

# 现代汉语字词构词 能力计算工具

使用说明

2025 年 6 月

中国 北京

目录

一、主要功能 ..... 1

二、使用方法 ..... 1

三、操作步骤 ..... 1

    1. 安装程序 ..... 1

    2. 软件操作 ..... 8

    3. 词表查看 ..... 10

## 一、主要功能

WordBuiding 是一款计算汉语字词组合能力的软件。汉字组合能力是指该汉字的构词数及所构成词语在语料库中的总频率。词语组合能力是指该词语所构成短语、成语等更高一级词汇单位的个数及其在语料库中的总频率。在本软件中，语料库为国家语委语料库，词频表为国家语委语料库 2000 万词语对统计，其中词频表只显示词频数量大于 50 的词语。

## 二、使用方法

本地安装后，使用用户名登陆，用户导入需要统计的“字/词表”，以及“词语频率表”即可得到字/词组合能力的统计结果。

## 三、操作步骤

### 1. 安装程序

①解压安装程序压缩包,找到并点击“Wordbuilding\_setup”→②根据页面提示点击‘下一步’进行安装→③勾选‘我同意.....’选项进入下一步

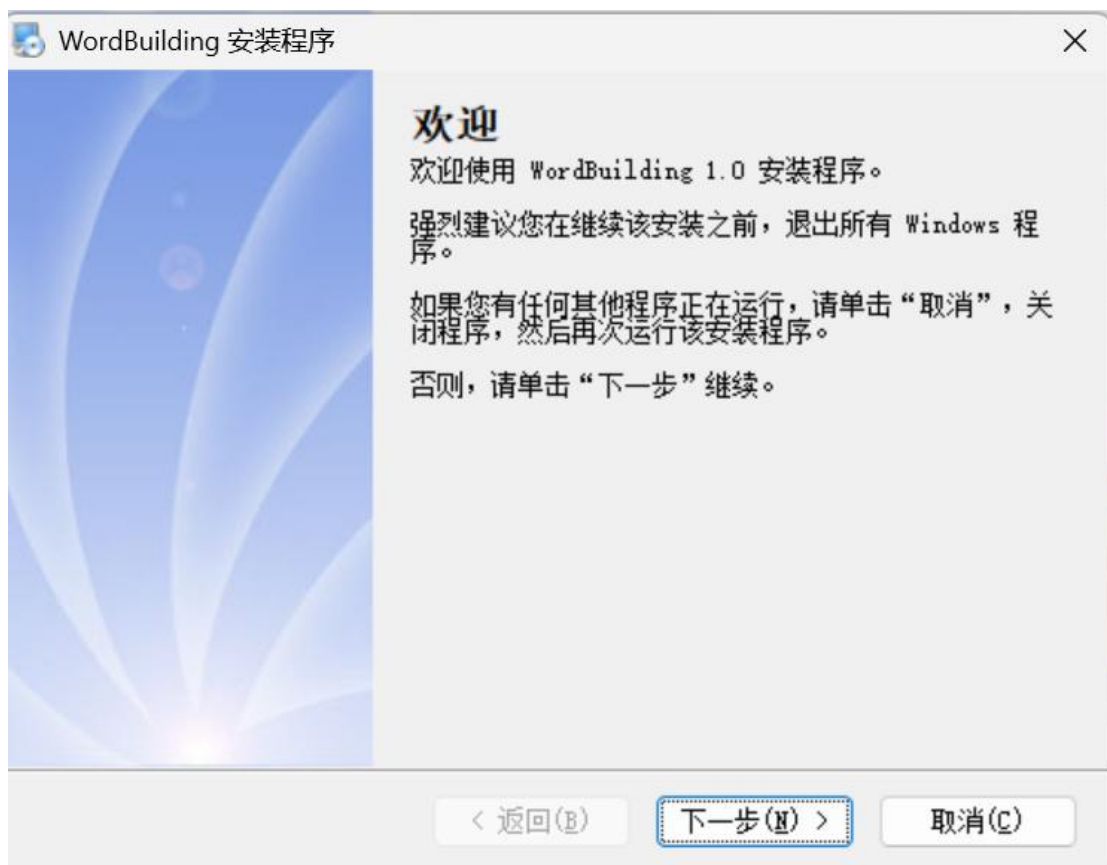


图 1 软件安装操作步骤界面一



图 2 软件安装操作步骤界面二

④填写用户信息：名称项通常为自动填入项，手动填写公司项。



图 3 软件安装操作步骤界面三

⑤选择安装地址，可点击‘更改’更换默认地址。



图 4 软件安装操作步骤界面四

⑥用户可修改文件夹名称，并修改软件使用权限。



图 5 软件安装操作步骤界面五

⑦点击‘下一步’进入下一步安装。

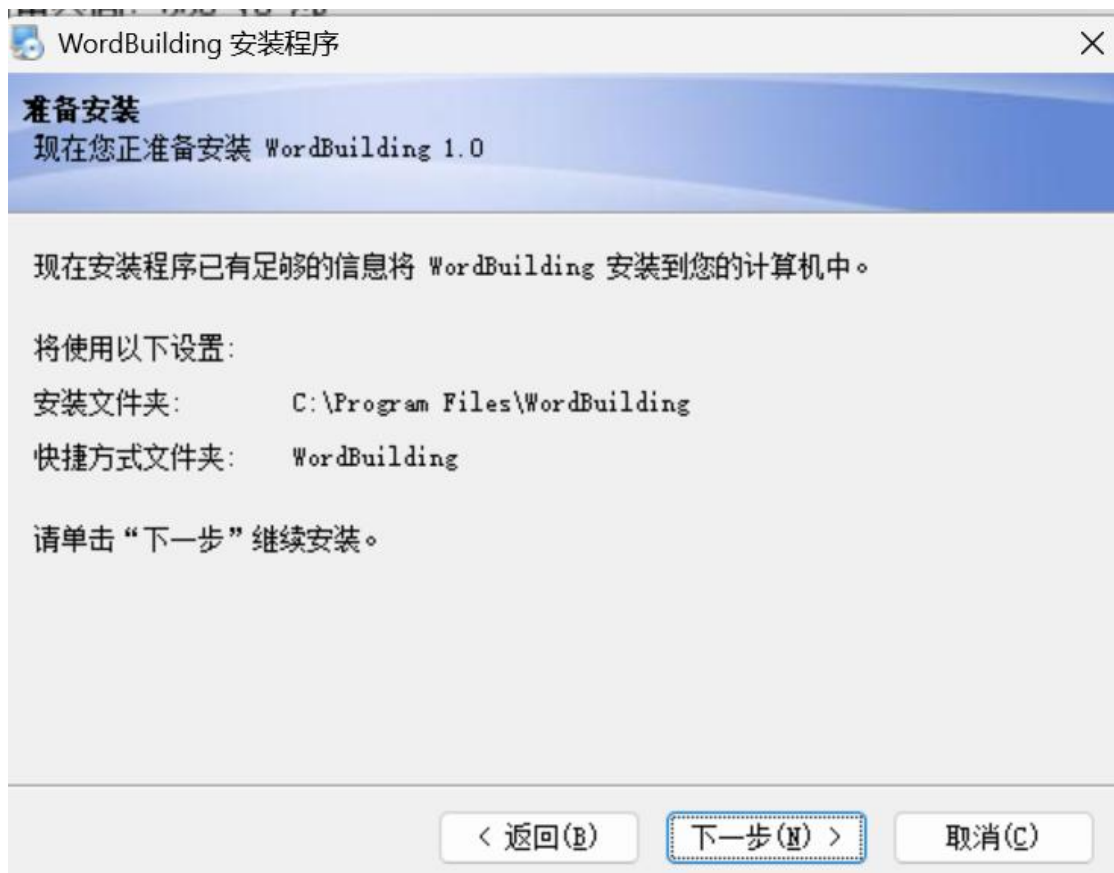


图 6 软件安装操作步骤界面六

⑧点击‘完成’安装完毕



图 7 软件安装操作步骤界面七

⑨点击任务栏左边“开始”图标，可以从“最近添加”处或根据字母顺序找到 WordLen 程序，点击 WordLen.exe 程序即可运行。



图 8 软件安装操作步骤界面八

## 2. 软件操作

### ①选择字\词表

注意：对字词表文件有以下要求：①文件类型为.txt。②内容编排上按照每个字/词为一行进行纵向成列排放。③txt 文件编码选择 ANSI 编码类型。④字词表的文件路径需存放于与程序同一根目录之下。

