****

**中文事件语料库构建方法研究**

需求规约

**Version 1.0**

2019.12

小组成员：

1651290 夏宇宁，1751139 程伊楠，1651701 魏楠，

1751188 李佳诺，1751694 王菲凡，1751924 秦琳博，

1750981 陈磊

所在院系： 软件学院

学科专业： 软件工程

指导老师： 杜庆峰

1. 介绍

信息抽取任务泛指在非结构化的文本中获取结构化的数据的过程，通常包含命名实体识别（Named Entity Recognition）、关系抽取（Relation Extraction）与事件提取（Event Extraction）三种任务，任务间层层递进。命名实体识别任务是在文本中查找每个提及的命名实体并标记 其类型的过程，关系抽取任务是对文本实体之间的语义关系进行分类，事件提取是提取实体之间参与的事件信息。

如何针对中文提出一个可行的事件语料扩充方案是亟待解决的问题。 本课题以政治事件为对象，主要探讨如何利用基于句法的谓词-论元提取模型与字典，构建事件标注模型，从而获得中文事件语料库，并结合机器学习事件提取模型探索验证语料库可靠性的方法，根据语料库评估结果迭代优化事件标注模型，以创建高质量中文事件语料库。

1. 整体描述

2.1项目特点

本课题将利用短语句法解析树提取句式中“谓语-论元结构”，利用成型的事件分类体系构造事件触发词字典，两者结合构建事件信息标注模型。例如利用政治领域已存在的事件分类体系CAMEO，构造触发词字典，标注政治事件触发词与事件元素，获取丰富的政治事件语料信息；充分研究当前语料评价标准的前提下，结合人工标注与机器学习事件提取模型， 验证语料的可靠性；对研究结果进行分析与总结。

2.2项目特色

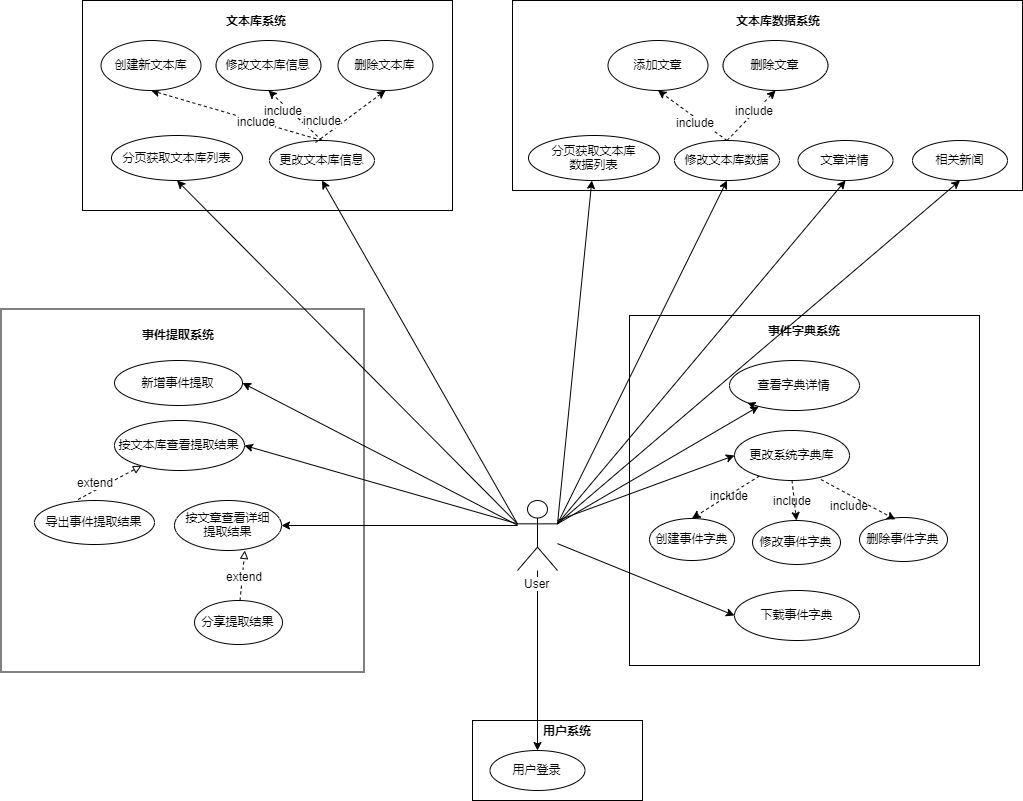
项目特色方面，从语料库扩展的角度，本课题将对基于句法的模式匹配的事件语料标注技术展开深入分 析，并重点关注其在应用字典匹配过程中与事件元素本身的映射关系，从横向上拓展了事件 语料提取技术的领域宽度。

