

作业布置

■ 以下两个选一个来完成:

➤ 编写一个信息检索系统

· 参考: <http://www.rhymezone.com/shakespeare/>



Browse: [Comedies](#), [Tragedies](#), [Histories](#), [Poetry](#),
[Help](#), [Coined words](#), [Most popular lines](#)

Find word or phrase:

Search

☐ Word or phrase ☒ Keywords ☐ Start a line

Keyword search results:

Limit to: [All](#), [Comedies](#), [Tragedies](#), [Histories](#), [Poetry](#)

Alas, poor Yorick! I knew him, Horatio: a fellow ➤ [Hamlet: V, i](#)
Yours; for I will never love that which my friend hates. ➤ [Much Ado About Nothing: V, ii](#)
You may think I love you not: let that appear ➤ [Much Ado About Nothing: III, ii](#)
Out, damned spot! out, I say!--one: two: why, ➤ [Macbeth: V, i](#)
When first I raised the tempest. say, my spirit, ➤ [The Tempest: V, i](#)
With love's light wings did I o'er-perch these walls; ➤ [Romeo and Juliet: II, ii](#)
Is this a dagger which I see before me, ➤ [Macbeth: II, i](#)
I love you with so much of my heart that none is ➤ [Much Ado About Nothing: IV, i](#)

➤ 编写一个简单的搜索引擎

· 参考百度或GOOGLE

作业布置

基于课程学习内容，实现简单的搜索引擎功能（界面可视化不做要求，可终端输出），要求实现以下基本功能：

1. 拼写检查（参考最小编辑距离原理）
2. 倒排索引
3. 使用TF/IDF或者VSM进行文档排序

注: page文件夹是从百度百科词条中抽取的文本，可以作为文档语料

可供参考的文章：

<https://conf.splunk.com/files/2017/slides/a-trip-through-the-splunk-data-ingestion-and-retrieval-pipeline.pdf>

<http://norvig.com/spell-correct.html>

作业布置

Thank you !