Available online at www.sciencedirect.com
ScienceDirect
Information and Software Technology 50 (2008) 3–9
INFORMATION AND SOFTWARE TECHNOLOGY
www.elsevier.com/locate/infsof

An analysis of the most cited articles in software engineering journals – 2001

Claes Wohlin *

Department of Systems and Software Engineering, School of Engineering, Blekinge Institute of Technology, PO Box 520, S-372 25 Ronneby, Sweden

Available online 13 October 2007

Abstract

Citations and related work are crucial in any research to position the work and to build on the work of others. A high citation count is an indication of the influence of specific articles. The importance of citations means that it is interesting to analyze which articles are cited the most. Such an analysis has been conducted using the ISI Web of Science to identify the most cited software engineering journal articles published in 2001. The objective of the analysis is to identify and list the articles that have influenced others the most as measured by citation count. An understanding of which research is viewed by the research community as most valuable to build upon may provide valuable insights into what research to focus on now and in the future. Based on the analysis, a list of the 20 most cited articles is presented here. The intention of the analysis is twofold. First, to identify the most cited articles, and second, to invite the authors of the most cited articles in 2001 to contribute to a special section of Information and Software Technology. Three authors have accepted the invitation and their articles appear in this special section. Moreover, an analysis has been conducted regarding which authors are most productive in terms of software engineering journal publications. The latter analysis focuses on the publications in the last 20 years, which is intended as a complement to last year's analysis focusing on the most cited articles in the last 20 years [C. Wohlin, An Analysis of the Most Cited Articles in Software Engineering Journals – 2007, Information and Software Technology 49 (1) 2–11]. The most productive author in the last 20 years is Professor Victor Basili.

Keywords: Citations; Bibliometric research; Software engineering

1. Introduction

Citations are a common way of judging the most influential work in different fields. The most cited articles often provide new insights, open a new avenue of research, or provide a significant summary of the state-of-the-art in an area. Citations are a way to show how researchers build their work on existing research to evolve research further. Basically, they are the backbone of research and hence articles and authors being cited frequently deserve acknowledgment for their contribution. The analysis presented here is intended to highlight which articles are most cited in software engineering in a given year. The analysis is intended to complement the work on ranking the most published scholars and institutions as done by Tse et al. [7].

The objective of the analysis presented here is to list the most cited journal articles in the field of software engineering recorded as published in 2001. The analysis is based on the ISI Web of Science [2]. The ISI web covers the major journals in the field, creating a web of references that ensures that the overall picture obtained from the web is likely to give a representative view of the most cited articles. In particular, it means that references from journals included in the ISI Web of Science also gets included in the web and hence included in information available from the created web of references.

The analysis is published as a list of the 20 most cited articles, or in case of ties the actual number of articles may be higher, for example, two articles may be tied for position 20 and hence the list would contain 21 articles. Authors of the top ranked articles have been invited to

* Tel.: +46 457 385820; fax: +46 457 27125.
E-mail address: Claes.Wohlin@bth.se

0950-5849/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.infsof.2007.10.002

write a new article for this special issue. They have been asked to either write a follow-up article given that the previous work attracted high attention or some current research that they are conducting.

Two main issues should be kept in mind:

- The possible selection of journals is limited to what is available through the ISI Web of Science.
- The focus is on software engineering. This means that the actual selection of journals is based on selecting journals that are perceived as mainly publishing software engineering articles. All articles in these journals have been analyzed. This implies that no judgment has been made whether a specific article is within software engineering or not. The main reason being that it would not make the results from the analysis replicable and it would also create discussions of the borderlines between fields such as software engineering, computer science and information systems.

The intention is for the analysis provided here to be conducted and published on a yearly basis in a special issue or special section of Information and Software Technology.

The article is structured as follows. Section 2 presents the ISI Web of Science and the actual selection of journals in the analysis. In Section 3, the analysis method is described. This includes some information about the ISI Web of Science in relation to the analysis method and a presentation of how the top 20 list was generated. Section 4 presents the results in relation to the top 20 list. In Section 5, an analysis of the most productive authors of software engineering journal articles is presented. A short summary of the findings is provided in Section 6.

