

Survey

孙伟松

2021.09.10

Survey vs. Literature Review

- According to the definition of survey paper provided by IEEE Communications Surveys & Tutorials journal (one of the best CS journals), "The term survey, as applied here, **is defined to mean a survey of the literature**. A survey article should provide a comprehensive review of developments in a selected area".
- In ACM Computing Survey (another prestigious CS journal), survey paper is described as “A paper that **summarizes and organizes recent research results in a novel way** that integrates and adds understanding to work in the field. A survey article **emphasizes the classification of the existing literature**, developing a perspective on the area, and evaluating trends.”
- In Elsevier journal of Computer Science Review, you will see that “Critical review of the relevant literature“ is **required a component** of every typical survey paper.

<https://academia.stackexchange.com/questions/15080/literature-review-versus-literature-survey-what-is-the-difference>

Systematic Literature Review

A systematic literature review (often referred to as a systematic review) is a means of **identifying**, **evaluating** and **interpreting** all available research relevant to a **particular research question**, or **topic area**, or **phenomenon of interest**. Individual studies contributing to a systematic review are called *primary* studies; a systematic review is a form of *secondary* study.

Systematic reviews aim to present a fair evaluation of a research topic by using a **trustworthy**, **rigorous**, and **auditable** methodology.

Reasons for Performing Systematic Literature Review

What benefits can we gain from systematic literature review?

- To **summarize the existing evidence concerning a technology** e.g. to summarize the empirical evidence of the benefits and limitations of a specific agile method. [e.g., technology development]
- To **identify any gaps in current research** to suggest areas for further investigation. [e.g., room for improvement]
- To **provide a framework/background** in order to appropriately position new research activities. [e.g., new direction]
- To **examine the extent to which empirical evidence supports/contradicts theoretical hypotheses**, or even to assist the generation of new hypotheses. [e.g., test coverage]

The Review Process

A systematic literature review involves several **discrete activities**.

The stages in a systematic review into three main phases:

- Planning the Review;
- Conducting the Review;
- Reporting the Review.

The Review Process—Planning the Review

The stages associated with planning the review are:

1. Identification of the need for a review

The need for a systematic review **arises from the requirement of researchers** to summarize all existing information about some phenomenon in a thorough and unbiased manner.

2. Commissioning a review

Sometimes an organization requires information about a specific topic but does not have the time or expertise to perform a systematic literature itself.

3. Specifying the research question(s)

Specifying the research questions is the most important part of any systematic review. The review questions drive the entire systematic review methodology.

4. Developing a review protocol

A review protocol **specifies the methods that will be used to undertake a specific systematic review**. A pre-defined protocol is necessary to reduce the possibility of researcher bias.

5. Evaluating the review protocol

The protocol is a critical element of any systematic review. Researchers must agree a procedure for evaluating the protocol.

The Review Process—Conducting the Review

The stages associated with conducting the review are:

1. Identification of research

The aim of a systematic review is to **find as many primary studies** relating to the research question as possible using an unbiased search strategy.

2. Selection of primary studies

Once the potentially relevant primary studies have been obtained, they need to be **assessed for their actual relevance**.

3. Study quality assessment

In addition to general inclusion/exclusion criteria, it is considered critical to **assess the “quality” of primary studies**.

4. Data extraction and monitoring

The objective of this stage is to design **data extraction forms to accurately record the information** researchers obtain from the primary studies.

5. Data synthesis

Data synthesis involves **collating and summarizing the results** of the included primary studies.

The Review Process—Reporting the review

The stages associated with reporting the review are:

1. Specifying dissemination mechanisms

It is important to communicate the results of a systematic review effectively. Academics usually assume that dissemination is about reporting results in academic journals and/or conferences.

2. Formatting the main report

In a technical report or in a section of a PhD thesis, a journal or conference paper.

3. Evaluating the report

Journal articles will be peer reviewed as a matter of course. Experts review PhD theses as part of the examination process.

Example: A Systematic Literature Review of Automated Query Reformulations in Source Code Search

Studies show that software maintenance costs up to 80% of the total budget. As a part of maintenance, software developers often resolve critical bugs to ensure the reliability of their software. They might also need to add new features to their software on a regular interval to stay competitive in the market. These bugs and features are reported as change requests (i.e., technical documents written by software users). Developers consult these documents to implement the required changes in the software code. As a part of change implementation, they often choose a few important keywords from a change request as an ad hoc query. Then they execute the query with a code search engine (e.g., Lucene) and attempt to find out the exact locations within the software code that need to be changed.

