面向科技情报服务的内容安全关键技术体系研究*

张昱　吴振峰
（中国科学技术信息研究所，北京 100038）

摘要：保障内容安全是做好科技情报服务的重要基础工作。开展面向科技情报服务的内容安全关键技术体系研究，对于保障科技信息资源内容安全、提供优质科技情报服务具有重要意义。针对科技情报服务全过程面临的内容安全防护需求，提出了以内容比对数据等基础保障为支撑，涵盖内容安全审核、开源风险监测、智能监控预警3个维度的内容安全关键技术体系，设计并研发了相关的内容安全防护功能。该研究可为提升科技情报服务全过程内容安全管理水平和效率提供参考借鉴。

关键词：科技情报服务；内容安全；技术体系
中图分类号：TP309　DOI：10.3772/j.issn.1673-2286.2024.06.009
引文格式：张昱，吴振峰. 面向科技情报服务的内容安全关键技术体系研究[J]. 数字图书馆论坛，2024，20（6）：83-90.

科技情报服务以情报判断为中心对数据和信息进行处理、组织和解释，以揭示其潜在的知识，并转化为可执行利用的决策支持或决策引导[1]。随着互联网和信息技术快速发展，信息来源趋向大众化、多媒体化，科技情报服务内容与方式发生深刻转变，同时信息具有渠道丰富、形式多样、时效性强等特点，导致信息内容真实性和安全性存在辨析难点[2-3]。以开源信息为例，各种类型的开源信息不仅具有公开性和可得性，还呈现数据规模大、来源广、流动性强等特点，这就导致科技情报服务过程中信息违法违规、重要信息泄露、内容造假等内容安全风险增大，科技情报服务的不确定性增加。开源情报视角下，互联网信息安全研究主要集中在数据采集、存储、跨境流动等方面，对于信息内容自身的合法合规性、敏感性等方面的技术研究较少[4]。因此，开展面向科技情报服务的内容安全关键技术体系研究，对于保障科技信息资源内容安全、提供优质科技情报服务具有重要意义。

在此背景下，本研究针对科技情报服务过程中面临的内容安全防护需求，在调研典型问题、关键技术、应用实践等方面研究与实践现状的基础上，从内容安全审核、开源风险监测、智能监控预警等方面梳理面向科技情报服务的内容安全关键技术体系，以期为科技情报领域的内容安全研究和实践提供借鉴。

1　相关研究与实践现状

1.1　面向科技情报服务的典型信息安全问题研究现状

面向科技情报服务的信息安全研究取得一些进展。通过对相关文献进行梳理，发现典型的信息安全问题主要体现在违法违规、信息泄露、内容造假等方面。

收稿日期：2023-08-23
*本研究得到高校人文社会科学重点研究基础重大项目“产业技术创新智能情报分析平台研究”（编号：22JJD870005）资助。

针对违法违规相关问题，罗娇等[5]认为科学数据安全是科技情报安全的重要组成部分，从知识产权视角将科技情报中的常见数据安全问题总结为产权缺位、主权缺失、共享受阻3个方面，并分别提出了科学数据管理的相关建议，保障科学数据的知识产权安全；朱世强等[6]开展了基于人工智能的内容安全发展战略研究，并指出阿里巴巴、百度、中国信息通信研究院等机构基于大规模标注数据和人工智能技术，从涉政、暴恐、色情等多个维度识别多媒体内容中的违法违禁内容。针对信息泄露相关问题，高美玲等[7]认为科技情报的信息安全深刻影响国家科技安全，将信息泄露划分为主动泄密、过失泄密、被动泄密3种类型，并指出信息泄露的原因主要包括学术交流细节的过度披露、大数据关联分析、情报机构的恶意攻击与信息收集等；Abbas[8]认为泄密人员通常使用图像融合技术与密码技术，对情报信息进行加密变换，生成看似正常且难以辨识的密码图，通过社交媒体传递泄密信息。针对内容造假相关问题，苏鹏等[9]围绕科技情报中的虚假情报信息研究，引入“信息迷雾”理论，分析了科技情报中“信息迷雾”的概念、形成及威胁，并探索了科技情报“信息迷雾”的应对策略；Nadia等[10]认为具有极高欺骗性的深度伪造技术借助深度学习算法，制作或修改视频、音频、图片、文本内容，以呈现出高度逼真但与实际不符的事物，加剧了社交媒体虚假信息的传播。

