

# A Literature Survey on Empirical Evidence in Software Engineering

Ekaterina Pek

Software Languages Team, ADAPT Lab  
Universität Koblenz-Landau, Germany

Ralf Lämmel

Software Languages Team  
Universität Koblenz-Landau, Germany

# Research context

- Software re-engineering
- Software reverse engineering
- Program comprehension
- Specifically:
  - API usage analysis
  - Language usage analysis

# Research questions

- How often do Software Engineering papers use corpora—collections of empirical evidence?
- What is the nature and characteristics of the used corpora?
- Does common contents occur in the used corpora?

# Chosen venues

- European Conference on Software Maintenance and Reengineering (CSMR)
- Inter. Symposium on Empirical Software Eng. and Measurement (ESEM)
- International Conference on Program Comprehension (ICPC)
- International Conference on Software Maintenance (ICSM)
- Working Conference on Mining Software Repositories (MSR)
- Working Conference on Source Code Analysis and Manipulation (SCAM)
- Working Conference on Reverse Engineering (WCRE)

# Papers of the survey

Year	Conference	# papers	
		total	long
2012	CSMR	30	30
2012	ESEM	43	24
2012	ICPC	23	21
2011	ICSM	36	36
2012	MSR	29	18
2011	SCAM	19	19
2011	WCRE	47	27
	Total	227	175

# Choice of venues

We choose these conferences because i) they cover software engineering topics that, based on our experience, we expect to make use of empirical evidence; ii) they cover ground related to our expertise and research focus on software reverse/re-engineering and program comprehension with ESEM as notable addition for broader coverage of empirical software engineering research; iii) the conferences are of comparable size. In our survey, we use only long papers. We choose to analyze only conference proceedings, because while journal articles may adhere to the best practices, conference proceedings arguably contain the most common practices of research in the community—and we are interested in the latter.

# Related work

Name	Ref	Year	# used		Period	# papers			Focus	Coding schema
			j	c		total	sel.	rel.		
Glass et al.	[5]	2002	6	0	1995–1999	—	369	369	Characteristics of SE research	Topics, research approaches and methods, theoretical basis, level of analysis
Sjøberg et al.	[6]	2005	9	3	1993–2002	5453	103	103	Controlled experiments	Extent, topic, subjects, task and environment, replication, internal and external validity
Zannier et al.	[7]	2006	0	1	1975–2005	1227	63	44	Empirical evaluation: quantity and soundness	Study type, sampling type, target and used population, evaluation type, proper use of analysis, usage of hypotheses
Kitchenham et al.	[8]	2009	10	3	2004–2007	2506	33	19	Systematic reviews	Inclusion and exclusion criteria, coverage, quality/validity assessment, description of the basic data
<b>Our study</b>		<b>2013</b>	<b>0</b>	<b>7</b>	<b>2011/2012</b>	<b>227</b>	<b>175</b>	<b>175</b>	<b>Empirical evidence</b>	<b>Emergед classification</b>

Name	Ref	Year	# used		Period	# papers		
			j	c		total	sel.	rel.
Glass et al.	[5]	2002	6	0	1995–1999	—	369	369
Sjøberg et al.	[6]	2005	9	3	1993–2002	5453	103	103
Zannier et al.	[7]	2006	0	1	1975–2005	1227	63	44
Kitchenham et al.	[8]	2009	10	3	2004–2007	2506	33	19
<b>Our study</b>		<b>2013</b>	<b>0</b>	<b>7</b>	<b>2011/2012</b>	<b>227</b>	<b>175</b>	<b>175</b>



Ref	Year	# used		Period	# papers			Focus
		j	c		total	sel.	rel.	
[5]	2002	6	0	1995–1999	—	369	369	Characteristics of SE research
[6]	2005	9	3	1993–2002	5453	103	103	Controlled experiments
[7]	2006	0	1	1975–2005	1227	63	44	Empirical evaluation: quantity and soundness
[8]	2009	10	3	2004–2007	2506	33	19	Systematic reviews
	<b>2013</b>	<b>0</b>	<b>7</b>	<b>2011/2012</b>	<b>227</b>	<b>175</b>	<b>175</b>	<b>Empirical evidence</b>

# papers		Focus	Coding schema
sel.	rel.		
369	369	Characteristics of SE research	Topics, research approaches and methods, theoretical basis, level of analysis
103	103	Controlled experiments	Extent, topic, subjects, task and environment, replication, internal and external validity
63	44	Empirical evaluation: quantity and soundness	Study type, sampling type, target and used population, evaluation type, proper use of analysis, usage of hypotheses
33	19	Systematic reviews	Inclusion and exclusion criteria, coverage, quality/validity assessment, description of the basic data
<b>175</b>	<b>175</b>	<b>Empirical evidence</b>	<b>Emerged classification</b>

# Originality

Existing surveys focus on **specific forms or characteristics of SE research to be analyzed with a predefined schema**. For instance, Kitchenham et al. surveyed SE journals and conferences to find out adoption rate of systematic literature reviews. Similarly, Sjøberg et al. sought to find and analyze existing controlled experiments in SE research. By contrast, we (first to our knowledge) seek to discover whatever empirical evidence is used to facilitate SE research and we allow our coding schema to emerge from the data. We follow the idea of Grounded Theory (GT) as understood by Glaser.

