密码学大作业一：单表代换统计分析工具

PB21071417 陈柯宇

实验原理

1. 允许用户使用密钥字，结合用户设置的密钥字与字母频率统计，给出明文。（工具只是辅助破译，让用户去破解）
2. 用`pyside6`设计gui，与用户交互。

实现过程

最初想法

最一开始，我是想做一个自动化比较高的一个自动破译工具，结合字符频率，双字母频率，三字母频率，单词频率等等，做一个智能的工具。

但是我发现这样行不通，在结合单字母统计结果与双字母统计结果，三字母统计结果等等不同统计结果统一分析时，如果出现冲突（例如单字母分析时，密文'a'根据统计结果为'e'，但是双字母分析时，密文'a'根据统计结果为't'，这时该以哪个统计结果为准），我们需要设置优先级，但是以哪一个统计结果优先级高，是没办法判断的，不同的文本，设置相同的优先级，错误率会很高。

但是我还是有一个思路，对于单字母统计结果与双字母统计结果，三字母统计结果，单词统计结果等等统计结果，我们可以选择去相信部分可靠的统计结果，即选择认为密文中出现频率最高的字符对应明文为'e'，结合这个结果，在双字母统计结果中确定与'e'有关的双字母组合，如'he'，'th'，'er'等4-5个频率，以此类推，这样可以建立几个非常可信的映射，对于其他字符的映射，我们可以利用单词，结合大字典进行碰撞（比如已经确定了一个单词的字符个数、不同位置已经确定的明文字符，我们用字典去碰撞，找出符合这些部分字符的明文单词，得到其他未知密文字符的明文映射）。但是我不知道这样是否行得通，也没有这个能力，也问了其他同学的实现方法，于是抛弃了这个想法。

最终实现方法

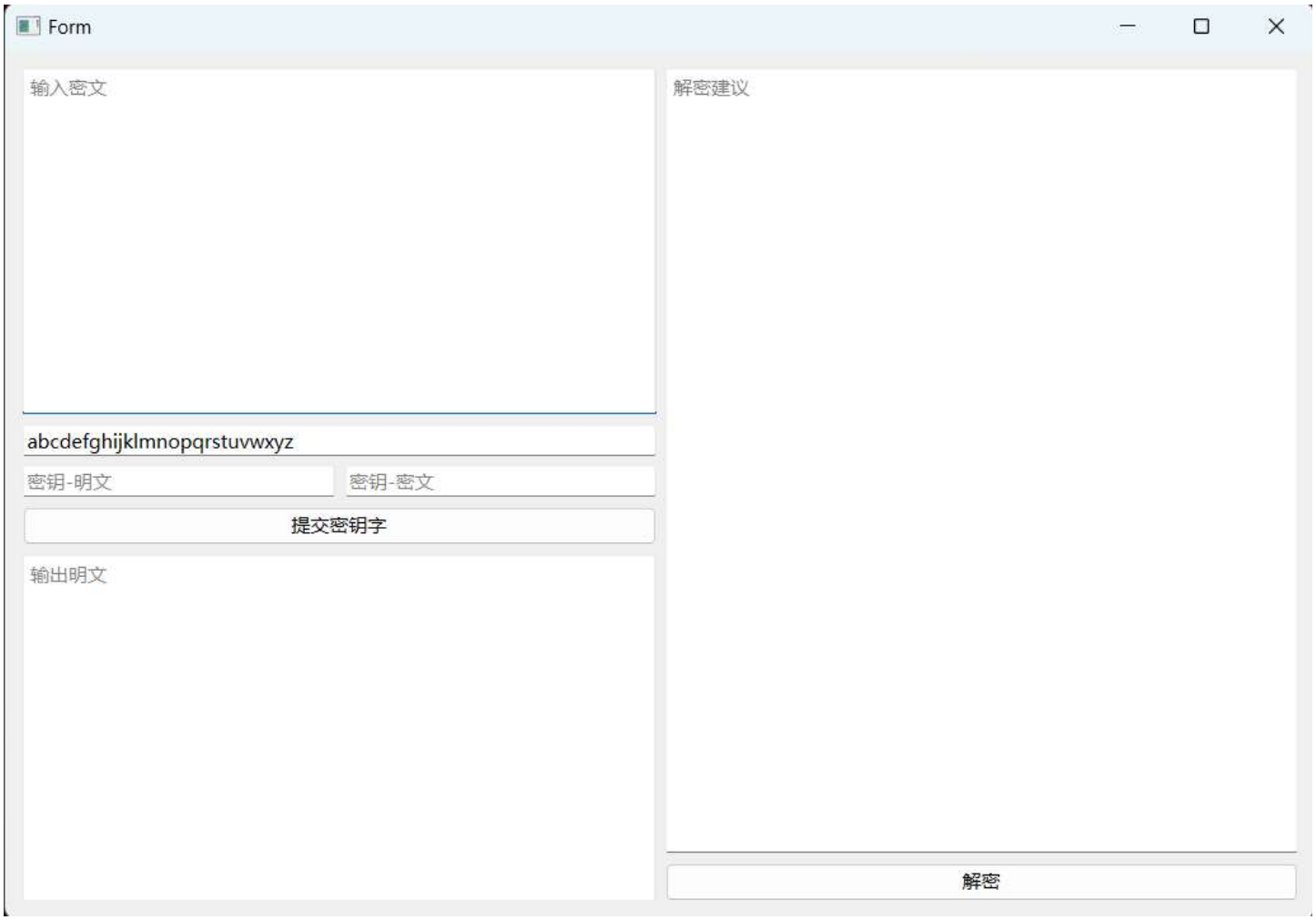
1. 在主程序中设置一个全局变量字典，用于接收用户的密钥字，并且这个字典并不包含用户没有设置的密钥字。
2. 将ppt中给出的字符概率按照概率从大到小设置为一个概率列表。
3. 对密文中的字符进行频率统计，从大到小设置为一个频率列表。
4. 将密钥字字典中的明文-密文映射同时在频率列表与概率列表中删除。
5. 再把改变后的频率列表与概率列表结合起来，形成一个“不可靠的明文-密文映射字典”。
6. 结合用户设置的密钥字字典与“不可靠的明文-密文映射字典”，对密文进行翻译，并且把针对密文做出的统计结果输出，形成“破译建议”。
7. 由用户自己观察当前的破译建议与“明文”（即在当前的明文-密文映射下的翻译），优化更新密钥字，再重复以上4-6步骤，直到用户认为破译结果是正确的。

