

# 第一章 研究方法

工程是人类借助科学技术改造世界的实践活动。工程前沿指具有前瞻性、先导性和探索性，对工程科技未来发展有重大影响和引领作用的关键方向，是培育工程科技创新能力的重要指引。根据前沿所处的创新阶段，工程前沿可分为侧重理论探索的工程研究前沿和侧重实践应用的工程开发前沿。

2023 年度全球工程前沿研究采用专家与数据多轮交互、迭代遴选研判的方法，通过专家研判与数据分析深度融合，在 9 个领域共遴选出 93 个工程研究前沿和 94 个工程开发前沿，并重点解读 28 个工程研究前沿和 28 个工程开发前沿。各领域前沿数量分布如表 1.1 所示。

工程前沿研究基本流程包括三步：数据对接、数据分析和专家研判。数据对接，主要是领域专家和图书情报专家依据各领域的技术体系，制定论文和专利数据检索式，明确数据挖掘的范围；数据分析，主要是通过共被引聚类形成文献聚类主题、共词聚类形成专利地图，获得前沿主题；专家研判，主要是通过前沿主题筛选、前沿名称修订、专家研讨等方法逐步筛选确定前沿。同时，为弥补因数据挖掘算法局限性或数据滞后所导致的前沿性不足，鼓励领域专家结合定量分析结果修正、归并、扩充前沿。研究实施流程如图 1.1 所示，

表 1.1 9 个领域前沿数量分布

领域	工程研究前沿 / 个	工程开发前沿 / 个
机械与运载工程	10	10
信息与电子工程	10	10
化工、冶金与材料工程	11	11
能源与矿业工程	12	12
土木、水利与建筑工程	10	10
环境与轻纺工程	10	10
农业	10	11
医药卫生	10	10
工程管理	10	10
合计	93	94

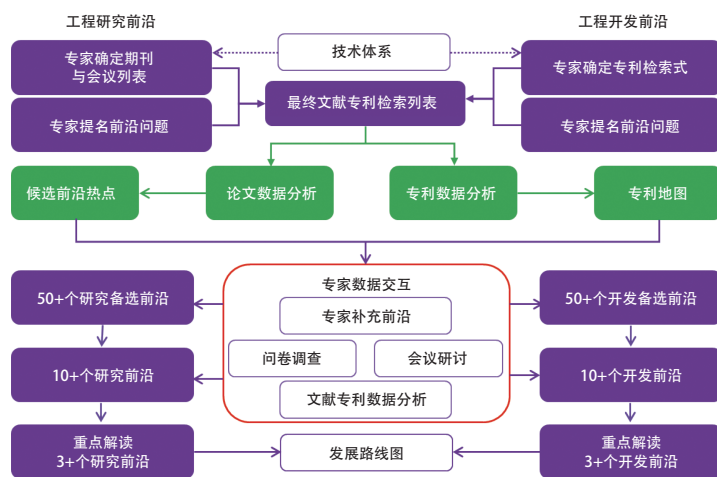


图 1.1 全球工程前沿研究流程

其中绿色部分以数据分析为主，紫色部分以专家研判为主，红色方框为专家与数据多轮深度交互的过程。

## 1 工程研究前沿遴选

工程研究前沿遴选包括两种途径：一是基于 Web of Science 数据库 SCI 期刊论文和会议论文数据，经数据挖掘聚类形成工程研究前沿主题；二是通过专家提名，提出工程研究前沿问题。以上结果经过专家研判论证、提炼得到备选工程研究前沿，再经过问卷调查和多轮专家研讨，遴选得出 9 个领域 93 个工程研究前沿。

### 1.1 论文数据获取与预处理

首先构建中国工程院 9 个学部领域技术体系与 Web of Science 学科的映射关系，获得每个领域对应的学术期刊和学术会议列表。经领域专家审核与修订后，确定本年度重点分析的 9 个领域共计 12 696 种学术期刊和 54 389 个学术会议。此外，针对 82 种综合性国际学术期刊，采用单篇文章归类的方法，即根据文章参考文献的主要归属学科来确定相关期刊中单篇文章的研究领域。

针对每个领域的期刊论文和会议论文，参照 Web of Science 高被引论文确定方法，综合考虑期刊论文和会议论文差别、出版年等因素，筛选出 2017—2022 年期间发表的被引频次位于前 10% 的高影响力论文（截至 2023 年 1 月），作为研究前沿分析的基础数据集。各领域数据源概况如表 1.1.1 所示。