# Grounded theory methodology

- Pilot study with a few explorative questions
- First pass over papers
  - Let the coding scheme emerge
- Second pass over papers
  - Revising paper coding for final scheme

# Coding scheme

- Information about corpora
- Forms of empirical research
- Self-classification
- Tools
- Structural signs of rigorousness / quality
- Reproducibility
- Assessment

Discuss how  
coding scheme  
emerged!

# Information about corpora

# Corpora -- usage

Almost all papers use a corpus of some sort. One out of six papers has more than one corpus. Most of the corpora consist of projects.

## Project-based corpora usage

ESEM	ICPC	WCRE	SCAM	CSMR	MSR	ICSM
58 %	81 %	81 %	84 %	87 %	89 %	94 %

Rather than images or traces ...

# Corpora -- contents

- Size: Half of the corpora, 99 cases, have three or less projects. 24 corpora that contain more than 10 projects. 8 papers with large corpora (with more than 100 projects).
- Languages. Most of the corpora are monolingual (147 cases); most of the remaining ones are bilingual (19 cases). As for the software language, 106 corpora contain projects written in Java, while C-like languages are used in 50 corpora (in C-like languages we include C, C++, C#).



# Corpora -- contents

- Code form: In 125 cases, corpora consist of source code; in 15 cases—of binaries. In the rest of the cases, code of the projects is not used, something else is in focus (developers, requirements, etc.).
- Access: In 128 cases, corpora consist only of open-source projects; in 12 cases, corpora consist only of projects not available publicly (e.g., industrial software); in 9 cases, corpora are self-written. The remaining cases mix access forms.

# Corpora -- contents

---

Project	# corp
JHotDraw	15
JEdit	12
Ant	11
ArgoUML	11
Eclipse	11
Firefox	10
Vuze/Azureus	8
Linux kernel	6
Lucene	6
Mozilla	6
Hibernate	5

---

# Corpora -- contents

An average project-based corpus consists of source code of three open-source projects, written in Java. Eclipse or its sub-parts is used in one out of eight papers using project-based corpora. The projects used in at least five papers are JHotDraw, JEdit, Ant, ArgoUML, Firefox, Vuze/Azerus, Linux kernel, Lucene, Mozilla, and Hibernate. Within the corpus, bug reports, defects, tests, and traces can be in the focus of the study.

## Popular projects

WCRE	SCAM	ESEM	ICSM	CSMR	ICPC	MSR
11 %	11 %	13 %	14 %	23 %	24 %	28 %
Eclipse JEdit	Lynx Minix	Eclipse	ArgoUML	Eclipse	JEdit	Firefox Eclipse

# Corpora -- sources

One out of four project-based corpora uses an established dataset, previous work, or online repository as a source of the projects. There is no common frequently used dataset or repository. Only SourceForge shows moderately frequent usage.

**Usage of corpora sources**

ESEM	SCAM	ICPC	ICSM	WCRE	CSMR	MSR
13 %	16 %	24 %	25 %	26 %	30 %	39 %

# Corpora -- sources

---

Type	# papers							
	Total	CSMR	ESEM	ICPC	ICSM	MSR	SCAM	WCRE
Established dataset	20	5	0	2	6	5	0	2
Previous work	13	2	3	2	1	1	2	2
Online repository	12	3	0	1	2	2	1	3
Total	43	9	3	5	9	7	3	7
Percentage	25	30	13	24	25	39	16	26

---

# Corpora -- evolution

- 52 papers use evolution of the projects in their research
- Forms of evolution measure: “version” (21 times), “revision” (11), “commit” (10), “release” (11).
- On the average, papers mentioning commits use 3,292 commits; papers with revisions—18,870 revisions; with versions— 10 versions; with releases —10 releases.

# Corpora -- evolution

- There are 46 papers that mention a time span of their study. In 36 cases, the unit of the time span is a year and on the average such papers are concerned with a 8-year span.
- We found 23 papers to mention what version control system was involved in the study. CVS is mentioned 11 times, SVN—11 times, Git and Mercurial—4 and 2 times respectively.

