

## 项目结题验收单

专家验收表（主持人所在单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	基于知识图谱的图书馆借阅问答系统设计与实现			
主持人	席亚军	职务/职称	副研究馆员	
所在单位	（加盖单位公章）塔里木大学图书馆			
专 家 意 见	<p>本项目自立项以来，项目组成员按照各自分工与项目计划安排，顺利完成了本课题项目的研究和实践工作，主要完成了以下工作：</p> <p>1. 图书馆数据集构建。本文对图书馆知识图谱的构建首先是通过豆瓣读书、图书馆官网、百度百科进行原始数据采集，得到结构化数据和非结构化数据，并使用训练好的模型对非结构化数据进行处理得到结构化数据。</p> <p>2. 图书馆知识图谱构建。首先根据本文特点来进行图书馆知识图谱的设计，尽量使结构便于后期系统的搭建。其次对知识图谱数据的生成做实体抽取、知识融合、知识存储的一系列操作。最后对生成的图书馆知识图谱使用 Cypher 语句进行可视化展示。</p> <p>3. 基于知识图谱的图书馆智能服务系统搭建。本文设计并开发图书馆智能服务系统，后端使用 Python，数据库使用 Neo4j，前端使用 Flask 框架进行展示。系统功能在智能问答的基础上增加了图书查询、新闻公告、本馆介绍，联系等功能，可以更加便捷的服务读者，具备一定的实用价值。</p> <p>4. 撰写论文 2 篇，1 篇已经录用，1 篇还在盲评；申请软件著作权 1 项（已提交版权局）；申请实用新型专利 2 项；参与编写教材 2 部。</p>			
专家签字	苏晔	周春宏	陈竺	李新
职务/职称	研究馆员	研究馆员	研究馆员	副研究馆员



项目编号: 2022057

## CALIS 全国农学文献信息中心研究项目 结题报告

项目名称: 基于知识图谱的图书馆借阅问答系统设计与实现

项目关键词: 知识图谱 智能问答系统 自然语言处理 深度学习

项目单位(盖章): 塔里木大学图书馆

通信地址: 新疆阿拉尔市塔里木大学图书馆 843300

项目主持人: 席亚军

联系电话: 13565142267

电子邮件: 187170403@qq.com

提交日期: 2023年5月20日

# 目 录

1 绪论.....	1
1.1 研究背景.....	1
1.2 国内外现状.....	2
1.3 研究的目的及意义.....	3
1.4 研究方法及内容.....	3
1.4.1 研究思路.....	3
1.4.2 研究方法.....	4
1.4.3 研究内容.....	4
1.5 相关技术概述.....	5
1.5.1 技术路线.....	5
1.5.2 自然语言处理技术.....	7
1.5.3 知识图谱技术.....	8
2 图书馆数据集的构建.....	9
2.1 数据源.....	9
2.1.1 数据采集.....	9
2.1.2 数据预处理.....	10
2.2 图书馆实体抽取.....	11
2.2.1 命名实体识别概念.....	11
2.2.2 基于 BERT 改进的命名实体识别模型.....	11
2.2.3 数据标注.....	12
2.2.4 模型搭建和训练.....	14
2.2.5 模型的评估.....	15
2.3 图书馆数据整合.....	15
3 图书馆知识图谱的构建.....	16
3.1 图书馆知识图谱的构建流程.....	16
3.2 图书馆知识图谱的设计.....	17
3.3 图书馆知识图谱的生成.....	18
3.3.1 实体抽取.....	18

3.3.2 知识融合 .....	19
3.3.3 知识存储 .....	20
3.4 图书馆知识图谱可视化 .....	21
4 图书馆智能服务系统 .....	22
4.1 需求分析 .....	22
4.1.1 功能需求分析 .....	22
4.1.2 其他需求分析 .....	23
4.2 系统架构 .....	24
4.3 系统模块 .....	25
4.3.1 图谱化展示 .....	25
4.3.2 问答模块 .....	26
4.3.3 其他模块 .....	27
4.4 系统测试 .....	29
4.4.1 系统准确性测试 .....	29
4.4.2 综合性测试 .....	29
5 结论与建议 .....	29
5.1 结论 .....	29
5.2 建议 .....	30
6 项目成果 .....	31
参考文献 .....	43

# 基于知识图谱的图书馆借阅问答系统设计与实现

关键词：知识图谱 智能问答系统 自然语言处理 深度学习

## 1 绪论

### 1.1 研究背景

目前，高校图书馆面临着用读者需求与图书馆服务供给不平衡、不充分的挑战，迫切需要向智能知识服务转型，但是我国图书馆智能服务领域的核心关键技术还较为薄弱，缺少自主研发的服务系统、算法、模型等<sup>[1]</sup>。2017年国务院颁布的《新一代人工智能发展规划》明确提出<sup>[2]</sup>，要大力发展知识计算引擎与知识服务技术，重点突破知识加工、深度搜索和可视交互核心技术。2017年《政府工作报告》中都提出要大力发展人工智能<sup>[3]</sup>。近几年对于机器学习、深度学习、知识图谱等方面的研究成为企业以及国家研究的热点。在大数据和人工智能技术迅速发展的背景下，图书馆服务也由信息化向智能化转变<sup>[4]</sup>。

图书馆服务是一种利用图书馆馆藏的资源与设施，为读者提供相关文献与资料查询的一系列活动，因此其也被称为图书馆读者工作<sup>[5]</sup>。图书馆的服务要遵循“读者第一、服务至上”的原则，以此来开展图书馆的服务工作，在现代图书馆中合理运用智能化技术进行图书馆的智能服务是当前阶段图书馆发展的主要方向，也是基于知识图谱的图书馆智能服务改革和发展的重要要求之一。所以，为促使人工智能和知识图谱等技术在图书馆智能服务的运用中得到有效保障，对于读者文献信息的需求，利用好图书馆中的资源为读者提供相应的文献与信息<sup>[6]</sup>。本文立足于现有技术，对图书馆智能化服务中的人工智能技术的运用做了深入探索。面对人工智能的影响和冲击，图书馆应当积极应对、客观看待、理性参与、及时变革，借助新技术实现舒适的空间环境、精准的需求定位、深层的知识发现、科学的资源管理、个性的读者服务以及良性的互动交流<sup>[7]</sup>。

针对上述问题，本文提出了一种基于知识图谱的图书馆智能服务系统，该系统通过构建知识图谱来描述图书馆资源的各种属性和关系，并以此为基础来实现智能化服务。本文旨在通过实验和评估证明该系统的可行性和优越性，从而为图书馆提供更加

高效、精准的服务。基于知识图谱的图书馆智能服务系统的意义主要是提高服务质量和效率，满足读者需求，推动图书馆事业的发展，使图书馆更加智能化、信息化。基于知识图谱的图书馆智能服务系统的研究背景和意义都非常重要，它可以帮助图书馆实现更加精准、个性化的服务，提高读者的满意度和使用体验，推动图书馆事业的发展。

## 1.2 国内外现状

目前，国内外对基于知识图谱的图书馆智能服务的研究已经取得了一定的进展，各国学者也对智能图书馆提出了可行的方案。2017年是人工智能应用与发展“元年”，国内外均将其视为重要技术前沿和战略高地。2016年底，信息技术研究和分析公司Gartner将人工智能作为2017年十大重要科技发展趋势之一<sup>[8]</sup>。但是，现有研究在实现服务智能化和提高服务质量方面仍然存在不足之处，例如资源描述不够精准、智能推荐算法效果不佳等。因此，本文旨在探索一种基于知识图谱的新型图书馆智能服务系统，以此来提高图书馆服务的效率和质量。

在国内，一些大型图书馆已经开始了基于知识图谱的智能服务系统的研究和应用。例如，首都图书馆建设街区自助图书馆，应用RFID技术的小型自助图书馆于东莞市落地，可为市民提供全天图书借还服务<sup>[9]</sup>。清华大学是最早将基于人工智能的问答引擎融入图书馆服务平台的国内高校之一<sup>[10]</sup>，其研发的“小图”主要基于AIML（人工智能标记语言）和搜索引擎技术，提供业务咨询和馆藏搜索。南京大学图书馆<sup>[11]</sup>、西安交通大学图书馆<sup>[12]</sup>等在内的诸多图书馆也基于检索、关键词匹配等技术构建了基于微信平台的问答系统。国家图书馆的“智慧图书馆”项目，上海图书馆的“知识图谱服务平台”等项目。各种基于知识图谱的图书馆项目对图书馆中的数据进行分析和挖掘，构建出一个知识图谱，实现了对图书馆服务的智能化服务。

在国外，一些知名的图书馆如美国国会图书馆、英国国家图书馆等也都在积极探索基于知识图谱的智能服务系统。例如，美国国会图书馆的“知识图谱服务平台”项目，英国国家图书馆的“知识图谱与智慧图书馆”项目。麻省理工学院(MIT)计算机科学系人工智能实验室开发的START(<http://start.csail.mit.edu>)可以回答多个领域的自然语言问题、AnswerBus多语系自动问答系统，可以回答英语、法语、西班牙语、德语和葡萄牙语的问题<sup>[13]</sup>。这些项目通过对知识图谱中的信息进行分析和挖掘，实现了对图书馆服务的智能化服务，并且已经开始了商业化应用。

综上所述，基于知识图谱的图书馆智能服务系统是一种利用知识图谱技术实现图

书馆智能服务的系统，它可以帮助图书馆实现更加精准、个性化的服务，提高读者的满意度和使用体验。随着技术的不断发展和应用的不断深入，基于知识图谱的图书馆智能服务系统将会在未来发挥越来越重要的作用。

### 1.3 研究的目的及意义

当前能投入实际应用的中文场景图书馆借阅问答系统比较少，能进行精准问答的系统更是少之又少。知识图谱的结构化知识形式为问答系统提供了一种新思路，随着自然语言处理技术和深度学习的发展，提高自然语言泛化能力与知识库检索能力，研发基于知识图谱的图书馆借阅智能问答系统成为新的研究课题。但当前基于知识图谱的图书馆借阅领域问答系统在实际应用场景中还存在一些挑战：1) 图书馆借阅种类繁多，数据高度分散；2) 非结构化知识抽取困难，标注训练语料缺乏，知识图谱构建成本较高且覆盖范围不够广泛；3) 由于自然语言的复杂性，问句理解以及问句到知识图谱的语义链接困难。

基于上述问题，本项目通过爬虫技术获取分散的图书馆借阅数据，然后采用基于深度学习方法的三元组抽取技术以实现知识图谱的半自动化构建，同时提升自然语言问句理解和泛化能力，设计并实现基于知识图谱的图书馆借阅智能问答系统。智能问答符合时代用户发展的需求，提升了高校图书馆参考咨询服务的发展，实践了高校图书馆服务与技术融合创新的社会职责。对于读者，以自然语言交互，简单易用，提供技术支持，激发用户的深度咨询行为；对于高校图书馆馆员，辅助了参考咨询服务，有助于馆员提供更完善的服务。对我国当代高校图书馆的发展具有重要的促进意义。

### 1.4 研究方法及内容

#### 1.4.1 研究思路

基于以上研究背景和国内外研究现状，本文在大量阅读国内外相关文献和学习资料的基础上，确定研究方案，研究基于知识图谱的图书馆借阅智能问答系统的设计与实现，技术路线如图 1-1 所示。首先采集领域相关数据，并进行数据预处理，根据领域特点和数据内容构建领域本体；然后在本体的基础上构建图书馆借阅知识图谱，设计基于知识图谱的知识问答算法，最后实现图书馆借阅智能问答原型系统的开发与应用。

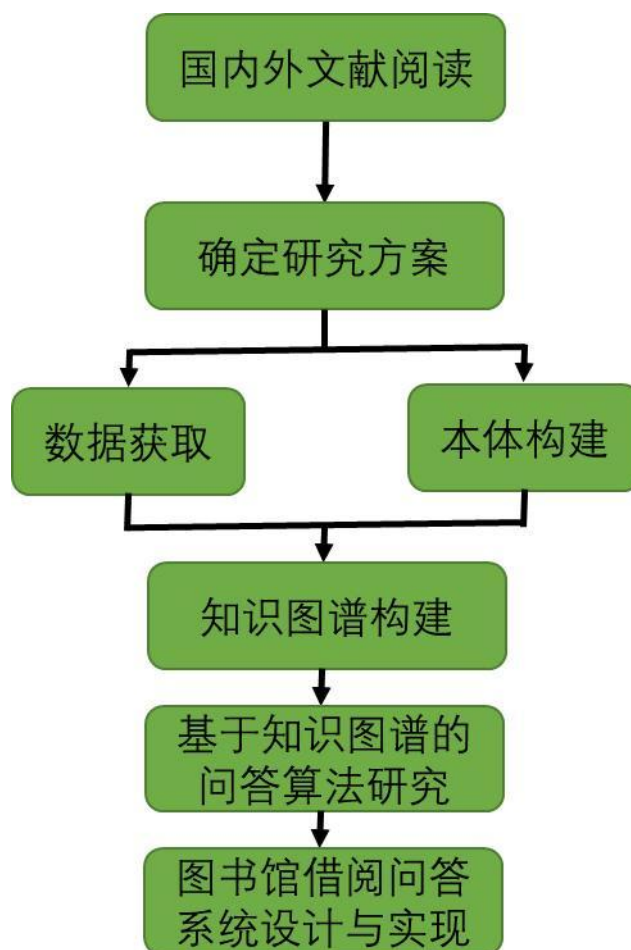


图 1-1 技术路线图

#### 1.4.2 研究方法

(1) 文献分析法。查找与阅读国内外相关文献，通过前人的研究，为自己对知识图谱、对自然语言处理、深度学习以及图书馆借阅问答系统设计科学认知和深入研究奠定一定的理论基础。

