

ICS 33.030
CCS M21

团 体 标 准

T/CAAAD 014-2024 T/CCSA 281-2023

互联网广告 基于生成式人工智能的 创意素材元数据管理指南

Guide to metadata management for Internet advertising AIGC creative materials

2024 - 12 - 01 发布

2025 - 01 - 01 实施

中国广告协会

中国通信标准化协会

发布

版权声明

本文件的版权归中国通信标准化协会和中国广告协会共同所有，任何单位和个人未经许可，不得进行技术文件的纸质和电子等任何形式的复制、印刷、出版、翻译、传播、发行、合订和宣贯等，也不得未经允许采用其具体内容编制中国通信标准化协会和中国广告协会以外各类标准和技术文件。如有以上需要请与版权所有方联系。

邮箱：IPR@ccsa.org.cn digitalad@china-caa.org

电话：010-62302847 010-65924878

目次

前言 III

引言 IV

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 2

5 概述 2

5.1 生成式人工智能广告创意素材元数据 2

5.2 素材元数据管理 2

6 元数据生成管理 2

7 元数据存储管理 3

7.1 元数据存储方式 3

7.2 需考虑的因素 3

8 元数据使用管理 3

8.1 访问途径 3

8.2 使用场景 3

8.3 需考虑的因素 4

9 元数据存储管理 4

10 安全性管理 5

10.1 安全原则 5

10.2 针对性安全管理措施 5

附录 A（资料性）元数据信息字段 6

附录 B（资料性）元数据存储方式 8

附录 C（资料性）生成式人工智能创意素材元数据管理系统方案 10

附录 D（资料性）星际文件系统（IPFS） 14

参考文献 16

前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国广告协会和中国通信标准化协会共同提出并分别归口。

本文件起草单位：中国广告协会、中国信息通信研究院、利欧集团数字科技有限公司、国家广告研究院、北京快手科技有限公司、北京风行在线技术有限公司、北京国双科技有限公司、悠易互通（北京）广告有限公司、联通在线信息科技有限公司、北京回旋加速网络科技有限公司、OPPO广东移动通信有限公司。

本文件主要起草人：姚轶珺、杨正军、霍焰、周崧弢、顾明毅、谷晨、潘冲、郑超、朱淑媛、龚涛、付艳艳。

引 言

近年来，随着数据规模、算法效率和计算能力的不断发展和提高，互联网上越来越多的内容通过生成式人工智能技术所创建，由此产生的数据量占比在快速的提升，国内互联网广告的创意素材生成也已经开始使用该技术。

根据《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》、《中华人民共和国广告法》、《互联网信息服务管理办法》、《互联网信息服务深度合成管理规定》、《互联网广告管理办法》、《互联网信息服务算法推荐管理规定》、《生成式人工智能服务管理暂行办法》等法律法规的相关规定，广告素材的生产者和使用者应当落实信息安全主体责任，建立健全算法机制机理审核、科技伦理审查、信息发布审核、数据安全、个人信息保护、反电信网络诈骗等管理制度，具有安全可控的技术保障措施。与传统的人工制作方式相比较，人工智能方式具有数量庞大，生成迅速，内容多样化等特性，这就给素材内容的管理工作带来了新的挑战。

对于数量极其庞大的生成式人工智能广告素材来说，素材所对应的元数据是所包含内容的技术摘要，通过元数据我们可以掌握素材的生产者，生产时间，生产设备，生产要素等一系列相关信息。基于这些信息我们可以采用技术手段对该类广告素材进行有效的管理，以满足法律法规的要求。

法律法规鼓励行业组织加强行业自律，建立健全行业标准、行业准则和自律管理制度。为了适应行业发展的要求，推动新技术的良性应用，在国家主管职能部门的指导下，由行业协会组织行业品牌企业、主导媒体和互联网广告公司等，对生成式人工智能创意素材元数据的管理提出了指导意见，确定了相关技术的应用发展方向。

