# 行业咨询分析系统

## 一、确定业务需求与需求捕获

### 系统目标

行业资讯分析系统旨在帮助企业或用户从多个信息源（如新闻网站、社交媒体、行业报告）中采集、整理、分析和可视化行业资讯，辅助战略决策。

### 核心功能

1. 自动采集指定来源的行业资讯。

2. 对资讯进行分类、关键词提取、情感分析等处理。

3. 提供资讯检索与高级筛选功能。

4. 为用户推送个性化资讯摘要。

5. 提供可视化分析报告（如行业趋势图、热词云等）。

## 二、需求特性表

|  |  |
| --- | --- |
| 编号 | 特性描述 |
| FEAT01 | 管理员具备对资讯来源的全面管理权限，可灵活添加新的资讯源（如指定网站URL、配置RSS订阅链接或接入第三方API接口）、删除不再使用的资讯源，以及根据实际需求调整资讯源的抓取参数（如抓取频率、内容格式等）。 |
| FEAT02 | 系统内置定时任务模块，能够按照预设时间间隔（如每小时/每天）自动抓取已配置的资讯源中的最新行业资讯内容，并将抓取结果以结构化格式存储至本地数据库中，确保数据可追溯性和长期可用性。 |
| FEAT03 | 系统通过集成自然语言处理（NLP）技术，对抓取的资讯内容进行深度分析，包括提取核心关键词、自动分类至预设行业类别（如科技、金融、医疗等）、识别文本情感倾向（如正面、负面、中性），并生成可用的分析元数据。 |
| FEAT04 | 用户可通过多维度筛选条件（如关键词匹配、时间范围限制、情感类型过滤、分类标签选择等）对存储的资讯进行快速检索，并支持排序、分页展示等高级功能，以提升信息获取效率。 |
| FEAT05 | 用户可对感兴趣的资讯进行个性化操作，包括一键收藏至个人书签库、添加自定义标签以便分类管理、以及撰写评论与其他用户互动，所有操作均记录于用户历史记录中。 |
| FEAT06 | 系统基于用户行为数据（如浏览历史、收藏偏好、评论内容等）构建用户画像，并通过智能算法生成个性化资讯推荐列表，优先展示与用户兴趣高度匹配的内容，提升用户体验。 |
| FEAT07 | 管理员或授权用户可通过可视化工具生成行业分析报告，支持自定义数据维度（如资讯来源分布、关键词热度、情感趋势等），并可导出为PDF、Excel等格式，便于分享与决策参考。 |
| FEAT08 | 用户可订阅每日或每周的行业摘要报告，系统根据用户偏好自动汇总相关资讯内容，并通过邮件或消息推送功能将报告发送至用户指定地址，确保用户及时获取关键信息。 |
| FEAT09 | 系统提供多语言界面支持（如中文、英文、西班牙语等），并兼容Web端、移动端（iOS/Android）等多种访问设备，确保用户在不同场景下均可无缝使用核心功能。 |
| FEAT10 | 系统生成并自动向用户推送行业摘要报告。 |

