

2024 年全国大学生计算机作品赛

参赛作品说明书

概要文档

Web 智寻 IntelliFind--

参赛作品名称：基于用户行为数据的网站评分算法系统

学 校：电子科技大学

学院（系）：信息与软件工程学院

专业班级：信息与软件工程学院

学生姓名：刘治学、李航宇、郑笑宇、李远行、吉劲帆

指导老师：白忠建

完成时间：2023.04.13

目录

1 作品概述.....	3
1.1 前言.....	3
1.2 项目目标.....	3
1.3 项目背景.....	4
1.4 项目命名.....	4
1.5 创意描述.....	5
1.6 特色描述.....	5
2 需求分析与建模.....	6
2.1 应用对象分析.....	6
2.2 功能性需求.....	6
2.3 非功能性需求.....	7
3 实施方案与可行性研究.....	8
3.1 可行性分析.....	8
3.1.1 市场可行性.....	8
3.1.2 技术可行性.....	8
3.1.3 资源投入可行性.....	9
3.2 实施方案.....	11
3.2.1 项目系统结构和模块设计.....	11
4 结语.....	14

1 作品概述

1.1 前言

在科技快速发展的今天，网络成为了人们工作生活绝不可少的一环，而其中每个人的日常生活都离不开网站的使用。网站是人们获取并交互信息的主要途径，一个网站性能的好坏，很大程度上影响了人们的体验。

根据 2015 年 Aberdeen Group 的研究报告，对于 Web 网站，1 秒的页面加载延迟相当于少了 11% 的页面观浏览，相当于降低了 16% 的顾客满意度。如果从金钱的角度计算，就意味着：如果一个网站每天挣 10 万元，那么一年下来，由于页面加载速度比竞争对手慢 1 秒，可能导致总共损失 25 万元的销售额。同时该报告中指出分析了超过 150 个网站和 150 万个浏览页面，发现页面响应时间从 2 秒增长到 10 秒，会导致 38% 的页面浏览放弃率。

由此可见，网站性能与业务目标有着直接的关系，对网站进行性能测试非常重要。再结合用户的实际体验，启动一个软件如果很卡，就不太想用了，如果在中间使用时再很卡时，下次再想使用的欲望就会强烈减少，甚至会产生排斥心理。

但是对于大多数站点管理者而言，并没有太多时间和精力自己去检测网站的性能和用户体验问题，他们往往希望有一个专门的网站，能够检测用户在网站体验中存在的问题，并作出相应评分，甚至提出相关的优化建议，这样就能客观的反映出网站性能的好坏，站点的管理员也更能针对性地优化和修复网站性能上的问题。因此一个合适能够检测用户行为数据的网站体验评分算法是必不可少的。

因此我们分析项目可行性并进行实践开发，提出能够检测用户行为数据并反馈给用户的智寻系统。

1.2 项目目标

智寻系统目标达成以下几点：

1. 接收输入的用户行为数据，通过算法分析得出结果，并反馈给用户。
2. 以直观清晰的方式布局反馈的结果报告，让用户直观的感受网站的体验问题和优化建议。
3. 模拟真实的网站，考虑工程上的更多的细节问题，让本系统在工程角度上更加完善和强大。

1.3 项目背景

党的十八大以来，习近平总书记高度重视网络安全和信息化工作，从信息化发展大势和国际国内大局出发，就网信工作提出了一系列新思想新观点新论断，深刻回答了一系列方向性、根本性、全局性、战略性重大问题，形成了内涵丰富、科学系统的习近平总书记关于网络强国的重要思想，为做好新时代网络安全和信息化工作指明了前进方向、提供了根本遵循。

习近平总书记关于网络强国的重要思想，坚持马克思主义立场观点方法，立足人类进入信息社会这一崭新时代背景，站在我们党“过不了互联网这一关，就过不了长期执政这一关”的政治高度，准确把握信息化变革带来的机遇和挑战，从国际和国内、历史和现实、理论和实践的结合上，深入回答了为什么要建设网络强国、怎样建设网络强国的一系列重大理论和实践问题，深化了我们党对信息时代共产党执政规律、社会主义建设规律、人类社会发展规律的认识，丰富了习近平新时代中国特色社会主义思想的科学内涵。

