# Systematic Mapping Studies

Isaac Griffith

CS 6620 - Empirical Software Engineering
Department of Computer Science
Idaho State University

ROAR

# Outline

At the end of today's lecture you should be able to

- Plan and conduct an SMS
- Identify and use appropriate search strategies
- Identify and use appropriate inclusion/exclusion criteria
- Execute data extraction and classification
- Understand the various threats to validity for mapping studies
- Structure and produce a mapping report

ROAR

# Planning the Mapping

In this stage all relevant decisions for conducting the mapping are made. This stage consists of:

❶ Need Identification and Scoping

❷ Study Identification

❸ Data extraction and classification

❹ Visualization

❺ Threats to Validity

ROAR

# Need Identification

SLRs and SMSs differ in their high-level goals and some parts of the process. For an SMS common research goals have been identified:

- "To examine the extent, range, and nature of research activity"
  - identify the extend different practices are studied and reported in the literature

- "To determine the value of undertaking a full systematic review"
  - by classifying paper types it allows us to determine if a full review can be undertaken
  - by structuring the area we can refine research questions

- "To summarize and disseminate research findings"
  - provides new grad students/researchers a useful tool to orient them to the area.

- "To identify research gaps in the existing literature"
  - identifies the number of studies and where a need for further research exists

ROAR

# Scoping

- This is all about the identification of research questions.

- SMS questions should be broader in scope than those for an SLR and should help us understand what is currently known about a topic.

- Example:
  - "What do we know about software product management?", which can be refined to:
    - "What research questions in software product management are being addressed?"
    - "What original research exists in the intersection of software product management and cloud (service) environment?"
    - "What areas in software product management require more research?"

ROAR

# Scoping

Additional questions beyond the topic coverage can be asked.

- Questions regarding the following are asked:
  - publication venue and publication types
  - research methods used
  - research trends
  - locale demographics of published papers
- Additionally other topic related information can be asked, for example
  - "What are existing approaches that combine static and dynamic quality assurance techniques and how can they be classified?"
  - "In which sources and in which years were approaches regarding the combination of static and dynamic quality assurance techniques published?"
  - "Is any kind of evidence presented with respect to the combination of quality assurance techniques and if so, which kind of evidence is given?"

ROAR

# Study Identification

❶ Choosing the search strategy

❷ Developing the search

❸ Evaluate the search

❹ Inclusion and Exclusion

❺ Quality Assessment

ROAR

# Search Strategies

There are three primary search strategies that can be used:

- Manual Search
- Database Search
- Snowball Sampling

There is no one specific recipe that works best, but a combination of the approaches tends to be better than any single approach.

Suggestion:

- Manual + Snowball -> If you have an idea of the research area already
- Database + Snowball -> If you are new to the area

ROAR

# Manual Search

- Using knowledge of the research area, a manual search can be conducted in the following locations:
  - Prominent Author's webpages
  - References of existing literature reviews
  - Known Conferences for the research area
  - Known Journals for the research area

- The goal is to obtain a representative sample of the papers from which snowball sampling can then be used.
  - The tradeoff between manual and database search is
    - Less false positives up front with manual search than with database search
    - More rounds of snowball sampling required for manual search
    - Higher chance of missing key primary studies with manual search if the original sample was not representative enough

ROAR

# Database Search

If you do not have enough knowledge to conduct a manual search, a database search will be necessary. This proceeds as follows:

- Develop your search string
- Database selection
- Search execution

Note that there is always a tradeoff between:

- finding all relevant primary studies
- being overwhelmed with false positives to be removed manually

# Search Development

- Similar to SLRs use the PICO(C) strategy, but focus on the P and I for SMS

- Utilize standards from the IEEE, ISO/IEC, and the Swebok to identify related and relevant keywords

- To improve your search string:
  - Consult experts in the area
  - Identify known keywords from other papers
  - Use an encyclopedia and/or thesaurus to identify additional keywords/synonyms

- Consider precision and noise
  - If a search string returns too many irrelevant hits, consider restricting further

ROAR

# Search String

- Use the research questions to create sets of keywords
- You should have one set per item of your PICO(C) question criteria
- For each item in each set, identify applicable synonyms for each keyword and add to the set.
- Build a Boolean search string from the selected keywords
  - **OR** together keywords within a set
  - surround sets with parentheses
  - **AND** together the sets
- If necessary, refine the expression to improve your results.
- You should analyze the sensitivity of the results of your search to refine the search string.

