

# Datago Tech. Ltd.

DataGo Ahead in FinTech!

2021

# Text Analytics in Finance



Massive Text Data  
Millions per day, billions in history



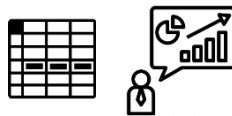
Text Analytics with AI

Is there any **bad news** on stocks in my portfolio?

What are the top 10 stocks attract **most comments** in social media?

How is the **disclosure quality/ESG rating** of this stock?

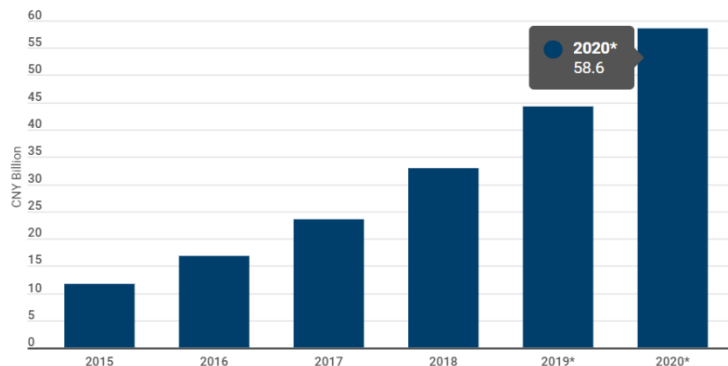
Which public companies' has **positive management's tone** in annual report?



Machine/Human Readable Analytics

# Financial Text Analytics in China

- Chinese text analytics is a billion-dollar market that grows rapidly
  - ▣ Not only valuable for investment in China



“with roughly 1.3 billion consumers, ..., make China an **attractive equity market** , also a **wealth of alternative data**. ”  
*Alternative Data in China* by Man Group (AUM: \$140 Billion)

- Lack of prominent vendors as in the West

REFINITIV®  
DATA IS JUST  
THE BEGINNING®



Bloomberg

MSCI



AlphaSense




RavenPack



Sentieo

# Challenges

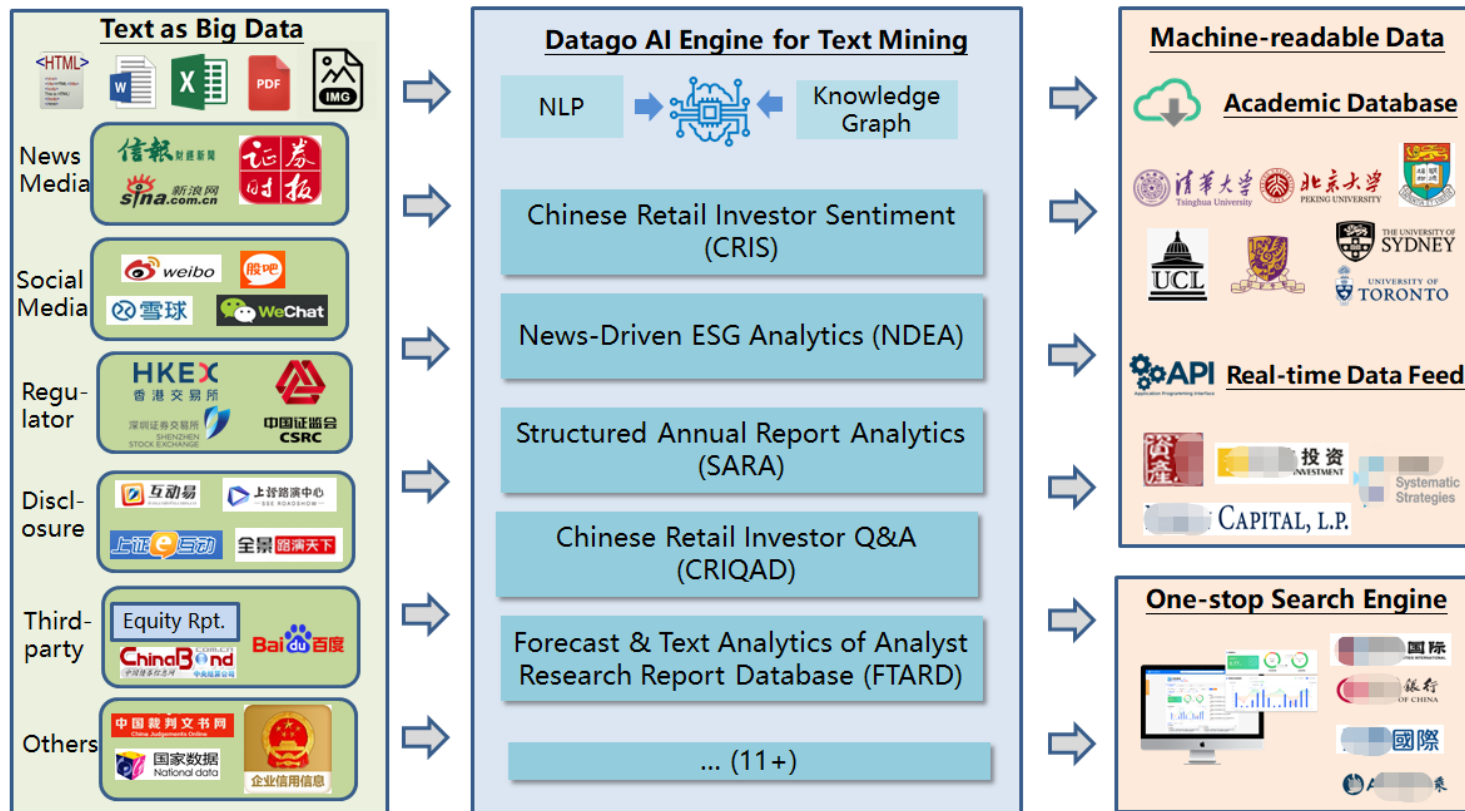
- Unique market with local sources of text data in China
  - ▣ Global vs. Local

