2020 年臺灣國際科學展覽會
優勝作品專輯

作品編號　**190004**

參展科別　電腦科學與資訊工程

作品名稱　偵測注音文密碼強度之研究

得獎獎項　大會獎：三等獎

就讀學校　臺北市立第一女子高級中學

指導教師　蕭旭君、黃芳蘭

作者姓名　劉孟瑋、吳晏婷

關鍵詞　注音文、密碼

作者簡介

我是劉孟瑋（左），今年就讀北一女中二年級。因緣際會下認識晏婷，找到一個願意指導我們且很用心的教授，同時投入與資訊安全相關的專題。這份專研能完成，要感謝許多我身邊的人，我的好同伴、常常幫助我的老師們、同學還有支持我的父母。雖然做這份專研十分的辛苦，但完成時卻可以感受到強烈的成就感！！希望之後也可以秉持著這一份對專研熱忱，去做任何一件辛苦的事！

我是吳晏婷（右），今年就讀北一女中二年級。在做這一次專研的過程中，我學到許多東西。像是在這一次專研中，我知道注音的美妙之處（雖然我常常使用它XD）、團隊合作的重要以及完成一份研究是多麼的困難。希望之後還有機會可以跟孟瑋或班上的其他人合作！

摘要

密碼為我們生活中常見的一項工具，它可以為我們保護我們的隱私，以免受到別人侵犯，但也同時間出現許多問題，例如：若密碼太過單純會使密碼的功能失去意義，讓其他人可以輕易存取使用者的資料。

在臺灣，注音文密碼是十分常用的密碼，因為它乍看像亂碼，其實有一定的規律存在，而這些密碼，卻容易被判斷為安全的密碼，因此，我們希望可以將這個問題改善。

在研究中，我們先研究密碼強弱，再探討注音文密碼中在一般密碼中的比例，最後達成我們希望的目標—寫出一個可以偵測加入注音文策略的程式。在研究中，我們亦討論各個密碼演算法的優點及缺點，以找出可以最準確判斷的程式，並在研究的最後，提出可以將這套方法擴充至其他語言或是輸入法的可能性。

Abstract

Passwords are a common tool in our daily lives. They can prevent our privacy from being revealed to others, but many problems have occurred as well. For instance, if the password a user sets can easily be guessed by others, its functions of is lost.

In Taiwan, Zhuyin passwords have been widely on the Internet. These passwords look like garbled codes at first, but there is a pattern in fact. If hackers learn these patterns, it will be easy for them to break these passwords. Unfortunately, the password meters that websites use while setting passwords does not consider these Zhuyin patterns. As a result, we intend to solve the problem by developing a program concerning these passwords.

In the project, we first researched on the definition of strong passwords and weak passwords. Then we discussed the percentage of Zhuyin passwords beyond all passwords. Finally, we reach the objective we set up—coding a program concerning the Zhuyin password strategies. In the research, we also compared the strengths and weaknesses of different password algorithms for the purpose of determining the most accurate password algorithm. At the end of the project, we also proposed the possibility of expanding the system to other languages and input methods.

壹、 前言

一、　　研究動機

之前在看自己的社群網站時曾經看到一篇工程師貼出的文章，文章的內容詢問為何”ji32k7au4a83”常常被當作密碼使用？一看電腦鍵盤才知道，這一組密碼是用注音文打出來的，而對應的中文是「我的密碼」。這讓我們產生了興趣，會想要這樣做是因為注音文打出來看起來乍看像亂碼，但注音文密碼並沒有被列入這些判斷程式中。因此，我們希望可以寫出一個程式來偵測密碼是否為注音文。

二、　　研究目的

因為發現各個網站建置新帳號密碼時，沒有偵測注音文密碼的機制，因此希望寫出一個偵測是否使用注音文作為密碼的程式，協助使用者設定更為安全的密碼。

貳、 研究方法及過程

一、　　研究設備及器材

（一）　　硬體

- 筆記型電腦*2

CPU ：IntelCore™i5-8250U 1.80GHz。

RAM：8.00GB

作業系統：Windows10 家用版 x64

用途：撰寫程式、架構開發

（二）　　軟體及工具

1. Code::Blocks ：

程式撰寫介面，使用版本為 17.12，是一個免費、開源、跨平臺的整合式開發環境，使用 C++開發，其使用了外掛程式架構，其功能可以使用外掛程式自由地擴充。目前 Code::Blocks 主要針對開發 C／C++程式而設計。

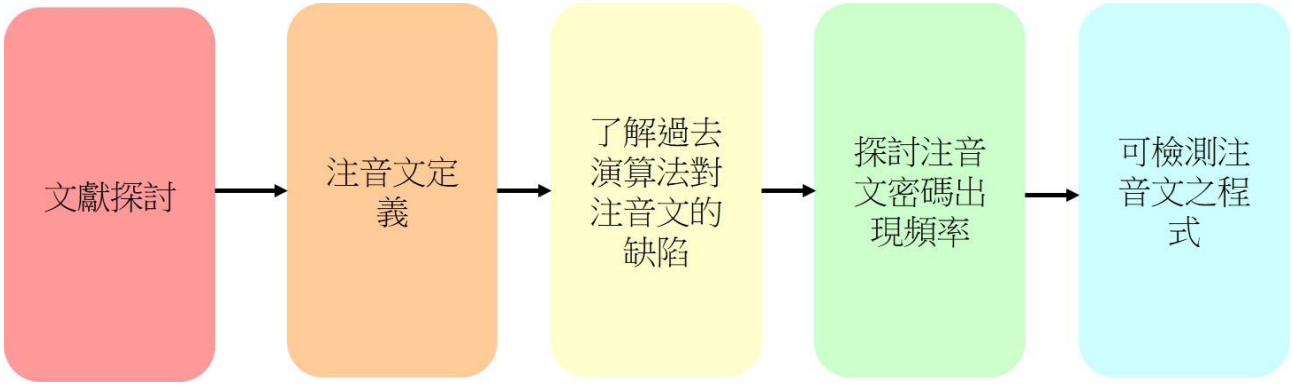
2. C++ ：

程式語言，使用版本為 C++14，是一種被廣泛使用的電腦程式設計語言。它是一種通用程式設計語言，支援多重程式設計模式，例如程序化程式設計、資料抽象化、物件導向程式設計、泛型程式設計和設計模式等。

3. Ubuntu：

Linux 系統，使用版本為 19.04，是以桌面應用為主的 Linux 發行版，Ubuntu 由 Canonical 公司發布，他們提供商業支援，是一個自由軟體。

二、 研究架構



三、 名詞解釋

本研究內容將多次提到以下詞語，為使閱讀者更加了解內容，在此解釋如下：

（一） 注音文密碼

為一轉換後之密碼，經過兩次轉換，第一次轉換是將英文或數字字元利用鍵盤的排列轉換成注音字符，如 a 可轉換成ㄇ，b 可轉換成ㄖ，3 可以轉換成ˇ。第二次轉換則是將第一次轉換後的注音轉換為中文單字。例如：ji3g4gp6，轉換為注音後是「ㄨㄛˇㄕˋㄕㄣˊ」，中文轉換為「我是神」，即為合法注音文密碼。

又因在文獻探討中，並沒有相關的注音文密碼定義，因此，我們將注音文密碼定義為：若在一組密碼中有兩個注音單元且兩者連續，即為注音文密碼。

（二） 注音單元

為可以轉換為中文單字的單體。如：臺灣中的「臺」，其注音為ㄊㄞˊ，而ㄊ+ㄞ+ˊ即為合法注音單元。

四、 文獻探討

在世界上，常用的中文拼音法多數為羅馬拼音，因此，有許多的密碼研究都與羅馬

拼音相關，因此，我們想知道究竟有沒有其他有關注音輸入法的相關論文。

由於注音為臺灣特有之輸入法，我們決定先從臺灣博碩士論文知識加值系統開始查詢是否有相關注音文的論文。密碼方面，我們發現大多數的論文都著重在密碼的解密或是加密，也有一些論文關於深度學習或是生物相關的內容。在注音文方面，因為查詢注音文會使搜尋太狹隘，所以我們先查詢「注音」兩字。發現每一篇論文不脫離教育議題，如：如何讓小一學生學好注音符號？等題目。後來，我們將範圍縮小為「注音文」，發現有一篇名叫「國語注音文詞轉換之研究」的論文對我們的研究可能有幫助。在調閱紙本論文後，我們發現此論文與我們欲達成之方向有些許偏差，它的內容是探討如何將注音轉換為最準確且最通順的中文語句，而我們希望可以將中文字轉換為注音。正好與內容為相反方向。閱讀完這篇論文後，我們得出的結論是：臺灣目前還沒有將密碼與注音結合在一起的文獻。

我們將搜索方向轉換為其他語言與密碼的關連，於是搜尋到” Birthday, Name and Bifacial-security: Understanding Passwords of Chinese Web Users”（ Ding Wang & Ping Wang , 2019），此論文探討主要語言為中文的使用者設定密碼的方式，得出結論：這些使用者主要用他們的生日、名字來設定他們的密碼。其中比較有趣的部份是羅馬拼音密碼與中文的關係，羅馬拼音密碼如” woaini”，轉換成中文即為「我愛你」是一個十分常見的密碼，在 Have I Been Pwned（以下簡稱為 HIBP）中發現已經外洩 85,895 次。HIBP 網站是一個允許用戶檢查個人數據是否被重複使用或是外洩的一個網站，此網站收集許多網站的外洩資料，目前蒐集 41 個外洩網站的資料，一些著名的網站，如 dropbox、linkedin、淘寶網，皆名列其中。另一個拼音密碼” woaini1314”也是，它可以分解為兩個部分，第一部分” woaini”為「我愛你」，而” 1314”在中文的諧音中為「一生一世」，在 HIBP 中也有 17,902 次的外洩紀錄。也是屬於不安全的密碼，因此，在閱讀完這篇論文之後，我們決定參考他們的方式，來對注音文密碼進行研究。