表 1.1.1 各领域数据源概况

序号	领域	期刊 / 种	会议 / 个	高影响力论文 / 篇
1	机械与运载工程	547	3 325	112 450
2	信息与电子工程	1 010	22 572	228 672
3	化工、冶金与材料工程	1 235	4 850	310 860
4	能源与矿业工程	946	2 893	160 920
5	土木、水利与建筑工程	406	1 428	104 330
6	环境与轻纺工程	1 371	1 645	245 146
7	农业	1 463	1 282	190 750
8	医药卫生	4 890	14 887	524 174
9	工程管理	828	1 507	60 673

### 1.2 论文主题挖掘

基于基础数据集，利用共被引方法对高影响力论文进行聚类分析，获得每个领域的前沿聚类主题，每个聚类主题由一定数量的核心论文组成。其中，2017—2020 年出版的期刊论文和会议论文，按照核心论文的数量、总被引频次、平均出版年、常被引论文占比依次筛选，每个领域获得 35 个不相似的文献聚类主题；2021—2022 年出版的期刊论文和会议论文，按照核心论文的数量、总被引频次、常被引论文占比依次筛选，每个领域获得 25 个不相似的文献聚类主题。以上聚类分析中，如果各领域聚类主题有交叉，则递补不交叉的聚类主题，对于没有聚类主题覆盖的学科，按照关键词进行定制检索和挖掘，最终筛选得到 9 个领域 756 个备选研究热点（包括相似和不相似主题），如表 1.2.1 所示。

表 1.2.1 各领域文献聚类结果

序号	领域	聚类主题 / 个	核心论文 / 篇	备选研究热点 / 个
1	机械与运载工程	12 404	49 484	98
2	信息与电子工程	23 529	101 924	68
3	化工、冶金与材料工程	30 400	122 129	66
4	能源与矿业工程	17 173	71 528	89
5	土木、水利与建筑工程	11 051	46 128	129
6	环境与轻纺工程	25 575	101 630	89
7	农业	19 645	76 697	76
8	医药卫生	52 636	215 909	66
9	工程管理	6 137	23 325	75

### 1.3 研究前沿确定与解读

与论文数据处理挖掘同步，领域专家基于专业背景知识，并结合其他综合性科技情报信息，如科技动态、科技政策、新闻报道等进行分析判断，提出工程研究前沿问题，并将其融入前沿确定的每个阶段。

在数据对接阶段，图书情报专家将领域专家提出的研究前沿问题转化为检索式，作为初始数据源的重要组成部分；在数据分析阶段，针对没有文献聚类主题覆盖的学科，领域专家提供关键词、代表性论文或代表性期刊，用于定制检索和挖掘；在专家研判阶段，领域专家对照文献聚类结果进行查漏补缺，对于未出现在数据挖掘结果中而专家认为重要的前沿进行第二轮提名，图书情报专家提供数据支撑。最终，领域专家对数据挖掘和专家提名的工程研究前沿素材进行归并、修订和提炼，而后经过问卷调查和多轮会议研讨，每个领域遴选出 10 余个工程研究前沿。

工程研究前沿确定后，各领域依据发展前景、受关注度选取 3（或 4）个重点研究前沿，邀请前沿方向的权威专家从国家和机构布局、合作网络、发展趋势、研发重点等角度详细解读前沿。

## 2 工程开发前沿遴选

工程开发前沿遴选同样包括两种途径：一是基于 Derwent Innovation 专利检索平台，对 9 个领域 53 个学科组中被引频次位于各学科组前 10 000 的高影响力专利家族进行文本聚类，获得 53 张专利地图，领域专家从专利地图中解读出备选工程开发前沿；二是通过专家提名，提出工程开发前沿问题。在这两种途径获得的备选开发前沿基础上，通过多轮专家研讨和问卷调查，最终遴选产生每个领域 10 余个工程开发前沿。

### 2.1 专利数据获取与预处理

在数据对接阶段，基于 Derwent Innovation 专利数据库，采用德温特世界专利索引(DWPI)手工代码、《国际专利分类表》(IPC 分类)、美国专利局分类体系(UC)等专利分类号和特定的技术关键词，初步构建 9 个领域 53 个学科组的专利数据检索范围及检索策略。领域专家对专利检索式删减、增补和完善，并提名备选前沿主题，图书情报专家将其转化为专利检索式。以上两部分检索式整合后确定 53 个学科组的专利

检索式，在 2017—2022 年“DWPI 和 DPCI（德温特专利引文索引）专利集合”中检索（专利引用时间截至 2023 年 1 月），进而获得相应学科的专利文献。最后对检索得到的百万量级专利文献根据“年均被引频次”和“技术覆盖宽度”指标进行筛选，综合评估得到每个学科前 10 000 个专利家族。

