Worlds Apart: Industrial and Academic Focus Areas in Software Testing

Article *in* IEEE Software · October 2017

DOI: 10.1109/MS.2017.265100233

2 authors:

Vahid Garousi
Queen's University Belfast
199 PUBLICATIONS **5,162** CITATIONS

SEE PROFILE

Michael Felderer
German Aerospace Center (DLR)
331 PUBLICATIONS **4,648** CITATIONS

SEE PROFILE

Worlds apart:

Industrial and academic focus areas in software testing

Vahid Garousi
Hacettepe University

Michael Felderer
University of Innsbruck, and QE LaB Business Services

MOTIVATION: PRACTITIONERS AND RESEARCHERS ARE FAR FROM EACH OTHER

The global software industry and the software engineering (SE) academia are two large communities. However, unfortunately, the level of joint industry-academia collaborations (IAC) in SE is very low, compared to the amount of activities in each of the two communities [1]. This is especially the case for software testing, which is a hot topic in SE research, reflected by the large amount of testing-related papers at recent ICSE conferences, and practice, reflected by the large amount and popularity of industrial testing conferences. The lack of mutual perception between industry and academia hurts both sides: researchers have fewer insights on the problems that are important to practitioners, while practitioners fail to learn what researchers have already discovered that might be useful to them.

The issue of IAC in SE has been an important topic since the early years of SE (in the 1960's). In an applied field such as SE, industrial impact and relevance [2] are of outmost importance. For example, there are projects such as the ACM SIGSOFT Impact project (www.sigsoft.org/impact) which have measured and analyzed the impact of SE research on practice.

But unfortunately, honestly speaking, many SE practitioners and researchers and especially those in software testing are not talking to each other (adequately). Various reasons have been discussed by researchers and practitioners for such a lack of motivation for collaborations between industry and academia [1], such as: each side having different objectives, industrial problems lacking scientific novelty or challenges, and the low applicability and scalability problems of the solutions developed in academia [3]. Also, a recent survey of 3,000 employees in Microsoft showed that unfortunately many practitioners do not find many of the top-cited SE research papers relevant or useful to their everyday challenges [2]. For the SE research community to have a meaningful future, there is a critical need to better connect industry and academia. There are even indications that the gap between industry and academia is even bigger in software testing compared to other areas of SE [4].

Many of the (academic) SE conferences include in their 'call for papers' (CFPs) phrases such as: "*this conference will bring together practitioners and researchers together working in the area of [a SE sub-area]*", but honestly, many conferences fail to really achieve that. There has been some success in certain conferences, e.g., the industry tracks of the IEEE/ACM ICSE and ICST (International Conferences on Software Engineering and International Conference on Testing, Verification and Validation, respectively), but still a lot more should be done to 'really' bring practitioners and researchers together.

To address the above challenges, the authors and their colleague have recently embarked on a number of studies, e.g., a systematic review of challenges and best practices in IACs in SE [1], how to select the right topics for IACs in software testing [5], and sharing their experience and success stories in IACs [6]. The two authors have worked as software engineers, academic researchers and also consultants for more than 30 years combined (in three countries: Canada, Turkey and Austria), and have given numerous talks on research and industrial conferences (26/18 and 30/15 research/industrial conference talks by the two authors, respectively). Through their experience, the authors have observed that practitioners and researchers often focus on different issues and this is a major reason for weakness of the connection between industry and academia in testing. Thus, to address the need for more IACs in SE, the authors focus on software testing as a representative area of SE, and aim at comparing the focus areas of industry and academic in software testing, as represented by the titles of the talks from a set of selected conferences in each of the two communities (industry and academia).

The lack of collaborations and mutual perception between industry and academia hurts both sides.

Such a comparison will help us shed light on the root-cause of the problem (low IACs) and to answer the question how we can improve things in this regard.

Several existing papers are slightly related to our work, e.g., [4, 7]. A 2004 position paper [4] entitled "*The (im)maturity level of software testing*" pointed out the large gap between the state-of-the-art in software testing literature, and the state of software testing practice and pointed out the need for further industrial empirical research in software testing. An

interesting recent 2016 work [7] analyzed the synergies of academic and industrial software testing conferences, for example the Program Committee (PC) composition of those venues. It also classified and compared the conferences based on where they can be placed on a purely academic-joint-purely industrial scale. Our goal in this paper is quite different while complementary to those studies as we analyze the focus areas of the testing researchers and practitioners in their respective conferences by looking at the titles of talks given at conferences.

