

# 海外藏中国文物知识管理与服务平台

## 需求规格说明

### 计科 2102

#### 1.1 概述

本文旨在说明海外藏中国文物知识管理与服务平台的功能需求和非功能需求。系统一共分为 5 个子系统: 知识图谱构建子系统、知识服务子系统 (Web 端)、掌上博物馆 (手机端)、问答子系统、后台管理子系统。

#### 2.1 子系统功能需求

##### 知识图谱构建子系统:

1) 数据爬取: 爬取海外博物馆网站的中国文物信息 (至少爬取指定的 5 家博物馆的所有中国文物, 博物馆网址列表附后), 包括文物名字、文物图片、年代、介绍等信息。按照规定格式保存下载的数据。具体数据格式附后。

2) 数据建模: 将爬取的数据转化为三元组形式。

3) 数据补充: 根据需要从互联网爬取数据, 对现有数据进行补充, 如相关数据的基础信息缺失, 需要从其他来源进行爬取或人工补充。例如, 从百度百科下载数据, 补充现有数据, 如书画作家信息等。

4) 数据存储: 将根据步骤 2 中建模好的三元组数据保存到 Virtuoso 或 Neo4j 图数据库中, 发布成链接开放数据 (选做), 可用于关系图谱、时间轴等知识图谱可视化、问答等功能开发; 全部数据 (用户数据和文物数据等) 需要保存到数据库 (mysql 或其他数据库) 中。

##### 海外文物知识服务子系统:

1) 数据浏览: 支持多种形式展示的浏览功能。①提供基本的筛选、排序功能, 可按照文物类型、文物年代、博物馆等多种基础信息进行索引、筛选、排序方式浏览文物信息以方便用户的使用。②提供查看文物详情功能, 显示文物的详细数据, 如文本、图像等信息, 点击文物图片, 可以进行放大缩小。③相关文物推荐功能, 在该文物页面显示相关文物, 相关规则自定, 如相似主题、同一作者、图像内容相似等。可参考克利夫兰、大英博物馆等网站。

2) 数据查询: 支持文物的简单查询功能和高级查询功能。简单查询根据输入的关键字, 如文物名称、博物馆名称、文物年代等进行查询。高级查询可以对文物的多个字段进行限定查询, 可参考克利夫兰博物馆。

3) 数据可视化显示: 将构建的文物知识图谱可视化, 以从不同角度等可视化效果, 展示文物知识。可参考的可视化方案如下: ①包含结点、边的力导向图知识图谱展示, 可参考全历史网站、历史人文大数据平台、Neo4j 图数据库等。②文物时间轴: 按照时间轴的方式、展示各个时段的文物信息、时间等信息。

4) 用户个人信息管理: 用户可以注册登录该系统, 设置用户名密码等个人信息。可扩展针对用户使用的多种功能, 如最近浏览记录、收藏、评论等功能 (选做)。

##### 知识问答子系统:

1) 简单问答: 支持单实体单属性的问答, 至少支持 10 类以上问题, 例如文物的收藏地、时代、材质、类型、介绍、书画作者等。

2) 复杂问答: 选做。

3) 闲聊: 选做。

### **掌上博物馆：**

1) 文物浏览：显示文物的基本信息、图片，播放讲解等（如果提供了讲解音视频）；可以进行按照关键字搜索的简单搜索。

2) 用户交互：用户可以对单个文物点赞、留言评论等功能、用户可以为单个文物上传自己拍摄的相关照片（例如，用户参观博物馆，拍摄了该文物的照片，然后上传）、可以上传针对某个文物的讲解（选做）、视频（选做）等。

3) 以图搜图：可以上传图片或直接拍摄一个照片，根据上传图片特征搜索相关文物。

4) 用户个人信息管理：用户可以注册登录该系统，设置用户名密码等个人信息。用户信息可以和海外文物知识服务子系统共用或单独使用一套用户系统。

5) 用户个人动态（选做）：用户可以发表动态，上传文字和图片，分享文物或博物馆等，其他用户可以点赞、评论等。（类似微博）

### **后台管理子系统：**

1) 用户管理：管理后台管理子系统用户、掌上博物馆用户、知识服务子系统用户的基础信息。如：对用户信息的增删查改功能；用户权限的管理，如登录、点赞、评论等用户功能的管理。例如，如果掌上博物馆用户发表不良评论，则停止其发表评论的权限。