除此之外，我們也參考了” From *Very Weak to Very Strong* : Analyzing Password-Strength Meters”（ Xavier de Carné de Carnavalet & Mohammad Mannan , 2014）此論文來撰寫我們的作品。在這篇論文中，它首先探討大多數的密碼強度偵測器中會有什麼樣的限制，像是長度的限制或是密碼中大小寫字母的限制等。接著，它提到一些可以破解密碼的程式，像是 JtR 等，以及說明曾經有發生重大資料洩漏事件的網站，同時也提到它們建立的 Leet 資料庫，所謂的「Leet」，就是有一些人在設定密碼時，會將其中的幾個字母改變成相似的字母，但本質仍不變。著名的例子如：p@$$w0rd 就是 Leet 之例子。最

後，他們比較世界常用網站的密碼準則，及推測它們可能使用的演算法。

因為在前面探討的文獻中並沒有關於注音文密碼的相關定義，因此，我們定義了「注音文密碼」，原先將注音文密碼定義為只要有一個注音單體就稱作注音文密碼，但發現單單用一個注音單體會容易出現誤判的現象，舉例來說，現有一注音單體：ㄇ+ㄧ+四聲，轉換成鍵盤上對應的字元為 au4，即為合格注音單體，若密碼為 au40001，它就不是真正的注音文密碼。因此，我們將注音文定義為一個詞，如果我們用詞語來判斷的話，會使得錯誤率降低許多，因此我們決定如果密碼有注音單元且兩個注音單元連續的話，這個密碼就定義為注音文密碼。

五、 注音文密碼的不安全性

曾經在一個介紹密碼強度演算法的一篇網路文章中提到設定密碼時可以將中文字詞轉換成鍵盤的注音，原因是這樣可以容易記住又不易被駭客猜到。但其實注音文密碼有一定的規則，只要有心的人，去學習這套規則，就可以輕易將這些密碼破除。

舉幾個典型的例子，如 ji32k7au4a83，轉換成中文則為「我的密碼」，在 HIBP 網站中顯示已經被揭露外洩 141 次。另一個密碼例子是 ji394su3，轉換成中文為「我愛你」，也已經外洩 21709 次。由此可知，注音文密碼實在非常不安全。

六、 各個網站密碼準則

我們從 Alexa Internet[註 1]中選出臺灣人常用的十大網站以及世界上常用十大網站，測驗注音文密碼的強度。我們選用的密碼是 ji32k7au4a83、ji32k7 au4a83、i32k7au4a83j，而 ji32k7au4a83 轉換成注音文為ㄨㄛˇ ㄉㄜ• ㄇㄧˋ ㄇㄚˇ，翻成中文是「我的密碼」，這三個密碼代表的分別為單純注音文、注音文中加一個空白以及注音文的排列組合。考量增加一個空白在注音文中，是因為在研究這些網站的準則時，發現有一些網站是利用空白來判斷強度，排列組合是因為我們亦發現有一些人會將原先想設置的密碼字元，稍微換一些順序，以增加強度，以這個例子中，我們將注音文密碼 ji32k7au4a83 中的「j」從最前方移至最後方。

以下為比較結果：

台灣十大常用網站		ji32k7au4a83	ji32k7 au4a83	i32k7au4a83j	是否偵測注音文
1	google.com	強	強	強	無法偵測
2	youtube.com	與（1）共用密碼系統	與（1）共用密碼系統	與（1）共用密碼系統	與（1）共用密碼系統
3	pixnet.net	僅會顯示密碼過短	僅會顯示密碼過短	僅會顯示密碼過短	無法偵測
4	ettoday.net	不會顯示強度，但會在變更前發送手機驗證碼	不會顯示強度，但會在變更前發送手機驗證碼	不會顯示強度，但會在變更前發送手機驗證碼	不會顯示
5	yahoo.com	不會顯示強度，會在變更密碼前要求再次登入	不會顯示強度，會在變更密碼前要求再次登入	不會顯示強度，會在變更密碼前要求再次登入	不會顯示
6	ltn.com.tw	沒有會員制度	沒有會員制度	沒有會員制度	沒有會員制度
7	google.com.tw	強	強	強	無法偵測
8	setn.com	沒有會員制度	沒有會員制度	沒有會員制度	沒有會員制度
9	facebook.com	中	強	強	無法偵測
10	momoshop.com.tw	不會顯示	不會顯示	不會顯示	不會顯示

2019.10 資料

（圖一）臺灣十大常用網站比較結果

世界十大常用網站		ji32k7au4a83	ji32k7 au4a83	i32k7au4a83j	是否偵測注音文
1	google.com	強	強	強	無法偵測
2	youtube.com	與（1）共用密碼系統	與（1）共用密碼系統	與（1）共用密碼系統	與（1）共用密碼系統
3	baidu.com	僅會顯示密碼過短	僅會顯示密碼過短	僅會顯示密碼過短	無法偵測
4	tmall.com	不會顯示	不會顯示	不會顯示	不會顯示
5	qq.com	不會顯示	不會顯示	不會顯示	不會顯示
6	sohu.com	不會顯示	不會顯示	不會顯示	不會顯示，但會在修改密碼前要求手機驗證
7	facebook.com	中	強	強	無法偵測
8	taobao.com	與（4）共用密碼系統	與（4）共用密碼系統	與（4）共用密碼系統	與（4）共用密碼系統
9	wikipedia.org	僅會輸出密碼太短	僅會輸出密碼太短	僅會輸出密碼太短	無法偵測
10	yahoo.com	不會顯示強度，會在變更密碼前要求再次登入	不會顯示強度，會在變更密碼前要求再次登入	不會顯示強度，會在變更密碼前要求再次登入	不會顯示

2019.10 資料

（圖二）世界十大常用網站比較結果

我們使用這三組密碼分別輸入每一個網站，依網站對密碼的回應情況，來判斷它是否會考慮到注音文的可能，我們將測試的結果紀錄如圖一與圖二的表格中。由以上表格可以知道，幾乎所有的網站都沒有將注音文密碼的可能性考慮到。Facebook 乍看之下是有考慮到，因為它所顯示的強度較其他網站還要弱，但是在加入一個空白之後，強度就從「中」變為「強」，若這個網站有加入注音文策略，則就不會因為一個密碼中的空白而改變顯示強度，因此，可以知道每一個網站幾乎都沒有考慮到注音文的策略。

由以上兩個表格我們更可以分析出，臺灣人常用的網站幾乎為新聞網站，而世界常用網站為購物網站，但兩者比較起來，反倒是臺灣前十大常用網站安全性較世界前十大常用網站的安全性高。這使我們忍不住思考，若這些網站使用者設置的密碼太過容易猜到，那不會很容易造成嚴重的後果嗎？

七、　　一般英文密碼強度研究

由於我們的研究涉及到密碼強度，因此我們從維基百科以及網路尋找資料，探討英文密碼如何定義其強度，以下是我們整理的結論。我們將弱密碼的顏色設為紅色，以警告他人不要使用這類型的密碼，而強密碼設定為綠色，請他人可以多利用這樣的方式設定密碼。

（圖三）弱密碼 v.s.強密碼

八、　　密碼強度檢測演算法

希望可以了解世界上主流的密碼強度檢測演算法為何，於是進行研究，以下介紹幾個較為常見的演算法。

為了使研究結果更具參考性，我們從 github 下載資料庫，並從中挑選幾組注音文密碼，與一般會被各個網站認定為強密碼的密碼比較。我們選用的密碼是由 google 在設定密碼時，會在設定密碼欄下出現的那一串建議設定的密碼（我們將那十組密碼認定為強密碼），取十組並與注音文密碼做比較。看最後的結果為何。

aipdxZVU3qyu5WE	xnZtfvwUmws6Q8U	6dFni9dgerF8biZ	Lv96yJHxcn9h9Tw	8ZaQHXHUdb3Khe8
2YQAEksXriJU9ge	d3ZCNAgYiik3HvQ	qYUaewD85QCJRLE	JCLcTqdQKzGK9AG	3LwRjAqjb6jGWF2

（圖四）十組強密碼

密碼	轉為注音	轉換為中文	種類
vm6jo3xu4	ㄒㄩˊㄨㄟˇㄌㄧˋ	徐偉立	名字
ej03xu3admin	ㄍㄨㄢˇㄌㄧˇadmin	管理admin	中文+英文
1u3ru41p3	ㄅㄧˇㄐㄧˋㄅㄣˇ	筆記本	日常用字
ji394su3	ㄨㄛˇㄞˋㄋㄧˇ	我愛你	日常用字
ji3cp3gj94	ㄨㄛˇㄏㄣˇㄕㄨㄞˋ	我很帥	日常用字

（圖五）選用注音文密碼及其種類

在前面的文獻探討中，我們發現使用者也可能將原本想設定的某一些字元做些改動（也

就是 leet），如將一些小寫字元改為大寫，因此，我們希望可以模擬這些使用者將注音文密碼的幾個字元改動，加入注音文密碼的行列，並做比較。

原密碼	調整後	轉為注音	種類
vm6jo3xu4	Vm6jo3xu4	ㄒㄩˊㄨㄟˇㄌㄧˋ	名字
ej03xu3admin	Ej03xu3admin	ㄍㄨㄢˇㄌㄧˇadmin	中文+英文
1u3ru41p3	1U3ru41p3	ㄅㄧˇㄐㄧˋㄅㄣˇ	日常用字
ji394su3	Ji394su3	ㄨㄛˇㄎˋㄋㄧˇ	日常用字
ji3cp3gj94	Ji3cp3gj94	ㄨㄛˇㄏㄣˇㄕㄨㄎˋ	日常用字

