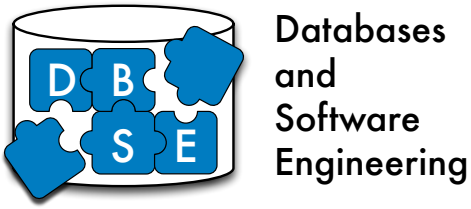
University of Magdeburg

School of Computer Science



Master's Thesis

Supporting Execution and Reporting of Systematic Literature Reviews

Author:

Cheryl Mariam Jacob

March 11, 2019

Advisors:

Prof. Dr. rer. nat. habil. Gunter Saake

Prof. Dr.-Ing. Thomas Leich

M.Sc.Yusra Shakeel

Department of Computer Science

Jacob, Cheryl Mariam:
Supporting Execution and Reporting of Systematic Literature Reviews
Master's Thesis, University of Magdeburg, 2019.

Statutory Declaration

I hereby declare that I have written this master thesis independently without external help and have used only those sources and aids that are mentioned in the references of the work. Literature and definitions which were obtained from other works such as journals and papers are labelled with their source.

11. March 2019

.....
Signature

Abstract

Researchers conduct Systematic Literature Review (SLR) to extract all the existing evidence in their area of research. There is a rapid growth in the number of studies published in the field of software engineering in the past years. SLR helps the researchers to identify and synthesize all the existing evidence relevant to the research question(s). Despite all the effectiveness, SLR is time-consuming and requires a lot of effort. Hence there is a rise in the need for automatic approaches that help the researchers with their study. Data extraction and summarization is one of the crucial steps in the process. Hence we conduct a systematic literature review to gather all the related information and analyze the different approaches used in automating the data extraction and summarization phase. Based on the findings of the review, a semi-automated approach is developed as a prototype to extract and summarize the data found during a review. The validity of the results is examined based on the performance measures such as precision, recall and Recall-Oriented Understudy for Gisting Evaluation (ROUGE) scores. The results show that the extraction and summarization can be partially automated. However, there are some limitations to the approach that has to be further investigated to provide better results to the researchers.

Acknowledgements

First of all, I would like to thank Prof. Gunter Saake for providing me with the opportunity to work on my master thesis within his working group.

It gives me great pleasure in acknowledging Prof. Thomas Leich for his inputs and support throughout my thesis.

I am deeply grateful to my supervisor Yusra Shakeel for her guidance on the topic and for helping me structure the thesis in the right way. I also thank her for her constant support with multiple meetings and interactions that helped me in building the concept of the thesis.

Last but not the least I thank my parents and my husband for their continuous support from the beginning of my master's studies, their prayers have helped me a lot to achieve my goals.

Contents

List of Figures

List of Tables

List of Code Listings

Acronyms

BOW	Bag of Words
EC	Exclusion Criteria
FN	False Negatives
FP	False Positives
HTML	Hypertext Markup Language
IC	Inclusion Criteria
IDF	Inverse Document Frequency
LSA	Latent Semantic Analysis
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
PDF	Portable Document Format
QA	Quality Assessment
regex	Regular Expression
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
RQ	Research Question
SLR	Systematic Literature Review
SVD	Singular Value Decomposition
TF	Term Frequency
TF-IDF	Term Frequency - Inverse Document Frequency
TN	True Negatives
TP	True Positives

VTM Visual Text Mining

1. Introduction

SLR is a well-defined methodology used to gather all the available evidence that address the defined RQs [A. Kitchenham, 2004]. SLRs were already being used and proven in the medical field. Performing SLRs in the software engineering domain require different guidelines, since it is generally less rigorous and has relatively lesser empirical research when compared to the research in the medical field [A. Kitchenham, 2007]. The traditional review methods might not be comprehensive and hence can introduce a persistent bias [Mallett et al. [2012]. SLR is emerging as a useful methodology in this regard by reducing the bias [Mallett et al., 2012]. A. Kitchenham [2007] mentioned that an SLR mainly consists of three phases: planning, conducting and reporting. SLRs could be complex since a large number of primary studies are being investigated and execution would require a significant amount of time and effort [Feng et al., 2017]. Hence, researchers are eagerly involved in proposing approaches to semi-automate different steps of the process.

To perform SLRs efficiently and effectively, different methods are being proposed, which will support the various steps of the process. The data extraction and reporting phase of an SLR involves the crucial tasks of identifying and reporting direct evidence to answer the defined research questions [A. Kitchenham et al., 2009]. The researchers have a tedious and time-consuming task of going through the studies to answer their research questions Carver et al. [2013]. Hence, any approach useful for reviewers to perform these steps effectively can accelerate the process.

1.1 Motivation of this Thesis

Literature review is an important process for researchers and students. Compared to the traditional method of literature survey, SLR ensures inclusion of most of the relevant evidence in the particular topic. It helps in finding gaps in the research and helps in obtaining unbiased results [Mallett et al., 2012]. Poorly conducted reviews can mislead any experimental study [Piper, 2013]. Performing an SLR is difficult and time consuming [Hassler et al., 2016]. Hassler et al. [2016] also identified lack of tool support for data extraction, data synthesis and analysis of qualitative

data. Marshall and Brereton [2013] also states that SLRs can be time consuming, error prone and hence requires tool support. Insufficient tool support has created hindrance in performing SLRs [Zhang et al., 2011]. We identified that not much work has been done on the extraction and reporting phase when compared to the other phases of the process. In the reporting phase, the outcome of an SLR is mainly reported with graphs and charts to show clustering and classification of documents. At the end of the process, researchers have the tedious task of reading all the selected studies. These researches do not emphasize enough about the need to summarize the text by extracting the information needed by the researchers.

1.2 Goal of this Thesis

In this thesis, we focus on studying the state-of-the-art of data extraction and reporting phase. We also propose a semi-automatic approach to perform and evaluate text extraction and summarization. It concentrates on the execution and reporting of SLRs, where we study the various methods of text mining that have been proposed to support meta-data extraction, extraction of particular sections of the paper and summarization of the information by using keyword search, inclusion/exclusion criteria and quality assessment criteria. This thesis aims at answering the research questions in the summarization process. The quality of the summary is assured by including the quality assessment criteria in the summary. Hence, it creates a multi-document summary of the results.

The goal of this thesis is organized and explicated as follows:

- Determine the current state-of-the-art of extraction and reporting phase of SLRs. The aim is to study the methodologies that were previously proposed to support the data extraction and reporting of SLRs. We also identify and work on the areas that need further improvement.
- Propose a semi-automatic approach to support the steps involved in extracting and reporting the results of SLRs. The approach consists of three tasks. The first task involves implementing a web application to extract the predefined sections of the paper. The data extraction is done by using keyword search and regular expressions to extract predefined sections of the paper. The second task consists of extracting the meta-data of a particular study (author name, title, creation date, publisher, etc.) which is done by using python scripts. The extracted data is stored in the form of a table for further reference. The next task is to perform text summarization on the selected studies to get an overview of the paper contents. The summary is extracted to ensure that the various quality assessment criteria stated by the researchers are satisfied. It also creates a general summary of the paper to extract the most vital contents of the document. The extracted data along with the text summary is then reported on the web application.
- Evaluation of the extraction of predefined sections will be carried out by calculating the precision and recall values. Evaluation for meta-data extraction is done manually. Evaluation for the summarization process is carried out

by using intrinsic and extrinsic methods. The intrinsic evaluation is done by using a question game. These questions are based on the quality assessment criteria stated in the SLR process. The scoring is based on the percentage of questions correctly answered in the summary. The extrinsic evaluation is done by selecting an SLR study that has been already conducted. We extract the primary studies used in that SLR as the input to our tool. The multi-document summary generated by the tool will be compared with the results stated in the SLR using ROUGE scores.

1.3 Structure of the Thesis

In **Chapter two** we provide the background information. We describe the various fundamental research concepts concerning our research topic before building the concept of the thesis. We explain the entire process of SLRs in detail along with other concepts, such as, Text mining and Natural language processing.

In **Chapter three** we describe the literature review and related work. We explain how the systematic literature review was carried out in each step with all the details required for further research or repetition purposes.

In **Chapter four** we provide details of the developed approach. The implementation is divided into three parts. The extraction of the meta-data, extraction of predefined section, and data summarization. This implementation supports the SLR methodology by implementing a semi-automatic approach to support the data extraction and reporting phase of the SLR.

In **Chapter five** we explain the evaluation process. In this chapter we explain the evaluation of the different parts of our implementation. Experiences and problems faced during implementation is also discussed.

In **Chapter six** we summarize the overall findings of this thesis by interpreting them and showing how we have succeeded in adhering to the goals mentioned in the introduction.

In **Chapter seven** we mention the future work that can be performed to improve our proposed approach.

2. Background

In this section, we describe the context to the problem that is addressed in this thesis. It is used to appropriately position the direction of our study. It explains the entire process of SLR in detail along with other concepts that support our research, such as, text mining and natural language processing. We also explain some of the text summarization algorithms and the similarity measures used during the summarization process.

2.1 Systematic Literature Review

A. Kitchenham [2007] introduced the guidelines for performing SLRs in software engineering. It follows a set of systematic steps to find all relevant papers related to the topic of interest. It is the process consisting of aggregating, checking, reviewing the data from the already existing evidences. Unlike the traditional literature review which may omit some important research, SLR makes sure that all the available relevant studies are included [Okoli and Schabram, 2010]. In an SLR, the RQs are formulated to gather all the evidence that helps in answering the questions. SLR uses transparent approaches to make sure that the results are unbiased. In addition to answering the questions, the results will describe the state-of-the-art of literature, limitations in the existing study, and guide the researchers in planning further research. Another important aspect of SLRs is that the researchers can repeat and reproduce it for further research.

A. Kitchenham [2004] highlighted the following key points of SLR:

- It makes sure that the results of the review are unbiased.
- The steps of the literature are mentioned beforehand. It also mentions the Inclusion Criteria (IC)/Exclusion Criteria (EC) and quality checks to ensure good quality of the literature review.
- It helps in finding the gaps in research and bridging them so that further research can be carried.

- The goal of the review is to gather all the evidence by summarizing the process and reporting it.
- Poorly conducted systematic literature reviews can mislead the research.
- The RQ must be as specific as possible. A vague question will provide vague answers that might not answer the RQ.
- If the research is a quantitative study, it is possible to combine the results using meta-analytic techniques.
- The RQs are set according to the topic of interest.
- It has a well defined search strategy.
- Explicitly defines the search keywords, boolean search string and the search venues which can be used for further reference.
- The IC/EC are clearly defined so that the research remains focused on the topic of interest.
- QA criteria are set to check the quality or reliability of the information from the primary studies.

The review process consists of three discrete phases as mentioned by A. Kitchenham [2007].

- Planning Phase
- Conducting Phase
- Reporting Phase

The phases mentioned above are explained in the following sub-sections. They are depicted in the Figure 2.1 on page 7.

2.1.1 Planning Phase

The first step towards a scientific review is identifying the RQs. It comprises of the researcher finding gaps in the area of the research and to come up with the RQ that needs to be answered [A. Kitchenham, 2007]. The researcher must ensure the need to perform the SLR on the stated RQs by performing an initial scoping review. Before starting the SLR process, the researcher must ensure that there is no other SLR with the same RQs. The planning phase makes sure that there is enough evidence available to conduct the research. The planning phase is very important to ensure the legitimacy of the results. Hence, a review protocol containing all the necessary data for the further steps is created. This predefined review protocol is necessary to remove all kinds of bias from the review and to make sure that other researchers can repeat and reproduce the SLR. To support the idea, a systematic literature review is conducted by adhering to the following steps.

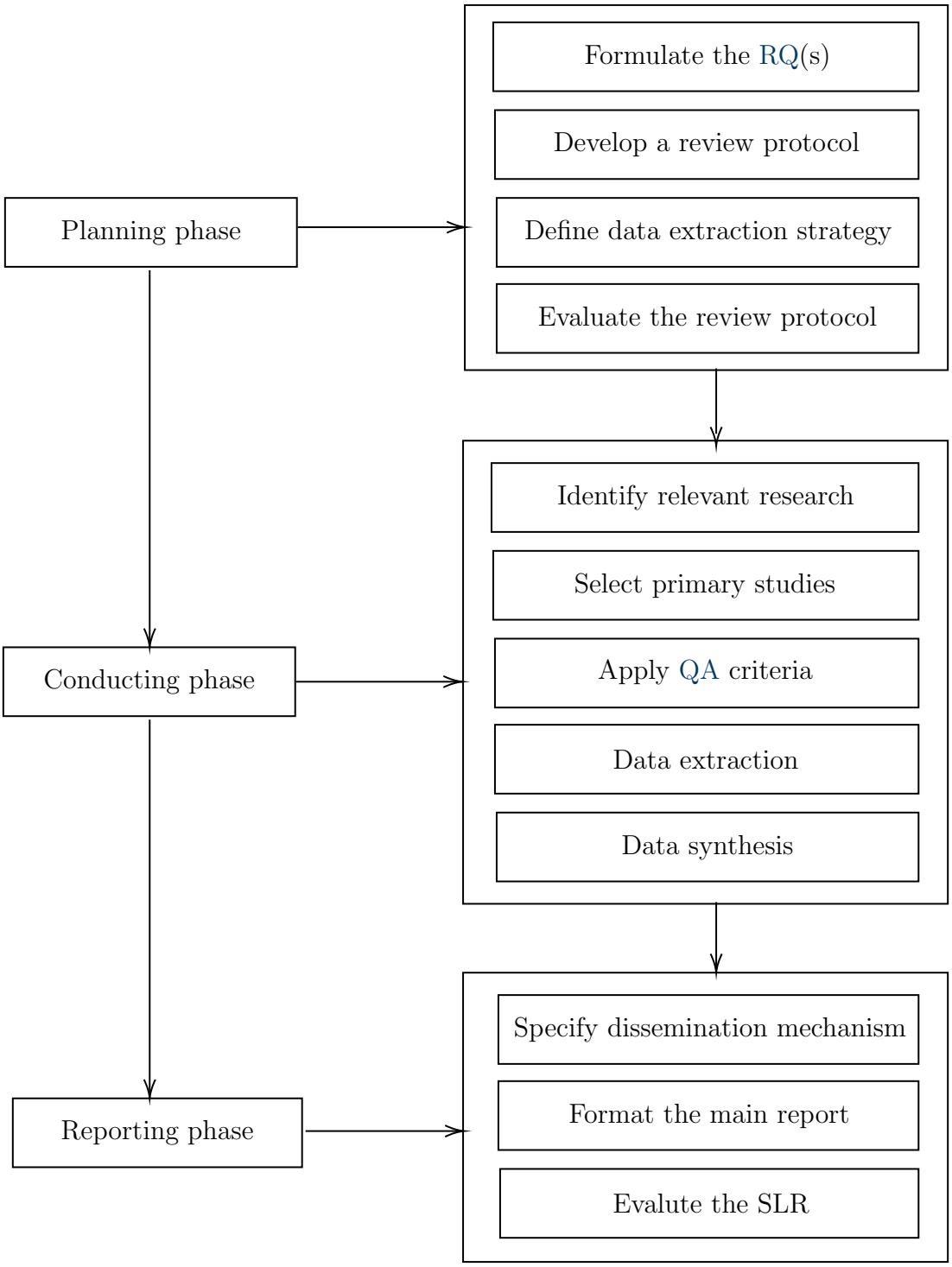


Figure 2.1: SLR phases, adapted from A. Kitchenham [2007].

Formulate the RQ(s):

A well formulated RQ will guide the SLR in answering the question the researcher needs to answer. They are formulated from the problem statement found from the initial scoping review. A study can have multiple RQs. Usually, the main question is broken down into multiple sub-questions. Specifying the RQ is the first methodological step that has to be undertaken by the researcher. To ensure the quality of the RQ, some guidelines have to be followed while formulating the RQ. The following guidelines are adapted from Mason [2014]:

- Clear: It must leave no space for ambiguity or confusion.
- Focused: It should be narrow enough to answer the underlying problem statement.
- Concise: It should be briefly written in few words. It should also be comprehensive by dealing with all the aspects of the study.
- Complex: It shouldn't be too simple which can be answered just by a normal search engine. It should involve investigation and evaluation by the researcher to arrive at the conclusion.
- Arguable: It must be open to debate and disagreement.

Develop a Review Protocol:

It is the plan that describes the proposed procedure to conduct the review. It should be planned before starting the review. It guides the researcher throughout the entire process of review. A review protocol is necessary to reduce bias in the review process and also to restrict overlap with already existing reviews. The different stages of the review protocol are mentioned below:

- **Define the Search Strategy:**

  The search strategy defines the process of searching the studies relevant to the RQ [Aromataris and Riitano, 2014]. It can be an organized structure of the predefined terms used to search a database which consists of the keywords from the RQs to retrieve the relevant [Vassar et al., 2016]b. The search strategy can be a keyword search, boolean search string, searching for exact phrases, wild-card search, citation search [Aromataris and Riitano, 2014]. The search is done either on the full text or by having restrictions to the headline or abstract of the document. The search strategy ensures that the publication bias is reduced as much as possible [Aromataris and Riitano, 2014]. Publication bias occurs when positive findings are more likely to be published faster than negative findings [Begg and Berlin, 1988]. Hence, a literature review with only published data might be biased. Therefore, it makes sure that the results include unpublished reports as well. The entire process is made transparent and documented carefully so that the results can be used for further research.

- **Defining Keywords:**

  The most important words from the RQ that best describe the RQ are defined as keywords. These words are believed to express the meaning and sense of the RQ. While defining the keywords, their synonyms are also taken into consideration. It is always a good practice to search using related words Mian et al. [2005]. This makes sure that the keywords retreive all the studies that are related to the RQ.

- **Defining the Boolean Search String:**

  SLR emphasizes on the usage of the boolean search string in the search strategy [Hassler et al., 2016]. The boolean operators such as "AND" , "OR" , "NOT" are combined along with the keywords to define the meaning of the keywords in order to produce more relevant results. It helps the researchers to immensely narrow down the search results [Dinet et al., 2004]. Boolean search strings can be used in any search engine including the digital libraries.

- **Selecting the Search Venues:**

  Search venues are the digital libraries that contain a huge collection of online databases. Digital libraries provide us with more authoritative and credible results that are not easily accessible on a normal search engine like Google [Behrends, 2012]. IEEE Explore, Springer Link, ACM Digital Library, ResearchGate, Google Scholar are few of the digital libraries.

- **Stating the IC/EC:**

  IC and EC are required to rule in or rule out the results of a search [Salkind, 2010]. IC define the subject that the study needs to address compulsorily. It can also define the characteristics of the study. For example:

  – The study must be conducted on a particular year

  – The study describes a particular topic.

  EC define the characteristics that shouldn't be present in a study while choosing the results. For example, ignore the studies that are not written in English. The studies found from the boolean search are analyzed to check for these criteria. Only those studies that comply with these criteria are selected.

  Stating the QA criteria:

  It is a checklist of factors that need to be addressed in the study. These factors help us in extracting good quality information by stating factors such as:

  – The aim of the study.

  – Limitations of the study.

  The researchers need to manually check the quality of the result. Each result is given a score based on how well they adhere to these criteria. A minimum acceptance score is set to determine the quality of the study. The quality of the study is evaluated based on this QA score.

Define Data Extraction Strategy:

The selected primary studies should be processed to find the relevant information that answers the RQs. The data extraction strategy should be carefully defined since it is an important step in extracting the relevant information. The researchers define the process of extracting information to avoid any bias in further steps [Petersen et al., 2015]. The data can be extracted textually or graphically depending on the nature of the results. Quantitative data can be analyzed and depicted graphically. The qualitative data is usually presented textually but it can also be depicted graphically by finding relations between various data and classifying them accordingly into clusters, for example, document maps.

Evaluate the Review Protocol:

Evaluating the review protocol is one of the crucial steps in the process. The findings are aggregated and reviewed by the internal and external members to validate the correctness of the protocol. The validity of the protocol is checked by the review team as well as external reviewers who hold expertise in the domain [Keele et al., 2007]. Validation includes checking the search string, the IC, EC, and the QA criteria [Tacconelli, 2010]. This reduces the possibility of researcher bias since the protocol is checked to cover all the information required to conduct next stages of an SLR.

2.1.2 Conducting Phase

This section depicts the actual implementation of the review. The protocol that is set during the planning phase is followed in the conducting phase. The search is carried out following the well-defined set of procedures. The conducting phase involves identifying the relevant research, selection of primary studies, and applying the QA criteria. After the results are set, data extraction is carried out by the researcher as the final step in the conduction phase.

Identify the Relevant Research:

The previously mentioned steps are followed to identify the relevant studies. The search strategy mentioned is used as defined in the review protocol. As a result of this process, relevant studies are found from the defined search venues. Snowballing methods are also used on the studies to ensure that no relevant studies are missed out [Badampudi et al., 2015]. This step ensures the completeness of the review protocol. There are two types of snowballing that has to be conducted.

- Forward snowballing: The researchers check for studies that have cited the selected studies in their research. This ensures that all the recent findings related to the selected studies are documented [Badampudi et al., 2015].
- Backward snowballing: This process looks into the references of each paper to find pieces of evidence that support the idea of the studies [Badampudi et al., 2015].

The researchers document the entire process with the results in each step to ensure transparency and reuse of the SLR.

Select Primary Studies:

The studies obtained are revisited to check whether they satisfactorily answer the RQs. The researchers read specific parts of the studies such as title, abstract, introduction, and keywords to get a better understanding of the study. The researchers will screen the studies by checking for bias and applying the IC/EC [A. Kitchenham, 2007]. This step needs to be carried out very carefully to avoid missing out on any important evidence.

Apply QA criteria

Applying the QA criteria is one of the challenging steps in the process since there are no definite rules stated. The QA questions have to be set such that it avoids bias and helps in improving the quality of the results [A. Kitchenham, 2007]. The researchers apply these QA questions on the set of selected studies and create a score for each study based on every QA question stated [Zhou et al., 2015]. The scores are summed up to get a final total which depicts the quality of each paper.

Data Extraction

Once the primary studies are selected, the data needed to answer the RQ has to be extracted. The data extraction strategy stated in the planning phase will help us in avoiding the bias. The data is extracted into data extraction forms [A. Kitchenham, 2007]. The quantitative data about the findings during the process is extracted along with the qualitative data which helps the researchers in answering the RQ. The extracted data is carefully checked with the help of external examiners to find whether it has sufficient evidence to answer the RQ by avoiding all kinds of biases.

Data Synthesis

It involves answering the RQ and summarizing the results found during the SLR process. It also outlines the qualitative results obtained and a short description of the abstract and introduction of the study.

- **Answering the RQs:** The final step of the review involves documenting the results of the SLR. These primary studies are used to answer the RQ. This is the main goal of the entire SLR process [A. Kitchenham, 2007]. It is necessary to document the results of the SLR process for future research. It is used by the researchers and other potentially interested parties to study the methodology and answer further questions arising from the report [A. Kitchenham, 2007]. The report is not only disseminated as a journal or conference paper, but it can also be distributed as a short summary leaflet, press release, posters, magazines or web pages so that it is available to the practitioners.
- **Summarization:** It is the process of reporting the findings of qualitative and quantitative data. The quantitative results are summarized in the form of tables and charts whereas the qualitative results are summarized in the form of text [A. Kitchenham, 2007]. This step can be semi-automated by using supervised and unsupervised approaches. The qualitative summarization can be supervised by indicating the vital focus points of the summary. It can also be

unsupervised by allowing the algorithm to create an extractive or abstractive summary.

2.1.3 Reporting Phase

This step defines a structure to document the entire process and the findings from all phases of the SLR process. The findings include the description of primary studies, results of quantitative summaries, details of meta-analysis etc.

Specify Dissemination Mechanisms

SLR can be reported in many ways such as a press report, leaflet, posters, web-pages, magazines or direct communication [A. Kitchenham, 2007]. The mechanism must be decided beforehand while preparing the protocol. It is decided based on the background of the expected users. If the results are intended to be used by the researchers, then it is reported in the form of a technical paper or a journal.

Format the Main Report

The results can be formatted as a journal, conference paper or a technical report such as a thesis. The journal or a conference paper usually consists of limited information and hence the results are also formatted into a technical report to include all the information in detail.

Evaluate the SLR

Evaluation of the SLR requires one or more person with similar competencies to check for validity and suitability of the research for publication. The technical reports and journal are usually peer reviewed by experts in the particular research topic.

2.2 Text Mining

Text mining refers to extraction of meaningful information from text. It helps in finding hidden patterns and trends in the document and ensures structured output [Shinde and Gill, 2014]. Text mining is a multidisciplinary field that involves various other fields, such as, text analysis, information retrieval, machine learning, data mining, clustering, classification, data extraction, semantic analysis, and text summarization [Babar and Rit, 2013]. It deals with unstructured and fuzzy text and hence makes the process of text mining very complex [Babar and Rit, 2013].

The common approach of text mining is to group the documents based on their similarities. The grouping is done by using text mining techniques such as clustering and classification. It also converts the unstructured and fuzzy text into a structured intermediate format which in turn can be used to perform analysis and extract data [Babar and Rit, 2013].

2.3 Text Summarization

Text summarization is a process where we reduce the size of long texts by extracting the important concepts of the document to reduce the time and effort spent on reading and understanding long text [Mani et al., 1999]. Summarization is usually done by using graph based algorithm to select the important sentences [Patil et al., 2015]. The sentences with the highest ranking are selected accordingly. There are various text summarization libraries that can help us achieve the goal.

The summarization can be extractive summarization or abstractive summarization. Extractive summarization is done by identifying the important sentences from the given text and concatenating them to create a summary. Whereas abstractive summarization mimics the paraphrasing of the corpus instead of simply extracting the sentences [Carenini and Cheung, 2008]. This is harder to implement since it creates a condensed human-like summary. The goal is to select those sentences that can create a meaningful summary and help the users to get the accurate gist of the paper. The summary can also be a multi-document summary by summarizing multiple related documents into a single file.

2.3.1 Aided Text Summarization:

Aided summarization can be either Machine Aided Human Summary (MAHS) where the humans take help from the machine or Human Aided Machine Summary (HAMS) where humans check the summary after the machine summarizes it [Nadella, 2015]. Aided text summarization refers to the process where the necessary paragraphs, context or key-phrases required in the document summary are mentioned beforehand [Nadella, 2015]. The key-phrases are ranked by giving them numerical scores. The key-phrases that are above a particular threshold are considered in the summarization process. Aided summarization algorithms use machine learning techniques to compare the given document and a human-generated input. The algorithm should be able to include the most significant sentences by analyzing the human-generated input. There can be features such as "the first and last line of every paragraph should be included", "the images should be included", "the minimum number of words in a sentence" etc.

2.3.2 Unaided Text Summarization:

The unaided algorithms give us the required results but they also carry a lot of overhead and training data. Unaided algorithms remove the need for training data. These algorithms determine the n-grams in the text. According to the rank given to the n-grams central key-phrases are extracted to build the sentences [Mihalcea and Tarau, 2004]. There are various libraries in python that help us in automatic document summarization. The number of sentences extracted is limited by using a threshold. The final summary must be informative enough to include all the features of the text.

N-Grams : An n-gram is a contiguous sequence of elements from a document. These elements can be letters, words, syllables etc. It is named as unigram (if the size is 1), bigram (if the size is 2) or trigram (if the size is 3).

2.4 Automating Text Summarization

Automatic text summarization is a process of automatically reducing the length of texts by preserving their original meaning [Beliga, 2014]. It makes use of data mining and machine learning techniques. The main idea of the process is to extract the information which represents the entire document. This helps the researchers by reducing the time and effort taken to study the different documents. In this section, we describe the various tools and methods used in extractive and unaided text summarization.

2.4.1 Summarization with NLTK

NLTK is one of the methods used to perform automatic text summarization. Loper and Bird [2002] mentioned the following steps used in the NLTK. In the first step, the text is split into sentences. These sentences are pre-processed and tokenized. For each of these tokens, the weighted frequency of occurrence is calculated. The text is sorted based on the sentences having the highest sum of weighted frequencies. The algorithm filters out the words with very low frequencies and very high frequencies to remove the noise. Hence, it creates an unaided summary of the document.

2.4.2 TextRank

Mihalcea and Tarau [2004] introduce TextRank as a graph-based ranking model written based on the PageRank algorithm. PageRank is an algorithm used by Google to rank the pages that show up on the search page. The number of links to the pages and the quality of the pages are checked to rank the pages. Similarly, TextRank uses an algorithm to find the similarity between sentences and find how one sentence could be connected to another as shown in Figure 2.2. This algorithm is completely language dependent and totally unsupervised [Mihalcea and Tarau, 2004]. TextRank helps in summarizing the document by preserving the original meaning of the document [Mihalcea and Tarau, 2004]. It uses undirected and weighted graphs to model the algorithm. It creates a vertex for each unit extracted from the input text. These vertices are connected to each other with edges, the weights assigned to these edges is the measure of the strength of the connections.

Keyword Extraction:

Keyword extraction is the process of extracting the words that best describe the text [Beliga, 2014]. These keywords are extracted by counting the frequency of words. It is one of the most important tasks in processes such as text mining, information retrieval, NLP, and extractive summarization [Beliga, 2014]. The user can define the number of keywords to be extracted. These keywords are always lexical units. The highest ranked keywords are selected to represent the text.

Sentence Extraction:

The process of sentence extraction is similar to the keyword extraction. The similarity between each sentence to every other sentence is calculated. Figure 2.2, adapted from Mihalcea and Tarau [2004] on page 15 shows us the similarity between the sentences in an article. The sentences are ranked based on the similarity and the highest ranked sentence is selected. Recursively the other related sentences are selected.

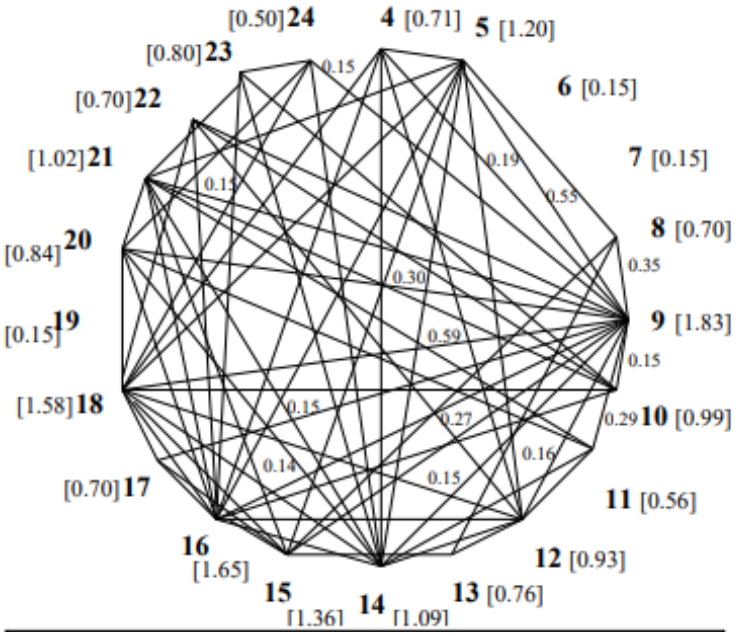


Figure 2.2: Sentence similarity, adapted from Mihalcea and Tarau [2004].

2.4.3 Textum

Torres et al. [2013] defines an algorithm called "Textum" to automatically locate the results from the paper. It makes use of text mining techniques to automate the summarization of unstructured documents. The process consists of 4 stages, which is depicted in Figure 2.3 on page 16. The algorithm imports the text and performs sentence segmentation. In the next step, it automatically finding attributes which helps in classification of the sentences. The algorithm uses 17 different attributes on every sentence to find the highest ranked sentences. These attributes include different measures such as calculating frequency, assigning weights to the keywords, calculating the length of the sentence, finding citations and finding sub-headings of the section. The classification was conducted by using rule-based methods as well as by machine learning approaches. It uses precision and recall to evaluate the outcome of the classification. The summarization is carried out by highlighting the selected sentences in their paragraphs, hence it provides the researchers with an option to read the entire paragraph in order to better understand the concept behind the selected sentences.

2.4.4 LSA Summarizer

Steinberger [2004] makes use of Latent Semantic Analysis (LSA) to identify semantically important sentences. It is a generic text summarization algorithm written to find the similarity between the original document and its summary. It creates an M*N matrix containing the weighted term-frequency vector and the sentences. The algorithm derives a latent semantic structure from the M*N matrix. Singular Value Decomposition (SVD) is applied to this matrix to create independent base vectors or concepts. SVD helps in finding relationships between these vectors to create semantic clusters of the terms and sentences. Recurring patterns will be combined

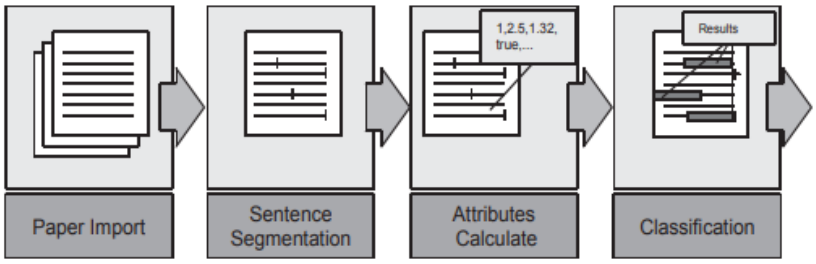


Figure 2.3: Textum algorithm, adapted from Torres et al., [2013].

into a single cluster, so that each singular vector represents a particular concept. The magnitude of these vectors describes the importance of each of the sentences. Hence the most important features are selected in the summarization process.
SVD of an M*N matrix A is calculated in Equation 2.1

$$A = U \sum V^T \tag{2.1}$$

where U and V are left singular and right singular vectors respectively. $\sum$ is a single-valued non-negative diagonal matrix in descending order [Steinberger, 2004]. Steinberger [2004] also stated that the LSA summarizer outperforms other summarizing methods and aims to improve the lemmatization process.

