Part A Methodology

An engineering front refers to the key direction that is forward-looking, leading, and exploratory. It has a major influence and a leading role in the future development of engineering science and technology and serves as an important guide for cultivating the capabilities for innovation in the field of engineering science and technology. The front is focused on the theoretical research or application development of engineering science and technology. Engineering fronts are divided into engineering research fronts and engineering development fronts. In this research, the engineering front identification is based on public data and expert research; therefore, it does not involve nonpublic domains.

Underpinned by experts' evaluation and data, the 2021 Global Engineering Fronts project has adopted multiround interactions between experts and data for iterative research and analysis, realizing a deep integration of judgments of experts and data analyses. In 2021, 93 global engineering research fronts and 93 global engineering development fronts are selected, with 28 engineering research fronts and 28 engineering development fronts being listed as the key focus of interpretation. The distribution of engineering research and engineering development fronts among the nine fields is shown in Table 1.1.

The research of fronts consists of three stages: data preparation, data analysis, and expert review. In the data preparation stage, domain, library, and information experts revise the initial literature and patent data to clarify the scope of data mining. In the data analysis stage, the co-citation clustering method is used to obtain clustered literature topics and ThemeScape patent maps. In the expert review stage, the fronts are gradually selected and determined through patent map interpretation, expert panel discussions, questionnaire surveys, and other methods. Then the list of the Top 10 fronts is modified, and the front naming is improved based on the performance of the front in literature or patent data. To address the problem of lacking of novelty due 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 engineering fronts. A flowchart of the operation procedure of the Global Engineering Fronts project is illustrated in Figure 1.1, in which the green, purple, and red boxes indicate the data analysis, expert research, and multiround iterative interactions between experts and data, respectively.

1 Identification of engineering research fronts

The identification of the engineering research fronts is performed in two steps. The first step involves determining the clustered literature topics through the clustering method of co-citation according to the SCI journal papers and data of

Table 1.1 Distribution of engineering research and engineering development fronts among 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	11
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	10	10
Total	93	93

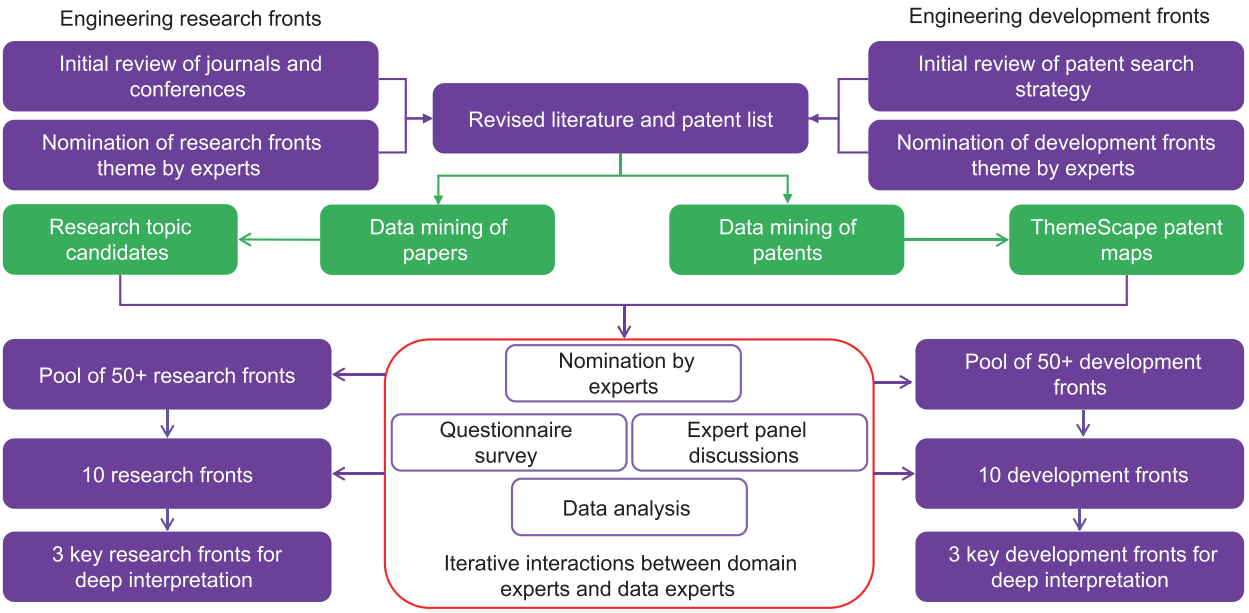

Figure 1.1 Operation procedure of the Global Engineering Fronts project

conference proceedings collected from the Web of Science Core Collection of Clarivate. The second step is defining the engineering research fronts through expert nomination. Alternative engineering research fronts that were identified through expert argumentation and refinement went through questionnaire surveys and multiple rounds of expert discussions, yielding 93 engineering research fronts in the nine fields.

