Chapter 1 Methodology

Underpinned by experts' evaluation and data, the 2019 Global Engineering Fronts project adopts multi-round interactions between experts and data for iterative research and analysis, realizing a deep integration of the experts' subjective judgments and objective data analyses. This project selected 93 global engineering research fronts and 94 global engineering development fronts in 2019, with 28 engineering research fronts and 28 engineering development fronts being listed as the current focus of interpreting according to principles, such as development prospects and the degree of attention. The number distribution of engineering fronts of the nine fields is shown in Table 1.1.

The specific research methods consist of three stages, namely data exchange, data analysis, and expert review. During the data exchange stage, the interaction of domain experts and library and information experts plays a significant role in defining the scope of data mining. In the data analysis stage, research hotspots and ThemeScape maps featured by data are obtained through clustering methods, and engineering hotspots are determined through expert research. In the expert review stage, the results of fronts are obtained through methods such as expert discussions and questionnaires. In addition, to fix the problem of insufficient research data owing to algorithm limitations or lags in data mining, experts from different fields were encouraged to check the results of the data analysis to fill in the gaps and nominate the engineering fronts. The specific operation procedure is shown in Figure 1.1, in which green and purple boxes indicate data analysis and expert research steps, respectively.

1 Identification of engineering research fronts

In this report, the identification of the engineering research fronts is performed mainly in the following two ways. The first is defining the literature clustering theme through the clustering method of co-citation according to the SCI journal papers and data of conference proceedings from the *Web of Science Core Collection* of *Clarivate Analytics*. The second is defining the engineering research fronts through expert nomination. Alternative engineering research fronts that were identified through expert argumentation and refinement went through questionnaires and multiple rounds of expert discussions, yielding approximately ten engineering research fronts in each field.

1.1 Acquisition and preprocessing of paper data

Clarivate Analytics mapped the fields of *Web of Science* and the nine academic division fields of the CAE and obtained a list of journals and conferences in each field. After the correction

Table 1.1 Number distribution of engineering fronts of the nine fields

Field	Number of engineering research fronts	Number of engineering development fronts
Mechanical and Vehicle Engineering	10	10
Information and Electronic Engineering	10	10
Chemical, Metallurgical, and Materials Engineering	11	12
Energy and Mining Engineering	12	12
Civil, Hydraulic, and Architectural Engineering	10	10
Environmental and Light Textile Engineering	10	10
Agriculture	10	10
Medicine and Health	10	10
Engineering Management	11	10
Total	93	94

