Journal of Intelligent Pervasive and Soft Computing

journal homepage: https://scirep.institute/journals/index.php/jipsc/

Worldwide Knowledge Dissemination in Chemistry

Faiza Qayyum[1], and Tanvir Afzal [1,*]

[1] Capital University of Science and Technology Islamabad, 44000, Pakistan.

[*]Correspondence:mafzal@jinnah.edu.pk

Abstract

The identification of knowledge-producing countries plays a pivotal role in the indication of such countries whose work has contributed towards subsequent research, and the same can be deemed as a hallmark of quality. Citation analysis is the most quintessential indicator to analyze the influence of scientific publications and research documents as well as in appraising or ranking the scientific brilliance of a particular author, institution or country. This dissertation identifies trends in chemistry publications production globally. The knowledge producer, consumer and emerging countries are identified on the basis of journals appearing in Microsoft Academic Search (MAS). The data set is ascertained from MAS on the basis of citations. A total of 1800,000 top citations from 18 different categories of chemistry are examined. The origins are classified and envisioned on Google Maps depending upon the first author's organizational associations. The result of research strongly establishes that the U.S.A, Germany, UK, France, and China collectively constitute a category of leading data producers. However, Asian and African countries fall in the category of knowledge consumers.

Keywords: *Citation analysis, knowledge dissemination, producers, consumers, chemistry;*

1. Introduction

In the contemporary world (due to the growing concentration of publications), it is essential for the research community to be well aware of global progress and the scientific status of countries across the globe to acknowledge those countries which produce a large number of publications, as well as to explore the reasons for less contribution in the publication by any country. The leading origin (country) of any domain (i.e. chemistry) can be traced by tracking the global trends since Periodicals are the main indicators of the new evolving ideas in any discipline [1]. Citation analysis is usually utilized to appraise the global influence of any particular author, institution and scientific document. It is utilized as a tool to trace and discover the developments of data dissemination. Configurations discovered by utilizing the citation analysis assist the researchers in developing scientific policies, tracing academic leadership, and academic rankings of authors, institutions, cities or journals. Periodicals count and citation rates are harnessed by academic rankings to measure data production and consumption globally[2,3].

Citation Analysis will greatly benefit the researchers as it will assist them in discovering an appropriate research environment for their area of research and, at the same time, get more prospects. Determination of data movement can only be done by the production of data meaning thereby when a periodical is published and consumed by others who then quote the idea so proposed in the periodical as only the number of

periodicals and the citation count do not provide any clue about the fact that who is really quoting a specific scientific document. An inventive examination and visualization are conducted in this research which will produce a wide-ranging study regarding data-producing and consuming countries. The network-based methodology introduced by Marc. B et al. [3] are the pioneer methodologies that disclose who quotes whom and who is quoted and harness it to discover the data movement. The network-based methodology has the following restrictions:

1. The said analysis is restricted to the time period between 2007 and 2009, which was recovered from Thomson Reuters' Web of Science. Therefore, it does not qualify to be the examination of recent years.
2. Thirteen million research papers were from diversified areas and were not concentrated on a single field. In contrast, our research paper emphasizes only one sphere, i.e. Chemistry, for all-encompassing examination.
3. Web of Science data set does not entail a vast range of books or conference publications, and it contains publications only in the English language.

Our research concentrates specifically on the domain of chemistry. As Chemistry is admittedly the mother of all sciences, therefore, it is of prime importance for the research fraternity to study the movement of data in Chemistry. Incidentally, enormous depository of data based on the dataset retrieved from the MAS. The MAS gives the details of a domain regarding different classifications such as Top Authors, Publications, Conferences Journals, Keywords, and their location. Therefore, we have harnessed the MAS as a database for our research. The MAS classify the publications of the Chemistry in different categories like, analytical chemistry, biochemistry, chromatography and so on. We will examine and visualize the Chemistry produced and consumed in this paper.

The rest of the paper is organized as follows. Section 2 contains related work. In Section 3 detail discussion of the methodology of our research. Section 4 contains an evaluation and discussion about results. In Section 5 we draw conclusion based on our findings.

2. Literature Review

In this section, a brief review of worldwide knowledge dissemination is presented.

