INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Experimental evaluation in computer science II: A quantitative study, 12 years later

Jacques Wainer   *Claudia G. Novoa Barsottini*
Danilo Lacerda
Leandro Rodrigues Magalhães de Marco

Technical Report - IC-09-02 - Relatório Técnico

January - 2009 - Janeiro

Experimental evaluation in computer science II: A quantitative study, 12 years later

Jacques Wainer Claudia G. Novoa Barsottini Danilo Lacerda
Leandro Rodrigues Magalhães de Marco

Abstract

This paper repeats part of the analysis performed in the 1995 paper "Experimental evaluation in computer science: A quantitative study" by Tichy and collaborators, for 147 papers randomly selected from the ACM, published in the year 2005. The papers published in 2005 are classified in the following way: 4% theory, 17% empirical, 4.7% hypothesis testing, 3.4% other, and 70% design and modeling (using the 1995 paper categories). Within the design and modeling class, 33% of the papers have no evaluation. The numbers of the 2005 sample are very similar to the original figures for the 1995 sample, which shows that Computer Science research has not increased significantly its empirical or experimental component.

1 Experimental evaluation in Computer Science

Tichy and collaborators [15] evaluated 400 articles published in 1993, 50 of them randomly selected papers published by ACM in 1993 and the rest systematically selected from a few journals in systems and software engineering, and classified the research reported in the paper in five categories (quoting [15] definitions):

- formal theory: articles whose main contributions are formally tractable propositions, e.g., lemmata and theorems and their proofs.
- design and modelling: systems, techniques, or models, whose claimed properties cannot be proven formally. Examples include software tools, performance prediction models, and complex hardware and software systems of all kinds. The papers in this class were further classified in the categories 0%, 0-10%, 10-20%, 20-50%, and +50%, according to the proportion of the paper that was dedicated to the evaluation of the new system, technique, or model.
- empirical work: articles that collect, analyze, and interpret observations about known designs, systems, or models, or about abstract theories or subjects (as this paper does). The emphasis is on evaluation, not on new designs or models.
- hypothesis testing: articles that define hypotheses and describe experiments to test them.

- others: articles that do not fit any of the four categories above, e.g., surveys.

Tichy and collaborators found that for the random sample 12% of the articles were in the theory class, 70% were design and model, 2% were empirical, 2% were hypothesis testing, and 14% were classified as others.

Tichy was particularly worried that from the random sample, 40% of the papers which propose a new system, model, framework, and so on, (the project and modeling class), lacked any evaluation (0% of space dedicated to the evaluation). That proportion was 50% for the papers in the software engineering journals. In comparison, the papers published in Optical Engineering (OE) and Neural Computation (NC), which were used as examples of engineering journals, only 15% and 12% of the papers that proposed new ideas lack the required evaluation.

According to Tichy:

> The low ratio of validated results appears to be a serious weakness in computer science research. This weakness should be rectified for the long-term health of the field.

The paper ends with a call for better standards in performing and publishing Computer Science (CS) research, with a greater emphasis on the evaluation of claims and designs, and a less emphasis on the creation of new systems, models, algorithms, frameworks, and so on.

This paper tries to evaluate if the Computer Science community has answered Tichy's calls. In particular, we repeated the evaluation of 147 randomly selected from all papers published by the ACM, including ACM journals, conferences, and workshop, and available in the ACM digital library, for the year 2005.

We understand that there are there are two main criticism to using the ACM published papers as representative of the whole CS research. The first one is that not all Computer Science subareas are well represented in ACM conferences and journals. Artificial Intelligence is one of such areas. The second main criticism is that most of the research published in by ACM are conference and workshop papers, which are usually short, and describe not fully developed ideas. One would expect journal papers would contain a more substantial evaluation of the first ideas presented at a conference or workshop.

We agree with both criticisms. And our answer to them is the same - indeed our sampling has bias, but there is no better alternative as a sample of CS research. The main problem in *selection bias* - any inclusion of a *particular* journal or set of journals, or non ACM conferences into the analysis can be criticised as distorting the conclusions toward a particular direction. For example, the inclusion of papers from the non ACM journal "Empirical software engineering," whose papers have probably a high content of empirical research, would only be meaningful if that journal is in some way "representative" of the whole of CS journals. We do not know *a priori* of any journal that is "representative" of CS research.

The only solution to avoid the selection bias is to evaluate a *random sample* from some "representative" sub-population of the whole set CS research. In the case of this paper, the sub-population was all the 2005 ACM published papers. As we discussed above, this is not a totally "representative" sub-population, but we do not believe that there is a better one.

