

# **Mapping of Publications Productivity on Breast Cancer: A Study Based on Clarivate Analytics – Web of Science Database**

**Ambhore Sagar P.<sup>1</sup>; Malve Priyanka R.<sup>2</sup>**

Research Associate, Dept .of Library & Information Science, Dr. Babasaheb Ambedkar Marathwada, University, Aurangabad. (MS) India<sup>1</sup>; Librarian, Sanghvi Kesahri College of Arts and Commerce, Chinchwad, Pune. (MS) India<sup>2</sup>  
ambhore.sagar02@gmail.com, malvepriyanka635@gmail.com

## **ABSTRACT**

*The Present Study analyzed research productivity in Breast Cancer for a period of 05 years between 2017 and 2021. Web of Science database a service from Clarivate Analytics has been used to download citation and source data. Bibexcel and Histcite application software have been used to present the datasets. Analysis part focuses on the parameters like citation impact at local and global level, influential authors and their total output, ranking of contributing institutions and countries. In addition to this scientographical mapping of data is presented through graphs using VOSviewer software mapping technique.*

**KEYWORDS:** Citations, Bibexcel, Most Productive Authors, Scientometric Analysis, Histcite, VOSviewer, Breast Cancer.

## **INTRODUCTION**

The Present paper analyzed the scientometric examination of research papers contributed in Breast cancer is a malignant tumor (a collection of cancer cells) arising from the cells of the breast. Although breast cancer predominantly occurs in women, it can also affect men. This article deals with breast cancer in women. Breast cancer and its complications can affect nearly every part of the body.

The breast is made up of different tissue, ranging from very fatty tissue to very dense tissue. Within this tissue is a network of lobes. Each lobe is made up of tiny, tube-like structures called lobules that contain milk glands. Tiny ducts connect the glands, lobules, and lobes, carrying milk from the lobes to the nipple. The nipple is located in the middle of the areola, which is the darker area that surrounds the nipple. Blood and lymph vessels also run throughout the breast. Blood nourishes the cells. The lymph system drains bodily waste products. The lymph vessels connect to lymph nodes, the small, bean-shaped organs that help fight infection. Groups of lymph nodes are located in different areas throughout the body, such as in the neck, groin, and abdomen. Regional lymph nodes of

the breast are those nearby the breast, such as the lymph nodes under the arm.

Cancer begins when healthy cells in the breast change and grow out of control, forming a mass or sheet of cells called a tumor. A tumor can be cancerous or benign. A cancerous tumor is malignant, meaning it can grow and spread to other parts of the body. A benign tumor means the tumor can grow but will not spread.

Breast cancer spreads when the cancer grows into adjacent organs or other parts of the body or when breast cancer cells move to other parts of the body through the blood vessels and/or lymph vessels. This is called a metastasis.

