Ten Years of Search Based Software Engineering: A Bibliometric Analysis

Fabrício Gomes de Freitas and Jerffeson Teixeira de Souza

Optimization in Software Engineering Group (GOES.UECE)
State University of Ceará, Avenue Paranjana, 1700, Ceará, Brazil
fabriciogf.uece@gmail.com, jeff@larces.uece.br

Abstract. Despite preceding related publications, works dealing with the resolution of software engineering problems by search techniques has especially risen since 2001. By its first decade, the Search Based Software Engineering (SBSE) approach has been successfully employed in several software engineering contexts, using various optimization techniques. Aside the relevance of such applications, knowledge regarding the publication patterns on the field plays an important role to its understanding and identity. Such information may also shed light into SBSE trends and future. This paper presents the first bibliometric analysis to SBSE publications. The study covered 740 publications of the SBSE community from 2001 through 2010. The performed bibliometric analysis concerned mainly 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 bibliometric laws in SBSE, such as Bradfords and Lotka.

Keywords: sbse research analysis, bibliometric, authorship pattern.

1 Introduction

Optimization approaches have been applied to solve software engineering problems since the 1970s [1]. The early works were mostly concerned of solving software testing problems, and, in particular, test data generation. Until the 1990s, some sporadic works also used search techniques in software estimations and software management.

In 2001, the SEMINAL (Software Engineering using Metaheuristic INnovative ALgorithms) workshop was organized to discuss the wider use of optimization methods in the software engineering context [2]. Also in 2001, a special issue of the Information and Software Technology journal was devoted to the application of search methods in software engineering. A paper [3] stated the validity of the approach in the software engineering context and the term "Search Based Software Engineering" (SBSE) was coined to identify such approach. Since then, the SBSE approach has received increasing attention. The frequency and diversity of SBSE applications has increased significantly, and it is now considered a consolidated research field [4].

M.B. Cohen and M. Ó Cinnéide (Eds.): SSBSE 2011, LNCS 6956, pp. 18–32, 2011.

From 2001 to 2010, the number of works relating to the field has increased considerably. As an attempt to cover the state-of-the-art status of the field, survey papers on SBSE have been presented [4][5]. Beyond the information summarized in such works, mainly concerned on the applications and problems tackled, publication patterns of the field should also be analyzed. This sort of information may be used as a way to improve the understanding of the field, recognize its research community, and identify its trends, among others.

Given the importance of such aspects for the development of the SBSE field, this work presents the first bibliometric analysis of the area. The main contributions of this paper are to:

1. Provide facts about the growth of the field, regarding both publications and authors. For completeness' sake, the analysis is performed in a year-on-year basis, and includes the discussion of several related metrics.
2. Indicate rankings of the SBSE literature. This aspect includes the most cited papers, the most prolific authors, and the journal with most SBSE papers.
3. Show the distribution of the SBSE publications among the available sources. Additionally, an analysis of the number of publications in conference proceedings against the number of journal articles is performed.
4. Present and discuss the level of collaboration among researchers in the SBSE literature. The collaboration is analysed regarding the amount the authors in the papers, and also as in relation to the cooperation among different universities and countries. The level of participation of researches institute and companies is also analysed.
5. Verify the application of two bibliometric laws in SBSE: Bradford's and Lotka's laws.
6. Compare the behaviour of the field in its first official decade against the previous period (since 1970 until 2000), and supply estimates for the next years regarding some bibliometric metrics, based on the previous behavior.

The paper is organized as follows. Section 2 describes the methodology used in this study, including the definition of the categories analysed, as well as the data used. Section 3 shows the bibliometric analysis for the four categories analysed (Publications, Sources, Authorship, and Collaboration) in the 2001-2010 period. Section 4 presents and discusses the estimates for the next years of the area. Section 5 briefly compares the metrics for the decade 2001-2010 with publication patterns of the period before 2001. Finally, Section 6 discusses conclusions and states future works.

2 Methodology

The bibliometric analysis presented in this study is divided on four main segments. Each segment represents a group of statistics related to a bibliometric aspect. The segments are described on Section 2.1.

In Section 2.2, the data used for the study is explained, together with its source.

2.1 Segments

Publications. This category covers information about the SBSE published works. The number of publications of a research field is an important indicative of its

development. The evolution of this amount through the years is also a significant figure. The most cited works are also presented and discussed.

Sources. One aspect related to the publications is the type of venue where the works are published. In this context, the distribution of the SBSE publications among sources such as conference proceedings, journals, books, etc., is analyzed. In addition, this category also examines the core journals of the field, i.e., the venues that published the most amounts of SBSE articles. This data is used to verify the Lotka's bibliometric law [6] of scientific productivity.