(2) 比较分析法。对深度学习与传统知识挖掘进行比较，进而从技术、成本、利用方面，对基于知识图谱的图书馆借阅问答系统设计与实现的要求进行论证。对深度学习与传统知识挖掘方法的不同之处，结合高校图书馆数字资源的价值、特点构建塔里木大学图书馆知识图谱的借阅问答系统方案。

#### 1.4.3 研究内容

- 1、相关理论和技术的研究分析
  - (1) 本体
  - (2) 知识抽取
  - (3) 知识存储



- (4) 知识融合与知识推理
- (5) 属性链接

## 2、图书馆借阅资源知识图谱构建

- (1) 基于规则的半结构化知识抽取
- (2) 基于 BERT-BiLSTM-CRF 的实体关系联合抽取
- (3) 基于 Neo4j 的知识存储

## 3、基于知识图谱的知识问答算法

- (1) 问句实体识别
- (2) 属性连接
- (3) 问答查询
- (4) 问句相似性计算

## 4、图书馆借阅问答原型系统设计与实现

- (1) 开发环境
- (2) 系统架构
- (3) 系统的设计
- (4) 系统的实现及测试

### 1.5 相关技术概述

#### 1.5.1 技术路线

本文是基于知识图谱的图书馆智能服务系统，主要依赖于以知识图谱技术构建的图书馆知识图谱。本文的技术路线图设计如图 1-2 所示。

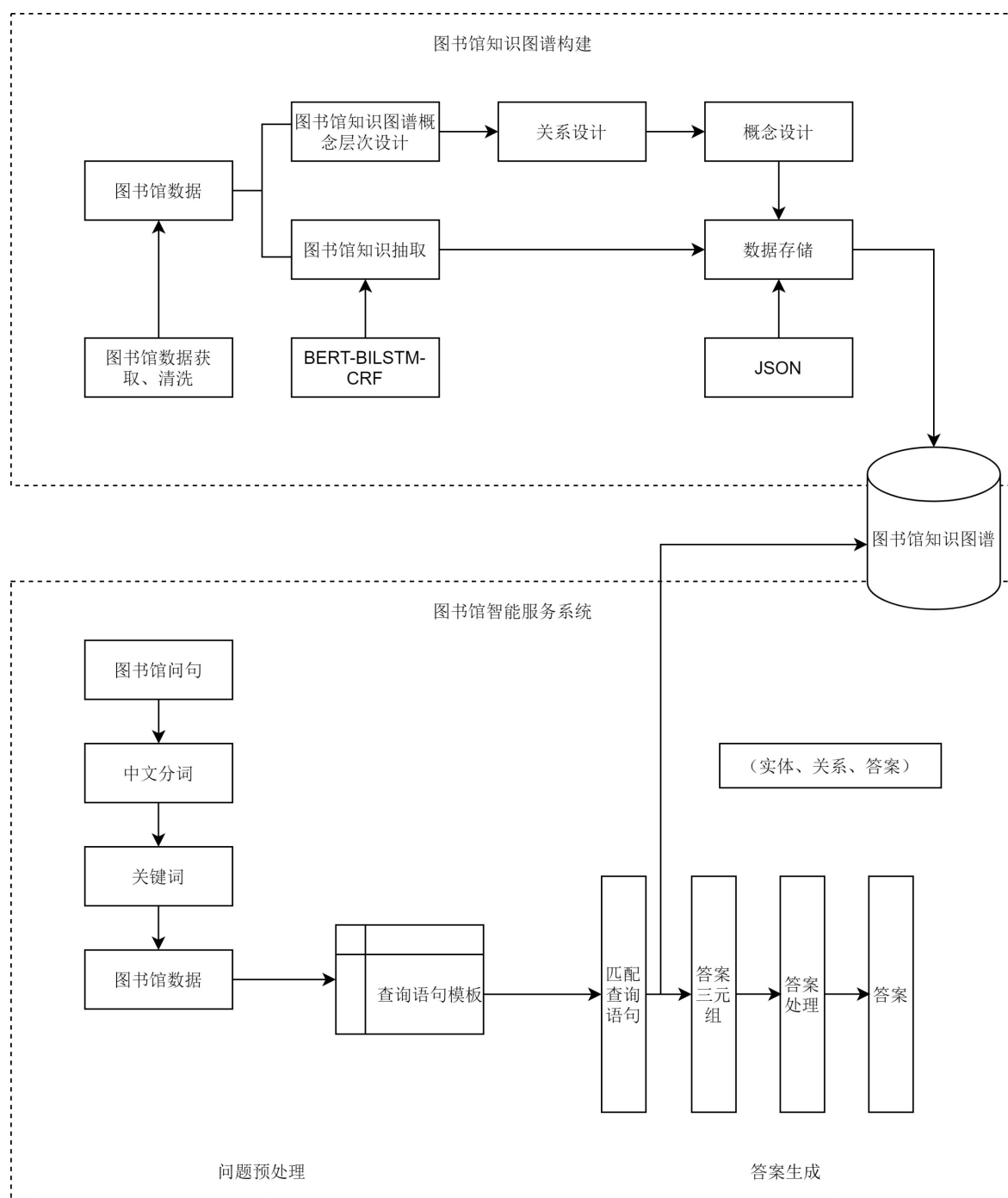


图1-2 技术路线图

知识图谱构建：首先需要收集图书馆相关的知识和数据，通过自然语言处理的人工智能核心技术，将这些数据转换成可供机器理解的知识图谱模型。在这一过程中，可以使用自然语言处理工具，如中文分词、实体识别、关系抽取等技术，对文本进行处理和解析，然后构建出实体和关系之间的知识图谱<sup>[14]</sup>。

知识图谱存储：将构建好的知识图谱存储到数据库中，以便于后续的查询和处理。在这一过程中，可以使用图数据库或关系型数据库等技术进行存储。例如，可以使用

Neo4j、HugeGraph 等图数据库进行存储<sup>[15]</sup>。

**知识图谱查询：**实现基于知识图谱的图书馆智能服务系统的核心功能之一，包括图书推荐、问答系统、数据可视化等功能。在这一过程中，可以使用 SPARQL 查询语言等技术进行查询。例如，可以通过 SPARQL 查询语言查询知识图谱中与读者输入相关的实体和关系<sup>[16]</sup>，然后给出相应的推荐、答案或可视化展示。

**读者画像：**通过对读者历史借阅记录和反馈信息的分析，建立读者画像，从而更好地进行个性化推荐和服务。在这一过程中，可以使用机器学习等技术进行数据挖掘和分析。例如，可以使用协同过滤算法、推荐系统等技术，对读者兴趣和行为进行分析，从而给出更加个性化的推荐和服务。

**系统架构设计：**在技术路线的最后阶段，需要对系统的整体架构进行设计和实现。在这一过程中，需要考虑系统的可扩展性、可维护性、安全性等方面。可以采用分布式系统、微服务架构等技术，来提高系统的性能和可靠性。

综上所述，基于知识图谱的图书馆智能服务系统的设计和实现需要涉及到多个技术领域，包括自然语言处理、数据库技术、机器学习、分布式系统等。在实践中，需要综合运用这些技术，不断优化和完善系统的功能和性能，以提供更加智能化、个性化的图书馆服务。

### 1.5.2 自然语言处理技术

自然语言处理(NLP)是一种人工智能技术，旨在让计算机理解和处理人类语言<sup>[17]</sup>。自然语言处理是人工智能的核心技术，实体关系提取和语义分析均属于典型的自然语言处理工作。

自然语言处理技术已经广泛应用于许多领域，包括机器翻译、智能客服、智能推荐、自动摘要等。其中，最具代表性的应用是机器翻译，它已经成为许多跨国公司必备的技能之一。机器翻译的核心技术是自然语言处理，它使用计算机程序来理解和生成自然语言，并将其转换为计算机可以理解的形式<sup>[18]</sup>。

自然语言处理模块主要包括分词、词性标注、命名实体识别和依存句法分析等技术<sup>[19]</sup>。其中，分词技术将自然语言文本分割成一个个单独的词汇；词性标注技术将每个单词标注为其词性；命名实体识别技术用于识别出自然语言文本中的实体名称，如作品名称、作者名等；依存句法分析技术则用于分析句子中各个词语之间的依存关系。

自然语言处理技术可以用于文本分类、情感分析、问答系统、文本生成等领域。例如，在问答系统中，自然语言处理技术可以帮助程序理解读者的问题并给出相应的

答案<sup>[20]</sup>。在文本生成领域，自然语言处理技术可以根据读者的需求生成符合要求的文本。

虽然自然语言处理技术的应用前景广泛，但是它的实现非常复杂。它需要从输入的自然语言中提取有用的信息，并将其转换为计算机可以理解的形式。同时，自然语言处理技术还需要处理不同语言、不同文化之间的差异，因此它的应用受到语言和文化背景的影响。

综上所述，自然语言处理技术是人工智能领域的一个重要分支，它的应用前景非常广泛，但实现起来非常复杂。

### 1.5.3 知识图谱技术

知识图谱，本质上是一种语义网络，用于存储和组织各种类型的知识和信息。知识图谱可以用于各种应用领域，包括搜索引擎、智能客服、数据挖掘、机器学习等。本文的知识图谱构建模式结构图如图 1-3 所示。

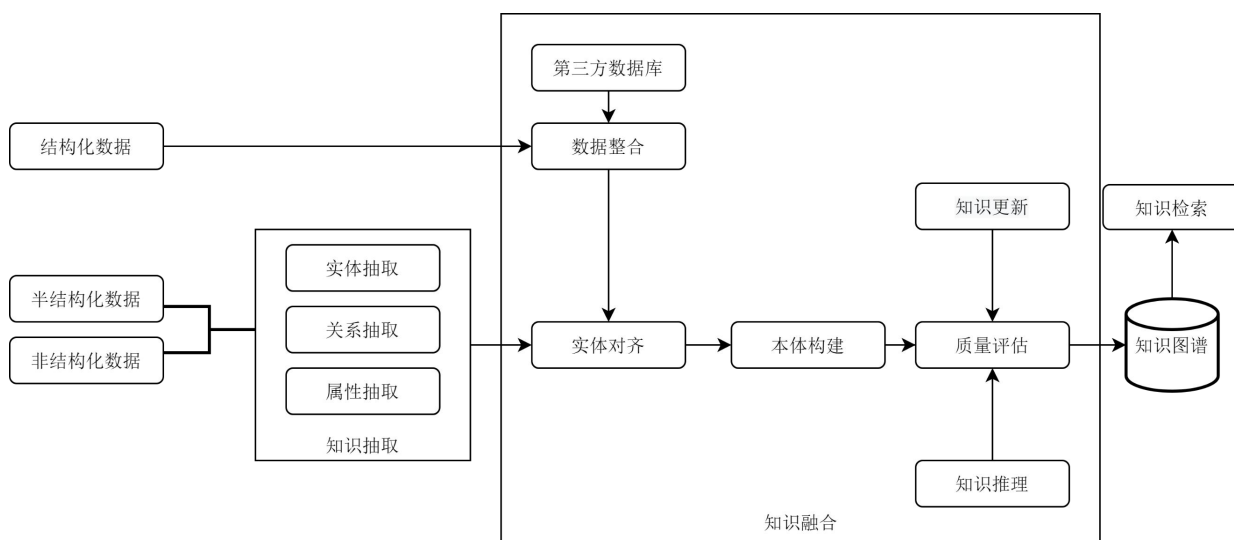


图1-3 知识图谱构建模式的结构图

知识图谱构建模块主要利用文本挖掘和知识图谱构建技术<sup>[21]</sup>，从大量的文献、书籍和其他数据源中抽取出图书馆服务相关的知识，并将其组织成一个可扩展的知识图谱。图书馆知识图谱应包括图书馆服务的相关知识、资源信息、图书馆组织机构、管理规则等内容涵盖了图书馆服务的方方面面。知识图谱的构建需要用到自然语言处理、实体识别、关系抽取等技术，将提取出的知识转化为计算机可理解的形式，并将其组织成一个结构化的图谱<sup>[22]</sup>。

知识图谱体系架构分为五部分，第一部分是知识建模，针对图书馆的应用属性，

知识特点对知识图谱结构进行设计。第二部分是知识存储，需要构建高效的存储和查询结构，本文采用图数据库，在构建图书馆知识图谱的时候本文采用三元组的形式进行表示，如<实体 1, 关系, 实体 2>、<实体、属性, 属性值>。第三部分是知识抽取，需要从半结构化和非结构化数据中获取有效的结构化数据。第四部分是知识融合，在得到实体和实体间关系后就要通过知识融合来增加知识的逻辑关系，从而提升知识的应用价值。第五部分是知识计算以及知识应用。

## 2 图书馆数据集的构建

### 2.1 数据源

#### 2.1.1 数据采集

通过查阅相关资料，确定本文图书馆知识图谱的数据以豆瓣读书数据库、百度百科、学校图书馆官网的数据为主，其包含了图书馆的读者服务、借阅服务、馆藏分布、图书馆活动、图书馆设备、学科知识、等方面的信息，以满足图书馆智能服务系统知识图谱构建时对数据质量的高要求。