（圖六）調整過的注音文密碼

我們調整密碼的方式為：將密碼中的第一個英文字母調為大寫。例如，現有一組密碼：ji32k7au4a83，就將它調整為 Ji32k7au4a83。

（一） 簡單演算法

簡單演算法顧名思義就是依照密碼長度、字母、數字、符號及其他為評分依據的演算法，舉例來說，以長度而言，若密碼長度小於等於 4 字元即可得到 5 分，若密碼長度為 5 到 7 字元即可得到 10 分，若長度大於八字元即可得到 25 分，但這一個方法有一個缺點，若使用者輸入注音文密碼即可得到極高的分數。

以下為我們利用上述標準訂定的評分規準：

長度	字母	數字
5分：小於等於4字元	0分：無字母	0分：無數字
10分：5到7字元	10分：全部都是小（大）寫字母	10分：1個數字
25分：大於等於8字元	25分：小大寫字母皆有	20分：大於1個數字

符號	加分事項
0分：無符號	2分：字母和數字
10分：一個符號	3分：字母、數字和符號或是大小寫字母、數字
25分：大於等於一個符號	5分：大小寫字母、數字和符號

（圖七）評分規準

項目 密碼	vm6jo3xu4	ej03xu3admin	1u3ru41p3	ji394su3	ji3cp3gj94
長度	25 分	25	25	25	25
字母	10 分	10	10	10	10
數字	20 分	20	20	20	20
符號	0 分	0	0	0	0
加分	2 分	2	2	2	2
總評	57 分	57 分	57 分	57 分	57 分
項目 密碼	Vm6jo3xu4	Ej03xu3admin	1U3ru41p3	Ji394su3	Ji3cp3gj94
長度	25 分	25	25	25	25
字母	25 分	25	25	25	25
數字	20 分	20	20	20	20
符號	0 分	0	0	0	0
加分	3 分	3	3	3	3
總評	73 分	73 分	73 分	73 分	73 分

注音文密碼總平均：65 分

（圖八）利用簡單演算法實測注音文密碼

項目 密碼	aipdxZVU3qyu5WE	xnZtfvwUmws6Q8U	6dFni9dgerF8biZ	Lv96yJHxcn9h9Tw	8ZaQHXHUdb3Khe8
長度	25 分	25	25	25	25
字母	25 分	25	25	25	25
數字	20 分	20	20	20	20
符號	0 分	0	0	0	0
加分	3 分	3	3	3	3
總評	73 分	73 分	73 分	73 分	73 分
項目 密碼	2YQAEksXriJU9ge	d3ZCNAgYiik3HvQ	qYUaewD85QCJRLE	JCLcTqdQKzGK9AG	3LwRjAqjb6jGWF2
長度	25 分	25	25	25	25
字母	25 分	25	25	25	25
數字	20 分	20	20	20	20
符號	0 分	0	0	0	0
加分	3 分	3	3	3	3
總評	73 分	73 分	73 分	73 分	73 分

強密碼總平均：73 分

（圖九）利用簡單演算法實測強密碼

項目 密碼	注音文密碼	強密碼
長度	25分	25分
字母	17.5分	25分
數字	20分	20分
符號	0分	0分
加分	2.5分	3分
總評	65分	73分

（圖十）利用簡單演算法實測注音文密碼與強密碼之比較

分數>=90: 非常安全

80<=分數<90: 安全

70<=分數<80: 非常強

60<=分數<70: 強

25<=分數<50: 一般

0<=分數<25: 非常弱

（圖十一）簡單演算法強度劃分標準

由圖九和圖十我們可以知道，利用簡單演算法檢測注音文密碼以及強密碼時，注音文密碼雖有較強密碼稍弱的成績（而且在經過調整後，分數跟強密碼相同），但強密碼的強度並沒有被認定為「非常安全」，所以，可以確認簡單演算法並不是一個準確測試密碼強度的演算法。

（二） 加分演算法

為簡單演算法稍加改良之後的演算法，在引入加分的同時引入減分。因為其算法有一點複雜，所以直接進行三個密碼的比較。密碼依然選用前述強密碼以及前面的注音文密碼做平均。因為此演算法有使用到加分項以及減分項，因此，我們在表格中將加分項的底色改為綠色，減分項為紅色。以下為其中一個例子：

項目 \ 密碼	vm6jo3xu4	ej03xu4admin	1u3ru41p3	ji394su3	ji3cp3gj94	Vm6jo3xu4	Ej03xu4admin	1U3ru41p3	Ji394su3	Ji3cp3gj94
長度（n*4）	9*4=36	12*4=48	9*4=36	8*4=32	10*4=40	9*4=36	12*4=48	9*4=36	8*4=32	10*4=40
大寫字母數（(n-大寫字母)*2）	(9-0)*2=18	(12-0)*2=24	(9-0)*2=18	(8-0)*2=16	(10-0)*2=20	(9-1)*2=16	(12-1)*2=22	(9-1)*2=16	(8-1)*2=14	(10-1)*2=18
小寫字母數（(n-小寫字母)*2）	(9-6)*2=6	(12-9)*2=6	(9-4)*2=10	(8-4)*2=8	(10-6)*2=8	(9-5)*2=8	(12-8)*2=8	(9-3)*2=12	(8-3)*2=10	(10-5)*2=10
數字（數字個數*4），但數字個數<n	3*4=12	3*4=12	5*4=20	4*4=16	4*4=16	3*4=12	3*4=12	5*4=20	4*4=16	4*4=16
符號（符號數*6）	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0
位於非頭尾數字、符號（非頭尾數字符號*2）	2*2=4	3*2=6	3*2=6	3*2=6	3*2=6	2*2=4	3*2=6	3*2=6	3*2=6	3*2=6
最低條件得分[註3]（最低條件條目數*2）	3*2=6	3*2=6	3*2=6	3*2=6	3*2=6	4*2=8	4*2=8	4*2=8	4*2=8	4*2=8
只有字母（-字母數）	0	0	0	0	0	0	0	0	0	0
只有數字（-數字數）	0	0	0	0	0	0	0	0	0	0
重複字元數	0	0	0	0	0	0	0	0	0	0
連續大寫字母（-2*大寫字母）例如輸入AUB，減2*2	0	0	0	0	0	0	0	0	0	0
連續小寫字母（-2*小寫字母）例如輸入auc，減2*2	(-2)*3=(-6)	(-2)*6=(-12)	(-2)*1=(-2)	(-2)*2=(-4)	(-2)*3=(-6)	(-2)*2=(-4)	(-2)*5=(-10)	(-2)*1=(-2)	(-2)*1=(-2)	(-2)*2=(-4)
連續數字（-2*連續數字數）例如輸入102，減2*2	0	(-2)*1=(-2)	(-2)*1=(-2)	(-2)*2=(-4)	(-2)*1=(-2)	0	(-2)*1=(-2)	(-2)*1=(-2)	(-2)*2=(-4)	(-2)*1=(-2)
正序或逆序字母，如：ABC，但要連續三個才減分（-2*發生次數）	0	0	0	0	0	0	0	0	0	0
正序或逆序數字，如：123，但要連續三個才減分（-2*發生次數）	0	0	0	0	0	0	0	0	0	0
正序或逆序符號（依鍵盤順序而言）但要連續三個才減分	0	0	0	0	0	0	0	0	0	0
總評（加分項+減分項）	76分	88分	92分	76分	88分	80分	92分	94分	80分	92分

注音文密碼總平均：85.8 分

（圖十二）利用加分演算法實測注音文密碼

項目 \ 密碼	aipdxZVU3qyu5WE	xnZtfvwUmws6Q8U	6dFni9dgerF8biZ	Lv96yJHxcn9h9Tw	8ZaQHXHUdb3Khe8	2YQAEksXriJU9ge	d3ZCNAgYiik3HvQ	qYUaewD85QCJRLE	JCLcTqdQKzGK9AG	3LwRjAqjb6jGWF2
長度（n*4）	15*4=60	15*4=60	15*4=60	15*4=60	15*4=60	15*4=60	15*4=60	15*4=60	15*4=60	15*4=60
大寫字母數（(n-大寫字母)*2）	(15-5)*2=20	(15-4)*2=22	(15-2)*2=26	(15-4)*2=22	(15-7)*2=16	(15-7)*2=16	(15-7)*2=16	(15-9)*2=12	(15-10)*2=10	(15-6)*2=18
小寫字母數（(n-小寫字母)*2）	(15-8)*2=14	(15-9)*2=12	(15-7)*2=16	(15-7)*2=16	(15-5)*2=20	(15-6)*2=18	(15-6)*2=18	(15-4)*2=22	(15-4)*2=22	(15-6)*2=18
數字（數字個數*4），但數字個數<n	2*4=8	2*4=8	3*4=12	4*4=16	3*4=12	2*4=8	2*4=8	2*4=8	1*4=4	3*4=12
符號（符號數*6）	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0	0*6=0
位於非頭尾數字、符號（非頭尾數字符號*2）	2*2=4	2*2=4	2*2=4	4*2=8	1*2=2	1*2=2	2*2=4	2*2=4	1*2=2	1*2=2
最低條件得分[註2]（最低條件條目數*2）	4*2=8	4*2=8	4*2=8	4*2=8	4*2=8	4*2=8	4*2=8	4*2=8	4*2=8	4*2=8
只有字母（-字母數）	0	0	0	0	0	0	0	0	0	0
只有數字（-數字數）	0	0	0	0	0	0	0	0	0	0
重複字元數	0	0	0	0	0	0	0	0	0	0
連續大寫字母（-2*大寫字母）例如輸入AUB，減2*2	(-2)*3=(-6)	0	0	(-2)*1=(-2)	(-2)*4=(-8)	(-2)*4=(-8)	(-2)*2=(-4)	(-2)*6=(-12)	(-2)*5=(-10)	(-2)*2=(-4)
連續小寫字母（-2*小寫字母）例如輸入auc，減2*2	(-2)*6=(-12)	(-2)*6=(-12)	(-2)*4=(-8)	(-2)*2=(-4)	(-2)*2=(-4)	(-2)*3=(-6)	(-2)*2=(-4)	(-2)*2=(-4)	(-2)*1=(-2)	(-2)*1=(-2)
連續數字（-2*連續數字數）例如輸入102，減2*2	0	0	0	(-2)*1=(-2)	0	0	0	(-2)*1=(-2)	0	0
正序或逆序字母，如：ABC，但要連續三個才減分（-2*發生次數）	0	0	0	0	0	0	0	0	0	0
正序或逆序數字，如：123，但要連續三個才減分（-2*發生次數）	0	0	0	0	0	0	0	0	0	0
正序或逆序符號（依鍵盤順序而言）但要連續三個才減分	0	0	0	0	0	0	0	0	0	0
總評（加分項+減分項）	96分	102分	118分	122分	106分	98分	106分	96分	94分	112分