互联网广告 基于生成式人工智能的创意素材元数据管理指南

1 范围

本文件提供规定了互联网广告场景中生成式人工智能创意素材元数据的生成、存储、使用和销毁四个生命阶段以及元数据安全管理的指导和建议。

本文件适用于在互联网广告领域,搭建生成式人工智能创意素材元数据管理模块或者平台的情形。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 25069-2022 信息安全技术 术语

GB/T 34090.1-2017 互动广告 第一部分:术语概述

GB/T 37043-2018 智慧城市 术语

GB/T 41867-2022 信息技术 人工智能 术语

3 术语和定义

GB/T 25069-2022、GB/T 34090.1-2017、GB/T 37043-2018和GB/T 41867-2022界定的以及下列术语和定义适用于本文件。

3.1

令牌 token

计算机身份认证中执行某些操作的权利对象。

3.2

生成式人工智能 generative artificial intelligence

生成式人工智能是利用复杂的算法、模型和规则,从大规模数据集中学习,以创造新的原创内容的人工智能技术。

3.3

生成式人工智能技术 generative artificial intelligence technology

生成式人工智能技术,是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术。

3.4

元数据 metadata

定义和描述其他数据的数据。

4 缩略语

下列缩略语适用于本文件。

API	应用程序编程接口	Application Programming Interface
HTTP	超文本传输协议	Hyper Text Transfer Protocol
HTTPS	超文本传输安全协议	Hypertext Transfer Protocol Secure
JSON	一种数据交换格式	JavaScript Object Notation

5 概述

5.1 生成式人工智能广告创意素材元数据

基于生成式人工智能的广告创意素材除了包含素材的内容信息外,还包含描述素材内容以及创作素材内容相关的信息数据,我们称之为元数据。元数据并不是生成式人工智能广告素材所独有,人工创作素材也包含元数据,但生成式人工智能方式包含的元数据信息量更多更广,得益于创作过程完全自动化,这类元数据可以把整个创作过程记录下来以用于复现素材内容。

- 与传统素材元数据相比,基于生成式人工智能的广告创意素材元数据具有以下特点:
- a) 包含更多的生产信息,例如与生成式人工智能相关的参数信息;
 - b) 根据记录的生产过程信息,复现素材内容;
 - c) 提示词可作为产生内容的一种标识;
 - d) 当发生知识产权法律纠纷时,作为一种自证的证据;
 - e) 集中管理的元数据信息,可以生成统计报表,供相关组织和人员使用;
 - f) 做为政府部门对素材监管的依据。

5.2 素材元数据管理

素材元数据管理覆盖元数据的整个生命周期,包括元数据生成,元数据存储,元数据使用,元数据销毁四个阶段。素材元数据生命阶段图如图 1 所示。

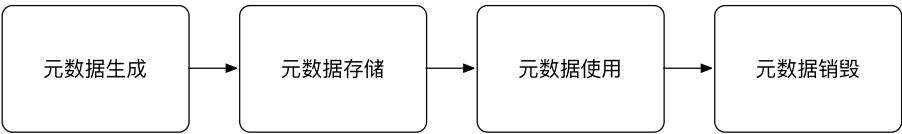


图 1 素材元数据生命阶段图

6 元数据生成管理

根据用户输入的提示信息和相关设置,在使用人工智能模型生成素材的同时记录元数据信息或者元数据索引信息。记录元数据时需要考虑以下几个方面的因素:

- a) 充分考虑元数据所需要包含的与生成式人工智能相关的要素信息,例如生产时间、生产地点、生产工具、生产人员和软硬件信息等。常用的信息字段参见附录 A。
注:常用的信息字段参见附录 A。
- b) 宜在创意素材生成后对已记录元数据进行过滤,去除不符合法律法规规定的部分;
- c) 针对生成式人工智能生成素材及元数据信息生成速度快和数量庞大的特点,充分考

虑软硬件计算资源方面的需求。

7 元数据存储管理

7.1 元数据存储方式

根据元数据信息存储位置的不同，元数据的存储方式包括：

- a) 嵌入式存储；
- b) 引用式存储。

引用式存储又可分为索引嵌入式和索引合并式两种存储方式。各种存储方式具有不同的特点，相关说明参见附录 B。

7.2 需考虑的因素

元数据的存储管理需要考虑以下几个方面的因素：

- a) 元数据存储位置；
- b) 创意素材是否包含元数据信息；
- c) 元数据存储宜采用加密方式，防止被窃取者轻易读取数据；
- d) 访问元数据信息存储介质时采取的安全控制措施；
- e) 针对人工智能生成素材数量庞大的特点，考虑存储容量的需求。

