

A Systematic Review of Interaction in Search-Based Software Engineering — *Review Protocol* —

Aurora Ramírez*, José Raúl Romero* and Christopher L. Simons†

November 5, 2017

Abstract

This document is provided as additional material for the paper “A Systematic Review of Interaction in Search-Based Software Engineering”, where the role of the human software engineer interaction within search-based approaches is reviewed. Here, the followed review protocol is described in detail, including the search strings and databases used to search candidate papers and the inclusion/exclusion criteria applied to select primary studies. The steps carried out to conduct the data extraction process are described too.

1 Introduction

The design and development of complex, large-scale software systems can be non-trivial and challenging for the software engineer to perform. In an attempt to assist him/her, formulating software development activities as optimisation problems or knowledge discovery has enabled the application of computationally intelligent search, involving, for example, a range of metaheuristic search approaches. Such search-based software engineering (SBSE) [1] approaches have attracted significant research attention in recent years. Indeed, attempts have been made to support the software engineer in many cognitively challenging development activities by applying SBSE approaches across a range of lifecycle activities, e.g. see [2].

The motivation of this review is founded on the following aspects:

- a recent increasing research interest in interactive SBSE resulting in a wide range of search techniques and human interaction mechanisms;
- the apparent lack of an up-to-date comprehensive picture i.e. to the best of our knowledge, there has been no systematic review of interactive SBSE conducted hitherto, and;

*Dept. of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain. Email: {aramirez, jrromero}@uco.es

†Dept. of Computer Science and Creative Technologies, University of the West of England, Bristol, United Kingdom. Email: chris.simons@uwe.ac.uk

- a desire to enable researchers to develop interactive SBSE approaches based on best available knowledge.

The purpose of this document is to fully describe the review methodology and the evaluation process followed to conduct a systematic literature review (SLR) on interactive search-based software engineering (iSBSE). This protocol follows the recommendations of recognized guidelines to conduct SLRs within the Software Engineering field [3, 4]. Furthermore, Annex A provides a detailed description of the search queries and the results obtained after their execution in different search engines in conformance to this review protocol.

2 Formulation of research questions

The following research questions (RQ) have been identified:

RQ1: In what ways has interactivity been adopted within search-based software engineering?

RQ2: Which findings about search techniques and interactive approaches along the complete development cycle can be extracted from the current state-of-the-art?

RQ3: To what extent do the detected gaps hamper human interaction in SBSE?

RQ4: What are the emerging trends and how might they be addressed in the future?

3 Selection of primary studies

3.1 Search strategy

List of possible terms. Terms have been extracted from the title and keywords of well-known papers in the area, and complemented with synonyms and other similar words.

- Interactivity, interactive, user-centered/centred, user-interaction, human-in-the-loop.
- Search-based Software Engineering, Search Based Software Engineering.
- optimization/optimisation.

Additionally, the ACM Computing Classification System [5] and the IEEE Taxonomy [6] have been consulted to derive a list of phases within the software life cycle: requirements, architecture, design, development, testing, debugging, verification, maintenance and evolution.

Search strings. For automated search, and having considered the aforementioned terms, the following two search strings have been constructed to execute the queries.

First search string:

```
(interactivity OR interactive OR human-in-the-loop OR
user-interaction OR user-centered OR user-centred) AND
(search-based OR "search based") AND "software engineering"
```

Second search string:

```
(interactivity OR interactive OR human-in-the-loop OR
user-interaction OR user-centered OR user-centred) AND software
AND (requirements OR architecture OR design OR development OR
testing OR debugging OR verification OR maintenance OR evolution)
AND (search OR search-based OR optimisation OR optimization)
```

Within the manuscript content, only the following fields are queried:

- Title
- Abstract
- Keywords

3.2 Databases and data sources for automated search

3.2.1 Data sources with specific references on the field

SBSE repository: http://crestweb.cs.ucl.ac.uk/resources/sbse_repository

3.2.2 Digital libraries

ACM Library: <http://dl.acm.org>

IEEE Xplore: <http://ieeexplore.ieee.org/Xplore/home.jsp>

ScienceDirect: <http://www.sciencedirect.com>

SpringerLink: <https://www.springer.com/la/librarians/librarian-tools>

3.2.3 Citation databases

Scopus: <https://www.scopus.com>

ISI Web of Knowledge: <https://www.webofknowledge.com>

3.2.4 Citation search engines

DBLP: <http://dblp.uni-trier.de>

Google Scholar: <https://scholar.google.com>

Additionally, general search engines like Google (<http://www.google.com>) can be used to collect grey literature.

3.3 Data sources for manual search

For manual searches, once the automated queries have been executed, the list of authors, and the most relevant journals and conferences, will be extracted. This list will be used to refine the results by querying journal and conferences websites and private websites from the authors.