中心思想：该工具只是辅助破译，主要的破译工作还是要用户自己去完成，主要由用户不断去完善优化密钥字字典，得到最终结果。

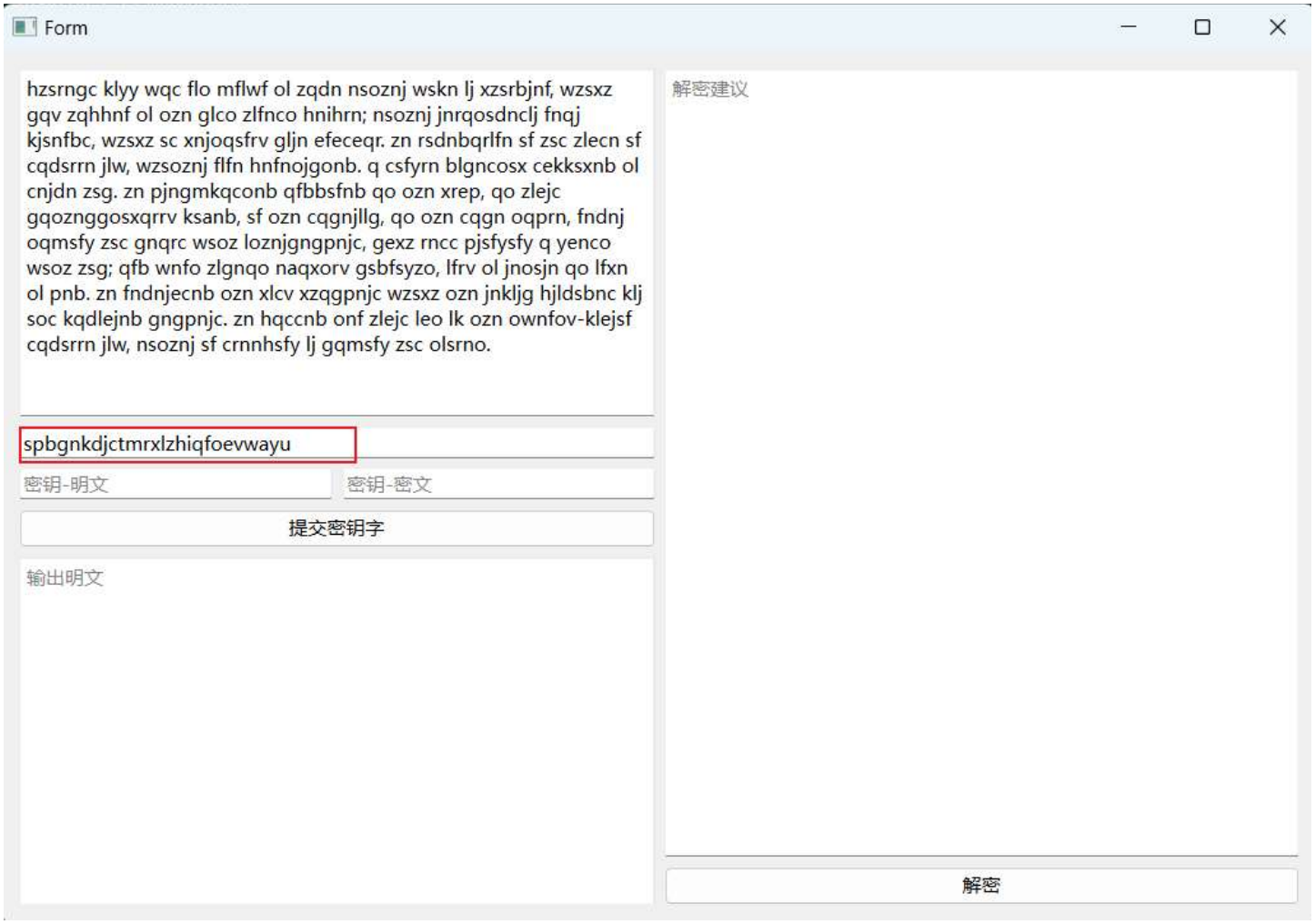
缺陷：如果密文中除了空格之外的字符超过26个，会出现错误。所以我们在测试ppt中给出的密文时，想先把大写字母变成小写。

