中国微观经济数据查询系统

使用说明

http://microdata.sozdata.com

北京搜知数据科技有限公司

2019年12月



目录

一、	背景	3
_,	系统架构	4
三、	使用说明	5
3	.1 登录与权限	5
	3.1.1 登录	5
	3.1.2 权限	6
3	.2 专题数据库	7
	3.2.1 工业企业数据库	7
	3.2.2 海关企业数据库1	5
	3.2.3 绿色发展数据库2	0
	3.2.4 创新企业发展数据库2	6
3	.3 跨库检索3	2
	3.3.1 单年匹配3	3
	3.3.2 时间序列匹配3	4



一、背景

经济数据在学术研究中起到重要支撑地位,而且随着经济学研究的量化趋势,经济数据的地位愈发重要。近些年来以经济数据为基础的实证研究大放光彩,目前已成为国内外经济前沿研究的主要形式。美国经济学会的最高奖克拉克奖,在1995年前80%的获得者主要研究领域为理论,20%为实证,而到了2006-2016年之间,仅有33%的获奖者主要研究领域为理论,其余67%均为实证,主要研究领域已由理论研究转为实证研究。

经济数据包含宏观、中观和微观数据,其中微观数据蕴含着丰富的信息,是宝贵的经济社会资源,具有重要的经济价值、社会价值和学术研究价值(许宪春等 2018)。布兰迪斯大学教授 Jefferson 指出"过度依赖汇总数据进行研究可能会模糊很多重要细节······需要依靠微观数据的研究来探悉结果表现以及背后的决定因素"。相对于宏观数据或行业数据,微观数据的优势非常明显:第一,微观面板数据包含了更多信息,例如企业的所有制、规模和出口等状态,这些信息对于企业行为的研究不可或缺;第二,微观面板数据同时包含了时间维度和个体维度,有助于解决计量经济学中的个体异质性问题,更容易保证估计的一致性;第三,微观面板数据增加了观测值个数,估计更有效率。对于产业组织理论、企业理论、公司金融、国际贸易、收入分配和劳动供给等研究领域来说,其研究所用数据主要就是微观数据。实际上,大规模的微观数据已经开始逐渐在学术研究中显示出重要性。对《经济研究》刊发文章统计结果显示,基于微观数据的实证研究文章比例由 2009-2013 年间的 35.8%上升到 2014-2018 年的 42.2%(甘犁、冯帅章, 2019)。

目前国内的微观数据库主要有针对家庭和居民个体的中国家庭收入调查 (CHIP)、中国综合社会调查 (CGSS)、中国健康与养老追踪调查 (CHARLS)、中国家庭金融调查 (CHFS)、中国家庭动态调查 (CFPS)等和针对企业的中国工业企业数据库、中国海关进出口数据库等。



虽然利用微观企业数据产生了一大批前沿学术成果,但以中国工业企业数据库为代表的企业微观数据,由于存在着一些系统误差与数据问题,使用之前,研究者常常需要花费大量精力对数据进行处理,这给使用者带来了极大的困扰。北京搜知数据科技有限公司(EPS 数据)针对研究者的"痛点",经过大量市场调研,花费大量人力进行专业化清洗、处理微观企业数据,打造出"中国微观经济数据查询系统(http://microdata.sozdata.com/)",让使用者不受数据所惑,更专注于科研本身。"中国微观经济数据查询系统力求为以实证研究为代表的经管学科学术研究的繁荣添砖加瓦。

参考文献:

[1] 许宪春, 张钟文, 余航, 叶银丹. 中国政府微观调查数据如何向学术研究开放 [J]. 经济学报, 2018, 5(04):1-13.

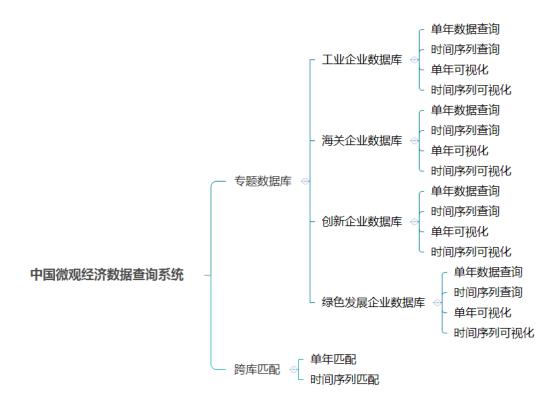
[2] 甘犁, 冯帅章. 以微观数据库建设助推中国经济学发展——第二届微观经济数据与经济学理论创新论坛综述[J]. 经济研究, 2019, 54 (04): 204-208.