2) 信息审核：审核用户发表的留言、图片、音视频等内容。对于审核不通过的内容会被屏蔽，同时对该用户进行一定程度惩罚，如禁止发评论等。自动审核功能：对于留言，可以设置敏感词，当用户提交留言出现敏感词时，进行自动屏蔽；对于图片，需要判断是否为不良图片，如果审核不通过，则不显示。人工审核功能：后台管理员或审核人员人工检查用户提交的文本、图片、音视频。

3) 数据管理：管理 1-3 中涉及的所有数据，对所有数据可以进行增删改查等操作，支持单个操作和批量操作。

4) 数据备份和恢复：支持数据库的备份和恢复。实现手动备份恢复或定时备份功能。支持查看所有的备份和恢复记录，显示记录时间等相关功能。可通过点击备份记录来将数据库恢复到该备份记录点上。

5) 日志管理：查看和检索后台管理子系统的操作日志。记录包括管理员等用户对后台数据的操作记录，数据库的备份还原等记录。

## **3.1 知识图谱构建子系统其他需求**

### **系统要达到的目标**

为其他组提供丰富准确的文物信息

采用爬虫技术爬取网页中的国内文物信息，并处理成结构化数据，使其他子系统的开发人员可以使用本组爬取到的数据进行关系图谱和时间轴等知识图谱的可视化，并呈现在 Web 端以及 APP 端

平台数据应结合其他小组的需求通过人工等方式进行数据补充

### **运行环境**

在 Windows 环境下使用 Scrapy 爬虫获取海外文物信息的知识图谱数据，使用 Virtuoso 或 Neo4j 图数据库，用知识与知识之间的联系进行知识图谱可视化

### **角色定义-开发者**

开发者使用爬虫在指定的博物馆官网爬取数据，并对爬取到的数据进行数据清洗和格式化，以固定格式返回给需要数据的小组

### **角色定义-管理者**

管理者使用爬取到的相关数据使用图数据库进行可视化操作，并生成开源链接，将可视化图片和链接展现给读者

## 数据来源

根据名单要求，本小组爬取的网站分别是：

5	香港中文大学 中国文化研究所	<a href="https://lingnancrf.ics.cuhk.edu.hk/ancient-books">https://lingnancrf.ics.cuhk.edu.hk/ancient-books</a>
6	上海博物馆	<a href="https://www.shanghaimuseum.net/mu/frontend/pg/lib1/index">https://www.shanghaimuseum.net/mu/frontend/pg/lib1/index</a>
7	俄国艾尔米塔什博物馆	<a href="https://hermitagemuseum.org/wps/portal/hermitage/explore/artworks/?lng=zh">https://hermitagemuseum.org/wps/portal/hermitage/explore/artworks/?lng=zh</a>
8	国立东方艺术博物馆（都灵）	<a href="https://www.maotorino.it/en/collections/online-catalogue/?src=&amp;portal=mao&amp;catalog_collection=131#filters-1">https://www.maotorino.it/en/collections/online-catalogue/?src=&amp;portal=mao&amp;catalog_collection=131#filters-1</a>