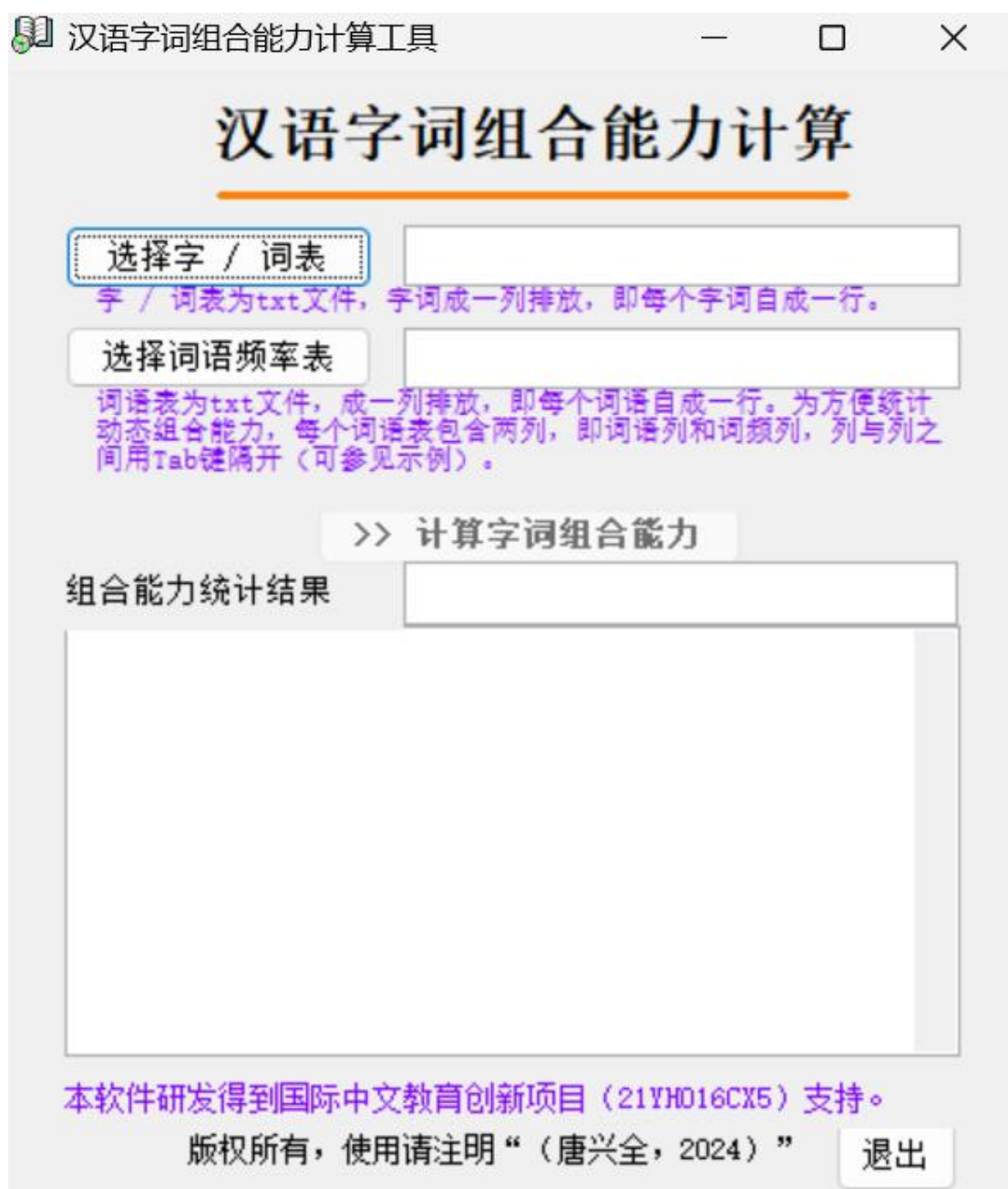


图 9 软件使用操作步骤界面一

②选择词语频率表。

压缩包提供词语频率表的示例，该词频表来源于对国家语委语料库 2000 万词语料库的统计，词频表中包含的是出现次数在 50 次以上的词语；50 次以下的词语未在该词频表中列出。

注意：对词语频率表文件有以下要求：①文件类型为.txt。②内容编排上为词语列和相应的词频列，列与列之间使用‘Tab’键隔开。③txt 文件编码选择 ANSI 编码类型。④词语频率表的文件路径需存放于与程序同一根目录之下。



图 10 软件使用操作步骤界面二

③成功选择文件后，点击计算字词组合能力，选择统计结果存放位置。

### 3. 词表查看

在保存位置下查找文件名称为“ WordBuilding”的.xls 类型的文件。

注意：如需同一位置下存放多个计算文件，需将已保存的“ WordBuilding.xls”文件重

命名存放。

计算结果将按照“字词”、“独现频率”、“组合数”、“组合频率”、“总频率”、“组合实例”六部分分列展示。

| #  | A  | B      | C   | D      | E      | F                                                                                       |
|----|----|--------|-----|--------|--------|-----------------------------------------------------------------------------------------|
| 1  | 字词 | 独现频率   | 组合数 | 组合频率   | 总频率    | 组合实例                                                                                    |
| 2  | 啊  | 0.0414 | 1   | 0.0008 | 0.0422 | 啊呀                                                                                      |
| 3  | 场  | 0.0024 | 2   | 0.0013 | 0.0037 | 场小,白桦里                                                                                  |
