

赛题 04

数据标注平台

目录

1.	组织管理	1
1.1.	团队介绍	1
1.2.	职责分配	3
1.3.	团队建设目标	5
1.4.	团队项目管理	6
1.4.1.	项目进度管理	6
1.4.2.	项目测试管理	7
1.4.3.	项目资源管理	7
1.4.4.	项目成本管理	9
2.	业务分析	10
2.1.	开发思想	10
2.2.	开发方法	10
2.3.	系统功能性需求	14
2.4.	系统非功能性需求	14
2.4.1.	安全性	14
2.4.2.	美观性	15
2.4.3.	直观性	15
2.4.4.	可变性	15

1. 组织管理

1.1. 团队介绍

本外包团队名为标注先锋五人队，致力于为发包商开发、实现完美的作品。本组是一支年轻富有活力的团队，由5个富有工作热情、责任心、踏实肯干的大学生领衔。本团队热衷于数据挖掘与文本分析和处理相关外包项目的开发与应用，秉承学习、创新、团结、拼搏的服务宗旨，科学有效地提高综合实力以及竞争力。目前正在利用数据挖掘和数据处理的知识向评论分析与反馈领域拓展。团队成员具体情况如下：

表 1- 1 团队成员介绍表

姓名	性别	团队角色	组别
Xia	男	项目经理	项目管理组
Yang	男	技术经理	项目开发组
Fang	男	技术组长	项目开发组
Chen	男	UI 设计经理	UI 设计组
Yan	男	客户关系经理	项目管理组

本外包团队致力于为客户提供细致周到的服务，立志于在创新、开发及服务领域，恪守信誉、提高声望，为客户提供方便、快捷的服务，坚持以“诚信为本，客户至上”为原则，坚持现场办公，以用户需求为核心，为用户挖掘潜在需求，可在第一时间掌握各地用户的最新需求，并及时反馈给用户。

本团队一致认为，一个真正优秀的团队中，在有着明确的分工的同时，又相互有交集。这使得每个人都能放眼整个项目，既能自我定义目标与工作，又能在相互交流中，对项目进行改进和兼容。同时这种相互促进的方式，在能高效完成各自工作的同时，也能在交流互助中提升每个人的总综合实力。

1.2. 职责分配

本组按照软件工程的思想，将项目整体开发过程进行任务细分，把整个项目研发过程分成不同的环节，主要分为项目经理、技术经理、技术组长、UI 设计经理、客户关系经理。

同时，本队根据每位成员性格、能力上的不同，并结合心理学角度和方法，初步分配成员适合的模块，分配成员相应的角色，具体分配如下：

● 项目经理

项目经理是整个项目组织中的核心角色，负责整个项目的实施。项目经理将负责所有的管理工作，对系统将承担最终的职责。项目经理将参与日常的系统实施管理，监控项目的进度。项目经理负责向领导小组汇报开发进度和开发相关的问题。

由于该项目建设涉及方面较多，时间紧，工作量大，因此为了更好地完成本项目的实施工作，项目经理负有如下职责：

- ① 领导项目团队，制定项目管理计划与主要里程碑。
- ② 在现场担任现场实施的总负责人，协调人员安排、问题协商与解决等工作。
- ③ 参与项目全程，负责组织构建项目开发的推进。
- ④ 审批软件项目研发的工作准则与工作流程。
- ⑤ 负责交付工作，保证项目的正常运行。
- ⑥ 组织召开项目的周会、例会和技术碰头会等工作。

● 技术经理

技术经理作为技术组的主要负责人，精通 html、js 语言，对 B/S 体系结构框架较为熟悉，对数据库关系模式的设计优化见解独到，并组织组内成员完成技术文档的撰写，主要负有如下职责：

- ① 负责整个软件项目的规划设计、项目组织架构的建立和调整。
- ② 小组任务的下达和跟踪，执行和监控项目。
- ③ 制定项目技术方针，领导团队完成快速原型的设计工作。

● 技术组长

技术组长在数据挖掘及自然语言处理领域、机器学习领域涉猎较深，负责整个项目技术路线的实施，主要负有如下职责：

- ① 规划成本和风险模型，并对其进行评估
- ② 不同平台的数据爬取
- ③ 数据的统计分析
- ④ 项目数据库表格的建立

● UI 设计经理

UI 设计经理作为 UI 界面设计的主力技术人员，拥有扎实的 HTML5、CSS、JavaScript 基础，熟悉前后端的数据交互形式，主要负有如下职责：主要负有如下职责：