Alternatives for the sub-population are: IEEE published papers (for the year 2005), or the set of papers published in journals indexed by the Web Of Science, or journals indexed by Scopus, or papers indexed by the DBLP.

Of course, we have no data to support our claim that among these possible sub-populations, the ACM published research is the "most representative." The Web of Science or Scopus indexed sub-populations are solely or mainly of journal and it seems clear that most of the CS production is published in conferences and workshops. The IEEE sub-population seems to us even more bias towards some of CS sub-areas than the ACM. Finally we do not know how to evaluate the bias of the DBLP indexing service.

If the reader does not agree with our intuitions that the ACM is the "most representative" sub-population of CS research, then this paper should not be read as an evaluation of the CS field as a whole, but only as an evaluation of the ACM published papers.

1.1 Related results

The analysis of how scientific research is performed in a particular domain by analysing aspects of the publications in that domain, is not common in Computer Science, as a whole, and more common for sub-areas of CS.

Glass and collaborators [7, 12] analyzed 628 Computer Science papers from 1995 to 1999 and classified them regarding different dimensions. Of interest to this paper is the "research approach" dimension, whose possible values are descriptive research, formulative research or evaluative research. The work [7] finds that 9.88% of the research is descriptive, 10.98% is evaluative and 79.15% is formulative.

If one considers that design and modeling class used in this paper, corresponds at least in part, with Glass's formulative approach to research, then one can conclude that from 70% to 80% of the published results in CS from 1993 to 1999 are mainly the proposal of a *new* system, model, algorithm, process, taxonomy, framework, and so on.

In specific domains, Prechelt [11] evaluates 190 articles on neural network published in 1993 and 1994. In the same line as Tichy [15], the author discover that only 8% of the papers present results on more than one realistic data set, and 1/3 of the papers have no quantitative comparison to previously known algorithms.

In CSCW, Pinelle and Gutwin [10] classify all papers published in the ACM CSCW conferences from 1994 to 1998 regarding the goals and methods of evaluation of the systems proposed. Wainer and Barsottini [16] classify all the papers published in the ACM CSCW conferences from 1998 to 2004 into categories similar to the ones used by Tichy [15], whether the paper proposes a new system (and how it is evaluated), whether it test a hypothesis, whether it describes an existing collaborative work/entertainment environment, or whether the paper has no empirical component (or in terms of [15], it is a design and project with 0% of evaluation).

Software engineering is an area in which bibliographic research seems more frequent. Dyba and collaborators [5] analyse the statistical power of the controlled experiments reported in 103 articles, published in the period 1993-2002, in nine journals and three major conferences in software engineering. Glass and collaborators [8] use the same dimensions of analysis used in [7, 12] to classify 369 journal papers in software engineering published in the period 1995-1999 randomly selected. Again, using the "research approach" dimension, they concluded that 55% of the papers follows the "formulative" approach. Other papers [9, 13, 2] discuss other aspects of the literature in Software Engineering.

2 Method

We follow to the best of our abilities exactly the methods used by Tichy [15]. The categories described in that work: formal theory, design and modeling, empirical work, hypothesis testing, and others, are called **major categories**. The divisions within the design and modeling were called **minor categories**.

The method followed in this paper, in a simplified way, was:

1. CGNB, DL and LRMM read all the 50 articles evaluated in [15] and discussed the classifications attributed in [15] in order to understand how to classify the papers in the major categories.
2. we randomly selected 200 articles published by ACM and available in the ACM portal[1]. The selection of the articles was performed in October of 2006.
3. from the 200 selected files, we removed the ones that were not in scientific articles. We obtained 147 scientific articles.
4. each of the 147 articles were attributed to two reviewers among CGNB, DL and LRMM.
5. each of CGNB, DL, and LRMM evaluated around 100 articles independently using [15] major and minor classifications.
6. articles with the same major classification by its two reviewers were considered closed.
7. articles with diverging major or minor classifications, or articles for which at least one of the reviewers was not sure of his/her own classification, were discussed by its two evaluators. If there was no agreement between them, JW would also evaluate it and either the most common class was attributed to the paper, or if there were three different classifications, JW's classification would be attributed to the paper.

The main difference from our methodology to Tichy's is that we decided on using only two reviewers for each paper, whereas in Tichy's paper, all four authors reviewed the random sample set, but the final results reported refer to the classification of a single reviewer. The other reviewers classification was used to measure the classification error (of the reported results due to a single reviewer). In our case a we used at least two evaluations to reduce the subjectivity of the classification. Second, we started with independent classifications, but if there was a disagreement between the reviewers, we allowed for a joint discussion in an attempt to reach a final classification. Again, the goal was to reach a classification with some control on the subjectivity.