In 2005, Tsai et al. examined the global trends of publications in three specific journals, International Journal of Science Education (IJSE), Journal of Research in Science Teaching (JRST), and Science Education (SE), for the time period of five years (1998-2002). A total of 802 papers were analyzed on the basis of the author's nationality. Points were assigned to the countries according to the relevant position of the author in the research paper. Their final results revealed that most of the periodicals were from English-speaking countries, the USA, UK and Canada. However, their study was limited to only three journals[5]. In 2008, Marc B. Gurein et al. analyzed the global trends of publications in ophthalmology for the time period of five years (2002 to 2006). A total of 7754 articles from five journals of ophthalmology are extracted from the Medline search engine. The abstract of the articles is chosen to check its relevance to ophthalmology. Final results were enlisted by considering the total population and research expenditures of the country. Out of 67 countries, the US, followed by UK and Japan, produced a major part of articles in journals. On the other hand, Singapore, Ireland and Australia are considered emerging countries. However, their work was limited to a small dataset [6]. In 2007, Robert J.

W. Tijssen examines the participation of Africa-based authors in international scientific literature via citation analysis. The dataset is extracted from Thomson Scientifics for the time period of 24 years (1980 to 2004). The results were obtained on the basis of the author's position on the paper. The final results revealed that African authors are not actively contributing to research due to the lack of research funds. Their study is limited to only one country[7].In 2013, Mazloumin et al. analyzed global excess scientific production trends. The basic idea was to analyze who quotes whom and who is quoted and harness it to discover the flow of data. The duration period was of 8 years (200 to 2009). The dataset consists of 13 million journal papers, and 80 million citations between them were extracted from Thomson Reuters' Web of Science (WoS)

data. A total of 45 countries were identified who produced more than 1,000 Web of sciences papers per year. The leading knowledge producer countries were North America, Europe and Australia, and knowledge consumer countries were Africa, Asia and South Africa. While there collected, 13 million research papers were from diversified areas and were not concentrated on a single field.[8].In 2015, Bipin et al. examined the geographical trends of publications in "Political Geography" and "Religion" periodicals through bibliometric measures. Top 1500 papers were extracted from which 900 were from Political geography and 600 from Religion were selected. The duration period is from 2005 to 2013. The source of dataset extraction was ScienceDirect. The final results revealed, in both periodicals leading procures were UK (44%)) and USA (38.5%). However, their study was limited to a small dataset [9]. In 2014, Ramesh Kauri et al. examined the "Pearl" journal to analyze the authors' pattern.

Various analysis techniques are applied to identify the regions of large productions, such as authorship pattern, Domain wise distribution of citations, Age of citations, most prominent journals cited and different sources of citations etc. the duration period was of 2 years (2009 to 2012). The final results showed that most of the work is written by single authors [10]. In 2004, Chaomei Chen et al. examined the dissemination of knowledge in "tissue engineering" via patent citation. The research also focuses on the improvement and understanding of knowledge diffusion and technology transfer. Based on this study US country is identified as the largest publication producer country, followed by the UK and Canada[11]. In 2015, Fei-chang Ma et al. examined the global trends in publications of global translational medicine research. The analysis was performed on data extracted from the Pubmed database for the duration of 1992 to 2012. The 20 leading producer countries and institutes were analyzed from which USA, Germany and Russia were on the top[12]. In 2015, Grosetti et al. examined a multilevel analysis of publications. The articles were analyzed from the Web of Sciences (WoS) for the time period 1987 to 2007. The authors' addresses were geocoded in order to trace the location. The results revealed the trends both globally and within countries[13].

3. Methodology

The all-inclusive examination is carried out to trace the source and source of data in the field of Chemistry. Empirical observations of references in chemistry are made on the basis of the MAS. Chemistry data movement is concentrating particularly on data movement across the globe. In this portion, we first define the data sets of chemistry publications which were utilized by us and the formation of the geo-visualization based on references. Our system comprises of following five major steps:

1. **Dataset collection:** This step entails retrieval of the dataset from the MAS for the Chemistry domain.
2. **Determination of geographical location:** In this step, we traced the geographical location of the author by utilizing Wikipedia.
3. **Country-wise grouping of research journals:** In this step, all the research journals belonging to a particular country are collected in a cluster.
4. **Identification of producers, consumers and Emerging countries.**: In this step, all the groups are placed on Google Map using google visualization developer as per their specific countries.
5. **Publication Based on Institutional Affiliation to Respective Countries on Google Map**

Trends are determined to harness similar data groups for further examination. Figure 1 describes the overall flow of methodology.

