



RESSET 文本分析客户端-小锐助手

使用说明

RESSET 技术部

2024 年

目录

1	平台介绍	4
2	平台首页	4
3	使用方式	5
4	登录方式	7
5	文本数据	8
5.1	中国上市公司财经文本	8
5.1.1	股票代码选择	8
5.1.2	时间设置	10
5.1.3	关键字设置	10
5.1.4	查询结果展示	12
5.2	政府工作文本数据	12
5.2.1	政府地区选择	13
5.2.2	时间设置	13
5.2.3	关键字设置	14
5.2.4	查询结果	14
5.3	美国上市公司财经文本	14
5.3.1	股票代码选择	15
5.3.2	时间设置	16
5.3.3	关键字设置	16
5.3.4	查询结果展示	17
5.4	财经资讯文本	17
5.4.1	输入查询条件	18
5.4.2	关键字设置	18
5.4.3	查询结果展示	19
6	词频搜索	20
6.1	输入关键词检索	20
6.2	导入模板检索	21
7	文本特征	23
7.1	文本信息	24
7.2	情感分析	25
7.3	可读性分析	27
7.4	自定义文本特征	28
7.5	特色指标	33
7.5.1	描述性创新	33
7.5.2	风险因子	34
7.5.3	竞争性因子	36
7.5.4	前瞻性因子	37
7.5.5	区块链概念因子	39

7.5.6	人口老龄化指标	41
7.5.7	员工待遇因子	42
7.5.8	贸易战潜在风险	43
7.5.9	管理会计因子	44
8	文本挖掘.....	46
8.1	文本预处理	46
8.1.1	文本转换	47
8.1.2	文本分词	48
8.1.3	去停用词	50
8.1.4	特征词库	50
8.1.5	文本向量化	51
8.2	机器学习	53
8.2.1	相似度计算	53
8.2.2	文本分类	56
8.2.3	文本聚类	59
8.2.4	主题模型	60

小锐助手

1 平台介绍

RESSET 文本分析客户端（简称小锐助手）是一种基于计算机技术的自动化文本分析工具，它可以帮助用户从文本中提取有价值的信息，并对其进行深入的分析 and 处理。客户端基于 NLP（自然语言处理）技术，使用先进的算法和模型来对文本进行文本转换、识别、分类、情感分析、实体建模等任务。

文本预处理：对文本清洗、分词、词性标注等预处理操作。

特征提取：通过词频统计、词性标注、命名实体识别等技术实现。

模型训练：以识别文本中的模式和趋势。

情感分析：以了解用户对某个主题或产品的态度和情感。

实体建模：识别出其中的实体关系和语义关系。

标签和分类：以方便用户进行组织和查询。

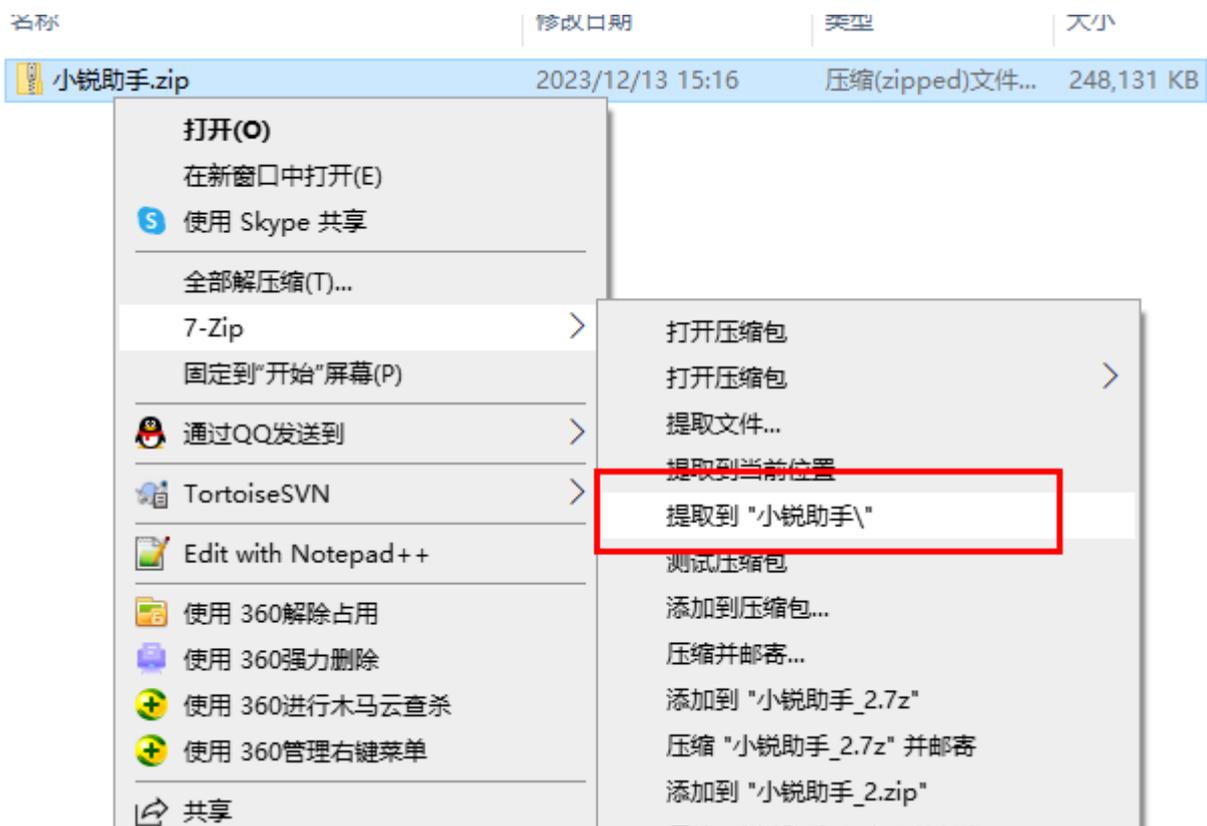
可视化分析：帮助用户直观地展示和分析文本数据。

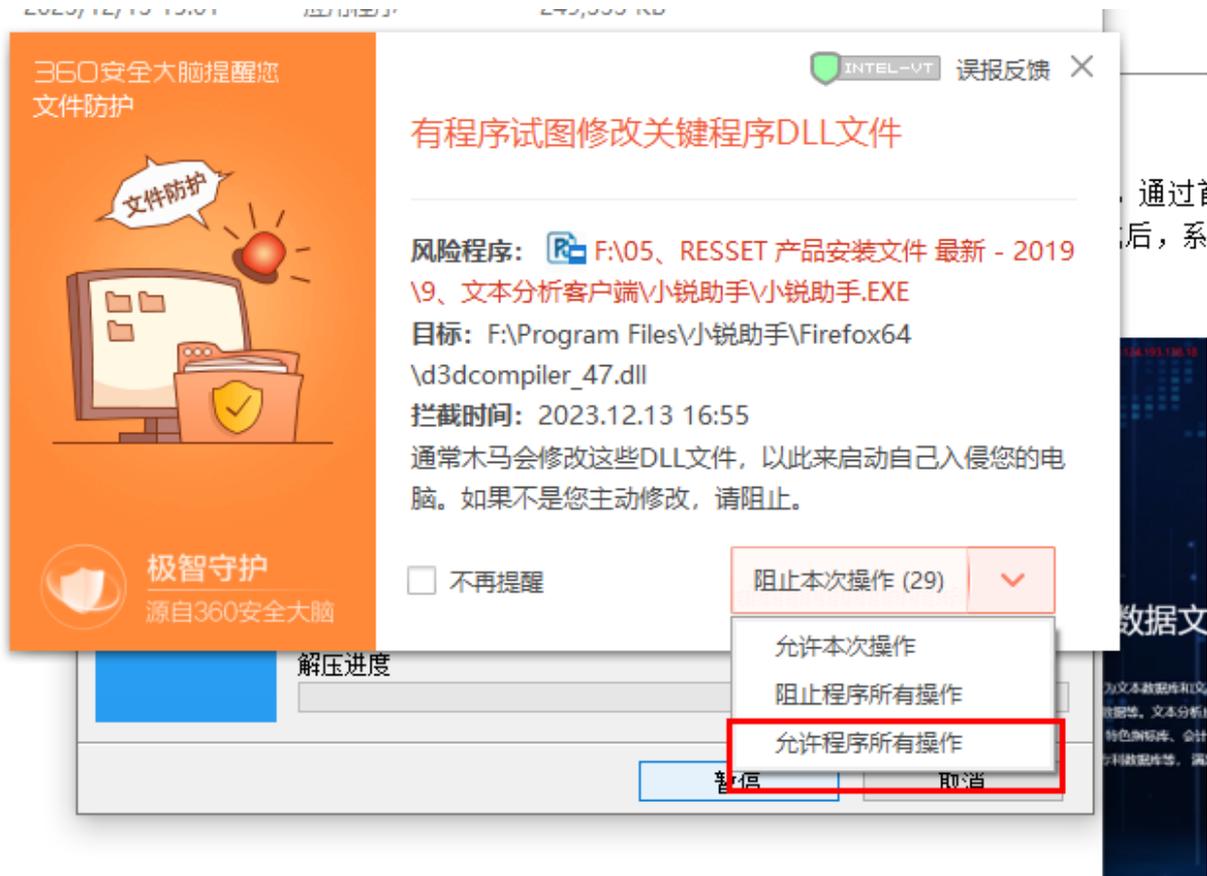
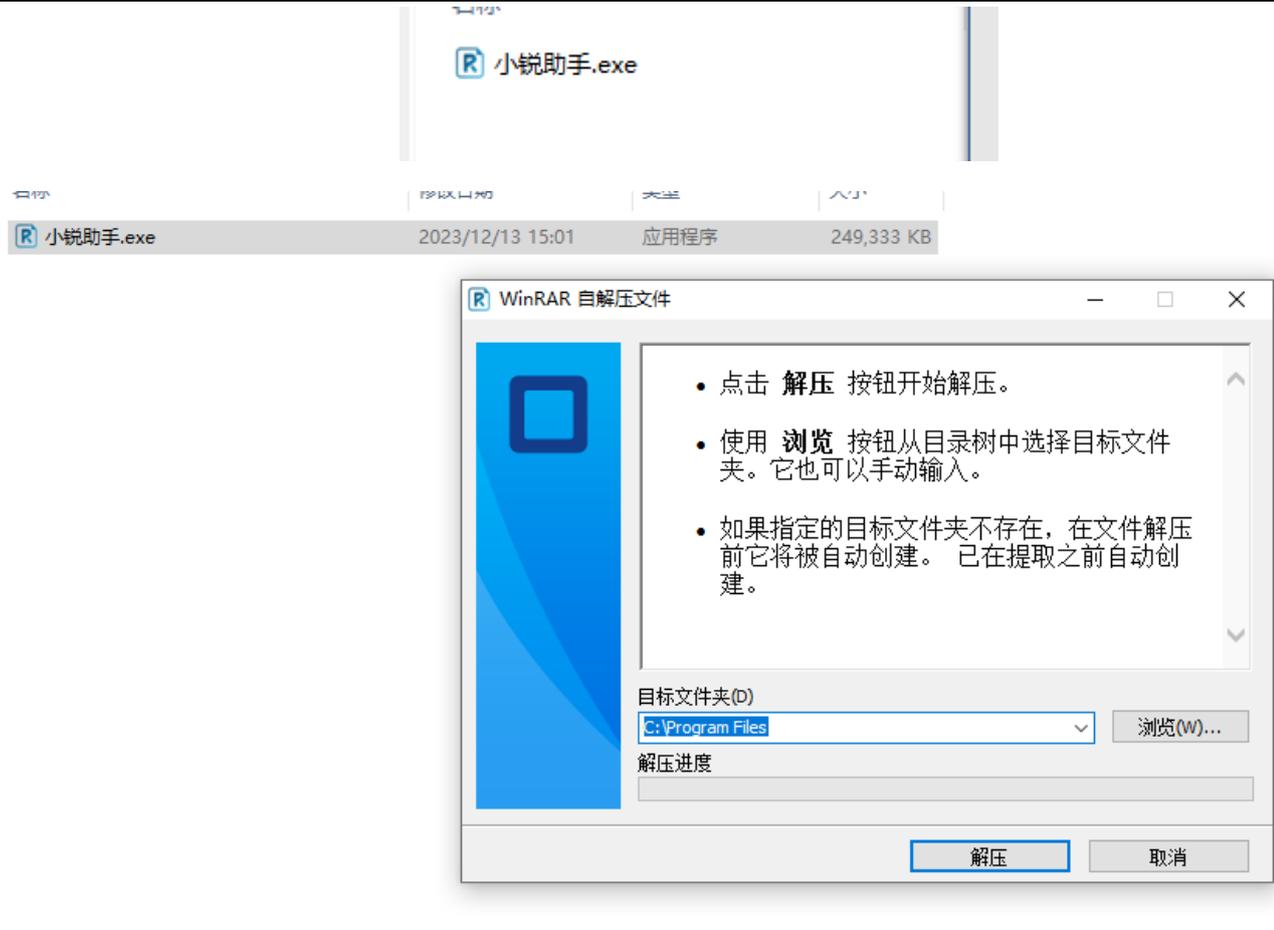
2 平台首页



3 使用方式

下载位置，访问 [RESSET 财经文本智能分析平台](#)，通过在系统首页左下角点击小锐助手，点击下载。下载完成后，将其解压，然后双击小锐助手.exe 进行安装，安装过程中 360 安全防卫会出现弹窗，选择允许程序所有操作即可，安装完成后，系统桌面会自动生成小锐助手快捷方式，双击即可登录。





4 登录方式

用户通过双击桌面小锐助手快捷方式打开小锐助手，在登录窗口中输入用户信息。**小锐助手**账号同 BS 模式账号一致。



备注：校内使用小锐助手，需要学校已购买或开通财经文本智能分析平台 B/S 模式账号，通过在登录窗口输入本校账号及密码后点击登录，助手自动识别 IP 进入小锐助手；也可通过 CARS 登录方式登录小锐助手。

5 文本数据

5.1 中国上市公司财经文本

文本数据起止时间： 2001-至今

数据库子库包含：定期报告（季报和年报）、问询函及回复说明、IPO 招股说明书、内部控制评价报告、业绩说明会全文、社会责任报告、上市公司重大事项公告、审计报告、风险管理业务公告以及上市公司典型案例。

以定期报告-年度报告分类为例进行操作。



5.1.1 股票代码选择

代码选择分为三种方式：全市场股票、从股票池选择、股票代码导入。

全市场股票：用户可选择全市场股票即选择全部的股票代码，不需要单独选择股票代码；

股票代码选择

全市场股票 从股票池选择 股票代码导入

从股票池选择：选择左侧行业分类，点击 **筛选**，股票代码进入中间待选区，用户可以通过点击 **全选** 待选区代码/单个选择代码将代码自动移至已选区，如图：

第一步：输入查询条件

股票代码选择

全市场股票 从股票池选择 股票代码导入

--行业选择-- 筛选 <input type="checkbox"/> 股票市场分类 <input type="checkbox"/> 上市状态 <input type="checkbox"/> 证监会行业分类2012版 <input type="checkbox"/> 中信行业分类 <input type="checkbox"/> 中证指数行业分类 <input type="checkbox"/> 中证指数_证监会行业分类2012版 <input type="checkbox"/> 申万行业分类2014版	--待选区5150个-- 全选 请输入股票名称/代码/拼音 <input type="text"/> 000001-平安银行 000002-万科A 000003-PT金田A 000004-国华网安 000005-ST 星源 000006-深振业A 000007-*ST 全新	已选0个 全部删除
---	--	------------------

注：点击待选字段选定，点击已选字段删除

注：点击待选字段选定，点击已选字段删除。

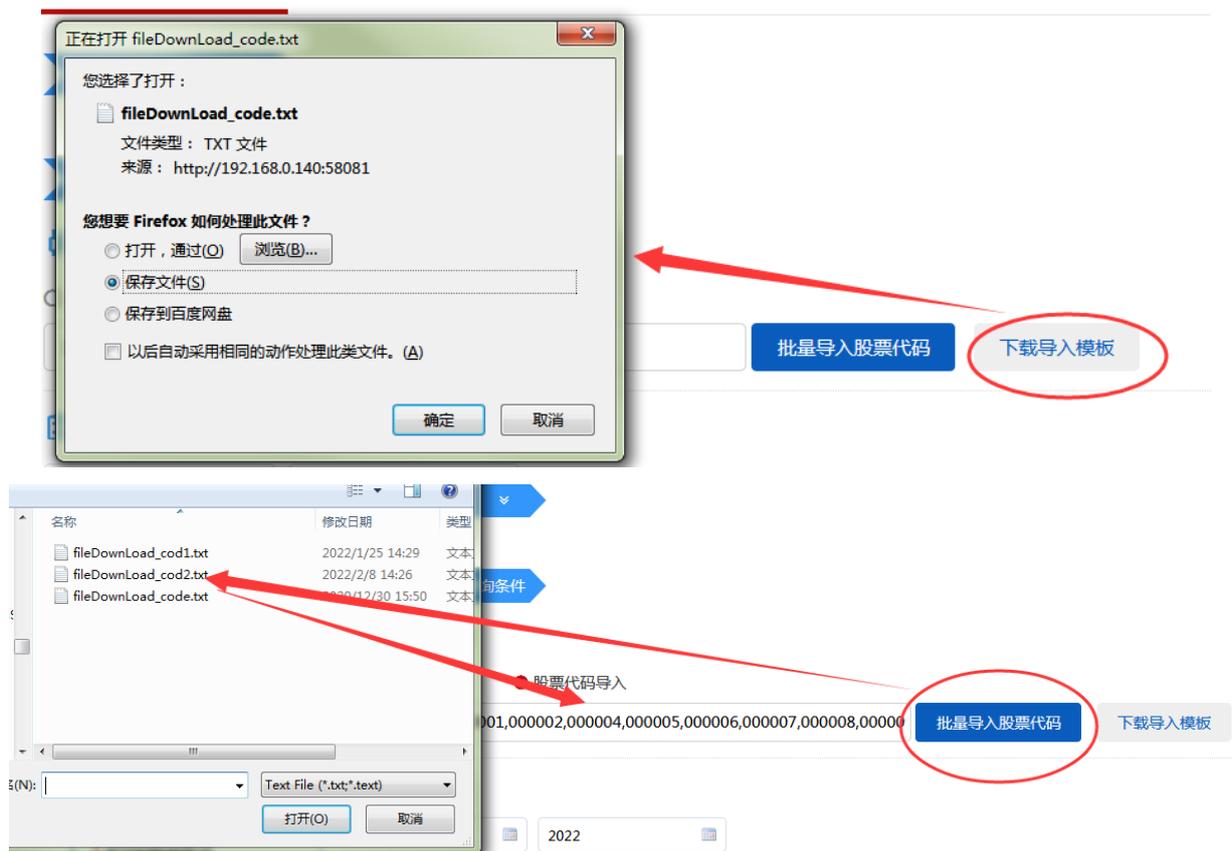
股票代码导入：用户需要先下载导入模板，在模板文件加入需要的股票代码信息，编辑完成后，点击批量导入股票代码，导入的代码会在股票代码框中显示，如图：

