Part A Methodology

Engineering is the practical activity in which human beings transform the world with the help of science and technology. An engineering front refers to the key direction that is forward looking, leading, and exploratory. It has a major influence and plays 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. According to the innovation stage of the front, engineering fronts are categorized into engineering research fronts which focuses on theoretical exploration and engineering development fronts which focuses on the practical application of engineering science and technology.

The Global Engineering Fronts 2023 project has adopted multi-round interactions between experts and data for iterative research and analysis, realizing a deep integration of judgments of experts and data analyses. In 2023, 93 engineering research fronts and 94 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, the domain, library, and information experts formulate the literature and patent search strategies according to the technological systems to define the scope of data mining. In the data analysis stage, the co-citation clustering method is used to obtain clustered literature topics and the co-word clustering method to obtain the ThemeScape patent maps. In the expert review stage, the fronts are gradually selected and determined through front topic selection, front name revision, expert panel discussions, and other methods. To address the 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 revise, combine, and expand the engineering fronts. A flowchart of the operation procedure of the Global Engineering Fronts project is illustrated

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	11
Medicine and Health	10	10
Engineering Management	10	10
Total	93	94

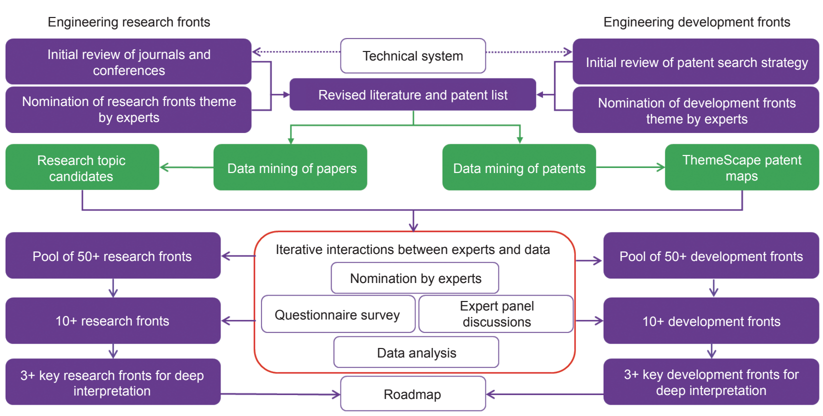

Figure 1.1 Operation procedure of the Global Engineering Fronts project

in Figure 1.1, in which the green, purple, and red boxes indicate the data analysis, expert research, and multi-round iterative interactions between experts and data, respectively.

1 Identification of engineering research fronts

The identification of the engineering research fronts is performed using two methods. The first approach involves determining the clustered literature topics through the co-citation clustering method according to the SCI journal papers and data of conference proceedings collected from the *Web of Science Core Collection*. The second method is calling for the engineering research themes through expert nomination. Candidate 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

Firstly, the mapping relationship between the technology systems of the nine academic divisions of the CAE and Web of Science disciplines was constructed, and the lists of academic journals and academic conferences corresponding to each field were obtained. After the correction and supplementation by domain experts, the sources for data analysis in the nine fields were determined to be 12 696 journals and 54 389 conferences. In addition, for articles from 82 multidisciplinary sciences journals, 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 2017 and 2022 were retrieved (the cut-off date of the citations was January 2023). For each field, the differences between journals and conferences, the publication year, and so on were comprehensively considered. By processing journals and conference proceedings separately, high-impact papers that are ranked among the top 10% of the citations were selected as the original dataset for the analysis of the research fronts, referring to the highly cited papers in Web of Science, as shown in Table 1.1.1.

Table 1.1.1 Overview of data sources in each field

No.	Field	Number of journals	Number of conferences	Number of top high-impact papers
1	Mechanical and Vehicle Engineering	547	3 325	112 450
2	Information and Electronic Engineering	1 010	22 572	228 672
3	Chemical, Metallurgical, and Materials Engineering	1 235	4 850	310 860
4	Energy and Mining Engineering	946	2 893	160 920
5	Civil, Hydraulic, and Architectural Engineering	406	1 428	104 330
6	Environmental and Light Textile Engineering	1 371	1 645	245 146
7	Agriculture	1 463	1 282	190 750
8	Medicine and Health	4 890	14 887	524 174
9	Engineering Management	828	1 507	60 673

1.2 Mining of clustered literature topics

Through the co-citation clustering analysis of the high-impact papers in the aforementioned nine data datasets, all the clustered literature topics in the nine fields were obtained and each clustered topic consists of a certain amount of core papers. The topics of the journal and conference papers published during 2017–2020 were selected according to the number of core papers, total number of citations, mean publication year, and proportion of consistently cited papers. Consequently, 35 diverse literature topics were extracted. The topics of the journal and conference papers published during 2021–2022 were selected according to the number of core papers, total number of citations, and proportion of consistently cited papers. Thereafter, 25 different clustered literature topics were obtained. Overlapping topics were replaced by topics that did not intersect with other fields. In addition, subjects that were not covered by clustered topics were extracted separately by keywords. Finally, 756 clustered literature topics in the nine fields were obtained (Table 1.2.1).

