目录

[第二章 NLP项目需求分析 2](#_Toc8942_WPSOffice_Level1)

[2.1. 文档介绍 2](#_Toc32429_WPSOffice_Level2)

[2.1.1. 文档目的 2](#_Toc32429_WPSOffice_Level3)

[2.1.2. 文档范围 2](#_Toc5277_WPSOffice_Level3)

[2.1.3. 读者对象 2](#_Toc1395_WPSOffice_Level3)

[2.1.4. 参考文档 2](#_Toc27760_WPSOffice_Level3)

[2.1.5. 术语与缩写解释 2](#_Toc22188_WPSOffice_Level3)

[2.2. 产品介绍 2](#_Toc5277_WPSOffice_Level2)

[2.2.1. 产品面向的用户群体 2](#_Toc5968_WPSOffice_Level3)

[2.2.2. 产品应当遵循的标准或规范 3](#_Toc137_WPSOffice_Level3)

[2.2.3. 产品范围 3](#_Toc2012_WPSOffice_Level3)

[2.2.4. 产品中的角色 3](#_Toc7070_WPSOffice_Level3)

[2.3. 产品的功能性需求 3](#_Toc1395_WPSOffice_Level2)

[2.3.1. 功能性需求分类 3](#_Toc19227_WPSOffice_Level3)

[2.3.2. m Feature M 4](#_Toc1159_WPSOffice_Level3)

[2.4. 产品的非功能性需求 6](#_Toc27760_WPSOffice_Level2)

[2.4.1. 用户界面需求 6](#_Toc10250_WPSOffice_Level3)

[2.5. 软硬件环境需求 6](#_Toc22188_WPSOffice_Level2)

[2.6. 产品质量需求 6](#_Toc5968_WPSOffice_Level2)

# 第二章 NLP项目需求分析

## 文档介绍

### 文档目的

**本文档适用于开发人员对产品的开发内容的了解和开发要求的熟知。**

### 文档范围

**本产品适用于开发人员，便于对产品的理解和开发。**

### 读者对象

**产品开发人员**

### 参考文档

《**CMMI书写规范文档**》

**《nlp项目-开发计划》**

### 术语与缩写解释

|  |  |
| --- | --- |
| **缩写、术语** | **解 释** |
| **Nlp (Neuro-Linguistic Programming)** | **神经语言程序学** |

## 产品介绍

**本产品用于对自然语言的处理和分析。便于学者理解。**

### 产品面向的用户群体

**研究自然语言的学者；**

**学习自然语言的学生**

### 产品应当遵循的标准或规范

**文档书写应遵循CMMI书写规范。开发源码遵循阿里巴巴书写规范。源文件命名方式遵循驼峰命名。**

### 产品范围

**本产品适用于学习语言的学生，有效的分析出语言的词性，依存关系等便于理解。**

### 产品中的角色

|  |  |
| --- | --- |
| 角色名称 | 职责描述 |
| 项目经理 | 对整个产品进行计划、督促 |
| 需求分析师 | 撰写《用户需求分析书》和《需求管理规格说明书》 |
| 系统架构师 | 对系统进行架构设计 |
| 配置管理员 | 对开发环境进行配置管理，撰写《配置管理计划书》 |
| 测试员 | 对程序进行测试，并攒写测试报告 |
| 质量的分析与度量 | 撰写《质量的分析与度量计划》 |

## 产品的功能性需求

### 功能性需求分类

|  |  |
| --- | --- |
| **功能类别** | **子功能** |
| **数据维护** | **数据保存** |
| **数据取用** |
| **分词处理** | **分词标注** |
| **实体抽取** |
| **实体关系** |
| **依存关系** |
| **关键词提取** |
| **摘要提取** |
| **文本分类** |