## **REVIEW OF LITERATURE**

**Khparde V.S and etal (2014)** Mapping of Library and Information Science Research Based on LISA (2006-2010) observed their article the top 30 most productive Indians and foreign journals contributing to Indian library and information science output together contributed 95.34 % share in the cumulative publication output of Indian in library and information research during 2006-2010 the cumulative share of these 30 journals showed decrease from 99.04 % in 2006-2008 to 86.95 % in 2009-2010. **Muneer Ahmad and Dr. M Sadik Batcha (2019)** Study analyzed research productivity in Journal of Documentation (JDoc) for a period of 30 years between 1989 and 2018. **Ranveer V B and Khparde V.S (2016)** Citations analysis of Ph. D theses on Mathematics submitted to Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (Maharashtra) India 6410 citations had taken to observe the frequency distribution of different forms of literature used by researcher in Mathematics heavily depend on periodicals sources for their studies. Only top ten form of literature is shown in study it was found that the journal are the major form of media used with citation count, of 4457 (69.53%) of total literature used, where as remaining sources account for 1953 (30.46%) therefore periodicals are the first hand original and latest information in the subject and are most preferred channel of information use. Amongst citations of non-periodical sources book occupy 1299 (20.27%) of citations, while the bulletin occupy 158 (2.46%) remaining 126 (1.97%) documents are other documents and 101 documents is not mentioned their form (**S. Batcha, 2018**) discussed thoroughly about scientometric output of cardiovascular disease of SAARC countries and offers a powerful set of methods and measures for studying the structure and process of research communication. The paper examines the research trend, authorship, collaborative pattern and activity index of five SAARC Countries regarding the disease which amounts to about 24.8% of deaths in SAARC countries. The result of the paper reveals that India is a leader country among SAARC nations having major research output followed by Pakistan in cardiovascular disease research. The paper also deliberated that USA, England and Australia are the top collaboration countries which has done collaboration with SAARC nations. (**Khan, 2001**) attained the scientometric analysis of the DESIDOC Journal of Library and Information Technology from 2010 to 2014. The data from website of journal analysed covers mainly the number of articles, authorship pattern, geographical distribution and types of documents cited. The research reveals that out of 307 contributions, 119 (38.76%) are contributed by single authors while the remaining 188 (61.24%) have joint authors. Furthermore the research finds that maximum contributions are from India. (**M.Sadik Batcha & Muneer Ahmad, 2017**) analysed comparative analysis of Indian Journal of Information Sources and Services (IJISS) and Pakistan Journal of Library and Information Science (PJLIS) during 2011-2017 and studied various aspects like year wise distribution of papers, authorship pattern & author productivity, degree of collaboration pattern of Co-Authorship , average length of papers , average keywords, etc and found 138(94.52%) of contributions from IJISS were made by Indian authors and similarly 94(77.05) of contributions from PJLIS were done by Pakistani authors. Papers by Indian and Pakistani Authors with Foreign Collaboration are minimal (1.37% of articles) and (4.10% of articles) respectively. (**M. S. Batcha, 2017**) analysed the research publication output in the field of robotic

technology and shows that the robotic technology is a progressive field increasing the publication output from single digit to 513 year after year during the period from 1990 to 2016. The results shows that developing countries like USA, UK and Germany gives the most output on robotic technology related research. Yet major proportion of contribution (36.30%) is from USA. English language is the most preferred for the research amounting (87.70%) followed by German. The Prolific authors in the field of robotic technology are highly found from USA among them the contribution by Bloss R is appreciable and author from Japan, Dario P competes with more number of publications in the study.

## OBJECTIVES

The main objective of the study is to consider on the mapping of 6998 articles published by the Breast Cancer during the period of 2017-2021 and the specific objectives are to identify and carry out the following factors

- To examine the annual publications output of Breast Cancer.
- To gauge publication density through mapping of top 30 authors, countries and institutions based on their number of research papers.
- Find out the top 30prolific authors, institutions and countries.

## DATA SOURCE AND METHODOLOGY

The data for the present study were downloaded from the Clarivate analytics-Web of Science database in July 2021. A total of 6998 research publications was downloaded from 2017-2021. The data downloaded were enhanced with different parameters like title, authors, years, countries, and research institutions. Furthermore, the downloaded data were analyzed by using Bibexcel, Histcite, and Vosviewer software applications.

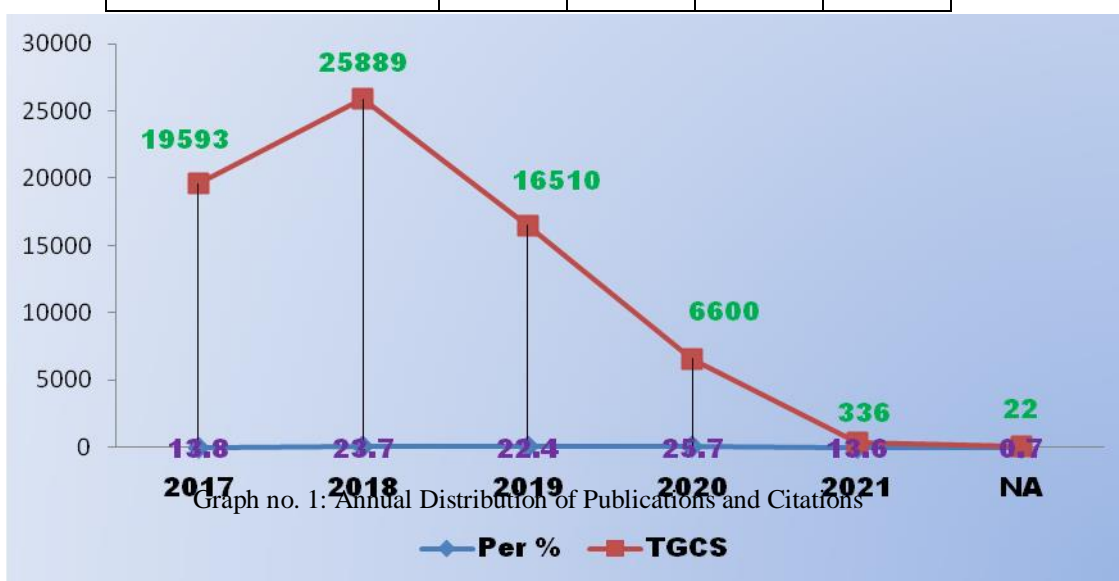
**Table No. 1. Details of the Important Points of the Data Sample During 2017 to 2021.**