ANALYSIS GOAL AND APPROACH

The goal of this study is to conduct an exploratory assessment between the focus areas of industry versus academia in software testing for the purpose of characterizing the potential root-causes for relatively low IACs in this area, from the point of view of practitioners and researchers, with the aim of helping to increase the level of collaborations.

Figure 1 depicts the analysis method that we used in this study. From the list of all available venues and conferences in software testing, to manage our data extraction effort, we sampled three well-known dedicated leading industrial conferences (GTAC, EuroSTAR and STAREast) and two leading academic research conferences (ICST and ISSTA), the acronyms of which are shown in Figure 1. The conferences were carefully selected based on the long-standing academic and industrial experience of the authors in both the industrial and research communities, based on their representativeness and popularities in the communities. The "STAR" family of conferences are among the most popular industrial software testing conferences world-wide and held annually in the US (STAREast and STARWest), in Canada (STARCanada) and also in Europe (EuroSTAR) since 1980's and 1990's. GTAC is Google's flagship test conference and is held annually since 2006. The ICST is supported by the IEEE and held since 2008. The ISSTA is supported by the ACM (Association for Computing Machinery) and held since 1975 (first as a workshop with a different name: http://historywiki.acm.org/sigs/SIGSOFT-ISSTA).

We collected all the talk titles from the main conferences of the selected venues held in 2013 and 2014 (excluding their "satellite" events such as workshops). The extraction of talk titles resulted in 354 industrial talks and 340 academic talks. For transparency and replication, the entire data set is available at https://goo.gl/zK6KYw.

We report next two types of analyses using the data to address the study goal: (1) word-cloud visualization to see the focus areas in a big picture, and (2) qualitative analysis of a subset of titles.

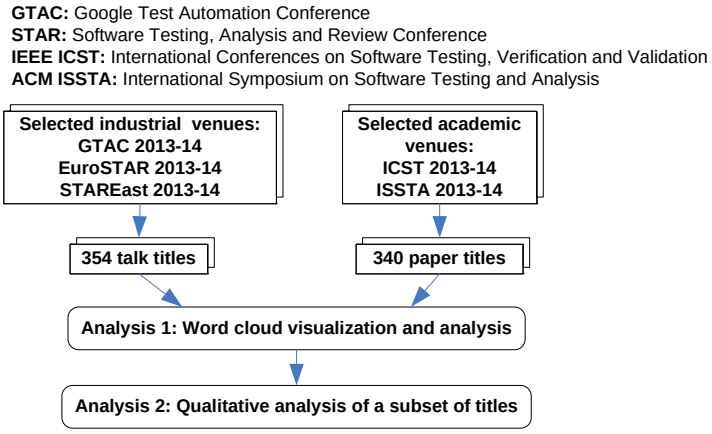

Figure 1- Research method

WE ARE TALKING ABOUT QUITE DIFFERENT THINGS

Figure 2 depict two word clouds that show the focus areas of industry and academia in software testing based on the talk titles of the selected conferences. An online tool named Wordle (www.wordle.net) was used to generate these word clouds. Beside the word clouds themselves, Figure 2 also lists the ten most common phrases. The visualization and top phrases both indicate a slight mismatch in focus areas. The top three phrases for the industry are: *(test) automation*, *mobile*, and *agile testing*, while for the academia they are: *model (-based testing)*, *combinatorial (testing)*, and *automated (testing)*.

At a first glance, we can say that both communities focus on test automation (automated testing). However, a closer look into the actual list of talk titles shows that when practitioners talk about test automation, they mostly refer to automating the test execution phase, whereas academics present talks on 'automated' approaches in support of testing (for instance, test-case design). To show this phenomenon with actual example talk titles, Table 1 shows three example titles from each pool showing the disparate focus w.r.t. automated testing. Whereas industry focuses on automated test execution

('implementing automation', 'virtualization') as well as its management ('outsourcing') and application in specific domains ('mobile devices', 'financial institutions'), academia has focused in these cases on automated test generation ('automated search for probabilistic test profiles'), test evaluation ('automated generation of oracles') and debugging ('automated program repair') and by taking optimization techniques (search-based approaches) into account. In summary, we can say that the topics of interest to testing practitioners (based on the given dataset) seem to not interest testing researchers and vice versa!