第一步：输入查询条件

股票代码选择

全市场股票 从股票池选择 股票代码导入

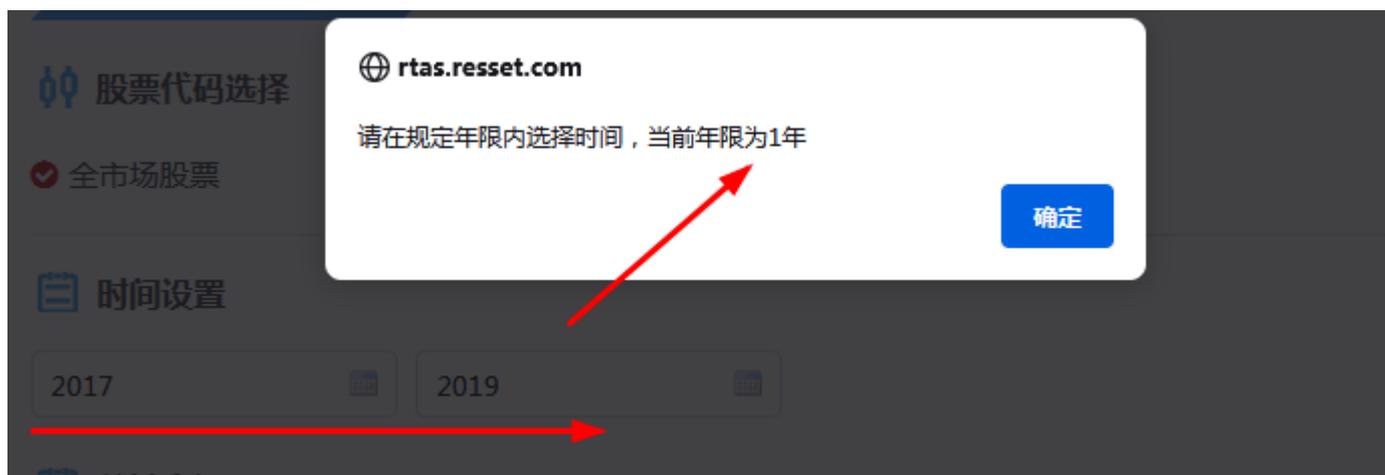
股票代码 **批量导入股票代码** **下载导入模板**



5.1.2 时间设置

时间范围设置: 系统后台内设置时间范围, 如果设置年份超出设定范围, 在下一步点击查询时则会给出相应的提示, 如图:

时间设置



5.1.3 关键字设置

关键字设置可对标题/全文进行筛选, 设置关键字: 包含任一关键字 (-分隔)、包含所有关键字 (-分隔) 以及不包含任一关键字 (-分隔)。

标题：包含任一关键字（-分隔）

关键字设置

标题 全文

标题

标题

标题

点击选取股票，用户可以在弹出框中选择关键字信息，如图：

-检索-

--行业选择--

- 股票市场分类
- 上市状态
- 证监会行业分类2012版
- 中信行业分类
- 中证指数行业分类
- 中证指数_证监会行业分类2012版
- 申万行业分类2014版

注：点击待选字段选定，点击已选字段删除

待选4761个

请输入股票名称/代码/拼音

- 000018-神城A退
- 000019-深粮控股
- 000020-深华发A
- 000021-深科技
- 000022-深赤湾A
- 000023-深天地A
- 000024-招商地产
- 000025-特力A

已选4个

- 000014-沙河股份
- 000015-PT中浩A
- 000016-深康佳A
- 000017-深中华A

标题

关键字设置

标题 全文

标题

标题

标题

全文：

关键字设置

标题 全文

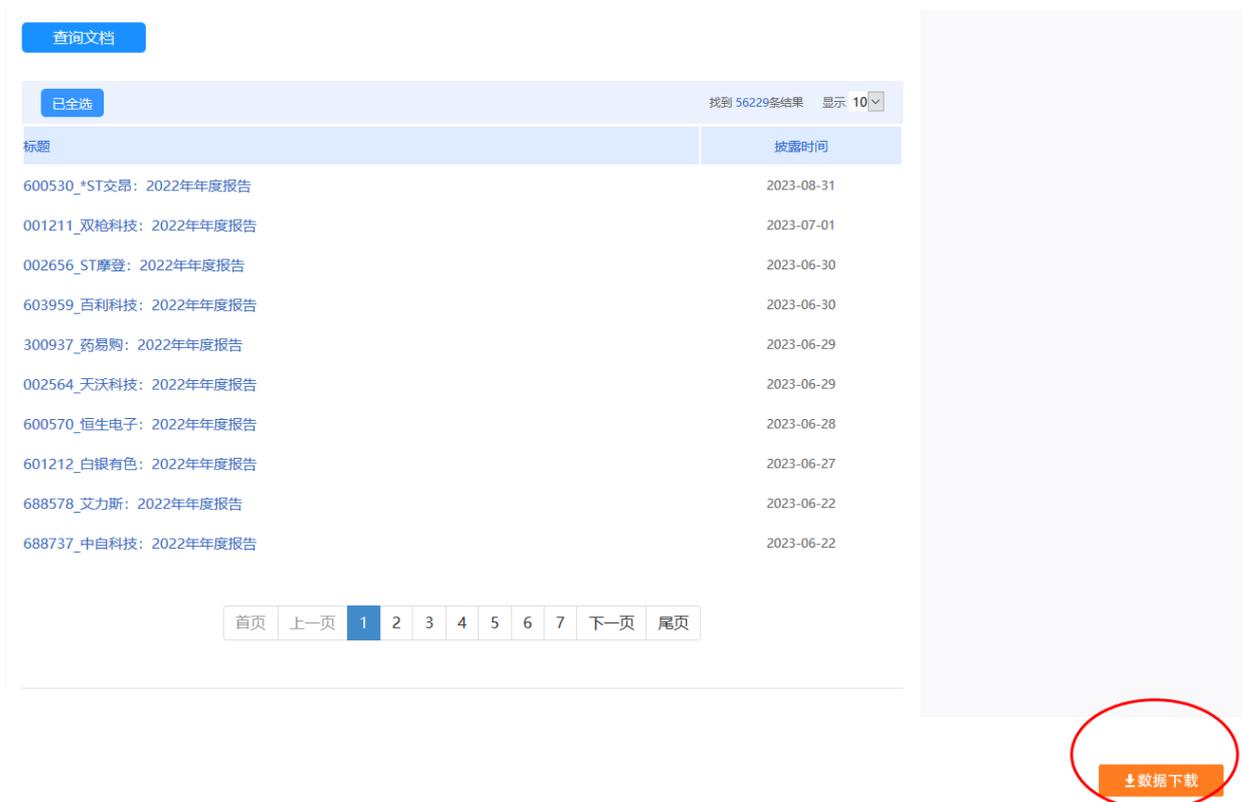
全文

全文

全文

5.1.4 查询结果展示

综上所述操作选择完成后，点击  按钮，结果即可显示，用户可以点击  按钮，进行查询结果下载，如图：



5.2 政府工作文本数据

政府工作文本数据库目前包括国务院、省级以及地级市政府工作报告三大数据源。政府工作报告是外界了解政府的主要信息来源，包含着宏观政策、发展规划、资源配置等重要信息。文本的主要内容有对过去一年工作业绩与问题的回顾性总结，对未来一年政府工作的预期目标，以及具体实施的方案与安排，是一个具有施政纲领性质的政策性文本，是记录政府履职的重要政策性工具。

文本数据起止时间： 1954-至今

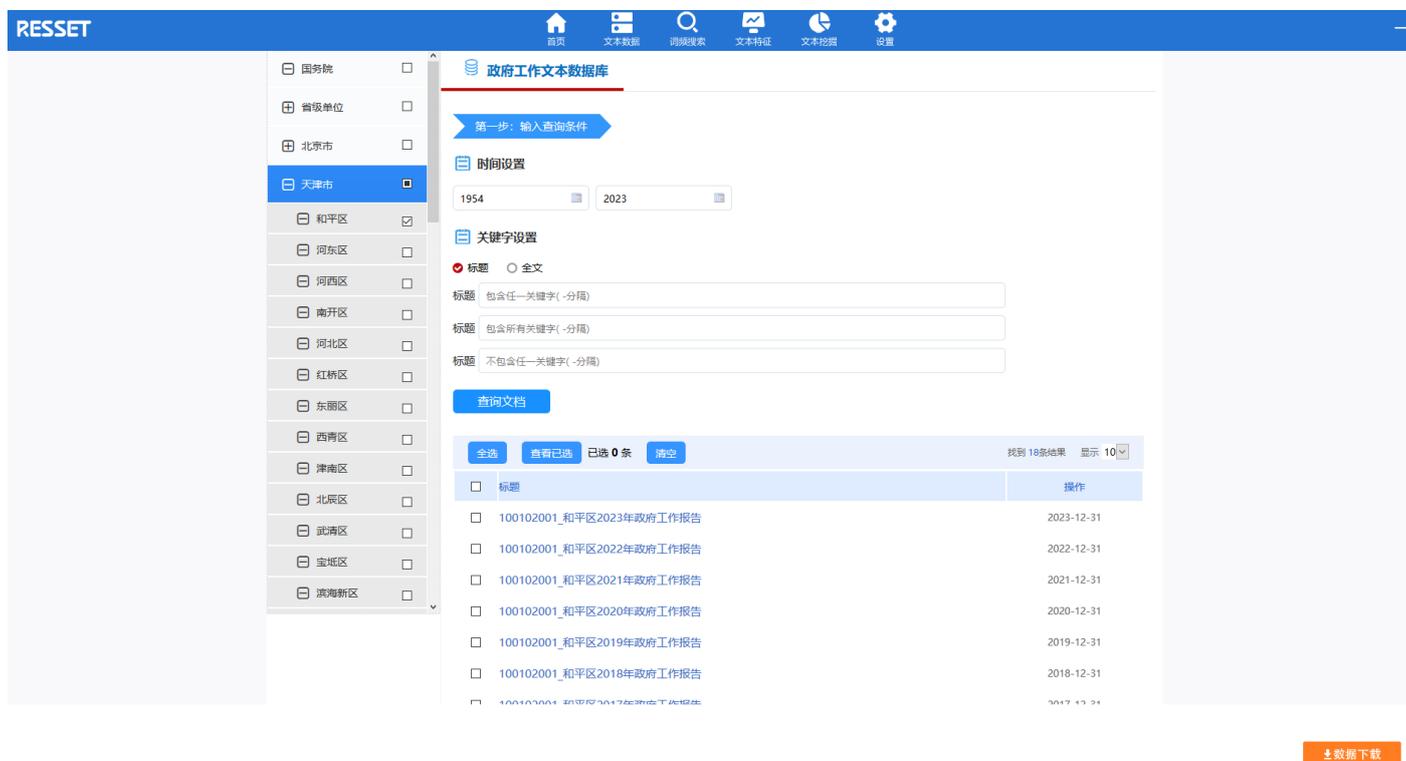
文本数据包含： 国务院、省级以及地级市政府工作报告。其是外界了解政府的主要信息来源，包含着宏观政策、发展规划、资源配置等重要信息。

文本分析： 数据库子库。



5.2.1 政府地区选择

选择政府工作报告（国务院），如图：



5.2.2 时间设置

时间设置：系统内设置时间范围为1年，如果设置年份超出1年，下一步点击查询时则会给出相应的提示，如图：



时间设置

2020  2021 

5.2.3 关键字设置

关键字设置可分：

标题： 包含任一关键字（-分隔）、包含所有关键字（-分隔）、不包含任一关键字（-分隔）；

全文： 包含任一关键字（-分隔）、包含所有关键字（-分隔）、不包含任一关键字（-分隔）。

关键字设置

标题 全文

标题

标题

标题

5.2.4 查询结果

点击查询按钮，用户可以对查询结果进行下载。如图：

查询文档

已选 10 条

找到 18 条结果 显示 10 

<input checked="" type="checkbox"/>	标题	操作
<input checked="" type="checkbox"/>	100102001_和平区2023年政府工作报告	2023-12-31
<input checked="" type="checkbox"/>	100102001_和平区2022年政府工作报告	2022-12-31
<input checked="" type="checkbox"/>	100102001_和平区2021年政府工作报告	2021-12-31
<input checked="" type="checkbox"/>	100102001_和平区2020年政府工作报告	2020-12-31
<input checked="" type="checkbox"/>	100102001_和平区2019年政府工作报告	2019-12-31
<input checked="" type="checkbox"/>	100102001_和平区2018年政府工作报告	2018-12-31
<input checked="" type="checkbox"/>	100102001_和平区2017年政府工作报告	2017-12-31
<input checked="" type="checkbox"/>	100102001_和平区2016年政府工作报告	2016-12-31
<input checked="" type="checkbox"/>	100102001_和平区2015年政府工作报告	2015-12-31
<input checked="" type="checkbox"/>	100102001_和平区2014年政府工作报告	2014-12-31

导出数据下载

5.3 美国上市公司财经文本

文本数据起止时间： 1954-至今；

文本数据包含**年报、季报、招股说明书**；

文本分析：**数据库子库**。



5.3.1 股票代码选择

代码选择分为两种方式：从股票池选择、股票代码导入。

从股票池选择：用户可以在左侧代码待选区，点击单个/全选，代码进入已选区，如图：

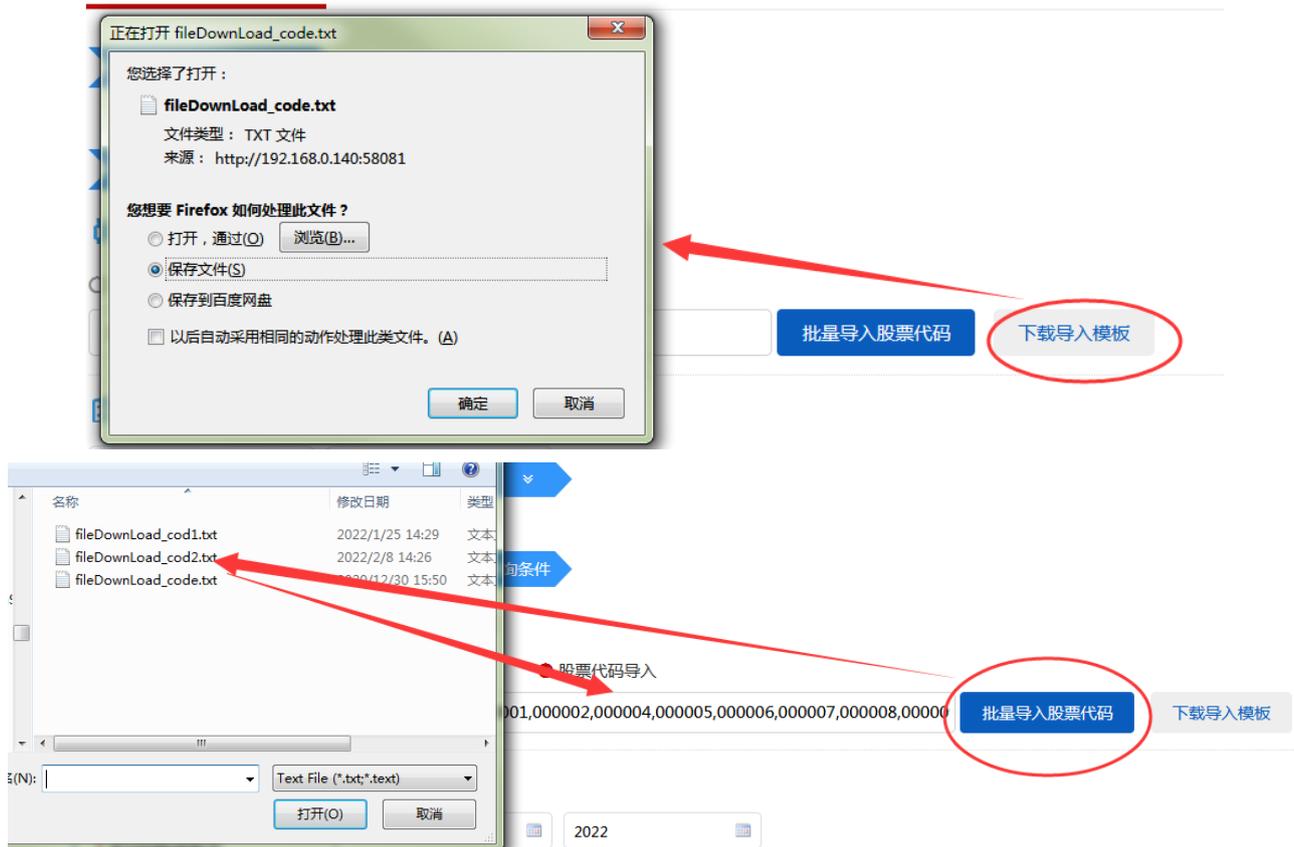


注：点击待选字段选定，点击已选字段删除

注：点击待选字段选定，点击已选字段删除。

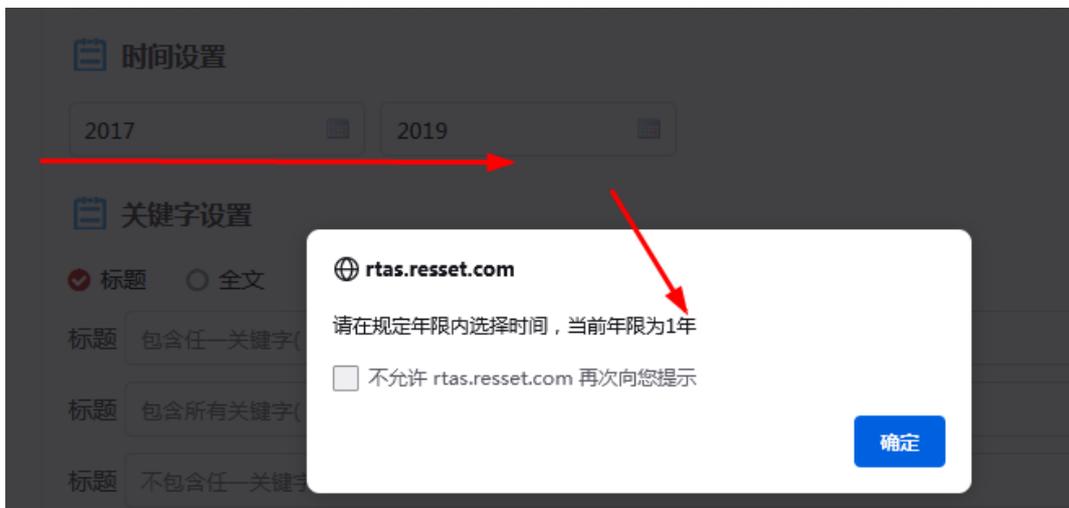
股票代码导入：用户可以通过先下载导入模板，在模板文件加入需要的代码信息，编辑完成后，点击批量导入股票代码，导入的代码会在股票代码框中显示，如图：





5.3.2 时间设置

时间范围设置：系统内设置时间范围为1年，如果设置年份超出1年，在下一步点击查询时则会给出相应的提示，如图：



5.3.3 关键字设置

关键字设置可对标题/全文进行筛选，设置关键字：包含所有关键字、包含任一关键字以及不包含任一关键字。

标题：

关键字设置

标题 全文

标题 包含所有关键字,(空格分隔)

标题 包含任一关键字,(空格分隔)

标题 不包含任一关键字,(空格分隔)

全文:

关键字设置

标题 全文

全文 包含所有关键字,(空格分隔)

全文 包含任一关键字,(空格分隔)

全文 不包含任一关键字,(空格分隔)

5.3.4 查询结果展示

综上所述操作选择完成后，点击 [查询文档](#) 按钮，查询结果即可显示，用户可以对查询结果进行下载，如图：

The screenshot shows the search results interface. At the top, there is a blue button labeled "查询文档". Below it, there are several filters: "全选", "查看已选", "已选 10 条", and "清空". On the right side, it says "找到 234 条结果" and "显示 10". The main content is a table with two columns: "标题" and "披露时间". The table contains 10 rows of data, each with a checkbox in the "标题" column and a date in the "披露时间" column. At the bottom right, there is a red circle around a button labeled "数据下载".

