



使用Bing词典建立自己的Anki 单词库（带语音图片）



lbaby · 3 天前

缘起

前两天时间线上@赛门喵 分享了麦克米伦7000高频词记忆卡，这是非常有用的内容，Anki也是强大的工具，对喵同学非常感谢！

知

写文章

比如下边的内容，就可以处理成卡片

新概念英语 第一册 共793词

序号 英语 词性 汉语

asffada

1 excuse n. 借口，理由；低劣的样品； v. 原谅

2 me pron. 我（宾格）

3 yes a. 是的； ad. 是，是的； v. 是

4 is v. 是

5 this a. pron. 这，这个

6 your pron. 你的，你们的

7 handbag n. 手提包

8 pardon n. 原谅，赦免； v. 宽恕，原谅

9 it pron. 它

10 thank you 谢谢你。

11 very much 非常地

12 pencil n. 铅笔； v. 用铅笔写

准备工作

安装Chrome

使用搜索引擎可以找到，因为谷歌被墙，推荐翻不了的同学去新浪下。

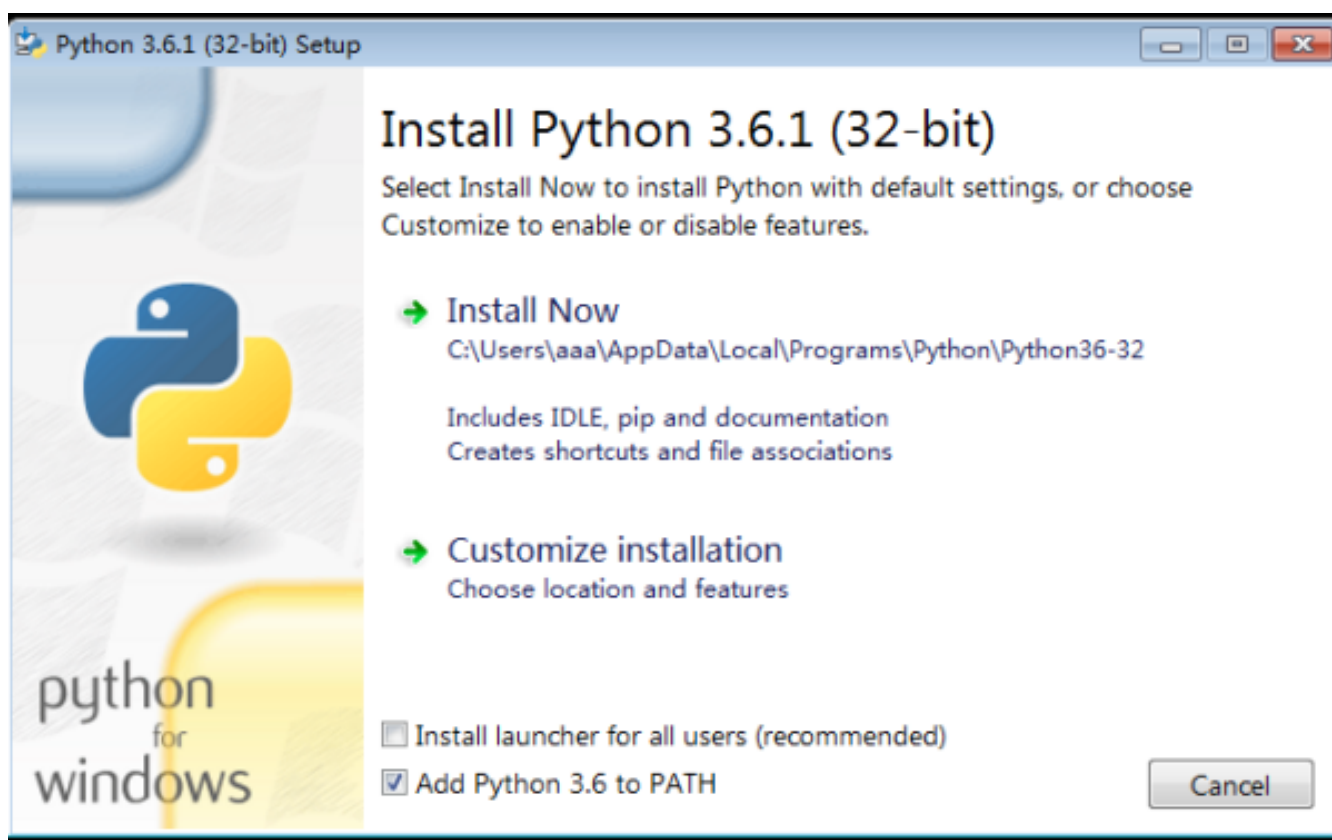
安装python及依赖包

知

写文章

- Python

去官网 [Welcome to Python.org](https://www.python.org/) 下载python 3.6, windows 如下勾选:



