

Towards an Evidence-Based Understanding of Electronic Data Sources

Lianping Chen
Lero, University of Limerick
Limerick, Ireland
lianping.chen@lero.ie

Muhammad Ali Babar
IT University of Copenhagen
Copenhagen, Denmark
malibaba@itu.dk

He Zhang
NICTA, UNSW
Sydney, Australia
he.zhang@nicta.com.au

Systematic Literature Reviews and Systematic Mapping Studies are relatively new forms of secondary studies in software engineering. Identifying relevant papers from various Electronic Data Sources (EDS) is one of the key activities of conducting these kinds of studies. Hence, the selection of EDS for searching the potentially relevant papers is an important decision, which can affect a study's coverage of relevant papers. Researchers usually select EDS mainly based on personal knowledge, experience, and preferences and/or recommendations by other researchers. We believe that building an evidence-based understanding of EDS can enable researchers to make more informed decisions about the selection of EDS. This paper reports our initial effort towards this end. We propose an initial set of metrics for characterizing the EDS from the perspective of the needs of secondary studies. We explain the usage and benefits of the proposed metrics using the data gathered from two secondary studies. We also tried to synthesize the data from the two studies and that from literature to provide initial evidence-based heuristics for EDS selection.

Evidence-Based Software Engineering, systematic literature review; systematic mapping study; primary study selection, search engine selection, database selection

1. INTRODUCTION

Secondary studies, such as Systematic Literature Reviews (SLR) and Systematic Mapping Studies (SMS), are the core methods of Evidence-Based Software Engineering (EBSE) (Kitchenham et al., 2004), which is expected to help practitioners to make informed decisions about technology selection and adoption (Dyba et al., 2005). Recently, there has been a growing interest in performing SLR and SMS in Software Engineering (SE). Retrieval of relevant papers is an important concern in any SLR and SMS (Dieste et al., 2009). This activity determines the coverage of the relevant papers to be considered in an SLR or SMS. Automatic search, i.e., executing search strings on Electronic Data Sources (EDS¹), is the dominant method for identifying relevant papers². However, researchers' selection of EDS is usually based on their personal preferences or ad hoc experiences, rather than an evidence-based decision on the basis of evidence. We assert that building an evidence-based understanding of EDS is essential for improving the current state of the practice in identifying primary studies through automatic searches. This

paper reports our initial efforts in reaching such an evidence-based understanding of EDS.

2. ELECTRONIC DATA SOURCES (EDS)

We have collected a list of EDS from a library of published SLR and SMS maintained by our research group. The following are the most often used EDS by SE researchers:

1. IEEE Xplore (IEEE);
2. ACM Digital Library (ACM);
3. EI Compendex & Inspec (EI);
4. ISI Web of Science (WoS);
5. CiteSeer (CS);
6. Google Scholar (GS);
7. ScienceDirect (SD);
8. SpringerLink (SL);
9. Wiley InterScience (WIS);
10. SCOPUS (SCOPUS);
11. Kluwer Online (KO)³.

We classify EDS into two main categories: index engines and publishers' sites. Index engines mainly index the work published by various publishers. Some popular index engines are WoS, EI, GS, SCOPUS, and CS. Publishers' sites refer to the online literature

¹ EDS is singular as well as plural based on the context in which it is used.

² Another method for identifying relevant studies is manual search, which refers to the search performed by manually browsing the journals or conference proceedings. Manual search has several advantages over automatic search; however, due to the vast volume of available literature, manual search is usually too time-consuming.

³ It should be noted that KO has been merged with SL

databases provided by the publishers to facilitate easy retrieval of the published literature. Some popular publishers' sites are IEEE, ACM, SD, SL, and WIS. Some EDS (e.g., ACM) can be considered both functions of index engines and publisher's sites.

3. METRICS FOR CHARACTERIZING EDS

In the context of primary study search, precision⁴ and recall⁵ are two widely used metrics in SLR. The two metrics are usually used for measuring the effectiveness of a search string when applied to a particular EDS or a particular set of EDS, and for measuring the effectiveness of the whole search phase. However, they are not adequate for measuring the effectiveness of EDS directly.