1.1 Acquisition and preprocessing of paper data

Clarivate 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 and supplementation by domain experts, the sources for data analysis in the nine fields were determined to be 12 215 journals and 44 153 conferences. In addition, for articles from 72 multidisciplinary sciences journals such as *Nature* and *Science*, the field of each article was reassigned to the most relevant subject area according to the subjects cited in its references. Accordingly, the articles and conference papers published between 2015 and 2020 were retrieved (the cut-off date of the citations was January 2021).

For each field, Clarivate comprehensively considered the differences between journals and conferences, the publication year, and so on. Next, the list of aforementioned papers was retrieved and extracted. By processing journals and conference proceedings separately, papers with high impact that are ranked among the top 10% of the citations were selected as the original dataset for the analysis of the research hotspots, as shown in Table 1.1.1.

1.2 Mining of clustered literature topics

Through the co-citation clustering analysis of the top 10% highly cited papers in the aforementioned nine data datasets, all the clustered literature topics in the nine fields were obtained. The topics of papers published during 2019–2020 were selected according to the number of core papers, total number of citations, and proportion of consistently cited papers. Thereafter, 25 different literature topics were obtained. The topics of the papers published before 2019 were selected according to the mean publication year of core publications and the proportion of consistent citations. Consequently, 35 diverse 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 by keywords. Finally, 775 clustered literature topics in the nine fields were obtained (Table 1.2.1).

1.3 Determination and interpretation of research fronts

While processing and mining the paper data, domain experts put forward research front issues by a comprehensive analysis of data pertaining to science and technology news and national strategic layouts of different countries, and integrated them into each stage of front determination.

In the data preparation stage, the library and information experts transform the front research questions raised by the domain experts into search formulas, which are an important part of the initial data source. In the data analysis stage, for subjects that are not covered by clustered literature topics, the domain experts provide keywords, representative papers, or representative journals to support Clarivate for customized search and mining. In the expert review stage, the domain experts will check for omissions based on the clustered literature results provided by Clarivate and conduct the second round of nomination for fronts that do not exist in the data mining results but are considered important. Library and information experts provide data support. Finally, the domain experts merge, revise, and refine the engineering research front topics obtained through data mining and expert nomination. Subsequent to questionnaire surveys and multiple rounds of conference discussions, approximately 10 engineering research fronts were selected for each field.

In each field, three key research fronts were selected according to the development prospects and the significance.

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	521	2 779	73 481
2	Information and Electronic Engineering	987	18 590	204 705
3	Chemical, Metallurgical, and Materials Engineering	1 188	4 068	274 485
4	Energy and Mining Engineering	616	2 338	115 816
5	Civil, Hydraulic, and Architectural Engineering	576	1 154	63 930
6	Environmental and Light Textile Engineering	1 345	1 288	207 518
7	Agriculture	1 484	1 093	204 873
8	Medicine and Health	4 685	11 583	476 629
9	Engineering Management	813	1 260	50 986

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

No.	Field	Number of topics	Number of top 10% highlycited papers	Number of alternative engineering research hotspots
1	Mechanical and Vehicle Engineering	8 158	33 822	114
2	Information and Electronic Engineering	20 495	88 795	67
3	Chemical, Metallurgical, and Materials Engineering	28 481	117 286	66
4	Energy and Mining Engineering	12 763	54 261	91
5	Civil, Hydraulic, and Architectural Engineering	7 223	31 099	102
6	Environmental and Light Textile Engineering	22 872	94 186	94
7	Agriculture	22 006	89 460	92
8	Medicine and Health	49 923	211 212	67
9	Engineering Management	5 147	21 001	82

Authoritative experts in the front direction were invited to interpret the fronts in detail from the perspectives of national and institutional layout, cooperation networks, development trends, and R&D priorities.

2 Identification of engineering development fronts

The identification of the engineering development fronts is primarily performed using two methods. First, based on the Derwent Innovation patent database of Clarivate, the top 10 000 patent families of 53 subjects in the nine fields with high citations were clustered, and 53 ThemeScape maps were obtained. The domain experts interpreted the alternative engineering development fronts from these maps. The second approach involves nomination by expert or patent analysis by small peer group. The alternative development fronts obtained through these two methods went through questionnaire surveys and several special seminars. Consequently, approximately 10 engineering development fronts were identified in each field.

