**1. 项目概述与目标回顾**

**2. 前期调研与数据源选择**

**3. 爬虫程序与数据收集**

**4. 威胁情报平台的构建**

**4.1 系统架构概述**

威胁情报平台的后端设计围绕数据的高效管理和查询进行了详细规划。平台支持对结构化数据（如IOC指标）和非结构化数据（如安全文章）的统一管理。后端接口设计涵盖了数据总量统计、增量数据更新、数据分类查询等核心功能，确保系统能够高效地处理大量数据并为前端展示提供强大的支持。

**4.2 后端接口设计**

为了确保平台的数据交互与管理的高效性，后端接口的设计遵循了规范化、模块化和高可扩展性的原则。以下是主要功能模块的接口设计：

**4.2.1 非结构化数据模块**

非结构化数据模块主要负责安全文章及其相关数据的管理和查询。该模块通过提供相关的接口，帮助用户查询特定日期范围内的非结构化数据总量和新增数据量。

1. **接口名称**：/api/unstructured-data/total

* **功能描述**：查询指定日期范围内的非结构化数据总量。
* **请求方法**：GET
* **请求参数**：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **参数名** | **类型** | **必填** | **默认值** | **格式** |
| start\_date | String | 否 | 当天 | |  | | --- | |  |  |  | | --- | | YYYY-MM-DD | |
| end\_date | String | 否 | 当天 | |  | | --- | |  |  |  | | --- | | YYYY-MM-DD | |

* **响应示例**：

|  |
| --- |
| {  "code": 200,  "message": "success",  "data": {  "total": 1500,  "list": [  {  "id": 7,  "sourceName": "360威胁情报中心",  "securityArticleLink": "https://example.com/12345",  "securityArticleLinkHash": "md5-abc123",  "securityArticleObsPath": "/security/articles/12345",  "createdTime": "2023-04-11T07:01:28.000+00:00"  }  ]  }  } |

1. **接口名称**：/ api/unstructured-data/fetched-total

* **功能描述**：查询指定日期范围内的新增非结构化数据量。
* **请求方法**：GET
* **请求参数**：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **参数名** | **类型** | **必填** | **默认值** | **格式** |
| start\_date | String | 是 | 当天 | |  | | --- | |  |  |  | | --- | | YYYY-MM-DD | |
| end\_date | String | 是 | 当天 | |  | | --- | |  |  |  | | --- | | YYYY-MM-DD | |
| {  "code": 200,  "message": "success",  "data": {  "fetched\_total": 200,  "list": [  {  "id": 7,  "sourceName": "360威胁情报中心",  "securityArticleLink": "https://example.com/12345",  "securityArticleLinkHash": "md5-abc123",  "securityArticleObsPath": "/security/articles/12345",  "createdTime": "2023-04-11T07:01:28.000+00:00"  }  ]  }  } | | | | |

**4.2.2 结构化数据模块**

结构化数据模块主要处理与威胁情报相关的结构化信息（如域名、IP、URL等IOC指标）。该模块提供了数据总量统计和类型分类统计接口，方便用户根据不同的威胁类型进行查询。

**1 .接口名称**：/api/structured-data/total

* **功能描述**：查询指定日期范围内的非结构化数据总量。
* **请求方法**：GET
* **请求参数**：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **参数名** | **类型** | **必填** | **默认值** | **格式** |
| start\_date | String | 是 | 当天 | |  | | --- | |  |  |  | | --- | | YYYY-MM-DD | |
| end\_date | String | 是 | 当天 | |  | | --- | |  |  |  | | --- | | YYYY-MM-DD | |

* **响应示例**：

|  |
| --- |
| {  "code": 200,  "message": "success",  "data": {  "structured\_total": 5000,  "list": [  {  "iocType": "DOMAIN",  "value": "malicious-domain.com",  "createdTime": "2023-04-11T08:00:00.000+00:00",  "malwareHash": "sha256-xyz789",  "malwareFamily": "APT29",  "iocComment": "钓鱼域名"  }  ]  }  } |