| Category | News                                                                              | Retail investor                                                                   | Investor Q&A                                                                        | Online Search | Other                                                                               |
|----------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|---------------|-------------------------------------------------------------------------------------|
| Global   |  | #wallstreetbets                                                                   |  | Google        |  |
| Local    |  |  | Transcripts?                                                                        | Baidu 百度      |  |

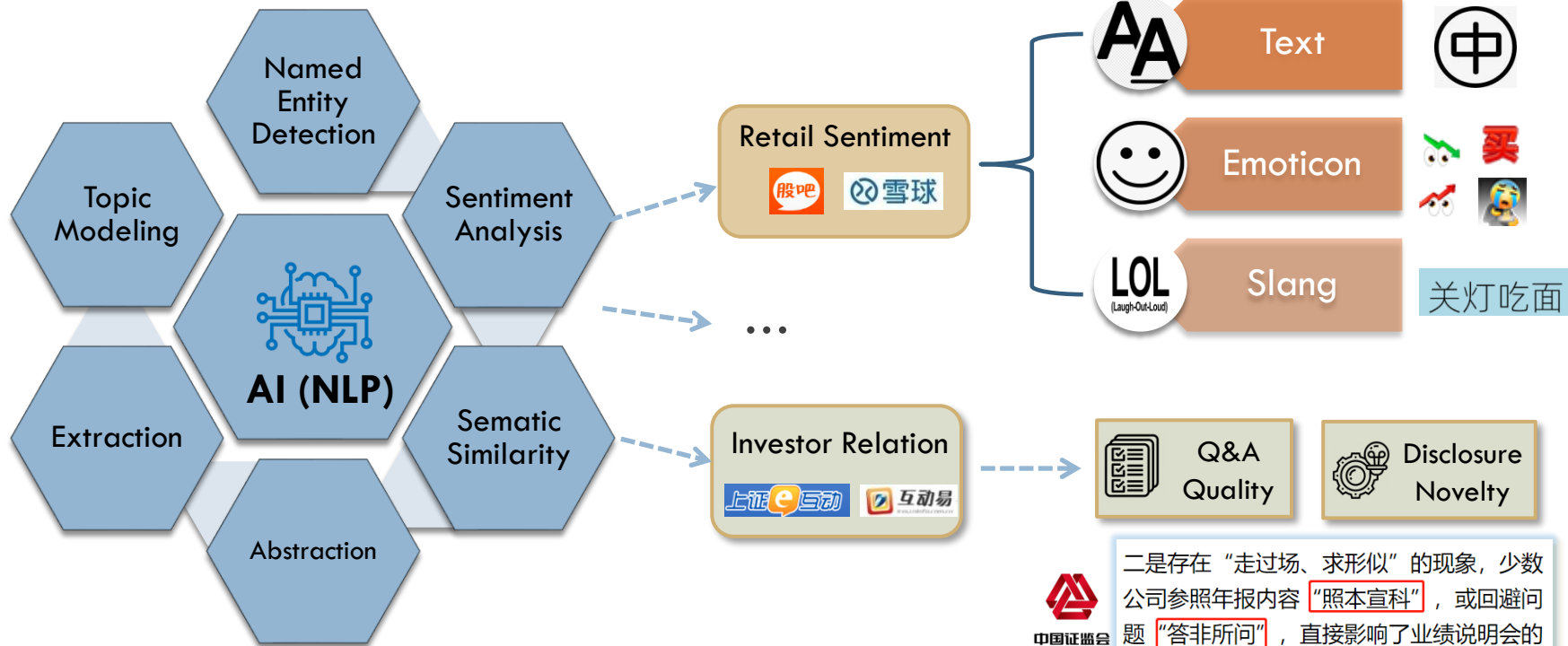
- Chinese language raises extra challenges
  - ▣ More difficult with less technical resources publicly available