标题	披露时间
<input checked="" type="checkbox"/> ABR_ABR_20211231Annual report	2021-12-31
<input checked="" type="checkbox"/> ABBV_ABBV_20211231Annual report	2021-12-31
<input checked="" type="checkbox"/> AB_AB_20211231Annual report	2021-12-31
<input checked="" type="checkbox"/> ABM_ABM_20211031Annual report	2021-10-31
<input checked="" type="checkbox"/> ABC_ABC_20210930Annual report	2021-09-30
<input checked="" type="checkbox"/> ABMD_ABMD_20210331Annual report	2021-03-31
<input checked="" type="checkbox"/> ABCB_ABCB_20201231Annual report	2020-12-31
<input checked="" type="checkbox"/> ABCL_ABCL_20201231Annual report	2020-12-31
<input checked="" type="checkbox"/> ABG_ABG_20201231Annual report	2020-12-31
<input checked="" type="checkbox"/> ABBV_ABBV_20201231Annual report	2020-12-31

5.4 财经资讯文本

文本数据起止时间： 2000-至今；

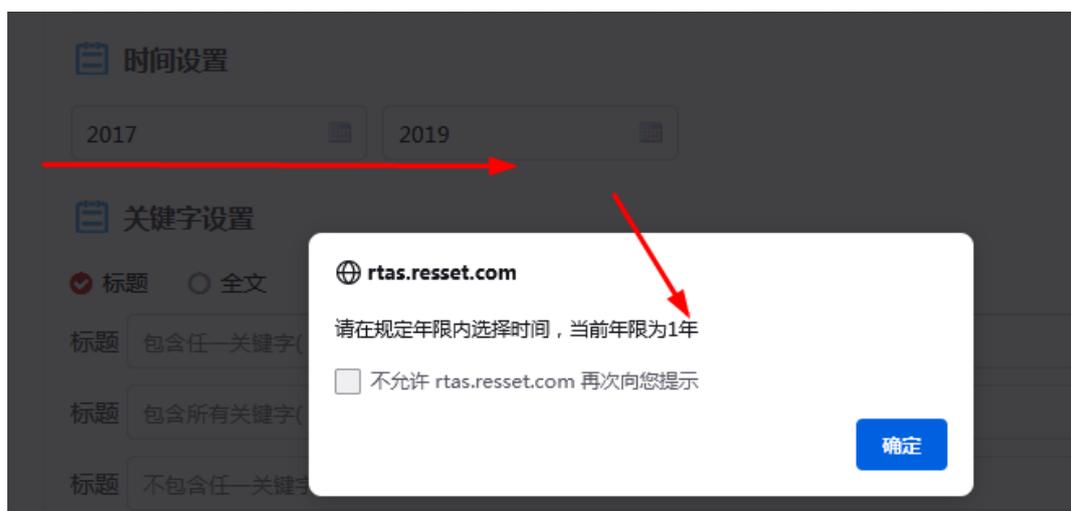
财经资讯包含：证券日报、证券时报、智通财经、中财网、中国经济网、中国经济报、中国证券报、上海证券报、财经网、新华社、凤凰新闻、今日头条、腾讯财经等 30 多个具有影响力的财经资讯；

文本分析：数据库子库。



5.4.1 输入查询条件

时间范围设置：系统后台内设置时间范围，如果设置年份超出设定范围，在下一步点击查询时则会给出相应的提示，如图：



5.4.2 关键字设置

关键字设置可对标题/全文进行筛选，设置关键字：包含所有关键字、包含任一关键字以及不包含任一关键字。

标题：

关键字设置

标题 全文

标题

标题

标题

全文：

关键字设置

标题 全文

全文 包含所有关键字,(空格分隔)

全文 包含任一关键字,(空格分隔)

全文 不包含任一关键字,(空格分隔)

5.4.3 查询结果展示

综上所述操作选择完成后, 点击 [查询文档](#) 按钮, 查询结果即可显示, 用户可以对查询结果进行下载, 如图:

查询文档

全选 查看已选 已选 10 条 清空 找到 49115 条结果 显示 10

<input checked="" type="checkbox"/> 标题	披露时间
<input checked="" type="checkbox"/> *ST庞大被证监会立案调查 遗留问题远不止信息披露违规	2023-05-28
<input checked="" type="checkbox"/> 奥贝胆酸NASH适应症上市申请被推迟 相关赛道药物"上市第一名"之争横生变数	2023-05-28
<input checked="" type="checkbox"/> 上市6周登顶畅销榜TOP1 海信电视U7K为"爆款王"打样	2023-05-28
<input checked="" type="checkbox"/> 信息披露不准确沈阳化工收警示函 2022年净利润现较大额亏损	2023-05-28
<input checked="" type="checkbox"/> 森源电气回复深交所问询函:新实控人能够对公司实施有效、稳定的控制	2023-05-28
<input checked="" type="checkbox"/> 具备开发800Gbps及以上砷酸锂调制器芯片能力 光库科技助推芯片国产化	2023-05-28
<input checked="" type="checkbox"/> <股东来了>(2023)甘肃片区燃高校 走进校园百余场	2023-05-28
<input checked="" type="checkbox"/> 2023数博会"东数西算"高端对话在贵阳成功召开	2023-05-28
<input checked="" type="checkbox"/> 2023数博会"数据二十条"安全产业发展论坛成功举办	2023-05-28
<input checked="" type="checkbox"/> <股东来了>(2023)*投教万里 星聚天山*系列投教活动走进伊力特	2023-05-28

首页 上一页 1 2 3 4 5 6 7 下一页 尾页

数据下载

6 词频搜索



6.1 输入关键词检索

用户可以通过在输入关键词框内输入单个、多个关键词进行检索，多个关键词用逗号分开，本助手支持中文、英文文档解析。

使用方法：1、输入关键词；2、选择自定义文件夹；3、点击检索；4、显示分析结果；5、下载分析结果文件。



RESSET 词频搜索

秒传海量文本 本地分析挖掘

Q

关键词模板下载
自定义文件夹
D:\360安全浏览器下载\1111
中文

导出TXT
导出表格

文件	时间	大小	字数	总词数	总句数	关键词	词频
ST全新: 关于对深交所 (2021) 第225号年报问询函回复的公告6f036587fb9e4bc9847c623f60883630.p...	2024-05-14	67.07KB	27322	14500	905	财务	31
ST全新: 关于对深交所 (2021) 第225号年报问询函回复的公告6f036587fb9e4bc9847c623f60883630.p...	2024-05-14	67.07KB	27322	14500	905	经济	8
ST全新: 独立董事关于对深交所公司部年报问询函 [2022] 第 162 号回复之独立意见08c6a88b6dc842e0...	2024-05-14	5.10KB	1971	1096	70	财务	2
ST全新: 独立董事关于对深交所公司部年报问询函 [2022] 第 162 号回复之独立意见08c6a88b6dc842e0...	2024-05-14	5.10KB	1971	1096	70	经济	2
ST全新: 诚公律师事务所对全新好2021年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7....	2024-05-14	35.96KB	14251	8115	515	财务	2

首页 上一页 1 2 3 下一页 尾页

6.2 导入模板检索

点击关键词模板下载，设置关键词，关键词整理完成后，双击输入框上传关键词文件，选择自定义文件夹，点击检索，查看分析结果并支持下载。

RESSET 词频搜索

秒传海量文本 本地分析挖掘

Q

关键词模板下载
自定义文件夹
中文



打开

关键词文件

名称	修改日期	类型
关键词模版.txt	2023/10/12 10:23	TXT

请输入关键词，多个关键词用逗号分开,双击导入关键词

关键词模板下载 自定义文件夹 中文

RESSET 词频搜索

秒传海量文本 本地分析挖掘

财务,经济

关键词模板下载 自定义文件夹 中文

RESSET 词频搜索

秒传海量文本 本地分析挖掘

RESSET 词频搜索

秒传海量文本 本地分析挖掘

财务,经济



关键词模板下载

自定义文件夹
D:\360安全浏览器下载\1111

中文

导出TXT

导出表格

文件	时间	大小	字数	总词数	总句数	关键词	词频
ST全新: 关于对深交所 (2021) 第225号年报问询函回复的公告6f036587fb9e4bc9847c623f60883630.p...	2024-05-14	67.07KB	27322	14500	905	财务	31
ST全新: 关于对深交所 (2021) 第225号年报问询函回复的公告6f036587fb9e4bc9847c623f60883630.p...	2024-05-14	67.07KB	27322	14500	905	经济	8
ST全新: 独立董事关于对深交所公司部年报问询函 (2022) 第 162 号回复之独立意见08c6a88b6dc842e0...	2024-05-14	5.10KB	1971	1096	70	财务	2
ST全新: 独立董事关于对深交所公司部年报问询函 (2022) 第 162 号回复之独立意见08c6a88b6dc842e0...	2024-05-14	5.10KB	1971	1096	70	经济	2
ST全新: 诚公律师事务所对全新好2021年年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7...	2024-05-14	35.96KB	14251	8115	515	财务	2

[首页](#)
[上一页](#)
[1](#)
[2](#)
[3](#)
[下一页](#)
[尾页](#)

7 文本特征

RESSET 小锐助手

[首页](#)
[文本数据](#)
[词频搜索](#)
[文本特征](#)
[文本挖掘](#)
[设置](#)

文本特征分析 自动化统计文本中的文本指标

- 常用文本分析方法
- 支持任意文本文件
- 自定义词汇和字典
- 一键导出分析结果

文本信息

- 解析文件大小
- 分析文件页数
- 准确统计文本句子数
- 提供文本字数和词汇数

情感分析

- 提供常用情感词库
- 支持自定义情感词库
- 提供百分比的语感分析
- 支持多种格式自定义文本

可读性分析

- 提供常用可读性词库
- 支持自定义的词库
- 支持多种格式自定义文本
- 一键导出分析结果

自定义文本特征

- 指标构建
- 词典定制
- 特征计算
- 分析结果导出

特色指标

描述性指标

风险因子

竞争因子

新增因子

区域概念因子

人口老龄化指标

员工离职因子

舆情监测风险

管理会计因子

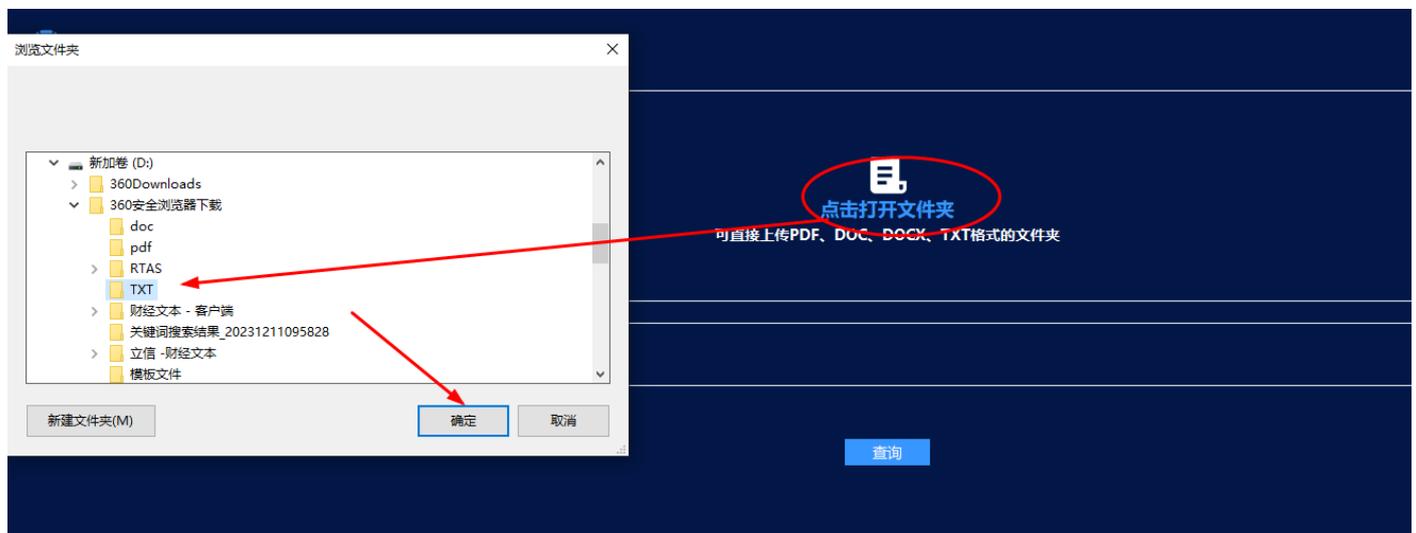
锐思数据

7.1 文本信息



文本信息功能：1、解析文件大小；2、分析文件页数；3、准确统计文本句子数；4、提供文本字数和词汇数。

操作：上传 PDF、DOC、DOCX、TXT 文件，系统自动生成文件基本信息。用户可以对解析结果进行导出。



文本特征

文本信息

情感分析

可读性分析

自定义文本特征

文本信息 统计文件的一些基本信息

点击打开文件夹

可直接上传PDF、DOC、DOCK、TXT格式的文件夹

已选择数据文件:D:\360安全浏览器下载\1111

已分析: 7/7 文件分析完成

查询

导出TXT 导出表格

文件名	上传时间	文件大小	总字数	总词数	总句数	总页数
ST全新: 关于对深交所 (2021) 第225号年报问询函回复的公告6f036587f69e4bc9847c623f60883630.pdf.txt	2024-04-12	67.87KB	26835	15206	1871	30
ST全新: 独立监事关于对深交所公司部年报问询函 (2022) 第 162 号回复之独立意见08c6a88b5dc842e0be1d0c8261743a8c...	2024-04-12	5.13KB	1923	1101	122	3
ST全新: 诚公律师事务所对全新好2021年年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7.pdf.txt	2024-04-12	36.34KB	13936	8286	969	18
ST国华: 致同会计师事务所 (特殊普通合伙) 关于深圳国华网安科技股份有限公司2021年年报问询函的回复d0c057d60d4d447...	2024-04-12	94.37KB	35953	20772	2498	42
昌平区2022年政府工作报告c1e521a152664f5bb462f83b3de4d17.txt	2024-04-12	42.84KB	15306	8589	955	1

首页 上一页 1 2 下一页 尾页

功能简介

1. 上传PDF、DOC、DOCK、TXT文件, 系统自动生成文件基本信息
2. 统计TXT文件时页数为1.

7.2 情感分析

文本特征分析 自动化统计文本中的文本指标

- 常用文本分析方法
- 支持任意文本文件
- 自定义词汇和字典
- 一键导出分析结果

文本信息

解析文件大小

分析文件页数

准确统计文本句子数

提供文本字数和词汇数

情感分析

提供常用情感词库

支持自定义情感词库

提供百分比的语调分析

支持多种格式自定义文本

可读性分析

提供常用可读性词库

支持自定义的词库

支持多种格式自定义文本

一键导出分析结果

自定义文本特征

指标构建

词典定制

特征计算

分析结果导出

特色指标

描述性指标 风险因子 竞争性因子 前瞻性因子 区块链概念因子

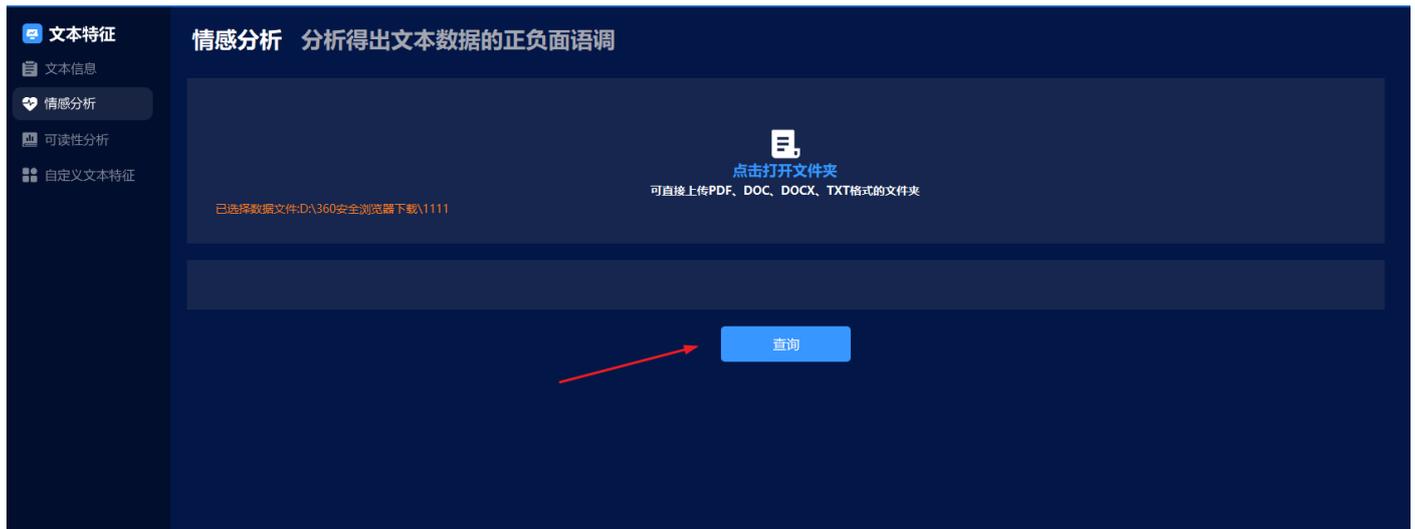
人口老龄化指标 员工待遇因子 贸易标准在风险 管理会计因子

功能简介: 根据情感词典计算文本的情感极性, 检索文本中所包含的积极情感词和消极情感词的个数。

最终文本的情感得分为 $TONE = (\text{积极情感词个数} - \text{消极情感词个数}) / (\text{总情感词个数})$ 。

情感得分取值范围为[-1,1], 情感得分越接近 1, 文本的情感极性越偏向于积极; 反之, 情感的分越接近-1, 文本的情感极性越偏向于消极。

功能操作：点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，上传完成后，点击查询按钮进行文件解析，用户可对解析结果进行预览、导出。

