

poem composer 使用说明

计研 111 何巍 2011210839

hwhf6699@gmail.com 13810485194

在 deployment 目录下，有 composer.jar，是可执行的 jar 文件。还有以下几个文件：

- 1) pingshuiyun.txt: 平水韵表
- 2) pattern.txt: 格律表
- 3) poem.txt: 唐诗语料库
- 4) collocation.txt: 搭配表，预处理后的结果
- 5) topic.txt: 主题模型下的 term-topic 分布，预处理后的结果

注：以上的文件均为 GBK 编码，在 linux 下打开的话需要编辑器支持 GBK 编码。

运行环境：要求操作系统装有 JAVA 运行环境，我在 Windows7 和 Ubuntu 12.04 下测试，程序均能正确运行。在 Linux 下运行，请确保系统安装有 GBK，如对于 Ubuntu 系统，可参考 http://blog.sina.com.cn/s/blog_a5b3ccfd0101a0u9.html。

预处理

注意：预处理的结果已提供，分别是 deployment 目录下 collocation.txt 和 topic.txt。因为预处理可能的耗时比较长，所以不建议进行重新进行预处理！建议直接运行 `java -jar composer.jar` 打开 gui 界面。

在进行预处理时，需要准备好 pingshuiyun.txt, pattern.txt 和 poem.txt 这 3 个输入文件，而 collocation.txt 或 topic.txt 为结果输出文件。

抽取搭配词

在命令行下，进入 deployment 目录，输入以下命令：

```
java -jar composer.jar -traincollo [-pingshuiyunpath path] [-patternpath path] [-poempath path]
[-collopath path]
```

其中 -pingshuiyunpath, -patternpath, -poempath 和 -collopath 均为可选参数，默认值分别是 pingshuiyun.txt, pattern.txt, poem.txt 和 collocation.txt。用户可以通过指定这几个选项来修改文件位置。

示例：

```
java -jar compose.jar -traincollo
```

```
java -jar compose.jar -traincollo -collopath mycollocation.txt
```

训练主题模型

在命令行下，进入 deployment 目录，输入以下命令：

```
java -jar composer.jar -traincollo [-pingshuiyunpath path] [-patternpath path] [-poempath path]
[-topicpath path] [-topicnum number]
```

其中-pingshuiyunpath, -patternpath, -poempath, -topicpath 和 -topicnum 均为可选参数，默认值分别是 pingshuiyun.txt, pattern.txt, poem.txt, collocation.txt 和 10.

示例：

```
java -jar compose.jar -traintopic
```

```
java -jar compose.jar -traintopic -topicnum 20
```

```
java -jar compose.jar -traintopic -topicpath mytopic.txt -topicnum 20
```

集句

在进行集句时，需要准备好 pingshuiyun.txt, pattern.txt, poem.txt, collocation.txt 和 topic.txt 等 5 个文件。

在命令行下，进入 deployment 目录，输入以下命令：

```
java -jar composer.jar [-pingshuiyunpath path] [-patternpath path] [-poempath path] [-collopath
path] [-topicpath path]
```