S.No.	Details about Sample	Observed Values
1	Duration	2017-2021
2	Collection Span	05 years
3	Total No. of Records	6998
4	Total No. of Authors	33865
5	Frequently Used Words	7493
6	Document Types	17
7	Languages	9
8	Contributing Countries	122
9	Contributing Institutions	7256
10	Institutions with Sub Division	18741
11	Total Local Citation Scores	0
12	Total Global Citation Scores	68950

**Discussion and Result**

**Table No. 2. Annual Distribution of Publications and Citations:**

Sr. No.	Publication Year	Recs	Per %	TLCS	TGCS
1	2017	968	13.8	0	19593
2	2018	1658	23.7	0	25889
3	2019	1571	22.4	0	16510
4	2020	1799	25.7	0	6600
5	2021	951	13.6	0	336
6	NA	51	0.7	0	22
<b>Total</b>		<b>6998</b>	<b>100</b>		<b>68950</b>



**Graph no. 1: Annual Distribution of Publications and Citations**

The table no. 2. Reveals that the numbers of research documents published from 2017 to 2021 are gradually increased. According to the publication output from the table no. 2. The year wise distribution of research documents, 2020 has the highest number of research documents 1799 (25.7%) with 6600 of total global citation score values and being prominent among the 05 years output and it stood in first position. The year 2018 has 1658 research documents and it stood in second position with 0 of total local citation score and 25889 of total global citation score were scaled. It is noticed that the increase in publications may not create impact on citation score yet the quality matters on total local citation scores and on total global citation scores. Graph number one mull over the year wise publications and depicts the citation score. It clearly indicates on the fact that the increased publication rate is not bringing the increased citation rate.

**Table No. 3. Publication output of Top 30 Authors and Citation Score.**

Sr. No.	Author	Recs	Percent	TLCS	TGCS	TGCS/t
1	Fasching PA	49	0.7	0	2034	661.28
2	Schneeweiss A	32	0.5	0	2668	730.72
3	Easton DF	25	0.4	0	1605	383.65
4	Milne RL	23	0.3	0	1631	384.3

5	Loibl S	20	0.3	0	2904	861.13
6	Harbeck N	20	0.3	0	1862	540.2
7	Hopper JL	20	0.3	0	1593	357.23
8	Rugo HS	19	0.3	0	3516	973.62
9	Iwata H	19	0.3	0	2789	788.08
10	Curigliano G	19	0.3	0	1508	446.9
11	Terry MB	19	0.3	0	1484	334.45
12	Cortes J	19	0.3	0	1437	495.42
13	Untch M	17	0.2	0	1516	485.73
14	Andre F	15	0.2	0	2227	677.33
15	Winer EP	14	0.2	0	2927	787.98
16	Loi S	13	0.2	0	1847	515.93
17	Goldgar DE	11	0.2	0	1838	405.7
18	Im SA	10	0.1	0	3111	840.73
19	Antoniou AC	10	0.1	0	1423	335.07
20	Mayer IA	8	0.1	0	1697	511.25
21	Jemal A	8	0.1	0	1434	366.37
22	Conte P	7	0.1	0	1526	377.53
23	Barrios CH	6	0.1	0	1808	471
24	Newman LA	6	0.1	0	1443	366.48
25	Baselga J	6	0.1	0	1410	294.1
26	Dieras V	4	0.1	0	1752	438.03
27	Schmid P	4	0.1	0	1649	471
28	Senkus E	3	0	0	1497	342.15
29	DeSantis CE	3	0	0	1419	360.87
30	Sauer AG	2	0	0	1417	360.2

