Computer Science and Application 计算机科学与应用, 2026, 16(6), 222-231
Published Online June 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.166222

Hans汉斯

基于多源数据融合的氢能产业链决策服务系统设计与实现

刘力鸣[1]，刘　超[1]，陈晓玲[1*]，贾　超[2]

[1]吉林省科技创新研究院，吉林 长春
[2]吉林省翰德人才咨询有限公司，吉林 长春

收稿日期：2026年5月15日；录用日期：2026年6月19日；发布日期：2026年6月29日

摘　要

为提升氢能产业数据价值，实现数据服务支撑氢能产业智能决策，研发了氢能产业链决策服务系统。该系统应用数据治理技术，将氢能产业政策、专利和文献等多源数据融合；采用自然语言处理算法，分析融合数据的语义特征；通过知识图谱技术，呈现氢能产业发展趋势；借助熵权TOPSIS算法，监测产业发展情况。系统应用氢能产业的数据服务、计算服务和信息服务三层架构，实现了产业链信息导航、关键技术图谱和产业技术监测功能，支撑了氢能产业科技创新发展和技术攻关前瞻研究。

关键词

多源数据，氢能产业，决策服务，系统实现

Design and Implementation of a Hydrogen Energy Industry Chain Decision Support System Based on Multi-Source Data Fusion

Liming Liu[1], Chao Liu[1], Xiaoling Chen[1*], Chao Jia[2]

[1]Jilin Provincial Information Institute of Jilin Provincial Information Institute of Science and Technology, Changchun Jilin
[2]Jilin Province Hande Talent Consulting Co., Ltd., Changchun Jilin

Received: May 15, 2026; accepted: June 19, 2026; published: June 29, 2026

[*]通讯作者。

文章引用: 刘力鸣, 刘超, 陈晓玲, 贾超. 基于多源数据融合的氢能产业链决策服务系统设计与实现[J]. 计算机科学与应用, 2026, 16(6): 222-231. DOI: 10.12677/csa.2026.166222

Abstract

To enhance the value of data in the hydrogen energy industry and enable data-driven support for intelligent decision-making, a Decision Support System for the Hydrogen Energy Industry Chain has been developed. The system employs data governance techniques to integrate multi-source data including industrial policies, patents, and literature; utilizes natural language processing algorithms to analyze semantic features of the aggregated data; leverages knowledge graph technology to visualize development trends in the hydrogen energy sector; and applies the entropy-weighted TOPSIS algorithm to monitor industry progress. Built upon a three-tier architecture comprising data services, computing services, and information services, the system provides functions such as industry chain navigation, key technology mapping, and technological monitoring, thereby supporting scientific innovation and forward-looking research in hydrogen energy technology.

Keywords

Multi-Source Data, Hydrogen Energy Industry, Decision Support Services, System Implementation

1. 引言

氢能产业是技术门槛高、资本需求高、市场规模要求高的新兴产业赛道，世界各国的研究规划尚处在分析论证或总体设计阶段。2019 年，我国首次在《政府工作报告》[1]中提及氢能产业发展，同步推出多项举措推动其发展进程；2025 年 1 月 1 日起施行的《中华人民共和国能源法》[1]明确规定，“国家积极有序推进氢能开发利用，促进氢能产业高质量发展”。当前，我国氢能领域的政策基础环境已逐步构筑完善，氢能产业正式迈入产业化发展快车道，有望成为我国下一个重要经济增长点。

国外学者依托产业经济数据，对氢能未来发展趋势展开预测[2] [3]，系统梳理产业发展困境，规划长远发展路径[4]，并提出构建氢能知识平台、创新运营模式等具体方向，以助力氢能产业发展[5]。国内学者则通过深度挖掘氢能领域政策文本与专利数据，精准剖析产业发展现状与未来走向，为政策优化制定提供针对性建议[6]-[8]，同时，借助 AI 等前沿技术手段分析氢能产业面临的现实问题，进而提出切实可行的应对策略[9] [10]。