- ① 前端页面 UI 设计
- ② 前端页面代码实现
- ③ 负责对外交流和发言。
- ④ 随时根据变化关注并调整项目进程，发布项目公告。
- ⑤ 和客户进行需求方面的沟通，分析需求现状。

● 客户关系经理

客户关系经理是项目团队中至关重要的一员，负责与客户之间建立和维护有效的沟通和关系。在项目实施过程中，客户关系经理将承担以下职责：

- ① 客户沟通与需求分析
- ② 项目计划与沟通
- ③ 问题解决与协调
- ④ 客户满意度管理

⑤ 客户培训与支持

⑥ 客户关系维护

1.3. 团队建设目标

在软件开发项目中，我们特别注重团队的建设，让团队为实现一个共同的目标而努力。在项目开发的过程中，我们的团队发挥出了极高的工作效率，培养了团队的默契和凝聚力。我们的建设主要分为以下几个方面：

● 良好的沟通氛围和交流环境

与项目成功有直接关系的三个因素分别为客户的积极参与、明确的需求表达和学校的大力支持。这三要素全部依赖于良好的沟通的技巧。要营造良好的沟通分为和交流环境，本团队采用了多种方式：

- ① 加强团队成员之间的沟通。成员之间由于价值观、性格等方面的差异可能会产生各种冲突，人际关系往往会陷入紧张的局面，甚至有可能出现敌视情绪。为此，团队在开发过程中做到了充分沟通，引导团队成员调整心态和准确定位角色，把个人目标与项目目标结合起来。
- ② 促使团队成员尽快熟悉工作环境。团队成员与周围环境之间也会产生不和谐，如对开发团队采用的信息技术不熟悉等。针对这种情况，项目经理在初期及时帮助团队成员熟悉工作环境，学习并掌握相关的技术，以利于软件项目目标的及时完成。
- ③ 加强软件开发团队与其他部门之间的沟通。在软件开发过程中，开发团队与其他部门也会产生各种各样的矛盾冲突，这需要软件开发经理与这些部门的管理者进行很好的沟通和协调，为软件开发团队争取更充足的资源与更好的环境。

● 同室作战

在整个项目的开发期间，项目组中的成员全部集中到统一工作地点。我们认为虽然现在通信技术和现代化办公设施已先进到可以远程办公，但是项目组成员

集中在一起工作更容易使团队成员全身心的投入到项目中来,也使团队成员增加了交流和默契。此外,作为一个大型的软件开发项目,成员之间相互的沟通信息量非常之大,而且要及时准确表达信息的最有效方法,就是面对面的沟通。同时,我们在办公室显著展示项目进度计划,并且定期对其进行更新,这样能让所有成员都非常清楚当前项目所处的状态,而且也能让项目经理及时发现问题纠正偏差。

● 劳逸结合

在软件开发过程中,由于需求不明确、技术日新月异、项目工期紧等客观原因,项目组每位成员压力都很大,经常要加班加点的工作。为了缓解高压环境下的工作压抑感,帮助组员减轻压力,我们定期组织一些娱乐活动。在关键里程碑完成的时候,进行总结和庆祝,如组织聚餐唱歌看电影等。这样在大家放松紧张的心情的同时,有利于小组成员之间的相互理解和支持,增强了团队的凝聚力和战斗力。