**三、建立用例模型**

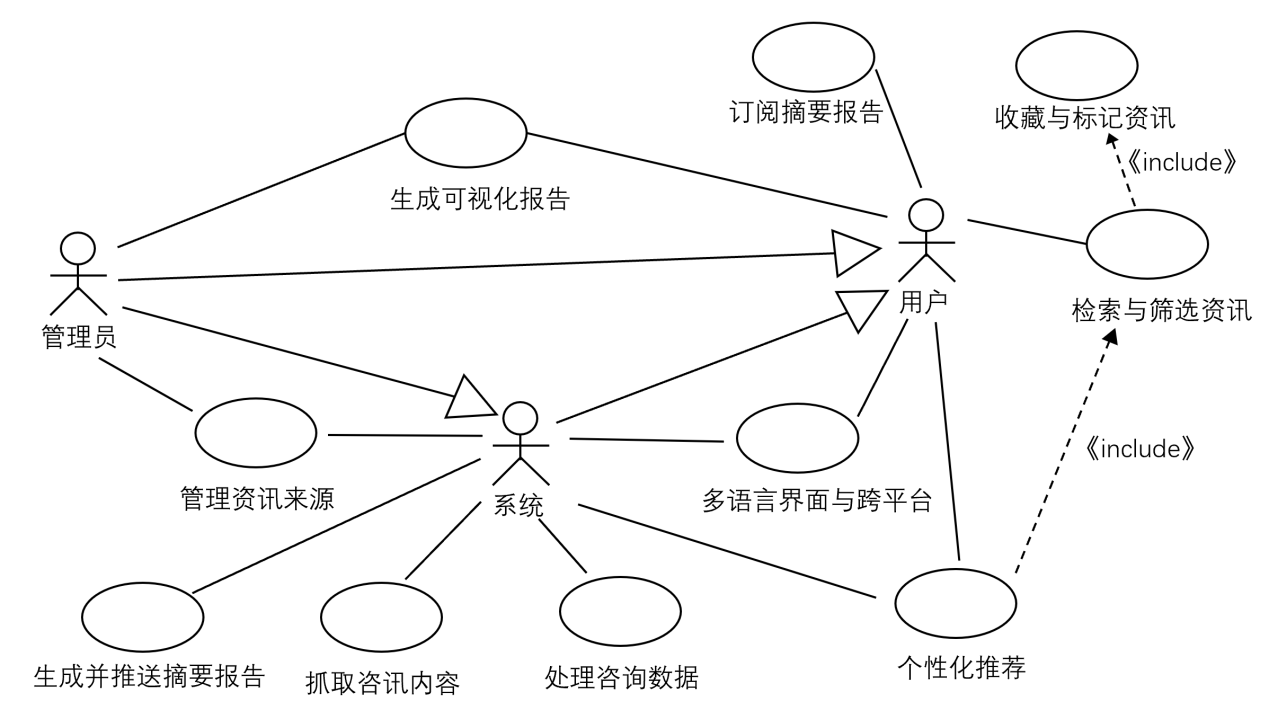
1、识别参与者：

管理员、用户、系统

2、合并特性获得用例

|  |  |  |
| --- | --- | --- |
| 参与者 | 特性 | 用例 |
| 管理员 | FEAT01,管理员可以添加、删除、配置资讯来源（如网站、RSS、API）。 | UC01.管理咨询来源 |
| FEAT07,管理员或用户可以生成可视化的行业分析报告。 | UC02.生成可视化报告 |
| 系统 | FEAT02,系统定时抓取最新行业资讯，并存储到本地数据库中。 | UC03.抓取咨讯内容 |
| FEAT03,系统对抓取的内容进行自然语言处理，包括关键词提取、分类、情感分析等。 | UC04.处理咨询数据 |
| FEAT06,系统可以根据用户行为生成个性化推荐。 | UC05.个性化推荐 |
| FEAT09,系统提供多语言界面与跨平台访问能力（Web/移动端）。 | UC06.多语言界面与跨平台 |
| FEAT10,系统生成并自动向用户推送行业摘要报告。 | UC07.生成并推送摘要报告 |
| 用户 | FEAT04,用户可以按关键词、时间、情感类型等条件进行资讯检索与筛选。 | UC08.检索与筛选资讯 |
| FEAT05,用户可以收藏资讯、添加标签、编写评论。 | UC09.收藏与标记咨讯 |
| FEAT08,用户可以订阅每日/每周行业摘要报告，系统自动推送。 | UC10.订阅摘要报告 |

3、绘制用例图



4、用例描述

|  |  |  |
| --- | --- | --- |
| 用例编号 | UC01 | |
| 用例名称 | 管理资讯来源 | |
| 用例概述 | 管理员通过后台界面添加、删除和配置资讯来源，设置抓取参数 | |
| 主参与者 | 管理员 | |
| 后置条件 | 系统可抓取并分析经过来源筛选的咨讯内容 | |
| 基本事件流 | 步骤 | 活动 |
| 1 | 管理员进入后台界面 |
| 2 | 管理员进行添加、删除、配置等操作 |
| 补充说明 | 包含来源有效性检测和批量导入/导出功能，需记录操作日志。 | |

