

REVIEW

Open Access

A mapping study of the Brazilian SBSE community

Wesley KG Assunção^{1*}, Márcio de O Barros², Thelma E Colanzi^{1,3}, Arilo C Dias-Neto⁴, Matheus HE Paixão⁵, Jerffeson T de Souza⁵ and Sílvia R Vergílio¹

* Correspondence: wesleyk@inf.ufpr.br

¹Federal University of Paraná (UFPR),
DInf, CP: 19081, CEP: 81531-980
Curitiba-PR, Brazil

Full list of author information is
available at the end of the article

Abstract

Research communities evolve over time, changing their interests for specific problems or research areas. Mapping the evolution of a research community, including the most frequently addressed problems, the strategies selected to propose solution for them, the venues on which results observed from applying these strategies are published, and the collaboration among distinct groups may provide lessons on actions that can positively influence the growth of research in a given field. To this end, this paper presents an analysis of the Brazilian SBSE research community. We present our major research groups focusing on the field, the software engineering problems most addressed by them, the search techniques most frequently used to solve these problems, and an analysis of our publications and collaboration. We could conclude that the Brazilian community is still expanding, both geographically and in terms of publications, and that the creation of a national workshop focusing on the research field was a keystone to allow this growth.

Keywords: Mapping studies; Bibliometrics analysis; Brazilian SBSE community

Introduction

In the technical literature, we find studies that report an increasing number of works and diversity of addressed Software Engineering (SE) areas in the Search Based Software Engineering (SBSE) field (Harman et al. 2009). By analyzing the SEBASE^a repository, we observe, as it happens with many other research fields, that the first SBSE works were published in European and North-American conferences: 166 out of 184 conference papers published up to 2005 were reported on European and North-American venues. Research was expanded to Asia and Oceania on the second part of the last decade and 62 out of 454 papers published up to 2009 were reported on conferences held on these places. More recently, a strong participation of South America has been perceived, essentially concentrated on Brazil; only one conference paper was published in a venue in Argentina and we did not identify other South American country in which SBSE papers were published.

A quick scan on SEBASE shows that 6% of its publications (of 1,093 papers as for the date) are accredited to Brazilian authors. This expressive number is mainly due to the organization, in the last three years, of an event in the area: the Brazilian Workshop on Search Based Software Engineering (WESB). Since then (2010), an increase in

the number of SBSE papers published by Brazilian authors can be perceived (Freitas and Souza 2011; Colanzi et al. 2012) and the SBSE field is growing rapidly in Brazil.

On the other hand, Brazil has gained international attention due to hosting large scale events, such as 2016 Olympic Games and FIFA World Cup 2014, and has attracted investments from different sources. This aroused interest in Brazilian research. This growing interest and the increasing number of Brazilian works in the SBSE field serve as motivation to this paper, which presents the SBSE Brazilian community and maps its production. This mapping aims to answer two main research questions.

- **RQ1:** What are the research groups working with SBSE in Brazil? This question allows the identification of existing groups: institutions and regions of the country; addressed SE areas; search-based techniques used; and number of researchers;
- **RQ2:** How is the production of the Brazilian community? This question allows acquiring knowledge about the field in Brazil. Collected data are analyzed considering the same categories used in a former bibliometrics analysis of the SBSE field (Liu et al. 2012): publication, sources, authorship and collaboration.

The main contributions of the research and analysis conducted and reported in this paper are presenting the Brazilian researchers and groups working on SBSE and providing an overview of the works produced by them. Furthermore, secondary contributions of this work include: 1) discussing some trends and challenges to this community; 2) contributing to consolidate the area in Brazil; and 3) allowing greater cooperation among its members and international researchers.

The paper is organized in sections. The next section encompasses related work, the methodology adopted to conduct the systematic mapping, as well as the mapping results. The section named Conclusions contains our final remarks.

Review

The systematic mapping is reported in this section. First, we present related work. After, we describe the methodology adopted in our study, including how data were collected and the categories used in the analysis. Finally, the collected data are analyzed to answer, respectively, the research questions RQ1 and RQ2.

Background

The field of SBSE is devoted to the application of search based algorithms to support different Software Engineering activities. In the last years, we observe a growing number of works in this field, reporting the use of such algorithms for software bug fixing (Emer et al. 2002), project management (Barreto et al. 2008; Braz and Vergilio 2004), process composition (Magdaleno 2010), refactoring, software slicing and comprehension (Harman 2007b), cloud engineering (Harman et al. 2012b), software repair (Goues et al. 2013), reverse engineering (Harman et al. 2013), and so on. Some of the former areas already have surveys and systematic reviews consolidating the works performed by researchers in their subfield, among them software testing (Afzal et al. 2009; Yoo and Harman 2012), software design (Räihä 2010) and software requirements (Zhang et al. 2008; Pitangueira et al. 2013).

Considering this interest, a great number of SBSE surveys have been published as bibliometrics analysis, mapping studies, or systematic literature reviews. For instance, in (Freitas and Souza 2011), the authors describe the first bibliometrics analysis to SBSE publications. A bibliometrics analysis is a quantitative analysis that describes patterns of publications within a given field or body of literature. This study covered 677 publications of the SBSE community from 2001 through 2010. They focused in four categories: Publication, Sources, Authorship, and Collaboration. Additionally, estimates for the next years of several publication metrics are given. The study also analyzed the applicability of bibliometrics laws in SBSE.

In (Colanzi et al. 2012), the authors present results of a mapping they had performed in order to provide an overview of the SBSE field in Brazil. The main goal is to map the Brazilian SBSE community by identifying the most active researchers, focus of published works, venue, and frequency of publications. It used the Brazilian Symposium on Software Engineering as a basis to identify works relating optimization and software engineering. Having a broader scope, the analyses presented in the present paper are based on a repository maintained by the international community, in which prediction and clustering works are also included (besides works on optimization). Due to this, different results are obtained when compared to (Colanzi et al. 2012), mainly considering authors and publications, now evaluated by using bibliometrics analysis. Thus, a major contribution of the present work, which was not presented previously, is the identification of the research groups and the collaboration links among them. New authors and groups were also identified, which is very important to expand and consolidate the Brazilian SBSE community.