# Overview of Business



# Optimized AI Model for Local Sources



中国证监会  
CSRC

二是存在“走过场、求形似”的现象，少数公司参照年报内容“照本宣科”，或回避问题“答非所问”，直接影响了业绩说明会的质量和沟通效果；

# Rich Corpus for **Chinese Financial** Text

- 3 millions+ proprietary training data labeled by business professionals



Cross-checked



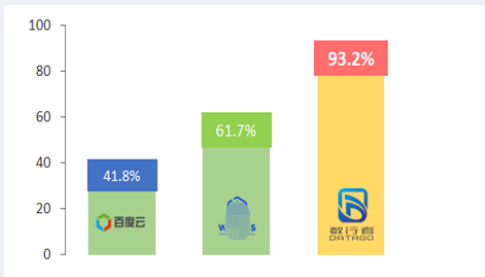
Comprehensive

跳水 =



Specialized

Accuracy of Sentiment Analysis for **Financial** News



**Accurate**

acc. > 90%



**Fast**

0.01s for 1 doc.



**Robust**

Availability: > 99.9%

# Datago Text Analytics Products(SaaS)

## Accurate, Prompt, Reliable

- Real-time analytics with high reliability

## Comprehensive

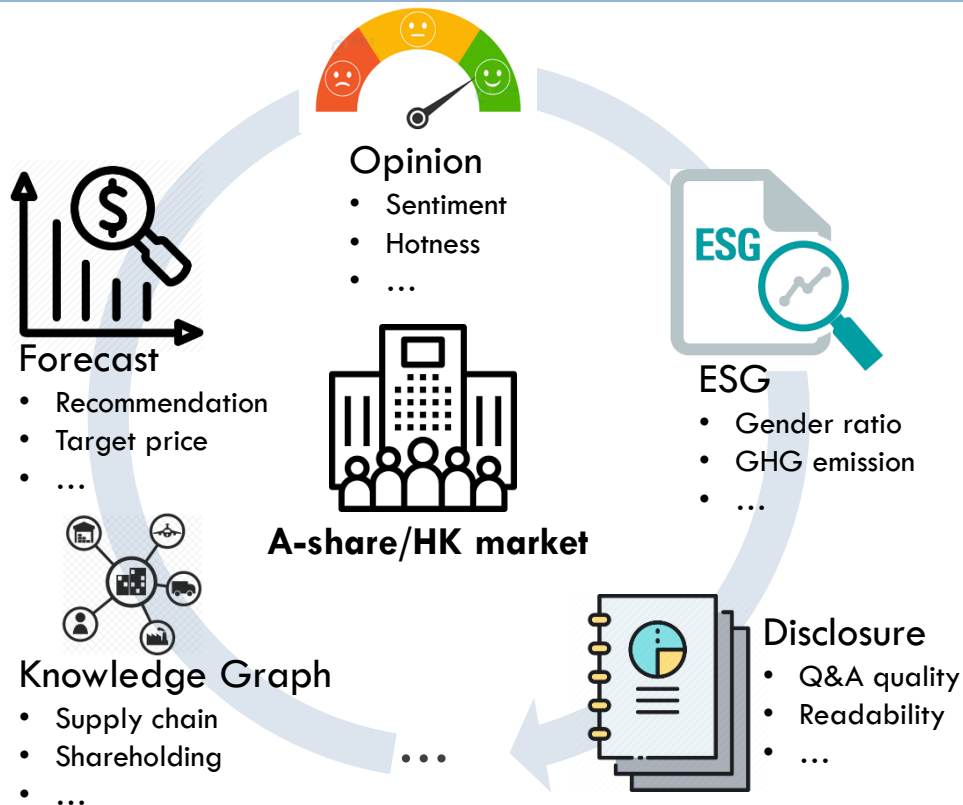
- 1000+ sources with 10 years+ history
- Retail sentiment, analyst sentiment, news sentiment, management tone

