

How to write a research abstract

Research Methods - Barbara Russo

SwSE - Software and Systems Engineering

Beck Kent

- Founder of the agile method of development
- eXtreme Programming
- Agility in writing a paper
- A long experience in reviewing articles

K.Beck's personal view

- Write the programme Committee
- One startling sentence
- Argument: problem, defence, related work
- Abstract

Write the programme Committee

- PC made of experts in the field
- PC's desk piled with papers
- 30 seconds to grab their attention
- If the paper is of broad interest there must be a spark in it
- If your paper is technical they may not be familiar with it. It must readably present the novel aspects of the work

One startling sentence

- Temptation to write everything of your work
 - Boil your message down to one startling sentence
 - Danger
 - Open to criticisms
- Example
 - “Network garbage collection is fast and easy”
 - The sentence is clear and open to discussion

Argument: problem, defence, related work

- Stand for the validity of your startling sentence
- Convince of the truth of your sentence

Divide your paper into four sections

- First section: the problem to be solved
 - Why is a problem
 - Importance to solve it
- Second section: describe the problem
 - Related work justifying the problem
 - Implementation details

Divide your paper into four sections

- Third section: solution of the problem
 - Discuss all the reasonable counter arguments
- Fourth section: what other people have done in the area
 - convince of the novelty

Abstract

- A four sentence summary of the conclusion of the paper
- First sentence: state the problem
- Second sentence: why the problem is a problem
- Third sentence: the startling sentence
- Fourth sentence. Implication of the startling sentence

Exercise - Find the 4 sentences

Abstract

In this paper, we focus on the evaluation of the factors that impact on the introduction of Open Source Software (OSS) by means of the analysis of Open Data Standards (ODS) generation. In this sense, we model the generation of Open Data Standards as a self reinforcing mechanism that perpetrates through time. To perform such analysis, we use urn models, models that are typically deployed when modelling path dependent processes. First, we perform the identification of the variables that impact on the generation effect, in particular Network Externalities that arise due to the presence of a large number of files and size of the files created. Second, we evaluate the urn selection process through a multi-urn schema. The main findings are a confirmation of the importance of Network Externalities as reported by theory and the importance of past historical file generation for the subsequent file generation process.

New abstract

Exercise - Find the 4 sentences

Former abstract

In this paper, we focus on the evaluation of the factors that impact on the introduction of Open Source Software (OSS) by means of the analysis of Open Data Standards (ODS) generation. In this sense, we model the generation of Open Data Standards as a self-reinforcing mechanism that perpetrates through time. To perform such analysis, we use urn models, models that are typically deployed when modelling path dependent processes. First, we perform the identification of the variables that impact on the generation effect, in particular Network Externalities that arise due to the presence of a large number of files and size of the files created. Second, we evaluate the urn selection process through a multi-urn schema. The main findings are a confirmation of the importance of Network Externalities as reported by theory and the importance of past historical file generation for the subsequent file generation process.

New abstract

The decision about the adoption of Free/Libre/Open Source Software (FLOSS) is a key issue in Small and Medium Enterprises (SMEs). Indeed, very often such organisations don't have the resources needed to fully evaluate the migration from existing legacy systems. To help the decision process of these organisations, we propose a preliminary study about an instrument based on the analysis of files' generation of targeted data standards. We model the file generation process as a self-reinforcing mechanism through the usage of urn models.

By applying the instrument to a large dataset in the office automation field, we found a first confirmation about the importance of network externalities as reported by theory and the importance of past historical file generation for the subsequent file generation process

What is the state of the art? Mapping studies and Systematic Literature Review

Research Methods - Barbara Russo

SwSE - Software and Systems Engineering

Types of studies

- Primary studies
- Secondary studies : mapping studies and systematic literature reviews
- Tertiary studies : research synthesis

Systematic mapping studies

- Main goal: providing an **overview** of a research area
- Identify the **quantity and type** of research and results available within it

Reference article: Kai Petersen, Robert Feldt¹, Shahid Mujtaba^{1;2}, Michael Mattsson, Systematic Mapping Studies in Software Engineering EASE2008

Exercise

- Which is which?