Authorship. Aside the works of a field, the researchers authoring these works should also be studied. In order to do this, the year by year amount of active, new and cumulative authors on SBSE is presented. The distribution of the number of works published by authors is used to evaluate the validity of the Bradford's bibliometric law [7] in the SBSE field. A ranking of the most prolific authors is also discussed.

Collaboration. The level of collaboration among SBSE authors is covered in this category. The analysis comprises two aspects: the number of authors per paper, and collaboration among groups. The cooperation among groups includes internal collaboration, when two or more groups in the same country collaborate in the paper, and external collaboration, when there are authors from more than one country.

2.2 Data Source

The source of the publications is one important aspect for a bibliometric analysis. One alternative is to use academic databases, but them may not include all the works from a field. For the SBSE field, this problem can be avoided by the use of the SBSE Repository from SEBASE [8], which is a tool that provides a comprehensive list of SBSE publications. Indeed, by covering works from different sources, it includes papers that may not appear in a particular database. The list is actively updated, including updates suggested by the SBSE community. Such repository is appropriate as source of information for this study, since it portrays the wider status of the field.

The inclusion of a work may take time. Then, more recent works may not appearr in the database. The publication data online in late 2010 covered 667 publications in 2001-2010. Those data were used for all the segments, but "Publications". For such segment, we were able to get an ongoing updated list in June 2011 with 740 works.

The repository is formed by a list of publications, and tools for search and ordering. For our analysis, scripts were used in order to extract the required information. Some information necessary for the study were not available in the repository. The citations data were collected for each work in Scopus and Google Scholar. Authors' affiliations were obtained by the related information in each work.

3 SBSE Bibliometric Analysis

3.1 Publications

The number of publications in a field is a central information of its development. Table 1 presents, on a year-on-year basis, the evolution of the number of SBSE

publications. The cumulative amount by each year is also presented. The contribution of a year on the total amount and the growth of the quantity against the previous year are also indicated.

Table 1. Number of works in each year between 2001 and 2010, including cumulative amount

Year	Quantity	%	Growth	Cumulative	%	Growth
2001	24	3.24%	-	24	3.24%	-
2002	30	4.05%	25.00%	54	7.30%	125.00%
2003	37	5.00%	23.33%	91	12.30%	68.52%
2004	45	6.08%	21.62%	136	18.38%	49.45%
2005	54	7.30%	20.00%	190	25.68%	39.71%
2006	61	8.24%	12.96%	251	33.92%	32.11%
2007	83	11.22%	36.07%	334	45.14%	33.07%
2008	127	17.16%	53.01%	461	62.30%	38.02%
2009	126	17.03%	-0.79%	587	79.32%	27.33%
2010	153	20.68%	21.43%	740	100.00%	26.06%
2001-2010	740	100.00%	-	740	100.00%	-

As shown in Table 1, the quantity of works by year was continuously increasing since 2001. Indeed, in the first three years after 2001, for example, the growth rate between sequential years was higher than 20%. Also, significant growth rates of 36.07% and 53.01% are found in 2007 and 2008, respectively. In 2009 there was a negativegrowth of 0.79%. In 2010, the number of works published was 21.43% higher than the previous year (153 against 126). In addition, the works published in 2010 correspond to 20.68 of all works, being the first year in number of publications. This shows that the field progress was still in considerable development in 2010.

The cumulative amounts on the first years indicate that the SBSE community achieved the first hundred publications in 2004, i.e., three years after its formal creation. The next hundred publications level was reached only two years after 2004, in 2006. Then, the next hundred publications were achieved in 2007, i.e., only one year after the previous landmark. These findings are a way to indicate the acceleration on the number of SBSE publications in the period.

Table 1 indicates that the cumulative number of publications in 2007 was equivalent to 45.14% of the total. This means that more than half (54.86%) of the SBSE publications were published through 2008 and 2010. Therefore, more than 50% of the SBSE works were published in the last 30% years of the considered period, which indicates the concentration of the most advances in recent years.

A final analysis in Table 1 regards the cumulative growth acquired in 2010, which was of 26.06%. This figure is an additional indication that the SBSE field is still expanding at significant rates, even after 10 years.

Table 2 below presents a ranking of the most cited publications of the decade. The order in the list is set according to the citation count in Scopus. Additionally, the citations given by Google Scholar are also shown, in order to present a more complete citation scenario, given that Google Scholar cover publications such as books, chapters, thesis, technical reports, and other types of publication that are not present in academic databases. Since this bibliometric analysis is concerned with the time period until 2010, the data shown in Table 2 the citation statics until December 31th, 2010.