总之，现有研究虽已完成氢能领域政策与产业数据的初步汇总，但数据层面仍存在维度单一、体量有限、权威性不足等问题，且相关支撑系统普遍面临技术滞后、设计欠合理、用户体验不佳的困境。为达成“以科技数据驱动科技创新、以产业数据助力科学决策”的核心目标[11]，本研究特研发氢能产业链决策服务系统(简称“系统”)。该系统将应用自然语言处理、机器学习等先进技术，对氢能产业海量多源异构数据进行有效融合，进而构建关键技术图谱与产业节点评价模型，以此清晰呈现产业链关键技术动态与产业发展趋势，并实时监测产业节点运行情况，最终为氢能产业创新发展与科技攻关提供精准的数据服务及科学的决策支撑，助力氢能产业实现高质量发展、激发技术创新活力。

2. 研究方法和关键技术

在氢能计算服务功能模块中，研究采用潜在狄利克雷分配(Latent Dirichlet Allocation, LDA) [12]主题

[1]https://www.xinhuanet.com/politics/2019lh/2019-03/16/c_1124242390.htm

模型对文本数据进行深度挖掘，以精准提取关键信息；同时运用熵权法与优劣解距离法(Technique for Order Preference by Similarity to Ideal Solution, TOPSIS)相结合的算法，构建科学的产业节点评价模型，提升评价结果的客观性与可靠性。在信息服务功能模块，则通过知识图谱相关技术，系统性构建覆盖多维度信息的氢能产业知识图谱，为用户提供结构化、可视化的信息查询与整合服务。

2.1. 潜在狄利克雷分配主题模型

LDA 主题模型是自然语言处理与文本挖掘领域应用广泛的无监督学习模型，能够高效挖掘文本数据中的潜在主题结构。作为三层贝叶斯概率模型，LDA 的核心架构可描述为“文档(D) - 主题(Z) - 词(W)”三级结构：其核心逻辑在于，单个文档由多个主题混合生成，每个主题对应一组词的概率分布，而文档的生成过程则严格遵循概率抽样机制。该模型的损失函数具体如公式(1)所示：

$$p(D|\alpha,\beta)=\prod_{d=1}^{M}\int p(\theta_d|\alpha)\left(\prod_{n=1}^{N_d}\sum_{Z_{d,n}} p(Z_{d,n}|\theta_d)\,p(w_{d,n}|z_{d,n},\beta)\right)\mathrm{d}\theta_d \tag{1}$$

其中，D 代表语料数据集合、α 为主题层面的狄利克雷(Dirichlet)参数、β 为词层面的狄利克雷(Dirichlet)参数、θ_d 表示编号为 d 的文档的主题分布、$z_{d,n}$ 代表编号为 d 的文档中第 n 个单词所属的主题、$w_{d,n}$ 则对应编号为 d 的文档中第 n 个单词。

若以 K 代表所有主题集合、D 代表所有文档集合、N 代表单篇文档所含的全部词，且其余字母含义与公式(1)保持一致，则 LDA 算法的生成流程可概括如下：

(1) 针对任意文档 $d\in D$，由超参数 α 采样得到该文档的主题分布 θ_n。

(2) 针对文档 d 中的每个主题 z，由主题分布 θ_d 采样得到对应词的主题指派 $z_{d,n}$。针对所有主题，由超参数 β 采样得到每个主题对应的词分布 $\Phi_{z_{d,n}}$。

(3) 针对文档 d 中的每个词 $w_{d,n}$，由主题的词分布 $\Phi_{z_{d,n}}$ 与词的主题指派 $z_{d,n}$ 共同采样生成。

重复上述步骤迭代优化，最终得到更优的超参数 α 和 β(流程示意图见图 1)。具体而言，对于文档集合 D 中的每一篇文档，需通过最大似然估计函数求解超参数 α 与 β(二者均服从对称狄利克雷先验分布)，以最大化文档集合的生成概率 $p(D)$；为避免模型过拟合，进一步采用“困惑度模型结合似然分数”的方法，确定单篇文档的最优主题数量。

