

Research Design, and Methodology Spreadsheet: **Social Network Mining from Natural Language Text and Event Logs**

1. Research Methodology:

The general research approach for the *R-AR-Verifier V2* is the design science research approach based on Hevner, March, Park, & Ram (2008). The proposed approach resulted in an iterative process framework consisting of the following steps:

1. *Problem Identification*: The process begins by identifying and defining a specific problem and a corresponding research question for a predefined topic and context. The problem should be relevant and significant, requiring a novel solution.
2. *Objectives and Outcomes*: The outcome should provide clear approaches, framework and technical solutions that solve the problem of 1. and can well be evaluated.
3. *Design and Development*: This involves the creation of innovative artefacts as potential solutions. These artefacts can be systems, processes, models, methods, or algorithms.
4. *Relevance and Rigor*: Relevance ensures that the designed artefact addresses the identified problem, while rigour ensures that the artefact is based on theoretical foundations and is validated through systematic methods.
5. *Evaluation*: The designed artefact is evaluated to assess its effectiveness, efficiency, and impact.
6. *Communication and Dissemination*: Findings are communicated through publications, presentations, and other forms of dissemination. This allows other researchers and practitioners to benefit from the knowledge and insights gained from the research.
7. *Reflection and Improvement*: Agile, iterative approach. Figure out limitations, future work, and reviews/ feedback by experts to always improve or enhance the current solution.

2. Systematic Literature Review

The systematic literature review was based on the structure and guidelines by (Kitchenham, 2004) and can be found in the *Related Work* section of the paper.

The reasons for performing such a systematic review are:

- Summarize existing work/ evidence of evaluated, published tested ideas/ approaches technologies.

- Recognition of research gaps
- Identification and mentioning of new research ideas/ aspects

Process:

The process is based on the steps mentioned by (Kitchenham, 2004) and is illustrated in Figure 1.

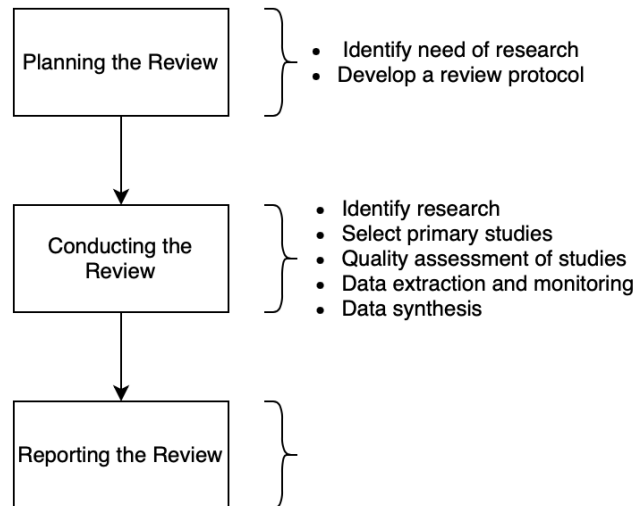


Figure 1: Systematic Literature Review Process
Source: Own creation based on (Kitchenham, 2004)

Execution, Results:

The process of a structured literature analysis was executed separately for each feature implemented in Social Network Mining from Natural Language Text and Event Logs. On the one hand, every feature requires separate research questions, keywords and literature searches on the other hand the goal of this research was to improve the R-AR-Verificator not only from a functional perspective but also from a design perspective. A modular design is created splitting every component into a separate module with a corresponding feature and functional description based on studies, approaches, ideas, and technologies mentioned in the literature search. Thereby the planning of the review is the same for all research topics, only the review conduction differs.

1. Planning:

Table 1: Planning Strategy

Task	Subtasks	Explanation/ Comment
Need for a Systematic Review	Review Objectives	<ul style="list-style-type: none"> - Finding relevant publications for the specific topic - Finding implementation approaches and open source code that can be adapted
	Sources were searched to identify primary studies	Databases: <ul style="list-style-type: none"> - dblp