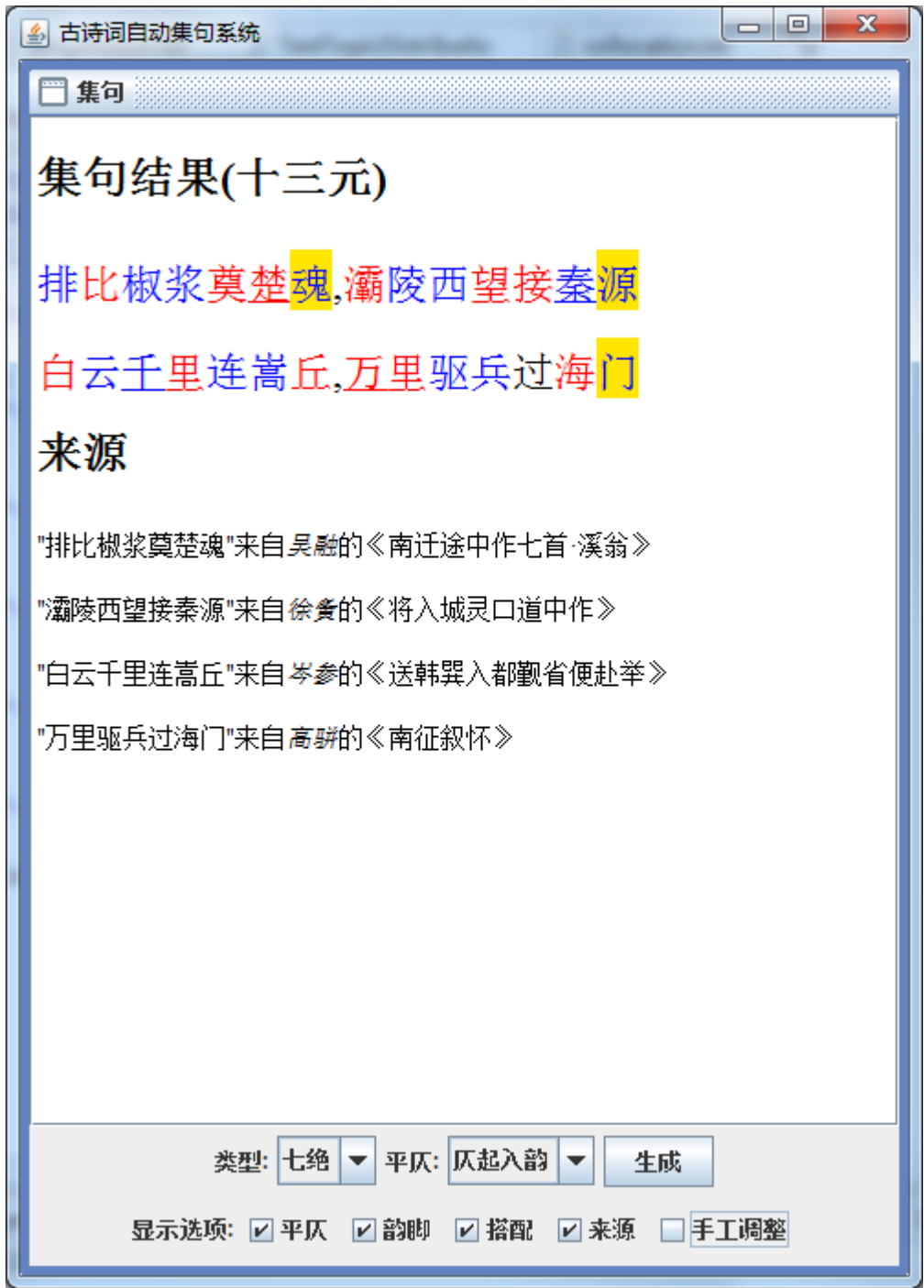
其中-pingshuiyunpath, -patternpath, -poempath, -collopath 和 -topicpath 均为可选参数，默认值分别是 pingshuiyun.txt, pattern.txt, poem.txt, collocation.txt 和 topic.txt.