2.4.5 LexRank Summarizer

LexRank summarizer is a graph based algorithm which uses either weighted cosine similarity or TF-IDF function to find the similarity between sentences [Erkan and Radev, 2004]. The algorithm also uses centroid-based summarization which identifies the central sentences that describe the main topic of the document [Erkan and Radev, 2004]. The centroid of each document consists of the words that have high TF-IDF scores. The sentences with the highest number of words from the centroid are selected to create the summary. This algorithm sets a threshold frequency and builds an edge between the two nodes (sentences) if the similarity is above the threshold. The similarities between the sentences are calculated using a modified cosine equation shown in Equation 2.2

$$\text{idf modified cosine(x,y)} = \frac{\sum_{w \in x,y} tf_{w,x} tf_{w,y} (idf_w)^2}{\sqrt{\sum_{xi \in x} tf_{xi,x} (idf_{xi})^2} * \sqrt{\sum_{yi \in y} tf_{yi,y} (idf_{yi})^2}} \tag{2.2}$$

Modified Cosine formula, adapted from Erkan and Radev, [2004]

where x and y represent the two different sentences , $\text{tf}_{w,s}$ is the frequency of occurrences of the word w in the sentence s. idf_w is the inverse document frequency of the word. They also use the PageRank scheme to rank the sentences. The result is the subset of the nodes that have the highest frequency.

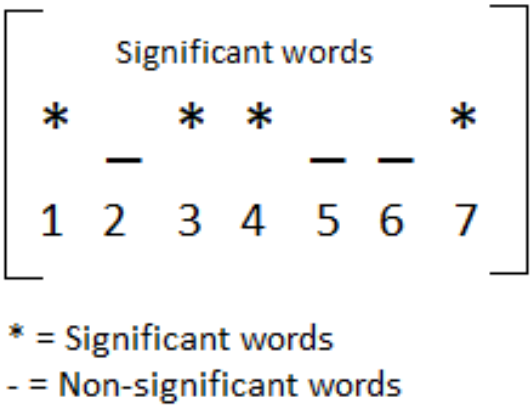


Figure 2.4: Significance Factor, adapted from Luhn [1958].

2.4.6 Luhn summarizer

Luhn summarizer was made to automatically create abstracts for documents [Luhn, 1958]. The document is divided into sentences and the "significance factor" is calculated based on the similarity between the words. The highest ranking sentences are selected in the automatic abstraction process. A limit is set to the distance between two words in order to consider them significant. As a result of this process, multiple clusters are created. The "significance factor" is calculated to establish the extent of the cluster. It is calculated based on the number of significant words in a sentence. A sentence is bracketed by selecting the two significant words at the corners. Only the parts of the sentence that are under the brackets are selected for summarization as shown in Figure 2.4. A limit is set for the number of insignificant words in between the significant words. If the count is beyond the limit, the sentence is discarded.

The "significance factor" is calculated based on the square of the number of significant words in a cluster divided by the total number of words in that cluster, Equation 2.3.

$$\text{Significance factor} = \frac{(\text{Number of Significant words})^2}{\text{Total number of words}} \tag{2.3}$$

2.4.7 Aided Text Summarization Algorithm

The aided algorithm is carried out by taking into account the user inputs which describe the type and topic of the text that needs to be extracted. The supporting input can either be keywords or sentences. If the input is a sentence then the algorithm will find the important keywords from the sentence. The input is compared with the tokenized and parsed text that needs to be summarized. Similarity calculation measures, such as, Cosine similarity, Levenshtein distance and TF-IDF are used to find the most important sentences. The sentences with the highest similarity or the least distance will be selected to create the summary.

2.4.8 Automatic Keyword Extraction

Automatic keyword extraction is the process of automatically extracting words that best describe the document [Beliga, 2014]. They are extracted from a controlled

vocabulary to represent the most relevant information from the document. It uses N-grams and stop-words to control the extraction. Extracting keywords manually can be expensive and time-consuming. Hence, machine learning approaches are used to create this data-set. The keywords can be extracted from the title, abstract, introduction and conclusion. These keywords can be used in building the document index which further helps in classifying and clustering the different sentences of the document [Beliga, 2014]. The extractive summarization mainly uses keyword extraction to build the summary.

2.5 Natural Language Processing

NLP is the field of understanding and generating human created natural language to interact with the computer [Manaris, 1998]. Many machine language algorithms are applied to process the natural language. The input text is analyzed to gather the "features" and a real-valued weight is assigned to each of the features [Khan et al., 2016]. The machine learning algorithm automatically focuses on the most important features to understand the input data. Users can automate things like data summarization, sentiment analysis, and classification by using NLP
The idea of NLP is to split the problem or sentences into smaller pieces to build a pipeline. The text that is split is then tokenized into words or tokens. Each token is fed into a classification model to extract the parts of speech. The processed text undergoes "lemmatization" to understand the most basic form of each sentence. The stop words are identified and filtered out. The next step is to find the dependencies between all the words in our text and group the related sentences together. The final step is the co-reference resolution, which means finding the meaning of pronouns, such as it, he, and she [Geitgey, 2018]. It is a very important step in NLP and hence helps us in creating a very meaningful summary as shown in Figure 2.5.

NLTK:

NLTK is a suite of libraries that help in processing human language data that are used in natural language processing. It helps in performing the NLP tasks such as tokenizing, removing stop words, parsing, classification, and semantic analysis Loper and Bird [2002].

2.5.1 Tokenization

The NLTK provides us with a set of libraries to help us with tokenization. Tokenization is a process where the input text is split into meaningful elements such as sentences, words or paragraphs [Webster and Kit, 1992]. These elements are called tokens. Tokenization can also clean the text by taking care of amalgams and locutions. It is an important step in natural language processing. It is necessary to clean the text before summarization or any other text analysis. This process is also known as lexical analysis that is often combined with a parser.

2.5.2 Parsing

Parsing helps in identifying the exact meaning of a sentence or a document. It helps in defining the structure and relation between different sentences by creating

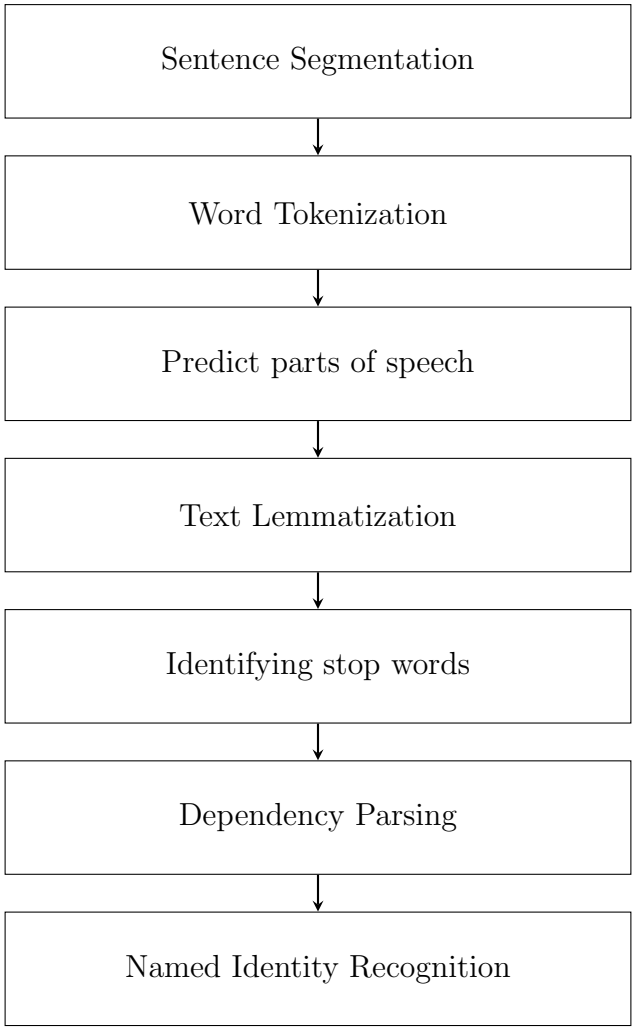


Figure 2.5: Process of NLP.

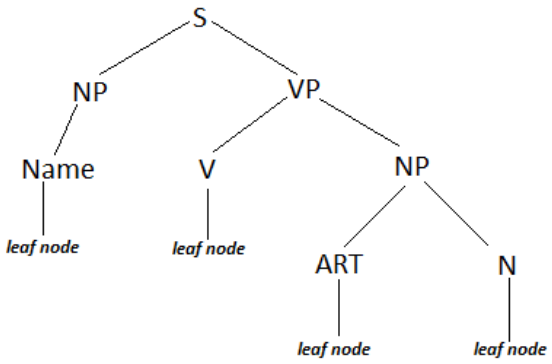


Figure 2.6: Example for a parse tree.

a parse tree that depicts the semantics and syntactic relation between the sentences. The tokenized text is sent to the parser to check whether the tokens form allowable expressions in accordance to the context-free grammar. The parse is then compiled, interpreted or translated to form meaningful output [Kübler et al., 2009].

Parse tree : The parse tree identifies the different parts of speech for every sentence. A single sentence can have more than one parse tree. Each sentence (S) is broken down into noun phrases (NP) and verb phrases (VP) which is further split into respective parts of speech such as noun (N), verb (V), article (ART) and so on. The leaf nodes in the parse tree constitute the words of the sentence as shown in Figure 2.6.

2.5.3 Word Segmentation

Word Segmentation is the process of splitting the input text into separate words. This method is also used in creating Bag of Words (BOW) in data mining. BOW is helpful in feature extraction from text segments. It consists of every word with its respective frequency of occurrence [Webster and Kit, 1992]. Hence it can be used in calculating the term frequency.

Example: BOW = “Word1”:1, “Word2”:2, “Word3”:1, “Word4”:1, “Word5”:2;

2.5.4 Stop Words

Stop words are the non-informative supportive words that are used to build a sentence. Words such as “is”, “are”, “the” are some of the stop words [Geitgey, 2018]. NLTK is programmed to filter such words from the sentences before processing them. Eliminating them reduces the overhead during processing.

2.6 Similarity Measures:

Similarity measure in text mining is the distance between the objects based on their features [Vijaymeena and Kavitha, 2016]. Two sentences are compared to get

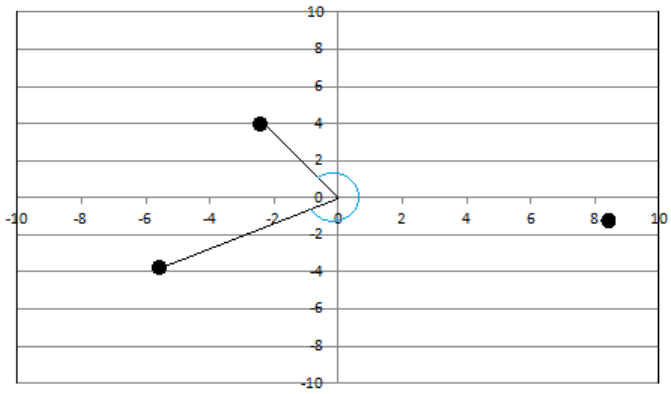


Figure 2.7: Vector space model, adapted from Shulga [2017].

a similarity measure between them. This concept can be used in clustering and classification of the documents or sentences. Cosine similarity is the most common similarity measure that is used in the field of machine learning and text mining.

Similarity Measure between Sentences

The ability to find the similarity between the sentences is a very important task in text mining. Given two sentences the similarity measure should check for semantically similar sentences by considering the natural language constraints. It is used to define a similarity rank between two sentences [Achananuparp et al., 2008].

Cosine Similarity:

The sentences are converted into vectors to form a vector space model. Each vector consists of the vectorized form of a sentence with the corresponding word frequency of each word in that sentence. Each sentence is a point in the vector space model [Achananuparp et al., 2008]. The distance between the two points is calculated by finding the cosine of the angle between them as shown in Figure 2.7. If the angle is 180 degrees, then the cosine of 180 is 0, which means the sentences are least similar. If the angle is 0 degree, then the cosine of 0 is 1, that makes the sentences most similar. Each word represents a dimension of the model. Two sentences (Vector A and B) are then compared with the formula in Equation 2.4.

$$similarity = cos(\theta) = \frac{A.B}{||A||.||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{2.4}$$

Term Frequency - Inverse Document Frequency

TF-IDF describes the importance of a word in the document. It is the weighting factor that has a higher value if the frequency of the word is higher in a document. This measure is commonly used in text mining and information retrieval.

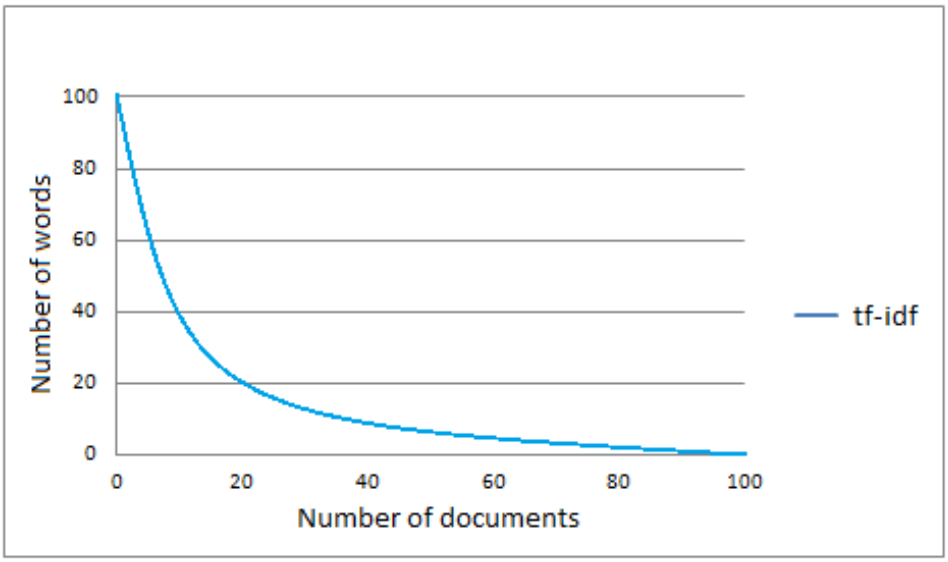


Figure 2.8: Example graph for TF-IDF.

Term frequency is simply the count of a word in a document. Inverse document frequency checks the importance of the word across all documents. It is the logarithmic inverse of the fraction of documents that contain the word divided by the total number of documents [Achananuparp et al., 2008]. TF-IDF shown in Equation 2.5 is the product of Term Frequency (TF) and Inverse Document Frequency (IDF). Hence a higher value of TF and a low-frequency of number of documents contributes to the high value of TF-IDF as shown in Figure 2.8. This helps in filtering out on common terms between documents.

$$tf - idf(t, d, D) = tf(t, d) * idf(t, D) \tag{2.5}$$

D - The set of documents
d - Single document
t - Term

We see that when the number of words are high and the inverse document frequency is low, the TF-IDF score is high and vice versa.

3. Related Work

In this section, we provide details of SLRs we performed that follows the guidelines published by A. Kitchenham [2007]. This literature review focuses on two diverse study types. The first focuses on the semi-automatic methods used in the data extraction and reporting phase of the SLR. The latter focuses on how the QA criteria can be applied to text summarization in a SLR.

3.1 Systematic Literature Review

The SLR consists of three phases namely, Planning, Conducting and Reporting. In the planning phase, we define the RQs, define a search string and state the IC/EC. During the conduction phase, we identify the primary studies, perform study selection, QA and data extraction. Finally, in the reporting phase, we answer the RQs mentioned in the planning phase and document it.

3.1.1 Planning Phase

Planning phase is one of the most important phases that defines and structures the literature review. In this section, we define the RQs, search strategy, IC/EC, QA criteria and the method to be considered for data extraction.

RQs

RQ1: How text mining techniques are used to automate data extraction from the primary studies?

We essentially begin with finding papers that describe different text mining techniques that can help us in the extraction and summarization of the studies. Here we summarize the different text mining methods, such as, clustering, classification, and visualization. We mainly focus on the data extraction and reporting phase of the SLR.

Name	URL
GoogleScholar	https://scholar.google.com
ScienceDirect	https://www.sciencedirect.com
SpringerLink	https://link.springer.com
IEEE Xplore	https://ieeexplore.ieee.org
ACM Digital Library	https://dl.acm.org/

Table 3.1: List of libraries.

RQ2: What are the different methods used to summarize data on a set of primary studies for SLRs in software engineering domain?

The extracted data is summarized in the reporting phase. Reporting is usually done using graphical representations of the data that is needed for further analysis. Here we describe how the different papers summarize or report the data so that it can help the researcher's needs.

Search Strategy

We define the strategy used to perform the search. Automatic keyword search along with snowballing is used to find relevant primary studies.

Search Keywords

We define the keywords that are required to perform the search. The search keywords must include all the keywords from the RQ. This assures that all the relevant papers concerning the RQ are found.

Keywords: Systematic literature review, data summarization, automation, methods, data extraction.

Search String

The boolean search string is written by taking into account the keywords that were defined and all their relevant synonyms. It is written using a combination of boolean operators such as such as AND, OR, NOT, and AND NOT operations and brackets.

String: (Method OR Automate OR Approach) AND ("Data Mining" OR "Text Mining") AND ("Data Summarization" OR "Text Summarization" OR "Data Extraction") AND ("Systematic Literature Review" OR SLR)

Search Venues

The boolean search string is used in the different digital libraries to find the relevant studies. Table 3.1 on page 24 depicts the selected libraries.

IC and EC

Here, we define the various criteria that must be considered before selecting the primary studies. The criteria are written based on the type of studies required. We make sure that the paper satisfies the IC and EC.

IC Criteria:

- The study must address an approach to automate data extraction/reporting in the SLR process : The focus of our approach is based on the automation of data extraction and hence, we look for studies that support our idea.
- Study is conducted between 2007 to 2018 : The "Guidelines for performing Systematic Literature Reviews in Software Engineering" by A. Kitchenham [2007] was published in the year 2007. This study is considered to be the foundation for SLRs in the software engineering domain. Therefore, we find studies conducted after 2007.
- Paper reviewed and officially published in a journal/workshop/conference : Since the papers are officially reviewed, the results will be more precise and proven.
- Articles related to software engineering domain : Since our literature review focuses on the software engineering domain, we filter out the other prominent domains, such as, the medical field.

EC:

- Papers not written in English : Our study and implementation is focused on data extraction in English. Since the studies and tools in other languages will not help our study.
- Abstracts and PowerPoint presentations : These studies are not structured appropriately and hence lacks detailed information of the topic.
- Technical reports and Bachelor/Master/PhD thesis : Since the studies aren't officially reviewed, there might be scope for irregularities in the result.
- Articles with unknown publication type or publisher : Publications with unknown source cannot be considered true to form.

QA

One of the ways to find the quality of the study is to rigorously check how the studies uphold the QA criteria. A set of questions are framed to check the quality of the primary studies. The checklist consists of questions that answer the different stages of the review which are stated below as mentioned by A. Kitchenham [2007].

- Design

- Conduct
- Analysis
- Conclusions

The following questions were set based on the guidelines mentioned by A. Kitchenham [2007]. We also adapt the checklist written by Dybå and Dingsøyr [2008]. They address the quality criteria in all the different stages stated by A. Kitchenham [2007]. Q1 answers the design stage. Q2 answers to the conducting stage. Q4 answers the analysis stage. Finally, Q3 and Q5 answer the questions regarding the conclusion stage.

QA Criteria:

Q1. Are the aims clearly stated?
If the primary study clearly states the objectives and aims of the study, then the paper is awarded a score of 1. If the aim is not at all mentioned then the score is 0. If the aim is defined but not clear, it can get a score between 0-1.

Q2. If the primary study uses a particular technology, is it clearly defined?
If the primary study uses a particular technology and it is explained clearly, it is awarded a score of 1. If the study does not use any particular technology it is still awarded a score of 1. If the primary study uses a particular technology but it is not explained, it is then awarded a score of 0.

Q3. To what extent does the result mention the external validity or future research?
If the studies can benefit the researchers outside the scope of the current study, then a score of 1 is assigned. Else, it is assigned a score of 0.

Q4. Is the purpose of the study clearly defined?
The study is awarded a score of 1 if the purpose and motivation are clearly defined. If the purpose is not mentioned, then it is awarded a score of 0.

Q5. Are the problems, negative results or the limitations mentioned?
If the paper mentions the negative aspects along with the other pros, then the study is awarded 1. Else it is awarded a 0. Every individual study is given 5 different scores based on each of the quality criteria. These scores are summed up to get the final score (Total=5) of each study. Higher the score, better is the quality of the study.

Data Extraction

The primary studies that are selected are used for extracting data from the predefined data fields. The data extraction is mainly done using text mining methods like clustering and classification. This helps us in finding trends and similarities between the documents.
The extracted data is reported using graphical representations. It can be presented in the form of various charts or tables making it easier for researchers to derive a better conclusion.

Attributes Extracted
Title
Authors
Year
Abstract
Introduction
Conclusion

Table 3.2: Data extraction.

Library	Number of Studies found	Number of studies selected
Springer	72	0
Science Direct	243	0
Google Scholar	1330	3
IEEE	742	0
ACM	5726	0
Total	8113	3

Table 3.3: Number of selected studies.

3.1.2 Conducting Phase

In this section, we demonstrate the implementation of the methodology that was described in the planning phase. We look into the details of how the primary studies were selected and processed through the selection process.

Identifying Relevant Research

The search was conducted in various online libraries using the search string that was defined in the planning phase. We found a total of 8113 results in the libraries listed in the Table 3.3. The results were intensively searched manually to check for all the IC/EC to derive the final set of primary studies.

Selection of Primary Studies

The primary studies that satisfy all the search criteria were mainly found on Google Scholar. The studies derived were then used for snowballing to find more relevant papers. The result of this search can be found in Table 3.4 in page 29. Table 3.4 contains the results of various steps, such as, entering the boolean search string in the libraries, selection of papers based on keywords, performing snowballing and checking the IC/EC. This entire process with the number of results in each step is depicted in the Figure 3.1 in page 28.

QA

To check the quality of the study that has been selected, we apply the QA criteria that were previously mentioned in the methodology stated by A. Kitchenham [2007].

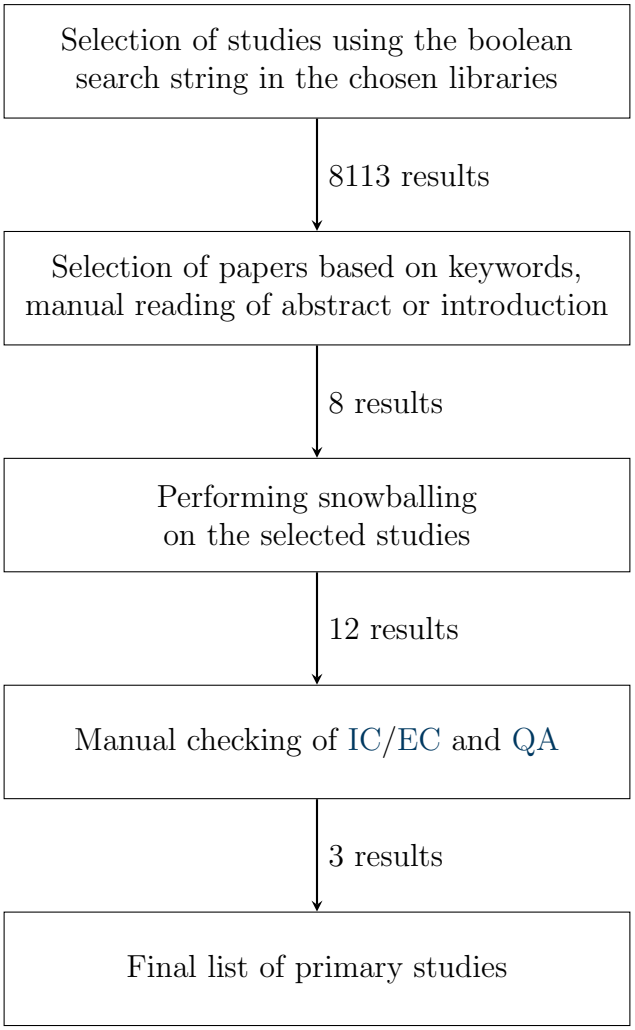


Figure 3.1: Selection of primary studies.

ID	Primary Study Name	Publisher	Year	Citation
PS1	Using Information Visualization and Text Mining to Facilitate the Conduction of Systematic Literature Reviews	Springer Berlin Heidelberg	2013	[Fabbri et al., 2013]
PS2	SLuRp – A Tool to Help Large Complex Systematic Literature Reviews Deliver Valid and Rigorous Results	EAST'12 - Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies	2012	[Bowes et al., 2012]
PS3	SLR-Tool - A Tool for Performing Systematic Literature Reviews.	ICSOFT 2010 - Proceedings of the Fifth International Conference on Software and Data Technologies, Volume 2.	2010	[Fernández-Sáez et al., 2010]

Table 3.4: List of primary studies.

Each study that has been selected must have a certain degree of agreement to the QA criteria. Each question gets a score in between 0-1. The primary study is ranked based on a total score of 5. Higher scores reflect better quality of studies. The Table 3.5 on page 30 shows the primary studies along with their scores. The scores are calculated based on the QA criteria defined previously.

Each primary study has a score in the range of 4.0 - 5.0 out of 5.0 which illustrates how well the primary studies conform to each of the QA criteria. Each study shows a small amount of disagreement to at least one of the QA criteria.

According to Fabbri et al. [2013] there are no limitations or negatives mentioned in the entire scope of the paper. Hence it is assigned a score of 0 to Q5. The study by Bowes et al. [2012] hasn't clearly mentioned how the SLuRp tool can be used outside the scope of the paper. Therefore it receives a score of 0.5 for its external validity. Similarly, the study by Fernández-Sáez et al. [2010] also doesn't clearly state its external validity and hence receives a score of 0.5. All the other quality criteria are satisfied completely and hence received a score of 1.

Data Extraction

We extract the meta-data of the study which includes data about the publications and the details of the authors. We also extract specific sections of the study, such as, abstract, introduction and conclusion to get a better idea about the paper. Since this review mainly focuses on finding the text mining techniques used in SLR, we extract sentences from these studies to formulate the summary.

3.1.3 Reporting Phase

This section describes a detailed answer to the RQs that were stated during the planning phase. The 3 papers that satisfied all our criteria is carefully studied to

ID	Primary Study	Q1	Q2	Q3	Q4	Q5	Total(/5)
PS1	Using Information Visualization and Text Mining to Facilitate the Conduction of Systematic Literature Reviews. [Fabbri et al., 2013]	1.0	1.0	1.0	1.0	0.0	4.0
PS2	SLuRp – A Tool to Help Large Complex Systematic Literature Reviews Deliver Valid and Rigorous Results. [Bowes et al., 2012]	1.0	1.0	0.5	1.0	1.0	4.5
PS3	SLR-Tool - A Tool for Performing Systematic Literature Reviews. [Fernández-Sáez et al., 2010]	1.0	1.0	0.5	1.0	1.0	4.5

Table 3.5: QA of primary studies.

answer these questions. In the latter part, it also reports how the entire SLR process was carried out in all the three phases. Based on the report, we identify the areas in the study that need improvement which will be dealt in the further chapters.

3.1.4 Answering the RQs

RQ1: How text mining techniques are used to automate data extraction from the primary studies?

Data extraction is mainly done by using various text mining methods. Here we describe how the different studies use text mining techniques such as clustering, and classification to automate the process of data extraction.

The "StArt" tool proposed by Fabbri et al. [2013] uses text mining techniques like similarity calculation to extract the data. It also deals with a large amount of data to find hidden patterns and trends through visualization. It makes use of information visualization to integrate the information from all the papers. It extracts quantitative data such as the percentage of studies that passed the IC/EC and the number of studies found using the search string.

The study also describes the usage of a text mining tool called "PEx". The documents selected for the data extraction are converted into ASCII format using text from different parts of the articles (title, abstract, references etc). PEx helps in processing the abstract by removing stop words, calculating term frequency and hence, creates clusters of the documents.

The "SLuRp" tool proposed by Bowes et al. [2012] supports data aggregation and statistical analysis using SQL statements. It makes use of online databases to query the necessary studies and uses Java to semi-automate the data extraction. SLuRp provides the users with a coding form to record the extracted qualitative data and

a performance form to record the quantitative data from each study. It is executed by using classification and mappings.

This paper also states another tool by Felizardo et al. [2011] which uses Visual Text Mining (VTM) techniques to automate the clustering and classification of the papers. This tool makes extensive use of text mining and visualization techniques.

Fernández-Sáez et al. [2010] describes the usage of "SLR tool" that uses classification schemes to classify the primary studies to perform data synthesis and analysis. The tool takes care of all the stages of the planning phase, checks for the IC/EC and then uses the classification scheme to allow the researchers to specify the category and subcategory according to their choice. The visual data that is obtained here is used to synthesize the data and extract conclusions out of the primary studies. It also makes use of clustering to find patterns and trends in the studies.

RQ2: What are the different methods used to summarize data on a set of primary studies for SLRs in software engineering domain?

The extracted data is summarized and reported in the reporting phase. Here we describe how the different papers summarize/report the data so that it can help the researcher's needs. Summarization can help in better visualization and also helps in deriving a conclusion for the particular study.

Fabbri et al. [2013] mentions "StArt" as a tool that emphasizes the need for data summarization. This tool also creates a link to the full text of each paper used in the summarization. Here the reporting is done by displaying the data through visualization and Excel formatted reports. The final report that contains the extracted data from the primary studies is presented by using graphics and spreadsheets. "Traceability" is one among the most important topics of interest within the paper. It depicts how the studies are related to one another. The researchers can easily group the studies based on one of the categories such as year, author, and keyword by selecting the appropriate category. The categorization is done by using text mining method such as similarity calculation. For each pair, the similarity is calculated and a similarity score is assigned by using a vector processing model proposed by Salton [1989]. The paper also mentions another tool "PEx" proposed by Maldonado et al. [2007]. It creates a abstract by removing stop words, calculating term frequency. It calculates similarity measures to create a document map by using VTM.

Bowes et al. [2012] describes the tool named "SLuRp". It records the data from all the studies by automatically aggregating and displaying the results into reports. It makes use of two main methods, tabular and graphical. The graphical results are illustrated using different types of charts or plots. The graphical representation brings out the tendency of the paper to incline towards the bias in the results. The tabular results are presented in different ways such as Hypertext Markup Language (HTML) or LaTeX. It is used when there are qualitative data that can be summarized using mathematical operations. The extracted data will be linked to the paper using citation links. It will be also linked to the exact paragraph from where the data is extracted. This helps the researchers to study the entire paragraph when the extracted data is within their scope of interest. The data that is reported usually consists of the percentage/number of papers that pass or fail the quality criteria and IC/EC.

Fernández-Sáez et al. [2010] mention that the summarization or reporting must consider all the information related to all the 3 phases of the review. The SLR tool mentioned here follows the same process. During the summarization, the extracted data is transformed into visual data where it allows the researchers to get a better idea and faster conclusion of the studies. The data is exported into Excel files for further analysis. It creates a general summary of all the documents and also creates summaries according to the categories selected during classification by generating different charts. It further generates reference files such as BibTeX/EndNote which is exported into a document file.

In all the papers, the extracted data is automatically transformed into graphical representations like charts and tables. These graphical representations provide better and faster understanding to the researchers about the primary studies that are being used. They help the researchers in finding the trend and in getting an overall meta-data about the studies being used.

3.2 QA Check

To support our idea of including QA in the summarization, we intended to perform another SLR. A initial review was conducted to see if there exists any research that uses the QA criteria mentioned in the planning phase of the SLR process as the evaluation criteria in the summarization process. The results from all the search venues were carefully studied based on the keywords, manual reading of abstract and introduction. There were no studies that matched our criteria. Based on the outcome of the SLR, it can be said that there are no researches aimed at using the QA criteria during the text summarization. Hence it was not possible to continue the research or to answer the RQ. This helps in building the concept for this thesis. Adding the QA criteria as a measure to evaluate the outcome of the summary can be very helpful in assessing the information that the researchers are looking for. Using the QA criteria as the input, a supervised algorithm can be built to support the need. Henceforth, the focus of this thesis will include the use of QA criteria in the text summarization phase of the SLR.

3.3 Summary

In this section, we illustrated the entire SLR process. We look into the data in each and every step of the process. The initial search was conducted to find relevant studies that described different methods to automate data extraction and summarization. The search which initially started with 8113 results were filtered out to get 3 results that matched our distinct criteria. When we look into the reporting phase of our SLR, we see that summarization is done through visualization or through reporting the extracted data based on the different processes of SLR. We see the need for summarizing the studies textually so that the researcher gets a brief idea about a particular study. We also see the need for answering the QA questions during the summarization. The studies that were found during the SLR process mentions summarization as a part of the SLR process but doesn't emphasize the need and importance to support the automation of this feature.

To evaluate the quality of the summarized data, we have written QA criteria which makes the summary to undergo a rigorous quality check. In the further chapters, we focus on extracting the meta-data, predefined data fields from the studies and also provide a gist to the selected studies.

