

## **An Analysis of the Most Cited Articles in Software Engineering Journals - 2002**

Claes Wohlin

*Dept. of Systems and Software Engineering,  
School of Engineering,  
Blekinge Institute of Technology,  
PO Box 520, S-372 25 Ronneby, Sweden  
Claes.Wohlin@bth.se  
Phone: +46 457 385820, Fax: +46 457 27125*

### **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 2002. 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 2002 to contribute to a special section of Information and Software Technology. Six authors have accepted the invitation and their articles appear in this special section.*

## 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. [Tse06].

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 2002. The analysis is based on the ISI Web of Science [ISI3]. 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 write a new article for this special section. 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. A short summary of the findings is provided in Section 5.

## 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 [McBurney02]. The metrics provided by ISI are being used for determining impact factors for journals [Glänzel02], assess and drive bibliometric research [Zitt05], and support studies like ours in other fields such as medicine [Jones05]. 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 [Jones05]. 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 [Thomson05]. The database includes in total more than 8500 journals and some other sources, for example, for some years Lecture Notes in Computer Science are included. However, journals are here used as a reference to the content of the database. The journals cover three areas: science, social sciences and arts & 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 tool being closest to include also conferences and so forth is probably GoogleScholar, but it was decided to use ISI since it is better established and used in determining the impact factor for different journals.

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 2002 and remarks about journals when they appeared some other years, but not in year 2002. The journals in the database for 2002 are the journals actually used in the citation study for 2002.

**Table 1. Journals considered in the analysis.**

<b>Journal</b>	<b>2002</b>	<b>Remarks</b>
ACM Trans. on Software Engineering and Methodology (TOSEM)	Yes	
Annals of Software Engineering	Yes	Two issues from 2002
Automated Software Engineering	No	Not in ISI at all
Empirical Software Engineering	No	From 2003
IEE Proceedings of Software Engineering	No	Not in ISI at all
IEEE Software (Software)	Yes	
IEEE Trans. on Software Engineering (TSE)	Yes	
Information and Software Technology (IST)	Yes	
International Journal of Software Engineering and Knowledge Engineering	Yes	
Journal of Software Maintenance and Evolution - Research and Practice	Yes	From 2001
Journal of Systems and Software (JSS)	Yes	
Requirements Engineering Journal	No	From 2004
Software Architecture	No	Only in 2004
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	

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. Method**

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, 2002. A list of journals is provided to the search engine within the Web of Science. This generates a complete list of articles published in the journals in 2002. 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 February 5<sup>th</sup>, 2008, 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 2001. However, they appear when searching for articles in 2002. 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 2001 although the ISI Web of Science lists it as being published in 2002) then there is a risk that some articles will not be covered at all in the analysis. Articles removed from 2002 will not appear when searching the database for 2001, 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 2002**

The method described in Section 3 resulted in a list of the most cited articles in 2002 for software engineering. The list contains 22 articles with three articles tied for position 20 on the list. Two authors are represented twice on the list, namely Barbara Kitchenham and Rajkumar Buyya. The complete list can be found in Appendix A. An excerpt of the list is presented in Table 2, where the nine 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.