**2 .接口名称**：/api/structured-data/category

* **功能描述**：按威胁类型（DOMAIN/IP/URL）分类统计结构化数据。
* **请求方法**：GET
* **请求参数**：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **参数名** | **类型** | **必填** | **默认值** | **格式** |
| start\_date | String | 否 | 当天 | |  | | --- | |  |  |  | | --- | | YYYY-MM-DD | |
| end\_date | String | 否 | 当天 | |  | | --- | |  |  |  | | --- | | YYYY-MM-DD | |

* **响应示例**：

|  |
| --- |
| {  "code": 200,  "message": "success",  "data": {  "domain": {  "total": 3000,  "list": [  {  "iocType": "DOMAIN",  "value": "malicious-domain.com",  "createdTime": "2023-04-11T08:00:00.000+00:00",  "iocComment": "钓鱼域名"  }  ]  },  "ip": {  "total": 1500,  "list": [  {  "iocType": "IP",  "value": "192.168.1.1",  "createdTime": "2023-04-11T08:00:00.000+00:00",  "iocComment": "恶意IP"  }  ]  },  "url": {  "total": 500,  "list": [  {  "iocType": "URL",  "value": "http://malicious-url.com",  "createdTime": "2023-04-11T08:00:00.000+00:00",  "iocComment": "钓鱼链接"  }  ]  }  }  } |

**4.3 关键设计说明**

为了确保系统的高效性，接口设计遵循了以下原则：

* **分页查询设计**：所有分页查询接口均固定返回每页15条数据，前端仅需提供当前页码，无需指定页面大小。分页计算基于LIMIT 15 OFFSET (page-1)\*15实现。
* **性能优化**：针对高频查询字段（如created\_time、iocType）建立了复合索引，以提高查询效率。

**4.4 数据模型定义**

系统中使用了两种主要的数据实体：

1. **非结构化数据实体（ThreatIntelligencePO）**：该实体主要用于存储与安全文章相关的元数据，如来源名称、文章链接、创建时间等。
2. **结构化数据实体（ThreatIntelligence）**：该实体用于存储与威胁情报相关的IOC数据，包括威胁类型、指标值、恶意软件哈希、威胁描述等信息。

