Scientometric Study of the IEEE Transactions on Software Engineering 1980-2010

Brahim Hamadicharef

Abstract. In this paper a scientometric study on the IEEE Transactions on Software Engineering 1980-2010 is presented. Using the full records from the Thomson Reuters (ISI) Web of Science (WoS), the journal's bibliometric measures are examined in terms of growth of literature, authorship characteristics, country of origin, distribution of articles' citations and references, and finally graph network of the research collaborations. A keyword occurrence analysis was carried out on the articles' title and the results used to create TagClouds showing perceptually their research importance. Furthermore, these TagClouds were created for the full 3 decades, shorter 5-year periods and most recent years, providing insights into potential research trends and help to relate them to major historic contributions in software engineering.

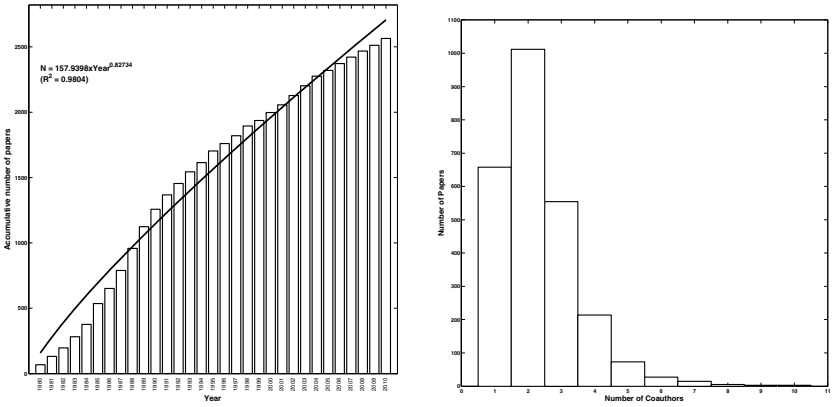
1 Introduction

The journal IEEE Transactions on Software Engineering (TSOFT) has been publishing article for more than 30 years and it was timely to carry out a scientometric study on its body of literature. The usefulness of such studies has been demonstrated as it allows not only to identify key bibliometric aspects of a particular research topic [1][2], a journal [3][4], or a series of conferences [5][6]. More than just providing bibliometric measures, it can also help young novice in their field to learn about eminent researchers in their field, identify key papers to read (typically reviews and surveys), journal to publish in. Furthermore, keyword analysis also allow to create network of research based on co-authorship (i.e. collaboration), research topics, etc.

The remainder of the paper is organized as follows. In Section 2 we detail the analysis methodology and results from the bibliometric study. Finally in Section 3, we conclude the paper.

Brahim Hamadicharef
Tiara, 1 Kim Seng Walk, Singapore 239403
e-mail: bhamadicharef@hotmail.com

F.L. Gaol et al. (Eds.): Proc. of the 2011 2nd International Congress on CACS, AISC 144, pp. 101–106.
springerlink.com



(a) (b)

Fig. 1 Growth of the TSOFT literature (a) and (b) Overall distribution of co-authorship

2 Methodology and Results

The Thomson Reuters ISI Web of Science (WoS) online interface was used to retrieve the records (with abstract) of the journal. Each year has one file with all the entries, saved as text files, and subsequently processed using MATLAB scripts. We present the growth of literature, authorship, citations, countries of origin, number of references, and international collaborations.

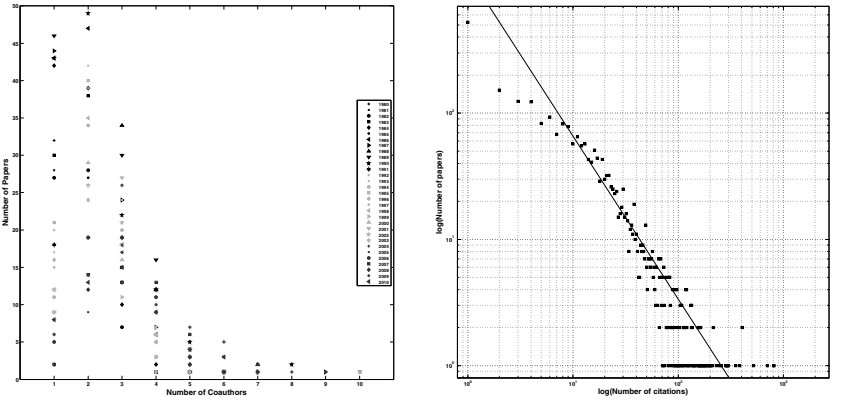
2.1 Growth of Literature

The growth of the TSOFT literature over the last 30 years (1980 to 2010) is shown in Figure 1(a). A quasi-linear growth can be observe, which is different from other studies [7], which usually showed power law characteristics.