2. Selection decisions

A key issue when looking at citations is what to count. This includes both, which publications and which references. When it comes to the publications, any analysis is constrained by the support given by different databases or search engines. In the analysis presented here, it was decided to use the ISI Web of Science. The actual count of citations is further discussed in Section 3.

2.1. Tool support

The selection of which tool to use to count citations has a major impact on the actual outcome and hence on the trustworthiness of the findings. It is worth noting that bibliometric research is a field of its own. In this field, publication patterns are studied including both descriptive (for example counting the number of publications from an organization) and evaluative (for example counting citations as a measure of impact). The Institute for Scientific Information (ISI) has been leading in the field since its establishment in 1961 [4]. The metrics provided by ISI are being used for determining impact factors for journals [1], assess and drive bibliometric research [10], and support studies like ours in other fields such as medicine [3]. Sample checks with some of the author's own publications confirmed that ISI strengths apply to the software engineering domain as well as in medicine [3]. Hence, the position of the ISI data as a leading source for bibliometric research and the actual use of it in other fields were decisive factors when determining to use this tool for this study of the most cited articles in software engineering.

2.2. Journals

The objective of the ISI as a database is to provide a comprehensive coverage of the most important and influential research. The information about ISI is based on [6]. The database includes in total more than 8500 journals and some other sources, for example, Lecture Notes in Computer Science. However, journals are here used as a reference to the content of the database. The journals cover three areas: science, social sciences and arts and humanities. The ISI staff reviews close to 2000 journals yearly, but only 10–12% makes it into the database. An interesting feature is how the references build a web. ISI captures the cited references and citation information that is included both from journals in the database and for those journals not included in the database but which are cited from journals included. This ensures a good coverage of citations and also that the data extracted provides a good picture of actual citations.

The objective was to make a selection of journals that provide as fair picture as possible of the most cited articles in software engineering. No database or tool support was found that was capable of also including all references in conferences papers and book chapters and hence the analysis is made under the assumption that journals provide a representative picture of the most cited type of articles and authors.

The first column in Table 1 lists the journals selected from the database as a suitable set of software engineering journals. However not all of these journals were available in the database. Thus, the table also includes information about journals actually in the database in 2001 and remarks about journals when they appeared some other years, but not in year 2001. The journals in the database for 2001 are the journals actually used in the citation study for 2001.

Several things may be observed from the table and some issues are worth commenting. A subset of journals has been in the database several years and appears every year. Some journals are added specific years and others are removed. Finally, some journals or issues of journals or specific articles can be found in the database based on that they are cited by articles in the database.

3. Methods

The analysis is done focusing on the science part of the ISI Web of Science. The citation search is conducted as follows. The search is conducted for one specific year, for example, 2001. A list of journals is provided to the search engine within the Web of Science. This generates a com-

Table 1
Journals considered in the analysis

Journal	2001	Remarks
ACM Trans. on Software Engineering and Methodology (TOSEM)	Yes	
Annals of Software Engineering	No	Some issues, but none from 2001
Automated Software Engineering	No	Not in ISI at all
Empirical Software Engineering	No	From 2003
IEE Proceedings of Software Engineering[a]	No	Not in ISI at all
IEEE Software (Software)	Yes	
IEEE Trans. on Software Engineering (TSE)	Yes	
Information and Software TechnoloSgy (IST)	Yes	
International Journal of Software Engineering and Knowledge Engineering	Yes	
Journal of Software Maintenance and Evolution – Research and Practice[b]	Yes	From 2001
Journal of Software Maintenance – Research and Practice	No	Until 2000
Journal of Systems and Software (JSS)	Yes	
Requirements Engineering Journal	No	From 2004
Software Architecture	No	Only in 2004
Software Engineering Journal	No	Last issue came in 1996
Software Process – Improvement and Practice	No	Not in ISI at all
Software Quality Journal	Yes	
Software Testing Verification & Reliability	Yes	
Software – Concepts and Tools	No	Not after 1998
Software – Practice and Experience (SPE)	Yes	

[a] These proceedings replaced Software Engineering Journal from 1997.
[b] The name of the journal changed in 2001.

plete list of articles published in the journals in 2001. It is then possible to sort the list based on the number of citations. This resulted in a sorted list of the most cited articles in the journals listed in Table 1.