In the context of SLR or SMS, researchers often search potentially relevant papers from multiple EDS. There may be rare cases, if any, where researchers rely on only one EDS for identifying relevant papers. Rather, researchers usually select a set of EDS. It is the combination of those selected EDS that provide researchers with an expected coverage of available relevant papers. Thus, light should be shed on the combination of several EDS as a whole instead of an isolated single EDS. Following this principle, we have proposed three metrics for characterizing EDS for helping better understanding of EDS by researchers performing secondary studies. The proposed three metrics are: overall contribution, overlap, and exclusive contribution. They are summarized as bellow.

- **Overall contribution** represents the percentage of the relevant papers returned by a certain EDS out of the total relevant papers. This metric can help identify the dominant EDS.
- **Overlap** indicates the papers that are returned by multiple EDS. The overlap can be represented in an overlap matrix.
- **Exclusive contribution** represents those papers that may be missed if a certain EDS is not searched. This metric can help researchers to decide as to which EDS should be omitted if they cannot afford to search all EDS because of limited resources. It can also provide some indication about the number of papers that may be omitted when a certain EDS is not searched.

4. TWO CASES

We have used two secondary studies performed in our research group as two cases. We have analyzed the search results obtained while performing the two studies using the abovementioned three metrics. The

first study is an SLR of Variability Management (VM) in software product lines ("SLR of VM"). The results of this study have been published in (Chen et al., 2009b). The second study is an SMS of Architectural Knowledge Management ("SMS of AKM"). It is at the stage of data analysis and reporting. In the following sections, we briefly describe each of these two cases.

4.1 SLR of VM (Case 1)

This study aimed to review the status of evaluation of VM approaches in software product lines (an overview of VM approaches is presented in (Chen et al., 2009a)). The search string⁶ used is as bellow:

<<software AND (product line OR product lines OR product family OR product families) AND (variability OR variation OR variant)>>

The search string was run on the following seven EDS: IEEE, ACM, CS, SD, EI, SL, and WoS. From all sources, 628 papers were found after removing the duplicates. Ninety-seven papers were finally selected.

The data of the metrics are presented in Table 1 to Table 3. Table 1 presents the overall contribution of each EDS. EI has the largest overall contribution (i.e., 42, 43.3%). This means that 43.3% relevant papers can be retrieved by only searching EI. Table 2 presents the overlap matrix. The number in each cell tells the number of papers returned from both EDS indicated by the cell's column and row header; hence, the number represents the extent of overlapping in terms of number of papers returned from both EDS. For example, the overlap between EI and WoS is 15 papers, which means that there are 15 papers' overlap between the results from EI and WoS. Table 3 presents the exclusive contribution. For example, the exclusive contribution of SL is 1 (1.01%), which means that 1 (1.01%) relevant paper will be missed if SL was not searched. If researchers cannot afford to search all databases, they can consider omitting SL first.

Table 1: Overall contribution of each EDS in Case 1

EDS	IEEE	ACM	CS	SD	EI	SL	WoS
Overall contribution	25	12	16	10	42	18	27
Percent	25.77	12.37	16.49	10.31	43.3	18.56	27.84

Table 2: The overlap matrix of Case 1

EDS	IEEE	ACM	CS	SD	EI	SL	WoS
IEEE	x				9		2
ACM		x	1		3		
CS		1	x	2	5	2	3
SD			2	x	6		6
EI	9	3	5	6	x	13	15
SL			2		13	x	11
WoS	2		3	6	15	11	x

⁴ Precision = no. of relevant documents retrieved by a search/no. of documents retrieved by that search

⁵ Recall = no. of relevant documents retrieved by a search/no. of existing relevant documents

⁶ The search string presented here is in an EDS-independent syntax. We have tried to develop an equivalent search string using the syntax of a particular EDS when we search that particular EDS.

Table 3: Exclusive contribution of each EDS in Case 1

EDS	IEEE	ACM	CS	SD	EI	SL	WoS
Exclusive contribution	16	8	8	3	10	1	7
Percent	16.49	8.25	8.25	3.09	10.31	1.03	7.22

4.2 SMS of AKM (Case 2)

This study is to provide a structure of the type of research reports and results that have been published in AKM by categorising them. The search string, constructed after piloting several candidate search strings, is as follows:

<<(software OR system OR systems) AND (architecture design rationale OR architecture design decision OR architectural design rationale OR architectural design decision OR architectural design decisions OR architectural knowledge OR architecture knowledge OR architectural decision OR architectural decisions OR architecture decision OR architecture decisions)>>

The search was run on the following seven EDS: ACM, IEEE, WoS, SD, EI, WIS, and SL. We received 1,962 results by all the searched EDS. After applying the inclusion and exclusion criteria, 130 papers were selected. The results from the analysis using the three abovementioned metrics were presented in Table 4, Table 5, and Table 6 respectively.

