

**本科生毕业设计[论文]**

**基于内容的中文期刊投稿推荐系统设计与实现**

院 系管理学院

专业班级 信息管理与信息系统1802班

姓 名李永涵

学 号U201815963

指导教师龚晓光

2022年5月26日

**学位论文原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于1、保密囗，在年解密后适用本授权书

2、不保密囗。

（请在以上相应方框内打“√”）

作者签名： 年 月 日

导师签名： 年 月 日

**摘 要**

对一些经验较为缺少的科研工作者而言，面临着众多的投递选择，有时难以挑选出一份最合适的科学刊物来发表他们的论文。投一份不合适的科学刊物，往往会导致拒稿、拖延或是文章发表后无人问津等情况。但是，对期刊投稿推荐的相关研究还比较少。

为了解决上述问题，本文设计并实现了中文期刊投稿推荐系统。该系统使用Java代码来实现基于内容的推荐算法，通过提取用户论文的标题、摘要和关键词的特征词来与论文库中的论文进行相似度计算，进而实现投稿期刊的推荐。相关数据的存储使用MySQL数据库，最终通过WPS加载项进行展示。用户只需要使用WPS加载项，系统即可向用户推荐适合投稿的中文期刊。该推荐系统分析将投稿期刊的关键内容与结构，与爬取到的期刊相关信息进行匹配，实现智能推荐。该系统基于论文内容，通过提取特征向量来建立推荐引擎从而进行推荐，通过分析文章的内容，向用户推荐适合文章投稿的中文期刊及相关期刊信息，以备用户投稿时进行参考。

**关键词：**基于内容的推荐系统；WPS加载项；中文期刊推荐

**Abstract**

For some inexperienced researchers, faced with numerous delivery options, it is sometimes difficult to select the most suitable scientific journal to publish their papers. Submitting to an inappropriate scientific journal often results in rejection, delays, or no-interest after the article is published. However, there are few related studies on journal submission recommendations.

In order to solve the above problems, this paper designs and implements a recommendation system for Chinese journal submissions. The system uses Java code to implement a content-based recommendation algorithm, and calculates the similarity with the papers in the paper database by extracting the feature words of the title, abstract and keyword of the user's paper, and then implements the recommendation of submitted journals. The storage of related data uses a MySQL database and is finally displayed through the WPS add-in. Users only need to use the WPS add-in, and the system can recommend Chinese journals suitable for submission to users. The recommendation system analyzes the key content and structure of submitted journals and matches the crawled journal-related information to achieve intelligent recommendation. Based on the content of the paper, the system establishes a recommendation engine by extracting feature vectors to make recommendations. By analyzing the content of the article, it recommends Chinese journals and related journal information suitable for article submission to users for reference when users submit articles.

**Key Words：**Content Based Recommendation System;WPS add-ins;Chinese Journal Recommendation

**目 录**

**[摘 要](#_Toc9875)** [I](#_Toc9875)

**[Abstract](#_Toc15694)** [II](#_Toc15694)

**[1 绪论](#_Toc30250)** [1](#_Toc30250)

[1.1 研究背景 1](#_Toc13854)

[1.2 研究目的与意义 2](#_Toc20885)

[1.3 主要工作 3](#_Toc23658)

[1.4 文章架构 3](#_Toc5805)

**[2 相关研究综述](#_Toc16461)** [5](#_Toc16461)

[2.1 推荐系统的应用与发展 5](#_Toc16421)

[2.2 基于内容的推荐算法 5](#_Toc30891)

[2.3 其他相关推荐算法 7](#_Toc30891)

**[3 中文期刊投稿推荐系统的设计](#_Toc15363)** [10](#_Toc15363)

[3.1 引言 10](#_Toc13534)

[3.2 系统功能分析 10](#_Toc7838)

[3.2.1 用户系统 10](#_Toc27415)

[3.2.2 投稿信息推荐 11](#_Toc5229)

[3.2.3 系统的后台管理 11](#_Toc10329)

[3.3 架构设计策略 11](#_Toc18497)

[3.4 数据库设计 12](#_Toc29493)

[3.4.1 E-R图及关系模型 12](#_Toc14659)

[3.4.2 数据库表设计 13](#_Toc7566)

[3.4.3 数据库备份策略 15](#_Toc27582)

**[4 推荐系统的搭建与测试](#_Toc29864)** [16](#_Toc29864)

[4.1 引言 16](#_Toc4457)

[4.2 测试环境介绍与搭建 16](#_Toc7903)

[4.3 系统架构 16](#_Toc27839)

[4.3.1 数据库连接 17](#_Toc31750)

[4.3.2 推荐引擎 17](#_Toc4205)

[4.4 UI模块与用户功能 18](#_Toc17844)

[4.5 对推荐方法的测试 20](#_Toc25310)

[4.5.1 推荐系统算法 20](#_Toc22285)

[4.5.2 测试数据来源 23](#_Toc20329)

[4.5.3 测试结果和分析 24](#_Toc32715)

[4.6 本章小结 25](#_Toc6576)

**[5 总结与展望](#_Toc9362)** [26](#_Toc9362)

[5.1 研究总结 26](#_Toc11563)

[5.2 研究展望 26](#_Toc16080)

**[致谢](#_Toc20994)****[28](#_Toc20994)**

**[参考文献](#_Toc8379)****[29](#_Toc8379)**

**1 绪论**

**1.1 研究背景**

由于网络信息的日益扩增，所有领域的科研工作者都面对着网络中已经出现的信息超载问题。这个问题使得使用互联网资源的人很难快速找到所需的信息。

论文的编辑与投稿对于科研工作者而言是十分关键的，科研工作者需要利用论文来对科研成果加以总结、记载、传播。在实际的科研工作中，研究人员需要在论文的撰写过程中消耗大量的精力与时间。在撰写完成后，还需要按照排版要求进行内容排版，随后还要进行投稿。

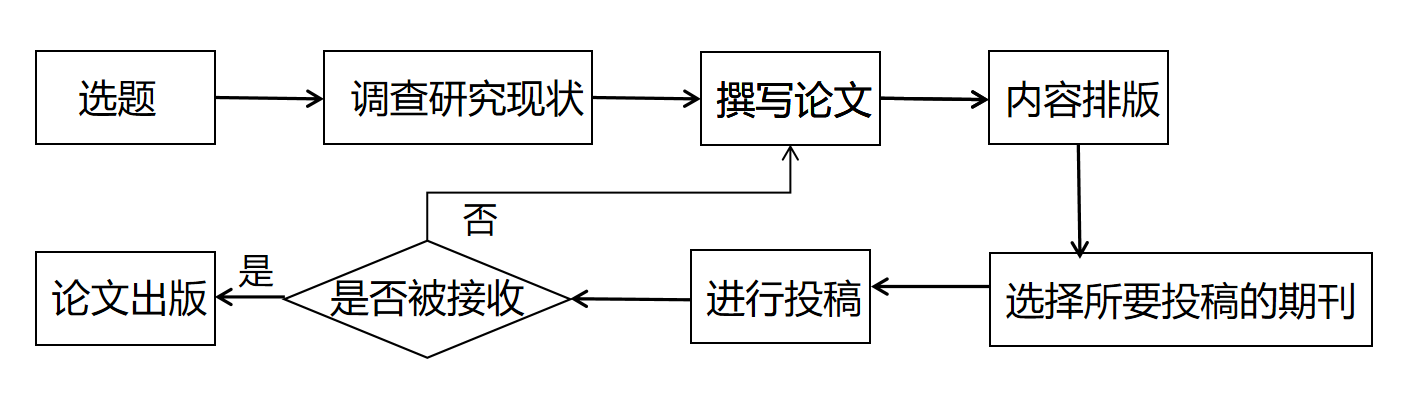


图1-1 期刊论文诞生流程

通常的解决方式是通过做索引擎查找所需要的信息，例如寻找所要投稿的期刊时，手动搜索期刊信息进行比对，自行凭经验进行投稿期刊的选择。互联网搜索引擎按照不同的需要，选用了相对应的效率最高的算法，在网络上自动地获取所需要的信息。再将所获取到的信息加以整合和汇总，并且按照信息的类别为客户提供相应的信息查询服务，最后再按照需要将所检索的相应信息内容进行呈现。首先客户提供关键词，而后互联网搜索引擎根据关键词，在海量的数据结果中有效地检索与这个关键词高度关联的数据信息并成为检索结论，其有效地解决了由于数据信息量过大而导致用户无法正确得到所需要的信息内容的问题。不过互联网搜索引擎也有缺点，如其搜索结果往往十分依赖客户提供的关键词。如果用户提供的关键词可以正确地表示其需要检索的目标特征，则系统就可以反馈比较准确的有关信息;但是，如果用户所提供的关键词不可以表示其需要检索的目标特征，或者出现了不相应的关键字，则系统也就无法了解用户的需要，也就无法提供比较正确的有关信息。在很多情况下，搜索引擎需要人工启发，推荐系统则能够主动推送，更加便捷、智能。

推荐系统属于信息过滤系统，基于数据分析和数据挖掘技术，在数据分析和挖掘技术基础上构建，根据从使用者中所获取的信息，向使用者提出可能感兴趣的信息。相比于传统搜索引擎，推荐系统并不需要使用者直接输入关键字，其通过调查使用者的兴趣爱好及倾向，进而开发其兴趣爱好点，进而促使用户找到自身需要的信息。因此个性化推荐系统也变成了现今信息时代检索信息的首选方式。