示例：

```
java -jar composer.jar
```

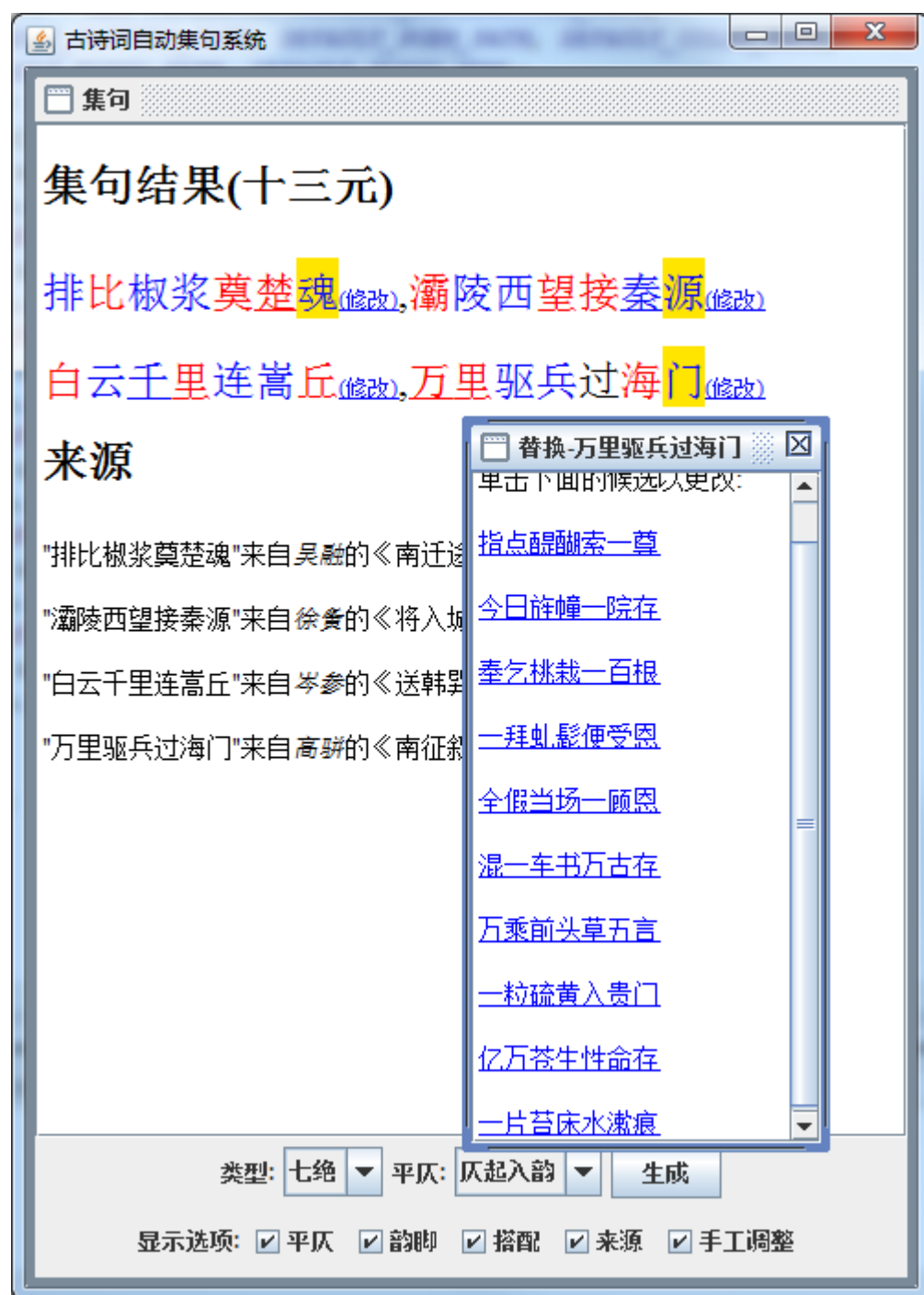
在 gui 界面中，可以在类型中选择“七律/五律/七绝/五绝”，在平仄中选择“仄起/平起/仄起入韵/平起入韵”来指定格律，然后点击“生成”产生集句结果。因为程序采用了缓存机制，

所以第 1 次生成集句结果的速度较慢(大约十秒左右)，以后的速度会加快。如下图所示：



上图中的“十三元”为集句诗所押的韵。可以通过下方的显示选项，设置是否标出平仄(蓝色为平，红色为仄，黑色为均可)，韵脚(黄色高亮部分)，搭配词组(下划线部分)，以及各句诗的出处。

我们可以手工对于程序集句的结果进行调整，在下方打开“手工调整”，此时各句的结尾处会增加一个蓝色的修改链接。单击此链接，可以在弹出的“替换”框中根据程序的推荐手工换一个更合适的句子，如下图所示：



单击“替换”框中的各候选句的蓝色链接以替换。