## Quantitative

- Sentiment: -1.0 ~ 1.0
- Recommendation: 1~5

## Unique

- Investor Q&A readability, novelty ...
- ESG for 4000+ A/HK stocks (~700 for MSCI)





# Market & Applications

- 50+ paying clients around the world
  - ▣ Hedge funds, institutes in US, UK ...
- One of most popular academic database
  - ▣ Support research in top journals
- Proved alternative data for investment
  - ▣ 5 out of top 10 hedge funds



Hedge fund



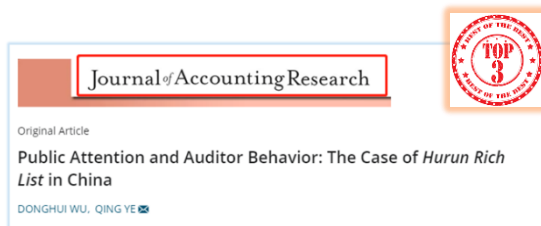
Broker



Bank



Institutes



# Our Team – Finance + Technology

## Management team



**Dr. LONG Yi (CEO)**

- 12 years+ experience in data mining,
- Guest lecturer of data mining, CUHK(SZ)
- Postdoctoral researcher in FinTech, CUHK
- HKU Ph.D,

## Co-founder and advisor



**T.J. Wong**

- Professor of USC
- **Former Dean** of the CUHK Business School
- UCLA Ph.D



**Tianyu ZHANG**

- Professor of CUHK
- Director of CUHK CIG
- **Cheung Kong Scholars**
- HKUST Ph.D



**Xiaoyang KANG**

- **30+ years experience** in investment
- Founder of Investoday Tech. Ltd
- CEO of Lighthouse Asset Management Ltd.

Frequent Contributors by Publications on  
Chinese Capital Market in Tier-1 Journals in 1999-2018

| # | Author             | Number of Papers | Total Citations | Average Citation per Paper |
|---|--------------------|------------------|-----------------|----------------------------|
| 1 | T.J. Wong          | 9                | 5,118           | 568.67                     |
| 2 | Yongxiang Wang     | 6                | 323             | 53.83                      |
| 2 | Donghui Wu         | 6                | 1,028           | 171.33                     |
| 2 | Tianyu Zhang       | 6                | 2,493           | 415.50                     |
| 5 | Phyllis Lai Lan Mo | 5                | 505             | 101.00                     |

# Improving Market Transparency

## 中国资本市场信息质量 暨上市公司信息透明度指数白皮书

光华-罗特曼信息和资本市场研究中心

感谢新财富在指数研制过程中的战略合作与支持，感谢所有参与本次指数通讯调查的分析师、机构投资者、上市公司董秘，感谢参与实地调研的上市公司，感谢香港数行者公司提供媒体数据和舆情结果。



活动主办单位：中国基金报、深圳益正金融服务有限公司

数据支持单位：数行者科技

# Going Beyond Finance ...

- Combat the spread of misinformation in local social media of Hong Kong



## 浸大開發AI揪出假新聞 設事實查核中心 每天篩選社媒消息

By 信報財經新聞 on December 30, 2020



浸大事實查核中心  
HKBU FactCheck Service

## 夥本地初創設數據庫

為加快篩選可疑內容的進度，中心短期內從社交平台着手，針對內容的熱度指數、流行度、情感指數等多種標準，為新聞數據做排序以供篩選。長遠而言，中心將夥拍本港初創數行者科技，開發本港首個虛假新聞研究數據庫，以本地語言為基礎，由人工智能演算法驅動的系統，自動偵測廣被瘋傳的可疑資訊。