实验测试

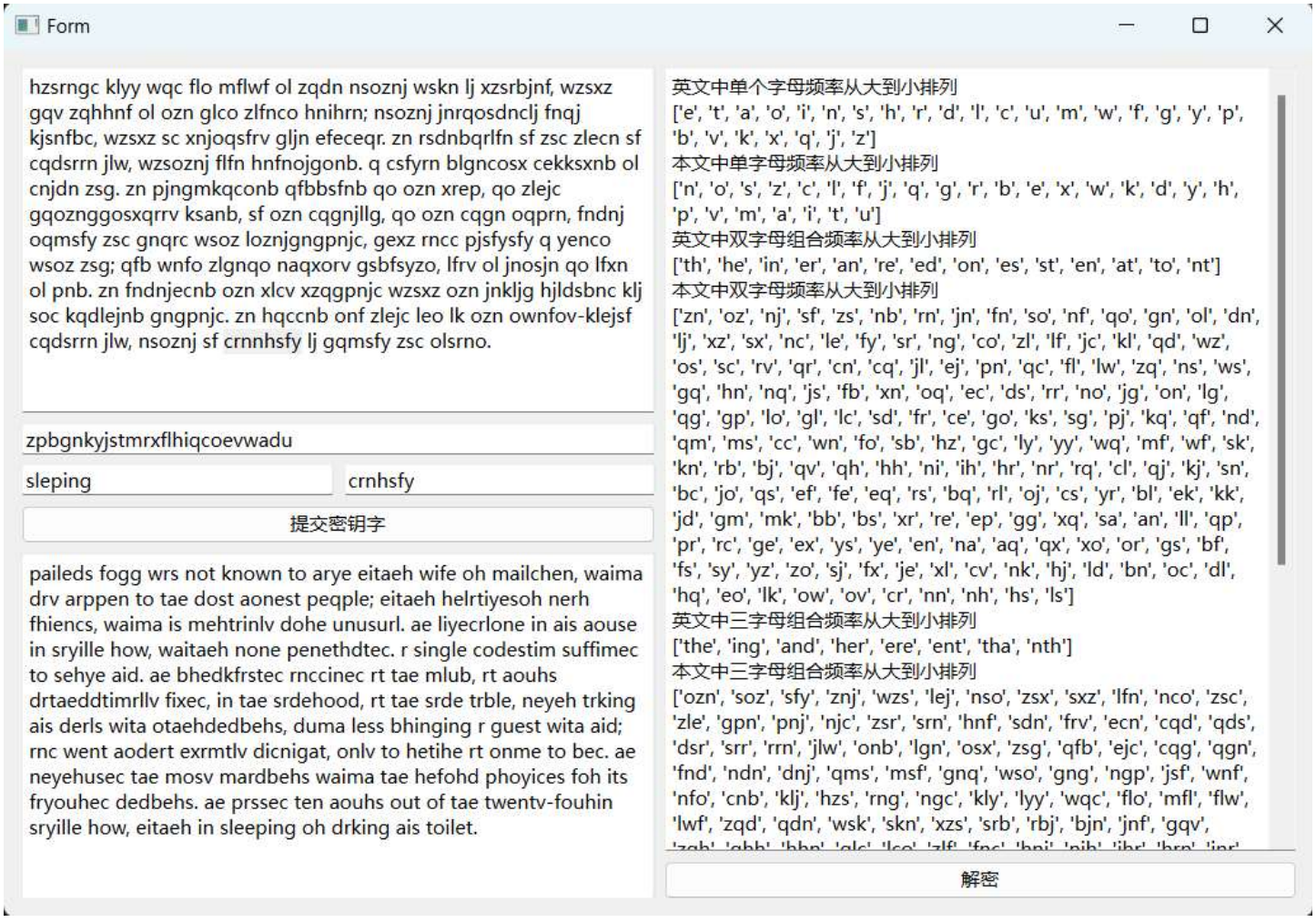
初始界面：



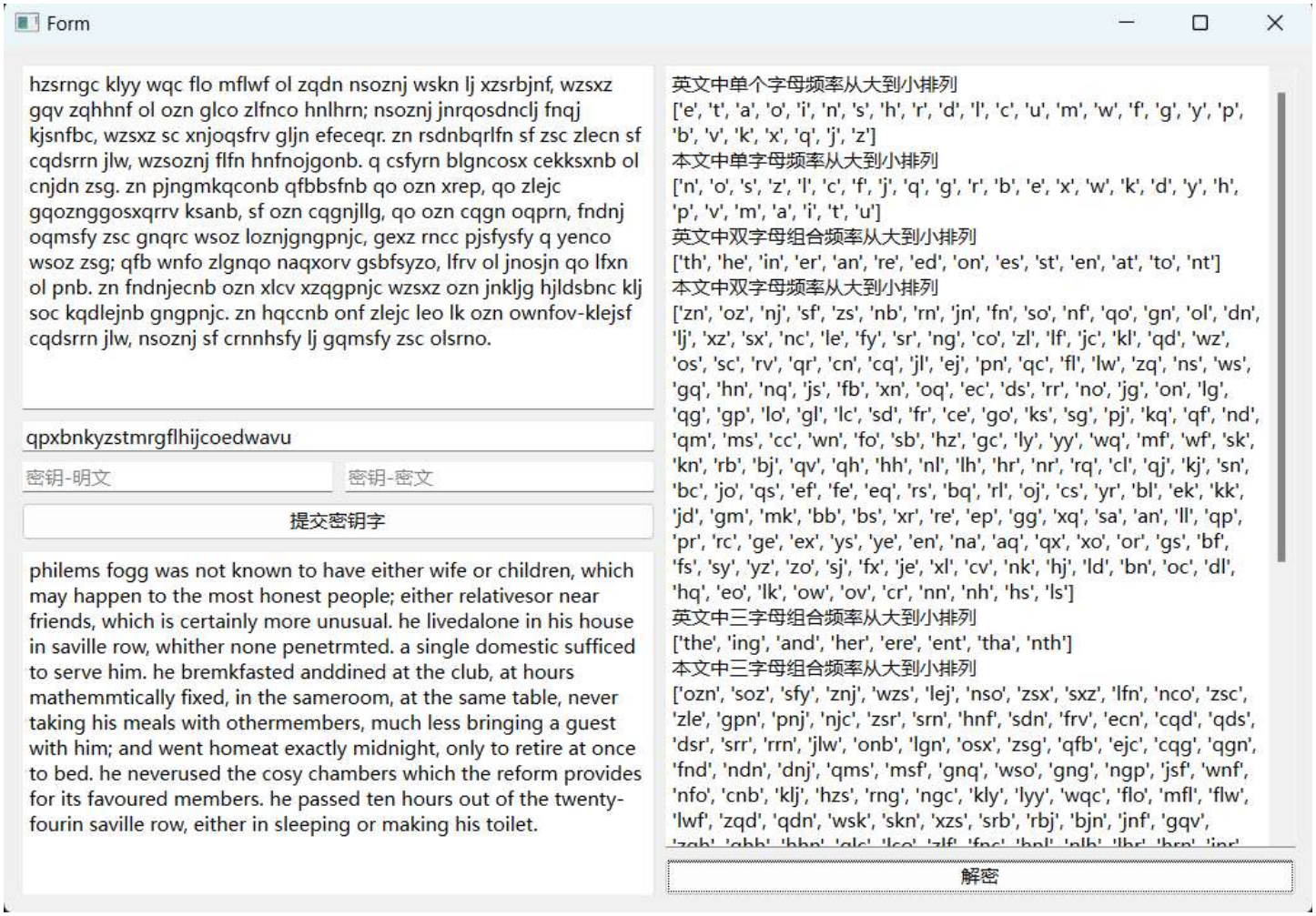
输入密文并且点击 `提交密钥字` 后的结果：



对照“明文”添加映射后的破译结果与破解建议：



不断添加密钥字得到明文：



实验总结

我认为我这个工具做的还蛮不错的，还从网上找了文本加密进行测试，但是还是很考验使用者的，需要使用者有一些灵感。

我学到了如何简单的设计一个GUI,并且把gui与后端结合起来，但是实验过程中其实还是走了一些弯路，写代码之前应该先要考虑好整个思路再写。

感谢陈嘉康同学给我的设计思路(让用户自己去破解)