以下是针对海外文物知识图谱系统的详细项目管理计划模板，包含关键管理要素和可执行方案：

---

海外文物知识图谱系统项目管理计划

一、项目概览

|  |  |
| --- | --- |
| 属性 | 说明 |
| 项目名称 | 海外文物知识图谱构建系统 |
| 项目周期 | 2025年4月15日 - 2025年6月3日（共7周） |
| 项目目标 | 完成3家博物馆数据爬取、构建三元组的知识图谱、实现可视化查询接口 |

二、项目范围管理

1. 功能范围（In-Scope）

- ✅ 支持大英博物馆/大都会博物馆/东京国立博物馆网站数据爬取

- ✅ 实现文物数据清洗与标准化（年代转换、单位统一）

- ✅ 构建基于Neo4j的知识图谱存储系统

- ✅ 提供基础SPARQL查询接口

- ✅ 数据质量校验覆盖率≥95%

2. 非功能范围（Out-of-Sarcope）

- ❌ 非中国文物的数据处理

- ❌ 博物馆实时数据同步

- ❌ 多语言支持（仅限中文）

三、项目时间计划（甘特图关键节点）

四、团队组织架构

|  |  |  |  |
| --- | --- | --- | --- |
| 角色 | 人数 | 职责 | 关键产出物 |
| 项目经理 | 1 | 整体进度把控、风险管理 | 项目计划书/周报/验收报告 |
| 爬虫开发工程师 | 1 | 网站爬取/反爬破解/数据采集 | 可运行爬虫程序/原始数据集 |
| 数据工程师 | 1 | 数据清洗/知识建模/质量校验 | 标准化数据集/RDF三元组文件 |
| 后端开发工程师 | 1 | 数据库设计/API开发 | 数据库Schema/GraphQL接口 |
| 测试工程师 | 1 | 自动化测试/压力测试 | 测试报告/Bug清单 |

五、风险管理计划

|  |  |  |  |
| --- | --- | --- | --- |
| 风险项 | 概率 | 影响 | 应对策略 |
| 目标网站反爬机制升级 | 高 | 严重 | 备用方案：Selenium模拟浏览器 |
| 文物年代信息缺失率超预期 | 中 | 中等 | 发百科数据自动补全工具 |
| Neo4j性能瓶颈 | 低 | 高 | 预备ArangoDB作为备选方案 |

六、沟通管理计划

1. 日常沟通

- 每日站会：15:30-16:00（微信）

- 紧急问题：Slack即时通知通道urgent

2. 文档管理

- 需求文档：Confluence系统

- 代码管理：Github仓库（每日17:00自动备份）

3. 里程碑会议

- 需求确认会：2025-05-05

- 中期评审会：2025-05-25

- 结项汇报会：2025-06-5

七、测试与验收计划

测试类型：

1. 爬虫稳定性测试：

- 连续72小时运行，故障率<0.5%

- 数据重复率≤3%

2. 知识图谱完整性测试：

- SPARQL查询响应时间<800ms（百万级数据）

- 实体关系覆盖率≥90%

3. 用户验收标准：

- 通过10个典型查询用例（如"展示清代瓷器清单"）

- 数据准确率≥98%

验收流程：

开发团队自测 → 测试组验证 → 客户UAT环境验收 → 签署验收确认书

八、交付物清单

1. 技术交付物：

- 可执行爬虫系统（Docker镜像）

- 知识图谱数据库备份文件

- API接口文档（Swagger格式）

2. 管理交付物：

- 项目总结报告（含性能指标）

- 系统运维手册

- 数据字典（含字段说明）

关键成功因素（CSF）

1. 反爬策略有效性：确保至少2家博物馆数据完整获取

2. 知识建模质量：人工抽检三元组合格率≥95%

3. 资源保障：标注团队需在7月前到位

4. 风险控制：建立每日数据采集量监控仪表盘

该计划可直接用于项目启动，建议配合使用Jira进行任务跟踪，并通过燃尽图监控开发进度。实际执行中需重点关注数据采集阶段的网站结构变化监测