1.2　基于人工智能的内容安全关键技术研究现状

内容安全技术能够提高内容甄别、保护等能力，是提供优质科技情报服务的重要基础保障。人工智能等先进技术快速发展，推进内容安全治理智能化、高效化、精准化，赋能内容安全关键技术研究[6]。通过对相关文献进行梳理，发现基于人工智能的内容安全关键技术研究主要体现在信息全生命周期保护、新技术安全应用、网络舆情管理等方面。信息全生命周期保护方面，刘少芳[11]通过梳理加拿大图书馆的研究数据安全管理模式，提出我国应完善多层级的研究数据安全管理政策体系、抓住数据生命周期的关键环节做好数据安全管理工作、加强研究数据平台建设保障数据安全等建议；吴振峰[12]指出隐私计算和区块链等关键技术被用于助力数据开放共享、加强数据隐私保护、降低数据泄露风险和服务数据价值流通等实际场景中，辅助信息生命周期各阶段中的数据权属界定和数据安全风险管控。新技术安全应用方面，李盼等[13]围绕机器学习算法模型安全问题，梳理了基于人工智能技术的算法安全评估、训练过程防御、测试推理过程防御、数据安全隐私保护等防护技术，为防御算法模型在训练和推理过程中的安全威胁提供参考；针对深度伪造技术谬用现象，大量虚假内容检测技术涌现，着力识别人工无法审核的内容，在虚假新闻检测、人脸资料审核等实际应用场景中发挥作用[14-15]。网络舆情管理方面，社会网络分析、话题检测与跟踪等关键技术可用于识别敏感信息和不规范内容。例如：可通过对文本数据进行语义网络分析，自动识别阴谋论文本[16]；吴振峰等[17]提出了一种网络新闻热点话题识别方法，辅助进行网络开源信息舆情监测。

1.3　面向行业服务的内容安全防护系统应用实践现状

随着大数据、人工智能等先进技术的加速创新并日益融入经济社会发展各领域，内容风险防控技术水平不断提高，面向多个行业及领域的内容安全防护软件系统落地应用。例如：随着国家信息安全政策的支持力度加大、用户需求增多，百度、网易等互联网企业分别推出了百度智能云、网易易盾等内容审核软件系统，借助人工智能和大数据技术，对文本、图片、视频等多媒体内容进行敏感性和规范性审核，覆盖涉政、涉黄、暴恐等多个审核维度，大幅降低了人工审核成本[18]；Akamai公司开发的EdgeAnalytics能够在网络上快速分析网络数据流中的可疑行为，EdgeScape能够对远程用户的上网方式和具体位置进行精确追踪定位[19]；英国边境管理、移民执法和签证及移民系统基于国家情报模型实现数据共享，并在此基础上实现情报主导，强化移民数据安全管控[20]。这些工作为面向科技情报服务的内容安全应用工具开发提供了有益借鉴。

综合来看，内容安全是信息安全领域中的一个重要分支，已有研究在剖析信息安全问题、加强新型防护技术手段应用、开发内容安全防护系统等方面取得一些进展，但是缺乏面向科技情报服务的内容安全关键技术体系的针对性设计。因此，本研究提出了以内容比对数据等基础保障为支撑，涵盖内容安全审核、开源风险监测、智能监控预警3个维度的内容安全关键技术体系，着力满足科技情报服务全过程的内容安全防护实

际需要。在此基础上，研发具有实际应用价值的内容安全防护技术工具。

2　内容安全关键技术体系结构

本研究所指面向科技情报服务的内容安全防护，是针对科技情报服务全过程所涉及的文本、图片、音视频等资源，开展内容审核，重点检测知识产权侵权、涉政涉黄等违法违规信息，党政、科技等重点领域的涉密或敏感信息，虚假新闻、网络谣言等内容造假信息，以及错别字词、标点符号错误等不规范表述，通过及时处置可能存在的风险隐患，确保科技情报服务全过程的信息内容安全，同时防止非法内容的传播和利用。为此，本研究提出了以内容比对数据等基础保障为支撑，涵盖内容安全审核、开源风险监测、智能监控预警3个维度的内容安全关键技术体系（见图1）。