综上所述，当研究者们把研究汇总为学术论文，向外发表时，可能会发现大部分的刊物与其研究领域不相关，或是相应的刊物选择过多。怎样迅速选定一种适合的刊物，对研究者投稿十分重要。所以，研究者们需要一种个性化的中文期刊投稿推荐系统，以协助他们挑选理想刊物。而借助中文期刊投稿推荐系统，用户们能够节约搜索投稿信息的时间。

**1.2 研究目的与意义**

推荐系统在现今已成为了协助人们处理信息量的快速增长问题的一种很好的解决方案，该技术最早是希望用于给用户生成意见，例如推荐歌曲、影片和商品等。不过，推荐系统也同样对科研人员的专业科研活动非常关键。在科研工作中，撰写论文是每位科研人员的重要工作之一，每位科研人员都期望能够把论文投到合适的地方。把论文投到最理想的刊物有赖于如下一些因素:刊物的知名度、刊物研究的内容、所投论文的题目类型等。

对在科研领域中有着大量的投稿经历的专家研究者和资深研究者们而言，综合分析投稿刊物的各项信息，进而挑选最适合的科学刊物是比较容易的，因为他们都很了解什么刊物才适合他们的论文。不过对很多经验较为缺少的科研工作者而言，由于他们并没有那么多的投稿经历，且也不能够准确了解新兴的科学刊物，面临着众多的投递选择，怎样挑选一份最合适的科学刊物来发表他们的论文相对比较困难。投一份不合适的科学刊物，往往会导致拒稿、拖延或是文章发表后无人问津等情况。但是，通过已有的论文信息来推荐投稿期刊的相关工作却不多。为期刊投稿设计一款推荐系统，能够协助用户挑选最合适的中文期刊进行投稿。该推荐系统分析将投稿期刊的关键内容与结构，与爬取到的期刊相关信息进行匹配，实现智能推荐。该系统基于论文内容，提取特征向量来建立推荐引擎从而进行推荐，通过分析文章的内容，向用户推荐适合文章投稿的中文期刊。在进行期刊投稿推荐时，会对期刊论文进行文本分析。这也为发现某期刊或某领域的研究热点提供了技术参考。

通过该系统，可以为用户投稿提供参考，降低投稿到不合适期刊的可能性。在遇到拒稿问题时，也可以通过该系统找到适合文章投稿的多个期刊。系统会给出这些期刊的相关信息，方便用户进行多方比对从而提高投稿的成功率。

**1.3 主要工作**

主要的研究内容为：

1. 解决采用什么论文信息，如何提取论文信息，对提取的论文信息应当做出如何的处理的问题，主要使用了TF-IDF算法来提取关键词并且通过Key-Value对来计算相似度。
2. 如何设计并实现系统的各个组成模块，主要设计并实现了基于WPS加载项的前端界面和通过java实现的推荐引擎。
3. 设计了三种测试方法对推荐结果进行了测试。以下是对于这些工作的介绍：

面对一篇待投稿的论文，其标题、摘要和关键词是明显区分其与其他论文的标识，通过将这些论文属性罗列出来就能得到初步的论文特征集合，并将论文特征集合写入数据库中。而这些论文特征都是文本信息，所以要提取代表文本信息的最小语素。因此使用了开源的分词工具提取文本信息的TF-IDF值，得到key-value对便于计算用户文章与论文库文章的相似度。对于相似度的计算方式，要进行多项设计并进行测试，去提升推荐的正确率。

**1.4 文章架构**

本论文共分为五个部分，具体内容如下：

第一部分为绪论。阐述论文研究背景、目的与意义，并介绍本文的主要工作。

第二部分介绍了本文应用的相关研究综述。对国内外有关推荐系统的研究文献进行总结。介绍了主要的技术，包括垂直搜素技术、基于内容的推荐系统、TF-IDF。

第三部分是本文工作的重要一环，设计了基于内容的中文期刊投稿推荐系统的。该部分的主要内容有：分析推荐系统所应具备的功能、设计推荐系统的架构以及对数据库的设计策略。

第四部分介绍了中文期刊投稿推荐系统的实现。包括开发环境介绍与搭建、基于内容的中文期刊投稿推荐系统的实现和UI界面的实现。随后对基于内容的中文期刊投稿推荐算法进行了三种设计并进行了测试。

第五部分总结了本文所做的工作，提出了本文所设计、实现的推荐系统的不足之处，以及改进方向，和对未来可以改善的工作做出的展望。

**2 相关研究综述**

**2.1 推荐系统应用与发展**

推荐系统能够主动推送，具有便捷、智能的特点。推荐系统属于信息过滤系统，基于数据分析和数据挖掘技术，在数据分析和挖掘技术基础上构建，根据从使用者中所获取的信息，向使用者提出可能感兴趣的信息。相比于传统搜索引擎，推荐系统并不需要使用者直接输入关键字，其通过调查使用者的兴趣爱好及倾向，进而开发其兴趣爱好点，进而促使用户找到自身需要的信息。因此个性化推荐系统也变成了现今信息时代检索信息的首选方式。

推荐系统在现今已成为了协助人们处理信息量的快速增长问题的一种很好的解决方案，该技术最早是希望用于给用户生成意见，例如推荐歌曲、影片和商品等。不过，推荐系统也同样对科研人员的专业科研活动非常关键。在科研工作中，撰写期刊论文是每位科研人员的重要工作之一，每位科研人员都期望能够把论文投到合适的地方。把论文投到最理想的刊物有赖于如下一些因素:刊物的知名度、刊物研究的内容、所投论文的题目类型等。对在科研领域中有着大量的投稿经历的专家研究者和资深研究者们而言，挑选最适合的科学刊物比较方便，因为他们都很了解什么才是适合他们论文的投稿刊物。不过对很多经验较为缺少的科研工作者而言，由于他们并没有那么多的投稿经历，且也不能够准确了解新兴的科学刊物，面临着众多的投递选择，怎样挑选一份最合适的科学刊物来发表他们的论文相对比较困难。此时，推荐系统就能为他们进行投稿推荐。

**2.2 基于内容的推荐算法**

信息检索的研究成果除垂直搜素引擎外，还有推荐系统。许多日常生活与工作中的信息都是以文字为载体，于是基于内容的推荐算法便应运而生并得到了广泛应用。基于内容的推荐方法在推荐系统应用十分广泛，且十分契合对文字类资源的推荐[3]。根据用户过去的选择提出建议（例如，在以网络为基础的电子商务推荐系统中，如果用户过去观看了一些网络电影，推荐系统可能会推荐他尚未在本网站上观看的近期网络电影）。基于内容的推荐还使用来自推荐对象的内容生成推荐；因此，可以分析某些内容，如文本、图像和声音。通过这种分析，可以建立对象之间的相似性，作为推荐类似于用户购买、访问、听到、查看和排名的项目的基础。基于内容的推荐方法的应用情况如下：使用用户行为数据进行推荐[4]，Pennock将用户下载的论文作为参考指标进行推荐[5]，利用引文数据进行论文推荐[6]等。基于内容的推荐方式并不需要人工分析，其工作原理是针对目标资源提取出特征向量，对用户资源提取出用户画像，随后进行比对。宋钰将单词相似度匹配转化为基于语义的匹配，实现了对推荐方式的创新和优化[11]。王冬晖设计的论文推荐引擎使用TF-IDF技术来计算待推荐文章摘要的特征向量，然后使用Softmax回归对特征向量分类，根据分类得分来对适合该篇论文的相关刊物进行推荐[12]。在推荐结果的评价方面，车丰在推荐系统给出推荐结果后，将推荐结果通过微信平台让用户进行评价，这一方面有益于提升推荐系统的准确率，另一方面也能便于提升推荐结果的新颖值[13]。要一璐通过对文章关键词的分析构建知识脉络,并在此基础上计算了文章的相关性[14]。

由于基于内容的推荐算法要求必须有基本的用户信息输入和推荐目标的信息，因此没有有冷启动问题。基于内容的推荐算法主要有两个方面的工作，其一是对目标事物的内容特征提取与用户特征提取（用户画像）。推荐过程是通过推荐引擎，在数据库中找寻与用户特征最为接近的目标事物。因此，构建起高效、准确的用户画像尤为关键，这是推荐系统的基石。以电影推荐为例，每一部电影都有许多标签信息与评价信息，相似的电影其相关信息也是相似的。推荐引擎通过对相关信息的分析，从而建立起一个关于电影的用户画像，进而在电影库中找到与用户画像相似的电影，最终将电影推荐给用户。

建立起目标特征向量和用户特征向量后，需要通过统计分析二者之间的相似度，得到相似度评分，取相似度评分最高的目标信息作为推荐结果推荐给用户。而要对文字信息进行处理进而建立特征向量，则需要对文本内容进行特征值提取。

词频和逆文档频率(TF-IDF)是一种评价某个词在文段中的重要程度的计算方法，可以用来提取特征词，并对特征词进行评分。词频(TF)是指某个词在文段中所出现的频数，但是相同频数对于不同文段长度来说所体现出的特征词重要性不同，所以需要进行归一化处理。而逆文档频率(IDF)是指某个词在文本库中出现频数，如果文本库中有大量文段含有该词，那么该词区分文段的能力就很弱。TF-IDF值是两个评价指标之积，通过IDF指标判断出某个词并非常用语，同时通过TF指标判断出该词在这一文段中频繁出现，那么该词很有可能便能作为该文段的特征词。因此，TF-IDF值越高的词语，其在该文段中的重要程度也越高。