### m Feature M

|  |  |
| --- | --- |
| **名称、标识符** | **分词标注** |
| **功能描述** | **给定用户输入的一段自然语言，对其进行中文分词与词性标注** |
| **优先级** | **高** |
| **输入** | **一段自然语言。例如：商品和服务** |
| **操作序列** | **输入一段自然语言，调用方法对其进行分词标注，最后返回输出** |
| **输出** | **[商品/n, 和/c, 服务/vn]** |
| **补充说明** | **对四家分词标注进行对比** |

|  |  |
| --- | --- |
| **名称、标识符** | **实体关系** |
| **功能描述** | **给定用户输入的一段自然语言，在正确识别实体的基础上，抽取它们之间的关系** |
| **优先级** | **高** |
| **输入** | **一段自然语言。例如：我和我的妈妈去了公园** |
| **操作序列** | **对实体（我、妈妈、公园）进行关系的提取** |
| **输出** | **实体与实体的关系；例如：我->妈妈：和；我、妈妈->公园：去了** |
| **补充说明** | **实体关系建立在实体抽取上，先期实体识别的结果会影响到关系抽取的结果** |

|  |  |
| --- | --- |
| **名称、标识符** | **文本分类** |
| **功能描述** | **可以用于新闻分类、简历分类、邮件分类、办公文档分类、区域分类等诸多方面。此外还可以实现文本过滤，能够从大量文本中快速识别和过滤出符合特殊要求的信息，可应用于品牌报道监测、垃圾信息屏蔽、敏感信息审查等领域。** |
| **优先级** | **高** |
| **输入** | **一段自然语言。例如：原标题：”光纤之父”高锟离世 享年84岁**  **据香港媒体报道，香港中文大学前校长、有“光纤之父”之称的高锟，今天（9月23日）下午在医院离世，享年84岁。**  **高锟光纤、宽带传送等科技成就极高，在国际学术领域获奖无数，并在2009年10月6日获得诺贝尔物理学奖，以表扬其“在光传输于纤维的光学通信领域突破性成就”。然而，高锟早于2004年时，证实患上阿兹海默症。**  **高锟在职业生涯中，持续钻研信号传送的研究，研发光纤通讯技术，并获得2009年诺贝尔物理学奖、爱迪生电信奖、马可尼国际奖、贝尔奖、巴伦坦奖章、利布曼奖等多个殊荣。1987年至1996年，高锟出任香港中文大学第三任校长，创立讯息工程学系、建筑学系，至2009年退休。**  **点击进入专题： 诺贝尔物理学奖得主高锟逝世 享年84岁**  **责任编辑：张义凌** |
| **操作序列** | **对文本进行分类** |
| **输出** | **输出分类结果。例如：教育** |
| **补充说明** |  |

|  |  |
| --- | --- |
| **名称，标识符** | **关键词提取** |
| **功能描述** | **在给定的一段语句中，跟这篇文章意义最相关（或者是出现频率比较高）的一些词语抽取出来** |
| **优先级** | **中** |
| **输入** | **一段文本：人工智能，它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能是计算机科学的一个分支，它企图了解智能的实质。** |
| **操作序列** | **先读取待提取的文本语料，调用函数返回权重比较大的关键词个数** |
| **输出** | **整段文本中权重比较大的语句：人工智能，智能，科学。。** |
| **补充说明** | **重要的词一定在文章中出现多次，但同时在各个文章中出现多次的词一定没有在只在同一篇文章中出现多次的词的重要性大。** |

|  |  |
| --- | --- |
| **名称，标识符** | **依存关系** |
| **功能描述** | **给定用户输入一段自然语言，在正确识别各分词标注的基础上，描述出各个词语之间的依存关系** |
| **优先级** | **中** |
| **输入** | **一段自然语言。例如：这是我的家** |
| **操作序列** | **描述词与词的支配与被支配的关系** |
| **输出** | **词与词的支配与被支配的关系 eg:家->这 （det）限定词修饰** |
| **补充说明** | **句法的分析必须在分词标注的基础上** |