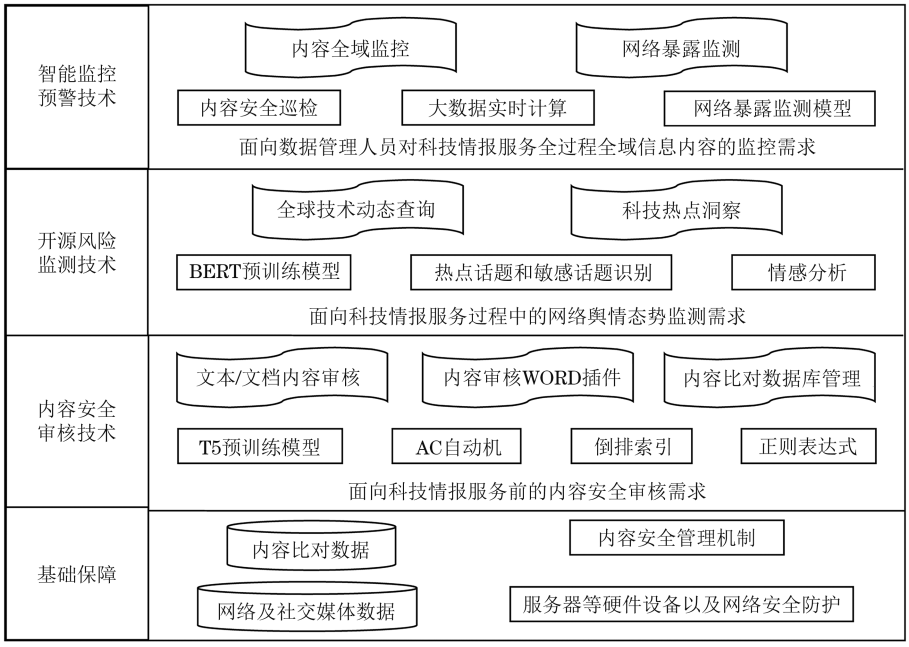

图1　内容安全关键技术体系结构

基础保障层是支撑内容安全关键技术体系的基础，包括基础数据资源、服务器等硬件设备及网络安全防护、内容安全管理机制。基础数据资源包括内容比对数据、网络及社交媒体数据等。其中：内容比对数据包括敏感信息库、规则库等，为内容安全审核技术提供参考依据和模型训练语料；网络及社交媒体数据包括科技领域开源网站、微信公众号等新媒体数据，为开源风险监测技术提供数据支撑。

内容安全审核技术以科技情报信息资源为基础，应用T5预训练模型、AC自动机、倒排索引等技术，通过对内容比对数据与科技情报信息资源进行检索和内容特征比对，快速自动检测党政、科技领域的敏感信息和错误表述，满足科技情报服务前的内容审核需求。

开源风险监测技术以网络及社交媒体数据为基础，应用BERT预训练模型、情感分析、热点话题和敏感话题识别等技术，开展开源情报专题分析，并获取相关资讯，强化科技情报服务过程中的网络舆情态势监测和风险内容屏蔽。

智能监控预警技术以科技情报服务全过程所涉及的信息资源为监测源，应用内容安全巡检、网络暴露监测模型、大数据实时计算等技术，融合内容安全审核技术和开源风险监测技术的分析结果，辅以人工筛选与深度分析，实现对风险信息内容的科学研判、分级预警和监控管理，以满足数据管理人员对科技情报服务全过程全域信息内容的监控需求。

3　内容安全主要关键技术

3.1　内容安全审核技术

对科技情报服务全过程所涉及的信息资源进行内

容安全审核，是确保科技情报信息资源的内容符合政治、法律、道德等层次要求，强化科技情报服务全过程内容安全管理的重要举措。内容安全审核是人工智能技术应用的热门方向，尽管相应软件包、建模工具日益增多，但是已有内容安全审核模型仍然面临文本语义理解不够深入、审核效率不够高等现实问题。预训练大语言模型、深度语义匹配模型等技术日新月异，已在文本自动纠错、领域知识挖掘等方面广泛应用。因此，本研究综合T5预训练模型、倒排索引等算法模型，提出了基于混合策略的内容安全审核技术实现流程（见图2）。

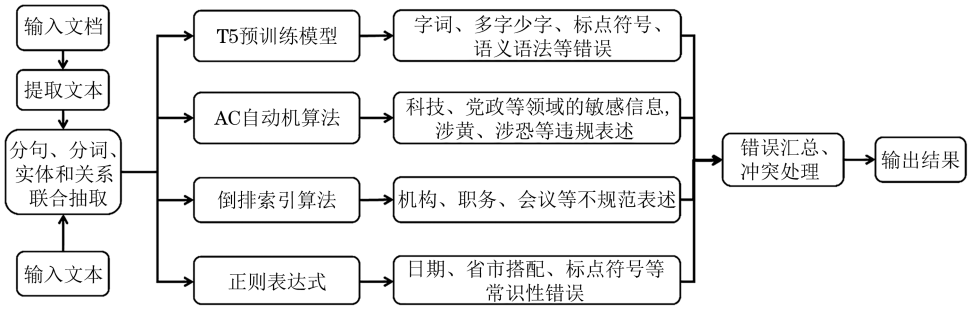

图2　内容安全审核技术实现流程

首先，对模型输入文本或文档进行预处理。若模型输入为文本，对文本进行分句、分词以及实体和关系联合抽取等预处理操作；若模型输入为WORD、PDF或图片格式的文档，应用主流的文档解析技术将文档转换为文本，进行上述相同的文本预处理操作。

