

# 第一章 研究方法

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南。根据前沿方向的侧重点是工程科技的理论研究还是应用开发，工程前沿分为工程研究前沿和工程开发前沿。本研究中，工程前沿基于公开数据和专家研判得出，不涉及非公开领域。

2020 年度全球工程前沿研究继续在以专家为核心、数据为支撑的原则下，采用专家与数据多轮交互、迭代遴选研判的方法，实现了专家研判与数据分析的深度融合，共遴选出 2020 年度 93 个工程研究前沿和 91 个工程开发前沿，并重点解读了其中的 28 个工程研究前沿和 28 个工程开发前沿。9 个领域的前沿数量分布如表 1.1 所示。

前沿研究按数据准备、数据分析、专家研判 3 个阶段分步实施。在数据准备阶段，领域专家和图书情报专家对初始论文、专利数据进行修订，明确数据挖掘的范围；在数据分析阶段，通过共被引聚类方法获得文献聚类主题和专利地图；在专家研判阶段，

通过专利地图解读、专家研讨、问卷调查等方法逐步筛选确定前沿，并结合前沿在论文或专利数据上的表现进一步调整 Top 10 前沿列表、完善前沿命名。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，鼓励领域专家对定量分析结果查漏补缺，提名前沿。研究实施流程如图 1.1 所示，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 工程研究前沿的遴选

本报告中，工程研究前沿的基础素材主要来自以下两种途径，一是科睿唯安基于 Web of Science 核心合集的 SCI 期刊论文和会议论文数据，通过共被引聚类方法获得文献聚类主题；二是专家提名备选工程研究前沿。两种途径获得的前沿经过专家论证、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。

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	10	10
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	11	9
医药卫生	10	10
工程管理	10	10
合计	93	91

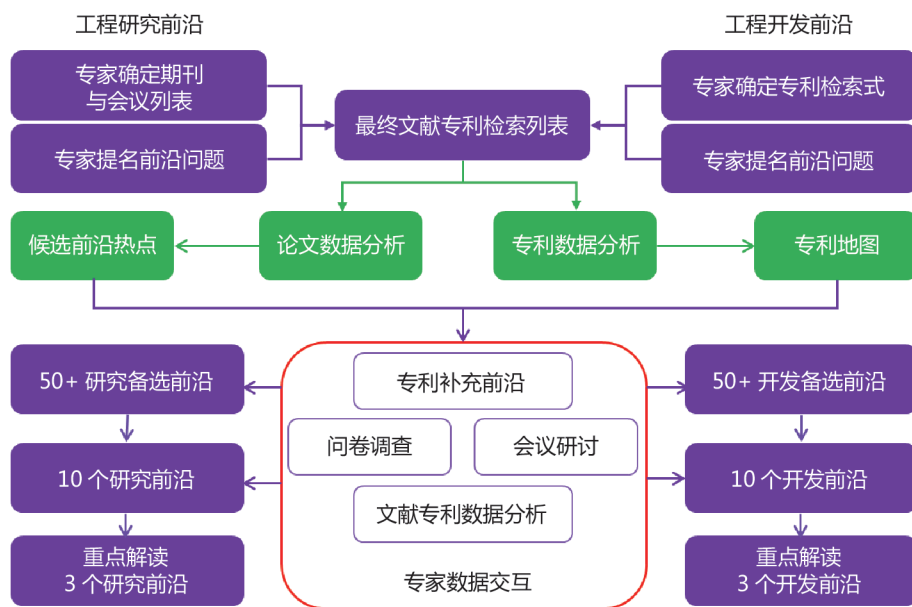


图 1.1 全球工程前沿研究实施流程

### 1.1 论文数据的获取与预处理

科睿唯安将 Web of Science 学科与中国工程院 9 个学部领域建立映射关系，获得每个领域对应的期刊和会议列表，经领域专家修订与补充，确定 9 个领域数据源共计 11 730 种期刊和 41 734 个会议。此外，对于 *Nature* 等 70 种综合学科的期刊，采用单篇文章归类的方法，即根据期刊内单篇文章的参考文献主要归属的学科来定义这篇文章的领域学科。在此基础上，检索得到 2014—2019 年上述期刊和会议收录的论文，论文引用时间截至 2020 年 2 月。

对于每个领域，科睿唯安综合考虑期刊和会议的差别、出版年等因素，对上述文献列表进行检索和数据挖掘，筛选出被引频次位于前 10% 的高影响力论文，作为研究前沿分析的原始数据集，如表 1.1.1 所示。