8 元数据使用管理

8.1 访问途径

一般可提供的元数据访问途径包括但不限于：

- a) 元数据直接操作
- b) API 接口访问
- c) 批量数据传输
- d) 统计数据的可视化访问

相关系统方案参见附录 C。

注：相关系统方案见附录 C

8.2 使用场景

元数据的使用场景如图 2 所示。

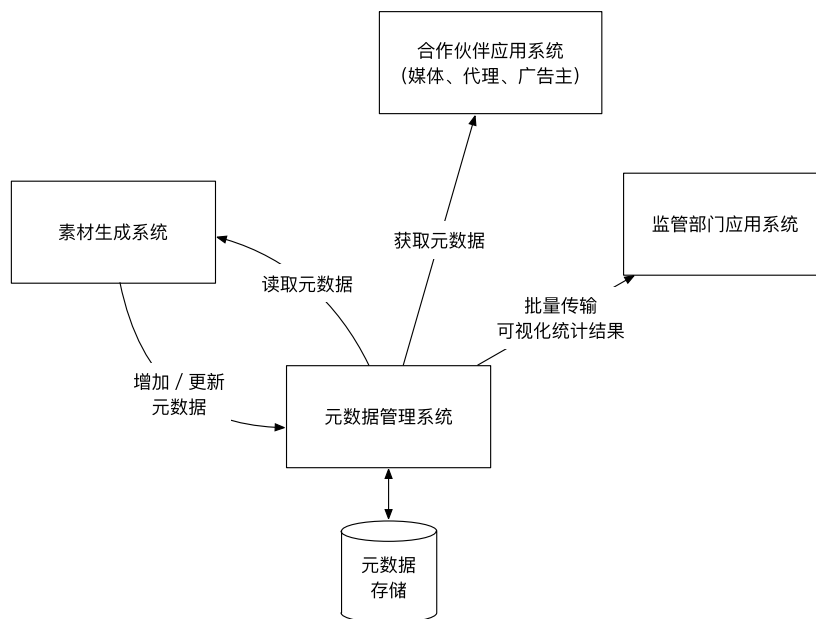


图2 元数据使用场景图

8.3 需考虑的因素

元数据的使用需要考虑以下几个方面的因素：

- 使用者提供访问接口服务，支持元数据录入，元数据查询等基础操作；
- 元数据管理系统宜具备用户和用户组管理功能，做到用户间的数据隔离；
- 根据需要，宜为元数据管理系统功能模块建立相应的用户访问控制权限；
- 元数据录入后，不再支持修改，宜仅支持数据追加功能，以保证数据不被篡改；
- 元数据访问涉及广告投放流程中的多个环节，宜根据实际情况优化访问方式，提高访问速度；
- 元数据传输宜采用加密方式，防止网络窃听；
- 操作元数据时宜同时保存操作日志；
- 发现包含违法信息的广告素材元数据时及时记录并通知其他使用者是至关重要的；
- 充分考虑元数据中涉及的个人信息处理，使之遵从相关法律法规规定。
- 为依法配合市场监督管理部门开展的互联网广告行业调查，宜通过技术手段及时提供真实、准确、完整的元数据信息。

9 元数据存储管理

对不再需要的元数据可进行安全、合规地消除或删除，销毁时需要考虑以下几个方面的因素：

- 了解并遵循相关的数据保护法规，充分考虑元数据的保存时间，使之符合使用需求；
- 建立和完善元数据的内部销毁政策；
- 对元数据销毁过程进行记录和审计，确保可追溯性。

10 安全管理

10.1 安全原则

生成式人工智能广告素材元数据在被生成、存储、使用、传输、提供、公开等处理过程中，需采取必要措施，确保数据处于有效保护和合法利用的状态，以及具备保障持续安全状态的能力。总体上宜遵循 T/CAAAD 001-2021 T/CCSA 329-2021 中关于数据安全部分的描述。