随后，通过综合应用T5预训练模型、AC自动机、倒排索引、正则表达式，开展内容安全审核。

（1）T5预训练模型在自然语言处理领域应用广泛，采用基于Transformer Encoder-Decoder结构的自监督预训练策略，具有可扩展性良好、模型参数少、性能优越等优点[21]。本研究基于自行构建的模型训练语料库对T5预训练模型进行微调，同时对模型输出打分策略进行改良，以快速识别字词错误、多字少字、标点符号错误、语义语法错误等信息，并给出纠错建议。

（2）AC自动机算法是常用的字符串多模式匹配算法，包括构建字典树、构建AC自动机、在Trie树上进行多模式匹配等主要步骤，具有高效、实时的优点。本研究以自行构建的内容比对数据库为基础，应用AC自动机快速识别党政、科技等领域的敏感信息，以及涉黄、涉恐等违规表述，并给出预警提示。

（3）倒排索引算法是现代搜索引擎的核心技术之一，具有信息检索效率高、检索复杂度低等优点。本研究以自行构建的内容比对数据库为基础，通过倒排索引快速检索和识别机构、职务、会议等不规范表述，并给出预警提示。

（4）正则表达式在字符串检索、抽取、替换等方面应用广泛，在本研究中用于识别日期、省市搭配、标点符号等常识性错误，并给出预警提示。本研究的正则表达式由字母、数字等普通字符，以及预定义的匹配模式、量词和边界匹配等特殊字符组成。

最后，模型汇总识别的错误并进行冲突处理，形成最终的审核结果。

面向科技情报服务前的内容安全审核需求，基于内容安全审核技术，开发文本内容审核、文档内容审核的应用功能以及内容审核WORD插件工具。基本逻辑是：以内容比对数据库为基础，应用内容安全审核模型对文本或文档进行多维度的内容审核，实现对错误表述和敏感信息的快速自动检测，筛选校正不规范文本和文档。内容审核的类型包括科技、党政等敏感信息和不规范表述，以及字词错误、多字少字错误、标点符号错误和日期、机构、人名、省市搭配等错误表述。

（1）文本内容审核功能。依次选取审核类型、审核领域，并提交拟审核内容，审核后根据风险类型对文本内容进行颜色高亮显示，审核结果涵盖风险等级、参考依据、修改建议等信息，方便参考修正。

（2）文档内容审核功能。与文本内容审核功能类似，选取审核类型、审核领域后，单篇或批量上传WORD、PDF、TXT、JPG、PNG、JPEG、TIF等格式的文档，审核结果将存储到WORD文档，并以批注方式提供风险等级、参考依据、修改建议等内容，方便参考修正。

（3）内容审核WORD插件。按照程序安装插件后，本地的WORD文档将具备内容安全审核相关功能，

在编辑WORD文档内容时可以实现边审边改。审核页面将展示存在的问题，并提供风险等级、参考依据、修改建议等信息，方便实时编辑修正。

（4）内容比对数据库管理。内设敏感信息库管理、规则库管理、账号管理、角色管理、历史记录信息管理、统计分析等子功能，支持根据实际需求对内容比对数据库进行更加细粒度的个性化配置，支撑算法模型迭代升级，辅助改善内容审核效果。

3.2　开源风险监测技术

网络及社交媒体数据是反映网络舆论状况的重要信息集合，是人们获取信息、洞察趋势的重要来源。针对内容安全技术相关研究对网络及社交媒体数据挖掘程度不够高的问题，开展基于网络及社交媒体数据的风险监测，有助于捕捉网络环境中关键信息所含有的隐性情报内容，并预测事件发展趋势，对于面向科技情报服务的内容安全工作具有重要的现实意义。因此，本研究以内容比对数据库、网络及社交媒体数据为基础，提出了以热点话题和敏感话题识别模型为核心的开源风险监测技术实现流程（见图3）。