二、系统架构

中国微观经济数据查询系统是一款以企业层面的微观数据为基础的,集数据查询、数据匹配、数据可视化于一体的数据查询系统。

中国微观经济数据查询系统目前由工业企业数据库、海关企业数据库、绿色 发展数据库、创新企业数据库四大专题数据库和跨库匹配模块构成。各专题数据 库包含了"单年数据查询"、"时间序列数据查询"、"单年可视化"、"时间序列可 视化"四大部分。同时,为方便用户进行跨学科、跨领域研究,我们增加了"跨 库匹配"模块,实现不同专题数据库间的企业匹配。





三、使用说明

3.1 登录与权限

3.1.1 登录

输入中国微观经济数据查询系统地址 http://microdata.sozdata.com/, 进入登录页面。系统有两种登录方式可供选择: 账户登录与 IP 登录。若用户自行购买了个人账户,则输入用户名和密码,点击"账户登录"按钮即可。若用户所在的机构采购或试用了中国微观经济数据查询系统,则无需输入用户名和密码,直接点击"IP 登录"按钮即可,需要说明的是"IP 登录"仅局限于本机构授权的IP 段范围。

若要购买中国微观经济数据查询系统或在使用过程中有什么问题,可拨打



010-85786021 转 8001(周一-周五 9:00-18:00) 或邮箱 service@sozdata.com 进行反馈。

用户登录 CUSTOMER LOGIN 用户名	
密码	
账户登录 当前IP:	

图 1 登录页

3.1.2 权限

根据用户的权限差异,我们对每个专题数据库下的年度数据进行了权限控制,包含:正式权限、试用权限和未开通权限。其中试用数据后面标注了试,用户可查询此年的数据,但只能下载前50条数据。未开通数据后面标注了 ,用户无法查看该数据库。



图 2 三种访问权限



3.2 专题数据库

目前,系统包含工业企业数据库、海关企业数据库、绿色发展数据库、创新企业数据库四个专题数据库,以后还将陆续增加其他专题数据库。各专题数据库 主要包含了企业基本信息与企业相关统计两部分指标,功能方面包含了单年查询、时间序列查询、单年可视化、时间序列可视化四大模块。

为在企业隐私与学术研究中寻得平衡,我们隐藏掉了企业组织机构代码、名称、通讯地址、法人代表等敏感信息,同时增加了企业标识码字段,用户可通过企业标识码来区分企业,从而达到学术研究的目的。

3.2.1 工业企业数据库

中国工业企业数据库提供了 1998-2013 年规模以上工业企业数据(1998-2006年为全部国有工业企业及年销售收入 500 万元及以上的非国有工业企业,2007-2010年调整为年主营业务收入达到 500 万元及以上的工业法人企业,2011年调整为年主营业务收入达到 2000 万元及以上的工业法人企业)。数据来源于国家统计局,依据《工业统计报表制度》而进行的工业调查统计。统计字段包含企业基本信息、企业生产与财务信息两大类别,其中企业生产与财务信息包含产业活动、企业生产销售及职工情况、资产负债表、损益表、现金流量表等。是研究中国产业与发展问题的重要数据集。

中国工业企业数据库具有如下几个特点:

1、时间跨度长

目前可提供 1998 年-2013 年,共计 16 年的数据,对于学者进行时间序列分析,研究中国制造业及制造业企业的发展变化过程、发展趋势与发展速度,描述不同阶段的发展状态,探究发展规律提供了有力的数据支撑。

2、企业涉及广



从 1998 到 2013 年,中国工业企业数据库一共包括了 400 多万个观测样本。按主营业务收入 (销售额) 在 500 万元以上的"规模以上"标准,每年观测样本数量从 1998 年 16 万家左右递增到 2010 年 40 万家左右。2011 年"规模以上"标准上调至 2000 万元以上后,其观测样本数量由 2011 年的 30 万家左右递增到 2013 年 34 万家左右。16 年观测期内,大约有 90 万家企业出现。是一个巨大的非平衡面板数据。

