



智能文本分析系统
Intelligent text analysis system

上海经禾信息技术有限公司
Shanghai Jinghe Information Technology Co., Ltd



智能文本分析系统
Intelligent text analysis system

“自定义文本”

“任意词统计”

“零代码分析”

CONTENTS

目录

Page **01** **研发背景**
Development background

Page **04** **功能介绍**
Text function

I 网页端功能
Web-side features

文本数据

I 客户端功能
Client capabilities

功能一：自由搜索

功能二：文本分析

功能三：文本挖掘

Page **19** **学术合作**
Academic cooperation

Page **20** **技术支持**
Technical support



智能文本分析系统
Intelligent text analysis system

研发背景

Development background

文本数据在语言和文字维度提供了大量的潜在信息价值，近年来受到较多关注。然而，对文本信息进行深度挖掘是一项极其复杂和专业的过程，让诸多使用者无从下手。

Altexs智能文本分析系统应学术需要，通过多年的大数据技术积累，提供了一个可以自定义、零代码和一键式文本分析的终端平台，以期让广大学者实现“文本自由”。

研发团队

Research team



团队

汇聚财经和计算机领域专家，知名高校教授团队。
财经专家学者：需求分析、产品设计、研究动态。
计算机专家：提供大数据解决方案、文本挖掘技术支持。



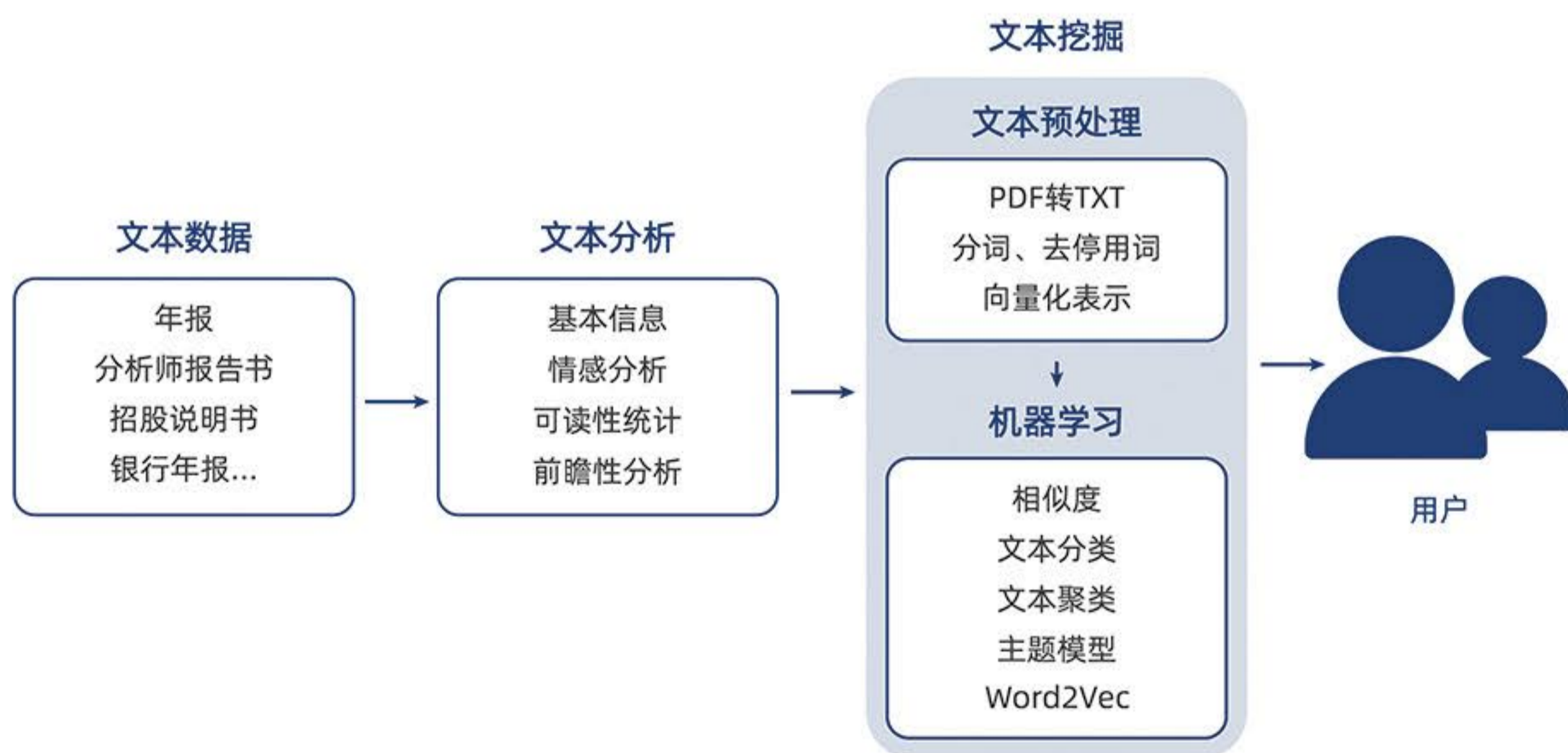
愿景

致力于成为领先的智能文本分析应用平台

文本功能

Text function

爱文本（AITEXTS）智能文本分析系统致力于为财经领域研究者提供精确、高效和可视化的文本分析工具，在文本数据、文本分析、文本挖掘三大领域打造零代码文本分析平台，让各类学者在文本领域的科研变得更简单！