词频T是一个词语w在其所处文段d中出现的频数，L是该文段除去重复词过后的文本长度，TF值的计算公式如公式2-1所示：

(2-1)

文段的总数目为*D，*含有词语w的文档数为*Di*。逆文档频率IDF值的计算公式如公式2-2所示：

(2-2)

则**，**

(2-3)

本文主要使用的就是基于内容的推荐算法，通过计算特征词的TF-IDF值来进行相关度的计算，进而实现推荐。

**2.3 其他相关推荐算法**

论文数据的日益扩增,信息超载等问题日益突出[9]。杨杰通过对网络结构进行分析，去除噪声数据的影响[10]。推荐算法是推荐系统的引擎，也是推荐正确率的关键影响因素，下面介绍一些其他的推荐算法：

首先介绍的是基于规则的推荐算法Grundy[1]提出了基于规则的推荐方法，Beel将基于规则的推荐用在论文推荐算法中[2]。在论文推荐算法中使用基于规则的推荐，简单直观，但缺点是人工分类信息步骤繁琐以及推荐结果缺少针对性。

随后介绍的是基于协同过滤的推荐方法。协同过滤的底层逻辑是如果两个用户相似，那么为他们推荐的信息也是相似的。例如，在淘宝购物时，顾客A购买了a商品，顾客B也购买了a商品。同时A顾客还购买了b商品，这时，基于基于协同过滤的推荐方法便会将b商品推荐给顾客B。这与传统的基于内容的推荐方式有很大的不同，第一点是基于内容的推荐算法是分析两个物品之间本身的相似度，例如猪肉和牛肉这类的替代品，而基于协同过滤的推荐方法则需要分析用户喜好，用户的选择才是两个目标的相似之处，例如经典的啤酒和尿布案例。这种相似性通常是通过外在评价来进行的，因此会产生一些问题。其一，冷启动问题：首先，当一个推荐场景刚刚搭建起来时，系统库中并没有评价信息，也就无法分析用户喜好也无法进行有效的推荐。其次，当系统库中新加入了一项事物过后，同样没有该事物与用户直接的喜好程度信息，也无法进行有效的推荐。刘坤对推荐系统中冷启动问题进行了研究，提出了结合机器学习，找到冷用户的最近关联项的方法解决冷启动问题[15]。其二，系统的扩展能力不足，随着系统的不断使用，系统库中的用户喜好信息和事物的关联信息会非常庞大，这便会为算法的计算造成难度，影响了系统的运行效率。总而言之，得益于其对用户喜好的分析，基于协同过滤的推荐方法应用十分广泛，大部分电商网站，如淘宝、京东；视频网站，如B站、抖音、快手；以及一些文字类的论坛，如百度贴吧、知乎等，都采用了基于协同过滤的推荐方式。在本文所涉及的推荐相关论文和所要投稿的期刊的应用场景中，采用该方法存在着用户评分积极性不高，稀疏性问题[7]等难点。郭冬萌根据数据记录优化了稀疏矩阵[16]。吴逍遥通过对用户评分熵值差异性的比较，对数据稀疏的问题进行改进[17]。

然后是基于图的推荐方法。实际的应用场景中，如对商品进行推荐、论文的参考文献等具有多种属性的事物间存在复杂的网络关系图，而对网络关系图的分析则可以采用基于图的推荐方法。基于图的推荐方法的思路是建立起关系图，用户和待推荐的目标都是图中的节点，其中的关系便是连接各个点的线，通过对各条线赋予权重，就可以设计算法对其关联度进行计算。在论文推荐领域，主要研究方法分为链路预测和图嵌入两个方向[8]。孟伟龙针对现有论文推荐方法考虑的论文特征信息较少，导致论文推荐准确率较低的问题，研究构建一种四层图模型来整合论文、作者、主题、关键词信息以及它们之间的联系，对能体现论文信息的各个部分都进行了充分的应用[18]。刘玲，黄丽蓉，刘胜宗对比了五种相关性度量方法的推荐效果,论文间的相关性通过引用关系进行量化[19]。孟凡琦通过统一的图模型整合各项论文信息，来判断论文的相关性和影响力[20]。丁芳媛基于主题共现图实现对论文新颖性和影响力的评价，进而实现推荐[21]。

基于参考文献的论文推荐系统和基于论文内容的推荐系统所研究的方向大为不同。基于参考文献的论文推荐系统主要分析的是待投稿论文的参考文献，以及参考文献作者的合著关系以及研究方向来为用户提供相关论文的推荐[22]。这种推荐方法按照推荐范围的不同，可以分为全局推荐和局部推荐[23]。

还有基于标签的推荐算法，该算法通过专家评分对事物进行标签化处理，再依据标签内容，对事物进行相似度比对。其主要问题在于标签设计的合理性。孙见山通过分析科研社交网络中产生的标签信息进行推荐[24]。

这些年来，计算机硬件和软件功能都有了更加显著的特征，各种技术如大数据技术、机器翻译和语音文本处理技术日趋成熟，而这些技术也与推荐系统进行了有效的融合。王连臣基于深度学习推荐模型，研究出了新的推荐模型架构并且实现了在线上使用该模型对期刊和会议实行推荐[25]。

在进行论文推荐时，有时会采用多种方法共同推荐。钱宇结合了基于内容和协同过滤的推荐方法，并研究实现了专家推荐算法，以提升推荐的准确率[26]。栗欢结合了社交网络中的特性以及传统的推荐算法，形成了复合推荐的新算法[27]。刘宇,朱文浩通过对文章内容进行分析并提取标签、进行赋权，实现了混合推荐[28]。

在未来的研究中，可以采用多种推荐方式协同推荐，以提升推荐系统的推荐正确率。可以使用基于图的推荐算法，对文章的参考文献进行分析，建立关系图，通过对参考文献所处期刊进行分析，可以更有效地进行推荐。与基于协同过滤地推荐算法结合，还可以提高推荐结果地新颖性。

**3 中文期刊投稿推荐系统的设计**

**3.1 引言**

中文期刊投稿推荐系统的设计场景是用户已经撰写并排版好的文章，此时为用户推荐与用户文章契合的中文期刊。为了满足用户的需求，期刊论文投稿系统中主要包括用户系统模块、系统数据库管理模块以及投稿信息推荐模块。系统的设计采用基于内容的推荐算法为用户进行推荐。

**3.2 系统功能分析**

为用户推荐适合投稿的中文期刊是中文期刊投稿推荐系统的题中应有之义。其系统组成包括用户系统、投稿信息推荐和系统的后台数据库管理这三大组成部分，设计的中文期刊投稿推荐系统功能图如图3-1所示**。**

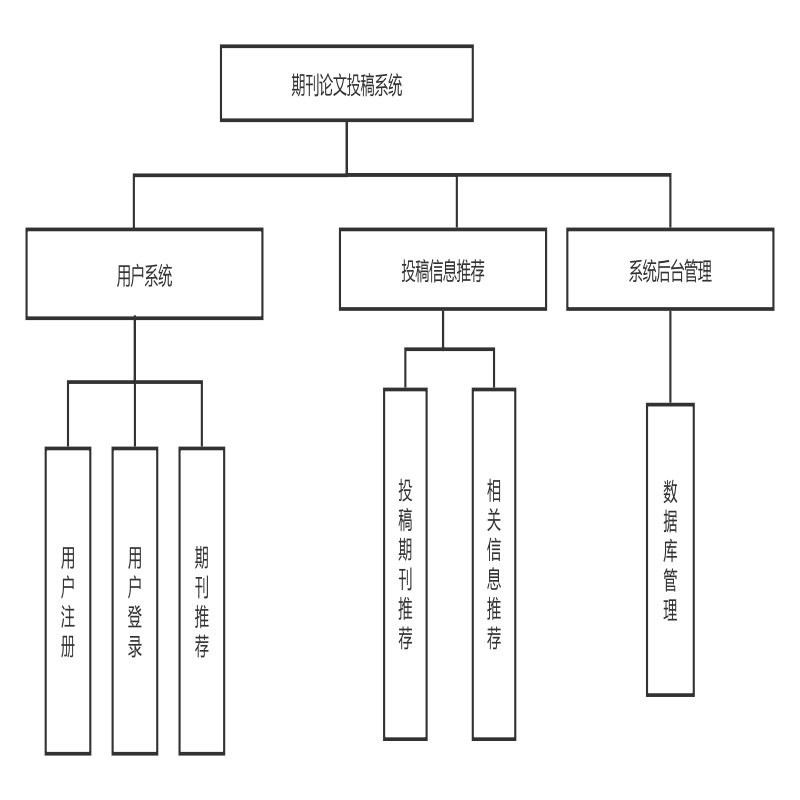


图3-1 中文期刊投稿推荐系统功能图

**3.2.1 用户系统**

用户系统作为用户与系统交互的界面，主要设计功能是方便用户使用推荐系统。该系统应该具有用户注册、用户登录和期刊推荐结果显示三个方面的功能。

用户使用中文期刊投稿推荐系统时，需要先进行注册，填写用户名和密码，随后进行登录，登陆后将会跳转到推荐界面。点击开始推荐，用户系统将会把用户已排版好的文章的标题、摘要和关键词写入数据库，并调用推荐模块。最终用户系统将会把推荐系统得到的推荐期刊从用户信息数据库中进行读取并显示。