|  |  |  |
| --- | --- | --- |
| 用例编号 | UC02 | |
| 用例名称 | 生成可视化报告 | |
| 用例概述 | 管理员或授权用户选择数据维度和可视化类型，系统生成可导出的行业分析报告。 | |
| 主参与者 | 管理员/高级用户 | |
| 前置条件 | 管理员或用户已登录系统并具有相应权限，进入“可视化报告生成”模块 | |
| 后置条件 | 生成的报告保存至系统数据库，可供导出或后续查看 | |
| 基本事件流 | 步骤 | 活动 |
| 1 | 系统显示可选择的数据维度（如时间段、行业类别、关键词等）和图表类型（柱状图、折线图、词云等） |
| 2 | 选择所需的维度、图表类型，并点击“生成报告”按钮 |
| 3 | 系统根据用户选择内容，进行数据分析并生成对应的可视化图表 |
| 4 | 系统展示生成的报告结果页面，用户可点击“导出PDF”或“导出Excel”进行保存 |
| 补充说明 | 支持PDF/Excel格式导出，可保存报告模板。 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | | UC03 | |
| 用例名称 | | 抓取资讯内容 | |
| 用例概述 | | 系统按照预设规则自动从配置的资讯源抓取最新内容，存储结构化数据至数据库。 | |
| 主参与者 | | 系统 | |
| 前置条件 | | 资源配置正常，且网络连接正常 | |
| 后置条件 | | 抓取内容成功存入数据库 | |
| 基本事件流 | 步骤 | | 活动 |
| 1 | | 系统读取预设的资讯源配置 |
| 2 | | 系统发起请求获取原始内容，根据规则解析标题、正文、发布时间等 |
| 3 | | 将结构化数据存入数据库，记录抓取状态 |
| 补充说明 | 包含异常监控和重试机制，记录抓取日志。 | | |

|  |  |  |
| --- | --- | --- |
| 用例编号 | UC04 | |
| 用例名称 | 处理资讯数据 | |
| 用例概述 | 系统对原始资讯进行NLP处理，生成分析元数据。 | |
| 主参与者 | 系统 | |
| 前置条件 | 原始咨讯已存入数据库 | |
| 后置条件 | 更新数据库内容 | |
| 基本事件流 | 步骤 | 活动 |
| 1 | 系统读取未处理的原始资讯文本 |
| 2 | 调用分类模型匹配预设分类体系 |
| 3 | 提取关键词和情感极性（正面/负面/中性） |
| 4 | 存储处理结果 |
| 补充说明 | 需预设分类体系和情感分析规则库。 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC05 | | |
| 用例名称 | 个性化推荐 | | |
| 用例概述 | 基于用户历史行为数据，通过推荐算法生成个性化资讯列表。 | | |
| 主参与者 | 系统 | | |
| 前置条件 | 用户已登录，用户行为数据大于0 | | |
| 后置条件 | 推荐列表缓存至用户会话 | | |
| 基本事件流 | | 步骤 | 活动 |
| 1 | 系统获取用户历史行为数据以及相似用户群组 |
| 2 | 计算资讯协同过滤评分（基于用户群偏好）和内容相似度评分（基于关键词向量）。 |
| 3 | 加权排序生成推荐列表 |
| 补充说明 | | 采用协同过滤+内容相似度混合算法。 | |

|  |  |
| --- | --- |
| 用例编号 | UC06 |
| 用例名称 | 多语言界面与跨平台 |
| 用例概述 | 系统提供多语言切换功能，并适配Web/移动端不同设备。 |
| 主参与者 | 系统 |
| 补充说明 | 采用响应式设计，语言包支持动态加载。 |

|  |  |
| --- | --- |
| 用例编号 | UC07 |
| 用例名称 | 生成并推送摘要报告 |
| 用例概述 | 系统自动汇总指定周期的关键资讯，推送给订阅用户。 |
| 主参与者 | 系统 |
| 补充说明 | 推送渠道包括邮件/站内消息。 |

|  |  |  |  |
| --- | --- | --- | --- |
| 用例编号 | UC08 | | |
| 用例名称 | 检索与筛选资讯 | | |
| 用例概述 | 用户通过组合条件检索资讯，支持结果排序和分页。 | | |
| 主参与者 | 用户 | | |
| 前置条件 | 用户进入搜索页面 | | |
| 后置条件 | 搜索结果页显示匹配的资讯列表 | | |
| 基本事件流 | | 步骤 | 活动 |
| 1 | 用户输入关键词，可选填分类、时间范围等筛选条件。 |
| 2 | 系统返回匹配结果（默认按相关性排序），分页每页20条。 |
| 3 | 用户点击“高级搜索”可使用布尔语法。 |
| 补充说明 | | 提供高级搜索语法和搜索历史记录。 | |