The Overcoming Vaccine Hesitancy in Hong Kong Project  
by Hong Kong Baptist University  
Report Series No. 11  
15-08-2021



An Observational Study to Understand Vaccine Hesitancy in Hong Kong  
Using Heterogeneous Data

[https://research.hkbu.edu.hk/f/page/20480/21910/\(EN\)OVH\\_Report\\_No.11.pdf](https://research.hkbu.edu.hk/f/page/20480/21910/(EN)OVH_Report_No.11.pdf)

# Thank you!

## DataGo Ahead in FinTech!



Email: [info@datago.com.hk](mailto:info@datago.com.hk)








Site: <http://datago.com.hk>

# Appendix

# Data Samples (Sentiment)

| 上市公司代碼    | 新聞情緒評分 | 新聞熱度評分 | 新聞綜合輿情評分 | 社交媒體情緒評分 | 社交媒體熱度評分 | 社交媒體綜合輿情評分 | 投資者互動回答覆蓋率 | 投資者互動回答品質評分 |
|-----------|--------|--------|----------|----------|----------|------------|------------|-------------|
| 000001.SZ | 52.44  | 98.12  | 52.2     | 62.49    | 89.01    | 63.37      | 0.76       | 40.10       |
| 000002.SZ | 60.85  | 99.57  | 61.73    | 49.79    | 95.98    | 49.96      | 0.70       | 48.48       |
| 000004.SZ | 28.48  | 39.38  | 27.19    | 28.15    | 50.84    | 27.18      | 0.90       | 66.92       |
| 000006.SZ | 15.56  | 54.14  | 14.47    | 43.23    | 50.82    | 42.67      | 0.93       | 75.20       |
| 000009.SZ | 68.93  | 84.52  | 70.19    | 53.51    | 91.39    | 53.64      | 0.94       | 59.87       |
| 000011.SZ | 48.65  | 52.99  | 47.18    | 25.35    | 76.54    | 24.6       | 0.91       | 53.13       |
| 000012.SZ | 80.54  | 70.63  | 82.55    | 60.06    | 79.7     | 60.5       | 0.99       | 62.81       |
| 000016.SZ | 77.87  | 84.16  | 79.73    | 64.59    | 89.98    | 64.96      | 0.95       | 62.49       |
| 000017.SZ | 35.85  | 8.38   | 35.82    | 62.43    | 25.63    | 62.53      | 0.98       | 66.46       |
| 000019.SZ | 58.48  | 26.93  | 58.8     | 44.88    | 50.96    | 44.19      | 0.98       | 69.29       |
| 000020.SZ | 56.05  | 19.7   | 57.24    | 73.57    | 12.15    | 74.56      | 0.50       | 35.34       |
| 000021.SZ | 79.27  | 78.23  | 81.08    | 39.57    | 89.74    | 38.9       | 0.95       | 62.98       |
| 000023.SZ | 29.32  | 30.24  | 27.47    | 59.15    | 5.12     | 59.42      | 1.00       | 60.12       |
| 000026.SZ | 62.5   | 68.49  | 63.56    | 46.96    | 46.09    | 46.95      | 0.98       | 86.53       |
| 000027.SZ | 68.85  | 74.3   | 70.44    | 49.82    | 77.4     | 50.46      | 0.94       | 82.90       |
| 000028.SZ | 75.76  | 90.63  | 77.45    | 54.99    | 76.5     | 55.78      | 0.94       | 62.28       |

# “AI” for Chinese Financial Text-1

|      |                                        |                                                                                     |               |
|------|----------------------------------------|-------------------------------------------------------------------------------------|---------------|
| 情感判别 | “该项新措施可降低 <b>风险</b> ，减少 <b>损失</b> ”    |  | 负面            |
|      | “英国经济“脱欧 <b>红利</b> ”恐难 <b>持久</b> ”     |  | 正面            |
| 实体识别 | “中美贸易战阴影笼罩什么是中 <b>美的</b> 新压舱石”         |  | 美的(SZ000033)  |
|      | “东南汽车与三菱汽车合作加 <b>深发展</b> 自主品牌”         |  | 深发展(SZ000001) |
|      | “经济进入良性循环轨道 <b>老百姓</b> 就业机会增多”         |  | 老百姓(SH603883) |
|      | “ <b>老百姓</b> 拟募集加码主业”                  |  | 老百姓(SH603883) |
| 事件分类 | 大湾区水域 <b>洪涝灾害</b> 跨区域 <b>救援</b> 实战演练举行 |  | 事故灾害          |