网页端功能

Web-side features

海量财经领域文本，覆盖各个来源巨量原始数据、几十种文本类型、千万条文本信息



获得各类重要文本数据



对文本精确归类和清洗



采用深度学习算法分析



多维呈现文本分析结果



专业团队交叉抽样校对

文本数据

关键指标

涵盖了海量文本的基本信息

基本
信息

词频
统计

收录亿条词汇词频、
支持自定义文本、
自定义关键词分析

结合已有研究和权威指标
对可读性进行衡量

可读性

文本数据

公司公告

文本库列表

| | | | |
|----------|---|---|--------|
| 年报 |  |  | 关联交易 |
| 管理层讨论与分析 |  |  | 利润分配 |
| 招股说明书 |  |  | 担保事项 |
| 交易异动 |  |  | 停牌公告 |
| 澄清公告 |  |  | 增发预案 |
| 股份增减持 |  |  | 质押冻结 |
| 违规违纪 |  |  | 高管变动 |
| 风险提示 |  |  | 抵押借贷 |
| 政策影响 |  |  | 收购兼并 |
| 股份激励 |  |  | 股份分置改革 |
| 独立董事意见 |  |  | 法律纠纷 |
| 社会责任报告 |  |  | 投资理财 |
| 资产重组 |  |  | 资金投向 |

文本数据

新闻报道



股吧评论



公司新闻

研究报告



公司研究



晨会报告



宏观策略



行业研究



债券研究



期货研究

公司研究



监管问询



业绩说明会



IPO路演



关键审计事项



投资者问答

综合文本



券商年报



银行年报

客户端功能 功能一：自由搜索

Client capabilities

海量财经领域文本，覆盖各个来源巨量原始数据、几十种文本类型、千万条文本信息

获得各类重要文本数据



可上传任意文本



支持任意词汇搜索



自定义词汇列表



一键分析结果导出



请输入关键词,多个关键词请用英文逗号,分开始: 风险,收益...