The list was generated on January 17th, 2007, and hence it may not be possible to exactly replicate the results presented below. The actual order of articles sorted according to citations may have changed due to the fact that new citations are made and hence are added to the database. It may also be the case that some articles are referenced frequently shortly after publication, while other articles may rise in citation as the years pass. The latter may be the case when a particular article is novel and opens a new avenue of research. Such articles may show an increase in citation as time goes by, since the article becomes accepted as a landmark in a specific area.

It should also be noted that some articles in the list appear as having been published in 2000. However, they appear when searching for articles in 2001. The reason for this may be that a specific issue from one year was actually released the year after. However, it was decided to not remove articles from the list, since this would cause other problems. If removing articles (due to that it shown as being published in 2000 although the ISI Web of Science lists it as being published in 2001) then there is a risk that some articles will not be covered at all in the analysis. Articles removed from 2001 will not appear when searching the database for 2000, and hence it was decided to use the list generated by the database.

Finally, self citations were removed. This was done manually, since no automatic way of filtering the citations was found. Self citation was defined as having at least one author in common with the original article. This means that research groups citing their own papers, but without any of the original authors have been kept.

Based on the rules above, the articles were sorted based on non-self citations and a "most cited" list emerged. When ties appear, the most cited article including self citation is listed first. However, the actual placement in terms of number in the list is not allowed to be affected by the self citations. It is worth mentioning that the removal of self citations only changed the internal order between articles slightly, and that the most cited articles were only marginally affected by removing the self citations. This is probably a result of that most researchers reference (for good reasons) their own work and hence no drastic changes in the list were observed when removing self citations.

4. Top 20 in 2001

The method described in Section 3 resulted in a list of the most cited articles in 2001 for software engineering. The list contains 20 articles with five articles tied for position 16 on the list. Only one author is represented twice on the list, namely Lionel Briand. The complete list can be found in Appendix A. An excerpt of the list is presented in Table 2, where the six most cited articles are listed. The table shows the rank of each article, basic information about the article and the number of times the article has been cited. The basic information contains the normal information provided when referencing research articles.

The number of citations is overall rather low compared with articles in other fields. However, this comparison depends on for example the size and definition of the field as well as to traditions such as whether to cite only journal articles or both journal and conference articles. When it comes to the overall rather low number of citations, it should be noted that conference proceedings are very com-

Table 2
Most cited software engineering articles in 2000 (top six)

Rank	Author Title of article Journal reference	Citations
1	F. Bellifemine, A. Poggi, G. Rimassa Developing multi-agent systems with a FIPA-compliant agent framework Software-Practice & Experience 31 (2) (2001) 103–128	48
2	G. Cugola, E. Di Nitto, A. Fuggetta The JEDI event-based infrastructure and its application to the development of the OPSSWFMS IEEE Transactions on Software Engineering 27 (9) (2001) 827–850	40
3	S.A. Deloach, M.F. Wood, C.H. Sparkman Multiagent systems engineering International Journal of Software Engineering and Knowledge Engineering 11 (3) (2001) 231–258	39
4	A.C. Myers, B. Liskov Protecting privacy using the decentralized label model ACM Transactions on Software Engineering and Methodology 9 (4) (2000) 410–442	27
5	F. Zambonelli, N.R. Jennings, M. Wooldridge Organisational rules as an abstraction for the analysis and design of multi-agent systems International Journal of Software Engineering and Knowledge Engineering 11 (3) (2001) 303–328	26
5	B. Ramesh, M. Jarke Toward reference models for requirements traceability IEEE Transactions on Software Engineering 27 (1) (2001) 58–93	26

mon in software engineering and references from conference papers to the journal papers are not included in the citation count. Moreover, the citation count is made after approximately five years (it depends slightly on when a specific article was made available) and the citations are expected to rise over the years. In particular, it should be noted that a citation means when an article has become published, then others should build on that work and also publish the article. This means that new research should be conducted and articles should have time to go through the process from submission to publication in a journal before resulting in a citation (as counted by the ISI Web of Science).

Based on the list in Table 2 and Appendix A, a set of authors was invited to contribute with an article to this special section of Information and Software Technology. No requirements were put on the authors regarding the content, but they were informed that all articles would go through the normal review process. In the process of inviting authors, it turned out that not all invited authors were able to contribute for different reasons. Anyway, three researchers accepted the invitation and their contributions can be found in this special section, although in some cases with other co-authors than their article listed for 2001.