|  |  |
| --- | --- |
| 用例编号 | UC09 |
| 用例名称 | 收藏与标记资讯 |
| 用例概述 | 用户收藏资讯条目，添加自定义标签和评论。 |
| 主参与者 | 用户 |
| 补充说明 | 支持标签云展示和评论互动。 |

|  |  |
| --- | --- |
| 用例编号 | UC10 |
| 用例名称 | 订阅摘要报告 |
| 用例概述 | 用户设置报告订阅偏好，管理系统推送设置。 |
| 主参与者 | 用户 |
| 补充说明 | 允许随时修改订阅。 |

6、划分用例优先级

|  |  |  |
| --- | --- | --- |
| 优先级 | 用例 | 说明 |
| 1 | UC03.抓取咨讯内容 | 系统基础功能，为其他分析/推荐功能提供数据源，直接影响系统核心价值 |
| UC04.处理咨询数据 | 资讯处理的核心环节，支撑检索、推荐等高阶功能 |
| UC08.检索与筛选资讯 | 用户最常用的核心功能，直接影响用户体验 |
| 2 | UC01.管理咨询来源 | 管理员高频操作，但仅影响后台维护效率 |
| UC05.个性化推荐 | 提升用户粘性的重要功能，但依赖用户行为数据积累 |
| UC02.生成可视化报告 | 重要增值服务，但主要面向高级用户 |
| 3 | UC09.收藏与标记咨讯 | 增强用户参与度，属于优化型功能 |
| UC10.订阅摘要报告 | 需集成消息推送系统，实现成本较高 |
| UC07.生成并推送摘要报告 | 自动化功能，依赖其他数据处理功能 |
| UC06.多语言节目与跨平台 | 影响范围广但开发成本高，适合系统成熟后扩展 |