10.2 针对性安全管理措施

10.2.1 存储方式

宜采用索引合并存储方式，可以最大限度的避免元数据被直接篡改（包括删除，修改，盗用）等问题。

10.2.2 风险防范

宜对元数据存储数据做冗余和灾备，防止单点事故，减少数据丢失的风险，并能提高访问性能。

10.2.3 稳定性

针对元数据使用的场景，宜建立开放的、分布式的、可参与的去中心化网络系统，降低中心服务带来的数据孤岛问题，通过分散风险和权力，提高网络整体的健壮性和安全性。相关技术参见附录 D。

10.2.4 安全培训

宜通过培训、教育及政策制定等方式，增强相关人员对数据保护尤其是关于元数据管理和废弃过程中的安全意识和技能。

附 录 A
(资料性)
元数据信息字段

A.1 素材创作信息

与广告素材生成相关的创作相关信息：

- a) 作者（工具使用者）姓名；
- b) 作者（工具使用者）ID；
- c) 和作者（工具使用者）相关的其它信息；
- d) 创作组织；
- e) 创作平台；
- f) 创作完成时间。

A.2 素材创作过程信息

使用AIGC工具时输入的参数信息：

- a) 文生图类：
 - 提示词；
 - 反向提示词；
 - 素材尺寸；
 - 采样方法；
 - 迭代步数；
 - CFG Scale；
 - 种子值。
- b) 图生图类：
 - 参考图；
 - 提示词；
 - 反向提示词；
 - 素材尺；
 - 缩放模式；
 - 采样方法；
 - 迭代步数；
 - CFG Scale；
 - 种子值。
- c) 图片后期处理类：
 - 参考图；
 - GFPAGAN visibility（图像清晰度优化）；
 - 缩放参数；
 - 去除背景方法；
 - 创建蒙板。
- d) 视频类：
 - 采样方法；
 - 迭代步数；
 - 图像尺寸；
 - 种子值；
 - 最大帧数；
 - 动画模式；
 - 提示词（含帧数信息）。

A.3 硬件设备信息

生成素材时所使用的硬件设备信息：

- a) 显卡信息；
- b) CPU 型号；
- c) 内存大小；
- d) 硬盘容量。

A.4 软件信息

生成素材时所使用的软件信息：

- a) 操作系统；
- b) 依赖软件库；
- c) 创作软件；
- d) 生成模型；
- e) 插件工具；
- f) 备案信息：
 - 备案编号。

A.5 其他信息

和互联网广告相关的其它信息：

- a) 内容分类；
- b) 广告发布自然人信息；
- c) 法人信息；
- d) 组织信息；
- e) 广告主信息；
- f) 广告主委托人信息；
- g) 素材版本信息。

附录 B
(资料性)
元数据存储方式

嵌入式存储方式

该存储方式将元数据信息存储在素材文件中，存储时与内容数据分开，存储在素材属性信息中，例如EXIF信息，PNG扩展信息等，可通过读取素材文件直接读取或修改元数据，见图B. 1



图 A. 1 元数据嵌入式存储方式

引用式存储方式

该存储方式将元数据信息与某个唯一索引绑定，然后将索引信息存储在素材文件中，而将元数据信息存储在其他位置，例如文件、数据库、云服务器上。在这种存储方式下，索引的保存也有两种方式，一种是保存在属性信息中（索引嵌入式），见图B. 2；另一种是将索引信息和内容数据信息合并存储在一起（索引合并式），见图B. 3。

- 将特定信息与内容数据合并存储在一起，需具备以下限制条件：
- 特定信息长度较短；
 - 根据算法，特定信息合并存储后不会对内容数据产生明显的影响，内容数据仍可正常使用；
 - 可以根据算法在内容数据中找到特定信息。

直接将元数据和内容数据合并存储在一起不满足上述条件。而由于元数据索引信息长度较短，最多只有几十个字节，并且加入到内容数据中后，素材文件还可以照常使用，也能方便的寻找到该信息，所以可以采用合并的方式存储。