Table 2. The 15 most cited SBSE works, ordered by Scopus

Ref.	Authors	Work	Scopus	GS
[9]	McMinn, Phil	Search-based software test data generation: a survey	209	382
[10]	Wegner, J., Baresel, A. and Sthamer, H.	Evolutionary test environment for automatic structural testing	162	282
[11]	Michael, C., McGraw, G. and Schatz, M.	Generating software test data by evolution	158	298
[12]	Clarke, J. et al.	Reformulating software engineering as a search problem	90	149
[3]	Harman, M. and Jones, B.	Search-based software engineering	82	163
[13]	Harman, M. et al.	Testability transformation	77	122
[14]	Li, Z., Harman, M. and Hierons, R.	Search algorithms for regression test case prioritization	75	126
[4]	Harman, M.	The current state and future of search based software engineering	72	171
[15]	Mitchell, B. and Mancondris, S.	On the automatic modularization of software systems using the bunch tool	71	105
[16]	Dolado, J.	On the problem of the software cost function	71	128
[17]	Cohen, M. et al.	Constructing test suites for interaction testing	69	120
[18]	Tonella, P.	Evolutionary testing of classes	67	138
[19]	Greer, D. and Ruhe, G.	Software release planning: An evolutionary and iterative approach	65	119
[20]	Bagnall, A., Rayward-Smith, V. and Whittley, I.	The next release problem	63	91
[21]	Canfora, G. et al.	QoS-aware replanning of composite Web services	54	137

From the 15 works presented in Table 2, 7 are related to software testing, including the 3 most cited works. Among these 7 publications, the test data generation issue is covered in 4, including the most cited work [9]. This observation may be related to the higher amount of software testing works in the SBSE publications. Nonetheless, the fact represents the high force and impact of the software testing area in SBSE in general, and of test data generation inside software testing.

General works also occurs vastly among the most cited publications [10][4]. Additionally, the most cited work is a survey [9]. Another main publication area on Table 2 is requirement engineering [19][20].

3.2 Source

As cited in Section 22, the data of 2010 for the following segments is not complete. Nevetheless, the analysis represents the state of SBSE up to 2010 with some works of the year. In order to further improve the publication analysis presented in Section 3.1, this section shows the distribution of the works among different sources. Table 3 presents the amount of 2001-2010 works in five main publications types, and it also shows the percentage contribution of each type in relation to all publications. In Table 4, the two main sources are examined yearly.

Table 3. Number of works from 2001 through 2010 by publication type

Type	Proceeding	Journal	Book/Chapter	Tech Report	Thesis	Other
Amount	469	140	6	23	26	3
%	70.32%	20.99%	0.89%	3.45%	3.90%	0.45%

Table 4. Works in proceedings and journal, and its relation (2001-2010)

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	01-10
Proceeding	8	24	33	30	38	33	63	80	89	71	469
Journal	16	3	2	7	11	18	16	32	22	13	140
Proceeding/ Journal	0.50	3.00	16.50	4.29	3.46	1.83	3.94	3.64	4.05	5.46	3.35

From Table 3, the source with more works published in the decade is conference proceedings (70.32%), followed by journals (20.99%). This preponderance of proceeding publications compared to journal articles is also observed in the Computer Science field as a whole [22]. The SBSE publication in books, including chapters, achieved only 6 works, which represents less than 1% (0.89%). This overall behavior is also in general consonant with the one presented in the overall Computer Science field [22]. The number of of works for technical reports and thesis, in the time span analyzed, were similar (3.45% and 3.90%, respectively).

The relation between the number of works published in conference proceedings and journals is a measure to be studied. In the decade under analysis, 2001 was the only year with this figure below 1.00, i.e., with more journal articles than proceedings papers. This result reflects the publication of the special issue of Information and Software Technology, in December 2001. In the subsequent years, the relation reached levels that, in general, were more than 2.00, with the exception of 2006 (1.83). In 2003, this measure got abruptly to more than 16.00. This observation was caused by the significantly increase of publication in conference in 2003 against 2002, which was not followed by articles in journals. From 2008 onwards, the values are increasing: 3.64, 4.05, 5.46; from 2008 to 2010, respectively. This trend is partially explained by the foundation of conferences dedicated to the area, which added a large number of papers. The specialized venues include the International Workshop on Search-Based Software Testing (SBST) in 2008, and the International Symposium on Search Based Software Engineering (SSBSE) in 2009.

Despite this predominance of proceedings publications, journal articles are generally taken as fundamental contributions to a field [22]. Thus, special attention to this publication type should be given. In Table 5, the top 11 journals with the most amounts of SBSE publications are presented.

As shown in Table 3, the top 3 journals are software engineering focused. Beyond that, the majority of the 11 journals are of software engineering. Aside general software engineering venues, as the top 3, journals of specific areas are also present, including software testing and requirements engineering. As a sign of the cross-disciplinarily of SBSE, journals of Operational Research and Soft Computing are also present. Among the venues on the ranking, three journals dedicated special issues to SBSE: Information and Software Technology (2001), Computers and Operations Research (2008), and Journal of Software Maintenance and Evolution (2008).