Additionally, technical reports and additional materials will be collected from publicly available sources like <https://arxiv.org>, if these materials are properly cited in a relevant peer-reviewed published work. These materials will not be cited in the review, but can be used as a data source for collecting additional references.

3.4 Study selection process

3.4.1 Inclusion criteria

Main criterion. The paper should propose an interactive search-based approach to solve a Software Engineering (SE) task. It can be decomposed into the following items/conditions:

1. The search model explicitly incorporates a user, who is required to perform at least one interaction to solve a task (evaluation, selection, ...) during the search process. Another valid approach is that the user participates once the search has finished but only if another iteration/algorithm is executed after that in order to integrate user's opinion in the overall process.
2. The paper should describe a methodology regarding the user's interaction. It should explain the role of the human in the search process, how his/her feedback is then integrated to the search process, or any other relevant information concerning the interactive action.
3. The search problem has to be defined in terms of a decision space, evaluation objectives and existing constraints, if any. The problem formulation can be either single-objective or multi-objective, or both.
4. The search problem should be (easily) framed in one or several phases of the software development life cycle (design, testing, maintenance...).
5. There is no restriction regarding the search technique applied to address the Software Engineering problem, except that it should be a computational method.
6. There is no restriction regarding the nature of the information provided by the user (quantitative, qualitative, weights...), but it should cause an effect on the search.
7. Theoretical proposals are also permitted.

At least all the following criteria must be satisfied to be included as a primary study: 1, 2, 4, 5 and 6.

Managing multiple variants of a same work. In case that the same authors have published several papers for the same interactive problem, the paper would be considered as a primary study for review only under the following conditions:

- If the problem is the same but the technique is different and novel, accept.

- If both the problem and the technique are the same but the authors report different findings with a new significant type of experimentation, accept.
- If there is a journal paper (satisfying inclusion criteria) as an extension of a previous conference paper, then reject the conference paper unless it provides different but significantly distinct experimental results.

Excluded variants (as primary studies) of the same work will be positively considered for counting and statistical purposes.

3.4.2 Exclusion criteria

1. The interactive algorithm does not solve a SE-specific optimisation problem.
2. The user only participates to (re)configure the parameters of the search algorithm before or after an independent run or he/she just acts as a decision maker at the end of the process, i.e. selecting a solution from the resulting set in multi-objective optimisation.
3. Papers written in a language other than English are discarded.
4. The full text of the manuscript cannot be reached.
5. There is no evidence that the publication has followed an accurate scientific peer-review rating process. More specifically, journals should appear in the Journal Citation Report, while conferences should be ranked in the CORE system¹.

4 Quality assessment process

4.1 Information gathering

For each paper under review, a number of data will be collected in the corresponding “Data Extraction Form”. It will contain information about the publication and its content. More specifically, each manuscript will be scrutinised to gather information about the problem formation, the search technique, the interactive approach, and the scope of the study.

For a detailed classification, see the document “Guide for the classification of interactive search-based studies” provided as additional material together with this review protocol.

4.1.1 Information gathering on the manuscript

- List of authors, including their name and affiliation.
- Title.
- Type of publication (journal, conference, workshop, technical report).
- Name of the journal/conference/workshop.

¹The only exception is the Symposium on Search Based Software Engineering, as it is a reference publication venue for the community.

- Publisher.
- Year of publication.

4.1.2 Information gathering on the problem formulation

- Type of software problem.
- Relation to any development practice e.g. agile, where apparent.
- Number of objectives if a multi-objective problem.
- Consideration of constraints.

4.1.3 Information gathering on the search technique

- Type of algorithm
- Single/Multi-objective approach².

4.1.4 Information gathering on the interactive approach

- Type of interactive algorithm.
- Type of feedback/information obtained from the user.
- Evaluation mechanism in terms of standard/non-standard metrics, scores, etc.
- About the time of user's intervention:
 - About the solutions shown.
 - Adjustment of the specific interaction time.
- Influence of the user's opinion along the search.

4.1.5 Information gathering on the experimental framework

- Type of study.
- Users' profile.
- Evaluation criteria.
- Case studies.
- Evidences publicly available about the information collected.
- Statistical tests.

4.2 Definition of quality instruments

Score mechanisms are not considered to evaluate the quality of each primary study.

²Notice that this item is not referring to the specific number of objectives in the problem formulation.

4.2.1 Managing missing data

There are multiple choices if missing information is in a paper: infer the information, contact the authors or omit this information.

In the review, all the papers will be fairly, objectively reviewed under the same conditions. Consequently, neither information will be presumed nor some authors will be contacted, since a potential response from all the authors cannot be guaranteed and, probably, it might affect the accuracy of the information collected from some papers. Thus missing information will be reflected in the paper but not considered.

4.2.2 Data extractors and role

1. Each paper is read by 2 people, and extracted data compared. If disagreement is significant, a third person will also read the paper.
2. After the data extraction is complete, all the people will read all the extraction forms and agree for a final data extraction report for each paper.