3、指标数量多

工业企业数据包括企业的两类信息:一类是企业的基本信息,另一类是企业生产销售与财务信息。企业的基本信息包括:法人代码、企业名称、法人代表、联系电话、邮政编码、具体地址、所属行业、注册类型、隶属关系等指标。企业的生产销售与财务信息则包括:工业总产值、工业销售产值、出口交货流动资产、应收账款、长期投资、固定资产、累计折旧、无形资产、流动负债、长期负债、实收资本、主营业务收入、主营业务成本、营业费用、管理费用、财务费用、营业利润、利税总额、广告费、研究开发费、工资总额、福利费总额、增值税、工业中间投入等指标。约有 200 个指标出现在 16 年的观测期中。

近年来,中国工业企业数据已成为国内外学者研究中国企业行为和绩效的主要数据库之一。伴随着计量经济学的引入,基于此数据库产生了很多研究成果并广泛发表在国内外著名的学术期刊上。研究主题涵盖产业组织理论、企业理论、公司金融、转型经济学、 国际贸易、劳动经济学和区域经济学等学科。

中国工业企业数据库模块包括单年数据查询、时间序列查询、单年可视化、时间序列可视化等四部分。

3.2.1.1 单年数据查询

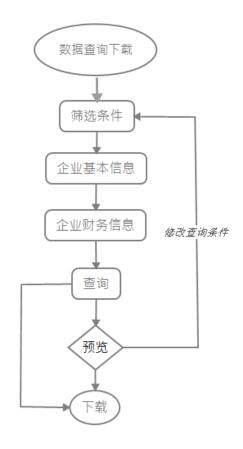
实现对1998-2013年间某一年度的全部企业数据按一定筛选条件进行的查询, 用户可便捷的下载到研究所需的数据内容。





图 3 工业企业数据库单年查询

单年数据查询中查询和下载数据流程如下:





由于每年的企业基本信息与财务信息可能发生变动,导致筛选条件、企业基本信息、企业财务信息等都有可能有所不同,这里以2013年为例说明。

筛选条件包含:企业所在地(省、地级城市两级)、国民经济行业(门、大、中、小四层)、企业登记注册类型、企业隶属关系、企业控股情况、自定义筛选(工业企业标识码、主要业务活动)。当每选中一个筛选条件时,右上角会实时显示"符合筛选条件的企业共计***家"。

企业基本信息包含企业基础信息、企业所属行业与主营业务、企业所属类别 三大类,每类下包含若干字段,字段可多选。企业基本信息或其下的每大类可全 选。基础信息中的"工业企业标识码"是系统对每家工业企业设定的唯一标识, 在不同年份出现的同一家企业具有相同且唯一的编码。

企业财务信息包含生产销售及职工情况、资产负债表、损益表三大类,每类下包含若干字段,字段可多选。企业财务信息或其下的每大类可全选。点击财务字段后面的 icon,可查看该指标的具体含义。

下载页面中可查看所选定的筛选条件与企业基本信息、企业财务信息等输出 字段,若不满意,可点击"修改查询条件"返回查询。

为保证数据下载速度,让下载更为流畅,防止恶意下载情况,我们将大于 5万条的文件进行了分页,每页 5万条(如果可以达到),每次只能下次一页(即 5万条)。用户可在"数据文件"下看到当前要下载文件的文件大小、数据量、输出格式等。也可预览前 50条数据。



数据文件				
当前页数据文件大小:	11.74KB			
当前页数据量:	50000行		*试用权限仅可下载前50行	
輸出格式:	CSV		⑥ 预览	
		12345	6 7 > 共7页	确定
)修改查询条件)下载	载

图 4 数据文件内容

数据文件的下载格式为学界常用的 CSV 格式。

3.2.1.2 时间序列查询

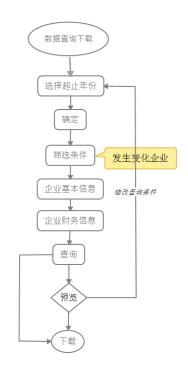
采用"序贯匹配法",对不同年份的企业进行匹配。用户可轻松获得 1998-2013 年任意起止年份间企业匹配后的面板数据。



图 5 工业企业数据库时间序列查询



时间序列查询中查询和下载数据流程如下:



这里大部分与前述 3. 2. 1. 1 单年查询一致,不再赘述。现仅说明几点注意的地方:

- (1) 筛选项中多出了"剔除发生变化企业",由于企业所在地、行业、登记注册类型、隶属关系、控股情况等都可能随着时间变化而变化,若要剔除这部分发生变化的企业,可选择相应的选项。同样的,基于这方面考虑,筛选项中的各筛选项均以企业的最新信息为基础。例如企业 A 在 2001 年所在地在江苏,2002年为上海,2003年为浙江,那么当确定起止时间为 2001-2003年后,企业 A 只在不限省份和浙江省中可以查阅到,在江苏与上海均不可查到。
- (2)企业财务信息中有"缺失情况",这里的意思是对于某些企业财务字段, 在所选的起止时间段内,很可能并不连续,会出现缺失。例如对于 2000-2006 年, 出口交货值这个字段在此时间段内每年均出现,但对于工业增加值,在 2004 年 缺失,其他年份具备。那么只有在"缺失情况"中选择选择"缺失 1 年",工业增加值这个字段才会显示出来。





图 6 缺失情况筛选

(3) 企业财务信息中的字段只能单选。

3.2.1.3 单年可视化

提供 1998-2013 年间每年财务指标的按企业所在地统计(省级数字地图、柱形-折线混合图)、按企业所属国民经济行业大类统计(条形图)、按企业控股情况统计(环形图)、按企业登记注册类型统计(环形图)、按企业隶属关系统计(条形图)。



图 7 工业企业数据库单年可视化



系统默认为有权限的最新年份的全国企业数量数据,用户可选择其他年份、其他统计指标。当用户点击地图上某一个省(自治区、直辖市)时,各图均自动切换到对应的省(自治区、直辖市)。除地图的其他图表,可点击右上角的 , 进行放大。

3.2.1.4 时间序列可视化

提供 1998-2013 年任意起止年份间财务指标的按企业所在地统计(折线图、面积图)、按企业登记注册类型统计(折线图、面积图)、按企业隶属关系统计(堆积柱形图)、按企业控股情况统计(堆积柱形图)。

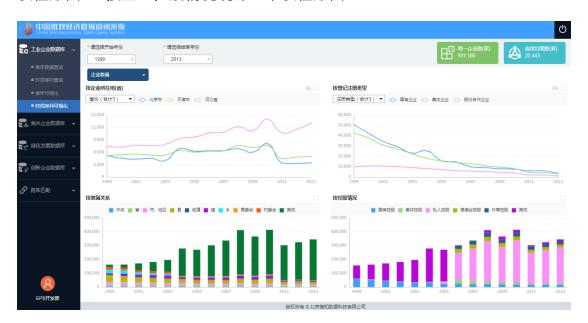


图 8 工业企业数据库时间序列可视化

用户可依次选择起止时间、统计指标,省份、登记注册类型,即可查看所需汇总数据的时间趋势。点击右上角的 ,进行放大。

右上角标注了对应起止时间内的全国"唯一企业数"和"连续出现数",便于用户把握整体情况。



3.2.2 海关企业数据库

海关企业数据库提供了 2000-2016 年每年发生进出口的中国企业,数据来源于进出口企业提交给海关总署的贸易数据,每年记录在 1000 万条以上。统计字段主要有企业基本信息、HS 商品 (8 位码)、收发货地、中转国、产销国、海关口岸、贸易方式、运输方式等,可用于研究中国企业的国际贸易行为。

海关企业数据库包括单年数据查询、时间序列查询、单年可视化、时间序列可视化等四部分。

3.2.2.1 单年数据查询

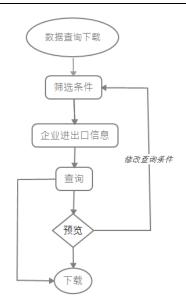
实现对 2000-2016 年间某一年度的全部数据按一定筛选条件进行的查询,用户可便捷的下载到研究所需的数据内容。



图 9 海关企业数据库单年数据查询

单年数据查询中查询和下载数据流程如下:





由于每年的企业进出口信息可能发生变动,导致筛选条件、企业进出口输出 字段等都有可能有所不同,特别是 2000-2006 年与 2007 年(含)之后的字段差 异较大。这里仅以 2006 年为例说明。

筛选条件包含:企业所在地(省、地级城市两级)、商品(HS2、4、6、8 四级)、收发货地(省、地级城市两级)、国别(大洲、国家两级)、关区、企业性质、贸易方式、运输方式、贸易流向(出口、进口)、自定义筛选(海关企业标识码)。当每选中一个筛选条件时,右上角会实时显示"符合筛选条件的企业共计***家"。企业进出口信息包含字段代码和字段名称,二者基本一一对应。在字段名称中会多出金额、数量、数量单位等字段。企业进出口信息字段可多选。企业进出口信息或其下的每大类可全选。下载页面中可查看所选定的筛选条件与企业进出口信息。若不满意,可点击"修改查询条件"返回查询。