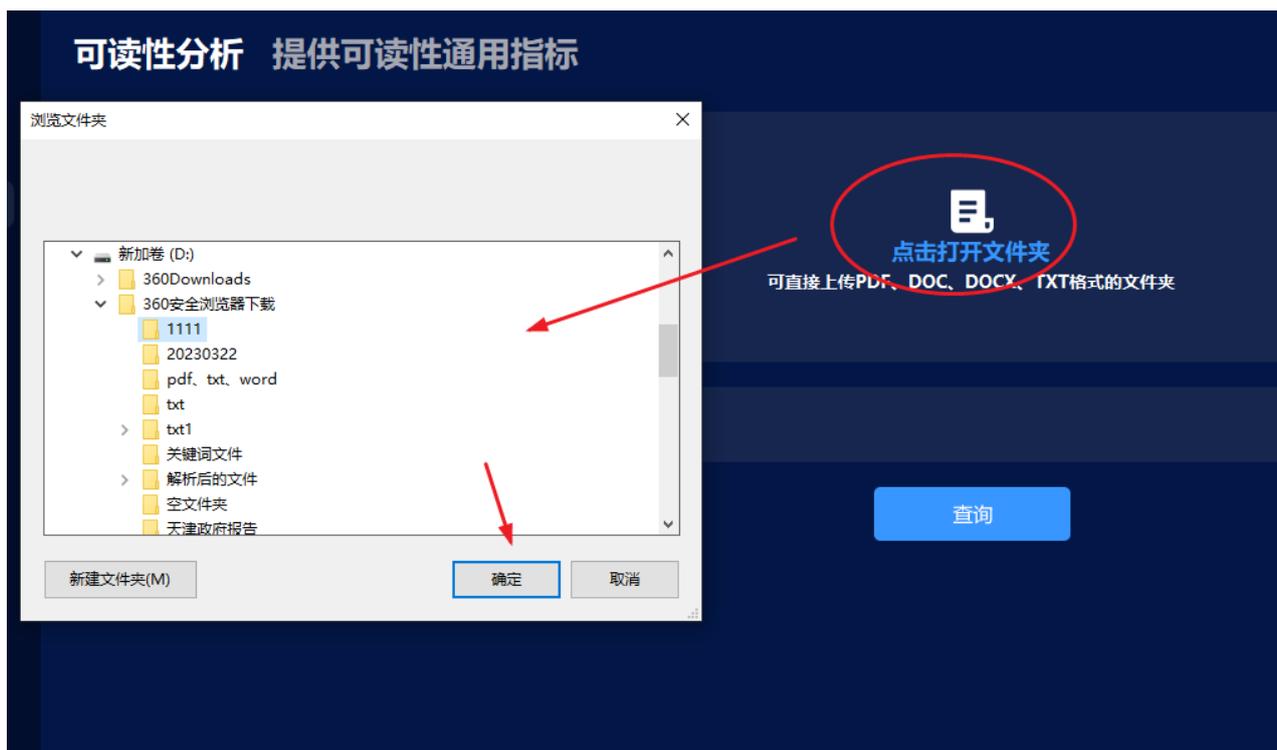


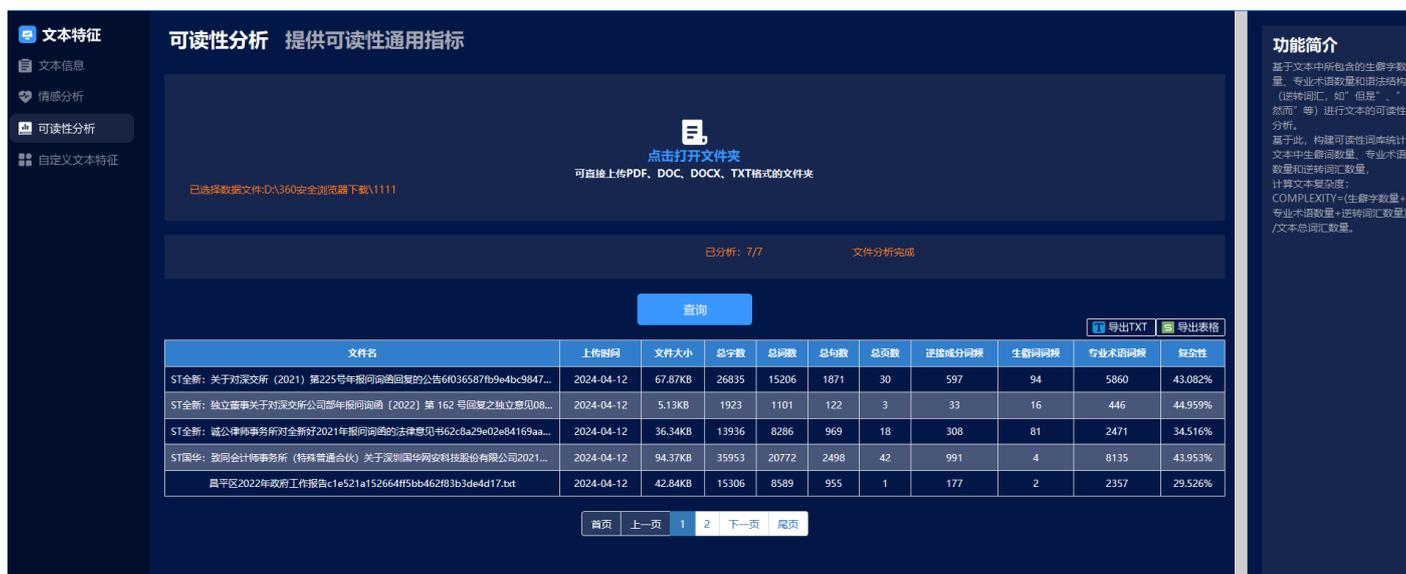
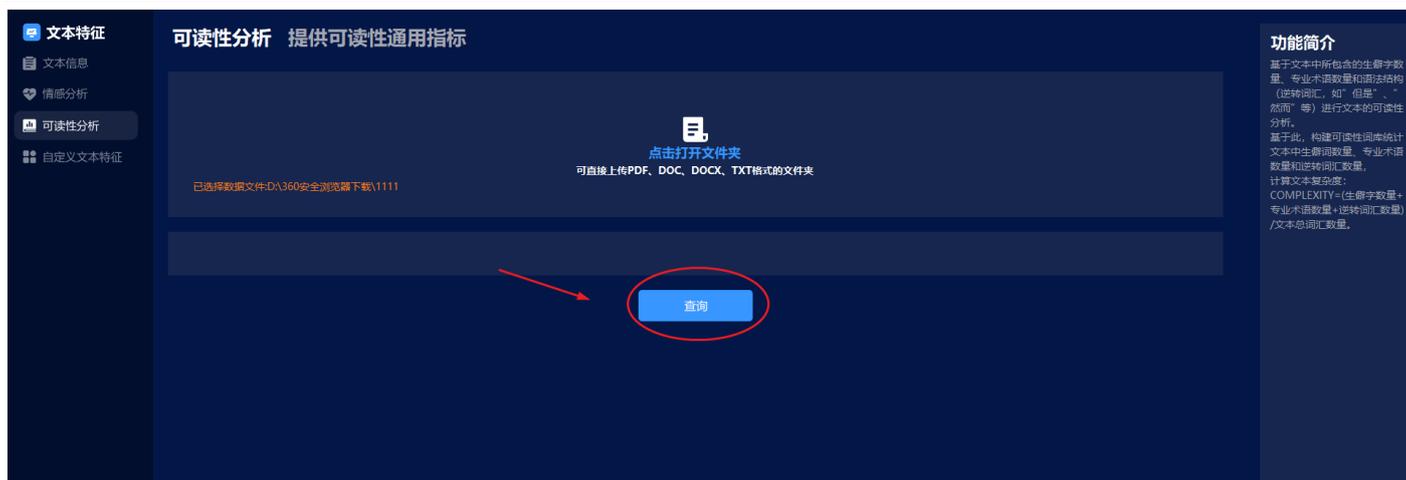
7.3 可读性分析



功能简介：基于文本中所包含的生僻字数量、专业术语数量和语法结构（逆转词汇，如“但是”、“然而”等）进行文本的可读性分析。基于此，构建可读性词库统计文本中生僻词数量、专业术语数量和逆转词汇数量，计算文本复杂度： $COMPLEXITY=(生僻字数量+专业术语数量+逆转词数量)/文本总词汇数量$ 。

功能操作：点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，上传完成后，点击查询按钮进行文件解析，用户可对解析结果进行预览、导出。



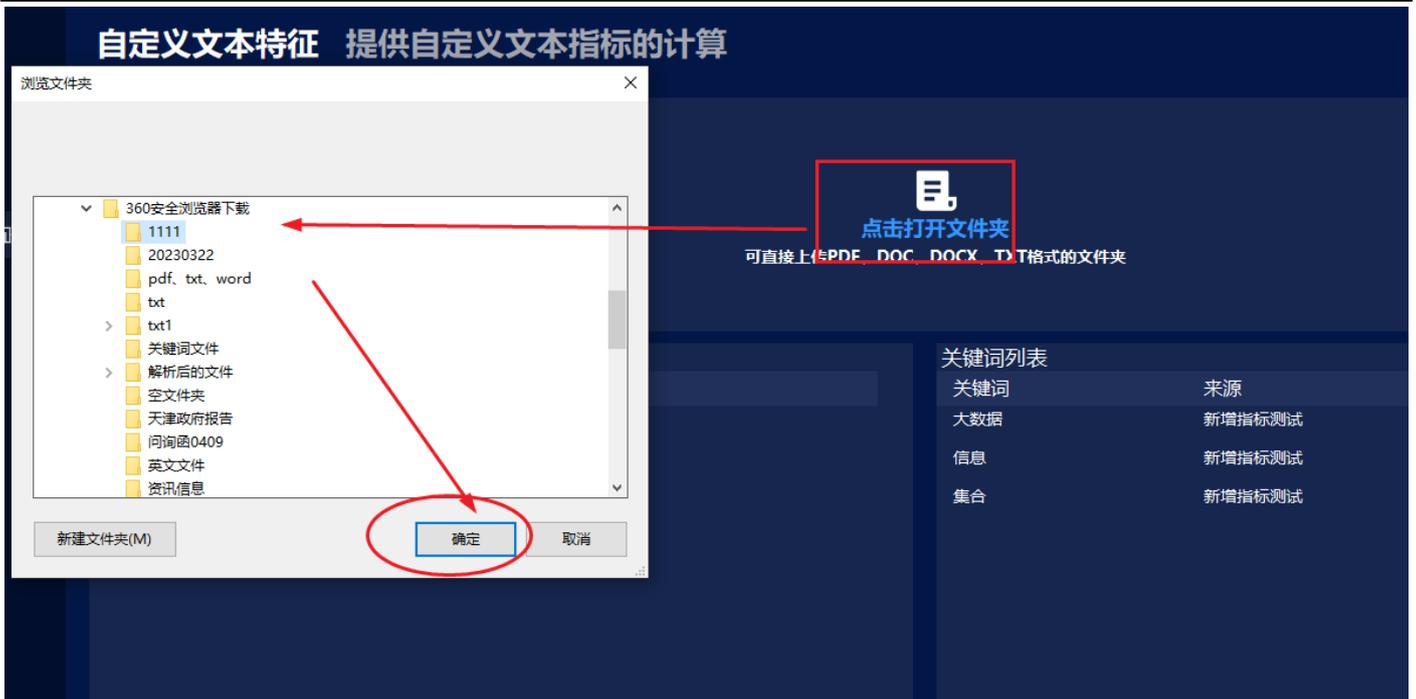


7.4 自定义文本特征

自定义文本特征提供“词典定制”和“特征计算”两大模块。其中，“特征词典定制”模块用于实际自定义指标构建，通过“种子词集输入”完成用户自定义文本指标的词典构建；“特征计算”模块为用户提供文本数据针对已有自定义文本指标的计算和下载功能。

功能步骤：

- ① 选中点击打开文件夹，上传需要解析的文档；



② 创建指标：用户可以通过在输入框内输入指标名称；

③ 点击添加指标按钮进行添加；



- ④ 选择新添加指标，用户可以通过在右侧输入框内输入单个或多个（用英文逗号隔开）关键词，点击, 关键词则显示在关键词列表区域内。





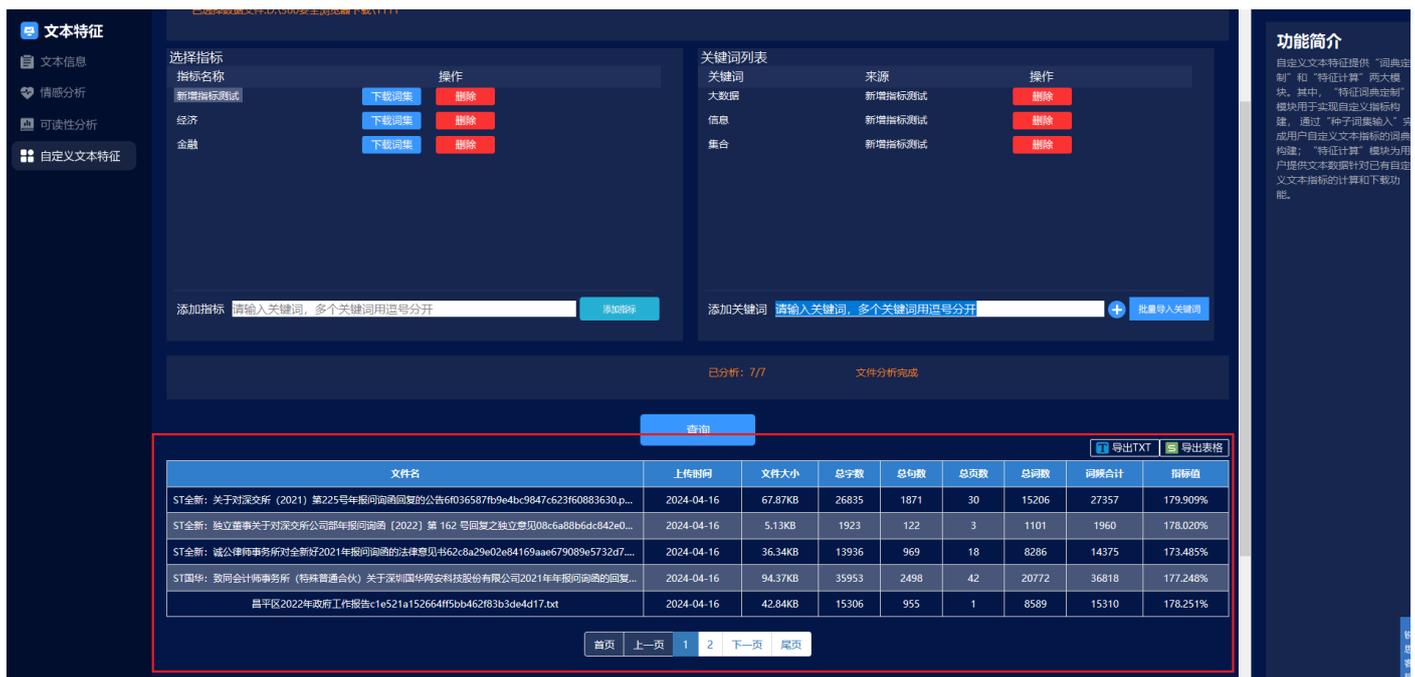
以上操作完成后，点击查询按钮进行文档解析，支持结果导出下载功能。



文档解析中：



结果展示:



7.5 特色指标



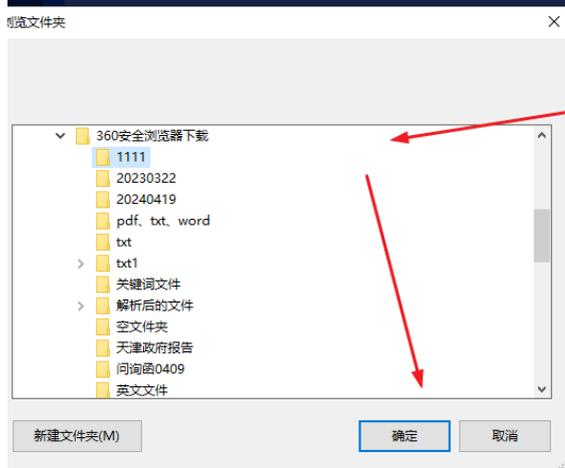
7.5.1 描述性创新

创新指标又称为描述性创新，是衡量企业创新水平的文本指标。描述性创新指标刻画了与技术创新有关的投入和产出描述性创新信息。创新作为企业的核心竞争力，是衡量企业价值和可持续发展能力的重要，指标创新性指标根据“创新”相关词汇总词频在公告中文本总词频的占比计算得出。



功能简介： 点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，小锐助手可将上传的文件转换成方便计算机处理的 TXT 格式文件进行解析。

描述性创新 统计分析得出创新因子指标



点击打开文件夹

可直接上传PDF、DOC、DOCX、TXT格式的文件夹

查询

描述性创新 统计分析得出创新因子指标



点击打开文件夹

可直接上传PDF、DOC、DOCX、TXT格式的文件夹

已选择数据文件: D:\360安全浏览器下载\1111

已分析: 7/7

文件分析完成

查询

导出TXT 导出表格

文件名	上传时间	文件大小	总字数	总句数	总页数	总词数	词频合计	指标值
ST全新: 关于对深交所(2021)第225号年报问询回复的公告6f036587fb9e4bc9847c623f60883630.pdf.txt	2024-05-14	67.87KB	26835	1871	30	15206	2	0.013%
ST全新: 独立董事关于对深交所公司部年报问询函〔2022〕第162号回复之独立意见08c6a88b6dc842e0be1...	2024-05-14	5.13KB	1923	122	3	1101	0	0.000%
ST全新: 诚公律师事务所对全新好2021年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7.pdf...	2024-05-14	36.34KB	13936	969	18	8286	0	0.000%
ST国华: 致同会计师事务所(特殊普通合伙)关于深圳国华网安科技股份有限公司2021年年报问询函的回复d0c...	2024-05-14	94.37KB	35953	2498	42	20772	140	0.674%
昌平区2022年政府工作报告c1e521a152664ff5bb462f83b3de4d17.txt	2024-05-14	42.84KB	15306	955	1	8589	32	0.373%