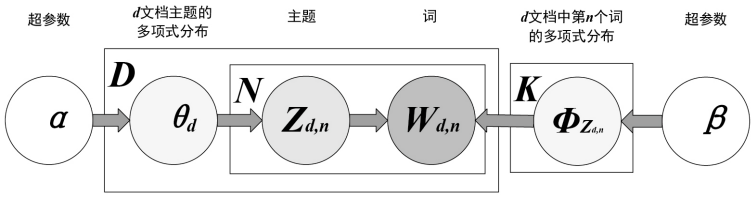

Figure 1. Schematic diagram of LDA model
图 1. LDA 模型示意图

在系统的氢能计算服务功能设计中，采用 LDA 主题建模算法，对氢能产业政策文本、专利文献及各类产业资讯数据展开深度挖掘。在此基础上，进一步结合高频关键词共现分析与主题强度演化追踪两种技术手段，从量化视角精准呈现政策核心关注点与技术突破方向的动态匹配关系；同时，通过多维度数据交叉分析，揭示政策导向、技术发展、市场热点三类创新要素间的协同演化规律，最终为氢能产业的政策优化调整与技术路径科学规划，提供具备数据支撑的决策依据。

2.2. 熵权－优劣解距离法的评价模型

熵权－优劣解距离法评价模型，是将熵权法与 TOPSIS 法有机融合而成的综合评价模型。其中，熵权法起到量化指标信息熵值的作用，以此反映各评价指标的相对离散程度；TOPSIS 法则聚焦多属性决策场景，用于实现评价对象的优选、排序、综合评估及决策支持[13]。

该模型核心逻辑为：先通过熵权法计算各评价指标客观权重，规避主观赋权偏差；再基于 TOPSIS 法构建评价体系，对各评价对象进行综合评分与排序。相较于单一评价方法，此模型兼具客观性强、数据适应性强的优势，且能输出评价对象与最优解、最劣解的贴近度等细节信息，为决策提供更丰富参考依据。具体计算过程如下：

第一步，数据标准化。采用极差法对原始数据矩阵 $X=\left(x_{ij}\right)_{mn}$ 进行标准化处理，其中 m 为样本数，n 为指标数，如公式(2)所示：

$$z_{ij}=\begin{cases}\dfrac{x_{ij}-\min x_j}{\max x_j-\min x_j} & \text{正向指标}\\ \dfrac{\max x_j-x_{ij}}{\max x_j-\min x_j} & \text{负向指标}\end{cases} \tag{2}$$

第二步，熵权法赋权。计算各评价指标权重，如公式(3)所示：

$$w_j=\frac{1-e_j}{\sum_{k-1}^{n}\left(1-e_k\right)} \tag{3}$$

其中，e_j 是第 j 项指标的信息熵，如公式(4)所示：

$$e_j=-\frac{1}{\ln m}\sum_{i=1}^{m}\frac{z_{ij}}{\sum_{i=1}^{m}z_{ij}}\ln\frac{z_{ij}}{\sum_{i=1}^{m}z_{ij}} \tag{4}$$

第三步，TOPSIS 模型运算。先构造理想解与负理想解，再计算各评价对象的相对接近度，若 D 表示欧氏距离，则如公式(5)所示：

$$C_i^{-}=\frac{D_i^{-}}{D_i^{-}+D_i^{+}},\quad 0\le C_i\le 1 \tag{5}$$

在系统的氢能计算服务功能中，采用熵权法结合 TOPSIS 模型，针对上游制氢、中游储运及下游应用环节构建多维度指标体系，涵盖创新投入、创新环境、创新绩效等核心指标。具体流程为：先通过熵权法计算各指标客观权重，消除传统赋权的主观偏差；再运用 TOPSIS 算法测算各环节与理想解的贴近度；最终量化评估显示。该方法有效揭示产业链的梯次发展特征，为资源配置提供具有空间异质性的优化路径。

2.3. 知识图谱技术

知识图谱是一种以图结构组织知识的语义网络，通过实体(Entity)、关系(Relation)与属性(Attribute)的三元组(Triple)形式，表征现实世界中的概念及其关联关系。其核心目标是让机器能够理解、推理并应用知识，主要构建步骤包括知识获取与抽取、知识表示与存储、知识融合与推理[14]，目前已在智能搜索与问答、垂直领域决策支持、推荐系统、自然语言处理等领域发挥重要作用。

在系统的信息服务功能中，采用知识图谱技术建立论文、技术、人才、机构等创新要素间的关联，可直观呈现氢能产业发展要素的复杂联系。该技术不仅能为产业技术创新提供基础框架、支持技术研发思路构建，还能为政策制定、投资决策等场景提供有价值的参考。