## 2.2 专利主题挖掘

在前面形成的专利家族数据基础上，针对 9 个领域 53 个学科组被引频次位于前 10 000 的高影响力专利家族，开展专利文本语义相似度分析，基于 DWPI 标题和 DWPI 摘要字段进行主题聚类，获得 53 张能快速直观呈现工程开发技术分布的 ThemeScape 专利地图，以关键词的形式展现所聚集专利的总体技术信息。

领域专家在图书情报专家的辅助下，从专利地图中提炼技术开发前沿、归并相似前沿、确定开发前沿名称，得到每个学科组的备选工程开发前沿。同时，为避免遗漏新兴的或交叉的前沿，领域专家重视专利地图中低频次、关联性较低的离群技术点的研判。

## 2.3 开发前沿确定与解读

在专利数据处理与挖掘的同时，领域专家基于专业背景知识并结合其他综合情报信息，如产业动态、科技政策、新闻报道等进行分析判断，提出开发前沿问题，并将其融入前沿确定的每个阶段。

在数据对接阶段，图书情报专家将领域专家提出的开发前沿问题转化为专利检索式，作为基础数据集的重要组成部分；在数据分析阶段，领域专家开展第二轮前沿提名，补充数据挖掘中淹没的专利量少、影响力尚未显现的新兴技术点；在专家研判阶段，领域专家研读高影响力专利，图书情报专家辅助领域专家从“高峰”“蓝海”和“孤岛”等多角度解读专利地图。最终，领域专家对专利地图解读结果与专家提名前沿进行归并、修订和提炼，得到备选工程开发前沿，而后通过问卷调查或多轮专题研讨，遴选出每个领域 10 余个工程开发前沿。

工程开发前沿确定后，各领域依据发展前景、受关注度选取 3（或 4）个重点开发前沿，邀请前沿方向的权威专家从国家和机构布局、合作网络、发展趋势、研发重点等角度详细解读前沿。

## 3 发展路线图

技术路线图是描绘技术未来发展趋势的重要工具。为强化工程前沿的学术引领作用，在本年度研究中，各领域深入分析重点工程研究前沿和重点工程开发前沿的发展方向、发展重点和发展趋势，以可视化的方式绘制该前沿未来 5~10 年的发展路线图。

## 4 术语解释

文献（论文）：包括 Web of Science 中经过同行评议的公开发布的研究性期刊论文、综述和会议论文。

高影响力论文：指被引频次在同出版年、同学科论文中排名前 10% 的论文。

文献聚类主题：对高影响力论文进行共被引聚类分析获得的一系列主题和关键词的组合。

**核心论文：**根据研究前沿的获取方式不同，核心论文有两种含义——如果是来自数据挖掘经专家修正的前沿，核心论文为高影响力论文；如果是来自专家提名的前沿，核心论文为按主题检索被引频次排名前10%的论文。

**论文比例：**某个国家或机构参与的核心论文数量占全部国家或机构产出核心论文数量的比例。

**施引核心论文：**指引用核心论文的文章。

**被引频次：**指某篇论文被 Web of Science 核心合集收录的所有论文引用的次数。

**平均出版年：**指对文献聚类主题中所有文献的出版年取平均数。

**常被引论文：**指引文速度排名前10%的论文。

**引文速度：**是一定时间内衡量累计被引频次增长速度的指标。在本研究中，每一篇文章的引文速度是从发表的月份开始，记录每个月的累计被引频次。

**高影响力专利：**每个学科依据 DPCI 年均被引频次排名前 10 000 的 DWPI 专利家族。

**核心专利：**根据开发前沿的获取方式不同，核心专利有两种含义——如果是来自专利地图的前沿，核心专利指高影响力专利；如果是来自专家提名的前沿，核心专利指按主题检索的全部专利。

**专利比例：**某个国家（作为专利优先权国家）或机构参与的核心专利数量占全部国家或机构产出核心专利数量的比例。

**ThemeScape 专利地图：**基于 Derwent Innovation 中的 DWPI 增值专利信息，通过分析专利文献中的语义相似度，将相关技术的专利聚集在一起，并以地图形式可视化展现，是形象反映某一行业或技术领域整体面貌的主题全景图。

**技术覆盖宽度：**指每个 DWPI 专利家族覆盖的 DWPI 分类的数量。该指标可以体现专利的领域交叉广度。

**中国工程院学部专业划分标准体系：**按照《中国工程院院士增选学部专业划分标准（试行）》确定，包含机械与运载工程，信息与电子工程，化工、冶金与材料工程，能源与矿业工程，土木、水利与建筑工程，环境与轻纺工程，农业，医药卫生，工程管理，共 9 个学部 53 个专业学科。