4. Methodology

In this section, we explain the approach we propose for the semi-automatic extraction and summarization process of the SLR. The implementation is mainly divided into three sub-parts as shown in Figure 4.1. To provide a complete report of the SLR process, we include meta-data extraction, extraction of the user-defined section of the paper and summarizing the entire document in our approach. The summarization is further divided into extractive summarization and keyword based summarization to execute our idea of answering the QA questions in the summarization. This makes sure that all the aspects of the document are reported by creating a multi-document summary for the given set of documents.

To perform SLRs efficiently and effectively, different methods are being proposed to support the various steps of the process. The reporting phase of SLR involves the crucial tasks of identifying and reporting direct evidence to answer the defined RQ. The researchers have a tedious and time consuming task of going through all the included studies to answer their RQ. Hence, any approach useful for reviewers to perform these steps effectively might accelerate the process. In this section, we explain the proposed idea to support data extraction and summarization of the information by using various text mining techniques. The sub-parts mentioned in Figure 4.1 were chosen for the implementation to help the researchers in speeding up the process of conducting SLRs.

- Meta-data of a Portable Document Format (PDF) file is the data that describe the attributes of the file [Brand et al., 2003]. The meta-data helps in extracting the basic information about the document. Milstead and Feldman [1999] mentioned that it also supports cataloging and indexing of the documents. It helps the researchers and other users to filter out the documents and hence making it easier to follow the SLR methodology. The extracted meta-data ensures that the information is stored and re-used for further research. It can also be used in search engines to find similar and relevant data for future work.
- Extracting specific sections of the PDF such as “Introduction”, “Abstract” and “Conclusion” can help the researchers in getting a deeper understanding of

the document. Most researchers read these sections to get the gist of the documents. These sections are believed to be the most important sections in a paper [Doumont, 2010].

- Summarizing the document is the main task of the proposal. A textual qualitative summary is created to help the researchers in understanding the document within a short period of time [Allahyari et al., 2017]. The extractive summarization helps in summarizing the paper as a whole, whereas the keyword based data extraction is done by using user-defined keyword search throughout the document to find particular topics that are interesting to the researchers. The keyword and its synonyms are searched in the primary studies to find relevant sentences. This ensures that all the aspects of the document are covered in the summary.

The implementation of lour proposed approach is carried out by using Python 3, a high level interpreted programming language [Sanner, 1999]. The project is made available as a web application using the Django web framework. Django is used for its features such as quick, transparent and high-quality code writing [Zublenko, 2016]. The project also uses Jinja as the templating engine to perform powerful automatic HTML escaping. The decision of using Python3 depend on the fact that it is associated with a huge collection of libraries for text mining and natural language processing, such as, NLTK, gensim, numpy, and scripy.

4.1 Meta-data Extraction

The meta-data of the document is the high-level information that describes the document. The meta-data of the documents help the researchers in extracting the title, authors, keywords, publishers, date of creation, creator, and producer. Franks [2006] states that the meta-data acts as a tool to classify and categorize the information that also helps in giving us an insight of the paper. This can help the researchers to further filter out some documents if it does not fulfill their criteria. Hence, we incorporate the process of extracting meta-data into our tool to help in effective categorization and to provide the users with the context of the study.

Python has a lot of different libraries to implement the meta-data extraction. The implementation is carried out by using the PyPDF2 python toolkit. PyPDF2 was used because of its simplicity and it satisfies our requirements of extracting only the meta-data. PyPDF2 can also be used to extract other data from the PDF file and manipulate them. We also explored other libraries such as pdfminer, pdfminer3k, and TIKA. They contain tons of other features along with the meta-data extraction. They are mainly used for PDF parsing and data analysis.

PyPDF2 retrieves the PDF file's meta-data information if it exists. PyPDF2 uses "PdfFileReader" to get the document information. We found that some PDF files use meta-data streams instead of dictionaries, and hence this function will not be able to access these meta-data streams. Every document has two types of properties, the raw property and the non-raw property. The non-raw property returns a "TextString" object and hence works well while displaying the results to the users. If

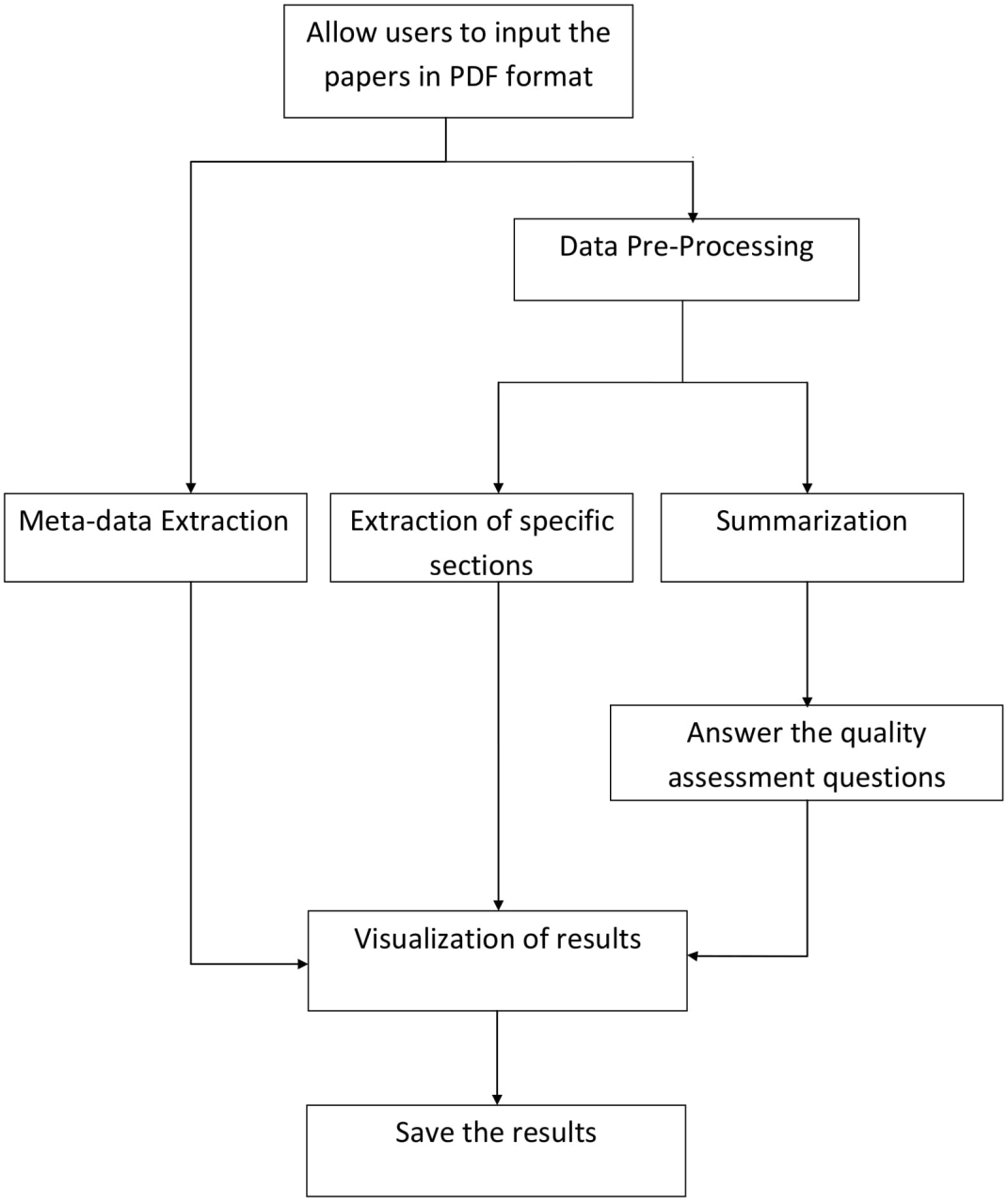


Figure 4.1: Proposed approach for summarization and data extraction.

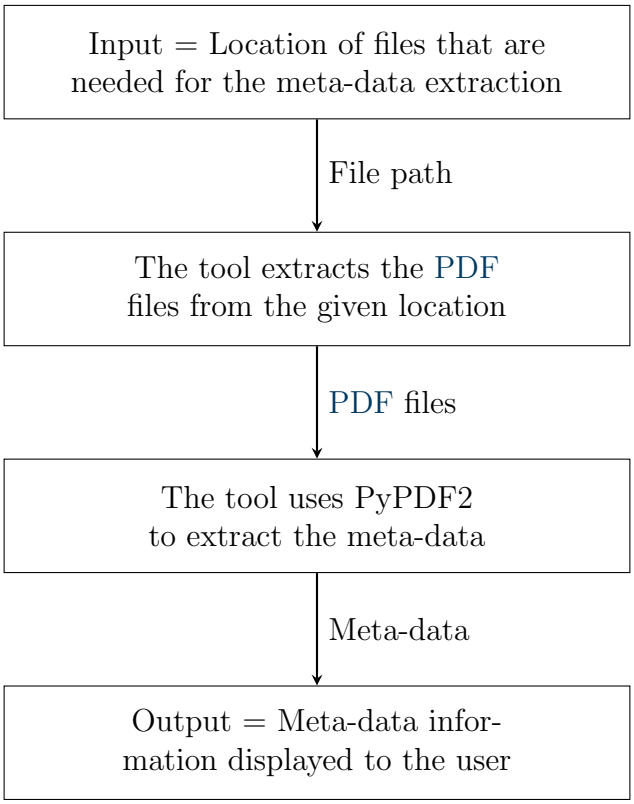


Figure 4.2: Meta-data extraction process.

PyPDF2 was unable to find any meta-data, a null string is returned. The extracted data is then stored in a file. Table 4.1 shows the data extracted from one of the input files. The properties such as "Title", "Author", "Subject", "Creator", "Producer", "Creation date", and "Keywords" can be extracted. Figure 4.2 describes the steps involved in the process.

Type of Meta-data	Meta-data
Title	SLR-TOOL A Tool for Performing Systematic Literature Reviews
Author	Ana M. Fernández-Sáez, Marcela Genero Bocco, Francisco P. Romero
Creation Date	D:20111007115732+02'00'
Creator	Toolkit http://www.activepdf.com
Keywords	Systematic Literature Review, SLR, Tool, Text mining
Producer	Acrobat Distiller 10.0.1 (Windows)

Table 4.1: Extracted meta-data.

4.2 Automatic Extraction of Specific Sections

This feature was implemented to help the readers who are interested in understanding some specific sections of the study. Doumont [2010] stated that the readers are primarily interested in the contents present in the introduction and conclusion sections of the study. It is also stated that the readers want to know the motivation and the outcome of the study and the abstract is considered to be in parallel to the introduction and conclusion. This indicates that these sections play a vital role in the initial filtering and understanding of the results. It is important to know why the work was carried out (motivation), what work has been done (introduction), and what are the findings of the study (conclusion).

Our tool helps in implementing this by extracting the sections "Abstract", "Introduction", and "Conclusion". The user selects one of these sections for every document according to their requirement. The extraction of these sections is carried out by using Regular Expression (regex) and the regex python library (re). The data cleaning is done to remove unwanted spaces and newline characters. The first step is to split the text by using the newline character. Listing 4.1 searches for one or more occurrence of the newline character.

Listing 4.1: Splitting the sentences

```
re.split(r'\n+', text)
```

Listing 4.2 searches for the occurrence of the selected sections in the subheadings.

Listing 4.2: Searching a section

```
(?<!\S)\d' +'.*'+ '\s*'+ section
```

Once the selected section is found, the corresponding paragraphs associated with it is extracted using Listing 4.3. The sentences are extracted until the next occurrence of another section is found.

Listing 4.3: Extracting the section

```
'(?<!\S)\d' +'.*'+ '\s*'+ '[A-Za-z]'
```

It makes use of negative look behind `(?<!\S)` which says that there are no white spaces before the `\d` character. When the algorithm finds the next section, the loop breaks and returns the specific section.
It was later observed that the sections can also begin with roman numbers. Hence, the algorithm was changed accordingly to search for roman numbers in the section labels as shown in Listing 4.4.

Listing 4.4: Extracting the section with roman numbers

```
'I|X|IV|V?I{0,3}' + '.*' + '\s*' + '[A-Za-z]+'
```

Figure 4.3 on page 40 depicts the step by step process used in extracting the sections.

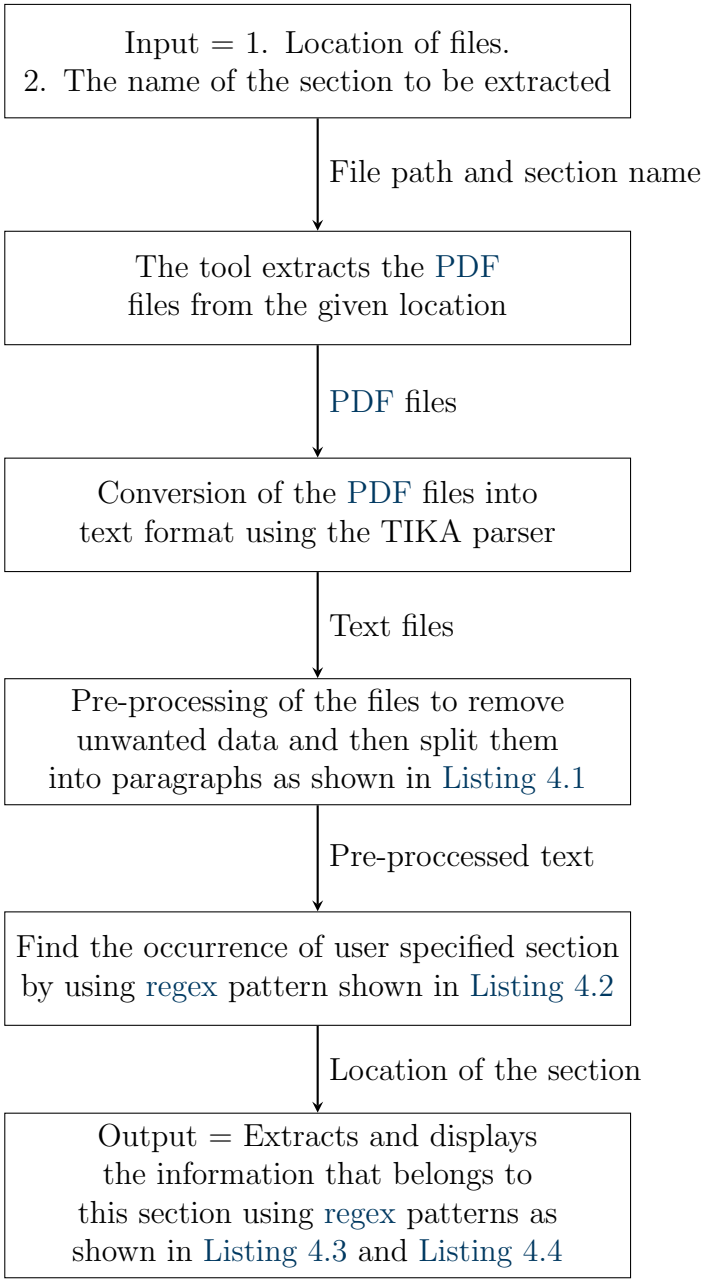


Figure 4.3: Process of extracting specific sections.

4.3 Semi-Automatic Summarization

Summarization is done on the entire document to present the most important concepts of the document. Summarization is independent of the structure or the syntax of the document. It helps the users to get a well structured document that contains the objectives and important features [Ouyang et al., 2009]. It integrates the core ideas of the document in a meaningful way ignoring the irrelevant information and by removing redundancies [Ouyang et al., 2009]. A good summary should be complete, accurate, and impartial. Due to the ever increasing amount of text data that are available, discovering relevant information is time-consuming. Hence, there is a great need to reduce this text and create summaries. Automatic summaries are also less biased than human written summaries [Torres-Moreno, 2014]. Therefore, we introduce the feature of text summarization in our tool.

The summarization is divided into extractive summarization and keyword based summarization. The extractive summarization algorithm automatically summarizes the document to include important concepts by checking the frequency of keywords. To support our idea of answering the QA questions, we use a keyword-based summarization algorithm to summarize the document. The process is semi-automatic because of the keywords that have to be manually added to answer the QA criteria.

Before proceeding with the implementation, the PDF studies are converted into text documents. Later, data cleaning is done to remove unwanted spaces and new line characters. The sections such as bibliography and acknowledgements are also removed from the text document. The summary is also free from unnecessary links that can hinder the quality of the summary as shown in Listing 4.5.

Figure 4.4 on page 42 shows us the initial steps of the process.

Listing 4.5: Removing links

```
re.findall('https?://(?:[-\w.]|(?:%[\da-fA-F]{2}))+', str(s))
```

4.3.1 Extractive Summarization

Extractive summarization is the process of extracting the sentences that best describe the document. It can be carried out by using various algorithms such as "TextRank", "LSA summarizer", "Luhn summarizer", "Textum", "LexRank summarizer", "gensim" etc. The algorithms that displayed the highest accuracy were used for the implementation. Figure 4.5 on page 43 shows the process involved in creating the extractive summary.

Appraoch 1: TextRank

Mihalcea and Tarau [2004] introduces TextRank as a graph-based ranking model written based on the PageRank algorithm. Mihalcea and Tarau [2004] mentioned that graph based algorithms and Google's PageRank algorithm have been running very successfully. The author argues that it works well because it decides the importance of a vertex by considering the global information from the entire graph. He concludes that the cohesive web model built by the tool approximates to the model created by humans in the process of summarization.

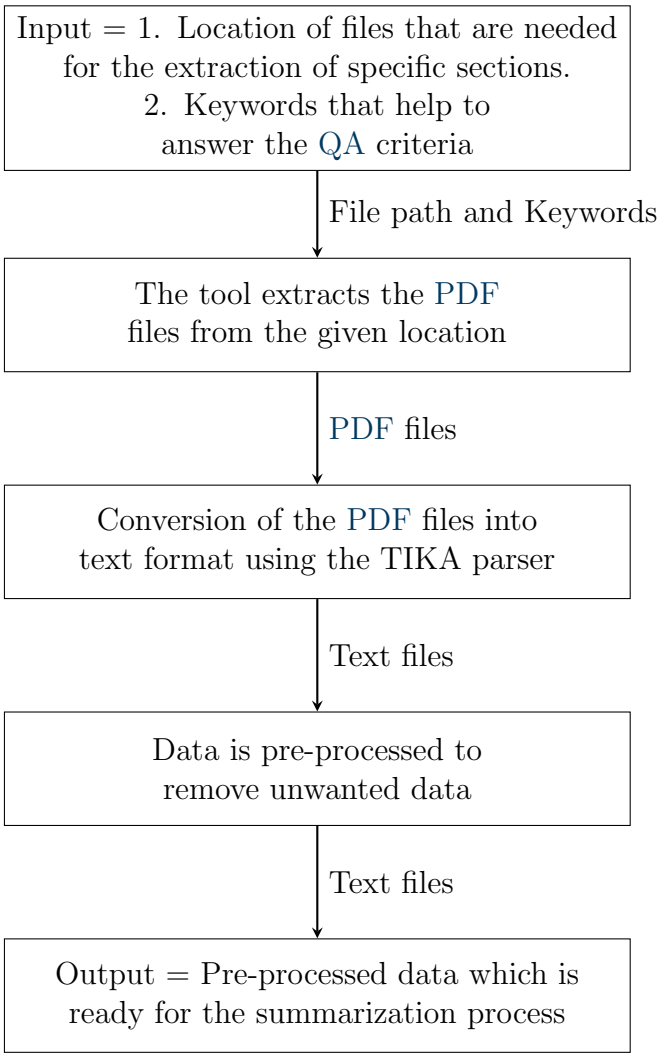


Figure 4.4: Data summarization.

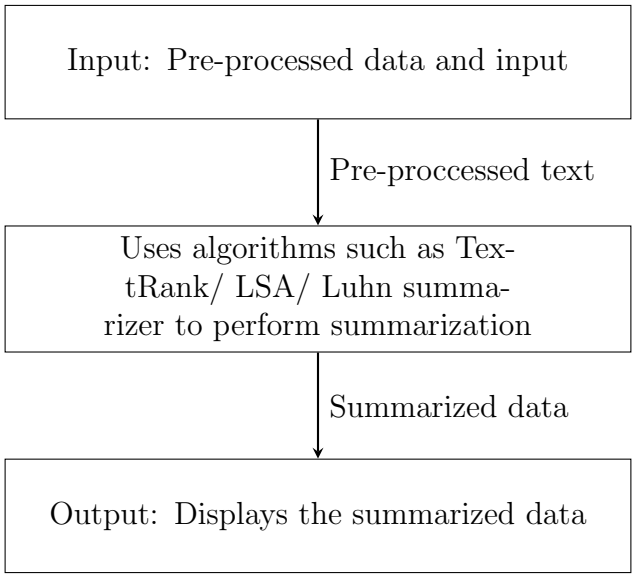


Figure 4.5: Process of extractive summarization.

TextRank helps our implementation by performing extractive summarization. TextRank uses different similarity measures to find the similarity between sentences and finds how one sentence could be connected to another as shown in Figure 2.2. It creates a vertex for each unit extracted from the input text. These vertices are connected to each other with edges. The weights assigned to these edges are the measure of the strength of the connections. The sentences are ranked based on the score given to each sentence. The top "n" sentences are selected for the summary. TextRank provides us the summary by preserving the original meaning of the document.

The document to be summarized is converted from PDF to text file using the TIKA parser to read all the contents of the file. This text is later used in the TextRank algorithm. The TextRank summarizer first splits the text into sentences at ".". These sentences are pre-processed to prepare it for the summarization Listing 4.6.

Listing 4.6: Clean text

```
sentences = clean_text(text, language)
```

These sentences are used to build a graph which denotes the similarity coefficient between the sentences as shown in Listing 4.7.

Listing 4.7: Build graph

```
graph = build_graph([sentence.token for sentence in sentences])
set_graph_edge_weights(graph)
```

All nodes with weights equal to 0 are considered unreachable and hence, removed from the graph as shown in Listing 4.8.

Listing 4.8: Remove unreachable nodes

```
remove_unreachable_nodes(graph)
```

If the graph is empty, the algorithm returns an empty string and ends the process as shown in Listing 4.9.

Listing 4.9: Check for empty graph

```
if len(graph.nodes()) == 0:
        return ' '
```

The nodes of the graph are ranked using PageRank to get a dictionary of sentences with their respective scores as shown in Listing 4.10.

Listing 4.10: Rank the sentences

```
scores = pagerank(graph)
add_scores_to_sentences(sentences, scores)
```

Based on the rank given to the sentences and the selection criteria, the most important sentences are extracted. Words can be sent into this method to tune the extraction process. Similarly, the "ratio" is used to reduce the number of sentences as shown in Listing 4.11.

Listing 4.11: Get the most important sentences

```
extracted_sentences = most_important_sentences(sentences,
ratio, words)
```

The extracted sentences are sorted in the order of their appearance in the original text as shown in Listing 4.12.

Listing 4.12: Sort the sentences

```
extracted_sentences.sort(key=lambda s: s.index)
```

There are chances that some unwanted sentences such as references are added to the summarization. To avoid such sentences, only those sentences with words more than 50 are selected in our summarization. Finally, the results are formatted by joining the sentences and returned to the user as shown in Listing 4.13.

Listing 4.13: Format the sentences

```
format_results(extracted_sentences)
```

Appraoch 2: LSA

Steinberger [2004] makes use of LSA to identify the semantically important sentences. It is a generic text summarization algorithm written to find the similarity between the original document and its summary. It uses TF and SVD to implement the summarization. Steinberger [2007] mentioned that the results of the summarizer were comparable to other well known summarizers. Steinberger [2004] also stated that in some set of documents, LSA summarizer outperforms other summarizing methods.

First step of the LSA implementation is to create a dictionary mapping of key (unique words) - value (row index) pairs as shown in Listing 4.14.

Listing 4.14: Dictionary mapping

```
for w in words if w not in stop_words
dict((w, i) for i, w in enumerate(unique_words))
```

Next step is to create a M*N matrix of |unique words|*|sentences| where the cells contain number of occurrences of words (rows) in sentences (cols). For every word that is not a stop word, the count is increased by 1 as shown in Listing 4.15.

Listing 4.15: Matrix creation

```
for col, sentence in enumerate(sentences):
  for word in map(stem_word, sentence.words):
     if word in dictionary:
        row = dictionary[word]
        matrix[row, col] += 1
```

Once the matrix is created, we compute the term-frequency for each sentence in the given matrix. Term frequency is normalized using a smoothing factor to mitigate the anomaly that is observed in longer documents. Higher term frequencies are observed in longer documents because it tends to repeat the same words over and over again. Hence a maximum word frequency is set for every sentence by numpy as shown in Listing 4.16.

Listing 4.16: Term-Frequency

```
for row in range(rows):
  for col in range(cols):
    frequency = matrix[row, col]/max_word_frequency
    matrix[row, col] = smooth + (1.0 - smooth)*frequency
```

SVD is applied to this matrix to create independent base vectors. SVD helps in finding relationships between these vectors to create semantic clusters of the terms and sentences. Recurring patterns will be combined into a single cluster. Each singular vector represents a particular concept. The magnitude of these vectors describes the importance of each of the sentences. Hence, the most important features are selected in the summarization process.

Approach 3: Luhn Summarizer

Luhn summarizer is an extractive summarizer that works on the idea that authors tend to repeat their words in the process of writing [Luhn, 1958]. This is taken as a means of emphasis on the importance of the words. If certain words are more often seen together with each other, it is considered to be more significant in the document. The algorithm has two phases. In the first phase, stop words are identified and the significant words are selected by using frequency analysis. A subset of common words that are not in the list of stop words are selected as significant words. The frequency of every significant word is calculated. Words having a frequency greater than a threshold will be selected as the most frequent words. It is shown in Listing 4.17

Listing 4.17: Significant words

```
significant_words = for w in words if w not in stop_words
words = model.most_frequent_terms(significant_words)
selected_words =  t for t in words if frequency(t) > n
```

The second step is to score each sentence. The frequency of the sentence is calculated based on a significance factor. It is the square of significant words divided by the total number of words in a sentence. If a sentence has 100 words and all the words are significant, then the significance factor is 100. It is shown in Listing 4.18.

Listing 4.18: Significance factor

```
if w in top_words:
    important_words += 1
 Significance factor= important_words^2 / total_number_of_words
```

In the final step, the sentences with the highest significance factor is used for the summarization.

4.3.2 Keyword based Summarization

To make our summary adhere to the QA criteria, we include keyword based summarization as an additional step. The summarization is done by extracting those sentences that have the keywords specified by the user. The algorithm takes the QA criteria and the study as input. The following QA criteria were used in the implementation.

- Q1. Extracting the the aim/goals.
- Q2. Extracting the limitations.
- Q3. Extracting the future work.
- Q4. Extracting the evaluation study.

Keyword extraction is done by removing stop words from the QA criteria to get the search keywords. We use "WordNet", a lexical database to find the synonyms for our keywords. WordNet groups together nouns, verbs, adjectives, and adverbs into sets called "synsets". WordNet consists of a hierarchical structure of words with super-subordinate relation. It makes use of semantic similarity to extract all the available sentences containing the keywords. Listing 4.19 shows how the synonyms are extracted for every keyword

Listing 4.19: Finding synonyms

```
for synset in wordnet.synsets(key):
    for name in synset.lemma_names():
        keywords.append(name)
```

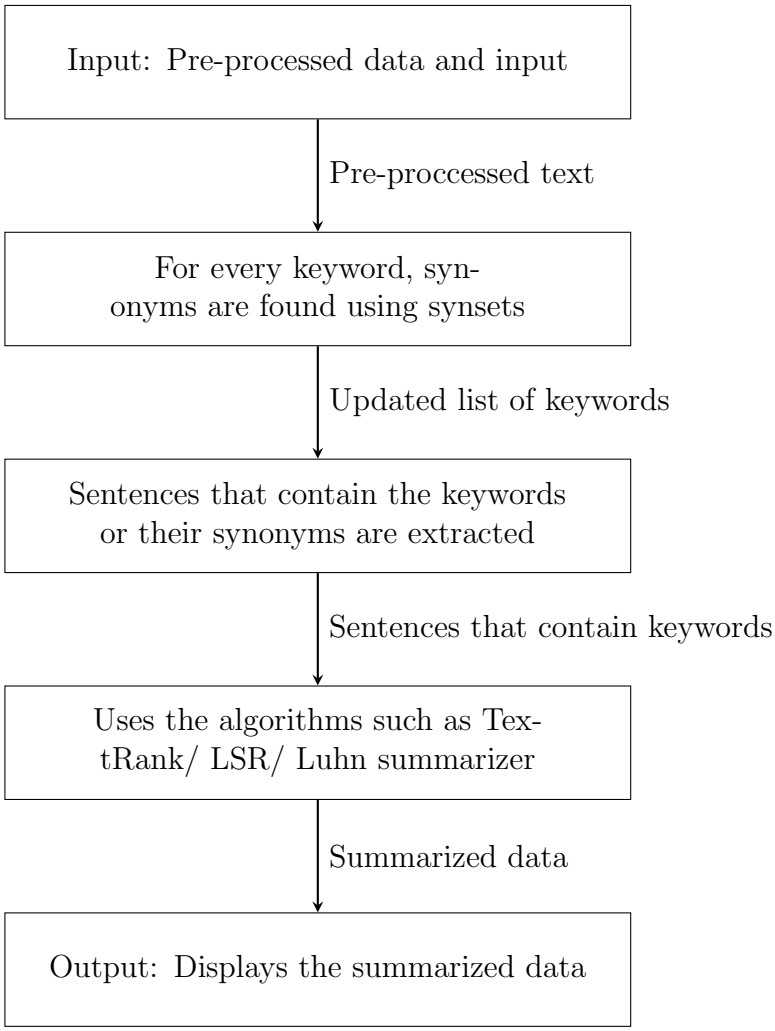


Figure 4.6: Process of keyword based summarization.

For every QA criteria, the keywords and their synonyms are searched in the original text. The corresponding sentences that match the keywords are extracted from the study. The sentences are grouped together for every QA criteria. This gives us a long list of sentences that match the keywords. To summarize the findings, the "TextRank" algorithm is applied again for every QA criteria to get the desired number of sentences. Figure 4.6 on page 47 depicts the steps involved in the keyword based summarization process.

4.4 Summary

In this chapter, we explained the implementation details of our approach. We started by explaining how the meta-data extraction was carried out. We then explained the details of extracting specified sections of the study. Finally, we explained how the data summarization was implemented. The two methods of data summarization are combined together for every study to get the desired results. The QA questions are successfully answered in the summarized text. The summary is well structured,

making it easy to read and understand the context of the study in a very short amount of time.

5. Evaluation

In this section, we explain the experimental setup and discuss the results of our approach. Evaluation is important to verify the significance of the approach for the researchers conducting the SLR. We discuss the results of the three different parts of our approach. We begin by discussing the results of meta-data extraction. Later, we evaluate the approach used to extract the specific sections by using precision and recall as the evaluation method. Finally, we discuss the results of summarization by using intrinsic and extrinsic methods.

5.1 Meta-data Extraction

The evaluation is performed by checking the different meta-data extracted from a number of studies. During the evaluation, it was found that some studies have incomplete meta-data if the author has not added the necessary information. Since our implementation makes use of python libraries to extract the meta-data, it is not possible to extract the meta-data if the author has not added the information. Table 4.1 shows an example of the extracted meta-data when the author has added complete. The evaluation is conducted on 25 different studies which are listed in Table 5.2. These studies were randomly selected from the digital libraries to check the performance of the tool in the real world scenario. The result is the percentage of correctly extracted meta-data. Table 5.1 shows us the results of evaluation. Figure 5.1 shows us the snapshot of the results obtained from the tool.

Advantages:

- The tool helps the readers in classifying the documents based on their meta-data.
- It helps in filtering out results without going through every document.

Disadvantages:

- If the PDF document is not associated with any meta-data, the tool returns a null value.

Type of meta-data	Percentage
Title	68%
Author	63%
Creation date	100%

Table 5.1: Evaluation of meta-data extraction.

Results of Evaluation

The results of evaluation show us that out of 25 studies, the title is extracted from 17 studies which gives us a success rate of 68%. Similarly, the name of the author was extracted from 16 studies (success rate of 63%) and the creation date was extracted from all the 25 studies (success rate of 100%). The extraction of some meta-data failed since the author did not add the meta-data information while creating the document. To overcome this problem, other approaches such as text mining using regex can be used to extract the meta-data from the text.