（1）数据预处理阶段。首先，对文本数据剔除噪声（如特殊字符、标点符号等）、分词、剔除停用词等，选取词频-逆文本频率（Term Frequency-Inverse Document Frequency，TF-IDF）值排名靠前的20个关键词作为文本的代表性关键词。随后，将所有文本数据的关键词TF-IDF向量组合成TF-IDF矩阵。考虑到新闻文本数据具有噪声多和维数高的特点，通过奇异值分解（Singular Value Decomposition，SVD）将TF-IDF矩阵进行压缩，得到矩阵 $\boldsymbol{Q} \in \boldsymbol{R}_{n\times q}$，剔除数据噪声和冗余信息，提高数据表达能力。

（2）最优话题数量的决策阶段。大多数情况下，对于一个真实数据集，真实的话题数量往往不明确，如果仅通过猜测指定任意的话题数量，往往会形成误导而不能获取可靠的话题识别结果。因此，本研究使用Krzanowski-Lai（KL）指标估计数据集的最优话题数量 k，如式（1）所示[22]。

$$\begin{cases} k = \arg\max C_q(k) \\ C_q(k) = \left| \dfrac{\mathrm{DIFF}_q(k)}{\mathrm{DIFF}_q(k+1)} \right| \\ \mathrm{DIFF}_q(k) = (k-1)^{\frac{2}{q}} S_q(k-1) - k^{\frac{2}{q}} S_q(k) \\ S_q(k) = \sum_{l=1}^{k} \sum_{s=1}^{n_l} \sum_{j=1}^{q} \left(x_{lsj} - \dfrac{1}{n_l} \sum_{s=1}^{n_l} x_{lsj} \right)^2 \end{cases} \tag{1}$$

式中：n_l 为第 l 个话题的样本数量；x_{lsj} 为矩阵 $\boldsymbol{Q}$ 中属于第 l 个话题的第 s 个样本对应的第 j 列的元素；DIFF为差分函数。

（3）热点话题和敏感话题识别阶段。针对已有的话题检测方法对数据的内在结构信息利用不够充分、话题检测的准确度总体不高等问题，本研究进一步改进了基于共享最近邻和马尔科夫聚类的网络新闻话题检测方法[17]，通过综合考虑数据样本共享最近邻的数量、秩次等信息定义数据样本间的关联强度，进一步构建共享最近邻图，使用马尔科夫聚类算法进行热点话题检测，在此基础上应用深度语义匹配技术将检测到的热点话题与自行构建的敏感信息库进行检索和比对，

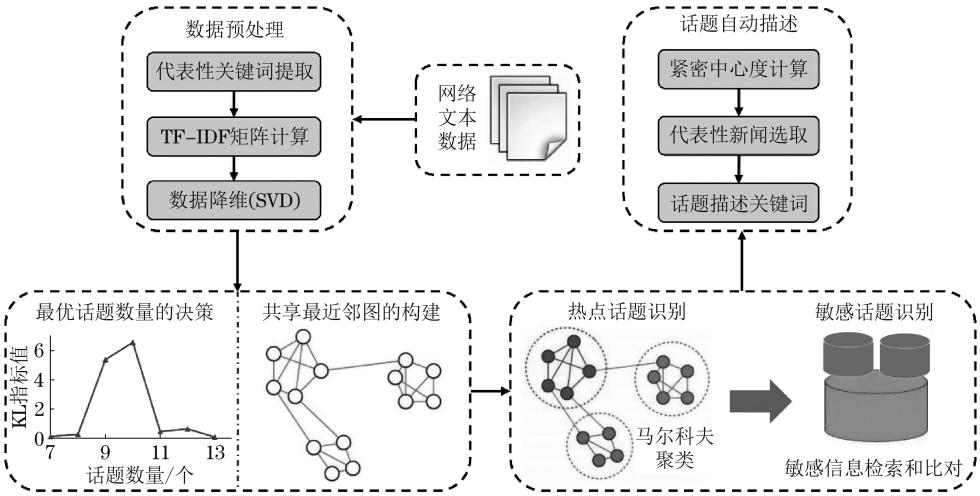

图3　开源风险监测技术实现流程

进而识别敏感话题。第 i 和 j 个数据样本间的关联强度 w_{ij} 定义如式（2）所示。

$$w_{ij}=\begin{cases}\dfrac{\max\limits_{v}\left\{K-0.5\left(r_{vi}+r_{vj}\right)\right\}}{\#\left(\left\{s\middle|S_i\cup S_j\right\}\right)}, & \text{第}i\text{和}j\text{个数据样本有关联关系}\\ 0, & \text{其他}\end{cases} \tag{2}$$

