the work studied the interplay of related subtopics by analyzing keyword clusters thus provides a better understanding of patterns, trends and other important factors as a basis for directing research activities, sharing knowledge and collaborating in the area of cloud computing research. *Scientometric Analysis of various subject areas focused on RGR, Dt, CC and MCC* The scientometric analysis of Indian engineering research output using Web of Science (WOS) database for the period 1999 to 2013 for identifying the research output in the field of engineering literature was taken by (Hosamani & Bagalkoti, 2014) . Study provided a comparative evaluation and performance of different types of scientometric indicators, such as number of publications, number of citations, *relative growth, doubling time,* activity index and collaboration from India. Karpagam, 2011 analysed the growth pattern of nanoscience and Nano technology literature 1990–2009.The study measures the performance based on several parameters, country annual growth rate, authorship pattern, collaborative index, collaborative coefficient, modified collaborative coefficient, subject profile, etc. Ponnudurai & Thilakar (2013) analysed the research output in Crop Science Research out during 1981- 2010 included year wise distribution research growth, relative growth rate, exponential growth, Asian Countries publications' share, citation impact, share of international collaborative papers and major collaborative partner countries patterns of research communication in most productive journals. Similar type of study was carried out by (Venkatesan, Gopalakrishnan , Gnanasekaran, 2013) on climate change studying 94756 records contributed worldwide over a period of 1999-2012. Doubling time was also calculated in the study. Ramiah & Kaliyaperumal (2014) used engineering index database covering a period of 2003-2012 to investigate growth and development of mobile technology in terms of publication output .The average number of publications published per year was 14456.7 and the highest number of publications 20318 were published in 2011. Authors from China have contributed maximum .The study found that the relative growth rates (RGR) has decreased from 2004 (0.98) to 2012 (0.13) in the span of 10 years. The doubling time (DT) has gradually increased from 0.71 in 2004 to 5.15 in 2012. Chitra (2012) analyses the growth pattern of Neuroscience literature during 1972 – 2011 (40 years) using Scopus retrieving 35869 records. The growth in the publication is studied through Relative Growth Rate and doubling time. The authorship pattern is measured by different collaboration parameters like collaborative index, degree of collaboration, collaborative coefficient and modified collaborative coefficient. The quality of the journal is assessed by SJR and SNIP. A total of 1291 Indian contributions covered in SCOPUS database were analysed the academic productivity of food and nutrition scientists in India during the period of 1960-2011. The research output is highly scattered as indicated by the average number of papers per institution and per states in India. The food and nutrition output is dominated by the two authored papers. Further, the study investigated Relative Growth Rate and Doubling Time authorship pattern, co-authorship pattern, highly prolific authors, highly published institutions and highly preferred journals by the scientists of India. Vellaichamy, Jeyshankar & Rao (2014).

III. PROBLEM

Scientometrics is one of the most important measures for the assessment of scientific productions. It is the quantitative study of the disciplines of science based on published literature and communication. This includes identifying emerging areas of scientific research, examining the development of research over time, or geographic and organizational distributions of research. The present study makes an endeavor to gauge and analyze *the country productivity, quality of the output, citation count, authorship collaboration and literature growth* on cloud computing by applying various scientometric parameters to the published literature.

IV. OBJECTIVES

A. To calculate total number of publications (TNP) and country ranking.
B. To calculate number of high quality paper's (NHQ).
C. To identify the citation impact of top ten countries and of India contributing to the field.
D. To find out Relative Growth Rate (RGR) and Doubling Time (Dt).
E. To find out authorship collaboration in terms of CI and MCC.

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

V. METHODOLOGY

The study was carried out using Web of Science, an international database searched from 20th march to 10th- April-2014 for all records of papers of top ten contributing countries in the field and India. In order to fulfill objective 1, the related database was visited and search term "cloud computing" within quotes was entered in the search box limiting the time span from 2009-2013. The result from all the countries were allowed to display in order to find out top productive countries in the field. The top ten countries were identified on the basis of highestnumber of publications retrieved from the database. Lastly India was included in the listing for comparison.

In order to fulfill objective 2, *"create citation report"* an option provided by the database was used for checking the "average citations". The results were sorted by *"Times cited highest to lowest"*. On the basis of earlier displayed results papers having citations more than double of average citations is calculated.