**3.2.2 投稿信息推荐**

投稿信息推荐系统是是本文所设计系统的主要功能，应具有从数据库中读取已经提取好的论文信息并处理好后得出推荐的中文期刊并写入数据库的功能，是整个期刊推荐系统的后端。该模块包括投稿期刊推荐和相关信息推荐。

投稿信息推荐模块对用户文章的标题、摘要和关键词进行分析，找出论文库中与该篇文章相似度高的文章并进行统计，将相似文章所属期刊最多的期刊作为推荐结果进行推荐。

**3.2.3 系统的后台管理**

为整个系统进行后台管理，由于推荐结果的数据来源均为系统后台，因此要对数据进行更新维护。主要部分为数据库管理。数据库是系统的数据来源，数据库的管理包括定时备份以及通过爬虫方法更新论文库。

**3.3 架构设计策略**

期刊投稿推荐系统采用三层架构模式进行设计，B/S模式的网络结构，将系统划分为三个层次，即：数据层、逻辑层和数据存储层，具体如图3-2所示。

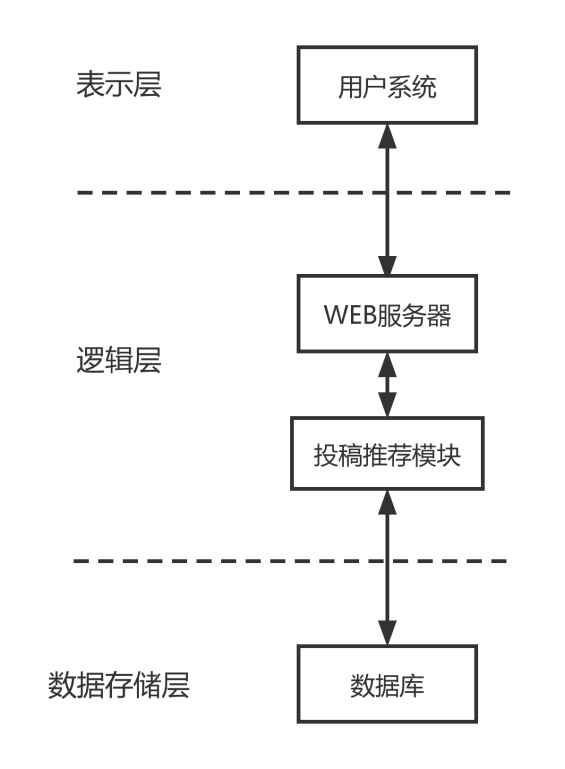


图3-2 系统结构示意图

表示层：用户系统位于表示层。中文期刊论文投稿推荐系统的前端界面采用WPS加载项来进行设计。作为整个系统的“门面”，其设计的要求是1.页面简单明确，可以很好地引导用户完成投稿期刊地推荐。2.功能完善，便于维护。3.页面尽量美观。其组成部分为系统注册页面、系统登录页面、推荐页面、期刊信息页面等，均为与用户关联的页面。

逻辑层：逻辑层是整个系统的“大脑”，包括web服务器与投稿推荐模块，是实现中文期刊投稿推荐功能的核心。其设计原则是：1.功能完善。2.尽可能提高运行效率。3.便于与表示层进行连接。

在本文中，逻辑层主要采用Eclipse IDE for Java Developers进行实现，其内容是实现中文期刊投稿推荐的主要功能，包括基于数据库内容进行推荐，以及查询页面的生成。

数据存储层：数据存储层采用采用MySQL数据库。其设计原则是：1.对网页爬虫所抓取的中文期刊论文信息数据，如期刊信息与每篇论文的标题、摘要和关键词以及从用户论文中提取的标题、摘要、关键词、特征向量等内容进行存储。这两部分数据要做到既要彼此分离，又要相互关联。为了保证系统数据的安全以及系统的正常运行，数据存储层需要一定的备份机制。

通过将系统划分为表示层、逻辑层和数据存储层这三个层次，各层次直接实现了逻辑上的分离，尤其是数据与显示实现了分离。层与层之间也做到了相互独立，当其中的一层发生改变时，只需确保接口的完整即可保障系统的运行，便于系统的敏捷开发和更新迭代。

**3.4 数据库设计**

**3.4.1 E-R图及关系模型**

通过分析，制定出E-R图如图3-4所示，关系模型为：

用户（用户id，用户喜好，用户名，密码）

提交论文（用户id，用户论文id）

用户论文（用户论文id，摘要，题目，关键词，推荐期刊名）

获取相似论文（用户论文id，期刊论文id，推荐结果）

论文（期刊论文id，摘要，题目，关键词，期刊名）

属于（期刊论文id，期刊id）

期刊（期刊id，期刊名，基本信息，出版信息，评价信息）

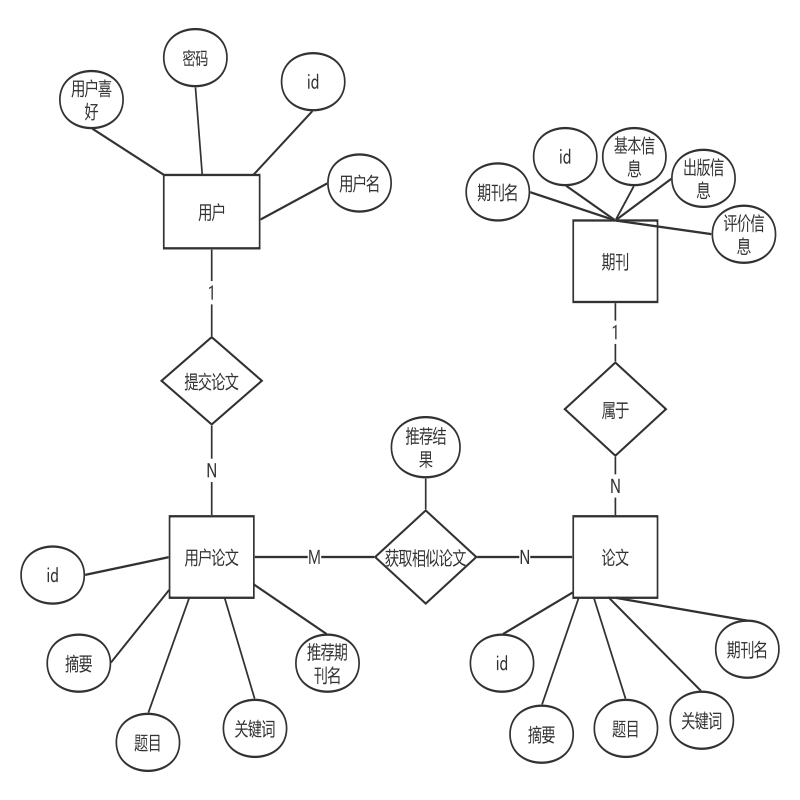
****

图3-3 E-R图

**3.4.2 数据库表设计**

Paper表记录了论文库中论文的id，摘要，关键词，标题以及论文所属的期刊。爬虫系统将这些内容爬取下来存储在这张表内，以供后续推荐系统使用。其中module\_id为外键。

表3-1 论文表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 名 | 类型 | 长度 | 是否非空 | 主键 | 注释 |
| id | BIGINT | 20 | 是 | 是 | 论文编号 |
| abstract | TEXT | 0 | 否 | 否 | 论文摘要 |
| kw | TEXT | 0 | 否 | 否 | 论文关键词 |
| title | TEXT | 0 | 否 | 否 | 论文标题 |
| module\_id | INT | 10 | 否 | 否 | 论文所属期刊id |

Paper\_logs表记录了用户提交论文的id，摘要，关键词，标题以及论文所属的期刊（默认为0，即待推荐论文集）。在WPS加载项中摘取所要进行推荐的文档的这些信息并存入该表，其中module\_id为外键，并为该字段建立索引。推荐系统将计算得出的Key-Value对写入pref\_list，并将最终推荐期刊名写入re\_module。

表3-2 用户上传论文表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 名 | 类型 | 长度 | 是否非空 | 主键 | 注释 |
| id | BIGINT | 20 | 是 | 是 | 用户上传论文编号 |
| abstract | TEXT | 0 | 否 | 否 | 用户上传论文摘要 |
| kw | TEXT | 0 | 否 | 否 | 用户上传论文关键词 |
| title | TEXT | 0 | 否 | 否 | 用户上传论文标题 |
| module\_id | INT | 10 | 否 | 否 | 论文所属期刊id（默认为0，即待推荐论文集） |
| re\_module1 | VARCHAR | 255 | 否 | 否 | 推荐期刊名1 |
| re\_module2 | VARCHAR | 255 | 否 | 否 | 推荐期刊名2 |
| re\_module3 | VARCHAR | 255 | 否 | 否 | 推荐期刊名3 |
| re\_module4 | VARCHAR | 255 | 否 | 否 | 推荐期刊名4 |
| re\_module5 | VARCHAR | 255 | 否 | 否 | 推荐期刊名5 |
| pref\_list | TEXT | 0 | 否 | 否 | Key-value对 |