## 爬取数据构成

文物 ID

文物名称

文物时代

文物作者

文物类型

文物归属部门

文物材料

文物国家归属

文物介绍

文物评价

文物 web 端原始链接

文物图片地址（获取最清晰的图片）

## 功能

将爬取的数据清洗,剔除不必要的信息,得到规范的爬取信息

根据文物名称判断实体是否已经存在于数据库中,剔除已存在的实体

统计并记录规范的爬取信息中缺失的部分,(如缺失年份、作者信息),保存此数据,以便之后补充

将每个文物的相关数据处理成实体

对结构化的数据处理成实体,抽取关系,构建成三元组形式

向已经存在的知识图谱里添加新实体和关系

对整体的实体信息和实体关系数据按照图数据库存储要求处理

将整体的实体信息和实体关系数据自动保存到云数据库

将图片数据存储到云数据库、国内免费图床

反馈给数据库管理员新增实体的数量和知识图谱中实体关系数量变化

用户通过接口获取知识图谱的数据

## 响应时间需求

程序应持续运行并输出响应日志,直到筛选出所有数据

## 可靠性需求

系统应 7\*24 小时提供服务,且能正确获取相关数据

## 系统安全性需求

系统要有较强的鲁棒性,不会因混入部分格式不规范的数据等各种原因导致程序崩溃

## 3.2 海外文物知识服务子系统其他需求

## **系统要达到的目标**

能为用户提供丰富准确的文物信息

平台要有良好的运行效率

## **运行环境**

使用者通过因特网登录到本系统中，查询相关信息

## **角色定义-使用者**

使用者是指在 Web 端使用本系统查询文物相关信息，使用本系统提供的功能的人员

### **场景 1：用户浏览相关文物信息**

主要参与者：用户

情景目标：浏览相关文物

前提条件：需要登入，联网

操作：用户进入网页浏览相关文物信息

异常：文物内容无法正常显示

使用频率：每天多次

使用方式：Web

次要参与者：系统后台

### **场景 2：用户按关键字搜索文物的基本信息**

主要参与者：用户

情景目标：通过搜索功能搜索到相应的文物信息

前提条件：需要联网

操作：

用户进入网页

在主界面点击搜索框

输入有关文物的信息

等待系统响应

点击文物查看信息

异常：

系统未响应

未找到与信息相关的文物

使用频率：每天多次

使用方式：Web

次要参与者：系统后台

次要参与者参与方式：检索与关键字相关的文物并展示

### **场景 3：数据可视化**

主要参与者：用户

情景目标：以可视化方式浏览文物信息

前提条件：需要联网

操作：

用户进入网页

选择某文物以可视化方式显示其信息

异常：系统未响应

使用频率：每天多次

使用方式：Web

次要参与者：系统后台

#### **场景 4：用户修改个人信息：**

主要参与者：用户

前提条件：需要联网

操作：

用户进入网页

用户进入用户列表

用户修改想要修改的信息（邮箱、密码、用户名等）

异常：系统未响应

使用频率：每天多次

使用方式：Web

次要参与者：系统后台

#### **场景 5：用户注册**

主要参与者：用户

前提条件：需要联网

操作：

用户进入登陆界面/注册界面

用户输入邮箱、密码及用户名

异常：系统未响应

使用频率：每天多次

使用方式：Web

次要参与者：系统后台

#### **场景 6：用户登录**

主要参与者：用户

前提条件：需要联网

操作：

进入登陆界面

填写邮箱密码

点击登录进行登录

异常：

系统未响应

填写未注册的账号，登陆失败

填写错误的密码，登陆失败

使用频率：每天多次

使用方式：Web

次要参与者：系统后台

#### **响应时间需求**

操作响应时间应该在五秒以内

#### **可靠性需求**

系统应 7\*24 小时提供服务，允许多人使用并正常运行

#### **系统安全性需求**

系统要有严格的权限管理，防止用户的非法操作对系统造成破坏或导致数据泄露

### **3.3 知识问答子系统其他需求**

#### **系统要达到的目标**

为用户提供准确简洁的回答  
能够保存用户对话的历史记录

### **运行环境**

本系统基于 Django 4.1.7 版本开发，要求 python 版本 $\geq 3.8.0$

### **角色定义-使用者**

使用者是指通过本系统与后台进行对话的人员

### **角色定义-管理者**

管理者是指在管理端执行上传、更新、修正信息，以及对本系统进行维护升级等相关功能的人员

### **场景 1：用户与系统进行对话**

主要参与者：用户

情景目标：获取相关文物的信息

前提条件：需要登入，联网

操作：

用户登入系统

用户选择某样文物

进入文物详细界面

用户启用对话框

用户在对话框中输入问题

系统进行相应回答

异常：

问题包含不合法消息

服务器无响应

使用频率：每天多次

使用方式：app 端/网页端

次要参与者：系统后台

### **场景 2：用户取回历史聊天记录**

主要参与者：用户

情景目标：用户再次启用对话框时能够显示历史聊天记录

前提条件：需要联网

操作：

用户登入系统

用户打开对话框

界面显示历史聊天记录

异常：

系统未响应

无历史聊天记录

使用频率：每天多次

使用方式：app 端/网页端