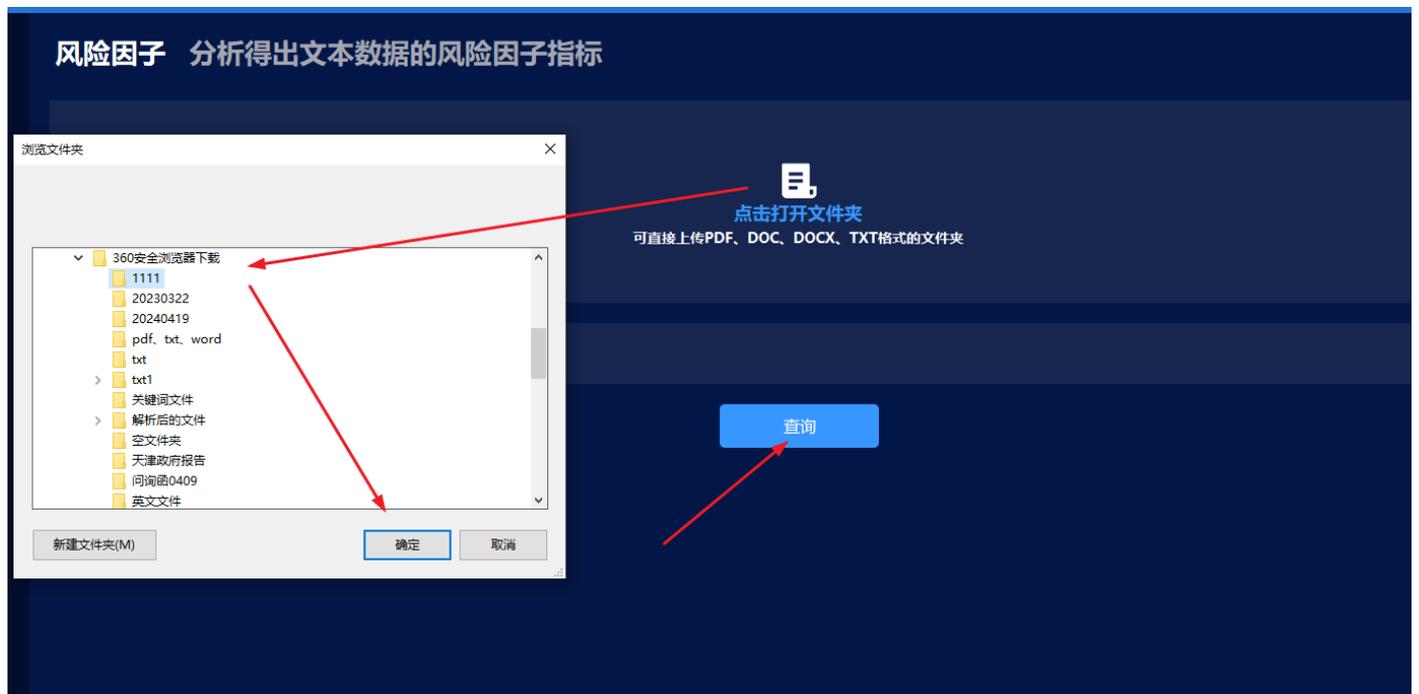
首页 上一页 1 2 下一页 尾页

7.5.2 风险因子

风险信息具有异质性与不确定性，证监会要求公司对风险进行“有意义”的披露，以帮助信息使用者更加准确的预测公司未来现金流，从而做出正确的投资决策。风险指标根据“风险”相关词汇总词频在公告中文本总词频的占比计算得出。



功能简介： 点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，小锐助手可将上传的文件转换成方便计算机处理的 TXT 格式文件进行解析。



风险因子 分析得出文本数据的风险因子指标

点击打开文件夹
可直接上传PDF、DOC、DOCX、TXT格式的文件夹

已选择数据文件:D:\360安全浏览器下载\1111

已分析: 7/7 文件分析完成

查询

导出TXT 导出表格

文件名	上传时间	文件大小	总字数	总句数	总页数	总词数	词频合计	指标值
ST全新: 关于对深交所 (2021) 第225号年报问询函回复的公告6f036587fb9e4bc9847c623f60883630.pdf.txt	2024-05-14	67.87KB	26835	1871	30	15206	254	1.670%
ST全新: 独立董事关于对深交所公司部年报问询函 (2022) 第 162 号回复之独立意见08c6a88b6dc842e0be1...	2024-05-14	5.13KB	1923	122	3	1101	23	2.089%
ST全新: 诚公律师事务所对全新好2021年年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7.pdf...	2024-05-14	36.34KB	13936	969	18	8286	102	1.231%
ST国华: 致同会计师事务所 (特殊普通合伙) 关于深圳国华网安科技股份有限公司2021年年报问询函的回复d0c...	2024-05-14	94.37KB	35953	2498	42	20772	196	0.944%
昌平区2022年政府工作报告c1e521a152664ff5bb462f83b3de4d17.txt	2024-05-14	42.84KB	15306	955	1	8589	32	0.373%

首页 上一页 1 2 下一页 尾页

7.5.3 竞争性因子

企业竞争战略指的是在企业总体战略的制约下，指导和管理具体战略经营单位的计划和行动。它是企业进行经营管理和内部资源配置的指南针，竞争战略的选择决定了企业最终表现结果的质量。竞争性指标根据“竞争性”相关词汇总词频在公告中文本总词频的占比计算得出。

竞争性因子 分析得出文本数据的竞争性因子数据指标

点击打开文件夹
可直接上传PDF、DOC、DOCX、TXT格式的文件夹

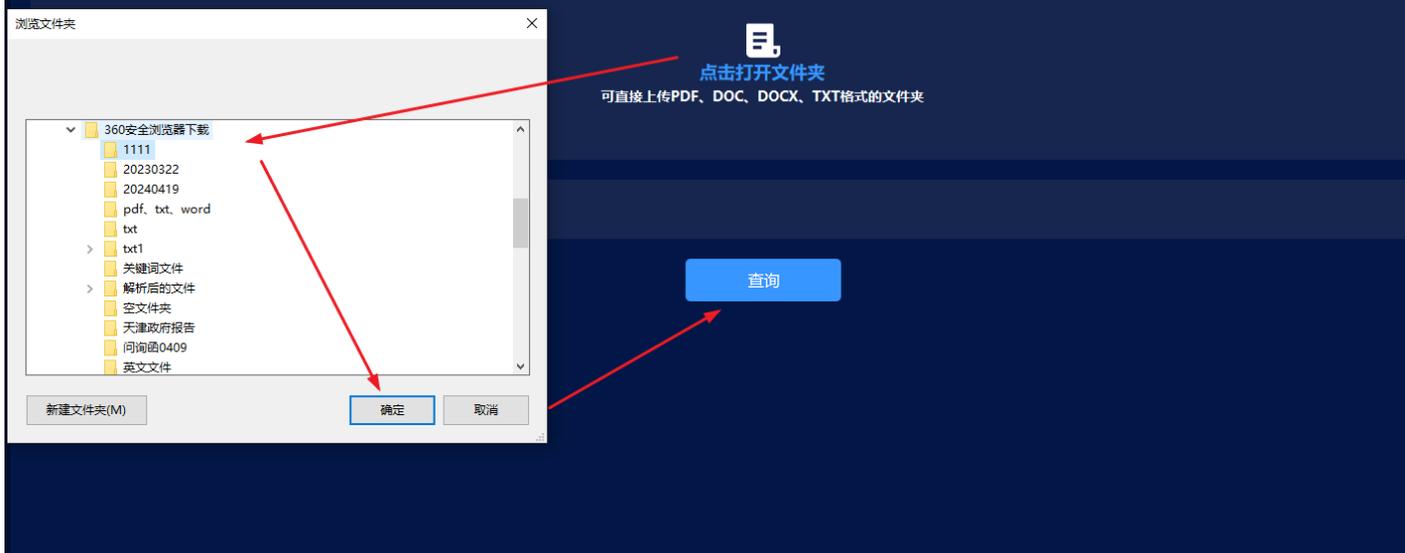
查询

功能简介

企业竞争战略指的是在企业总体战略的制约下，指导和管理具体战略经营单位的计划和行动。它是企业进行经营管理和内部资源配置的指南针，竞争战略的选择决定了企业最终表现结果的质量。竞争性指标根据“竞争性”相关词汇总词频在公告中文本总词频的占比计算得出。

功能简介： 点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，小锐助手可将上传的文件转换成方便计算机处理的 TXT 格式文件进行解析。

竞争性因子 分析得出文本数据的竞争性因子数据指标



竞争性因子 分析得出文本数据的竞争性因子数据指标

已选择数据文件: D:\360安全浏览器下载\1111

已分析: 7/7 文件分析完成

查询

导出TXT 导出表格

文件名	上传时间	文件大小	总字数	总句数	总页数	总词数	词频合计	指标值
ST全新: 关于对深交所(2021)第225号年报问询回复的公告6f036587fb9e4bc9847c623f60883630.pdf.txt	2024-05-14	67.87KB	26835	1871	30	15206	32	0.210%
ST全新: 独立董事关于对深交所公司部年报问询函〔2022〕第162号回复之独立意见08c6a88b6dc842e0be1...	2024-05-14	5.13KB	1923	122	3	1101	1	0.091%
ST全新: 诚公律师事务所对全新好2021年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7.pdf...	2024-05-14	36.34KB	13936	969	18	8286	3	0.036%
ST国华: 致同会计师事务所(特殊普通合伙)关于深圳国华网安科技股份有限公司2021年年报问询函的回复d0c...	2024-05-14	94.37KB	35953	2498	42	20772	193	0.929%
昌平区2022年政府工作报告c1e521a152664ff5bb462f83b3de4d17.txt	2024-05-14	42.84KB	15306	955	1	8589	188	2.189%

首页 上一页 1 2 下一页 尾页

7.5.4 前瞻性因子

前瞻性，指文本中包含与公司未来发展和业绩相关的信息含量。对投资者了解公司前景、预测发展趋势以及评估未来公司价值具有极大帮助。前瞻性指标根据“前瞻性”相关词汇总词频在公告中文本总词频的占比计算得出。



功能简介： 点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，小锐助手可将上传的文件转换成方便计算机处理的 TXT 格式文件进行解析。



前瞻性因子 分析得出文本数据的前瞻性因子数据指标

点击打开文件夹
可直接上传PDF、DOC、DOCX、TXT格式的文件夹

已选择数据文件:D:\360安全浏览器下载\1111

已分析: 7/7 文件分析完成

查询

导出TXT 导出表格

文件名	上传时间	文件大小	总字数	总句数	总页数	总词数	词频合计	指标值
ST全新: 关于对深交所 (2021) 第225号年报问询回复的公告6f036587fb9e4bc9847c623f60883630.pdf.txt	2024-05-14	67.87KB	26835	1871	30	15206	49	0.322%
ST全新: 独立董事关于对深交所公司部年报问询函 [2022] 第 162 号回复之独立意见08c6a88b6dc842e0be1...	2024-05-14	5.13KB	1923	122	3	1101	0	0.000%
ST全新: 诚公律师事务所对全新好2021年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7.pdf...	2024-05-14	36.34KB	13936	969	18	8286	6	0.072%
ST国华: 致同会计师事务所 (特殊普通合伙) 关于深圳国华网安科技股份有限公司2021年年报问询函的回复d0c...	2024-05-14	94.37KB	35953	2498	42	20772	95	0.457%
昌平区2022年政府工作报告c1e521a152664ff5bb462f83b3de4d17.txt	2024-05-14	42.84KB	15306	955	1	8589	63	0.733%

首页 上一页 1 2 下一页 尾页

7.5.5 区块链概念因子

区块链研究数据库是基于对中国上市公司公告进行整理、解析和分析，从而建立的专业数据库。区块链技术的集成应用在新的技术革新和产业变革中起着重要作用，是企业自主创新的重要突破口。本数据库通过分析上市公司在公告中涉及区块链及区块链相似词集的程度。

文本特征 区块链概念因子 分析得出文本数据的区块链概念因子数据指标

点击打开文件夹
可直接上传PDF、DOC、DOCX、TXT格式的文件夹

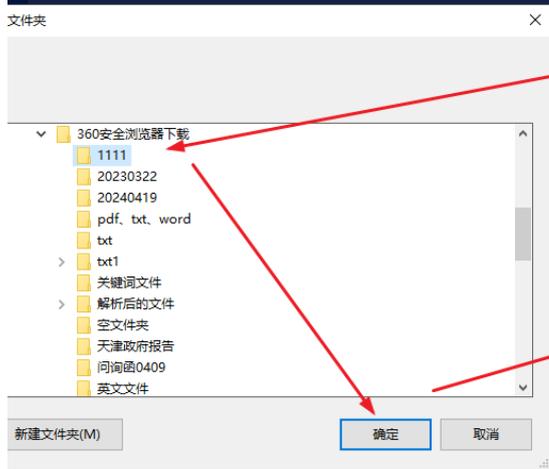
描述性创新
风险因子
竞争性因子
前瞻性因子
区块链概念因子
人口老龄化指标
员工待遇因子
贸易战潜在风险
管理会计因子

功能简介
区块链研究数据库是基于对中国上市公司公告进行整理、解析和分析，从而建立的专业数据库。区块链技术的集成应用在新的技术革新和产业变革中起着重要作用，是企业自主创新的重要突破口。本数据库通过分析上市公司在公告中涉及区块链及区块链相似词集的程度。

查询

功能简介: 点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，小锐助手可将上传的文件转换成方便计算机处理的 TXT 格式文件进行解析。

区块链概念因子 分析得出文本数据的区块链概念因子数据指标



点击打开文件夹

可直接上传PDF、DOC、DOCX、TXT格式的文件夹

查询

区块链概念因子 分析得出文本数据的区块链概念因子数据指标



点击打开文件夹

可直接上传PDF、DOC、DOCX、TXT格式的文件夹

已选择数据文件:D:\360安全浏览器下载\1111

已分析: 7/7

文件分析完成

查询

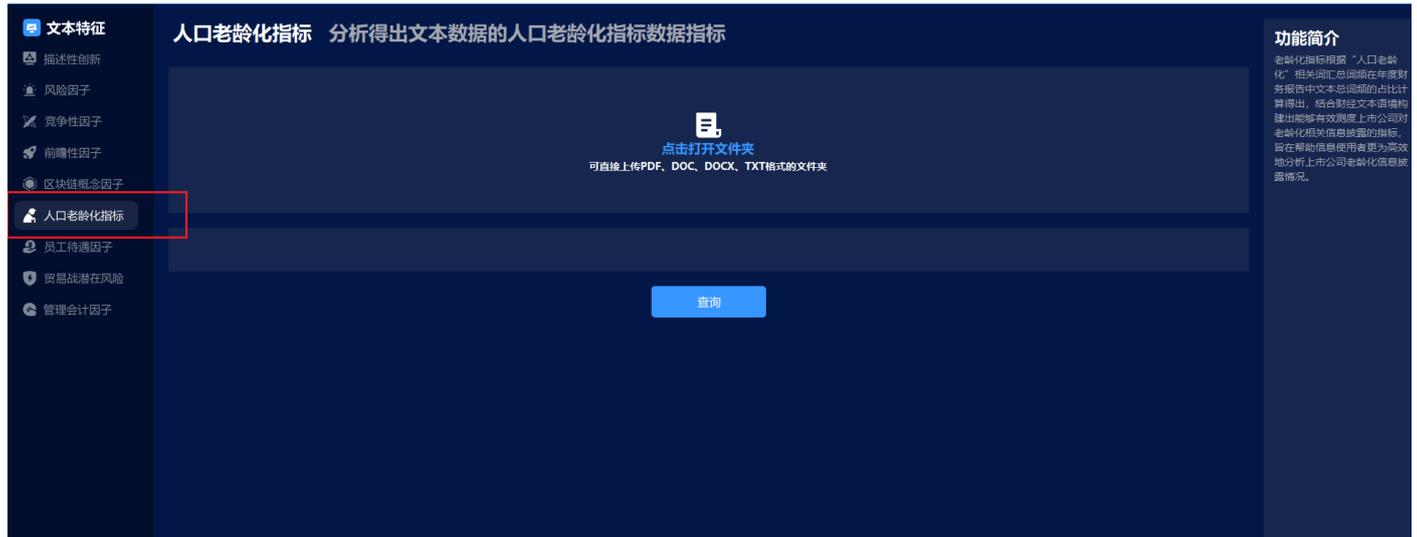
导出TXT

导出表格

文件名	上传时间	文件大小	总字数	总句数	总页数	总词数	词频合计	指标值
ST全新: 关于对深交所 (2021) 第225号年报问询函回复的公告6f036587fb9e4bc9847c623f60883630.pdf.txt	2024-05-14	67.87KB	26835	1871	30	15206	0	0.000%
ST全新: 独立董事关于对深交所公司部年报问询函 (2022) 第 162 号回复之独立意见08c6a88b6dc842e0be1...	2024-05-14	5.13KB	1923	122	3	1101	0	0.000%
ST全新: 诚公律师事务所对全新好2021年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7.pdf...	2024-05-14	36.34KB	13936	969	18	8286	0	0.000%
ST国华: 致同会计师事务所 (特殊普通合伙) 关于深圳国华网安科技股份有限公司2021年年报问询函的回复d0c...	2024-05-14	94.37KB	35953	2498	42	20772	0	0.000%
昌平区2022年政府工作报告c1e521a152664ff5bb462f83b3de4d17.txt	2024-05-14	42.84KB	15306	955	1	8589	1	0.012%

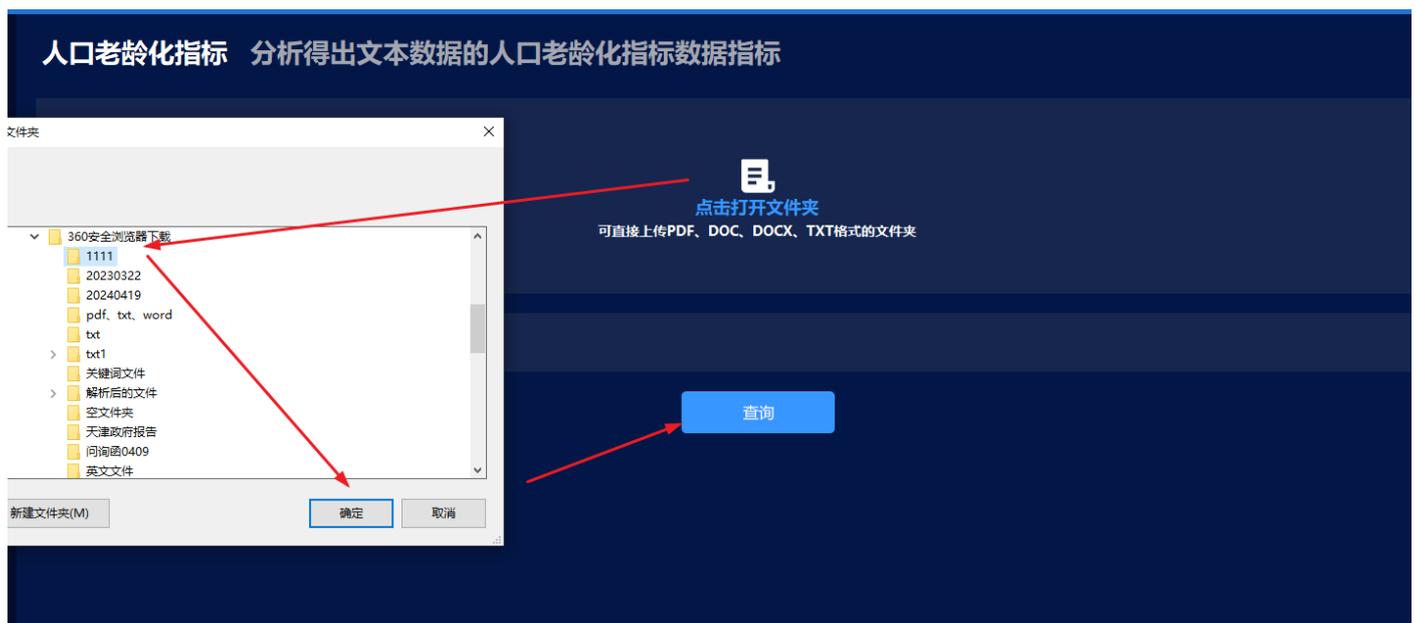
首页 上一页 1 2 下一页 尾页

7.5.6 人口老龄化指标



老龄化指标根据“人口老龄化”相关词汇总词频在年度财务报告中文本总词频的占比计算得出，结合财经文本语境构建出能够有效测度上市公司对老龄化相关信息披露的指标，旨在帮助信息使用者更为高效地分析上市公司老龄化信息披露情况。

功能简介： 点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，小锐助手可将上传的文件转换成方便计算机处理的 TXT 格式文件进行解析。