NHQ is based on the calculation of the citation per paper for different countries as the
pattern of citation varied from one country to another country. Papers that received more than twice the average citations have been considered as high quality papers.

$$NHQ\% = \left(\frac{number\ of\ high\ quality\ papers\ for\ a\ country}{total\ number\ of\ high\ quality\ papers}\right) \times 100$$

In order to fulfill objective 3, results were sorted by *"Times cited lowest to highest* " and then results provided were analyzed. The growth of publications was analysed by using two parameters Relative Growth Rate and Doubling time (Mahapatra 1985). RGR is a measure to study the increase in number of articles of time , for calculating RGR, cumulative output Log2 or lnN2 of one year is subtract from Log1 or lnN1 of the same year which is Log2 of previous year and is calculated as

$$RGR = (\ln N2 - \ln N1) / (t2-t1)$$

Where N2 and N1 are the cumulative number of publications in the years t2 and t1.Dt is calculated using following formula

$$Dt = \ln 2 / RGR$$

In order to fulfill objective 5 which is measuring author collaboration, it is measured using the formulas such as collaborative Index (CI) & Modified Collaborative Coefficient

$$CI = \frac{\sum_{j=1}^{A} jfj}{N} \qquad MCC = \frac{A}{A-1}\left\{1 - \frac{A=\sum_{j=1}^{A}(1/j)fj}{N}\right\}$$

Where;

A= total number of authors , N= total number of papers ,Fj= number of papers having j authors
J= number of authors in a paper.

VI. RESULTS

COUNTRY	2009	2010	2011	2012	2013	TNP
USA	24 (3.83%)	82 (13.09%)	143 (22.84%)	153 (24.44%)	224 (35.78%)	626
CHINA	2 (0.54%)	18 (4.9%)	48 (13.07%)	92 (25.06%)	207 56.40%)	367
AUSTRALIA	4 (2.79%)	8 (5.59%)	24 (16.78%)	48 (33.56%)	59 (41.25%)	143
ENGLAND	7 (5.22%)	14 (10.44%)	21 (15.67%)	32 (23.88%)	60 (44.77%)	134
TAIWAN	_	5 (3.93%)	21 (16.53%)	40 (31.49%)	61 (48.03%)	127
S.KOREA	1 (0.86%)	7 (6.03%)	26 (22.41%)	31 (26.72%)	51 (43.96%)	116
SPAIN	4 (4.3%)	3 (3.22%)	13 (13.97%)	31 (33.33%)	42 (45.16%)	93
GERMANY	7 (7.86%)	9 (10.11%)	21 (23.59%)	20 (22.47%)	32 (35.96%)	89

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

CANADA	2 (2.35%)	5 (5.88%)	12 (14.11%)	21 (24.7%)	45 (52.94%)	85
JAPAN	4 (5.19%)	11 (14.28%)	39 (50.64%)	10 (12.98%)	13 (16.88%)	77
INDIA	_	_	6 (27.27%)	7 (31.81%)	9 (40.9%)	22
Total	55 (2.92%)	162 (8.62%)	374 (19.9%)	485 (25.81%)	803 (42.73%)	**1879**

Table 1: Ranking & yearly publications of countries

Table 1 reveals the amount of publication on Cloud Computing during 5 years . USA has thehighest number of publications 626 (33.31%) followed by China 327 (17.4%). India have the least number of publications 22 (1.17%) followed by Japan 77 (4.09). There is a tremendous growth in the publications of Cloud Computing from 55 (2.92%) publications in 2009 it escalates to 803 (42.73%) in 2013. India did not have any publication in 2009 and 2010. During the period of study a total of 1879 publications were published.

COUNTRY	TNP	TNC	AC	NHQ	NHQ%
USA	626	3511	5.61	68	30.50
CHINA	367	790	2.15	49	21.97
AUSTRALIA	143	1117	7.81	16	7.17
ENGLAND	134	609	4.54	16	7.17
TAIWAN	127	268	2.11	17	7.62
S.KOREA	116	219	1.89	16	7.17
SPAIN	93	689	7.41	10	4.48
GERMANY	89	600	6.74	14	6.28
CANADA	85	351	4.13	8	3.60
JAPAN	77	123	1.60	7	3.14
INDIA	22	150	6.82	2	0.90
TOTAL	**1879**	**8427**	_	**223**	_

Table 2: Total publications, citations, average citation & high quality papers

TNP: Total number of publications **TNC:** Total number of citations **AC:** Average Citations **NHQ:** Number of high quality papers.