Industrial conferences

Word	Freq.
automation	34
mobile	29
agile	24
quality	18
cloud	13
performance	12
world	11
management	11
data	10
continuous	9

Research conferences

Word	Freq.
model	18
combinatorial	13
automated	10
web	8
analysis	8
applications	7
empirical	6
product	6
mutation	6
data	6

Figure 2- Focus areas of the industrial and research conferences and the most common phrases

Table 1-Three example talk/paper titles from each pool showing the different focus w.r.t. automated testing

Industry	Academia
• The challenges of big testing: automation, virtualization, outsourcing, and more • The importance of automated testing on real and virtual mobile devices • Designing and implementing automation at a large financial institution	• Using automated program repair for evaluating the effectiveness of fault localization techniques • Automated generation of oracles for testing user-interaction features of mobile apps • Adding contextual guidance to the automated search for probabilistic test profiles

Regarding the ten most common phrases, industrialists discuss testing in its relationship to software quality quite frequently. Also, management issues and testing in the context of agile development are hot topics. Mobile testing, cloud testing and performance are also popular issues. Finally, testing in context of continuous integration and delivery is frequently discussed in talks at industrial conferences.

On the other hand, researchers in academia feel excited by 'theoretically challenging' issues, e.g., combinatorial testing and search-based test-case design, while industrialists just want to find ways to improve effectiveness and efficiency of testing (e.g., better test automation) without the need for 'fancy' methods or techniques which they mostly find to be too complicated and hard to implement and deploy in practice [1]. Model-based testing is quite popular in academia and seems to have a limited usage in some industry sectors, e.g., for testing automotive and safety-critical software, but according to our data, it has no widespread penetration in all industry sectors. Also mutation testing is widely discussed in research but has very low industrial penetration. To sum up, the topics of interest in industry and academia are quite different. Thus we can state that the two groups "live in two different worlds", since their areas of interest and focus, as the talk titles show, are quite different and we have seen that this is one main reason for low interaction among the two communities.

Based on a systematic review of challenges and best practices in IACs in SE [1] and also the authors' discussion with their past/current industry partners, many practitioners complain about low applicability and scalability of testing techniques proposed in most research papers. Furthermore, a lack of cost-benefit analysis in most research papers, i.e., how much effort/time has to be spent to adopt/implement a testing technique in a project and how much real cost savings will be yielded as a result, is highlighted as a major limitation of research ideas by practitioners. The authors have had some initial success in conducting cost-benefit analysis of testing techniques in their joint IACs with partners, e.g., [8-10].

Another differentiating aspect that we have been hearing is the difference in writing formality by academics versus practitioners. Unfortunately, many practitioners believe that academic papers are too formal and hard to understand [2, 11]. On the other hand, practitioners usually strive to write things in the simplest form possible. Although discussing the root causes of this phenomenon is beyond the scope of this paper, we conducted an automated analysis on the readability of talk titles using a free online service (www.eadability-score.com). Since we did not have the transcript of all the industry talks, we could not conduct this analysis on the presentations' full texts. We expected to see a difference in readability of talk titles among the two groups. Figure 3 shows the various readability metrics which are well-known in the literature and, as we can see, this automated analysis confirms our expectation that academics talk titles are indeed harder to read than industrial ones. Note that, unlike the other metrics in this set, the 'Flesch-Kincaid Reading Ease' metric is a reverse metric in that, the higher its value, the easier it is to read the given text.

What interests testing practitioners does not seem to interest researchers and vice versa!

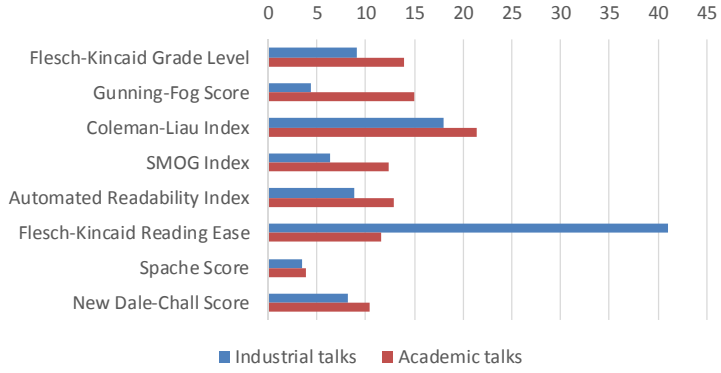

Figure 3- Various readability metrics of talk titles in the two groups

QUALITATIVE ANALYSIS OF A SUBSET OF TITLES

As discussed above, our data extraction yielded a set of 354 industrial talks and 340 academic talks. We review next a subset of those to get a sense of the topics each group is interested in. For this purpose, we sorted each set by the length of the talk title, as shown in Table 2 and Table 3. Each table lists 15 sampled titles from the industrial and research conference talks, respectively (those with shortest and longest titles).