Paper\_modules表记录了爬虫模块所爬取的所有期刊名和基本信息、出版信息和评价信息。待推荐的论文集id为0，不必记入此表。

表3-3 期刊表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 名 | 类型 | 长度 | 是否非空 | 主键 | 注释 |
| id | INT | 10 | 是 | 是 | 期刊编号 |
| name | TEXT | 0 | 是 | 否 | 期刊名 |
| fun\_inf | TEXT | 0 | 否 | 否 | 基本信息 |
| pub\_inf | TEXT | 0 | 否 | 否 | 出版信息 |
| eva\_inf | TEXT | 0 | 否 | 否 | 评价信息 |

推荐系统将推荐的论文编号，上传论文编号，推荐论文编号，推荐论文所属期刊编号，推荐的算法编号（便于后续设计的更新迭代）写入Recommendations表。其中paper\_id,paperlog\_id,module\_id为外键。

表3-4 推荐表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 名 | 类型 | 长度 | 是否非空 | 主键 | 注释 |
| id | BIGINT | 20 | 是 | 是 | 推荐结果编号 |
| paperlog\_id | BIGINT | 20 | 是 | 否 | 用户上传论文编号 |
| paper\_id | BIGINT | 20 | 是 | 否 | 推荐论文编号 |
| module\_id | INT | 10 | 否 | 否 | 推荐论文所属期刊id |

用户昵称和用户密码用于WPS加载项的用户注册和登录功能。

表3-5 用户表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 名 | 类型 | 长度 | 是否非空 | 主键 | 注释 |
| id | BIGINT | 20 | 是 | 是 | 用户编号 |
| name | VARCHAR | 255 | 否 | 否 | 用户昵称 |
| password | VARCHAR | 255 | 否 | 否 | 用户密码 |

**3.4.3 数据库备份策略**

考虑到数据库数据变化不大，备份策略为对数据库里面的每张表进行单独备份，每周进行一次差异备份，每月进行一次完全备份。同时，每月进行一次论文库的更新。

**4 推荐系统的搭建与测试**

**4.1 引言**

第三章介绍了中文期刊投稿推荐系统设计，在本章中将对该系统进行搭建并设计测试算法，测试不同算法的推荐准确率。

**4.2 测试环境介绍与搭建**

测试环境搭建时使用的软件环境如下：

1. Windows10,64位操作系统
2. Eclipse IDE for Java Developers - 2021-12
3. Java jdk 8.0
4. Apache-tomcat-9.0.55
5. MySQL 5.5.20
6. WPS Office 2019
7. Nodejs 16.14.0
8. Visual Studio Code 1.64
9. PHP 5.3.10

**4.3 系统架构**

整体开发的过程中，主要使用的开发工具是eclipse，采用maven web项目。采用MySQL数据库，作为系统的数据层，主要存储了用户数据库与论文数据库。在prom.xml文件的<dependency>标签中添加用于数据库连接与数据库方法调用的Jfinal框架，用于分词和提取用户特征和论文特征的ansj分词工具以及用于前后端交互的dwr框架。推荐系统的整体架构如图4-1所示。

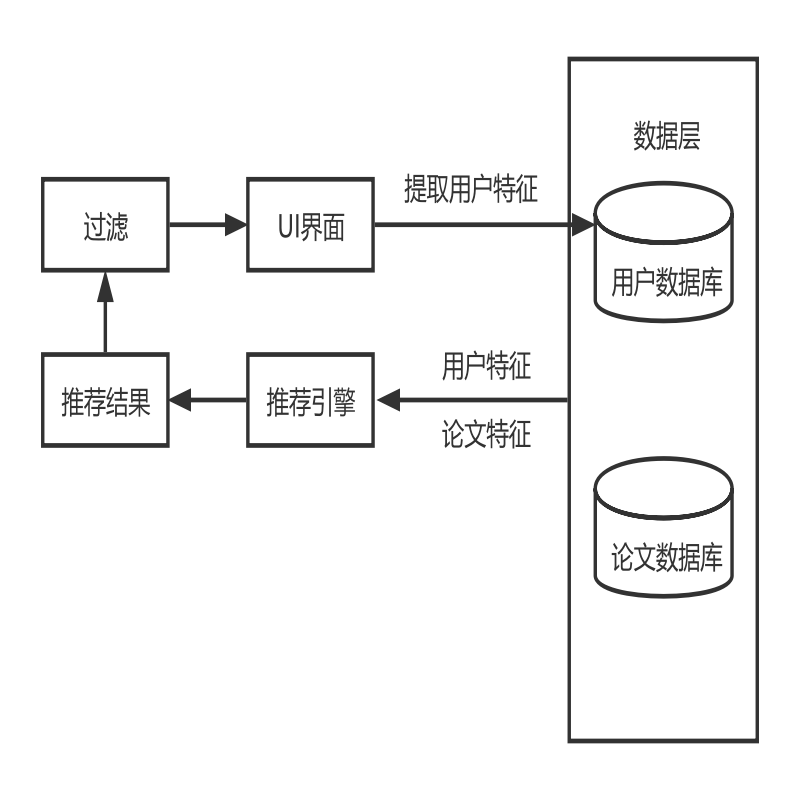


图 4-1 推荐系统的整体架构

**4.3.1 数据库连接**

通过Jfinal框架中的model与本地数据库进行交互。Model是Jfinal框架中最重要的组件之一，它充当MVC模式中的Model部分。通过继承Model，就可以马上拥有的众多便捷的操作数据库的方法。通过生成器，可自动生成Model、Base Model、Mapping Kit、Data Dictionary四类文件。抽象出的Base Model类中写明了数据库中各项的set和get方法。在Base Model各类中声明的dao静态对象方便查询操作。便于操作数据库与生成数据字典。

在实际使用时，在项目根目录下的res目录下，修改dbconfig.properties文件中有关数据库的配置，Jfinal框架会读取到对应数据库的表结构，实现与数据库的连接。

**4.3.2 推荐引擎**

推荐引擎通过读取数据库中论文库和用户论文库中论文标题、摘要和关键词

建立用户特征和论文特征计算相似度，对相似度进行排名得到的就是初步的推荐结果。将这一初步推荐结果写入数据库中的 recommendations表，将用户论文的关键词-TFIDF值对写入数据库中的paperlogs表。统计推荐论文所属的期刊，将同一期刊数最多的期刊名写入paperlogs表中。根据相应的推荐算法，对文章进行推荐。随后，通过对数据库的操作，统计recommendations表中这次推荐的论文所属期刊的论文数目，按论文数目对期刊进行降序排列。得到的结果如图4-1，图4-2所示：

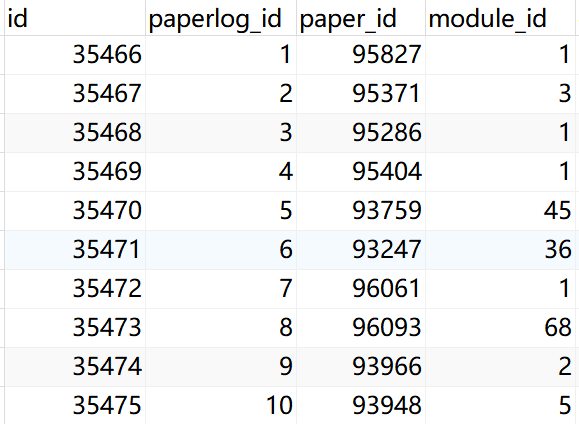


图4-1 recommendations表



图4-2 paperlogs表

**4.4 UI模块与用户功能**

前台展示页面部分使用jsp开发技术通过DWR框架与后台进行交互。

DWR（Direct Web Remoting）框架用于实现前端的web界面与后台java类的方法之间的交互。DWR工作原理是通过把Java类转化为JavaScript。

将dwr的jar包所对应的依赖写入pom.xml文件中。配置web.xml,写入dwr的servlet，在web.xml相同目录下创建dwr.xml文件并进行配置，主要就是将java代码中编写的业务类在dwr中进行注册，使之成为dwr的一个组件供dwr调用。

在加载jsp页面时，调用推荐引擎，实现推荐。在dwr.xml配置时，需要将要调用的java类写入，并声名转化为javascript后的名称，在网页端调用方法时，使用的是转化后的javascript方法。

用户界面主要用WPS加载项进行实现。WPS加载项(Add-ins)是一套用于扩展WPS应用的基于Web技术的解决方案。每个WPS加载项对应打开一个Web页面，通过调用Web中的JavaScript方法来完成它的功能逻辑。WPS加载项打开的网页可以与WPS应用直接交互，同时一个WPS加载项中的多个网页形成一个整体，数据可以相互传导。因为WPS加载项的底层是以 Chromium开源浏览器项目为基础进行的优化扩展，所以没有浏览器不兼容的问题。WPS加载项具备开发便捷、轻量、跨平台的特性，适配Windows/Linux操作系统。