优先级划分依据：

1. 核心基础功能（必须优先实现）：

UC03/UC04：构成系统数据处理管道

UC08：满足用户最基本需求

1. 重要增强功能（系统具备基本价值后实现）：

UC01/UC05/UC02：分别提升数据质量、用户体验和商业价值

3、优化扩展功能（可后续迭代）：

UC09/UC10/UC07/UC06：完善系统生态

**四、建立类模型**

1. 发现类

根据系统目标和需求特性，可以发现以下主要的类：

**Class1：管理员（Admin）：**

·负责管理资讯来源、生成报告、进行系统配置等。

**Class2：用户（User）：**

·负责浏览、收藏、检索资讯，管理个人订阅等。

**Class3：资讯（NewsItem）：**

·存储抓取到的资讯内容。

**Class4：资讯来源（NewsSourse）**

·描述资讯的来源，包括网站、API等。

**Class5：行业分析报告（AnalysisReport）**

·生成和存储行业分析报告。

**Class6：用户行为（UserBehavior）**

·记录用户的行为（如浏览、收藏等），供个性化推荐使用。

**Class7：推荐引擎（RecommendationEngine）**

·根据用户行为生成个性化的资讯推荐列表。

**Class8：定时任务（scheduledTask）**

·定时抓取资讯内容并存储。

**Class9：报告推送（ReportPush）**

·负责推送报告和摘要给用户。

**Class10：搜索引擎（SearchEngine）**

·实现资讯的检索和筛选功能。

后期完善系统功能时可选的类：

**Class11：多语言支持类（LanguageSupport）**

·负责系统的多语言切换功能，包括不同语言的支持，语言资源的管理等。

**Class12：设备适配类（DeviceAdaptation）**

·负责适配不同设备的界面需求，包括 Web 和移动端（iOS/Android）的不同显示。

1. 类关联分析

**Relationship1：管理员与资讯来源：**

**·**管理员能够添加、删除、配置资讯来源。

**Relationship2：用户与资讯：**

·用户可以收藏、标记和评论资讯。