Data extractors will follow the instructions indicated by the “Data extraction form”.

4.2.3 Resolving agreements and disagreements between data extractors

If there is a conflict between two data extractors, agreements and disagreements are documented. Then, all three data extractors may comment and vote.

5 Data extraction

5.1 Procedure

The three authors will participate in the data extraction process. Paper distribution will be noted in the “Candidate papers form”.

1. Search for papers using the search strings defined in Section 3.1.
 - (a) **Outcome.** List of papers from each database and data source.
2. Obtain a unique global list of papers by removing duplicated entries and irrelevant papers (title, keywords, abstract).
 - (a) **Outcome.** List of candidate papers.
3. Apply inclusion and exclusion criteria.
 - (a) **Outcome.** List of selected primary studies.
4. Extract data from each primary study using the “Data extraction form”.
 - (a) **Outcome.** Narrative synthesis of the paper highlighting the most relevant findings.
5. Screen the paper for potential missing references in Step 2. If any, apply Step 3 to the potential paper. If succeed, add to the list of selected primary studies.
 - (a) **Outcome.** List of selected primary studies.

5.2 Extraction strategy

Every step of the extraction process will be recorded using forms and spreadsheets. Different versions will be conveniently named and the final forms will be made publicly available. The following documents will be created during the process:

- Search results: list of papers found per database (see Annex A).
- Candidate papers: list of papers after applying the successive filters.
- Data extraction: information extracted for each primary study by the 2 extractors and the final decision for each field of the classification scheme (see Section 4.1).
- Final data extraction: summary of results in tabular form, so simple statistics can be obtained.
- Descriptive statistics: statistics about the authors, type of publication and publication year.
- Disagreements: information about the observed disagreements.

5.3 Validation process

1. Before collecting the search results:
 - (a) To reach an agreement regarding the Review protocol [discussion meeting].
 - (b) To check a pilot search to validate the search strings, all reviewers performing this task independently.
2. After collecting the search results, to check that the list of search results is complete, i.e. it includes all the distinct papers returned by each different database [all reviewers].
3. After selecting the primary studies:
 - (a) Having the inclusion and exclusion criteria been evaluated by more than one reviewer, to check inconsistencies and solve disagreements [discussion meeting].
 - (b) To perform data extraction to a random sample of primary studies (up to 3 papers) and check that the “data extraction form” is consistently valid, i.e. all the reviewers interpret its fields in the same way.
4. After extracting all the data from primary studies:
 - (a) To extract the relevant information of each paper, and a third reviewer to double-check both forms in order to solve disagreements.
 - (b) To check the final information extracted from each study and to solve any disagreement. If necessary, to meet for discussion.

6 Data synthesis

- Quantitative summary of manuscripts. Information like the year of publication, authors and publication type can be summarised to determine the scope of iSBSE. Tables and graphics can be used.
- Quantitative summary of results. Results can be quantitatively described for data fields such as types of techniques, cycle phase, empirical studies, etc. Tables showing percentages will be mostly used.
- Cross-validation of specific characteristics. References can be organised into tables and frequency plots to show whether different characteristics frequently appears together, as well as to identify existing gaps. It should be focused on the most interesting characteristics (interactive approach, experimental framework). Narrative text and basic statistics can be considered to summarise our findings.
- Qualitative summary of results. It can be regarded as a “Discussion section” containing:
 - Summary of results.
 - Principal findings, meaning of findings.
 - Strengths and weaknesses.
 - Recommendations.

A Results from search queries

A.1 General information

This annex summarises the outcomes obtained from each database, the limitations of the search engines and how specific problems when conducting the search have been solved:

- Queries have been conducted between 18th and 26th of July 2016.
- Queries have been revised on 9th of January 2017 to include published papers between July and December 2016.
- Queries have been revised on 8th of August 2017 to include additional terms in Search String #2.
- Results have been exported to BIBTEX files with an representative name: YYMMDD-Database-SearchStringCode. Example: 160718-ACM-1-1.bib.
- Basic information about each search result (author, title, publication, year) has been saved in the corresponding spreadsheet (*isbse-search-results.xlsx*).

A.2 SBSE repository

Webpage: http://crestweb.cs.ucl.ac.uk/resources/sbse_repository

Date: 18th July 2016

Field: Global Quick Search

Table 1: Search results from SBSE repository

ID	Query	# Papers
1	interactivity	2
2	interactive	30
3	human-in-the-loop	1
4	user-interaction	1
5	user-centered	2
6	user-centred	0
Total (distinct papers)		34

A.3 ACM Library

Webpage: <http://dl.acm.org>

Date: 18th (search string #1) and 23th (search string #2) July 2016.