		<ul style="list-style-type: none"> - Scopus - ACM - IEEEExplore - Springer Link <p>- Google Scholar is used to download additional resources found by the criteria below:</p> <ul style="list-style-type: none"> - Information of Experts in that specific area - Forward Backward search of found literature - Detailed look at high profiled authors in that area
	Inclusion/ exclusion criteria	<ul style="list-style-type: none"> - Papers that have already been downloaded and found in libraries are not downloaded again in the other libraries - Only use conference papers, journal articles and technical reports - Selecting Papers based on title and abstract - Keywords Search string should only be included in title or abstract
	Quality assurance	Quality checklist will be established
	Data extraction	Most important information of the literature will be extracted as well as research gaps will be elaborated
	Data synthesis	-----
	Data combination	Results from the extracted data will be combined logically to answer the research question
Review Protocol	Background	Based on the already existing R-AR-Verificator V1, new functionalities should be developed based on information from literature studies
	Research Question	has formulated in the literature review conduction of every topic
	Strategy	Define the databases to be searched with predefined keywords and search strings
	Timetable	The review is conducted from mid of April to mid of June and split overall tasks to be tackled

Own creation based on (Kitchenham, 2004)

In the systematic literature review generally, **literature is linked to the research question**. It will be tried to **identify research gaps** as well as to **find new ideas for further approach/ system features**.

2. Literature Review Conduction / 3. Literature Review Report

For the literature review conduction a topic, and a research question (RQ) are established. Furthermore, the search time window, the searched libraries as well as papers from my advisor, used keywords and search strings are mentioned.

Topic 01: Organizational Mining as part of Process Mining from Event Logs

RQ: *How to mine organizational information from event logs?*

Topic 02: Organizational Mining of Resources in Natural Language Text

RQ: How to mine organizational information from natural language text?

Table 2: Search Process Documentation of Topic 01

Time	Libraries/ Contacts	Keywords	Search Strings	Resulting Papers
20.04 - xx	<ul style="list-style-type: none"> - Libraries mentioned in the planning document and paper recommendations from experts. - Forward/ Backward search. - Author searches -> for example van der Aalst, Yang, ... Authors with a high amount of publications in that specific area. 	<p>Organizational Mining</p> <p>Social Network</p> <p>Business Process</p> <p>Organizational Structure</p> <p>Graph</p> <p>Process Model</p> <p>NLP</p>	<p>Organizational Mining from Event Logs</p> <ol style="list-style-type: none"> 1. "Organizational Mining" 2. "Social Network" AND "Business Process" <p>Organizational Mining from Natural Language Text</p> <ol style="list-style-type: none"> 1. "Organizational Structure" AND ("NLP" OR "Natural Language Processing") 2. ("Knowledge Graph") AND ("NLP" OR "Natural Language Processing") 	<p>Title: Organizational Mining dblp: 94 papers 7 papers downloaded 7 papers used</p> <p>Title: Organizational Mining Scopus: 134 papers 3 papers downloaded 2 papers used</p> <p>Title: "Organizational Mining" (Brackets have been used here since without brackets 8k results have been outputted) ACM: 10 papers 0 papers downloaded 0 papers used</p> <p>Title: Organizational Mining: IEEEExplore 17 papers 1 paper downloaded 1 paper used</p> <p>Title: Organizational Mining: SpringerLink 0 papers</p> <p>Title: Social Network AND Business Process dblp: 24 papers 2 papers downloaded according to title and abstract 1 papers used</p> <p>Scopus: 26 papers 2 papers downloaded 2 papers used</p> <p>ACM: 0 papers 0 papers downloaded 0 papers used</p> <p>IEEEExplore 5 papers 0 papers downloaded 0 papers used</p> <p>SpringerLink 66 papers 0 paper downloaded 0 papers used</p> <p>Title: Organizational Structure AND NLP OR Natural Language Processing dblp: 1 paper</p>

				<p>1 paper downloaded 1 paper used</p> <p>Scopus: 1 paper (same paper as from dblp) 0 paper downloaded 0 paper used</p> <p>ACM: 0 paper 0 paper downloaded 0 paper used</p> <p>IEEEExplore: 0 paper 0 paper downloaded 0 paper used</p> <p>SpringerLink: 1 paper (Appice 2017 paper) 0 paper downloaded 0 paper used</p> <p>Title: Knowledge Graph AND NLP OR Natural Language Processing</p> <p>dblp: 31 paper 3 paper downloaded 1 paper used</p> <p>Scopus: 35 paper 2 paper downloaded 1 paper used</p> <p>ACM: 2 paper 0 paper downloaded 0 paper used</p> <p>IEEEExplore: 5 papers 1 paper downloaded 1 paper used</p> <p>SpringerLink: Could not be done fully 0 papers</p>
--	--	--	--	---

References

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2008). Design science in information systems research. *Management Information Systems Quarterly*, 28(1), 6.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004), 1-26.