为保证数据下载速度,让下载更为流畅,防止恶意下载情况,我们将大于10万条的文件进行了分页,每页10万条(如果可以达到),每次只能下次一页(即10万条)。用户可在"数据文件"下看到当前要下载文件的文件大小、数据量、输出格式等。也可预览前50条数据。



数据文件		
当前页数据文件大小:	13.11KB	
当前页数据量:	100000行	*试用权限仅可下载前50行
輸出格式:	CSV	◎ 預覧
		< 1 2 3 4 5 6 7 80 > 共80页 确定
		◆ 修改查询条件 ▼ 下载 ▼ 压缩为.zip格式下载

图 10 海关企业数据库数据文件内容

数据文件的下载格式为学界常用的 CSV 格式。

3.2.2.2 时间序列查询

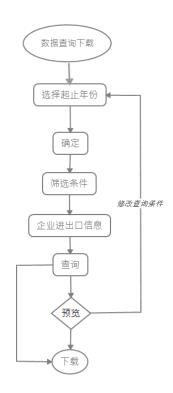
采用专业化清洗与匹配方式,对不同年份的企业进行匹配。用户可轻松获得 2000-2016年任意起止年份间企业匹配后的面板数据。



图 11 海关企业数据库时间序列查询



时间序列查询中查询和下载数据流程如下:



这里大部分与前述 3.2.2.1 单年查询一致,现仅说明几点注意事项:

- (1) 筛选条件的贸易流向中的"2-进出口",意味着该企业既发生了进口, 也发生了出口行为。
- (2)企业进出口信息中的字段代码与字段名称与筛选项关联,当某字段被筛选了,企业进出口信息中的该相关字段才可勾选。
 - (3) 企业进出口信息中的金额与数量,只能单选。

3.2.2.3 单年可视化

提供2000-2016年间每年进、出口额的按国别统计(世界数字地图、条形图)、按企业所在地统计(条形图)、按企业性质统计(环形图)、按 HS 商品统计(矩形树图)、按运输方式统计(柱形-折线混合图)、按贸易方式统计(环形图)。





图 12 海关企业数据库单年可视化

系统默认为有权限的最新年份的出口数据,用户可选择其他年份、进口指标。世界地图为中国当年进/出口最多的 top10 国家/地区。地图左侧为各省进/出口额排序,右侧为中国对各国进/出口额排序,下图分别为按企业性质、贸易方式分的进/出口比例、top20 商品。

3.2.2.4 时间序列可视化

提供2000-2016年任意起止年份间进、出口额的按国别统计(世界数字地图、 条形图)、按企业所在地统计(条形图)、按企业性质统计(堆积柱形图)、按贸 易方式统计(堆积柱形图)、按进出口金额统计(折线图)。





图 13 海关企业数据库时间序列可视化

3.2.3 绿色发展数据库

绿色发展数据库提供了1998-2012年中国工业企业的排放排污和环境治理等信息,数据来源于原中国环保部。统计字段主要有企业基本信息、生产信息、水环境、大气环境,内容涵盖了资源利用类指标(工业用水量、煤炭消费量)、污染排放类指标(工业废水排放量、二氧化硫排放量),污染治理类指标(废水治理设施数、氮氧化物去除量)等数十项指标信息,对于研究中国环境治理、产业结构升级具有重要基础性意义。

绿色发展数据库包括单年数据查询、时间序列查询、单年可视化、时间序列可视化等四部分。

3.2.3.1 单年数据查询

实现对 1998-2012 年间某一年度的全部数据按一定筛选条件进行的查询,用户可便捷的下载到研究所需的数据内容。



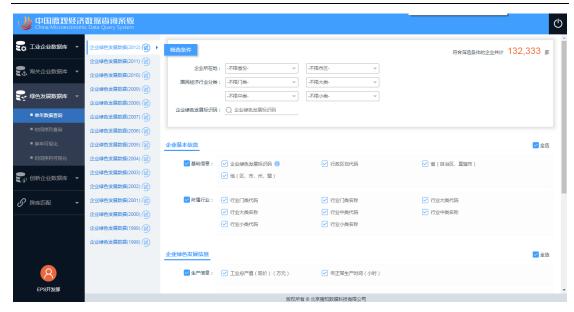
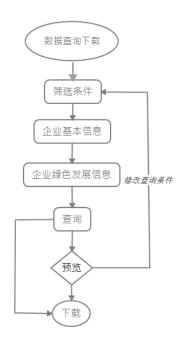