课题应用价值方面，首先，基于中文的事件提取在工业界目前应用不多，对基于句法的模式匹配事件标注技术的研 究有助于为文本摘要、信息提取等各种实际应用场景提供合适的技术解决方案；其次，对于语料的可靠性研究，有助于构建更加具备普适性与公开性的中文事件提取模型，方 便相关学者在语料训练上的选择，确保事件提取模型的鲁棒性。

1. 系统与其他系统的接口
2. 主要的功能需求描述

4.1文字描述

4.2用例模型



|  |
| --- |
| **用例:** 登录 |
| **ID：**UC1 |
| **参与者：** 用户 |
| **前置条件:** 用户拥有登录的账号 |
| **基本事件流:**  1. 用户在指定位置输入用户名和密码  2. 用户点击“登录”按钮  3. 如果用户输入的用户名与密码匹配：  3.1 用户成功登入系统  5. 如果用户输入的用户名与密码不匹配：  5.1 登录操作失败 |
| **后置条件：**用户可以使用系统提供的所有功能 |

|  |
| --- |
| **用例:** 查看词典详情 |
| **ID：**UC2 |
| **参与者：** 用户 |
| **前置条件:** 用户已经登录成功 |
| **基本事件流:**  1. 用户点击要查看的字典项的“查看”（字典）按钮从而向系统发出请求获取字典数据的请求  2. 系统从数据库中查找该字典项的内容  3. 如果用户的请求合法：  3.1 系统返回字典数据给用户 |
| **后置条件：**用户可以在页面上浏览某一字典项的信息 |
| **可选事件流：**  如果用户的请求不合法，系统返回查看字典失败提示 |

|  |
| --- |
| **基用例:** 更改系统字典库 |
| **ID：**UC3 |
| **参与者：** 用户 |
| **前置条件:** 用户已经登录成功 |
| **基本事件流:**  1. 根据用户点击的不同按钮（“修改”，“删除”，“创建事件词典”），进入不同的子用例 |
| **后置条件：**数据库需实时更新用户对字典库的修改 |
| **被包含的用例：**修改词典用例，删除事件词典用例，创建事件词典用例 |

|  |
| --- |
| **子用例:** 修改词典基本信息 |
| **ID：**UC4 |
| **参与者：** 用户 |
| **前置条件:** 用户已经登录成功 |
| **基本事件流:**  1. 用户点击某一字典项的“修改”（字典）按钮从而向系统发出修改字典数据的请求  2. 系统弹出修改字典的页面  3. 用户根据需求填写修改字典页面  4. 用户点击“确认修改”按钮  5. 若用户的修改请求合法：  5.1 系统将用户的修改写入数据库 |
| **后置条件：**数据库需实时更新用户对字典库的修改 |
| **可选事件流：**  若用户的修改请求不合法，系统返回修改字典错误提示 |

|  |
| --- |
| **子用例:** 创建事件词典 |
| **ID：**UC5 |
| **参与者：** 用户 |
| **前置条件:** 用户已经登录成功 |
| **基本事件流:**  1. 用户点击“创建事件词典”按钮从而向系统发出创建字典的请求  2. 系统弹出新建字典的页面  3. 用户根据需求填写新建字典页面  4. 用户点击“确认创建字典”按钮  5. 若用户的创建字典请求合法：  5.1 系统将用户的修改写入数据库 |
| **后置条件：**数据库需实时更新用户对字典库的修改 |
| **可选事件流：**  若用户的创建字典请求不合法，系统返回新建字典失败提示 |

|  |
| --- |
| **子用例:** 删除事件词典 |
| **ID：**UC6 |
| **参与者：** 用户 |
| **前置条件:** 用户已经登录成功 |
| **基本事件流:**  1. 用户点击某一字典项的“删除”（字典）按钮从而向系统发出删除字典的请求  2. 若用户的删除字典请求合法：  2.1 系统把该字典从数据库删除 |
| **后置条件：**数据库需实时更新用户对字典库的修改 |
| **可选事件流：**  若用户的删除字典请求不合法，系统返回删除字典失败提示 |