3. 系统设计与实现

围绕产业数据采集与治理、数据分析和挖掘、产业节点评价模型展开设计，应用成熟可靠的开源框架进行快速开发，部署在云环境中进行统一的系统运维管理。

3.1. 体系架构设计

通过分析业务和系统负载状况，将系统解耦成三层服务结构(如图 2 所示)。由于多源异构数据的数据格式、数据质量、到达时间等多方面差异性，数据的处理难度随数据体量增加呈指数型上升。为了构建统一数据平台，系统在数据服务层集成数据采集、数据清洗和数据仓储等功能。系统在计算服务层集成数据挖掘、数据接口、数据缓存服务，用于实现数据分析和数据传输交换等业务功能。计算服务层对外提供接口，为可视化平台以及其他扩展业务提供计算后的数据结果。系统在信息服务层集成数据可视化服务，包括氢能产业链信息导航、氢能产业关键技术图谱、氢能产业要素词云图和氢能产业链节点监测等模块。

Figure 2. System architecture diagram
图 2. 系统体系架构图

3.2. 氢能产业数据功能实现

氢能产业数据服务是集数据采集、数据清洗和数据仓储于一体的大数据处理服务，基于 Lambda 架构模式[15]实现批处理加实时处理的数据双通路架构，解决了大数据场景下高准确性和低延迟的矛盾，使系统整体兼顾准确性、实时性和容错性。在氢能产业数据功能实现中使用 LDA 主题模型作为数据预处理工具。针对多源异构数据中大量非结构化文本(如政策文件、技术文献、资讯报道等)，通过 LDA 主题模型进行潜在主题挖掘，从文本集合中自动提取氢能产业相关的核心主题(如制氢技术、储运装备、政策法规等)，并量化文本与各主题的关联概率。

在数据采集阶段，通过数据爬取工具从网页或 Application Programming Interface 中采集，并借助 Kafka-Python 将数据写入 Kafka 中。数据库变更通过 Debezium 捕获，日志文件通过 Filebeat 实时上传，确保多源数据的统一接入。

(1) 在数据处理阶段，按照数据类型进行批流差异化处理。在处理过程中先对采集的非结构化文本数

据进行预处理，包括分词、去停用词及词向量转换，构建文档–词项矩阵。采用 Gibbs 采样算法训练 LDA 模型，设定主题数 K，迭代优化得到主题–词项分布与文档–主题分布。通过困惑度指标评估模型收敛性，确定最优迭代次数。将输出的主题特征作为标签，关联至流式与批式数据处理流程，为 Flink 实时校验提供语义过滤规则，为 Spark 离线处理的维度关联提供主题维度，实现非结构化数据向结构化特征的转化。对于 Kafka 中流式数据内容，使用 Flink 进行数据有效性校验、敏感字段脱敏以及实时统计指标计算等消费活动，清洗后的数据分别写入 Kafka 实时 Topic 和 HBase 实时存储，供下游服务消费(如图 3 所示)。对于离线批数据使用 Spark 每 6 小时定时处理 HDFS 原始数据，进行维度关联、全量统计指标计算和数据质量审计，最终将结果以 Parquet 格式存储到 HDFS 并注册到 Hive 表中(如图 4 所示)。