It is interesting also to study which journals appear on the list. It turns out that most articles listed among the top 20 are published in the journals also included in the study presented yearly by the Journal of Systems and Software [7]. In addition, four entries on the list are from more specialized journals, i.e. International Journal of Software Engineering and Knowledge Engineering, and Journal of Software Maintenance and Evolution – Research and Practice, which is very similar to the outcome when analyzing the most cited articles from 2000 [9]. It can be noted that the most cited articles are published in IEEE Transactions on Software Engineering, although it is worth noting that one issue of the International Journal of Software Engineering and Knowledge Engineering has three articles on the list. However, it is probably worth noting that it is not a matter of articles in these journals getting cited more than others per se. It is more a matter of that articles that are likely to get highly cited are submitted to the journals. On the other hand, most software engineering researchers today have access to most journals in Table 1 electronically. For example, Information and Software Technology is accessible through ScienceDirect [5], which a large number of university libraries have access to and hence the researchers at those universities. This means that articles published in any of the journals listed in Table 1 stands a fair chance of being highly cited.

Nevertheless, the top 20 list is dominated by the IEEE Transactions in Software Engineering in a similar way as the analyses published in [8] for 1999 and in [9] for 2000. Seven out of 20 articles on the list is published in IEEE Transactions in Software Engineering. Three journals have three articles each on the list, one journal has two articles on the list and then finally two journals have one article each on the list.

5. Most productive authors in the last 20 years

In the last year's special issue [9], an analysis of the most cited articles in the last 20 years was provided in addition to the analysis of the most cited articles in 2000. As a complement to this view on the last 20 years, this section presents an analysis of the most productive software engineering authors in the last 20 years, where productive is measured in terms of journal publications in the journals available through the ISI Web of Science. In other words, in the journals listed in Table 1. The search was done for articles published in 1988–2007. The focus was on research articles published in English. In total, 9144 articles were published in the journals in Table 1. The analysis was conducted on

Table 3
Most productive authors

Rank	Authors	Number of articles
1	V. Basili and V.R. Basili	44
2	B. Kitchenham and B.A. Kitchenham	39
3	T.Y. Chen	35
3	T.M. Khoshgoftaar	35
3	C. Wohlin	35
6	R. Glass and R.L. Glass	32

August 6th, 2007. The main focus of the ISI Web of Science is to focus on articles and journals. However, it is possible to get a list of the top 100 authors in terms of articles published in the analyzed time frame. The top six is shown in Table 3, and the top 20 list can be found in Appendix B.

Several authors use different acronyms when writing articles as for example Basili, where he can be found as an author both as V. Basili and V.R. Basili. In the ISI Web of Science, he is listed with 32 articles as V.R. Basili and with 12 articles as V. Basili. This also points to a potential threat to the analysis. It is possible that authors have more articles than actually listed if they have published articles using different combinations of acronyms. However, the intention has been to try to identify all articles by the authors by searching on their surnames in the database, although minor deviations cannot be ruled out. Another potential threat is that people have changed surname during these 20 years. The latter is more or less impossible to handle in the analysis.

Based on this type of analysis it is also possible to study how different authors have contributed over the years. In other words, it is possible to study for example the half time splits that make up these 20 years by studying contributions between 1988 and 1997, respectively, between 1998 and 2007. These splits are for the top six authors in Table 4.

From Table 4, it is possible to see that several authors have many contributions in both periods, which illustrate a long and productive research career. Others have many more articles in one of the periods, which show that researchers have entered the field or they may have left the field. The latter does not seem to be the case for the researchers listed in Table 4. By studying different time periods, it is possible to see how different authors production of journal articles changes over time, and hence of course also the ranking in different time periods as for example illustrated in [7], where scholars are ranked based on their publications in a five year time period.