在这样的大背景下，网络中占据绝对地位的网站就更加需要改进自身。不仅需要改进内容，更加需要考虑到网站性能和体验上的问题，考虑到对网站浏览者，考虑到对用户的体验的维度。这样才能真正把网站的管理做好，才能真正坚持习近平总书记提出的网络强国的重要理论。

对于智寻系统，就是在这样的背景下，孕育而生，能够有效的帮助网站管理者优化改进自己的网站，这是顺应党的号召的良好表现。

1.4 项目命名

Web 智寻 IntelliFind 系统是一个基于用户行为数据的网站评分算法系统，为用户提供方便、直观的服务，并且响应了党提出的网络强国的重要战略。

Web，代表本系统采用 BS 架构，是一个 Web 网站系统，用户使用浏览器就可以方便地访问和使用。

“智寻”二字，代表本系统可以检测用户提供的日志，通过分析得出直观清晰的结果。用户通过将自己网站记录的日志文件传递给本系统，进而通过算法调用，最后计算得出这些用户行为的得分、优化建议等，然后以各种数据图形图、数据等直观的展现在用户面前，让用户能够针对自己网站的体验做出优化和修改。

IntelliFind，融合了单词 Intelligence 智能和 Find 寻找的含义，也是智寻二字的体现。

Web 智寻 IntelliFind 系统，致力于为用户提供可靠、方便、直观、清晰的检测、分析的服务。

1.5 创意描述

经过调研市面上其他网站分析系统，我们发现如下问题：

1. 对于非开发人员不友好。市面上有的系统在反馈解析结果时引用了大量的专业术语，不便于非专业人员的使用。
2. 网站功能老旧、界面粗糙或是操作复杂。
3. 不支持用户自定义，只能够按照自己的分析套路进行，容易误导用户。
4. 功能不完善，可视化界面较少，分析结果不准确。

1.6 特色描述

1. 支持自定义模式。用户可以根据自身需要。对不同问题设定不同权重，反映出自己更加注重哪一方面的问题，能够更好地帮助用户抓住网站问题所在，对症下药，寻求高效地答案。
2. 系统对于用户上传的 json 文件进行了预处理，经过了数据解析与拆分工作后再交由算法模型生成结果。保障了算法模型的运行效率，尽可能地减少用户等待的时间，提高用户体验。
3. 系统采用了新兴的算法大模型对用户上传文件进行解析，并且将解析结果尽可能地可视化呈现用户，用户能够更加直观地看出网站问题所在，更好地感受数字反映出的问题
4. 文件系统。系统设计了完善丰富地文件系统，能够很好地支持用户上传文件的请求，对于上传的文件，不仅数据库有上传记录，文件本身也统一保存在文件系统中。
5. 两层缓存保证数据查询效率。系统设计了基于 LRU 和 Redis 的两层缓存,极大地减少了数据查询的时间，提高了用户使用体验。
6. 分布式部署。web 服务和算法模型分布式部署在两个服务器上，减少了不同板块的相互影响。

2 需求分析与建模

2.1 应用对象分析

智寻系统的使用者主要有：

用户：以网站的管理人员为主，在日常管理网站的过程中经常会排查网站的各种问题，并且需要及时根据结果优化自身的网站，以便及时修改优化网站，为自己的客户带来更好的体验。同时这些网站管理者也是本网站的用户，他们可以在本网站当中查看以往的相关的检测信息、结果等。具体如图 2-1：