強密碼平均：105 分

（圖十三）利用加分演算法實測強密碼

[註 2]表格中有提到「最低條件條目數」，而最低條件條目為 1.密碼長度不小於 8 位　2.包含大寫字母　3.包含小寫字母　4.包含數字　5.包含符號

等級劃分如下：

80<=分數: 非常強

60<=分數<80: 強

40<=分數<60: 好

20<=分數<40: 弱

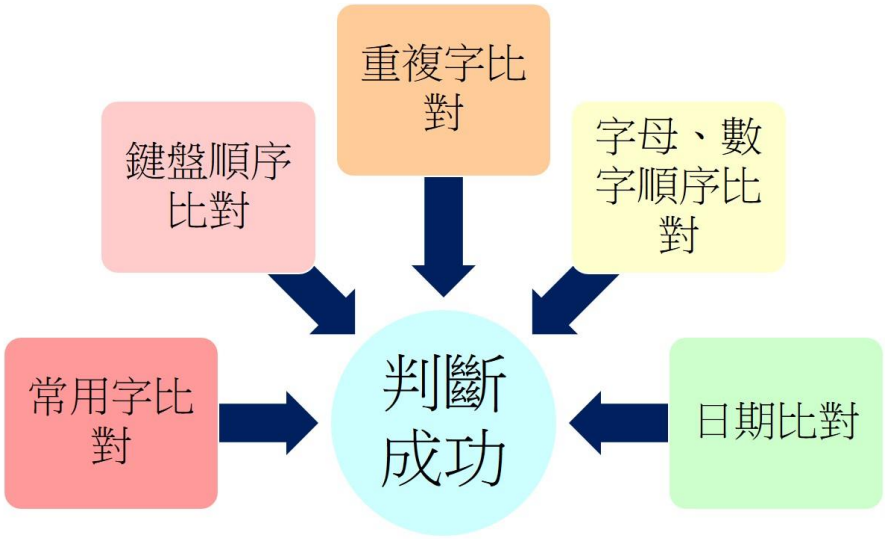
0<=分數<20: 非常弱

（圖十四）加分演算法強度劃分標準

由以上我們可以知道，雖強密碼的分數比注音文密碼的分數來的高，但也可以發現注音文密碼與強密碼的分數在經過平均後，強度皆被列在「非常強」的級別，所以，可以判斷加分演算法也沒有考量到注音文密碼的可能性。

（三）　zxcvbn 演算法

為改良後的密碼演算法，有鑑於前述兩種演算法的缺陷，所改良的新演算法。與其他的演算法不同的是，它加入了許多可能性，而非使用統計數字演算，主要分為五種，這五種方法分別為：常用字比對、鍵盤順序比對、重複字比對、字母、數字、順序比對、日期比對



（圖十五）zxcvbn 的比較方式

以下介紹可能會有疑問的名詞：

密碼複雜度：一個密碼對抗猜測或是暴力破解的有效程度，密碼複雜度越高，表示密碼愈難破解。

我們依然使用上述的強密碼以及上述的注音文密碼做比較。

項目	vm6jo3xu4	ej03xu4admin	1u3ru41p3	ji394su3	ji3cp3gj94
分數（滿分**4**分）	4分	4分	4分	3分	4分
猜測數（**log10**）	12.007	12.028	10.487	8.423	14.843
密碼複雜度	39.886	39.957	34.837	27.981	49.308
分體1	vm6（暴力）	e（暴力）	1（leet）	j（暴力）	j（暴力）
分體2	jo3（leet）	jo3（leet）	u（暴力）	i（常用）	i（常用）
分體3	xu4（leet）	xu3（leet）	3ru41（leet）	394（leet）	3c（暴力）
分體4		admin（常用）	p3（暴力）	su3（leet）	p3g（leet）
分體5					j9（暴力）
分體6					4（leet）

項目	Vm6jo3xu4	Ej03xu4admin	1U3ru41p3	Ji394su3	Ji3cp3gj94
分數（滿分**4**分）	4分	4分	4分	3分	4分
猜測數（**log10**）	12.715	12.264	11.195	8.659	16.024
密碼複雜度	42.238	40.741	37.190	28.765	53.229
分體1	Vm6（暴力）	E（暴力）	1（leet）	J（暴力）	J（暴力）
分體2	jo3（leet）	jo3（leet）	U（暴力）	i（常用）	i（常用）
分體3	xu4（leet）	xu3（leet）	3ru4l（leet）	394（leet）	3c（暴力）
分體4		admin（常用）	p3（暴力）	su3（leet）	p3g（leet）
分體5					j9（暴力）
分體6					4（leet）

注音文平均：3.8 分

（圖十六）利用 zxcvbn 演算法實測注音文密碼

項目　密碼	aipdxZVU3qyu5WE	xnZtfvwUmws6Q8U	6dFni9dgerF8biZ	Lv96yJHxcn9h9Tw	8ZaQHXHUdb3Khc8
分數（滿分4分）	4 分	4 分	4 分	4 分	4 分
猜測數（log10）	21.164	26	23.975	26.886	22.671
密碼複雜度	70.304	86.371	79.644	89.313	75.310
分體一	aip（常用密碼）	xnZ（暴力）	6dF（暴力）	Lv96yJHxcn9h9Tw（暴力）	8（暴力）
分體二	dxZ（鍵盤）	tfv（鍵盤）	ni（常用密碼）		ZaQ（鍵盤）
分體三	VU3（leet）	wU（暴力）	9dg（暴力）		HX（暴力）
分體四	qy（暴力）	mw（常用密碼）	erF（鍵盤）		Hud（常用密碼）
分體五	u5（leet）	s6Q8U（暴力）	8（暴力）		b3（leet）
分體六	WE（常用密碼）		biZ（常用密碼）		K（暴力）
分體七					he（常用密碼）
分體八					8（暴力）

項目　密碼	2YQAEksXriJU9ge	d3ZCNAgYiik3HvQ	qYUaewD85QCJRLE	JCLcTqdQKzGK9AG	3LwRjAqjb6jGWF2
分數（滿分4分）	4 分	4 分	4 分	4 分	4 分
猜測數（log10）	24.753	20.665	24.788	25.883	26.824
密碼複雜度	82.228	68.647	82.343	85.980	89.109
分體一	2YQ（暴力）	d3（leet）	qYU（暴力）	JCLcTqdQKzGK（暴力）	3LwRj（暴力）
分體二	AEks（常用密碼）	ZC（暴力）	a（常用密碼）		A（常用密碼）
分體三	Xr（暴力）	NAgY（常用密碼）	ew（常用密碼）		qjb6jGWF2（暴力）
分體四	iJU（鍵盤）	ii（常用密碼+重複）	D85QCJRLE（暴力）		
分體五	9ge（leet）	k（暴力）		9AG（leet）	
分體六		3H（leet）			
分體七		vQ（暴力）			
分體八					