# Corpora -- evolution

One out of three papers with project-based corpora uses evolution aspect in its research. In half of the cases, large-scale evolution is involved: several thousands commits/revisions or ten versions/releases of projects—often spanning several years of a project's lifetime.

## Evolution usage

ICPC	SCAM	ESEM	CSMR	ICSM	WCRE	MSR
14 %	16 %	21 %	33 %	33 %	33 %	56 %



# Corpora -- requirements

One out of five papers requires the projects of its corpus to have an ecosystem: a bug tracker, or a mailing list, or some kind of documentation. Other requirements focus on the size and language of the projects, application domain, development history, etc.

## Popular requirements

SCAM	MSR	ICSM	ICPC	ESEM	WCRE	CSMR
11 %	28 %	14 %	24 %	13 %	11 %	23 %
domain	ecosys	size	size	ecosys	ecosys	ecosys
lang	p.work					
size						

# Corpora -- tuning

We have detected few common actions applied to corpora during research: source code/binaries modification; execution of the tests on the corpus or of the corpus itself; filtering of the corpus contents. Altogether, one out of four papers contains signs of one of these actions.

## Popular actions

ESEM	MSR	SCAM	ICPC	ICSM	WCRE	CSMR
8 %	11 %	11 %	14 %	19 %	19 %	20 %
tests	run	modif.	modif.	tests	modif.	modif.

One out of four project-based corpora requires some manual effort.

## Manual effort

CSMR	ESEM	ICSM	SCAM	ICPC	MSR	WCRE
13 %	13 %	22 %	26 %	33 %	33 %	37 %

# Self-classification

# Self-classification

---

Type	#
case study	48
experiment	44
empirical study	22
evaluation	14
exploratory study	6
...	...

---

# Self-classification

Four out of five papers provide self-classification, but it might be vague. The most popular term, ‘case study,’ may be misused. Cf., “There is much confusion in the SE literature over what constitutes a case study. The term is often used to mean a worked example. As an empirical method, a case study is something very different.” [12]. Cf., “... our sample indicated a large misuse of the term case study.” [7]

## Self-classification

SCAM	MSR	CSMR	WCRE	ICPC	ESEM	ICSM
37 %	61 %	80 %	81 %	86 %	88 %	94 %

# Emerged forms

# Experiments

One out of ten papers contains an experiment. The majority of the experiments use project-based corpora; experiments often use questionnaires, usually two per experiment. An average experiment involves 16 students, often in two groups (by the received treatment or experience level); it consists of four tasks and lasts for an hour. One out of four experiments suggests some compensation to its participants; one out of four experiments is preceded by a pilot study.

ICPC and ESEM are the main source of experiments involving professionals.

Experiments						
MSR	SCAM	CSMR	WCRE	ICSM	ESEM	ICPC
0 %	0 %	3 %	7 %	8 %	21 %	38 %

# Questionnaire

Altogether, we have found 36 questionnaires in 24 papers. As mentioned, 20 questionnaires are used in experiments—to distinguish, we will refer to them as experiment-related and the other 16 we will qualify as experiment-unrelated.



# Questionnaire

More than half of the detected questionnaires are used in experiments—often as pretest and posttest questionnaires. The other half, experiment-unrelated questionnaires, are found in one out of twelve papers. Sizewise, on the average there is no difference between experiment-related and -unrelated questionnaires. Experiment-unrelated questionnaires usually involve professionals as participants—in contrast to experiment-related questionnaires that mostly use students. Typical requirements for participants in experiment-unrelated questionnaires have to do with experience or expertise. One out of three experiment-unrelated questionnaires are preceded by a pilot study.

## Experiment-unrelated questionnaires

MSR	CSMR	SCAM	WCRE	ICSM	ICPC	ESEM
0 %	3 %	5 %	7 %	8 %	19 %	25 %

# Literature survey

Literature surveys are quite rare: only one out of 35 papers contains it. On the average, a literature survey starts with few thousand papers to be filtered down to few dozens papers that will be analyzed. Usually, the first round of filtering is based on the title and abstract, then the full text of the papers is considered. There is not enough information to conclude about frequently used digital libraries or conferences/journals. Half of the surveys were following guidelines of systematic literature reviews by Kitchenham [22].

**Literature surveys**

ICSM	MSR	SCAM	WCRE	CSMR	ICPC	ESEM
0 %	0 %	0 %	0 %	3 %	5 %	13 %

# Comparisons

One out of three papers compares tools, techniques, approaches, metrics, etc.—half of the time, to evaluate what was introduced in the study. On the average, such evaluation involves one other entity. In the other half of the cases, the average number of compared entities is 3.