构建本文所需的知识图谱首先需要获取构建所需数据集<sup>[23]</sup>。本文所使用的数据集来源比较广泛，但是最主要的来源包括三个。首先是利用豆瓣读书平台来获取图书的各类相关信息，最终获取到的结构化图书信息数据，共计 92556 余条。其次是对获取的图书信息进行知识的补充完善，本文使用百度百科搜索引擎对图书的结构化和非结构化信息进行搜索补充，获取到前者提到的结构化数据中图书相关的信息。最后是获取学校官网馆藏分布、活动、设备、新闻等非结构化信息来对知识图谱进行补充。之所以选择这三个主要来源的原因主要又三点。首先是因为豆瓣读书平台图书数据量大，对于图书的作者、出版社、分类等属性描述比较完善，试验中也发现豆瓣读书平台的数据比较容易获取，技术上难度不是很大。其次是对图书属性等信息的补充，百度百科所能查询到的数据可以很好的对已经获取的数据进行补充，本文知识图谱的结构以图书名称为中心节点，可以较好的体现图书馆知识图谱的其他各类服务的实体关系。最后是为了对图书馆其他服务的数据集进行收集。对于本文收集的数据集，后期需要对实体进行消歧工作<sup>[24]</sup>，比如，书籍的 ISBN 号是唯一项，可以使用 ISBN 号来统一标识书籍，对于作者、评分等不是唯一的字段，需要对实体进行统一。

从以上提到的三个数据库中抓取的数据有结构化数据和非结构化数据<sup>[25]</sup>，这个数据集包含大量的文本数据，并且尽可能地覆盖各种不同的语言和主题。

结构化数据包括，作品、作者、分类、出版时间、评分、价格、出版社、丛书名、ISBN、区域、状态，如表 2-1 所示。

表 2-1 结构化实体类型

作品	作者	分类	出版时间	评分	价格	出版社	丛书名	ISBN	区域	状态
冬牧场	李娟	文学	2012-6	9	29.80 元	新星出版社	新经典 文库	9787513 305662	四楼 C区	借出
爱你就 像爱生 命	李银河	文学	2008-5	8.8	18.00 元	上海锦绣文 章出版社	王小波 作品全 集	9787806 859988	四楼 C区	在馆
城南旧 事	林海音	文学	2003-7	9	16.00 元	中国青年出 版社	现当代 文学名 著系列	9787500 652045	四楼 C区	借出
三体	刘慈欣	文学	2008-1	8.8	23.00 元	重庆出版社	科幻世 界	9787536 692930	四楼 C区	在馆

非结构化数据内容包括开放时间、分布、活动、方式、时间、设备、介绍等，如表 2-2 所示。

表 2-2 非结构化实体类型

开放时间	分布	活动	方式	时间	设备
周一至周日	总借阅台	新生入馆教育	微信公众号	2月28日	朗读亭
周一至周日	新书展区	试用知网研学平台	微信扫码	4月7日	自助售卖机

### 2.1.2 数据预处理

在自然语言任务处理中，中文与英文最直观的不同就是中文的词是由一串连续的字组成，词语之间没有像英文那样的天然的空格把词分隔开。并且处理后的数据集也会存在缺失和重复等问题，所以在使用数据之前需要对数据集进行数据预处理，对一些系统使用不到的数据简单的删除即可，对数据集中的缺失值筛选后补全来达到数据清洗的效果，对于重复数据进行数据整合，多个文件的数据集成为一个，对于冗余的数据进行数据减少操作，也减少就是相同或相识的数据。由于数据之后要用于标注任务，所以需要确定好实体的类型和边界。本文们在字层面进行 BIO 标注。

本文所收集的图书馆数据来自不同的平台，所以数据集的结构、类型会存在一定的差异。所有数据集准备完成后，数据集可能包含部分与本文无关的错误信息，如果

不进行剔除，极大可能会导致实验出错或者无法进行下去这会影响到实验的进度。比如，除了本文所涉及的图书馆信息之外，还有一部分数据不属于本文所开发系统，这些数据和图书馆数据混合在一起，也有些是数据不完整，缺失了部分数据。对于本文所提到的这部分错误数据的处理方法首先是利用 jieba<sup>[26]</sup> 分词库对收集好的数据集进行分词处理，之后在混合数据里对图书馆数据进行提取，以此来解决数据混合的问题。最后，通过对准备好的数据集的缺失数据项的筛选，查找到缺失数据项，与语料库进行比较，补全缺失数据，以此来解决数据不完整、缺失的问题。对所有数据集与处理完毕后，剩余的部分非结构化数据，本文采用 BERT-BILSTM-CRF 命名实体识别模型对其进行结构化处理，最终得到结构化图书馆数据集。

## 2.2 图书馆实体抽取

### 2.2.1 命名实体识别概念

NER (Named Entity Recognition)，中文称为命名实体识别<sup>[27]</sup>，是 NLP 中一项非常基础的任务。命名实体一般指的是文本中具有特定意义或者指代性强的实体，通常包括人名、地名、机构名、日期时间、专有名词等。而命名实体识别，就是要在文本中将这些实体标注出来。

命名实体识别任务是信息提取、问答系统、句法分析、机器翻译等诸多 NLP 任务的基础，其准确度决定了下游任务的效果，是 NLP 中非常重要的一个基础问题。比如在关系抽取任务中，命名实体识别起到了抽取实体的作用，其准确度直接影响了关系抽取的最终结果。

### 2.2.2 基于 BERT 改进的命名实体识别模型

BERT 使用了迁移学习的思想<sup>[28]</sup>，先用大量数据预训练出一个较为通用的模型，然后再针对本文具体的命名实体识别模型训练任务，最后只需要再使用较小的模型进行微调，减小训练成本。迁移学习通俗来讲，就是运用已有的知识来学习新的知识，核心问题是找到已有知识和新知识之间的相似性，使用预训练好的模型来抽取词、句子的特征，类似人类学习中的举一反三。

NLP 领域目前的趋势就是预训练，即先使用某种模型预训练一个语言模型出来，包括 ELMo、GPT、还有 BERT 都采用了预训练的语言模型。BERT 与其他预训练模型的区别在于 ELMo 对上下文进行双向编码<sup>[29]</sup>，但是 BERT 结合了两种方法都有点，对上下文进行双向编码。因此，本文使用 BERT-BILSTM-CRF 结合的模型。

BERT 的特点首先是模型非常的深，在训练双向语言模型时以较小的概率把少量的

词替成了 Mask 或者另一个随机的词，增加预测下一句的 loss，模型深而窄，12 层且中间层只有 1024，对比 Transformer 的中间层为 2048 减少了一半。其次是 Masked Language Model (MLM)，同时利用左侧和右侧的词语，该方法最早在 EMLo 中提出。

LSTM (Long-Short Time Memory) 模型最早由 Hochreiter<sup>[30]</sup> 提出，是一种特殊的循环神经网络。模型内部具体实现非常复杂，主要是为了把上下文的数据信息进行保存和更换。模型内部实现由一下公式说明。第一个公式是为了确定上衣时刻的单元状态到当前时刻可以保留多少，第二个公式表示输入门更新，第三个公式计算结果表示当前单元状态，最后两个公式表示输出门的计算。

$$f_t = \text{Sigmoid}(W_f \times [h_{t-1}, x_t] + b_f) \quad (2-1)$$

$$i_t = \text{Sigmoid}(W_i \times [h_{t-1}, x_t] + b_i) \quad (2-2)$$

$$o_t = \text{Sigmoid}(W_o \times [h_{t-1}, x_t] + b_o) \quad (2-3)$$

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_c \times [h_{t-1}, x_t] + b_c) \quad (2-4)$$

$$h_t = o_t * \tanh(C_t) \quad (2-5)$$

条件随机场 (Conditional Random Fields, CRF) 作为一种条件概率分布模型被用于命名实体识别<sup>[31]</sup>。由于直接使用 BERT 和 BILSTM 结合的模型对于输出结果中的标签并不一定是符合要求的，各个标签之间有一定的关系，所以本文在 BILSTM 之后还需要引入 CRF 层来把模型的精度进行提高，其具体的原理用以下公式来说明， $Z(x)$  表示归一化式子， $Z(x)$  和  $s(x, y)$  的具体的计算式子如下所示，最后一式代表 LSTM 的输出概率和转移概率，也就是 CRF 转移概率对应的具体值。

$$P(X/Y) = \frac{\exp(s(x, y))}{Z(x)} \quad (2-6)$$

$$Z(x) = \sum_y \exp(s(x, y)) \quad (2-7)$$

$$s(x, y) = \sum_i \text{Emit}(x_i, y_i) + \text{Trans}(y_{i-1}, y_i) \quad (2-8)$$

### 2.2.3 数据标注

由于本文训练模型需要使用标注数据，而联合标注与原始标注相比，相邻词语标签之间可能会存在依赖关系。这一问题可以通过标签转化的方式，把联合标注转化成原始标注解决，所以本文使用 BIO 标注。

BIO 标注，即把联合标注中的每种存在“跨字”情况的标签 X 转化成新的两种标签“B-X”“I-X”。B 即 Begin，被标注为“B-X”的字表示该字所在词片段类型为 X，且该字是词片段的起始字；I 即 In，“I-X”表示该字所在的词片段类型为 X，但该字是词



片段的起始字之后的字；O 表示该字不属于事先定义的任何词片段类型。

首先对原始数据进行处理，使用脚本文件对原始数据的每一段文字分割为一个文本文件，便于后期使用标注工具进行标注。

前期准备数据为一个文本文档，每段数据最后以回车符结尾，数据主要已图书为主，原始数据如图 2-2 所示。

- 1 《红楼梦》是清代作家曹雪芹所著的一部长篇小说，被誉为中国古代小说的巅峰之作，也是世界文学名著之一。小说描写了一个大家族的兴衰历程，主要围绕着贾宝玉、林黛玉、薛宝钗等人的爱情、
- 2 《活着》是一部由余华所著的小说，讲述了一个中国农民的悲惨人生。小说通过主人公福贵的生活经历，深刻地反映了中国农民在历史的变迁中所遭受的苦难和不幸，并探讨了人生的意义和价值。该
- 3 《1984》是乔治·奥威尔所著的一部反乌托邦小说，出版于1949年。小说描绘了一个虚构的极权主义社会，人民被剥夺了所有自由和隐私，由政府全面控制和操纵。小说主要讲述了主人公温斯顿·史密
- 4 《三体》是刘慈欣所著的一部科幻小说三部曲，包括《三体》、《黑暗森林》和《死神永生》。小说主要讲述了地球人类与三体人之间的交流和冲突，以及人类如何应对外星文明的挑战和威胁。小
- 5 《百年孤独》是哥伦比亚作家加西亚·马尔克斯所著的一部魔幻现实主义小说，出版于1967年。小说主要讲述了布恩迪亚家族七代人的故事，揭示了家族成员之间的复杂关系和命运的轮回。小说中运用
- 6 《哈利·波特》是美国作家J.K.罗琳所著的一系列魔幻小说，共7本，讲述了一个孤儿哈利·波特的成长历程，他在霍格沃茨魔法学校学习魔法，与好友罗恩和赫敏一起经历了一系列冒险和挑战，最终
- 7 《飘》是美国作家玛格丽特·米切尔所著的一部小说，出版于1936年。该小说背景设定在美国南北战争时期，讲述了一个爱情故事，主人公是一个名叫斯嘉丽的女性，她是一个富有的南方贵族，经历
- 8 《三国演义》是中国古典小说之一，描写了中国东汉末年至三国时期的历史。小说以三国时期的政治、军事、外交等大事为主线，穿插了大量的英雄故事和传奇故事，塑造了许多历史人物的形象，如
- 9 《房思琪的初恋乐园》是台湾作家林奕华所著的一部小说，出版于2016年。小说主要讲述了一个少女房思琪的成长故事，她在童年时期遭受了性侵犯，以及她在成长过程中所经历的挫折和困惑。小说
- 10 《动物农场》是美国作家乔治·奥威尔所著的一部反乌托邦寓言小说，出版于1945年。小说以动物为主人公，讲述了它们在动物农场中推翻人类统治、建立自己的政权，最终被一些猪头篡夺领导权。《
- 11 《福尔摩斯探案全集》是英国作家柯南道尔所著的一系列侦探小说，共包括4部短篇小说和4部长篇小说。小说主要讲述了福尔摩斯与华生医生一起解决各种案件的故事，福尔摩斯是一个极具才智和洞察
- 12 《白夜行》是日本作家东野圭吾的一部小说，于1997年出版。故事讲述了两个孩子之间的命运交错，一对孩子在复仇和爱情之间徘徊，最终达成了自己的目的。小说以第一人称的方式分别讲述了两个
- 13 《小王子》是法国作家安托万·德·圣-埃克苏佩里于1943年出版的一部童话故事。小说讲述了一名小王子从自己的星球上出发，寻找朋友和探索宇宙的故事。在他的旅途中，他遇到了各种各样的人和
- 14 安徒生童话故事集是由丹麦作家汉斯·克里斯蒂安·安徒生所著的一系列童话故事，首次出版于1835年。故事集包含了经典的童话故事，如《丑小鸭》、《小人鱼》、《卖火柴的小女孩》、《皇帝的新
- 15 《天龙八部》是中国著名作家金庸所著的一部武侠小说，于1963年开始连载，1969年完成。小说以北宋末年为背景，讲述了八个门派之间的恩怨情仇，展现了武林中的英雄豪杰和江湖中的荣辱悲欢。
- 16 《撒哈拉的故事》是法国作家圣-埃克苏佩里的一部小说，于1948年出版。小说讲述了一名飞行员在沙漠中遇到了一位小王子，小王子来自一个遥远的星球，他的旅途中遇到了各种各样的人和动物。
- 17 《呐喊》是中国现代文学家鲁迅所著的一部短篇小说集，于1923年出版。小说集共收录了15篇小说，涵盖了鲁迅的多种文学风格和主题，如《狂人日记》、《阿Q正传》、《药》等。小说集以现实主义
- 18 《杀死一只知更鸟》是美国作家哈珀·李所著的一部小说，于1960年出版。小说讲述了一个小女孩斯卡特在南方小镇长大的故事，她的父亲是一名律师，为一名黑人辩护，引起了当地白人的反感和歧视
- 19 《明朝那些事儿》是中国作家当年明月所著的一部历史小说，该小说讲述了明朝从建立到崩溃的历史，以及明朝时期的政治、军事、经济、文化等方面的变迁。小说通过生动的故事情节和丰富的历史
- 20 《失踪的女孩》是埃莱娜·费兰特的一本书籍，这本书讲述了一个13岁女孩失踪的故事。她的家人和朋友们都在寻找她，同时也探索了她失踪的原因和背后的秘密。

