

思考题：请为以下程序做模块设计

初始输入：若干文件，每个文件是一篇文章，由若干英文单词组成。例如：file1.txt, file2.txt, file5.txt, ...

调用输入：初始输入中的一个文件的文件名, 例如：file2.txt;

调用输出：从初始输入的若干文件中，找到和调用输入文件最相似的文件（除了它自己），并输出这个文件的文件名，例如：file5.txt。

注意：程序未来可能需要调整，比如，相似度的计算方法可能会调整，又如，调用输入可能允许输入一个新文件，再如，调用输出可能允许输出多个相似的文件，等等。



ChatGPT

为了完成这样的作业，你需要进行模块化设计，这样可以轻松地应对未来可能发生的变更。以下是一个高层次的模块化设计，它涉及到几个关键组件：

1. **文件读取模块 (File Reader):**
 - 功能：负责读取初始输入的所有文件，并将它们存储为可处理的格式。
 - 灵活性：可以调整以支持不同格式的输入文件。
2. **文本处理模块 (Text Processor):**
 - 功能：负责处理每个文件的文本内容，例如分词、标准化和去除停用词等。
 - 灵活性：可以根据需要更新和替换文本处理的方法。
3. **相似度计算模块 (Similarity Calculator):**
 - 功能：负责计算文件之间的相似度。
 - 灵活性：可以容易地替换相似度计算的算法，如余弦相似度、Jaccard相似度等。
4. **查询处理模块 (Query Handler):**
 - 功能：负责处理调用输入，找到并返回最相似的文件。
 - 灵活性：可以扩展以支持新文件的查询或一次返回多个相似文件。
5. **结果输出模块 (Result Outputter):**
 - 功能：负责将找到的最相似文件输出给用户。
 - 灵活性：可以扩展以支持多种输出格式和方法。
6. **主控制模块 (Main Controller):**
 - 功能：作为程序的入口点，协调上述所有模块的工作。
 - 灵活性：可以根据程序的变化调整工作流程。

注意事项

- 每个模块都应设计为尽可能独立，以便在不影响其他模块的情况下修改或替换。
- 通过定义清晰的接口和使用抽象类，可以在不同模块间实现松耦合。
- 在实现时，考虑使用设计模式，如工厂模式来创建相似度计算器，策略模式来定义不同的文本处理策略等。

这种模块化设计使得将来的需求变化能够通过最小的修改来实现，而不需要重写整个程序。