1. Searching Methods:

We applied three methods in order to meet the specified search string, which contains two main parts: RE keywords and NLP keywords as shown below:

Main Keywords	**Alternative Keywords**
Part I: Requirements engineering +	
Requirements elicitation, requirements analysis, requirements specification, requirements modeling, requirements modeling, requirements validation, requirements verification, requirements management, requirements traceability, requirements classification, requirements document, requirements specification.	
Part II: Natural language processing +	
NLP, statistical NLP, machine learning, deep learning, ~~information*~~ extraction, ~~information~~ retrieval^, text mining, text analysis, linguistic instruments, linguistic approaches.	

*^ We removed the term 'information' from the keywords list, in NLP part, to reduce the number of characters in the search string, since some of the DLs do NOT allow for long queries, also to save space for other important keyword related to RE.

The search methods are depicted in Figure 1.

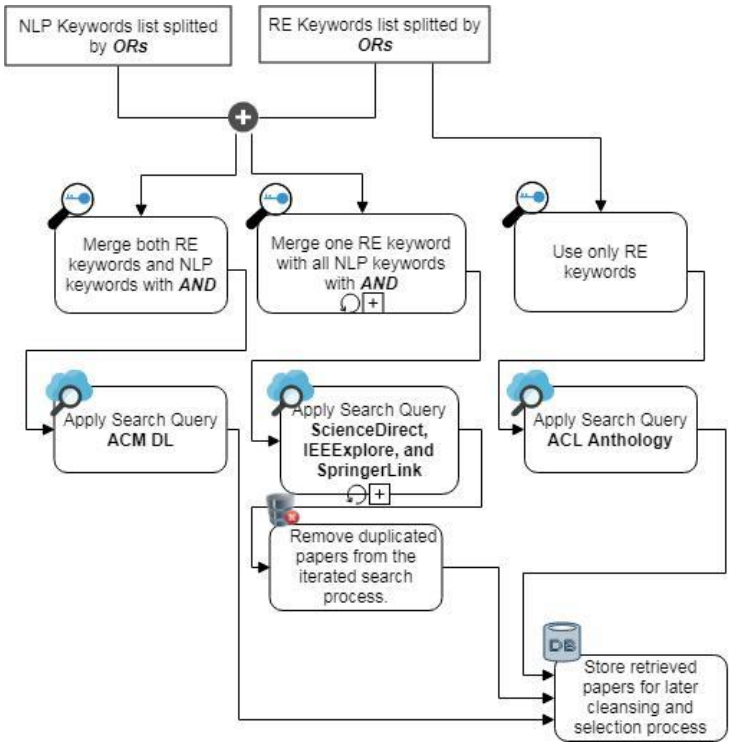

Figure 1 SMS Searching Methods.

Examples of the search strings for each method are:

1.1 Method 1: Direct Method

This method was used with ACM only, since it supports such long search query. The used search query is:

```
("requirements engineering"OR"Requirements elicitation" OR
"requirements analysis"OR"requirements
specification"OR"requirements modeling"OR" requirements
validation"OR"requirements verification"OR"requirements
management"OR"requirements traceability"OR"requirements
classification"OR"requirements document") AND

("Natural Language Processing"OR"NLP"OR"statistical
NLP"OR"machine learning"OR"deep
learning"OR"extraction"OR"retrieval"OR"text mining"OR"text
analysis"OR"linguistic instruments"OR"linguistic approaches")
```

1.2 Method 2: Iterative Method

This method was used with IEEExpoler, SpringeLink and ScienceDirect. Those DLs have limited search terms, and the only way to apply the search string is by breaking down the search string into several queries, for example:

```
("requirements engineering") AND

("Natural Language Processing"OR"NLP"OR"statistical
NLP"OR"machine learning"OR"deep
learning"OR"extraction"OR"retrieval"OR"text mining"OR"text
analysis"OR"linguistic instruments"OR"linguistic approaches")
```

1.3 Method 3: Partial Method

This method was used with ACL, a journal for computational linguistic studies. The journal website does not have advanced searching features for (e.g., applying filters or complex logical operators). So, we used only the parts related to RE (i.e., Part I), because all of the studies indexed by this journal applied some sort of NLP techniques and tools.

```
("requirements engineering"OR"Requirements elicitation" OR
"requirements analysis"OR"requirements
specification"OR"requirements modeling"OR" requirements
validation"OR"requirements verification"OR"requirements
management"OR"requirements traceability"OR"requirements
classification"OR"requirements document")
```

2. Searching Results

After the application of the three methods described in Section 1, we went through the inclusion and exclusion criteria *manually*, and in period of 12 days, we got an overall results of **803** studies which are related to NLP4RE out of **11,406** studies.

We saved the searching results into five independent EndNote libraries The details of the searching results per DLs are shown in Table 1. However, those are not the final numbers of the related studies since we need to verify the studies selection by the other team and then combine the studies from all the DLs into a single repository and remove the duplicated studies 'if any'. For the related reviews (i.e., Exclusion no. 4), we saved them into another folder to be validated by Dr Liping, as she requested.

Table 1. The details of the search results and methods

No.	Criteria	ACM (WA)	ScienceDirect (WA)	ACL Anthology (WA)	SpringerLink (MU)	IEEExpolre (MU)
E1	Exclude tables of contents, editorials, white papers, commentaries, extended abstracts, communications, books, tutorials, non-peer reviewed papers, and duplicated studies.	**Retrieved: 440** **Duplicated: 22** **Excluded: 23** **Included: 395**	**Retrieved: 5,100** **Duplicated: 2,536** **Excluded: 50** **Included: 2,514**	**Retrieved: 53** **Duplicated: 1** **Excluded: 0** **Included: 52**	**Retrieved: 5,135** **Duplicated: 930** **Excluded: 892** **Included: 3313**	**Retrieved: 678** **Duplicated: 76** **Excluded: 16** **Included: 586**
E2	Exclude short papers (i.e., papers that have less than 6 pages).	**Excluded: 108** **Included: 287**	**Excluded: 68** **Included: 2,446**	**Excluded: 15** **Included: 37**	**Excluded: 168** **Included: 3145**	**Excluded: 157** **Included: 429**
E3	Exclude papers written in a language that is not English.	**Excluded: 0** **Included: 285**	**Excluded: 1** **Included: 2,445**	**Excluded: 0** **Included: 37**	**Excluded: 9** **Included: 3136**	**Excluded: 1** **Included: 428**
E4	Exclude review papers (i.e., papers that report secondary research), but keep them separately for further consideration as related work. (These studies are should be saved into another folder for later inspection).	**Excluded: 8 (related: 1)** **Included: 276**	**Excluded: 176 (related: 29)** **Included: 2,240**	**Excluded: 0 (related: 0)** **Included: 37**	**Excluded: 133 (related: 45)** **Included: 3003**	**Excluded: 10 (related: 7)** **Included: 418**
E5	Exclude papers if their full text is not available on the searched database and cannot be found on the Internet.	**Excluded: 2** **Included: 274**	**Excluded: 5** **Included: 2,240**	**Excluded: 0** **Included: 37**	**Excluded: 1** **Included: 3002**	**Excluded: 0** **Included: 418**
E6	Exclude papers that are clearly not relevant to NLP4RE.	**Excluded: 168** **Included: 106**	**Excluded: 1946** **Included: 294**	**Excluded: 25** **Included: 12**	**Excluded: 2852** **Included: 150**	**Excluded: 178** **Included: 240**