图2-1 原始数据示意图

对原始文件以回车符进行分割文件后得到多个文本文件，每个文本文件的内容是数据集里的一个段落，每个文件数据较少，便于之后使用精灵标注助手进行标注，分割完成后的文件如图 2-3 所示。

名称	修改日期	类型	大小
10000.txt	2023/3/20 20:22	TXT 文件	1 KB
10001.txt	2023/3/20 20:22	TXT 文件	1 KB
10002.txt	2023/3/20 20:22	TXT 文件	1 KB
10003.txt	2023/3/20 20:22	TXT 文件	1 KB
10004.txt	2023/3/20 20:22	TXT 文件	1 KB
10005.txt	2023/3/20 20:22	TXT 文件	1 KB
10006.txt	2023/3/20 20:22	TXT 文件	1 KB
10007.txt	2023/3/20 20:22	TXT 文件	1 KB
10008.txt	2023/3/20 20:22	TXT 文件	1 KB
10009.txt	2023/3/20 20:22	TXT 文件	1 KB
10010.txt	2023/3/20 20:22	TXT 文件	1 KB

图2-2 分割文件示意图

之后使用精灵标注助手对处理好的数据进行人工标注，在精灵标注助手里新建标注项目，定义好图书、作者等实体类型，导入完成分割的文件后对每一个分割好的文件进行标注，标注方式为选中实体后设置实体的类型，依次完成所有文件的标注。标注完成后对标注文件导出，导出后的文件为多个 .ann 文件，文件内容格式为实体编号、

实体类型、实体起始位置、实体结束位置、实体，内容如图 2-4 所示。

最后需要使用脚本处理标注好的文件为模型可以识别的 BIO 标注文本，并将所有文件合并为一个文件，如图 2-5 所示。

T1	图书	1 3	三体
T2	作者	5 8	刘慈欣
T3	图书	24 26	三体
T4	图书	29 33	黑暗森林
T5	图书	36 40	死神永生

是	O
刘	B-作者
慈	I-作者
欣	I-作者
所	O

图2-3 标注文件示意图

图2-4 BIO标注文本示意图

### 2.2.4 模型搭建和训练

在搭建 BERT 模型之前，首先需要安装训练模型所必需的依赖库，包括 TensorFlow、Rnncell、NumPy、Scikit-learn 等。这些库可以通过 pip 命令进行安装。接下来可以从 Hugging Face 官网上下载 BERT 模型，也可以使用其他网站提供的下载链接。下载完成后，将模型文件解压缩到本地目录中，下载完模型后解压缩模型文件，并将其保存到训练模型的目录中。需要注意的是，下载的 BERT 模型文件比较大，可能需要一定的时间和存储空间。同时，不同的 BERT 模型可能适用于不同的任务和数据集，所以本文选择中文版本模型进行下载。

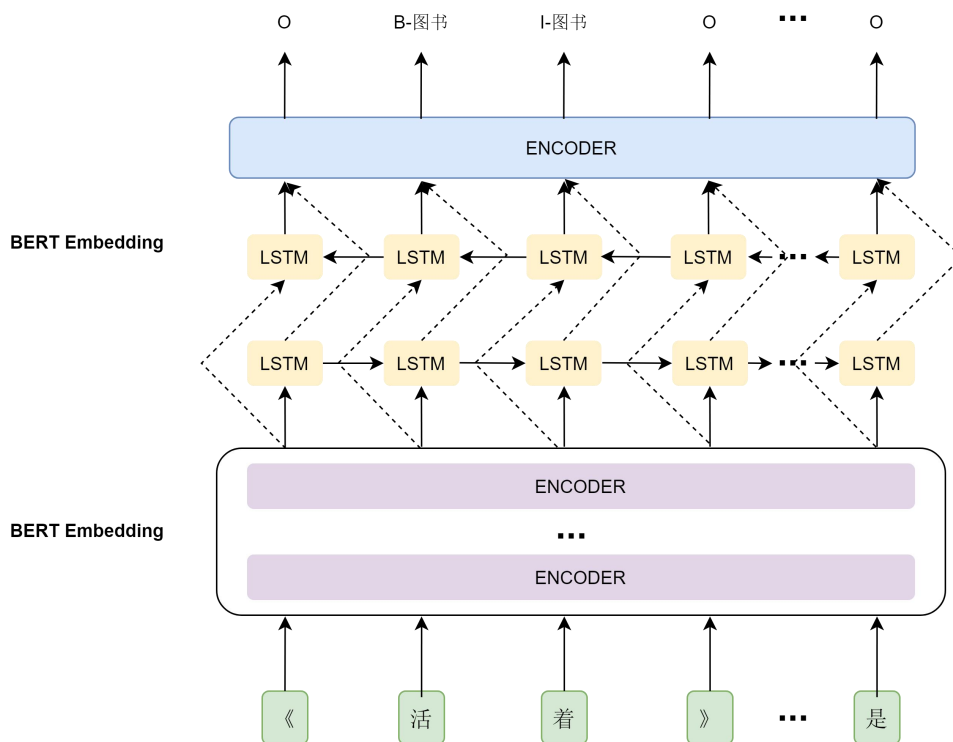


图2-5 模型的结构

搭建模型要依据用到的 BERT+Bi-LSTM+CRF 模型的结构，如图 2-6 所示，将 BIO 标注好的图书馆数据以一个标签对应一个字符的形式输入，embedding 将文本数据转换为词向量表示形式，然后再输入到 BERT 模型中进行处理。

之后启动项目开始训练 BERT 模型，在训练过程中，设置合适的参数，如学习率、批量大小、迭代次数等，使用随机梯度下降等优化算法来训练模型。在训练过程中需要注意使用正则化技术来减少过拟合的影响，使用调参技术来优化模型的性能，如调整批次大小、学习率等。

### 2.2.5 模型的评估

训练结束后会生成一个模型文件和训练日志，模型文件可以用于之后的命名实体识别处理数据。还可以利用梯度累加的时间换空间的策略，在本文算力资源有限的情况下，可以累加多个 batch 的梯度然后一起更新。最终模型实验结果对比数据如图 2-7 所示，充分说明本文选用模型合理。

表 2-3 实验结果对比

模型	F1	Precision	Recall
BERT	71.25	74.50	69.21
BERT-BILSTM-CRF	91.89	92.62	90.71

最后需要对模型的准确率进行测试，引入训练好的模型文件对“《三体》是刘慈欣所著的一部科幻小说三部曲”等非结构数据来进行识别，最终识别结果要以实体、实体起始位置、实体结束位置、实体类型的格式展示，测试后发现测试结果较为准确，测试结果如图 2-8 所示。

```
input sentence, please: 《三体》是刘慈欣所著的一部科幻小说三部曲。
2023-05-10 19:45:57.352763: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened
[{'word': '三体', 'start': 1, 'end': 3, 'type': '图书'}, {'word': '刘慈欣', 'start': 5, 'end': 8, 'type': '作者'}]
input sentence, please: 《活着》是一部由余华所著的小说。
[{'word': '活着', 'start': 1, 'end': 3, 'type': '图书'}, {'word': '余华', 'start': 8, 'end': 10, 'type': '作者'}]
```

图2-6 测试示意图

## 2.3 图书馆数据整合

使用实体识别模型处理非结构数据之后，使非结构化数据结构化。之后需要对处理好的非结构化数据和结构化数据进行合并，然后对整合到一起的结构化文件进行清

洗去重，最终处理后得到 8251 条完整的图书馆数据。

### 3 图书馆知识图谱的构建

#### 3.1 图书馆知识图谱的构建流程

知识图谱的构建方式主要有自顶向上和自底向上 2 种。自顶向下是指直接获取结构化数据存储到数据库中，而自底向上是指进行相对复杂的数据处理之后再存储到数据库中，本文将图书馆知识图谱的构建流程，将采用自顶向下和自底向上相结合的方式来进行。

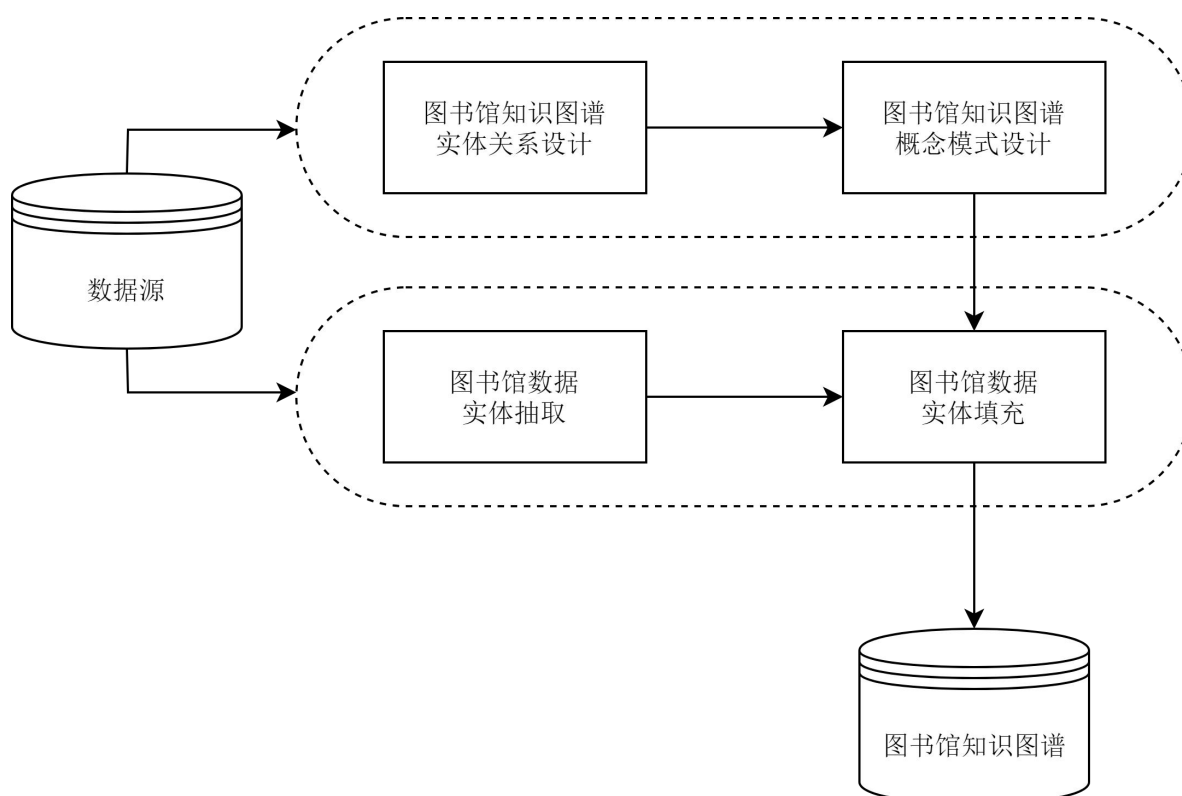


图3-1 图书馆知识图谱的构建流程图

图书馆知识图谱的构建，分为两大模块，如图 3-1 中图书馆知识图谱的构建流程所示。一是图书馆知识图谱的关系模式层设计模块，对图书馆知识图谱的服务类型和实体的关系进行设计，之后形成图书馆知识图谱的概念模式。二是图书馆知识图谱的生成模块，根据前期设计好的图书馆知识图谱的关系模式，将处理好的数据集中的实体进行填充，之后按顺序完成图书馆知识图谱的数据提取、数据融合和数据存储。在进行图书馆知识图谱的构建过程中，需要对网络获取的结构化数据处理之后直接存储到数据库中，属于自顶向下的构建方式，而非结构化数据则进行结构化处理之后存储到数据库中，属于自底向上的构建方式。图书馆知识图谱构建流程如图 3-1 所示。

### 3.2 图书馆知识图谱的设计

图书馆知识图谱的设计是为了将图书馆各类服务的数据整合在一起，以便更好地进行查询和分析。其中，作品实体是指图书馆中所有的图书，作者实体是指图书的作者等。在图书馆的知识体系中包含着作品、作者、出版社、区域等多个实体类别。每种实体类别都是为之后图书馆智能服务做数据支撑，所有的实体和实体以及之间的关系组成了一个复杂的图书馆知识图谱体系。

表 3-1 实体类型

实体类型	名称	数量
Book	作品	2139
Person	作者	1125
Category	分类	15
Date	出版时间	1879
score	评分	42
price	价格	321
Press	出版社	178
BookTitle	从书名	198
ISBN	ISBN	2139
State	状态	2
OpenHours	开放时间	1
Distribution	分布	51
Activity	活动	40
Mode	方式	10
Time	时间	10
Equipment	设备	38
Region	区域	16
Introduce	介绍	47

本文以图书为核心，以图书的作者、出版社、出版时间、分类等信息的组成关系为主，扩展其他图书馆相关的设备、开放时间、活动等实体，根据图书概念层次，可以确定图书馆知识图谱中实体的类别包括 7 类服务，18 种实体。具体数量如表 3-1 所示。