強密碼平均：4 分

（圖十七）利用 zxcvbn 演算法實測強密碼

項目	強密碼	注音文密碼
分數（滿分4分）	4分	3.8分
猜測數（log10）	24.3609	11.8645
密碼複雜度	80.9249	39.4135

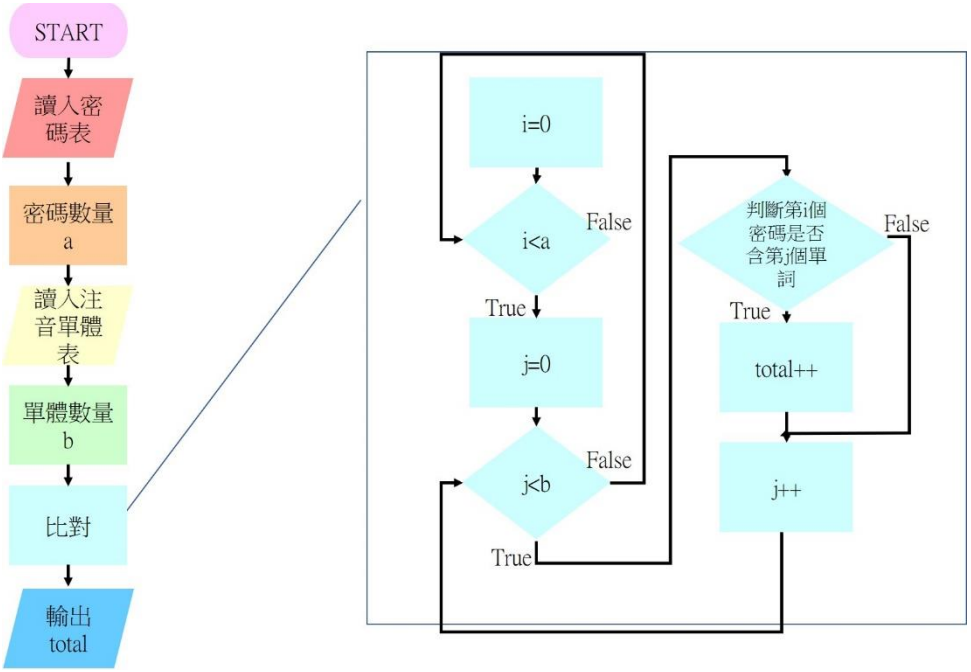
（圖十八）利用 zxcvbn 演算法比較注音文及強密碼

由上表可以得知，zxcvbn 將注音文的級數（平均後）及一般強密碼的級數列為不同的級別，卻仍然得到 3.8 分的分數，與所設定的強密碼相去不遠。但因為注音文密碼的安全性並沒有如此的高，因此，我們決定以 zxcvbn 的程式作為藍圖，加以改良。

九、 注音文密碼探討

（一） 注音文出現頻率分析程式撰寫

為瞭解注音文在密碼中的常用程度，我們做了一個內容為 100 組密碼的密碼表，內含有八個注音文密碼，為「vm6jo3xu4」、「ej03xu3admin」、「1u3ru41p3」、「ji394su3」、「ji3cp3gj94」、「ji32k7」、「au4a83」以及「ji32k7au4a83」，轉換為中文分別為「徐偉立」、「管理 admin」、「筆記本」、「我愛你」、「我很帥」、「我的」、「密碼」以及「我的密碼」。並從網路上找到中文常用字詞表（約 350 字），在把字詞表轉換成注音後，比對程式對注音文的分類是否正確，來檢測我們程式的準確度。



（圖十九）程式想法流程圖

我們總共嘗試了三種方法，來比較出程式對注音文分類準確率最高的方法。

1. 方法一

我們利用網路上下載下來的中文常用字詞表，作為注音的比對表。密碼資料使用那一百筆資料。

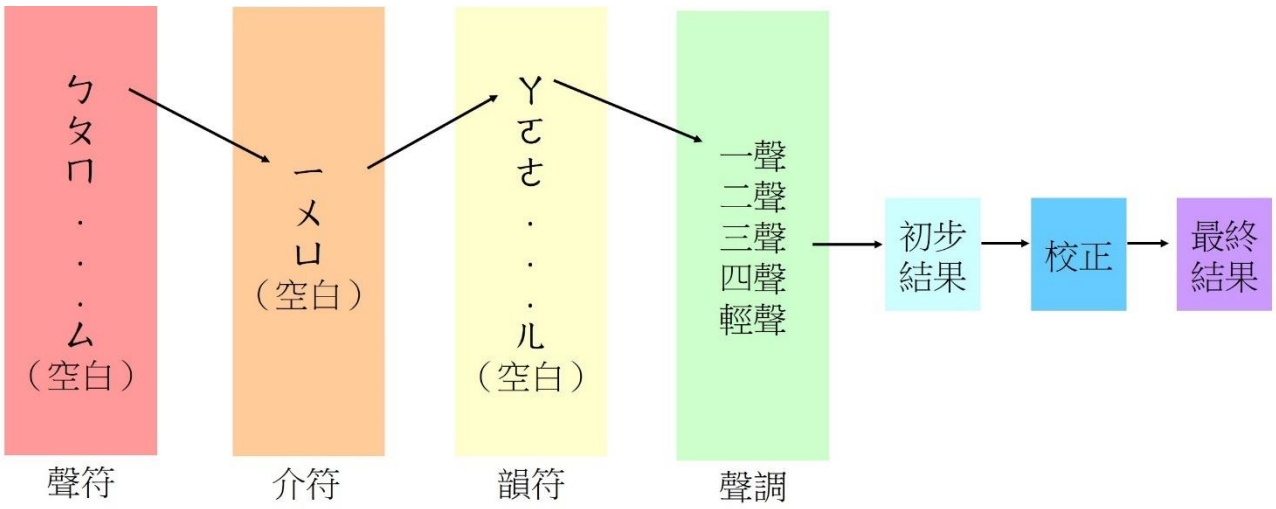
2. 方法二

把原先的注音比對表改成是注音單元，若一組密碼中有兩個注音單元，則將此密碼定義為注音文密碼。如：若密碼為 ji32k7，而在比對資料中剛好有 ji3 以及 2k7 兩個注音單元，就要將分數加一，但若密碼為 ji32k8，因為比對資料中沒有兩個可供此密碼比對的注音單元，則分數就不會加一。使用的密碼資料依然是之前的那一百筆密碼資料。

3. 方法三

注音比對表一樣是注音單元，但將判斷條件改為若有兩個注音單元，且兩個注音單元相連，則定義為注音文密碼。如：現有一密碼 ji32k7 以及 ji3aa2k7，而比對資料中有 ji3 以及 2k7 兩個單元。這兩組密碼在方法二中皆會被列為注音文，但因為第二組密碼的單元分開，因此不被列為注音文密碼，第一組密碼則會被列為注音文密碼，因為兩個注音單元是連續的。密碼資料依然是一百筆的密碼資料

**注音單元表處理方式：我們將注音符號分為四組：第一組是聲符，也就是ㄅ~ㄙ，第二組是介符，ㄧ、ㄨ、ㄩ，第三組是韻符，ㄚ~ㄦ，而最後一組是聲調，就是所謂一聲、二聲、三聲、四聲及輕聲。我們寫一個程式，讓這些符號可以依照我們所編的組別進行排列，並依照鍵盤順序將它們轉換成英文字母，例如：ㄅ+ㄨ+ㄛ+一聲，在轉換成英文字母後即為：2ji+空格。但排列出來的結果並不是所有都為合法的注音單元，例如：ㄅ+ㄧ+ㄛ+一聲，轉換成英文字母後會成為 1ui+空格，卻不是合法的注音單元，因此，我們還需要進行校正。我們的校正方法是利用手動的方式，將不合法的注音單元刪除。



（圖二十）注音比對表處理流程圖

編號	注音	英文對應	編號	注音	英文對應
1	ㄅ	1	22	ㄗ	y
2	ㄆ	q	23	ㄘ	h
3	ㄇ	a	24	ㄙ	n
4	ㄈ	z	25	•	7
5	ㄉ	2	26	ㄧ	u
6	ㄊ	w	27	ㄨ	j
7	ㄋ	s	28	ㄩ	m
8	ㄌ	x	29	ㄚ	8
9	ˇ	3	30	ㄛ	i
10	ㄍ	e	31	ㄜ	k
11	ㄎ	d	32	ㄝ	,
12	ㄏ	c	33	ㄞ	9
13	ˋ	4	34	ㄟ	o
14	ㄐ	r	35	ㄠ	l
15	ㄑ	f	36	ㄡ	.
16	ㄒ	v	37	ㄢ	0
17	ㄓ	5	38	ㄣ	p
18	ㄔ	t	39	ㄤ	;
19	ㄕ	g	40	ㄥ	/
20	ㄖ	b	41	ㄦ	-
21	ˊ	6			

（圖二十一）注音及英文字母鍵盤對照表

（圖二十二）英文字母鍵盤及注音鍵盤對照圖

參、 研究結果及討論

一、 偵測注音文的程式撰寫結果討論

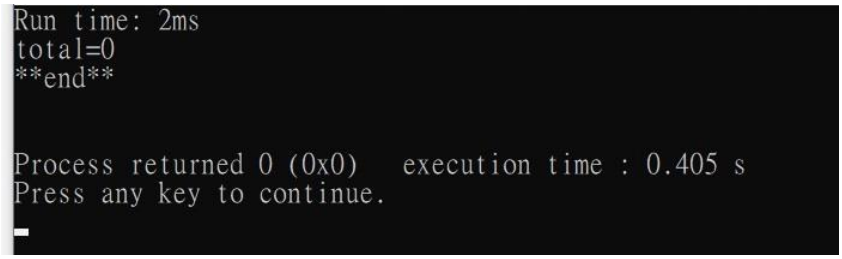
（一） 程式準確率評比公式

為可以更加容易比較三者方法的差異，我們制訂一個程式準確率評比公式，以每一個密碼為一分，若輸出一個分類正確的密碼，則總分加一分，若答錯，則不扣分也不加分。例如：以此密碼表來說，若它輸出有七個注音文密碼，只有錯一個分類，則得到 99 分。滿分為 100 分，最低為 0 分。

方法一：

使用比對表：常見中文字詞表

判斷依據：若密碼中有與比對表中相同字詞，即為注音文密碼。



（圖二十三）方法一輸出

由上圖我們可以發現，這一個方法的誤差十分的大，因為它只有在與比對內容相同時，才會被稱作是注音文密碼，但是若比對表中並沒有包含時，則就不會被包含在內。舉例來說：現有一組密碼 au4a83，翻成中文為「密碼」，但若比對表中並沒有