Table 4: Overall contribution of each EDS in Case 2

EDS	IEEE	ACM	SL	SD	EI	WoS	WIS
Overall Contribution	53	47	15	7	85	63	1
Percent	40.77	36.15	11.54	5.38	65.38	48.46	0.77

Table 5: The overlap matrix of Case 2

EDS	IEEE	ACM	SL	SD	WoS	EI	WIS
IEEE	×	14			30	46	
ACM	14	×			6	11	
SL			×		15	14	
SD				×	6	6	
WoS	30	6	15	6	×	59	
EI	46	11	14	6	59	×	
WIS							×

Table 6: Exclusive contribution of each EDS in Case 2

EDS	IEEE	ACM	SL	SD	EI	WoS	WIS
Exclusive Contribution	3	32	0	1	9	3	1
Percent	2.31	24.62	0	0.77	6.92	2.31	0.77

5. A SYNTHESIS OF EXISTING EVIDENCES

We tried to synthesise the existing evidences about the three metrics. We performed an ad hoc literature search to find evidences related to EDS selection. The work by Bailey et al. (Bailey et al., 2007) and Dybå et al. (Dyba et al., 2007) were found. Bailey et al. mainly focuses on one (i.e., overlaps) of the three proposed metrics. The data reported by Bailey et al. covers three

SLRs in three different areas. Dybå et al. reported that articles retrieved from SD, SL, KO, and WIS were also returned by either WoS or EI Compendex based on their SLR in agile software development. Based on this evidence, which covers six SLRs in six different research areas, we had the following findings:

1. Very little overlap was found between different publishers' sites.
2. The results overlap between index engines and publishers' sites usually exist.
3. The results overlap among index engines usually exist, and sometimes, the overlap is significant.
4. If EI and WoS are searched, no significant amount of relevant papers will be missed if SL is not searched.
5. IEEE and ACM usually have considerable exclusive contributions.
6. EI and WoS usually have considerable exclusive contributions, and high overall contributions.

It is worth to note that Bailey et al. (Bailey et al., 2007) concluded that "very little overlap was found" between the EDS. However, when we looked into their data presented in the paper, we found that the overlap does exist between index engines and publishers' sites (e.g., GS has considerable overlap with ACM and IEEE). The overlap between different search engines exists as well. The overlap between different publishers' sites is rare. We assert that classifying EDS into index engines and publishers' sites is necessary to increase the clarity of the data analysis and to draw concrete and fine-grained conclusions.

As believed by many researchers (including the authors), the suitable selection of EDS is topic specific because, e.g., area specific conferences and journals are published through particular publishers. However, the existing evidence indicates that certain rules/patterns do exist (e.g., as summarised above). Given the limited number of cases studied in this paper, we could not claim the above items as rules in this stage; however, our findings imply that some underlying rules/patterns may exist, which encourages us to collect more evidence for further validation.

6. DISCUSSION

6.1 The Usage of the Measurement Results

The results from analysing the search results using the proposed three metrics can help increase researchers' evidence-based understanding of EDS for conducting secondary studies. For example, the results can provide researchers with evidence-based information for EDS selection related questions such as: (a) Which EDS to search for relevant papers? (b) How many relevant papers will be missed if a certain EDS was not searched? (c) How to prioritize different EDS during

search? The measurement results using the three metrics from a particular SLR or SMS will directly benefit researchers in the area in which an SLR or SMS is performed. Even though the value of the metrics can be varying over time and topic, and unknown before the search, the synthesis of the metrics from multiple SLRs still show some observable phenomena (e.g., as summarised in Section 5). They can be useful for researchers doing SLRs in new topics. At least they can be used as evidence-based heuristics. To give a more concrete understanding, we describe some scenarios here.