Systematic literature reviews in global software development

- Context: There has been an increase in research into global software development (GSD)
- Objective: The aim of this research is to catalogue GSD ... in order to identify the topics covered, the active researchers, the publication vehicles, and to assess the quality of the SLRs identified.
- Method: We performed a broad automated search ... dealing with GSD. We differentiate between ... studies and papers reporting those studies. Data relating to each of the following was extracted and synthesised from each study: authors and their affiliation at the time of publication, the journal or conference in which the paper was published, the quality of each study and the main GSD study topic.
- Results: Twenty-four GSD ... studies and 37 papers reporting those studies were identified. Major GSD topics covered include: (1) organisational environment, (2) project execution, and (3) project planning and control. The main research groups are based in Brazil (17), Ireland (8), and Sweden (7).
- Conclusions: GSD ... studies are most frequently reported in the International Conference on Global Software Engineering and IEEE Software; the two most popular topics for research are risk factors due to the organisational environment and the development process. The most active researchers are based in Brazil. The quality of the ... studies has not changed over time

Requirements prioritisation in software engineering

- In this paper, we report about a ... study in software requirements prioritisation with a specific focus on empirical studies.
- The results show that the interest from the research community is clustered around the more recent years. The majority of the studies are about the validation of research or solution proposals. We report the prevalence of studies on techniques and methodologies while there is a scarce interest in the strict evaluation of tools that could be beneficial to industry.
- In most of the empirical studies we found a bottom-up approach, centering on the techniques and on accuracy as the dependent variable, as well as on functional requirements as the main research focus. Based on the results, we provide recommendations for future research directions

Understanding requirement prioritization techniques

- Software is becoming progressively more integral part of day-to-day life. Developing the software that meet stakeholders' need is the ultimate goal in today's environment. As the complexity of software increases so does the requirements. There are many requirements which should be fulfilled in the given time duration on the other hand some requirements should be considered first to reduce the risks. Hence, proper gathering and prioritizing requirements may leads to the successive development of the software. In literature there are number of techniques which focus on requirement prioritization problem. This paper presents the comparative study of various requirement prioritization techniques.

Exercise

- Navigate to find a primary study on requirements

Systematic mapping studies

- Start with a **research goal** or a theme
- Derive **research questions**
- The research questions are **general** as they aim at **discovering research trends** e.g. publication trends over time, topics covered in the literature

Systematic mapping studies

- The **outcome** is an inventory of papers on the topic area, mapped to a **classification** and **counting** contributions in relation to the categories of that classification
- Sometimes **venues**
- What? How many? Where?