As one can see from the tables, the talks at the industrial venues are mostly hands-on tutorial-like presentations presented to provide mini-training for the conference attendees on certain topics, e.g., talks #4, 6 in Table 2 (on Android testing and dynamic testing tools). Many industry talks fall into the category of best practices and lessons learned, e.g., talks #2, 3. Some talks are about team building and human factors in this context, e.g., talks #7, 8, 353. We also notice talks on the relationship between testing and process issues, e.g., talks #3, 7, 8, 9. Also talks on testing specifically in current domains of interest such as mobile or robotics are common in industrial conferences, e.g., talks #4, 350, 352.

On the other hand, papers and presentations at the academic testing conferences are more theory-focused (e.g., #337 in Table 3 on 'Mutant subsumption graphs' or #6 on test-case generation using 'constraint programming') as opposed to being hands-on and practical as found in industry events. There are many systematic empirical studies, e.g., talks #1, 5, 7, 10, which sometimes even present an evaluation in an industrial context, e.g., talk #5. Also search-based software testing approaches are popular at industrial conferences, e.g., talks #5, 9. Furthermore, model-based test generation (e.g., using state machines), e.g., talks #4, 7, is a prominent topic. Also security testing, e.g., talks #1, 3, 4, and dynamic analysis techniques are also presented, e.g., talks #2, 336.

Table 2- 15 sampled talk titles from the industrial conferences (shortest and longest ones)

#	Talk title
1	Free tests are better than free bananas: using data mining and machine learning to automate real-time production monitoring

2	Mobile quality assurance: what functional and non-functional testers need to know about advanced best practices
3	From request to delivery - using agile methods to trace customer requirements and improve quality
4	Espresso, spoon, Wiremock, oh my! (or how I learned to stop worrying and love Android testing)
5	Next gen automation: abstracted language and tool agnostic approaches to reduce quality costs
6	AddressSanitizer, ThreadSanitizer and MemorySanitizer: dynamic testing tools for C++
7	Questioning auditors questioning testing, or how to win friends and influence auditors
8	How we transformed the traditional software QA by getting rid of the central QA group
9	The art of testing transformation: blending technology with cutting-edge processes
10	The challenges of BIG testing: automation, virtualization, outsourcing, and more
…	
350	Robotic testing
351	Visual testing
352	Mobile testing
353	Team building
354	Testing me

Table 3- 15 sampled talk titles from the research conferences (shortest and longest ones)

#	Talk title
1	Empirical investigation of the web browser attack surface under cross-site scripting: An urgent need for systematic security regression testing
2	Make it work, make it right, make it fast: building a platform-neutral whole-system dynamic binary analysis platform
3	Test generation and evaluation from high-level properties for common criteria evaluations - the TASCCC testing tool
4	Generic approach for security error detection based on learned system behavior models for automated security tests
5	A search-based approach for cost-effective software test automation: decision support and an industrial case study
6	Test generation for robotized paint systems using constraint programming in a continuous integration environment
7	Assessing quality and effort of applying aspect state machines for robustness testing: a controlled experiment
8	On an embedded software design architecture for improving the testability of in-vehicle multimedia software
9	Search-based testing of relational schema integrity constraints across multiple database management systems
10	Threats to the validity and value of empirical assessments of the accuracy of coverage-based fault locators
…	
336	Collecting a heap of shapes
337	Mutant subsumption graphs
338	Reconstructing core dumps
339	Crowdsourcing GUI tests
340	Declarative mocking

Furthermore, we noticed that industrial talks often use humorous titles, most probably to attract audience [12] and set informal tones in the talks, e.g., "*Free tests are better than free bananas: using data mining and machine learning to automate real-time production*" and, " "*Espresso, Spoon, Wiremock, Oh my! (or how I learned to stop worrying and love Android testing)*".

In summary, we see from the analysis of titles that the topics present at industrial and research conferences are quite different. Industrial conferences generally place a strong emphasis on tutorial-like talks, management, organizational and process aspects, whereas research conferences focus on techniques, e.g., for search-based testing, model-based testing or dynamic analysis, but also on systematic empirical evaluation. A possible limitation of our analysis of titles is that the consideration of whitepapers and other types of grey literature could lead to refined industrial and academic topic lists.