### 1.2 论文主题挖掘

通过对上述 9 个领域前 10% 的高影响力论文进行共被引聚类分析，得到每个领域的全部文献聚

类主题。对于平均出版年在 2018—2019 年的聚类主题，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，获得 25 个不相似的文献聚类主题；对于平均出版年在 2018 年之前的聚类主题，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，获得 35 个不相似的文献聚类主题。其中，如果各领域聚类主题有交叉，则递补不交叉的聚类主题。此外，对于没有聚类主题覆盖的学科按关键词进行定制检索和挖掘。最终筛选得到 9 个领域 800 个备选研究热点，如表 1.2.1 所示。

### 1.3 研究前沿的确定与解读

在论文数据处理与挖掘的同时，领域专家基于对其他数据如科技新闻、各国战略布局等的综合分析，提出研究前沿问题，并将其融入前沿确定的每个阶段。

在数据准备阶段，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，这是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

**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	512	2641	70 748
2	信息与电子工程	958	17 418	199 347
3	化工、冶金与材料工程	1144	3939	253 221
4	能源与矿业工程	594	2181	105 674
5	土木、水利与建筑工程	560	1075	56 402
6	环境与轻纺工程	1326	1174	186 022
7	农业	1167	951	70 293
8	医药卫生	4675	11 163	445 940
9	工程管理	794	1192	45 716

**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	7596	31 816	144
2	信息与电子工程	19 294	85 292	64
3	化工、冶金与材料工程	26 703	111 032	65
4	能源与矿业工程	11 621	49 722	95
5	土木、水利与建筑工程	6133	27 449	135
6	环境与轻纺工程	20 849	86 909	85
7	农业	7784	32 821	72
8	医药卫生	47 145	202 238	65
9	工程管理	4675	19 012	75

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，用于支撑科睿唯安进行定制检索和挖掘。在专家研判阶段，领域专家对照科睿唯安提供的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 个左右工程研究前沿。

各领域依据发展前景、受关注度选取 3 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 工程开发前沿的遴选

工程开发前沿的基础素材也来自两种途径：一是科睿唯安基于 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 个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；二是专家提名备选工程开发前沿。两种方式获得的备选开发前沿通过多轮专家研讨和问卷调查，获得每个领域 10 个左右工程开发前沿。

### 2.1 专利数据的获取与预处理

在数据准备阶段，科睿唯安基于 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进行删减、增补和完善，并提名备选前沿主题，图书情报专家转化为专利检索式。科睿唯安将以上两部分检索式进行整合，确定 53 个学科组的专利检索式，在 DWPI 和德温特专利引文索引 (DPCI) 数据库中检索，获得相应学科的专利文献。专利检索时间范围为 2014—2019 年，专利引用时间截至 2020 年 2 月。

为了进一步聚焦专利文献，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

对 9 个领域 53 个学科组被引频次位于前 10 000 的高影响力专利开展专利文本语义相似度分析，基于 DWPI 标题和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辅助下，从专利地图提炼技术开发前沿、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，为避免遗漏新兴前沿，领域专家尤其注重专利地图中低频次、关联性较低的技术空白点的解读。

### 2.3 开发前沿的确定与解读

在专利数据处理与挖掘的同时，领域专家基于对其他数据如科技新闻、各国战略布局等的综合分

析，提出开发前沿问题，并将其融入前沿确定的每个阶段。

在数据准备阶段，图书情报专家将领域专家提出的关键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”和“蓝海”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，每个领域遴选出 10 个左右工程开发前沿。

各领域依据发展前景、受关注度选取 3 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 术语解释

**文献 (论文)：**包括 Web of Science 中经过同行评议的公开发表的研究型期刊论文、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，核心论文指高影响力论文；如果是来自专家提名的前沿，核心论文指按主题检索被引频次排前 10% 的论文。图片中的圆点大小根据核心论文数量绘制。

**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献。

**被引频次：**指论文被科睿唯安 Web of Science 核心合集收录的论文引用的次数。

**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。

**引文速度：**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，每一篇文献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。

**常被引论文：**指引文速度排名前 10% 的论文。

**高影响力专利：**每个学科被引频次排名前 10 000 的专利家族。

**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。

**专利地图：**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，是形象地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：**包含中国工程科学技术（含农业、医学）的 9 个学部所涵盖的 53 个专业学科，按照《中国工程院院士增选学部专业划分标准（试行）》确定。