# 

|  |  |
| --- | --- |
| **名称，标识符** | **摘要提取** |
| **功能描述** | **能根据需求灵活控制摘要长度，自动抽取关键信息，形成摘要结果。可用于内容理解、内容分发、智能写作等多种应用** |
| **优先级** | **中** |
| **输入** | **央广网北京2月28日消息 据中国之声《新闻和报纸摘要》报道，国务院总理李克强2月27日向第五届中德创新大会致贺信。**  **李克强在贺信中表示，当前新一轮科技革命和产业变革席卷全球，科技创新正深刻改变着人类的生产生活方式。中德科技创新合作开创了大国科技合作的先例，为两国务实合作装上了大功率“引擎”。**  **李克强指出，中国经济发展正处在新旧动能转换和结构升级的关键时期。我们将贯彻落实新发展理念，深入实施创新驱动发展战略，促进大众创业、万众创新上水平，加快建设创新型国家。希望中德双方汇集众智、增进共识，深化科技创新交流合作，推动两国经济社会健康发展，为全球经济注入新动力。**  **中德政府间科技合作协定签订40周年暨第五届中德创新大会27日在京举行。两国科技、企业、政府等各界300余名代表出席。** |
| **操作序列** | **央广网北京2月28日消息 据中国之声《新闻和报纸摘要》报道，国务院总理李克强2月27日向第五届中德创新大会致贺信。李克强在贺信中表示，当前新一轮科技革命和产业变革席卷全球，科技创新正深刻改变着人类的生产生活方式。中德科技创新合作开创了大国科技合作的先例，为两国务实合作装上了大功率“引擎”。中德政府间科技合作协定签订40周年暨第五届中德创新大会27日在京举行。** |
| **输出** | **整段文本中权重比较大的语句：人工智能，智能，科学。。** |
| **补充说明** | **能根据需求灵活控制摘要长度，自动抽取关键信息，形成摘要结果** |

## 产品的非功能性需求

### 用户界面需求

|  |  |
| --- | --- |
| **需求名称** | **详细要求** |
| **登陆界面** | **界面美观、友好。背景为黑色，登陆栏为蓝色，输入密码错误时可以弹出提示框。** |

## 软硬件环境需求

|  |  |
| --- | --- |
| **需求名称** | **详细要求** |
| **数据库服务器** | **数据库：Mysql 5.6**  **内存：8G**  **操作系统：windows10**  **磁盘存储容量：500G** |
| **WEB应用服务器** | **JDK版本：jdk1.8**  **WEB服务器：Apacher Tomcat 8.5.35**  **操作系统：windows10** |
| **数据备份服务器** | **数据库：Mysql 5.6**  **内存：8G** |
| **操作系统** | **Windows10** |

## 产品质量需求

|  |  |
| --- | --- |
| **主要质量属性** | **详细要求** |
| **强壮性** | **任何操作，包括用户操作、系统交互等过程中的错误和异常，只能在相关业务或者模块中产生影响，并当操作正确时能正常响应** |
| **可靠性** | **数据库插入使用数据锁，加强事务管理** |
| **安全性** | **有关网络环境、操作系统、数据备份、权限管理等方面的安全性要达到高有效性系统的要求，具体要求待定。**  **应用程序的安全性，主要参考运维管理”中的“安全管理”。同时，要满足下面的条件:**  **1)密码类重要信息在系统内不能存在明文;2) 单点登录过程中的各种口令不能被窃取;**  **3) 单点登录环境下，在凭证的传递及凭证的反向验证过程中不能被第三人恶意截获。**  **4) 能防止SQL注入、脚本攻击等常见的WEB攻击手段。** |
| **性能、效率** | **系统响应时间：2s以内** |
| **易用性** | **操作流程简单** |
| **可扩展性** | **可扩展其他功能模块** |
| **可移植性** | **使用Java语言开发，尽量遵从各种Java规范，以易于在各种服务器上进行移植;数据库表结构、SQL 语句等，在不必要时，尽量不使用数据库特性;协议方面尽量使用常用的格式。** |
| **兼容性** | **界面对于市场上主流浏览器的兼容** |
| **强壮性** | **任何操作，包括用户操作、系统交互等过程中的错误和异常，只能在相关业务或者模块中产生影响，并当操作正确时能正常响应** |
| **可靠性** | **数据库插入使用数据锁，加强事务管理** |
| **安全性** | **有关网络环境、操作系统、数据备份、权限管理等方面的安全性要达到高有效性系统的要求，具体要求待定。**  **应用程序的安全性，主要参考运维管理”中的“安全管理”。同时，要满足下面的条件:**  **1)密码类重要信息在系统内不能存在明文;2) 单点登录过程中的各种口令不能被窃取;**  **3) 单点登录环境下，在凭证的传递及凭证的反向验证过程中不能被第三人恶意截获。**  **4) 能防止SQL注入、脚本攻击等常见的WEB攻击手段。** |