**5. 进展与成果**

**5.1 后端接口实现成果**

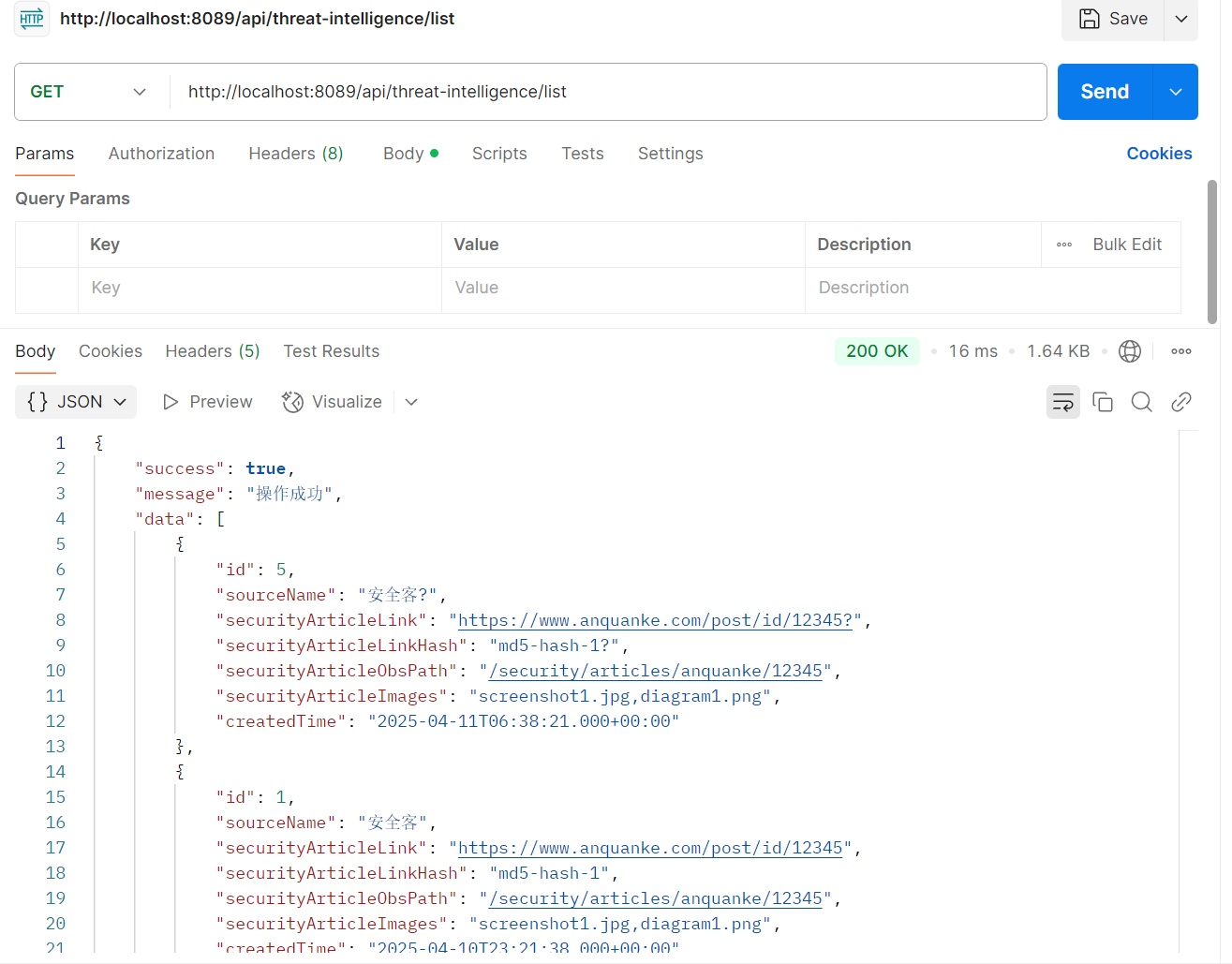
本阶段已完成了威胁情报平台后端的主要接口开发，涵盖非结构化数据和结构化数据的统计与分类功能。接口均已通过测试，能够正常返回预期结果。

**非结构化数据接口：**

* /api/unstructured-data/total：支持查询指定日期范围内的非结构化数据总量。
* /api/unstructured-data/fetched-total：支持查询每日新增数据量。

**结构化数据接口：**

* /api/structured-data/total：实现结构化数据总量统计。
* /api/structured-data/category：支持根据 IOC 类型（DOMAIN、IP、URL）分类统计数据。



后端接口实现

**5.2 数据存储与查询成果**

* 系统已建立 **MySQL 数据存储表**，支持爬虫采集数据的持久化保存。
* 已通过测试，能够高效完成按日期、按威胁类型的查询与过滤。
* 数据表结构覆盖了 **非结构化数据（文章来源、链接、发布时间）** 与 **结构化数据（IOC 类型、指标值、恶意软件哈希等）**。