在设计知识图谱时，需要考虑实体之间的关系，将数据中重复的以及多余的数据去除，处理好图书馆知识图谱数据的冗余问题才能更好的提高系统的工作效率。例如：作品实体和作者实体之间的关系是作者和作品关系，同一个作者和书籍实体间出现两次关系的数据要去除。此外，需要考虑实体的属性和关系的属性，例如：作品实体的属性可以包括作品名称、作者实体的属性包括作者名称等。图书馆知识图谱以三元组方式表示知识，并且实体间可有多个关系，所有的关系共同构成图书馆知识图谱知识网络，例如，作品“三体”作者“刘慈欣”，作品“三体”状态“在馆”等，一个实体可以与多个实体连接。

对于图书馆知识图谱关系层的设计，需要考虑每种实体之间的关系，关系涉及到双向关系和单向关系两种。比如分类和介绍只需要知道分类的介绍是什么，而不需要知道介绍的分是什么，所以实体之间是单向关系；分类和作品之间需要双向都可以查到，所以实体之间是双向关系。根据图书馆知识图谱关系层设计的类别综合分析基于知识图谱的图书馆智能服务系统后整理出图书馆知识图谱实体关系，共包含 18 类实体，25 种关系。图书馆知识图谱实体间关系 E-R 图如图 3-2 所示。

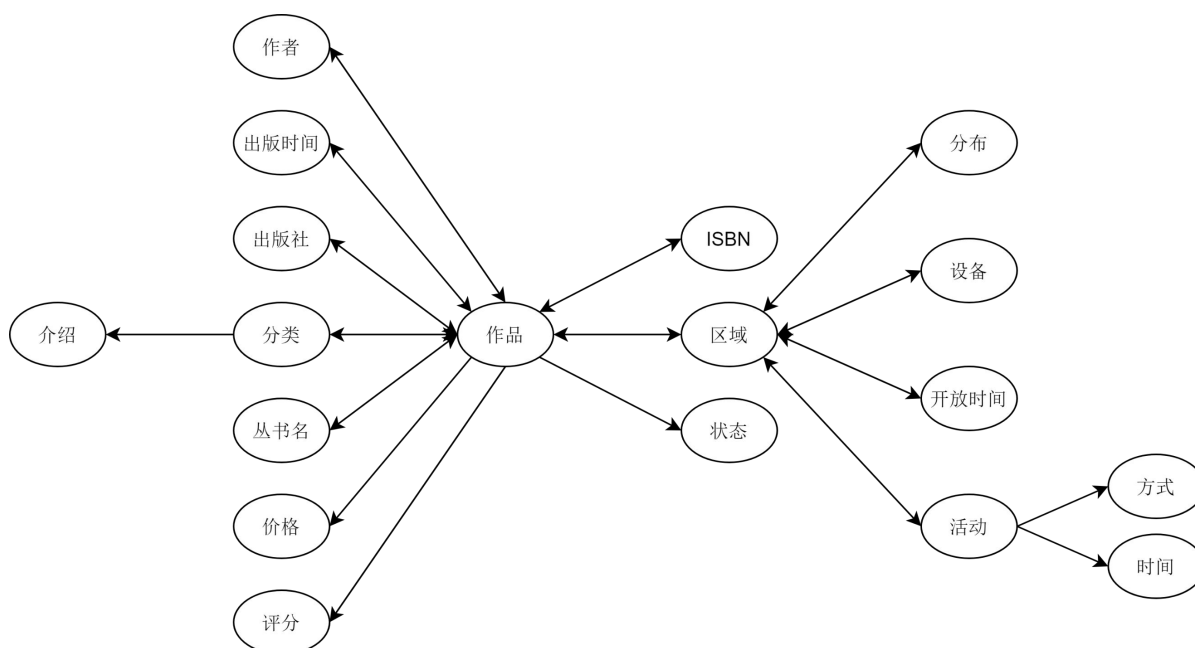


图3-2 实体关系E-R图

### 3.3 图书馆知识图谱的生成

#### 3.3.1 实体抽取

在构建知识图谱之前，需要从原始数据中提取实体信息，实体抽取在构建图书馆

知识图谱数据的过程中是一个非常重要的任务。由于获取的数据还包括非结构化数据。对非结构化的知识提取分为三个阶段：命名实体识别、属性提取和实体关系提取阶段。

首先，在命名实体识别步骤中，本文采用了基于 BERT 的命名实体识别模型，用于从原始文本中提取作品和作者等实体。该模型是基于预训练的模型，进行命名实体识别时利用编写的脚本代码对输入的文本进行识别，最终输出一些被标记好的实体。

其次，在识别出实体后，进一步使用词性标注的方式为每个词标注好属性。

最后，根据实体之间的关系来进行实体关系抽取，抽取他们的关系，通过语义角色标注和句法依存分析，得到实体之间的属性，前期训练的模型可用于分析短语各个组成部分之间的依赖关系。

通过这种方法，可以将图书馆的非结构化数据转换为结构化数据，并将其存储在结构化数据中。在实体关系提取阶段通过对自然语言依赖性的句法分析，得出每个词之间的关系，并得出三元组形式。

### 3.3.2 知识融合

本文对于抽取的实体整合成三元组形式之后，需要将所有三元组数据存储到知识图谱中。在本文中，采用了 Neo4j 图数据库来进行图书馆知识图谱的存储，构建方法采用脚本导入。将图书馆各类服务的实体和实体间关系以及实体属性存储在图数据库中，通过图数据库查询语句可以直观的对图书馆知识图谱的信息来进行展示和分析。

因此，本文以标准名称将其所有实体别名分类为子类别，以便于之后对关系的整合。知识融合包括知识表示、知识推理和知识更新三方面。知识表示需要把前期处理好的图书馆结构化数据进行转换，最后变成计算机可以识别的数据形式。本文选用图数据库，所以需要节点来表示实体，边则代表实体之间的关系，本文对实体进行了数据库中的节点和边的转换，所有的结构化和非结构化数据融合到一起共同组成了一个完整的用于图书馆知识图谱构建的数据集。知识推理则是要利用原先已知的信息来对新的知识进行推理而得出新的结论。知识更新则是对原有的图书馆数据集进行补充和完善的过程。

知识融合过程较为复杂，但是完成知识融合后才可以使构建的图书馆知识图谱更加完整和可靠。完成知识融合操作后，接下来需要对图书馆知识图谱数据进行存储，由于存储过程采用脚本文件进行导入，所以需要将处理好的结构化数据数据转换为 TXT 文件，为下一步的知识存储做准备，在转换的过程中需要注意的是，必须为每个实体设立标签类型，TXT 数据示例图如图 3-3 所示。

我生活的种种模式##赫尔伯特.A.西蒙(美)##作者##作品##作者##1  
 多样性红利##斯科特·佩奇(Scott E. Page)##作者##作品##作者##1  
 今天也要用心过生活##松浦弥太郎##作者##作品##作者##1  
 学习究竟是什么##万维钢##作者##作品##作者##1  
 人生模式##阳志平##作者##作品##作者##1  
 听听那冷雨##余光中##作者##作品##作者##1  
 曾国藩的正面与侧面##张宏杰##作者##作品##作者##1  
 谈修养##朱光潜##作者##作品##作者##1  
 经济人的末日：极权主义的起源##[美]彼得·德鲁克##作者##作品##作者##1  
 权力##[美]杰弗瑞·菲佛##作者##作品##作者##1  
 自我分析##[美]卡伦·霍尼##作者##作品##作者##1  
 创新者的窘境##[美]克莱顿·克里斯坦森##作者##作品##作者##1  
 动机与人格##[美]马斯洛##作者##作品##作者##1  
 瞬变##[美]奇普·希思[美]丹·希思##作者##作品##作者##1  
 如何有效阅读一本书##[日]奥野宣之##作者##作品##作者##1  
 麦肯锡教我的写作武器##[日]高杉尚孝##作者##作品##作者##1  
 沟通圣经##[英]尼基·斯坦顿##作者##作品##作者##1

图3-3 TXT数据示例图

### 3.3.3 知识存储

知识图谱的存储是整个系统中最关键的一环。在本文中，使用了 Neo4j 作为知识图谱的存储数据库。因为数据结构上知识图谱的数据存储在关系型数据库中并不合适，数据庞大关系复杂，使用关系型数据库构建和查询的效率很低。而 Neo4j 是一种高性能的图数据库，可以支持复杂的图查询和图分析操作，满足知识图谱的操作要求，比较适用于本文需求。Neo4j 导入数据的方式有使用 CSV 文件导入、编程语言导入、工具导入等等。具体使用哪种导入方式导入数据取决于数据量的大小和导入数据的长短，综合比较后本文采用编程语言的方式导入数据。使用编程语言驱动的方式导入数据的要求是图数据库已经打开在本地或者服务器运行。知识存储构建图书馆知识图谱主要是创建节点和边来组织数据，节点表示知识图谱中的实体，边表示知识图谱中实体间的关系，关系可以有方向，关系的两端表示开始节点和结束节点。

对于编程语言本文选择 Python 语言进行数据入库操作，主要利用 py2neo 库来对数据库操作，常用的方法有 Graph、Node、Relationship，也就是 Python 与 Neo4j 的接口，安装 py2neo 后可以使用 Python 操作 Neo4j 数据库。首先确保 Neo4j 数据库已经在运行，之后采用脚本的方式读取文档每一行数据后分割数据，接着匹配实体类型后生成创建语句创建实体及其属性和关系。节点的建立要用到 Node，建立节点时需要注意定义节



点的类型以及一个基本属性。节点间关系的建立用到 Relationship，节点间关系时有向的，建立关系时必须定义一个起始节点和一个结束节点，有种特殊情况关系的起始和结束节点时同一个节点时被允许的。如果建立关系的时候，指定的起始节点和结束节点不存在的时候，数据库会自动在建立关系的同时将起始节点和结束节点也建立。脚本读取完数据集文件的所有行之后存储结束。最终生成数据库实体和关系种类如图 3-4 所示。

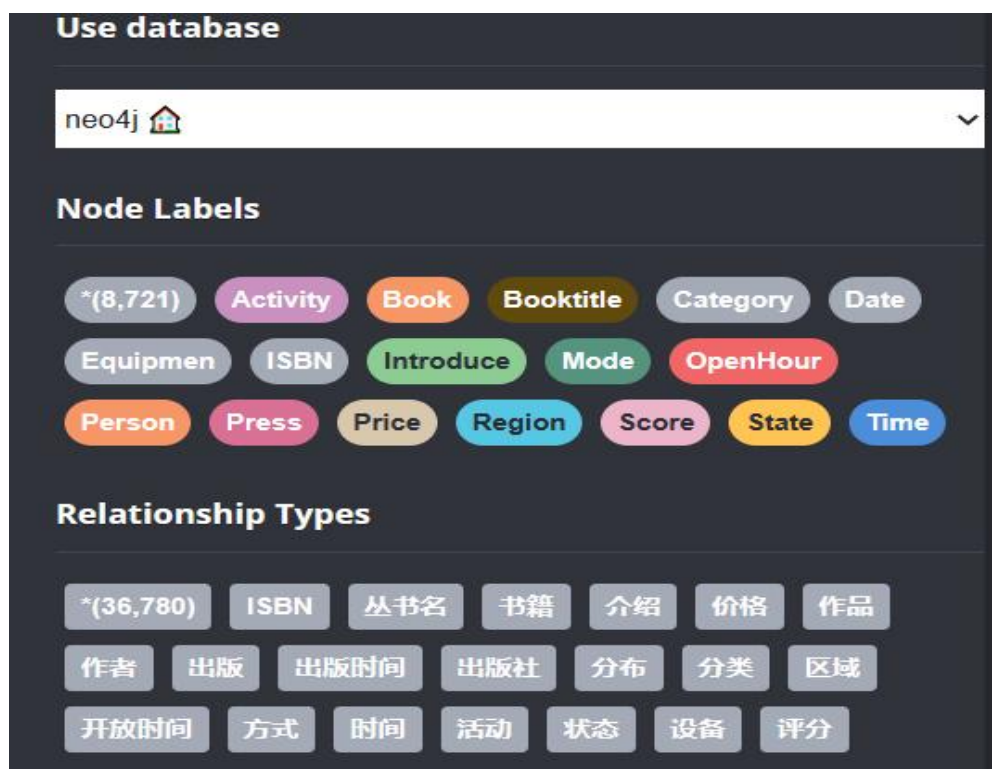


图3-4 数据库实体和关系

### 3.4 图书馆知识图谱可视化

在完成了数据导入之后，也就完成图书馆知识图谱的知识存储，适合本文图书馆的知识图谱构建完毕，接下来使用 Cypher 语句对其进行可视化。知识图谱的可视化是将知识图谱呈现给读者的一个重要方式。在本文中，采用了基于 Web 的可视化方案。本文构建的图书馆知识图谱节点较多，查询时候查询结果显示比较慢。因此，以图 3-5 为例，实体类别以颜色区分来展示部分知识图谱数据。图中展示了图书馆节点的相关连接关系，展示了部分图书、作者、分类等实体，详细展示了图书馆知识图谱的实体的关系及属性，不同颜色的圆圈表示不同种类的实体节点，节点和节点之间通过线条连接，线条上的标签用来表示节点间的关系。

本文对于图书馆知识图谱的可视化操作主要利用 Neo4j 查询语句来对图数据库进

行查询，如果查询实体对应关系的实体是什么可以采用  $\text{MATCH (n)-[:\%REL0\%]->(m)}$   $\text{WHERE n.Name = "\%ENT0\%" RETURN m}$  语句，语句的 REL0 表示实体间的关系，ENT0 表示给定的实体名称，查询时需要用实体和关系替换。查询之后返回的数据为查询到的实体在数据库中的名称，结果以图谱方式展现，由于系统设计包括图书查询界面，所以可以在系统开发完成后来进行可视化查询。