# “AI” in The Market-1

三级情感评分(正面, 负面, 中性)

事件分类

06:59:02

## 连亏两年却要分红2亿! \*ST罗普涨停 深交所关注函来了

情绪: 利好 属性: 分红

连亏两年的\*ST罗普, 却在一季度罕见大手笔分红, 周五股价涨停了! 6月13日晚间, \*ST罗普公告称一季度拟分红2.01亿元(含税); 6月14日上午, 深交所紧急下发关注函, 要求公司说明是否符合现金分红条件, 未来6个月内是否存

### 涉及个股

\*ST罗普 4.68 -3.70% [加入自选](#)

相关公司识别

### 个股小工具

个股行情 资金流向 概念板块 个股资讯  
全部公告 重大事项 龙虎榜单 十大股东

0 已阅 152

# “AI” in The Market -2

个股资讯 实时热点: 贵州茅台 ↑ TCL科技 ↑ 京东方A ↑ 光大银行 ↓

拼音/代码/名称

**老百姓 (603883.SH)** 买

客户端 桌面快捷方式 加入自选股

77.74 ↑ 2.22 +2.94%

今开: 76.18 成交量: 2.85万手 振幅: 4.98%

最高: 79.94 成交额: 2.22亿元 换手率: 0.70%

最低: 76.18 总市值: 317.75亿 市净率: 7.65

昨收: 75.52 流通市值: 316.70亿 市盈率TTM: 52.80

2021-02-08 11:30:00

资讯与公告 个股资讯 行业资讯 公司公告 年度报告 中期报告 一季度报告 三季度报告 理财师解读 公司号

## 个股相关资讯:

- 2021-02-05 20:38 江苏紧盯老百姓的“保命钱” 开展社会保险基金审计
- 2021-02-05 19:13 一心堂(002727.SZ)披露收购宜宾老百姓药房门店资产及其存货的进展
- 2021-02-04 21:47 荆州市委书记: 春节各单位院子打开让老百姓停车、上洗手间
- 2021-02-04 20:51 荆州市委书记: 春节期间各单位院子全免费开放, 不能搞得老百姓回家过年, 停车位找不到洗手间找不着!
- 2021-02-04 19:58 央行: 继续加强规范存款创新产品 守护好老百姓的钱袋子
- 2021-02-04 09:10 陕西省政协医药卫生界委员远程医疗: 让老百姓享受到便捷的医疗服务
- 2021-02-03 19:25 农业农村部召开全国春季田管暨春耕备耕工作视频会议
- 2021-02-03 17:38 张忆东: 现在老百姓配置股票就像20年前配置房地产 方兴未艾
- 2021-02-03 14:22 人行重庆银管部调查: 老百姓对房价下降预期减弱
- 2021-02-03 12:56 当选后第一件事, 就是解决老百姓最着急的事

## 一心堂(002727.SZ)披露收购宜宾老百姓药房门店资产及其存货的进展

榕隆汇 2月5日 | 一心堂(002727.SZ)公布, 公司2020年第14次总裁办公会讨论同意全资子公司四川一心堂医药连锁有限公司收购宜宾市老百姓(603883.SZ)大药房有限责任公司(“宜宾老百姓药房”, 该宜宾老百姓药房与上市公司老百姓大药房连锁股份有限公司无关联关系)所持有7家门店资产及其存货, 并发布《关于全资子公司四川一心堂医药连锁有限公司收购宜宾市老百姓大药房有限责任公司门店资产及其存货的公告》(公告编号: 2020-205号)。

2021-02-05 和讯网

## 鹤壁农村推行“电采暖、气做饭、房保温”, 老百姓的生活变了样

自从2017年成功入选由中央财政支持的北方地区冬季清洁取暖试点城市, 鹤壁市开始以“电采暖、气做饭、房保温、煤清零”的技术路线, 在全市农村范围内开展清洁取暖工作, 老百姓自此告别了烧煤生活, 据了解, 这种低温空气源热泵机...