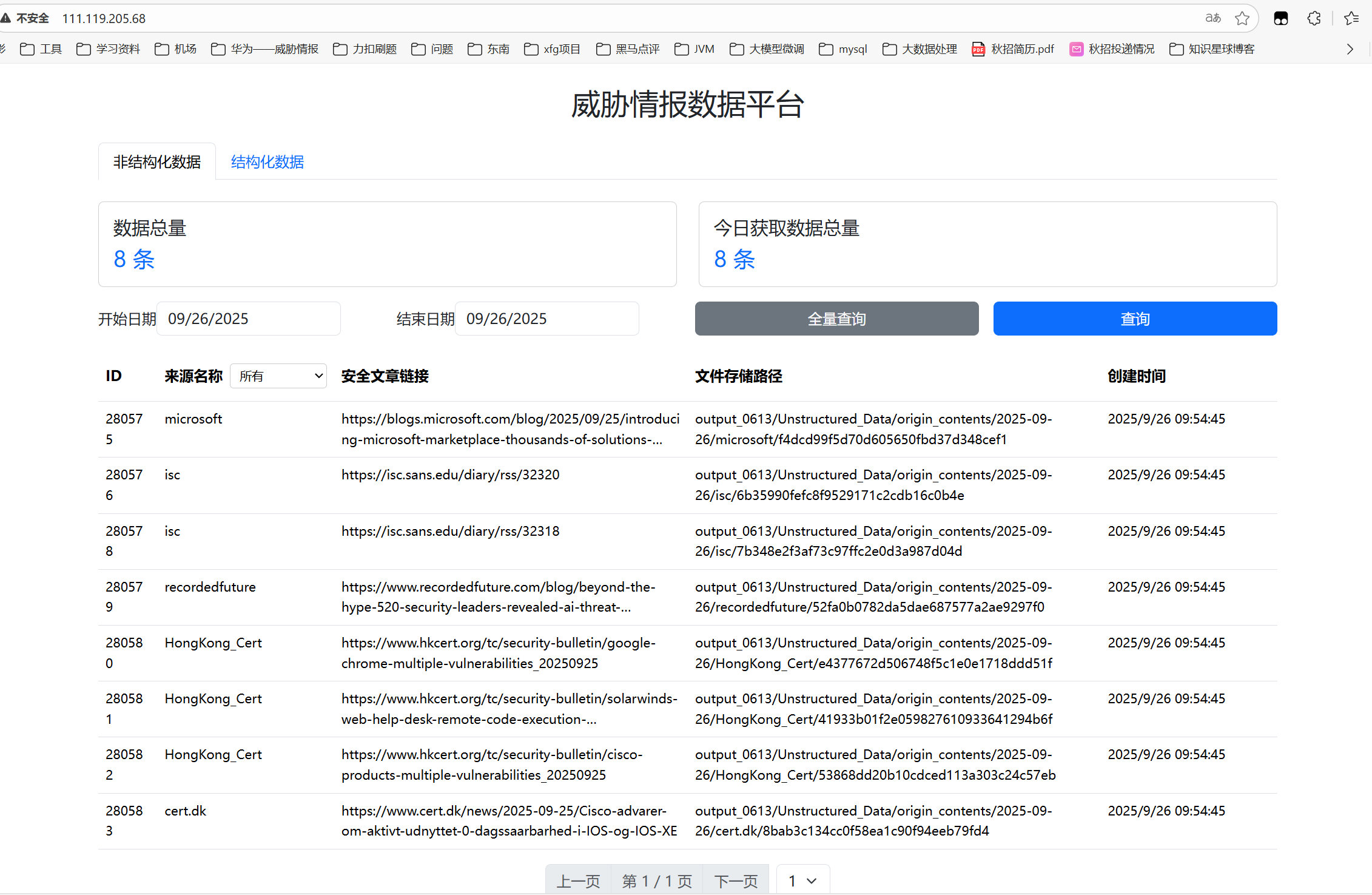
数据库存储

**5.3 前端展示功能**

平台前端已实现核心功能模块，能够直观展示威胁情报数据，主要包括：

1. **非结构化数据管理界面**

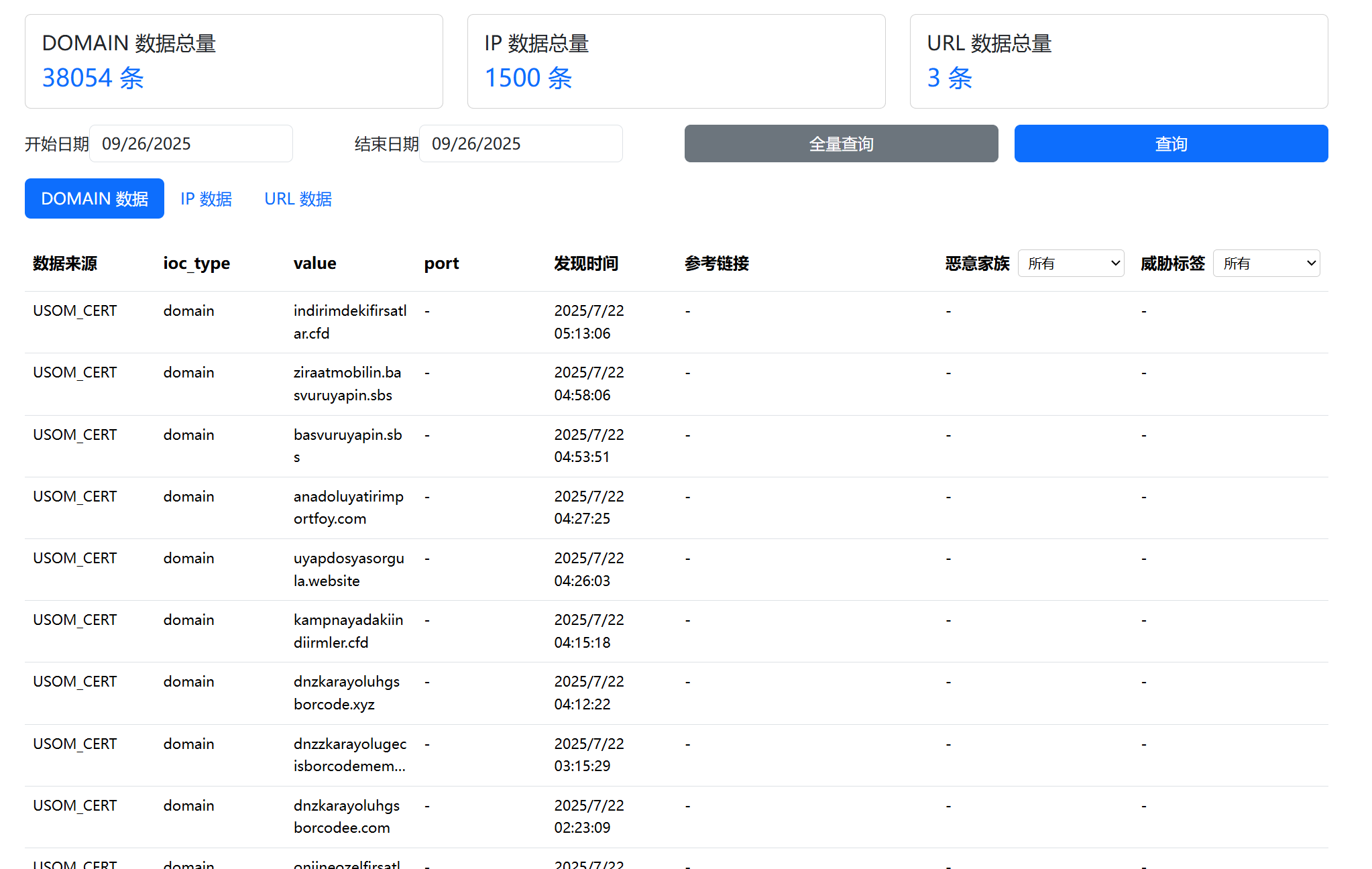
* 展示非结构化威胁情报数据总量和每日新增情况。
* 支持日期筛选，动态更新统计结果。



威胁情报每日新增展示

1. **结构化数据管理界面**

* 按 DOMAIN / IP / URL 类型展示 IOC 数量。
* 支持全量查询/按照日期查询，展示每日新增数据



结构化数据分类展示

**5.4 系统集成成果**

目前，系统已基本实现从 **数据爬取 → 数据清洗 → 数据存储 → 后端接口 → 前端展示** 的完整闭环。

* 爬虫采集的数据能够自动导入数据库。
* 后端接口进行统计与分类，并提供给前端调用。
* 前端页面能够完整展示统计结果与文章信息，形成了初步可用的威胁情报展示平台。