图 A. 2 元数据引用式存储方式，索引存储在属性信息中

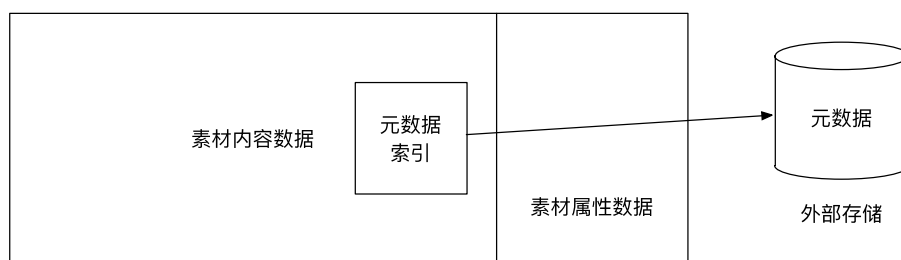


图 A.3 元数据引用式存储方式，索引存储在内容数据中

盲水印式的元数据存储技术

基于盲水印的元数据存储是一种索引合并存储方式，主要是指将元数据索引做为水印信息，通过特定的合并算法，将其和素材内容信息存储在一起的存储方式。该类存储算法具有以下特性：

- 隐蔽性，即水印对人眼不可见
- 健壮性，即使经过各种人为修改如裁剪、压缩、旋转等操作依然可以提取水印信息
- 不易移除性，即使盲水印被发现，也不易被移除或破坏。

多种元数据存储方式的对比

不同存储方式的对比参见表B.1

表 A.1 元数据不同存储方式的比较

对比项	嵌入式	引用式	
		索引嵌入式	索引合并式（盲水印）
信息存储位置	文件属性数据区域	外部	外部
素材包含信息	元数据数据	元数据索引	元数据索引
查询便捷程度	方便，可直接查询	不方便，到外部查询	不方便，到外部查询
安全性	低，内容可直接读取	中等，内容不可直接读取	高，内容需有特定算法读取
防篡改	可篡改内容	可篡改为其它错误索引	不能

附 录 C
(资料性)
生成式人工智能创意素材元数据管理系统方案

系统介绍

方案架构

典型方案架构见图C. 1

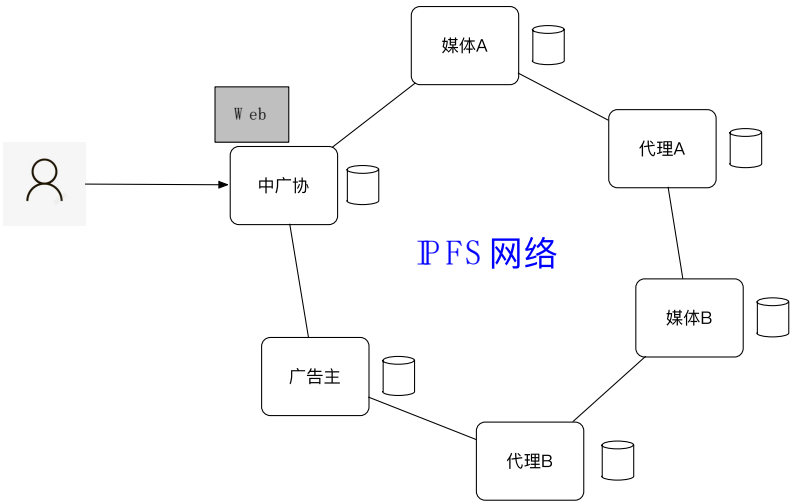


图 B. 1 元数据管理系统架构图

架构设计需求

- 元数据管理系统架构设计的需求包括：
- 采用 IPFS 网络架构，满足互联网广告流通环节中各方都可以参与的目标；
 - 不同组织的元数据信息可就近进行分布式存储；
 - 元数据信息存入相应结点后不可修改，可根据 HASH 值直接查询和读取；
 - 在被授权的情况下，进行元数据信息查询。