[1]portal.acm.org

CGNB, DL, and LRMM are graduate students with different experiences with the CS literature. DL and LRMM are more experienced with image processing literature, and CGNB, with medical informatics. JW is a tenured faculty with 15 years of experience in CS areas such as Artificial Intelligence, Medical Informatics, and Computer Supported Collaborative Work.

2.1 Random selection

We decided on selecting a sample from the papers published in 2005 because, by October of 2006, when this research started, it was unclear how many of the papers published in 2006 were already available in the ACM digital library.

The articles were selected using the following procedure. From the ACM Portal page, we navigated to the advance search for ACM Digital Library (not the ACM Guide to the Computer Literature), and used the following search criteria: `Published since January 2005`,`Published before December 2005`, and `any type of publication`. The search, made in October 2006, returned 16,162 articles. [2] We wrote a program that collects the pdf links in all the pages of the search result. A second program randomly selected 200 entries from the list of links, and downloaded the corresponding pdf.

2.2 Number of articles selected

The number of articles selected was determined by the number of articles [15] used. Using the same query specified in section 2.1 for 1993, we discovered that 5,095 articles from 1993 are available from the ACM digital library. If we consider that this is the total number of articles published by ACM in 1993, then Tichy reviewed around 1% of the ACM papers of that year. We followed that same heuristics, and given that the query to the ACM digital library described above, returned 16,162 articles published in 2005, we decided to randomly select 200 articles.

After a first analysis, we removed 53 of them because they were not scientific articles, but publications like table of contents, abstracts, editorials, news, and so on. Although the final number of selected papers (147) is smaller than our goal of 1% of the total number of published papers in the period, we decided not to add any more papers.

2.3 Confidence intervals

The confidence intervals are all at a confidence level of 90%. We use the adjusted Wald confidence interval for proportions [1]. Given that from a population of n, x of them are "success", the adjusted Wald interval computes the proportion of successes as

$$p = \frac{x + z_\alpha^2/2}{n + z_\alpha^2}$$

[2] At the time of the writing of this article, the same search returns 16.188 articles.

Class	2005	
Theory	6	(4%)
Empirical	26	(17%)
Hypothesis	7	(4.7%)
Other	5	(3.4%)
Design total	103	(70%)
0%	34	(23%)
0-10%	10	(6.8%)
10-20%	22	(14%)
20-50%	31	(21%)
>50%	6	(4%)
Total	147	

Table 1: Totals for the classification of the 147 papers.

(instead of the standard proportion x/n) and compute the confidence interval using the standard Wald formula

$$CI = p \pm z_\alpha \sqrt{\frac{p(1-p)}{n + z_\alpha^2}}$$

where z_α is the z-critical value for the α confidence level.

3 Results

Table 1 is the main result of this paper. It shows the totals (and percentages) for articles classified into one of the major and minor classes.

Of the selected papers, 13 were journal papers, (papers numbered 5, 12, 15, 24, 28, 33, 45, 56, 83, 94, 95, 97, 113 in Appendix A), 2 were published on SIG bulletin, (papers 4 and 29), and the rest was conference papers.

Appendix A lists the 147 papers. The abbreviations used are: Proc. for Proceedings, Int. for International, Conf. for Conference, Symp. for Symposium, and Ws. for Workshop. Appendix B lists the final classification for each of the 147 papers.

Of the selected papers, 13 were journal papers (papers numbered 5, 12, 15, 24, 28, 33, 45, 56, 83, 94, 95, 97, 113 in Appendix A), 2 were published on SIG bulletin (papers 4 and 29), and the rest was conference papers.

3.1 Difficulties

The main problem with this form or research is the subjectivity of the classification. For 90 of the 147 papers, the two original reviewers, independently, arrived to the same classification.

Of the remainder 57 papers for which there was no independent agreement, 42 were resolved by the discussion of the two original evaluators. Among these 42, 15 (36%) were divergence between adjacent minor classifications, for example, one would classify the paper

	0-10%	10-20%	20-50%	>50%	Emp.	Hyp.	Other	Theory
Design 0%	1	12	1	0	7	5	5	0
Design 0-10%		6	3	0	1	0	0	1
Design 10-20%			11	0	2	1	2	0
Design 20-50%				8	2	0	0	0
Design >50%					1	0	0	0
Empirical						1	4	0
Hypothesis							2	0
Other								0

Table 2: Confusion matrix for the 57 papers with some divergence of classification

as Design and Modeling with 10% to 20% evaluation, and the other Design and Modeling, with 20% to 50% evaluation. Another 9 (21%) were both classified as Design and Modeling, but the minor classifications were not adjacent. Finally, of the 42, 18 (42%) were classified independently in different major categories.