|  |
| --- |
| **用例:** 下载事件字典 |
| **ID：**UC7 |
| **参与者：** 用户 |
| **前置条件:** 用户已经登录成功 |
| **基本事件流:**  1. 用户点击某一字典项对应的“下载字典”按钮  2. 系统提示用户选择字典下载的本地位置  3．下载字典到本地文件 |
| **后置条件：**用户可以在页面上浏览从某一文本中提取到的事件信息 |
| **可选事件流：**  若字典下载到本地的过程中出现错误，系统显示下载错误提示信息 |

|  |
| --- |
| **用例:** 新增事件提取 |
| **ID：**UC8 |
| **参与者：** 用户 |
| **前置条件:** 用户已经登录成功 |
| **基本事件流:**  1. 用户按要求填写新增事件提取的列表，包含：  - 选择提取所用到的文本库  - 填写事件提取的名称  - 选择提取级别  - 选择提取字段  - 选择提取事件所用字典  2. 用户点击“开始提取”，向系统提交事件提取申请 |
| **状态：**在用户提交事件提取申请后，其执行状态分为：排队中，执行中，已完成，已取消 |
| **后置条件：**系统实时更新事件提取的状态 |

**状态**的用语规范

|  |
| --- |
| **用例:** 按文本库查看事件提取结果 |
| **ID：**UC9 |
| **参与者：** 用户 |
| **前置条件:** 用户已经登录成功，且该事件提取的状态是 已完成 |
| **基本事件流:**  1. 用户点击某一文本库对应的“查看事件提取结果”图标，向系统发出查看某文本库内事件提取结果的请求  2. 系统从数据库中查找分析结果  3. 系统按照文章分行展示分析结果  - 文本ID  - 标题  - 正文  - 事件数量 |
| **后置条件：**用户可以查看每一篇文章的详细事件提取结果 |

|  |
| --- |
| **用例:** 查看某文章的详细事件提取结果 |
| **ID：**UC10 |
| **参与者：** 用户 |
| **前置条件:** 用户已经进入了按文本输出结果的事件提取页面 |
| **基本事件流:**  1. 用户点击某一文本事件提取的“详情”按钮，向系统发出查看某文本内事件提取详细结果的请求  2. 系统从数据库中查找分析结果  3. 系统按照事件分行展示分析结果  - 实施者  - 受访者  - 触发词  - 时间  - 地点  - 事件提取 |
| **后置条件：**用户可以在页面上浏览从某一文本中提取到的事件信息 |
| **被扩展的用例：**分享事件提取结果用例，收藏事件提取结果用例 |

|  |
| --- |
| **用例:** 分享事件提取结果 |
| **ID：**UC11 |
| **参与者：** 用户 |
| **前置条件:** 用户已经进入了文本事件详细信息页面 |
| **基本事件流:**  1. 用户点击“分享”按钮  2. 系统提示用户选择需要分享的媒介  3. 用户选择想要分享的媒介  4. 系统将该文本的事件提取结果分享到对应的媒介 |
| **后置条件：**事件提取信息正确显示在用户分享到的媒介上 |

|  |
| --- |
| **用例:** 收藏事件提取结果 |
| **ID：**UC12 |
| **参与者：** 用户 |
| **前置条件:** 用户已经进入了文本事件详细信息页面 |
| **基本事件流:**  1. 用户点击“收藏”按钮  2. 系统将该文本事件提取的详细信息添加到收藏夹 |
| **后置条件：**用户可在收藏夹内浏览该文本的事件提取结果 |

|  |
| --- |
| **用例:** 事件提取结果导出 |
| **ID：**UC13 |
| **参与者：** 用户 |
| **前置条件:** 该事件提取的状态是 已完成 |
| **基本事件流:**  1. 用户点击“导出结果”按钮  2. 系统提示用户选择结果导出的本地位置  3．导出结果到本地文件 |
| **后置条件：**用户可以在页面上浏览从某一文本中提取到的事件信息 |
| **可选事件流：**  若结果导出到本地的过程中出现错误，系统显示导出错误提示信息 |

|  |
| --- |
| **用例:** 分页获取文本库列表 |
| **ID：**UC14 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  1. 用户点击“我的质析”按钮，向系统发出查看文本库的请求  2. 系统从文本库中查找结果  3．系统按照分页格式展示文本库查找结果，并按以下标签分列展示： "id": 文本库ID  "name": 文本库名称  "desc": 文本库描述  "num": 文本库数据总数  "create\_time": 创建时间 |
| **后置条件：**用户可以在页面上浏览文本库中存储的文本列表信息 |
| **可选事件流：**  若文本库中暂无存储信息，则在展示框中不显示文本信息 |