● 共同的工作框架、规范和纪律约束

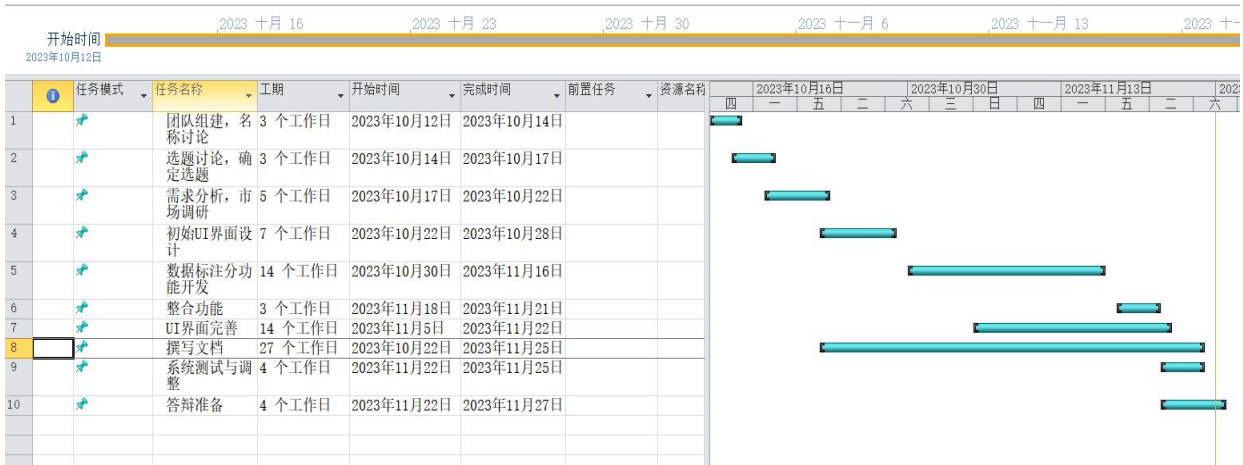
项目的开发是创造性的工作,但要有必要的开发纪律。本团队建立了共同的工作框架,使团队成员知道如何达到目标;建立规范使各项工作有标准可以遵循;建立一定的纪律约束可以保证计划的正常执行。

1.4. 团队项目管理

1.4.1. 项目进度管理

工程项目进度管理是工程项目管理的重要组成部分,也是工程项目管理工作中矛盾比较突出的问题。项目进度管理通过控制手段使实际进度符合计划进度的要求,每隔一定时间将进度的实际值与进度的计划值做比较,若发现实际进度偏离了计划进度,我们就采取相应的措施,纠正偏差,调整进度,使项目按期完成。以下是项目计划表以及甘特图设计:

表 1- 2 项目计划表以及甘特图



1.4.2. 项目测试管理

作为项目开发的重要环节，项目测试对项目建设的成败起着举足轻重的作用。随着软件开发规模的增大、复杂程度的增加，以寻找系统中的错误为目的的测试工作就显得更加困难。因此，为了尽可能多地在系统被正式应用之前找出错误，生产出高质量的产品，就必须对系统进行严格的测试工作。

从项目的生存周期看，测试往往指对系统的测试，这样做的优点是被测对象明确，测试的可操作性相对较强。但是，由于测试的依据是规格说明书、设计文档和使用说明书，如果设计有错误，测试的质量就难以保证。即使测试后发现是设计的错误，这时修改的代价是相当昂贵的。因此，较理想的做法应该是对系统的开发过程以及各阶段形成的结果分别进行严格地审查。

为了确保投诉线索智能筛查系统的顺利完成，并保证其中各子系统质量，我们将严格按照制定的测试计划进行软件测试工作。

测试的组织基本与开发工作同步，由项目经理和客户经理建成测试小组，对已完成的模块进行功能测试。在此期间，先由技术开发人员编写测试大纲，明确测试的内容和测试通过的准则，设计完整合理的测试用例，以便系统整合后进行全面测试。

1.4.3. 项目资源管理

项目资源是存在于项目开发过程中，能够被项目开发者所利用的一切资源，不论是劳动力、材料、设备，还是资金、技术等形成生产力的各种要素，以后是图书、网络资源等等，只要我们需要，这些都是我们项目开发中的资源；项目资源管理是针对项目资源的利用、协调和往来交流所开展的活动。

本团队在接手该项目之前，我们充分了解了自身条件，在项目过程中，对可获得的资源进行了整合，将项目资源分为以下三个部分：

(1) 企业资源

在我们确定所选命题之后，详细了解企业要求，并且详细了解企业的各种资料，为项目提供初始基础。

(2) 文献资源

文献资源分为一下两个部分：互联网资源以及图书馆资源。

万能互联网可以说拥有无限的资源，随着互联网及其应用的发展，不断形成的数据也是一类由互联网衍生而来的重要的人造资源。互联网作为一种战略性新兴产业不断融入到经济社会发展过程中，推各种事业的发展进步。而这种资源正是可以被我们自由免费的使用。