然后把Python加入可执行路径:

在 计算机--右键-->高级系统设置->高级（环境变量）->XXX的用户变量，选中PATH，在最后添加

```
;C:\users\aaa\AppData\Local\programs\Python\Python36-32
```

注意，前边的分号不可以少，按你实际安装的Python路径填写，不要硬抄上边的，完成点击确定。

然后，按开始按键，输入cmd回车，然后在命令提示符里输入

```
pip install splinter urllib3
```

等待完成就可以了。如果cmd报错命令没找到，检查你的环境变量是不是搞错了

- Chromedriver

随后搜索chromedriver 包，国内最方便的是Taobao的Mirror

[npm taobao org/mirrors/](https://npm.taobao.org/mirrors/)

知

写文章

解压后可以放在python路径下，如果不放在python路径下，则需要像python那样，把chromedriver 所在路径加入到PATH中去

完成以后，可以打开一个cmd窗口，输入chromedriver 应该有类似以下输出：

```
Starting ChromeDriver 2.29.461591 (62ebf098771772160f391d75e589dc567915b233) on i
Only local connections are allowed.
```

Mac & Linux

- Python

Mac下没有实验，从官网下载默认安装应该问题不大

Linux 下打开终端使用（??? 用Linux 的同学应该知道怎么处理）

```
sudo apt-get install python3
```

- 依赖库：

Mac 下按下Ctrl + Space 调出spotlight(或点击Finder右上角调出)，在输入框里输Terminal，运行终端，在终端里输入：

```
sudo pip3 install splinter urllib3
```

Linux 命令同上

- Chromedriver

从taobao的 Mirror下载chromeDriver 包

npm.taobao.org/mirrors/...

解压以后放在你能找到的目录下或家目录下，然后进入终端，输入：

```
chmod +w ~/.bash_profile
vi ~/.bash_profile
```

进入vi以后按i键进入编辑模式，把光标移到最后（使用上下左右键移动）

添加两行：

```
export PATH=$PATH:/usr/local/bin
```

知

写文章

按esc , 然后输入:wq 回车以后退出vi

Linux 下操作类似, 只是修改的文件要换成~/.bashrc

修改文件以后, 重新进入terminal, 输入chromedriver 未报错就可以认为成功了。

语音文件

142000个英语单词语音库, 下载后解压到任意目录

```
ed2k://|file|5B142000%E4%B8%AA%E5%8D%95%E8%AF%8D%E8%AF%AD%E9%9F%B3%E5%BA%93%5D.%v
ed2k://|file|5B142000%E4%B8%AA%E5%8D%95%E8%AF%8D%E8%AF%AD%E9%9F%B3%E5%BA%93%5D.%v
```

从Bing抓取单词和释义

得到抓取脚本

从下边的地址拿到bing2anki.py

github.com/lbaby/codeGa...

能够处理的文件格式

每个单词（或词组一行），代码尝试去找字母的组合，直到找到一个单词(词组)的边界，找到边界以后就不管本行还有没有其他单词了，具体例子见下：

```
word #ok word
1 word #ok word
单词 word #ok word
单词 word n. thank you#ok word, n. and following rejected
单词 word CET4 #ok word, CET4 rejected
单词 thank you , word#ok, phrase thank you accepted ", " and following rejected
单词 word CET #NOT ok, phrase "word CET" accepted
CET 单词 word #NOT ok, CET accepted, all following rejected
```

就是说，你想要处理的单词或词组，需要把要处理的单词和词组在每一行的最前边，后跟一个空格然后跟一个不是单词的内容

windows 抓取

知

写文章

```
type words.txt | python bing2anki.py c:\path\to\voice
```

words.txt 是你需要生成卡片的单词列表，c:\path\to\voice是你解压的voice文件夹目录

Mac&Linux 抓取

打开终端，输入

```
cat words.txt | python3 bing2anki.py /path/to/voice
```

words.txt 是你需要生成卡片的单词列表，/path/to/voice是你解压的voice文件夹目录

如果你的words.txt是gbk的，则有可能出现以下错误：

```
UnicodeDecodeError: 'utf-8' codec can't decode byte 0xd0 in position 0: invalid c
```