设计说明

系统架构基于IPFS协议，采用去中心化方式，各个结点可方便的加入或者退出，方便不同组织参与以及共享元数据信息；

IPFS协议使用加密HASH来确保文件内容（元数据信息）的真实性和完整性，使得恶意篡改和删除变得困难；

IPFS协议采用内容寻址方式来访问文件（元数据信息），使得访问更便捷，同时减少了链接失效问题；

由于IPFS协议是开源的、社区维护的和模块化的，用户可以根据自己喜欢的技术、需求和价值定制IPFS，以达到元数据管理的特殊需求；

功能说明

功能层次见图C. 2

用户（组）管理	权限管理	元数据录入	元数据查询	元数据统计	API接口	
IPFS实现		IPFS访问API	数据库		Web服务	
操作系统						
网络			硬件服务器			

图 B.2 软件功能层次

元数据管理系统功能包含：

- 以 IPFS 实现为元数据信息提供基本存储功能；
- 提供一个基于 Web 服务的管理系统，根据需要部署在 IPFS 的某个或者多个结点上；
- 该管理系统提供基础的用户管理，权限管理，元数据操作以及 API 访问功能等模块；
- 数据库提供除元数据信息以外的其它信息的存储、查询功能；
- 管理系统可根据需求添加其它功能模块；

使用场景

元数据管理系统可应用场景包括：

- a) 多方参与场景。该场景适用于参与互联网广告投放流程中的多个组织，涉及广告主，广告代理商，媒体等。在这个场景中，每个组织是其中的一个或多个存储结点，只要有一个结点（组织）存储了某个广告素材元数据，它就可以被其他结点（组织）访问；
- b) 单独使用场景。该场景满足单一组织独立管理元数据的场景，例如，媒体独立搭建系统，所有创意素材由媒体自行创作，素材元数据由媒体独立使用，媒体可在多个物理地点或者IDC建立存储结点。

API 访问接口

获取访问Token

获取访问Token的接口信息如下：

- a) 接口协议：HTTPS；
- b) 数据格式：JSON；
- c) 请求方式：POST；
- d) 功能说明：根据输入的用户名密码，获取访问 TOKEN；
- e) 输入参数：见表 C.1；

表 B.1 获取访问 Token

参数名称	类型	请求类型	是否必填	说明
login_name	string	body	是	访问用户名
password	string	body	是	访问密码

- f) 输出内容：见示例。

示例：

```
{
  "code": 0,
  "message": "OK",
  "token": "xxxxxxxxxxxxxxxxxxxxxx",
}
```

```
“time”: “2023-09-01 18:00:00”,
}
```

读取元数据

- 读取元数据的接口信息如下：
- a) 接口协议：HTTPS；
 - b) 数据格式：JSON；
 - c) 请求方式：POST；
 - d) 功能说明：根据输入的元数据 HASH，获取对应的元数据信息；
 - e) 输入参数：见表 C. 2；

表 B. 2 读取元数据

参数名称	类型	请求类型	是否必填	说明
token	string	header	是	访问token
meta_hash	string	body	是	元数据hash值

- f) 输出内容：见示例。

示例：

```
{
“code”: 0,
“message”: “OK”,
“meta_info”: {...},
“time”: “2023-09-01 18:00:00”,
}
```

录入元数据

- 录入元数据的接口信息如下：
- a) 接口协议：HTTPS；
 - b) 数据格式：JSON；
 - c) 请求方式：POST；
 - d) 功能说明：保存输入的元数据，返回对应元数据 HASH 值；
 - e) 输入参数：见表 C. 3；

表 B. 3 录入元数据

参数名称	类型	请求类型	是否必填	说明
token	string	header	是	访问token
meta_info	json	body	是	元数据

- f) 输出内容：见示例。

示例：

```
{
“code”: 0,
“message”: “OK”,
“meta_hash”: “xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx”,
“time”: “2023-09-01 18:00:00”,
}
```

追加元数据信息

- 追加元数据的接口信息如下：
- a) 接口协议：HTTPS；
 - b) 数据格式：JSON；
 - c) 请求方式：POST；

- d) 功能说明：根据输入的元数据 HASH 和新的内容，添加后生成新的元数据信息，加入到该元数据链上，原有元数据信息仅作为数据保存，不再作为查询结果，同时保存修改记录；
- e) 输入参数：见表 C. 4；

