**“汉语词汇推荐系统（实验系统）”帮助文档**

**1 推荐系统简介**

本推荐系统通过用户的个性化信息，以搜索引擎和翻译系统为数据来源，向用户主动推荐个性化词汇。

1. **使用说明**

**2.1安装程序**

1. 在本机下载并安装python

下载地址： <https://www.python.org/downloads/>，下载操作系统相应的版本后安装，windows安装应使用3.7及以上版本

1. 打开<https://langtech.fun/account/register/>页面，点击右下角“下载”，下载install\_and\_getword.rar，并解压到install\_and\_getword文件夹
2. 在本机安装支持软件包，以及特定版本的chrome浏览器和驱动

安装方法：双击install\_and\_getword文件夹中的setup.bat

Tips：

安装时，如果弹出对话框询问是否以管理员权限运行脚本，请按“是”。

使用时，chrome浏览器如果提示更新，请拒绝更新。

1. 重启电脑
	1. **采集数据**
2. 使用chrome浏览器，输入**用户名和密码**登录百度网页（这一步是为了让浏览器记录cookie，当需要验证时可以使用cookie登录）
3. 关闭chrome浏览器（全部实例）
4. 双击install\_and\_getword文件夹中的getMyDataWord\_run.bat，按要求输入百度账户名和百度账户密码，会弹出chrome浏览器网页，并自动执行浏览跳转，如果需要验证，按要求进行信息验证
5. 等待chrome自动关闭，此时myQueryWordList.txt保存在install\_and\_getword\getMyDataWord\collect\_static\txt中

Tips：如果程序提示少量handshake failure，ssl\_error,请忽略

**2.3 上传数据**

1. 打开浏览器，输入www.langtech.fun:8080/ckb/index，词汇推荐系统会展示随机推荐词汇。



图1 随机推荐结果

1. 使用firefox浏览器打开<https://langtech.fun/account/register/>,进行用户注册，**用户名和密码**可以不使用百度账户名和密码
2. 打开<https://langtech.fun/account/login/>进行登录
3. 点击上传文件按钮，上传install\_and\_getword\getMyDataWord\collect\_static\txt中的myQueryWordList.txt
4. 也可以在登录状态下，点击右侧“上传文件”，上传预先准备好的个人（检索）语料，在词义关系列表框中选择表示聚合关系的语义词典，点击浏览器地址栏，按回车键更新页面，系统将按照个人语料返回推荐结果。



图2 登录状态的推荐结果示例

词云主要呈现三种信息。一是词语的频次信息。词语的大小直观地反映词语出现频次的多少；当鼠标移入节点时，显示词语在语料库中的具体出现频次。二是推荐词内部扩展关系。对无扩展关系的词进行无重复随机着色，对存在扩展关系的词，依据扩展关系着以相同的颜色。比如，由基础词“产生”扩展出“生产”，后者根据前者着色。三是拓展词信息。区分个人语料中存在基础词和新荐的拓展词，对推荐的扩展词着以不同的颜色，而基础词退化为灰色，以凸显新荐词。在图3中点击“显示”，9个新荐词以彩色形式展示，其余词退化为背景色。



图3 拓展词展示

在推荐过程中，用户据自己的学习情况设置停用词，右键单击词语节点将该词语移入停用词表，系统在之后的推荐过程中将过滤该词；对于已经存在的停用词，可以在右侧边栏输入并移出停用词表。

为了便于用户学习，在右侧边栏中点击“显示”按钮，将显示基础词与拓展词的语义关系对，点击“隐藏”按钮将隐藏这些语义关系对，点击“乱序”，将随机打乱这些具有语义关系的词。用户也可以在左侧“今日推荐词”栏中点击某个词语，系统使用相同颜色突出显示具有语义关系的词，其他背景词退化为灰色。



图4 关系词展示