Systematic Literature Reviews have also been performed in the context of SBSE. A Systematic Literature Review (SLR) is “*a means of evaluating and interpreting all available research relevant to a particular research question, topic area or phenomenon of interest*” (Dybå et al. 2005). For instance, in (Afzal et al. 2009), the authors present a systematic review that examined existing work into non-functional search-based software testing (NFSBST). They analyzed types of non-functional testing targeted using search techniques, different fitness functions used in distinct types of NFSBST, and challenges in the application of these techniques. The systematic review was based on a set of 35 papers published from 1996 to 2007. The results showed that search techniques have been applied for non-functional testing of several aspects, including execution time, quality of service, security, usability, and safety. A large number of search techniques were found to be applicable to those problems, including simulated annealing, *tabu* search, Genetic Algorithms (GAs), Ant Colony Optimization (ACO), grammatical evolution, Genetic Programming (GP) (and its variants), and swarm intelligence methods. The review reported on different fitness functions used to guide the search for each of the identified aspects.

(Ali et al. 2010) describe the results of a SLR that aimed at characterizing how empirical studies have been designed to investigate search-based software testing (SBST) cost-effectiveness and what empirical evidence is available in the literature regarding SBST cost-effectiveness and scalability. The authors also provided a framework that drives the data collection process and can be the starting point for guidelines on how SBST techniques can be empirically assessed. According to the authors, the intent would be aiding future researchers doing empirical studies in SBST by providing an

unbiased view of the body of empirical evidence and by guiding them in performing well-designed and executed empirical studies.

Finally, (Harman 2007a) describes a survey that analyzed a set of works on the application of optimization techniques in software engineering. The paper briefly reviews widely used optimization techniques and the key ingredients required for their successful application to software engineering, providing an overview of existing results in eight software engineering application domains. The paper also describes the benefits that are likely to accrue from the growing body of work in this area and provides a set of open problems, challenges and areas for future work.

Methodology

The main goal of a mapping study is to provide an overview of a research area, identifying the quantity and type of research and results available within it (Petersen et al. 2008). In our case, the goal is to provide an overview of the SBSE area in Brazil. To this end, seven researchers were involved in performing the mapping steps, according to the strategy delineated by (Petersen et al. 2008), in order to answer the research questions presented on Introduction. Next, we describe how these steps were conducted.

Conducting search and screening of works

The search for relevant publications was conducted by using the SBSE Repository SEBASE. This choice is due to two main reasons. The first one is that academic databases, such as IEEEExplorer and ACM Digital Library, may not include all the works from a specific field. The second one is that the chosen repository is considered a comprehensive base in the SBSE field and has been used as a reference by a similar work (Freitas and Souza 2011). It contains publications covering different sources and is frequently updated. However, we think those academic databases are very important, and to avoid significant omissions in our collected data, we conducted a verification step. Such steps and other ones included in our methodology are as follows.

In a first step, we asked the researcher responsible for maintaining and updating the SEBASE repository for a list of papers published by Brazilians. Then, the list was checked, considering a single inclusion criterion: at least one author is a Brazilian researcher. We define a Brazilian researcher as someone who works or studies in a Brazilian research institute, typically a state-held university or a research laboratory for a public or private company. This definition rules out research performed by Brazilians working for companies and universities which reside outside the Brazilian territory, except if those researchers have, in some point, returned to their home country and participated in local research to the extent required to be author in at least one paper. At the end of this step, a list of authors was generated. After this, we conducted a verification step to increase the level of confidence in the collected data. We searched in the DBLP site^b by other works from each author present in the generated list. In such step new authors were found and also searched.

As a result of all steps conducted and mentioned above, 73 works were obtained. It is important to notice that the search included only papers published before 6th February 2013.

Classification schema, data extraction and mapping

According to our research questions, two analyses were conducted considering a classification schema with different categories. On regard of the first research question, the following categories were considered:

- Research groups: we identify the main institutions appearing in the selected papers and the regularity of publications. From this, we identify the main research groups;
- Software engineering areas: this aspect reports the main interest areas of the identified groups. The areas were identified using the ACM Computing Classification System^c. Areas containing one or two papers were grouped in the category named Other subjects;
- Search-based techniques: this dimension shows what are the preferences and expertise of each group, and what are the most commonly used search-based techniques in each area. We consider categories such as classical and metaheuristic. We also decided to classify the metaheuristics according to the kind of evolutionary algorithm used and the number of objectives. In this analysis the surveys were excluded, as well as some other papers that do not clearly mention the applied algorithm.

On regard of the second research question, the analysis was based on the same categories adopted in the bibliometrics analysis described in (Freitas and Souza 2011):

- Publications: this aspect reports on the number of publications produced by Brazilian researchers working on the SBSE field;
- Sources: this aspect is related to the venue in which our selected papers were published, including conference proceedings, academic journals, books, and technical reports;
- Authorship: this is related to the number of Brazilian researchers entering or leaving the SBSE field;
- Collaboration: here we address the level of collaboration within the community, measured as the number of papers with more than one author, even from the same institution.

According to this classification schema, the following information was extracted for each paper and copied to a worksheet for further analysis: authors; institutions; publication year and venue; city/state; SE area; search-based technique used, and number of citations at Google Scholar^d. After this, the data were analyzed on regard of different views to answer the research questions. These views are presented in the next section.

SBSE Brazilian community

This section contains an analysis of the obtained information in order to answer RQ1 (Introduction). First we analyze the main areas and techniques addressed by the Brazilian community and after this we identify the main research groups, as well as their expertise and interests.

Software engineering areas

Brazilian SBSE papers were grouped by SE area according to our classification schema (Classification schema, Data Extraction and Mapping). The selected areas are: Testing, Management, Requirements, Software Design, and Other Subjects. Five surveys about SBSE were published by Brazilian researchers, but these surveys were not included in our analysis on regard of the SE area.