Example of outcome - Bubble plot

Bubble Are Counts

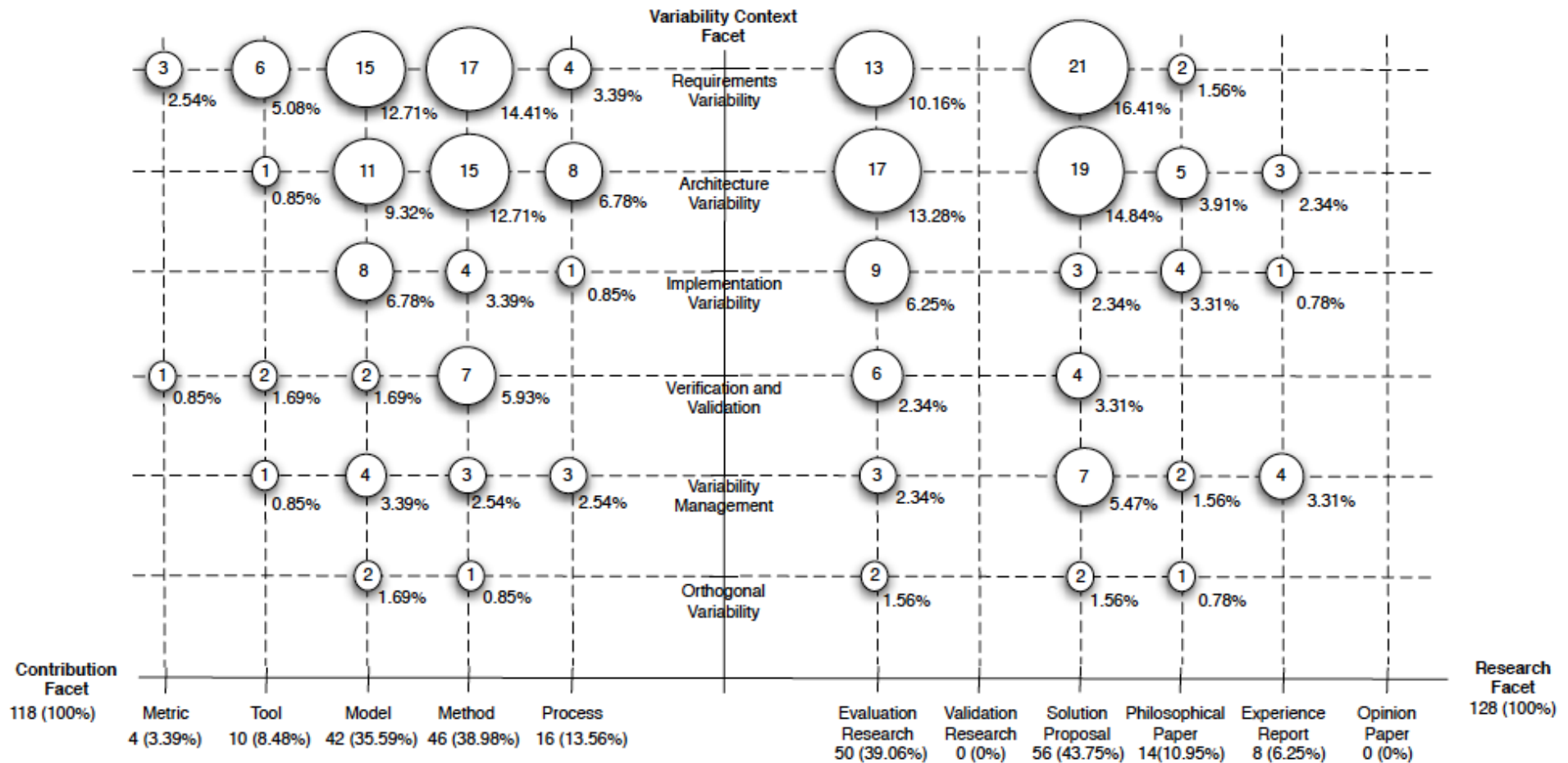


FIGURE 3: Visualization of a Systematic Map in the Form of a Bubble Plot

The systematic mapping study process

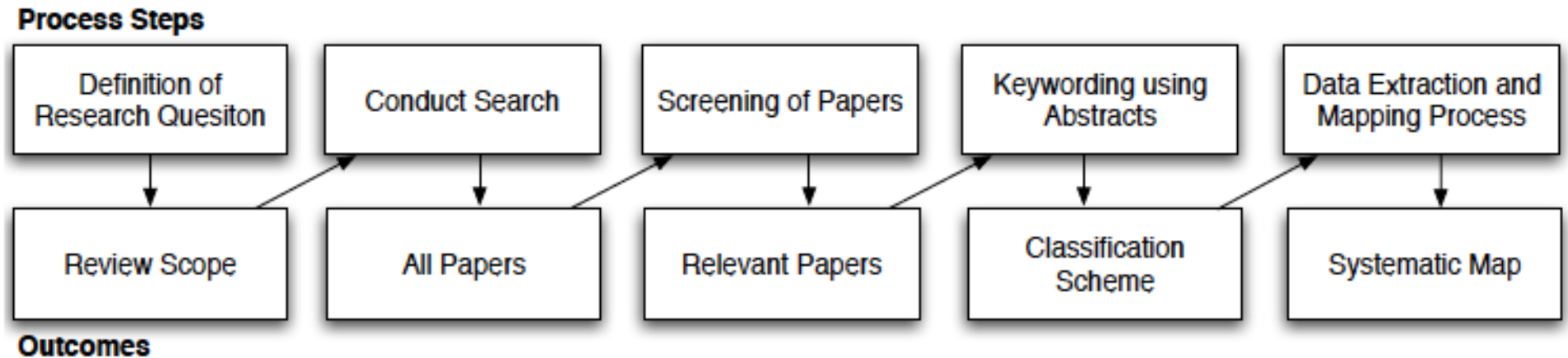


FIGURE 1: The Systematic Mapping Process

Definition of Research Goals and Questions

- Formulate your goal for a mapping study
- Derive your research questions that answer the goal