**Relationship3：用户与用户行为：**

·用户行为数据影响个性化推荐。

**Relationship4：用户行为与推荐引擎：**

·推荐引擎通过用户行为数据来生成推荐。

**Relationship5：管理员与行业分析报告：**

·管理员可以生成行业分析报告。

**Relationship6：用户与报告推送：**

·系统自动推送报告摘要到用户指定的地址。

**Relationship7：搜索引擎与资讯：**

·搜索引擎根据用户输入的条件对资讯进行检索和筛选。

1. 细化类的属性

**管理员（Admin）：**

**·管理员ID：**唯一标识管理员。

**·用户名：**管理员的用户名。

**·权限:**管理员的权限级别。

**·操作日志：**记录管理员的所有操作。

**用户（User）：**

**·用户ID：**唯一标识用户。

**·用户名：**用户的名字。

**·邮箱：**用户的联系邮箱。

**·订阅偏好：**用户对行业摘要的订阅偏好。

**·收藏资讯：**用户收藏的资讯列表。

**·行为记录：**用户的浏览历史、评论、收藏等行为数据。

**资讯（NewsItem）：**

**·资讯ID：**唯一标识资讯。

**·标题：**资讯的标题。

**·内容：**资讯的正文内容。

**·来源：**资讯的来源（如网站、API等）。

**·发布时间：**资讯的发布时间。

**·分类：**资讯的行业分类（如科技、金融、医疗等）。

**·情感分析：**正面、负面或中性情感分析。

**·关键词：**从资讯内容中提取的关键词。

**·抓取时间：**资讯的抓取时间。

**资讯来源（NewsSourse）**

**·来源ID：**唯一标识资讯来源。

**·URL：**指定网站或RSS订阅链接。

**·API接口：**第三方API接口。

**·抓取参数：**抓取频率、内容格式等配置信息。

**行业分析报告（AnalysisReport）**

**·报告ID：**唯一标识报告。

**·报告名称：**报告的名称。

**·生成时间：**报告生成的时间。

**·内容：**报告的具体内容（可能包括趋势图、热词云等）。

**·导出格式：**导出报告的格式（如PDF、Excel等）。

**用户行为（UserBehavior）**

**·行为ID：**唯一标识一条用户行为记录。