Software testing is the area that receives most attention from researchers: 47% of all works are dedicated to testing. This also happens in the international scenario (Harman et al. 2009). Most works address test data generation. Other works focus on test case prioritization and selection, integration testing, selection of testing strategies, test case allocation and test case selection and evaluation.

Testing is followed by Management (18%) and Requirements (11%). Management is an area in which search-based approaches were applied to optimize both the software development process and the product. Works address different tasks related to project planning, including scheduling, task and resource allocation. Other works are related to selection of project portfolios and software technologies.

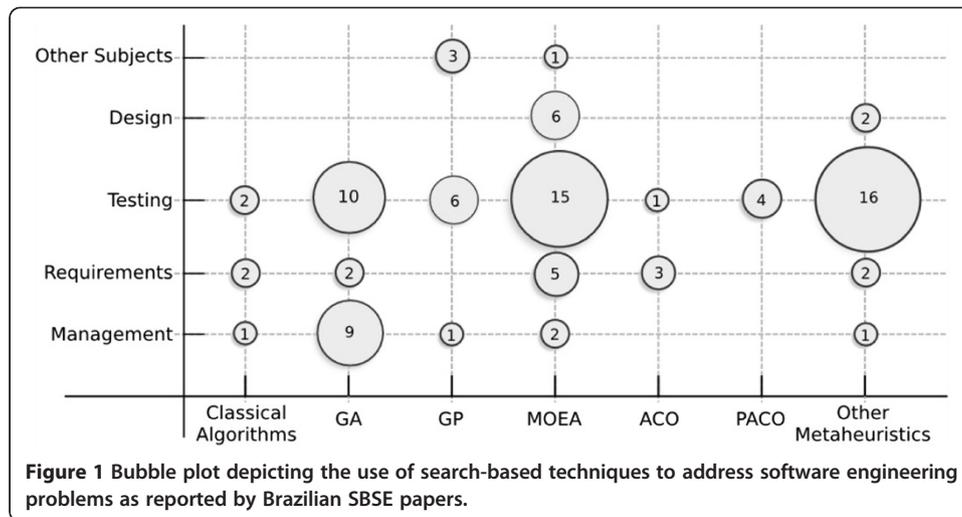
Among the works on Requirements, the Next Release Problem has been addressed by Brazilian researchers. The goal is to find an ideal set of requirements for a software release considering different objectives, such as customer's requests, resource constraints, and requirement characteristics. Most recently, requirements prioritization has also been addressed (in 2011 and 2012).

Since 2010 works involving other SE areas have appeared in the Brazilian scenario. However, we can also observe that the interest in Testing, Management and Requirements has not decreased. Therefore, this change only indicates that new application SE areas are emerging. 8% of the papers are about Software Design. In 2010 the first paper on this area was published addressing web services composition. From 2012 other works have addressed architecture refactoring and software module clustering. Other subjects, such as refactoring, software reliability prediction, component search, component clustering, database management and SBSE evaluation, have also been addressed recently. They represent 10% of all selected papers.

Search-based techniques

In this section, the SBSE papers are analyzed to determine the preferred search-based algorithms used by Brazilian researchers. Evolutionary techniques are preferred: Multi-Objective Evolutionary Algorithms (MOEAs), Genetic Algorithms (GA) and Genetic Programming (GP) represent 31%, 22% and 11%, respectively. NSGA-II is the most used MOEA, but SPEA2 and MOCcell are also frequently used. Besides evolutionary techniques, 22% of the studies have employed other metaheuristics, including Simulated Annealing, Hill Climbing and GRASP. Furthermore, 4 papers (4%) used Ant Colony Optimization (ACO) and 4 papers employed Pareto ACO (PACO) (4%). Classical techniques are applied by 5% of the papers: in this category Greedy and Branch-and-Bound are the most used.

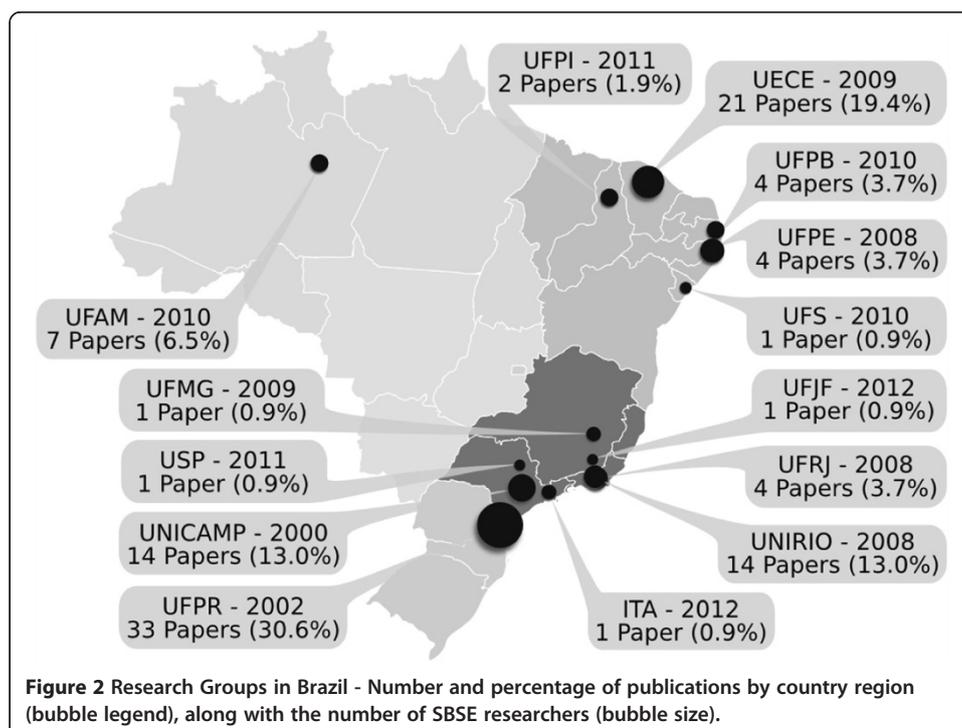
Figure 1 presents the number of SBSE papers that use some kind of optimization algorithm per SE area. MOEAs were applied in papers of all SE areas. Considering all papers, all kinds of optimization algorithms were used to solve problems related to Testing. In this case, other metaheuristics and MOEAs were the most used. For



Management, the most applied algorithm was GA. For Requirements, multi-objective algorithms are generally applied; ACO, GA, Simulated Annealing and GRASP were also applied. Finally, for Design, MOEAs were the most used.