Unfortunately, even the experienced developers often fail to choose the right search queries. As a consequence, the developers often experience difficulties in detecting the appropriate locations within the code and spend the majority of their time in numerous trials and errors. There have been **many studies that attempt to support developers in constructing queries by automatically reformulating their ad hoc queries**. In this systematic literature review, we carefully **select 70 primary studies on query reformulations from 2,970 candidate studies**, perform an in-depth qualitative analysis using the Grounded Theory approach, and then **answer six important research questions**. We analyze the methodologies, evaluation approaches, and challenges of these primary studies, and provide the statistical evidence on the conducted researches **during the last 15 years**.

Our investigation has reported several major findings. First, to date, eight major methodologies (e.g., term weighting, query-term co-occurrence analysis, thesaurus lookup) have been adopted in query reformulation. Second, the existing studies suffer from several major limitations (e.g., lack of generalizability, vocabulary mismatch problem, weak evaluation, the extra burden on the developers) that might prevent their wide adoption. Finally, we **discuss several open issues** in search query reformulations and **suggest multiple future research opportunities**.

Example: Research Questions

Table 1. Research Questions

Ref#	Question
General questions	
RQ₁	Which methods, algorithms and data sources have been used for automated query reformulations targeting code search in the literature?
RQ₂	Which methods, metrics or subject systems have been used to evaluate and validate the researches on automated query reformulations?
RQ₃	What are the major challenges of automated query reformulations intended for code search? How many of them have been solved to date by the literature?
Statistical questions	
RQ₄	How much activities of research on automated query reformulations have been performed to date? What are the venues that these researches got published at?
Focused questions	
RQ₅	What are the differences and similarities between query reformulations for local code search and query reformulations for Internet-scale code search?
RQ₆	What are the scopes for future work in the area of automated query reformulation targeting the code search?

Example: Primary Studies Search

	Initial Search	Impurity Removal	Filtered by Title (C1&C2)	Filtered by Abstract (C1&C2)	Combined	Duplicate Removal	Filter by Full Texts (C3)	Final Selection
(2004–2018)								
ACM Digital Library	176	114	40	16	+			56
CrossRef	152	101	27	7				
DBLP	216	179	74	10				
Mendeley	203	123	59	30				
Google Scholar	200	173	81	18				
IEEE Explore	432	299	65	41				
ProQuest	137	127	68	38				
ScienceDirect	22	14	04	04				
SpringerLink	313	265	47	03				
Web of Science (ISI)	999	912	91	26				
Wiley Online Library	21	10	04	02				
Total:	2,871	2,317	562	195	109	93	62	56
(2018–2021)								
Google Scholar	99	99	85	30	30	27	14	14

Sources to be Searched

Source	Responsible
Information and Software Technology (IST)	Kitchenham
Journal of Systems and Software	Kitchenham
IEEE Transactions on Software Engineering	Kitchenham
IEEE Software	Kitchenham
Communications of the ACM (CACM)	Brereton
ACM Surveys	Brereton
Transactions on Software Engineering Methods (TOSEM)	Brereton
Software Practice and Experience	Budgen & Kitchenham
Empirical Software Engineering Journal (ESEM)	Budgen
IEE Proceedings Software (now IET Software)	Kitchenham
Proceedings International Conference on Software Engineering (ICSE 04, 05, 06, 07)	Linkman & Kitchenham & Brereton
Proceedings International Seminar of Software Metrics (Metrics04, Metrics05)	Kitchenham & Brereton
Proceedings International Seminar on Empirical Software Engineering (ISESE 04, 05, 06)	Kitchenham & Brereton

Fig. 4. Selection of primary studies

"If I have seen a little farther than others, it is because I have stood on the shoulders of giants."

--Newton

Thank you for your listening!

Projects

1. A Survey of Code Search
2. A Survey of Code Summarization
3. A Survey of Test Case Prioritization

孙伟松 : weisongsun@smail.nju.edu.cn

张犬俊 : quanjun.zhang@smail.nju.edu.cn