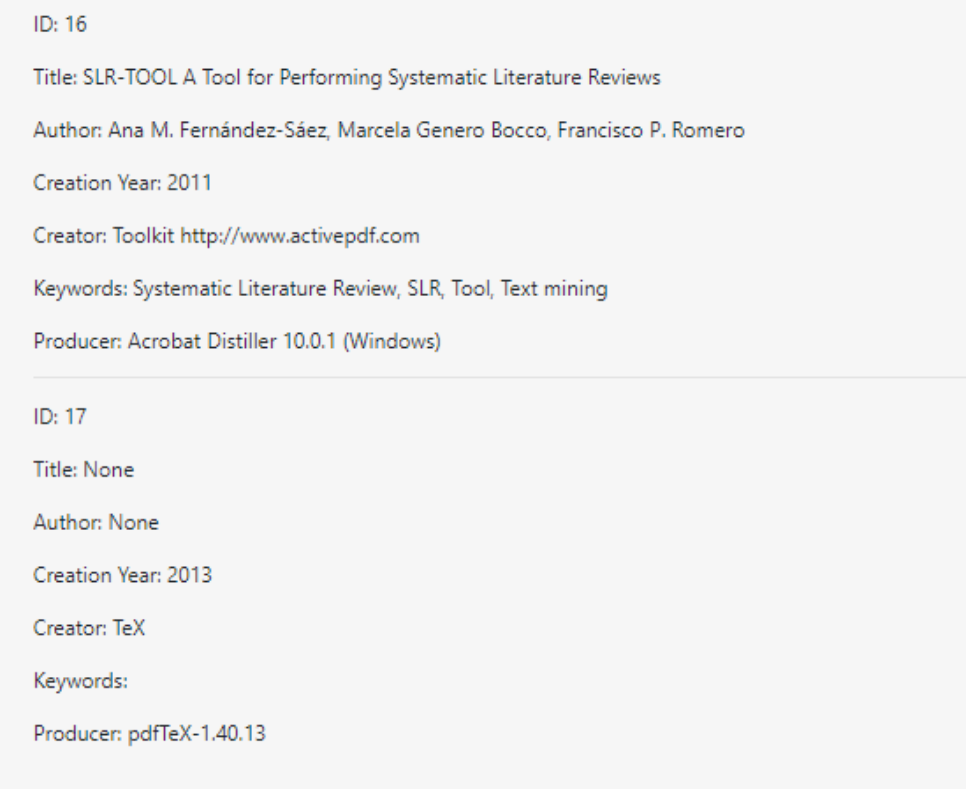


Figure 5.1: Snapshot of meta-data extraction.

5.2 Extraction of Specific Sections

To evaluate the extraction of specific sections, the algorithm was tested on 25 different PDF files. Table 5.2 gives the list of studies used in the implementation. These studies belong to the software engineering domain and were randomly selected from

Title	Year
Text mining- state of art	2000
Procedures for Performing Systematic Reviews by Barbara Kitchenham	2004
SUMMAC: a text summarization evaluation	2004
Performing Systematic Literature Reviews in Software	2006
Applying Systematic Reviews to Diverse Study Types: An Experience Report	2007
Systematic Mapping Studies in Software Engineering	2008
Motivation in Software Engineering- A systematic literature review	2008
Strength of Evidence in Systematic Reviews in Software Engineering	2008
Problem-Based Learning for Engineering Students: An Evidence-Based Comparative Study	2009
Ontology- Based Knowledge Management System and Application	2011
BNOSA: A Bayesian network and ontology based semantic annotation framework	2011
Light-weight tool support for semantic annotation of textual documents	2011
SLR-TOOL A Tool for Performing Systematic Literature Reviews	2011
SLuRp – A Tool to Help Large Complex Systematic Literature Reviews Deliver Valid and Rigorous Results	2012
Using Information Visualization and Text Mining	2012
Automatically Locating Results to Support Systematic Reviews in Software Engineering	2013
Graph-based semantic annotation for enriching educational content with linked data	2013
Tools to Support Systematic Literature Reviews in Software Engineering- A Mapping Study libre	2013
Text Mining and Semantics: a Systematic Mapping Study	2013
Tools to Support Systematic Reviews in Software	2014
Ontology-based annotation and retrieval of services in the cloud	2014
Quality Assessment of Systematic Reviews in SE a tertiary study	2015
Integrating Library Instruction into the Course Management System for a First-Year Engineering Class	2015
Semantic Web in data mining and knowledge discovery: A comprehensive survey	2016
PDFdigest: an Adaptable Layout-Aware PDF-to-XML Textual Content Extractor for Scientific Articles	2018

Table 5.2: List of studies used for meta-data extraction and extraction of specific sections.

the digital libraries to check the performance of the tool. Figure 5.2 shows us a snapshot of the extracted section. Every extracted section is given a score from 0-1. Evaluation measures such as precision and recall were used to check the quality of extraction. The data is classified based on the following metrics.

- True Positives (TP) – The extracted paragraph matches the ground truth.
- False Negatives (FN) – The paragraph is not extracted.
- False Positives (FP) – Paragraph is wrongly labeled.
- True Negatives (TN) – Nothing is extracted because there was no data present.

Precision Equation 5.1 and Recall Equation 5.2 are calculated by the following formulas respectively.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{5.1}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{5.2}$$

If the extraction is complete it is awarded a score of 1. If the extraction failed completely, it is awarded a score of 0. If there is a partial extraction, it is awarded a score in between 0-1.

The following tables show the different confusion matrices and Table 5.6 shows the final values obtained during evaluation.

Confusion Matrices for the Extraction of Specific Sections

TP= 20	FP=0
FN=5	TN=0

Table 5.3: Confusion Matrix for Extracting Abstract.

TP= 15	FP=2
FN=8	TN=0

Table 5.4: Confusion Matrix for Extracting Introduction.

TP= 15	FP=3
FN=5	TN=2

Table 5.5: Confusion Matrix for extracting Conclusion.

Justification of Results

Table 5.3 shows that among 25 studies, the tool successfully extracted the exact paragraph out of 20 studies (TP) and failed to extract the desired result for 5 (FN) other studies. The extractions failed since the abstract of the study was not recognized as a new paragraph in some of the studies. Some other studies had formats that were not recognized by the tool such as a tabular format. It also had different text format for the heading (Example : A B S T R A C T).

Table 5.4 shows that among 25 studies, the tool successfully extracted the exact paragraph for 15 studies (TP) and failed to extract the desired result for 8 other studies (FN). The tool also extracted the wrong data for 2 studies (FP). The extraction failed since the introduction sections had a different sub-heading name (FN). In a few studies, the tool tried to extract the introduction from the table of contents and hence failed in extracting the information (FN). We also encountered partial results when the text from the footer came in between the paragraph. Hence it was considered as (FP). When there was no introduction section in the study, the tool found the keyword elsewhere in the study and extracted wrong information (FP).

Table 5.5 shows that among 25 studies, the tool successfully extracted the exact paragraph for 15 studies (TP) and failed to extract the desired result for 5 other studies (FN). The tool also extracted the wrong data for 3 studies (FP). There were no data present in 2 papers and hence it was not extracted (TN). The tool failed to extract the information when the conclusion section had a different name (FN). The tool also tried to extract the conclusion from the table of contents and hence failed to extract the text (FN). Few studies had a brief conclusion section within the abstract and hence our tool extracted the wrong data (FP). When there was no conclusion section in the study, the tool found the keyword elsewhere in the study and extracted wrong information again (FP).

Results of evaluation

Section	Precision	Recall
Abstract	1.00	0.80
Introduction	0.87	0.65
Conclusion	0.83	0.75

Table 5.6: Evaluating the extraction of sections.

The precision and recall is calculated based on Equation 5.1 and Equation 5.2. Table 5.6 depicts the values of precision and recall using the confusion matrix for every

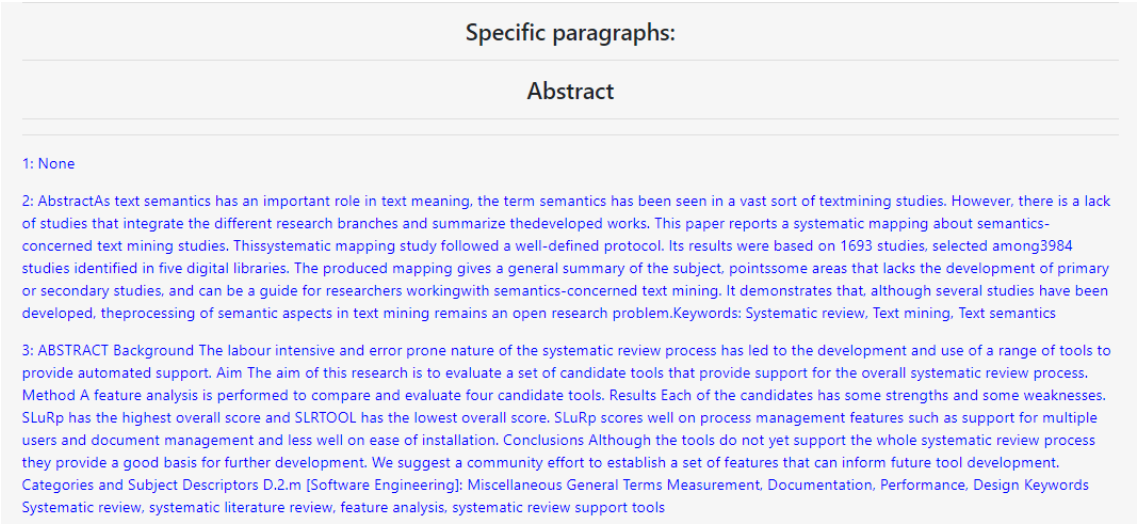


Figure 5.2: Snapshot of extraction of specific sections.

section. Higher precision values relate to a lower rate of wrongly extracted sentences (FP). Whereas, higher recall values relate to a lower rate of falsely extracted null values (FN). High scores for both show that the summarizer is returning accurate results with a majority of positive results.

Advantages:

- It helps the researchers in quickly reading the important sections of the paper.
- It provides readers with the motivation and outcome of the study.

Disadvantages:

- If the section has a different name, the tool fails to extract it.
- It could also fail when the paper has an unknown format.

5.3 Summarization

The summarization is evaluated by considering two different SLR studies and extracting all the primary studies used in the study. These primary studies are used as the input to our tool. The tool summarizes all these studies to create a multi-document summary. Figure 5.3 shows us a snapshot of the summarized data. The result of this summarization process is compared with the results of the SLR study. The evaluation is done by using intrinsic and extrinsic methods. We selected two well written SLR studies where the results of the studies were clearly stated. Table 5.7 and Table 5.8 show the two different SLR study examples. The SLR study is tagged as the reference paper and their primary studies as the input.

We use ROUGE score for the intrinsic evaluation which is one of the widely used measures of evaluating the results of automatic summarization [Graham, 2015]. It works by comparing the generated summary against the reference summary. ROUGE-N

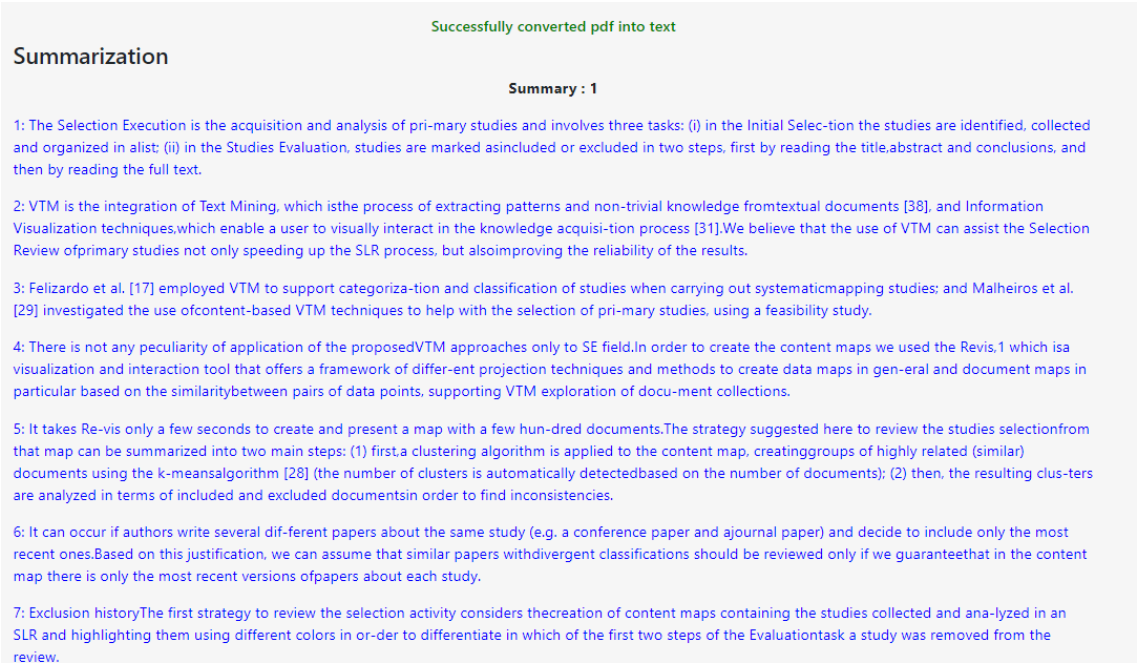


Figure 5.3: Snapshot of the summarized data.

refers to the overlapping of N-grams between the generated summary and the reference summary [Lin 2004]. ROUGE-1 refers to the overlapping of unigrams between the two summaries. Whereas, ROUGE-2 refers to the overlapping of bigrams between the two summaries. The formula used for the calculation of the ROUGE score is explained using the simplified equation in Equation 5.3.

$$\frac{\sum_r \sum_s match(gram_{s,c})}{\sum_r \sum_s count(gram_s)} \tag{5.3}$$

Where, $\sum_r$ refers to every reference summary and $\sum_s$ refers to every sentence in the generated summary. It loops through every n-gram in a single reference summary and counts the number of times a matching n-gram is found in the candidate summary. The denominator counts the total number of n-grams across all the reference summaries.

Extrinsic evaluation is carried out using the question game. In this method, a set of questions are prepared that are expected to be answered by the summary. The quality of the extracted summary is decided based on whether these questions are answered. We use some of the QA questions stated during the SLR to perform the extrinsic evaluation.

5.3.1 Intrinsic Evaluation

In this section, we use the ROUGE scores to compare the summaries created by using the various summarizers. We use ROUGE-1 to compare the summaries. Since we are comparing a human-written summary with the machine-generated automatic summary, we believe that it is more appropriate to compare unigrams rather than

Title	Year	Type
Effective Engineering Information Literacy Instruction: A Systematic Literature Review	2018	Reference
Effect of Guided Research Experience on Product Design Performance: A Pilot Study	2003	Input
Effect of library instruction on undergraduate electrical engineering design projects	2007	Input
Problem-Based Learning for Engineering Students: An Evidence-Based Comparative Study	2008	Input
Using Videos to Teach the Ethical Use of Engineering Information	2008	Input
Continuing library instruction via on-line tutorials	2009	Input
Enhancing Engineering Students' Knowledge of Information Literacy and Ethics through an Interactive Online Learning Module	2010	Input
Challenges for information literacy education at a university of technology	2010	Input
One or Many? Assessing different delivery timing for information resources relevant to assignments during the semester. A work-in-progress	2012	Input
Work in Progress: Collaboration for Quality: a Librarian-Faculty Partnership to Assess Students' Information Literacy in Freshman Engineering	2013	Input
If the Engineering Literature Fits, Use It! Student Application of Grey Literature and Engineering Standards	2015	Input
Integrating Library Instruction into the Course Management System for a First-Year Engineering Class: An Evidence-Based Study Measuring the Effectiveness of Blended Learning on Students' Information Literacy Levels	2015	Input
Using Direct Information Literacy Assessment to Improve Mechanical Engineering Student Learning - A Report on Rubric Analysis of Student Research Assignments	2015	Input

Table 5.7: SET 1 of studies used for summarization.

Title	Year	Type
Tools to Support Systematic Literature Reviews in Software Engineering: A Mapping Study	2013	Reference
Automated Information Extraction from Empirical Software Engineering Literature: Is that possible?	2007	Input
Using Context Distance Measurement to Analyze Results across Studies	2007	Input
An Approach Based on Visual Text Mining to Support Categorization and Classification in the Systematic Mapping	2010	Input
Analysing the Use of Graphs to Represent the Results of Systematic Reviews in Software Engineering	2011	Input
Using Visual Text Mining to Support the Study Selection Activity in Systematic Literature Reviews	2011	Input
Linked Data approach for selection process automation in Systematic Reviews	2011	Input
SLuRp – A Tool to Help Large Complex Systematic Literature Reviews Deliver Valid and Rigorous Results	2012	Input
Using GQM and TAM to evaluate StArt – a tool that supports Systematic Review	2012	Input
A visual analysis approach to validate the selection review of primary studies in systematic reviews	2012	Input
Towards Evidence-Based Ontology for Supporting Systematic Literature Review	2012	Input
A Federated Search Approach to Facilitate Systematic Literature Review in Software Engineering	2012	Input
Slrtool: A Tool to Support Collaborative Systematic Literature Reviews	2014	Input

Table 5.8: SET 2 of studies used for the summarization.

bigrams. Table 5.9 and Table 5.10 shows the results of our intrinsic evaluation on two different SLR studies.

The tool uses the ROUGE score calculator provided by the sumy evaluation framework for text summaries.

Algorithm	ROUGE score
TextRank summarizer	59%
LSA summarizer	64%
Luhn summarizer	69%

Table 5.9: ROUGE score - SET 1.

Algorithm	ROUGE score
TextRank summarizer	65%
LSA summarizer	63%
Luhn summarizer	64%

Table 5.10: ROUGE score - SET 2.

If the ROUGE score is 69%, it means that 69% of the generated summary matches the reference summary. The ROUGE score reduces when we use ROUGE-2 since there are lesser possibilities of having bigrams while comparing human-written summary with the machine-generated summary.

5.3.2 Extrinsic Evaluation

In this section, we use the QA criteria as the questions in the question game. Every summary is manually checked to see if the questions are answered in the summary. These answers are then evaluated using precision and recall as the evaluation metrics.

- True Positive - The data was accurately extracted.
- False Negative - The data is present but not extracted.
- False Positive - The wrong data is extracted.
- True Negative - The data was not present and hence not extracted.

The following QA criteria were used:

- Q1. Extracting the aim/goals of the study.
- Q2. Extracting the limitations of the study.
- Q3. Extracting the future work.
- Q4. Extracting the evaluation study.

Confusion Matrices for SET 1

TP = 8	FP = 3
FN = 1	TN = 0

Table 5.11: Confusion matrix for evaluating first set with Q1.

TP = 4	FP = 4
FN = 1	TN = 3

Table 5.12: Confusion matrix for evaluating first set with Q2.

TP = 9	FP = 1
FN = 2	TN = 0

Table 5.13: Confusion matrix for evaluating first set with Q3.

TP = 5	FP = 3
FN = 4	TN = 0

Table 5.14: Confusion matrix for evaluating first set with Q4.

Justification of the Results for SET 1

The evaluation is carried out on 12 SLR studies from the first set of studies. Table 5.11 shows that the tool extracted Q1 (Aim of the study) from 8 studies (TP) and failed to extract it from 1 study (FN). The tool also extracted the wrong information for Q1 from 3 studies (FP). Table 5.12 shows that the tool extracted Q2 (Limitations of the study) from 4 studies (TP) and failed to extract it from one study. The tool also extracted the wrong information for Q2 from 4 studies (FP). Table 5.13 shows that the tool extracted Q3 (Future work of the study) from 9 studies (TP) and failed to extract it for 2 other studies (FN). The tool also extracted the wrong information for Q3 from 1 study (FP). Table 5.14 shows that the tool successfully extracted Q4 (Evaluation study) from 5 studies (TP) and extracted the wrong data from 3 studies (FP).

FNs occurred since one of the PDF file was in a non-searchable format. In some other studies the tool failed to extract the data when they were present under unknown sections (FN). Similarly, some sentences were extracted from the study when the algorithm found the specific keywords but the keywords actually belonged to a different context (FP).

Results of Extrinsic Evaluation for SET 1

QA criteria	Precision	Recall
Q1	0.80	0.88
Q2	0.50	0.80
Q3	0.90	0.82
Q4	0.63	0.55

Table 5.15: Precision and recall - SET 1.

Table 5.15 shows the precision and recall values for the stated QA for the first set of the studies. It was calculated using Equation 5.1, Equation 5.2, and the above confusion matrices. It can be said that SET-1 has an average precision of 0.71 and an average recall of 0.76. High precision in Q1 and Q3 relates to a low rate of wrongly extracted sentences (FP). Whereas, high recall in Q1, Q2, and Q3 relates to a low rate of falsely extracted null values (FN). High scores for both in Q1, Q3 show that the summarizer is returning accurate results with a majority of positive results. Similarly, lower precision value in Q2 relates to a higher rate of wrongly extracted sentences (FP) and lower recall in Q4 relates to a higher rate of falsely extracted null values (FN).

Confusion Matrices for SET 2

TP = 10	FP = 2
FN = 0	TN = 0

Table 5.16: Confusion matrix for evaluating second set with Q1.

TP = 6	FP = 5
FN = 1	TN = 0

Table 5.17: Confusion matrix for evaluating second set with Q2.

TP = 6	FP = 3
FN = 3	TN = 0

Table 5.18: Confusion matrix for evaluating second set with Q3.

TP = 5	FP = 6
FN = 1	TN = 0

Table 5.19: Confusion matrix for evaluating second set with Q4.

Justification of the Results for SET 2

The evaluation is carried out on 12 SLR studies from the second set of studies. Table 5.16 shows that the tool extracted Q1 (Aim of the study) from 10 studies (TP) and extracted the wrong information for Q1 from 2 studies (FP). Table 5.17 shows that the tool extracted Q2 (Limitations of the study) from 6 studies (TP). It also failed to extract the limitations from 1 study (FN). The tool also extracted the wrong information for Q2 from 2 studies (FP). Table 5.18 shows that the tool extracted Q3 (Future work of the study) from 6 studies (TP). It failed to extract the future work from 3 studies (FN). It extracted the wrong information for Q3 from 3 studies (FP). Table 5.19 shows that the tool successfully extracted Q4 (Evaluation study) from 5 studies (TP) and extracted the wrong data from 6 studies (FP). It failed to extract the data for Q4 from 1 study by providing null results (FN).

FPs occurred when the tool extracted sentences that had the specified keywords but they actually belonged to a different context. Similarly, FNs occurred when the information was present in the study with different keywords that were not known to the tool. Hence, the tool failed to extract these results.

Results of Extrinsic Evaluation for SET 2

QA criteria	Precision	Recall
Q1	0.83	1.00
Q2	0.54	0.86
Q3	0.66	0.66
Q4	0.45	0.83

Table 5.20: Precision and recall - SET 2.

Table 5.20 shows the precision and recall values for the stated QA for the second set of the studies. It was calculated using Equation 5.1, Equation 5.2, and the above confusion matrices. It can be observed that SET-2 has an average precision of 0.62 and an average recall of 0.84. High precision in Q1 relates to a low rate of wrongly extracted sentences (FP), and high recall in Q1, Q2 and Q4 relates to a low rate of falsely extracted null values (FN). High scores for both in Q1 show that the summarizer is returning accurate results with a majority of positive results. Similarly, lower precision value in Q4 relates to a higher rate of wrongly extracted sentences (FP).

Advantages:

- The summarization process helps the readers in getting a gist of the long documents.
- It helps us in answering the QA questions.

Disadvantages:

- Since the QA criteria are generic, it may exclude some features that are specific to the study.
- The algorithm might extract the wrong data because the context of the text is not considered during the keyword based extraction.

5.4 Threats to Validity

In this section, we discuss the limitations of the proposed methodology. During the extraction of meta-data, it was observed that there are few files that do not have any meta-data associated with them. During the creation of these files, if the authors could not added the correct meta-data, this approach fails to get the necessary output.

Similarly, during the extraction of specific sections of the paper, we realized that various papers have different formats. For example, we came across papers that used roman numbers for numbering the sections instead of numbers and hence, we changed the regex to include the same. The structure of the PDF documents vary immensely and hence makes it difficult to define a certain set of regex for extraction. We have also come across spelling mistakes and different naming conventions while extracting the specific sections. Such unexpected formats could not be handled by the tool and hence, this method might fail to extract exact results if the study is not well formatted.

During the summarization process, it was found that different algorithms work well with different studies. Therefore, we have mentioned three different algorithms which summarize the best. Additionally, in the keyword based summarization, there are some limitations with the choice of the QA criteria. Certain aspects might not be covered since the QA criteria that are stated are more generic. Hence, some topics specific to the study could be omitted if the extractive summarization did not include them in the results. The tool might also extract the wrong data because our tool does not consider the context of the keywords in the study during the keyword based extraction.

5.5 Summary

In this chapter, we performed the evaluation of our approach using various criteria. We started by stating the procedure to evaluate the tool. We analyzed the results by using the evaluation methods to draw meaningful conclusions. We also stated the threats to validity and hence concluded the chapter. During the evaluation, we discussed the difficulties faced during the extraction of meta-data, extraction of specific sections as well as the summarization process. The advantages and shortcomings of the different methods were also mentioned. The extraction of meta-data provided us with results above 63%. The result also showed 100% accuracy in extracting meta-data of one of the papers. The extraction of specific sections provided us with the highest precision of 1.00 and the lowest precision of 0.83. It has also shown the highest recall of 0.80 and the lowest recall of 0.65. They depict how well the tool has performed in extracting the data. The process of summarization has also been evaluated using extrinsic and intrinsic methods with two different data-sets. The intrinsic evaluation provided us with ROUGE scores all above 63%. The extrinsic evaluation provided us with precision above 0.50 (except for Q4) and recall above 0.66 (except for Q4). We have also come across precision as high as 0.90 and recall as high as 1.0. We have seen that different algorithms work well with different data-sets. For example, the Luhn summarizer worked well with SET 1 but during the evaluation of SET 2, TextRank showed better results. The results of the evaluation

depict that the tool can be used as a prototype to automate the data extraction and reporting of the SLR.

6. Conclusion

A. Kitchenham [2007] mentions that a research is valid when it is performed thoroughly and completely in a fair manner. An "SLR synthesizes existing work in a manner that is fair and seen to be fair" - A. Kitchenham [2007]. SLR methodology protects the research against different kinds of bias and helps in finding real effects by combing the data from all the studies.

A. Kitchenham [2007] states that one of the most common reason to undertake an SLR is to summarize the existing evidence available in the particular area of research. The researchers who perform the SLR has to differentiate between the information which supports their study and the information which does not support their study. In this thesis, we began by analyzing the existing approaches to extract and summarize the data. Based on the findings, we conclude that there is some research made in this regard, specially on the reporting phase but it lacks the functionality of automatic multi-document summarization of the results. Therefore we proposed an approach by defining the RQs that focus on efficiently summarizing the findings.

Our approach aims at reducing the time and effort taken by the researchers in extracting and summarizing the findings of their research. The method we proposed is based on text mining techniques, such as, NLP, keyword extraction and sentence extraction. Our approach focuses on data summarization of multiple documents as well as help in meta-data extraction for better filtering and extraction of specific sections of the study to provide the users with additional information regarding the study. Summarization is implemented as an extractive summary which includes extraction of certain user specified information by using keyword based summarization to answer the QA criteria. The sentences were scored based on different criteria and the sentences with the highest similarity were chosen to compose the summary.

To evaluate the tasks, we used precision and recall to check the completeness in extracting the user defined sections. Similarly in evaluating the summary, intrinsic and extrinsic methods were used. During the intrinsic evaluation, two well written SLR studies were taken into consideration. Their primary were applied to the tool as inputs to perform summarization. The summary provided by the tool was compared

to the results of the original SLR study. The comparison was done using ROUGE scores. Similarly, the extrinsic evaluation was carried out by using the question-answer game which uses the QA criteria as the questions to check if the tool answered all those questions.

Findings of the Study:

- During the SLR we found out that there is not enough research on automating the data extraction and summarization phases of the SLR.
- Due to the increase in the number of studies published in the field of software engineering in the past years, there is an increasing need to automate the data extraction and summarization phase of the SLR.
- We found out that the various aspects of the study, such as, the meta-data, specific sections, the summary of the study can be extracted using regex and different text mining techniques. Our tool successfully extracted these aspects from the studies and also show satisfactory results during evaluation.
- The evaluation was carried out by using evaluation measures, such as, precision, recall, and ROUGE scores.
- Hence, we can reduce the time and effort taken by the researchers by using our tool that semi-automates the process of data extraction and summarization.

However, to improve the validation of the tool, we recommend some further improvements in the next section. We conclude this thesis work by stating the shortcomings of our approach as well as stating a few methodologies that could be implemented in the future to overcome the shortcomings of this approach.

7. Future Work

In this section, we suggest few steps that can be taken to improve our tool. During the evaluation, we came across several aspects that can be improved in the future. The most important improvement needed is to increase the accuracy of the results.

Having empty meta-data in the papers have reduced the accuracy of extraction. We suggest using methods that use regex to extract the meta-data from the contents of the studies if the meta-data is not available. They make use of complex regex to improve the accuracy and might work as a more effective process.

The extraction of specific sections of the studies can also be improved by considering the different formats of the papers. Hence, we suggest to incorporate a method that distinguishes between the different study formats and use extraction methods that are specific to each type of study. We believe that this will help us in achieving better precision during extraction. The tool can also be extended to find the other sections of the paper, such as, methods and results.

We also suggest improving the summarization process by various methods. The keyword based summarization can be improved by adding more QA criteria which are more specific to the study. We can also improve the summarization process by using abstractive summarization methods as an alternative to the existing extractive summarization method. The tool can be extended to include these features to get a better summary.