2021-02-05 人民政协网

## “元朝规定十户共用一把菜刀, 老百姓如何做饭?” 评论区是真的吗? 哈哈哈哈哈

那些说自己没有情商的男生, 都是为自己强行女生找借口。运动场上的灰太狼, 都能闭着眼睛瞎话, 明明嗷嗷比狼羊羊脸大还群, 灰太狼就是认定懒羊羊是草原最肥的羊, 是不是比很多男生都强, 哈哈哈哈哈 穿背心漏肩, 也挡不...

2021-02-05 网易财经

## 荆州市委书记: 春节各单位院子打开让老百姓停车、上洗手间

昨天(2月2日)上午, 荆州市委书记倪光中深入医院、社区、农村、福惠点等, 调研冬季常态化疫情防控工作, 强调要突出精准防控, 不搞层层加码, 严格按照规定执行防控任务, 保障全市人民度过一个平安祥和的新春佳节, 强调突出精准...

2021-02-04 澎湃新闻

## 央行: 继续加强规范存款创新产品 守护好老百姓的钱袋子

央行: 继续加强规范存款创新产品 守护好老百姓的钱袋子

2021-02-04 网易财经

## 5批次化妆品检出禁用物质 老百姓天津某店登经营单位

5批次化妆品检出禁用物质 老百姓天津某店登经营单位

2021-02-04 中国经济网

# Advantage of Datago NLP Tech.

| News                | Datago (score)   | Baidu AI  | Others    |
|---------------------|------------------|-----------|-----------|
| 宇通客车：卓越的公司被低估的客车龙头  | Positive(0.47)✓  | Positive✓ | Negative✗ |
| *ST美丽股东王仁年持有的公司股份流拍 | Negative(-0.17)✓ | Positive✗ | Positive✗ |
| 华盛达或遭虚假陈述民事赔偿诉讼     | Negative(-0.8)✓  | Neutral✗  | Negative✓ |

# Comprehensive Analytcsis

**数行者科技**  
DATAGO TECHNOLOGY LIMITED

<https://finance.sina.com.cn/roll/2019-01-22/doc-ihqfskcn9259972.shtml> **AI智能诊断**

数据显示, 2018年全年共召回新能源汽车13.57万辆, 召回比例高达13.46%; 三年年龄的新能源二手车保值率仅为47.5%。据国家市场监督管理总局统计数据显示, 2018年包括比亚迪(腾势)、力帆、众泰(云100、云100s)、路虎、华晨宝马(530le)、北汽新能源、江淮汽车七大企业召回了新能源汽车(包括纯电动车和插电式混合动力车)。具体各家车企, 北汽新能源累计召回6.94万辆, 众泰汽车召回3.13万辆, 比亚迪(腾势)召回1.01万辆, 特斯拉召回8905辆, 力帆汽车召回6431辆, 江淮汽车召回4248辆, 路虎中国召回3406辆, 华晨宝马召回2001辆。

舆情得分: -0.79

**强烈利空**

相关事件: 产品异常, 0.592 [更多>>](#)

相关公司:  
特斯拉(TSLA), 0.8442  
江淮汽车(600418), 0.4422 [更多>>](#)

相关行业: 28\_汽车, 0.977 [更多>>](#)

相似新闻:

| # | 标题                            | 新闻来源     | 相似度得分      |
|---|-------------------------------|----------|------------|
| 1 | 新能源汽车背后: 召回率达13% 三年车龄贬值过半     | 证券日报(北京) | 0.99129176 |
| 2 | 新能源汽车高速发展背后 召回率高达13%          | 第一财经     | 0.9903729  |
| 3 | 新能源汽车高速发展背后 召回率高达13%、三年车龄贬值过半 | 腾讯财经     | 0.9852063  |

Summarization

Sentiment analysis  
(from -1 to +1)