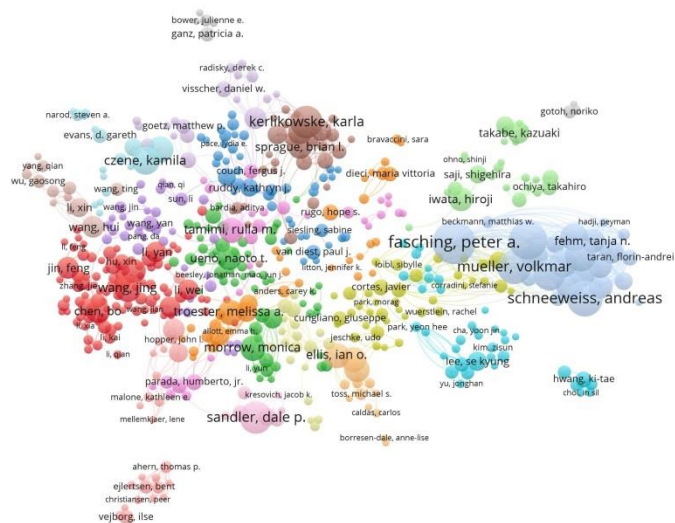


Figure 1: Showing Highly Prolific Authors

**Analysis of the Publication Output of Top 30 Authors**

Table No. 3. and figure 1 displays the ranking of authors of research articles. In the rank analysis the authors who have published more than 02 articles or more are considered into account to avoid a long list. It was observed that there is total of 33865 authors for 6998 records and it shows the top 30 most productive authors during 2017-2021. Fasching, P. A. published 49 (0.7%) articles with 2034 TGCS articles, followed by Schneeweiss, A. 32 (0.5%) with 2668 TGCS articles. The data set clearly depicts that no matter how many publications that an author brings out yet the quality publications alone shows impact in the form of total local citations score and total global citations score.

**Table No. 4. Ranking of Institutions and their Research Performance**

Sr. No.	Institution	Recs	Percent	TLCS	TGCS
1	Harvard Med Sch	160	2.3	0	4623
2	Univ Texas MD Anderson Canc Ctr	159	2.3	0	4711
3	Unknown	159	2.3	0	128
4	Mem Sloan Kettering Canc Ctr	150	2.1	0	5871
5	NCI	117	1.7	0	2659
6	Univ Calif San Francisco	112	1.6	0	3939
7	Dana Farber Canc Inst	110	1.6	0	5303
8	Mayo Clin	110	1.6	0	3805
9	Fudan Univ	106	1.5	0	1933
10	Univ Toronto	97	1.4	0	2479
11	Stanford Univ	96	1.4	0	2722
12	Univ N Carolina	93	1.3	0	2216
13	China Med Univ	92	1.3	0	1314
14	Karolinska Inst	86	1.2	0	2393
15	Sun Yat Sen Univ	81	1.2	0	823
16	Univ Washington	80	1.1	0	1877
17	Univ Cambridge	79	1.1	0	2732
18	Univ Penn	77	1.1	0	2381
19	Univ Michigan	76	1.1	0	2747
20	Brigham & Womens Hosp	75	1.1	0	1263
21	Nanjing Med Univ	74	1.1	0	861
22	Univ Calif Los Angeles	74	1.1	0	3125
23	Seoul Natl Univ	71	1	0	2847
24	Univ Milan	71	1	0	2418
25	Univ Southern Calif	68	1	0	1440
26	Duke Univ	67	1	0	1382
27	Univ Calif San Diego	62	0.9	0	1460
28	Univ Pittsburgh	62	0.9	0	3203
29	Emory Univ	61	0.9	0	1317
30	Fred Hutchinson Canc Res Ctr	60	0.9	0	1355