ROAR

# Database Selection

- A single database is usually not enough to identify your primary studies.

- Using multiple databases will inevitably result in duplicates, but a greater sample of the population as well.

- For CS/SE Studies you should use at least the following:
  - IEEExplore
  - ACM Digital Library

- You should use 2 or more general indexes such as:
  - INSPEC/Compendex
  - Web of Science

ROAR

# Database Selection

- Additionally, you can use the following indexes (but they are not necessary)
    - ScienceDirect
    - SpringerLink
- **Do NOT use Google Scholar** since the results are not replicable
- Realize: you cannot find all primary studies for a given topic, and the what is found is simply a sample

ROAR

# Search Execution

- Keep a detailed record or the search findings.
  - Number found
  - Included studies
  - Excluded studies
  - Duplicate studies
  - Snowball sampling is simply a manual search for primary studies based on a selection of identified primary studies.
  - There are two types of snowball sampling: Forward and Backward

ROAR

# Snowball Sampling

**Backward Snowballing**

- Search the references of a primary study for new primary studies
- Can be done either by:
  - Searching actual paper reference section, or
  - Searching through listed references on online paper listing

**Forward Snowballing**

- Search the items citing a primary study for new primary studies
- Can be done either by:
  - Using Google scholar cited-by listings
  - Using "Citations" section at paper listing
- This process continues as new studies are added until no new studies are added

ROAR

# Search Evaluation

To evaluate the search you need the following:

- A search string to refine
- A known set of papers from the research area, which are within your population (at a minimum 10 papers)
    - Ask a known expert in the area for a selection of papers
    - Identify key researchers in the area and select papers from their websites/CVs

Search strings are evaluated by

- Executing the search in the selected databases
- Checking if the results contain the known set of papers

# When to Stop Searching

- Search is a time-consuming process, and we will never be able to find all of the papers.

- To save time, adopt search stoppage criteria (and don't forget to note them in the report)
  - If using database and another complementary search (manual or snowball)
    - Stop when the manual or snowball does not return a certain number of selected studies (i.e., < 4 new studies added to the list of primary studies)
  - Use a time budget (based on funding or time constraints) and create a list of reviewed studies and a list of not considered studies

ROAR

# Inclusion and Exclusion

- Similar to SLRs we need to define inclusion/exclusion criteria to select our primary studies.

- SMSs usually have less restrictive criteria than an SLR, and tend to focus on:
  - The relevance of the topic of the article -> does it answer the research questions
  - The venue of publication and type of publication
  - The time period considered (typically 10 years, usually excluding the current partial year)
  - Requirements on evaluation (avoid if research has not yet reached maturation for evaluation)
  - Restrictions with respect to language

# Inclusion/Exclusion Examples

**Inclusion Criteria**

- "English peer-reviewed articles in conferences or journals published until Dec. 2011"
- "Articles that focus on software product lines"
- "Articles that provide some type of evolution of existing software artifacts, included among the terms selected"
- "Context, objectives, and research method are reasonably present"

**Exclusion Criteria**

- "Articles that are not related to software product lines"
- "Articles that do not imply evolution of any software artifact"
- "Non-peer reviewed publications"
- "Articles that are not written in English"
- "Context, objectives, or research method are manifestly missing"

# Study Selection

To improve the reliability of the selection process there a few approaches

- Objective criteria for decision (e.g. measure agreement)
  - Measure agreement using Cohen's kappa measure of inter-rater agreement
  - Use Think-Aloud procedure to raise level of agreement to a specified mark
- Decision rules
  - Include papers any reviewer has marked "include" or "uncertain" even if it has been marked "exclude" by another reviewer
  - Include only papers that have been marked "include" or "uncertain" by all reviewers, and exclude papers marked "exclude" by all reviewers or marked "exclude" by some and "uncertain" by others.