## Comparisons

ICPC	MSR	SCAM	WCRE	ESEM	ICSM	CSMR
19 %	22 %	26 %	30 %	33 %	36 %	47 %

# Tools

# Tools

One out of four papers introduces a new tool; another one out of four papers uses some home-grown tooling. Almost three out of four papers use existing tools.

The most popular standard tool, Eclipse—an IDE and a platform for plug-in development—is used in one out of seven papers. Other popular tools cater for source code analysis, clone detection, evolution analysis, data mining, statistics, quality analysis, document classification.

## Home-grown tooling

ICPC	ESEM	CSMR	ICSM	MSR	WCRE	SCAM
5 %	17 %	20 %	33 %	33 %	33 %	42 %

## Introduced tools

ESEM	MSR	ICPC	SCAM	CSMR	WCRE	ICSM
4 %	11 %	19 %	21 %	30 %	37 %	44 %

# Use of existing tools

---

Tool	# papers
Eclipse <sup>1</sup>	25
R project <sup>2</sup>	16
CCFinder <sup>3</sup>	6
Understand <sup>4</sup>	6
Weka <sup>5</sup>	6
ConQAT <sup>6</sup>	4

---

---

Tool	# papers
MALLET <sup>7</sup>	4
ChangeDistiller <sup>8</sup>	3
CodeSurfer <sup>9</sup>	3
Evolizer <sup>10</sup>	3
RapidMiner <sup>11</sup>	3
RECODER <sup>12</sup>	3

---

<sup>1</sup> <http://eclipse.org>

<sup>2</sup> <http://www.r-project.org/>

<sup>3</sup> <http://www.ccfinder.net/>

<sup>4</sup> <http://www.scitools.com/>

<sup>5</sup> <http://www.cs.waikato.ac.nz/ml/weka/>

<sup>6</sup> <https://www.conqat.org/>

<sup>7</sup> <http://mallet.cs.umass.edu/>

<sup>8</sup> <http://www.ifi.uzh.ch/seal/research/tools/changeDistiller.htm>

<sup>9</sup> <http://www.grammatech.com/products/codesurfer/overview.html>

<sup>10</sup> <http://www.ifi.uzh.ch/seal/research/tools/evolizer.html>

<sup>11</sup> <http://rapid-i.com/content/view/181/190/>

<sup>12</sup> <http://sourceforge.net/projects/recoder/>

# Structural signs of rigorousness/quality

# Structural signs of rigorousness/ quality

Half of the papers use research questions to structure their study. One out of seven papers uses a “Goal-Question-Metric” approach and/or formulate (null) hypotheses to structure their research. One out of seven papers provides an explicit set of definitions of the terms used in the study. Threats to validity are discussed in three out of five papers.

The following three combinations of structural signs describe at least half of the papers in each conference, except for WCRE, where only 44 % of papers are covered by these combinations.

## No structural signs

ICPC	MSR	WCRE	ESEM	ICSM	CSMR	SCAM
14 %	17 %	19 %	21 %	22 %	27 %	53 %

## Both research questions and threats to validity

ICSM	WCRE	SCAM	CSMR	MSR	ICPC	ESEM
8 %	11 %	16 %	20 %	22 %	24 %	42 %

## Only threats to validity

ESEM	MSR	SCAM	WCRE	CSMR	ICPC	ICSM
4 %	11 %	11 %	15 %	17 %	19 %	31 %



# Reproducibility

# Reproducibility

We judged one out of six papers to be reproducible with respect to the used corpus and tools. We did *not* assess whether enough details were provided to re-conduct the research itself.

## Judged to be reproducible

ICSM	WCRE	SCAM	ICPC	ESEM	CSMR	MSR
3 %	4 %	16 %	19 %	25 %	27 %	33 %

# Assessment

# Assessment

We have low confidence in one out of eleven papers that we have coded. In the rest, half of the time we are moderately confident and half of the time—highly confident in the results.

## High confidence

---

WCRE	SCAM	ICSM	ICPC	ESEM	CSMR	MSR
15 %	26 %	42 %	52 %	54 %	60 %	78 %

---

## Moderate confidence

---

MSR	CSMR	ESEM	ICPC	ICSM	SCAM	WCRE
17 %	33 %	33 %	43 %	53 %	58 %	67 %

---

## Low confidence

---

ICPC	ICSM	MSR	CSMR	ESEM	SCAM	WCRE
0 %	6 %	6 %	7 %	13 %	16 %	19 %

---

# Answers to the research questions

- I The overwhelming majority of surveyed conference papers use corpora—collections of empirical evidence.
- II The majority of the corpora consist of projects and can be characterized by size, code form, software language, evolution measures, requirements, and applied tunings.
- III There are no frequently used projects or corpora across all the papers. We have detected though some pattern of project recurrence with low frequency.