Fifteen papers required the evaluation from JW. Of these, 4 were divergences on minor adjacent categories, 3 were divergence on minor non-adjacent categories, and 8 were divergence on major categories.

Table 2 is the confusion matrix for the 57 papers for which there was no independent agreement by the two original evaluators. The entry 12 in the line Design 0% and column D. 10-20% indicates there for 12 papers there was at least **one** classification as Design and Modeling with 0% of evaluation and **one** as Design and Modeling with 10-20%. Of the largest figures in table 2, the confusion between Design 0% and Design from 10 to 20% seems due to divergences on what evaluation is - both reviewers agreed that the paper proposes a new entity, but they disagree on whether there is or not evaluation of the entity in the paper. The second largest confusion is between Design 10-20% and Design 20-50% which seems to indicate differences in each reviewer's methods of calculating the area dedicated to evaluation. Finally, there are a large number of confusions are between Design 0% and the Empirical and Hypothesis classes. That confusion is very puzzling because Empirical and Hypothesis papers are mainly about evaluation, and a Design 0% has no evaluation at all.

4 Conclusions

This research classified 147 randomly selected papers published by ACM in 2005 in 5 classes proposed by Tichy and collaborator [15]. The classes describe different forms of research in Computer Science and in some way different epistemologies of CS.

The Theory class reflects a Mathematical view of CS, in which a research contribution is mainly in the form of proof of theorems. The Hypothesis Testing class reflects a Natural Sciences and Popperian view of CS, in which a hypothesis is clearly states and an experiment is performed that confirms or disprove the hypothesis. The Empirical and the Design and Modeling classes reflect an Engineering view of CS. The Design and Modeling class includes papers that propose a **new** system, algorithm, program, models, methodology, framework (which we will collectively refer as an entity), and optionally evaluates this new entity. The Empirical class include papers that evaluate entities proposed by others. This three epistemologies for CS - as a form of Mathematics, as a form of Natural Sciences and as a form of Engineering, has been discussed by other authors [3, 6, 4].

Under this view, this paper evaluated how CS is done, or what kind of science Computer Science is, by analysing what is published as CS research (by ACM in the year 2005).

Computer Science follows mainly an Engineering epistemology - most CS research is the proposal of **new** entities. Our research points that the Empirical and the Design and Modeling classes amount to 87% (17% + 70%) of the papers published by ACM in 2005. Theory papers are 4% of the total and Hypothesis Testing papers are 4.7% of the total.

But even if one assumes that CS mostly follows an Engineering epistemology, the amount of evaluation reported in the papers is low. Of the 129 papers in the "Engineering epistemology" class (Empirical + Design and Modeling), 34 (26%) have no evaluation. If we focus on the Design and Modeling category, a third of the papers that propose a new entity, do not evaluate them at all! If one arbitrarily determines that at least one fifth of the space of paper that proposes a new entity should be dedicated to evaluating it, then only 36% of the papers satisfy this mild requirement.

Tichy [15] evaluated all the papers published in 1993 in Optical Engineering, which he classifies as a Engineering journal, and in that sample only 15% of the Design and Modelling papers had no evaluation at all, and 67% of the Design and Modeling papers had at least one fifth of the space dedicated to evaluation. Thus the figures for CS published in 2005, regarding evaluation, are worse than they were for an Engineering journal in 1993.

One can argue that the low amount of evaluation is typical of Conference papers in CS, and that journal papers would probably have more rigorous standards. In this research we did not try to verify this hypothesis directly, but within our random sample of 147 papers, there is a sub-sample of 13 journal papers, distributed in the following categories: 1 paper in Other, 2 papers in Empirical, 1 in Design 0%, 2 in Design 0-10%, 4 in Design 10-20%, 2 in Design 20-50%, and 1 in Design >50%. Thus, in the sub-sample 10% of the Design and Modeling papers have no evaluation, and 30% have at least a fifth for evaluation. Unfortunately, since the size of the sub-sample is small, the confidence interval for these proportions are large. The difference between the 10% without evaluation published in journal and the global 34% is not statistically significant with 90% confidence.

Thus, if one assumes that the journal Optical Engineering is representative of the Engineering publications (in 1993), then although CS follows in most part an Engineering epistemology, it still falls behind in the rigor of empirical evidence it demands from its research, in comparison with other Engineering domains.

The comparison of the results from 1993 [15] and the results from 2005 shows that the distribution of papers among the different categories has changed very little. Table 3 lists the proportions of each classification for both years, with a 90% confidence interval. Thus, one can only state, with 90% confidence, that the number of empirical papers increased in the period. If one pays less attention to the confidence intervals, one may find a small trend towards more evaluation in Design and Modelling papers.