表 B. 4 追加元数据信息

参数名称	类型	请求类型	是否必填	说明
token	string	header	是	访问token
meta_fields	json	body	是	新的内容
meta_hash	string	body	是	元数据hash值

- f) 输出内容：见示例。

示例：

```
{
  "code": 0,
  "message": "OK",
  "time": "2023-09-01 18:00:00",
}
```

附 录 D
(资料性)
星际文件系统 (IPFS)

概述

星际文件系统 (InterPlanetary File System, 简称IPFS) 是一套用于组织和传输数据的模块化协议, 他是根据内容寻址和点对点网络的原则重新进行设计的。由于IPFS是开源的, 所以IPFS有多种实现。虽然IPFS有多个使用案例, 但其主要案例是以去中心化的方式发布数据(文件、目录、网站等)。

系统组成

一个遵循IPFS协议实现的系统主要包含以下子系统, 见表D. 1。

表 C. 1 IPFS 实现的子系统列表

作用	子系统
表示和组织数据	CIDs, IPLD, UnixFS, MFS, DAG-CBOR, DAG-JSON, CAR files
内容路由, CID和IP地址间的关联	Kademlia DHT, Delegated routing over HTTP, Bitswap, mDNS
传输数据	Bitswap, HTTP Gateways, Sneakernet, Graphsync
数据和结点寻址	Multiformats
IPFS和HTTPS的桥梁	IPFS Gateways, Pinning API Spec
点对点连接	libp2p (TCP, QUIC, WebRTC, WebTransport)
可变性和动态命名	IPNS (Interplanetary Naming System), DNSLink

工作机制

数据表示和寻址

内容标识符(CID)

在 IPFS 中, 数据被分成多个块, 每个块会被分配一个称为内容标识符(CID)的唯一标识符。通常, CID 是通过将数据的哈希值与其编解码器相结合计算得到的。

星际关联数据 (IPLD)

IPFS 使用星际关联数据(IPLD)来处理 cid 和内容寻址数据。IPFS 使用 IPLD 表示内容寻址数据(如文件目录和其他层次结构)之间的关系, 采用了称为 Merkle DAG 的有向无环图(DAG)。使用 IPLD 实现常用功能, IPFS 能够提供一种更定制化的、特定的机制来表示和寻址文件、目录及其符号链接, 称为 UnixFS。

内容可寻址存档 (CAR) 文件

IPFS使用内容可寻址归档 (CAR) 文件来存储和传输IPLD内容寻址数据的序列化归档。CAR文件类似于TAR文件, 因为它们是为存储内容寻址数据的集合而设计的。

内容路由

内容路由指的是IPFS在网络中到何处去找到给定CID的方式; 具体来说, 就是哪些网络对等点提供了您所请求的CID。换句话说, 节点不能仅仅通过CID在网络中查找数据; 它需要网络上对等体的IP地址和端口信息。为了实现内容路由, IPFS使用了以下子系统: 分布式哈希表、Bitswap、HTTP上的委托路由和mDNS。

数据传输

除了路由数据之外，IPFS网络中的结点必须有效地分发和交付内容寻址数据，同时考虑到网络中有些结点已经拥有数据副本，而其他结点没有数据副本，却需要一个。为了处理数据的传输，IPFS使用以下子系统：Bitswap、IPFS HTTP网关和人工传递。

参 考 文 献

- [1] 中华人民共和国数据安全法
 - [2] 中华人民共和国个人信息保护法
 - [3] 中华人民共和国网络安全法
 - [4] 中华人民共和国广告法
 - [5] 互联网信息服务管理办法
 - [6] 互联网信息服务深度合成管理规定
 - [7] 互联网广告管理办法
 - [8] 互联网信息服务算法推荐管理规定
 - [9] 生成式人工智能服务管理暂行办法
 - [10] T/CAAAD 001-2021 T/CCSA 329-2021 互联网广告数据应用和安全技术要求
-