**Table 2. Most cited software engineering articles in 2002 (top nine).**

Rank	Author	Citations
	Title of article	
	Journal reference	
1	Kitchenham BA, Pfleeger SL, Pickard LM, Jones PW, Hoaglin DC, El Emam K, Rosenberg J	64
	Preliminary guidelines for empirical research in software engineering	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (8): 721-734, AUG 2002	
2	Krauter K, Buyya R, Maheswaran M	53
	A taxonomy and survey of grid resource management systems for distributed computing	
	SOFTWARE-PRACTICE & EXPERIENCE 32 (2): 135-164 FEB 2002	
3	Mockus A, Fielding RT, Herbsleb JD	42
	Two case studies of open source software development: Apache and Mozilla	
	ACM TRANSACTIONS ON SOFTWARE ENGINEERING AND METHODOLOGY Volume: 11 (3): 309-346 JUL 2002	
4	Jackson D	31
	Alloy: A lightweight object modelling notation	
	ACM TRANSACTIONS ON SOFTWARE ENGINEERING AND METHODOLOGY Volume: 11 (2): 256-290 APR 2002	
5	Collberg CS, Thomborson C	26
	Watermarking, tamper-proofing, and obfuscation - Tools for software protection	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (8): 735-746 AUG 2002	
6	Baker M, Buyya R, Laforenza D	25
	Grids and Grid technologies for wide-area distributed computing	
	SOFTWARE-PRACTICE & EXPERIENCE Volume: 32 (15): 1437-1466 DEC 2002	
7	Briand LC, Melo WL, Wust J	22
	Assessing the applicability of fault-proneness models across object-oriented software projects	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (7): 706-720 JUL 2002	
8	Plasil F, Visnovsky S	19
	Behavior protocols for software components	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (11): 1056-1076 NOV 2002	
9	Glass RL, Vessey I, Ramesh V	18
	Research in software engineering: an analysis of the literature	
	INFORMATION AND SOFTWARE TECHNOLOGY Volume: 44 (8): 491-506 JUN 2002	

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 common 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, six 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 2002.

It is interesting also to study which journals appear on the list. It turns out that all 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 [Tse06]. This is slightly different from previous years [Wohlin07] and [Wohlin08], where some articles from other journals also made it into 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 [ScienceDirect], 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 [Wohlin07] for 2000 and in [Wohlin08] for 2001. Nine out of 22 articles on the list is published in IEEE Transactions in Software Engineering. Two journals have four articles each on the list, one journal has three articles on the list and then finally two journals have one article each on the list.

## **5. 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.

A top 20 list for year 2002 has been presented. Some authors of the most cited articles in 2002 were invited to contribute to a special section of Information and Software Technology. Six such articles can be found in this issue.

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 or section of the journal. This means that next year's analysis will focus on the most cited software engineering journal articles in 2003.

## 6. 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>).

## 7. References

[Glänzel02] W. Glänzel and H. F. Moed, “State of the Art Report: Journal Impact Measures in Bibliometric Research”, *Scientometrics*, Vol. 52, No. 2, pp. 171-193, 2002.

[ISI3] <http://isi3.isiknowledge.com> (last visit September 29, 2008)

[Jones05] 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, pp. 53-65, 2005.

[McBurney02] M. K. McBurney and P. L. Novak, “What Is BiblioMetrics and Why Should You Care?”, *Proceedings of the Professional Communication Conference*, pp. 108-114, 2002.

[ScienceDirect] <http://www.sciencedirect.com> (Last visit September 29, 2008)

[Thompson05] <http://www.thomson.com/scientific/scientific.jsp> (last visit September 29, 2008)

[Tse06] T. H. Tse, T. Y. Chen and R. L. Glass, “An Assessment of Systems and Software Engineering Scholars and Institutions (2000-2004)”, *Journal of Systems and Software*, Vol. 79, No.6, pp. 816-819, 2006.

[Wohlin07] C. Wohlin, “An Analysis of the Most Cited Articles in Software Engineering Journal – 2000”, *Information and Software Technology*, Vol. 49, No. 1, pp. 2-11, 2007.

[Wohlin08] C. Wohlin, “An Analysis of the Most Cited Articles in Software Engineering Journals – 2001”, *Information and Software Technology*, Vol. 50, No. 1-2, pp. 3-9, 2008..

[Zitt05] M. Zitt, S. Ramanan-Rahary and E. Bassecouard, “Relativity of Citation Performance and Excellence Measures: From Cross-Field to Cross-Scale Effects of Field-Normalisation”, *Scientometrics*, Vol. 63, No. 2, pp. 373-401, 2005.

## Appendix A: Top 20 cited software engineering articles in 2002

The list in Table A1 shows a ranking of the most cited software engineering articles in 2002. The citations count was made February 5, 2008.