Table 5. Ranking of journals with the most number of SBSE publications

Journal	#	%
Information and Software Technology	23	16.43%
Journal of Systems and Software	12	8.57%
IEEE Transactions on Software Engineering	10	7.14%
Computers and Operations Research	7	5.00%
Software Testing, Verification and Reliability	5	3.57%
Applied Soft Computing	4	2.86%
IEEE Transactions on Reliability	4	2.86%
Information Sciences	4	2.86%
Journal of Software Maintenance and Evolution	3	2.14%
Requirements Engineering	3	2.14%
Soft Computing	3	2.14%

Table 6 shows the distribution of SBSE articles among journals. In this case, the data shows how many journals have published each number of papers.

Table 6. Number of journals that published each article amount

Article amount	23	12	10	7	5	4	3	2	1
Journals	1	1	1	1	1	3	3	14	34

By examining the data in Table 6, the Bradford's Law [7] is applied. This bibliometric law states that the journals that publish works from a field may be grouped in three categories, each with roughly a third of the publications. The relation among the number of journals in each group is estimated to be of $1{:}n{:}n^2$, which means that among groups it is necessary to look into exponentially more journals to find the same number of articles. The first group is composed by the core journals, i.e., the main venues for the field. The next group has journals with average number of papers. The last category, the long tail, is formed by several journals, each with few publications. From Table 6, the top three journals play as core journals, since they correspond to 45 articles, which is 32.14% of all. The next group is found in the next 12 journals (41 articles, or 29.28%). In order to represent the last articles, the 43 remaining journal are necessary. Then, the Bradford relation in SBSE is 3:12:43.

3.3 Authorship

Table 7 shows statistics regarding the number of publishing authors in the SBSE field. The "New" column shows the number of authors publishing for the first time in each year, considering the time span from 2001 to 2010. The percentage participation and growth of this aspect are also presented. The "Active" column refers to the amount of different authors publishing in a given year. Finally, the numerical relation between new and active authors is indicated.

The number of authors joining the SBSE community has significantly increased each year since 2001. Therefore, beyond the strong increase in the number of publications (as shown in Table 1), the number of new authors also firmly increased. The increase in the number of publications could be merely because of regular authors. Alternatively, in fact, more authors continue to publish SBSE works for the

first time each year. Indeed, 2010 alone was responsible for more than 10% (12.32%) of the new authors in the decade. The renewing factor was even higher since 2008, with 17.25% in 2008 and 17.63% in 2009.

Table 7. Statistics for new and active authors per year (2001-2010)

Year	New	%	Growth	New/ Works	Active	Growth	New/ Active
2001	49	6.36%	-	2.04	49	-	1.00
2002	43	5.58%	-12.24%	1.43	60	22.45%	0.72
2003	54	7.00%	25.58%	1.46	83	38.33%	0.65
2004	48	6.23%	-11.11%	1.07	87	4.82%	0.55
2005	51	6.62%	6.25%	0.98	100	14.94%	0.51
2006	67	8.69%	31.37%	1.12	116	16.00%	0.58
2007	95	12.32%	41.79%	1.14	162	39.66%	0.59
2008	133	17.25%	40.00%	1.08	230	41.98%	0.58
2009	136	17.63%	2.26%	1.10	236	2.61%	0.58
2010	95	12.32%	-30.15%	1.07	168	-28.81%	0.57
2001-2010	771	100%	-	-	-	-	-

An additional study concerning the new authors is to analyze this aspect taking into account the number of works in each year. As expected, the highest value is found in 2001, since all authors, in the time span, are considered new in that year. The amount in general reduces through the years. An observation that must be highlighted is that in 2010, despite the fewer new authors, the average amount of new authors per works is similar to the presented in previous years. This also occurs in the relation between new and active authors: in 2010 there were less new authors, but they corresponded to 57% of the authors in that year, indicating the dynamism of the SBSE community.

In addition to the new authors, the number of active authors is also an aspect of interest. From Table 7, one can observe that this number has increased since 2001, with exception to 2010, when the number of active authors was 28.81% lower than in 2009. However, 2010 stands as the third year with most active authors, which is a sign of the yet strong recent activity of the SBSE field.

Using the information of new authors in each year given in Table 7, the cumulative amount per year can be calculated. The evolution of such statistic is presented in Figure 1. Until 2005, the cumulative amount achieved 245 (31.78%). This means that the last half of the decade was responsible for about 70% of the authors.