Although it is interesting to observe such growth, one useful bibliometric measure to calculate is its *obsolescence*. In the 70s, researchers established mathematical law for describing the temporal loss of utility [8], an indicator of obsolescence [9] called half-life h (associated to an annual loss), that can be calculated for a set of documents. Using the 30 years of records, the half-life (h) for the TSOFT literature is equal to $h = 15.62$ with an annual loss of 4.34%.

2.2 Authorship

In Figure 1(b) the distribution of co-authors is shown with a peak at two co-authors. Looking at the (color-coded) yearly basis distribution revealed more details, as shown in Figure 2(a). For e.g., the contributions from recent years show a slightly larger number of co-authors compared to early years. In the 80s, in particular, a large number of articles had only one or two co-authors, whereas in recent years



(a) (b)

Fig. 2 Yearly distribution of co-authorship (a) and (b) Citations of the TSOFT literature

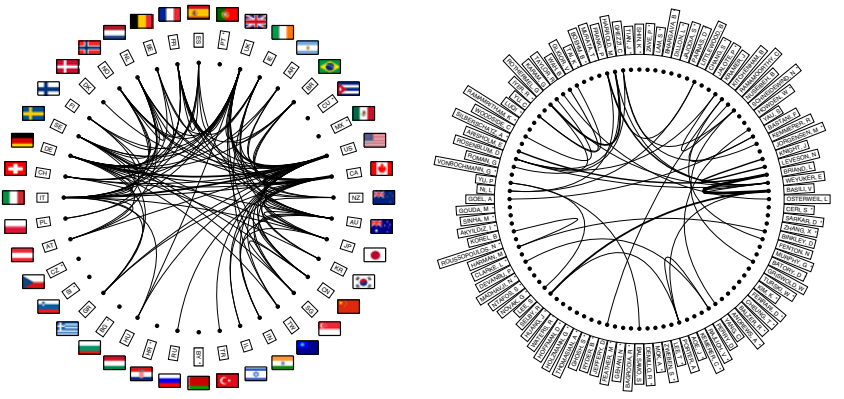
this seems to increase more papers having 5, 6 or 7 co-authors. One paper on *Detection of Mutual Inconsistency in Distributed Systems* has the highest number (10) of authors [10].

2.3 Citations

In Figure 2(b), the citation distribution (i.e. the number of citations versus the number of papers) is shown as a loglog plot. Such typical distribution can be fitted with a regression model of the form $y = a \times x + b$, with a=-1.32 and b=7.20 (goodness-of-fit R^2= 0.90). Notice that the model is a power law but shown as linear due to the loglog scales. The most cited paper is cited 811 times, TSOFT papers are on average cited 22.22 times (median 9.00 times). Some 13.38% of these publications have never been cited.

2.4 Countries of Origin

There are in total 46 countries contributing to the TSOFT literature (WoS has 1632 entries with the field country of origin) and include USA (932, 57.11%), Canada (108, 6.62%), Italy (103, 6.31%), U.K. (82, 5.02%), Germany (56, 3.43%), France (42, 2.57%), Norway (35, 2.14%), Japan (30, 1.84%), Taiwan (21, 1.29%), Australia (19, 1.16%), India (18, 1.10%), Israel (17, 1.04%), China (14, 0.86%), Belgium (13, 0.80%), Korea (13), Spain (11, 0.67%), Sweden (11), Greece (10, 0.61%), etc. In Figure 4(a), the graph of international collaborations is shown, with link between countries. Of all countries, the main strength is between the USA and Canada, and with U.K., Italy, Germany and France. In Figure 3(b), the international collaboration is shown at researchers' level (100 top researchers publishing in TSOFT).



(a) (b)

Fig. 3 TSOFT (a) countries collaboration and (b) researchers collaboration network

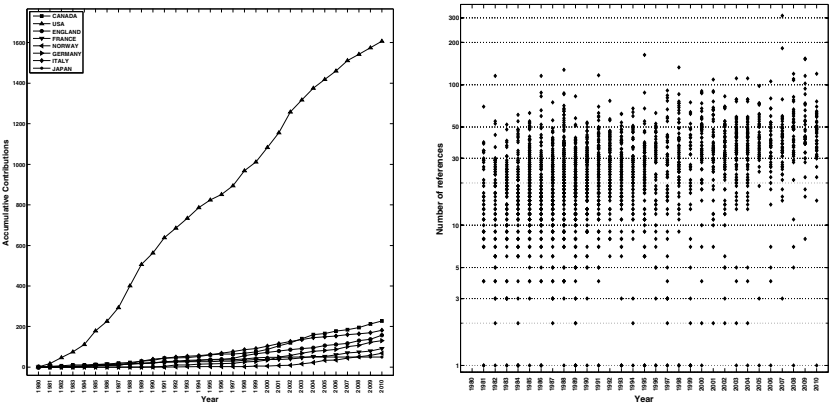
The evolution of accumulated contributions per countries is shown in Figure 4(a). From the very early years, the USA has the most significantly contributed to the TSOFT literature. It is worth noticing the increasing contributions from Canada, U.K, Italy, Germany and France, as well as Norway and Japan.