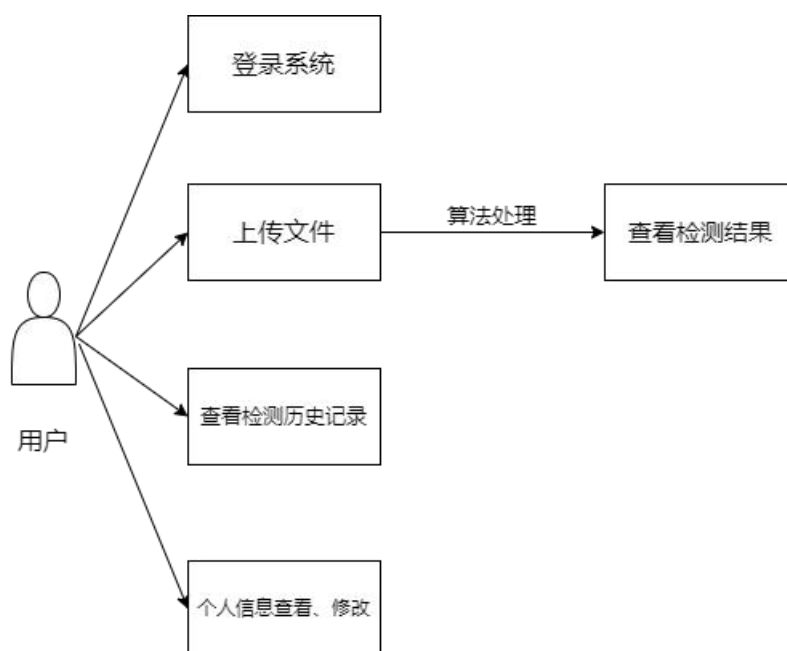


图 2-1 网站主要功能

2.2 功能性需求

功能简介

作为一款力求模拟真实的网站评分系统，需要解析用户传入的数据，这些数据绝大部分就是用户自身网站的日志信息，里面就包含了网站体验过程中的各种问题，然后通过调用算法，最终得出清晰直观的结果，这就是一条完整的功能线。具体如下：

1. 登录功能：为了实现记录不同用户的检测历史和用户信息等功能，故设定登

录功能，包括注册、邮箱验证、忘记密码找回、图形验证码等真实网站具有的流程。同时为了让用户使用十分简易方便，登录的所有功能均直接在同一页面，用户只需要直观的自行选择即可。

2. 用户个人信息查看、修改：对于有登录的网站而言，用户自然可以设置用户的用户名、邮箱、地址、个性签名、头像等，使得用户之间鲜明的区别开来。
3. 检索历史检测记录：对用户的不同次检测，在数据库中存储了当次检测的相关数据，例如相应问题的数据字段、检测得分等等。用户点击即可跳转到对应结果页面。
4. 上传文件：核心功能。实现了用户从上传文件到后台读取、分析、检测文件，并通过算法调用，得出最终检测结果并渲染成为检测结果页面的过程。

2.3 非功能性需求

1. 可用性需求：系统应易于使用，功能明确，操作简单，用户能够轻松理解如何使用网站，并进行文件的上传和结果的展示。
2. 性能需求：系统应快速处理传入文件中的内容，并通过算法分析，能够快速响应并返回结果，渲染在网站上。结果应当在 10 秒以内呈现给用户。
3. 可维护性需求：系统应易于维护和更新，有良好的文档和代码结构，支持快速诊断和解决问题，同时能支持性能优化。
4. 可扩展性需求：系统应易于扩展和定制，能够容易地添加新的功能或更新已有功能，未来能够支持新的文献类型和识别技术。

3 实施方案与可行性研究

3.1 可行性分析

3.1.1 市场可行性

在开发本系统以前，我们对市面上已有的产品进行了调研，目前市面上有很多网站性能检测网站和服务可用性检测网站，例如百川云、GTmetrix、Pingdom、Uptime Robot 等，但是这些产品如果照搬到本系统中的需求当中，还是有很多难以达到要求的地方，故做如下分析：

1. 业务倾向不同：例如，百川云是一家专注于网络安全的厂商，GTmetrix 是一家提供托管服务的公司。业务倾向的不同导致了这些产品在实际运作的时候，对于本项目而言，这些网站的业务就不是很适合了。
2. 不支持个性化需求：对于本项目，各个网站的管理员是我们的用户，本系统希望为不同的用户提供个性化的服务，例如展示解析历史变化图，存储历史报告等等。上述的网站基本都不支持与本项目业务相关的个性化服务。
3. 数据隐私和安全性的要求：由于本系统会处理用户的网站日志数据，因此对数据隐私和安全性的要求可能会与市面上的产品有所不同。系统需要得到用户的信赖，确保用户的数据得到保护，而现有的一些产品可能并未专门考虑这些方面的需求，因此需要定制化开发。
4. 需要适应快速迭代和定制化开发：由于本系统需要根据不同用户的需求提供个性化的服务，因此需要具备快速迭代和定制化开发的能力。市面上的通用产品较难满足这种灵活性和定制化的开发需求，因此需要建立自己的开发团队或寻找专业的定制化开发服务提供商。