The increase of Empirical papers is interesting. It shows that there was an increase of

Class	1993	2005
Theory	6.1 - 21.8	2.0 - 7.8
Empirical	0 - 9.1	13.1 - 23.5
Hypothesis	0 - 9.1	2.4 - 8.7
Other	7.6 - 24.1	1.5 - 6.9
Design total	58.5 - 79.4	63.5 - 75.8
0%	20.5 - 41.5	17.9 - 29.3
0-10%	2.0 - 15.5	4.0 - 11.1
10-20%	6.1 - 21.8	10.7 - 20.5
20-50%	13.8 - 33.0	16.1 - 27.1
>50%	0 - 6.1	2.0 - 7.8

Table 3: 90% confidence interval (using the adjusted Wald method) for the proportions (in percentages) for each class, for 1993 and 2005.

reuse in Computer Science research. Empirical papers have an emphasis in evaluation an *existing* program, algorithm, model, framework, and so on. Thus, from 1993 to 2005 there was a small increase in research that evaluates an already existing entity.

If one agrees with Tichy's call for more evaluation in Computer Science research, in the conclusions of [15] and specially in [14] (as we do), then the comparison of the 1993 results with the 2005's is discouraging - very little has changed since 1993.

References

[1] A. Agresti and B. A. Coull. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52:119–126, 1998.

[2] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4):571–583, 2007.

[3] P. J. Denning. Is computer science science? *Commun. ACM*, 48(4):27–31, 2005.

[4] P. J. Denning. Computing is a natural science. *Commun. ACM*, 50(7):13–18, 2007.

[5] T. Dyba, V. B. Kampenes, and D. I. Sjoberg. A systematic review of statistical power in software engineering experiments. *Information and Software Technology*, 48(8):745–755, 2006.

[6] A. H. Eden. Three paradigms of computer science. *Minds Mach.*, 17(2):135–167, 2007.

[7] R. L. Glass, V. Ramesh, and I. Vessey. An analysis of research in computing disciplines. *Commun. ACM*, 47(6):89–94, 2004.

[8] R. L. Glass, I. Vessey, and V. Ramesh. Research in software engineering: an analysis of the literature. *Information and Software Technology*, 44(8):491–506, 2002.

[9] E. Mendes. A systematic review of web engineering research. In *2005 International Symposium on Empirical Software Engineering, ISESE 2005*, pages 498–507, 2005.

[10] D. Pinelle and C. Gutwin. A review of groupware evaluations. In *Proceedings of WET ICE 2000*, pages 86–91. IEEE Computer Society, 2002.

[11] L. Prechelt. A quantitative study of experimental evaluations of neural network learning algorithms: Current research practice. *Neural Networks*, 9(3):457–462, 1996.

[12] V. Ramesh, R. Glass, and I. Vessey. Research in computer science: An empirical study. *Journal of Systems and Software*, 70(1-2):165–176, 2004.

[13] D. Sjoeberg, J. Hannay, O. Hansen, V. Kampenes, A. Karahasanovic, N.-K. Liborg, and A. Rekdal. A survey of controlled experiments in software engineering. *IEEE Transactions on Software Engineering*, 31(9):733 – 753, 2005.

[14] W. F. Tichy. Should computer scientists experiment more? *Computer*, 31(5):32–40, 1998.

[15] W. F. Tichy, P. Lukowicz, L. Prechelt, and E. A. Heinz. Experimental evaluation in computer science: A quantitative study. *Journal of Systems and Software*, 28(1):9–18, 1995.

[16] J. Wainer and C. Barsottini. Empirical research in cscw - a review of the acm/cscw conferences from 1998 to 2004. *Journal of the Brazilian Computer Society*, 13:27–36, 2007.