|  |
| --- |
| **基用例:** 更改文本库信息 |
| **ID：**UC15 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  根据用户点击的不同按钮（“创建文本库”，“删除文本库信息”，“修改文本库信息”），进入不同的子用例。 |
| **后置条件：**用户可以在页面上浏览文本库中修改后的文本列表信息 |
| **可选事件流：** |

|  |
| --- |
| **子用例:** 修改文本库信息 |
| **ID：**UC15 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  1. 用户点击“修改”按钮，向系统发出修改文本库目标文本信息的请求  2. 系统在屏幕上显示待输入的文本框，等待用户输入待修改的id  3．用户输入待修改的文本id  4. 系统修改文本库中的id对应的目标文本  5. 若修改成功，系统前端则直接显示修改后的文本库信息；若修改失败，则维持原始文本库信息 |
| **后置条件：**用户可以在页面上浏览文本库中修改后的文本列表信息 |
| **可选事件流：** |

|  |
| --- |
| **子用例:** 删除文本库信息 |
| **ID：**UC16 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  1. 用户点击“删除”按钮，向系统发出删除文本库目标文本信息的请求  2. 系统在屏幕上显示待输入的文本框，等待用户输入待删除的id  3．用户输入待删除的文本id  4. 系统删除文本库中的id对应的目标文本  5. 若删除成功，系统前端则直接显示修改后的文本库信息；若删除失败，则维持原始文本库信息 |
| **后置条件：**用户可以在页面上浏览文本库中删除目标文本后的文本列表信息 |
| **可选事件流：** |

|  |
| --- |
| **子用例:** 创建文本库 |
| **ID：**UC17 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  1. 用户点击“创建”按钮，向系统发出创建文本库文本信息的请求  2. 系统在屏幕上显示待输入的文本框，包括两个标签：  "name"：新文本库名称  "desc"：新文本库描述  3．用户输入待输入的新文本库名称和描述  4. 系统创建文本库中的目标文本  5. 若创建成功，系统前端则直接显示创建后的文本库信息；若创建失败，则维持原始文本库信息 |
| **后置条件：**用户可以在页面上浏览文本库中创建目标文本后的文本列表信息 |
| **可选事件流：** |

|  |
| --- |
| **用例:** 分页获取文本库数据列表 |
| **ID：**UC18 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  1. 用户点击“查询”按钮，向系统发出查询文本库文本信息的请求  2. 系统在屏幕上显示待输入的文本框，包括三个标签：  "lib"：文章id  "page"：要查询的页数  "size"：要查询的数量  3．用户输入待输入的目标文本库文章的id和页数和数量  4. 系统查找文本库中的目标文本  5. 若查找成功，系统前端则直接显示目标文本库信息；若查找失败，则维持原始文本库信息 |
| **后置条件：**用户可以在页面上浏览文本库中目标文本中的文本列表信息 |
| **可选事件流：** |

|  |
| --- |
| **基用例:** 修改文本库数据 |
| **ID：**UC19 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  根据用户点击的不同按钮（“添加文章”，“删除文章”），进入不同的子用例。 |
| **后置条件：**用户可以在页面上浏览文本库中修改后的文本列表信息 |
| **可选事件流：** |

|  |
| --- |
| **子用例:** 添加文章 |
| **ID：**UC20 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  1. 用户点击“添加文章”按钮，向系统发出添加文章到文本库的请求  2. 系统在屏幕上显示待输入的文本框，包括两个标签：  "lib"：要添加文本的文本库id  "body"：csv文件与字段的对应关系  3．用户输入要添加文本的文本库id、csv文件与字段的对应关系  4. 系统修改文本库中的相应文本  5. 若添加成功，系统前端则直接显示添加后的文本库信息；若添加失败，则维持原始文本库信息 |
| **后置条件：**用户可以在页面上浏览文本库中添加文章后的文本列表信息 |
| **可选事件流：** |

|  |
| --- |
| **子用例:** 删除文章 |
| **ID：**UC21 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  1. 用户点击“删除文章”按钮，向系统发出文本库中删除目标文章的请求  2. 系统在屏幕上显示待输入的文本框，包括两个标签：  "lib"：要删除文本的文本库id  "id"：要删除文章id  3．用户输入要删除文本的文本库id、要删除文章id  4. 系统修改文本库中的相应文本  5. 若删除成功，系统前端则直接显示删除后的文本库信息；若添加失败，则维持原始文本库信息 |
| **后置条件：**用户可以在页面上浏览文本库中删除文章后的文本列表信息 |
| **可选事件流：** |