把命令改成

```
cat words.txt | iconv -c -f gbk -t utf-8 | python3 bing2anki.py /path/to/voice
```

抓取需要一定的时间，由于webdriver开启天量的短链接，如果不限制抓取速度，会有可能耗尽操作系统端口（操作系统：我能怎么办，我也很绝望啊），因此限制抓取速度为一个单词5秒左右，完成以后，脚本会自动退出。

生成卡片

安装模板

安装anki2以后，从下边的地址拿到模板（感谢@leaflyer）：

[github.com/lbaby/codeGa...](https://github.com/lbaby/codeGarden)

[单词卡片模板__1 Recite.apkg](#)

双击导入到Anki里

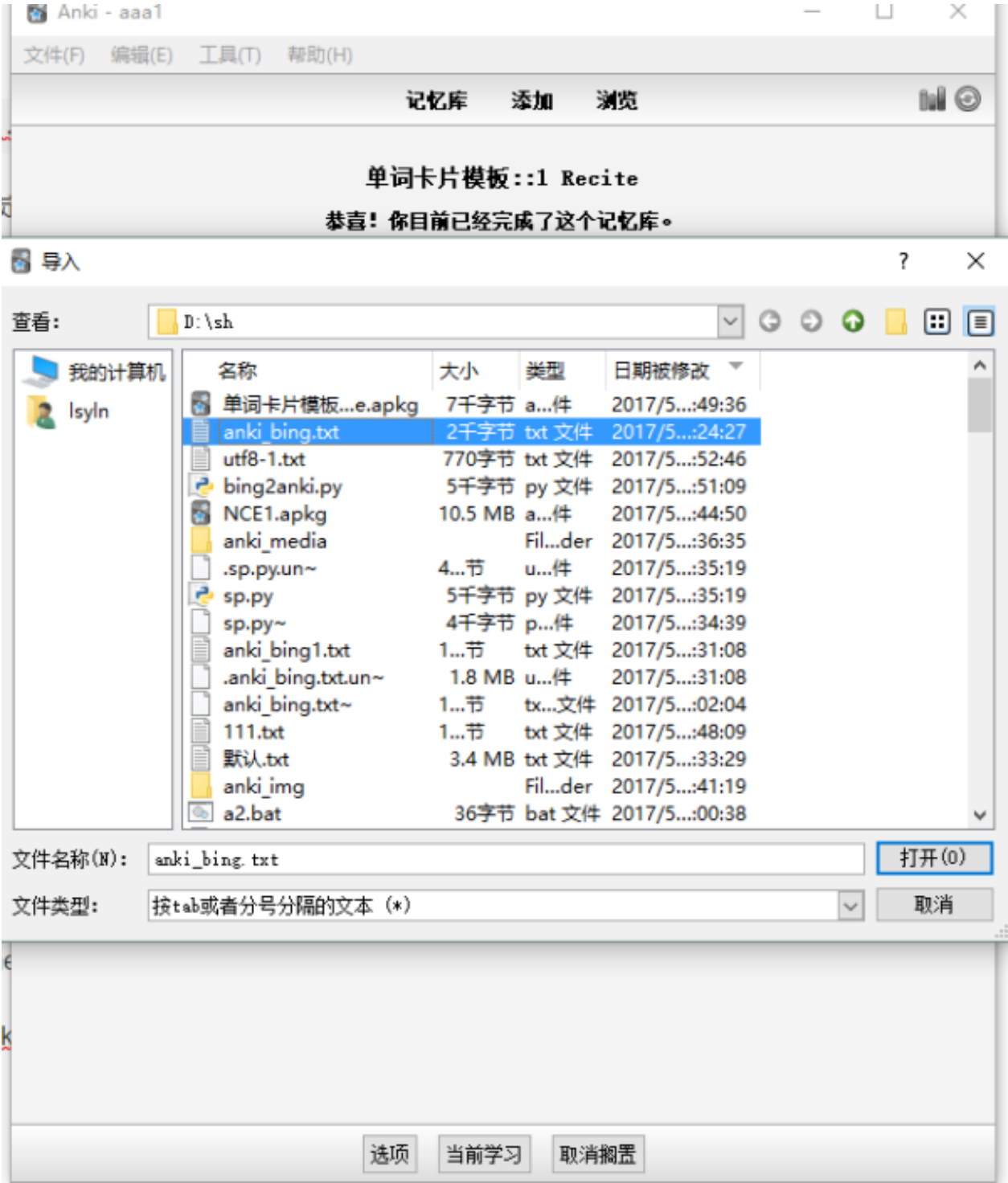
导入单词卡片

从Bing抓取结束以后，会在当前目录下生成一个anki_bing.txt和下载的媒体（图片和发音）文件夹anki_media