分析一下

📁【打开本地关键词文件】 📄【模板下载】

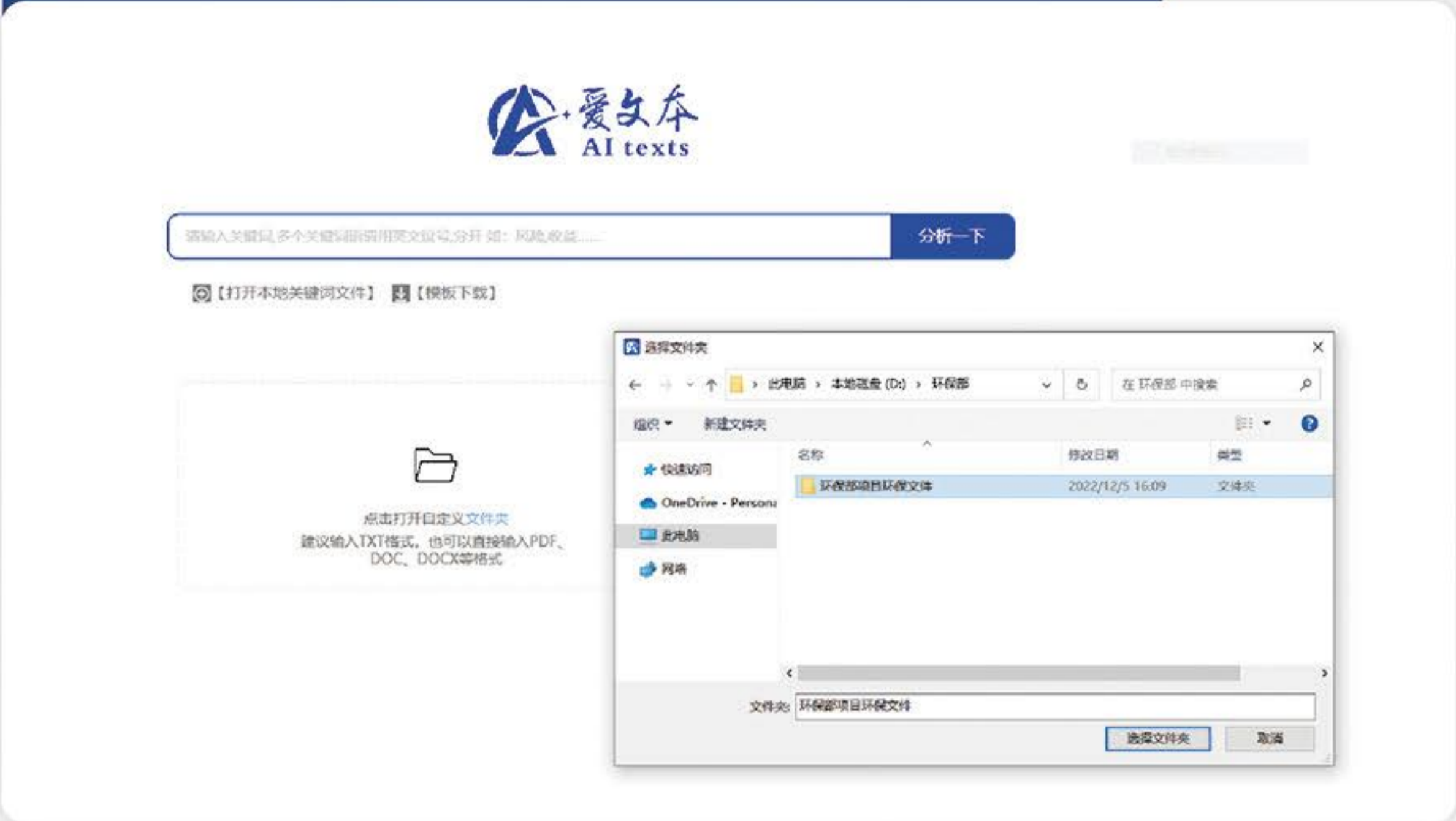


点击打开自定义文件夹
建议输入TXT格式,也可以直接输入PDF、DOC、DOCX等格式



选择系统已有文本
选择系统已有的文本文件进行词频分析

可自定义文本，支持TXT、PDF、DOC、DOCX等文件格式



可选择系统已有文本



系统提供的各类文本年限达20+年

返回

| 文件名称 | 会计年度 | 文件大小 | 当前状态 | 操作 | 下载路径 |
|----------|------|--------|------|------|--------|
| A股上市公司年报 | 2001 | 0.088G | 未下载 | ↓ 下载 | 📄 打开路径 |
| A股上市公司年报 | 2002 | 0.123G | 未下载 | ↓ 下载 | 📄 打开路径 |
| A股上市公司年报 | 2003 | 0.15G | 未下载 | ↓ 下载 | 📄 打开路径 |
| A股上市公司年报 | 2004 | 0.18G | 未下载 | ↓ 下载 | 📄 打开路径 |
| A股上市公司年报 | 2005 | 0.185G | 未下载 | ↓ 下载 | 📄 打开路径 |
| A股上市公司年报 | 2006 | 0.247G | 未下载 | ↓ 下载 | 📄 打开路径 |
| A股上市公司年报 | 2007 | 0.353G | 未下载 | ↓ 下载 | 📄 打开路径 |
| A股上市公司年报 | 2008 | 0.381G | 未下载 | ↓ 下载 | 📄 打开路径 |
| A股上市公司年报 | 2009 | 0.438G | 未下载 | ↓ 下载 | 📄 打开路径 |

提供精准词汇、扩展词汇、词频统计等诸多核心指标


风险
分析一下

【打开本地关键词文件】 【模板下载】

重新选择数据 已选数据: 券商年报_2021 导出表格

| 序号 | 文件名 | 股票代码 | 会计年度 | 词汇名称 | 词汇类别 | 词汇组别 | 词频统计 | 句子示例 | 出现该关键词的句子数 |
|----|----------------|--------|------|------|------|------|------|--------------|------------|