|  |
| --- |
| **用例:** 文章详情 |
| **ID：**UC22 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  1. 用户点击“文章详情”按钮，向系统发出查看目标文章详情的请求  2. 系统在屏幕上显示待输入的文本框，包括两个标签：  "lib"：要查看文章所属文本库的id  "id"：文章id  3．用户输入要查看文章所属文本库的id、目标文章id  4. 系统查找文本库中的相应文本  5. 若查找成功，系统前端则直接显示删除后的文本库信息；若查找失败，则维持原始文本库信息 |
| **后置条件：**用户可以在页面上浏览文本库中查找到文章后的文章详情信息 |
| **可选事件流：** |

|  |
| --- |
| **用例:** 相关新闻 |
| **ID：**UC23 |
| **参与者：** 用户 |
| **前置条件:** 用户登录成功 |
| **基本事件流:**  1. 用户点击“相关新闻”按钮，向系统发出查看相关新闻的请求  2. 系统在屏幕上显示待输入的文本框，包括两个标签：  "lib"：要查看文章所属文本库的id  "id"：文章id  3．用户输入要查看文章所属文本库的id、目标文章id  4. 系统查找文本库中的相应文本  5. 若查找成功，系统前端则直接显示删除后的文本库信息；若查找失败，则维持原始文本库信息 |
| **后置条件：**用户可以在页面上浏览文本库中查找到文章后的文章详情信息 |
| **可选事件流：** |

1. 非功能需求描述

6.其他需求等

6.1 易用性

界面风格统一，其友好的用户交互设计，让用户更方便地获取自己想要的信息。

尽量减少用户的学习成本，方便操作，操作流程合理，操作完成时有统一规范的提示信息。

支持没有计算机使用经验、计算机使用经验较少及有较多计算机使用经验的用户均能方便地使用本系统

系统具有一定的容错和抗干扰能力，在非硬件故障或非通讯故障时，系统能够保证正常运行，并有足够的提示信息帮助用户有效正确地完成任务。

系统具有一定的灵活性，可以提示用户如何操作

系统的通用性、操作简单易学，实用性强

网站功能精简，不含冗余恼人的其它功能。

6.2 可维护性

可维护性是指在不影响系统其他部分的情况下修改现有系统功能中问题或缺陷的能⼒。开发⼈员创建和设计系统架构时，为了提高系统的可维护性，必须考虑以下几个方⾯的要素：低耦合、高内聚合系统。

本系统将采用严格的软件工程的规范进行开发，并采用良好的设计模式保证系统各模块之间的低耦合 及模块之间的高内聚。

本系统的所有代码将会被详细注释，对于系统所有代码，我们会生成详尽的技术文档。对于系统开发过程可能出现的报错，我们将以文档的方式详细罗列报错码及对应的报错信息。

根据兼职行业的现状，本系统不仅要考虑现有的数据集成和功能集成，还要充分考虑因时新需求导致的对现有功能的新要求。

因而系统的设计方案应易于理解

相关的文档资料应清晰、完整、一致

同时要有良好的功能、数据可扩展性以及对环境适应能力

6.3 可扩展性

系统建成后，应在现行系统上不需要做⼤的改动或不影响整个系统结构，就可以增加功能模块，这就必须在系统设计时留有接口，使其具有可扩展性和维护性，这样就方便在后期的维护过程中根据用户的需求添加相应的功能，同时也不会影响系统其他功能模块的正常运行。

除此之外，还应当保证所有文档说明的详细、完整度，使其他接收者或新加入开发者能够快速地了解系统的情况，并且正确地在原有基础上进行更新和加强。

6.4 稳定性

网站能够正常完成所有的功能需求，且保证服务器长时间运行无障碍。故障修复时间一般控制在3小时内，除非出现恶性攻击、病毒破坏、系统资源不足、硬件故障、操作系统崩溃网络中断等情况，不允许出现内存泄露。要求系统可以承受告警风暴和大数据量的冲击，不出现内存溢出等现象。

6.5 灾难恢复

即使黑客入侵随意篡改数据，或者发生灾难造成数据丢失，但由于区块链系统中，任一节点的损坏或者失去都会不影响整个系统的运作，并且可在链上取下当前准确数据，因此对数据恢复尤为容易