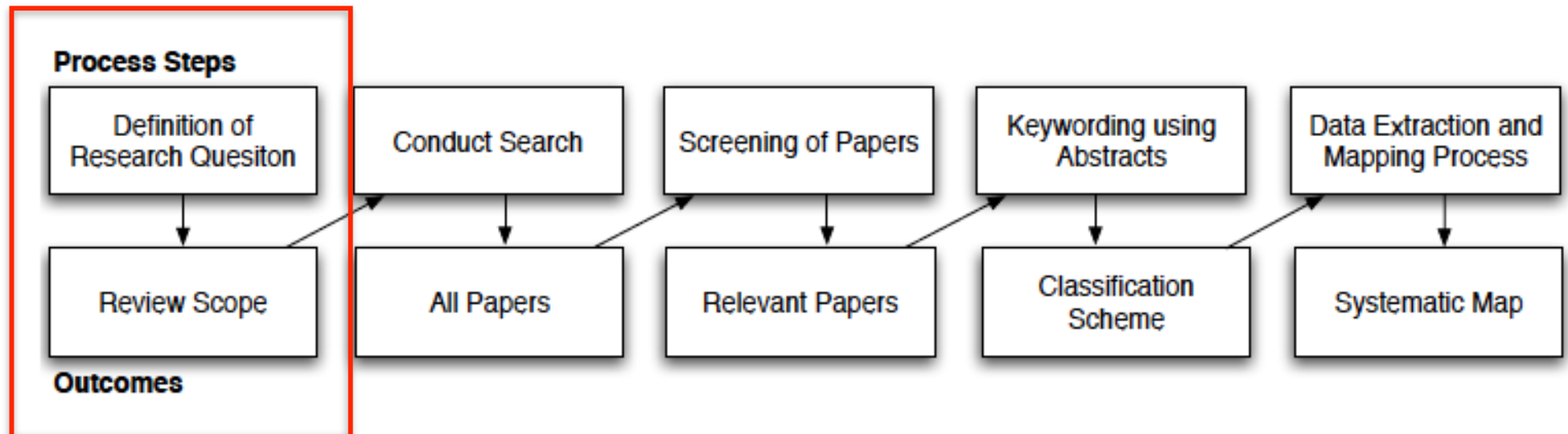


FIGURE 1: The Systematic Mapping Process

Example from Wortmann et al.

- Goal: Characterise **Model-based systems engineering** is a key enabler for the **complex systems of systems** of Industry 4.0
 - RQ1: What are the expected **benefits** from applying MBSE to contribute to Industry 4.0?
 - RQ2: Which Industry 4.0 **concerns** are addressed through applying MBSE?
 - RQ3: Which MBSE **tools and languages** are used in Industry 4.0 and which concerns do they address?
 - RQ4: What are the most frequently **applied research methods** in the context of applying MBSE to Industry 4.0?

Example from Bailey et al

- Goal: empirical knowledge of the **effectiveness** of the object-oriented paradigm
 - RQ1: Which **journals** include papers on software design?
 - RQ2: What are the most investigated **OO design topics** and how have these **changed over time**?
 - RQ3: What are the most frequently **applied research methods**, and in what **study context**?

Conduct search on Primary Studies

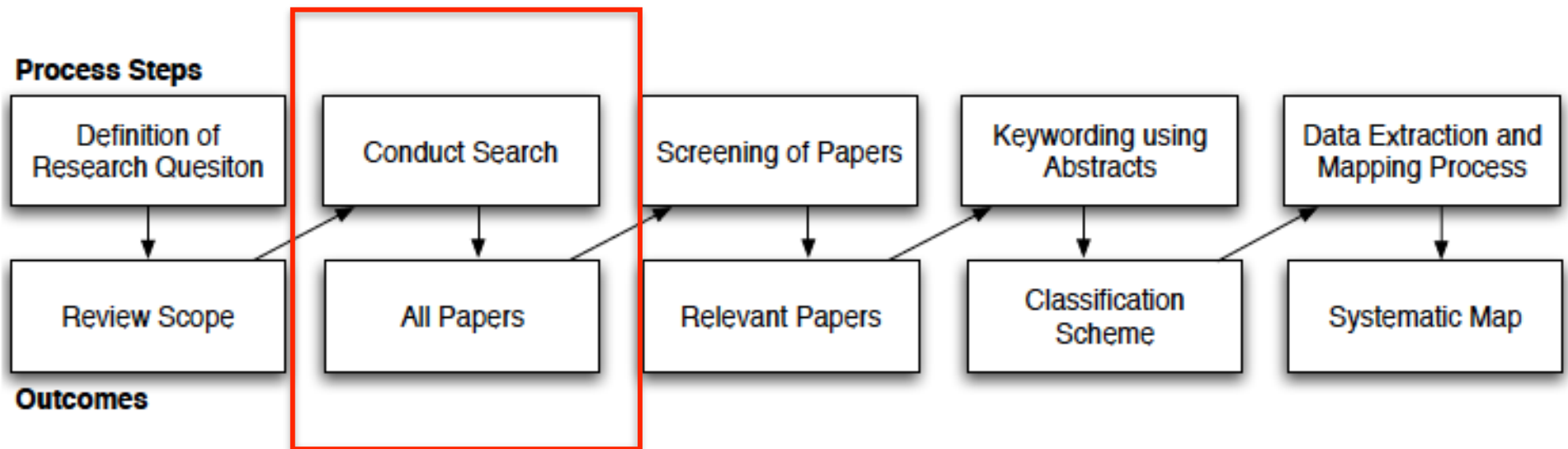


FIGURE 1: The Systematic Mapping Process

What are the search strings?