A References for the 147 articles

Article number	Reference
1	Proc. 2005 Conf. on Designing for User eXperience, article No. 16
2	Proc. 2005 ACM SIGMETRICS Int. Conf. on Measurement and modeling of computer systems, pages 315-326
3	Proc. 1st Int. Ws. on open source data mining & frequent pattern mining implementations, pages 46-55
4	ACM SIGCSE Bulletin, volume 37, issue 4, pages 64-68
5	Mobile Networks and Applications, volume 10, issue 3, pages 355-370
6	Proc. 3rd Int. Ws. on Middleware for grid computing, pages 1-6
7	Proc. 2005 ACM SIGCHI Int. Conf. on Advances in computer entertainment technology, pages 176-179
8	Proc. 37th Conf. on Winter simulation, pages 2554-2560
9	Proc. 2005 Ws. on End-to-end, sense-and-respond systems, applications and services, pages 43-48
10	Special interest tracks and posters 14th Int. Conf. on World Wide Web, pages 946-947
11	Proc. 33rd annual ACM SIGUCCS Conf. on User services, pages 122-129
12	Ubiquity, volume 6, issue 14, page 1
13	Proc. second Australasian Conf. on Interactive entertainment, pages 11-18
14	Proc. 7th Int. Conf. on Electronic commerce, pages 41-44
15	Mobile Networks and Applications, volume 10, issue 1-2, pages 115-131
16	Proc. 14th ACM Int. Conf. on Information and knowledge management, pages 632-639
17	Proc. thirty-seventh annual ACM Symp. on Theory of computing, pages 49-56
18	Proc. 2005 IEEE/ACM Int. Conf. on Computer-aided design, pages 335-342
19	Proc. 3rd ACM Int. Ws. on Wireless mobile applications and services on WLAN hotspots, pages 33-38
20	Proc. 21st spring Conf. on Computer graphics, pages 137-145
21	Proc. 2005 Ws. on Software and compilers for embedded systems
22	Proc. tenth ACM SIGPLAN Int. Conf. on Functional programming, pages 204-215
	Continues on next page

– continued from previous page	
Number	**Reference**
23	Proc. 42nd annual Conf. on Design automation, pages 71-76
24	Wireless Networks, volume 11, issue 6, pages 719-728
25	Proc. 37th Conf. on Winter simulation, pages 336-345
26	Proc. 43rd annual Southeast regional Conf. - Volume 2, pages 300-305
27	Proc. 10th annual SIGCSE Conf. on Innovation and technology in computer science education, pages 291-295
28	Mobile Networks and Applications, volume 10, issue 6, pages 985-995
29	ACM SIGPLAN Notices, volume 40, issue 11, pages 18-28
30	Proc. 7th ACM SIGMM Int. Ws. on Multimedia information retrieval, pages 193-200
31	Proc. 2005 Conf. on Asia South Pacific design automation, pages 1264-1267
32	Proc. 3rd Int. Ws. on Traceability in emerging forms of software engineering, pages 8-13
33	ACM Transactions on Information Systems (TOIS), volume 23, issue 3, pages 267-298
34	CHI '05 extended abstracts on Human factors in computing systems, pages 1272-1275
35	Proc. first Int. Ws. on Advances in model-based testing, pages 1-7
36	Proc. 2005 ACM SIGCHI Int. Conf. on Advances in computer entertainment technology, pages 51-60
37	Proc. 31st Int. Conf. on Very large data bases, pages 613-624
38	Proc. 20th annual ACM SIGPLAN Conf. on Object oriented programming, systems, languages, and applications, pages 231-245
39	Proc. 2005 ACM Symp. on Applied computing, pages 868-873
40	Proc. twenty-first annual Symp. on Computational geometry, pages 370-371
41	Proc. 10th Int. Conf. on Intelligent user interfaces, pages 260-262
42	Proc. 2005 Int. Symp. on Low power electronics and design, pages 155-160
43	Proc. 2005 Int. Conf. on Augmented tele-existence, pages 172-179
44	Proc. 2005 ACM Symp. on Applied computing, pages 1250-1251
45	ACM SIGMOBILE Mobile Computing and Communications Review, pages 2-14
46	Proc. 15th ACM Great Lakes Symp. on VLSI, pages 244-247
47	Proc. 2005 Conf. on Genetic and evolutionary computation, pages 2083-2088
48	Proc. 14th Int. Conf. on World Wide Web, 463-470
	Continues on next page