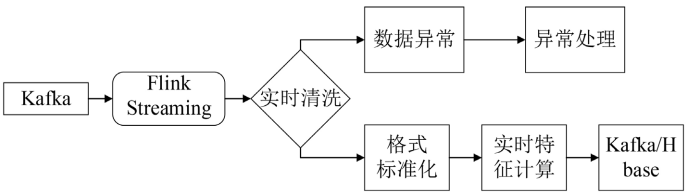

Figure 3. Streaming data processing flow
图 3. 流式数据处理流程

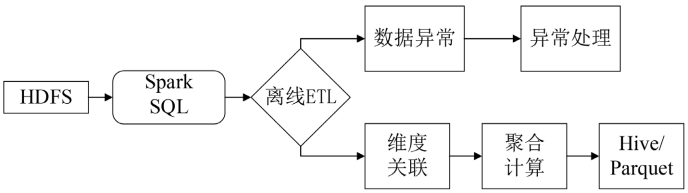

Figure 4. Batch data processing flow
图 4. 批式数据处理流程

(2) 在数据存储阶段，系统采用分层存储策略，对原始数据、实时数据和聚合数据采用不同的存储方式(如表 1 所示)。针对原始数据需长期存储全量内容的特点，选用 HDFS 分布式架构实现海量存储与高容错，搭配 Parquet 列式格式减少离线分析 IO 开销，此选型具备低成本、高查询效率的特点，适配高查询需求的业务。针对实时数据存储数据到达量大的特点，采用 Kafka 对接 Flink 实时消费数据，HBase 存储近 72 小时状态数据，数据实时写入 Kafka 与 HBase 闭环，HBase 毫秒级响应高并发请求，自动清理过期数据降成本，满足高并发读写，贯通流式处理链路。针对聚合数据量较大、需要频繁查询的特点，选用 Cassandra 存结构化指标，Elasticsearch 提供多维检索。此技术选型针对复杂查询效率更高，分布式架构保障 TB 级数据容错，并支持实时报表分析。

Table 1. Lists the tiered storage policies
表 1. 分层存储策略表

序号	存储类型	技术方案	数据特征
1	原始数据	HDFS + Parquet	全量原始数据，长期存储
2	实时数据	Kafka + HBase	最新 3 天数据，高并发读
3	聚合数据	Cassandra + Elasticsearch	预计算指标数据，快速查询

(3) 在数据服务阶段，Presto 联合查询 HBase 和 Hive，支持实时和历史数据的统一访问。SparkSQL 用于复杂分析，而 Superset 则提供可视化展示。通过这种分层处理机制，系统实现了毫秒级实时响应与

复杂离线分析的无缝整合，满足大规模数据处理的多样化需求。

3.3. 氢能计算服务功能实现

系统的产业链信息导航、智能检索、研究热点和关键技术图谱功能需要预先处理海量数据，进行分类和关键词提取等操作。为了提升数据计算效率，系统在氢能计算服务中实现了数据挖掘、缓存和 API 接口的功能。

系统基于 Elasticsearch 分布式搜索引擎构建了高效的多维信息抽取体系。使用 Bulk API 将数据仓库中的元数据批量导入 Elasticsearch 集群，利用其倒排索引机制实现数据的高速存储与检索。通过语义分析等业务操作后，将数据结果根据不同的类型进行持久化处理(如图 5 所示)。

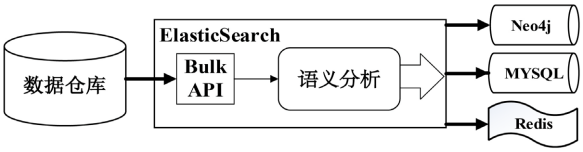

Figure 5. Data processing diagram
图 5. 数据处理示意图

语义分析流程如下：首先，建立 LDA 主题模型，提取文本中的关键词和主题。然后，构建以主题为节点的隐马尔可夫链(Hidden Markov Model, HMM)，利用 Pagerank 算法评估技术主题的重要性。最后，设定阈值，筛选出得分高于阈值的主题，作为产业关键技术主题。

3.4. 氢能信息服务功能实现

氢能信息平台是面向用户展示的数据可视化服务平台，主要包含了产业链信息导航、产业关键技术图谱和产业节点监测功能。信息平台采用前后端分离的开发框架，前端使用 React + TypeScript + Webpack + AntV，后台使用 Django REST Framework + Celery，微服务部署在 Docker + Kubernetes+Jenkins 的云环境中(如图 6 所示)。

Figure 6. Technical architecture diagram
图 6. 系统技术架构图

在产业链信息导航功能实现中，首先，将氢能产业划分为上游制氢、中游储运和下游应用三个层级，包含了 10 个门类 28 个方向的氢能产业节点；其次，基于氢能计算服务提供的数据接口，使用 Elasticsearch 检索产业节点作为关键词，分类提取每个节点的政策动态、资讯动态、市场趋势、机构信息和人才专家等创新要素资讯；最后，根据关键词将提取结果与产业节点建立对应关系(如图 7 所示)。

Figure 7. Examples of hydrogen energy industry chain information navigation
图 7. 氢能产业链信息导航实例