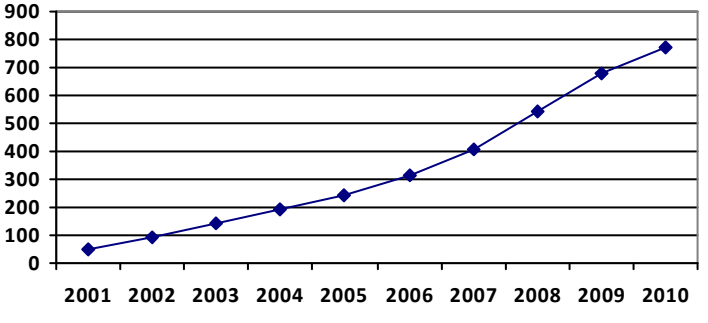

Fig. 1. Cumulative evolution of number of authors, between 2001 and 2010

In Table 8 below, the distribution of publications among authors is presented.

Table 8. Number of authors that published each number of works

Works	1	2	3	4	5	6	7	8	9	10	11	12	14	16	18	19	20	24	27	86
Authors	477	126	58	27	27	12	9	6	8	4	1	4	2	1	1	1	3	2	1	1

Authors who published only one work represent the larger group, formed of 477 researches (61.87% of the SBSE community). Researchers that published two works also add up to more than a hundred, being in total 126, or 16.34%. As the number of publication increases, in general the number of authors with that amount reduces. In fact, there is a bibliometric law to study this phenomenon. The Lotka's law indicates that the number of authors publishing n works is roughly$1/n^2$ of the amount that published one work. In fact, the exponent in the fraction is specific for each field, and also may vary over time, but number 2 is generally used. It should be highlighted that the law is only an estimate and its accuracy may depend on the scientific field under analysis and the considered time span.

As a result from Lotka's law, the expected number of authors with one publication is 60%. By the previous analysis, it can be observed that this roughly occurs in the SBSE field (61.87%). In addition, the number of authors with two publications is expected to be $(1/2^2)$*60%, equals to 15%, which also can be considered valid for the SBSE community (16.34%). For three publications, the number of authors, according to Lotka's law, should be $(1/3^2)$*60%, about 6.67%. In SBSE, there are 8.69% in this group. The number of authors with 4 publications, 27, is also cosistent to the expected value from Lotka's law (25). For 10 publications, the amount also nearly conforms to the law, with 4 authors ($1/10^2$*60% from 771). For authors with more publications, however, the Lotka's law does not apply. For 20 publications, for instance, the expected number of authors is 1, while there are 3 authors in SBSE.

Among the authors, a ranking of the most prolific ones can be formed. Table 10 indicates a list ordered with regard to the total number of publications, including conference works, journal articles, books chapters, etc. The percentage participation in relation to all SBSE publications is also given. In order to present more complete information, the amount and position of each author concerning only publications in conference proceedings and journal articles is also presented.

From Table 10, the most prolific author has participated in more than 10% (11.15%) of the SBSE publications. For the 2nd position, the percentage participation is of 3.50%, indicating a large gap between the 1st position and this one. On the other hand, from the 2nd spot onwards, the difference among the sequential positions is not so large, which indicates a smooth distribution among the authors. The first and second authors with most publications appear alone in their positions. That also happens with positions 5, 6, 7, and 10. In the other positions, some authors share the same spot: there are two authors in 3rd, three in 4th, two in 8th, and three in 9th.

Given the information in Table 10, it is possible to compare the main order, that takes into account all publications, with the specific ranking for conferences and journals. Aside from authors in 1st, 5th, and one in 9th, the spots of the authors differs among the different rankings. For instance, authors in 9th place in the total ordering appear in 3rd in journal articles.

A final analysis of Table 10 can be performed by taking information from Table 2, that concerns on the most cited publications. Among the 15 most cited publications, 10 have at least one of the most prolific authors shown in Table 9. If we consider the top 10 most cited publications, the participation of one of the most prolific authors achieves 8. This indicates that in SBSE there is a correlation between the groups of authors from the most cited publications and the most prolific authors group

Table 9. The ranking of SBSE authors with more than 10 publications in 2001-2010

Rank	Author	Works	%	Conference Proceedings (rank)	Journal Articles (rank)
1	Mark Harman	86	11.15%	61 (1)	20 (1)
2	Xin Yao	27	3.50%	19 (2)	4 (5)
3	John Clark	24	3.11%	15 (5)	7 (2)
3	Robert Mark Hierons	24	3.11%	16 (4)	7 (2)
5	Andrea Arcuri	20	2.59%	15 (5)	1 (8)
5	Joachim Wegener	20	2.59%	15 (5)	5 (4)
5	Massimiliano Di Penta	20	2.59%	17 (3)	3 (6)
8	Phil McMinn	19	2.46%	15 (5)	4 (5)
9	Giuliano Antoniol	18	2.33%	16 (4)	2 (7)
10	Enrique Alba	16	2.08%	13 (6)	3 (6)
11	Andre Baresel	14	1.82%	11 (7)	3 (6)
11	Francisco Chicano	14	1.82%	11 (7)	3 (6)
13	Günther Ruhe	12	1.56%	5 (12)	6 (3)
13	Shin Yoo	12	1.56%	7 (10)	2 (7)
13	Spiros Mancoridis	12	1.56%	9 (8)	3 (6)
13	Stefan Wappler	12	1.56%	9 (8)	1 (8)
17	Per Kristian Lehre	11	1.43%	7 (10)	1 (8)