# Study Selection Process
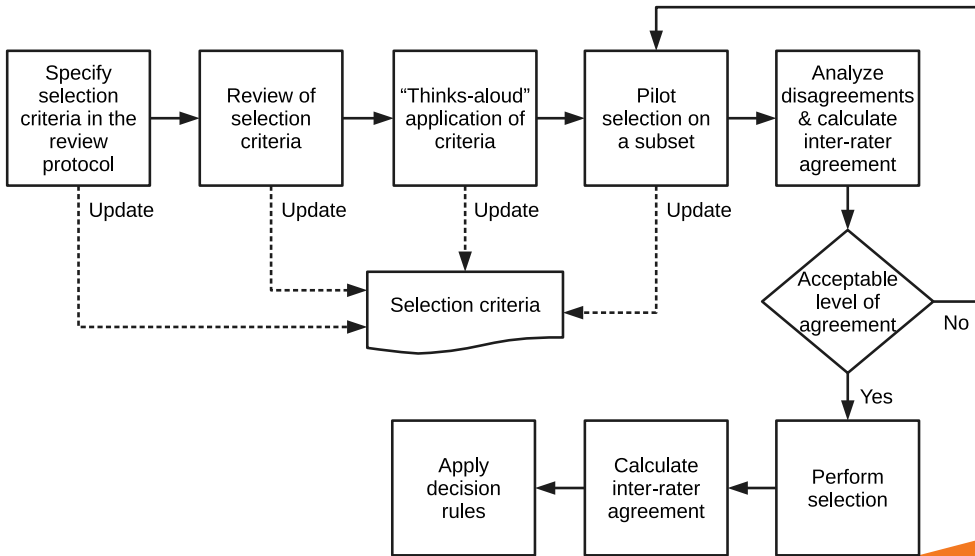
# Quality Assessment

- Not typically done for an SMS like it is for an SLR

- Can be useful to assure that sufficient information is available to actually extract the information

- Note: quality assessment should not pose high requirements for an SMS

ROAR

# **Data Extraction/Classification**

❶ Data Extraction

❷ Classification
- – Topic-independent classification
- – Topic-specific classification

ROAR

# Data Extraction

- Two approaches when there is more than one researcher, the additional researcher(s) may
  1. check the outcome (most common and best strategy)
  2. conduct all the data extractions independently

- If needed a consensus meeting can be held

- Whether agreement can be measured depends on if the papers are classified into a common classification scheme

ROAR

# Topic-independent Classification

Most frequently applied facets for classification

- Venue
- Research Type
- Research Method

Others

- Contribution Type

ROAR

# Venue Classification

- Peer Reviewed
  - Journal article (refereed)
  - Review article
  - Book section
  - Conference proceedings
- Non-refereed
  - Non-refereed Journal Articles
  - Book sections
  - Non-refereed conference proceedings
  - Scientific books
  - Book
  - Edited book (conf proc. or spec iss)

- Professional Communities
  - Trade Journal
  - Articles in professional manuals
  - Professional proceedings
  - Published development or research report
- General Public
  - Popularized article, newspaper
  - Popularized monograph

ROAR

# Venu Classification

- Thesis
  - B.Sc
  - M.Sc
  - MPhil
  - Doctoral Dissertation
- Public artistic and design act
  - Published work of art
  - Public partial realization of art
- Audiovisual material, software
  - Audiovisual material
  - ICT Software
- Patents
  - Granted patent
  - Invention disclosure

# Research Type Classification

- Validation Research
  - Techniques investigated are novel and have not yet been implemented in practice. Techniques used are for example experiments, i.e., work down in the lab.

