

Part A Methodology

An engineering front is defined as 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, engineering front identification is based on public data and expert research; hence, it does not involve nonpublic domains.

Underpinned by evaluation of experts and data, the 2020 Global Engineering Fronts project has adopted multiround interactions between experts and data for iterative research and analysis, realizing the deep integration of judgments of experts and data analyses. In 2020, 93 global engineering research fronts and 91 global engineering development fronts were selected, with 28 engineering research fronts and 28 engineering development fronts listed as the key focus for interpretation. The distribution of engineering research and engineering development fronts among the nine fields is presented in Table 1.1.

Research on 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, co-citation clustering method is used to obtain clustered literature topics and ThemeScope 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 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 fill in the gaps and nominate engineering fronts. A flowchart of the operating 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 engineering research fronts is performed in two steps. The first step involves determining the clustered literature topics through clustering method of co-

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	10	10
Energy and Mining Engineering	12	12
Civil, Hydraulic, and Architectural Engineering	10	10
Environmental and Light Textile Engineering	10	10
Agriculture	11	9
Medicine and Health	10	10
Engineering Management	10	10
Total	93	91

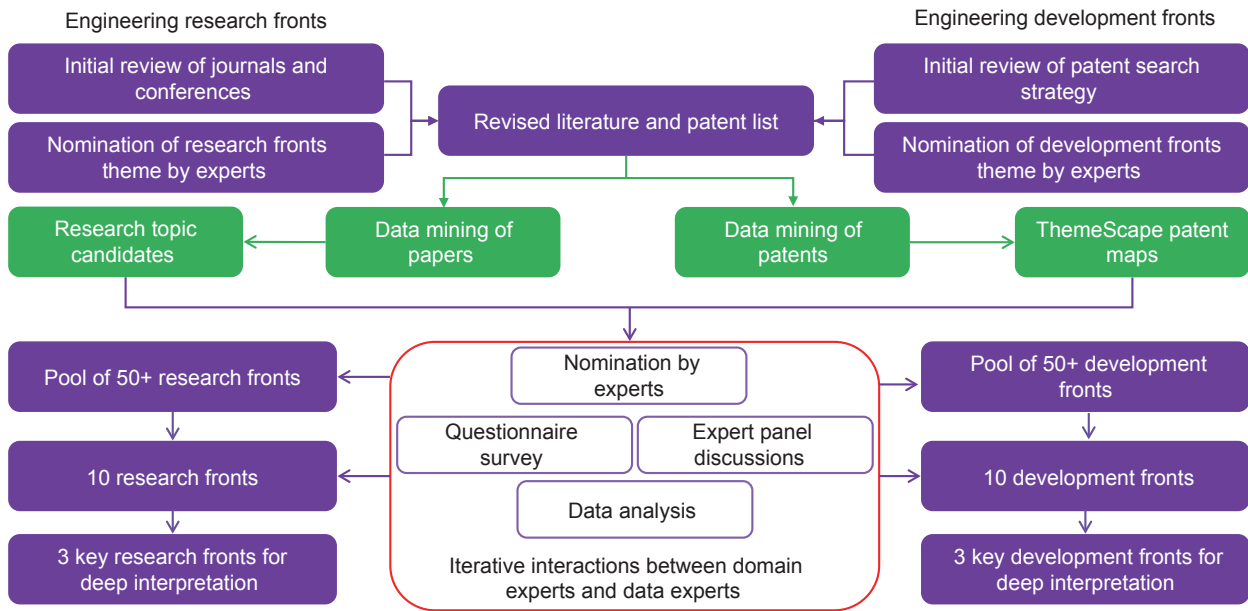


Figure 1.1 Operation procedure of the Global Engineering Fronts project

citation according to the SCI journal papers and data of 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 the *Web of Science* and nine academic division fields of the CAE and obtained a list of journals and conferences in each field. After correction and supplementation by domain experts, the sources for data analysis in the nine fields were determined to be 11 730 journals and 41 734 conferences. For the articles from 70 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 2014 and 2019 were retrieved (the cut-off date of the citations was February 2020).

For each field, *Clarivate* comprehensively considered the differences between journals and conferences, publication year, and so on. Next, the list of aforementioned papers was retrieved and extracted. By processing journals and conference proceedings separately, the papers with high impact that are ranked among the top 10% of highly cited papers were selected as the original dataset for the analysis of research hotspots, as presented 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 datasets, all the clustered literature topics in the nine fields were obtained. The topics of papers published during 2018–2019 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 2018 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, 800 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 present 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 check for omissions based on the clustered literature results provided by *Clarivate* and conduct a second round of nominations 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

Table 1.1.1 Number of journals and conferences in each field and the 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	512	2 641	70 748
2	Information and Electronic Engineering	958	17 418	199 347
3	Chemical, Metallurgical, and Materials Engineering	1 144	3 939	253 221
4	Energy and Mining Engineering	594	2 181	105 674
5	Civil, Hydraulic, and Architectural Engineering	560	1 075	56 402
6	Environmental and Light Textile Engineering	1 326	1 174	186 022
7	Agriculture	1 167	951	70 293
8	Medicine and Health	4 675	11 163	445 940
9	Engineering Management	794	1 192	45 716

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	7 596	31 816	144
2	Information and Electronic Engineering	19 294	85 292	64
3	Chemical, Metallurgical, and Materials Engineering	26 703	111 032	65
4	Energy and Mining Engineering	11 621	49 722	95
5	Civil, Hydraulic, and Architectural Engineering	6 133	27 449	135
6	Environmental and Light Textile Engineering	20 849	86 909	85
7	Agriculture	7 784	32 821	72
8	Medicine and Health	47 145	202 238	65
9	Engineering Management	4 675	19 012	75

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

2 Identification of engineering development fronts

The identification of 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 on 53 subjects in the nine fields with high citations were clustered, and 53 ThemeScape maps were obtained. The domain experts interpreted alternative engineering development fronts from these maps. The second approach involves nominations by an expert or patent analysis by a 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 initial patent data retrieval scope and search strategies for the 53 disciplines in 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 the library and information experts. *Clarivate* integrated the above two parts of the search formulas, determined the patent search formulas of the 53 disciplines, searched the DWPI and Derwent Patent Citation Index databases, and obtained the patent literature of the corresponding disciplines. The

retrieved patents were published between 2014 and 2019; the cut-off date of the citations was February 2020.

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 was 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 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 fronts 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, and yet to show their influence. 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 layouts, 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 the *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, then 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 year 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% based on citation velocity.

Highly cited patents: The top 10 000 patent families cited in each discipline.

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, then the core patent refers to the highly cited patent; if it arises from the front nominated by domain experts, then the core patent refers to all patents obtained by topic search.

Percentage of published patents: The proportion of published patents in which a 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 patents to gather the patents of related technologies.

Technical coverage width: This is measured by the number of DWPI Manual Codes to which each patent family belongs. 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)*.