3.4 Collaboration

After the previous study regarding the authorship in SBSE, the cooperation among the authors is discussed in this section. Firstly, Table 10 shows statistics about the number of authors per paper. Papers with more than five authors were grouped together with the ones with five authors. The yearly evolution of the percentage participation of each group is shown in Figure 2.

Table 10. Authorship pattern in the period 2001-2010

Authors	1	2	3	4	5+
# Works	87	230	209	101	40
%	13.04%	34.48%	31.33%	15.14%	6.00%

Considering the decade 2001-2010, the predominant group of collaboration is the one with 2 authors (34.48%). The next group, with 3 authors, has a similar rate (31.33%). The number of publications with one author, 87, corresponds to 13.04%,

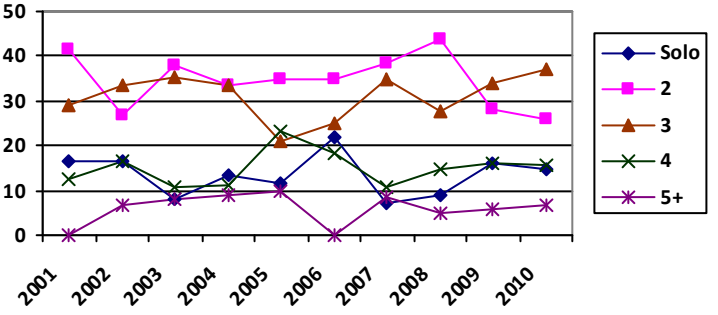

Fig. 2. Authorship percentage evolution per year

which indicates a collaboration rate of 86.96%, i.e., nearly 87% of SBSE publications have more than one author. In Computer Science, this collaboration rate reached 86% [23]. From Figure 2, the predominance of papers with 2 authors is observed through the years. Indeed, it happened in seven of the ten years. However, since 2008 the group with 2 authors is in a trend of reduction, while, simultaneously, the group with 3 authors is increasing. With the exception of 2007, the group with 5 or more author is the small one in the years.

Table 11 shows the collaborative level, i.e., the average of authors per paper, in each year, including the average for the decade. For better visualization of the evolution, the data are also presented in Figure 3.

Table 11. Collaborative level (CL) of SBSE, i.e., the average amount of authors per paper

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	01-10
CL	2.37	2.80	2.89	2.73	2.86	2.40	2.78	2.64	2.71	2.75	2.70

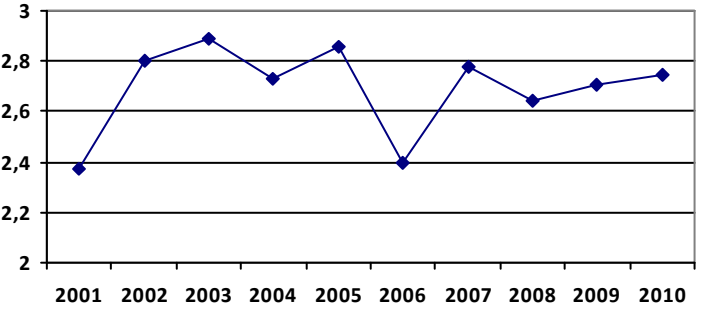

Fig. 3. Evolution of SBSE collaborative level by year (2001-2010)

As show in Table 11 and Figure 3, the average number of authors per paper strongly increases between 2001 and 2002. Until 2005, the value remained near 2.8, and in 2006 reduced to about the same level of 2001. Given the increase trend of papers with 3 authors alongside the decrease of papers with 2 authors since 2008, the average amount shows an increase trend since 2008.

