

博物馆网站数据采集子系统

需求分析报告



2020-4-1

目录

[1.简介 1](#_Toc37248017)

[1.1目的 1](#_Toc37248018)

[1.2范围 1](#_Toc37248019)

[1.3阅读人群 1](#_Toc37248020)

[2. 总体描述 1](#_Toc37248021)

[2.1项目来源和背景 1](#_Toc37248022)

[2.2产品前景 1](#_Toc37248023)

[2.3产品功能 2](#_Toc37248024)

[2.4用户特征 2](#_Toc37248025)

[2.5一般性限制 2](#_Toc37248026)

[2.6假设和依赖 3](#_Toc37248027)

[3.详细需求 3](#_Toc37248028)

[3.1 功能需求 3](#_Toc37248029)

[3.2 外部接口需求 4](#_Toc37248030)

[3.3 性能需求 4](#_Toc37248031)

[3.4 质量需求 5](#_Toc37248032)

[3.5 系统运行网络环境 5](#_Toc37248033)

[4.实现约束 5](#_Toc37248034)

[4.1系统的实现约束 5](#_Toc37248035)

1. 简介

1.1目的

本文档描述了博物馆应用平台系统中，博物馆网站数据采集子系统的相关功能和必要需求。

1.2范围

博物馆网站数据采集子系统是一个能够快速及时的捕获所需博物馆网站的信息的系统，包括博物馆的基本介绍、参观信息（开放时间等）、展览信息、教育活动、经典藏品信息，学术研究信息等，信息定时更新。对于爬取信息进行加工和过滤，抽取需要内容，如展览主题，展览时间、介绍等。

1.3阅读人群

阅读对象是本项目的其他组的开发、测试与维护人员。

2.总体描述

2.1项目来源和背景

本项目是博物馆网站信息采集系统，由于博物馆网站众多，本产品致力于智能高效的搜集网站信息，删除不必要的信息，并保留关键信息，减少查询博物馆的工作量。

其他小组成员通过编辑筛选条件可以找出所需博物馆的信息，并且排除多余的垃圾信息，实现查找、阅读的高效与便捷。2.1产品前景

2.2产品前景

本产品用于快速采集博物馆网站信息，能够筛选有用的消息如开馆时间、展品信息、展览信息、教育活动信息、学术研究信息等，产品有筛选准确快速的优势、导入数据库查询定位快速准确、持续更新信息。通过产品说明介绍，能快速掌握使用方法，使用简单方便，仅需要掌握电脑基本操作。

2.3产品功能

1）数据爬取：爬取全国一级博物馆（130家左右）的网站信息，包括博物馆基本的介绍、参观信息（开放时间等）、展览信息、教育活动、经典藏品信息、学术研究信息等，对于展览信息可以定时更新。

2）数据加工：对于爬取的信息进行过滤和加工，抽取需要的内容。例如：对于展览页面，要得到展览主题、展览时间、展览地点、展览介绍等信息。

3）数据导入：采用合适的方式保存抽取的数据，能够导入到数据库中。

4）数据更新：支持数据的持续更新。例如：根据情况，每天或每周爬取一次新的数据，更新原有数据。

2.4用户特征

|  |  |
| --- | --- |
| 用户类型 | 说明 |
| 其他小组开发人员 | 用爬取来的信息进行其他功能构建的小组成员 |

2.5一般性限制

***环境***：运行系统应在windows 10及其以上本版

***设计和实现约束***：开发文档遵循IEEE开发标准

***开发语言***：Python

2.6假设和依赖

无

3.详细需求

3.1 功能需求

* 博物馆网站数据采集子系统的整体功能用例图见图1。
* 用例的简要描述见表1、表2。

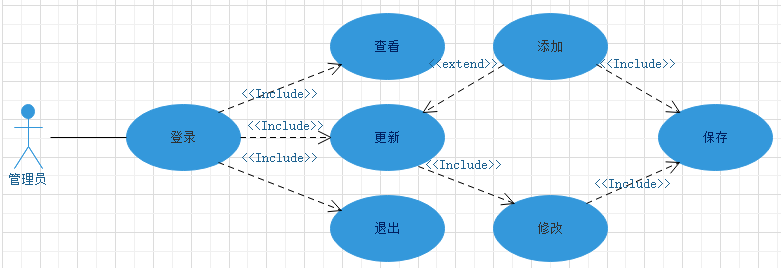


图1.整体功能用例图

表1.Actor一览表

|  |  |
| --- | --- |
| Actor | 可选操作 |
| 管理员 | 登录系统，查看信息表，更新信息表，更新信息表，保存信息，退出系统。 |

表2.用例一览表

|  |  |
| --- | --- |
| 用例中文名称 | 简要用例描述 |
| 登录 | 管理员输入账号密码登录系统。 |
| 查看 | 管理员选择查看已有的信息表。 |
| 更新 | 管理员选择更新信息表。 |
| 退出 | 管理员选择退出系统。 |
| 添加 | 当信息表为空时，管理员向表中添加数据。 |
| 修改 | 管理员用新加工后的数据替换信息表中的数据。 |
| 保存 | 管理员保存信息表。 |

3.2 外部接口需求

* 硬件接口

无

* 软件接口

无

* 通信接口

无

3.3 性能需求

* 系统对管理员请求的最大响应时间是10秒钟，在此时间内将响应结果显示在屏幕上；
* 系统的加载时间不大于5秒钟；
* 系统支持的客户端人数为10人。

3.4 质量需求

* 可用性：系统可以使用，每天更新1次；
* 可扩展性：系统可增加新的功能，所需时间不超过2天；
* 安全性：系统不会影响计算机其他软件的使用；
* 可靠性：系统运行过程中，不会发生内存泄漏和进程死锁，系统无故障运行时间达到500小时以上，如果在运行过程中发生错误，系统一般不会回到发生错误前的状态；
* 可维护性：从系统运行过程中查找、修复错误的时间预期不超过1天；
* 易操作性：对所有博物馆的信息页面布局采用的模板不超过3个，管理员能轻松地对数据进行更新。

3.5 系统运行网络环境

本系统运用于小组成员内部的数据交流，通过网络进行数据传输，在建立的数据库中进行筛选，通过专用接口传输到app中，展示出来。

4.实现约束

4.1系统的实现约束：

1. 操作系统：Windows10、Linux
2. 开发平台：Python
3. 数据库：Sql server、My Sql