Research groups

To map the research groups working with SBSE in Brazil, we first identified the institution of each author appearing in the selected papers. For each institution, we counted the number of published papers, as well as the frequency and regularity of publications. The result is depicted in Figure 2. This figure presents the institutions, number and percentage of papers published, as well as the year of the first publication. Papers



having more than one author from different groups were counted for all groups. The size of the circles represents the number of authors.

We can observe that the greatest number of authors is in UFPR (in Curitiba/Paraná, with 25 authors) and in UECE (in Fortaleza/Ceará, with 16 authors), the same universities that have the greatest number of publications. In addition, UNICAMP has nine authors, UFPE seven, UNIRIO and UFRJ six, followed by UFAM, UFPB and UFPI with four. ITA and UFMG have three researchers, UFS two, UFJF and USP only one. The first paper was published by the group from UNICAMP at 2000. Three institutions (out 14) published the first paper at 2010, when the first edition of WESB was promoted; they were followed by two institutions in 2011 and two 2012. Six institutions have published only one or two papers. In our analysis characterizing the research groups such institutions were not included, since we consider that they do not have a consolidate group with clear interests in SBSE. Hence, they do not appear in Figure 3 which contains information about the SE areas and techniques addressed per research groups.

Taking into account the SE area, the groups of UECE, UFPR and UNIRIO have publications in almost all areas. For UECE we observe a great interest for Requirements, followed by Testing. Testing is also the most addressed area at UFPR, UNICAMP, and UFAM. UFPR also has interest in other areas, such as Software Design and Management. The major interests of UNIRIO are Management and Software Design. Management is also addressed by UFPB and UFRJ. UFPE has publications only in Management.

Regarding the search-based techniques, we observe that MOEAs are the preferred technique used by UECE and UFPR. ACO is employed only by UECE. UFPR, as well as UNICAMP, also have several publications using Other Metaheuristics. The multi-objective algorithm PACO was only used by UFPR. This institution also has an expressive number of GP-based works. The research group at UNIRIO has concentrated its efforts on GA. UNICAMP, UFPB and UFPE have preference for GA. UFAM also prefers GA, besides Classical Algorithms. Finally, UFRJ published papers only with Classical Algorithms.

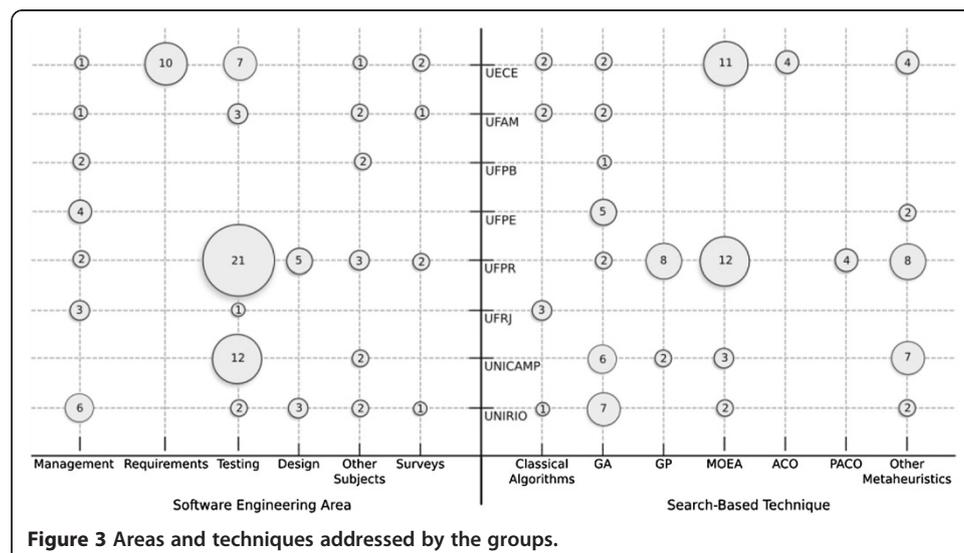


Figure 3 Areas and techniques addressed by the groups.

Production analysis

This section presents a quantitative analysis considering the production of the Brazilian community, using the categories from bibliometrics analysis (Liu et al. 2012), in order to answer RQ2 (Introduction).

Publications

Clearly, the most significant piece of information to characterize the development of SBSE in Brazil relates to the number of SBSE papers published by the community. Table 1 shows exactly the number of SBSE papers published by the Brazilian community, from 2000 to 2012, on a year-on-year basis. Also, it presents the cumulative amount by each year, the contribution of a particular year on the total amount and the growth compared to the previous year.

As can be seen in Table 1, in total, SBSE Brazilian researchers have published, up to 2012, 98 works, which represent 6.67% of all papers reported in the SEBASE repository. That overall result places Brazil as the fourth country with more SBSE papers, behind UK (25%), USA (16%) and Canada (8%). Table 1 shows there was a significant acceleration on the amount of publications in 2010, with more than 70% of the papers being published from 2010 to 2012. That behavior can be at least partially attributed to WESB, which started in 2010 and accounted for 9 papers in that year. However, the other 14 works published in 2010, which represents almost all papers published up to 2008, show this significant growth reached beyond this important national event.