A. Appendix

A.1 Search Results of the SLR

Title	Year
Text Mining to Understand the Influence of Social Media Applications on Smartphone Supply Chain	2018
Towards Accurate Detection of Offensive Language in Online Communication in Arabic	2018
Automatic Construction of E-Government Services Ontology from Arabic Webpages	2018
Performance evaluation of density-based clustering methods	2009
A Rich Arabic WordNet Resource for Al-Hadith Al-Shareef	2017
Vision for SLR tooling infrastructure: Prioritizing value-added requirements	2017
Using Ensemble StackingC Method and Base Classifiers to Ameliorate Prediction Accuracy of Pedagogical Data	2018
Discovering implicit intention-level knowledge from natural-language texts	2009
From Business Intelligence to semantic data stream management	2016
Improving Performance of Text Summarization	2015
Feature extraction approaches from natural language requirements for reuse in software product lines: A systematic literature review	2015
GraphSum: Discovering correlations among multiple terms for graph-based summarization	2013
Data mining and integration for predicting significant meteorological phenomena	2010
Hardcore Gamer Profiling: Results from an unsupervised learning approach to playing behavior on the Steam platform	2018
Detecting discussion communities on vaccination in twitter	2017
ATSSI: Abstractive Text Summarization Using Sentiment Infusion	2016
A systematic mapping study on text analysis techniques in software architecture	2018
Noun retrieval effect on text summarization and delivery of personalized news articles to the users desktop	2010
Ranking Sentences for Keyphrase Extraction: A Relational Data Mining Approach	2014
Temporal Data Mining and Visualization for Treatment Outcome Prediction in HIV Patients	2012
SCIFNET: Stance community identification of topic persons using friendship network analysis	2016
Topic knowledge map and knowledge structure constructions with genetic algorithm, information retrieval, and multi-dimension scaling method	2014
An Enhanced Fuzzy Clustering and Expectation Maximization Framework based Matching Semantically Similar Sentences	2015
Automating the Extraction of Static Content and Dynamic Behaviour from e-Commerce Websites	2017
Topic identification based on document coherence and spectral analysis	2011
Graph-based Arabic NLP Techniques: A Survey	2018
Usability evaluation methods for the web: A systematic mapping study	2011
Web data extraction, applications and techniques: A survey	2014
Investigating the relationship between price, rating, and popularity in the Blackberry World App Store	2017
Data stream mining for predicting software build outcomes using source code metrics	2014
Performance Analysis of Graph based Keyphrase Extraction metrics for uncertain User-generated data	2018
Challenges and Techniques for Testing of Big Data	2016
Testing embedded software: A survey of the literature	2018
A systematic literature review: Opinion mining studies from mobile app store user reviews	2017
Big Data and forensics: An innovative approach for a predictable jurisprudence	2018
Automated discovery of multi-faceted ontologies for accurate query answering and future semantic reasoning	2013
A comparison of automated training-by-example selection algorithms for Evidence Based Software Engineering	2018

Approximation of COSMIC functional size to support early effort estimation in Agile	2013
A Histogram Method for Summarizing Multi-dimensional Probabilistic Data	2013
A systematic literature review of software requirements reuse approaches	2018
Adaptive affinity propagation method based on improved cuckoo search	2016
Warning system for online market research-Identifying critical situations in online opinion formation	2011
An improved methodology on information distillation by mining program source code	2007
Aspect based Summarization of Context Dependent Opinion Words	2014
Identifying the Effective Parameters for Vertical Handover in Cellular Networks Using Data Mining Techniques	2016
Towards Extended Data Mining: An Examination of Technical Aspects	2018
Information Inference in Scholarly Communication Infrastructures: The OpenAIREplus Project Experience	2014
A Systematic Mapping Study driven by the margin of error	2018
ClowdFlows: Online workflows for distributed big data mining	2017
Greedy Algorithms for Granular Computing Problems in Spatial Granulation Technique	2017
A survey of the applications of text mining in financial domain	2016
Optimistic Multi-granulation Rough Set Based Classification for Medical Diagnosis	2015
Summarizing User Opinions: A Method for Labeled-data Scarce Product Domains	2015
Business Intelligence Model to Analyze Social Media Information	2018
Learning to Predict the Need of Summarization on News Articles	2013
A Binary-based MapReduce Analysis for Cloud Logs	2016
A multi-disciplinary review of knowledge acquisition methods: From human to autonomous eliciting agents	2016
Static analysis of android apps: A systematic literature review	2017
Data analytics of urban fabric metrics for smart cities	2018
A Method of Extracting The Semi-structured Data Implication Rules	2018
SwiftRank: An Unsupervised Statistical Approach of Keyword and Salient Sentence Extraction for Individual Documents	2017
Application of clustering methods for detecting critical acute coronary syndrome patients	2018
A systematic literature review on methods that handle multiple quality attributes in architecture-based self-adaptive systems	2017
Knowledge Based Summarization and Document Generation using Bayesian Network	2016
Pattern-based Mining in Electronic Health Records for Complex Clinical Process Analysis	2017
A Study on Ontology Based Abstractive Summarization	2016
Automated browsing in AJAX websites	2011
Ant colony heuristic for user-contributed comments summarization	2017
Mining the Czech Insolvency Proceedings Data	2014
Long range dependence in texts: A method for quantifying coherence of text	2017
A grid-based approach for enterprise-scale data mining	2007
Finding association rules in semantic web data	2012
Machine learning based heterogeneous web advertisements detection using a diverse feature set	2018
Social context summarization using user-generated content and third-party sources	2018
SyMSS: A syntax-based measure for short-text semantic similarity	2011
Prediction of Online Lectures Popularity: A Text Mining Approach	2016
Software requirements selection and prioritization using SBSE approaches: A systematic review and mapping of the literature	2015
Multi-document summarization using closed patterns	2016
A Flexible Keyphrase Extraction Technique for Academic Literature	2018
Topic Detection Approaches in Identifying Topics and Events from Arabic Corpora	2018
Dependable large scale behavioral patterns mining from sensor data using Hadoop platform	2017
A survey on opinion mining and sentiment analysis: Tasks, approaches and applications	2015
Extractive multi-document text summarization using a multi-objective artificial bee colony optimization approach	2018
Social networking data analysis tools & challenges	2018
Performance Evaluation of Distributed Association Rule Mining Algorithms	2016
Data mining agent conversations: A qualitative approach to multiagent systems analysis	2013
Extracting reusable design decisions for UML-based domain-specific languages: A multi-method study	2016
Grid-enabling data mining applications with DataMiningGrid: An architectural perspective	2008
A discretization algorithm based on Class-Attribute Contingency Coefficient	2008
A data mining approach to knowledge discovery from multidimensional cube structures	2013
Requirements monitoring frameworks: A systematic review	2016
An automated Psychometric Analyzer based on Sentiment Analysis and Emotion Recognition for healthcare	2018
Requirements traceability technologies and technology transfer decision support: A systematic review	2018
A short text modeling method combining semantic and statistical information	2010
Traffic big data prediction and visualization using Fast Incremental Model Trees-Drift Detection (FIMT-DD)	2016
Integrating statistical and lexical information for recognizing textual entailments in text	2013
Summarizing data using a similarity based mountain method	2008
A framework for automated construction of resource space based on background knowledge	2014
A noun-based approach to feature location using time-aware term-weighting	2014
Re-discover Values of Data Using Data Jackets by Combining Cluster with Text Analysis	2017
Optimization for Automatic Personality Recognition on Twitter in Bahasa Indonesia	2018
Effectiveness of template detection on noise reduction and websites summarization	2013
A two phased service oriented Broker for replica selection in data grids	2013
A new hybrid semi-supervised algorithm for text classification with class-based semantics	2016
Extracting lists of data records from semi-structured web pages	2008

Semi-supervised Aspect Based Sentiment Analysis for Movies Using Review Filtering	2016
Development of a human error taxonomy for software requirements: A systematic literature review	2018
Big Data Analytics for Behavior Monitoring of Students	2016
New research directions for data and knowledge engineering: A philosophy of language approach	2008
A Service-oriented Framework for Integration of Domain-specific Data Models in Scientific Workflows	2013
Extending the SOA paradigm to e-Science environments	2011
Enhancing diversity and coverage of document summaries through subspace clustering and clustering-based optimization	2014
Multimodal latent topic analysis for image collection summarization	2016
A periodicity-based parallel time series prediction algorithm in cloud computing environments	2018
A multiview approach for intelligent data analysis based on data operators	2008
Generating automatic linguistic descriptions with big data	2017
Advanced knowledge-based systems	2010
Ensemble Classifier for Mining Data Streams	2014
Identifying streaming frequent items in ad hoc time windows	2013
Prognosis of Dementia Employing Machine Learning and Microsimulation Techniques: A Systematic Literature Review	2016
A Framework to Formulate Customer Taste from Unstructured Review Data	2016
Linguistic Resumes in Software Engineering: The Case of Trend Summarization in Mobile Crash Reporting Systems	2016
COWB: A cloud-based framework supporting collaborative knowledge management within biomedical communities	2016
Social information discovery enhanced by sentiment analysis techniques	2018
Software reliability modeling based on ISO/IEC SQuaRE	2016
Potential and limitations of the ISBSG dataset in enhancing software engineering research: A mapping review	2014
A tool for design pattern detection and software architecture reconstruction	2011
Open source software ecosystems: A Systematic mapping	2017
DSS from an RE Perspective: A systematic mapping	2016
Common motifs in scientific workflows: An empirical analysis	2014
Highly-cited papers in software engineering: The top-100	2016
Big Data for Internet of Things: A Survey	2018
Modeling and Management of Big Data: Challenges and opportunities	2016
A Data Cube Model for Analysis of High Volumes of Ambient Data	2012
Automatic text categorization based on content analysis with cognitive situation models	2010
Identification of SLR tool needs – results of a community workshop	2016
Applying Deep Learning for Arabic Keyphrase Extraction	2018
Evaluation of E-Commerce Product Reviews Based on Structural, Metadata, and Readability Characteristics	2017
Information extraction for search engines using fast heuristic techniques	2010
A local variance-controlled reversible data hiding method using prediction and histogram-shifting	2010
Systematic literature review of ensemble effort estimation	2016
DAWQAS: A Dataset for Arabic Why Question Answering System	2018
Knowledge Acquisition for Electronic Health Records on cloud	2017
Chunked N-Grams for Sentence Validation	2015
Pattern graph tracking-based stock price prediction using big data	2018
A Systematic Literature Review Comparing Primary and Community Health Care Indicators and Measurement Frameworks	2017
Multi-modal sliding window-based support vector regression for predicting plant water stress	2017
A systematic review of systematic review process research in software engineering	2013
Multiple criteria mathematical programming for multi-class classification and application in network intrusion detection	2009
Complex Data-driven Predictive Modeling in Personalized Clinical Decision Support for Acute Coronary Syndrome Episodes	2016
A systematic review on security in Process-Aware Information Systems – Constitution, challenges, and future directions	2014
Communication and personality profiles of global software developers	2015
Business Intelligence and Analytics in Small and Medium-sized Enterprises: A Systematic Literature Review	2017
Reproducibility and replicability of software defect prediction studies	2018
An overview of the Applications of Natural Language to Information Systems	2013
On the value of a prioritization scheme for resolving Self-admitted technical debt	2018
A systematic literature review of software visualization evaluation	2018
Identification of Pathophysiological Subclinical Variances During Complex Treatment Process of Cardiovascular Patients	2018
LoCo CoCo: Automatically constructing coordination and communication networks from model-based systems engineering data	2017
Image steganography using uncorrelated color space and its application for security of visual contents in online social networks	2018
SparseHC: A Memory-efficient Online Hierarchical Clustering Algorithm	2014
Search-based software library recommendation using multi-objective optimization	2017
Early software defect prediction: A systematic map and review	2018
Automatically maintaining navigation sequences for querying semi-structured web sources	2007
A Study of Process Mining-based Business Process Innovation	2016
An Improved SRL Based Plagiarism Detection Technique Using Sentence Ranking	2015

Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts	2018
A recommender system of reviewers and experts in reviewing problems	2016
Opinion Ensembling for Improving Economic Growth through Tourism	2017
Automatic Chinese Multiple-Choice Question Generation for Human Resource Performance Appraisal	2018
Document Clustering Using Hybrid XOR Similarity Function for Efficient Software Component Reuse	2013
Automatically maintaining wrappers for semi-structured web sources	2007
Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques	2017
Past and future of software architectures for context-aware systems: A systematic mapping study	2018
Automatic validation of requirements to support multidimensional design	2010
On Continent and Script-Wise Divisions-Based Statistical Measures for Stop-words Lists of International Languages	2016
An Overview of Software Functionality Service: A Systematic Literature Review	2017
Outcomes from Indexing Initiatives of Medical Imaging DICOM Metadata Repositories. A Secondary Analysis	2018
Why good data analysts need to be critical synthesists. Determining the role of semantics in data analysis	2017
Synchronization-based clustering on evolving data stream	2018
The effect of automatic concern mapping strategies on conceptual cohesion measurement	2016
Knowledge management initiatives in software testing: A mapping study	2015
Processing and mining complex data streams	2014
A systematic literature review on the barriers faced by newcomers to open source software projects	2015
Source Code Driven Enterprise Application Decomposition: Preliminary Evaluation	2015
Large scale opinion mining for social, news and blog data	2017
Multi-Document Summarization Using K-Means and Latent Dirichlet Allocation (LDA) – Significance Sentences	2018
Software product lines traceability: A systematic mapping study	2017
Mining software repositories for comprehensible software fault prediction models	2008
Exploring hypergraph-based semi-supervised ranking for query-oriented summarization	2013
Development of Review Rating and Reporting in Open Journal System	2017
Ontology-enriched multi-document summarization in disaster management using submodular function	2013
Language independent web news extraction system based on text detection framework	2016
Efficient sequential pattern mining with wildcards for keyphrase extraction	2017
A component recommender for bug reports using Discriminative Probability Latent Semantic Analysis	2016
Identifying relevant studies in software engineering	2011
An indicative opinion generation model for short texts on social networks	2018
A survey on trends of cross-media topic evolution map	2017
Contents	2015
subject index	2007
An Integrated Framework of Knowledge Transfer and ICT Issues in Co-creation Value Networks	2016
Software startup engineering: A systematic mapping study	2018
Exploring quality measures for the evaluation of feature models: a case study	2017
Key Author Analysis in Research Professionals' Relationship Network Using Citation Indices and Centrality	2015
Integration of Cloud computing and Internet of Things: A survey	2016
Towards High-Dimensional Computational Steering of Precomputed Simulation Data using Sparse Grids	2011
Effect of developer collaboration activity on software quality in two large scale projects	2016
Designing a NIALM in Smart Homes for Cognitive Assistance	2013
Geoblood: A Web Based Tool for Geo-analysis of Biological Data	2016
Repairing inconsistent dimensions in data warehouses	2012
TFRP: An efficient microaggregation algorithm for statistical disclosure control	2007
The thematic and citation landscape of Data and Knowledge Engineering (1985–2007)	2008
Granular computing on information tables: Families of subsets and operators	2018
Knowledge pairing systems in granular computing	2017
Search Based Software Engineering: Review and analysis of the field in Brazil	2013
Fast Emulation of Self-organizing Maps for Large Datasets	2015
Research on data fusion algorithm and anti-collision algorithm based on internet of things	2018
A Theoretical Multi-level Privacy Protection Framework for Biomedical Data Warehouses	2015
The usage of ISBSG data fields in software effort estimation: A systematic mapping study	2016
An Open Framework for Dynamic Big-data-driven Application Systems (DBDDAS) Development	2014
Systematic Literature Review on the Anonymization of High Dimensional Streaming Datasets for Health Data Sharing	2015
Criteria in AHP: A Systematic Review of Literature	2015
Security aspects in healthcare information systems: A systematic mapping	2018
Matching parse thickets for open domain question answering	2017
Quality attributes and quality models for ambient assisted living software systems: A systematic mapping	2017
Pre Processing of Twitter's Data for Opinion Mining in Political Context	2016
Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations	2017
Automotive software engineering: A systematic mapping study	2017
Revisiting distance-based record linkage for privacy-preserving release of statistical datasets	2015
Self-adaptive processing graph with operator fission for elastic stream processing	2017
Data embedding using pixel value differencing and diamond encoding with multiple-base notational system	2012
Reversible data hiding for high quality images using modification of prediction errors	2009
Using Semantic Resources in Image Retrieval	2016
Evaluating and selecting software packages: A review	2009

Measuring the health of open source software ecosystems: Beyond the scope of project health	2014
Feasibility Study of Social Network Analysis on Loosely Structured Communication Networks	2017
Quo Vadis computer science: From Turing to personal computer, personal content and collective intelligence	2008
Links between the personalities, styles and performance in computer programming	2016
A secure kNN query processing algorithm using homomorphic encryption on outsourced database	2017
Identifying and characterizing change-prone classes in two large-scale open-source products	2007
Improving the performance of question answering with semantically equivalent answer patterns	2008
A hierarchical semantic-based distance for nominal histogram comparison	2013
Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems	2016
Exploring the links between software development task type, team attitudes and task completion performance: Insights from the Jazz repository	2018
Understanding the attitudes, knowledge sharing behaviors and task performance of core developers: A longitudinal study	2014
A functional interpretation of linguistic summaries of data	2012
Discovering the core semantics of event from social media	2016
Cognitive memory-inspired sentence ordering model	2016
A systematic mapping study of search-based software engineering for software product lines	2015
Explaining classifier decisions linguistically for stimulating and improving operators labeling behavior	2017
Exploiting relevance, coverage, and novelty for query-focused multi-document summarization	2013
Assessing data analysis performance in research contexts: An experiment on accuracy, efficiency, productivity and researchers' satisfaction	2018
View-based model-driven architecture for enhancing maintainability of data access services	2011
Design science research contribution to business intelligence in the cloud — A systematic literature review	2016
Manufacturing execution systems: A vision for managing software development	2015
Designing service-based applications in the presence of non-functional properties: A mapping study	2016
An Enhanced xAPI Data Model Supporting Assessment Analytics	2018
Performance evaluation of parallel strategies in public clouds: A study with phylogenomic workflows	2013
Software clone detection: A systematic review	2013
Big Data Analytics in the Cloud: Spark on Hadoop vs MPI/OpenMP on Beowulf	2015
A framework for multidimensional design of data warehouses from ontologies	2010
Sentiment analysis: A review and comparative analysis of web services	2015
Privacy by diversity in sequential releases of databases	2015
Limiting disclosure of sensitive data in sequential releases of databases	2012
A systematic mapping study about socio-technical congruence	2018
MSR4SM: Using topic models to effectively mining software repositories for software maintenance tasks	2015
Full-fledged semantic indexing and querying model designed for seamless integration in legacy RDBMS	2018
An Architecture for Integrating Genetic and Clinical Data	2014
The relationship between design patterns and code smells: An exploratory study	2016
Systematic literature review of machine learning based software development effort estimation models	2012
Reversible data hiding based on an adaptive pixel-embedding strategy and two-layer embedding	2016
Indonesian Named-entity Recognition for 15 Classes Using Ensemble Supervised Learning	2016
Migration of the Individuals	2016
Modeling query-document dependencies with topic language models for information retrieval	2015
Near-synonym substitution using a discriminative vector space model	2016
A Chinese time ontology for the Semantic Web	2011
Data analysis based on discernibility and indiscernibility	2007
Towards comprehending the non-functional requirements through Developers' eyes: An exploration of Stack Overflow using topic analysis	2017
Optimization for Automatic Personality Recognition on Twitter in Bahasa Indonesia	2018
Effectiveness of template detection on noise reduction and websites summarization	2013
A two phased service oriented Broker for replica selection in data grids	2013
A new hybrid semi-supervised algorithm for text classification with class-based semantics	2016
Extracting lists of data records from semi-structured web pages	2008
Semi-supervised Aspect Based Sentiment Analysis for Movies Using Review Filtering	2016
Development of a human error taxonomy for software requirements: A systematic literature review	2018
Big Data Analytics for Behavior Monitoring of Students	2016
New research directions for data and knowledge engineering: A philosophy of language approach	2008
A Service-oriented Framework for Integration of Domain-specific Data Models in Scientific Workflows	2013
Extending the SOA paradigm to e-Science environments	2011
Enhancing diversity and coverage of document summaries through subspace clustering and clustering-based optimization	2014
Multimodal latent topic analysis for image collection summarization	2016
A periodicity-based parallel time series prediction algorithm in cloud computing environments	2018
A multiview approach for intelligent data analysis based on data operators	2008
Generating automatic linguistic descriptions with big data	2017
Advanced knowledge-based systems	2010
Ensemble Classifier for Mining Data Streams	2014
Identifying streaming frequent items in ad hoc time windows	2013
Prognosis of Dementia Employing Machine Learning and Microsimulation Techniques: A Systematic Literature Review	2016
A Framework to Formulate Customer Taste from Unstructured Review Data	2016

Linguistic Resumes in Software Engineering: The Case of Trend Summarization in Mobile Crash Reporting Systems	2016
COWB: A cloud-based framework supporting collaborative knowledge management within biomedical communities	2016
Social information discovery enhanced by sentiment analysis techniques	2018
Software reliability modeling based on ISO/IEC SQuaRE	2016
Potential and limitations of the ISBSG dataset in enhancing software engineering research: A mapping review	2014
A tool for design pattern detection and software architecture reconstruction	2011
Open source software ecosystems: A Systematic mapping	2017
DSS from an RE Perspective: A systematic mapping	2016
Common motifs in scientific workflows: An empirical analysis	2014
Highly-cited papers in software engineering: The top-100	2016
Big Data for Internet of Things: A Survey	2018
Modeling and Management of Big Data: Challenges and opportunities	2016
A Data Cube Model for Analysis of High Volumes of Ambient Data	2012
Automatic text categorization based on content analysis with cognitive situation models	2010
Identification of SLR tool needs – results of a community workshop	2016
Applying Deep Learning for Arabic Keyphrase Extraction	2018
Evaluation of E-Commerce Product Reviews Based on Structural, Metadata, and Readability Characteristics	2017
Information extraction for search engines using fast heuristic techniques	2010
A local variance-controlled reversible data hiding method using prediction and histogram-shifting	2010
Systematic literature review of ensemble effort estimation	2016
DAWQAS: A Dataset for Arabic Why Question Answering System	2018
Knowledge Acquisition for Electronic Health Records on cloud	2017
Chunked N-Grams for Sentence Validation	2015
Pattern graph tracking-based stock price prediction using big data	2018
A Systematic Literature Review Comparing Primary and Community Health Care Indicators and Measurement Frameworks	2017
Multi-modal sliding window-based support vector regression for predicting plant water stress	2017
A systematic review of systematic review process research in software engineering	2013
Multiple criteria mathematical programming for multi-class classification and application in network intrusion detection	2009
Complex Data-driven Predictive Modeling in Personalized Clinical Decision Support for Acute Coronary Syndrome Episodes	2016
A systematic review on security in Process-Aware Information Systems – Constitution, challenges, and future directions	2014
Communication and personality profiles of global software developers	2015
Business Intelligence and Analytics in Small and Medium-sized Enterprises: A Systematic Literature Review	2017
Reproducibility and replicability of software defect prediction studies	2018
An overview of the Applications of Natural Language to Information Systems	2013
On the value of a prioritization scheme for resolving Self-admitted technical debt	2018
A systematic literature review of software visualization evaluation	2018
Identification of Pathophysiological Subclinical Variances During Complex Treatment Process of Cardiovascular Patients	2018
LoCo CoCo: Automatically constructing coordination and communication networks from model-based systems engineering data	2017
Image steganography using uncorrelated color space and its application for security of visual contents in online social networks	2018
SparseHC: A Memory-efficient Online Hierarchical Clustering Algorithm	2014
Search-based software library recommendation using multi-objective optimization	2017
Early software defect prediction: A systematic map and review	2018
Automatically maintaining navigation sequences for querying semi-structured web sources	2007
A Study of Process Mining-based Business Process Innovation	2016
An Improved SRL Based Plagiarism Detection Technique Using Sentence Ranking	2015
Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts	2018
A recommender system of reviewers and experts in reviewing problems	2016
Opinion Ensembling for Improving Economic Growth through Tourism	2017
Automatic Chinese Multiple-Choice Question Generation for Human Resource Performance Appraisal	2018
Document Clustering Using Hybrid XOR Similarity Function for Efficient Software Component Reuse	2013
Automatically maintaining wrappers for semi-structured web sources	2007
Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques	2017
Past and future of software architectures for context-aware systems: A systematic mapping study	2018
Automatic validation of requirements to support multidimensional design	2010
On Continent and Script-Wise Divisions-Based Statistical Measures for Stop-words Lists of International Languages	2016
An Overview of Software Functionality Service: A Systematic Literature Review	2017
Outcomes from Indexing Initiatives of Medical Imaging DICOM Metadata Repositories. A Secondary Analysis	2018
Why good data analysts need to be critical synthesists. Determining the role of semantics in data analysis	2017
Synchronization-based clustering on evolving data stream	2018
The effect of automatic concern mapping strategies on conceptual cohesion measurement	2016
Knowledge management initiatives in software testing: A mapping study	2015

Processing and mining complex data streams	2014
A systematic literature review on the barriers faced by newcomers to open source software projects	2015
Source Code Driven Enterprise Application Decomposition: Preliminary Evaluation	2015
Large scale opinion mining for social, news and blog data	2017
Multi-Document Summarization Using K-Means and Latent Dirichlet Allocation (LDA) – Significance Sentences	2018
Software product lines traceability: A systematic mapping study	2017
Mining software repositories for comprehensible software fault prediction models	2008
Exploring hypergraph-based semi-supervised ranking for query-oriented summarization	2013
Development of Review Rating and Reporting in Open Journal System	2017
Ontology-enriched multi-document summarization in disaster management using submodular function	2013
Language independent web news extraction system based on text detection framework	2016
Efficient sequential pattern mining with wildcards for keyphrase extraction	2017
A component recommender for bug reports using Discriminative Probability Latent Semantic Analysis	2016
Identifying relevant studies in software engineering	2011
An indicative opinion generation model for short texts on social networks	2018
A survey on trends of cross-media topic evolution map	2017
Survey on mining subjective data on the web	2012
Extracting abstract and keywords from context for academic articles	2018
Mining Chinese social media UGC: a big-data framework for analyzing Douban movie reviews	2016
PatSearch: an integrated framework for patentability retrieval	2018
The big data system, components, tools, and technologies: a survey	2018
Big data analytics: a survey	2015
Mining unstructured content for recommender systems: an ensemble approach	2016
Survey on using constraints in data mining	2017
A review of data mining using big data in health informatics	2014
Sourcerer: mining and searching internet-scale software repositories	2009
Ontology of core data mining entities	2014
Syntactic sentence compression in the biomedical domain: facilitating access to related articles	2007
Scaling up data mining algorithms: review and taxonomy	2012
Interpretation of text patterns	2018
A survey of sentiment analysis in social media	2018
Network text analysis of conceptual overlap in interviews, newspaper articles and keywords	2013
Profiling Web users using big data	2018
Theory-driven or process-driven prediction? Epistemological challenges of big data analytics	2017
Review of social media analytics process and Big Data pipeline	2018
Homomorphic Pattern Mining from a Single Large Data Tree	2016
An architectural framework for developing a recommendation system to enhance vendors™ capability in C2C social commerce	2018
Are we done with business process compliance: state of the art and challenges ahead	2018
Social media and political communication: a social media analytics framework	2013
Investigations into Data Ecosystems: a systematic mapping study	2019
Keyphrase Extraction Using Knowledge Graphs	2017
Understanding big data themes from scientific biomedical literature through topic modeling	2016
Mining local and tail dependence structures based on pointwise mutual information	2012
Discovering outlying aspects in large datasets	2016
Discovering and characterizing political elite cliques with evolutionary community detection	2013
A systematic review of provenance systems	2018
Data Change Exploration Using Time Series Clustering	2018
How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis	2018
A synthetic data generator for online social network graphs	2016
An emotional polarity analysis of consumers™ airline service tweets	2013
A pruning strategy to improve pairwise comparison-based near-duplicate detection	2019
Controlled experiments on the web: survey and practical guide	2009
State-of-the-art in biomedical literature retrieval for clinical cases: a survey of the TREC 2014 CDS track	2016
Spatiotemporal traffic network analysis: technology and applications	2018
User characterization for online social networks	2016
Harvesting and analysis of weak signals for detecting lone wolf terrorists	2013
Learning from streaming data with concept drift and imbalance: an overview	2012
Actionable Knowledge As A Service (AKAAS): Leveraging big data analytics in cloud computing environments	2015
#nowplaying Madonna: a large-scale evaluation on estimating similarities between music artists and between movies from microblogs	2012
Learning from imbalanced data: open challenges and future directions	2016
Cluster-based trajectory segmentation with local noise	2018
Personalized Information Seeking Assistant (PiSA): from programming information seeking to learning	2017
High-level event recognition in unconstrained videos	2013
Accuracy evaluation of methods and techniques in Web-based question answering systems: a survey	2018
Manufacturing process data analysis pipelines: a requirements analysis and survey	2019
Verification of lack of emergent behavior in extending a social network of agents	2015
Multimodal biomedical image retrieval using hierarchical classification and modality fusion	2013
The many faces of data-centric workflow optimization: a survey	2018

Robust non-negative matrix factorization via joint sparse and graph regularization for transfer learning	2013
Optimal Compressed Sensing and Reconstruction of Unstructured Mesh Datasets	2018
Autonomic workload performance tuning in large-scale data repositories	2018
SVM-based robust image watermarking technique in LWT domain using different sub-bands	2018
Meta-analysis of evaluation methods and metrics used in context-aware scholarly recommender systems	2019
Optimization of condition-based maintenance using soft computing	2017
A Methodological Framework for the Integrated Design of Decision-Intensive Care Pathways an Application to the Management of COPD Patients	2017
Novel approach for estimating solubility of solid drugs in supercritical carbon dioxide and critical properties using direct and inverse artificial neural network (ANN)	2017
Bayesian learning of inverted Dirichlet mixtures for SVM kernels generation	2013
Hybrid optimization with cryptography encryption for medical image security in Internet of Things	2018
Application of the fuzzy min-max neural network to fault detection and diagnosis of induction motors	2013
Large-Scale Time Series Analytics	2019
Early warning analysis for social diffusion events	2012
A framework of identity resolution: evaluating identity attributes and matching algorithms	2015
HierFlat: flattened hierarchies for improving top-down hierarchical classification	2017
Bionic knowledge and information reuse methodology for uncertainty minimization in product design	2018
Classification of M-QAM and M-PSK signals using genetic programming (GP)	2018
Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques	2018
Emotion classification of social media posts for estimating peoples reactions to communicated alert messages during crises	2014
Comparing sighted and blind users task performance in responsive and non-responsive web design	2018
Principles of data mining	2007
Introduction to data mining	2007
Artificial intelligence: a modern approach	2016
A method of inquiry	2008
Data mining: concepts and techniques	2011
The WEKA data mining software: an update	2009
Engineering design: a systematic approach	2013
Computer architecture: a quantitative approach	2011
Strategic management: A stakeholder approach	2010
Adam: A method for stochastic optimization	2014
Introductory econometrics: A modern approach	2015
The descriptive phenomenological method in psychology: A modified Husserlian approach.	2009
Truth and method (Bloomsbury revelations)	2013
Qualitative research design: An interactive approach	2012
The economic approach to human behavior	2013
A method and server for predicting damaging missense mutations	2010
Data Mining: Practical machine learning tools and techniques	2016
The ecological approach to visual perception: classic edition	2014
Simulation and the Monte Carlo method	2016
Foundations in sociolinguistics: An ethnographic approach	2013
Research methods for business: A skill building approach	2016
Analyzing real-time PCR data by the comparative CT method	2008
The finite element method in electromagnetics	2015
Caring: A relational approach to ethics and moral education	2013
Method in social science: revised 2nd edition	2010
Top 10 algorithms in data mining	2008
Clinical trials: a practical approach	2013
Software metrics: a rigorous and practical approach	2014
Measurement, design, and analysis: An integrated approach	2013
Intercultural communication: A discourse approach	2011
An introduction to computational fluid dynamics: the finite volume method	2007
Voices of the mind: Sociocultural approach to mediated action	2009
Practical HPLC method development	2012
The comparative method: Moving beyond qualitative and quantitative strategies	2014
Igneous petrogenesis a global tectonic approach	2007
Robust statistics: the approach based on influence functions	2011
The montessori method	2013
Machine learning: An artificial intelligence approach	2013
Saliency detection: A spectral residual approach	2007
Auditing and assurance services: an integrated approach	2012
Elementary numerical analysis: an algorithmic approach	2017
The finite element method: linear static and dynamic finite element analysis	2012
Critical discourse studies: A sociocognitive approach	2015
Strategic management: theory: an integrated approach	2014
Behavioural ecology: an evolutionary approach	2009
Empathy: A social psychological approach	2018
Explaining science: A cognitive approach	2010
Method in theology	2017
Social Change and Crime Rate Trends: A Routine Activity Approach (1979)	2016

Project management: a managerial approach	2011
Using surveys to value public goods: the contingent valuation method	2013
Distributed optimization and statistical learning via the alternating direction method of multipliers	2011
Project management: a systems approach to planning, scheduling, and controlling	2017
Field effect transistor manufacturing method	2010
Aircraft Design: A Conceptual Approach 5e and RDSWin STUDENT	2012
A topical approach to life-span development, 3E	2007
Computer networks: a systems approach	2007
Tourism: A community approach (RLE Tourism)	2013
Universal sample preparation method for proteome analysis	2009
Semiconductor device and manufacturing method thereof	2010
The method of second quantization	2012
Attention and interpretation: A scientific approach to insight in psycho-analysis and groups	2013
Document analysis as a qualitative research method	2009
Method of fabricating oxide semiconductor device	2008
New rules of sociological method: A positive critique of interpretative sociologies	2013
Semiconductor device and manufacturing method thereof	2010
Vegetation description and data analysis: a practical approach	2011
Control system synthesis: a factorization approach, part II	2011
Computational complexity: a modern approach	2009
The split Bregman method for L1-regularized problems	2009
ADADELTA: an adaptive learning rate method	2012
The exergy method of thermal plant analysis	2013
Internationalisation in industrial systems a network approach	2015
Predicting and changing behavior: The reasoned action approach	2011
STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method	2012
Psychoanalytic treatment: An intersubjective approach	2014
Mathematical statistics: A decision theoretic approach	2014
The finite element method for engineers	2008
Sociological work: Method and substance	2017
System identification: a frequency domain approach	2012
A primer on the Taguchi method	2010
Data mining with big data	2014
Critical ethnography: Method, ethics, and performance	2011
Fused nanocrystal thin film semiconductor and method	2009
Tools of the mind: The Vygotskian approach to early childhood education	2007
The descriptive phenomenological psychological method	2012
Data mining: concepts, models, methods, and algorithms	2011
Solving frontier problems of physics: the decomposition method	2013
Multimodality: A social semiotic approach to contemporary communication	2009
Learning to rank: from pairwise approach to listwise approach	2007
Industrial Technological Development (Routledge Revivals): A Network Approach	2015
Ensemble empirical mode decomposition: a noise-assisted data analysis method	2009
Validity and reliability of the experience-sampling method	2014
The analysis of the self: A systematic approach to the psychoanalytic treatment of narcissistic personality disorders	2013
Comprehensive approach to modeling and simulation of photovoltaic arrays	2009
Programming massively parallel processors: a hands-on approach	2016
Attention and self-regulation: A control-theory approach to human behavior	2012
A diagnostic approach for going beyond panaceas	2007
Plants and microclimate: a quantitative approach to environmental plant physiology	2013
A constructivist approach to teaching	2012
The method of volume averaging	2013
Programming the finite element method	2013
A step-by-step approach to using SAS for factor analysis and structural equation modeling	2013
The method of weighted residuals and variational principles	2013
Linear system theory: the state space approach	2008
Shocks and frictions in US business cycles: A Bayesian DSGE approach	2007
Network medicine: a network-based approach to human disease	2011
The Delphi method for graduate research	2007
Discovering knowledge in data: an introduction to data mining	2014
Hayes, Andrew F.(2013). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression?Based Approach. New York, NY: The Guilford Press	2014
Network economics: A variational inequality approach	2013
Field-effect transistor and method for manufacturing the same	2008
Electrospinning: a fascinating method for the preparation of ultrathin fibers	2007
Method for making electronic devices using metal oxide nanoparticles	2009
Matching NLO QCD computations with parton shower simulations: the POWHEG method	2007
Thin-film device and method of fabricating the same	2011
Mail and Internet surveys: The tailored design method–2007 Update with new Internet, visual, and mixed-mode guide	2011