式中：S_i 和 S_j 分别为与第 i 和 j 个样本距离较近的 K 个样本构成的集合，每个集合中的样本按照距离从小到大排序；$\#\left(\left\{s\middle|S_i\cup S_j\right\}\right)$ 为 S_i 和 S_j 并集的元素数量；r_{vi} 和 r_{vj} 分别为第 v 个样本在 S_i 和 S_j 中的位置，$v\in\left\{s\middle|S_i\cup S_j\right\}$。

（4）话题自动描述阶段。使用紧密中心度来刻画每个话题中文本内容的重要程度，其中紧密中心度的取值越大表示对应的文本内容越具有代表性。选取紧密中心度取值排名前10的文本内容，经过分词、剔除停用词处理后，进一步选取TF-IDF值排名前5的关键词来描述相应话题。

面向科技情报服务过程中的网络舆情态势监测需求，基于开源风险监测技术，实现全球技术动态查询、科技热点洞察的应用功能。基本逻辑是：瞄准重点技术领域，依托Hadoop、Spark等大数据处理平台，借助大数据实时计算技术，构建开源网络及社交媒体数据的实时采集、清洗、存储、分析、应用技术处理框架，建设大规模开源数据资源；借助热点话题和敏感话题识别模型、BERT预训练模型、情感分析等技术，实时监测全球技术动态，识别科技热点话题和敏感话题、重点领域安全风险，帮助用户及时掌握科技领域最新网络舆情态势，强化科技情报服务过程中的内容安全保障。

（1）全球技术动态查询功能。基于Elasticsearch的分布式大规模数据存储与快速检索技术构建全球技术动态查询栏目，用户可以按照国别、领域、时间等快速查询战略与政策、未来产业、新一代信息技术等方向的进展。此外，构建基于大模型技术的全球技术动态监测助手，与用户进行实时咨询互动，以自动问答方式反馈用户问题，并支持自动生成技术报告。

（2）科技热点洞察功能。借助热点话题和敏感话题识别、情感分析等技术，提供按照主题、关键词、时间等检索科技热点话题和敏感话题的功能，并通过主题词网络图、实时脉络、热点文章和风险关联文章等版块细粒度展示每个话题的详情。

3.3 智能监控预警技术

智能监控预警技术是保障内容安全的有力手段。因此，本研究面向科技情报服务的个性化内容安全任务、生成可读性强的内容安全检测报告等实际需求，以大数据实时计算技术为支撑，提出了以内容安全巡检为核心的智能监控预警技术实现流程（见图4），实现对科技情报服务全过程所涉及的信息资源内容的全域扫描，自动生成涵盖敏感信息、不规范表述等风险的内容安全检测报告，在保证审核质量和效率的基础上极大减少人工干预，辅助做好科技情报服务全过程的内容建设及安全管理。

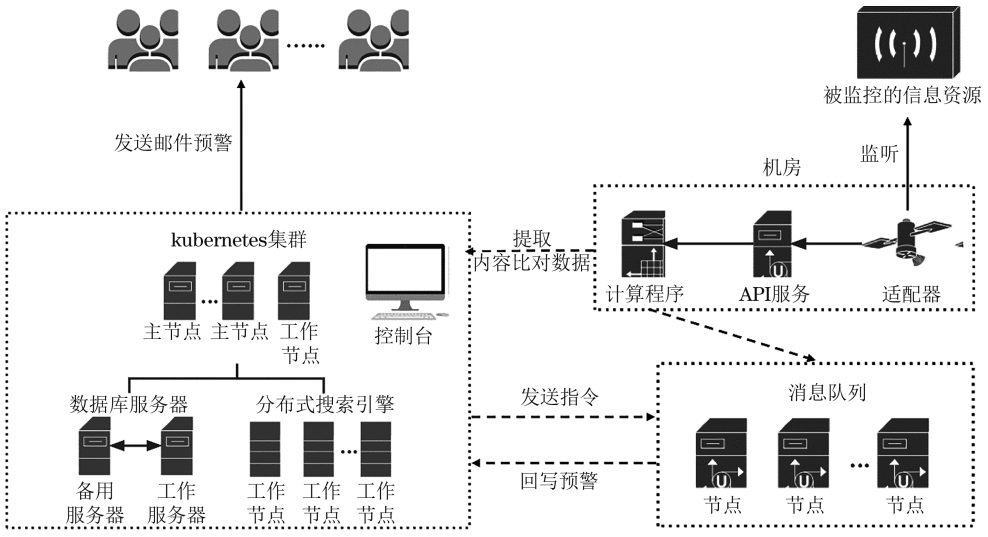

图4 智能监控预警技术实现流程

首先，在集群服务器上使用Scrapy分布式爬虫框架、Hadoop和Spark等大数据处理架构、Kafka分布式发布订阅消息系统、Redis数据库、Elasticsearch分布式搜索和分析引擎等，实现科技情报信息资源的分布式高效处理，以及实时聚合计算秒级响应[23]。

