软件工程综合实验要求

海外藏中国文物知识管理与服务平台

北京化工大学

二零二三年十二月

### 1、知识图谱构建子系统

编写程序从海外博物馆网站爬取中国文物信息，进行加工处理，构建海外文物知识图谱。主要包括以下功能：

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3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。

### 2、海外文物知识服务子系统

编写Web端程序，使用知识图谱构建子系统获取的数据，实现数据浏览、查询、可视化等服务。主要包括以下功能：

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](https://www.allhistory.com/relationindex）、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）。

### 3、知识问答子系统

使用知识图谱构建子系统获取的数据或互联网百科数据等，实现知识问答服务。主要包括以下功能：

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。建议使用基于大语言模型的知识图谱检索增强技术。

2）复杂问答：选做。

3）闲聊：选做。

### 4、掌上博物馆

编写手机端App，主要支持文物展示、用户交互等功能。选取知识图谱构建子系统获取中的单类文物作为掌上博物馆的展示内容，例如瓷器、玉器、青铜器或中国画等。主要包括以下功能：

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）

注意：可自行设计增加更多功能。可参考<https://www.dpm.org.cn/> Creative.html#app。

### 5、后台管理子系统

编写Web端程序，管理系统中的所有数据。主要包括以下功能：

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、音视频等内容的功能。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。

### 附录1：知识图谱构建子系统参考资料

（1）CIDOC-CRM参考文档：[https://cidoc-crm.org/html/cidoc\_crm\_v7.1.1\_with\_translations.html#E1](https://cidoc-crm.org/html/cidoc_crm_v7.1.1_with_translations.html" \l "E1)

（2）建模工具Karma：<https://github.com/usc-isi-i2/Web-Karma>

该工具网络资源较少，建议直接参考项目wiki

（3）Neo4j图数据库，用来管理三元组

（4）Virtuoso数据库，用来管理三元组<https://github.com/openlink/virtuoso-opensource>

### 附录2：数据爬取格式及要求

（1）数据格式要求

所有爬取的数据以utf-8编码保存为csv格式。建议使用csv，pandas等包进行读写，防止出现格式错误等问题。各组应自行检验导出csv文件可正常使用。

（2）数据内容要求

应至少包含以下数据，文物基础数据（网站提供），如标题、时代、介绍等；文物对应的详情页面的URL；文物图片（原图，爬取请注意检查，部分网站可能需要手动解析原图地址）；原图下载链接。

需要提交一个说明文档，包括英文字段对应的中文说明，如time：文物时代。

### 附录3：海外文物知识服务子系统参考资料

（1）查询：

克利夫兰博物馆：搜索界面包括简单搜索和高级搜索（advanced search）https://www.clevelandart.org/art/collection/search

（2）时间轴：

时间轴工具：https://timeline.knightlab.com/

全历史时间轴：https://www.allhistory.com/

（3）知识图谱可视化：

全历史关系图谱：（链接为单个主题示例，更多内容参考该网站-关系图谱）https://www.allhistory.com/relation?networkId=5cf8e3ecefe5550001f1435c

（4）关系图谱：

历史人文大数据平台：https://dhc.library.sh.cn/relation?uri=http%3A%2F%2Fdata.library.sh.cn%2Fentity%2Fperson%2F4lnpsgh342b2n1ru

（5）关系图谱可视化工程参考（理解展示形式，实现方法自拟） <https://emiliorizzo.github.io/vue-d3-network/>

### 附录4：数据爬取网站

每团需爬取3家海外博物馆。为平衡各组爬取数量，请按照下表爬取。如需调换，需要两组组长同意后，上报助教老师

表1 网站分配

|  |  |  |  |
| --- | --- | --- | --- |
| 团号 | 爬取博物馆序号（对应见下表） | | |
| 1团 | 1 | 11 | 17 |
| 2团 | 2 | 10 | 16 |
| 3团 | 3 | 9 | 15 |
| 4团 | 4 | 8 | 14 |
| 5团 | 5 | 7 | 13 |
| 6团 | 6 | 12 | 18 |

表2 海外博物馆名称、网址和中国文物数量

|  |  |  |
| --- | --- | --- |
| 序号 | 博物馆名 | 网址 |
| 1 | 大英博物馆 | <https://www.britishmuseum.org/collection/search?keyword=chinese&place=China&view=grid&sort=object_name__asc&page=1> |
| 2 | 大都会 | <https://www.metmuseum.org/art/collection> |
| 3 | 史密斯 | <https://www.si.edu/openaccess> |
| 4 | 波士顿美术馆 | <https://www.mfa.org/collections/search> |
| 5 | 费城艺术博物馆 | <https://www.philamuseum.org/collections/search.html> |
| 6 | 弗利尔美术馆 | [https://www.freersackler.si.edu/collections/ https://asia.si.edu/collection-area/chinese-art/](https://www.freersackler.si.edu/collections/) |
| 7 | 美国自然历史博物馆 | <https://anthro.amnh.org/collections> |
| 8 | 纳尔逊-阿特金斯艺术博物馆 | <https://art.nelson-atkins.org/collections> |
| 9 | 旧金山亚洲艺术博物馆 | <http://searchcollection.asianart.org/view/objects/asimages/23140?t:state:flow=925e5439-97d7-4a9b-9733-62cdb565a1d9> |
| 10 | 克利夫兰 | <https://www.clevelandart.org/art/collection/search> |
| 11 | 普林斯顿大学艺术博物馆 | <https://artmuseum.princeton.edu/search/collections> |
| 12 | 西雅图艺术博物馆 | <https://www.seattleartmuseum.org/> |
| 13 | 圣地亚哥艺术博物馆 | <https://collection.sdmart.org/objects-1/portfolio?records=50&query=Creation_Place2%20has%20words%20%22china%22&sort=9> |
| 14 | 明尼阿波利斯艺术博物馆 | <https://collections.artsmia.org/> |
| 15 | 芝加哥艺术博物馆 | <https://www.artic.edu/collection?place_ids=China> |
| 16 | 宾夕法尼亚大学考古与人类学博物馆 | <https://www.penn.museum/> |
| 17 | 哈佛大学 | <https://harvardartmuseums.org/collections?q=chinese> |
| 18 | 亚伯达大学 | <https://search.museums.ualberta.ca/> |

注：**国外博物馆**一般可在博物馆网站文物搜索栏中使用“chinese”，“china”等关键字进行筛选，还可在网站上提供的location等字段进行筛选，以得到中国文物。