Comparative plant ecology: a functional approach to common British species	2014
Transparent conductive film and method for manufacturing the same	2011
Kanski's Clinical Ophthalmology E-Book: A Systematic Approach	2015
Dual reciprocity boundary element method	2012
Feature selection for knowledge discovery and data mining	2012
Development through life: A psychosocial approach	2017
Planning health promotion programs: an intervention mapping approach	2016
Electromagnetic simulation using the FDTD method	2013
The econometric approach to efficiency analysis	2008
Connectionist speech recognition: a hybrid approach	2012
Biogeography: an ecological and evolutionary approach	2016
Techniques for nuclear and particle physics experiments: a how-to approach	2012
Data analysis: A model comparison approach	2011
The psychology of humor: An integrative approach	2018
Molecular evolution: a phylogenetic approach	2009
ZnO thin film transistor and method of forming the same	2007
Sociolinguistics: Method and interpretation	2008
Thin film transitor substrate and method of manufacturing the same	2008
The nesC language: A holistic approach to networked embedded systems	2014
The psychology of religion: An empirical approach	2018
The improvement guide: a practical approach to enhancing organizational performance	2009
Constructing social research: The unity and diversity of method	2010
An introduction to logic and scientific method	2013
Crisis communications: A casebook approach	2016
Thin film transistor and method of manufacturing thin film transistor	2010
Vapor-liquid equilibria using UNIFAC: a group-contribution method	2012
Global marketing: A decision-oriented approach	2007
Statistical analysis: a computer oriented approach	2014
Targeting cancer cells by ROS-mediated mechanisms: a radical therapeutic approach?	2009
Semiconductor device, electronic device, and method of manufacturing semiconductor device	2009
Abnormal psychology: An integrative approach	2011
Stress Inoculation Training: A preventative and treatment approach	2017
Nonparametric statistics: A step-by-step approach	2014
Semiconductor device, electronic device, and method of manufacturing semiconductor device	2009
Creating significant learning experiences: An integrated approach to designing college courses	2013
Petroleum formation and occurrence: a new approach to oil and gas exploration	2012
Method meets art: Arts-based research practice	2015
Toyota production system: an integrated approach to just-in-time	2011
Electronic materials: the oligomer approach	2008
Numerical solution of partial differential equations by the finite element method	2012
Oxide semiconductor thin film transistor and method of manufacturing the same	2007
Associative memory: A system-theoretical approach	2012
Muscle Biopsy: A Practical Approach: Expert Consult; Online and Print	2013
Modern Raman spectroscopy: a practical approach	2013
Aphid ecology an optimization approach	2012
Family stress management: A contextual approach	2016
Human resource management: A contemporary approach	2007
Semiconductor device and method for manufacturing the same, and electric device	2009
A semi-empirical approach to projecting future sea-level rise	2007
The text mining handbook: advanced approaches in analyzing unstructured data	2007
Systematic literature reviews in software engineering a systematic literature review	2009
Semiconductor device, and display device, driving method and electronic apparatus thereof	2014
On the controllable soft-templating approach to mesoporous silicates	2007
ABINIT: First-principles approach to material and nanosystem properties	2009
Basic principles of plasma physics: a statistical approach	2018
Experimental pulse NMR: a nuts and bolts approach	2018
Contemporary hermeneutics: Hermeneutics as method, philosophy and critique	2017
Educational data mining: A survey from 1995 to 2005	2007
Thin film transistor having an etching protection film and manufacturing method thereof	2008
Organic electroluminescent display device having a planarizing layer and manufacturing method thereof	2010
Phenomenological psychology: Theory, research and method	2007
A scaling normalization method for differential expression analysis of RNA-seq data	2010
A tractable approach to coverage and rate in cellular networks	2011
Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach: A Signal Processing Approach	2012
A green approach to the synthesis of graphene nanosheets	2009
A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species	2011
The Monte Carlo method for semiconductor device simulation	2012
An Interior-Point Method for Large-Scale -Regularized Least Squares	2007
Shape optimization by the homogenization method	2012
Practical model-based testing: a tools approach	2010
Air pollution control: A design approach	2010
METHOD OF FABRICATING ZnO FILM AND THIN FILM TRANSISTOR ADOPTING THE ZnO FILM	2007

Open source GIS: a GRASS GIS approach	2013
A holistic lexicon-based approach to opinion mining	2008
Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval	2013
Data mining techniques: for marketing, sales, and customer relationship management	2011
Supervised descent method and its applications to face alignment	2013
Structure, transmission type liquid crystal display, reflection type display and manufacturing method thereof	2007
Method and apparatus for browsing using multiple coordinated device sets	2011
Thin film transistor including low resistance conductive thin films and manufacturing method thereof	2009
Data mining with decision trees: theory and applications	2008
ML-KNN: A lazy learning approach to multi-label learning	2007
Introduction to mathematical systems theory: a behavioral approach	2013
From the editors: Common method variance in international business research	2010
A general and simple method for obtaining R2 from generalized linear mixed?effects models	2013
An Introduction to Management Science: Quantitative Approach	2018
Analysis of physiological systems: The white-noise approach	2012
Drug combination studies and their synergy quantification using the Chou-Talalay method	2010
Functions of innovation systems: A new approach for analysing technological change	2007
High-throughput functional annotation and data mining with the Blast2GO suite	2008
Clinical prediction models: a practical approach to development, validation, and updating	2008
A voltage and frequency droop control method for parallel inverters	2007
Art, mind, and brain: A cognitive approach to creativity	2008
A new multipoint method for genome-wide association studies by imputation of genotypes	2007
Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories	2007
Choose your method: A comparison of phenomenology, discourse analysis, and grounded theory	2007
Sources of method bias in social science research and recommendations on how to control it	2012
Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences	2010
Semiconductor device with an active layer containing zinc oxide, manufacturing method, and electronic device	2012
High-performance lithium-ion anodes using a hierarchical bottom-up approach	2010
Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty: an end-result study using a new method of result evaluation	2014
Stability by Liapunov's Direct Method with Applications by Joseph L Salle and Solomon Lefschetz	2012
The LMTO method: muffin-tin orbitals and electronic structure	2012
Network analysis and synthesis: a modern systems theory approach	2013
A first course in the finite element method	2011
The biopsychosocial approach to chronic pain: scientific advances and future directions.	2007
The numerical method of lines: integration of partial differential equations	2012
Volcanic successions modern and ancient: A geological approach to processes, products and successions	2012
Method of fusing biomaterials with radiofrequency energy	2007
Designing for situation awareness: An approach to user-centered design	2016
Data mining static code attributes to learn defect predictors	2007
Method and apparatus for browsing using alternative linkbases	2011
In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro	2007
Hierarchical control of droop-controlled AC and DC microgrids. A general approach toward standardization	2011
Word representations: a simple and general method for semi-supervised learning	2010
Pixel structure of active matrix organic light-emitting diode and method for fabricating the same	2007
Mechanics of composite materials: a unified micromechanical approach	2013
Evolution and structure of the Internet: A statistical physics approach	2007
A course in robust control theory: a convex approach	2013
Experience sampling method: Measuring the quality of everyday life	2007
Dialogue and the development of children's thinking: A sociocultural approach	2007
Faster and better: A machine learning approach to corner detection	2010
The behaviour change wheel: a new method for characterising and designing behaviour change interventions	2011
The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices	2010
The institutional logics perspective: A new approach to culture, structure, and process	2012
Aviation automation: The search for a human-centered approach	2018
H-infinity optimal control and related minimax design problems: a dynamic game approach	2008
Substrate for growing wurtzite type crystal and method for manufacturing the same and semiconductor device	2014
Calculus on manifolds: a modern approach to classical theorems of advanced calculus	2018
Focusing-oriented psychotherapy: A manual of the experiential method	2012
Reciprocally convex approach to stability of systems with time-varying delays	2011
MEEP: A flexible free-software package for electromagnetic simulations by the FDTD method	2010
The rainbow of desire: The Boal method of theatre and therapy	2013
Computer-aided verification of coordinating processes: the automata-theoretic approach	2014
The skilled helper: A problem-management and opportunity-development approach to helping	2013
A flexible and accurate genotype imputation method for the next generation of genome-wide association studies	2009
Educational data mining: a review of the state of the art	2010
What form transforms? A constructive-developmental approach to transformative learning	2009
RELION: implementation of a Bayesian approach to cryo-EM structure determination	2012
The impact of supply chain integration on performance: A contingency and configuration approach	2010
The medical care costs of obesity: an instrumental variables approach	2012

TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts	2009
The finite element method in heat transfer and fluid dynamics	2010
Entropy generation minimization: the method of thermodynamic optimization of finite-size systems and finite-time processes	2013
Stress, health, and the social environment: A sociobiologic approach to medicine	2013
Method and apparatus for displaying network-based deal transactions	2008
High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method	2007
Method and system for managing distributed content and related metadata	2007
Distributed GraphLab: a framework for machine learning and data mining in the cloud	2012
A universal method to produce low work function electrodes for organic electronics	2012
Know thyself and become what you are: A eudaimonic approach to psychological well-being	2008
Distributed control of robotic networks: a mathematical approach to motion coordination algorithms	2009
Abstract state machines: a method for high-level system design and analysis	2012
Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels	2009
Semiconductor device in which zinc oxide is used as a semiconductor material and method for manufacturing the semiconductor device	2009
What reviewers should expect from authors regarding common method bias in organizational research	2010
The how of happiness: A scientific approach to getting the life you want	2008
Global supplier development considering risk factors using fuzzy extended AHP-based approach	2007
The (G? G)-expansion method and travelling wave solutions of nonlinear evolution equations in mathematical physics	2008
Purposeful sampling for qualitative data collection and analysis in mixed method implementation research	2015
Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach	2008
A systematic literature review of empirical evidence on computer games and serious games	2012
Research methods in applied settings: An integrated approach to design and analysis	2011
Discriminant analysis of principal components: a new method for the analysis of genetically structured populations	2010
Teaching learning-based optimization: a novel method for constrained mechanical design optimization problems	2011
Using the framework method for the analysis of qualitative data in multi-disciplinary health research	2013
Touch screen device, method, and graphical user interface for determining commands by applying heuristics	2009
Reinventing project management: the diamond approach to successful growth and innovation	2007
LnCuO (S, Se, Te) monocrystalline thin film, its manufacturing method, and optical device or electronic device using the monocrystalline thin film	2008
Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on	2010
Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework.	2011
Automated solution of differential equations by the finite element method: The FEniCS book	2012
Management of hyperglycaemia in type 2 diabetes: a patient-centered approach. Position statement of the American Diabetes Association (ADA) and the	2012
An open systems approach to quantum optics: lectures presented at the Université Libre de Bruxelles, October 28 to November 4, 1991	2009
System, method, and computer program product for providing location based services and mobile e-commerce	2009
A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer	2007
Listening to Young Children, Expanded Third Edition: A Guide to Understanding and Using the Mosaic Approach	2017
Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach	2016
A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective	2008
The Finite Element Method in Engineering: Pergamon International Library of Science, Technology, Engineering and Social Studies	2013
Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach-2010 revision	2010
A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals	2009
Method of delivery and pregnancy outcomes in Asia: the WHO global survey on maternal and perinatal health 2007-08	2010
Management of hyperglycemia in type 2 diabetes, 2015: a patient-centered approach: update to a position statement of the American Diabetes Association	2015
Thin film etching method	2011
Obesity controlling method	2009
Phase field method	2010
Method of dry etching oxide semiconductor film	2010
Dry etching method for oxide semiconductor film	2011
Semiconductor apparatus and method of manufacturing the same	2010
Somalia: A new approach	2010
Method of securing vertebrae	2010

Method for manufacturing adjustable lens	2014
Sound library and method	2016
Color EL display and method for producing the same	2012
Surgical access apparatus and method	2012
Customer identification system and method	2007
Method for manufacturing semiconductor device	2013
Identification verification system and method	2010
Deconvolution: A wavelet frame approach	2007
Information exploration systems and method	2010
The cognitive approach to language	2008
Organic light emitting display (OLED) and its method of fabrication	2010
Optimization modelling: a practical approach	2007
Web data mining: exploring hyperlinks, contents, and usage data	2007
Apparatus and method for authentication and method for registering a person	2007
Reliability sensitivity method by line sampling	2008
The continual reassessment method for multiple toxicity grades: a Bayesian quasi?likelihood approach	2007
Method and apparatus for data normalization	2008
Method and apparatus for multipath processing	2008
Bulldozer: An approach to multithreaded compute performance	2011
Bone defect repair device and method	2014
Thin film transistor having oxide semiconductor layer and manufacturing method thereof	2009
Wireless power distribution system and method	2016
The retrograde approach to coronary artery chronic total occlusions: a practical approach	2012
A new approach to funding social enterprises	2012
Method for transferring product service records	2008
A new constrained independent component analysis method	2007
System and method for seamless switching	2009
Curviliner spinal access method and device	2012
Client proximity detection method and system	2011
Method and system for fall detection	2012
System and method for managing application alerts	2007
Spring clip and method for assembling same	2010
Discontinuous Galerkin methods: general approach and stability	2009
A variational approach to copositive matrices	2010
System and method for telephonic voice authentication	2007
An evolutionary method for complex-process optimization	2010
System and method for convenience gaming	2013
A numerical method for incompressible non-Newtonian fluid flows based on the lattice Boltzmann method	2007
Modelling thrombosis using dissipative particle dynamics method	2008
Disk system and method of updating firmware	2007
Method for fabricating pixel structure of active matrix organic light-emitting diode	2009
Method for controlling LED-based backlight module	2007
An intersectional approach to men's health	2012
Adjustable implant and method of use	2011
Expandable support device and method of use	2008
System and method for interactive contests	2014
Preparation of ZnO nanorods through wet chemical method	2007
Meld: A declarative approach to programming ensembles	2007
Signal transmission multiple antenna method and device	2008
Generating device, encryption device, decryption device, generating method, encryption method, decryption method, and computer program product	2018
Method of manufacturing an optical composite	2011
An enrichment method to detect low concentration formaldehyde	2008
A multiple object tracking method using Kalman filter	2010
Method of fabricating mobility enhanced CMOS devices	2007
A hybrid approach to private record linkage	2008
A method of converting Z-number to classical fuzzy number	2012
Antivirus protection system and method for computers	2011
A multilevel approach to geography of innovation	2010
A discontinuous?Galerkin?based immersed boundary method	2008
An approach for practical multiobjective IMRT treatment planning	2007
Data synchronization method and system between devices	2012
Distributed network data storage system and method	2007
NOA: a novel Network Ontology Analysis method	2011
System and method for cloud enterprise services	2016
Method for coating articles by mold transfer	2011
Virtual machine configuration system and method thereof	2008
Method of making a floorboard and method of making a floor with the floorboard	2007
Green entrepreneurship: A method for managing natural resources?	2008
Development of Finite Element Limiting Analysis Method and Its Applications in Geotechnical Engineering [J]	2007
Multicriteria evaluation of inner climate by using MOORA method	2008

Method of sealing tissue using radiofrequency energy	2013
Method for forming an endoscope articulation joint	2011
Method and apparatus for sharing viewing preferences	2009
Randomization based probabilistic approach to detect trojan circuits	2008
A contraction theory approach to stochastic incremental stability	2009
Method and system for distributing multiple dragged objects	2009
An ontology-centric approach to sensor-mission assignment	2008
A wavelet optimization approach for ECG signal classification	2012
Wireless communication apparatus and wireless communication method	2014
Exposure apparatus and method for producing device	2008
System and method for data stream processing	2012
Method and system for synchronization of content rendering	2008
An ontology-based approach to text summarization	2008
X-ray imaging apparatus and its control method	2008
Webpage processing method and system for mobile terminal	2009
Battery controller and method for controlling a battery	2007
Method and system for network time protocol forwarding	2009
Method and apparatus for controlling wireless power transmission	2015
Method and apparatus for processing a dynamic webpage	2008
Multi-level cell memory device and method thereof	2011
Method and system for encapsulating variable-size packets	2008
Through-silicon via and method for forming the same	2010
Sealing device and method for a processing system	2010
Presence Detector and Method for Estimating an Audience	2010
Emotion as motion: Asymmetries in approach and avoidant actions	2007
Landscape phenology: an integrative approach to seasonal vegetation dynamics	2009
Method of controlling transmit power of uplink channel	2013
Device and method for analyzing an information signal	2011
Method and system for applying input mode bias	2010
Store and forward switch device, system and method	2007
Method for connection reconfiguration in cellular radio network	2009
Implantable biomaterial and a method of producing same	2015
New approach to q-Euler polynomials of higher order	2010
Method and apparatus for medical intervention procedure planning	2008
Benchmarking: a method for continuous quality improvement in health	2012
Method and apparatus for continuous guidance of endoscopy	2014
Thin film transistor, method of manufacturing the same, and flat panel display having the same	2012
Apparatus and method for cross axis parallel spectroscopy	2011
System and method for managing copyrighted electronic media	2008
Disc recording medium, recording method, disc drive device	2007
Gemba Kaizen. A commonsense, low-cost approach to management	2007
Two-layer shallow water system: a relaxation approach	2009
Method for sorting and displaying a multiple user database	2007
A stochastic programming duality approach to inventory centralization games	2009
Articular disc prosthesis and method for treating spondylolisthesis	2010
Experimental verification of the flexibility-based damage locating vector method	2007
System and method for a merchant loyalty system	2012
Method and system for producing repeating spatial forces	2010
Transmit-only and receive-only Bluetooth apparatus and method	2007
System for and method of generating image annotation information	2007
Control system and method for a concrete vehicle	2010
A scaling mitigation approach during direct contact membrane distillation	2011
National innovation systems: the emergence of a new approach	2011
Expandable fusion device and method of installation thereof	2014
Total knee prosthesis and method for total knee arthroplasty	2011
WLAN communication system and method with mobile base station	2009
System and method for generating a flickering flame effect	2007
System and method for mobile device push communications	2014
The treatment of desmoid tumors: a stepwise clinical approach	2012
HECTAR: a method to predict subcellular targeting in heterokonts	2008
Deep learning approach for active classification of electrocardiogram signals	2016
A metalearning approach to processing the scope of negation	2009
Smartcard transaction method and system using fingerprint recognition	2009
Digital broadcasting system and method of processing data	2010
School choice and increasing performance difference: A counterfactual approach	2013
A rural high school's collaborative approach to school improvement	2009
Pediatric nonalcoholic fatty liver disease: a multidisciplinary approach	2012
Family mealtimes: a contextual approach to understanding childhood obesity	2012
Can the cognitive load approach make instructional animations more effective?	2007
III-nitride current control device and method of manufacture	2012
Client, brokerage server and method for providing cloud storage	2014

Semiconductor device including active layer made of zinc oxide with controlled orientations and manufacturing method thereof	2011
Method and system for providing a voice mail message	2008
Thin film transistor including selectively crystallized channel layer and method of manufacturing the thin film transistor	2013
Method and apparatus for facilitating refinement of a search	2007
Method of determining voltage compensation for flash memory devices	2008
Fuzzy set and cache-based approach for bug triaging	2011
Method and apparatus for computer modified magnetic resonance imaging	2009
Robust finite-time control approach for robotic manipulators	2010
The suitability of the configuration approach in entrepreneurship research	2009
Boostmap: An embedding method for efficient nearest neighbor retrieval	2008
A fast and accurate computational approach to protein ionization	2008
Exploring consumer boycott intelligence using a socio-cognitive approach	2010
Approach for standardization of off-grid electrification projects	2009
Method and apparatus for intravascular imaging and occlusion crossing	2009
Method and apparatus for structured streaming of an XML document	2007
A semismooth Newton method for Tikhonov functionals with sparsity constraints	2008
Reflexivity, the picturing of selves, the forging of method	2007
Video encoding method, video decoding method, video encoding apparatus, video decoding apparatus, video encoding program, and video decoding program	2008
Turn-On Fluorescent Sensor for Hg2+ via Displacement Approach	2008
Device and method for measuring six degrees of freedom	2013
Enhancing the performance of differential evolution using orthogonal design method	2008
Delineation of ECG characteristic features using multiresolution wavelet analysis method	2012
Method and apparatus for creating a unique audio signature	2011
Method and apparatus for an augmented reality user interface	2013
Thermally contained/insulated phase change memory device and method	2011
On the homotopy analysis method for non-linear vibration of beams	2009
Conversation control apparatus, conversation control method, and programs therefor	2010
An entropy-based approach for measuring complexity in supply chains	2010
Convergent approach for commercial synthesis of gefitinib and erlotinib	2007
Method and device for manipulating color in a display	2012
Sediment cascades: an integrated approach	2010
A new approach to homogenize daily radiosonde humidity data	2011
Embedded fractional nonlinear supercapacitor model and its parametric estimation method	2010
Sudden cardiac death in the young: a clinical genetic approach	2007
A parametric finite element method for fourth order geometric evolution equations	2007
A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time'approach	2011
Light emitting device and method of driving the same	2015
Method, device, and system for multiplexing of video streams	2013
An efficient calibration method for continental?scale land surface modeling	2008
Method for manufacturing valve unit, and valve unit	2018
System and method of monitoring and controlling application files	2007
A multi-scale, multi-wavelength source extraction method: getsources	2012
System and method for location based matching and promotion	2008
Identifying decision strategies: A process-and outcome-based classification method	2008
Programmed method: developing a toolset for capturing and analyzing tweets	2014
Mobile networking system and method using IPv4 and IPv6	2010
S/kademlia: A practicable approach towards secure key-based routing	2007
Method for streaming XPath processing with forward and backward axes	2007
A simple method for directional transcriptome sequencing using Illumina technology	2009
Volume control system and method for a mobile communication device	2008
A population dynamics approach for the dispatch of distributed generators	2011
System and method for processing currency bills and tickets	2014
High-power-capable circularly polarized patch antenna apparatus and method	2013
Multireference state-specific Mukherjee's coupled cluster method with noniterative triexcitations	2008
Material design and structural color inspired by biomimetic approach	2012
System and method for positioning a mobile drive unit	2015
Decentralized power control scheme in femtocell networks: A game theoretic approach	2009
A method for designing strong S-Boxes based on chaotic Lorenz system	2010
System-in-package having integrated passive devices and method therefor	2010
A diagnostic approach for electro-mechanical actuators in aerospace systems	2009
Crack cocaine: A practical treatment approach for the chemically dependent	2014
Method and device for augmented reality message hiding and revealing	2015
Systematic coarse-graining of molecular models by the Newton inversion method	2010
Balloon catheter having metal balloon and method of making same	2013
Ceramic honeycomb structural body and method of manufacturing the same	2010
Patient matching surgical guide and method for using the same	2014
Antioxidant capacity of sulfated polysaccharides from seaweeds. A kinetic approach	2011
System and method for adaptively managing pages in a memory	2007

A note on He's homotopy perturbation method for van der Pol oscillator with very strong nonlinearity	2007
Personal name resolution in email: A heuristic approach	2008
Method and system for monitoring an activity of a user	2011
Screening for hemochromatosis by measuring ferritin levels: a more effective approach	2008
Grenade explosion method a novel tool for optimization of multimodal functions	2010
Multi-level semiconductor module and method for fabricating the same	2008
Synthesis of nanoparticles of CoxFe (3? x) O4 by combustion reaction method	2007
System and method for storing items and tracking item usage	2009
The amenity value of English nature: a hedonic price approach	2014
Dual drainage pathway shunt device and method for treating glaucoma	2007
System and method for money management in electronic trading environment	2010
System and method for providing asset management and tracking capabilities	2012
Costing of cloud computing services: A total cost of ownership approach	2012
Water diplomacy: A negotiated approach to managing complex water networks	2012
Information processing system, information processing device, and information processing method	2009
The 'prime-ome': towards a holistic approach to priming	2015
Q method and surveys: Three ways to combine Q and R	2009
Method and communication device for expanded coverage in a mesh network	2012
Apparatus and method to provide emergency access to bone marrow	2010
A novel method for measuring semantic similarity for XML schema matching	2008
Orally dissolving strips: A new approach to oral drug delivery system	2013
Lens generating a batwing-shaped beam distribution, and method therefor	2013
A multiscale stochastic finite element method on elliptic problems involving uncertainties	2007
Method, apparatus and system for virtualized peer-to-peer proxy services	2010
Self-luminous display device and driving method of the same	2011
Vertically stacked field programmable nonvolatile memory and method of fabrication	2007
System and method for recommending programming to television viewing communities	2011
The Anisakis allergy debate: does an evolutionary approach help?	2012
Semiconductor white light emitting device and method for manufacturing the same	2008
Method to form ultra high quality silicon-containing compound layers	2011
Method and apparatus for voice control of a television control device	2007
Method and apparatus for improving throughput in a wireless local area network	2008
Synthesis and electrochemical properties of olivine LiFePO4 prepared by a carbothermal reduction method	2008
Microbial reduction of graphene oxide by Escherichia coli: a green chemistry approach	2013
Vehicle navigation device and method of displaying POI information using same	2007
Empirical study of machine learning based approach for opinion mining in tweets	2012
A stochastic geometry approach to coexistence in heterogeneous wireless networks	2009
Method and system for distributing and updating software in wireless devices	2009
Decision making in natural resource management: a structured, adaptive approach	2013
Quantum dot as a spin-current diode: A master-equation approach	2007
A trust-semantic fusion-based recommendation approach for e-business applications	2012
Application of the intermediate derivatization approach in agrochemical discovery	2014
Surrogate science: The idol of a universal method for scientific inference	2015
Development of an in planta method for transformation of alfalfa (Medicago sativa)	2008
Mobile device and method for receiving and processing program-accompanying data	2008
Method for manufacturing thin film transistor using multi-tone mask	2012
Resistive memory cell random access memory device and method of fabrication	2007
Boethius and Dialogue: Literary Method in the Consolation of Philosophy	2014
Method and system for location management and location information providing system	2009
Lyapunov differential equation approach to elliptical orbital rendezvous with constrained controls	2011
Drug target prediction and repositioning using an integrated network-based approach	2013
Synthesis of ZnO nanoparticles using surfactant free in-air and microwave method	2011
A method to objectively assess swallow function in adults with suspected aspiration	2011
A framework for a personalized surgical approach to ovarian cancer	2015
Nitric oxide donor composition and method for treatment of anal disorders	2007
A Lagrangian Approach for the Incompressible Navier?Stokes Equations with Variable Density	2012
Towards a new supraregulatory approach to environmental assessment in Northern Canada	2007
Public health policy research: making the case for a political science approach	2011
A successive clutter-rejection-based approach for early detection of diabetic retinopathy	2011
System and method for providing a tile management controller	2018
Building communication theories: A socio/cultural approach	2013
System and method of controlling variables using a radial control menu	2011
System and method for linking multiple entities in a business database	2011
Method and apparatus for an integrated identity security and payment system	2007
Adaptive unknown input observer approach for aircraft actuator fault detection and isolation	2007
Introduction to the quantum Yang-Baxter equation and quantum groups: an algebraic approach	2013
Method and apparatus for blood glucose testing from a reversible infusion line	2007
Malignant pheochromocytomas and paragangliomas the importance of a multidisciplinary approach	2011
An evolutionary grey, hop, skip, and jump approach: generating alternative policies for the expansion of waste management	2015
Method and system for noninvasive face lifts and deep tissue tightening	2013
Two-step supercritical dimethyl carbonate method for biodiesel production from Jatropha curcas oil	2010

Apparatus and method of biometric determination using specialized optical spectroscopy systems	2011
Arctic ice?ocean simulation with optimized model parameters: Approach and assessment	2011
Super-resolution method for face recognition using nonlinear mappings on coherent features	2011
Synthesis and characterization of lithium manganese phosphate by a precipitation method	2010
Boundary conditions at fluid-permeable interfaces in porous media: A variational approach	2009
The anatomic approach to primary, revision and augmentation anterior cruciate ligament reconstruction	2010
System and method for interfacing with heterogeneous network data gathering tools	2012
System and method for determining an abused sensor during analyte measurement	2011
Intelligent apparatus, system and method for financial data computation and analysis	2007
Method and system for providing supervisory control over wireless phone usage	2007
Power save mode-based operating method and device in WLAN	2018
Method of using expandable vein ligator catheter having multiple electrode leads	2008
Model predictive control approach for guidance of spacecraft rendezvous and proximity maneuvering	2012
Corporate foresight and innovation management: A portfolio-approach in evaluating organizational development	2010
Synthesis of magnetite nanoparticles by surfactant-free electrochemical method in an aqueous system	2012
Inserting apparatus and method with controlled, master cycle speed-dependent actuator operations	2008
Method and apparatus for conducting electronic commerce transactions using electronic tokens	2007
System and method for producing variable information documents using undetermined data sources	2010
Professional development of teachers for computer-supported collaborative learning: A knowledge-building approach	2009
Analysis of 3D solids using the natural neighbour radial point interpolation method	2007
Modeling graphene in the finite-difference time-domain method using a surface boundary condition	2013
Digital sound synthesis by physical modeling using the functional transformation method	2012
A group decision support approach to evaluate experts for R&D project selection	2008
Computer-processor based interface for telepresence system, method and computer program product	2012
Spectral difference method for compressible flow on unstructured grids with mixed elements	2009
Indian bank efficiency and productivity changes with undesirable outputs: A disaggregated approach	2014
System and method for extension of group buying throughout the internet	2009
Solving multi-level multi-objective linear programming problems through fuzzy goal programming approach	2010
Contaminated rivers: a geomorphological-geochemical approach to site assessment and remediation	2007
Method and device for organizing user provided information with meta-information	2010
Trust, power and transaction costs in B2B exchanges A socio-economic approach	2008
A novel method of evaluating performance characteristics of a self-excited induction generator	2009
Evaluation and analysis method for natural gas hydrate storage and transportation processes	2008
Assessment of human deoxynivalenol exposure using an LC MS/MS based biomarker method	2012
A new approach of pellet formation of a filamentous fungus Rhizopus oryzae	2007
Phase adjustment apparatus and method for a memory device signaling system	2010
Fatigue management system and method for hybrid nonvolatile solid state memory system	2015
Identification of cell types from single-cell transcriptomes using a novel clustering method	2015
Integration of augmented reality and GIS: A new approach to realistic landscape visualisation	2008
Sexual behavior that is out of control: A theoretical conceptual approach	2008
Evaluation of a multiplex PCR assay as an alternative method for Listeria monocytogenes serotyping	2010
Machine and method for redeeming currency to dispense a value card	2011
Snapshot of the eukaryotic gene expression in muskoxen rumen a metatranscriptomic approach	2011
System and method for providing service or adding benefit to social networks	2010
Moving applications to the cloud: an approach based on application model enrichment	2011
Frost retardation of an air-source heat pump by the hot gas bypass method	2008
A Parameter-Dependent Approach to Robust Filtering for Time-Delay Systems	2008
Treatment system for surgery and control method of treatment system for surgery	2012
Posteromedial supine approach for reduction and fixation of medial and bicondylar tibial plateau fractures	2008
Co-innovation of family farm systems: A systems approach to sustainable agriculture	2014
Method and apparatus for providing a web-based active virtual file system	2008
Method and apparatus for selectively applying interference cancellation in spread spectrum systems	2010
Method and system for powering an electronic device via a wireless link	2011
A PLS-based statistical approach for fault detection and isolation of robotic manipulators	2012
Method for terminal-assisted interference control in a multi-carrier mobile communication system	2010
Rank-geofm: A ranking based geographical factorization method for point of interest recommendation	2015
Method and device for improving spatial and off-axis display standard conformance	2012
Job display control method	2009
Review of the establishment of nitro group charge method and its applications	2009
Method, systems, and computer program products for implementing function-parallel network firewall	2011
Appropriate technology A comprehensive approach for water and sanitation in the developing world	2009
A preemptive goal programming method for aggregating OWA operator weights in group decision making	2007
Method and apparatus for impairment correlation estimation in a wireless communication receiver	2011
Is entrepreneurial success predictable? An ex?ante analysis of the character?based approach	2008
Generalized differential quadrature finite element method for cracked composite structures of arbitrary shape	2013
A generalized single linkage method for estimating the cluster tree of a density	2010
The competitiveness of Korea as a developer of hydrogen energy technology: the AHP approach	2008
Network traversal method for establishing connection between two endpoints and network communication system	2011
An optimized Schwarz method with two?sided Robin transmission conditions for the Helmholtz equation	2007

How are parental psychological control and autonomy?support related? A cluster?analytic approach	2009
Device, method, and graphical user interface for manipulating selectable user interface objects	2014
A robust directional saliency-based method for infrared small-target detection under various complex backgrounds	2013
Method and system for power factor correction using constant pulse proportional current	2008
Method, apparatus and computer program product for creating a wireless docking group	2013
System and method for filtering and organizing items based on common elements	2009
Method and system for high performance data metatagging and data indexing using coprocessors	2012
System and method for issuer originated payments for on-line banking bill payments	2011
Assessing spatial attributes of forest landscape values: an internet-based participatory mapping approach	2008
Super?refractory status epilepticus: an approach to therapy in this difficult clinical situation	2011
A simple filter-based approach to surface enhanced Raman spectroscopy for trace chemical detection	2012
Method, apparatus and article for computational sequence generation and playing card distribution	2009
Repeater, excitation light supply device used for the same, and excitation light supply method	2016
Apparatus and method for providing and obtaining product information through a broadcast signal	2008
Household energy consumption versus income and relative standard of living: a panel approach	2007
Apparatus and method for transmitting a reference signal in a radio communication system	2013
Developer feeding member, developing apparatus, process cartridge and developer feeding member mounting method	2007
The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming	2008
A connectionist approach to the organization and continuity of working models of attachment	2007
A knowledge elicitation approach to the measurement of team situation awareness	2017
Structure and magnetic properties in Mn doped SnO2 nanoparticles synthesized by chemical co-precipitation method	2008
Systems pathology approach for the prediction of prostate cancer progression after radical prostatectomy	2008
Universal approach to predicting heat transfer coefficient for condensing mini/micro-channel flow	2013
Economic valuation of environmental benefits from wastewater treatment processes: An empirical approach for Spain	2010
Method and system for calibrating a multi-mode, multi-standard transmitter and receiver	2015
Alport syndrome and thin glomerular basement membrane nephropathy: a practical approach to diagnosis	2009
Maskless fabrication of concave microlens arrays on silica glasses by a femtosecond-laser-enhanced local wet etching method	2010
Distributed State Estimation and Energy Management in Smart Grids: A ConsensusInnovations Approach	2014
System and method for patient positioning for radiotherapy in the presence of respiratory motion	2010
Image-blur correction device and corresponding method for controlling compensation within an octogon region	2012
An integrated approach to the prediction of chemotherapeutic response in patients with breast cancer	2008
Identification and modeling of process damping in turning and milling using a new approach	2010
Indoor navigation system and method	2011
Method for improving web performance by adapting servers based on client cluster characterization	2007
System and method for determining optimal atrioventricular delay based on intrinsic conduction delays	2007
Ten-year recurrence rates in young women with breast cancer by locoregional treatment approach	2009
Past, present and future research on multiple identities: Toward an intrapersonal network approach	2014
Method for predicting airflow rates	2009
Equipment and method for identifying, monitoring and evaluating equipment, environmental and physiological conditions	2011
The role of epistasis in the manifestation of heterosis: a systems-oriented approach	2007
Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in Arabidopsis thaliana	2007
Acoustical panel comprising interlocking matrix of set gypsum and method for making same	2008
Mutational analysis in cytological specimens of advanced lung adenocarcinoma: a sensitive method for molecular diagnosis	2007
Method and apparatus for implementing multiple push buttons in a user input device	2008
A new method for dealing with residual spatial autocorrelation in species distribution models	2012
System and method for speech enhancement	2007
System and method for directory services and e-commerce across multi-provider networks	2008
An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer	2014
The variational inference approach to joint data detection and phase noise estimation in OFDM	2007
Minimally invasive medical implant and insertion device and method for using the same	2007
Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs	2013
System and method of maximizing integrated circuit manufacturing yield with context-dependent yield cells	2010
Application of He's variational iteration method to nonlinear Jaulent Miodek equations and comparing it with ADM	2007
A fast inverse consistent deformable image registration method based on symmetric optical flow computation	2008
Method, system, and program for specifying multidimensional calculations for a relational OLAP engine	2011
Exposure apparatus and device manufacturing method	2007
Efficient and reliable broadcast in intervehicle communication networks: A cross-layer approach	2010
Image display apparatus and control method	2012
The hp?d?adaptive finite cell method for geometrically nonlinear problems of solid mechanics	2012
Feasibility and safety of a new robotic thyroidectomy through a gasless, transaxillary single-incision approach	2010
Versatile Method for the Synthesis of 4?Aminocyclopentenones: Dysprosium (III) Triflate Catalyzed Aza?Piancatelli Rearrangement	2010
Self-selective electroless plating: An approach for fabrication of functional 1D nanomaterials	2008