本文使用WPS加载项作为UI界面，配合其他论文排版插件使用，使用户从撰写论文到论文排版到寻找期刊投稿这一过程全部都可以在WPS实现，十分便捷。其主要功能为：读取所要进行推荐的论文的标题、摘要、关键词并写入数据库，显示推荐系统所得出的推荐期刊以及用户登录、注册。在此主要介绍读取待推荐论文标题、摘要和关键词的方法。首先，需要用户对文档进行标准排版，即按期刊论文的标准对论文进行排版并将文档以标题命名。读取文件名以获取论文标题。摘要和关键词的提取需要对文章结构进行分析，以摘要为例：首先得到文档自然段数，然后对文章所有自然段进行遍历，通过Range.Text属性提取自然段内容，并通过判断段落开头字符是否为“摘 要”或“摘要”等字符进行筛选。通过PHP连接MySQL数据库，随后将用户论文的标题、摘要、关键词写入数据库。

用户使用时先进行注册和登录，随后进入推荐界面，点击开始推荐后，通过打开外部链接的方式打开推荐模块的推荐界面，实现对推荐模块的调用，推荐模块将推荐结果写入数据库，推荐界面再从数据库中将推荐结果读取并显示出来，推荐的系统用户用例图如图4-3所示。用户可以自行选择得到3个或5个期刊推荐结果（默认为3个）。在得到推荐的期刊后，用户可以点击相应的期刊，查看该期刊的基本信息。其运行界面如图4-4所示。

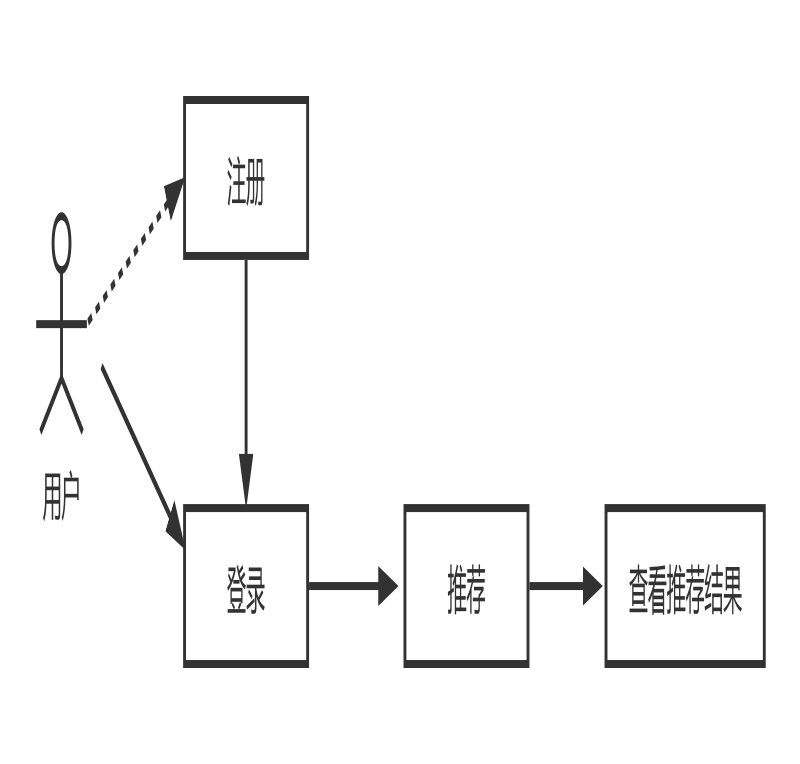


图4-3 推荐的系统用户用例图

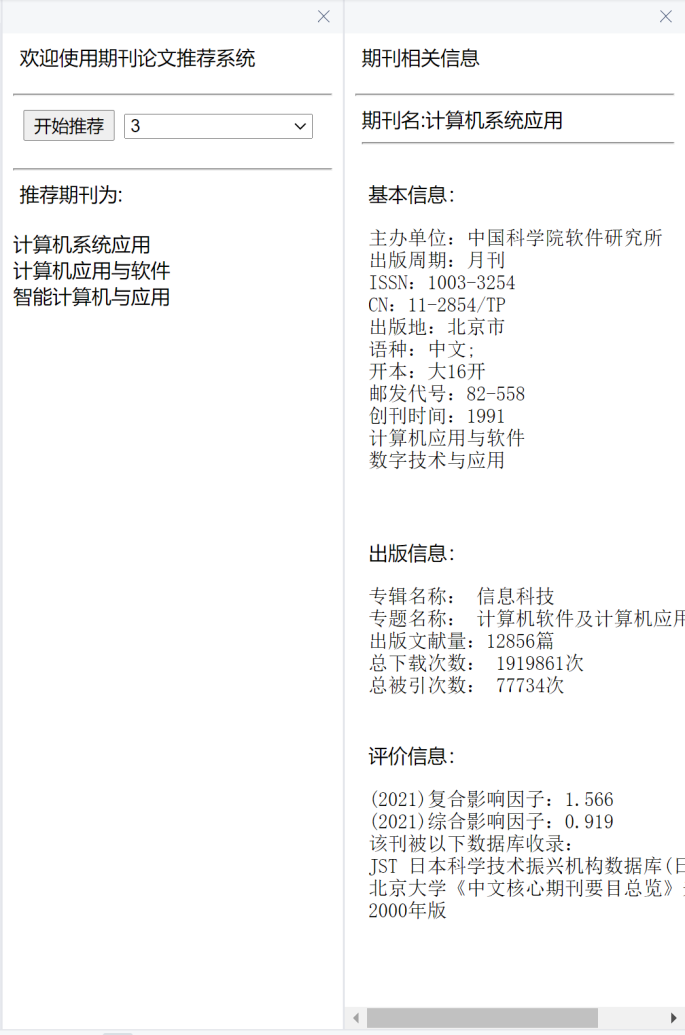


图4-4 推荐页面

**4.5 对推荐方法的测试**

**4.5.1 推荐系统算法**

投稿推荐系统工作原理主要为一方面从用户文章中记录文章信息，从而计算用户偏好，进而建立起用户模型。另一方面同样对论文库中的论文进行计算，建立起推荐对象模型。最后通过相似度计算，将推荐结果显示给用户。其主要流程如图4-5所示。

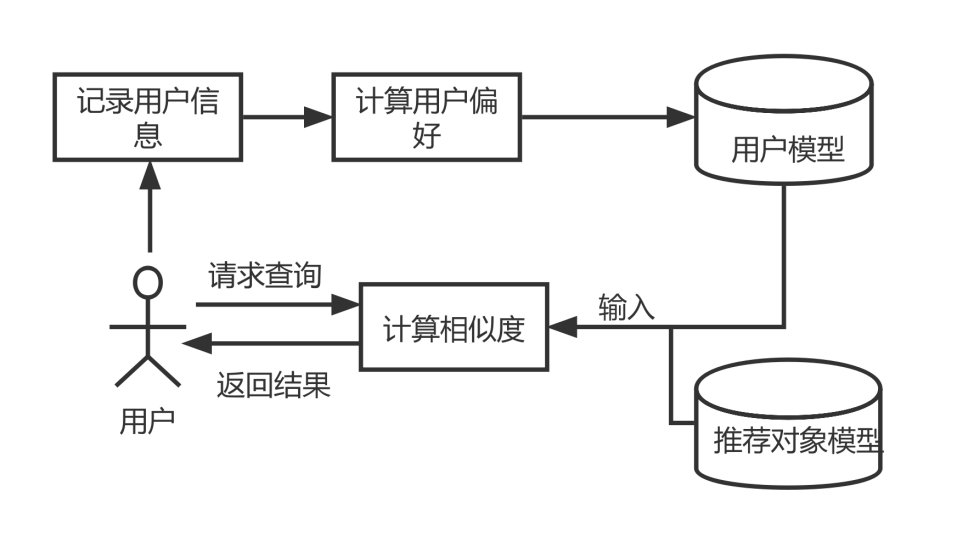
****

图4-5 中文期刊投稿推荐系统工作原理

而其中最重要的部分就是推荐算法，即在推荐引擎中对用户偏好，以及用户特征向量与论文库中论文特征向量相似度计算。由于论文信息主要以文字内容的形式，所以采用基于内容的推荐方法。即计算即将投稿的的期刊论文文本与其他已经投稿的期刊论文的相似性，得到与所投论文相似的已投论文集，随后将已投论文集中数量前三的期刊名作为推荐的期刊进行显示。

使用基于内容的推荐方式进行推荐有两个基本步骤。第一步是获取并处理用户的基础数据。为解决这个问题，可以将用户论文的题目、摘要、关键词摘录出来，并从中提取出能代表论文主要内容的特征词。

论文的题目、摘要和关键词都是文本属性，其最小语素是单个词语。所以其特征词便是能代表此文段含义的最小语素集。要得到这样的集合，就需要对文段进行分词处理。在分词处理的过程中，需要对停用词和部分常用词和无用词进行过滤，因为这些词语对文段的实际含义影响不大。而特征词的数目也需要定义，过多的特征词会使得分词过程毫无意义，一旦达到了所需的特征词数目，就应停止计算。

在这方面的研究已经比较成熟，有许多开源的分词工具可供选择。本文通过java语音进行推荐系统的实现，因此选择了集成在maven项目中的ansj分词工具。这种分词工具采取n-Gram、CRF和HMM这三种模型来进行分词，分词效率十分可观。同时，可以自行进行语料库的定义，便于提取出适合特定领域特点的文本的特征词。这项工具还对特征词的词性进行了分类，本文的应用环境中，名词和动词对论文的区分度相较于形容词更高。除此之外，还可以通过TF-IDF提取特征值。