图 14 绿色发展数据库单年数据查询

单年数据查询中查询和下载数据流程如下:



由于每年的企业基本信息与绿色发展信息可能发生变动,导致筛选条件、企业基本信息、企业绿色发展信息等都有可能有所不同,这里以2012年为例说明。

筛选条件包含: 企业所在地(省、地级城市两级)、国民经济行业(门、大、



中、小四级)、自定义筛选(绿色发展企业标识码)。当每选中一个筛选条件时, 右上角会实时显示"符合筛选条件的企业共计***家"。

企业基本信息包含企业基础信息、企业所属行业两大类,每类下包含若干字段,字段可多选。企业基本信息或其下的每个大类可全选。基础信息中的"企业绿色发展标识码"是系统对每家绿色发展企业设定的唯一标识,在不同年份出现的同一家企业具有相同且唯一的编码。

下载页面中可查看所选定的筛选条件与企业基本信息、企业绿色发展信息等输出字段,若不满意,可点击"修改查询条件"返回查询。

为保证数据下载速度,让下载更为流畅,防止恶意下载情况,我们将大于 5万条的文件进行了分页,每页 5万条(如果可以达到),每次只能下次一页(即 5万条)。用户可在"数据文件"下看到当前要下载文件的文件大小、数据量、输出格式等。也可预览前 50条数据。



图 15 绿色发展数据库数据文件内容

数据文件的下载格式为学界常用的 CSV 格式。



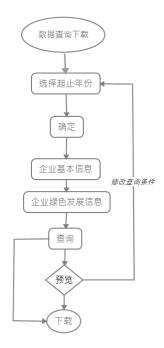
3.2.3.2 时间序列查询

采用专业化清洗与匹配方式,对不同年份的企业进行匹配。用户可轻松获得 1998-2012年任意起止年份间企业匹配后的面板数据。



图 16 绿色发展数据库时间序列查询

时间序列查询中查询和下载数据流程如下:





这里大部分与前述 3.2.3.1 单年查询一致,现仅说明几点事项:

(1)企业绿色发展信息中有"缺失情况",这里的意思是对于某些企业绿色发展字段,在所选的起止时间段内,很可能并不连续,会出现缺失。例如对于2001-2005年,工业粉尘排放量这个字段在2004年缺失,那么只有选择"缺失1年"才能显示出这个指标。

企业绿色发展信息			í	缺失情况	-各年无缺失- ▼
生产信息:	○ 工业总产值(现价)(万元)				各年无缺失 缺失1年 缺失2年
水环境:		○ 其中:新鮮用水量(吨) ○ 废水治理设施处理能力(吨日) ○ 其中:当年新增设施去除的(千克) ○ 氦氮(排放量)(千克)	○ 工业废水排放量(吨)○ 氨氮(去除量)(千克)	売) 🕡	缺失3年 缺失4年 缺失5年
大气环境:	□ 工业煤炭消费量(吨)□ 燃料油消费量(不含车船用)(吨)	○ 其中:燃料煤消费量(吨)○ 其中:重油(吨)	○ 原料煤消费量(吨)○ 柴油(吨)		

图 17 绿色发展数据库缺失情况

- (2) 企业绿色发展信息中的字段只能单选。
- (3)由于统计口径的变动,造成了某些指标的不连续,我们利用统计口径变动的公式,补充了部分数据。具体统计指标换算如下:

化学需氧量 (产生量) =化学需氧量 (去除量) +化学需氧量 (排放量)

氨氮 (产生量)=氨氮 (去除量)+氨氮 (排放量)

二氧化硫产生量=二氧化硫去除量+二氧化硫排放量

氮氧化物产生量=氮氧化物去除量+氮氧化物排放量

烟(粉)尘排放量=烟尘排放量+工业粉尘排放量

烟(粉)尘产生量=烟尘去除量+烟尘排放量+工业粉尘去除量+工业粉尘排放量



具体补充数据如下(蓝色底纹):

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
化学需氧量 (去除量)															
化学需氧量 (产生量)															
氨氮 (去除量)															
氨氮 (产生量)															
二氧化硫去除量															
二氧化硫产生量															
氮氧化物去除量															
氮氧化物产生量															
烟(粉)尘排放量															
烟(粉)尘产生量															

3.2.2.3 单年可视化

提供 1998-2012 年间每年企业绿色发展指标的按企业所在地统计(省级数字地图、柱形-折线混合图)、按企业所属国民经济行业大类统计(条形图)、TOP5 行业大类(环形图)、TOP5 地区(环形图)。