WHAT WE CAN DO TO IMPROVE THINGS

What we see and have personally experienced is that, having different disparate focus areas (as we showed in our short analysis) is one major reason which prevents wider IACs in software testing and other areas of software engineering. When different topics interests practitioners and researchers, it is less likely that they would collaborate.

Effort is needed from both sides for them to come 'closer' to each other (in terms of focus areas), to increase the level of joint collaboration and to ensure win-win situations for both sides. Many generic best practices to improve IACs in SE have been presented in the past, e.g., in [1], which should be carefully studied and utilized. For instance, researchers should focus on industry problems rather than abstract/artificial challenges, e.g. as we did in our recent efforts [5, 6], and industrialists should value the research results and be open for IAC. This can be supported by studies performing empirical evaluation in industrial contexts, e.g., via surveys or case studies, thus providing a link between industrial and academic topics in software testing. On one hand, these studies integrate the industrial point of view into research studies. On the other hand, empirical studies in an industrial context can also be presented as best practices and lessons learned at industrial

conferences, which are especially welcome at these venues. In return, industrialists should raise sophisticated industrial challenges where solutions from research can be expected in discussions with researchers or on special tracks of research conferences or workshops.

Researchers are advised to use the principles of Action Research (AR) [13] in their research, especially when collaborating with industry, to ensure that the research problems are chosen from the actual needs of the industry. We proposed a grounded-theory-based approach for doing so in a recent work [5] which was based on our past experience in 15+ IAC projects in testing [6]. The approach was applied in an ongoing IAC between the first author and a major Turkish defense software company to derive the topics for several mutually-attractive joint projects, as listed below:

- Need for more test automation for several test groups
- Assessing and improving an in-house test automation framework for test group Q
- Need to establish a systematic, effective and efficient GQM-based measurement program for the testing department
- Need for assessment and improvement of test process maturity using TMMI and TPI-Next
- Need for bi-directional knowledge transfer from/to international venues and organizations in the aviation industry

The research-intensive SE conferences (such as ICSE and ICST) have to undertake more measures to become more 'interesting' from the industry's perspective and to attract more practitioners. Implementing such measures may not be easy, but to ensure meaningful collaborations, many steps have to be taken by both sides. Lately, a few promising developments in this direction have been taken on a number of research conferences, e.g., the *Software Engineering in Practice (SEIP)* track of the ICSE, the industry track of the ICST, and a special workshop dedicated to IAC in software testing (the Workshop on Testing: Academia-Industry Collaboration, Practice and Research Techniques, TAIC PART). On industrial testing conferences, one good example in the right direction to increase IAC is the annual 'Software Quality Days' event in Austria which hosts a scientific track with Springer proceedings. This track even has a best industrial-experience paper award and many accepted papers in this track have a major case-study component, e.g., [14, 15].

CONCLUSIONS AND ROAD AHEAD

While the conferences that we have selected are among the most popular in the two categories (industrial and academic), looking at other conferences in the two categories would probably yield similar results.

Analysis of talk titles in industrial versus research conferences showed that practitioners and researchers are focusing on quite different things in testing. Having different focus areas is a major reason which leads to (very) low relative IAC in software testing, and also other areas of software engineering, in almost every country (the authors have had first-hand experience in Canada, Turkey and Austria). Effort is needed from both sides to increase the level of joint collaborations and to ensure win-win situations for both sides. Researchers should also be aware of the challenges of their industrial partners, and choose problems and topics that are novel, feasible, industrially relevant, and potentially impactful [16]. By studies such as this one, the authors are continuing their efforts to bring practitioners and researchers in software engineering, in general, and software testing in particular, closer to each other so that they can mutually benefit each other much more than they actually do today.

Researchers should focus on industry problems rather than abstract/artificial challenges, and industrialists should value the research results

REFERENCES

[1] V. Garousi, K. Petersen, and B. Özkan, "Challenges and best practices in industry-academia collaborations in software engineering: a systematic literature review," *Information and Software Technology*, vol. 79, pp. 106–127, 2016.

[2] D. Lo, N. Nagappan, and T. Zimmermann, "How practitioners perceive the relevance of software engineering research," presented at the Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering, Bergamo, Italy, 2015.

[3] L. Briand, "Embracing the Engineering Side of Software Engineering," *IEEE Software*, vol. 29, pp. 96-96, 2012.