| 4  | 爱  | 0.0198 | 22  | 0.0463 | 0.0661 | 爱憎,可爱,爱国,热爱,爱人,喜爱,爱好,恋爱,爱回斯坦,爱护,爱国主义,亲爱的,爱上,敬爱,心爱,爱吃,爱好着,友爱,慈爱,爱尔兰,相爱,疼爱                |
| 5  | 安  | 0.0043 | 26  | 0.0615 | 0.0658 | 安排,安全,不安,安慰,安静,安顿,延安,安定,西安,安培,安群,治安,安置,长安,公安机关,安心,公安,公安局,安宁,平安,大兴安岭,安定团结,印更安人,安详,天安门,保安 |
| 6  | 按  | 0.0342 | 6   | 0.0274 | 0.0616 | 按照,按劳分配,按时,按钮,按着,按往                                                                     |
| 7  | 暗  | 0.0042 | 9   | 0.0136 | 0.0178 | 黑暗,暗暗,暗示,暗中,阴暗,暗自,昏暗,暗淡,明暗                                                              |
| 8  | 八  | 0.0261 | 20  | 0.0298 | 0.0569 | 十八,八十,八路军,八月,二十八,八百,一九七八,八旗,四面八方,一九八二,一九八三,八路,八一,一九八一,八字,第八,一九八〇,四十八,一九五八,一九八四          |
| 9  | 巴  | 0.0011 | 14  | 0.0208 | 0.0219 | 尾巴,巴黎,嘴巴,巴西,淋巴,费尔巴哈,巴掌,下巴,古巴,巴勒斯坦,淋巴结,巴基斯坦,巴金,巴尔扎克                                      |
| 10 | 把  | 0.3113 | 5   | 0.0081 | 0.3194 | 把握,有把握,火把,投机倒把,把戏                                                                       |
| 11 | 爸  | 0.0014 | 1   | 0.0144 | 0.0144 | 爸爸                                                                                      |

图 11 部分计算结果展示界面