实验1：C++源程序的压缩和解压

1. 实验要求

（一）压缩器

1. 为了提高C++源程序的可读性，C++程序在书写过程中加入了空行、空格、缩进、注释等。假设你想牺牲可读性，以节省磁盘空间，那么你可以存贮一个删除了所有不必要空格和注释的C++源程序的压缩文本。

2.由于C++源程序是由一些具有特定功能的单词组成的，因此我们可以采用转换为编码的方法来进一步实现减小源文件大小的压缩功能。

为了实现这一效果，我们可以先把C++源代码中的各类单词（记号）进行拼装分离，并进行编码的转换。

说明： C++语言包含了几种类型的单词（记号）：标识符，关键字，数（包括整数、浮点数），字符串、注释、特殊符号（分界符）和运算符号等【详细的单词类别及单词组成规则见另外的文件说明】。

3.把具有功能1和功能2的C++源程序压缩器实现，并得到压缩文本。

（二）解压器

1.能够把通过（一）压缩器压缩得到的C++源程序压缩文本进行重新解压并还原出除了被删除的所有不必要空格和注释的C++源程序。

（三）要求应用程序应为Windows界面，在一个界面中可以实现压缩与解压两个功能。

该界面应该具备的详细功能有：

1.打开一个C++源文件，并可以浏览该源程序。

2.压缩所打开的C++源程序

3.查看或浏览得到的压缩文本

4.打开一个C++源程序的压缩文本，并可以浏览该压缩文本。

5.解压该压缩文本，并可以浏览所解压出来的源程序。

1. 代码设计
2. 表
3. 读取.cpp文件
4. 导出.bin文件
5. 读取.bin文件
6. 导出为.cpp文件
7. Windows界面
8. 测试设计
9. 正确率
10. 压缩率
11. 遇到的问题
12. char类型与汉字

先不考虑这个问题

1. cpp文件的编码格式有可能是utf-8，有可能是ANSI

要求用户统一为ANSI

如果要处理utf-8编码，则需要：

第一步：[(45条消息) warning C4828: 文件包含在偏移 0x725 处开始的字符，该字符在当前源字符集中无效(代码页 936)\_shlyyy的博客-CSDN博客\_文件包含在偏移](https://blog.csdn.net/ArthurHai521/article/details/123834153)  
第二步：[(45条消息) C++控制台输出中文时乱码的解决方案\_ufgnix0802的博客-CSDN博客\_c++输出中文乱码](https://blog.csdn.net/qq135595696/article/details/121071282)

1. 实验结果
2. 总结
3. 附录1：源代码
4. 附录2：测试数据