**·行为类型：**浏览、收藏、评论等行为类型。

**·资讯ID：**行为关联的资讯ID。

**·时间戳：**行为发生的时间。

**·内容：**与行为相关的内容（如评论内容）。

**推荐引擎（RecommendationEngine）**

**·推荐ID：**唯一标识一个推荐结果。

**·生成时间：**推荐生成的时间。

**·推荐内容：**推荐的资讯列表。

**·用户ID：**与该推荐结果相关的用户ID。

**·推荐依据：**推荐的依据（如用户历史行为、协同过滤等）。

**定时任务（scheduledTask）**

**·任务ID：唯一标识定时任务。**

**·任务类型：任务的类型（如资讯抓取）。**

**·执行频率：任务的执行频率（如每天、每小时）。**

**·下次执行时间：下一层执行时间。**

**报告推送（ReportPush）**

**·推送ID：**唯一标识一个推送记录。

**·推送时间：**报告推送的时间。

**·推送内容：**报告的内容。

**·推送渠道：**邮件、站内消息等。

**搜索引擎（SearchEngine）**

**·搜索ID：**唯一标识一个搜索记录。

**·查询关键词：**用户输入的搜索关键词。

**·筛选条件：**如时间范围、分类标签等。

**·排序方式：**按相关性、时间等排序方式。

**多语言支持类（LanguageSupport）**

**·语言列表：**支持的语言列表。

**·当前语言：**当前语言设置。

**设备适配类（DeviceAdaptation）**

**·设备类型：**设备的具体类型（如Web、iOS、Android）。

**·界面布局：**根据设备调整布局。

1. 细化类的操作

**管理员（Admin）：**

**·添加资讯来源（）：**将新的资讯来源添加到系统中。

·**删除资讯来源（）：**删除不再使用的资讯来源。

·**修改资讯来源（）：**修改资讯来源的抓取参数。

**用户（User）：**

**·检索资讯（）：**根据条件（如关键词、情感类型等）检索资讯。

·**收藏资讯（）：**收藏感兴趣的资讯。

·**标记资讯（）：**为资讯添加自定义标签。

·**订阅报告（）：**根据用户的订阅偏好订阅报告。