When one considers the languages used in these papers, the analysis showed that 80.62% works were published in English, which accounts for 79 out of the 98 papers. The other 19 papers (19.38%) were published in Portuguese. Those results show that SBSE research in Brazil happens both at national (papers published in the local workshop) and international (papers published in conferences and journals abroad) levels. The local workshop works as leverage for papers in early development stages, engaging

Table 1 Number of works in each year between 2000 and 2012, including cumulative amount

Year	Quantity	%	Growth	Cumulative	%	Growth
2000	1	1.02%	-	1	1.02%	-
2001	1	1.02%	0%	2	2.04%	100.00%
2002	2	2.04%	100.00%	4	4.08%	100.00%
2003	2	2.04%	0%	6	6.12%	50.00%
2004	3	3.06%	50.00%	9	9.18%	50.00%
2005	2	2.04%	-33.33%	11	11.22%	22.22%
2006	2	2.04%	0%	13	13.26%	18.18%
2007	3	3.06%	50%	16	16.32%	23.07%
2008	6	6.12%	100.00%	22	22.45%	37.50%
2009	4	4.08%	-33.33%	26	26.53%	18.18%
2010	23	23.47%	475%	49	50.0%	88.46%
2011	27	27.56%	17.39%	76	77.55%	55.10%
2012	22	22.45%	-18.51%	98	100.00%	28.94%
2000-2012	98	100.00%		98	100.00%	

researchers in discussions with their partners and providing feedback that will ultimately help the paper to reach for larger venues.

Table 2 presents the ranking with the 15 most cited publications, based on Google Scholar. Not surprisingly, 7 out of the 15 most cited papers relate to testing, which is consistent with the ratio of papers published by area. However, the most cited paper tackled a management problem (team allocation). In general, as expected, the overall impact of the papers is still narrow, since the huge majority of the published papers have less than three years.

Source

When considering the venue where those SBSE works have been published, the analysis showed that around 75% (74 papers) of the publications were published in conference proceedings, with 20 articles (20.04%) in journals, 2 posters (2.04%), 1 (1.02%) book chapter and 1 (1.94%) technical report. This result, when compared with the behavior of the overall SBSE world community (Freitas and Souza 2011), which published around 70% of their works in conference proceedings and 20% in journals, shows a local preference to conferences, which, once again, can be attributed to the presence

Table 2 The 15 most cited SBSE works, ordered by Google Scholar

Ref.	Work	SE Area	GS
(Barreto et al. 2008)	Staffing a Software Project: a Constraint Satisfaction and Optimization-based Approach	Management	72
(Bueno and Jino 2002)	Automatic Test Data Generation for Program Paths using Genetic Algorithms	Testing	49
(Oliveira et al. 2010)	GA-based Method for Feature Selection and Parameters Optimization for Machine Learning Regression applied to Software Effort Estimation	Management	47
(Bueno and Jino 2000)	Identification of Potentially Infeasible Program Paths by Monitoring the Search for Test Data	Testing	39
(Costa et al. 2005)	Modeling Software Reliability Growth with Genetic Programming	Testing	36
(Harman et al. 2012a)	Search Based Software Engineering: Techniques, Taxonomy, Tutorial	Survey	32
(Braz and Vergilio 2004)	Using Fuzzy Theory for Effort Estimation of Object-Oriented Software	Management	28
(Costa et al. 2007)	Exploring Genetic Programming and Boosting Techniques to Model Software	Testing	26
(Braga et al. 2008)	A GA-Based Feature Selection and Parameters Optimization for Support Vector Regression Applied to Software Effort Estimation	Management	24
(Costa et al. 2010)	A Genetic Programming Approach for Software Reliability Modeling	Other Subjects	20
(de Carvalho et al. 2010)	A Symbolic Fault-Prediction Model Based on Multiobjective Particle Swarm Optimization	Testing	20
(Emer and Vergilio 2003)	Selection and Evaluation of Test Data Based on Genetic Programming	Testing	18
(Souza et al. 2010)	The Human Competitiveness of Search Based Software Engineering	Other Subjects	18
(Bueno et al. 2007)	Improving Random Test Sets using the Diversity Oriented Test Data Generation	Testing	16
(Freitas and Souza 2011)	Ten Years of Search Based Software Engineering: A Bibliometric Analysis	Survey	16

and value Brazilian SBSE researchers give to WESB, which published almost 25% of all SBSE Brazilian works (Table 3). In addition to this conference, Brazilian researchers have shown a relevant presence in the International Symposium on Search-Based Software Engineering (SSBSE), publishing 14 papers in total. The other four conferences with most works can be seen in Table 3.

In terms of journal venues, SBSE Brazilian publications have been significantly spread with 3 journals publishing 2 articles, and 14 journals publishing a single paper.

Authorship

In Table 4, it is presented statistics concerning the Brazilian SBSE publishing authors. It shows, in different columns, the number of “New” authors publishing for the first time in each year, the percentage participation and growth of these new authors, the number of “Active” authors per year and the numerical relation between new and active authors.

Taking into account the 98 analyzed papers, there are 90 Brazilian authors. The number of authors joining the Brazilian SBSE community increased significantly in 2010, with a growth of 200% compared to 2009. In that year, 21 out of the 41 publishing authors were new. Even with a decrease of authors since then, the national SBSE community had 41 active authors in 2012. Since there is no data available on the number of SBSE active authors in the world in 2012, we can compare the number from 2010 (Freitas and Souza 2011). In that year, there were 168 SBSE active authors overall, which indicates that the Brazilian community contributed with 21.4% of all publishing authors in 2010, representing the 36 active Brazilian authors in that year.

When considering authors individually, as can be seen in Table 5, 38 of those authors published only one paper and 25 of them published two, with only five authors publishing more than 10 papers. This scenario may help to confirm we have a community growing in the last 3 years, as presented in Table 4.

Table 6 shows the Brazilian researchers with more works published, including conference and journal articles, up to 2012. It also presents the number of citations at Google Scholar (taking into account only the papers selected for our mapping study), percentage of participation in relation to all Brazilian SBSE publications, the amount and position of each author concerning only publications in conference proceeding and journal articles. The table shows that the two most prolific authors account for almost half of all publications (36 out of the 73).