| 1 | 10.华泰证券股份有限公司 | 600918 | 2021 | 风险 | 精准词汇 | 1 | 696 | 六、前瞻性陈述的风险 | 325 |
| 2 | 10.华泰证券股份有限公司 | 600918 | 2021 | 风险管理 | 扩展词汇 | 1 | 162 | 公司关于公司未来发展 | 100 |
| 3 | 10.华泰证券股份有限公司 | 600918 | 2021 | 信用风险 | 扩展词汇 | 1 | 58 | 货币流动性风险较少, 种 | 42 |
| 4 | 10.华泰证券股份有限公司 | 600918 | 2021 | 高风险 | 扩展词汇 | 1 | 5 | 提升境外衍生品产品开发 | 5 |
| 5 | 10.华泰证券股份有限公司 | 600918 | 2021 | 金融风险 | 扩展词汇 | 1 | 1 | 有效防范化解金融风险 | 1 |
| 6 | 10.华泰证券股份有限公司 | 600918 | 2021 | 风险投资 | 扩展词汇 | 1 | 1 | 根据公司及子公司中 | 1 |
| 7 | 10.华泰证券股份有限公司 | 600918 | 2021 | 李总风险 | 扩展词汇 | 1 | 1 | 公司业务部门、分支机 | 1 |
| 8 | 10.华泰证券股份有限公司 | 600918 | 2021 | 总风险 | 扩展词汇 | 1 | 1 | 公司业务部门、分支机 | 1 |
| 9 | 104.浙商证券股份有限公司 | 601878 | 2021 | 风险 | 精准词汇 | 1 | 531 | 10%的一般风险准备金 | 225 |

功能二：文本分析

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



基本信息

读取文件大小

分析文本页数

精确计算文本句子数

提供字数和词汇数



情感分析

可上传任意文本

提供常用情感词典

自定义情感词典

多维度正负面语调



可读性分析

基于中文文本语境

借鉴权威研究成果

提供通用可读性指标

一键分析结果导出



前瞻性分析

可自定义选择符号
拆分句子

精确计算拆分句子数

提供常见前瞻性词语

分析句子前瞻性

文本分析

基本信息-统计各类文本的重要信息

提交 处理中 完成

点击打开文件或者文本
建议上传PDF格式的文件或文本, 也可以直接上传TXT、DOC、DOCX等格式的文件或文本

使用说明:

1. 上传PDF、DOCX、DOC、TXT文件, 系统自动生成文件的基本信息。
2. 统计TXT文件时字数设置为0

字段名称解释:

1. 总字数1: 文本总字数, 包括英文和数字, 单位: 个
2. 总字数2: 文本总字数, 不包括英文和数字, 单位: 个
3. 总词数1: 文本总词数, 包括停用词, 停用词是指英文单词、标点符号和无意义中文单词 (例如: 哎呀, 不仅, 而且) 等, 单位: 个
4. 总词数2: 文本总词数, 不包括停用词, 停用词是指英文单词、标点符号和无意义中文单词 (例如: 哎呀, 不仅, 而且) 等, 单位: 个
5. 句子数: 文本总句子数, 单位: 个
6. 页数: PDF、DOCX、DOC文件的总页数

开始

情感分析-三套词典保障情感倾向准确性

提交 处理中 完成

点击打开文件或者文本
建议上传TXT格式的文件或文本, 也可以直接上传PDF、DOC、DOCX等格式的文件或文本

选择情感词典 (必选项)

AI texts情感词典 IM情感词典 知大情感词典

自定义词典(添加的文件默认已选择) + 添加词典

说明: 根据用户自定义的情感倾向词典, 每次可上传一个情感倾向词典文件, 系统会自动将词典中的词汇与情感倾向词典进行匹配, 分为积极、中性、较为积极、较为消极、非常积极、非常消极, 则需要上传五次统计的情感倾向词典列表

使用说明:

1. 自动计算中文文本的情感倾向特征。
2. 可定制包括IM词典, 知大大词典和自定义词典在内的词典词典, 运用多种方法计算文本语气。

字段名称解释:

1. 积极词指数: 基于IM/知大大/自定义词典计算文本中的积极词指数, 单位: 个
2. 消极词指数: 基于IM/知大大/自定义词典计算文本中的消极词指数, 单位: 个
3. TONE1: TONE1 = (积极词指数 - 消极词指数) / 文本词数, TONE1值越大, 表示文本信息越积极
4. TONE2: TONE2 = (积极词指数 + 消极词指数) / (积极词指数 + 消极词指数), TONE2值越大, 表示文本信息越积极

开始

可读性分析-轻松获取主流可读性指标

提交 处理中 完成

点击打开文件或者文本
建议上传TXT格式的文件或文本, 也可以直接上传PDF、DOC、DOCX等格式的文件或文本

常用词典列表 下载
常用字表 下载
次常用字表 下载

使用说明:

1. 结合中文文本可读性的已有研究, 自动统计文本中的一些常用可读性指标。

字段名称解释:

1. 总字数1: 文本总字数, 包括英文和数字, 单位: 个
2. 总字数2: 文本总字数, 不包括英文和数字, 单位: 个
3. 总词数1: 文本总词数, 包括停用词, 停用词是指英文单词、标点符号和无意义中文单词 (例如: 哎呀, 不仅, 而且) 等, 单位: 个
4. 总词数2: 文本总词数, 不包括停用词, 停用词是指英文单词、标点符号和无意义中文单词 (例如: 哎呀, 不仅, 而且) 等, 单位: 个
5. 句子数: 文本总句子数, 单位: 个

开始

功能三：文本挖掘 – 预处理

本步骤是文本挖掘前的预处理，每一步处理都可以得到相应的过程文件，后一步的处理需要基于前一步的文件处理结果，如不需要手动调整文本挖掘的预处理过程文件，可直接进行一键分析。