- Primary studies are identified by keyword search
 - Search strings on scientific databases or browsing, or
 - Manually in conference proceedings or journal publications
- Derive keywords from the RQs

Identify search string - PICO

- PICO: Population, Intervention, Comparison and Outcomes
- **Structure keywords and eventually search them in the text with PICO**

Kitchenham and Charters 2007

PICO: Population

- It may refer to a specific role, a category of stakeholders, an application area or an industry group
- Example: published literature on “Edge networks”

PICO: Intervention

- It is the **methodology/tool/technology/procedure** that addresses a specific issue, for example, technologies to perform specific tasks such as requirements specification, system testing, or software cost estimation
- Example: empirical studies on “Fog computing” (e.g., end-user clients to carry out ‘big’ storage)

PICO: Comparison

- It is the **methodology/tool/technology/procedure** with which the research is being compared
- It is often referred to as the “**control**” **treatment**
- Example of control treatment: “cloud data centres”

PICO: Outcome

- It may relate to **factors of importance** to practitioners such as improved reliability, reduced production costs, and reduced time to market
- Example: distributed decision

Search strings

- Start with search strings (from RQs) in different databases or
- Start with the reference lists of a starting set of papers
- Iterate the process until you are not able to find relevant papers

Select your databases

- IEEE
- ACM
- Scopus
- Google scholar
- dblp
- Microsoft research
- Elsevier
- Springer

API for mining publication DBs

- Available APIs
- <http://libguides.mit.edu/apis>
- Examples:
- <http://ieeexplore.ieee.org/gateway/>
- <http://api.elsevier.com/documentation/SCOPUSSearchAPI.wadl#simple>

API query in Python - example

- <http://kitchingroup.cheme.cmu.edu/blog/2015/04/03/Getting-data-from-the-Scopus-API/>

Now you have a pool of papers!

Screening of papers

- Inclusion criteria
- Exclusion criteria

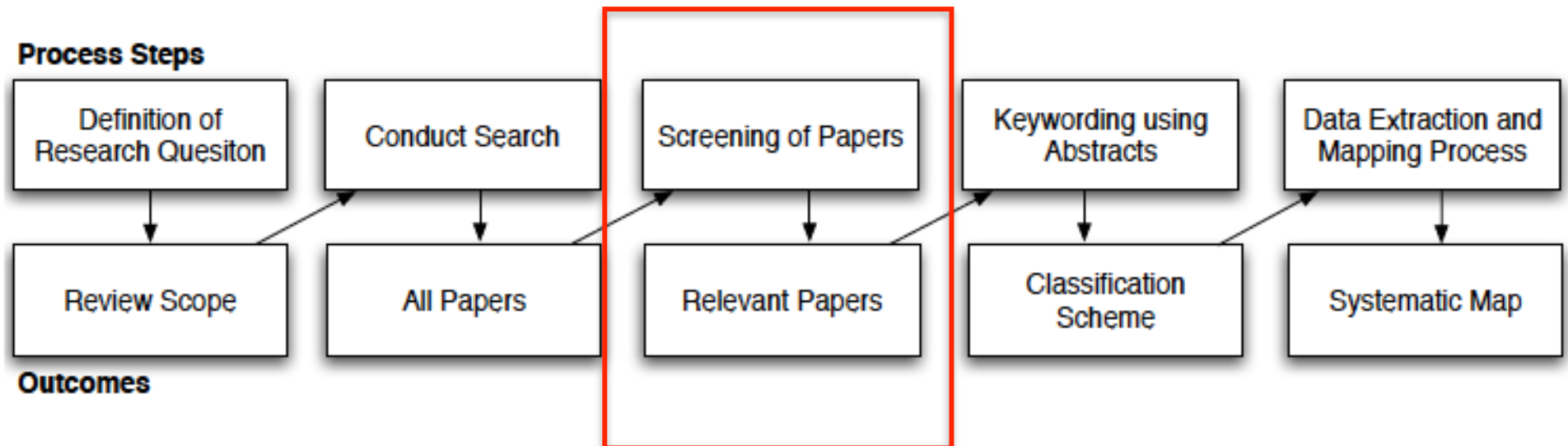


FIGURE 1: The Systematic Mapping Process

Selection

- Inclusion
 - State the criteria for inclusion and use snowballing on citations
- Exclusion
 - State criteria for exclusion

Example - inclusion

- Example of Inclusion criteria
 - Peer-reviewed studies published in journals, conferences, and workshops
 - Studies are accessible electronically
 - Contributing MBSE to Industry 4.0

Exclusion - exclusion

- Example of Exclusion criteria:
 - Studies not available in English
 - Studies not peer-reviewed, e.g., books, slides, websites
 - Short papers of less than two pages, such as calls for papers, editorials, or curriculums
 - Studies where Industry 4.0 is mentioned as future application, related work, or broad context only, e.g., papers on Internet of Things (IoT) or Cyber-Physical Systems (CPS) mentioning Industry 4.0 as a possible use case only

Reading Strategies to investigate I/E criteria

- Titles only
- Titles and keywords
- Titles, keywords, and abstracts

- In doubt read the full paper

Snowballing

- Forward snowballing: articles that have cited the articles found in the search
- Backward snowballing: articles extracted from the reference lists

Now you have a pool of **relevant** papers!