「密碼」此詞，則就不會被列入。除此之外，也有些人會將密碼設定為自己的名字，而名字是絕對不會被包含在常用字詞表中的，因此，方法二就出現了。

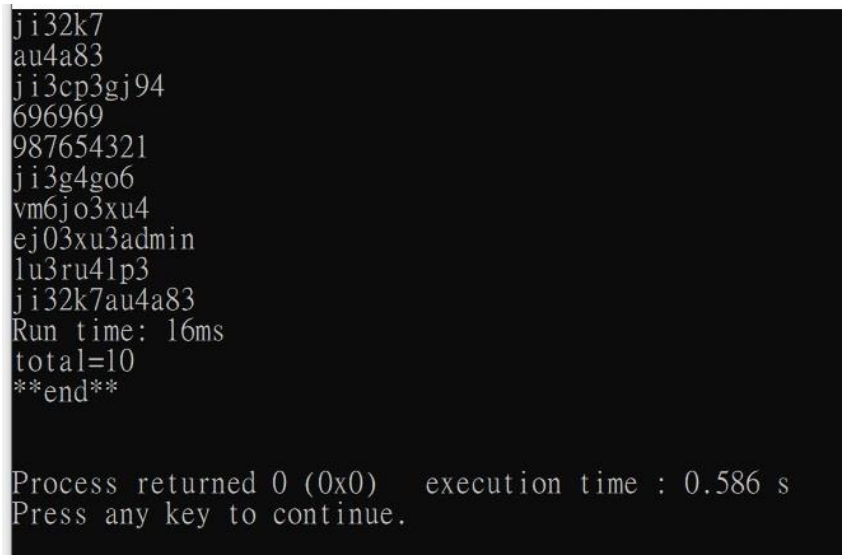
方法一總分：92 分。

原因：沒有檢測到任何一個注音文，也沒有將任何一組非注音文密碼歸類為注音文。

方法二：

使用比對表：注音單元表

判斷依據：若一個密碼中有兩個注音單元，即為注音文密碼。



（圖二十四）方法二輸出

由上圖可以發現，這一百組中裡面預設的八組注音文密碼都有被檢測出來，但還夾雜一些非注音文的密碼。例如：若密碼為 ji32k7，比對資料有 ji3 以及 2k7 兩個注音單元，則就不會有錯誤，但如圖中密碼 987654321，因比對資料中剛好有 87 以及 54 注音單元，87 可以轉換為注音的ㄚ+輕聲，而 54 可以轉換為ㄓ+四聲，兩者都是合法的注音單元，符合判斷條件，於是被判斷為注音文，但此密碼很明顯非注音文。

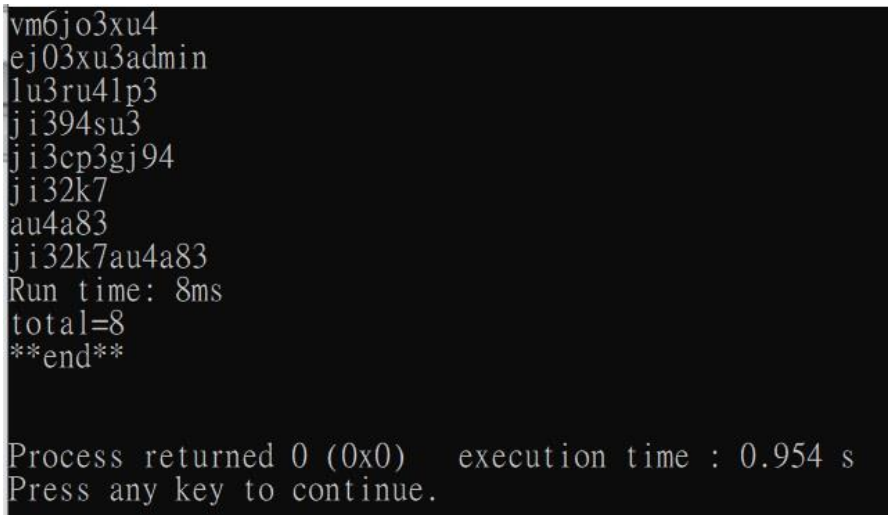
方法二總分：98 分

原因：有檢測到全數注音文，但有兩個非注音文密碼被列為注音文密碼。

方法三：

使用比對表：注音單元表

判斷依據：若密碼中有兩個注音單元，且兩注音單元連續，則為注音文密碼。再加上組合不能重複出現，這兩個條件。



（圖二十五）方法三輸出

由上圖，我們可以知道，這一次的八個注音文密碼不但被檢測出來了，還沒有其他的密碼參雜在其中，因此，我們認為這是我們最準確檢測注音文密碼的方式。

方法三總分：100 分

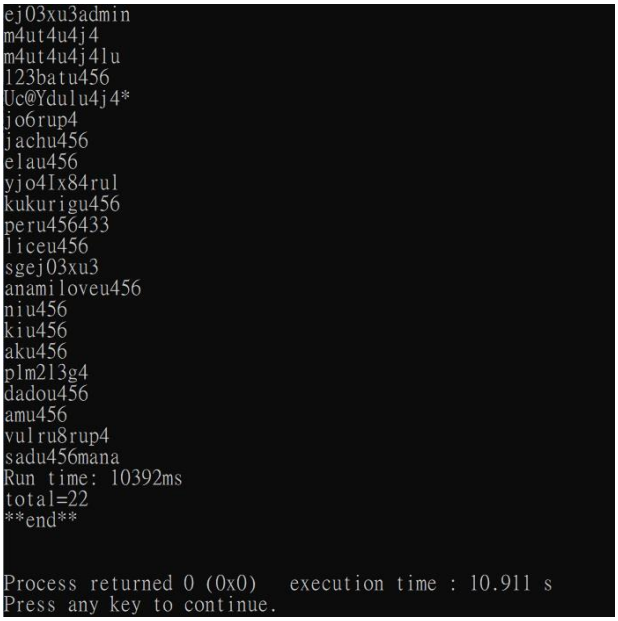
原因：全部的密碼皆正確。

（二） 三種方法之比較

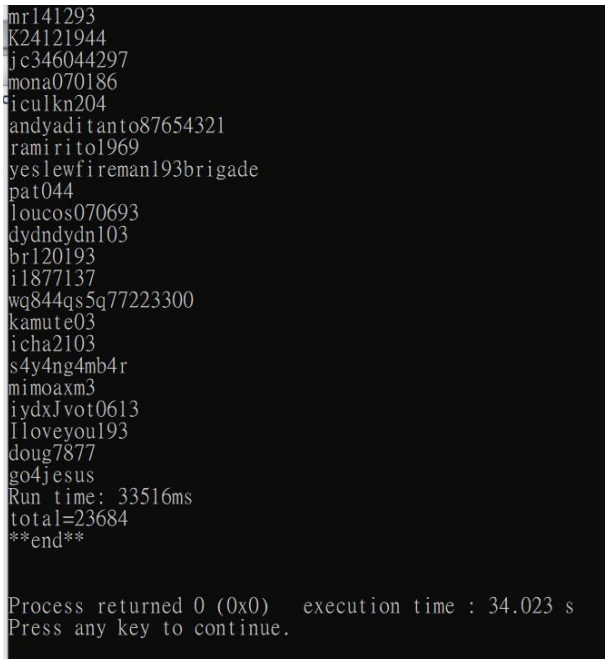
	方法一	方法二	方法三
比對表	中文常用字詞表	注音單元表	注音單元表
依據	符合中文常用字詞表的其中一詞	兩個注音單元	兩個注音單元 兩個注音單元連續 不能出現重複的組合
總分	92分	98分	100分

（圖二十六）三種方法比較

（三）　　注音文密碼比率分析



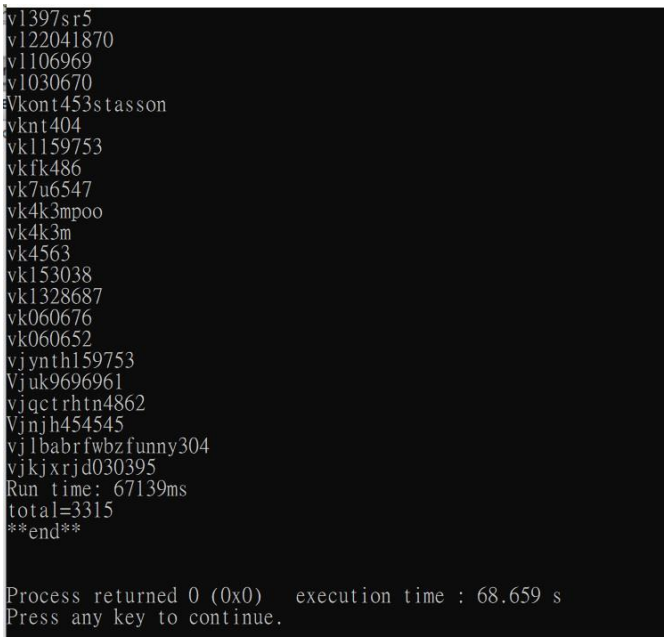
（圖二十七）比率分析方法一輸出



（圖二十八）比率分析方法二輸出

兩者比較後，我們可以發現方法一所檢出的密碼數很明顯的比方法二檢出的少，猜測可能的原因是因為使用的比較表為常用字表，如果使用者將自己的名字設定為密碼，那就不是常用字詞表會出現的字詞，就不會被列入。另外的一個可能是因為

方法二使用的是注音單體表，若密碼中有兩個注音單體就會被列為注音文，因此，會比較容易產生誤判的現象。



（圖二十九）比率分析方法三輸出

我們將它與方法二比較後，發現方法三所輸出的密碼數較方法二少，推測可能是因為多新增了好幾個條件的關係。由上圖我們可以知道，方法三所認定的注音文密碼數有 3315 組，大概佔整個密碼表的 0.3%。

