

拓尔思数家资讯大数据云服务平台

产品背景

当前,全球大数据正进入加速发展时期,应用创新不断迈向新高度,“跨界、融合、基础、突破”是大数据落地的关键,大数据在“产业化、行业化、智能化”方面不断赋能各行各业。基于互联网海量公开数据的价值挖掘,已成为产业数字化转型和内容价值提炼的共性需求。

为解决各行业领域对内容资讯数据的场景服务需求,拓尔思推出数家资讯大数据云服务平台——以内容资产为核心的新一代数据型资讯业务云平台。“数家”以大数据+AI为支撑,面向各行业用户提供个性化的全网资讯数据分析和知识服务。如数家平台可为政务信息化建设提供数据支撑,提高政务精细化管理服务水平;可以为传媒用户实现内容素材的供给、选题热点的研判、生产知识辅助和传播效果评价等,以提升媒体“四力”建设;除此以外,数家面向B端用户提供行业情报和产业数据汇聚,以全业务链数据的整合与治理,为实现数据资产化和数字化转型赋能。

产品概述

数家通过大数据主流计算框架和基于深度学习的自然语言处理技术,以“多维知识库+智能语义”的方式对内容进行结构化标引,融合多领域对资讯的应用需求,为用户提供高信源、精加工、面向多元场景的资讯内容服务。



产品总体架构

产品功能

内容供稿云服务

将数字报、网站、新闻客户端、微信、微博、第三方平台、短视频平台的内容数据进行结构化分类整合，以“云服务”的方式满足用户对互联网资讯内容的数据需求，可实现内容精细化检索、个性化订阅、基础渠道数据供给、特色资源数据供给，可根据用户实际数据需求分析自动对接数据到用户的应用系统中。



线索发现云服务

面向全网数据，通过线索发现技术实时发现互联网高价值内容，并通过多维知识库及智能语义技术进行线索聚合、追踪分析、发展趋势预测，包含实时聚焦、网民关注、微博爆料、突发事件、自然灾害、近期政策、近期会议，历史今天等功能模块。



热点分析云服务

利用大数据分析技术从海量的互联网数据中发现当前资讯传播热点，实现对不同行业、不同地域的热点信息聚类。所有热点数据提供基础属性要素的提取，多层级的分析下钻和深度语义分析，包括行业热点、地域热点、英文热点、趋势预测、头版头条、微博热点、微信热点、排行榜单等模块。支撑内容创作的选题研判和热点趋势的多角度掌控。



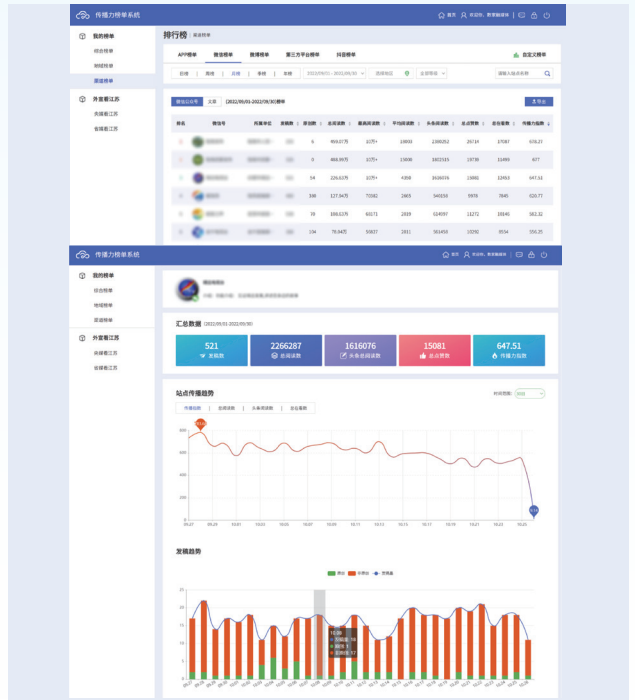
传播分析云服务

基于报纸、网站、微博、微信、客户端、第三方公众平台抓取到的数据，运用原创内容识别、文本相似度分析等大数据分析技术，结合稿件传播影响力数据模型和指标体系，对原创稿件的全网传播效果进行持续性的实时分析，包括传播转载分析、阅读互动分析、传播轨迹可视化分析和多维报表自动生成等。



榜单分析云服务

对海量互联网传播数据进行实时监测，基于科学系统的评价指标体系，构建权威榜单，实现全渠道、多维度、跨周期的站点排行、账号排行、内容排行、定制化专题排行及跨渠道机构综合排行。为用户衡量宣传主体的传播力、发掘优质内容提供依据，有效助力内容传播运营。