Field: Abstract

Notes: The queries have been adapted according to the following observations:

- The search has been restricted to “The ACM Guide to Computing Literature”.
- This database can return complete proceedings, which have been discarded before computing any inclusion/exclusion criteria.
- ACM does not support OR/AND syntax. In a given line, the option “matches all” is equivalent to “AND”, whereas “matches any” is equivalent to “OR”.
- It is not clear whether each line in the advanced search field is considered as “AND” or “OR” (it might depend on the use of “matches all”/“matches any”).
- Special characters (*) are not supported and words using “-” are split sometimes. In other situations, “-” is not considered. For example, the engine does not distinguish between “search-based software engineering” and “search based software engineering”. Other example: “user-interaction” returns results where only “interaction” appears if a field different from “abstract” is chosen.
- This database does not support more than one field (abstract, title, keywords).
- Matching terms are highlighted but they do not always correspond to the search terms. For example, “interactive+“search-based software engineering” returns papers where the term “interaction” appears instead of “interactive”.

Search string transformations: The search strings have been modified as follows:

String #1 should be converted into N independent “matches all” combinations:

- #1.1: interactivity “search based software engineering”. Removed items: 3 proceedings, 1 duplicated result (Marculescu et al., 2013).
- #1.2: interactive “search based software engineering”. Removed items: 3 proceedings, 1 duplicated result (Marculescu et al., 2013).
- #1.3: “human-in-the-loop” “search based software engineering”.
- #1.4: “user-interaction” “search based software engineering”.
- #1.5: “user-centered” “search based software engineering”.
- #1.6: “user-centred” “search based software engineering”.

For string #2, expanding all the possible combinations seems to be impractical. Some frequent combinations have been considered. For instance, it seems that interactive and interactivity provide the same results for this database, so only the first one will be used. The search string has been split for each software phase and only the term “search” is used because it is the most generic one:

- #2.1: interactive search “software requirements”.
- #2.2: interactive search “software design”. Removed items: 1 duplicated result (Maltz, 2013), 1 proceedings.
- #2.3: interactive search “software testing”. Removed items: 3 proceedings, 3 duplicated results (Petke, 2015; Marculescu et al., 2013; Harrold et al., 1998).
- #2.4: interactive search “software maintenance”.
- #2.5: interactive search “software architecture”.
- #2.6: interactive search “software development”.

- #2.7: interactive search “software debugging”.
- #2.8: interactive search “software verification”.
- #2.9: interactive search “software evolution”.

Table 2: Search results from ACM Library

ID	Query	# Papers
1.1	recordAbstract:(+interactivity + “search based software engineering”)	8
1.2	recordAbstract:(+interactive + “search based software engineering”)	8
1.3	recordAbstract:(+ “human-in-the-loop” + “search based software engineering”)	0
1.4	recordAbstract:(+ “user-interaction” + “search based software engineering”)	1
1.5	recordAbstract:(+ “user-centered” + “search based software engineering”)	0
1.6	recordAbstract:(+ “user-centred” + “search based software engineering”)	0
2.1	recordAbstract:(+interactive + search + “software requirements”)	9
2.2	recordAbstract:(+interactive + search + “software design”)	27
2.3	recordAbstract:(+interactive + “search based software testing”)	6
2.4	recordAbstract:(+interactive + “search based software maintenance”)	2
2.5	recordAbstract:(+interactive + “search based software architecture”)	6
2.6	recordAbstract:(+interactive + “search based software development”)	14
2.7	recordAbstract:(+interactive + “search based software debugging”)	0
2.8	recordAbstract:(+interactive + “search based software verification”)	0
2.9	recordAbstract:(+interactive + “search based software evolution”)	0
Total (distinct papers)		86

A.4 DBLP

Webpage: <http://dblp.uni-trier.de/>

Date: 18th July 2016

Field: “Search for publications” (it seems that only the title is considered as search field)

Notes: The queries have been adapted due to the following observations:

- Instructions from the webpage:
 - case-insensitive prefix search: default. e.g., sig matches “SIGIR” as well as “signal”.

- exact word search: append dollar sign (\$) to word. e.g., graph\$ matches “graph”, but not “graphics”.
- phrase search: connect words by a dot (.). e.g., inform.retriev.tech.
- boolean and: separate words by space. e.g., codd model.
- boolean or: connect words by pipe symbol (|). e.g., graph|network.
- boolean not: prepend word by minus sign (-). e.g., knuth -don.
- Problems with the use of connectors and special characters:
 - The use of “AND” and “phrase search” seems to be incompatible.
 - The use of “AND”/“OR” and more than one word with “-” (or one word with two “-”) seems to be incompatible.
 - The user of too many “AND” and words with “-” does not provide any results.

Search string transformations: The search strings have been modified as follows:

String #1 after applying the transformations becomes:

- #1.1: interactivity|interactive search based software engineering
- #1.2: interactivity|interactive search software
- #1.3 human-in-the-loop search software
- #1.4 user-interaction search software
- #1.5 user-centered search software
- #1.6 user-centred search software

Similar to ACM library, only the most frequent combinations have been considered for string #2. Notice that the search is limited to publication title, so generic terms and too long strings would probably not work well.