在产业关键技术图谱功能实现中，根据关键技术、前沿技术、现代技术和颠覆性技术对数据进行划分，实现了产业关键技术图谱、研究趋势图、研究热点词云图和领先地区柱状图功能(如图 8 所示)。关键技术图谱整合了论文、技术、人才及机构四维节点数据，依托氢能计算服务构建的 Neo4j 图数据库存储结构化元组信息，通过 ECharts.js 组件实现多维度关系映射与动态交互可视化。研究趋势、研究热点和词云图从多个角度对产业节点进行分析，使用 Echarts-Wordcloud 等前端组件，调用氢能计算服务的数据接口实现。

Figure 8. Example of key technology map of hydrogen energy industry
图 8. 氢能产业关键技术图谱实例

在产业节点监测功能实现中，系统使用了基于熵权法结合 TOPSIS 的评价模型，结合企业分布和专利分布的综合性关键技术指标(如表 2 所示)，对产业链上下游的关键技术节点发展情况进行评估(如图 9 所示)。针对上游制氢、中游储运及下游应用环节构建多维度指标体系，涵盖创新投入、创新环境、创新绩效等核心指标。首先基于熵权法计算各指标客观权重，消除传统赋权的主观偏差，然后运用 TOPSIS 算法测算各环节与理想解的贴近度，最后量化评估显示，该方法有效揭示产业链的梯次发展特征，为资源配置提供空间异质性优化路径。

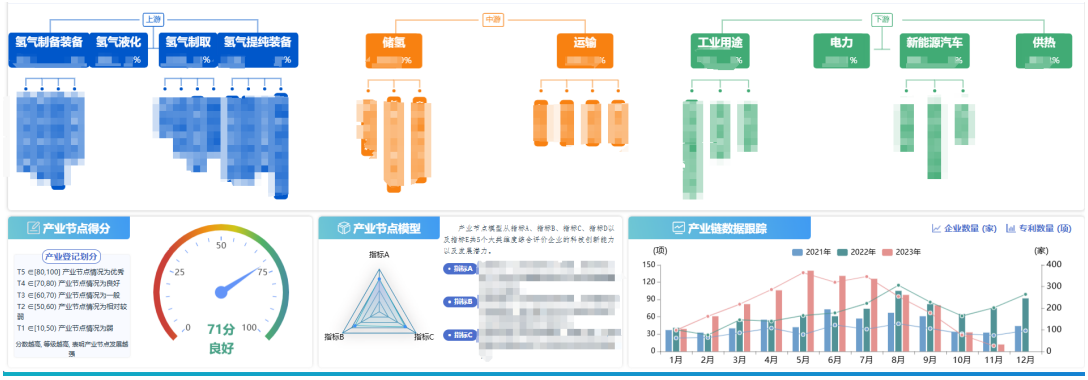

Figure 9. Example of node monitoring function in hydrogen energy industry
图 9. 氢能产业节点监测功能实例

Table 2. Index table of evaluation model
表 2. 评估模型指标表

序号	一级指标名称	序号	二级指标名称	二级选取原因
1	创新投入	1	科技研发人员数量	统计目标单位专职从事氢能相关技术研发的人员总数，直观反映企业创新人力投入规模。
		2	研发机构数量	计算目标单位设立或参与的氢能技术研发实验室、研究中心等机构数量，体现创新资源集聚程度。
		3	科技企业数量	统计氢能产业链上下游从事关键技术研发与应用的科技企业数量，衡量产业创新主体活跃度。
		4	资金投入金额	汇总目标单位在氢能技术研发、设备购置等方面的年度资金投入，量化创新资金支持力度。
2	创新绩效	5	发明专利申请授权数量	统计目标单位获得授权的氢能领域发明专利数量，体现核心技术原创能力与技术壁垒高度。
		6	实用新型专利申请授权数量	统计目标单位授权的氢能技术改进类专利数量，反映技术应用创新的实践成果。
		7	核心期刊论文产出数量	统计目标单位人员在氢能领域核心期刊发表的学术论文数，展现前沿技术研究深度与理论创新水平。
3	创新环境	8	科技政策数量	统计地方或国家针对氢能产业发布的专项政策文件数量，体现政策环境支持密度。
		9	政策力度	通过政策文件中资金补贴、税收优惠等量化条款赋值打分，评估政策对产业发展的实际推动强度。