Adaptive pattern recognition based controller apparatus and method and human-factored interface therefore	2010
Fit-for-purpose biomarker method validation for application in clinical trials of anticancer drugs	2010
System and method for radio communication between an implantable medical device and an external base unit	2010
Method and system for managing performance of data transfers for a data access system	2007
The critical pathway for deceased donation: reportable uniformity in the approach to deceased donation	2011
Highly diastereo-and enantioselective aldol reaction of methyl ?-isocyanoacetate: A cooperative catalysis approach	2011
Multiconfiguration second?order perturbation theory approach to strong electron correlation in chemistry and photochemistry	2012
Antioxidant and anti-inflammatory properties of Capsicum baccatum: from traditional use to scientific approach	2012
Development of an advanced A-TIG (AA-TIG) welding method by control of Marangoni convection	2008
Welding method and welding system with determination of the position of the welding torch	2014
Spine distraction implant and method	2014
Bridging the gaps between theory and practice: a service niche approach to urban sustainability indicators	2008
A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data	2013
Inferring earthquake physics and chemistry using an integrated field and laboratory approach	2012
Method for conducting financial transactions	2013
Segmentation based classification of 3D urban point clouds: A super-voxel based approach with evaluation	2013
Systematic approach to pediatric ocular trauma	2007
Method of detection of signals using an adaptive antenna in a peer-to-peer network	2007
Understanding texture changes of high pressure processed fresh carrots: A microstructural and biochemical approach	2007
MnOx/TiO2 composite nanoxides synthesized by deposition-precipitation method as a superior catalyst for NO oxidation	2010
A socio-ecological approach to physical activity interventions in childcare: a systematic review	2014
Nine-hospital study comparing broth microdilution and Etest method results for vancomycin and daptomycin against methicillin-resistant Staphylococcus aureus	2009
An enactive approach to understanding acting	2007
Pricing strategies: a marketing approach	2011
Method and apparatus for interacting with a visually displayed document on a screen reader	2007
Demonstration of intrinsic efflux activity of Escherichia coli K-12 AG100 by an automated ethidium bromide method	2008
Mobile station and communication control method	2010
A simple method for estimating interactions between a treatment and a large number of covariates	2014
Indirect approach to invariant point determination for SLR and VLBI systems: an assessment	2007
Comparison of total mercury and methylmercury cycling at five sites using the small watershed approach	2008
Understanding the lived experiences of patients: application of a phenomenological approach to ethics	2010
Algebraic approach to pseudospin symmetry for the Dirac equation with scalar and vector modified Pöschl-Teller potentials	2009
Method for automatic georeferencing aerial remote sensing (RS) images from an unmanned aerial vehicle (UAV) platform	2011
System and method for using image data in connection with configuring a universal controlling device	2010
A new catalytic method for the preparation of bis-indolyl and tris-indolyl methanes in aqueous media	2007
Experimenting for sustainability transitions: A systematic literature review	2016
Preparation and photocatalytic activity of Cu-doped ZnO thin films prepared by the sol gel method	2012
System and method for analysis of respiratory cycle-related EEG changes in sleep-disordered breathing	2007
Method and program for performing baseline correction of amplification curves in a PCR experiment	2013
Method of improving orientation and color balance of digital images using face detection information	2009
Method and system for searching an information retrieval system according to user-specified location information	2008
A systematic approach to the development of research-based web design guidelines for older people	2007
Towards an integrated approach to Design for X : an agenda for decision-based DFX research	2010
Portable electronic device, method, and graphical user interface for displaying electronic lists and documents	2012
Synthesis and photocatalytic activity of WO3 nanoparticles prepared by the arc discharge method in deionized water	2008
Determination of total dietary fiber (CODEX definition) by enzymatic-gravimetric method and liquid chromatography: collaborative study	2010
Safety of contraceptive method use among women with systemic lupus erythematosus: a systematic review	2009
Method of manufacturing a disk drive	2012
Laser speckle contrast analysis: a new method to evaluate peripheral blood perfusion in systemic sclerosis patients	2014
Interactive, topic-based visual text summarization and analysis	2009
Multi-thread bone screw and method	2011
Energy and environmental systems planning under uncertainty an inexact fuzzy-stochastic programming approach	2010
Method for fabricating group-III nitride devices and devices fabricated using method	2008
Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs	2011
Algorithmic approach to pushback design based on stochastic programming: method, application and comparisons	2010

System for distributing and selecting audio and video information and method implemented by said system	2014
An immersed boundary lattice-Boltzmann method for the simulation of the flow past an impulsively started cylinder	2008
Sensitive drug distribution system and method	2010
Method and system for performing peer-to-peer communication between stations within a basic service set	2011
Occipito-cervical stabilization system and method	2011
Method and system for location-aware authorization	2010
Method, radio system, and base station	2013
System and method for the manufacture of semiconductor devices by the implantation of carbon clusters	2012
A method for selective enrichment and analysis of nitrotyrosine-containing peptides in complex proteome samples	2007
Ten years of a biomimetic approach to the copper (II) radical site of galactose oxidase	2007
Catalytic asymmetric dearomatization by transition-metal catalysis: a method for transformations of aromatic compounds	2016
Soy-castor oil based polyols prepared using a solvent-free and catalyst-free method and polyurethanes therefrom	2013
System and method for deployable templates	2014
Structure and method for improved stress and yield in pFETs with embedded SiGe source/drain regions	2008
Two-pion exchange electromagnetic current in chiral effective field theory using the method of unitary transformation	2009
A new approach to prepare Ti3+ self-doped TiO2 via NaBH4 reduction and hydrochloric acid treatment	2014
Method and system for supporting the communication of presence information among computing devices of a network	2009
A lifetime versus a graft life approach redefines the importance of HLA matching in kidney transplant patients	2009
Method and system for detecting an end of transaction for contactless transactions on a mobile device	2008
Ultrasonographic tissue characterisation of human Achilles tendons: quantification of tendon structure through a novel non-invasive approach	2010
Gaming system and method of operating a gaming system having a bonus participation bidding sequence	2010
System and method for organizing data	2011
System and method for virtual packet reassembly	2007
Performance characteristics and clinical utility of an enzymatic method for the measurement of glycated albumin in plasma	2007
System and method for determining reentrant ventricular tachycardia isthmus location and shape for catheter ablation	2007
Dual-axis electron tomography: a new approach for investigating the spatial organization of wood cellulose microfibrils	2007
Method of measuring biological information using light and apparatus of measuring biological information using light	2007
A co-clinical approach identifies mechanisms and potential therapies for androgen deprivation resistance in prostate cancer	2013
Multiple molecular architectures of the eye lens chaperone ?B-crystallin elucidated by a triple hybrid approach	2011
Method and apparatus for delivering epinephrine	2007
Apparatus and method for electrocrushing rock	2012
A simple method for designation of urban ventilation corridors and its application to urban heat island analysis	2010
Display device and driving method thereof	2014
Method of treating contrast-induced nephropathy	2015
Method and system for analyzing and predicting the performance of computer network using time series measurements	2007
Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index	2009
Comparing non?hierarchical governance in action: the Open Method of Co?ordination in pensions and information society	2007
Highly Active and Stable Co3O4/ZSM-5 Catalyst for Propane Oxidation: Effect of the Preparation Method	2013
Fullerenols as a new therapeutic approach in nanomedicine	2013
Surgical management of ameloblastoma: Conservative or radical approach	2011
Forecasting method selection in a global supply chain	2012
Method and apparatus for detecting respiratory disturbances	2009
Method of operating a vehicle safety system	2013
Method of forming memory cell access device	2013
Highly efficient visible light TiO2 photocatalyst prepared by sol gel method at temperatures lower than 300C	2011
The solution of the variable coefficients fourth-order parabolic partial differential equations by the homotopy perturbation method	2009
Exploring worldviews and their relationships to sustainable lifestyles: Towards a new conceptual and methodological approach	2012
Piecewise linear regression: A statistical method for the analysis of experimental adsorption data by the intraparticle-diffusion models	2010
Non-destructive stringer inspection apparatus and method	2007
Gaming device and method having free activation mode and free activation mode with free activation retrigger	2010
Control synthesis of rutile TiO 2 microspheres, nanoflowers, nanotrees and nanobelts via acid-hydrothermal method and their optical properties	2011
Method and apparatus for tracking three-dimensional movements of an object using a depth sensing camera	2013

Further studies on control synthesis of discrete-time TS fuzzy systems via augmented multi-indexed matrix approach.	2014
Iron-Mediated Direct Suzuki?Miyaura Reaction: A New Method for the ortho-Arylation of Pyrrole and Pyridine	2010
An improved HPLC ICPMS method for determining inorganic arsenic in food: application to rice, wheat and tuna fish	2012
A statistical approach to modelling permafrost distribution in the European Alps or similar mountain ranges	2012
System and method for detecting account compromises	2013
Method and device for a touchless interface	2011
On?line identification and quantification of mean shifts in bivariate processes using a neural network?based approach	2007
Electric appliance and a control method thereof	2013
Time-domain reflectometry method and its application for measuring moisture content in porous materials: A review	2009
MIM capacitor structure and method of manufacture	2007
End user oriented workflow approach including structured processing of ad hoc workflows with a collaborative process engine	2011
Method for improving rate-distortion performance of a video compression system through parallel coefficient cancellation in the transform	2008
Condition-based fault tree analysis (CBFTA): A new method for improved fault tree analysis (FTA), reliability and safety calculations	2007
Severe ovarian hyperstimulation syndrome after gonadotropin-releasing hormone (GnRH) agonist trigger and freeze-all approach in GnRH antagonist protocol	2014
System and method for synchronized media distribution	2012
Method and circuitry to generate a reference current for reading a memory cell, and device implementing same	2008
Method, system, and computer readable medium for reading and programming flash memory cells using multiple bias voltages	2014
The prevalence and impact of overuse injuries in five N orwegian sports: Application of a new surveillance method	2015
Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory	2010
Network with mobile terminals as browsers having wireless access to the internet and method for using same	2009
Tissue engineering for the lower urinary tract: a review of a state of the art approach	2007
Method and composition for polishing a substrate	2007
Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction	2010
Method in a Communication Network for Distributing Vehicle Driving Information and System Implementing the Method	2008
Method for making a semiconductor device having a high-k gate dielectric layer and a metal gate electrode	2007
Method for printing a visual printer calibration test pattern	2009
NURBS?enhanced finite element method for Euler equations	2008
Dissolvable barrier for downhole use and method thereof	2014
Connectivity is a two-way street the need for a holistic approach to fish passage problems in regulated rivers	2009
Method and device for filtering body fluid	2008
Modelling runoff and erosion for a semi-arid catchment using a multi-scale approach based on hydrological connectivity	2009
System and method for enterprise-wide dashboard reporting	2008
Nucleation of electroactive ?-phase poly (vinilidene fluoride) with CoFe2O4 and NiFe2O4 nanofillers: a new method for the preparation of multiferroic nanocomposites	2011
Message distribution system, server, mobile terminal, data storage unit, message distribution method, and message distribution computer program product	2007
(Meth) acrylate derivative, polymer and photoresist composition having lactone structure, and method for forming pattern by using it	2007
Synchronized X-ray/breathing method and apparatus used in conjunction with a charged particle cancer therapy system	2011
Telecommunication and multimedia management method and apparatus	2012
System and method for evaluating heart failure based on ventricular end-diastolic volume using an implantable medical device	2009
Method, system, and computer readable medium for facilitating a transaction between a customer, a merchant and an associate	2010
Body height estimation from head and face dimensions: a different method	2010
Method and apparatus for processing electronic dispute data	2011
System and method for communicating data using efficient fast fourier transform (FFT) for orthogonal frequency division multiplexing (OFDM)	2012
A method for guidance and control of an autonomous vehicle in problems of border patrolling and obstacle avoidance	2011
Intrusion detection system and network flow director method	2009
A family-centered approach to planning and measuring the outcome of interventions for children with attention-deficit/hyperactivity disorder	2007
Method and apparatus for relay zone bandwidth allocation	2011
Low-cost disposable tourniquet cuff apparatus and method	2012

Reconstruction of a human cornea by the self-assembly approach of tissue engineering using the three native cell types	2010
transmission apparatus, data receiving apparatus, data communication method, data transmission method, received-data processing method, and computer program	2009
Application of data mining to network intrusion detection: classifier selection model	2008
Method and apparatus for providing voice and data services in a mobile communication system with various overlapped access networks	2011
System and method providing multi-tier applications architecture	2008
Validation of an HPLC-UV method for sorafenib determination in human plasma and application to cancer patients in routine clinical practice	2009
An integrative perceptual approach for teaching Chinese characters	2007
Low-force electrochemical mechanical processing method and apparatus	2007
Educational data mining	2012
Unannounced standardised patients in real practice: a systematic literature review	2007
System and method for guiding entity-based searching	2013
A facile hydrazine-assisted hydrothermal method for the deposition of monodisperse SnO 2 nanoparticles onto graphene for lithium ion batteries	2012
Apparatus and method for providing a marketing service	2012
Functional tricuspid regurgitation at the time of mitral valve repair for degenerative leaflet prolapse: the case for a selective approach	2011
Three dimensional NAND device and method of charge trap layer separation and floating gate formation in the NAND device	2014
Video segment management and distribution system and method	2011
System and method for transponder-enabled account transactions	2010
Role of zinc interstitials and oxygen vacancies of ZnO in photocatalysis: a bottom-up approach to control defect density	2014
Refactoring to rich internet applications. a model-driven approach	2008
Perioperative results of 214 laparoscopic adrenalectomies by anterior transperitoneal approach	2008
A data-driven approach to a priori SNR estimation	2011
Method and System for Accessing, Diagnosing and Treating Target Tissue Regions Within the Middle Ear and the Eustachian Tube	2009
A domain decomposition approach to implementing fault slip in finite?element models of quasi?static and dynamic crustal deformation	2013
gene testing combined with single nucleotide polymorphism microarray preimplantation genetic diagnosis for aneuploidy: a novel approach in optimizing pregnancy	2011
A cluster centers initialization method for clustering categorical data	2012
Characterization of the zinc-containing metalloprotease encoded by zpx and development of a species-specific detection method for Enterobacter sakazakii	2007
Method and system for determining a location of a wireless transmitting device and guiding the search for the same	2010
On cloud storage and the cloud of clouds approach	2012
Display device and method for manufacturing the same	2009
An empirical study of the applications of data mining techniques in higher education	2011
Evaluation of knowledge management performance: An organic approach	2015
Method and system for VoIP over WLAN to Bluetooth headset using ACL link and sniff for aligned eSCO transmission	2008
Identifying security bug reports via text mining: An industrial case study	2010
Estimating non-market environmental benefits of the Conversion of Cropland to Forest and Grassland Program: A choice modeling approach	2007
Matching sensors to missions using a knowledge-based approach	2008
Method and apparatus for performing retro peritoneal dissection	2014
Method of reflecting on another device a change to a browser cache on a handheld electronic device, and assocaited device	2011
Method for estimating the range of a motor vehicle	2013
A design method for 3D origami based on rotational sweep	2009
Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors	2016
Method and apparatus for treatment of intracranial hemorrhages	2015
Diet modification to reduce phosphorus surpluses: A mass balance approach	2007
Call forwarding control device and method of call management	2009
System and method for credential delegation using identity assertion	2009
A group theoretic approach to construct cryptographically strong substitution boxes	2013
Water heater and method of controlling the same	2008
A modified failure mode and effects analysis method for supplier selection problems in the supply chain risk environment: A case study	2013
A qualitative method proposal to improve environmental impact assessment	2013
Potential energy surfaces from high fidelity fitting of ab initio points: the permutation invariant polynomial - neural network approach	2016
A comprehensive approach to evaluate nutritional status in Crohn's patients in the era of biologic therapy: a case-control study	2007
An evolving method for solar-grade silicon production: solvent refining	2012

Physical Vapor Deposition of [EMIM][Tf2N]: A New Approach to the Modification of Surface Properties with Ultrathin Ionic Liquid Films	2008
A topology optimisation for three-dimensional acoustics with the level set method and the fast multipole boundary element method	2014
Handover control method, mobile station and communication control apparatus	2007
Rapid preparation of silver nanorods and nanowires by a microwave-polyol method in the presence of Pt catalyst and polyvinylpyrrolidone	2007
Data read method of magnetic random access memory	2007
An integrated approach to delivering exposure-based treatment for symptoms of PTSD and depression in OIF/OEF veterans: Preliminary findings	2012
Apparatus and method for a server deterministically killing a redundant server integrated within the same network storage appliance chassis	2008
Method for providing reliable session communication within a network	2007
System and method for strategic power supply sequencing in a computer system with multiple processing resources and multiple power supplies	2008
DNA Photography: An Ultrasensitive DNA?Detection Method Based on Photographic Techniques	2007
Quantification of F2-isoprostanes as a reliable index of oxidative stress in vivo using gas chromatography mass spectrometry (GC-MS) method	2009
A variational approach to the numerical simulation of hydraulic fracturing	2012
System and method for clinical strategy for therapeutic pharmacies	2010
Systemic approach for the maintenance of complex structural systems	2008
Method and apparatus for placing sensors using 3D models	2007
Active matrix type display apparatus and driving method thereof	2013
Infectious diseases of animals and plants: an interdisciplinary approach	2011
A multi-phase covering Pareto-optimal front method to multi-objective scheduling in a realistic hybrid flowshop using a hybrid metaheuristic	2009
Systems and method for simple scale-out storage clusters	2013
Polymorphic phase transition among the titania crystal structures using a solution-based approach: from precursor chemistry to nucleation process	2014
Flow valve and method	2009
System and method to control a rotary-linear actuator	2007
Whole Versus the Piecemeal Approach to Evaluating Soy,2	2010
Method and apparatus for correlating events in a network	2009
Inkjet recording method	2010
Synthesis of samarium-and nitrogen-co-doped TiO2 by modified hydrothermal method and its photocatalytic performance for the degradation of 4-chlorophenol	2009
Synthesis of delaminated LDH: A facile two step approach	2010
The texture profile method	2017
Endoscope Assembly and Method of Performing a Medical Procedure	2009
A micromechanical approach to investigate asphalt concrete rutting mechanisms	2012
Smart antenna beamforming device in communication system and method thereof	2009
Method and device for sound detection and audio control	2012
Computer-implemented method and system for combining keywords into logical clusters that share similar behavior with respect to a considered dimension	2010
Playing it safe: Assessing cumulative impact and social vulnerability through an environmental justice screening method in the south coast air basin, California	2011
An information gain-based approach for recommending useful product reviews	2011
Determinants of equity share prices in India: A panel data approach	2012
Antireflection film, polarizing plate, method for producing them, liquid crystal display element, liquid crystal display device, and image display device	2011
Application of the Phosphoramidate ProTide Approach to 4 '-Azidouridine Confers Sub-micromolar Potency versus Hepatitis C Virus on an Inactive Nucleoside	2007
Vehicle-drive control system and method and program therefor	2007
System and method of remote surveillance and applications therefor	2010
A multi-objective approach to supply chain visibility and risk	2014
Method for unlocking a locked computing device and computing device thereof	2012
System and method of proxy authentication in a secured network	2010
Method and system for a three dimensional facial recognition system	2010
Making sense in participation: An enactive approach to social cognition	2008
A new RNS based DA approach for inner product computation	2013
System and method for managing virtual worlds mapped to real locations in a mobile-enabled massively multiplayer online role playing game (MMORPG)	2017
Calibration of the empirical likelihood method for a vector mean	2009
Hybrid immune-genetic algorithm method for benefit maximisation of distribution network operators and distributed generation owners in a deregulated environment	2011
A Study of Computing Undergraduates Undertaking a Systematic Literature Review	2011
Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review	2018
Spoken Language Recognition With Prosodic Features	2013
A Low-Cost Real-Time Research Platform for EMG Pattern Recognition-Based Prosthetic Hand	2015
A Sign-Component-Based Framework for Chinese Sign Language Recognition Using Accelerometer and sEMG Data	2012
A Novel Chinese Sign Language Recognition Method Based on Keyframe-Centered Clips	2018

Software project management approaches for global software development: a systematic mapping study	2018
SLR: Semi-Coupled Locality Constrained Representation for Very Low Resolution Face Recognition and Super Resolution	2018
An empirical investigation of the software integration success factors in GSD environment	2017
Systematic Mapping of Process Mining Studies in Healthcare	2018
Success Factors for Software Outsourcing Partnership Management: An Exploratory Study Using Systematic Literature Review	2017
Iterated Posterior Linearization Smoother	2017
Study of Senone-Based Deep Neural Network Approaches for Spoken Language Recognition	2016
Decision Support System for Risk Assessment and Management Strategies in Distributed Software Development	2017
Quantitative Determination of Cyfluthrin in N-Hexane by Terahertz Time-Domain Spectroscopy With Chemometrics Methods	2010
Posterior Linearization Filter: Principles and Implementation Using Sigma Points	2015
Antenna Placement Optimization for Distributed Antenna Systems	2012
Optimizing the Performance of Spoken Language Recognition With Discriminative Training	2008
High accuracy gene signature for chemosensitivity prediction in breast cancer	2015
Review of Statistical and Analytical Degradation Models for Photovoltaic Modules and Systems as Well as Related Improvements	2018
Euler Elastica Regularized Logistic Regression for Whole-Brain Decoding of fMRI Data	2018
Cooperative Localization Using Posterior Linearization Belief Propagation	2018
Grid integration of wind generation considering remote wind farms: hybrid markovian and interval unit commitment	2017
Genetic Improvement of Software: A Comprehensive Survey	2018
A Survey of App Store Analysis for Software Engineering	2017
Frequency-Hopping Signal Parameters Estimation Based on Orthogonal Matching Pursuit and Sparse Linear Regression	2018
Dynamic MR Image Reconstruction Separation From Undersampled	2014
A Brain-Computer Interface Based on a Few-Channel EEG-fNIRS Bimodal System	2017
A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction	2018
A fast image matching technique for the panoramic-based localization	2016
A Size Estimation Model for Board-Based Desktop Games	2017
Semi-supervised Domain Adaptation on Manifolds	2014
On the Correlation of Sensor Location and Human Activity Recognition in Body Area Networks (BANs)	2018
Optimal Transport for Domain Adaptation	2017
Sticky Projections-A Model-Based Approach to Interactive Shader Lamps Tracking	2016
Regression Methods for Virtual Metrology of Layer Thickness in Chemical Vapor Deposition	2014
A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors	2016
Predicting Vertical Acceleration of Railway Wagons Using Regression Algorithms	2010
Exact Feature Extraction Using Finite Rate of Innovation Principles With an Application to Image Super-Resolution	2009
Automatic Prediction of Children's Reading Ability for High-Level Literacy Assessment	2011
Airborne Vehicle Detection in Dense Urban Areas Using HoG Features and Disparity Maps	2013
Texture Classification and Visualization of Time Series of Gait Dynamics in Patients With Neuro-Degenerative Diseases	2018
Regenerator site selection for mixed line rate optical networks	2014
Fast Bayesian JPEG Decompression and Denoising With Tight Frame Priors	2017
Bridging the Gap: From Research to Practical Advice	2018
On the Privacy Risks of Virtual Keyboards: Automatic Reconstruction of Typed Input from Compromising Reflections	2013
Cross-Domain Recognition by Identifying Joint Subspaces of Source Domain and Target Domain	2017
Discriminative Structured Feature Engineering for Macroscale Brain Connectomes	2015
Robust LSTM-Autoencoders for Face De-Occlusion in the Wild	2018
Domain Invariant Transfer Kernel Learning	2015
Weakly Supervised Photo Cropping	2014
Spatial Logistic Regression for Support-Vector Classification of Hyperspectral Imagery	2017
Scatter Component Analysis: A Unified Framework for Domain Adaptation and Domain Generalization	2017
A Systematic Literature Review on Fault Prediction Performance in Software Engineering	2012
Motionlets Matching With Adaptive Kernels for 3-D Indian Sign Language Recognition	2018
On Dynamic Effects Influencing IGBT Losses in Soft-Switching Converters	2011
Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints	2012
Constrained Manifold Learning for Hyperspectral Imagery Visualization	2018
Linear Disentangled Representation Learning for Facial Actions	2018
Dictionary Learning-Based Feature-Level Domain Adaptation for Cross-Scene Hyperspectral Image Classification	2017
Classifying Discriminative Features for Blur Detection	2016
Magnetic Resonance RF Pulse Design by Optimal Control With Physical Constraints	2018
An Exemplar-Based Multi-View Domain Generalization Framework for Visual Recognition	2018
Empirical Evaluation of the Impact of Object-Oriented Code Refactoring on Quality Attributes: A Systematic Literature Review	2018

Brain-Wide Genome-Wide Association Study for Alzheimer's Disease via Joint Projection Learning and Sparse Regression Model	2019
Optimized relay placement to federate segments in wireless sensor networks	2010
Deep Learning of Transferable Representation for Scalable Domain Adaptation	2016
A Classification and Comparison Framework for Cloud Service Brokerage Architectures	2018
OInduced: An Efficient Algorithm for Mining Induced Patterns From Rooted Ordered Trees	2011
Integrated Low-Rank-Based Discriminative Feature Learning for Recognition	2016
Robust Single-Shot Fringe Projection Profilometry Based on Morphological Component Analysis	2018
Probabilistic Real-Time Thermal Rating Forecasting for Overhead Lines by Conditionally Heteroscedastic Auto-Regressive Models	2017
Stacked Sparse Autoencoder Modeling Using the Synergy of Airborne LiDAR and Satellite Optical and SAR Data to Map Forest Above-Ground Biomass	2017
Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition	2016
Robust Visual Knowledge Transfer via Extreme Learning Machine-Based Domain Adaptation	2016
Multilevel and Session Variability Compensated Language Recognition: ATVS-UAM Systems at NIST LRE 2009	2010
Generating Query Facets Using Knowledge Bases	2017
Clustering-Based Ensembles as an Alternative to Stacking	2014
Structured Domain Adaptation	2017
Gait Rhythm Fluctuation Analysis for Neurodegenerative Diseases by Empirical Mode Decomposition	2017
Dynamic Programming for Instance Annotation in Multi-Instance Multi-Label Learning	2017
An efficient error concealment algorithm for H.264/AVC using regression modeling-based prediction	2010
Improved Modeling of Cross-Decoder Phone Co-Occurrences in SVM-Based Phonotactic Language Recognition	2011
Deep Domain Generalization With Structured Low-Rank Constraint	2018
A Dynamic Discretization Approach for Constructing Decision Trees with a Continuous Label	2009
Learning the LMP-Load Coupling From Data: A Support Vector Machine Based Approach	2017
Domain Generalization and Adaptation Using Low Rank Exemplar SVMs	2018
No-Reference Quality Assessment of Tone-Mapped HDR Pictures	2017
A Knowledge-Based Approach to Arterial Stiffness Estimation Using the Digital Volume Pulse	2012
Blind Domain Adaptation With Augmented Extreme Learning Machine Features	2017
Hybrid CS-DMRI: Periodic Time-Variant Subsampling and Omnidirectional Total Variation Based Reconstruction	2017
Multi-Modal Sensing Techniques for Interfacing Hand Prostheses: A Review	2015
Damped Posterior Linearization Filter	2018
Dual Low-Rank Pursuit: Learning Salient Features for Saliency Detection	2016
EQAR: Effective QoS-Aware Relay Node Placement Algorithm for Connecting Disjoint Wireless Sensor Subnetworks	2011
Machine Learning Methods for Attack Detection in the Smart Grid	2016
Transmission Contingency-Constrained Unit Commitment With High Penetration of Renewables via Interval Optimization	2017
Impact From Dynamic Line Rating on Wind Power Integration	2015
Local Minimax Learning of Functions With Best Finite Sample Estimation Error Bounds: Applications to Ridge and Lasso Regression, Boosting, Tree Learning, Kernel Machines, and Inverse Problems	2009
Attributed Social Network Embedding	2018
Domain Invariant and Class Discriminative Feature Learning for Visual Domain Adaptation	2018
Resource Allocation in SWIPT Networks Under a Nonlinear Energy Harvesting Model: Power Efficiency, User Fairness, and Channel Nonreciprocity	2018
Sparse Spatio-Spectral LapSVM With Semisupervised Kernel Propagation for Hyperspectral Image Classification	2017
Face Spoof Detection With Image Distortion Analysis	2015
S3DRGF: Spatial 3-D Relational Geometric Features for 3-D Sign Language Representation and Recognition	2019
Score-Independent Audio Features for Description of Music Expression	2008
Training CNNs for 3-D Sign Language Recognition With Color Texture Coded Joint Angular Displacement Maps	2018
Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction	2019
Robust Fringe Projection Profilometry via Sparse Representation	2016
Time-Aware Multivariate Nearest Neighbor Regression Methods for Traffic Flow Prediction	2015
A New Feature Selection Scheme Using a Data Distribution Factor for Unsupervised Nominal Data	2008
Domain-Specific Image Analysis for Cervical Neoplasia Detection Based on Conditional Random Fields	2011
Online Handwritten Shape Recognition Using Segmental Hidden Markov Models	2007
Upstream Ultrasonic Level Based Soft Sensing of Volumetric Flow of Non-Newtonian Fluids in Open Venturi Channels	2018
A Generative Data Augmentation Model for Enhancing Chinese Dialect Pronunciation Prediction	2012
Adaptive Greedy Dictionary Selection for Web Media Summarization	2017
HPCMP CREATE-AV Quality Assurance: Lessons Learned by Validating and Supporting Computation-Based Engineering Software	2016
Robust Frequency-Hopping Spectrum Estimation Based on Sparse Bayesian Method	2015
Push-Pull: Deterministic Search-Based DAG Scheduling for Heterogeneous Cluster Systems	2007
Applications of Deep Learning and Reinforcement Learning to Biological Data	2018
A Sparse Plus Low-Rank Exponential Language Model for Limited Resource Scenarios	2015
Joint Denoising/Compression of Image Contours via Shape Prior and Context Tree	2018
Learning With Augmented Features for Supervised and Semi-Supervised Heterogeneous Domain Adaptation	2014