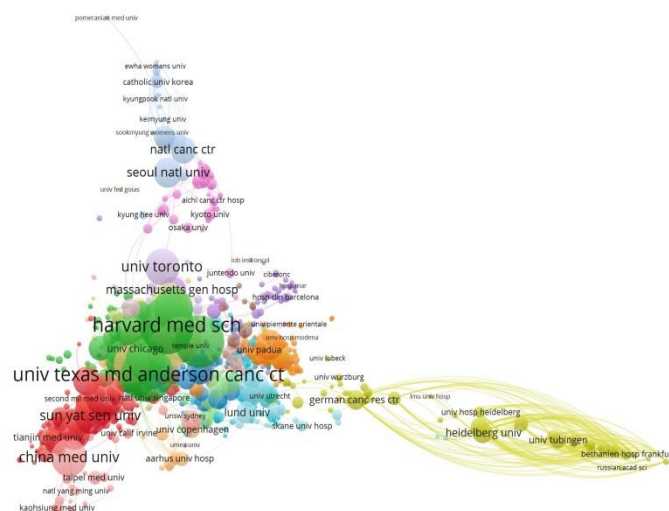


Figure 2: Collaboration of Institutions and their clusters

### Analysis of the Publication Output of Top 30 Institutions

The individualities of 30 most productive institutions were analyzed in this part, Institutions which published more than 31 and above publications have considered as highly productive institutions. Table IV summarizes articles, the global citation score, local citation score and average author per paper of the publications of these institutions. In total, 7256 institutions, including 18741 subdivisions published 6998 research papers during 2017-2021. The topmost thirty prolific institutions involved in this research have published 31 and more research articles. The institution “Harvard Med Sch” holds the first rank and the institution published 160 (2.3%) research papers with 0 local and 4623 global citation scores. The second rank holds by “Univ Texas MD Anderson Can. Ctr” the institution published 159 (2.3%) research papers with 0 local and 4711 global citation scores respectively.

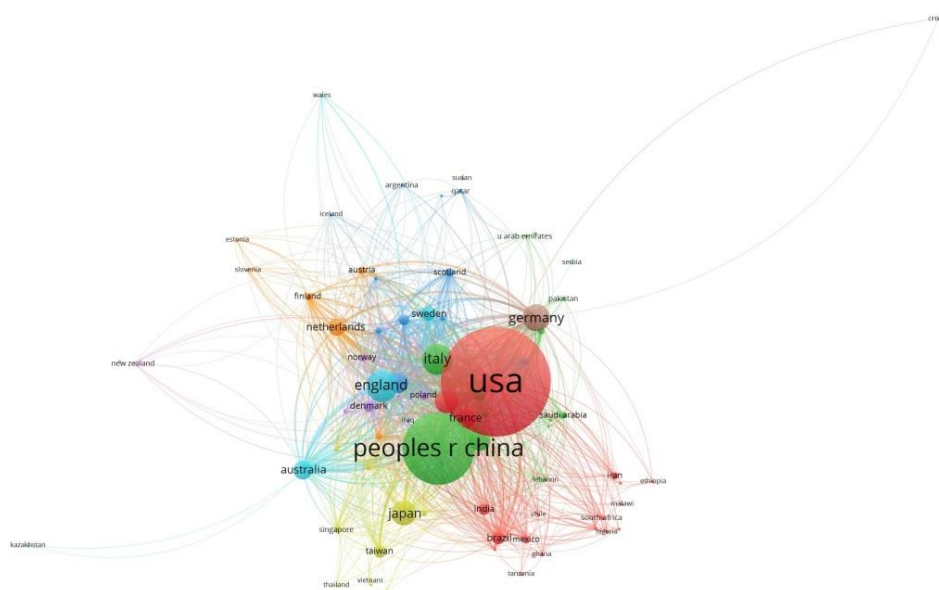
Table No. 5. Distribution of the Publication Output of Top 30 Countries

Sr. No.	Country	Recs	Percent	TLCS	TGCS
1	USA	2464	35.2	0	39155
2	China	1387	19.8	0	15059
3	UK	527	7.5	0	12751
4	Italy	448	6.4	0	8832
5	Germany	384	5.5	0	9875
6	Japan	335	4.8	0	6238
7	Canada	307	4.4	0	7523
8	South Korea	287	4.1	0	6865
9	Spain	258	3.7	0	7309
10	France	252	3.6	0	10145
11	Australia	241	3.4	0	6968
12	Netherlands	211	3	0	5154
13	Sweden	170	2.4	0	3766
14	Unknown	160	2.3	0	128
15	Brazil	137	2	0	4262



**Mapping of Publications Productivity on Breast Cancer: A Study Based on Clarivate Analytics – Web of Science Database**