4. 结语

该系统通过大数据和人工智能技术，实现多源异构数据的融合与集中管理，构建关联共享的智能化

服务平台，整合氢能产业链的多维度信息，为政府、企业和投资者提供精准的决策支持，助力产业在复杂市场环境中稳健发展。通过实时监测产业链节点数据，系统帮助科技管理者全面掌握产业链完整性及节点强弱情况，为政策制定和技术创新提供科学依据，有效推动了氢能产业的高质量发展和持续升级。

未来研究将拓展数据融合维度，引入国际氢能市场交易、企业运营动态等实时数据流，构建跨国别产业对标分析体系；深化 AI 技术应用，蒸馏大语言模型提升系统维护效率，开发产业技术成熟度预测与风险预警模型；优化知识推理机制，探索基于时序图谱的动态推演算法，实现产业链断点识别与补链路径智能推荐。

基金项目

2024 年吉林省科技发展计划项目重点研发，项目名称“基于决策驱动的科技资源智能感知系统研发与应用”(项目编号：20240302071GX)。

参考文献

[1] 中华人民共和国能源法[Z]. http://www.npc.gov.cn/npc/c2/c30834/202411/t20241108_440884.html, 2024-11-08.

[2] Popel', O.S. and Tarasenko, A.B. (2022) Promising Directions for Hydrogen Energy Development in Russia. *Applied Solar Energy*, **58**, 152-158. https://doi.org/10.3103/s0003701x22010133

[3] Behrendt, F. (2025) The Future of Industrial Hydrogen: Renewable Sources and Applications for the Next 15 Years. *Clean Energy*, **9**, 3-8. https://doi.org/10.1093/ce/zkae103

[4] Ziobrowski, Z. and Rotkegel, A. (2024) Assessment of Hydrogen Energy Industry Chain Based on Hydrogen Production Methods, Storage, and Utilization. *Energies*, **17**, Article 1808. https://doi.org/10.3390/en17081808

[5] Dawood, F., Shafiullah, G.M. and Anda, M. (2020) A Hover View over Australia's Hydrogen Industry in Recent History: The Necessity for a Hydrogen Industry Knowledge-Sharing Platform. *International Journal of Hydrogen Energy*, **45**, 32916-32939. https://doi.org/10.1016/j.ijhydene.2020.09.122

[6] 黄晓林, 胡锡晟, 黄卉, 等. 中国氢能源产业政策量化分析及区域布局研究[J]. 科技情报研究, 2021, 3(2): 83-95.

[7] 付强, 杨洸, 金辉, 等. 中国氢能产业链技术现状及发展趋势[J]. 油气与新能源, 2024, 36(4): 19-30.

[8] 李建林, 李光辉, 郭丽军, 等. “十四五”规划下氢能应用技术现状综述及前景展望[J]. 电气应用, 2021, 40(6): 10-16.

[9] 王嘉炜, 妙丛. 基于大数据分析的加氢站智能管理系统[J]. 科技创新与应用, 2022, 12(17): 27-30.

[10] 张卫, 刘延杰. 基于大数据视角的新兴产业环境分析——以氢能产业为例[J]. 智库理论与实践, 2023, 8(4): 87-100.

[11] 付强, 陈晓玲, 李沫, 等. 基于深度整合的科技资源系统的设计与实现[J]. 吉林大学学报(信息科学版), 2022, 40(6): 924-929.

[12] Blei, D.M., Ng, A. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, **3**, 993-1022.

[13] 喻啸林, 周建华. 熵权-TOPSIS 法在高校网络舆情监测预警中的应用[J]. 电脑与信息技术, 2025, 33(1): 69-73.

[14] 杨航, 张啸成, 张永刚. 基于实体类别信息的数据分析与关系抽取模型构建[J]. 吉林大学学报(理学版), 2025, 63(2): 428-436.

[15] 张炯娜, 王若瑶, 何晓茜. 基于 AIS 算法和 Lambda 的医学图书馆推荐系统构建[J]. 信息技术, 2024(12): 80-85, 92.