- #2.1: interactivity|interactive search software requirements
- #2.2: interactivity|interactive optimization|optimisation software requirements
- #2.3: interactivity|interactive search software design
- #2.4: interactivity|interactive optimization|optimisation software design
- #2.5: interactivity|interactive search software testing
- #2.6: interactivity|interactive optimization|optimisation software testing
- #2.7: interactivity|interactive search software maintenance
- #2.8: interactivity|interactive optimization|optimisation software maintenance
- #2.9: interactivity|interactive search software architecture
- #2.10: interactivity|interactive optimization|optimisation software architecture
- #2.11: interactivity|interactive search software development
- #2.12: interactivity|interactive optimization|optimisation software development
- #2.13: interactivity|interactive optimization|optimisation software debugging
- #2.14: interactivity|interactive search software debugging

- #2.15: interactivity|interactive optimization|optimisation software verification
- #2.16: interactivity|interactive search software verification
- #2.17: interactivity|interactive optimization|optimisation software evolution
- #2.18: interactivity|interactive search software evolution

Table 3: Search results from DBLP

ID	Query	# Papers
1.1	interactivity interactive search based software engineering	1
1.2	interactivity interactive search software	14
1.3	human-in-the-loop search software	0
1.4	user-interaction search software	0
1.5	user-centered search software	2
1.6	user-centred search software	0
2.1	interactivity interactive search software requirements	1
2.2	interactivity interactive optimization optimisation software requirements	0
2.3	interactivity interactive search software design	7
2.4	interactivity interactive optimization optimisation software design	5
2.5	interactivity interactive search software testing	5
2.6	interactivity interactive optimization optimisation software testing	0
2.7	interactivity interactive search software maintenance	0
2.8	interactivity interactive optimization optimisation software maintenance	0
2.9	interactivity interactive search software architecture	0
2.10	interactivity interactive optimization optimisation software architecture	1
2.11	interactivity interactive search software development	0
2.12	interactivity interactive optimization optimisation software development	0
2.13	interactivity interactive optimization optimisation software debugging	0
2.14	interactivity interactive search software debugging	0
2.15	interactivity interactive optimization optimisation software verification	0
2.16	interactivity interactive search software verification	0
2.17	interactivity interactive optimization optimisation software evolution	1
2.18	interactivity interactive search software evolution	2
Total (distinct papers)		20

A.5 IEEE Xplore

Webpage: <http://ieeexplore.ieee.org/Xplore/home.jsp>

Date: 19th July 2016

Field: Metadata only

Notes: Instructions for command search:

- Search operators: AND, OR, NOT, NEAR, ONEAR
- Search: Metadata only
- There is a maximum of 15 search terms.

Search string transformations: String #2 is too long for this search engine and should be split. An independent search is conducted for each software phase. Common part of the query:

```
(interactivity OR interactive OR human-in-the-loop OR
user-interaction OR user-centered OR user-centred) AND (search
OR search-based OR optimization OR optimisation) AND...
```

- 2.1: ... “software requirements”. Removed items: 1 (IEEE Standard Glossary of Software Engineering Terminology).
- 2.2: ... “software design”. Removed items: 1 (IEEE Standard Glossary of Software Engineering Terminology).
- 2.3: ... “software testing”. Removed items: 1 (Proceedings Symposium SBSE 2010).
- 2.4: ... “software maintenance”. Removed items: 1 (IEEE Standard Glossary of Software Engineering Terminology).
- 2.5: ... “software architecture”. Removed items: 1 (Result from 2017).
- 2.6: ... “software development”. Removed items: 3 (IEEE Standard Glossary of Software Engineering Terminology, result from 2017, 1 conference proceedings).
- 2.7: ... “software debugging”.
- 2.8: ... “software verification”.
- 2.9: ... “software evolution”.

A.6 ISI Web of Knowledge

Webpage: <https://www.webofknowledge.com>

Date: 19th July 2016

Field: Topic

Table 4: Search results from IEEE Xplore

ID	Query	# Papers
1	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search-based OR “search based”) AND “software engineering”	13
2.1	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimization OR optimisation) AND “software requirements”	1
2.2	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimization OR optimisation) AND “software design”	22
2.3	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimization OR optimisation) AND “software testing”	21
2.4	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimization OR optimisation) AND “software maintenance”	24
2.5	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimization OR optimisation) AND “software architecture”	35
2.6	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimization OR optimisation) AND “software development”	28
2.7	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimization OR optimisation) AND “software debugging”	1
2.8	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimization OR optimisation) AND “software verification”	1
2.9	(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimization OR optimisation) AND “software evolution”	2
Total (distinct papers)		122

Notes: Information from the webpage:

- Boolean connectors: AND, OR, NOT, SAME, NEAR
- Field Tags: Topic (TS), Title (TI), Author (AU), Author Identifiers (AI), Group Author (GP), Editor (ED), Publication Name (SO), DOI (DO), Year Published (PY), Address (AD), Research Area (SU), ISSN/ISBN (IS).