- Evaluation Research
  - Techniques are implemented in practice and an evaluation of the technique is conducted. That means, it is shown how the technique is implemented in practice (solution implementation) and what are the consequences of the implementation in terms of benefits and drawbacks (implementation evaluation). This also includes to identify problems in industry.

- Solution Proposal
  - A solution for a problem is proposed, the solution can be either novel or a significant extension of an existing technique. The potential benefits and the applicability of the solution is shown by a small example or a good line of argumentation.

# Research Type Classification

- Philosophical Paper
  - These papers sketch a new way of looking at existing things by structuring the field in form of a taxonomy or conceptual framework.

- Opinion Paper
  - These papers express the personal opinion of somebody whether a certain technique is good or bad, or how things should been done. They do not rely on related work and research methodologies.

- Experience Paper
  - Experience papers explain on what and how something has been done in practice. It has to be the personal experience of the author.

# Contribution Type

- Method
- Metric
- Model
- Process
- Tool

# Research Type Classification

| | R1 | R2 | R3 | R4 | R5 | R6 |
|---|---|---|---|---|---|---|
| **Conditions** | | | | | | |
| Used in practice | T | | T | F | F | F |
| Novel solution | | T | F | | F | F |
| Empirical Evaluation | T | F | F | T | F | F |
| Conceptual framework | | | | | T | F |
| Opinion about something | F | F | F | F | F | T |
| Authors' experience | | | T | | F | T |
| **Decisions** | | | | | | |
| Evaluation Research | X | | | | | |
| Solution Proposal | | X | | | | |
| Experience Paper | | | X | | | |
| Validation Research | | | | X | | |
| Philosophical Paper | | | | | X | |
| Opinion Paper | | | | | | X |

# Research Method Classes

**Evaluation Research**

- Industrial Case Study
- Controlled experiment with practitioners
- Practitioner targeted survey
- Action research
- Ethnography

**Validation Research**

- Simulation as an empirical method
- Laboratory experiments (machine or human)
- Prototyping
- Mathematical analysis and proof of properties
- Academic case study (e.g. with students)

ROAR

# Topic-specific Classification

- A classification scheme related to the topics can be one of:
  - Emergent from the study and selected papers
  - Based on existing literature

- If available it is useful to start with an existing baseline classification
  - Provides for comparability between mapping studies
  - Can use IEEE, ISO/IEC, or Swebok classifications

- Additionally, we can extract keywords from abstracts, titles, author keywords, introductions and conclusions to construct a new classification scheme
  - This works using open coding to assign labels to papers
  - We then structure these labels/codes merging and renaming as necessary
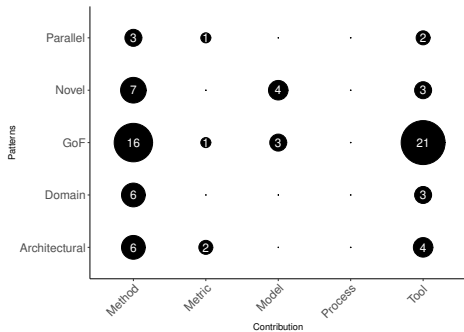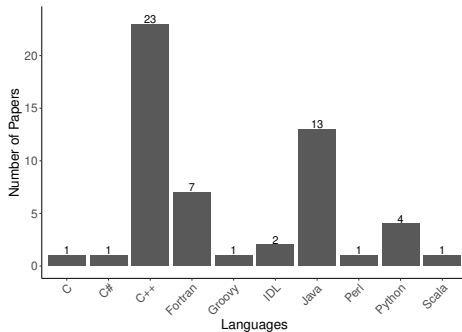  - We then count the numbers of papers with labels and plot them on a bubble plot.