我們在觀察了所輸出的密碼中，發現有一些密碼挺有意思的，比如說，我們在 github 所下載的 000webhost 的密碼資料庫中，有發現一組密碼是 ej03xu3admin，000webhost 是一個可以讓使用者簡單架設專屬網站的一個網站。在這一組密碼表中所發現的密碼有趣的地方是，它並不是單單由注音所組成，而是混雜了英文，前面的「ej03xu3」的部分，轉換成中文是「管理」，後面的英文「admin」，則是中文管理的意思。

在這一組密碼表中，我們也印證了方法一可能不行的原因：那就是，我們在密碼表中發現有人使用自己的名字當作密碼。例如：我們在密碼表中有發現一組密碼：「vm6jo3xu4」，這一組密碼轉換為中文為一個人的名字，推測應該是「徐偉立」。

二、 考量注音文密碼的密碼強度程式

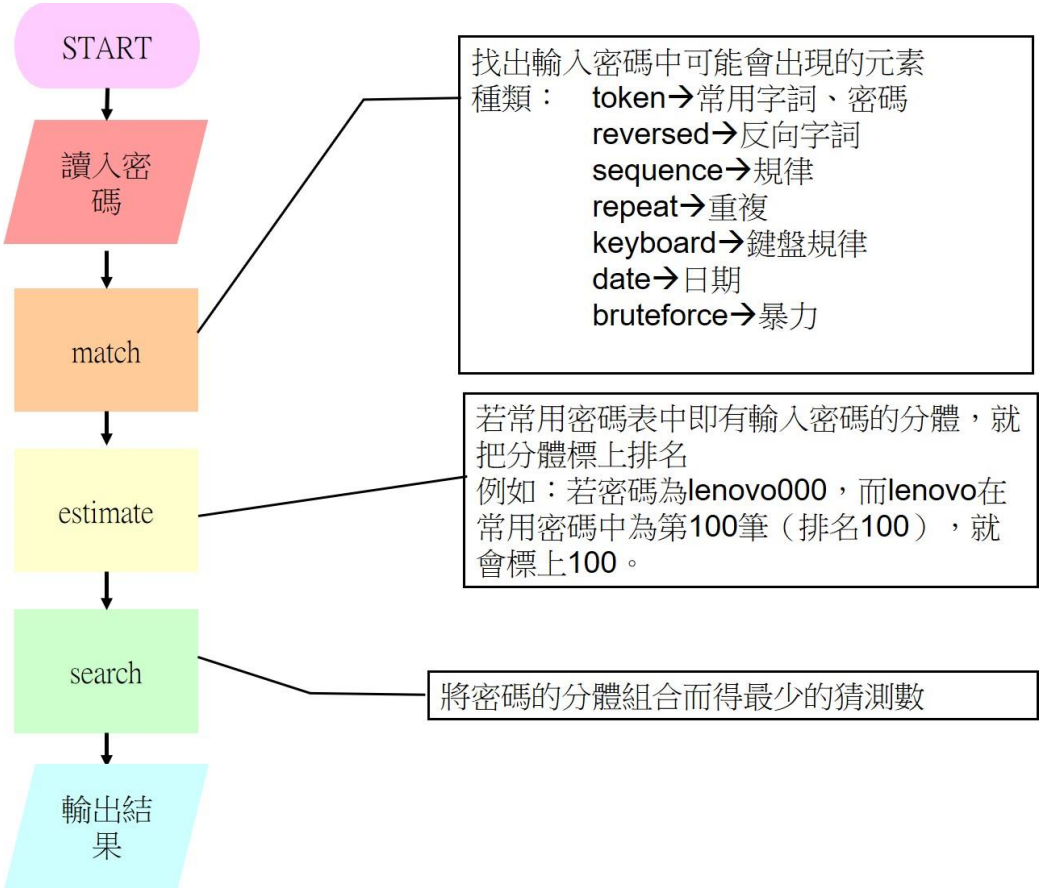
我們分成好幾步驟撰寫這份程式。首先，我們先利用我們所研究出來的英文

密碼強度定義撰寫程式，我們先用研究結果中的順序字元、重複字元以及長度中下手，寫出一個最初級的密碼判斷小程式。

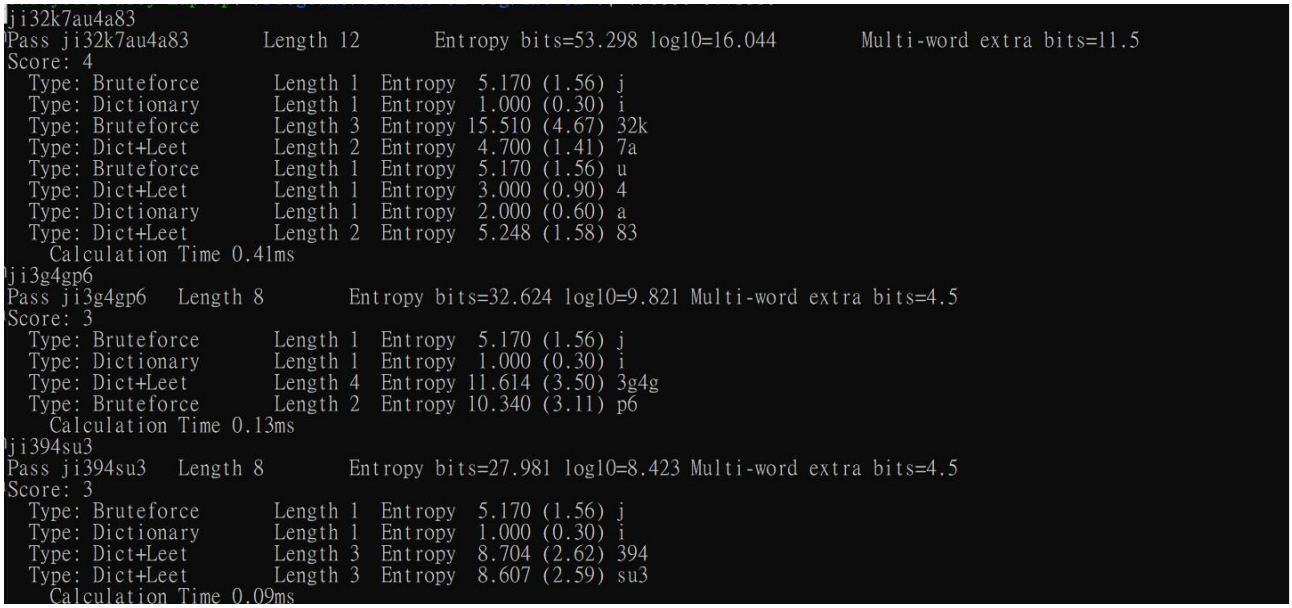
```
12345678
Improvements:
numerical order

123456
Improvements:
too short
numerical order

zxcvbn
Improvements:
too short
keyboard order

abcdefg
Improvements:
too short
alphabetical order
```

（圖三十）小程式輸出

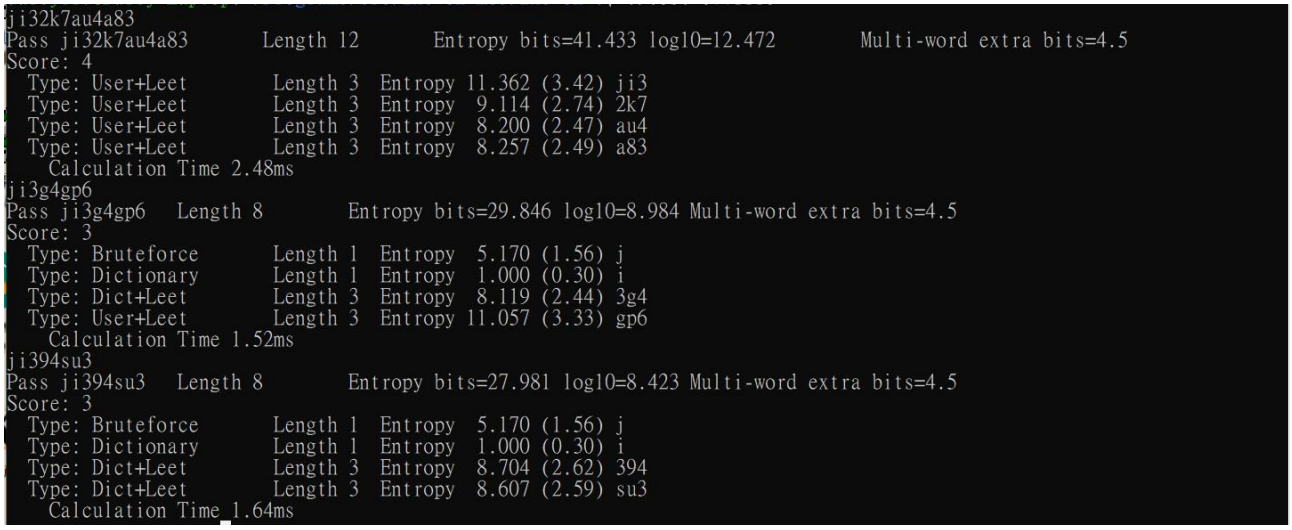
這個小程式可以判斷出長短以及鍵盤順序、字母順序。因為會使用到常用字詞的比對，所以，我們決定用 zxcvbn 的程式進行更改。



（圖三十一）zxcvbn 程式架構



（圖三十二）未加入注音比對表的 zxcvbn 輸出



（圖三十三）加入注音比對表的 zxcvbn 輸出

如圖所示，我們輸入的密碼為 ji32k7au4a83，而在注音單元判斷檔案中，含有「ji3」、「2k7」、「au4」以及「a83」這四個單元。而結果顯示出我們的程式中可以判斷這組密碼含有這些單元。