Search string transformations: String #2 returns an excessive number of non-related references. It might be caused by the length of the string and possible conflicts between logical operators. The following template is used instead:

```
(interactivity OR interactive OR human-in-the-loop OR
user-interaction OR user-centered OR user-centred) AND (search
OR search-based OR optimisation OR optimization) AND ...
```

- 2.1: ... “software requirements”.

- 2.2: ... “software design”.
- 2.3: ... “software testing”.
- 2.4: ... “software maintenance”.
- 2.5: ... “software architecture”. Removed items: 2 (results from 2017).
- 2.6: ... “software development”. Removed items: 2 (results from 2017).
- 2.7: ... “software debugging”.
- 2.8: ... “software verification”.
- 2.9: ... “software evolution”. Removed items: 2 (results from 2017).

Table 5: Search results from ISI Web of Knowledge

ID	Query	# Papers
1	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search-based OR "search based") AND "software engineering") SU=Computer Science	15
2.1	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimisation OR optimization) AND "software requirements") SU=Computer Science	6
2.2	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimisation OR optimization) AND "software design") SU=Computer Science	60
2.3	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimisation OR optimization) AND "software testing") SU=Computer Science	9
2.4	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimisation OR optimization) AND "software maintenance") SU=Computer Science	18
2.5	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimisation OR optimization) AND "software architecture") SU=Computer Science	60
2.6	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimisation OR optimization) AND "software development") SU=Computer Science	69
2.7	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimisation OR optimization) AND "software debugging") SU=Computer Science	0
2.8	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimisation OR optimization) AND "software verification") SU=Computer Science	2
2.9	TS=((interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND (search OR search-based OR optimisation OR optimization) AND "software evolution") SU=Computer Science	1
Total (distinct papers)		207

A.7 ScienceDirect

Webpage: <http://www.sciencedirect.com>

Date: 19th (search string #1) and 24th (search string #2) July 2016

Field: Abstract, Keywords, Title

Notes: The following items have been selected:

- Refine your search: journals, books
- Research field: Computer Science
- Time period: All years

It should be noted that the advanced search only allows using two text field (connected by AND). Nesting logic connectors inside one of them does not have the desired effect.

Search string transformations: The search strings have been modified as follows:

Text fields for search string #1:

- Text field #1: interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred
- Text field #2: “search-based software engineering” OR “search based software engineering”

For search string #2, the “Expert Search” accepts the complete text, but it provides 3.063 results. Here, it is better to use the “Advanced search” and combine the first text field with a second including as much terms as possible but connected by “OR”:

Text field #1: interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred

- 2.1: Text field #2: “software requirements selection” OR “software requirement prioritization” OR “software requirements prioritisation”.
- 2.2: Text field #2: “search-based software design” OR “search based software design”.
- 2.3: Text field #2: “search-based software testing” OR “search based software testing”.
- 2.4: Text field #2: “search-based software maintenance” OR “search based software maintenance”.
- 2.5: Text field #2: “search-based software architecture” OR “search based software architecture”.
- 2.6: Text field #2: “search-based software development” OR “search based software development”.
- 2.7: Text field #2: “search-based software debugging” OR “search based software debugging”.

- 2.8: Text field #2: “search-based software verification” OR “search based software verification”.
- 2.9: Text field #2: “search-based software evolution” OR “search based software evolution”.

Table 6: Search results from ScienceDirect

ID	Query	# Papers
1	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“search-based software engineering” OR “search based software engineering”)[All Sources(Computer Science)]	34
2.1	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“software requirements selection” OR “software requirements prioritization” OR “software requirements prioritisation”)[All Sources(Computer Science)].	3
2.2	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“search-based software design” OR “search based software design”)[All Sources(Computer Science)].	2
2.3	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“search-based software testing” OR “search based software testing”)[All Sources(Computer Science)].	9
2.4	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“search-based software maintenance” OR “search based software maintenance”)[All Sources(Computer Science)].	0
2.5	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“search-based software architecture” OR “search based software architecture”)[All Sources(Computer Science)].	0
2.6	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“search-based software development” OR “search based software development”)[All Sources(Computer Science)].	0
2.7	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“search-based software debugging” OR “search based software debugging”)[All Sources(Computer Science)].	0
2.8	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“search-based software verification” OR “search based software verification”)[All Sources(Computer Science)].	0
2.9	TITLE-ABSTR-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) and TITLE-ABSTR-KEY(“search-based software evolution” OR “search based software evolution”)[All Sources(Computer Science)].	0
Total (distinct papers)		43

A.8 Scopus

Webpage: <https://www.scopus.com>

Date: 20th July 2016

Field: Article Title, Abstract, Keywords

Observations: The following items are also selected to restrict the search:

- Document Type: All
- Date range: All years to Present
- Subject Areas: Physical Sciences

3 results for search string #1 are discarded (conference proceedings).