人口老龄化指标 分析得出文本数据的人口老龄化指标数据指标

点击打开文件夹
可直接上传PDF、DOC、DOCX、TXT格式的文件夹

已选择数据文件:D:\360安全浏览器下载\1111

已分析: 7/7 文件分析完成

查询

导出TXT 导出表格

文件名	上传时间	文件大小	总字数	总句数	总页数	总词数	词频合计	指标值
ST全新: 关于对深交所(2021)第225号年报问询回复的公告6f036587fb9e4bc9847c623f60883630.pdf.txt	2024-05-14	67.07KB	27322	905	30	14500	0	0.000%
ST全新: 独立董事关于对深交所公司部年报问询函【2022】第162号回复之独立意见08c6a88b6dc842e0be1...	2024-05-14	5.10KB	1971	70	3	1096	0	0.000%
ST全新: 诚公律师事务所对全新好2021年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7.pdf...	2024-05-14	35.96KB	14251	515	18	8115	0	0.000%
ST国华: 致同会计师事务所(特殊普通合伙)关于深圳国华网安科技股份有限公司2021年年报问询函的回复d0c...	2024-05-14	92.57KB	35803	1457	42	20089	0	0.000%
昌平区2022年政府工作报告c1e521a152664ff5bb462f83b3de4d17.txt	2024-05-14	42.84KB	15306	955	1	8589	1	0.012%

首页 上一页 1 2 下一页 尾页

7.5.7 员工待遇因子

RESSET 小锐助手

首页 文本数据 词频报表 文本特征 文本挖掘 设置

文本特征 员工待遇因子 分析得出文本数据的员工待遇因子数据指标

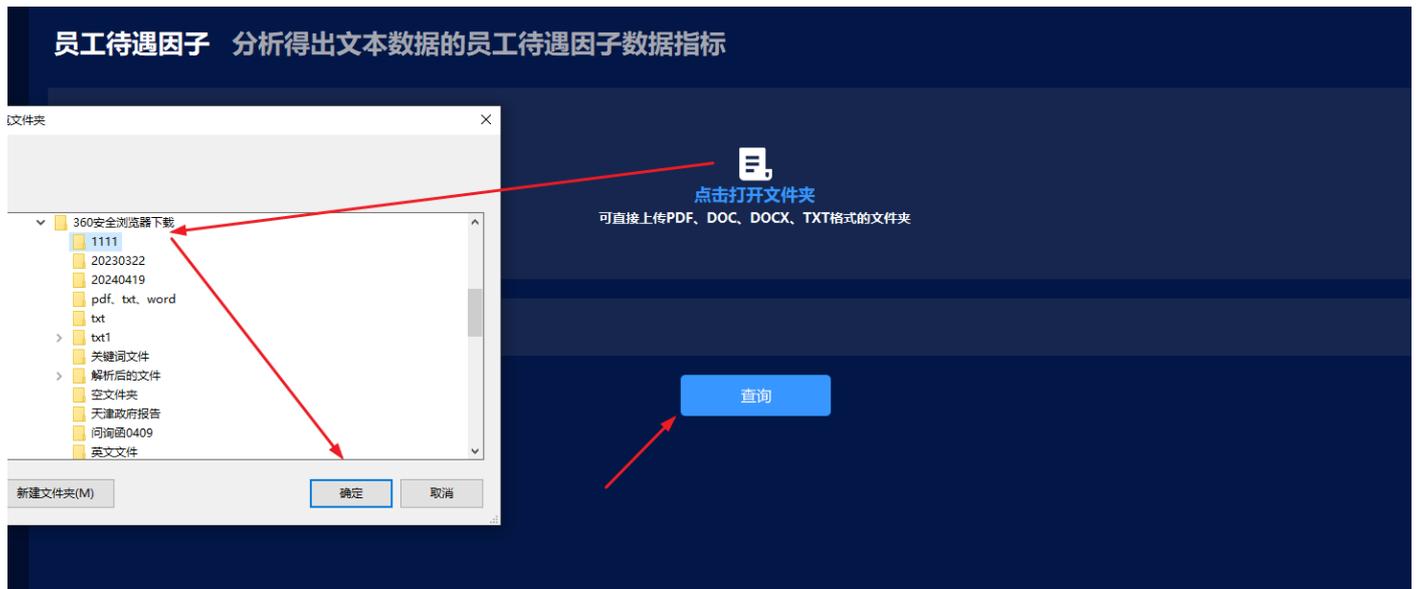
点击打开文件夹
可直接上传PDF、DOC、DOCX、TXT格式的文件夹

查询

功能简介
员工待遇因子是衡量企业对员工友好程度的文本指标。本数据库基于马斯洛需求理论选取出7个员工友好维度进行员工友好度指标的构建。具体来讲，7个员工友好维度包括：基本权益、员工福利、健康与安全、人文关怀、沟通、晋升、自我实现。

员工待遇因子是衡量企业对员工友好程度的文本指标。本数据库基于马斯洛需求理论选取出7个员工友好维度进行员工友好度指标的构建。具体来讲，7个员工友好维度包括：基本权益、员工福利、健康与安全、人文关怀、沟通、晋升、自我实现。

功能简介： 点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，小锐助手可将上传的文件转换成方便计算机处理的 TXT 格式文件进行解析。



7.5.8 贸易战潜在风险



企业微观层面的、随时间变化的、用于衡量中国上市公司感知到的贸易战不确定性指标。

功能简介： 点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，小锐助手可将上传的文件转换成方便计算机处理的 TXT 格式文件进行解析。

贸易战潜在风险 分析得出文本数据的贸易战潜在风险因子数据指标

可直接上传PDF、DOC、DOCX、TXT格式的文件夹

已选择数据文件:D:\360安全浏览器下载\1111

已分析: 7/7 文件分析完成

查询

导出TXT 导出表格

文件名	上传时间	文件大小	总字数	总句数	总页数	总词数	词频合计	指标值
ST全新：关于对深交所（2021）第225号年报问询回复的公告6f036587fb9e4bc9847c623f60883630.pdf.txt	2024-05-14	67.07KB	27322	905	30	14500	0	0.000%
ST全新：独立董事关于对深交所公司部年报问询函〔2022〕第162号回复之独立意见08c6a88b6dc842e0be1...	2024-05-14	5.10KB	1971	70	3	1096	0	0.000%
ST全新：诚公律师事务所对全新好2021年报问询函的法律意见书62c8a29e02e84169aae679089e5732d7.pdf...	2024-05-14	35.96KB	14251	515	18	8115	0	0.000%
ST国华：致同会计师事务所（特殊普通合伙）关于深圳国华网安科技股份有限公司2021年年报问询函的回复d0c...	2024-05-14	92.57KB	35803	1457	42	20089	0	0.000%
昌平区2022年政府工作报告c1e521a152664ff5bb462f83b3de4d17.txt	2024-05-14	42.84KB	15306	955	1	8589	0	0.000%

首页 上一页 1 2 下一页 尾页

7.5.9 管理会计因子

文本特征 管理会计因子 分析得出文本数据的管理会计因子数据指标

描述性创新

风险因子

竞争性因子

前瞻性因子

区块链概念因子

人口老龄化指标

员工待遇因子

贸易战潜在风险

管理会计因子

功能简介

管理会计指标是衡量企业管理会计活动的指标。根据指标相关词汇总词频在年度财务报告中文本总词频的占比计算得出。

可直接上传PDF、DOC、DOCX、TXT格式的文件夹

查询

管理会计指标是衡量企业管理会计活动的指标。根据指标相关词汇总词频在年度财务报告中文本总词频的占比计算得出。

功能简介： 点击打开文件夹位置，可直接上传含 PDF、DOC、DOCX、TXT 格式的文件夹，小锐助手可将上传的文件转换成方便计算机处理的 TXT 格式文件进行解析。

管理会计因子 分析得出文本数据的管理会计因子数据指标

管理会计因子 分析得出文本数据的管理会计因子数据指标

已选择数据文件:D:\360安全浏览器下载\1111

已分析: 7/7 文件分析完成

查询

导出TXT 导出表格

文件名	上传时间	文件大小	总字数	总句数	总页数	总词数	词频合计	战略管理	预算管理	成本管理	运营管理	投融资管理	绩效管理	风险管理
ST全新: 关于对深交所 (2021) 第225号年报问询函回复的公告6f036587f...	2024-05-14	67.07KB	27322	905	30	14500	0	0.000%	0.000%	0.000%	0.021%	0.062%	0.000%	0.000%
ST全新: 独立董事关于对深交所公司部年报问询函 [2022] 第 162 号回复...	2024-05-14	5.10KB	1971	70	3	1096	0	0.000%	0.000%	0.000%	0.000%	0.091%	0.000%	0.000%
ST全新: 诚公律师事务所对全新好2021年报问询函的法律意见书62c8a29e...	2024-05-14	35.96KB	14251	515	18	8115	0	0.000%	0.000%	0.000%	0.025%	0.074%	0.000%	0.000%
ST国华: 致同会计师事务所 (特殊普通合伙) 关于深圳国华网安科技股份有...	2024-05-14	92.57KB	35803	1457	42	20089	0	0.000%	0.040%	0.000%	0.065%	0.005%	0.010%	0.000%
昌平区2022年政府工作报告c1e521a152664ff5bb462f83b3de4d17.txt	2024-05-14	42.84KB	15306	955	1	8589	1	0.000%	0.070%	0.000%	0.023%	0.000%	0.035%	0.012%

首页 上一页 1 2 下一页 尾页

8 文本挖掘

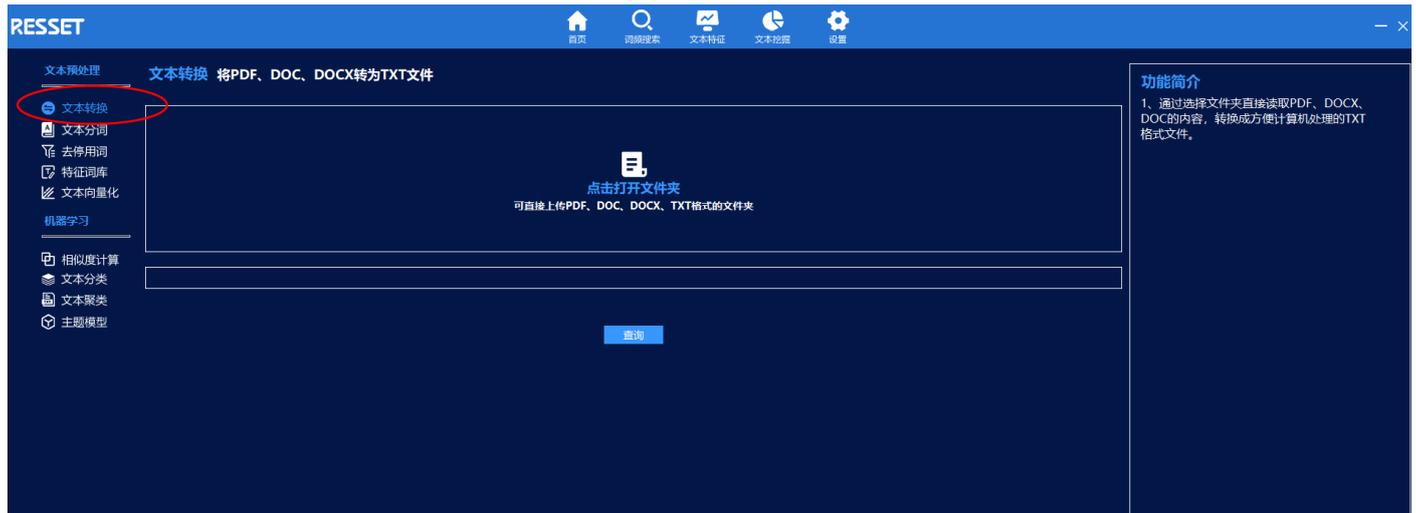


8.1 文本预处理

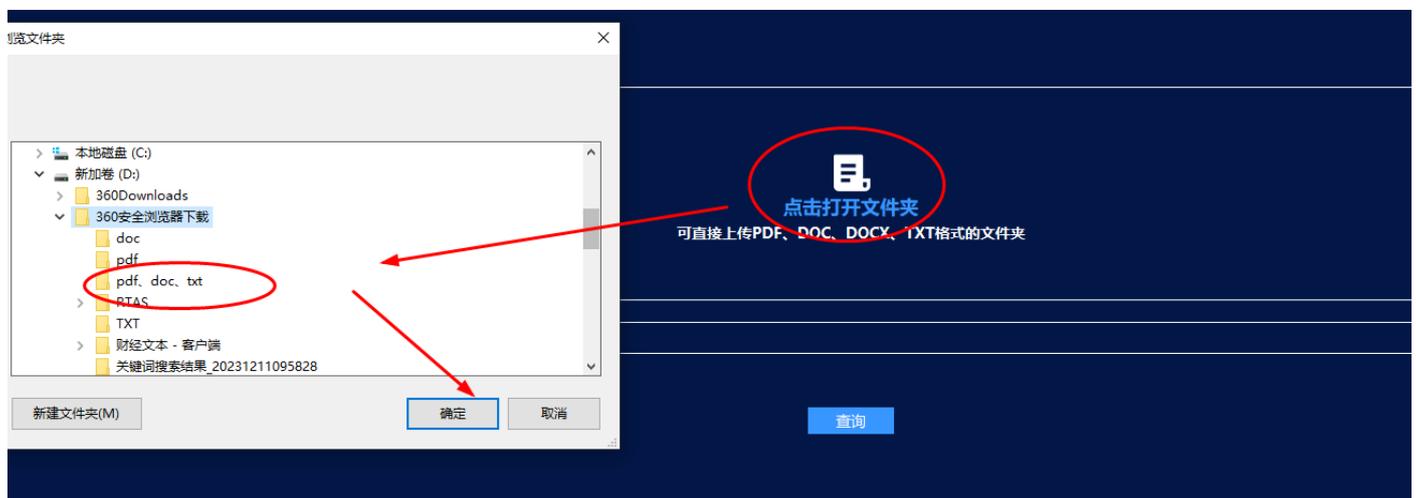


文本预处理：本步骤是文本挖掘的预处理过程，每一个小步骤都可以得到对应的结果文件，后一步可以根据前一步的分析结果文件进行设定，也可以自定义进行设定，既支持多指标流程化也支持独立分析单个指标。

8.1.1 文本转换



功能简介：通过选择文件夹直接读取 PDF、DOCX、DOC 的内容，转换成方便计算机处理的 TXT 格式文件。



文本预处理 **文本转换** 将PDF、DOC、DOCX转为TXT文件

文本转换
文本分词
去停用词
特征词库
文本向量化

机器学习
相似度计算
文本分类
文本聚类
主题模型

点击打开文件夹
可直接上传PDF、DOC、DOCX、TXT格式的文件夹

已选择数据文件:D:\360安全浏览器下载\pdf、doc、txt

已分析: 5/5 文件分析完成

查询

导出TXT 导出表格

文件名	上传时间	文件大小	总字数	总词数	总句数	总页数
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	2023-12-18	23.01KB	11357	5163	459	5
会计教学中批判性思维教学法运用策略分析.pdf.txt	2023-12-18	11.47KB	4147	2568	271	1
初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	2023-12-18	2.36KB	832	832	243	1
技术参数.docx.txt	2023-12-18	4.81KB	1815	1083	109	1
本科生批判性思维能力及高影...双一流”建设高校的实证研究.pdf.txt	2023-12-18	39.12KB	17730	8860	467	8

首页 上一页 1 下一页 尾页

8.1.2 文本分词

功能简介：上传需要的文本文件，自动输出分词结果。

分词模型分文三种：**精准模式分词、全模式分词和搜索引擎模式。**

》**精准模式分词：**将文本精确地切分成词语，适用于文本分析和挖掘等任务。在精确模式下，会对待分析的句子进行全词匹配，找出所有可能的词语，然后依据词典中的词语频率计算出最可能切分结果，分词结果不存在冗余。

文本预处理 **文本分词** 提供可视化分词工具，自动得出分词结果

文本转换
文本分词
去停用词
特征词库
文本向量化

机器学习
相似度计算
文本分类
文本聚类
主题模型

点击打开文件夹
可直接上传TXT格式的文件夹

已选择数据文件:F:\Program Files\小锐助手\customdata\6d8996fa1c5349b9969f18244408b88d

精准模式分词 全模式分词 搜索引擎分词

查询

文件为上一步上传解析后的结果文件。

文本预处理 **文本分词** 提供可视化分词工具，自动得出分词结果

文本转换
 文本分词
 去停用词
 特征词库
 文本向量化
 机器学习
 相似度计算
 文本分类
 文本聚类
 主题模型

精准模式分词 全模式分词 搜索引擎分词

已分析: 5/5 文件分析完成

查询

文件名	上传时间	文件路径	文件大小	总字数	总词数
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	2023-12-18	customdata/29b2f6861a3a492b8a60df4622326226/PBL联合CBS教学模式对...实...	23.01KB	11357	5163
会计教学中批判性思维教学法运用策略分析.pdf.txt	2023-12-18	customdata/29b2f6861a3a492b8a60df4622326226/会计教学中批判性思维教学...	11.47KB	4147	2568
初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	2023-12-18	customdata/29b2f6861a3a492b8a60df4622326226/初中物理科学探究中培养学...	2.36KB	832	832
技术参数.docx.txt	2023-12-18	customdata/29b2f6861a3a492b8a60df4622326226/技术参数.docx.txt	4.81KB	1815	1083
本科生批判性思维能力及高影...双一流”建设高校的实证研究.pdf.txt	2023-12-18	customdata/29b2f6861a3a492b8a60df4622326226/本科生批判性思维能力及高...	39.12KB	17730	8860

》**全模式分词**：将文本中所有可能词语都分解出来，适用于一些对速度要求比较高的场景，分词结果存在冗余。例如：“中华人民共和国”，分词结果为：“中华”，“中华人民”，“中华人民共和国”，“华人”，“人民”等。

文本预处理 **文本分词** 提供可视化分词工具，自动得出分词结果

文本转换
 文本分词
 去停用词
 特征词库
 文本向量化
 机器学习
 相似度计算
 文本分类
 文本聚类
 主题模型

精准模式分词 全模式分词 搜索引擎分词

已分析: 5/5 文件分析完成

查询

文件名	上传时间	文件路径	文件大小	总字数	总词数
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	2023-12-18	customdata/1a145fcc6da3439d884c7e7b99db8b18/PBL联合CBS教学模式对.....	23.01KB	11357	5826
会计教学中批判性思维教学法运用策略分析.pdf.txt	2023-12-18	customdata/1a145fcc6da3439d884c7e7b99db8b18/会计教学中批判性思维教学...	11.47KB	4147	3082
初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	2023-12-18	customdata/1a145fcc6da3439d884c7e7b99db8b18/初中物理科学探究中培养学...	2.36KB	832	791
技术参数.docx.txt	2023-12-18	customdata/1a145fcc6da3439d884c7e7b99db8b18/技术参数.docx.txt	4.81KB	1815	1237
本科生批判性思维能力及高影...双一流”建设高校的实证研究.pdf.txt	2023-12-18	customdata/1a145fcc6da3439d884c7e7b99db8b18/本科生批判性思维能力及高...	39.12KB	17730	9873