Inter-Swarm Content Distribution Among Private BitTorrent Networks	2013
Incomplete Multisource Transfer Learning	2018
Multi-Camera Saliency	2015
Spectrally Constrained Unimodular Sequence Design Without Spectral Level Mask	2018
System Monitoring with Metric-Correlation Models	2011
Salient Object Detection via Structured Matrix Decomposition	2017
On the Projection of PLLRs for Unbounded Feature Distributions in Spoken Language Recognition	2014
Spatio-Temporal Feature-Extraction Techniques for Isolated Gesture Recognition in Arabic Sign Language	2007
Domain Transfer Learning for MCI Conversion Prediction	2015
Decay Rate Estimators and Their Performance for Blind Reverberation Time Estimation	2014
A review of applications of search based software engineering techniques in last decade	2016
A Study of Computing Undergraduates Undertaking a Systematic Literature Review	2011
Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review	2018
Spoken Language Recognition With Prosodic Features	2013
A Low-Cost Real-Time Research Platform for EMG Pattern Recognition-Based Prosthetic Hand	2015
A Sign-Component-Based Framework for Chinese Sign Language Recognition Using Accelerometer and sEMG Data	2012
A Novel Chinese Sign Language Recognition Method Based on Keyframe-Centered Clips	2018
Software project management approaches for global software development: a systematic mapping study	2018
SLR: Semi-Coupled Locality Constrained Representation for Very Low Resolution Face Recognition and Super Resolution	2018
An empirical investigation of the software integration success factors in GSD environment	2017
Systematic Mapping of Process Mining Studies in Healthcare	2018
Success Factors for Software Outsourcing Partnership Management: An Exploratory Study Using Systematic Literature Review	2017
Iterated Posterior Linearization Smoother	2017
Study of Senone-Based Deep Neural Network Approaches for Spoken Language Recognition	2016
Decision Support System for Risk Assessment and Management Strategies in Distributed Software Development	2017
Quantitative Determination of Cyfluthrin in N-Hexane by Terahertz Time-Domain Spectroscopy With Chemometrics Methods	2010
Posterior Linearization Filter: Principles and Implementation Using Sigma Points	2015
Antenna Placement Optimization for Distributed Antenna Systems	2012
Optimizing the Performance of Spoken Language Recognition With Discriminative Training	2008
High accuracy gene signature for chemosensitivity prediction in breast cancer	2015
Review of Statistical and Analytical Degradation Models for Photovoltaic Modules and Systems as Well as Related Improvements	2018
Euler Elastica Regularized Logistic Regression for Whole-Brain Decoding of fMRI Data	2018
Cooperative Localization Using Posterior Linearization Belief Propagation	2018
Grid integration of wind generation considering remote wind farms: hybrid markovian and interval unit commitment	2017
Genetic Improvement of Software: A Comprehensive Survey	2018
A Survey of App Store Analysis for Software Engineering	2017
Frequency-Hopping Signal Parameters Estimation Based on Orthogonal Matching Pursuit and Sparse Linear Regression	2018
Dynamic MR Image Reconstruction Separation From Undersampled	2014
A Brain-Computer Interface Based on a Few-Channel EEG-fNIRS Bimodal System	2017
A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction	2018
A fast image matching technique for the panoramic-based localization	2016
A Size Estimation Model for Board-Based Desktop Games	2017
Semi-supervised Domain Adaptation on Manifolds	2014
On the Correlation of Sensor Location and Human Activity Recognition in Body Area Networks (BANs)	2018
Optimal Transport for Domain Adaptation	2017
Sticky Projections-A Model-Based Approach to Interactive Shader Lamps Tracking	2016
Regression Methods for Virtual Metrology of Layer Thickness in Chemical Vapor Deposition	2014
A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors	2016
Predicting Vertical Acceleration of Railway Wagons Using Regression Algorithms	2010
Exact Feature Extraction Using Finite Rate of Innovation Principles With an Application to Image Super-Resolution	2009
Automatic Prediction of Children's Reading Ability for High-Level Literacy Assessment	2011
Airborne Vehicle Detection in Dense Urban Areas Using HoG Features and Disparity Maps	2013
Texture Classification and Visualization of Time Series of Gait Dynamics in Patients With Neuro-Degenerative Diseases	2018
Regenerator site selection for mixed line rate optical networks	2014
Fast Bayesian JPEG Decompression and Denoising With Tight Frame Priors	2017
Bridging the Gap: From Research to Practical Advice	2018
On the Privacy Risks of Virtual Keyboards: Automatic Reconstruction of Typed Input from Compromising Reflections	2013
Cross-Domain Recognition by Identifying Joint Subspaces of Source Domain and Target Domain	2017
Discriminative Structured Feature Engineering for Macroscale Brain Connectomes	2015
Robust LSTM-Autoencoders for Face De-Occlusion in the Wild	2018

Domain Invariant Transfer Kernel Learning	2015
Weakly Supervised Photo Cropping	2014
Spatial Logistic Regression for Support-Vector Classification of Hyperspectral Imagery	2017
Scatter Component Analysis: A Unified Framework for Domain Adaptation and Domain Generalization	2017
A Systematic Literature Review on Fault Prediction Performance in Software Engineering	2012
Motionlets Matching With Adaptive Kernels for 3-D Indian Sign Language Recognition	2018
On Dynamic Effects Influencing IGBT Losses in Soft-Switching Converters	2011
Image reconstruction from highly undersampled space data with joint partial separability and sparsity constraints	2012
Constrained Manifold Learning for Hyperspectral Imagery Visualization	2018
Linear Disentangled Representation Learning for Facial Actions	2018
Dictionary Learning-Based Feature-Level Domain Adaptation for Cross-Scene Hyperspectral Image Classification	2017
Classifying Discriminative Features for Blur Detection	2016
Magnetic Resonance RF Pulse Design by Optimal Control With Physical Constraints	2018
An Exemplar-Based Multi-View Domain Generalization Framework for Visual Recognition	2018
Empirical Evaluation of the Impact of Object-Oriented Code Refactoring on Quality Attributes: A Systematic Literature Review	2018
Brain-Wide Genome-Wide Association Study for Alzheimer's Disease via Joint Projection Learning and Sparse Regression Model	2019
Optimized relay placement to federate segments in wireless sensor networks	2010
Deep Learning of Transferable Representation for Scalable Domain Adaptation	2016
A Classification and Comparison Framework for Cloud Service Brokerage Architectures	2018
OInduced: An Efficient Algorithm for Mining Induced Patterns From Rooted Ordered Trees	2011
Integrated Low-Rank-Based Discriminative Feature Learning for Recognition	2016
Robust Single-Shot Fringe Projection Profilometry Based on Morphological Component Analysis	2018
Probabilistic Real-Time Thermal Rating Forecasting for Overhead Lines by Conditionally Heteroscedastic Auto-Regressive Models	2017
Stacked Sparse Autoencoder Modeling Using the Synergy of Airborne LiDAR and Satellite Optical and SAR Data to Map Forest Above-Ground Biomass	2017
Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition	2016
Robust Visual Knowledge Transfer via Extreme Learning Machine-Based Domain Adaptation	2016
Multilevel and Session Variability Compensated Language Recognition: ATVS-UAM Systems at NIST LRE 2009	2010
Generating Query Facets Using Knowledge Bases	2017
Clustering-Based Ensembles as an Alternative to Stacking	2014
Structured Domain Adaptation	2017
Gait Rhythm Fluctuation Analysis for Neurodegenerative Diseases by Empirical Mode Decomposition	2017
Dynamic Programming for Instance Annotation in Multi-Instance Multi-Label Learning	2017
An efficient error concealment algorithm for H.264/AVC using regression modeling-based prediction	2010
Improved Modeling of Cross-Decoder Phone Co-Occurrences in SVM-Based Phonotactic Language Recognition	2011
Deep Domain Generalization With Structured Low-Rank Constraint	2018
A Dynamic Discretization Approach for Constructing Decision Trees with a Continuous Label	2009
Learning the LMP-Load Coupling From Data: A Support Vector Machine Based Approach	2017
Domain Generalization and Adaptation Using Low Rank Exemplar SVMs	2018
No-Reference Quality Assessment of Tone-Mapped HDR Pictures	2017
A Knowledge-Based Approach to Arterial Stiffness Estimation Using the Digital Volume Pulse	2012
Blind Domain Adaptation With Augmented Extreme Learning Machine Features	2017
Hybrid CS-DMRI: Periodic Time-Variant Subsampling and Omnidirectional Total Variation Based Reconstruction	2017
Multi-Modal Sensing Techniques for Interfacing Hand Prostheses: A Review	2015
Damped Posterior Linearization Filter	2018
Dual Low-Rank Pursuit: Learning Salient Features for Saliency Detection	2016
EQAR: Effective QoS-Aware Relay Node Placement Algorithm for Connecting Disjoint Wireless Sensor Subnetworks	2011
Machine Learning Methods for Attack Detection in the Smart Grid	2016
Transmission Contingency-Constrained Unit Commitment With High Penetration of Renewables via Interval Optimization	2017
Impact From Dynamic Line Rating on Wind Power Integration	2015
Local Minimax Learning of Functions With Best Finite Sample Estimation Error Bounds: Applications to Ridge and Lasso Regression, Boosting, Tree Learning, Kernel Machines, and Inverse Problems	2009
Attributed Social Network Embedding	2018
Domain Invariant and Class Discriminative Feature Learning for Visual Domain Adaptation	2018
Resource Allocation in SWIPT Networks Under a Nonlinear Energy Harvesting Model: Power Efficiency, User Fairness, and Channel Nonreciprocity	2018
Sparse Spatio-Spectral LapSVM With Semisupervised Kernel Propagation for Hyperspectral Image Classification	2017
Face Spoof Detection With Image Distortion Analysis	2015
S3DRGF: Spatial 3-D Relational Geometric Features for 3-D Sign Language Representation and Recognition	2019
Score-Independent Audio Features for Description of Music Expression	2008
Training CNNs for 3-D Sign Language Recognition With Color Texture Coded Joint Angular Displacement Maps	2018

Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction	2019
Robust Fringe Projection Profilometry via Sparse Representation	2016
Time-Aware Multivariate Nearest Neighbor Regression Methods for Traffic Flow Prediction	2015
A New Feature Selection Scheme Using a Data Distribution Factor for Unsupervised Nominal Data	2008
Domain-Specific Image Analysis for Cervical Neoplasia Detection Based on Conditional Random Fields	2011
Online Handwritten Shape Recognition Using Segmental Hidden Markov Models	2007
Upstream Ultrasonic Level Based Soft Sensing of Volumetric Flow of Non-Newtonian Fluids in Open Venturi Channels	2018
A Generative Data Augmentation Model for Enhancing Chinese Dialect Pronunciation Prediction	2012
Adaptive Greedy Dictionary Selection for Web Media Summarization	2017
HPCMP CREATE-AV Quality Assurance: Lessons Learned by Validating and Supporting Computation-Based Engineering Software	2016
Robust Frequency-Hopping Spectrum Estimation Based on Sparse Bayesian Method	2015
Push-Pull: Deterministic Search-Based DAG Scheduling for Heterogeneous Cluster Systems	2007
Applications of Deep Learning and Reinforcement Learning to Biological Data	2018
A Sparse Plus Low-Rank Exponential Language Model for Limited Resource Scenarios	2015
Joint Denoising/Compression of Image Contours via Shape Prior and Context Tree	2018
Learning With Augmented Features for Supervised and Semi-Supervised Heterogeneous Domain Adaptation	2014
Inter-Swarm Content Distribution Among Private BitTorrent Networks	2013
Incomplete Multisource Transfer Learning	2018
Multi-Camera Saliency	2015
Spectrally Constrained Unimodular Sequence Design Without Spectral Level Mask	2018
System Monitoring with Metric-Correlation Models	2011
Salient Object Detection via Structured Matrix Decomposition	2017
On the Projection of PLLRs for Unbounded Feature Distributions in Spoken Language Recognition	2014
Spatio-Temporal Feature-Extraction Techniques for Isolated Gesture Recognition in Arabic Sign Language	2007
Domain Transfer Learning for MCI Conversion Prediction	2015
Decay Rate Estimators and Their Performance for Blind Reverberation Time Estimation	2014
A review of applications of search based software engineering techniques in last decade	2016

A.2 Results of the Tool

A.2.1 Results of Meta-data Extraction

ID: 1

Title: Problem-Based Learning for Engineering Students: An Evidence-Based Comparative Study

Author: Cynthia Hsieh; Lorrie Knight

Creation Year: 2009

Creator: Elsevier

Keywords:

Producer: None

ID: 2

Title: None

Author: None

Creation Year: 2015

Creator: Adobe InDesign CS6 (Windows)

Keywords:

Producer: Adobe PDF Library 10.0.1

ID:3

Title: Applying Systematic Reviews to Diverse Study Types: An Experience Report

Author: Tore Dybå

Creation Year: 2007

Creator: PScript5.dll Version 5.2.2

Keywords:

Producer: Acrobat Distiller 8.1.0 (Windows)

ID: 4

Title: ESELAW-ISBN-978-9974-8379-3-5.pdf

Author: FamiliaDias

Creation Year: 2013

Creator: PDFCreator Version 1.4.1

Keywords:

Producer: GPL Ghostscript 9.05

ID: 5

Title: BNOSA: A Bayesian network and ontology based semantic annotation framework

Author: Quratulain Rajput

Creation Year: 2011

Creator: Elsevier

Keywords:

Producer: Acrobat Distiller 8.1.0 (Windows)

A.2.2 Extraction of Specific Sections

Specific paragraphs: Introduction

1: prepare students to understand the role of engineers and their activities in society. The library workshop introduces students to the library and its resources for research. Traditionally, a small group of librarians divide the instructional responsibilities over a series of comparable sessions. The students are given a library

assignment to be completed during the class.Over the last few years, the librarians have observed a decrease in student attention and less competence among students in the successful completion of their library assignments. The librarian decided to take a proactive approach to change the downward trend. They felt, based on their observations and review of the literature, a more active learning environment would enhance the students' library classroom experience. They proposed the use of a type of problem-based learning (PBL) in the classes. However,the engineering professors were concerned that a changed process might not help students reach their desired learning outcomes.Therefore, the librarians designed a study to compare the students' mastery of the learning outcomes using the lecture-based vs. the problem-based method. The goal was to develop empirical evidence to convince the engineering professors that the new approach would result in positive outcomes.LITERATURE REVIEW a review of the literature suggests that many librarians are aware of the potential positive outcomes of employing problem-basedJanuary 2008 25mailto:chsieh@pacific.edumailto:lknight@pacific.edulearning in bibliographic instruction. This active, participatory teaching style is consistent with the goals of information literacy in several ways. As an inquiry-based form of instruction, PBLclosely parallels the Association of College and ResearchLibraries Information Competency Standards1 for defining and purposefully resolving an information need. Similarly, both PBLand IL share an overarching goal of instilling skills and abilities for lifelong learning. In addition, the theoretical bases that underlie both IL and PBL recognize and address the influence of learning styles on student acquisition and retention of skills and knowledge.A number of studies in the literature of both librarianship and education describe the methods of problem-based learning.Students are given a realistic scenario without a single correct answer. They are guided through a process of analyzing the problem, researching their alternatives, and presenting a recommended solution. Advocates of PBL feel this provides real-world and real-time learning opportunities that replicate the type of problem solving students will use throughout their academic and professional lives.2 Although PBL is not firmly established in library instruction, there are several reports of successful endeavors. Snavely3 reviews several studies about the use ofPBL in library instruction and describes how the lessons learned can lead to ongoing improvement. PBL also fosters collaboration with subject faculty that makes library user education more integral to the course content and therefore more valued by students and professors.4 Pelikan5 notes that PBL is similar to a"reference interview with a twist" in that it allows librarians to partner with students in developing research strategies and selecting databases.There is increasing evidence that supports the belief that active, experiential educational experiences are more transferable than passive, lecture-based instruction.6 This is especially true for today's students, who prefer a self-directed, technological approach to information retrieval. Although faculty and students may believe that technological fluency is equivalent to information competency, librarians recognize the distinction and address it in user education.7 Problem-based learning offers librarians a chance to capitalize on students' pre-existing confidence in theInternet while overlaying it with skills in evaluation and critical thinking.8 Carder, Willingham, and Bibb9 use focused cases to encourage students to identify their own information scenarios and strategies. They conclude that this approach makes information literacy more relevant rather than

"just another hoop to jump." For the current generation of college students,learning is more valid when it builds upon an existing knowledge base and style.10 A literature review in engineering education also revealed an extensive interest in engineering students'learning styles and a general discontent with the conventional lecture-based instructional style."There is increasing evidence that supports the belief that active, experiential educational experiences are more transferable than passive,lecture-based instruction."In an attempt to assess engineering students' preferred learning styles, many studies focus on the role of personality

2: Pearson, 2012), 145–57.

3: The volume of research that needs to be considered by software engineering (SE) practitioners and researchers is constantly expanding. In many areas it has become almost impossible for the individual to read, critically evaluate, and synthesize the state of current knowledge, let alone keep updating this on a regular basis. As a result, reviews have become essential tools for anyone who wants to keep up with the new evidence that is accumulating in his or her field of interest. Reviews are also required to identify areas where the available evidence is insufficient and further studies are required. However, because of the often poor quality of traditional narrative reviews, there has recently been an increasing focus on formal methods of systematically reviewing studies, to produce explicitly formulated, reproducible, and up-to-date summaries of the effects of health care and social science interventions [5], [6], [9], [15], [24], [26]. The recent focus on evidence-based software engineering (EBSE) [8], [19] has resulted in an increased interest in systematic reviews (SRs) within the field of SE as a means of evaluating and interpreting the available research relevant to a particular research question, topic area, or phenomenon of interest [1], [18]. Recent examples of such reviews include [9], [12], [16], [28], EBSE aims to improve decision making related to software development and maintenance by integrating current best evidence from research with practical experience and human values [19]. This is an ambitious aim, particularly because the gap between research and practice can be wide. EBSE seeks to close this gap by encouraging a stronger emphasis on methodological rigor while focusing on relevance for practice. EBSE involves five steps to reach its aim [8]:

A.2.3 Results of Summarization

Results of Extractive Summarization:

Summarization

Summary : 1

1: The Selection Execution is the acquisition and analysis of primary studies and involves three tasks: (i) in the Initial Selection the studies are identified, collected and organized in a list; (ii) in the Studies Evaluation, studies are marked as included

or excluded in two steps, first by reading the title,abstract and conclusions, and then by reading the full text.

2: VTM is the integration of Text Mining, which is the process of extracting patterns and non-trivial knowledge from textual documents [38], and Information Visualization techniques,which enable a user to visually interact in the knowledge acquisition process [31].We believe that the use of VTM can assist the Selection Review of primary studies not only speeding up the SLR process, but also improving the reliability of the results.

3: Felizardo et al. [17] employed VTM to support categorization and classification of studies when carrying out systematic mapping studies; and Malheiros et al. [29] investigated the use of content-based VTM techniques to help with the selection of primary studies, using a feasibility study.

4: There is not any peculiarity of application of the proposed VTM approaches only to SE field.In order to create the content maps we used the Revis,1 which is a visualization and interaction tool that offers a framework of different projection techniques and methods to create data maps in general and document maps in particular based on the similarity between pairs of data points, supporting VTM exploration of document collections.

5: It takes Revis only a few seconds to create and present a map with a few hundred documents.The strategy suggested here to review the studies selection from that map can be summarized into two main steps: (1) first,a clustering algorithm is applied to the content map, creating groups of highly related (similar) documents using the k-means algorithm [28] (the number of clusters is automatically detected based on the number of documents); (2) then, the resulting clusters are analyzed in terms of included and excluded documents in order to find inconsistencies.

6: It can occur if authors write several different papers about the same study (e.g. a conference paper and a journal paper) and decide to include only the most recent ones.Based on this justification, we can assume that similar papers with divergent classifications should be reviewed only if we guarantee that in the content map there is only the most recent versions of papers about each study.

7: Exclusion history The first strategy to review the selection activity considers the creation of content maps containing the studies collected and analyzed in an SLR and highlighting them using different colors in order to differentiate in which of the first two steps of the Evaluation task a study was removed from the review.

8: (a) Each point represents a primary study and the color indicates if a study has been included (blue points) or excluded (red points).

9: The squares represent regions (clusters) where the content of the studies (i.e. title, abstracts and keywords) are similar; (b) the color indicates the history of the selection activity,red points are studies excluded in the first step (reading abstracts), green points are studies excluded in the second step (reading full text) and blue points are included studies.

10: The included study (blue point) is similar in content to other excluded studies and isolated in terms of references, indicating that its inclusion should be reviewed.

11: Case studies: investigating the use of VTM techniques to support the selection review activity.The examples used in Section 3 (i.e. SRLs 1–4) contain a rather small number of primary studies (dozens of articles), however, in real SRLs a large number of candidate studies are considered –hundreds and even thousands.

Results of Keyword based Summarization

Question Game Summary : 1

Question: 1

01: It takes Revis only a few seconds to create and present a map with a few hundred documents.The strategy suggested here to review the studies selection from that map can be summarized into two main steps: (1) first,a clustering algorithm is applied to the content map, creating groups of highly related (similar) documents using the k-means algorithm [28] (the number of clusters is automatically detected based on the number of documents); (2) then, the resulting clusters are analyzed in terms of included and excluded documents in order to find inconsistencies.

02: (a) Each point represents a primary study and the color indicates if a study has been included (blue points) or excluded (red points).

03: The squares represent regions (clusters) where the content of the studies (i.e. title, abstracts and keywords) are similar; (b) the color indicates the history of the selection activity,red points are studies excluded in the first step (reading abstracts), green points are studies excluded in the second step (reading full text) and blue points are included studies.

Question: 2

01: The purpose of EBSE is to inform researchers about empirical evidence that can be used to improve SE practice, however, Kitch-enham et al. [26] found out that relatively few SLRs provided ad-vice oriented to practitioners.The results of an independent assessment study conducted by da Silva et al. [9] showed that the main limitation constraining the use of SLRs in SE is that a large number of SLRs do not assess the quality of the underlying primary studies, confirming the pre-vious findings of Kitchenham et al.

02: This suggests that currently the attention given to research synthesis in SE is limited.K.R.

Question: 3

01: Our proposal focuses on extending that approach to supporting the review of the selection task

Question: 4

01: The Selection Execution is the acquisition and analysis of primary studies and involves three tasks: (i) in the Initial Selection the studies are identified, collected and organized in a list; (ii) in the Studies Evaluation, studies are marked as included

or excluded in two steps, first by reading the title, abstract and conclusions, and then by reading the full text.

02: incorporating two different VTM approaches to supporting the Selection Review task: (i) content map; and (ii) citation map visualizations.

03: Exclusion history: The first strategy to review the selection activity considers the creation of content maps containing the studies collected and analyzed in an SLR and highlighting them using different colors in order to differentiate in which of the first two steps of the Evaluation task a study was removed from the review.

Bibliography

[1] B A. Kitchenham. Procedures for performing systematic reviews. 2004. (cited on Page 1 and 5)

[2] B A. Kitchenham. Guidelines for performing systematic literature reviews in software engineering. 01 2007. (cited on Page xiii, 1, 5, 6, 7, 11, 12, 23, 25, 26, 27, and 65)

[3] B A. Kitchenham, O. Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. Systematic literature reviews in software engineering – a systematic literature review. *Information and Software Technology*, 51(1):7 – 15, 2009. ISSN 0950-5849. doi: https://doi.org/10.1016/j.infsof.2008.09.009. URL http://www.sciencedirect.com/science/article/pii/S0950584908001390. Special Section - Most Cited Articles in 2002 and Regular Research Papers. (cited on Page 1)

[4] Palakorn Achananuparp, Xiaohua Hu, and Xiajiong Shen. The evaluation of sentence similarity measures. pages 305–316, 2008. (cited on Page 21 and 22)

[5] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D Trippe, Juan B Gutierrez, and Krys Kochut. Text summarization techniques: a brief survey. *arXiv preprint arXiv:1707.02268*, 2017. (cited on Page 36)

[6] Edoardo Aromataris and Dagmara Riitano. Constructing a search strategy and searching for evidence. *Am J Nurs*, 114(5):49–56, 2014. (cited on Page 8)

[7] Samrat Babar and Rit. Text summarization: An overview. 10 2013. (cited on Page 12)

[8] Deepika Badampudi, Claes Wohlin, and Kai Petersen. Experiences from using snowballing and database searches in systematic literature studies. page 17, 2015. (cited on Page 10)

[9] Colin B Begg and Jesse A Berlin. Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 419–463, 1988. (cited on Page 8)

[10] Shawn Behrends. Libraries vs. google in the 21st century. 2012. (cited on Page 9)

[11] Slobodan Beliga. Keyword extraction : A review of methods and approaches. 2014. (cited on Page 14, 17, and 18)

[12] David Bowes, Tracy Hall, and Sarah Beecham. Slurp - a tool to help large complex systematic literature reviews deliver calid and rigorous results. pages 33–36, 09 2012. (cited on Page 29, 30, and 31)

[13] Amy Brand, Frank Daly, and Barbara Meyers. Metadata demystified. *Bethesda, MD*, pages 1–19, 2003. (cited on Page 35)

[14] Giuseppe Carenini and Jackie Chi Kit Cheung. Extractive vs. nlg-based abstractive summarization of evaluative text: The effect of corpus controversiality. pages 33–41, 2008. URL http://dl.acm.org/citation.cfm?id=1708322.1708330. (cited on Page 13)

[15] J. C. Carver, E. Hassler, E. Hernandes, and N. A. Kraft. Identifying barriers to the systematic literature review process. pages 203–212, Oct 2013. ISSN 1949-3770. doi: 10.1109/ESEM.2013.28. (cited on Page 1)

[16] Jerome Dinet, Monik Favart, and Jean-Michel Passerault. Searching for information in an online public access catalogue (opac): the impacts of information search expertise on the use of boolean operators. *Journal of Computer Assisted Learning*, 20(5):338–346, 2004. (cited on Page 9)

[17] J. Doumont. English communication for scientists. 2010. (cited on Page 36 and 39)

[18] Tore Dybå and Torgeir Dingsøyr. Strength of evidence in systematic reviews in software engineering. pages 178–187, 2008. (cited on Page 26)

[19] Günes Erkan and Dragomir R Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 22: 457–479, 2004. (cited on Page 16)

[20] Sandra Fabbri, Elis Hernandes, Andre Di Thommazo, Anderson Belgamo, Augusto Zamboni, and Cleiton Silva. Using information visualization and text mining to facilitate the conduction of systematic literature reviews. pages 243–256, 2013. (cited on Page 29, 30, and 31)

[21] Katia Romero Felizardo, Elisa Yumi Nakagawa, Daniel Feitosa, Rosane Minghim, and José Carlos Maldonado. An approach based on visual text mining to support categorization and classification in the systematic mapping. pages 34–43, 2010. URL http://dl.acm.org/citation.cfm?id=2227057.2227062. (cited on Page 31)

[22] Katia Romero Felizardo, Mehwish Riaz, Muhammad Sulayman, Emilia Mendes, Stephen G MacDonell, and Jose Carlos Maldonado. Analysing the use of graphs to represent the results of systematic reviews in software engineering. pages 174–183, 2011. (cited on Page 31)

[23] Luyi Feng, Yin Kia Chiam, and Sin Kuang Lo. Text-mining techniques and tools for systematic literature reviews: A systematic literature review. pages 41–50, 2017. (cited on Page 1)

[24] Ana Fernández-Sáez, Marcela Genero, and Francisco Romero. Slr-tool - a tool for performing systematic literature reviews. pages 157–166, 01 2010. (cited on Page 29, 30, 31, and 32)

[25] Patricia Franks. Why metadata matters. *Information Management Journal*, pages 55–61, 01 2006. (cited on Page 36)

[26] Adam Geitgey. Natural language processing is fun. *Retrieved from https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e Google Scholar*, 2018. (cited on Page 18 and 20)

[27] Yvette Graham. Re-evaluating automatic summarization with bleu and 192 shades of rouge. pages 128–137, 2015. (cited on Page 54)

[28] Edgar Hassler, Jeffrey C. Carver, David Hale, and Ahmed Al-Zubidy. Identification of slr tool needs – results of a community workshop. *Information and Software Technology*, 70:122 – 129, 2016. ISSN 0950-5849. doi: https://doi.org/10.1016/j.infsof.2015.10.011. URL http://www.sciencedirect.com/science/article/pii/S0950584915001779. (cited on Page 1 and 9)

[29] Staffs Keele et al. Guidelines for performing systematic literature reviews in software engineering. 2007. (cited on Page 10)

[30] Wahab Khan, Ali Daud, Jamal A Nasir, and Tehmina Amjad. A survey on the state-of-the-art machine learning models in the context of nlp. *Kuwait journal of Science*, 43(4), 2016. (cited on Page 18)

[31] Sandra Kübler, Ryan McDonald, and Joakim Nivre. Dependency parsing. *Synthesis Lectures on Human Language Technologies*, 1(1):1–127, 2009. (cited on Page 20)

[32] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. *Text Summarization Branches Out*, 2004. (cited on Page 55)

[33] Edward Loper and Steven Bird. Nltk: the natural language toolkit. *arXiv preprint cs/0205028*, 2002. (cited on Page 14 and 18)

[34] Hans Peter Luhn. The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165, 1958. (cited on Page xiii, 17, and 45)

[35] J. C. Maldonado, M. Mendonca, R. Pinho, V. Malheiros, and E. Hohn. A visual text mining approach for systematic reviews. 00:245–254, 09 2007. ISSN 1938-6451. doi: 10.1109/ESEM.2007.21. URL doi.ieeecomputersociety.org/10.1109/ESEM.2007.21. (cited on Page 31)

[36] Richard Mallett, Jessica Hagen-Zanker, Rachel Slater, and Maren Duvendack. The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness*, 4(3):445–455, 2012. doi: 10.1080/19439342.2012.711342. URL https://doi.org/10.1080/19439342.2012.711342. (cited on Page 1)

[37] Bill Manaris. Natural language processing: A human-computer interaction perspective. *Advances in Computers*, 47:1–66, 12 1998. doi: 10.1016/S0065-2458(08)60665-8. (cited on Page 18)

[38] Inderjeet Mani, David House, Gary Klein, Lynette Hirschman, Therese Firmin, and Beth Sundheim. The tipster summac text summarization evaluation. pages 77–85, 1999. doi: 10.3115/977035.977047. URL https://doi.org/10.3115/977035.977047. (cited on Page 13)

[39] Christopher Marshall and Pearl Brereton. Tools to support systematic literature reviews in software engineering: A mapping study. pages 296–299, 2013. (cited on Page 2)

[40] George Mason. How to write a research question. 2014. URL http://writingcenter.gmu.edu/writing-resources/wc-quick-guides. (cited on Page 8)

[41] Paula Gomes Mian, Tayana Conte, Ana Candida Cruz Natali, Jorge Calmon de Almeida Biolchini, and Guilherme Horta Travassos. A systematic review process for software engineering. 2005. (cited on Page 9)

[42] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. 2004. (cited on Page xiii, 13, 14, 15, and 41)

[43] Jessica Milstead and Susan Feldman. Metadata: Cataloging by any other name. *Online-Weston then Wilton*, 23:24–31, 1999. (cited on Page 35)

[44] Sravan Nadella. Automatic text summarization using importance of sentences for email corpus. 2015. (cited on Page 13)

[45] Chitu Okoli and Kira Schabram. A guide to conducting a systematic literature review of information systems research. 2010. (cited on Page 5)

[46] You Ouyang, Wenji Li, and Qin Lu. An integrated multi-document summarization approach based on word hierarchical representation. pages 113–116, 2009. (cited on Page 41)

[47] Aarti Patil, Komal Pharande, Dipali Nale, and Roshani Agrawal. Automatic text summarization. *International Journal of Computer Applications*, 109(17), 2015. (cited on Page 13)

[48] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1 – 18, 2015. ISSN 0950-5849. doi: https://doi.org/10.1016/j.infsof.2015.03.007. URL http://www.sciencedirect.com/science/article/pii/S0950584915000646. (cited on Page 10)

[49] Rory J Piper. How to write a systematic literature review: A guide for medical students. 2013. (cited on Page 1)

[50] Neil J Salkind. Encyclopedia of research design. 1, 2010. (cited on Page 9)

[51] Gerard Salton. Automatic text processing: The transformation, analysis, and retrieval of information by computer. 1989. (cited on Page 31)

[52] Michel F. Sanner. Python: a programming language for software integration and development. *Journal of molecular graphics modelling*, 17 1:57–61, 1999. (cited on Page 36)

[53] Minakshi R Shinde and Parmeet C Gill. Pattern discovery techniques for the text mining and its applications. 2014. (cited on Page 12)

[54] Dima Shulga. Introduction to text representation and similarity. 2017. URL https://towardsdatascience.com/introduction-to-text-representation-and-similarity-b5dd3fd71737. (cited on Page xiii and 21)

[55] Josef Steinberger. Using latent semantic analysis in text summarization and summary evaluation. 2004. (cited on Page 15, 16, and 44)

[56] Josef Steinberger. Lsa-based multi-document summarization. 2007. (cited on Page 44)

[57] Evelina Tacconelli. Systematic reviews: Crd's guidance for undertaking reviews in health care. *The Lancet Infectious Diseases*, 10(4):226, 2010. (cited on Page 10)

[58] José Torres, Daniela Cruzes, and Laís N. Salvador. Automatically locating results to support systematic reviews in software engineering. 01 2013. (cited on Page xiii, 15, and 16)

[59] Juan-Manuel Torres-Moreno. Automatic text summarization. pages 179–217, 09 2014. doi: 10.1002/9781119004752.ch6. (cited on Page 41)

[60] Matt Vassar, Paul Atakpo, and Melissa J Kash. Manual search approaches used by systematic reviewers in dermatology. *Journal of the Medical Library Association: JMLA*, 104(4):302, 2016. (cited on Page 8)

[61] MK Vijaymeena and K Kavitha. A survey on similarity measures in text mining. 2016. (cited on Page 20)

[62] Jonathan J. Webster and Chunyu Kit. Tokenization as the initial phase in nlp. pages 1106–1110, 1992. doi: 10.3115/992424.992434. URL https://doi.org/10.3115/992424.992434. (cited on Page 18 and 20)

[63] He Zhang, Muhammad Ali Babar, and Paolo Tell. Identifying relevant studies in software engineering. *Information and Software Technology*, 53(6):625–637, 2011. (cited on Page 2)

[64] You Zhou, He Zhang, Xin Huang, Song Yang, Muhammad Ali Babar, and Hao Tang. Quality assessment of systematic reviews in software engineering: A tertiary study. pages 14:1–14:14, 2015. doi: 10.1145/2745802.2745815. URL http://doi.acm.org/10.1145/2745802.2745815. (cited on Page 11)

[65] Ekaterina Zublenko. Why django is the best web framework for your project. 2016. (cited on Page 36)