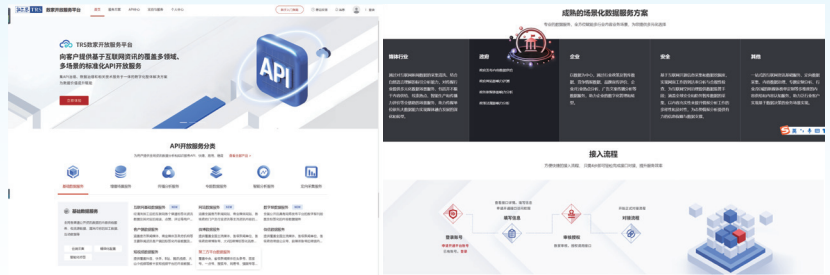
专题跟踪云服务

以事件或专题为中心，构建专属的事件专题分析服务，包括事件的总览、关键因子探索、话题漂移追踪、事件溯源、传播节点分析等一系列深度挖掘的内容语义分析场景。



● 标准化API开放云服务

提供基于互联网资讯的覆盖多领域、多场景的标准化API开放服务。数家将多年积累的业务经验和技術优势以API的方式对外开放，采用一站式接入服务，提供专业的技术支持与API文档、调试工具及示例，全程为开发加速。目前平台提供200余种API接口，涵盖基础数据服务、增值场景服务、智能分析服务、传播分析服务、专题分析服务、定向采集服务，全方位赋能多元内容业务场景，为用户提供丰富的选择。



产品优势

● 高效精准的数据资源

在并行采集基础上实现子模块的二级调度，以保证大规模采集平台的高时效。

● 海量资讯全覆盖采集

覆盖1100+家数字报、30万+媒体和政府网站采集点、3000+APP客户端、数百万的微信公众号和准全量的微博账号、涵盖头条、百家号、企鹅号、抖音、快手、微信视频号、梨视频等分发平台，63个语种15万+境外采集点。全库数据超过八百亿条。

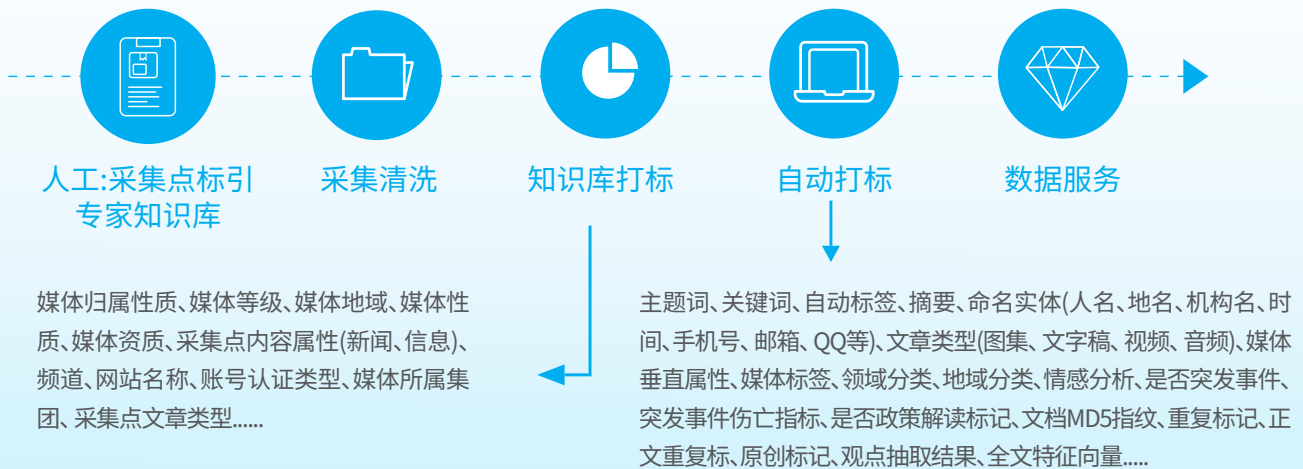
数字报刊 1100+国内报刊 央媒报刊 行业产业报 省级媒体报刊 地市级报刊 区县级报刊 商业报刊..... 企事业单位报刊.....	新闻网站 30万+网站采集点 中央重点新闻网站 互联网新闻信息服务单位 互联网新闻信息稿源单位 地方新闻网站 商业门户网站 各级政府网站 垂直行业网站.....	资讯客户端 3000+资讯APP 央媒资讯客户端 地方资讯客户端 行业资讯客户端 商业资讯客户端 政府机构客户端 垂直服务客户端 团体机构客户端.....	主流社交媒体 微信公众号+微博 微信公众号数据 每日200万+ 微博账号发文数据 每日5000万+	第三方平台 20+三方平台 今日头条 一点资讯 企鹅号 百家号 大鱼号 新浪看点 搜狐号 网易号 澎湃号 大风号 趣头条 抖音 快手 A站 B站 秒拍 火山 腾讯视频 优酷 梨视频
境外数据 63语种15万采集点	境外通讯社 Facebook	境外报刊 Twitter	境外广播电视台 Youtube	境外综合媒体 Instagram.....