16	Taiwan	127	1.8	0	3540
17	Switzerland	124	1.8	0	6253
18	India	122	1.7	0	1638
19	Belgium	115	1.6	0	3893
20	Denmark	109	1.6	0	2847
21	Poland	89	1.3	0	3182
22	Mexico	77	1.1	0	1843
23	Norway	76	1.1	0	1690
24	Finland	72	1	0	1384
25	Iran	72	1	0	490
26	Portugal	72	1	0	2290
27	Austria	68	1	0	1801
28	Saudi Arabia	65	0.9	0	552
29	Turkey	65	0.9	0	651
30	Israel	61	0.9	0	2825



**Figure 3: Showing Ranking of Country wise Distribution**

**Analysis of the Publication Output of Top 30 Countries**

Table No. 5. and figure 3 displays the publication output of the top twenty five countries by number of papers and USA acquired 1st rank among the top thirty countries under consideration with its total global citation score 39155 .In all 122 countries participated in research during 2017 and 2021. The countries that rank between 2<sup>nd</sup> and 30<sup>th</sup> position are China, UK, Italy, Germany, Japan, Canada, South Korea Spain France Australia Netherlands Sweden Brazil Taiwan Switzerland India Belgium Denmark Poland Mexico Norway Finland Iran Portugal Austria Saudi Arabia Turkey Israel. We have found by using this country mapping analysis that there are nodes with clarity of linking between each node, which indicates that there are countries linking and associated with other associated countries. It could be identified that the country wise analysis the following countries USA, China, UK, Italy, Germany were identified the most productive country based on the number of research papers published.



## **CONCLUSION:**

The number of papers published in Communication and mass media has gradually increased during 2017-2021 and the study has shown that 6998 research documents have been published in Breast Cancer during the period. It could be identified that the author's wise analysis were acknowledged the most prolific authors based on the number of research papers contributed. It could be identified that the institutions wise analysis were acknowledged the most prolific institutions based on the number of research papers output they published. It could be identified that the country wise analysis the following countries USA, China, UK, Italy, Germany, were identified the most productive country based on the number of research papers published.

## **REFERENCES**

- [1] Ahmad, M., Batcha, M. S., Wani, B. A., Khan, M. I., & Jahina, S. R. (2017). Research Output of Webology Journal (2013-2017 ): A Scientometric Analysis. *International Journal of Movement Education and Social Science*, 7(3), 46–58.
- [2] Batcha, M. S. (2017). Research Output Analysis on Robotic Technology: A Scientometric Study. *Indian Journal of Information Sources and Services*, 7(1), 25–31.
- [3] Batcha, M. S., Jahina, S. R., & Ahmad, M. (2018). Publication Trend in DESIDOC Journal of Library and Information Technology during 2013-2017: A Scientometric Approach. *International Journal of Research in Engineering, IT and Social Sciences*, 08(04), 76–82.
- [4] Khaparde Vaishali, V.B. Ranveer & Ambhore Sagar P(2014). Mapping Of Library And Information Science Research Based ON LISA (2006-2010), *Journal of Library, Information and Communication Technology (JLICT)* Vol. 6 Issue (1-2) June, 2014, pp.1-10.
- [5] Muneer Ahmad and Dr. M Sadik Batcha (2019) Mapping of Publications Productivity on Journal of Documentation 1989-2018: A Study Based on Clarivate Analytics – Web of Science Database *Library Philosophy and Practice*
- [6] Ranveer V. B (2014), Collection Building: A Scientometric Study, “*Knowledge Librarian*”- *An International Peer Reviewed Bilingual E-Journal of Library and Information Science* Volume: 01, Issue: 02, Nov. – Dec. 2014 eISSN NO. 2394- 2479, p.1-12.
- [7] Ranveer V B and Khaparde V.S (2016) citations analysis of Ph. D theses on Mathematics submitted to Dr.Babasaheb Ambedkar Marathwada University, Aurangabad (Maharashtra) India, *International Journal of Information Sources and Services*, vol. 3: July - August, 2016: Issue -4 ISSN 2349-428X (Print) pp.86-98.
- [8] Ranveer V B and Khaparde V.S (2016) citations analysis of Ph. D theses on Statistics submitted to Dr.Babasaheb Ambedkar Marathwada University, Aurangabad (Maharashtra) India, *International Journal of Digital Library Services*, Vol. 6, July – Sept. 2016, Issue - 3, ISSN:2250-1142 (Online), ISSN 2349-302X (Print) pp.105-117.