文本预处理



文本挖掘

文本预处理

清晰展示PDF一键向量化结果

| | |
|------|--------------------|
| 身份 | 1.46473426246024 |
| 边界 | 1.48425619144692 |
| 定距 | 3.86712678342564 |
| 雨段 | 4.55469261832265 |
| 术系 | 4.57284286651055 |
| 族群 | 3.92910938958016 |
| 群聚 | 3.86392823889605 |
| 聚层 | 3.46439401792722 |
| 聚层区 | 4.55754771354323 |
| 空间 | 0.42023356340949 |
| 人类 | 0.643909066078698 |
| 人类学 | 3.17240552154828 |
| 研究 | 0.0322600353042901 |
| 重庆 | 1.15582839612517 |
| 重庆大学 | 1.85057976956705 |

| 文件名 | 大小 | 路径 | 时间戳 |
|-------------------------|-------|--|---------------------|
| 论李斯特《奥伯曼》 | | | 2022-08-26 11:54:36 |
| 身份与边界：巴赫 | | | 2022-08-26 11:54:51 |
| 神经网络缺失型突 | | | 2022-08-26 11:54:52 |
| 连续前缀航平面组 | | | 2022-08-26 11:54:52 |
| 鑫广安公司生猪养殖成本扩张战略研究.txt | >99KB | C:\Users\jinghe\Desktop\文本向量化_2022-08-26 11时50分56... | 2022-08-26 11:54:57 |
| 针对“程科”采取的毒化行为及机理研究.txt | 255KB | C:\Users\jinghe\Desktop\文本向量化_2022-08-26 11时50分56... | 2022-08-26 11:54:58 |
| 长周期波作用下炉面块体雷诺数效应的研究.txt | 79KB | C:\Users\jinghe\Desktop\文本向量化_2022-08-26 11时50分56... | 2022-08-26 11:54:58 |

分词功能处理过程

分词功能结果预览

可读性分析过程

可读性分析结果图

| 名称 | 单位 | 得分 | 权重 | 得分 | 权重 | 得分 | 权重 | 得分 | 权重 | 得分 | 权重 |
|-----------|----|-------|------|-------|------|-------|------|-------|------|-------|------|
| 1. 句长 | | 20.70 | 0.16 | 25.00 | 0.17 | 15.00 | 0.11 | 10.00 | 0.07 | 5.00 | 0.04 |
| 2. 词频 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 3. 句法复杂度 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 4. 词汇多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 5. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 6. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 7. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 8. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 9. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 10. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 11. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 12. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 13. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 14. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 15. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 16. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 17. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 18. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 19. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |
| 20. 句法多样性 | | 10.00 | 0.08 | 15.00 | 0.11 | 20.00 | 0.15 | 25.00 | 0.19 | 30.00 | 0.23 |

文本挖掘 – 一键分析

本步骤可实现从源文件导入到结果输出的一键文本挖掘，文本预处理将内嵌在处理过程中，不单独提供，如不需要手动调整文本挖掘的预处理过程文件，只需要文本挖掘结果，可直接运行本步骤。

机器学习

| | | |
|--|----------|--|
| | 相似度计算 | 支持 一键 自定义向量化 相似度分析 包含 欧式 余弦 Jacard 距离 |
| | 文本分类 | 包括: ①设计阶段、②应用阶段 |
| | 文本聚类 | 计算出文件夹中文本的聚类模型 |
| | 主题模型 | 计算出文件夹中文本的LDA模型 |
| | Word2Vec | 根据Word2Vec模型计算词语相似度 |

文本挖掘

机器学习

相似度计算功能支持下载系统自由向量化文本（TFIDF文件）

提供准确的向量化相似度分析

The screenshot shows the 'Vector Similarity Analysis' (向量化相似度分析) interface. It features a progress bar at the top with stages: '提交' (Submitted), '处理中' (Processing), and '完成' (Completed). The main area includes a file upload section with a '点击上传文件夹' (Click to upload folder) button and a note: '支持上传向量化TXT文本格式的文件，或包含向量化TXT文本文件的文件夹' (Supports uploading vectorized TXT text files or folders containing vectorized TXT text files). Below this is a link to '下载系统自有向量化文本 (TFIDF文件)' (Download system's own vectorized text (TFIDF files)). A '选择相似度类型' (Select similarity type) section offers three options: '余弦距离' (Cosine distance), '欧氏距离' (Euclidean distance), and 'Jaccard 距离' (Jaccard distance). A '使用指南' (Usage guide) sidebar on the right provides instructions: 1. Upload vectorized text collection; 2. Calculate similarity between two files. It also includes a '字段说明' (Field description) section stating: '相似度：采用欧式距离、余弦距离、Jaccard 距离计算的文本相似度' (Similarity: Text similarity calculated using Euclidean distance, cosine distance, and Jaccard distance). A '开始' (Start) button is located at the bottom right.