在实际使用中，通过在数据库的用户论文表中创建喜好特征词map，记录“提取的特征词-TFIDF值”这样的Key-Value对，便于后续处理。以本文为例，提取的喜好特征词map为：{"0":{"设计":2072.3318633398494,"中文":3914.50861438,"技术":1381.5516883728487,"推荐":3684.1383153853712,"信息":1378.717435475,"期刊":1842.0689178304644,"实现":1381.5518682695142,"系统":4374.913679847,"论文":3208.6674788461332,"基于":3223.627900472942}}，可见对特征词的提取比较准确。此处采用map形式进行记录是为了便于后续系统的更新迭代。其基本形式为{“0”:{},”1”:{}}，其中数字代表的是期刊id，0代表该论文属于未发表的待推荐文章集，1以及后续其他id代表所对应的期刊，{}中即为Key-Value对。后续设计时可将论文参考文献的题目，摘要，关键词进行分析，将Key-Value对记入该参考文献所属期刊id后的括号中。随后在该期刊内部寻找相似论文，并对相似论文多的期刊进行推荐。本文并未采用这种设计方式，其原因主要在于并非所有参考文献都属于中文期刊，故这种推荐方式并不适用，但在扩大推荐范围后，与本文推荐方式综合使用会提高推荐的准确性。

第二步是判断两篇论文相似。为解决这个问题，需对从两篇论文提取出来的特征词匹配程度进行合理量化。我采用的量化标准为：将从两篇文章中提取出的特征值求交集，交集中的特征词即为两篇文章的关联词，将关联词的TFIDF值之积的加总即为两篇论文的相似度，如公式（4-1）所示：

***​*** （4-1）

例如：两篇论文，分别提取出特征词与TFIDF值如下：论文A：“系统”：100，“推荐”：80，“实现”：40；论文B：“程序员”：100，“推荐”：90，“编程”：30。两篇文章只有一个共同特征词“推荐”，故相似度为：80\*90=7200。

随后，根据匹配分值得到推荐的论文，将其论文id和所属期刊id记录到数据库的推荐表中，将所属期刊数最多的期刊名作为最后的推荐结果。

已有的论文推荐系统设计中只使用了摘要这一文章信息[17]。本文设计了三种测试方式。方式1：对只对摘要进行分析，两篇文章摘要的相似度即为这两篇文章的相似度，a为A和B两篇文章摘要的重合特征词的集合，即：

***​***（4-2）

方式2：对标题、摘要、关键词进行分析，即将文章的标题、摘要和关键词合在一起作为文本集合，随后进行TFIDF值的提取以及相似度的计算，b为A和B两篇文章文本集合的重合特征词集合，即：

***​***（4-3）

方式3：分别对对标题、摘要、关键词进行分析，并单独提取特征词。将标题相似度的百分之五十与摘要相似度的百分之三十以及关键词相似度的百分之二十之和作为最终的文章相似度。其中，a为A和B两篇文章摘要的重合特征词的集合，c为文章标题的重合特征词集合，d为文章关键词的重合特征词集合，即：

|  |  |
| --- | --- |
|  | （4-4） |

**4.5.2 测试数据来源**

测试数据是通过垂直搜索技术在知网库中进行爬取。垂直搜索技术产生于互联网时代一些领域对信息搜索引擎的效率的需求。相比于传统的搜索引擎，垂直搜索引擎的主要特点是为了特定的专业领域而设计的，具有个性化与专业性。为了满足不同领域的需求，垂直搜索引擎的个性化和专业化水平不断提升，同时搜索结果越发精确。垂直搜索引擎的高个性化和专业化水平，提高了信息的全面率和准确率。其中爬虫模块对互联网中特定专业领域的特定信息进行爬取。分析模块主要分析网页结构，将所需的网络数据通过网络爬虫写入网页数据库中。索引模块通过建立索引的方式将网页数据库中的数据导入索引数据库。用户在进行查询搜索时，检索模块通过索引模块将索引库中的相关内容返回给用户，如图4-6所示。

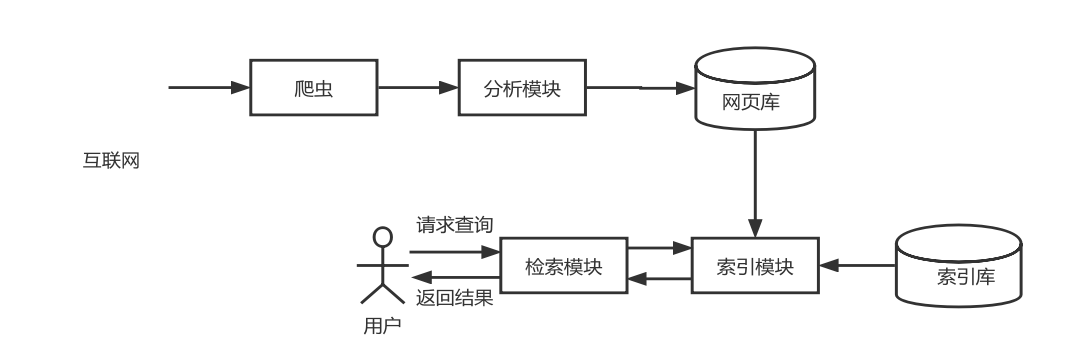


图4-6 垂直搜索引擎流程图

测试所所需的数据为用户待推荐文章和论文库文章的标题、摘要、关键词，以便建立用户特征向量和论文特征向量。论文库文章的相关信息则通过网络爬虫的方法提取。测试数据选用信息科技学科中计算机软件及计算机应用下74个期刊合计776881篇文章。选择其中五个期刊，将每个期刊中最新的二十篇文章从论文库中删去并放入用户论文库中作为用户论文数据。

**4.5.3 测试结果和分析**

通过推荐系统得出推荐的三个期刊，如果推荐的三个期刊中包含该篇论文所处的期刊，则视为推荐成功。按三种不同的推荐方式对用户数据库中的100篇论文进行推荐，其准确率评价指标如图4-7所示。

图4-7 各测试方式的准确率评价指标

从中可以发现，相比于只进行文章摘要的相似度计算，通过将论文的标题摘要和关键词整合为论文信息集合，随后再进行分词并计算相似度的推荐正确率有明显的提升。究其原因是论文标题和关键词同样是区分文章的重要指标，能够在一定程度上反应文章的研究方向。

但对于分别对文章标题、摘要和关键词进行特征词提取并分别计算其相似度最后按权重得到论文相似度的方法却不能稳定提升推荐正确率。主要的原因有以下几点：一是单独提取特征词时，文本长度对TFIDF的值影响较大。尤其是对于关键词，由于关键词仍然不是最小语素，所以仍然采用提取特征词和TFIDF值的方式加以分析，但由于关键词的文本长度很短，所以提取出的TFIDF值与建立论文信息集合的值相差较大。通过直接对关键词提取特征词并加以分析的方法不一定准确。二是对文章标题、摘要和关键词的相似度赋予的权重可能不够准确。文章的标题、摘要和关键词都可以对文章加以区分，但区分的程度不同。对这三部分所应占的权重还需进一步分析。

**4.6 本章小结**

本章就中文期刊投稿推荐系统的实现进行了介绍，包括用户系统和投稿推荐模块。介绍了数据库的连接方法，推荐算法的实现用户界面的搭建等。随后对推荐系统的三种不同推荐方法进行了测试。

**5 总结与展望**

**5.1 研究总结**

综上所述，本论文设计并实现了基于内容的中文期刊投稿推荐系统。主要的研究内容为：

1. 采用什么论文信息，如何提取论文信息，对提取的论文信息应当做出如何的处理，主要使用了TF-IDF算法来提取关键词并且通过Key-Value对来计算相似度。
2. 如何设计并实现系统的各个组成模块，主要设计并实现了基于WPS加载项的前端界面和通过java实现的推荐引擎。
3. 设计了三种测试方法对推荐结果进行了测试。

就结果而言，基于内容的中文期刊投稿推荐系统可以为初入科研领域、缺少中文期刊投稿经验的科研工作者提供基础的投稿期刊推荐。但该系统还有一些不足有待后续改善，主要在于：对标题、摘要、关键词的权重分配需要进一步研究；推荐方法单一；没有对期刊论文语料库进行自定义，分词时可能得到无含信息量充足的特征词；参考文献作为一篇论文与其他论文的交集，也提供了许多相似研究课题的文章，可以作为推荐的研究方向等；提取出的特征词数目是直接拟定的，特征词数目对推荐结果的影响还需要加以考量。期待以后进行改进；不同词性的特征词通常所能代表的论文信息不同，一般来讲，名词和动词相较于形容词、副词更能区分不同的论文。因此可以考虑赋予不同词性不同的权重，或者对一些词性进行过滤；对期刊没有进行分级。

**5.2 研究展望**

本文所设计并实现的推荐系统是实时推荐的，在推荐引擎进行推荐时，计算得出了论文库中论文信息的临时特征向量，该计算过程需要对论文库中所有论文进行计算，计算量很大。而在给出推荐结果后，这个临时特征向量库便不复存在了，当进行下一次推荐时，需要重新计算，影响了推荐的效率。后续设计时，可以将论文特征向量计算后写入数据库，实现离线推荐。这样的设计同样也可应用于实时推荐的总数据库，当论文库中有论文更新时，只对新论文进行特征向量的计算。这样可以大幅度提升推荐系统的工作效率，做到以空间换时间。同时还有分布式计算等技术，可以提升推荐系统的运行效率。