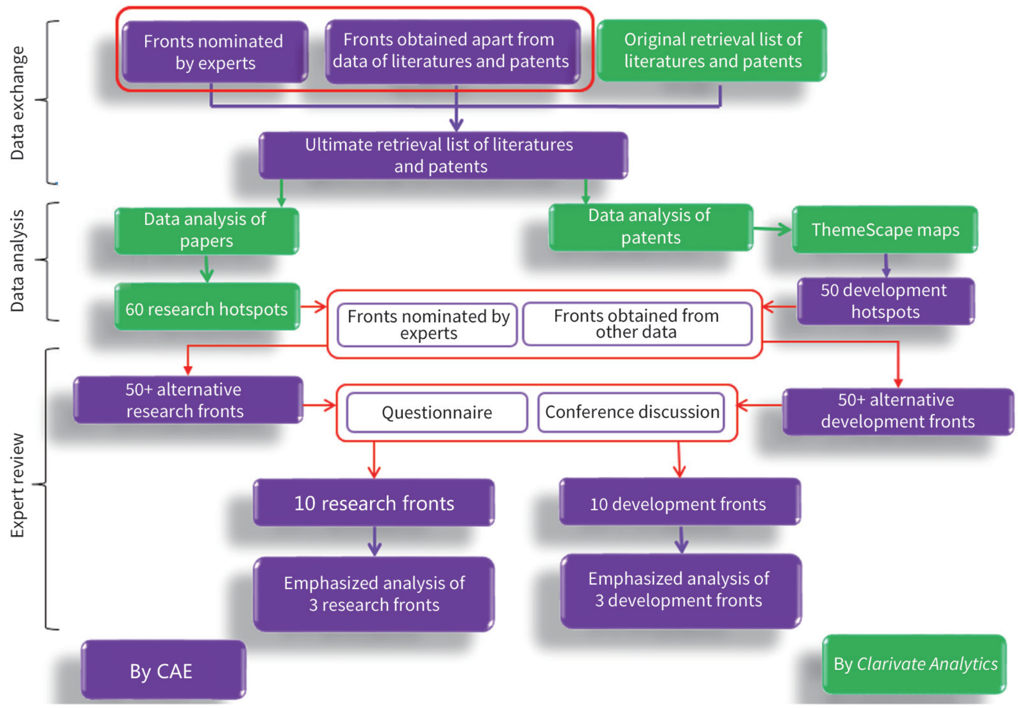

Figure 1.1 Operation procedure of Global Engineering Fronts project

and supplementation by domain experts, the sources for data analysis in the nine fields were determined to be 10 817 journals and 24 330 Conferences. In addition, for articles from 70 journals with integrative subjects (such as *Science*), the field of each article was identified according to the subjects of its references. In this way, the articles and conference papers published between 2013 and 2018 were retrieved (the cut-off date of the citations was February 2019).

For each field, *Clarivate Analytics* comprehensively considered the differences between journals and conferences, the publication year, etc. Next, the paper lists mentioned above were retrieved and extracted. By processing journals and conference proceedings separately, papers with high impact and rank among the top 10% of the citations were selected as the original data set for the analysis of the research hotspots, as shown in Table 1.1.1.

1.2 Acquisition and selection of literature clustering topics

Through the citation clustering analysis of the top 10% highly-cited papers in the above nine data sets, all the literature clustering topics in the nine fields were obtained. The topics of papers published during the year 2017–2018 were selected according to the number of core papers, the total number of citations, and the proportion of consistently cited papers. After that, 25 different literature topics were obtained. In addition to the criteria for the selection of the topics of papers published in 2017–2018, the papers published before 2017 were selected according to the mean publication year of core publications and the proportion of consistent citations. As a result, 35 different literature topics were extracted. Overlapping topics were replaced by topics that did not intersect with other fields. In addition, subjects that were not covered by clustering topics were extracted separately according to keywords. Finally, 806 literature clustering topics in the nine fields were obtained (Table 1.2.1).

1.3 Expert review

The identification of the research fronts by the experts is an important supplement to data mining. In the data exchange stage, domain experts proposed research fronts issues, and library and information experts codified these issues for data

Table 1.1.1 Number of journals and conferences in each field and number of top 10% highly-cited papers

No.	Field	Number of journals	Number of conferences	Number of top 10% highly-cited papers
1	Mechanical and Vehicle Engineering	457	1 768	38 676
2	Information and Electronic Engineering	986	9 632	109 507
3	Chemical, Metallurgical, and Materials Engineering	1 128	2 313	219 081
4	Energy and Mining Engineering	226	785	440 641
5	Civil, Hydraulic, and Architecture Engineering	359	512	28 384
6	Environmental and Light Textile Engineering	1 003	605	93 524
7	Agriculture	1 575	975	105 523
8	Medicine and Health	4 328	7 059	392 142
9	Engineering Management	755	681	32 927

Table 1.2.1 Statistics of co-citation clustering results in each field

No.	Field	Number of topics	Number of top 10% highly-cited papers	Number of alternative engineering research hotspots
1	Mechanical and Vehicle Engineering	6 720	29 960	138
2	Information and Electronic Engineering	16 816	76 015	67
3	Chemical, Metallurgical, and Materials Engineering	26 563	116 361	68
4	Energy and Mining Engineering	10 624	47 860	100
5	Civil, Hydraulic, and Architectural Engineering	5 594	25 867	115
6	Environmental and Light Textile Engineering	18 486	80 850	90
7	Agriculture	7 663	33 909	81
8	Medicine and Health	46 264	203 487	63
9	Engineering Management	4 240	18 321	84

mining, which was an important part of the data sources analysis. In the data analysis stage, domain experts provided keywords, representing papers or journals for the subjects that did not belong to the clustering topics, which were used to support the *Clarivate Analytics*' search. In the expert review stage, the domain experts compared and checked the literature clustering results provided by *Clarivate Analytics*. Clustering topics that did not appear in the results of data mining but were considered important by experts underwent a second round of nomination and were supported by data provided by the library and information experts. Finally, domain experts merged, revised, and refined the engineering research front topics obtained through data mining and expert nomination and selected 93 engineering research fronts through network questionnaires, academic questionnaires, and multiple rounds of discussion.