ROAR

# Keywording

ROAR
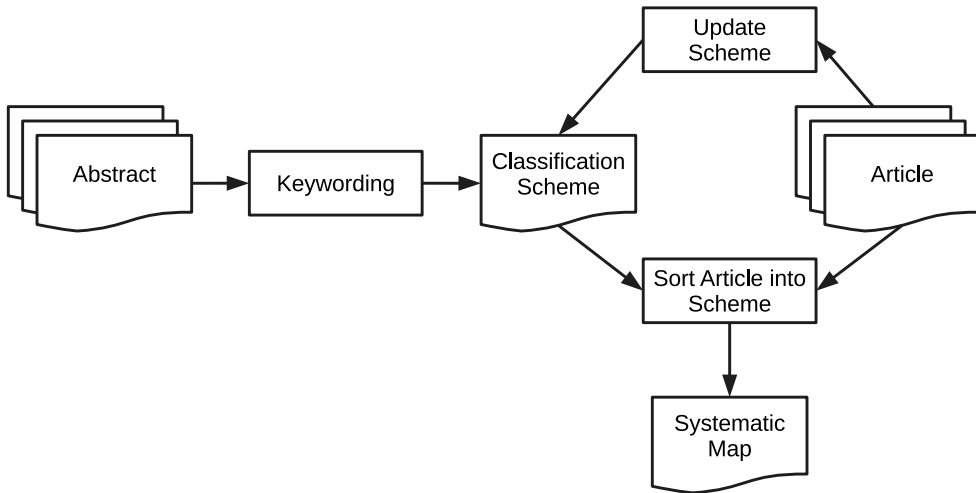
# Visualization

- Most common visualizations are: Bubble Plots, Bar Charts, and Pie Charts, but all visualization types are useful.



**Bubble Plot**



**Bar Chart**

# **Threats to Validity**

- Descriptive Validity - The extend to which observations are described accurately and objectively.
  - Poorly designed data extraction forms and recording of data

- Theoretical Validity - Determined by our ability of being able to capture what we intend to capture. Confounding factors such as biases and selection of subjects play an important role.
  - Publication Bias, as negative or new and controversial views may not be published
  - Potential research bias in the selection of studies and reporting of the data.
  - The quality of the sample of studies obtained with respect to the targeted population.

- Generalizability
  - External Generalizability - generalizability between groups or organizations
  - Internal Generalizability - generalizability within a group

ROAR

# Threats to Validity

- Interpretive Validity - when the conclusions drawn are reasonable given the data, and hence maps to conclusion validity.
    - A Threat is Researcher Bias.

- Reliability - reliability of the conclusions drawn in relation to the data collected, e.g. due to a possible bias of the researchers in the interpretation of that data

- Repeatability - Requires detailed reporting of the research process.

# Conducting the Mapping

- Execute the process defined during the planning phase
- Remember to record the information at all stages
- Note that this process is iterative and may require one or more revisions
- It is useful to use tools to record data
  - Spreadsheets
  - Reference management tools

# **Reporting the Mapping**

Structuring your SMS report:

- Introduction
  - provide information on the background of the topic studied
  - describe the need for the mapping and highlight the usefulness
- Related Work
  - provide and overview of existing secondary and tertiary studies in the area
- Research Method
  - present the following in separate subsections
    - research questions
    - search
    - study selection
    - data extraction (and quality assessment)
    - analysis and classification
    - validity evaluation

ROAR

# Reporting the Mapping

- Results
  - present the outcomes of the study and structure the section with respect to the mapping questions

- Discussion/Conclusions

- Appendix
  - present the included and excluded borderline papers

ROAR

# Evaluate the Mapping

- Use the checklists provided by Petersen et al (2015)
  - Improve the quality of your report
  - Ensure that you did not miss anything
  - To evaluate other SMSs

ROAR

# Are there any questions?

ROAR