Table 4
Most productive authors and their splits

Rank	Authors	1988–1997	1998–2007	Number of articles
1	V. Basili and V.R. Basili	23	21	44
2	B. Kitchenham and B.A. Kitchenham	14	25	39
3	T.Y. Chen	9	26	35
3	T.M. Khoshgoftaar	17	18	35
3	C. Wohlin	2	33	35
6	R. Glass and R.L. Glass	16	16	32

6. Summary

The analysis here is intended to highlight and acknowledge the articles attracting most citations within software engineering. Insights into what is viewed as important to build upon may provide valuable insights into both what research is important and where the field of software engineering is heading. The ISI Web of Science has been used to identify the most cited software engineering journal articles. The analysis has this year been focused on year 2001 and the most productive software engineering researchers in the last 20 years.

A top 20 list for year 2001 has been presented. Some authors of the most cited articles in 2001 were invited to contribute to a special section of Information and Software Technology. Three such articles can be found in this issue.

Furthermore, a list ranking the most productive authors of software engineering journal articles was presented. The productivity was measured as the number of journal publications based on searching the ISI Web of Science [6]. The analysis revealed that Professor Victor Basili has been the most productive researcher in terms of journal publications in the last 20 years. By dividing the 20 years into two periods of 10 years each, it was possible to study how the production of journal articles has changed over time. For example, the analysis shows that Professor Basili has contributed almost with an equal number of publications in both periods with 23 articles, respectively, 21 articles while others have a more uneven distribution over the 20 years. The latter may be due to people entering or leaving the field.

The analysis of the most cited software engineering journal articles in a specific year has been made to a yearly tradition for Information and Software Technology, and to invite the most cited authors to contribute to a special issue of the journal. This means that next year's analysis will focus on the most cited software engineering journal articles in 2002. In addition, the intention is to provide some extra analysis every year as we have this year studied the most productive researchers in the last 20 years.

Acknowledgment

This work was partly funded by The Knowledge Foundation in Sweden under a research grant for the project "Blekinge – Engineering Software Qualities (BESQ)" (http://www.bth.se/besq).

Appendix A. Top 20 cited software engineering articles in 2001

The list in Table A1 shows a ranking of the most cited software engineering articles in 2001. The citations count was made 17 January 2007.

Table A1
Ranking of most cited articles

Rank	Author Title of article Journal reference	Citations
1	F. Bellifemine, A. Poggi, G. Rimassa Developing multi-agent systems with a FIPA-compliant agent framework Software-Practice & Experience 31 (2) (2001) 103–128	48
2	G. Cugola, E. Di Nitto, A. Fuggetta The JEDI event-based infrastructure and its application to the development of the OPSSWFMS IEEE Transactions on Software Engineering 27 (9) (2001) 827–850	40
3	S.A. Deloach, M.F. Wood, C.H. Sparkman Multiagent systems engineering International Journal of Software Engineering and Knowledge Engineering 11 (3) (2001) 231–258	39
4	A.C. Myers, B. Liskov Protecting privacy using the decentralized label model ACM Transactions on Software Engineering and Methodology 9 (4) (2000) 410–442	27
5	F. Zambonelli, N.R. Jennings, M. Wooldridge Organisational rules as an abstraction for the analysis and design of multi-agent systems International Journal of Software Engineering and Knowledge Engineering 11 (3) (2001) 303–328	26
5	B. Ramesh, M. Jarke Toward reference models for requirements traceability IEEE Transactions on Software Engineering 27 (1) (2001) 58–93	26
7	S.G. Eick, T.L. Graves, A.F. Karr, J.S. Marron, A. Mockus Does code decay? Assessing the evidence from change management data IEEE Transactions on Software Engineering 27 (1) (2001) 1–12	25
8	C.C. Michael, G. McGraw, M.A. Schatz Generating software test data by evolution IEEE Transactions on Software Engineering 27 (12) (2001) 1085–1110	24
9	B. Bauer, J.P. Muller, J. Odell Agent UML: A formalism for specifying multiagent software systems International Journal of Software Engineering and Knowledge Engineering 11 (3) (2001) 207–230	21
10	J.J. Dolado On the problem of the software cost function Information and Software Technology 43 (1) (2001) 61–72	20
11	C.L. Hsu, T.S. Wu, T.C. Wu New nonrepudiable threshold proxy signature scheme with known signers Journal of Systems and Software 58 (2) (2001) 119–124 Sep 1 2001	19
12	J. Wegener, A. Baresel, H. Sthamer Evolutionary test environment for automatic structural testing Information and Software Technology 43 (14) (2001) 841–854 Sp. Iss. SI	18
12	L.C. Briand, C. Bunse, J.W. Daly A controlled experiment for evaluating quality guidelines on the maintainability of object-oriented designs IEEE Transactions on Software Engineering 27 (6) (2001) 513–530	18
12	D. Whitley An overview of evolutionary algorithms: practical issues and common pitfalls Information and Software Technology 43 (14) (2001) 817–831 Sp. Iss. SI	18
15	J. Boyland Alias burying: Unique variables without destructive reads Software-Practice & Experience 31 (6) (2001) 533–553	17
16	B.A. Kitchenham, R.T. Hughes, S.G. Linkman Modeling software measurement data IEEE Transactions on Software Engineering 27 (9) (2001) 788–804	15
16	N. Chapin, J.E. Hale, K.M. Khan, J.F. Ramil, W.G. Tan Types of software evolution and software maintenance Journal of Software Maintenance and Evolution-Research and Practice 13 (1) (2001) 3–30	15
16	L.C. Briand, J. Wust Modeling development effort in object-oriented systems using design properties IEEE Transactions on Software Engineering 27 (11) (2001) 963–986	15
16	D.S. Wallach, A.W. Appel, E.W. Felten SAFKASI: A security mechanism for language-based systems ACM Transactions on Software Engineering and Methodology 9 (4) (2000) 341–378	15
16	E.M. Clarke, S. Jha, W. Marrero Verifying security protocols with Brutus ACM Transactions on Software Engineering and Methodology 9 (4) (2000) 443–487	15