Search string transformations: The complete string #2 can be used in the search field but it provides 752 results. The alternative is to proceed as with other databases:

```
(interactivity OR interactive OR human-in-the-loop OR
user-interaction OR user-centered OR user-centred) AND (search
OR search-based OR optimisation OR optimization) AND ...
```

- 2.1: ...“software requirements”. Removed items: 2 (conference proceedings)
- 2.2: ...“software design”. Removed items: 6 (conference proceedings).
- 2.3: ...“software testing”.
- 2.4: ...“software maintenance”. Removed items: 1 (conference proceedings).
- 2.5: ...“software architecture”. Removed items: 9 (conference proceedings, duplicated entries, results from 2017).
- 2.6: ...“software development”. Removed items: 39 (conference proceedings, duplicated entries, results from 2017).
- 2.7: ...“software debugging”.
- 2.8: ...“software verification”.
- 2.9: ...“software evolution”.

A.9 SpringerLink

Webpage: <https://www.springer.com/la/librarians/librarian-tools>

Date: 25th July 2016

Field: Global search (“search by”)

Notes: The queries have been adapted according to the following observations:

- Tips for advanced search (copied from the webpage):

Table 7: Search results from Scopus

ID	Query	# Papers
1	TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("search-based software engineering" OR "search based software engineering") AND (LIMIT-TO(SUBJAREA,"COMP"))	23
2.1	(TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("software requirements") AND TITLE-ABS-KEY(search OR search-based OR optimisation OR optimization)) AND (LIMIT-TO(SUBJAREA,"COMP"))	4
2.2	(TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("software design") AND TITLE-ABS-KEY(search OR search-based OR optimisation OR optimization)) AND (LIMIT-TO(SUBJAREA,"COMP"))	91
2.3	(TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("software testing") AND TITLE-ABS-KEY(search OR search-based OR optimisation OR optimization)) AND (LIMIT-TO(SUBJAREA,"COMP"))	34
2.4	(TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("software maintenance") AND TITLE-ABS-KEY(search OR search-based OR optimisation OR optimization)) AND (LIMIT-TO(SUBJAREA,"COMP"))	16
2.5	(TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("software architecture") AND TITLE-ABS-KEY(search OR search-based OR optimisation OR optimization)) AND (LIMIT-TO(SUBJAREA,"COMP"))	44
2.6	(TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("software development") AND TITLE-ABS-KEY(search OR search-based OR optimisation OR optimization)) AND (LIMIT-TO(SUBJAREA,"COMP"))	56
2.7	(TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("software debugging") AND TITLE-ABS-KEY(search OR search-based OR optimisation OR optimization)) AND (LIMIT-TO(SUBJAREA,"COMP"))	1
2.8	(TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("software verification") AND TITLE-ABS-KEY(search OR search-based OR optimisation OR optimization)) AND (LIMIT-TO(SUBJAREA,"COMP"))	4
2.9	(TITLE-ABS-KEY(interactivity OR interactive OR human-in-the-loop OR user-interaction OR user-centered OR user-centred) AND TITLE-ABS-KEY("software evolution") AND TITLE-ABS-KEY(search OR search-based OR optimisation OR optimization)) AND (LIMIT-TO(SUBJAREA,"COMP"))	3
Total (distinct papers)		218

- The search function is not case sensitive. A query “zuze” or “Zuse” produces the same result.
- The query searcher for complete words.
- The “*” is used for wildcards. A query for “Medic*” lists all matches on words starting with “Medic”.
- The ISBN may be entered with or without the “-”.
- Author names are listed by “Last name, F.” in the database.
- Other points to be taken into account:
 - Logical connectors (AND, OR) are not recognised.
 - “Search based software engineering” is equals to “search-based software engineering”.
- Additional items have been selected:

Table 8: Search results from SpringerLink

ID	Query	# Papers
1.1	interactivity “search based software engineering”	2
1.2	interactive “search based software engineering”	64
1.3	human-in-the-loop “search based software engineering”	9
1.4	user-interaction “search based software engineering”	15
1.5	user-centered “search based software engineering”	2
1.6	user-centred “search based software engineering”	0
2.1	interactive “software requirements selection”	2
2.2	interactive “software requirements priorit*”	50
2.3	interactive “search based software design”	0
2.4	interactive “search based software testing”	13
2.5	interactive “search based software maintenance”	4
2.6	interactive “search based software architecture”	0
2.7	interactive “search based software development”	1
2.8	interactive “search based software debugging”	0
2.9	interactive “search based software verification”	0
2.10	interactive “search based software evolution”	0
Total (distinct papers)		124

- Content type: journals, online available, online content (PDFs).
- Topics: Computer Science.
- Language: English.
- Search Field: “Search by” (not restricted to title).