1) Suppose a researcher is going to do a SLR in VM (it is not uncommon that multiple SLRs will be conducted in the same area, e.g., there are multiple SLRs in cost estimation), a sub topic of VM, or an area tightly related to VM, the data of the three metrics in Case 1 will be useful to her, especially when she is new to the area.

2) Bjørnson and Dingsøyr (Bjørnson and Dingsøyr, 2008) performed a SLR on knowledge management in software engineering. They searched IEEE, ACM, EI and WoS only based on their opinion that articles retrieved from SD, SL, KO, and WIS are also returned by either WoS or EI Compendex. Based on the existing evidence, their opinion appears to be incorrect. However, based on the existing evidence (as summarised in Section 5), the readers will have a clue that the study may have missed some papers due to not searching SD, SL, KO, and WIS, but only very limited number of papers may have been missed.

3) According to our interviews with some researchers worldwide, some of them have problems to access SL. Thus, searching SL is difficult for them. Based on the synthesised results, the researchers will know that if they do not search SL, probably only very limited number of papers might be missed if they search EI and WoS. The researchers can decide whether or not to search SL according to the requirement of literature coverage of their review and their available resources.

6.2 Cautions

There are chances that the EDS with low exclusive contributions may return few important papers that will not be returned by any other searched EDS. In the case where the results need to be exhaustive, the EDS with low exclusive contributions may still need to be searched. Changes of EDS (e.g., changes in indexing scope) need to be considered when using the results from studies performed before the EDS changes. In practice, selection of EDS also needs to take other factors into consideration (e.g., usability of EDS).

7. SUMMARY

Retrieval of relevant papers is an important problem in any evidence-based discipline (Dieste et al., 2009). The selection of EDS can have significant effect on the

coverage of relevant papers for a secondary study. However, currently, the selection of EDS mainly relies on researchers' preferences and ad hoc experiences, rather than, evidence-based decision. This paper reports our initial efforts aimed at improving this situation. We have proposed a set of three metrics to characterize EDS. We have then presented two cases where we applied the proposed metrics. We also have synthesised the data from our two cases and that from the literature. The results can serve as initial evidence based heuristics for EDS selection. We have also discussed how the results from the analysis using these three metrics can be used. This work may increase the researchers' awareness of the characteristics of EDS and the impact of their combination on search performance.

We assert that if researchers report the results from their SLR or SMS for the proposed three metrics as by-products of their SLR (only very limited extra effort is required per our experience), the results will benefit researchers interested in that particular area where the SLR or SMS is performed. The synthesised results from those SLR or SMS will gradually form an evidence base that can help researchers to make evidence-based decisions on EDS selection, which in turn will advance the state of practice of secondary studies in software engineering.

Acknowledgement: This work is partly supported by SFI under grant no. 03/CE2/I303_1.

8. REFERENCES

- Bailey, J., Cheng, Z., Budgen, D., Charters, S. & Turner, M. (2007) Search Engine Overlaps : Do they agree or disagree? *2nd Int'l Workshop on Realising Evidence-Based Softw. Engineering*.
- Bjørnson, F. O. & Dingsøyr, T. (2008) Knowledge management in software engineering: A systematic review of studied concepts, findings and research methods used. *Information and Software Technology*, 50, 1055-1068.
- Chen, L., Babar, M. A. & Ali, N. (2009a) Variability Management in Software Product Lines: A Systematic Review. *the 13th International Software Product Line Conference*. San Francisco, CA, USA.
- Chen, L., Babar, M. A. & Cawley, C. (2009b) A Status Report on the Evaluation of Variability Management Approaches. *13th International Conference on Evaluation and Assessment in Software Engineering*. UK, British Computer Society.
- Dieste, O., Grimán, A. & Juristo, N. (2009) Developing search strategies for detecting relevant experiments. *Empirical Software Engineering*, 14, 513-539.
- Dyba, T., Dingsøyr, T. & Hanssen, G. K. (2007) Applying Systematic Reviews to Diverse Study Types: An Experience Report. *ESEM*.
- Dyba, T., Kitchenham, B. A. & Jorgensen, M. (2005) Evidence-based software engineering for practitioners. *Software, IEEE*, 22, 58-65.
- Kitchenham, B., Dyba, T. & Jorgensen, M. (2004) Evidence-Based Software Engineering. *ICSE*. IEEE.