》**搜索引擎模式**：在精确模式的基础上，对长词再次使用全模式进行切分，最终得到最可能的切分结果，适用于搜索引擎等场景。

文本预处理 **文本分词** 提供可视化分词工具，自动得出分词结果

文本转换
 文本分词
 去停用词
 特征词库
 文本向量化
 机器学习

精准模式分词 全模式分词 搜索引擎分词

已分析: 5/5 文件分析完成

查询

文件名	上传时间	文件路径	文件大小	总字数	总词数
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	2023-12-18	customdata/31a462f731eb422ab283ea22c91c8c20/PBL联合CBS教学模式对...实...	23.01KB	11357	5758
会计教学中批判性思维教学法运用策略分析.pdf.txt	2023-12-18	customdata/31a462f731eb422ab283ea22c91c8c20/会计教学中批判性思维教学...	11.47KB	4147	2795
初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	2023-12-18	customdata/31a462f731eb422ab283ea22c91c8c20/初中物理科学探究中培养学...	2.36KB	832	832
技术参数.docx.txt	2023-12-18	customdata/31a462f731eb422ab283ea22c91c8c20/技术参数.docx.txt	4.81KB	1815	1119
本科生批判性思维能力及高影...双一流”建设高校的实证研究.pdf.txt	2023-12-18	customdata/31a462f731eb422ab283ea22c91c8c20/本科生批判性思维能力及高...	39.12KB	17730	9906

8.1.3 去停用词

功能简介: 在文本分析和文本挖掘的任务中，只关注有价值、有意义的文本信息，去停用词是为了将一些无价值意义的词语（也被称为停用词）剔除，停用词不仅对挖掘文本信息没有帮助，而且会降低文本挖掘的效率，增加数据中无用的噪声。

上传文件使用文本转换后的文件，用户也可以重新上传新的文件进行解析。

文本预处理 **去停用词** 智能剔除分词后的停用词，得到全新分词结果

文本转换
 文本分词
 去停用词
 特征词库
 文本向量化
 机器学习

精准模式分词 全模式分词 搜索引擎分词

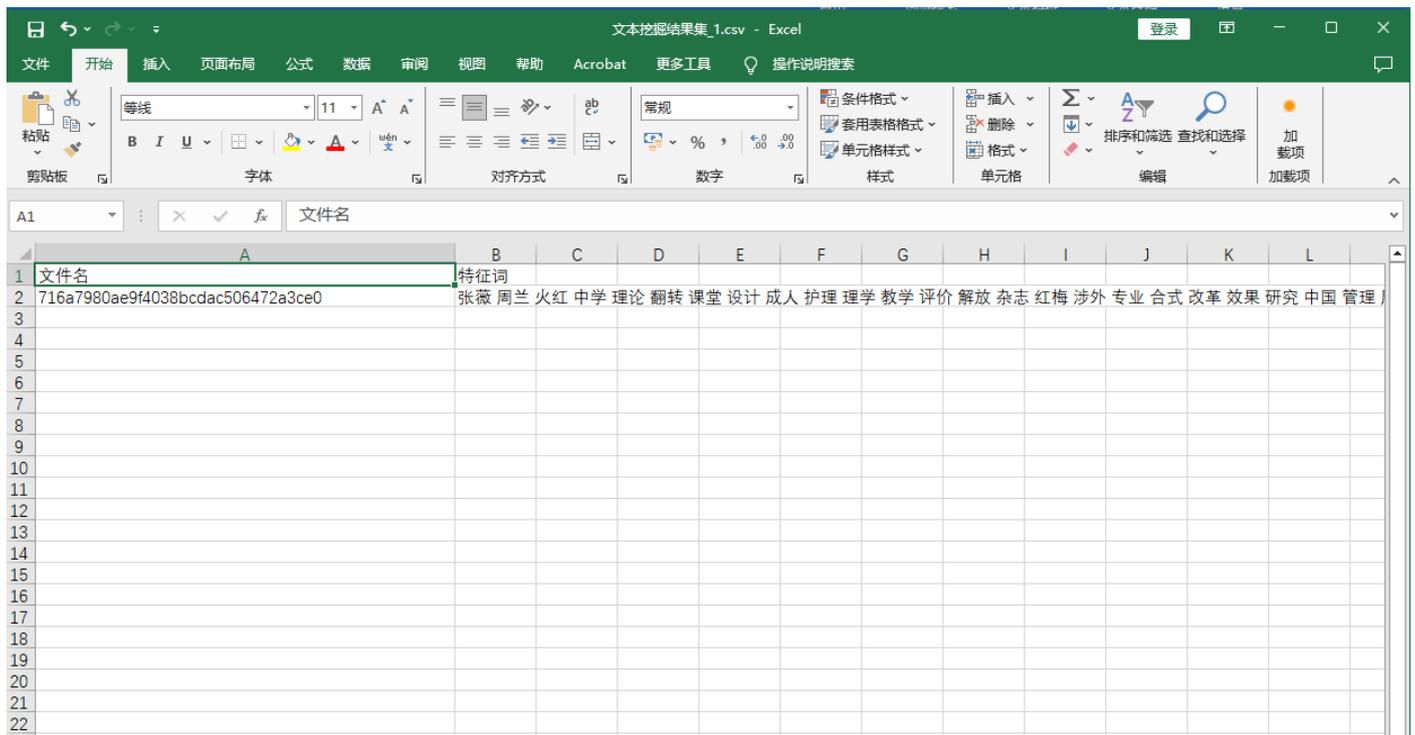
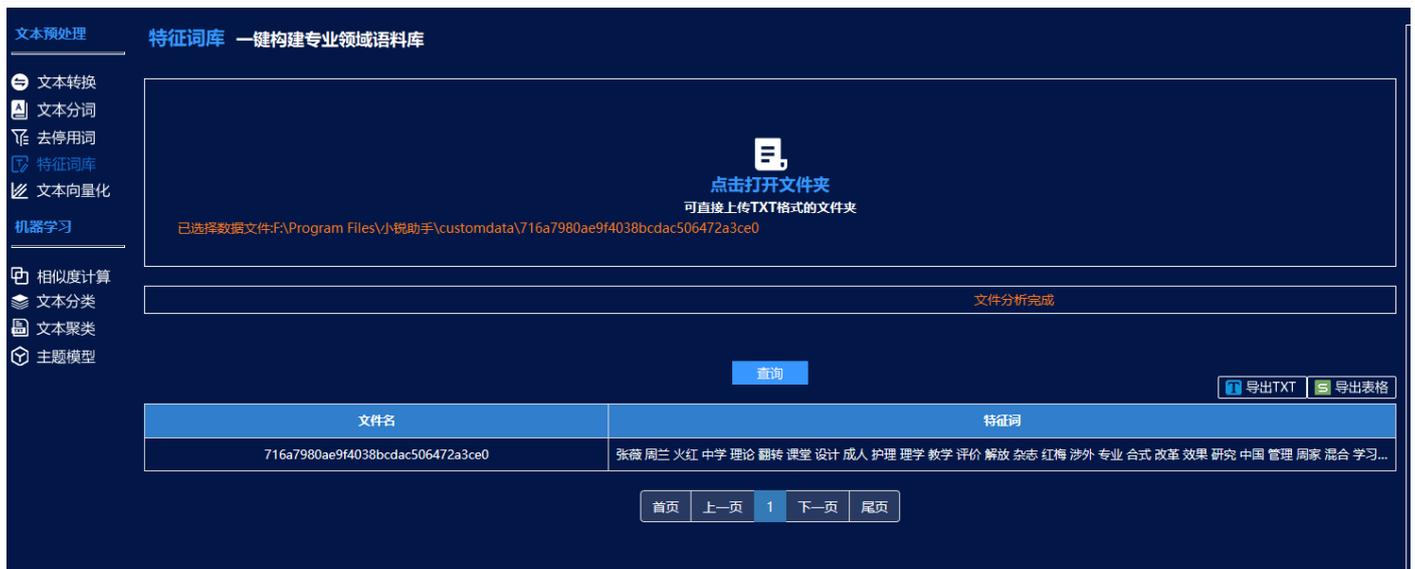
已分析: 5/5 文件分析完成

查询

文件名	上传时间	文件路径	文件大小	总字数	总词数
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	2023-12-18	customdata/716a7980ae9f4038bcdac506472a3ce0/PBL联合CBS教学模式对...实...	32.56KB	12542	2975
会计教学中批判性思维教学法运用策略分析.pdf.txt	2023-12-18	customdata/716a7980ae9f4038bcdac506472a3ce0/会计教学中批判性思维教学...	15.70KB	4597	1412
初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	2023-12-18	customdata/716a7980ae9f4038bcdac506472a3ce0/初中物理科学探究中培养学...	3.21KB	832	42
技术参数.docx.txt	2023-12-18	customdata/716a7980ae9f4038bcdac506472a3ce0/技术参数.docx.txt	6.12KB	1885	602
本科生批判性思维能力及高影...双一流”建设高校的实证研究.pdf.txt	2023-12-18	customdata/716a7980ae9f4038bcdac506472a3ce0/本科生批判性思维能力及高...	55.97KB	19733	5832

8.1.4 特征词库

功能简介: 特征词指的是在文本中具有一定信息量或表征文本特征的词语，这些词语通常被用来描述文本的主题、情感或其他重要的属性。根据用户上传文件及内嵌特征词库提取文本中的特征词。



8.1.5 文本向量化

功能简介: 用户上传文件合集，根据用户选择的算法计算出文本中词的向量值，进一步可以得到整篇文档的向量表示，可用于后续的文本挖掘计算。

文本预处理 | 文本向量化 | 数值向量来表示文本的语义

- 文本转换
- 文本分词
- 去停用词
- 特征词库
- 文本向量化

机器学习

- 相似度计算
- 文本分类
- 文本聚类
- 主题模型

已选择数据文件:F:\Program Files\小锐助手\customdata\17d3d47cf947b0735db9191eb

TF-IDF TFC LTC 文件分析完成

查询 导出表格

文件名	词	向量值
Feature.txt	张薇	0.0006931471805599453
Feature.txt	周兰	0.0006931471805599453
Feature.txt	火红	0.0006931471805599453
Feature.txt	中学	0.0006931471805599453
Feature.txt	理论	0.0006931471805599453

首页 上一页 1 2 3 4 5 6 7 下一页 尾页

文本预处理 | 文本向量化 | 数值向量来表示文本的语义

- 文本转换
- 文本分词
- 去停用词
- 特征词库
- 文本向量化

机器学习

- 相似度计算
- 文本分类
- 文本聚类
- 主题模型

已选择数据文件:F:\Program Files\小锐助手\customdata\17d3d47cf947b0735db9191eb

TF-IDF TFC LTC 文件分析完成

查询 导出表格

文件名	词	向量值
Feature.txt	一幕	0.03162277660168379
Feature.txt	一步	0.03162277660168379
Feature.txt	一段	0.03162277660168379
Feature.txt	一流	0.03162277660168379
Feature.txt	一节	0.03162277660168379

首页 上一页 1 2 3 4 5 6 7 下一页 尾页

文本预处理 | 文本向量化 | 数值向量来表示文本的语义

文本转换
文本分词
去停用词
特征词库
文本向量化
机器学习
相似度计算
文本分类
文本聚类
主题模型

点击打开文件夹
可直接上传TXT格式的文件夹
已选择数据文件:F:\Program Files\小锐助手\customdata\33f49d17d3d47cf947b0735db9191eb

TF-IDF TFC LTC 文件分析完成

查询 导出表格

文件名	词	向量值
Feature.txt	一幕	0.03162277660168379
Feature.txt	一步	0.03162277660168379
Feature.txt	一段	0.03162277660168379
Feature.txt	一流	0.03162277660168379
Feature.txt	一节	0.03162277660168379

首页 上一页 1 2 3 4 5 6 7 下一页 尾页

8.2 机器学习

RESSET 小锐助手

首页 文本数据 词频搜索 文本特征 文本挖掘 设置

文本挖掘 自动化多维度多文本的文本挖掘

步骤1: 文本预处理
本步骤是文本挖掘的预处理过程, 每一个小步骤都可以得到对应的结果文件, 后一步可以根据前一步的分析结果文件进行设定, 也可以自定义文件进行设定, 既支持多指标流程化也支持独立分析单个指标。

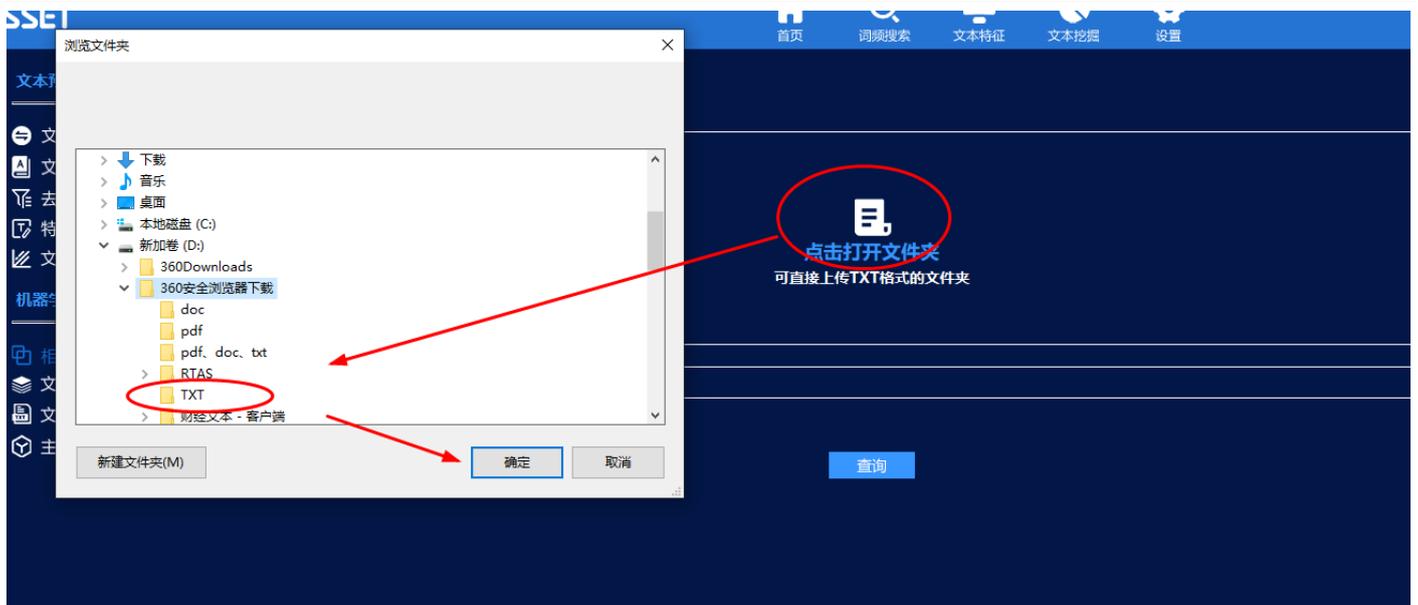
步骤2: 机器学习
本步骤包含文本相似度、文本分类、文本聚类和主题模型等功能, 支持步骤一的无缝结合, 也支持自定义文本进行一键式流程化分析。

立即体验 立即体验 立即体验 立即体验 立即体验 立即体验 立即体验 立即体验

机器学习: 本步骤包含文本相似度、文本分类、文本聚类和主题模型等功能, 支持步骤一的无缝结合, 也支持自定义文本进行一键式流程化分析。

8.2.1 相似度计算

功能简介: 上传文本文件合集, 利用 TF-IDF 算法计算每篇文档中的向量表示, 选择不同的距离公式计算得出所有文档之间的两两相似度。



余弦距离:



欧氏距离:

文本预处理 相似度计算 算得出文件夹中每个文件之间的相似度

余弦距离
 欧氏距离
 Jaccard距离

已分析: 6/18 文件分析中...

查询 导出表格

文本预处理 相似度计算 算得出文件夹中每个文件之间的相似度

余弦距离
 欧氏距离
 Jaccard距离

已分析: 18/18 文件分析完成

查询 导出表格

文件1	文件2	相似度
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	100
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	中学语文教学培养学生思辨能...学语文批判性思维教学案例}.pdf.txt	9.235
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	会计教学中批判性思维教学法运用策略分析.pdf.txt	9.662
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	7.966
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	基于思辨能力培养的中学语文...学语文批判性思维教学案例}.pdf.txt	10.414

首页 上一页 1 2 3 4 5 6 7 下一页 尾页

Jarcard 距离:

文本预处理 相似度计算 算得出文件夹中每个文件之间的相似度

余弦距离
 欧氏距离
 Jarcard距离

文件上传完成!