3.1.2 技术可行性

对于技术可行性，技术可行性需要考虑所选择的技术是否成熟，是否能够满足构成本系统的要求。我们将从前端、后端、算法三个方面阐述技术可行性。

1. 前端

- a) 通过 Vite + Vue3 构建项目前端模块，通过数据的响应式交互让用户能

Web 智寻 IntelliFind -- 基于用户行为数据的网站评分算法系统

够在操作时实时看到数据的更新。

- b) 通过 Element-Plus 和 ECharts 组件丰富页面内容，美化数据的展示形式，提升用户体验。
- c) 通过 Axios 发送异步请求，通过 Pinia 将 Token 暂存于本地，进而实现前后端分离式项目请求响应交互时的鉴权和认证。

2. 后端

- a) 使用基于 golang 的 gin 框架。golang 是 web 开发的常用编程语言之一，其执行速度效率接近 c++，其并发模型基于协程（goroutines）和通道（channels），这使得编写高并发服务器变得简单而高效。
- b) 使用 MySQL 数据库存储。MySQL 具备作为数据存储系统的多项优势，包括可靠性、性能、灵活性、成本效益、社区支持、扩展性和兼容性。
- c) 基于 Redis 和 LRU 的两级缓存。使用 LRU 算法作为项目的第一级缓存，使用 Redis 作为第二级缓存。这两种缓存在技术上已经较为成熟，使用较为广泛，认可度较高。

3. 算法

- a) 对于用户上传的文件，算法需要对其进行解析，最后返回响应的结果，我们将这个过程分为两个部分，人为解析和模型处理。
- b) 人为解析：通过 python 脚本解析用户提供的 json 文件，从中提取有用的数据字段，获得有效数据，交予模型处理。
- c) 模型处理：我们选择使用开源大模型 Llama2，并且经过官方的预训练和我们的训练，同时使用开源库 Transformers 运行模型，以做到保证效率和性能。

综上所述，本系统的技术可行性较高，市场上已经有比较成熟的技术栈可以支持实现该项目的各项功能，同时需要根据具体的业务需求和技术方案进行选择和整合。

3.1.3 资源投入可行性

对于资源投入的可行性分析，我们考虑以下几个方面：

1. 人力资源：开发智寻系统需要有相应的技术人员，包括前端、后端和算法部分人员，以及日常维护人员。在招募人员时考虑人员的专业技术和经验，以确保能够顺利完成项目。
2. 硬件资源：开发和运行智寻系统需要相应的硬件资源，包括服务器、计算机

等。需要根据系统的规模和预期的用户量来确定相应的硬件资源投入。

3. 软件资源：开发和运行这个系统需要使用相应的软件资源，包括操作系统、数据库、开发工具等。

4. 时间资源：开发一个完整的网站体验评分系统需要相当长的时间，包括需求分析、设计、开发、测试、部署等各个阶段。需要合理安排时间资源，确保项目能够按时完成。

综合考虑以上各个方面的资源投入，我们进行相应的成本预算和时间计划。人力资源而言，我们具有各类技术人员，前端两名、后端一名、算法一名、机动一名，分工明确，并且组长规定时间进度，队员按照要求进行开发和测试；硬件资源方面选用阿里云轻量级服务器进行初步开发，之后考虑用户数量扩大更换服务器；软件资源方面，选用大量开源技术，如 MySQL、Vue、Flask 等；时间资源方面我们严格指定了项目的开发时间，能按时保质保量的完成。