2 Identification of engineering development fronts

2.1 Selection of engineering development hotspots

The identification of the engineering development fronts is performed mainly in the following two ways. In the first method, the *Derwent Innovation* patent database of *Clarivate Analytics* was considered as the original data source. The matching relation between the patent classification of *Derwent* and the specialty division criteria system of the CAE's academic divisions was used to obtain the primary data for the analysis. Then, 53 subjects of the nine fields with at least 10 000 citations were clustered, and 53 ThemeScape maps were obtained. The domain experts interpreted the alternative engineering development fronts from these maps, while analyzing the field of engineering management as a separate subject group. The second approach involved nomination by expert or patent analysis by small peer. The alternative development fronts obtained through the above two methods went through questionnaire surveys and several special seminars. As a result, approximately ten engineering development fronts in each field were identified.

2.2 Acquisition and interpretation of the ThemeScape maps

Clarivate Analytics established the matching relation between the Derwent Manual Codes and the specialty division criteria system of the CAE's divisions. Then, the scope of the patent data retrieval and search strategies in the nine fields was determined. Domain experts deleted, supplemented, and improved the Derwent Manual Codes to determine the patent retrieval criteria of the 53 specialty groups. The retrieved patents were published between 2013 and 2018, and the cut-off date of the citations was February 2019.

Based on the *Derwent Innovation* patent platform retrieval, the annual average of the citation number and technical coverage width indicators were considered comprehensively, and the topics of the top 10 000 highly-cited patents, corresponding to each specialty group, were selected. Fifty-three patents from the ThemeScape maps were obtained by considering the semantic similarity of the patents' text, which is effective in displaying the distribution of the engineering development techniques.

Experts from various fields, with the assistance of library and information experts, selected the engineering development fronts from ThemeScape maps, merged similar fronts, and determined the final development fronts. Finally, they decided the alternative engineering development fronts of each specialty group. In addition, to avoid assessing the patent data mining merely by figures and facts, field group experts interpreted the data from patents with few citations and poor correlation in the ThemeScape maps.

2.3 Nomination of development fronts by experts

To compensate for the limitations of the algorithms or the lags in data mining, which might extract inappropriate development fronts, domain experts were encouraged to check the result and identify new development fronts. In the data exchange stage, domain experts proposed keywords and descriptions of development fronts, and library and information experts codified these fronts for patent retrieval, which formed an important part of the data source. In the data analysis stage, the results were double-checked and the ThemeScape map deviation was corrected to identify the fronts that were unpopular or marginal, which were then overlooked by the statistical data. In the expert review stage, domain experts made necessary further nominations for the fronts that were not listed in the data mining results but were considered to be important. This procedure was supported by the data provided by the library and information experts. In the end, domain experts merged, revised, and refined the data for the identification of the engineering development fronts obtained from data mining and expert nomination and acquired 351 candidate development fronts. Through network and academic questionnaires and multiple rounds of discussion, 94 engineering development fronts were selected.

3 Terminologies

Fronts: "Fronts" in global engineering research fronts refers to the main directions that should be followed for the development of future engineering technology. The front is an important guideline for cultivating innovation in engineering disciplines.

Literature (Papers): This includes peer-reviewed and published journal articles, reviews, and conference papers retrieved from *Web of Science*.

High-impact papers: The top 10% papers in terms of the number of citations, published in the same year and belonging to the same subject.

Literature clustering topic: A combination of topics and keywords obtained through a co-citation clustering analysis of high-impact papers.

Core papers: High-impact papers related to the engineering research fronts.

Citing papers: Collection of papers that have cited core papers.

Mean publication year: Average publication years for all papers in the literature clustering themes.

Citation velocity: An indicator used to measure the growth rate of the cumulative number of citations during a certain period of time. In this study, the citation velocity of each paper begins with the month of publication, and the cumulative number of citations per month have been recorded.

Consistently cited papers: Papers in the top 10% in terms of citation velocity.

ThemeScape map: A themed landscape representing the overall outlook of a specific industry or technical field. It is a visual presentation in the form of a map obtained by analyzing the semantic similarity of patents to gather the patents of related technologies.

Technical coverage width: It is measured by the number of Derwent Manual Codes to which each patent family belongs. This indicator can reflect the breadth of technology coverage of each patent.

Specialty division criteria system of the CAE's academic divisions: This includes 53 specialized fields covered by nine academic divisions of engineering science and technology. It is determined in accordance with the *Academic Divisions and Specialty Division Criteria of the Chinese Academy of Engineering for the Election of Academicians (for Trial Implementation)*.