图 18 绿色发展数据库单年可视化

系统默认为有权限的最新年份的全国企业数量数据,用户可选择其他年份、 其他统计指标。当用户点击地图上某一个省(自治区、直辖市)时,各图均自动 切换到对应的省(自治区、直辖市)。除地图的其他图表,可点击右上角的²⁰,



进行放大。

3.2.3.4 时间序列可视化

提供 1998-2012 年任意起止年份间企业绿色发展指标的按企业所在地统计 (线图、面积图)、按企业所在地企业数量 TOP5 (堆积柱形图)、按行业大类统计 (堆积柱形图)、按行业大类企业数量 TOP5 (堆积柱形图)。

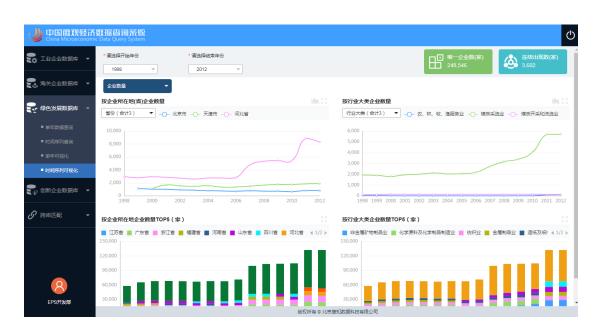


图 19 绿色发展数据库时间序列可视化

用户可依次选择起止时间、统计指标,省份/行业,即可查看所需汇总数据的时间趋势。点击右上角的 , 进行放大。

右上角标注了对应起止时间内的全国"唯一企业数"和"连续出现数",便于用户把握整体情况。

3.2.4 创新企业发展数据库

创新企业数据库提供了1998-2013年中国规模以上工业企业的专利申请和授权情况,数据来源于国家知识产权局。统计字段主要有企业基本信息、三种专利



申请、三种专利授权、三种专利非授权等信息,对于研究中国科技进步、实施创新驱动具有重大意义。

创新企业数据库包括单年数据查询、时间序列查询、单年可视化、时间序列可视化等四部分。

3.2.4.1 单年数据查询

实现对 1998-2013 年间某一年度的全部数据按一定筛选条件进行的查询,用户可便捷的下载到研究所需的数据内容。

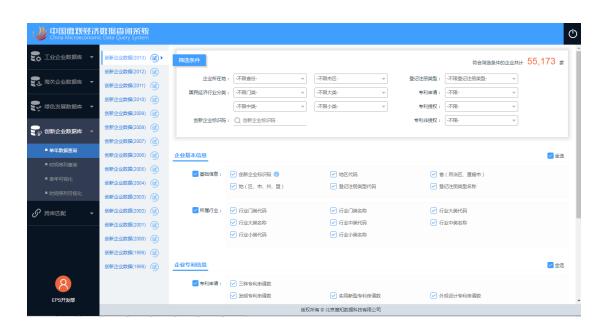
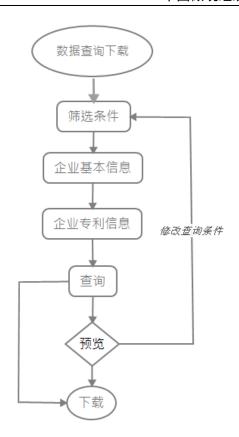


图 20 创新数据库单年数据查询

单年数据查询中查询和下载数据流程如下:





创新企业数据库中的基本信息与专利信息在 1998-2013 年保持恒定,这里以 2013 年为例说明。

筛选条件包含:企业所在地(省、地级城市两级)、国民经济行业(门、大、中、小四级)、登记注册类型、专利申请(不限、发明、实用新型、外观)、专利授权(不限、发明、实用新型、外观)、专利非授权(不限、发明、实用新型、外观)、自定义筛选(创新企业标识码)。当每选中一个筛选条件时,右上角会实时显示"符合筛选条件的企业共计***家"。

筛选条件中需要说明的一点是,当用户选择了专利申请、专利授权、专利非 授权中的筛选项时,查询结果仅显示此筛选项下有数据的记录。

企业基本信息包含企业基础信息、企业所属行业两大类,每类下包含若干字段,字段可多选。企业基本信息或其下的每个大类可全选。基础信息中的"创新企业标识码"是系统对每家创新企业设定的唯一标识,在不同年份出现的同一家