Collaboration

In regard of collaboration, which accounts for papers with more than one author, even if those authors are from the same institution, Table 7 shows statistics about the

Table 3 Ranking of conferences proceedings with the most number of SBSE publications

Conference proceeding	#	%
Brazilian Workshop on Search-Based Software Engineering (WESB)	24	24.48%
International Symposium on Search-Based Software Engineering (SSBSE)	14	14.28%
Annual Conference on Genetic and Evolutionary Computation (GECCO)	6	6.12%
International Conference on Tools with Artificial Intelligence (ICTAI)	4	4.08%
Brazilian Symposium on Software Engineering (SBES)	3	3.06%
International Workshop on Search-Based Software Testing (SBST)	3	3.06%

Table 4 Statistics for new and active authors per year (2000-2012)

Year	New	%	Growth	Active	Growth	New/active
2000	2	2.22%	-	2	-	1.00
2001	0	0%	-100.0%	2	0%	0
2002	2	2.22%	-	4	100.0%	0.50
2003	3	3.34%	50.0%	5	25.0%	0.60
2004	3	3.34%	0%	4	-20.0%	0.75
2005	1	1.11%	66.67%	5	25.0%	0.20
2006	0	0%	-100.0%	3	-40.0%	0
2007	3	3.34%	-	9	200.0%	0.33
2008	14	15.56%	366.67%	18	55.55%	0.77
2009	7	7.77%	-50.0%	14	7.70%	0.50
2010	21	23.33%	200.0%	41	192.85%	0.51
2011	15	16.66%	-28.57%	36	-12.19%	0.41
2012	19	21.11%	26.66%	41	13.88%	0.46
2000-2012	90	100.00%	-	-	-	-

number of authors per paper. Papers with 2 authors represent 29.59% of all published papers. Only 6 works (6.12%) had a single author. Around 35% of the papers had four or more authors, which shows a significant collaborative level. When compared to the collaborative results presented in (Freitas and Souza 2011), where, considering all SBSE publications from 2001 to 2010, only 21.14% of paper had more than 4 authors and more than 13% had only one. Those results confirm a higher collaborative attitude by the SBSE Brazilian community.

Conclusions

The Brazilian SBSE community had been recognized in the last few years as one of the most active in the world. Brazilian researchers have published search-based papers for many Software Engineering areas from 2000 to 2012, including software testing (54 papers), project management (14), requirements (14), and software design (8). They have also used a wide range of search algorithms on their approaches, including multi-objective evolutionary algorithms (26 papers), genetic algorithms (25), genetic programming (10), and ant colony optimization (8). Both in quantitative as well as qualitative terms, this research community has shown a recent and significant growth which qualifies it as an interesting case to be presented and reflected upon. We have identified 98 papers published in the field by 90 different authors from 2000 to 2012, summing up to more than 700 citations on Google Scholar. The number of publications (as well as collaboration between authors and the number of researchers engaging the field) has shown a strong increase since 2010 and more than 70% of the papers were published from this year on. In that sense, this paper discussed this community, by presenting the SBSE Brazilian researchers and groups and providing an overview of the works being produced by them.

Table 5 Number of authors that published each amount of works

Works	1	2	3	4	5	6	7	8	13	21	32
Authors	38	25	5	7	4	2	2	2	2	2	1

Table 6 The ranking of SBSE authors with more than 5 publications in 2000-2012

Rank	Author		Works	%	Citations (rank)	Conference Proc. (rank)	Journal articles (rank)
	Name	Affiliation					
1	Silvia Vergilio	UFPR	32	32.65%	272 (1)	21 (1)	10 (1)
2	Aurora Pozo	UFPR	21	21.42%	181 (2)	11 (3)	9 (2)
2	Jerffeson Souza	UECE	21	21.42%	120 (4)	18 (2)	3 (3)
3	Fabricio Freitas	UECE	13	13.26%	66 (7)	11 (3)	2 (4)
3	Marcio Barros	UNIRIO	13	13.26%	86 (6)	10 (4)	1 (5)
4	Camila Maia	UECE	8	8.16%	52 (8)	7 (5)	1 (5)
4	Thelma Colanzi	UFPR	8	8.16%	26 (10)	7 (5)	1 (5)
5	Ariilo Dias-Neto	UFAM	7	7.14%	6 (11)	6 (6)	0 (6)
5	Mario Jino	UNICAMP	7	7.14%	122 (3)	5 (7)	2 (4)
6	Paulo Bueno	UNICAMP	6	6.12%	116 (5)	4 (8)	2 (4)
6	Rafael Carmo	UECE	6	6.12%	45 (9)	5 (7)	1 (5)

This work served to highlight interesting aspects of the community, including the importance of the Brazilian Workshop on Search Based Software Engineering (WESB), which started in 2010 and helped significantly in the increase and consolidation of the Brazilian SBSE community. Other than that, the mapping showed a significant spread in the active SBSE groups in geographical terms, which gives a perspective of an even more dissemination in the future.

This paper may serve as motivation to other national or institutional research groups present their contributions to the SBSE field, which may help the international SBSE community to understand itself and use this knowledge to drive its own evolution. As the Brazilian SBSE community prepares to organize SSBSE in 2014, this work may serve both as a presentation as well as an invitation to this important event, which it is expected to strengthen and consolidate SBSE research in Brazil.

The behavior of Brazilian researchers in the last few year resembles that presented by the international research community in the early-years of the SBSE field: the first studies concentrated almost entirely in the application of heuristic search to software testing; later, some authors moved to address software requirements, design, and project management. While researchers from abroad are exploring these late areas intensively since 2001, each one of them representing about 10% of the world-wide published papers on SBSE from 2001 on, the domestic community caught up around 2008/2009. Comparing samples of very different sizes is always dangerous, but we observe that recent works produced by the Brazilian community tend to be more diverse in application areas than those produced by the international community as a whole. For instance, research in SBSE testing accounts for only 49% of the papers published by Brazilians since 2008, while they represent 61% of the papers published by the international community. Project management has also received more attention of the

Table 7 Authorship pattern in the period 2000-2012

Authors	1	2	3	4	5+
# Works	6	29	28	25	10
%	6.12%	29.59%	28.57%	25.52%	10.20%

Brazilian community (24% of the papers since 2008), while design and requirements receive about the same attention as in the international community.