故智寻系统的开发资源投入是可行的。

3.2 实施方案

3.2.1 项目系统结构和模块设计

1. 系统架构

在项目设计初期，本团队就对系统架构进行了设计，根据业务逻辑的不同层次，分别划分为用户层、业务层、技术层和存储层，如图 3-1：

- a) 用户层：用户层考虑了使用的用户群体，对于本项目而言，用户绝大多数都是各自网站的站点管理员。
- b) 业务层：智寻系统主要实现四大业务，分别对应登录功能、个人信息相关、查看检测历史和解析文件的相关业务，为业务层实现提供方法。
- c) 技术层：智寻系统采用四大关键技术。使用 Vue3 和 Gin 技术实现 Web 页面前后端的搭建，使用 Python 语言解析用户提供的 json 日志文件，并使用 Llama2 模型及相关技术进行模型处理返回响应的结果。
- d) 存储层：使用 MySQL 数据库分别构建用户信息存储数据库和用户检测存储数据库。

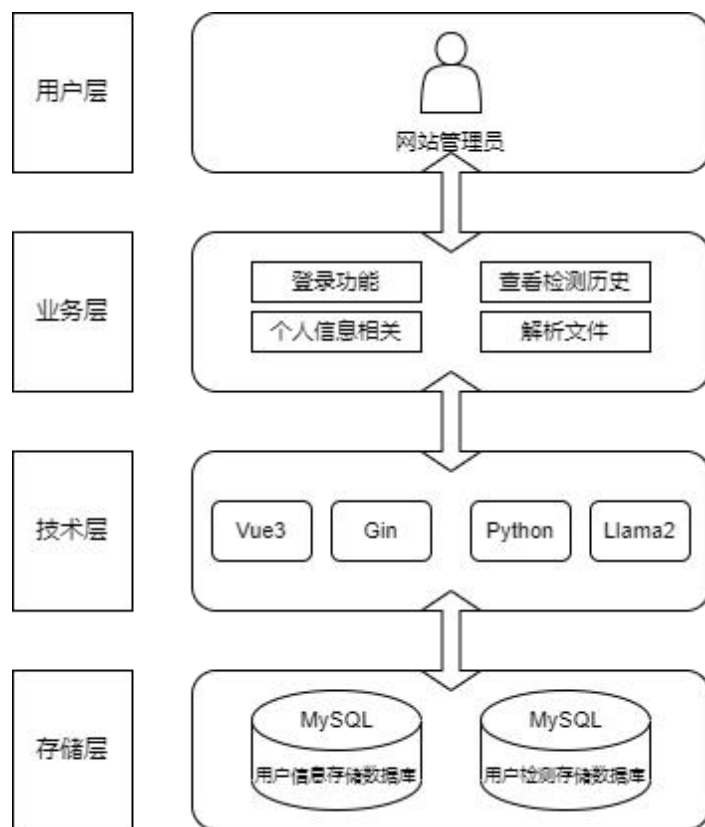


图 3-1 项目结构设计

2. 模块设计

模块设计如图 3-2 所示，我们按照功能将其分为登录、用户、解析、历史四大模块。

登录模块：提供了完善的登录工具链，包括账号登录，账号注册和修改密码的功能。账号登录要求用户输入账号、密码，之后通过图形验证码进行人类检测，最后登录；帐号注册要求用户输入用户名、密码和注册使用的邮箱，其中邮箱注册之后无法更改，并且通过我们绑定的自定义域名邮箱 xxx@lzx0626.me 进行邮箱验证，最后注册账号；修改密码同样要求用户通过邮箱验证，最后输入新密码进行重置。

用户模块：存储了用户的基本信息包括用户邮箱、注册时间，还有个性化信息包括姓名、城市、生日、性别、电话、头像等，为用户提供了自定义个性化的服务。

解析模块：接收用户传入的 json 文件，同时可以接收用户自定义输入的权重配置信息，然后通过算法调用，展示解析之后的结果，包括用户环境配置信息、权重配置信息、还有解析之后的各项得分和综合建议。