Analysis and classification

- **Read abstracts**
- **Look for keywords and concepts that reflects the contribution of the paper**
- **Identify the context of the research**
- **Combine keywords from different papers and come up with a set of categories for the underlying population**

- When abstracts are of too poor quality
 - Study also the introduction or conclusion

Example - Wortmann et al.

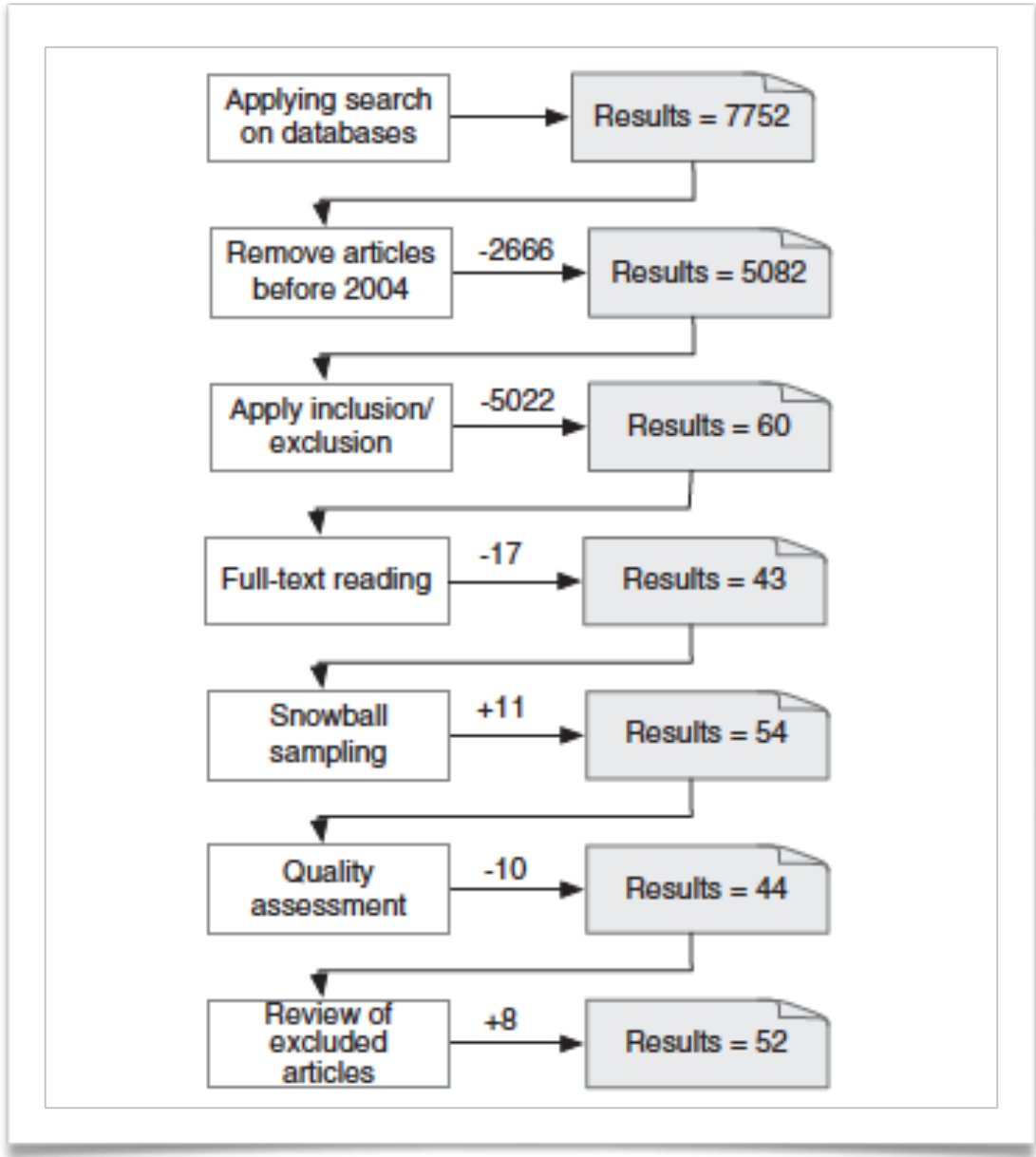
- Major facets
 - Factors of classification
 - Contribution Type
 - Research Type
 - **Industry 4.0 Concern**
 - Modeling Technique