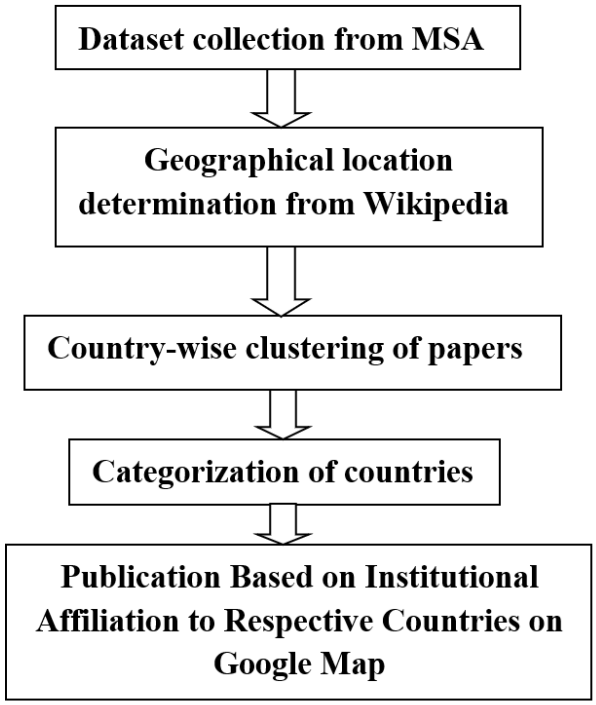

Fig. 1. Methodology Flow.

3.1. All-Inclusive dataset collection:

The examination of data sink and source countries in the domain of chemistry relies on the dataset gathered from the MAS. Our research is only restricted to Chemistry despite the availability of several groups of knowledge on the MAS. The MAS contains various types of publications, for instance, books, journal articles, conference proceeding papers, reports and thesis. The MAS classifies Chemistry publications into 18 different classes, and the top 100 papers were chosen from each of the 18 classes on the basis of references for tracing the trends of data dissemination in the said classes. Moreover, the top 1000 citations of all root papers are collected in the data set. The dataset so gathered contains knowledge regarding the authors, author's institute of association, publication date and venue. However, analysis is conducted by taking into consideration only the first author of a publication.

Table 1. Dataset Overview.

Domain	Categories	Root Papers	Citation
Chemistry	18	1800	1800000

3.2. Determination of Geographical Location

Wikipedia has been used by us to gather the author's geographical location by locating the author's university. To find out the geographic location of the author, each paper was geo-located, harnessing the

organizational association of the corresponding author as the MAS classified the authors only on the basis of organizational association; therefore, to geo locate the authors on Google map, the country of the author was required to be gathered. The country of each author belonging to a specific institute was gathered from Wikipedia.

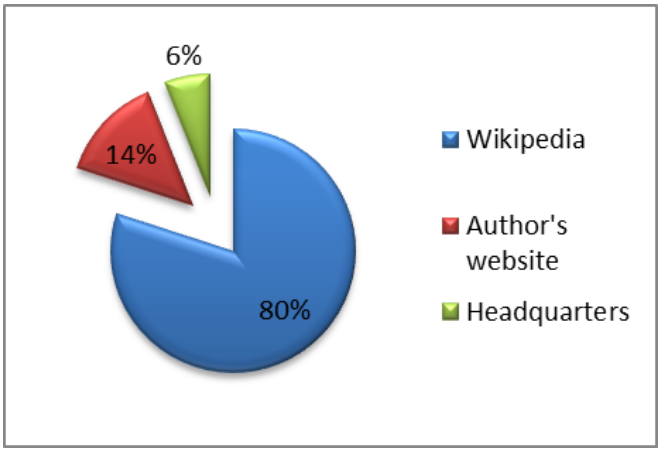

Fig. 2. Extracted Countries from Institutional Affiliation.

3.3. Country-wise clustering of research journals

In this step, we gathered the geographical location of each author in the dataset and the papers and their references were classified on the basis of the author's country. The authors who are associated with the same country were categorized into a single group.

3.4. Mapping publication on the basis of connection to a specific country on Google Map

By utilizing the Google visualization developer, each country in the dataset was mapped on a Google map. Every group is represented by a circle of radius equal to the number of publications connected to that specific country.

3.5. Trend determination from geo-located records

For determining the trends from geo-located records, the data set of the field of Chemistry gathered from the MAS was geo located on Google map. All the 18 classes of Chemistry were selected for examination, and about 1800000 citations were gathered. For each class top, 100 papers were chosen as Root papers. Top 1000 citations for each root paper were also gathered. Trends were located by utilizing the organizational associations of the only first author.