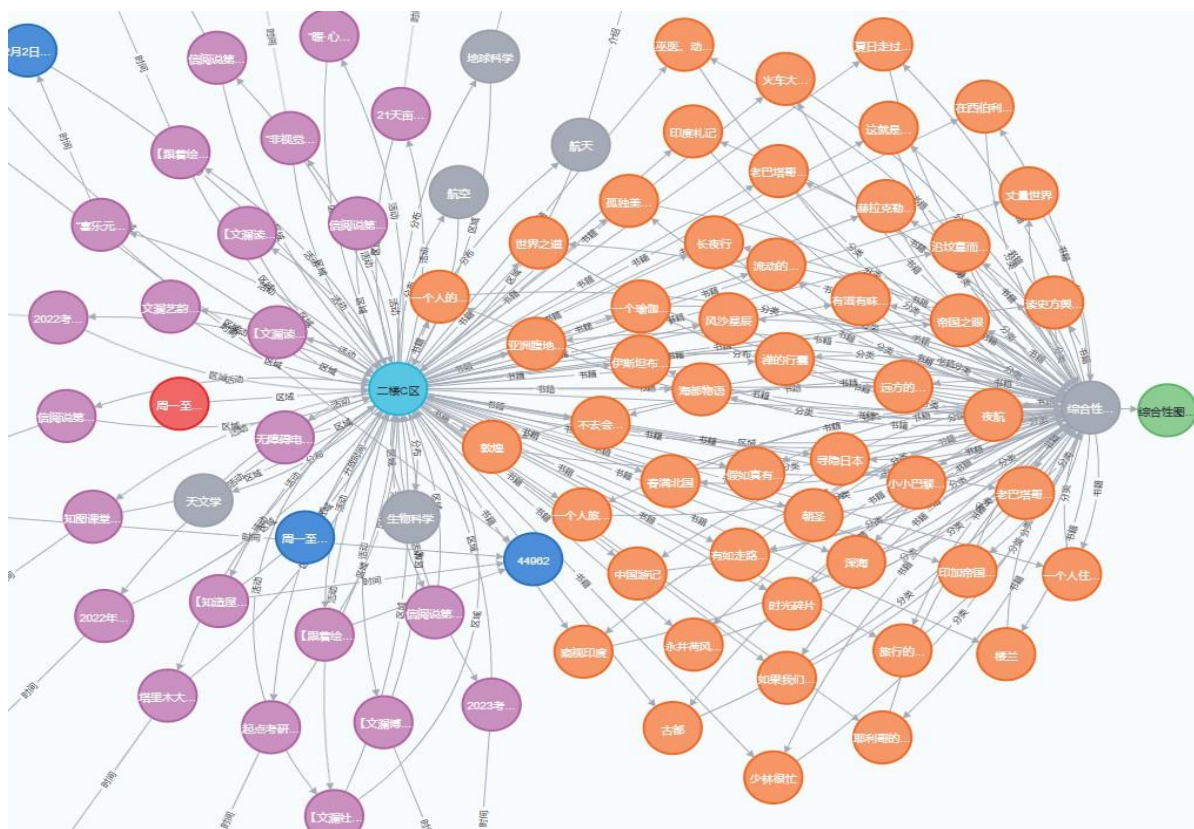


图3-5 图书馆知识图谱展示图

## 4 图书馆智能服务系统

### 4.1 需求分析

#### 4.1.1 功能需求分析

基于知识图谱的图书馆智能服务系统应该具备以下功能。

(1) 图书馆信息的搜索与展示：提供基于关键字、作者、出版社、书名等多种方式的图书信息查询，并在知识图谱中展示相关的实体和关系。

(2) 智能问答：利用自然语言处理和知识图谱技术，支持读者提出问题，系统能够基于图谱中的知识自动回答问题。

(3) 新闻公告：展示图书馆的通知文件以及活动新闻文件。

(4) 本馆概括：通过 Web 端展示图书馆的简介。

(5) 联系我们：图书馆的各种联系方式，方便读者与图书馆联系解决遇到的问题。

基于知识图谱的图书馆智能服务系统主要由自然语言处理模块、知识图谱构建模块、问题匹配和回答模块三部分构成。其中，自然语言处理模块用于将读者提出的自然语言问题转化为计算机可理解的形式，知识图谱构建模块用于构建图书馆的知识图谱，问题匹配和回答模块则利用知识图谱和机器学习等技术实现问题的匹配和回答。

问题匹配和回答模块主要包括问题匹配、答案抽取和答案生成三个步骤。问题匹配是指将读者提出的自然语言问题与知识图谱中的信息进行匹配，找到与问题相关的知识点。答案抽取是指从匹配到的知识点中抽取与问题相关的信息。答案生成是指根据抽取到的信息生成最终的答案，并将其展示给读者。

问题匹配是整个系统的核心步骤，其主要涉及到问句分类、实体识别、关系识别和问题相似度计算等技术。首先，问句分类技术可以将自然语言问题划分为不同的类型，如简单事实问题、定义问题、原因问题等。其次，实体识别技术可以从自然语言问题中识别出问题中涉及到的实体名称。再次，关系识别技术可以识别出问题中涉及到的实体之间的关系。最后，问题相似度计算技术可以将问题与知识图谱中的问题进行比较，找到与之最相似的问题，从而找到相应的答案。

答案抽取主要是从匹配到的知识点中抽取与问题相关的信息。这个过程主要依赖于关系抽取、实体抽取、属性抽取等技术。答案生成则是将抽取到的信息进行整合，生成最终的答案，并将其展示给读者。答案生成技术主要包括模板匹配、逻辑推理、自然语言生成等技术。

#### 4.1.2 其他需求分析

本文所做系统服务对象为想要咨询图书馆馆藏资源的读者。从读者的角度来看，读者在使用系统的时候一般来说对于系统的使用便捷性较为关注，而系统的内部运行流程不甚关心。本文面向读者的角度来进行设计，使系统的前端交互界面操作简易化。系统设计和实现的基于知识图谱的图书馆智能服务系统主要包括以下几个方面的功能。

(1) 图书推荐：根据读者的兴趣和需求，推荐适合的图书，同时提供相关的知识点和参考文献。

(2) 问答系统：读者可以通过输入问题，系统根据知识图谱中的知识点进行查询，并给出相应的答案和解释。

(3) 数据可视化：将知识图谱中的知识和关系进行可视化展示，方便读者了解和

掌握知识。

除了上述的功能需求，还需要考虑以下需求：

- (1) 读者友好性：系统应该简单易用，读者可以轻松找到所需信息。
- (2) 数据可靠性：图书馆的数据应该及时准确地更新，以保证知识图谱中的信息是最新的。
- (3) 系统性能：系统需要具备快速响应的能力，确保保证读者能够快速得到所需的信息。

## 4.2 系统架构

基于知识图谱的图书馆智能服务系统的架构主要分为 3 个层次，如图 4-1 所示。

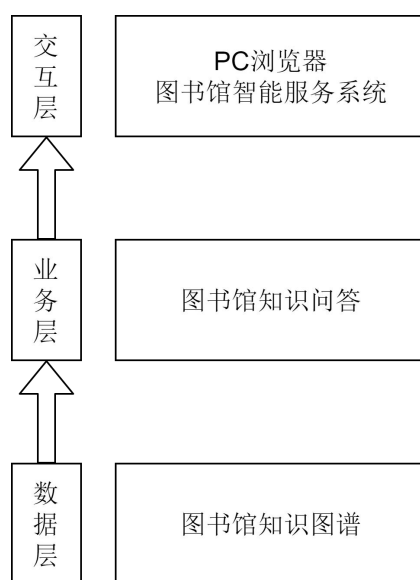


图4-1 系统的架构层次图

(1) 数据层指图书馆知识图谱，本文采用了图数据库来进行图书馆数据存储，将整理好的数据导入 Neo4j 图数据库为智能服务系统的问答等功能提供技术支撑。图书馆知识图谱包括图书信息、图书馆服务等 7 种类型 18 种实体数据，以及实体之间的关系数据。

(2) 业务层指图书馆知识问答，还包括其他智能服务业务。本文采用的方案首先是对问句进行分词处理后通过命名实体来进行实体的识别，再进行词性标注，得到打标签后的词语后加载问句模板来通过模板匹配的方法实现智能问答。利用的是双向最大匹配算法对问题集进行分词，查询语句使用构造的 Cypher 语句查询图书馆知识图谱数据库，返回结果。

(3) 交互层是指服务效果浏览器展示。本文采用 Flask 框架来实现可视化展示，

主要便于读者友好的使用智能服务系统，在前端方面设计图书信息查询、图书馆知识问答等功能，方便读者信息的使用。

基于知识图谱的图书馆智能服务系统的整体架构分为五个模块，分别是数据处理模块、知识图谱构建模块、后台处理模块、问答模块、前端展示模块，如图 4-2 所示。其中数据处理模块主要是对图书馆的原始数据进行清洗、过滤、分类等处理，为之后的知识图谱构建做好数据准备；知识图谱构建模块则是将实体及其关系构建为知识图谱，存储到 Neo4j 数据库中便于后台操作数据实现系统功能；问答模块主要是对图书馆中的实体进行抽取和识别，实现系统的问答功能；后台处理模块对前端数据处理返回前端结果；前端展示模块是通过图谱上的知识和自然语言处理技术实现读者问题的智能回答和图谱化展示。

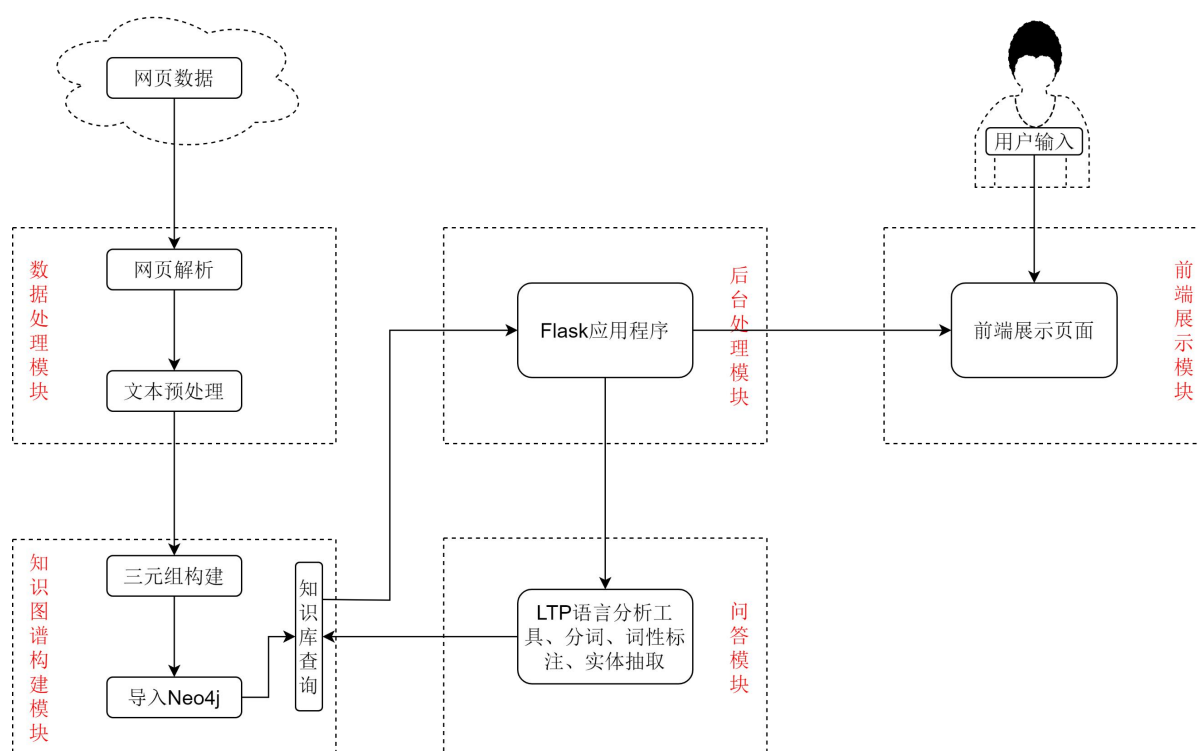


图4-2 系统整体架构图

## 4.3 系统模块

### 4.3.1 图谱化展示

为了提高读者的查询体验和效率，本系统设计了查询图谱化展示模块。该模块主要实现了通过读者输入关键词，展示与关键词相关的知识图谱，并且通过对知识图谱的可视化，为读者提供更加直观的查询结果。具体如图 4-3 所示。

该模块的实现主要包括以下步骤。

(1) 读者输入关键词：读者在系统中输入关键词，系统会自动匹配与该关键词相关的实体或属性。

(2) 查询知识图谱：系统根据读者输入的关键词，在图书馆知识图谱中查询与之相关的实体和属性，构建查询结果集合。

(3) 图谱可视化：通过图谱可视化技术，将查询结果集合可视化展示给读者。具体来说，该模块采用了常用的网络图谱绘制技术，为每个实体和属性分配不同的节点，并通过连线表示它们之间的关系。