其次，内容安全巡检计算节点包括主节点和工作节点，通过HTTP和消息队列RabbitMQ进行通信。主节点是内容安全防护的控制中心，负责指定具体内容巡检任务，维护内容比对数据库，配置巡检时间、邮件接收人等具体信息。为保证监控信息资源的网络和数据安全，主节点不获取任何用户的本地化数据，仅获取巡检结果。工作节点负责具体的巡检任务，应用程序编程接口（Application Programming Interface，API）服务、计算程序、适配器等组件间通过内网通信，最大限度减少机房的资源消耗。

具体而言，内容安全巡检API服务启动时通过HTTP协议获取主节点的内容比对数据并同步给计算程序，适配器启动时订阅RabbitMQ服务，等待主节点的执行指令；适配器接收到主节点下发的扫描指令后监听目标源数据，获取数据后调用API服务，API服务调用计算程序进行内容安全审核；API服务将审核结果推送到RabbitMQ服务，由主节点的订阅服务消费后自动生成内容安全检测报告，并通过邮件方式发送给数据管理员。

面向数据管理人员对科技情报服务全过程全域信息内容的监控需求，基于智能监控预警技术，实现内容全域监控、网络暴露监测的应用功能。基本逻辑是：针对科技情报服务全过程的信息内容，使用内容安全巡检模型、网络暴露监测模型等技术，深度关联分析科技情报服务信息内容、须重点关注的本地文本文档、开源网络及社交媒体数据等从不同渠道获取的信息，实现对风险信息内容的科学研判、分级预警和分类管理。

（1）内容全域监控功能。对科技情报服务全过程的信息进行内容安全审核、对开源网络及社交媒体数据进行开源风险监测，对内容安全审核结果和开源风险监测结果进行深度关联分析，生成详细的风险预警列表，并通过条形图、折线图等方式可视化展示监控数据风险情况，并支持面向监控数据情况的查看、检索、导出等功能。

（2）网络暴露监测功能。监控对象为需重点关注的本地化文本文档，通过批量上传本地化文本文档进行主动监控，使用桑基图和网络图可视化展示网络暴露传播链条和暴露图谱，同时提供网络暴露详情列表，辅助研判网络舆情事件发展态势，帮助数据管理人员及时防范重要数据和敏感信息泄露及不当使用。

4　结语

新一轮科技革命和产业变革深入发展，高新技术领域成为国际竞争最前沿和主战场。与此同时，我国发展进入战略机遇和风险挑战并存、不确定难预料因素增多的时期，信息安全威胁不断升级。重视科技情报服务全过程的信息安全问题，对于我国抢占科技竞争和未来发展制高点意义重大。本研究围绕面向科技情报服务的内容安全技术问题，提出了以内容比对数据等基础保障为支撑，涵盖内容安全审核、开源风险监测、智能监控预警3个维度的内容安全关键技术体系，为强化面向科技情报服务的内容安全研究与管理、提升内容安全防护能力提供参考借鉴。

根据对相关研究的梳理与分析，发现面向科技情报服务的内容安全关键技术体系的发展方向呈现出以下特点：①技术体系深度方面，从单一技术向多种技术综合集成、协同应用的体系化方向拓展，以更好地应对复杂的网络环境和多样的风险形式；②技术体系范畴方面，从简单防护向事前预警、事中干预、事后纠正的系统性防护转变，以更好地应对不断延伸的内容安全区域边界；③技术体系能力方面，从被动防护向主动防护转变，基于大模型等先进人工智能技术构建的前端展示、中端预警、后端分析的内容安全防护系统或软件工具将发挥重要作用。未来，将扩大研究范围、优化升级关键技术体系及配套数据体系、建设完备的内容安全防护系统，进行深入探索和改进，进一步提升面向科技情报服务的内容安全技术水平。

参考文献

[1]　李辉，张惠娜，侯元元，等. 情报3.0时代科技情报服务能力研究：基于工程技术视角的服务能力四层结构模型[J]. 情报理论与实践，2017，40（3）：1-4.

[2]　支凤稳，马小琪，曹明帅，等. 我国科技情报服务研究主题回顾与展望[J]. 中国科技资源导刊，2023，55（5）：11-21.

[3]　赵志耘. 论复杂信息环境下的科技情报卓智赋能[J]. 情报学报，

2022，41（12）：1229-1237.