2.5 Number of References

From 1980 to 2010 the yearly average number of references per article grew from 12.89 to 42.85 (median 9.96 to 41.07). The article with the greatest number of reference is by M. Jørgensen and M. Shepperd on *a systematic review of software development cost estimation studies* [11] with 311 references. On the same year, J. Hannay and colleagues also published a TSOFT paper on *a systematic review of theory use in software engineering experiments* [12] with 182 references.

2.6 Keywords Analysis

The tool WORDLE [13] was used to to create TagClouds from the keywords distribution of the TSOFT articles' title. As shown in Figure 5, there are many keywords which are all equally sized, showing their importance in the research published in TSOFT. The 30 years of contribution was split into smaller 5-year periods, to be able to compare different advancements in software engineering from the 80s, the 90s and the first decade after the year 2000 (e.g. Figure 5(a) and Figure 5(b)). Focus was given to the recent years as well as an overall picture for the 30 years of publications as shown in Figure 5(c).



(a) (b)

Fig. 4 Accumulated contribution to the TSOFT literature per countries (a) and (b) Distribution of the number of references

(a) 2000-2004 (b) 2005-2009 (c) All

Fig. 5 TagClouds of TSOFT articles' title

3 Conclusions

After 30 years of publications, it was timely to present a scientometric study of the IEEE Transactions on Software Engineering. In this paper, based on bibliometric measures, characteristics such as growth of literature, authorship, country of origin, citations, references, were detailed. A graph was created to present the journal's network of collaborations between authors. Keywords from on the articles' title, and in particular their occurrence, were analyzed and the results used to create TagClouds. These are showing the key research topics within the software engineering research field. Furthermore, using the full 3 decades, shorter 5-year periods and most recent years, insights into potential research trends could be investigate and thus help to relate these topics to major historic contributions in software engineering.

References

1. Patra, S.K., Mishra, S.: Bibliometric study of bioinformatics literature. Scientometrics 67(3), 477–489 (2006)
2. Chiu, W.-T., Ho, Y.-S.: Bibliometric analysis of tsunami research. Scientometrics 73(1), 3–17 (2007)
3. Hamadicharef, B.: Scientometric Study of the Journal NeuroImage 1992–2009. In: Proceedings of the 2010 International Conference on Web Information Systems and Mining (WISM 2010), Nanjing, China, October 23-24, pp. 201–204 (2010)
4. Thanuskod, S.: Journal of Social Sciences: A Bibliometric Study. Journal of Social Sciences 24(2), 77–80 (2010)
5. Hamadicharef, B.: Bibliometric Study of the DAFx Proceedings 1998-2009. In: Proceedings of the 13th International Conference on Digital Audio Effects (DAFx 2010), Graz, Austria, September 4–6, pp. 427–430 (2010)
6. Bartneck, C., Hu, J.: Scientometric Analysis of the CHI Proceedings. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2009), Boston, USA, April 4–9, pp. 699–708 (2009)
7. Hamadicharef, B.: Brain–Computer Interface (BCI) Literature – A Bibliometric Study. In: Proceedings of the 10th International Conference on Information Science, Signal Processing and their applications (ISSPA 2010), Kuala Lumpur, Malaysia, May 10–13, pp. 626–629 (2010)
8. Brookes, B.C.: The Growth, Utility, and Obsolescence of Scientific Periodical Literature. Journal of Documentation 26(4), 283–294 (1970)
9. Line, M.B.: The 'half-life' of periodical literature, apparent and real obsolescence. Journal of Documentation 26(1), 46–54 (1970)
10. Parker, D.S., Popek, G.J., Rudisin, G., Stoughton, A., Walker, B.J., Walton, E., Chow, J.M., Edwards, D., Kiser, S., Kline, C.: Detection of Mutual Inconsistency in Distributed Systems. IEEE Transactions on Software Engineering 9(3), 240–247 (1983)
11. Jørgensen, M., Shepperd, M.: A Systematic Review of Software Development Cost Estimation Studies. IEEE Transactions on Software Engineering 33(1), 33–53 (2007)
12. Hannay, J.E., Sjoberg, D.I.K., Dyba, T.: A Systematic Review of Theory Use in Software Engineering Experiments. IEEE Transactions on Software Engineering 33(2), 87–107 (2007)
13. Viegas, F.B., Wattenberg, M., Feinberg, J.: Participatory Visualization with Wordle. IEEE Transactions on Visualization and Computer Graphics 15(6), 1137–1144 (2009)