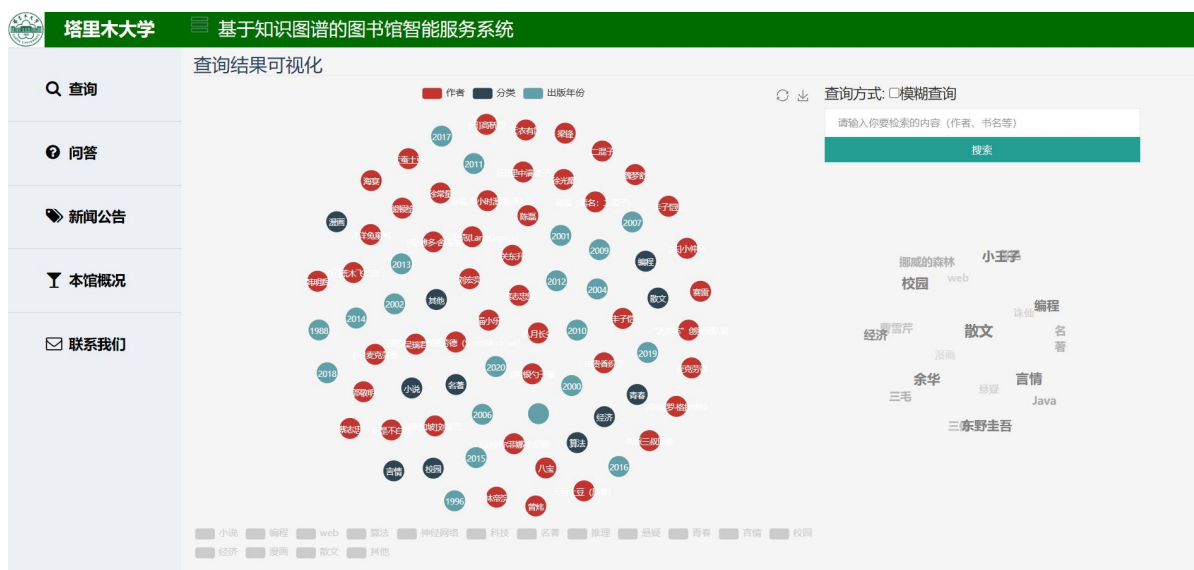


图4-3 查询模块展示图

### 4.3.2 问答模块

问答模块是本系统的重要功能之一，它能够根据读者提出的问题，快速、准确地返回问题的答案，具体实现如图 4-4 所示。

该模块的实现基于知识图谱技术和自然语言处理技术，具体步骤如下。

(1) 自然语言理解：将读者输入的自然语言转换成计算机可以理解的形式，即将读者的语言转化成计算机语言。该步骤主要使用基于深度学习的自然语言处理技术，包括词法分析、语法分析、命名实体识别等技术，将读者输入的自然语言转化为知识图谱上的查询语句。

(2) 查询图谱：系统根据读者输入的查询语句，在图书馆知识图谱中查询与之相关的实体和属性，构建查询结果集合。

(3) 答案生成：根据查询结果集合，生成针对读者提出的问题的答案。答案的生成主要基于知识图谱技术和自然语言生成技术，系统会将查询结果转化为自然语言，

输出给读者。



图4-4 问答模块展示图

本文开发的基于知识图谱的图书馆智能服务系统中的问答模块支持的问句类型多样，包括读者服务、借阅服务、开闭馆时间、馆藏分布、图书馆活动、图书馆设备、学科知识相关问题的问答。图书相关的具体的问答形式可以为图书的作者、图书的出版时间、图书的出版社、图书的丛书名、图书的价格、图书的评分、图书的 ISBN、图书的区域、图书的状态、作者的作品、出版社的出版、分类的书籍、区域的书籍、丛书书籍等。图书馆其他服务相关的具体的问答形式可以为设备的区域、区域的设备、区域的开放时间、活动的区域、活动的时间、活动的方式、分类的介绍、区域的分布、区域的活动等。

### 4.3.3 其他模块

基于知识图谱的图书馆智能服务系统的意义主要是提高服务质量和效率，满足读者需求，推动图书馆事业的发展，使图书馆更加智能化、信息化。系统的研究背景和意义都非常重要，它可以帮助图书馆实现更加精准、个性化的服务，提高读者的满意度和使用体验，推动图书馆事业的发展。因此，对于系统的模块设计，除了最重要的查询和问答模块之外，本系统也关注其他图书馆服务模块的设计还设计了新闻公告、本馆概括和联系我们三个简单模块方便读者使用。

新闻公告模块用来展示学校图书馆的通知公告、新闻动态、资源动态、培训与活动，可以方便读者及时了解图书馆活动的时间地点和内容，及时收到图书馆的放假通知以及周末开闭馆时间，避免读者前往图书馆后无法进入的情况发生，新闻公告模块

如图 4-5 所示。



图4-5 新闻公告模块展示图

本馆概括模块用来展示学校图书馆的简介，方便读者对图书馆的前世今生有所了解。首先介绍图书馆的建馆时间和历史和发展目标。其次介绍图书馆的收藏重点，以农科为优势，以生命为特色，结合其他学科形成多学科、多层次的综合性藏书体系。最后介绍图书馆的建筑面积、藏书数量、组成、服务等内容，页说明了图书馆发展目标。本馆概括模块如图 4-6 所示。



图4-6 本馆概括模块展示图

联系我们模块用来展示图书馆联系方式，对于各种问题的咨询方式做了详细介绍，读者可以选择到达现场咨询或者电话咨询等方式进行问题咨询。在图书馆遇到的问题



以及系统使用过程中遇到的问题都可以进行咨询。联系我们模块如图 4-7 所示。



图4-7 联系我们模块展示图

## 4.4 系统测试

### 4.4.1 系统准确性测试

在系统准确性测试中，本文选择了一些常见的图书馆问题，通过提问系统进行测试，如“三体的作者是谁？”、“二楼 A 区的开放时间是多久？”等。本文根据提读者的问题，从知识图谱中提取相关信息，生成答案并返回给读者。在测试中，本文考察了系统对于不同类型问题的回答准确性、速度以及健壮性等指标。实验结果表明，系统的回答准确率高达 85% 以上，响应速度在 3000ms 以内，满足读者使用要求。

### 4.4.2 综合性测试

为了进一步评估系统的性能，本文进行了综合性测试，测试主要包括邀请了一些图书馆读者，让他们通过系统提问并给出反馈意见。然后，本文从读者的角度评估系统的易用性、功能性、实用性等方面。

通过图书馆读者的测试和反馈意见，本文发现系统的准确性和速度得到了读者的高度认可，特别是系统能够快速、准确地回答读者的问题，解决了图书馆读者常见的疑惑和问题，提升了读者的满意度和使用体验。

## 5 结论与建议

### 5.1 结论

时至今日，基于知识图谱的各类新型智能服务模式已经有了长足的发展，并且在

人们生活的方方面面发挥着重要作用。本文立足实际，介绍了基于知识图谱的图书馆智能服务系统的技术路线，包括系统架构设计、核心技术和实现方法等。最终设计出一种基于知识图谱的图书馆智能服务系统，该系统可以帮助图书馆工作人员减轻工作压力，提高图书馆的服务效率和质量，为读者提供更加贴心细致的服务。

本系统以知识图谱技术为基础，结合自然语言处理技术进行开发。并在后续的开发中实现了系统界面可视化，完成了计划内任务，具体的开发内容如下。

(1) 图书馆数据集构建。本文对图书馆知识图谱的构建首先是通过豆瓣读书、图书馆官网、百度百科进行原始数据采集，得到结构化数据和非结构化数据，随后进行命名实体识别模型训练，最后使用训练好的模型对非结构化数据进行处理得到结构化数据。对处理好的结构化数据与未处理的结构化数据进行合并后得到图书馆知识图谱构建的完整数据集。

(2) 图书馆知识图谱构建。本文使用上述的图书馆数据集进行图书馆知识图谱的构建，首先根据本文特点来进行图书馆知识图谱的设计，尽量使结构便于后期系统的搭建。接着对知识图谱数据的生成做实体抽取、知识融合、知识存储的一系列操作。最后对生成的图书馆知识图谱使用 Cypher 语句进行可视化展示。

(3) 基于知识图谱的图书馆智能服务系统搭建。本文在上述任务完成的基础上来设计图书馆智能服务系统，后端使用 Python，数据库使用 Neo4j，前端使用 Flask 框架进行展示。本文设计的系统功能在智能问答的基础上增加了图书查询、新闻公告、本馆介绍，联系我们功能，可以更加便捷的服务读者，具备一定的实用价值。

本文通过系统测试和评估，证明了该系统具有较高的回答准确率、较快的回答速度和较高的读者满意度，在实际应用中表现良好，同时也说明系统的有效性和实用性，具有很好的应用前景和推广价值。或许在落地应用的情况下仍有诸多不足，但本文充分证实了该系统搭建的可行性，可为该方向提供一种思路。

## 5.2 建议

本文在系统搭建完成的同时，也注意到了系统中存在不足需要进一步完善来拓展其实用价值。

在图书馆知识图谱构建完成之后，后期需要考虑知识图谱的可持续扩充的问题，以及扩充的便捷性和稳定性，防止图书馆知识图谱在扩充后出现奔溃等。

对于图书查询和智能问答方面需要对回答准确率和响应时间的平衡做一定的调整，以期最短的时间内达到最高的准确率。

系统整体的稳定性是重中之重，需要防止因大量读者同时使用引起的大规模数据流导致系统奔溃也是后期急需解决的问题。同时，也需要对系统的跨平台及安全性等问题做出一定的弥补。

在今后的研究中，需要持续探索新型智能化图书馆服务上的创新性应用，对系统的功能以及性能作持续性的改进和优化。相信本文所提出的基于知识图谱的图书馆智能服务系统可以有更加广阔的发展空间，为新型智能化图书馆的实施提供专业技术支持，为图书馆智能服务提供更好的支持和帮助。

## 6 项目成果

论文：

1. 雷泽宇, 王志星, 席亚军. 数字图书馆知识图谱的构建[J]. 中文科技期刊数据库(全文版) 图书情报, 2023. (已录用)
2. 张云, 王志星, 席亚军. 图书馆智能服务系统的设计与实现[J]. 办公自动化. (已投稿)

软件著作权：

1. 席亚军. 基于知识图谱的图书馆个性化推荐系统. (已申请)

实用新型专利：

1. 席亚军. 可升降的图书馆还书车. (处于新案审查阶段)
2. 席亚军. 一种书籍分类信息读取设备. (处于新案审查阶段)

参与教材出版：

1. 常春, 席亚军. 信息组织[M]. 科学技术文献出版社, 2022.
2. 司莉, 席亚军. 信息组织原理与方法[M]. 武汉大学出版社, 2022.

## 论文:

- 雷泽宇, 王志星, 席亚军. 数字图书馆知识图谱的构建[J]. 中文科技期刊数据库(全文版) 图书情报, 2023. (已录用)

## 中文科技期刊数据库(全文版) 图书情报



主管单位: 科技部西南信息中心 主办单位: 重庆维普资讯有限公司 国内刊号: 50-9208/G 国际刊号: 1671-556X

### 录用通知书

雷泽宇 王志星 席亚军 同志:

您撰写的文章《数字图书馆知识图谱的构建》经专家和编辑审阅, 已被本刊录用。拟在《中文科技期刊数据库(全文版) 图书情报》2023年第08月发表。

特此通知, 此稿请勿他投!

查稿稿号: 1652625280579276800  
重庆维普资讯有限公司  
2023年05月06日

#### 重要提示:

- 为保证论文的刊发质量, 编辑部在不违背作者论文原意前提下, 有权对论文进行文字方面的修改(包括标题)。
- 作者文章文责自负, 本编辑部不承担任何连带责任。
- 本刊版权归重庆维普资讯有限公司所有。作者稿件一经录用, 均视为作者同意刊载以及同意在本刊合作的数据库及互联网站传播。如作者不同意文章被收录, 请在来稿时向本刊声明。

2. 张云, 王志星, 席亚军. 图书馆智能服务系统的设计与实现[J]. 办公自动化. (已投稿)

13:42

0.84 KB 5G 86

< 您好! 我的文... 共 13 封



马幕周社长 昨天

张云 ✓

作者同志: 文章编号: 10538, 您的论文《图书馆智能信息服务系统的设计与实现》已通过编辑部初审, 在送给专家外审前, 请您在5个工作日内务必做好以下事宜:

(1) 为了缩短稿件处理时间, 避免反复修改, 文章重复率不超过15%。

# 实用新型专利

## 1.席亚军. 可升降的图书馆还书车. (处于新案审查阶段)

**申请信息**

申请号/专利号	2023201220732	发明名称	可升降的图书馆还书车
申请日	2023-01-13	主分类号	B62B3/04
案件状态	新案审查	分案提交日	
主分类版本号	2006.01	副分类号	A47B63/00,B62B5/04,B62B5/06
副分类版本号	2006.01,2006.01,2006.01	分类日期	

**申请人**

姓名或名称	国籍或总部所在地	邮政编码	详细地址
塔里木大学	中国	843300	新疆维吾尔自治区阿克苏地区阿拉尔市塔里木大道东1487号

**发明人**

发明人姓名	席亚军,司春景,孙文彬
-------	-------------

## 2.席亚军. 一种书籍分类信息读取设备. (处于新案审查阶段)

**著录项目信息**

申请号/专利号: 2022232951000	发明名称: 一种书籍分类信息读取设备
申请日: 2022-12-09	主分类号: B65G 47/46
案件状态: 等待提案	分案提交日:

**申请人**

姓名或名称	国籍或总部所在地	邮政编码	详细地址
塔里木大学	--	--	--

**发明人/设计人**

发明人姓名: 席亚军、司春景、孙文彬
--------------------

**联系人**

**代理情况**

代理机构名称: 郑州明华专利代理事务所(普通合伙)	第一代理人: 叶朝辉
---------------------------	------------

**优先权**

**申请国际阶段**


**著录项目变更**



中国科学技术信息研究所研究生系列教材

# 信息组织

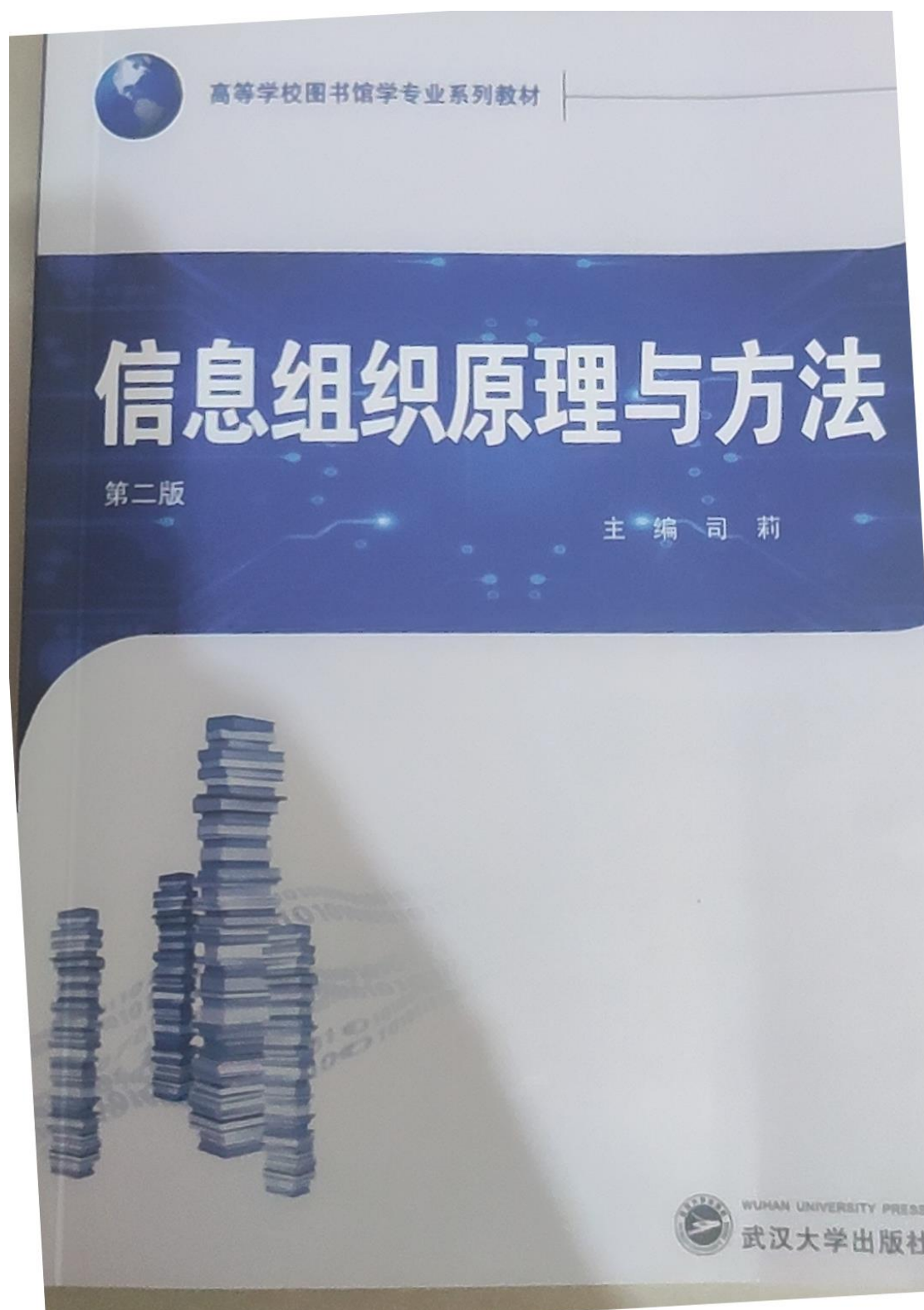
常 春 编著  
席亚军 副主编

 科学技术文献出版社  
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

· 北京 ·



2.司莉, 席亚军. 信息组织原理与方法[M]. 武汉大学出版社, 2022.





高等学校图书馆学专业系列教材

武汉大学规划教材建设项目资助出版

# 信息组织原理与方法

第二版

主 编 司 莉  
副主编 席亚军



WUHAN UNIVERSITY PRESS

武汉大学出版社

流水号

2023R11L0776833



受理签字: \_\_\_\_\_

审查签字: \_\_\_\_\_

## 计算机软件著作权登记申请表

软件基本信息	软件全称	基于知识图谱的图书馆个性化推荐系统				版本号	V1.0
	软件简称	图书馆个性化推荐系统				软件分类	应用软件
	软件作品说明	<input checked="" type="radio"/> 原创 <input type="radio"/> 修改（含翻译软件、合成软件） <input type="checkbox"/> 修改软件须经原权利人授权 <input type="checkbox"/> 原有软件已经登记 原登记号： 修改（翻译或合成）软件作品说明：					
开发完成日期	2023年03月30日						
发表状态	<input checked="" type="radio"/> 已发表 首次发表日期：2023年04月15日 首次发表地点：中国 新疆 阿克苏 <input type="radio"/> 未发表						
开发方式	<input type="radio"/> 单独开发 <input checked="" type="radio"/> 合作开发 <input type="radio"/> 委托开发 <input type="radio"/> 下达任务开发						
著作权人	姓名或名称	类别	证件类型	证件号码	国籍	省份/城市	成立/出生日期
	席亚军	自然人	居民身份证	132624198110026270	中国	河北 承德	1981年10月02日
	吴珍	自然人	居民身份证	652901197108196923	中国	新疆 阿拉尔	1971年08月19日
	司春景	自然人	居民身份证	13240219811002062X	中国	河北 承德	1981年10月02日

流水号

2023R11L0776833



权利说明	权利取得方式	<input checked="" type="radio"/> 原始取得 <input type="radio"/> 继受取得 ( <input type="radio"/> 受让 <input type="radio"/> 承受 <input type="radio"/> 继承 ) <input type="checkbox"/> 该软件已登记 (原登记号: ) <input type="checkbox"/> 原登记做过变更或补充 (变更或补充证明编号: )		
	权利范围	<input checked="" type="radio"/> 全部 <input type="radio"/> 部分 ( <input type="checkbox"/> 发表权 <input type="checkbox"/> 署名权 <input type="checkbox"/> 修改权 <input type="checkbox"/> 复制权 <input type="checkbox"/> 发行权 <input type="checkbox"/> 出租权 <input type="checkbox"/> 信息网络传播权 <input type="checkbox"/> 翻译权 <input type="checkbox"/> 应当由著作权人享有的其他权利 )		
软件鉴别材料	程序鉴别材料	<input checked="" type="radio"/> 一般交存: 提交源程序前连续的30页和后连续的30页 <input type="radio"/> 例外交存: 使用黑色宽斜线覆盖, 页码为: <input type="radio"/> 例外交存: 前10页和任选连续的50页 <input type="radio"/> 例外交存: 目标程序的连续的前、后各30页和源程序任选连续的20页		
	文档鉴别材料	<input checked="" type="radio"/> 一般交存: 提交任何一种文档的前连续的30页和后连续的30页 <input type="radio"/> 例外交存: 使用黑色宽斜线覆盖, 页码为: <input type="radio"/> 例外交存: 前10页和任选连续的50页		
软件功能和技术特点	硬件环境	开发: CPU类型: 酷睿 i5; 内存类型: DDR4; 内存容量: 8GB; 硬盘容量: 240GB; 运行: CPU: 1000MHz; 内存: 512MB; 硬盘空间: 200G; 宽带或者其他方式接入互联网		
	软件环境	开发: Windows 8 初级版bai (Starter) 64位操作系统; MyEclipse 运行: WINDOW 10以上操作系统, SQLSERVER 2000以上数据库; WindowsServer2003; IE7.0、Chromel10.0、Firefox4.0		
	编程语言	Java	源程序量	15061行
	主要功能和技术特点	开发目的: 协助用户进行基于知识图谱的图书馆个性化推荐管理      面向领域 / 行业: 图书馆管理 主要功能: 进行基于知识图谱的图书馆个性化推荐管理, 该系统能够实现的功能包括: 图书资料详细查找, 借阅证资料详细查找, 存档成功, 已借书籍续借处理, 借出与归还处理, 借出, 借阅入档等。软件采用导向型思路, 使用者可以轻松的对系统进行操作。      技术特点: 信息安全软件; 界面友好, 功能强大, 操作简便。软件同时提供多种辅助工具及系统安全维护系统, 用户还可以设定软件操作人员及操作权限。软件可以通过智能数据分析为有需要的用户或企业提升其工作效率和工作质量。		



流水号

2023R11L0776833



证书份数	<u>1</u> 份正本 <u>2</u> 份副本	
请确认所需要的计算机软件著作权登记证书副本数量。登记证书和副本数量之和不能超过软件著作权人的数量		
提交申请材料清单		
材料类型	材料名称	
申请表	打印签字或盖章的登记申请表	4页
身份证明文件	著作权人（1） - 席亚军的居民身份证复印件 著作权人（2） - 吴珍的居民身份证复印件 著作权人（3） - 司春景的居民身份证复印件 代理人乾升（重庆）知识产权服务有限公司的统一社会信用代码证书复印件	1页 1页 1页 1页
权利归属证明文件	合作开发合同或协议	2页
软件鉴别材料	程序鉴别材料 - 一般交存 文档鉴别材料 - 一般交存	60页 18页
其他材料		



## 参考文献

- [1] 黄永文,孙坦,赵瑞雪,等.大数据与人工智能背景下新型知识服务研究与实践[J].图书情报工作,2022,66(19):36-46.
- [2] 国务院印发《新一代人工智能发展规划》[J].广播电视信息,2017(08):17.
- [3] 吴建中.人工智能与图书馆[J].图书与情报,2017(06):1-5.
- [4] 施国良,谢泽宇,杨小莉.高校图书馆复杂网络构建与智慧化应用探索[J].图书情报工作,2019,63(23):106-112.
- [5] 王天明.智能化技术在图书馆服务管理中的应用[J].科技传播,2019,11(07):167-168.
- [6] 林梅芳.信息社会环境下高校图书馆读者服务转变方式[J].知识文库,2019(07):208+219.
- [7] 汪东升,游祎.我国图书馆领域人工智能研究现状与展望[J].图书馆研究与工作,2023(04):35-41.
- [8] Gartner.Gartner's Top 10 Technology Trends 2017[EB/OL].[2017-01-20].
- [9] 王展妮,张国亮.图书馆机器人应用研究综述[J].大学图书馆学报,2015,33(03):82-87.
- [10] 姚飞,纪磊,张成昱,等.实时虚拟参考咨询服务新尝试——清华大学图书馆智能聊天机器人[J].现代图书情报技术,2011(04):77-81.
- [11] 沈奎林,邵波,赵华.利用微信构建图书馆智能问答系统[J].图书馆学研究,2015(08):75-80.
- [12] 李丹.图书馆微信平台建设实践与思考[J].现代图书情报技术,2016(04):104-110.
- [13] 陆伟,戚越,胡潇戈,等.图书馆自动问答系统的设计与实现[J].情报工程,2019,5(02):5-16.
- [14] 郑晓霞.基于知识图谱的高校图书馆用户服务研究[J].图书情报导刊,2022,7(07):7-13.
- [15] 刘岍,李杨,段宏,等.知识图谱构建技术综述[J].计算机研究与发展,2016,53(03):582-600.
- [16] 冯志伟.自然语言处理的重要资源：“知识图谱”[J].外语学刊,2021(05):1-9.
- [17] 赵京胜,宋梦雪,高祥.自然语言处理发展及应用综述[J].信息技术与信息化,2019(07):142-145.
- [18] Tao Li, Xuhan Jin. Overview of Naming Entities Based on Natural Language Processing[J]. Academic Journal of Computing & Information Science,2022,5(1).
- [19] 袁里驰.基于 BERT-BiLSTM-CRF 的中文分词和词性标注联合方法[J/OL].小型微型计算机系统:1-9[2023-05-11].
- [20] 雷钧涵,杨毛佼,杜静,等.基于自然语言处理技术的智能问答平台构建[J].信息技术与标准化,2022(Z1):68-73.
- [21] 李飞.基于知识图谱的问答系统研究与实现[D].南京邮电大学,2022.
- [22] 王昊楠.基于知识图谱的中文图书问答系统的设计与实现[D].南京大学,2021.
- [23] 郭伟鹏,沈松雨.基于 BERT 模型的领域知识图谱构建研究[J].科技创新与应用,2022,12(36):16-19.
- [24] 徐增林,盛泳潘,贺丽荣,等.知识图谱技术综述[J].电子科技大学学报,2016,45(04):589-606.
- [25] 崔帅帅.基于深度学习的中草药知识图谱构建及应用[D].塔里木大学,2022.
- [26] 王玉宇,董广伟.基于 jieba 分词的网红图书馆问题与对策研究[J].江苏科技信



- [27] 息,2023,40(06):35-38.
- [28] 李冬梅,罗斯斯,张小平,等.命名实体识别方法研究综述[J].计算机科学与探索,2022,16(09):1954-1968.
- [29] 李东升,鲍玉来,刘建华,等.基于 BERT 的高校图书馆微信信息服务的命名实体识别方法[J].现代情报,2023,43(04):64-76.
- [30] Devlin J, Chang M W, Lee K, et al. BE R T: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. ar Xiv:1810.04805,2018.
- [31] AWAZ A, AKRAM U, SALAM A, et al. VGG-UNET for Brain Tumor Segmentation and Ensemble Model for Survival Prediction [C]//2021 International Conference on Robotics and Automation in Industry (ICRAI). Rawalpindi: IEEE 2021:1-6.
- [32] 帅亚琦,李燕,陈月月,等.基于 BERT-BILSTM-CRF 的慢性支气管炎中医医案实体识别[J].现代信息科技,2023,7(05):145-148+152.