4. Experiments and Results

Different steps are performed for the identification of Sources and Sink countries. The Root papers R_p for each country are extracted. C_t is the total number of citations received by a country (for total R_p). The number of internal citations (citations received within a country) is given by Ci, and C_e represents external citations. C_e is calculated by subtracting the citations received within the country, i.e. internal citations, from the total citations obtained by the root papers.

$$C_e = C_t - C_i \quad (1)$$

This removes the effect of self-citations. The ratio R of external citation C_e to R_p of the total root papers is calculated for each country i.e.

$$R = C_e / R_p \quad (2)$$

The baseline scenario is defined to consider the country as a producer or consumer. IF the ratio of root papers to external citations R>= 1 country is a producer else, If the ratio of root papers of external citation R>=0 .75 and R=<0.99, then the country is an emerging knowledge producer country else Consumer country.

4.1. Results

Main knowledge-producing countries are the United States, Germany, the United Kingdom and China. Whereas Asian countries are mostly knowledge consumers. Moreover, the Developing world is largely dependent on the knowledge produced by the developed world. The number of papers and citations is directly proportional to the size of a country or city. Industrial countries perform better, but developing scientific powers like China and Japan are catching up quickly. The ratio R is relatively high in European countries. The cause of the low production ratio in developing countries is that many people immigrate to the developed nations' institutions in pursuit of higher education.

Analysis of knowledge-producing countries

Figure **3** examines **k**nowledge**-**producing countries. Figure-**3 also** provides listed top nine countries in the table**,** although **k**nowledge**-**producing countries are in large number. We can also examine in Figure **3** that **the** U.S.A. tops the list**,** and most of the European countries are **k**nowledge**-**producing countries in mathematics. **I**n our systems**,** knowledge**-**producing and consuming countries are determined. On the basis of citation, we have **developed the following rules:**

"IF ratio of Root Papers to External Citations (Rp/Ce) >= 1 Country Is Producer."

Where Rp signifies total root papers published in the given time frame collected for each country and Ce is External citations. Figure 3 pie chart shows the visualization of knowledge-producing countries and the percentage contribution of the top nine countries.

Country name	Ratio of R_p/C_e (R)
United States	1.3122
Germany	1.2887
Japan	1.1955
France	1.1863
Canada	1.1181
China	1.1627
United Kingdom	1.0951

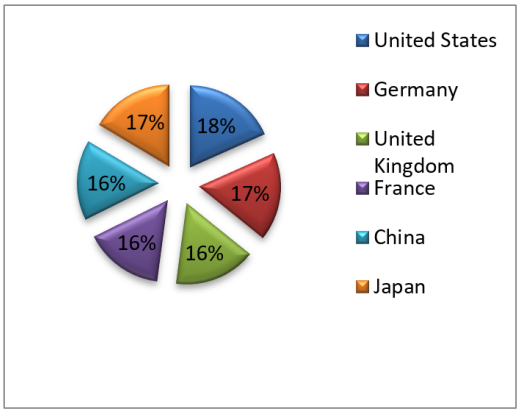

Fig. 3. Knowledge Producer Countries.

Analysis of Knowledge Consuming Countries

Figure 4 exhibits the knowledge-consuming countries. From the list of knowledge-consuming countries, we selected ten countries from the top of the list, although there are so many counties that consume knowledge.

Country name	Ratio of R_p/C_e (R)
India	0.3347
Italy	0.3490
South Africa	0.3174
Sri Lanka	0.2983
Zimbabwe	0.2411
Singapore	0.2157
Thailand	0.2190
Indonesia	0.2106
Colombia	0.1913

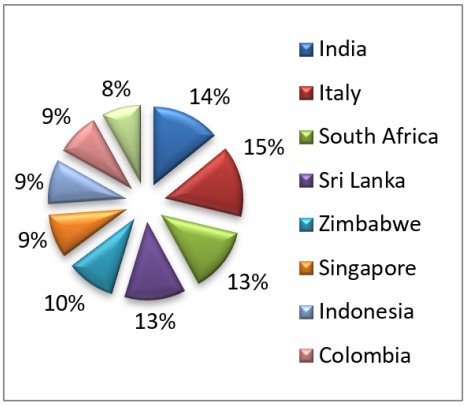

Fig. 4. Knowledge Consumer Countries.

Result Analysis of Emerging Knowledge Producing countries

Our system explored that some of the European countries like Italy, Sweden, Denmark, Iceland, and Belgium, while in Asia, Singapore and Russian Federation are the developing knowledge-producing countries. Figure 5 presents the details of these developing knowledge-producing countries and their contribution to knowledge production.

Country name	Ratio of R_p/C_e (R)
Russia	0.9963
Australia	0.9770
UAE	0.9897
Chille	0.9772
Oman	0.9760
Austria	0.8217
Brazil	0.7937
Hungary	0.7759

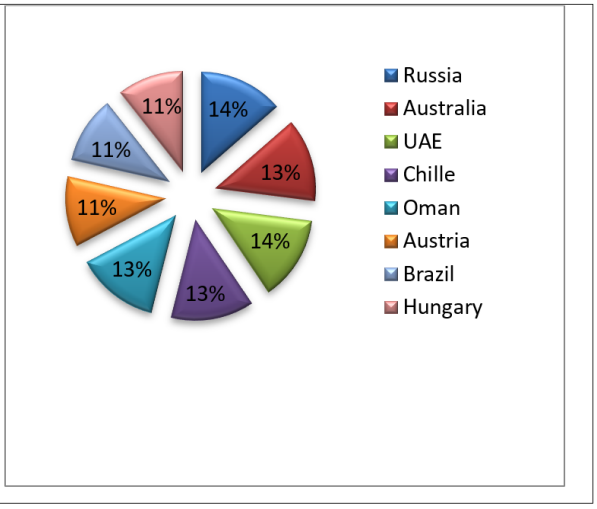

Fig. 5. Knowledge Consumer Countries.

5. Conclusion

By developing and utilizing this system, we came to the conclusion that to understand the flow of knowledge, it is essential to identify the knowledge-producing and consuming countries. Furthermore, through this study, we also concluded that knowledge is mainly produced in developed countries such as Europe, Japan and North America (U.S.A.), while knowledge is largely consumed in the developing countries, including most the Asian and African Countries. Some of these countries are developing in Knowledge like Russia, New Zealand Portugal etc. We have also studied that knowledge creation and diffusion measure the ecosystem of science which is the most vital asset of science. In future, this technique can be utilized to analyze the importance of knowledge-producing and consuming countries in any specific discipline or domain.

References

1. Kuri, R., & Hajje, V. (2014). Citation analysis of Pearl: A journal of library and information science. Asian Journal of Multidisciplinary Studies, 2(9), 8-13.
2. Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. Science, 178(4060), 471-479.
3. Ho, J. C., Saw, E. C., Lu, L. Y., & Liu, J. S. (2014). Technological barriers and research trends in fuel cell technologies: A citation network analysis. Technological Forecasting and Social Change, 82, 66-79.

4. Mazloumian, A., Helbing, D., Lozano, S., Light, R. P., & Börner, K. (2013). Global multi-level analysis of the 'Scientific Food Web'. Scientific reports, 3(1), 1-5.
5. Tsai, C. C., & Lydia Wen, M. (2005). Research and trends in science education from 1998 to 2002: A content analysis of publication in selected journals. International journal of science education, 27(1), 3-14.
6. Guerin, M. B., Flynn, T. H., Brady, J., & O'Brien, C. J. (2009). Worldwide geographical distribution of ophthalmology publications. International Ophthalmology, 29(6), 511-516.
7. Tijssen, R. J. (2007). Best of both worlds? Domestic relevance vs. international visibility of local science journals in developing countries. Centre for Science and Technology Studies (CWTS), The Netherlands. Available at: http://WWW. codes ria. org/IMG/pdf/RobertiTijssenipaper. pdf.
8. Mazloumian, A., Helbing, D., Lozano, S., Light, R. P., & Börner, K. (2013). Global multi-level analysis of the 'Scientific Food Web'. Scientific reports, 3(1), 1-5.
9. SETHI, B. B., & Maharana, B. (2015). Research Publication Trend of Two Periodicals, "Political Geography" and "Religion": A Bibliometric Perception.
10. Kuri, R., & Hajje, V. (2014). Citation analysis of Pearl: A journal of library and information science. Asian Journal of Multidisciplinary Studies, 2(9), 8-13.
11. Chen, C., & Hicks, D. (2004). Tracing knowledge diffusion. Scientometrics, 59(2), 199-211.
12. Ma, F. C., Lyu, P. H., Yao, Q., Yao, L., & Zhang, S. J. (2014). Publication trends and knowledge maps of global translational medicine research. Scientometrics, 98(1), 221-246.
13. Grossetti, M., Eckert, D., Gingras, Y., Jégou, L., Larivière, V., & Milard, B. (2014). Cities and the geographical deconcentration of scientific activity: A multilevel analysis of publications (1987–2007). Urban Studies, 51(10), 2219-2234.