Table 12. Collaboration aspect regarding authors' affiliation

Year	External	% year	Internal	% year	Out Univ.	% year
2001	1	4.17%	6	25.00%	6	25.00%
2002	3	10.00%	7	23.33%	7	23.33%
2003	4	10.81%	9	24.32%	6	16.22%
2004	8	17.78%	13	28.89%	8	17.78%
2005	6	11.54%	8	15.38%	3	5.77%
2006	7	11.67%	15	25.00%	8	13.33%
2007	12	14.46%	20	24.10%	9	10.84%
2008	22	17.89%	28	22.76%	16	13.01%
2009	20	15.38%	25	19.23%	17	13.46%
2010	6	6.82%	16	18.18%	6	6.82%
2001-2010	89	11.54%	147	19.02%	86	11.10%

The Table 12 above shows statistics about collaboration concerning authors' affiliations. The cooperation among countries occurred in more than 10% of the papers, with exception for 2001. In 2004 and 2008, the rate achieved more than 17%. Aside the percentage participation, it is important to highlight the increase in the amount of such collaboration throughout the years. The internal cooperation, i.e., with authors in different institutions of the same country, has achieved, except in 2005, more than 20%, which shows a strong internal collaboration. The other aspect counts the participation of research institutes outside universities and software companies. From Table 13, the percentage evolution of this cooperation shows that in general the rates in the last five years are lower than in the previous years: in 2001-2005, the value range from 5% to 25%; among 2006 and 2010 the rate varies from 6% to 13%. In the decade as a whole, the percentage is more than 10% (11.10%).

4 Further Analysis

In Section 4.1, we compare measures from SBSE between the period pre-2001 (1976-2000) with the 2001-2010 decade. In 4.2 estimates for the next five years are given. Due to size constraints, only some statistics are covered in this section.

4.1 SBSE Pre-2001

Table 13 regards authorship in SBSE pre-2001. During the 25 years on that period, the number of SBSE authors reached 122 against the 771 authors in the 2001-2010 decade (Table 7). The CL column indicates the collaborative level, i.e., the average number of authors per paper. The value for the pre-2001 period, 2.52, is similar to 2.70 for 2001-2010 (Table 11). In the number of authors per paper, an interesting change can be observed. Before 2001, the predominance was of 1 author (27.91%), followed by 2 authors per paper (26.74%). In the 2001-2010 time span (Table 10), the most predominant groups are "2 authors" (34.48%), and "3 authors" (31.33%).

Table 13. Authorship measures for SBSE pre-2001 (1976-2000)

Ative Authors	CL	Amount of authors per paper				
		1	2	3	4	5+
122	2.52	27.91%	26.74%	22.09%	17.44%	5.81%

4.1 Estimates

The estimation used the linear trend approach, since it was successful to estimate the values for 2010 based on the data from the previous years. Projections for the next 5 years of the number of active authors, number of works, and relation between conference proceeding papers and journal articles (C/J) are given in Table 14.

Table 14. Estimates of number of works and active authors for 2011-2015

Year	2011	2012	2013	2014	2015
Authors	179	191.9	204.8	217.7	230.6
Works	94.4	101.5	108.6	115.7	122.8
C/J	4,17	4,08	3,99	3,89	3,80

As shown in Table 14, the number of active authors is expected to achieve 200 by 2013. Such landmark occurred in 2008 and 2009 (Table 7), but the decreased observed in 2010 (168) reflects on the trend. Based on the estimations, 2011 is to have less than 100 publications. The low behavior in 2010 also explains this landmark in 2011. The relation between conference and journal publications has a decreasing trend. This indicates that the field is expected to have relatively more publications in journals.

5 Conclusions

SBSE publications have, in general, continuously increased since 2001. In 2009 however, there was a negative growth compared to 2008. In relation to all works published in 2001-2009, 2010 had a cumulative growth of 26.06%, which indicates the recent development on the field. Other sign of that expansion is that more than 50% of the publications in the decade were from the last three years. Among the most cited works, software testing is the predominant area. Among software testing works, the data generation problem stands as the most cited issue in SBSE.

Conference proceedings are the most common type of publication in SBSE. In the decade 2001-2010, the rate achieved 70.32%. Publication in books and chapters are weakly represented. The relation between conference proceeding and journal articles was 3.35 in the decade, i.e., in average 3.35 conference works for each journal article. The relation increased in the last three years of the decade. In journals that most published SBSE papers, there is a predominance of software engineering venues, but journals of operational research and soft computing are also present. The Bradford's Law seems to apply to SBSE, and the core journals for SBSE was identified as the top three venues.

The joining of new authors in SBSE is still significant. The year 2010 was responsible for 12% of the total in the decade. In the last five years of the decade, a rate of about 0.58 was observed in the participation of new authors in each year. The Lotka's Law was found to be applicable to the SBSE field in most cases. A correlation between the most prolific authors and authors with the most cited papers was also found.

The predominant number of author per paper in 2001-2010 is two. However, the evolution through the years indicates the reduction of that group and the ascending of papers with 3 authors. The collaboration among countries was observed in 11.54% of the papers. Cooperation among affiliations inside countries reached 19%, indicating the high level of cooperation in the SBSE field. The presence of authors from research institutes and software companies achieved 11.10% of publications in 2001-2010.