历史模块：罗列了用户的解析记录，点击详情可以跳转到某次的解析结果；另将最近七次的解析反馈数据绘制为折线图，展示用户解析结果的变化趋势。

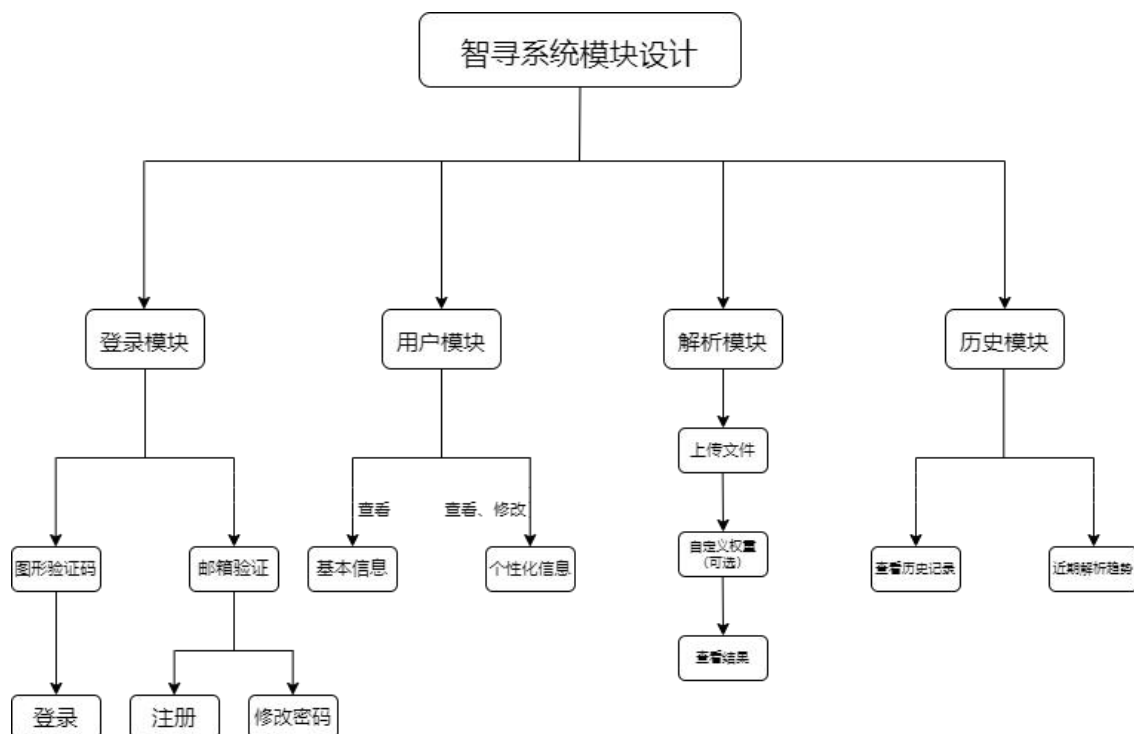


图 3-2 项目功能设计

其中解析模块的具体流程如图 3-3 所示：

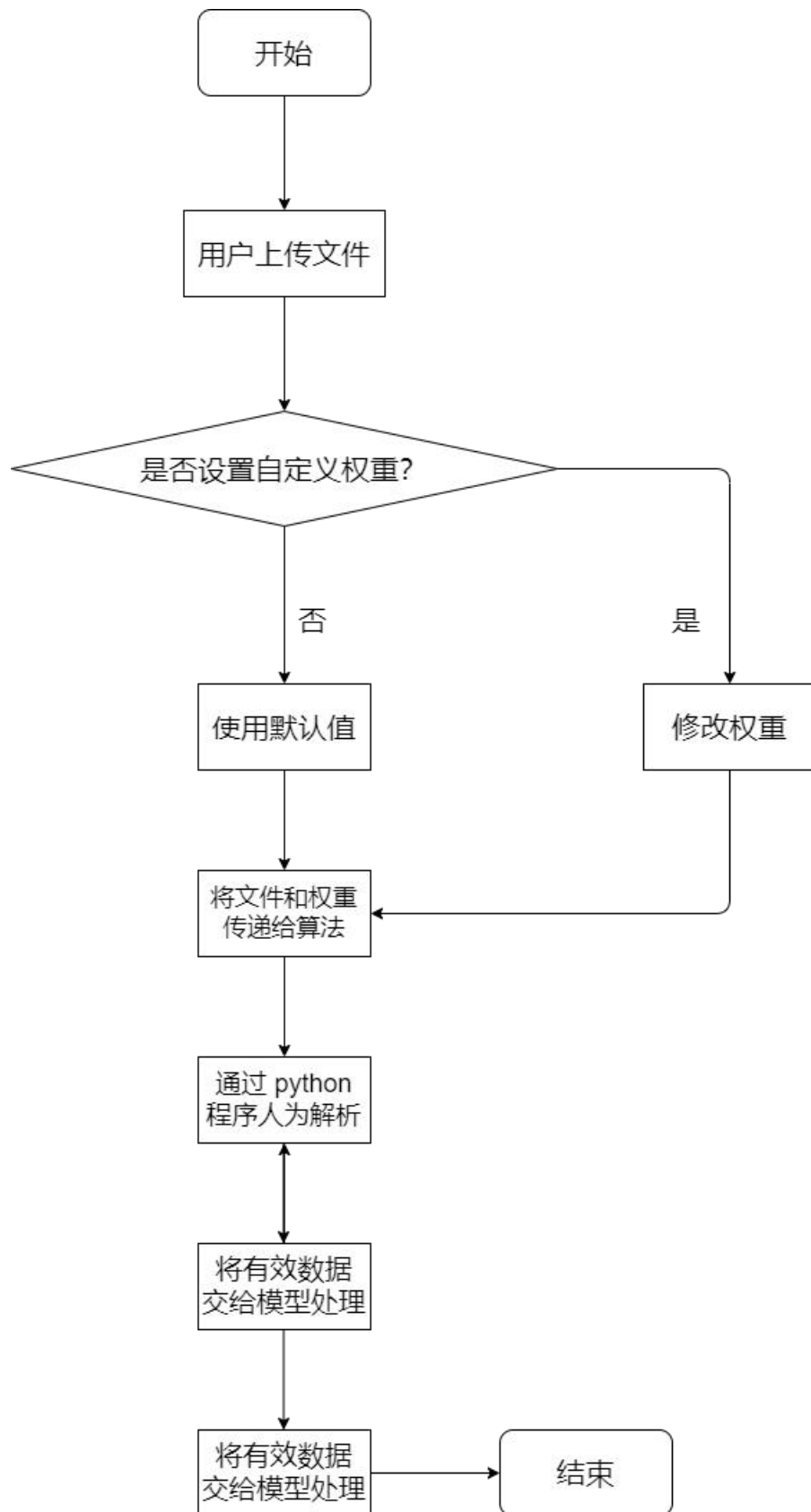


图 3-3 解析流程图

4 结语

Web 智寻 IntelliFind——基于用户行为数据的网站评分算法系统致力于利用用户行为数据来对网站进行评分，其目的在于帮助用户提高网站质量，调整其市场营销策略，针对用户行为进行更精准的营销活动，提高转化率；通过对网站进行评分和排序，可以鼓励网站运营者遵守良好的网络实践，减少误导性内容和不良信息的传播。此外，学者和市场研究人员可以使用这种基于用户行为的评分系统来分析网络趋势、用户偏好和行为模式，为相关研究提供数据支持。

此外，该项目也响应时代号召，深入贯彻习近平总书记关于科技创新的重要论述和对科技基础能力建设的重要指示，采用先进的技术开发产品，并提供给使用人员简单的操作方式，有助于科学教育和科普活动深入开展，创新发展理念深入人心，全民科学文化素质不断提升，尊重创新、鼓励创新、支持创新的社会氛围日益浓厚。

项目实现功能

- 用户管理
 - 用户登录
 - 用户注册
 - 重置密码
 - 更新用户个人信息
 - 注销登陆
- 文件解析
 - json 文件上传和保存
 - json 文件基本解析
 - 根据解析计算各项分值
 - ai 大模型分析 json 数据
 - 可视化解析结果
 - 解析结果存储
- 历史记录删除