企业具有相同且唯一的编码。

下载页面中可查看所选定的筛选条件与企业基本信息、企业专利信息等输出字段,若不满意,可点击"修改查询条件"返回查询。

为保证数据下载速度,让下载更为流畅,防止恶意下载情况,对于不同权限的用户做了不同的下载限制处理,同时对数据文件进行了分页,每页若干条,每次只能下次一页中的数据。用户可在"数据文件"下看到当前要下载文件的文件大小、数据量、输出格式等。也可预览前 50 条数据。

数据文件			
当前页数据文件大小:	12.17KB		
当前页数据量:	30000		*试用权限仅可下载前50行
输出格式:	CSV		⑥ 预览
		< 1 2	2 > 共2页
		◆ ◆	✓ 压缩为.zip格式下载

图 21 创新企业数据库数据文件内容

数据文件的下载格式为学界常用的 CSV 格式。

3.2.4.2 时间序列查询

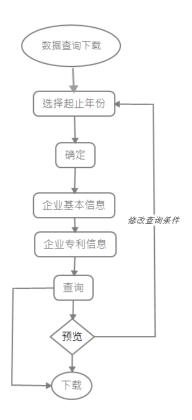
采用专业化清洗与匹配方式,对不同年份的企业进行匹配。用户可轻松获得 1998-2013年任意起止年份间企业匹配后的面板数据。





图 22 创新企业数据库时间序列查询

时间序列查询中查询和下载数据流程如下:



这里大部分与前述 3.2.4.1 单年查询一致,现仅说明几点事项:



- (1) 企业专利信息中的字段只能单选。
- (2)当用户在筛选条件里选择了专利申请、专利授权、专利非授权中的筛选项时,此筛选项所对应的年份数据必定是连续的,但企业专利信息里的指标则未必连续。例如当用户在筛选条件里选择了发明专利申请,但在"企业专利信息"里选择了实用新型专利授权,那么在查询出来的结果中,仅展示此起止年份里的各企业实用新型专利授权数,它未必是时间连续的,但这些企业必定在实用新型专利授权数中是时间连续的。

3.2.4.3 单年可视化

提供 1998-2013 年间每年专利指标的按企业所在地统计(省级数字地图、柱形-折线混合图)、按企业登记注册类型统计(条形图)、按企业所属行业大类统计(条形图)、TOP10企业(列表)。



图 23 创新企业数据库单年可视化

系统默认为有权限的最新年份的全国企业数量数据,用户可选择其他年份、 其他统计指标。当用户点击地图上某一个省(自治区、直辖市)时,各图、表格 均自动切换到对应的省(自治区、直辖市)。除地图的其他图表,可点击右上角



的 , 进行放大。

3.2.4.4 时间序列可视化

提供1998-2013年任意起止年份间企业专利指标的按企业所在地统计(线图、面积图)及其 top10企业列表、按企业所属行业大类统计(线图、面积图)及其 top10企业列表、按登记注册类型统计(线图、面积图)及其 top10企业列表。



图 24 创新企业数据库时间序列可视化

用户可依次选择起止时间、统计指标,省份/行业,即可查看所需汇总数据的时间趋势。点击右上角的,进行放大。

3.3 跨库检索

跨库匹配分为单年匹配和时间序列匹配,用户可选定工业企业数据库、海关企业数据库、绿色发展数据库间任意多个专题数据库的匹配,轻松获得截面和面板数据,方便用户进行跨学科、跨领域研究。



3.3.1 单年匹配

实现三个专题数据库间企业某一年度按一定筛选条件进行的查询,用户可便捷的同时下载企业某年的多专题截面数据。



图 25 跨库单年匹配

用户可任选两个或三个数据库进行匹配,通过点击生实现。



图 21 单年跨库匹配数据库选择

用户依次点击要匹配的数据、筛选条件、企业基本信息、企业其他信息等,



即可查询到所需数据。每次可最多下载5万条数据。

3.3.2 时间序列匹配

采用专业化清洗与匹配方式,对不同专题数据库中的企业进行跨库匹配。用 户可轻松获得任意起止年份间企业匹配后的面板数据。



图 26 跨库时间序列匹配

在时间序列的跨库匹配中,需要选择哪个数据库作为时间序列的衡量基准。

查询、下载步骤依次为选择基准数据、选择匹配数据、确定筛选条件(含剔除变化企业)、勾选企业基本信息(多选)、勾选企业其他信息(单选)。每次最大可下载5万条数据。