Regarding the future of the SBSE research and practice in Brazil, we identify some 'hot' trends by analyzing the papers published recently, including high-level design optimization, selection of refactoring strategies, evaluation of SBSE experimental studies, application of SBSE approaches to industrial cases, and using recent heuristic optimization approaches (such as Ant Colony Optimization) to address classic problems (such as the next release problem) in more complex cases. Thus far, no area has shown signs of losing interest from researchers, both in the international and the domestic communities, though research in software testing continues to be the leading trend. The trends discussed above may drive the next steps in research, which may include more interaction with local industry, more collaboration with international partners, a deeper understanding of certain classic problems (such as software module clustering and the next release problem), and customizing problems and their solutions to specificities of the Brazilian industry.

Endnotes

^aSEBASE SBSE Repository, http://crestweb.cs.ucl.ac.uk/resources/sbse_repository/.

^bwww.informatik.uni-trier.de/~ley/db/. It includes most relevant computer science conferences and journals, published by IEEE, ACM and LNCS, besides important book series.

^cwww.computer.org/portal/web/publications/acmsoftware

^dThe worksheet containing the list of the selected papers and their classification is available at: http://www.inf.ufpr.br/gres/apoio_en.html

Abbreviations

ACO: Ant colony optimization; GA: Genetic algorithm; GP: Genetic programming; GRASP: General responsibility assignment software patterns; MOCeLL: Mathematical olympiad cell; MOEA: Multi-objective evolutionary algorithm; NFSBST: Non-functional search-based software testing; NSGA-II: Non-dominated sorting genetic algorithm-II; PACO: Pareto ACO; SBSE: Search based software engineering; SBST: Search based software testing; SE: Software engineering; SLR: Systematic literature review; SPEA2: Strength pareto evolutionary algorithm 2; WESB: Brazilian workshop on search based software engineering.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors carried the bibliometrics analysis, collected, and analyzed the papers. WKGA analyzed the data (publications) per research group, generating the Brazil's map with all identified groups. MOB and ACDN searched for works related to this research to contextualize the contribution. TEC and SRV analyzed publications per software engineering area and search based techniques. MHEP and JTS analyzed publications per year and collaborations among researchers identified in our mapping study. At the end, all authors read and approved the final manuscript.

Author's information

WKGA: is an assistant professor at Federal Technological University of Paraná (UTFPR). He is currently a PhD candidate at the Postgraduate Program in Informatics (PPGInf) of Federal University of Paraná (UFPR). He also holds the MSc degree from PPGInf of UFPR. His main research interests are: model based software reuse, software testing, search-based software engineering, multi-objective optimization and bio-inspired computation.

MOB: is an associate professor at the Applied Informatics Department of the Federal University of Rio de Janeiro State (UNIRIO). He holds a Doctor degree in System Engineering and Computer Science from COOPE/UFRJ. His research concentrates on using optimization and simulation to describe and find (close to) optimal solutions to software engineering problems and bring forth insight about these problems. His main research interests within Software Engineering are software design and project management.

TEC: is an assistant professor at State University of Maringá. She is currently a PhD candidate at the Postgraduate Program in Informatics of Federal University of Paraná (UFPR). She holds the MSc degree in Computer Science and Computational Math from University of São Paulo (USP/São Carlos), Brazil. Her areas of interest are: software architecture, software product lines, search-based software engineering and multi-objective evolutionary algorithms.

ACDN: is an associate professor at Institute of Computing at the Federal University of Amazonas. He holds a Doctor degree in Systems Engineering and Computer Science from COPPE/UFRJ. He leads the Experimentation and Testing on Software Engineering Group (ExperTS) at the Federal University of Amazonas. Additional information can be found at <http://www.icomp.ufam.edu.br/arilo>.

MHEP: is currently a MSc student at the Postgraduate Program in Computer Science of the State University of Ceará (UECE). His current research focuses on the application of robust optimization to requirements engineering. His areas of interest are: software requirements, search based requirements engineering and evolutionary algorithms.

JTS: is a professor and the dean of graduate studies and research at State University of Ceará (UECE), Brazil. He holds a Ph.D. in Computer Science from the School of Information Technology and Engineering (SITE) of University of Ottawa, Canada. His main research interests are: search-based software engineering, multi-objective evolutionary algorithms, machine learning and software patterns.

SRV: received the MS (1991) and DS (1997) degrees from University of Campinas, UNICAMP, Brazil. She is currently at the Computer Science Department at the Federal University of Paraná, Brazil, where she has been a faculty member since 1993. She has been involved in several projects and her research interests are in the area of Software Engineering, such as: software testing, search based software engineering and software metrics.

Acknowledgements

We should like to thank CNPq, FAPERJ, FAPEAM and INCT-SEC for their financial support.

Author details

¹Federal University of Paraná (UFPR), DInf, CP: 19081, CEP: 81531-980 Curitiba-PR, Brazil. ²Applied Informatics Department, Federal University of Rio de Janeiro State, CEP: 22240-090 Rio de Janeiro-RJ, Brazil. ³Informatics Department (DIN), State University of Maringá (UEM), CEP: 87020-900 Maringá-PR, Brazil. ⁴Institute of Computing, Federal University of Amazonas, CEP: 69077-000 Manaus-AM, Brazil. ⁵Optimization in Software Engineering Group (GOES.UECE), State University of Ceará, Fortaleza, Ceará, Brazil.