**Table A1: Ranking of most cited articles.**

Rank	Author	Citations
	Title of article	
	Journal reference	
1	Kitchenham BA, Pfleeger SL, Pickard LM, Jones PW, Hoaglin DC, El Emam K, Rosenberg J	64
	Preliminary guidelines for empirical research in software engineering	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (8): 721-734, AUG 2002	
2	Krauter K, Buyya R, Maheswaran M	53
	A taxonomy and survey of grid resource management systems for distributed computing	
	SOFTWARE-PRACTICE & EXPERIENCE 32 (2): 135-164 FEB 2002	
3	Mockus A, Fielding RT, Herbsleb JD	42
	Two case studies of open source software development: Apache and Mozilla	
	ACM TRANSACTIONS ON SOFTWARE ENGINEERING AND METHODOLOGY Volume: 11 (3): 309-346 JUL 2002	
4	Jackson D	31
	Alloy: A lightweight object modelling notation	
	ACM TRANSACTIONS ON SOFTWARE ENGINEERING AND METHODOLOGY Volume: 11 (2): 256-290 APR 2002	
5	Collberg CS, Thomborson C	26
	Watermarking, tamper-proofing, and obfuscation - Tools for software protection	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (8): 735-746 AUG 2002	
6	Baker M, Buyya R, Laforenza D	25
	Grids and Grid technologies for wide-area distributed computing	
	SOFTWARE-PRACTICE & EXPERIENCE Volume: 32 (15): 1437-1466 DEC 2002	
7	Briand LC, Melo WL, Wust J	22
	Assessing the applicability of fault-proneness models across object-oriented software projects	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (7): 706-720 JUL 2002	
8	Plasil F, Visnovsky S	19
	Behavior protocols for software components	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (11): 1056-1076 NOV 2002	
9	Glass RL, Vessey I, Ramesh V	18
	Research in software engineering: an analysis of the literature	
	INFORMATION AND SOFTWARE TECHNOLOGY Volume: 44 (8): 491-506 JUN 2002	
10	Medvidovic N, Rosenblum DS, Redmiles DF, Robbins JE	17
	Modeling software architectures in the unified modeling language	
	ACM TRANSACTIONS ON SOFTWARE ENGINEERING AND METHODOLOGY Volume: 11 (1): 2-57 JAN 2002	

11	Kamiya T, Kusumoto S, Inoue K	16
	CCFinder: A multilinguistic token-based code clone detection system for large scale source code	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (7): 654-670 JUL 2002	
11	Smaragdakis Y, Batory D	16
	Mixin layers: An object-oriented implementation technique for refinements and collaboration-based designs	
	ACM TRANSACTIONS ON SOFTWARE ENGINEERING AND METHODOLOGY Volume: 11(2): 215-255 APR 2002	
11	Hall A, Chapman R	16
	Correctness by construction: Developing a commercial secure system	
	IEEE SOFTWARE Volume: 19 (1): 18-25 JAN-FEB 2002	
11	Rus I, Lindvall M	16
	Knowledge management in software engineering	
	IEEE SOFTWARE Volume: 19 (3): 26-38 MAY-JUN 2002	
15	Antoniol G, Canfora G, Casazza G, De Lucia A, Merlo E	15
	Recovering traceability links between code and documentation	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (10): 970-983 OCT 2002	
16	Kitchenham B, Pfleeger SL, McColl B, Eagan S	14
	An empirical study of maintenance and development estimation accuracy	
	JOURNAL OF SYSTEMS AND SOFTWARE Volume: 64 (1): 57-77 JAN 2002	
16	Kang KC, Lee J, Donohoe P	14
	Feature-oriented product line engineering	
	IEEE SOFTWARE Volume: 19 (4): 58-65 JUL-AUG 2002	
16	Tai KC, Lei Y	14
	A test generation strategy for pairwise testing	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (1): 109-111 JAN 2002	
16	Dobrica L, Niemela E	14
	A survey on software architecture analysis methods	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (7): 638-653 JUL 2002	
20	Deavours DD, Clark G, Courtney T, Daly D, Derisavi S, Doyle JM, Sanders WH, Webster PG	13
	The Mobius framework and its implementation	
	IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 28 (10): 956-969 OCT 2002	
20	Leroy X	13
	Bytecode verification on Java smart cards	
	SOFTWARE-PRACTICE & EXPERIENCE Volume: 32 (4): 319-340 APR 2002	
20	van der Linden F	13
	Software product families in Europe: The Esaps & Cafe projects	
	IEEE SOFTWARE Volume: 19 (4): 41-49 JUL-AUG 2002	