数家数据采集范围

● 数据精加工及智能化标引

采用专家规范化标引+机器自动标引相结合的方式,实现对内容资讯的“精加工”。所有采集源由人工整理配置,保证了数据源头的“纯净”和完整性;配备了专家级的知识标引团队,对行业知识进行梳理、构建了数十个不同维度的媒体知识库,保证数据标签的多维度和精准度,更适配高信源数据在各行业领域的应用场景。利用NLP技术实现内容智能化标引的核心功能,保证数家库中的数据全部具备智能化知识属性,为行业应用夯实基础数据。

所有采集数据经过统一打标:数据结构化→属性知识化→内容标签化→场景服务化



数据智能化标引

● 深度的内容分析

利用传统机器学习+深度学习框架和大数据分布式计算框架等,对内容进行数百个维度的统计和分析挖掘,以形成各行业所需的细化场景应用服务。

● 精准化的持续迭代

专业专职的数据分析师和算法工程师以每日迭代的工作模式,以保证所有的分析结果都尽可能的客观准确。

应用场景

目前,平台已服务客户逾百家,涉及媒体、政府、企业、安全等行业领域,满足不同领域不同数据类型的定制化数据需求。

媒体行业

通过对互联网新闻数据的采集清洗、结合自然语言理解的标引分析能力,对传媒行业提供多元化、全链路的数据场景服务,助力传媒单位依托大数据能力实现媒体融合发展的深化和转型。

应用场景包括:

- 智能生产服务
- 新闻线索发现
- 热点挖掘分析
- 新闻传播效果分析
- 个性化资讯内容供给
- 榜单分析服务
- 专题追踪服务

政务领域

为政府行业提供基于互联网数据的数据研判、内容标引、政策分析、宣传影响力评价、政务绩效评估和专题定制等业务场景的数据支撑服务,助力政府单位加快数据赋能的数字化改革。

应用场景包括:

- 政府发布内容数据供给
- 政府网站影响力评测
- 政务新媒体影响力分析
- 政策法规影响力分析

企业

以数据为中心,通过行业政策及智库数据、竞争情报数据、品牌宣传评价、企业/行业热点分析、广告文案传播分析等数据服务,助力企业的数字化管理和转型。

应用场景包括:

- 行业信息定制加工
- 智库数据与分析
- 品牌传播效果评估
- 企业风控预警
- 重点事件分析

安全

基于互联网开源信息采集和数据挖掘度,实现网宣工作的到达率分析与合规性检查、为互联网空间治理提供数据监管手段;涵盖全球安全和防务智库数据的采集,以内容充实性来提升情报分析工作的多样性和及时性,为态势情报分析提供有力的信息保障与数据支撑。

应用场景包括:

- 开源情报定制采集
- 信息加工处理
- 智库数据分析
- 互联网空间治理

其他

数家提供互联网资讯基础服务、定向数据采集、内容数据治理、专题定制分析、行业/区域的新媒体榜单定制等多维度的内容供给和内容认知服务,助力泛行业客户实现基于数据决策的业务场景实现。

应用场景包括:

- 互联网资讯基础服务
- 定向数据采集
- 内容数据治理
- 专题定制分析
- 新媒体榜单定制
- 多维度的内容供给和内容认知服务

DAAS 数据即服务

- 内容供给服务
- 一键转载服务
- 互动数据服务
- 政策数据服务
- 行业数据服务
- 背景素材供给
- 自定义数据适配供给服务
- 源地址自动解析内容服务
- 融媒体平台数据资源服务
- 个性化数据看板.....

SAAS 软件即服务

- 线索发现服务
- 热点分析服务
- 情报分析服务
- 趋势预测服务
- 宣传监管服务
- 素材智能匹配
- 传播分析服务
- 榜单定制服务
- 版权追踪服务
- 专题追踪服务
- 谣言识别服务
- 精准推荐服务
- 定制数据分析及可视化.....

KAAS 知识即服务

- 每日自动简报服务
- 专题追踪报告服务
- 传播分析报告服务
- 知识体系构建服务
- 知识库及语料接口服务
- 机器写作服务
- 专家咨询服务.....

数家服务能力