Received: 29 October 2013 Accepted: 21 April 2014

Published: 6 May 2014

References

- Afzal W, Torkar R, Feldt R (2009) A systematic review of search-based testing for non-functional system properties. *Inf Softw Technol* 51(6):957–976
- Ali S, Briand LC, Hemmati H, Panesar-Walawege RK (2010) A systematic review of the application and empirical investigation of search-based test case generation. *IEEE Trans Softw Eng* 36(6):742–762
- Barreto A, Barros M, Werner CML (2008) Staffing a software project: a constraint satisfaction and optimization-based approach. *Comput Oper Res* 35(10):3073–3089
- Braga PL, Oliveira ALI, Meira SRL (2008) A GA-based feature selection and parameters optimization for support vector regression applied to software effort estimation. In: *Proceedings of the ACM Symposium on Applied Computing (SAC '08)*. Butterworth-Heinemann, New York, USA, pp 1788–1792
- Braz M, Vergilio S (2004) Using fuzzy theory for effort estimation of object-oriented software. In: *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'04)*. IEEE Computer Society, Washington, USA, pp 196–201
- Bueno P, Jino M (2000) Identification of potentially infeasible program paths by monitoring the search for test data. In: *Proceedings of the fifteenth IEEE International Conference on Automated Software Engineering (ASE'00)*. French, Grenoble, pp 209–218
- Bueno P, Jino M (2002) Automatic test data generation for program paths using genetic algorithms. *Int J Softw Eng Knowl Eng* 12(06):691–709
- Bueno P, Wong W, Jino M (2007) Improving random test sets using the diversity oriented test data generation. In: *Proceedings of the 2nd International Workshop on Random Testing (WRT'07)*. ACM New York, NY, USA, pp 10–17
- Colanzi TE, Vergilio SR, Assunção WKG, Pozo A (2012) Search based software engineering: review and analysis of the field in Brazil. In: *Journal of Systems and Software*. Elsevier Science Inc., New York, NY, USA, pp 1–15
- Costa EO, de Souza GA, Pozo ATR, Vergilio SR (2007) Exploring genetic programming and boosting techniques to model software reliability. *IEEE Trans Reliab* 56(3):422–434
- Costa E, Pozo A, Vergilio S (2010) A genetic programming approach for software reliability modeling. *IEEE Trans Reliab* 59(1):222–230
- Costa EO, Vergilio SR, Pozo A, Souza G (2005) Modeling software reliability growth with genetic programming. In: *Proceedings of the 16th IEEE International Symposium on Software Reliability Engineering (ISSRE '05)*. IEEE Computer Society, Washington, DC, USA, pp 171–180
- de Carvalho AB, Pozo A, Vergilio SR (2010) A symbolic fault-prediction model based on multiobjective particle swarm optimization. *J Syst Softw* 83(5):868–882
- Dybå T, Kitchenham B, Jorgensen M (2005) Evidence-based software engineering for practitioners. *IEEE Softw* 22(1):158–165
- Emer M, Vergilio S (2003) Selection and evaluation of test data based on genetic programming. *Softw Qual J* 11(2):167–186
- Emer M, Vergilio S, GPTesT (2002) A testing tool based on genetic programming. In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'02)*. San Francisco, CA, USA, pp 1343–1350
- Freitas FG, Souza JT (2011) Ten years of search based software engineering: a bibliometric analysis. In: *Proceedings of the Third International Conference on Search Based Software Engineering (SSBSE'11)*. Springer, Berlin Heidelberg, pp 18–32
- Goues CL, Forrest S, Weimer W (2013) Current challenges in automatic software repair. *Softw Qual J* 21(3):421–443

- Harman M (2007a) The current state and future of search based software engineering. In: *Future of Software Engineering (FOSE '07)*. IEEE Computer Society, Washington, DC, USA, pp 342–357
- Harman M (2007b) Search Based Software Engineering for Program Comprehension. In: *ICPC '07. Proceedings of the 15th IEEE International Conference on Program Comprehension*. IEEE Computer Society, Washington, DC, USA pp 3–13. doi:10.1109/ICPC.2007.35
- Harman M, Lakhota K, Singer J, Yoo S (2012a) Cloud engineering is search-based software engineering too. *J Syst Softw* 86(9):2225–2241
- Harman M, Langdon WB, Weimer W (2013) Genetic Programming for Reverse Engineering. 20th Working Conference on Reverse Engineering (WCRE'13). Koblenz, Germany
- Harman M, Mansouri SA, Zhang Y (2009) Search Based Software Engineering: A Comprehensive Analysis and Review of Trends Techniques and Applications. Tech. Rep. TR-09-03, King's College London
- Harman M, McMinn P, Souza JT (2012b) Search Based Software Engineering: Techniques, Taxonomy, Tutorial. *Empirical Software Engineering and Verification*. Springer, Berlin Heidelberg, pp 1–59
- Liu Z, Liu Y, Guo Y, Wang H (2012) Progress in Global Parallel Computing Research: A Bibliometric Approach. *Scientometrics*, Springer, Netherlands, pp 1–17
- Magdaleno A (2010) An optimization-based approach to software development process tailoring. In: *Proceedings of the Second International Symposium on Search-based Software Engineering (SSBSE'10)*. Benevento, Italy, pp 40–43
- Oliveira A, Braga P, Lima P, Cornelio M (2010) GA-based method for feature selection and parameters optimization for machine learning regression applied to software effort estimation. *Inf Softw Technol* 52(11):1155–1166
- Petersen K, Feldt R, Mujtaba S, Mattsson M (2008) Systematic Mapping Studies in Software Engineering. 12th International Conference on Evaluation and Assessment in Software Engineering (EASE'08). British Computer Society, UK, pp 1–10
- Pitangueira AM, Maciel RSP, Barros MO, Andrade AMS (2013) A systematic review of software requirements selection and prioritization using SBSE approaches. In: *Proceedings of the V Symposium on Search-based Software Engineering (SSBSE'13)*. Springer, Berlin Heidelberg, pp 188–208
- Räihä O (2010) A survey on search-based software design. *Comput Sci Rev* 4:203–249
- Souza JT, Maia C, Freitas FG, Coutinho D (2010) The human competitiveness of search based software engineering. In: *Proceedings of the Second International Symposium on Search-based Software Engineering (SSBSE'10)*. Benevento, Italy, pp 143–152
- Yoo S, Harman M (2012) Regression testing minimization, selection and prioritization: a survey. *Softw Test Verification Reliabil* 22:67–120
- Zhang Y, Finkelstein A, Harman M (2008) Search-based requirements optimisation: existing work & challenges. In: 4th International Conference on Requirements Engineering: Foundation for Software Quality (RE'2008). Springer, Berlin Heidelberg, pp 88–94

doi:10.1186/2195-1721-2-3

Cite this article as: Assunção et al.: A mapping study of the Brazilian SBSE community. *Journal of Software Engineering Research and Development* 2014 **2**:3.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com