次要参与者：系统后台

### **响应时间需求**

操作响应时间应该在五秒以内

### **可靠性需求**

系统应 7\*24 小时提供服务，允许多人使用并正常运行

### **系统安全性需求**

系统要有严格的权限管理，防止用户的非法操作对系统造成破坏或导致数据泄露

### **3.4 掌上博物馆其他需求**

#### **系统要达到的目标**

能为用户提供丰富准确的文物信息

应用要有良好的运行效率

#### **运行环境**

用户可以通过网络登录到本系统中查询相关信息

硬件要求：安卓 6.0 及以上

#### **角色定义-使用者**

使用者是指在 APP 使用本系统查询文物相关信息，使用本系统提供的功能的人员

#### **场景 1：用户对文物进行点赞评价**

主要参与者：用户

情景目标：对某文物进行点赞评论

前提条件：需要登入，联网

操作：

打开软件进入登陆界面

登录

用户选择某样文物

进入文物详细界面

用户进行点赞

用户进入评论区

用户写下评论

点击上传评论

评论上传成功

异常：

评论内容不合法，对不合法内容进行处理

服务器无响应

使用频率：每天多次

使用方式：app

次要参与者：系统后台

次要参与者参与方式：显示点赞状态，保存并展示评论

#### **场景 2：用户按关键字搜索文物的基本信息**

主要参与者：用户

情景目标：通过搜索功能搜索到相应的文物信息

前提条件：需要联网

操作：

用户进入 APP

在主界面点击搜索框

输入有关文物的信息

等待系统响应

点击文物查看信息

异常：

系统未响应

未找到与信息相关的文物

使用频率：每天多次

使用方式：app

次要参与者：系统后台

次要参与者参与方式：检索与关键字相关的文物并展示

### **场景 3：根据上传图片搜索相关文物**

主要参与者：用户

情景目标：通过图片搜索功能搜索到相应的文物信息

前提条件：需要联网

操作：

用户进行登录

点击以图搜图

用户给予软件读取权限

读取图片

根据图片匹配文物，打开存储文物信息的新页面

用户拖动屏幕，找到搜索的文物进入文物信息页面

异常：

用户没有给予读取存储权限

系统未响应

使用频率：每天多次

使用方式：app

次要参与者：系统后台

### **场景 4：用户修改个人信息：**

主要参与者：用户

前提条件：需要联网

操作：

用户进入 app

用户进入用户列表

用户修改想要修改的信息（邮箱、密码、用户名等）

异常：系统未响应

使用频率：每天多次

使用方式：app

次要参与者：系统后台

### **场景 5：用户注册**

主要参与者：用户

前提条件：需要联网

操作：

用户进入登陆界面/注册界面

用户输入邮箱、密码及用户名

异常：

注册信息不合法

系统未响应

使用频率：每天多次



使用方式：app

次要参与者：系统管理员

参与方式：审核注册信息

### **场景 6：用户登录**

主要参与者：用户

前提条件：需要联网

操作：

进入登陆界面

填写账号密码

点击登录进行登录

异常：

系统未响应

填写未注册的账号，登陆失败

填写错误的密码，登陆失败

使用频率：每天多次

使用方式：Web

次要参与者：系统后台

### **响应时间需求**

操作响应时间应该在五秒以内

### **可靠性需求**

系统应 7\*24 小时提供服务，允许多人使用并正常运行

### **系统安全性需求**

系统要有严格的权限管理，防止用户的非法操作对系统造成破坏或导致数据泄露

## **3.5 后台管理子系统其他需求**

### **系统要达到的目标**

能为用户提供丰富准确的文物信息

平台要有良好的运行效率

### **运行环境**

本系统部署在 Web 端，通过 URL 进行访问。后端使用 SpringBoot 微服务框架，Java 版本 11，前端使用 Vue 框架，Nodejs 版本为 18

### **角色定义-使用者**

使用者是指通过本系统进行后台系统管理的人员

### **场景 1：使用者管理后台用户信息**

主要参与者：使用者

情景目标：管理后台用户信息

前提条件：需要登入，联网

操作：

使用者登入系统，输入用户名和密码

使用者选择用户信息管理

使用者选择相对应的用户信息进行增删改查

使用者确认修改

系统进行相应回答

异常：

使用者输入不正确的用户名或者密码

服务器无响应

使用频率：每天多次

使用方式：Web

次要参与者：系统后台

## **场景 2：使用者管理后台数据信息**

主要参与者：使用者

情景目标：管理后台数据信息

前提条件：需要联网

操作：

使用者登入系统，输入用户名和密码

使用者选择数据信息管理

使用者选择相对应的数据信息进行增删改查

使用者确认修改

系统进行相应回答

异常：

使用者输入不正确的用户名或者密码

服务器无响应

使用频率：每天多次

使用方式：Web

次要参与者：系统后台

## **响应时间需求**

操作响应时间应该在五秒以内

## **可靠性需求**

系统应 7\*24 小时提供服务

## **系统安全性需求**

系统要有严格的权限管理，防止用户的非法操作对系统造成破坏或导致数据泄露