图书资源比较直接，在项目开发过程中，我们可以在图书馆查阅相关的资料，借由图书馆直接的资源，为项目开发提供知识基础。

(3) 人力资源

人是一切事物的核心。大千世界，一切发明，一切创造，都离不开人这一个核心事物。在项目整个的完成过程中，如何完全发挥“人”的作用是至关重要的。人力资源的合理有效利用，有利于提高整个项目完成效率和完成质量。人力资源中项目成员积极性的良好调动，对于整个项目的实践有着极其重要的作用。我们将由一下三个方面合理配置和调动团队的人力资源：

① 明确分配成员角色和职责。每个项目成员从最开始就有明确的分工，每个人有不同的项目任务，各司其职。在实际操作过程中，对于项目遇到的具体问题，我们先进行集中开会商讨，然后将问题的解决具体落实到人，明确该人对该问题负责。

② 人员配备管理计划。团队人数有限，每个人负责的具体问题难度不一，完成进度也不相同。根据每个成员的具体完成情况，我们会实行团队成员的交叉合作，相互帮助，提高项目完成速率。

③ 积极地沟通指导老师。指导老师作为项目的指导者，有较多的经验和较为完善的知识体系。项目构思和项目实践过程中遇到的问题，我们会询问指导老师，老师会提供专业性的指导建议，对于项目的具体实现有着有着重要的指导意义。

1.4.4. 项目成本管理

项目成本是围绕项目而发生的全部费用的总和，包括项目全生命周期各阶段的资源消耗。通过使用科学的方法来计划和控制项目成本，全面考虑问题和管理好项目成本，降低项目风险，从而达到项目高效益的目标。项目成本管理由以下过程组成：

(1) 项目资源计划过程

决定完成项目各项活动需要哪些资源，例如：劳动力、材料、设备、资金、技术等形成生产力的各种要素，以及以及各种资源的需求量。

(2) 项目成本估计过程

对原有的项目成本数据进行归纳分析，并结合回归统计的理论，估计完成项目各活动所需每种资源成本的近似值。

(3) 项目成本预算过程

具将项目内容划分为几个子项目或几个部分，便于同类工作在同组内执行，从而对项目成本进行有效的预算和管理控制。把估计总成本分配到各具体工作。

(4) 项目成本控制过程控制项目预算的改变。

项目成本管理由以上四个方面：项目资源计划、项目成本估计、项目成本预算、项目成本控制组成。四者相互协调作用。我们在项目的具体实现过程中，就在这四个方面对项目的成本进行了讨论控制。

2. 业务分析

2.1. 开发思想

对于数据标注平台，我们分四个阶段寻求系统的解决方案：

1. 阶段一：需求分析和规划

在这个阶段，我们将明确定义数据标注平台的目标用户，分析他们的需求，并进行技术可行性研究，以便为后续的开发工作奠定基础。

2. 阶段二：基本功能实现

在这一阶段，我们将建立用户注册和权限管理系统，实现数据导入和导出功能，同时提供基本的标注工具，确保用户能够轻松开始标注工作。

3. 阶段三：团队协作和审核流程

通过引入团队协作平台和审核流程，我们将加强用户之间的协作能力，提高标注结果的质量和一致性，为团队提供更有效的工作流程。

4. 阶段四：高级标注工具和自动化

在这个阶段，我们将引入更复杂的标注工具，如语义分割和关键点标注，并集成自动化工具，以提高标注效率和适应不同类型的标注任务。

5. 阶段五：性能优化和扩展性

通过性能优化和系统设计，我们将确保平台在大规模数据标注场景下保持高效运行，并考虑未来的扩展性，以应对不断增长的需求。

6. 阶段六：用户反馈和持续改进

在最后一个阶段，我们将积极收集用户反馈，不断迭代改进平台，以确保它始终能够满足用户的实际需求，并保持与技术和行业的最新发展同步。

2.2. 开发方法

对于数据标注平台的开发，我们项目组选择采用敏捷开发方法中的 DevOps 模型，结合增量式开发的特点，以及 Scrum 方法的团队协作和持续改进为辅助，以构建一个高效、灵活且用户需求驱动的系统。

DevOps 模型强调开发团队和运维团队之间的紧密合作，通过自动化流程、持续集成和持续交付，实现快速的软件交付和反馈循环。这与数据标注平台的需求相符，因为标注平台需要不断适应用户的标注需求和新的数据类型。DevOps 的