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

No.	Field	Number of clustered topics	Number of core papers	Number of candidate research hotspots
1	Mechanical and Vehicle Engineering	12 404	49 484	98
2	Information and Electronic Engineering	23 529	101 924	68
3	Chemical, Metallurgical, and Materials Engineering	30 400	122 129	66
4	Energy and Mining Engineering	17 173	71 528	89
5	Civil, Hydraulic, and Architectural Engineering	11 051	46 128	129
6	Environmental and Light Textile Engineering	25 575	101 630	89
7	Agriculture	19 645	76 697	76
8	Medicine and Health	52 636	215 909	66
9	Engineering Management	6 137	23 325	75

1.3 Determination and interpretation of research fronts

While processing and mining the paper data, domain experts put forward research front issues via a comprehensive analysis of data pertaining to science and technology news and policies, and integrated them into each stage of front determination.

In the data preparation stage, the library and information experts transform the front research issues raised by domain experts into search strategies, 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, domain experts provide keywords, representative papers, or representative journals for customized search and mining. In the expert review stage, domain experts will check for omissions based on the clustered literature results 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 or four key research 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.

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, the top 10 000 patent families of 53 subjects in the nine fields with highest citations were clustered, and 53 ThemeScape maps were obtained. The domain experts interpreted the candidate engineering development fronts from these maps. The second approach involves nomination by experts. The candidate development fronts obtained through these two methods went through multiple rounds of expert discussion and questionnaire surveys. Consequently, approximately 10 engineering development fronts were identified in each field.

2.1 Acquisition and preprocessing of patent data

In the data preparation stage, based on the *Derwent Innovation* patent database, the initial patent data retrieval scope and search strategies for the 53 disciplines of the nine fields were determined 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 patent search strategies to determine the patent retrieval criteria and nominated candidate front topics, which were then transformed into patent search strategies by library and information experts. The above two parts of the search strategies were integrated to determine the patent search strategies of the 53 disciplines, searched in the enhanced patent data—DWPI and Derwent Patent Citation Index (DPCI) collection, and obtained the patent literature of the corresponding disciplines. The retrieved patents were published between 2017 and 2022; the cut-off date of the citations was January 2023. 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 candidate engineering development fronts of each specialty group. To avoid missing emerging or interdisciplinary 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 fronts based on a comprehensive analysis of other data, such as science and technology news and policies, 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 domain experts into patent search strategies as an important part of the basic dataset. In the data analysis stage, 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, domain experts studied highly cited patents, and library and information experts assisted them in interpreting patent maps from multiple perspectives, such as “peaks”, “blue oceans”, and “islands”. Finally, 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 or four key development fronts were selected according to the development prospects and the significance. Authoritative experts in the front directions 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 Development roadmap

Technology roadmaps are an important tool to depict the development trend of technologies. To strengthen the academic leading role of the engineering fronts, the Global Engineering Fronts 2023 project conducted detailed analysis on the development focuses and trends of key engineering research fronts and key engineering development fronts for each field, and drew a development roadmap in a visual way for each front in the next five to ten years.

4 Terminologies

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

High-impact papers: Papers that rank 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 it originates from a front that is obtained from data mining and revised by experts, 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 strategy.

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*.

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

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

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.

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

Core patents: According to the different ways of obtaining a development front, core patents have two meanings. If they come from the fronts of the patent map, core patents refer to highly cited patents. If they are from the fronts nominated by domain experts, core patents refer to all patents obtained by topic search.

Percentage of published patents: The proportion of core patents in which a country (priority country) or institution participates among the total number of core 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 is specified in the *Specialty Division Criteria of the Academic Divisions of the Chinese Academy of Engineering for Member Election (Trial)*. It refers to 53 specialized disciplines covered by the nine academic divisions of engineering science and technology, including mechanical and vehicle engineering; information and electronic engineering; chemical, metallurgical, and materials engineering; energy and mining engineering; civil, hydraulic, and architectural engineering; environmental and light textile engineering; agriculture; medicine and health; and engineering management.