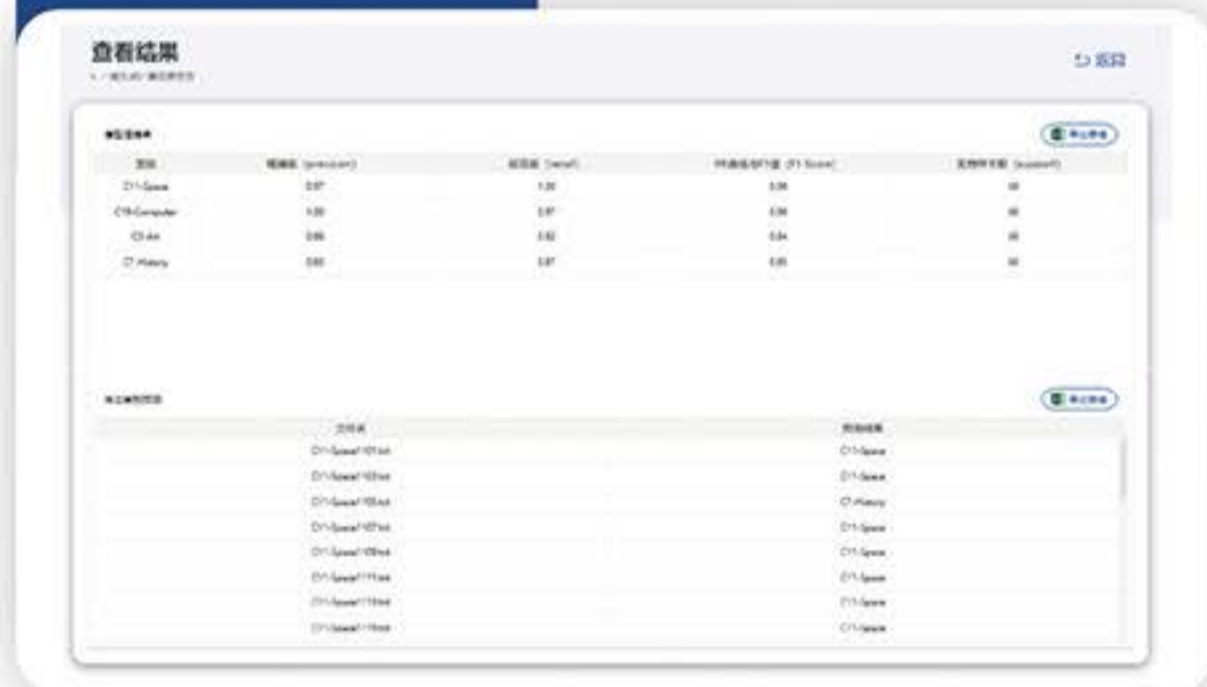
轻松下载系统的向量化文本

The screenshot displays the '轻松下载系统的向量化文本' (Easily download system's vectorized text) interface. At the top, there is a '公司公告' (Company Announcement) button. Below it, four document types are listed with corresponding icons: '年报' (Annual Report) with a calendar icon, '社会责任报告' (Social Responsibility Report) with a house icon, '管理层讨论与分析' (Management Discussion and Analysis) with a speech bubble icon, and '招股说明书' (Prospectus) with a document icon.