查询 导出表格

文本预处理 相似度计算 算得出文件夹中每个文件之间的相似度

文本转换
文本分词
去停用词
特征词库
文本向量化
机器学习

相似度计算
文本分类
文本聚类
主题模型

点击打开文件夹
可直接上传TXT格式的文件夹
已选择数据文件:D:\360安全浏览器下载\TXT

■ 余弦距离 ■ 欧式距离 Jaccard距离

已分析: 18/18 文件分析完成

查询

文件1	文件2	相似度
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	100.000
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	中学语文教学培养学生思辨能...学语文批判性思维教学案例》.pdf.txt	2.500
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	会计教学中批判性思维教学法运用策略分析.pdf.txt	2.000
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	0.000
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	基于思辨能力培养的中学语文...学语文批判性思维教学案例》.pdf.txt	2.222

导出表格

首页 上一页 1 2 3 4 5 6 7 下一页 尾页

文件1	文件2	相似度
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	100
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	中学语文教学培养学生思辨能...学语文批判性思维教学案例》.pdf.txt	2.5
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	会计教学中批判性思维教学法运用策略分析.pdf.txt	2
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	0
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	基于思辨能力培养的中学语文...学语文批判性思维教学案例》.pdf.txt	2.222
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	审辩式思维在创造力发展中的重要性.pdf.txt	0
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	审辩式思维培养水平的评估指标研究.pdf.txt	1.37
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	审辩式思维教育具有的现实意义.pdf.txt	0
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	审辩式思维能力及其测量.pdf.txt	1.887
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	审辩式思维观照下的初中文言文教学.pdf.txt	1.818
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	循本源抓关键 提升审辩式思维能力.pdf.txt	0
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	思维的批判性与大学的批判性思维教育.pdf.txt	1.562
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	批判性思维教学：意蕴、特征及建构路径.pdf.txt	1.786
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	批判性思维教育的一个议题：...究？罗素、杜威和图尔敏.pdf.txt	1.493
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	本科生批判性思维能力及高影...双一流”建设高校的实证研究.pdf.txt	1.37
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	浅析审辩式思维的培养策略.pdf.txt	1.923
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	美国斯坦福大学批判性思维教育及启示.pdf.txt	1.818
PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	认知偏差与批判性思维教育.pdf.txt	0
中学语文教学培养学生思辨能...学语文批判性思维教学案例》.pdf.txt	PBL联合CBS教学模式对...实习生批判性思维能力的影响.pdf.txt	2.5
中学语文教学培养学生思辨能...学语文批判性思维教学案例》.pdf.txt	中学语文教学培养学生思辨能...学语文批判性思维教学案例》.pdf.txt	100
中学语文教学培养学生思辨能...学语文批判性思维教学案例》.pdf.txt	会计教学中批判性思维教学法运用策略分析.pdf.txt	3.226
中学语文教学培养学生思辨能...学语文批判性思维教学案例》.pdf.txt	初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	0
中学语文教学培养学生思辨能...学语文批判性思维教学案例》.pdf.txt	基于思辨能力培养的中学语文...学语文批判性思维教学案例》.pdf.txt	2.222

8.2.2 文本分类

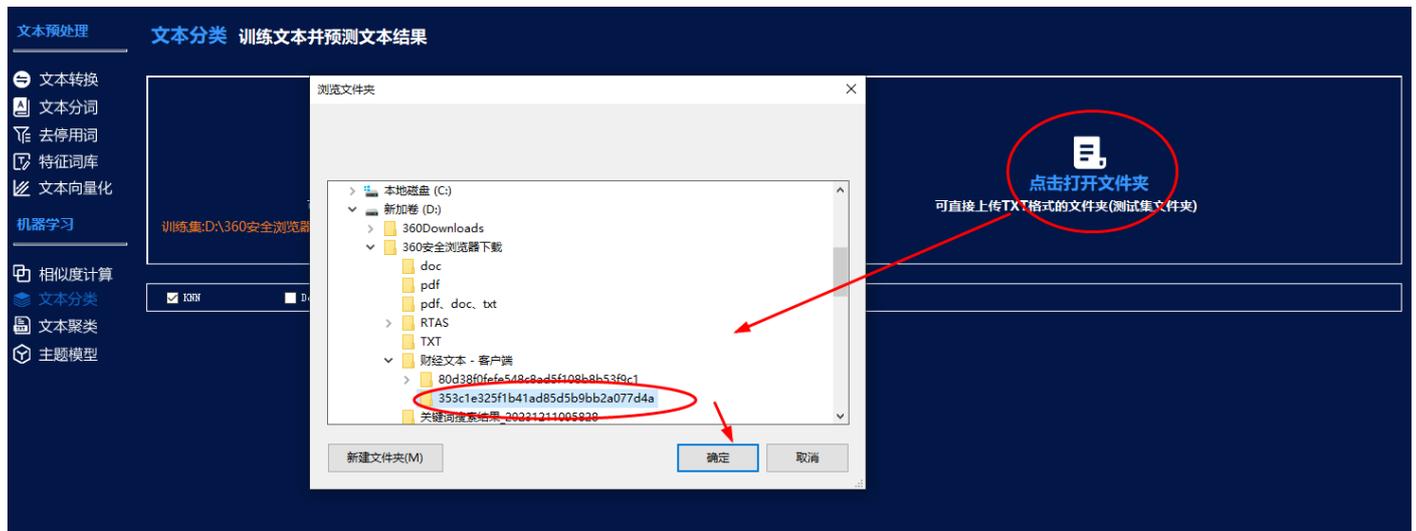
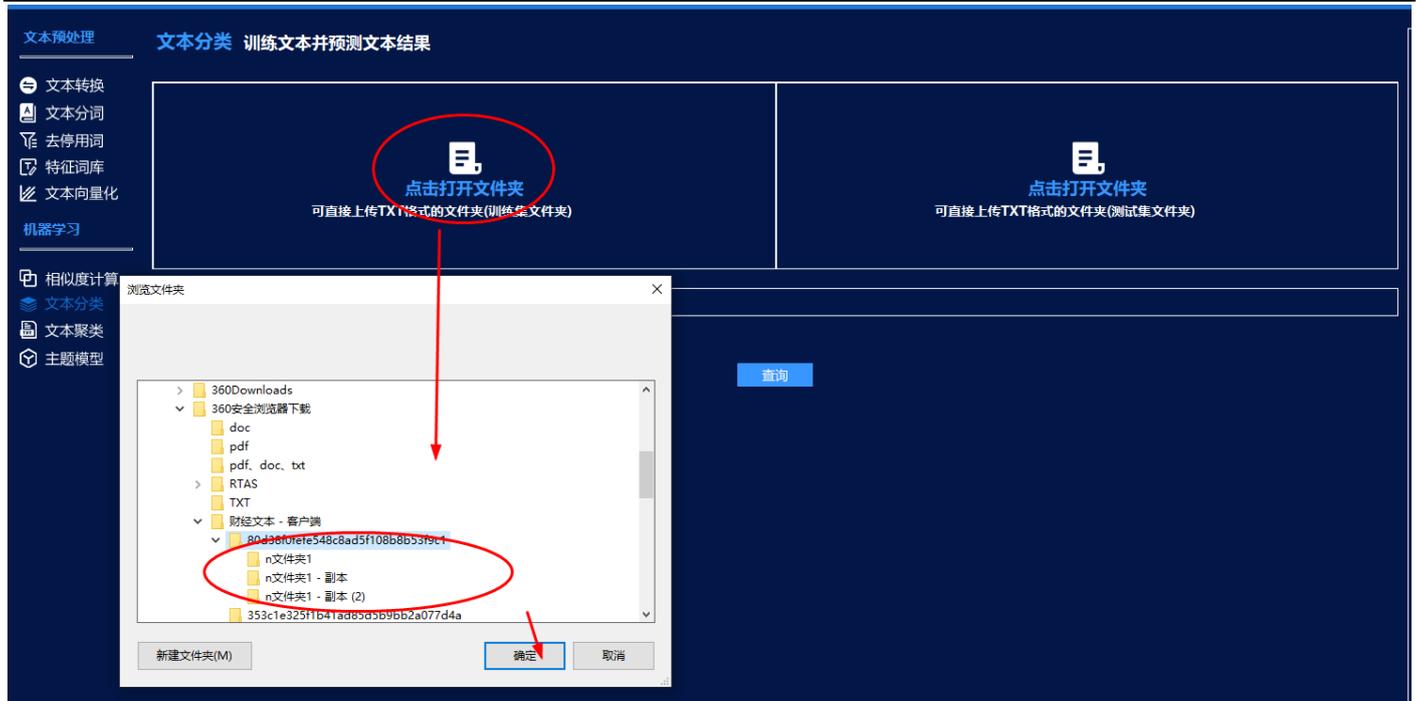
功能简介：

》用户上传训练集文件和测试文件夹。

训练集文件夹：包含若干个文件夹，这些若干文件夹名称为分类名称，每个分类文件夹中包含若干 txt 文件。

测试集文件夹：包含多个带预测的 txt 文件。

》提取文本特征得到文件特征矩阵，根据用户选择的分类算法建立相应模型，同时根据训练集文件夹文本数据训练模型，模型训练完成后输出测试集文件夹下文件的分类结果。



KNN:

文本预处理 文本分类 训练文本并预测文本结果

点击打开文件夹

可直接上传TXT格式的文件夹(训练集文件夹)

训练集:D:\360安全浏览器下载\财经文本 - 客户端\80d38f0fe548c8ad5f108b8b53f9c1

点击打开文件夹

可直接上传TXT格式的文件夹(测试集文件夹)

测试集:D:\360安全浏览器下载\财经文本 - 客户端\353c1e325f1b41ad85d5b9bb2a077d4a

KNN
 DecisionTree
 SVM
 文件分析完成

[查询](#) [导出表格](#)

文件	类别
万科A: 2021年第二季度报告.pdf.txt	n文件夹1 - 副本
乐普医疗: 300003乐普医疗业绩说明会、路演活动信息20210201.pdf.txt	n文件夹1 - 副本
疫情对一带一路沿线国家主要经济指标影响的灰色测度模型构建及预测 (6) - 副本.pdf.txt	n文件夹1 - 副本
疫情对一带一路沿线国家主要经济指标影响的灰色测度模型构建及预测 (7) - 副本.pdf.txt	n文件夹1 - 副本

[首页](#)
[上一页](#)
1
[下一页](#)
[尾页](#)

DecisionTree:

文本预处理 文本分类 训练文本并预测文本结果

点击打开文件夹

可直接上传TXT格式的文件夹(训练集文件夹)

训练集:D:\360安全浏览器下载\财经文本 - 客户端\80d38f0fe548c8ad5f108b8b53f9c1

点击打开文件夹

可直接上传TXT格式的文件夹(测试集文件夹)

测试集:D:\360安全浏览器下载\财经文本 - 客户端\353c1e325f1b41ad85d5b9bb2a077d4a

KNN
 DecisionTree
 SVM
 文件分析完成

[查询](#) [导出表格](#)

文件	类别
万科A: 2021年第二季度报告.pdf.txt	n文件夹1
乐普医疗: 300003乐普医疗业绩说明会、路演活动信息20210201.pdf.txt	n文件夹1 - 副本
疫情对一带一路沿线国家主要经济指标影响的灰色测度模型构建及预测 (6) - 副本.pdf.txt	n文件夹1 - 副本
疫情对一带一路沿线国家主要经济指标影响的灰色测度模型构建及预测 (7) - 副本.pdf.txt	n文件夹1 - 副本

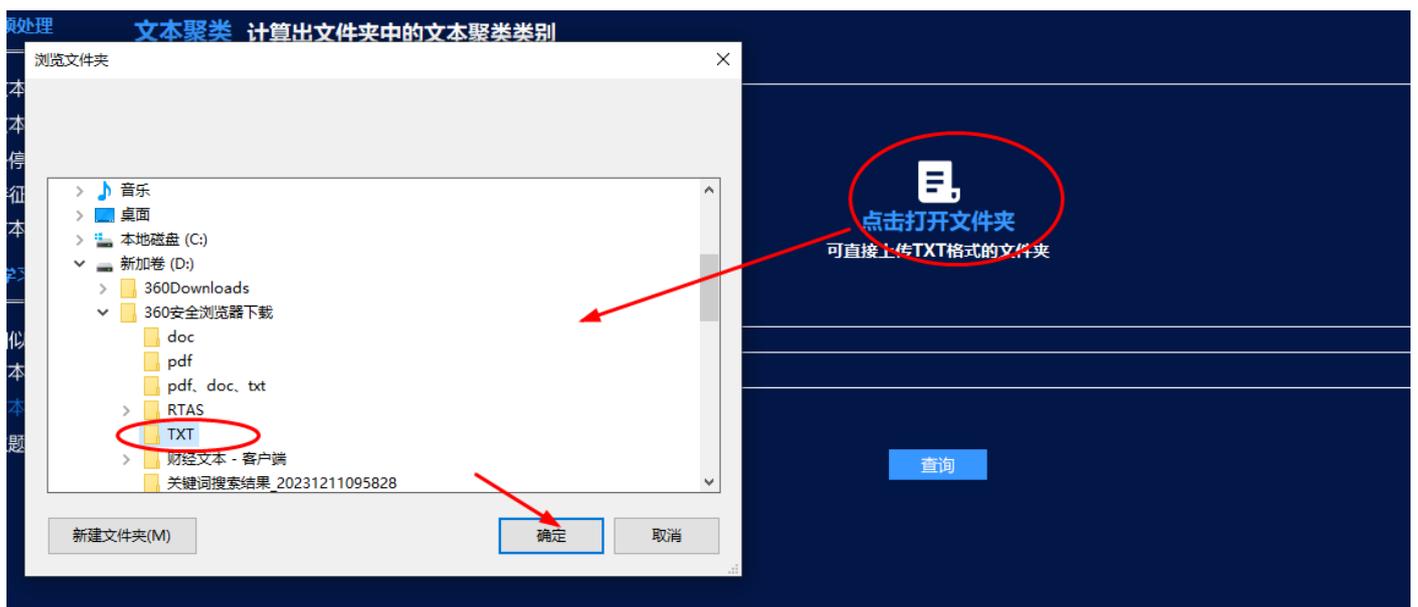
[首页](#)
[上一页](#)
1
[下一页](#)
[尾页](#)

SVM:



8.2.3 文本聚类

功能简介: 对用户上传的数据进行分词处理, 然后利用 tf-idf 算法获取 tf-idf 矩阵, 利用 PCA 选择重要的特征信息, 最后根据用户选择的聚类算法对所有文档进行聚类。



K-Means 算法:

文本预处理 文本聚类 计算出文件夹中的文本聚类类别

文本转换
文本分词
去停用词
特征词库
文本向量化

机器学习
相似度计算
文本分类
文本聚类
主题模型

已选择数据文件:D:\360安全浏览器\下载\TXT

点击打开文件夹
可直接上传TXT格式的文件夹

K-Means算法 DBSCAN算法 文件分析完成

查询

导出表格

文件	类别
PBL联合CBS教学模式对...实习生批判性思维能力的影...pdf.txt	类别1
中学语文教学培养学生思辨能...学语文批判性思维教学案...pdf.txt	类别1
会计教学中批判性思维教学法运用策略分析.pdf.txt	类别1
初中物理科学探究中培养学生...科版8年级“浮力”教...pdf.txt	类别1
基于思辨能力培养的中学语文...学语文批判性思维教学...pdf.txt	类别1

首页 上一页 1 2 3 4 下一页 尾页

DBSCAN 算法:

文本预处理 文本聚类 计算出文件夹中的文本聚类类别

文本转换
文本分词
去停用词
特征词库
文本向量化

机器学习
相似度计算
文本分类
文本聚类
主题模型

已选择数据文件:D:\360安全浏览器\下载\TXT

点击打开文件夹
可直接上传TXT格式的文件夹

K-Means算法 DBSCAN算法 文件分析完成

查询

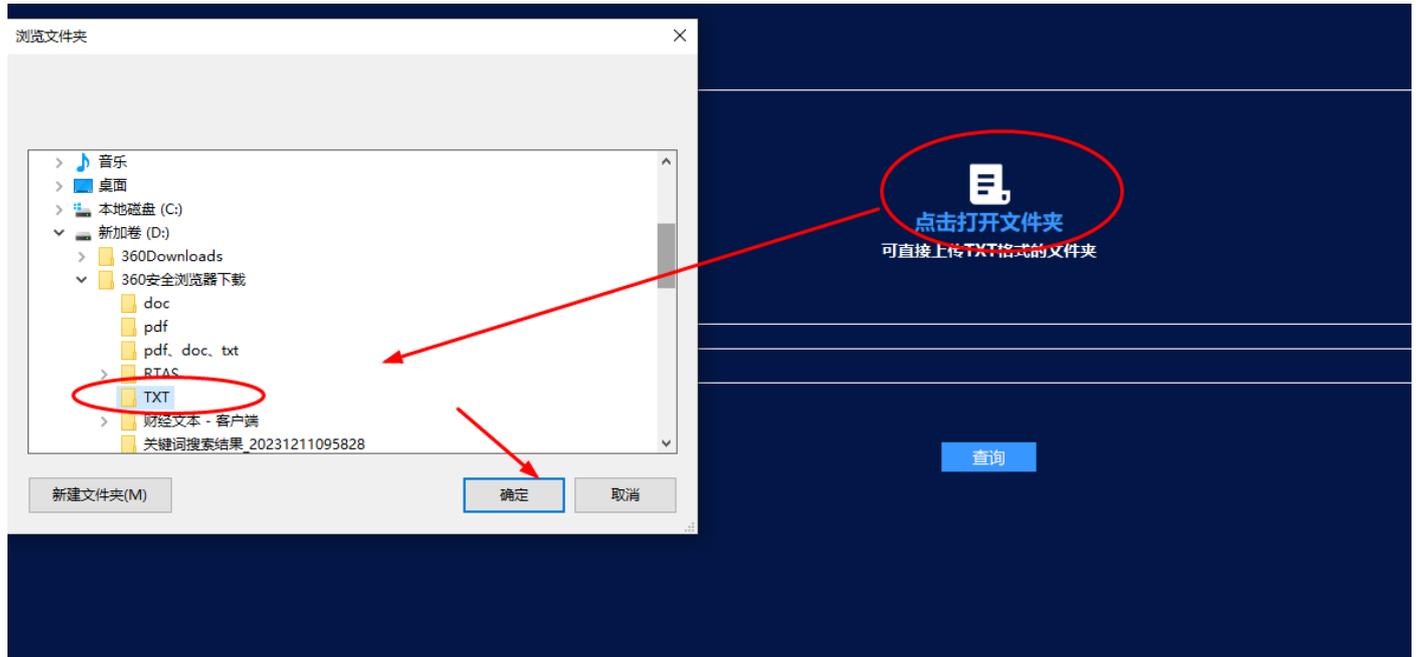
导出表格

文件	类别
PBL联合CBS教学模式对...实习生批判性思维能力的影...pdf.txt	类别0
中学语文教学培养学生思辨能...学语文批判性思维教学...pdf.txt	类别0
会计教学中批判性思维教学法运用策略分析.pdf.txt	类别0
初中物理科学探究中培养学生...科版8年级“浮力”教...pdf.txt	类别0
基于思辨能力培养的中学语文...学语文批判性思维教学...pdf.txt	类别0

首页 上一页 1 2 3 4 下一页 尾页

8.2.4 主题模型

功能简介: 根据上传的文档,完成分词、去停用词等预处理操作,根据文本预处理的结果,利用 TextBank 算法挖掘文本中主题词,返回每个文档中前是个主题词。



A1		文件
	A	B
1	文件	主题词
2	PBL联合CBS教学模式对...实习生批判性思维能力的影...pdf.txt	护理 教学模式 批判性 教学 思维能力 思维 学生 课堂 实习生 能力
3	中学语文教学培养学生思辨能...学语文批判性思维教学案例》.pdf.txt	批判性 教学 语文 思维 案例 阅读教学 学生 教育工作者 课堂 语言
4	会计教学中批判性思维教学法运用策略分析.pdf.txt	批判性 思维 学生 教学 问题 教师 基础 内容 教学法 过程
5	初中物理科学探究中培养学生...科版8年级“浮力”教学为例.pdf.txt	
6	基于思辨能力培养的中学语文...学语文批判性思维教学案例》.pdf.txt	教学 学生 语文 案例 能力 素养 批判性 学校 教师 学科
7	审辩式思维在创造力发展中的重要性.pdf.txt	
8	审辩式思维培养水平的评估指标研究.pdf.txt	学生 思维 教师 问题 水平 论证 教学 理由 结论 观点
9	审辩式思维教育具有的现实意义.pdf.txt	
10	审辩式思维能力及其测量.pdf.txt	思维 论证 评价 思维能力 参议员 批判性 药剂师 政府 医生 民主
11	审辩式思维观照下的初中文言文教学.pdf.txt	文言文 教学 思维 学生 观照 文本 教师 文化 流程 目标
12	循本源抓关键 提升审辩式思维能力.pdf.txt	
13	思维的批判性与大学的批判性思维教育.pdf.txt	思维 批判性 品质 逻辑 思想 理性 大学 前提 领域 准则
14	批判性思维教学：意蕴、特征及建构路径.pdf.txt	思维 批判性 学生 教学 教师 评价 建构 特征 师生关系 品质
15	批判性思维教育的一个议题：...究？罗素、杜威和图尔敏.pdf.txt	论证 情境 真理 批判性 观念 证据 思维 问题 理论 船主
16	本科生批判性思维能力及高影...双一流”建设高校的实证研究.pdf.txt	批判性 思维能力 学生 思维 影响力 课程 基线 科研 高校 能力
17	浅析审辩式思维的培养策略.pdf.txt	学生 思维 教学 课堂 灯泡 小组 科学 小学 课程标准 教师
18	美国斯坦福大学批判性思维教育及启示.pdf.txt	批判性 思维 课程 高校 学生 思维能力 主题 内容 名声 通识
19	认知偏差与批判性思维教育.pdf.txt	
20		
21		
22		
23		