[4] A. Bertolino, "The (Im)maturity level of software testing," *SIGSOFT Softw. Eng. Notes*, vol. 29, pp. 1-4, 2004.

[5] V. Garousi and K. Herkiloğlu, "Selecting the right topics for industry-academia collaborations in software testing: an experience report," in *IEEE International Conference on Software Testing, Verification, and Validation*, 2016, pp. 213-222.

[6] V. Garousi, M. M. Eskandar, and K. Herkiloğlu, "Industry-academia collaborations in software testing: experience and success stories from Canada and Turkey," *Software Quality Journal, special issue on Industry Academia Collaborations in Software Testing*, pp. 1-53, 2016.

[7] Á. Beszédes and L. Vidács, "Academic and Industrial Software Testing Conferences: Survey and Synergies," in *Proceedings of the International Conference on Testing: Academic and Industrial Conference - Practice and Research Techniques (TAIC PART)*, 2016.

[8] Y. Amannejad, V. Garousi, R. Irving, and Z. Sahaf, "A Search-based Approach for Cost-Effective Software Test Automation Decision Support and an Industrial Case Study," in *Proc. of International Workshop on Regression Testing, co-located with the Sixth IEEE International Conference on Software Testing, Verification, and Validation*, 2014, pp. 302-311.

[9] S. A. Jolly, V. Garousi, and M. M. Eskandar, "Automated Unit Testing of a SCADA Control Software: An Industrial Case Study based on Action Research," in *IEEE International Conference on Software Testing, Verification and Validation (ICST)*, 2012, pp. 400-409.

[10] S. Mohacsi, M. Felderer, and A. Beer, "Estimating the Cost and Benefit of Model-Based Testing: A Decision Support Procedure for the Application of Model-Based Testing in Industry," in *2015 41st Euromicro Conference on Software Engineering and Advanced Applications*, 2015, pp. 382-389.

[11] Multiple online authors, "Why are academic articles/journals/papers so hard to understand?," *http://english.stackexchange.com/questions/132564/why-are-academic-articles-journals-papers-so-hard-to-understand*, Last accessed: Sept. 2016.

[12] Unknown freelance authors, "Titles that Talk: How to Create a Title for Your Article or Manuscript," *https://www.freelancewriting.com/creative-writing/titles-that-talk/*, Last accessed: Sept. 2016.

[13] P. S. M. d. Santos and G. H. Travassos, "Action-research use in software engineering: An initial survey," in *Proceedings of the International Symposium on Empirical Software Engineering and Measurement*, 2009, pp. 414-417.

[14] A. N. Ghazi, K. Petersen, and J. Börstler, "Heterogeneous Systems Testing Techniques: An Exploratory Survey," in *Proceedings of International Conference on Software Quality Days*, D. Winkler, S. Biffl, and J. Bergsmann, Eds., ed: Springer, 2015, pp. 67-85.

[15] M. Felderer and A. Beer, "Using Defect Taxonomies to Improve the Maturity of the System Test Process: Results from an Industrial Case Study," in *Software Quality. Increasing Value in Software and Systems Development: 5th International Conference, SWQD 2013, Vienna, Austria, January 15-17, 2013. Proceedings*, D. Winkler, S. Biffl, and J. Bergsmann, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 125-146.

[16] A. Begel and T. Zimmermann, "Analyze this! 145 questions for data scientists in software engineering," presented at the Proceedings of the 36th International Conference on Software Engineering, Hyderabad, India, 2014.

AUTHOR BIOGRAPHIES

Vahid Garousi is an Associate Professor of Software Engineering in Hacettepe University in Ankara, Turkey and the managing director of Maral Software Consulting Corporation, based in Calgary, Canada. Previously, he worked as an Associate Professor at the University of Calgary, Canada from 2006 until 2014. His research interests include software testing, empirical software engineering, and improving industry-academia collaborations in software engineering. He was selected as a Distinguished Visitor (speaker) for the IEEE Computer Society's Distinguished Visitors Program (DVP) for the period of 2012-2015. Contact him at vahid.garousi@hacettepe.edu.tr.

Michael Felderer is a senior researcher at the University of Innsbruck, Austria. He holds a PhD and habilitation degree in computer science. His research interests include software and security testing, software processes, requirements engineering, empirical software engineering, as well as improving industry-academia collaboration. Besides his research activities, he transfers his research results into practice as a consultant and speaker on industrial conferences. Contact him at michael.felderer@uibk.ac.at.