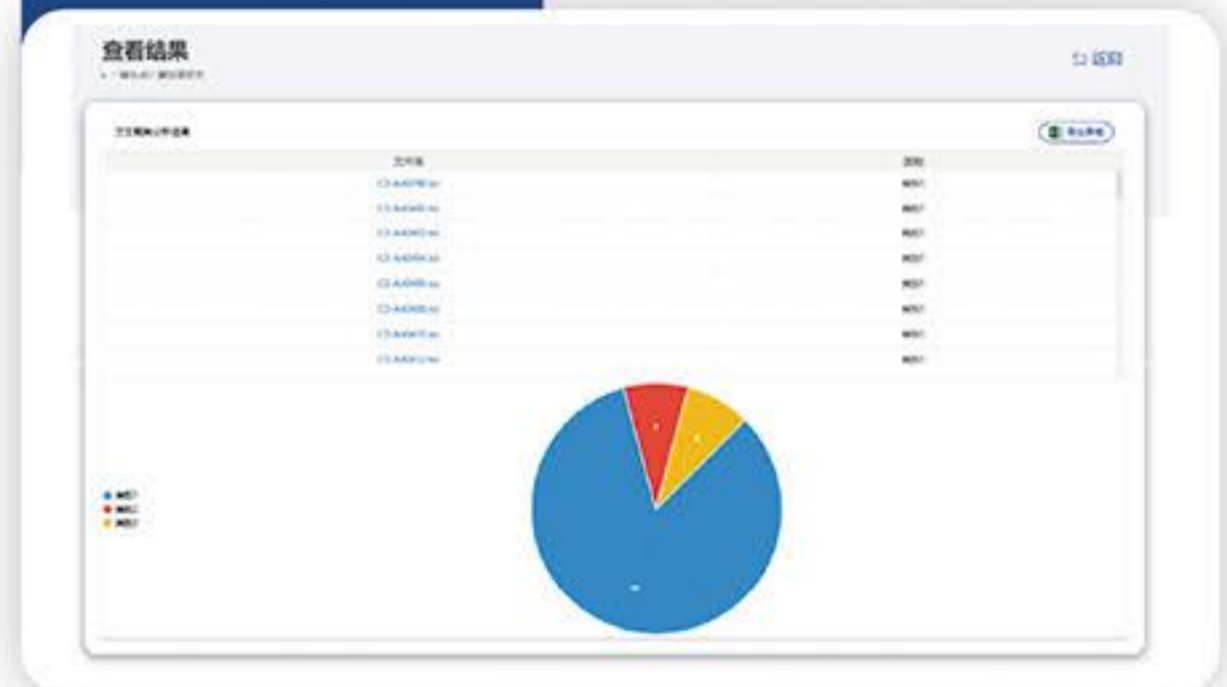
一键文本分类全景图



一键文本分类结果图



一键文本聚类结果图



主题模型分析全景图



Word2Vec分析全景图





智能文本分析系统
Intelligent text analysis system

学术合作

“爱文本（AItxts）智能文本分析系统”在文本数据、文本分析、文本挖掘智能化呈现上具有多年的业务实践和技术积淀。凭借领先的文本数字化和智能化产品，打造服务于科研、投资的多维场景解决方案，为客户创造价值。

我们将继续秉承开放合作、价值创造的原则，积极协同科学研究和投资实务等各领域的合作伙伴，为用户量身打造高价值、高水平的文本分析应用软件产品和解决方案。



技术支持

上海经禾信息技术有限公司是中国大陆领先的金融大数据服务企业，总部位于上海市宝山区华滋奔腾大厦（上海市移动互联网中心），公司创始人和股东均为具有多年从业经验的券商、金融机构和知名互联网企业的资深人士。

我们旨在利用自身的技术和经验，为客户提供大数据产品和服务，开发基于资本市场的数据库、软件产品，并将这些产品推向市场。

爱文本·会思想的文本

分析海量文本，挖掘数据价值

联系我们

市场部

联系人：刘经斌

电话：(+86) 021-66181082

手机：(+86) 17821816737

邮箱：liujingbin@aitexts.com

地址：上海市移动互联网中心（华滋奔腾大厦）A座18层



扫一扫，联系我们

产品网址

<https://www.aitexts.com>