Let's focus on

- Industry 4.0 Concern
 - Digital Representation
 - Failure Handling
 - Human Factors
 - Information Management
 - Integration
 - Processes
 - Product Modelling
 - (Re-) configuration
 - Verification & Validation
 - Visualisation

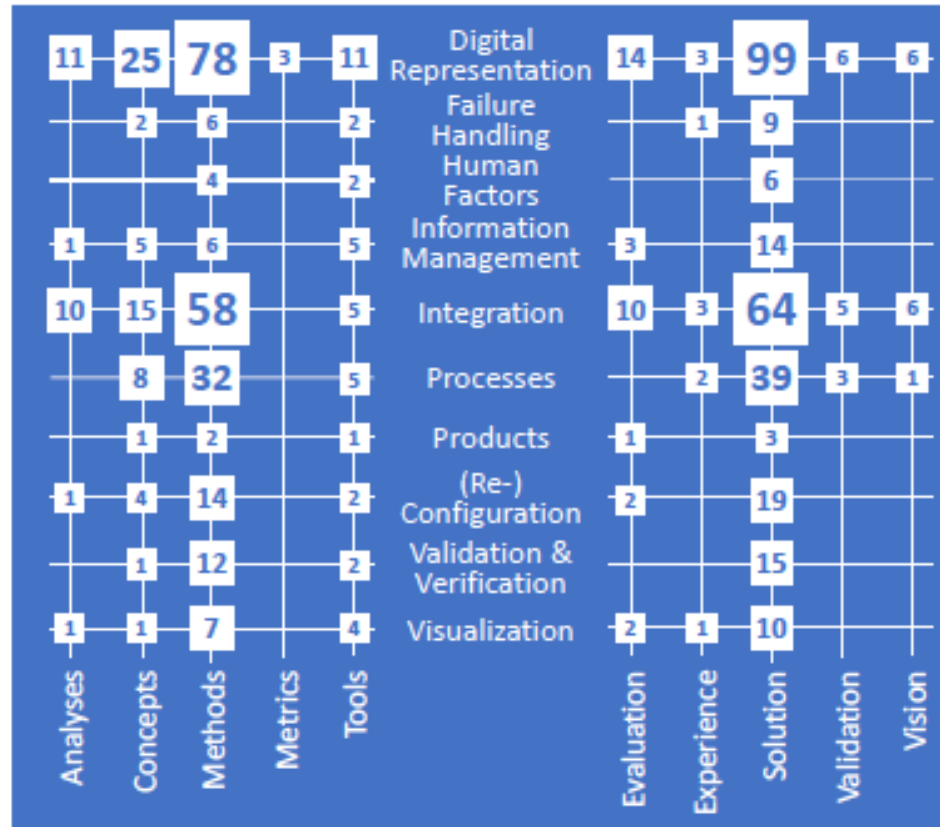
Analysis and classification

- Extract information for each item and represent it visually
- Optional: Group topics into themes and sub-themes
- Assign a paper to a factor or sub-factors
- Count papers per factor and visually plot the count



Kai Petersen, Sairam Vakkalanka, Ludwik Kuzniarz

Classification - Wortmann et al.



Industry 4.0 concerns by research type and contribution type

Data Extraction and Mapping of Studies (Systematic Map)