The comparison of the decade against the pre-2001 period shows that the predominant group of authors per paper went from 1 to 2. The collaboration level increased from 2.52 to 2.70, indicating more collaboration in 2001-2010. Estimates for the next 5 years using the linear trend show the expected continuous expansion of the field. Additionally, relatively, more journal articles are expected to be published.

Future works regard the coverage of more recent works uncovered in the last three segments, mainly from 2010. Further research includes bibliometric analysis specific on SBSE areas, such as software testing or requirements engineering. The identification of the most collaborative authors and countries may also be addressed.

Acknowledgments. The authors would like to thank Mark Harman for directing the SEBASE Project, and Yuanyuan Zhang for the organization of the SEBASE SBSE Repository.

References

1. Miller, W., Spooner, D.L.: Automatic Generation of Floating-Point Test Data. IEEE Transactions on Software Engineering 2(3), 223–226 (1976)
2. Harman, M., Jones, B.F.: The SEMINAL Workshop: Reformulating Software Engineering as a Metaheuristic Search Problem. ACM SIGSOFT Software Engineering Notes 26(6), 62–66 (2001)
3. Harman, M., Jones, B.F.: Search-based Software Engineering. Information & Software Technology 43(14), 833–839 (2001)
4. Harman, M.: The Current State and Future of Search Based Software Engineering. In: Proceedings of International Conference on Software Engineering/Future of Software Engineering 2007, Minneapolis, pp. 342–357 (2007)
5. Harman, M., Mansouri, S., Zhang, Y.: Search Based Software Engineering: A Comprehensive Analysis and Review of Trends Techniques and Applications. Technical report (2007)
6. Lotka, A.: The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences 16(12), 317–323 (1926)
7. Bradford, S.: Sources of information on specific subjects. Journal of Information Science 10(4), 173–180 (1985)
8. SEBASE SBSE Repository, `http://www.sebase.org/sbse/publications/repository.html`

9. McMinn, P.: Search-based software test data generation: a survey. Software Testing, Verification and Reliability 14(2), 105–156 (2004)
10. Wegener, J., Baresel, A., Sthamer, H.: Evolutionary Test Environment for Automatic Structural Testing. Information and Software Technology - Special Issue on Software Engineering using Metaheuristic Innovative Algorithms 43(14), 841–854 (2001)
11. Michael, C.C., McGraw, G., Schatz, M.A.: Generating Software Test Data by Evolution. IEEE Transations on Software Engineering 27(12), 1085–1110 (2001)
12. Clarke, J., Dolado, J.J., Harman, M., Hierons, R.M., Jones, B., Lumkin, M., Mitchell, B., Mancoridis, S., Rees, K., Roper, M., Shepperd, M.J.: Reformulating Software Engineering as a Search Problem. IEE Proceedings Software 150(3), 161–175 (2003)
13. Harman, M., Hu, L., Hierons, R., Wegener, J., Sthamer, H., Baresel, A., Roper, M.: Testability Transformation. IEEE Transaction on Software Engineering 30(1), 3–16 (2004)
14. Li, Z., Harman, M., Hierons, R.M.: Search Algorithms for Regression Test Case Prioritization. IEEE Transactions on Software Engineering 33(4), 225–237 (2007)
15. Mitchell, B., Mancoridis, S.: On the Automatic Modularization of Software Systems using the Bunch Tool. IEEE Transactions on Software Engineering 32(3), 193–208 (2006)
16. Dolado, J.J.: On the Problem of the Software Cost Function. Information and Software Technology 43(1), 61–72 (2001)
17. Cohen, M.B., Gibbons, P.B., Mugridge, W.B., Colbourn, C.J.: Constructing Test Suites for Interaction Testing. In: Proceedings of the 25th International Conference on Software Engineering, Portland, pp. 38–48 (2003)
18. Tonella, P.: Evolutionary Testing of Classes. In: Proceedings of the 2004 ACM SIGSOFT International Symposium on Software Testing and Analysis, Boston, pp. 119–128 (2004)
19. Greer, D., Ruhe, G.: Software Release Planning: An Evolutionary and Iterative Approach. Information & Software Technology 46(4), 243–253 (2004)
20. Bagnall, A.J., Rayward-Smith, V.J., Whittley, I.M.: The Next Release Problem. Information and Software Technology 43(14), 883–890 (2001)
21. Canfora, G., Di Penta, M., Esposito, R., Villani, M.L.: QoS-Aware Replanning of Composite Web Services. In: Proceedings of 2005 IEEE International Conference on Web Services, Orlando, pp. 121–129 (2005)
22. Franceschet, M.: The role of conference publications in CS. Communications of the ACM 53(12) (2010)
23. Franceschet, M., Constantine, A.: The effect of scholar collaboration on impact and quality of academic papers. Journal of Informetrics 4(4) (2010)