[4] 王宇航，郭涛，张潇丹，等. 互联网信息服务内容安全要求及评估框架研究[J]. 信息安全学报，2022，7（1）：27-39.

[5] 罗娇，刘细文. 知识产权视角下科学数据安全管理的策略选择[J]. 图书情报工作，2021，65（12）：38-46.

[6] 朱世强，王永恒. 基于人工智能的内容安全发展战略研究[J]. 中国工程科学，2021，23（3）：67-74.

[7] 高美玲，赵淦森. 科技情报中的信息安全问题分析与对策研究[J]. 中国科技期刊研究，2022，33（12）：1619-1627.

[8] ABBAS F H. Securing secret data using an enhanced blowfish encryption with image steganography using pixel indicator technique[D]. Dublin：National College of Ireland，2020.

[9] 苏鹏，王延飞. 对信息迷雾的情报观察：概念、形成与应对[J]. 情报理论与实践，2021，44（3）：6-12.

[10] NADIA M B，DANIEL L S. Aging in an era of fake news[J]. Current Directions in Psychological Science，2020，29（3）：316-323.

[11] 刘少芳. 网络安全视角下加拿大图书馆研究数据管理的启示[J]. 河南图书馆学刊，2023，43（11）：94-97.

[12] 吴振峰. 全球化数字化背景下开展网站系统信息内容安全防护的思考[J]. 全球科技经济瞭望，2023，38（6）：44-53.

[13] 李盼，赵文涛，刘强，等. 机器学习安全性问题及其防御技术研究综述[J]. 计算机科学与探索，2018，12（2）：171-184.

[14] TOLOSANA R，VERA-RODRIGUEZ R，FIERREZ J，et al. Deepfakes and beyond：a survey of face manipulation and fake detection[J]. Information Fusion，2020，64：131-148.

[15] ZHOU X Y，ZAFARANI R. A survey of fake news：fundamental theories，detection methods，and opportunities[J]. ACM Computing Surveys，2020，53（5）：1-40.

[16] MIANI A，HILLS T，BANGERTER A. Interconnectedness and（in）coherence as a signature of conspiracy worldviews[J]. Science Advances，2022，8（43）：eabq3668.

[17] 吴振峰，兰天，王猛猛，等. 基于共享最近邻和马尔科夫聚类的网络新闻话题检测方法[J]. 数据分析与知识发现，2022，6（10）：103-113.

[18] 黄孝章，童婷薇. 互联网第三方内容审核服务发展探析[J]. 北京印刷学院学报，2023，31（1）：50-56.

[19] 万国根. 面向内容的网络安全监控模型及其关键技术研究[D]. 成都：电子科技大学，2005.

[20] BOLT D. An inspection of the intelligence function of border force and immigration enforcement[EB/OL]. [2023-12-18]. https://assets.publishing.service.gov.uk/media/5a803e87ed915d74e33f94bd/ICIBI_inspection_intelligence_functions_Border_Force_IE_July_2016.pdf.

[21] COLIN R，NOAM S，ADAM R，et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. Journal of Machine Learning Research，2020，21：1-67.

[22] KRZANOWSKI W J，LAI Y T. A criterion for determining the number of groups in a data set using sum-of-squares clustering[J]. Biometrics，1988，44（1）：23-34.

[23] 刘智慧，张泉灵. 大数据技术研究综述[J]. 浙江大学学报（工学版），2014，48（6）：957-972.

作者简介

张昱，男，硕士，高级工程师，研究方向：管理科学与工程。
吴振峰，男，博士，助理研究员，通信作者，研究方向：自然语言处理与智能情报，E-mail：wuzf@istic.ac.cn。

Key Technology System of Content Security for Scientific and Technical Information Service

ZHANG Yu WU ZhenFeng
(Institute of Scientific and Technical Information of China, Beijing 100038, P. R. China)

Abstract: Ensuring content security is an important and basic work for scientific and technical information service. It is of great significance to carry out research on the key technology system of content security for scientific and technical information service to ensure the content security of scientific and technical information resources and provide high-quality scientific and technical information service. In view of the content security protection needs faced by the whole process of scientific and technical information service, a key technology system of content security is proposed, which is supported by basic guarantees such as content comparison data and covers three dimensions: content security review, open-source risk monitoring, and intelligent monitoring and early warning. The relevant content security prevention and control functions are designed and developed. This study can provide a reference for improving the level and efficiency of content security management in the whole process of scientific and technical information service.

Keywords: Scientific and Technical Information Service; Content Security; Technology System

（责任编辑：王玮）