Search string transformations: The search strings have been modified as follows:

Search string #1:

- 1.1: interactivity “search based software engineering”.
- 1.2: interactive “search based software engineering”. Removed items: 1 (“news”).
- 1.3 human-in-the-loop “search based software engineering”.
- 1.4 user-interaction “search based software engineering”.
- 1.5 user-centered “search based software engineering”.
- 1.6 user-centred “search based software engineering”.

Search string #2: The combination of words without logic connectors provides thousands of results (it seems that “OR” is applied). Therefore, the more practical solution is to consider specific combinations of words as in other databases:

- 2.1: interactive “software requirements selection”.
- 2.2: interactive “software requirements priorit*”. Removed items: 4 (scientific programme-abstracts, alphabetical listing, “P”, “R”).
- 2.3: interactive “search based software design”.
- 2.4: interactive “search based software testing”.
- 2.5: interactive “search based software maintenance”.
- 2.6: interactive “search based software architecture”.

- 2.7: interactive “search based software development”.
- 2.8: interactive “search based software debugging”.
- 2.9: interactive “search based software verification”.
- 2.10: interactive “search based software evolution”. Removed items: 1 (result from 2017).

A.10 Other search engines

These search engines are not accurate enough, so they are discarded. They provide results in different languages and from non-related research fields. Search fields cannot be customised and logic connectors cannot be considered.

Search engine: CiteSeerX

- Webpage: <http://citeseerx.ist.psu.edu>
- Date: 18th July 2016
- Field: Title, Keywords, Abstract

Search engine: Google Scholar

- Webpage: <https://scholar.google.com>
- Date: 18th July 2016

A.11 Manual search

A.12 List of authors

After the search, the following authors appear as important contributors to iSBSE:

- Blekinge Institute of Technology (Karlskrona, Sweeden):
 - Bogdan Marculescu
 - Robert Feldt
 - Simon Poulding
- Centro Ricerca e Sviluppo (Napoli, Italy)
 - Cosimo Birtolo
 - Paolo Pagano
 - Luigi Troiano
- École de Technologie Supérieure (Montréal, Canada):
 - Adnane Ghannem
 - Ghizlane El Boussaidi
- State University of Ceará, (Fortaleza, Brasil):

- Allysson A. Araújo
- Altino Dantas
- Jefferson T. de Souza
- Thiago d. N. Ferreira
- Italo Yeltsin

- University of Bahia (Salvador, Brazil):
 - Antônio M. Pitangueira

- University of Michigan (Michigan, United States of America):
 - Slim Bechikh
 - Mohamed W. Mkaouer
 - Marouane Kessentini

- University of Salerno (Salerno, Italy)
 - Gabriele Bavota
 - Filomena Carnevale
 - Andrea De Lucia

- University of Trento (Trento, Italy)
 - Francis Palma
 - Angelo Susi
 - Paolo Tonella

- University of the West of England (Bristol, United Kingdom):
 - Jim Smith
 - Chris Simons

Sources for additional searches by author:

- DBLP: 21th July 2016
- Google Scholar: 26th July 2016
- Personal webpages: 26th July 2016

A.13 Additional papers

The following papers have been found as references of other papers, or are known by the authors of the iSBSE review. They all have been also searched in online databases to confirm that they are accessible.

- Y. Lin, X. Peng, Y. Cai, D. Dig, D. Zheng, W. Zhao. “Interactive and Guided Architectural Refactoring with Search-Based Recommendation”. Proc. 24th ACM SIGSOFT Int. Symposium on Foundations of Software Engineering.2016 [indexed by ACM library].
- M.W. Mkaouer. “Interactive Code Smells Detection: An initial Investigation”. Proc. 8th Symp. on Search-Based Software Engineering. 2016. [indexed by Springer]
- T. Neubauer, C. Stummer. “Interactive Decision Support for Multiobjective COTS Selection”. 40th Annual Hawaii International Conference on System Sciences (HICSS). 2007. [indexed by IEEE Xplore].

References

- [1] M. Harman and B. F. Jones, “Search-based software engineering,” *Information and Software Technology*, vol. 43, no. 14, pp. 833–839, 2001.
- [2] M. Harman, S. A. Mansouri, and Y. Zhang, “Search-based software engineering: Trends, techniques and applications,” *ACM Computing Surveys*, vol. 45, no. 1, p. 11, 2012.
- [3] B. Kitchenham and S. Charters, “Guidelines for performing Systematic Literature Reviews in Software Engineering,” tech. rep., Keele University and Durham University, 2007.
- [4] B. A. Kitchenham, D. Budgen, and P. Brereton, *Evidence-Based Software Engineering and Systematic Reviews*. CRC Press, 2016.
- [5] *ACM Computing Classification System*, Revision 2012. Available on: <https://www.acm.org/publications/class-2012>.
- [6] *IEEE Taxonomy*, version 1.0, 2017. Available on: https://www.ieee.org/documents/taxonomy_v101.pdf.