我们找到anki的文件->导入，选中anki_bing.txt，点击打开

知

 写文章 ooo

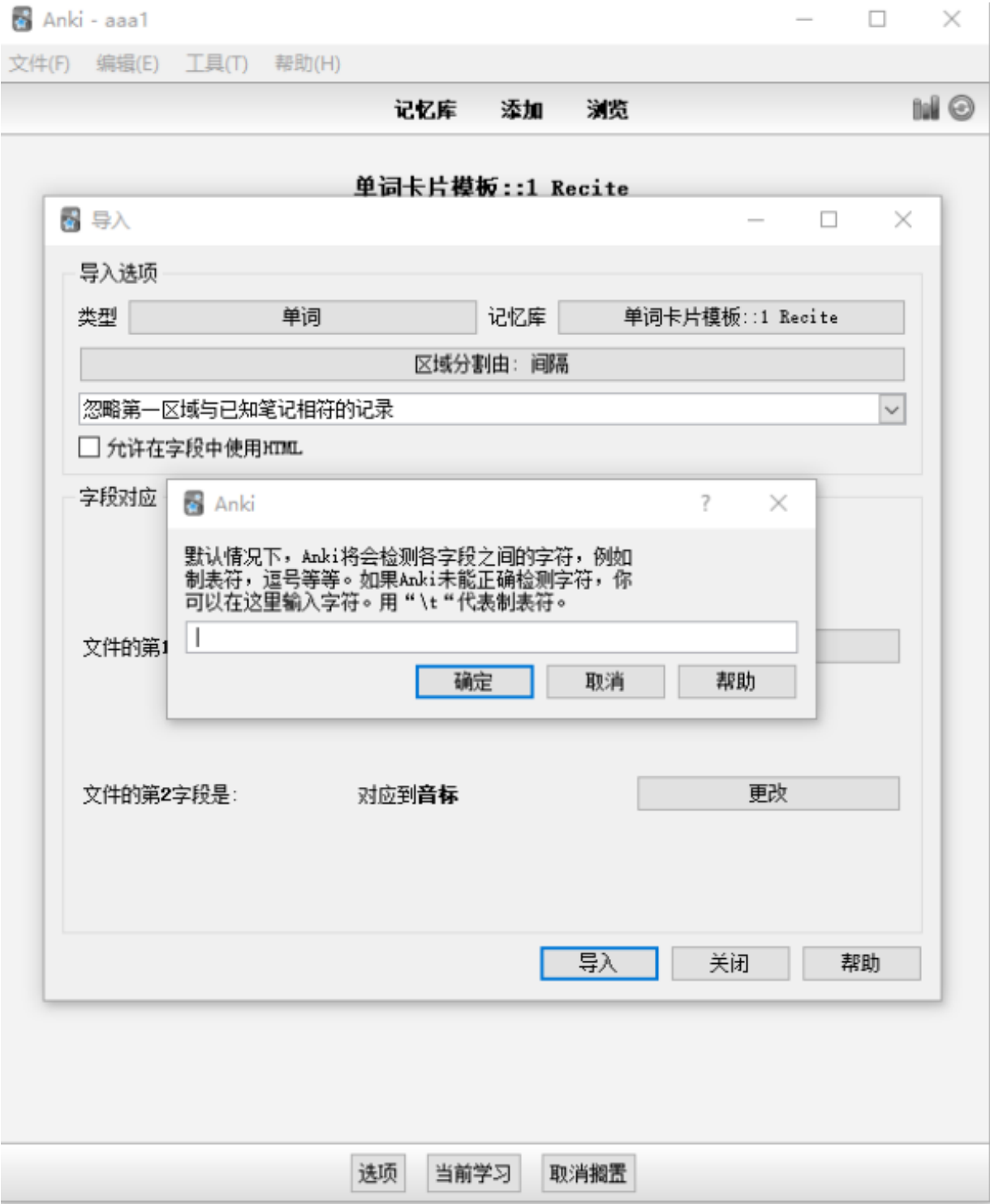


然后在导入选项里选择：

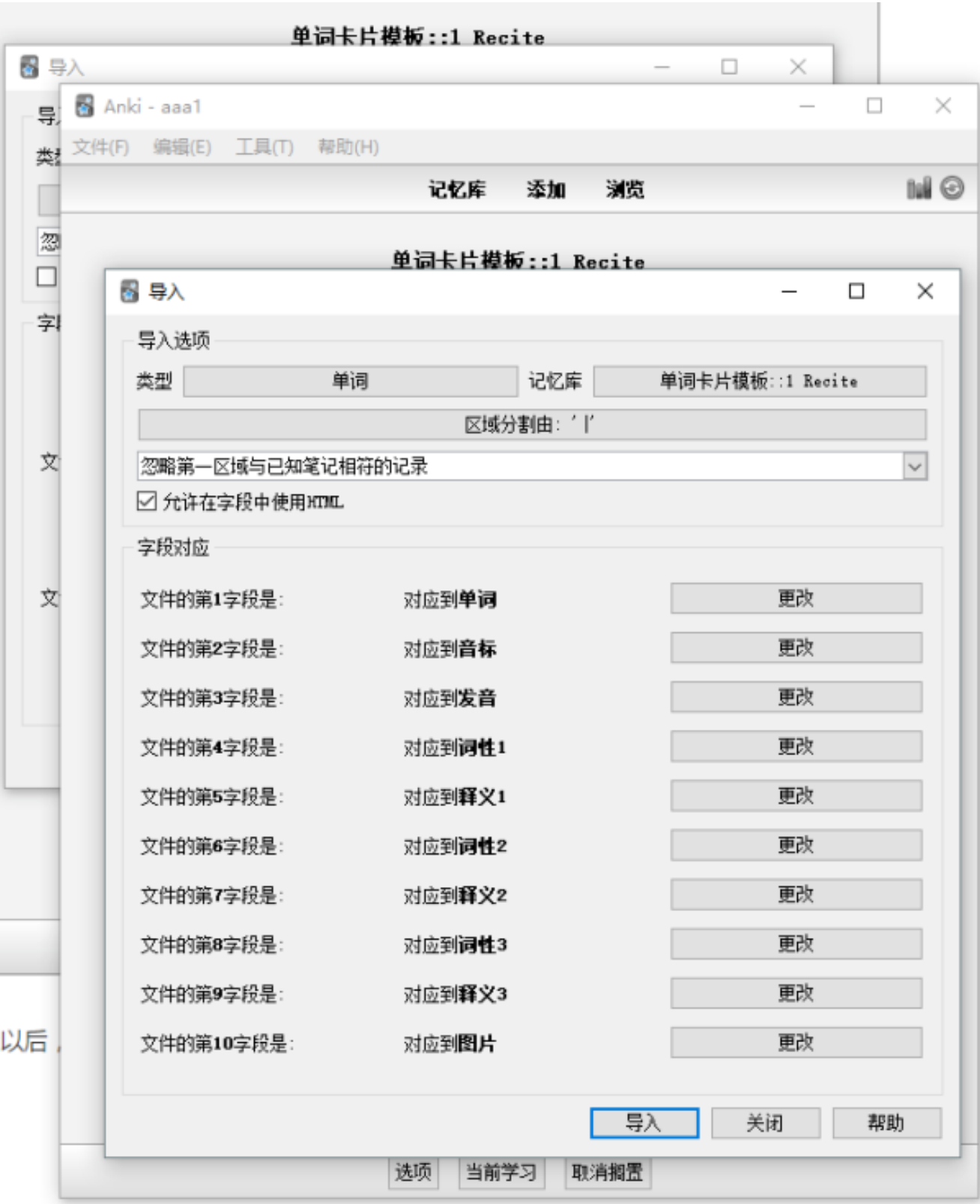
类型：单词

记忆库：按你的要求选

点击 区域分隔由间隔：， 在弹出的框里输入 |



点确定以后，选中“允许在字段中使用HTML”