– continued from previous page	
Number	**Reference**
49	Proc. IEEE Int. Conf. on Web Services (ICWS'05) - Volume 00, pages 759-766
50	Proc. 10th annual SIGCSE Conf. on Innovation and technology in computer science education, pages 133-137
51	Proc. 2005 joint Conf. on Smart objects and ambient intelligence & innovative context-aware services & usages and technologies, pages 283-286
52	Proc. 42nd annual Conf. on Design automation, pages 503-508
53	Proc. 31st Int. Conf. on Very large data bases, pages 970-981
54	Proc. 2005 Conf. on Graphics interface, pages 121-128
55	Papers presented at the 2005 Ws. on Wireless traffic measurements and modeling, pages 31-37
56	ACM SIGACCESS Accessibility and Computing, iIssue 83, pages 7-11
57	Proc. 2nd ACM Int. Ws. on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks, pages 168-174
58	Proc. tenth ACM Symp. on Access control models and technologies, pages 120-129
59	Proc. 6th Conf. on Information technology education, pages 1-5
60	Proc. fourth Int. joint Conf. on Autonomous agents and multiagent systems, pages 373-380
61	Proc. 2005 Conf. on Asia South Pacific design automation, pages 862-867
62	Proc. 2005 ACM SIGMOD Int. Conf. on Management of data, pages 718-729
63	Proc. 2005 joint Conf. on Smart objects and ambient intelligence & innovative context-aware services & usages and technologies, pages 213-218
64	Proc. 28th annual Int. ACM SIGIR Conf. on Research and development in information retrieval, pages 98-105
65	Proc. 11th Brazilian Symp. on Multimedia and the web, pages 1-3
66	Proc. eleventh ACM SIGKDD Int. Conf. on Knowledge discovery in data mining, pages 249-255
67	Proc. 22nd Int. Conf. on Machine learning, pages 73-80
68	Proc. 33rd annual ACM SIGUCCS Conf. on User services, pages 277-282
69	Proc. Conf. on Design, Automation and Test in Europe - Volume 2, pages 1238-1243
70	CHI '05 extended abstracts on Human factors in computing systems, pages 1509-1512
71	Proc. second Australasian Conf. on Interactive entertainment, pages 141-144
72	Proc. 2005 Int. Ws. on Mining software repositories, pages 1-5
	Continues on next page

– continued from previous page	
Number	**Reference**
73	Proc. 10th European software engineering Conf. held jointly with 13th ACM SIGSOFT Int. Symp. on Foundations of software engineering, pages 156-165
74	Proc. Conf. on Design, Automation and Test in Europe - Volume 2, pages 814-819
75	Proc. 7th Int. Conf. on Multimodal interfaces, pages 317-324
76	Proc. 2005 Int. Symp. on Wikis, pages 67-74
77	Proc. 19th Conf. computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction & citizens online & considerations for today and the future, pages 1-4
78	Proc. 4th Int. Symp. on Information and communication technologies, pages 236-242
79	Proc. 2005 ACM SIGCHI Int. Conf. on Advances in computer entertainment technology, article No. 4
80	Proc. 1st Int. Ws. on Utility-based data mining, pages 86-89
81	Proc. 10th European software engineering Conf. held jointly with 13th ACM SIGSOFT Int. Symp. on Foundations of software engineering, pages 1-10
82	Proc. 4th Int. Ws. on Task models and diagrams, pages 95-102
83	Mobile Networks and Applications, volume 10, issue 1-2, pages 233-249
84	Proc. 2005 IEEE/ACM Int. Conf. on Computer-aided design, pages 456-460
85	Proc. 2005 Conf. on Asia South Pacific design automation, pages 1047-1051
86	Proc. 2005 ACM Ws. on Storage security and survivability, pages 94-102
87	ACM SIGGRAPH 2005 Papers, pages 488-495
88	Proc. thirty-seventh annual ACM Symp. on Theory of computing, pages 331-337
89	Proc. 32nd ACM SIGPLAN-SIGACT Symp. on Principles of programming languages, pages 283-295
90	Proc. 14th ACM Int. Conf. on Information and knowledge management, pages 578-585
91	Proc. 42nd annual Conf. on Design automation, pages 652-657
92	Proc. 5th Int. Ws. on Software engineering and middleware, pages 39-45
93	Proc. 36th SIGCSE technical Symp. on Computer science education, pages 282-286
94	ACM Transactions on Programming Languages and Systems (TOPLAS), volume 27, issue 3, pages 441-476
95	ACM SIGOPS Operating Systems Review, volume 39, issue 1, pages 70-84
96	CHI '05 extended abstracts on Human factors in computing systems, pages 1037-1049
	Continues on next page