特点如下：

- 持续集成和持续交付：

强调通过自动化工具和流程，实现代码的频繁集成和交付，确保软件质量和快速部署。

- 自动化测试：

通过自动化测试工具，保障软件在不同阶段的质量，尤其对于标注平台这样需要高质量标注结果的系统。

- 持续监控和反馈：

强调通过监控系统的运行情况和用户反馈，快速发现和解决问题，确保系统的稳定性和用户满意度。

- 团队协作：

DevOps 强调开发团队和运维团队的紧密协作，确保开发出的软件能够顺利地部署和运行。

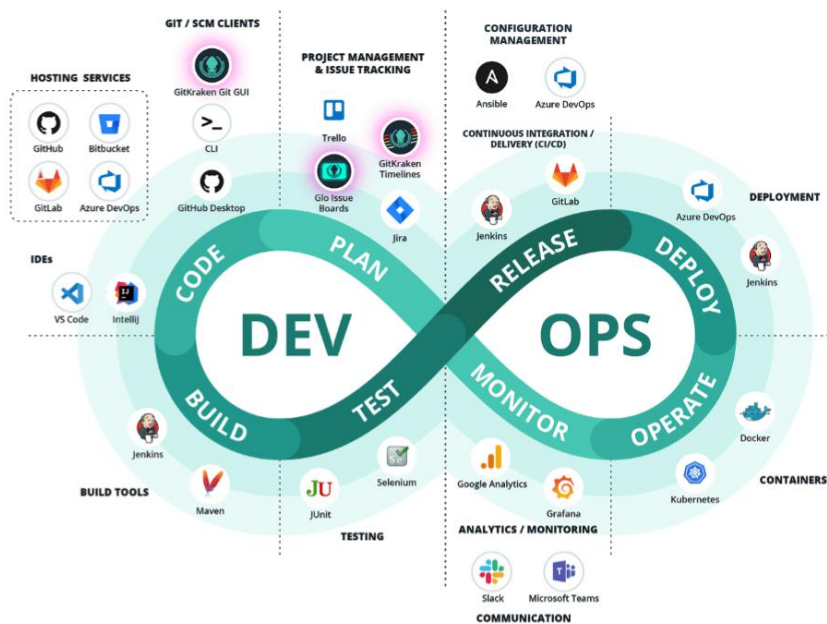


图 2- 1 DevOps 模型示意图

综上所述，DevOps 模型的特点很适合本系统的开发。DevOps 模型通过整合开发和运维活动，以及自动化关键流程，提高了软件开发和交付的效率和质量，使得团队更能适应变化和满足用户需求。

辅以 Scrum 方法，我们将项目划分为短周期的 Sprint，每个 Sprint 专注于实现特定的功能或解决特定的问题，通过每日站会和 Sprint 回顾，促进团队成员之间的沟通与合作。Scrum 的用户故事为基础的开发方法也符合数据标注平台需求，以确保系统能够满足用户的具体需求。

Scrum 编程方法特点如下：

- 快速迭代：

Scrum 强调迭代开发，通过每个 Sprint（冲刺）的短周期来快速响应变化的需求，使产品逐步完善。

- 用户故事为基础：

类似于 XP, Scrum 也注重以用户故事 (User Stories) 为基础进行分析设计和测试用例的制定, 确保满足用户需求。

● 持续测试:

Scrum 提倡对开发中的系统进行持续的、大规模的测试, 保障软件的质量和可用性。

● 鼓励重构:

与 XP 相似, Scrum 鼓励长时间的分析与设计, 并鼓励在项目进展中进行重构, 以提高系统的可维护性和适应性。

● 团队协作:

虽然 Scrum 没有像 XP 那样规定结对编程, 但它强调团队的自组织和协作, 通过 Daily Scrum 等活动促进团队成员之间的沟通与合作。

以下是 Scrum 开发方法中的主要角色定位及其作用图:

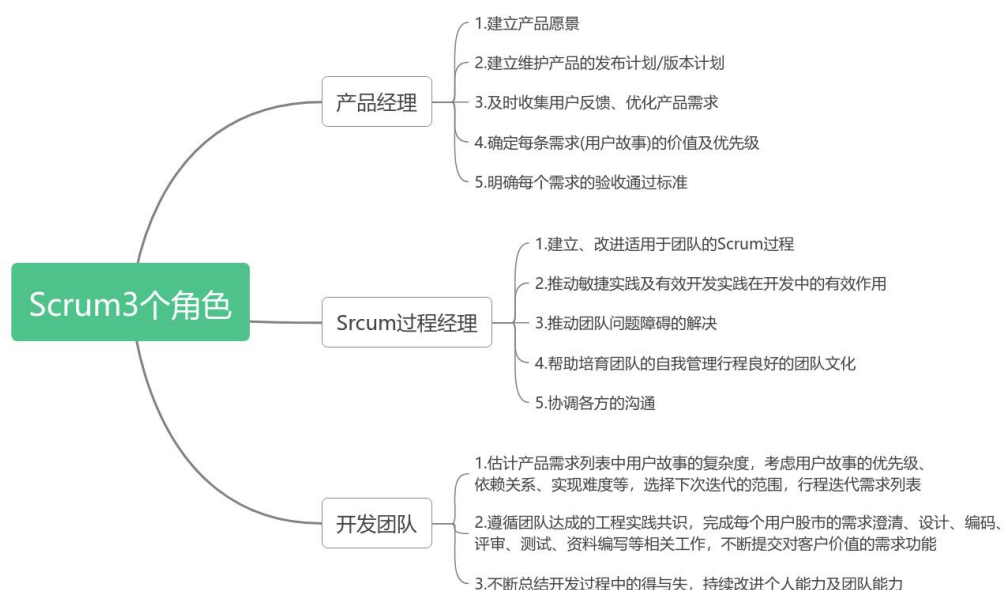


图 2-2Scrum 主要角色定位及其作用图

综合而言，我们选择 DevOps 模型作为主要的开发方法，以其快速、自动化和持续改进的特点，辅以 Scrum 方法的团队协作和用户需求驱动，以确保数据标注平台的高效、灵活和用户满意。

2.3. 系统功能性需求

该评论数据分析系统主要任务是从市政府 12345 在线投诉平台爬取投诉举报信息、对投诉线索数据进行格式化并提取关键信息、对投诉线索智能筛查、对投诉线索统计分析。具体分析需求如下表所示：

表 2- 1 系统功能需求分析表

系统功能需求分析表	
信息抽取标注模块	从文本中提取出有用的信息，为后续的信息处理和分析提供基础数据。
文本分类标注模块	是将文本划分为不同的类别，为后续的信息处理和分析提供基础数据。
图像文本标注模块	框选出各种场景图片中出现的文本，标注结果会作为文本检测模型的训练数据来优化模型效果。
图像分类模块	将图片按照一定的类别进行归类，为后续的信息处理和分析提供基础数据。

2.4. 系统非功能性需求

2.4.1. 安全性

本项目面向金融领域的专业人员，因此数据的安全性至关重要。为了防止敏感金融信息泄露，本系统将采用强大的加密技术，确保数据在传输和处理过程中始终得到保护。此外，系统还将具备强大的抗外部攻击能力，以确保金融数据不

受未经授权的访问，保障系统的正常运行。在金融行业，数据的安全性是系统设计的首要考虑因素，以防范潜在的风险和影响。

2.4.2. 美观性

良好的人机交互、严密的操作逻辑和简洁美观、有亲和力的界面整体设计往往更受用户的欢迎，所以该系统应使界面变得有个性化，使操作变得舒适、简单、自由。

2.4.3. 直观性

对于一个数据分析系统而言，往往将庞大的数据精简化，以一种直观乐于接受的方式将分析的结果展示出来，所以为了使使用者能够从该系统中直观地获取相应信息，本系统的展现形式应该具备直观性。

2.4.4. 可变性

本项目面向金融专业人员，对于金融文档的智能抽取方面，需要进行多维度的分类和提取，而在机器学习的智能抽取上可能会面临一些误差和不准确性。为了满足金融行业不断变化的需求，系统的设计必须具备高度的灵活性和可扩展性。系统将采用先进的机器学习算法，以便在出现误差或不准确情况时能够及时对系统进行强化学习和调整，确保系统持续适应金融信息抽取的新需求。在金融领域，可变性是确保系统长期有效运行和满足用户不断变化需求的重要因素。