点击导入等完成就行了。

添加媒体文件

刚才脚本生成的anki_media目录有一堆发音文件和图片，把它们按下边的路径全部复制到anki的媒体目录

%AppData%\Anki2

- **Mac(Command+shift+G然后输入)**

~/Library/Application Support/Anki2

- **Linux**

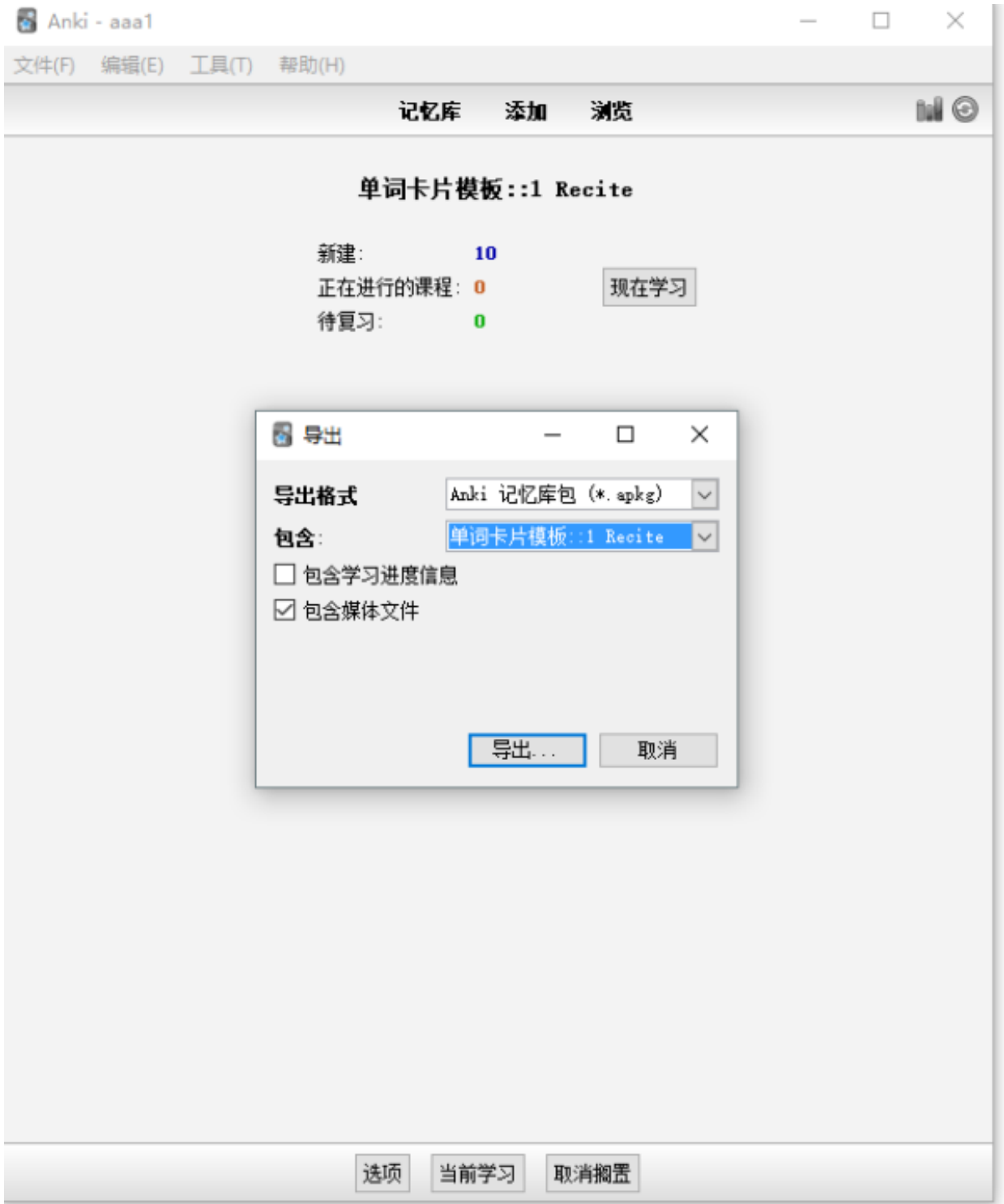
~/local/share/Anki2

然后找到你的Anki用户名的目录下有一个 collection.media目录，把发音和图片copy过来。

现在你的卡片已经可用了

分享卡片

Anki->文件->导出，命一个你喜欢的名字，导出，就大功告成了！



现在把你制作的卡片分享到anki论坛或share deck上，和朋友们一起学习吧！

Anki（知识记忆工具）



还没有评论



写下你的评论