**运行环境：**

Python 2.7

**主要技术：**

SPIMI索引构建

单一字符串压缩字典

Gamma编码压缩倒排索引

**有以下几种命令：**

构建倒排索引 ConstuctIndex <file in 1> ... <file in n> <file out>

<file in 1>到<file in n>表示n个需要构建索引的文件的相对路径

<file out>表示存储倒排索引的文件的相对路径

构建索引之后默认已读入索引

读入并解压倒排索引 ReadIndex <file>

<file>表示要读入的倒排索引的文件的相对路径

关键字的查找 Search <word>

<word>表示要查询的关键字，中间不能含空格

进行这个操作时需要先读入倒排索引

输出语料相关信息 Information

进行这个操作时需要先读入倒排索引

**文件中的倒排索引格式：**

一个int型的数DocNum，表示文档的数量

接下来DocNum行string表示文档的DOCNO，对应的id即为行数

一行string，表示字典的单一字符串

一个int型的数DicNum，表示词项的数量

接下来DicNum项，每项包括两行：

一行三个int型的数，前两个表示词项在单一字符串中的起始位置，第三个表示词项df

一行以文档间隔的形式，Gamma编码后，并用二进制流形式存储的文档索引