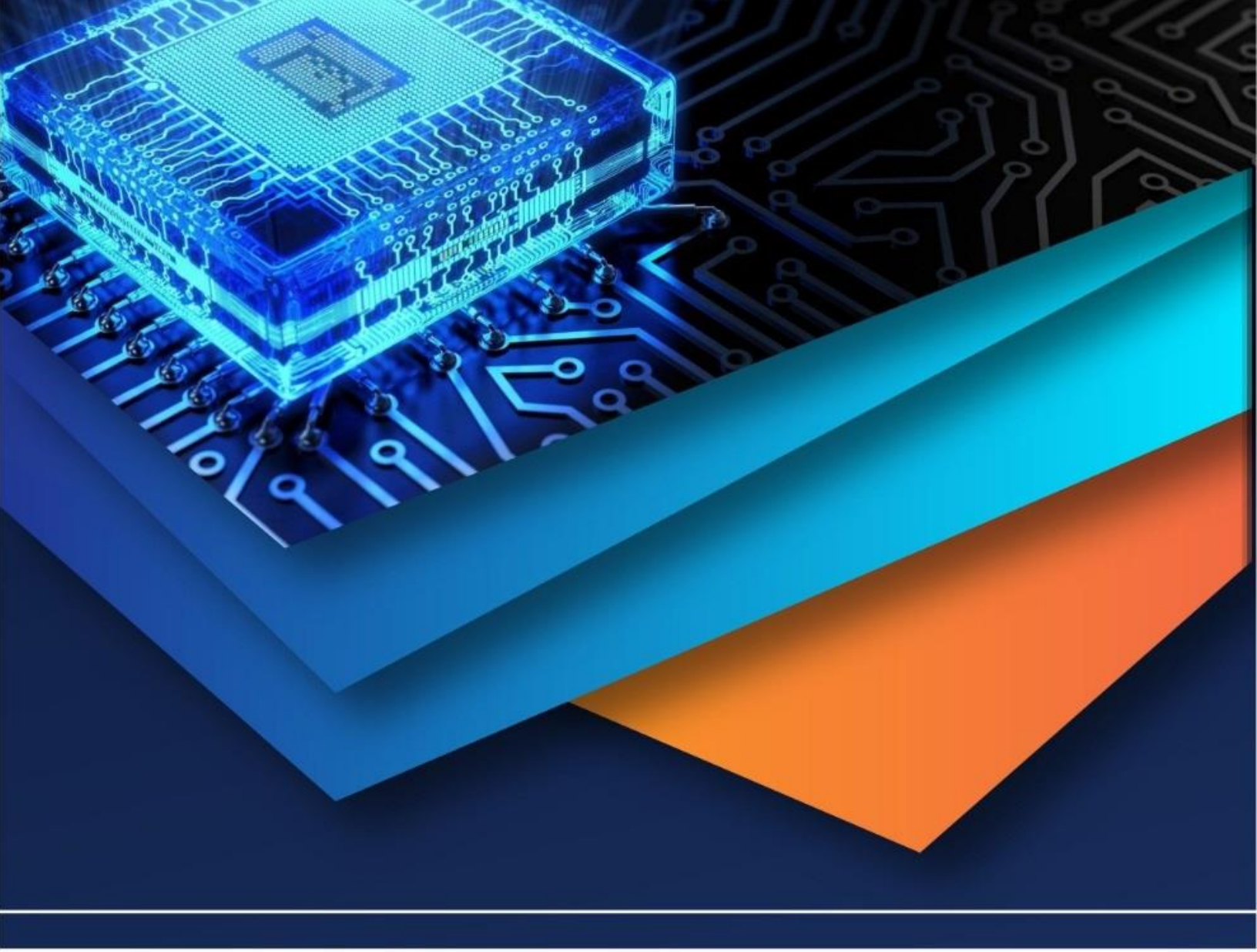
Table 2 reveals the total number of citations and high quality papers and shows that USA leads the list with 3511 citations followed by Australia with 1117 citations. Least number of citations is of Japan with 123 citations followed by India with 150 citations. In case of high quality papers, out of a total of 223 papers, USA is on the top with 68 papers followed by China with 49 papers. Least number of high quality papers is of India with 2 papers followed by Japan with 7 papers. Out of the total of 1879 papers only 223 (11.86%) are high quality papers.

VII. CONCLUSION

Considering the above certainties it is reasoned that the exploration yield in the field of distributed computing was higher i.e. 1879 amid the square year 2009-2013. Productions continues expanding each year. It increments from 55 distributions in 2009 to 803 productions in 2013 seeing a huge growth.USA is on the top having 626 distributions in five years and is trailed by China with 367 productions. India has the minimum productions just 22. RGR and Dt are contrarily corresponding i.e. rate of development of production was diminished and the relating Dt was expanded. The discoveries of aggregate references uncovers that aggregate

Volume 4 Issue IV, April 2016
ISSN: 2321-9653

number of references and astounding papers are 8427 and 223 individually. If there should be an occurrence of aggregate references and brilliant papers USA is up on the rundown with 3511 references and 68 great papers. Altogether references USA is succeeded by Australia with 1117 references and in the event of fantastic papers is succeeded by china with 49 top notch distributions. Out of the aggregate productions rate of brilliant papers is 11.86% just which demonstrates that lion's share of distributions are not refered to or refered to just a couple times. It is more clear from the outline of the appropriation of references where it is found that the papers with "0" reference are most noteworthy with 841 productions and is trailed by the distributions with "1" reference which are 321 altogether. To assess the creator cooperation Shared List (CI), and Changed Synergistic Coefficient (MCC) were utilized and demonstrated that 90% of the exploration yields were of collective in nature. In India the exploration in this field is childish stage. This might be because of non-accessibility of assets and steady preparing programs. Fortifying of preparing projects at institutional level, national and global level gets to be compulsory. The lacking on the commitment might be because of non accessibility of global coordinated effort.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY