Data Science and Engineering (DSE) Record, Volume 1, issue 1.

Biomedical Research Collaboration: Method of Parameter Extraction and Analysis using MEDLINE Database

Thiraphat Tanphiriyakun[1,2] and Phasit Charoenkwan[1,3]

[1] Data Science Program, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Orthopaedics, Chiang Mai University, Chiang Mai, Thailand
[3] College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
thiraphat_ta@cmu.ac.th

Abstract. This study proposed methods for collaboration information extraction using affiliation from MEDLINE database. The protocol enabled a large scale dataset extraction. The results show increasing trends of collaboration parameters year by year and in a higher Journal Quartile score.

Keywords: Biomedical Research Collaboration · Orthopaedics Research · Bibliometric Study

1 Introduction

A number of biomedical research publication continues to grow in the recent decade. Accordingly, a co-authorship and collaborative publication also have been shown in increasing trends[1]. Research collaboration is considered performance indicator to evaluate institute's performance[2]. However, MEDLINE, the largest biomedical research database does not provide collaboration information in abstract's affiliation search. This limitaon results in lack of quantitative bibliometric study in biomedical research. This study aimed to explore method and available parameters to assess collaboration in research publication using Orthopaedics publication dataset.

2 Literature reviews

International, national, and institutional collaboration are metrics which contain in research affiliation[3]. Many useful affiliation parser tools have been developed to retrieve important information such as institute of author and country of author[4] [5]. Nonetheless, we addressed the problems described as followed; (i) the Parsers allow processing in a single record, still no tools were able to process large amount on information in the same time; (ii) some inaccuracies regarding to abbreviation of country such as ITA: Italy and state name such as New Mexico and Mexico.

3 Materials and Methods

This study was waived from institutional ethic committee board approval. Affiliation information from MEDLINE database was retrieved via PubMed EDirect application programming interface (API)[6][7]. Data pre-processing and exploratory analysis were done using Python Programming Language version 3.7. Tableau 2019.3 (Tableau, Seattle, Washington, United States) was used for data visualisation. A dataset of 104,160 Orthopaedics articles from 3,922 journals between January 2010 to October 2019 were enrolled (Fig. 1). Journal Quartile (Q) information was obtained from ScimagoJr[8] database and matched with the published journals. The extraction processes were (i) adding country's abbreviation and states to the model; (ii) iterating the extraction process of the whole dataset (Fig. 2). The affiliation information in each record was splitted to list using a pipe character (|). Then, affiliations in the list were continuously parsed one by one using the PubMed parser tool[5]. The list of first authors, co-authors, list of countries, and list of institutes were obtained. Accordingly, collaboration parameters in each article were calculated.

4 Results

A set of 105,240 articles was obtained from MEDLINE database. We removed duplicated rows and missing data. The remaining 104,160 articles were proceeded to this study. The extracted parameters of collaboration were average number of author, institute, country, ratio of Orthopaedics and non-Orthopaedics authors of each article. The average value of collaboration parameters were listed by year and Journal Quartile (Q), shown in heatmap table (Fig. 3). The intensity in the heatmap visualization refers to mean of the value in each year. The increasing trends of collaboration were found in each year and higher ranking Journal Quartile. The ratio of Orthopaedics and non-Orthopaedics author was lower in higher ranking Journal Quartile.

5 Discussion and Conclusion

This study proposed methods to retrieve collaboration information from MEDLINE database.

The results showed positive trends of all collaboration parameters in the Orthopaedics publication from 2010 to 2019. These collaboration also increase in higher ranking Journal Quartile.

We addressed the issues of the previous Parser and were able to extract countries of author and institutes of author in a large dataset using modified version and iterating process (Fig. 1 and Fig. 2). This proposed method is useful for bibliometric biomedical researchers to extract collaboration information in MEDLINE database.

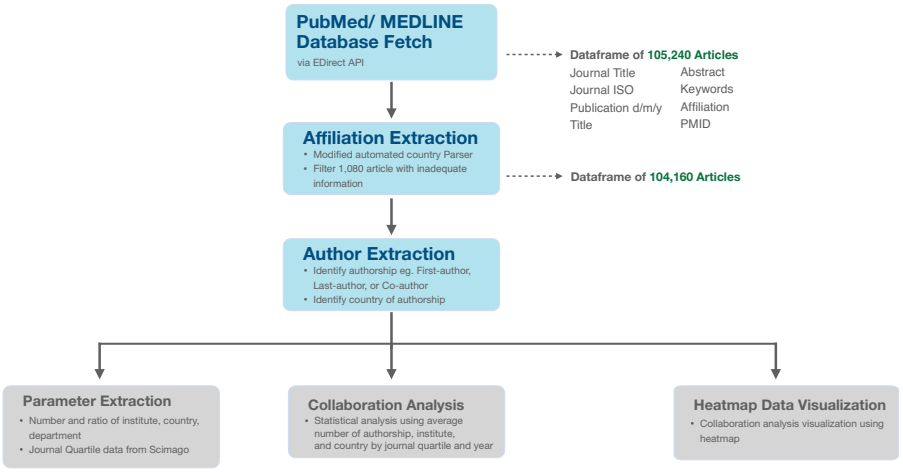

Fig. 1. The protocol of information retrieval in this study

Fig. 2. The process of affiliation information extraction

	Year	Quartile Q4	Q3	Q2	Q1
Average Number of Author in Article	2010	1.000	1.166	1.023	1.027
	2011	1.176	1.249	1.056	1.067
	2012	1.516	1.400	1.117	1.110
	2013	2.510	2.042	1.534	1.474
	2014	3.545	3.590	3.995	4.470
	2015	2.982	4.451	5.303	5.648
	2016	3.095	4.804	6.318	6.642
	2017	2.821	5.022	6.775	7.268
	2018	5.266	5.838	7.524	7.944
	2019	5.395	6.268	7.653	7.937
Average Number of Institute in Article	2010	1.000	1.056	1.010	1.013
	2011	1.032	1.144	1.003	1.031
	2012	1.143	1.149	1.053	1.042
	2013	1.643	1.286	1.169	1.156
	2014	1.753	1.776	1.959	2.161
	2015	1.493	2.000	2.267	2.617
	2016	1.546	2.101	2.696	2.899
	2017	1.509	2.150	2.738	3.062
	2018	2.349	2.521	3.010	3.285
	2019	2.425	2.701	3.014	3.357
Average Number of Country in Article	2010	1.000	1.017	1.006	1.002
	2011	1.008	1.010	1.003	1.009
	2012	1.056	1.024	1.010	1.011
	2013	1.143	1.042	1.032	1.032
	2014	1.260	1.134	1.156	1.221
	2015	1.109	1.158	1.199	1.341
	2016	1.111	1.182	1.245	1.404
	2017	1.171	1.187	1.286	1.434
	2018	1.505	1.310	1.400	1.454
	2019	1.569	1.383	1.368	1.466
Ratio of Orthopaedic vs Non-orthopaedic Author	2010		1.988	0.806	1.150
	2011	1.167	1.792	1.774	0.911
	2012	1.321	1.487	2.176	1.722
	2013	1.473	1.808	1.424	1.916
	2014	1.625	1.847	1.510	1.561
	2015	1.888	1.829	1.636	1.484
	2016	2.227	1.718	1.602	1.450
	2017	1.831	1.965	1.701	1.442
	2018	1.879	1.749	1.661	1.578
	2019	1.949	1.842	1.679	1.566

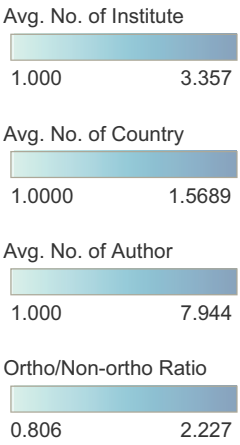

Fig. 3. The extracted collaboration parameters of Orthopaedics publication between January 2010 to October 2019 in MEDLINE database.

References

1. Fontelo, P., Liu, F.: A review of recent publication trends from top publishing countries. Systematic Reviews 7(1), 147 (sep 2018), https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-018-0819-1
2. Time: World University Rankings 2019: methodology | Times Higher Education (THE), https://www.timeshighereducation.com/world-university-rankings/world-university-rankings-2019-methodology
3. Elsevier: Research Metrics Guidebook | Research Intelligence | Elsevier, https://www.elsevier.com/research-intelligence/resource-library/research-metrics-guidebook
4. Torvik, V.I.: MapAffil: A bibliographic tool for mapping author affiliation strings to cities and their geocodes worldwide. D-Lib Magazine 21(11-12), 1–1 (2015)
5. Achakulvisut, T., Acuna, D.E., Cybulski, T., Hassan, T., Badger, T.G., H-Plus-Time, Brandfonbrener, D.: titipata/pubmed_parser: Pubmed Parser (oct 2016), https://zenodo.org/record/159504
6. Sayers, E.: A General Introduction to the E-utilities (2010)
7. Kans, J.: Entrez Direct: E-utilities on the UNIX Command Line - Entrez Programming Utilities Help - NCBI Bookshelf (2013), https://www.ncbi.nlm.nih.gov/books/NBK179288/
8. Scimago: Scimago Journal & Country Rank, https://www.scimagojr.com/