**资讯（NewsItem）：**

·**提取**关键词()**：**从资讯内容中提取关键词。

·**分类**()****：将资讯归类到相应的行业类别。

·**情感**分析()**：**对资讯进行情感分析。

**资讯来源（NewsSourse）：**

·**配置**抓取参数()**：**设置资讯来源的抓取参数，如频率和格式。

**行业分析报告（AnalysisReport）：**

**·生成报告()**：生成行业分析报告。

·**导出**报告()**：**将报告导出为不同格式（如PDF、Excel等）。

**用户行为（UserBehavior）：**

**·记录行为()**：记录用户的行为。

·**生成**用户画像()**：**根据行为数据生成用户画像。

**推荐引擎（RecommendationEngine）：**

**·生成推荐()**：根据用户画像生成个性化推荐。

·**推送**推荐()**：**将推荐结果推送给用户。

**定时任务（scheduledTask）：**

·**执行**任务()**：**按照设定的时间间隔执行抓取任务。

·**管理**任务()****：管理定时任务（添加、删除、修改）。

**报告推送（ReportPush）：**

**·推送报告()**：将报告通过邮件或消息推送给用户。

**搜索引擎（SearchEngine）：**

·**执行**搜索()**：**根据条件执行搜索并返回结果。

**多语言支持类（LanguageSupport）：**

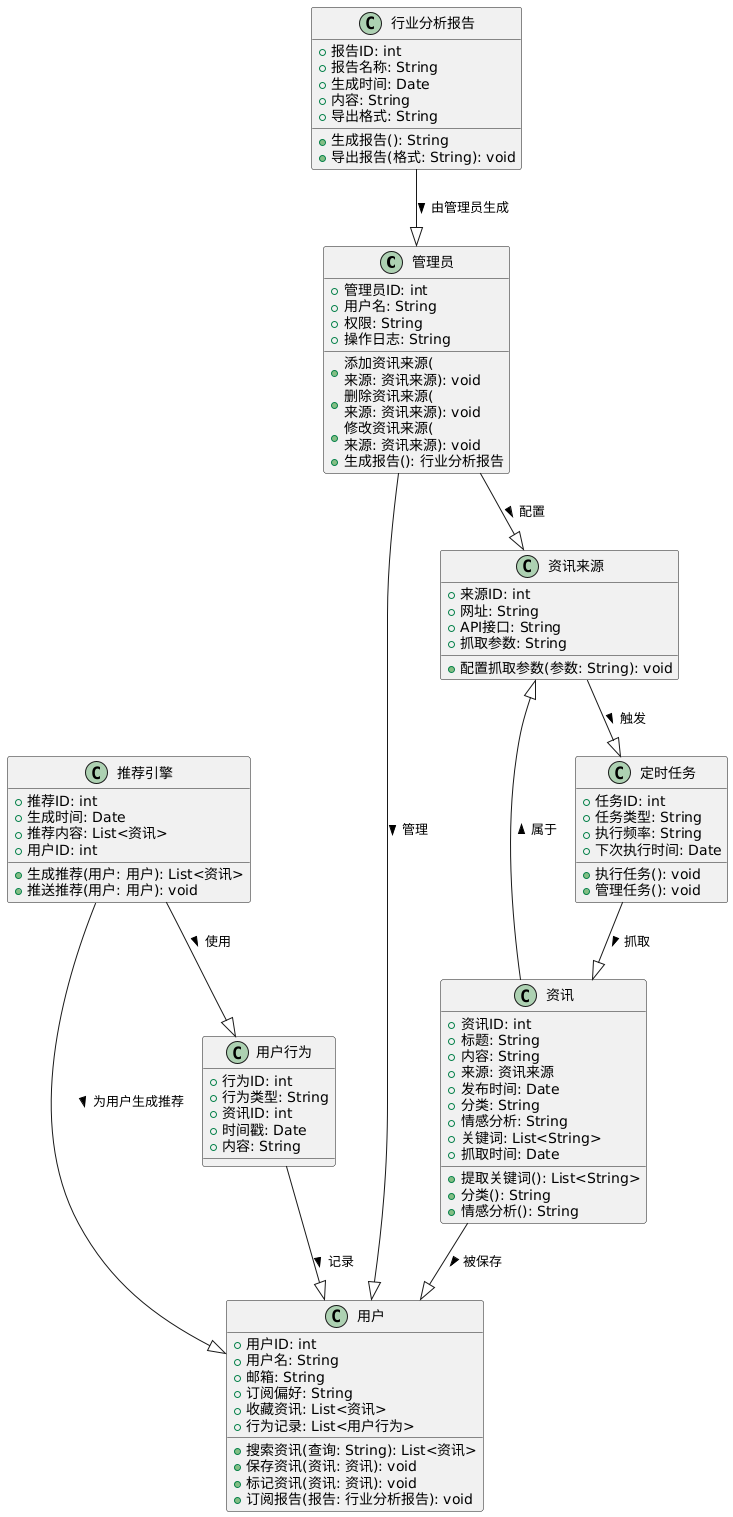
**·切换语言（）：**切换系统的显示语言。

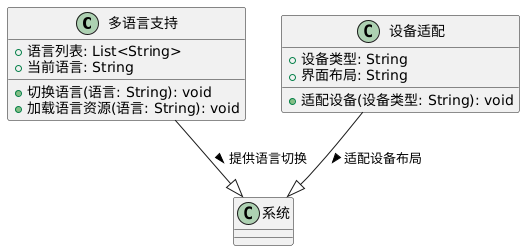
**·加载语言资源（）：**加载指定语言的翻译资源。

**设备适配类（DeviceAdaptation）：**

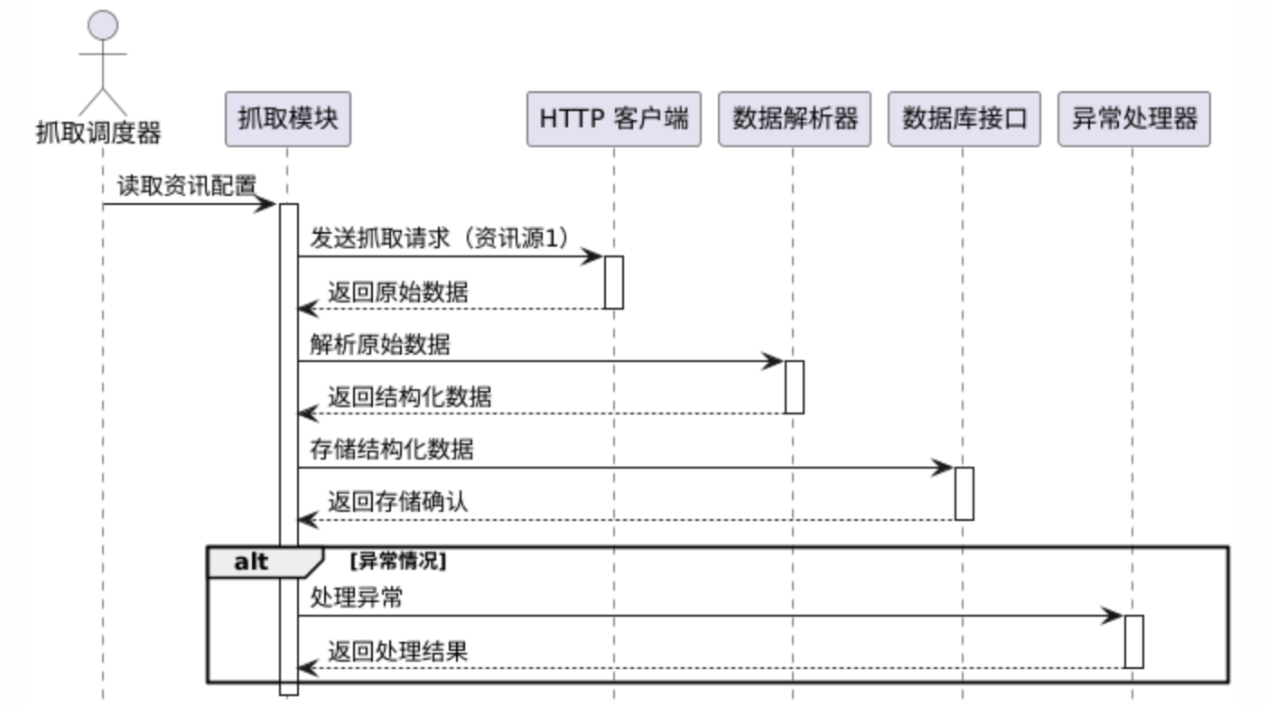
**·适配设备（）：**根据不同的设备调整界面布局。

1. 类图



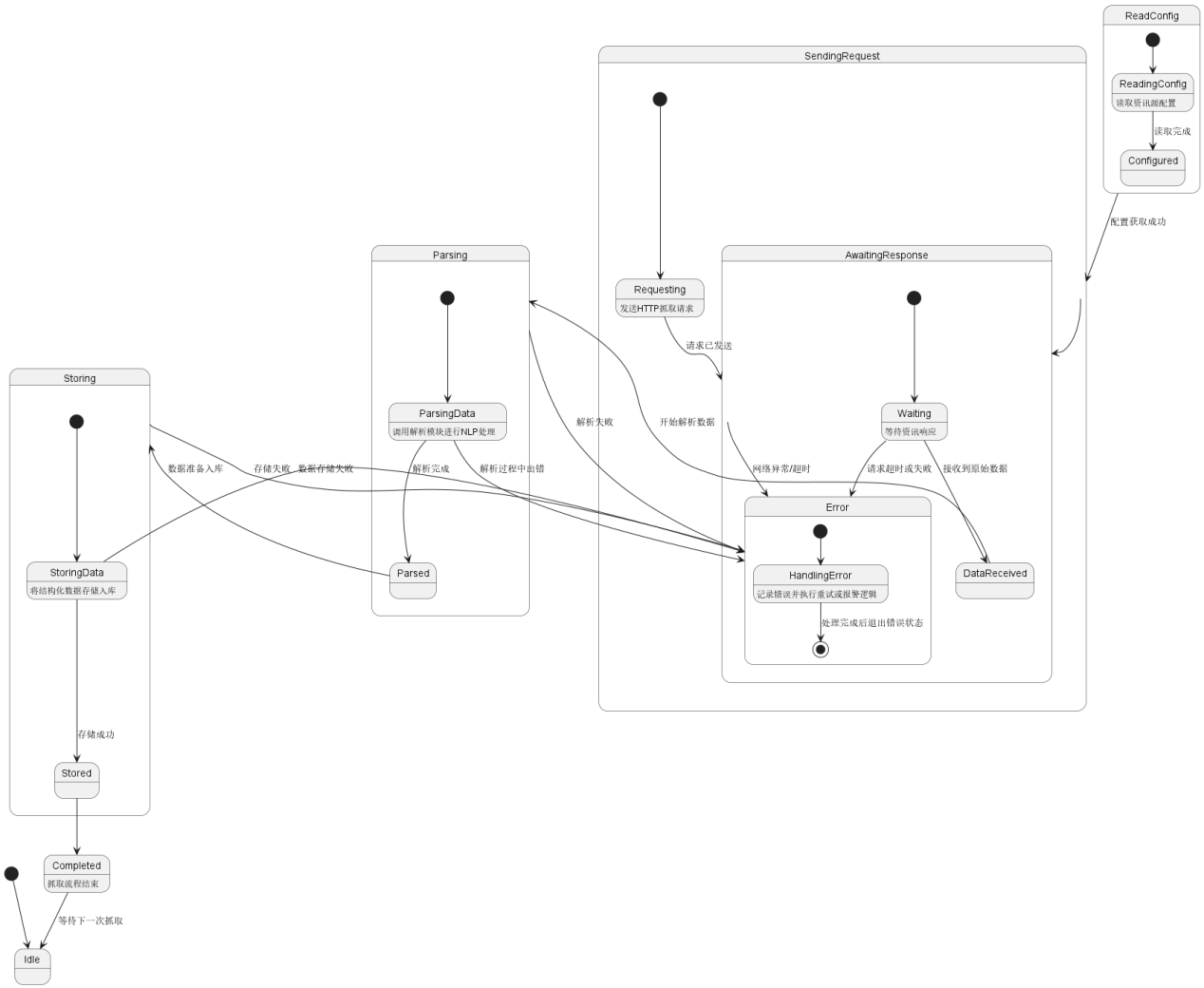


**五、建立交互图**



UC03.抓取咨讯内容

1. **建立状态图&用户界面设计**

****

状态图



用户界面