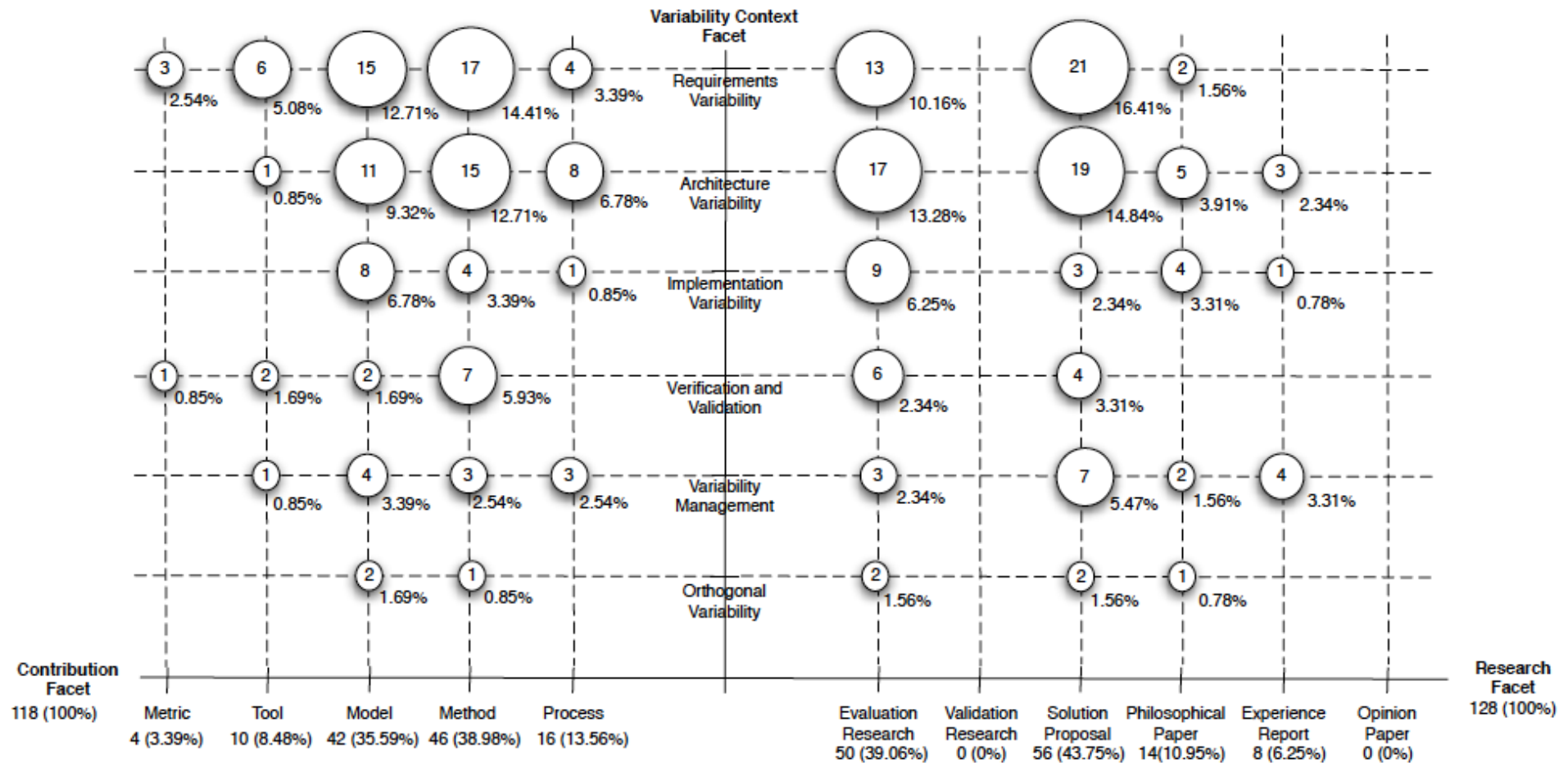


FIGURE 3: Visualization of a Systematic Map in the Form of a Bubble Plot

Discuss your results

- What are the themes that have been studied the most?
- What are the research gaps?

Systematic Literature Review

Research Methods - Barbara Russo

SwSE - Software and Systems Engineering

Systematic Literature review

- Systematic reviews aim at aggregating evidence and hence a very specific goal has to be formulated (e.g. whether an intervention is practically useful by industry) - Kitchenham et al.

Why LR

- To **summarise** the existing **evidence** concerning a treatment or technology
- To **identify** any **gaps** in current research in order to suggest areas for further investigation
- To provide a framework/background in order to appropriately **position new research activities**

The review protocol

- **Background.** The rationale for the survey.
- The **research questions** that the review is intended to answer
- The **strategy** that will be used to search for primary studies including search terms and resources to be searched
 - An **initial mapping study** can help determine an appropriate strategy

The review protocol

- **Study selection criteria.** Study selection criteria are used to determine which studies are included in, or excluded from, a systematic review.
 - It is usually helpful to **pilot** the selection criteria on a subset of primary studies.

The review protocol

- **Study selection procedures.** The protocol describes how the selection criteria will be applied e.g. how many assessors will evaluate each prospective primary study, and how disagreements among assessors will be resolved

The review protocol

- **Study quality assessment** checklists and procedures. The researchers should develop quality checklists to assess the individual studies.
- **Data extraction strategy.** This defines how the information required from each primary study will be obtained.

MP vs LR

- The search for studies in systematic maps is based on a topic area, while literature reviews are driven by specific research questions

Example in SE

- <http://guides.library.cornell.edu/c.php?g=356172&p=2497100>

Citations network

- Pajek tool
- <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>