Event classification

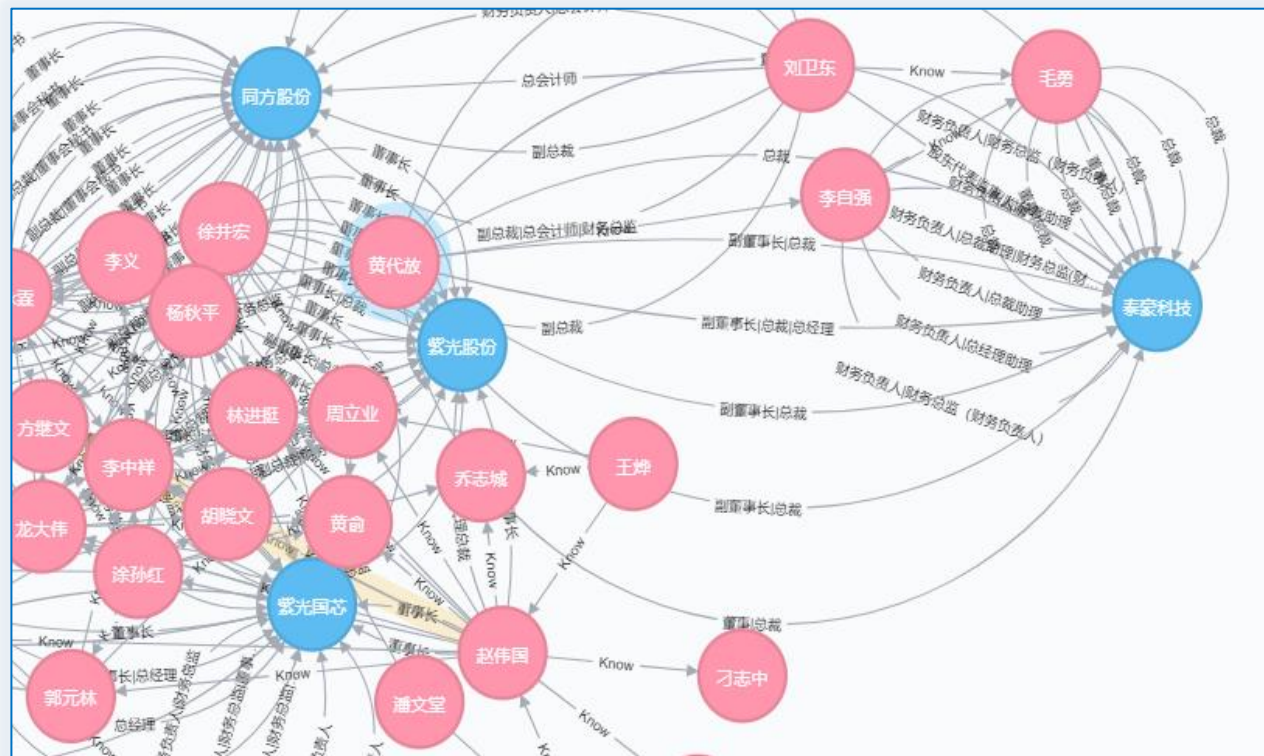
Firm detection

Industry detection

Similar News

# Well-developed Knowledge Graph

- ✓ Executive
- ✓ Analyst
- ✓ Fund manager
- ✓ Competitor
- ✓ Retail Investor
- ✓ Share holder
- ✓ Supplier
- ✓ Client



# Robust and Scalable Infrastructure

## Application

### Smart Report

脱水研报/公告  
研报/公告分类  
明星分析师/大V

### Rec-Sys

智能标签  
要闻精选  
个性推荐

### Graph+

知识梳理  
关系检索  
规则推理

### Search

垂直搜索  
多维聚合  
智能排序

### Tracking

实时监控  
风险预警  
热点追踪

### Customizing

文档数据化  
流程自动化  
智能投顾

## Tech.

### Information

Text extraction  
Picture Extraction  
Table Extraction  
Knowledge Extraction  
...

### Text Analysis

Sentiment  
Event classification  
Summarizing  
Trend detection  
...

### Info. Manage

Recommend  
Retrieval  
Tagging  
Tracking  
...

### Knowl. graph

Human Relation  
Firm relation  
Industry relation  
Inference  
...

### Toolkits

NLP tools  
ML tools  
Labeling tools  
Validation tools  
...

## Infra.

### Data Collection

Scrapy  
kafka  
flume  
...

### Data Storage

MySQL  
Hadoop  
HDFS  
elasticsearch  
...

### Compute Platform

Spark  
learn  
TensorFlow  
...

### Service Manage

docker  
Jenkins  
kubernetes  
...

# One-stop Search Engine



<https://www.datagotech.cn/>