2.1 Acquisition and preparation of the ThemeScape maps

In the data preparation stage, based on the Derwent Innovation patent database, Clarivate developed the initial patent data retrieval scope and search strategies for the 53 disciplines of the nine fields using the Derwent World Patents Index (DWPI) Manual Codes, International Patent Classification numbers, United States Patent Classification numbers , and other patent classification numbers, and specific technical keywords. Domain experts deleted, supplemented, and improved the DWPI Manual Codes to determine the patent retrieval criteria; further, the nominated alternative front topics were selected, which were then transformed into patent search formulas by library and information experts. Clarivate integrated the above two parts of the search formulas, determined the patent search formulas of the 53 disciplines, searched in the “enhanced patent data - DWPI and DPCI (Derwent Patent Citation Index)” collection, and obtained the patent literature of the corresponding disciplines. The retrieved patents were published between 2015 and 2020; the cut-off date of the citations was January 2021.

To further concentrate patent literature, the millions of patent documents were screened according to the annual average number of citations and technical coverage width indicators, thereby obtaining the top 10 000 patent families in each discipline.

2.2 Mining of patent topics

Semantic similarity analysis of patent texts were conducted for the top 10 000 highly cited patents on 53 disciplines in the nine fields. Based on literature topic clustering using DWPI titles and abstracts, 53 ThemeScape patent maps were obtained, which effectively display the distribution of the engineering development techniques and show the overall technical information of the collected patents in the form of keywords.

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 selected the alternative engineering development fronts of each specialty group. To avoid missing emerging fronts, domain experts interpreted the data from patents with few citations and poor correlation in the ThemeScape maps.

2.3 Determination and interpretation of development fronts

While processing and mining the patent data, domain experts identified issues on development front based on a comprehensive analysis of other data, such as science and technology news and national strategic layouts of different countries, and integrated them into each stage of front determination.

In the data preparation stage, the library and information experts transformed the key front issues raised by the domain experts into patent search formulas as an important part of the basic dataset. In the data analysis stage, the domain experts conducted the second round of front nomination

to supplement the emerging technology points that are significant, but have been submerged in data mining with few patents. In the expert review stage, the domain experts studied highly cited patents, and the library and information experts assisted them in interpreting patent maps from multiple perspectives, such as “peaks” and “blue oceans”. Finally, the domain experts merged, revised, and refined the interpreted results of the patent maps and fronts nominated by experts to obtain candidate engineering development fronts, and then selected approximately 10 engineering development fronts in each field through questionnaire surveys or multiple rounds of seminars.

In each field, three key development fronts were selected according to the development prospects and the significance. Authoritative experts in the front direction were invited to interpret the fronts in detail from the perspectives of national and institutional layout, cooperation networks, development trends, and R&D priorities.

3 Terminologies

Publications/Papers: This includes peer-reviewed and published journal articles, reviews, and conference papers retrieved from Web of Science.

High-impact papers: Papers that are in the top 10% in terms of citation frequency are considered to be of high impact, taking into account the year of publication and journal subject category.

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

Core papers: Depending on how the research front is obtained, core papers have two meanings. If the paper originates from a front revised by data mining experts, then the core paper is considered as a high-impact paper. If it comes from a front nominated by domain experts, the core paper is included in the top 10% of papers in terms of citation frequency obtained using the corresponding search formula.

Percentage of core papers: The proportion of core papers in which a country or institution participates among the total number of core papers produced by all countries or institutions.

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

Citation number: The number of times the paper has been cited by the Web of Science Core Collection of Clarivate .

Mean publication year: Average publication years for all papers among the clustered literature topics.

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

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

Highly cited patents: The around top 10 000 DWPI families ranked by the average annual DPCI citations.

Core patents: According to the different ways of obtaining the development front, core patents have two meanings. If it comes from the front of the patent map, the core patent refers to the highly cited patent; if it is from the front nominated by domain experts, the core patent refers to all patent obtained by topic search.

Percentage of published patents: The proportion of published patents in which a country (priority country) or institution participates among the total number of published patents produced by all countries or institutions.

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 value added DWPI information of patents to gather the patents of related technologies.

Technical coverage width: It is measured by the number of DWPI Classes to which each DWPI patent family covers. This indicator can reflect the breadth of the technology coverage of each patent.

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