

说明文档

编译信息

- 测试平台：Linux、Mac OS X
- 编程语言：C++
- 编译器：g++
- 编译器版本：6.1.12

程序设计结构

- 面向对象结构，将每个功能都单独分离成为一个类，然后进行运算。

各模块的功能说明

- *KeywordToken.cpp*、*KeywordToken.hpp*:
Token类：作为分词后的结果的承载物。
- *LevenshteinDistance.cpp*、*LevenshteinDistance.hpp*:
相似度计算器一：计算获得的两个tokens之间的莱文斯坦距离，作为最终相似度的一个部分。
- *VectorDistance.cpp*、*VectorDistance.hpp*:
相似度计算器二：计算获得的两个tokens所组成的n维向量，通过计算两个向量之间的夹角的cos值以获得，作为最终相似度的一个部分。
- *FileHelper.cpp*、*FileHelper.hpp*:
文件管理：负责管理文件的输入输出，并且会存储得到的tokens。
- *main.cpp*:
主函数：声明了所需要的对象、依次调用了所有的方法。

实现的功能

- 从指定路径输入的文件中获得相应的信息，最终输出0或1到指定文件路径。
- 使用 levenshtein 距离算法和 n 维向量夹角算法两套算法加权计算结果。

特色

1. 完成了比赛要求的所有功能
2. 使用了两套算法：levenshtein 距离算法和n维向量夹角算法综合计算结果。
3. 通过以上两种算法的加权平均结果，使得识别结果更加精确可靠。