Table B1
Ranking of most productive authors

Rank	Authors	Number of articles
1	V. Basili and V.R. Basili	44
2	B. Kitchenham and B.A. Kitchenham	39
3	T.Y. Chen	35
3	T. Khoshgoftaar	35
3	C. Wohlin	35
6	R. Glass and R.L. Glass	32
7	R. Lai	29
8	L. Briand and L.C. Briand	28
9	G. Rothermel	26
10	E. Weyuker and E.J. Weyuker	25
11	S.L. Pfleeger	24
12	G. Canfora	23
13	C.C. Chang	21
14	M.J. Harrold	20
14	M. Jorgensen	20
14	M. Shepperd and M.J. Shepperd	20
17	S.C. Chou	19
17	A. Cimitile	19
17	M. Harman	19
17	Luqi	19
17	P. Runeson	19
17	K. Saleh	19

Appendix B. Most productive software engineering researchers in 1988–2007

The list in Table B1 shows a ranking of the most productive researchers in software engineering based on articles published in 1988–2007. The list was generated on 6 August 2007 from the ISI Web of Science [2].

References

[1] W. Glanzel, H.F. Moed, State of the art report: journal impact measures in bibliometric research, Scientometrics 52 (2) (2002) 171–193.
[2] http://isi3.isiknowledge.com (last visit 4 october, 2007).
[3] A.W. Jones, Crème de la crème in forensic science and legal medicine: the most highly cited articles, authors and journals 1981–2003, International Journal of Legal Medicine 119 (2005) 53–65.
[4] M.K. McBurney, P.L. Novak, What is bibliometrics and why should you care?, Proceedings of the Professional Communication Conference, pp. 108–114, 2002.
[5] http://www.sciencedirect.com (last visit 8 October, 2007).
[6] http://www.thomson.com/scientific/scientific.jsp (last visit 4 october, 2007).
[7] T.H. Tse, T.Y. Chen, R.L. Glass, An assessment of systems and software engineering scholars and institutions (2000–2004), Journal of Systems and Software 79 (6) (2006) 816–819.
[8] C. Wohlin, An analysis of the most cited articles in software engineering journals – 1999, Information and Software Technology 47 (15) (2005) 957–964.
[9] C. Wohlin, An analysis of the most cited articles in software engineering journal – 2007, Information and Software Technology 49(1) 2–11.
[10] M. Zitt, S. Ramanan-Rahary, E. Bassecoulard, Relativity of citation performance and excellence measures: from cross-field to cross-scale effects of field-normalisation, Scientometrics 63 (2) (2005) 373–401.