– continued from previous page

Number	Reference
97	ACM Transactions on Graphics (TOG), volume 24, issue 2, pages 182-208
98	Proc. 7th Int. Conf. on Electronic commerce, pages 173-176
99	Proc. 1st ACM Ws. on Dynamic interconnection of networks, pages 28-32
100	CHI '05 extended abstracts on Human factors in computing systems, pages 1909-1912
101	Proc. 7th Ws. on Multimedia and security, pages 51-56
102	Proc. 6th ACM SIGCHI New Zealand chapter's Int. Conf. on Computer-human interaction & making CHI natural, pages 67-74
103	Proc. 10th Conf. on Theoretical aspects of rationality and knowledge, pages 5-21
104	Proc. 43rd annual Southeast regional Conf. - Volume 1, pages 162-167
105	Proc. 13th annual ACM Int. Conf. on Multimedia, pages 735-744
106	Proc. 14th ACM Int. Conf. on Information and knowledge management, pages 800-807
107	Proc. 2005 Conf. on Asia South Pacific design automation, pages 856-861
108	Proc. twentieth ACM Symp. on Operating systems principles, pages 31-44
109	Proc. 2005 ACM Conf. on Emerging network experiment and technology, pages 286-287
110	Proc. 13th annual ACM Int. Conf. on Multimedia, pages 872-881
111	Proc. 2005 ACM Symp. on Applied computing, pages 1328-1332
112	Proc. 2005 Conf. on Genetic and evolutionary computation, pages 1493-1500
113	ACM Transactions on Design Automation of Electronic Systems (TODAES), volume 10, issue 4, pages 673-689
114	Proc. 2005 ACM SIGMIS CPR Conf. on Computer personnel research, pages 128-138
115	Proc. 2005 ACM/IEEE Conf. on Supercomputing, page 70
116	Proc. 22nd Int. Conf. on Machine leaning, pages 625-632
117	Proc. twenty-first annual Symp. on Computational geometry, pages 190-197
118	Proc. 7th Int. Conf. on Electronic commerce, pages 839-841
119	Proc. 37th Conf. on Winter simulation, pages 2280-2289
120	Proc. 4th Int. Symp. on Information processing in sensor networks, article No. 3
121	Proc. 15th ACM Great Lakes Symp. on VLSI, pages 96-101
122	Proc. 2005 annual Conf. on European association of cognitive ergonomics, pages 123-130

Continues on next page

– continued from previous page	
Number	**Reference**
123	Proc. tenth ACM SIGPLAN Int. Conf. on Functional programming, pages 92-104
124	Proc. 2005 Conf. on Genetic and evolutionary computation, pages 687-694
125	Proc. 5th ACM/IEEE-CS joint Conf. on Digital libraries, pages 121-130
126	Proc. 15th ACM Great Lakes Symp. on VLSI, pages 426-430
127	ACM SIGGROUP Bulletin, volume 25, issue 1, pages 12-19
128	Proc. 10th annual SIGCSE Conf. on Innovation and technology in computer science education, pages 138-142
129	Proc. 3rd IEEE/ACM/IFIP Int. Conf. on Hardware/software codesign and system synthesis, pages 160-165
130	Proc. of th 2005 Conf. on Computer support for collaborative learning & learning 2005 & the next 10 years!, pages 602-606
131	Proc. 2005 ACM SIGCOMM Ws. on Economics of peer-to-peer systems, pages 116-121
132	CHI '05 extended abstracts on Human factors in computing systems , pages 2005-2008
133	Proc. 1st Int. Ws. on open source data mining & frequent pattern mining implementations, pages 26-35
134	Proc. 37th Conf. on Winter simulation, pages 2581-2584
135	Proc. 2005 Conf. on Applications, technologies, architectures, and protocols for computer communications, pages 49-60
136	Proc. 15th ACM Great Lakes Symp. on VLSI, pages 308-311
137	Proc. 10th Int. Conf. on Artificial intelligence and law, pages 146-154
138	Proc. 14th Int. Conf. on World Wide Web, pages 453-462
139	Proc. 2005 Conf. on Interaction design and children, pages 40-47
140	ACM SIGGRAPH 2005 Papers, pages 1208-1215
141	Proc. 37th Conf. on Winter simulation, pages 2068-2077
142	Proc. 2005 Conf. on Graphics interface, pages 33-40
143	Proc. 2005 Conf. on Interaction design and children, pages 56-61
144	Proc. 2nd ACM Int. Ws. on Vehicular ad hoc networks, pages 10-19
145	Proc. 10th European software engineering Conf. held jointly with 13th ACM SIGSOFT Int. Symp. on Foundations of software engineering, pages 306-315
146	Proc. 31st Int. Conf. on Very large data bases pages 1055-1066
147	Proc. Conf. on Design, Automation and Test in Europe - Volume 3, pages 153-158

B Classification of the 147 articles

Class	Articles
Theory	17 85 88 89 109 117
Empirical	2 7 8 16 35 55 56 64 76 83 92 93 99 111 112 116 120 122 129 130 131 133 134 135 139 141
Hypothesis	3 34 39 54 75 100 103
Other	12 14 68 70 114
Design 0%	1 9 10 13 19 22 25 26 37 38 40 44 46 57 58 59 60 63 71 72 80 82 86 95 96 98 102 104 118 119 125 127 137 143
Design 0-10%	4 6 11 32 45 49 65 94 128 138
Design 10-20%	5 18 21 27 31 33 43 53 69 73 90 97 105 107 113 121 123 126 136 140 145 146
Design 20-50%	15 20 23 28 30 36 41 42 47 48 50 51 52 61 62 66 67 74 77 79 84 87 91 101 106 108 110 115 124 132 147
Design >50%	24 29 78 81 142 144