此外，论文的其他属性如参考文献的文本信息和参考文献的作者等文章信息可能不适用基于内容的推荐算法，可以结合协同过滤和基于图的算法等进行综合推荐。

在此次对基于内容的中文期刊投稿推荐系统的设计与实现中所学习的推荐系统实现技术在进行其他推荐系统设计时仍然适用。受限于为用户的特定文章进行投稿推荐这一推荐情景，基于内容的推荐方式的最大优点：通过不断迭代从而提升为同一用户的推荐结果准确率这一优点未能完全发挥。

正如在前文中提到的那样，对key-value对的提取是存储于一个map中的，这意味着用户的喜好模型和用户画像是可以进行更新的。通过不断搜集用户喜好特征，可以对用户喜好列表进行完善。但用户喜好不是一尘不变的，可以通过衰减机制，降低时效性差的特征向量的权重来不断提升推荐结果的准确率。

同时采用多种推荐方式也能实现优势互补，从而提升推荐结果的准确度。前文以及提及了基于协同过滤的推荐算法受益于其对用户喜好的把控度好等优点，在如今的各行各业都有不错的应用，但基于协同过滤的推荐算法存在冷启动问题，即系统库中缺少初始信息时难以进行推荐的问题。而通过基于内容的推荐方法，可以得到一个初步的推荐关联。在此基础上在进行基于协同过滤算法的推荐，就能提升推荐结果的准确性于新颖性。

**致谢**

首先要感谢我的母校华中科技大学培养了我，不仅在学业上教会了很多知识，

还在生活方面指导了很多。在过去的四年校园生活中，有过很多困难和挑战，每一份经历对于我来说都是弥足珍贵的。在本科生涯的最后阶段，学位论文将为学校生涯告一段落，然后我将踏入社会进行全新的开始。在毕业设计的阶段，指导老师龚老师提供了很多的帮助，十分感谢他在我毕业设计项目中的指导。

最后要感谢的是我的家人，对于我一路走来的辛苦和自己做出的决定都提供了支持和帮助，这将是以后促使我前进的不断动力。也要感谢我的室友和大学中交到的朋友们，他们都教会了我很多生活中的方方面面。

**参考文献**

[1] E. Rich. User modeling via stereotypes[J]. Cognitive Science,1979,3(4): 329­354

[2] J.Beel.Towards effective research-paper recommender systems and user modeling based on mind maps[J]. arXiv preprint arXiv:1703.09109,2017,54(3): 31­55

[3] C.L.Giles, K.D.Bollacker,S.Lawrence.Citeseer:An automatic citation indexing system[C].Proceedings of the 3rd ACM Conference on Digital Libraries, Pennsylvania,USA,1998,89­98

[4] H.Van de Sompel, J. Bollen. An architecture for the aggregation and analysis of scholarly usage data[C]. Proceedings of the 6th ACM/IEEE­CS Joint Conference on Digital Libraries,Chapel Hill, USA,2006,298­307

[5] D. M. Pennock,E. Horvitz.Collaborative filtering by personality diagnosis:A hybrid memory and model­based approach[C]. Proceedings of the Sixteenth Conference on Uncertainty in Arti-ficial Intelligence,Ann Arbor,2013,48109­2110

[6] Q.He,J.Pei,D.Kifer,etal. Context-aware citation recommendation[C]. Proceedings of the 19th international Conference on World Wide Web,North Carolina,USA， 2010,421­430

[7] A. Vellino. Usage­based vs. citation­based methods for recommending scholarly research arti­cles[J]. arXiv e­prints,2013,arXiv­1303

[8] [Recommender systems survey](https://kns.cnki.net/kcms/detail/detail.aspx?filename=SJES13050900032704&dbcode=SJES&dbname=scholar_journal_SJES" \t "https://kns.cnki.net/kcms/detail/frame/kcmstarget)[J].J.Bobadilla,F.Ortega,A.Hernando,A. Gutiérrez  Knowledge-Based Systems . 2013

[9] 孙婧. 论文推荐系统研究[J]. 数码世界，2020，(01):237.

[10] 杨杰. 个性化推荐系统应用及研究[D].中国科学技术大学,2009.

[11] 宋钰. 基于ResearchGate中用户生成内容的学术论文推荐系统研究[D].东南大学，2018.

[12] 王冬晖. 基于内容的计算机科学论文推荐系统设计与实现[D].吉林大学，2017.

[13] 车丰. 基于排序主题模型的论文推荐系统[D].大连海事大学，2015

[14] 要一璐. 基于知识脉络的科技论文推荐[D].山西大学,2016.

[15] 刘坤. 推荐系统中用户冷启动问题研究[D].兰州大学,2017.

[16] 郭冬萌. 个性化推荐系统中的推荐算法研究[D].北京交通大学,2017.

[17] 吴逍遥. 基于用户相似性的推荐算法优化研究与实现[D].北京交通大学,2019.DOI:10.26944/d.cnki.gbfju.2019.000831.

[18] 孟伟龙. 基于图模型的论文推荐系统设计与实现[D].西北农林科技大学，2019.

[19] 刘玲，黄丽蓉，刘胜宗. 论文推荐系统的关键技术研究[J]. 湖南工程学院学报(自然科学版)，2017，27(04):43-47.

[20] 孟凡琦. 基于查询的论文参考文献个性化推荐系统研究[D].天津大学,2014.

[21] 丁芳媛. 基于新颖性和影响力的论文推荐方法研究[D].华南理工大学,2020.DOI:10.27151/d.cnki.ghnlu.2020.004279.

[22] 贺小雨. 基于多属性特征的论文推荐系统设计与实现[D].电子科技大学，2021.

[23] 孙婧. 基于引文网络图模型的论文推荐系统研究与应用[D].云南师范大学，2020.

[24] 景岐-蔻享学术论文推荐系统V1.0. 安徽省，安徽国科新材科技有限公司，2019-04-18.

[25] 王连臣. 基于深度神经网络的论文推荐系统研究[D].吉林大学，2019.

[26] 钱宇. 科研文献开放获取系统中的推荐算法研究与应用[D].浙江大学,2017.

[27] 栗欢. 基于社区发现和关联规则的论文混合推荐技术研究[D].华中师范大学,2015.

[28] 刘宇,朱文浩.基于内容和标签权重的混合推荐算法[J].计算机与数字工程,2020,48(04):773-777.



**本科生毕业设计（论文）任务书**

题 目 **基于内容的中文期刊投稿推荐系统**

**设计与实现**

（任务起止日期：2021年11月2日-2022年6月5日）

院 系 **管理学院**

专业班级 **信管1802班**

姓 名 **李永涵**

学 号 **U201815963**

指导教师 **龚晓光**

教研室（系、所）负责人 2021年10月28日审查

院（系）负责人 2021年11月2日批准

**任务书填写要求**

1. 填表请用五号宋体字编辑，签名须手写，A4纸双面打印。
2. 此任务书表格内容应由指导教师填写。
3. 此任务书最迟必须在毕业设计**开始前一周下达给学生**。

|  |
| --- |
| 课题内容：  知网文献期刊数据爬取，存储与自动更新。建立分析模块，将半结构化的网页信息分析处理成结构化的信息，方便后续建立索引。 文档关键信息提取算法与实现。通过自然语言处理（NLP）对文本数据进行分析处理，将收集到的原始数据处理成模型构造过程中能够使用的数据。 投稿推荐算法的设计与实现。将论文中的各项信息如作者、论文引用网络、研究领域、关键信息等内容进行关联形成网络模型。在进行论文搜索及推荐的操作时，将所提出的模型结合相应算法形成推荐结果。 服务器端WebAPI设计与开发。 Office web-addin与wps web-addin客户端开发，将期刊投稿系统生成插件，方便用户使用。 |
| 课题任务要求：  分析将投稿期刊的关键内容与结构，与爬取到的期刊相关信息进行匹配，实现智能推荐，并开发设计期刊投稿推荐系统 |
| 主要参考文献（由指导教师选定）  [1]贺小雨. 基于多属性特征的论文推荐系统设计与实现[D].电子科技大学,2021. [2]孙婧. 基于引文网络图模型的论文推荐系统研究与应用[D].云南师范大学,2020. [3]孙婧. 论文推荐系统研究[J]. 数码世界,2020,(01):237. [4]孟伟龙. 基于图模型的论文推荐系统设计与实现[D].西北农林科技大学,2019. [5]王连臣. 基于深度神经网络的论文推荐系统研究[D].吉林大学,2019. [6], 景岐-蔻享学术论文推荐系统V1.0. 安徽省,安徽国科新材科技有限公司,2019-04-18. [7]宋钰. 基于ResearchGate中用户生成内容的学术论文推荐系统研究[D].东南大学,2018. [8]刘玲,黄丽蓉,刘胜宗. 论文推荐系统的关键技术研究[J]. 湖南工程学院学报(自然科学版),2017,27(04):43-47. [9]王冬晖. 基于内容的计算机科学论文推荐系统设计与实现[D].吉林大学,2017. [10]车丰. 基于排序主题模型的论文推荐系统[D].大连海事大学,2015. |
| 同组设计者  无 |
| 指导教师签名：  年 月 日 |