除此之外，我們還可以看到在加入注音比對表之後，ji32k7au4a83 的密碼複雜度從原先的 53.298 降至 41.433，猜測數（經取 $\log_{10}$）從原先的 16.044 降至 12.472。因此，我們可以知道在加入注音文比對表之後，可以將注音文的猜測數減少。

我們修改程式的方法十分簡單，程式中有一段可以供使用者加入自己的字詞表供比對的程式碼，只要將已校正的注音字詞表加入就可以了。

項目	vm6jo3xu4	ej03xu4admin	1u3ru41p3	ji394su3	ji3cp3gj94
分數（滿分4分）	3分	3分	2分	3分	4分
猜測數（log10）	9.058	9.072	7.658	8.423	10.716
密碼複雜度	30.091	32.229	25.440	27.981	35.599
分體1	vm6（注音）	ej03（注音）	1u3（注音）	j（暴力）	ji3（注音）
分體2	jo3（leet）	xu3（注音）	ru4（注音）	i（常用）	cp3（注音）
分體3	xu4（注音）	admin（常用）	1p3（注音）	394（leet）	gj94（注音）
分體4				su3（leet）	

項目	Vm6jo3xu4	Ej03xu4admin	1U3ru41p3	Ji394su3	Ji3cp3gj94
分數（滿分4分）	3分	4分	2分	3分	4分
猜測數（log10）	9.359	10.003	7.658	8.659	11.017
密碼複雜度	31.091	33.229	25.440	28.765	36.599
分體1	Vm6（注音）	Ej03（注音）	1U3（注音）	J（暴力）	Ji3（注音）
分體2	jo3（leet）	xu3（注音）	ru4（注音）	i（常用）	cp3（注音）
分體3	xu4（注音）	admin（常用）	1p3（注音）	394（leet）	gj94（注音）
分體4				su3（leet）	

利用加入注音比對表的 zxcvbn 實測注音文密碼總平均：3.2 分

（圖三十四）利用已加入注音比對表之 zxcvbn 實測注音文密碼

項目　密碼	aipdxZVU3qyu5WE	xnZtfvwUmws6Q8U	6dFni9dgerF8biZ	Lv96yJHxcn9h9Tw	8ZaQHXHUdb3Khe8
分數（滿分4分）	4 分	4 分	4 分	4 分	4 分
猜測數（log10）	20.599	26	23.975	26.886	22.671
密碼複雜度	68.428	86.371	79.644	89.313	75.310
分體一	aip（常用密碼）	xnZ（暴力）	6dF（暴力）	Lv96yJHxcn9h9Tw（暴力）	8（暴力）
分體二	dxZ（鍵盤）	tfv（鍵盤）	ni（常用密碼）		ZaQ（鍵盤）
分體三	VU3（注音）	wU（暴力）	9dg（暴力）		HX（暴力）
分體四	qy（暴力）	mw（常用密碼）	erF（鍵盤）		Hud（常用密碼）
分體五	u5（leet）	s6Q8U（暴力）	8（暴力）		b3（leet）
分體六	WE（常用密碼）		biZ（常用密碼）		K（暴力）
分體七					he（常用密碼）
分體八					8（暴力）

項目　密碼	2YQAEksXriJU9ge	d3ZCNAgYiik3HvQ	qYUacwD85QCJRLE	JCLcTqdQKzGK9AG	3LwRjAqjb6jGWF2
分數（滿分4分）	4 分	4 分	4 分	4 分	4 分
猜測數（log10）	24.753	20.665	24.788	25.883	24.431
密碼複雜度	82.228	68.647	82.343	85.980	81.157
分體一	2YQ（暴力）	d3（leet）	qYU（暴力）	JCLcTqdQKzGK（暴力）	3Lw（leet）
分體二	AEks（常用密碼）	ZC（暴力）	a（常用密碼）		RjAqjb（暴力）
分體三	Xr（暴力）	NAgY（常用密碼）	ew（常用密碼）		6jG（注音）
分體四	iJU（鍵盤）	ii（常用密碼+重複）	D85QCJRLE（暴力）		WF2（注音）
分體五	9ge（leet）	k（暴力）		9AG（leet）	
分體六		3H（leet）			
分體七		vQ（暴力）			
分體八					

利用加入注音比對表的 zxcvbn 實測強密碼總平均：4 分

（圖三十五）利用已加入注音比對表的 zxcvbn 實測強密碼（未調整過大小寫）

項目	強密碼	注音文密碼
分數（滿分**4**分）	4分	3.1分
猜測數（**log10**）	24.0651	9.1623
密碼複雜度	79.9421	30.6424

（圖三十六）加入注音文的 zxcvbn 之注音文密碼與強密碼比較

項目	注音文密碼（前）	注音文密碼（後）
分數（滿分**4**分）	3.8分	3.1分
猜測數（**log10**）	11.8645	9.1623
密碼複雜度	39.4135	30.6424

（圖三十七）改動前後的 zxcvbn 實測注音文密碼比較

項目	強密碼（前）	強密碼（後）
分數（滿分**4**分）	4分	4分
猜測數（**log10**）	24.3609	24.0651
密碼複雜度	80.9249	79.9421

（圖三十八）改動前後的 zxcvbn 實測強密碼比較

由上三表可以知道，在加入注音比對表後輸入注音文密碼，注音文密碼所花費的猜測數會與未加入注音文比對表前差大約 1000 倍（因為其 $\log_{10}$ 所得之數為 3），在強密碼的部分，則差不到 10 倍（大約 1 倍）。因此，我們可以知道，這一個方法既可以讓注音文密碼被檢測到，但也不會影響其他非注音文密碼的檢測結果。

在加入注音文比對表之後，我們還有發現一件有趣的事情：那就是，有一些注音文密碼在輸入之後會與我們所預期的結果有些許的不同。我們原先以為是錯誤，後來在研究過 zxcvbn 的演算方法後，我們決定將這個結果保留，因為，zxcvbn 有一個機

制，它會先將密碼中可能有的元素加以排列組合後，計算這些不同的組合可能出現的猜測數。因此，我們判斷 zxcvbn 出現與我們預期不相同的結果是因為它經過演算後，認定輸出的結果所花費的猜測數較少。例如：當輸入 ji3g4gp6（我是神）的時候，輸出的會是「ji」、「3g4」以及「gp6」，但預期的輸出應該是「ji3」、「g4」以及「gp6」。

三、 強度的更改

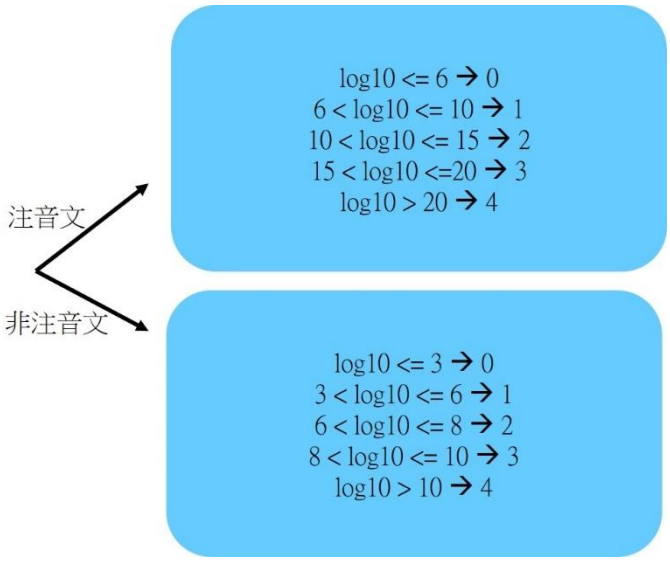
之前有曾經提到，注音文密碼的安全性並沒有想像的高，平均大概只有 2~3 分的強度，屬於中弱強度的密碼。因此，我們決定若偵測器遇到注音文密碼時，就用我們利用標準密碼計算得到的強度等級，同時在顯示器顯示「This is a Zhuyin password. Add a few words to make it longer and stronger.」希望可以提醒使用者他所設定的密碼為注音文密碼，若他希望設定注音文密碼，則他需要再加入一些字詞來讓他的密碼強度更強（也就是符合強密碼中的長密碼形式），或是使用者可以更改密碼，讓他的密碼不是注音文的形式。

標準注音文密碼：

密碼	轉換為中文	長度	猜測數（log10）	訂定強度
au4a83	密碼	6	5.255	0
1u3ru41p3	筆記本	9	7.658	1
ji32k7au4a83	我的密碼	12	12.472	2

（圖三十九）標準注音文密碼及其強度

以下為我們訂定的強度判斷方式：



（圖四十）更改後的 zxcvbn 強度判斷方式

訂定強度的方法十分簡單，以密碼長度作為標準，如標準密碼中的第一項，其密碼長度為 6，經轉換後相當於兩個中文字，因此，強度定為 0，再由標準密碼算出其猜測數經 $\log_{10}$ 的數字，取其整數值而得。

以下為執行結果：

```
ej03xu3admin
Pass ej03xu3admin       Length 12       Entropy bits=32.229 log10=9.702 Multi-word extra bits=2.8
  Type: User+Leet        Length 4  Entropy 10.103 (3.04) ej03
  Type: User+Leet        Length 3  Entropy  9.886 (2.98) xu3
  Type: Dictionary       Length 5  Entropy  9.490 (2.86) admin
    Calculation Time 4.85ms
Score: 2
This is a zhuyin password. Add a few words to make it longer and stronger.

lu3ru4lp3
Pass lu3ru4lp3  Length 9        Entropy bits=25.440 log10=7.658 Multi-word extra bits=2.8
  Type: User+Leet        Length 3  Entropy  5.392 (1.62) lu3
  Type: User+Leet        Length 3  Entropy 10.570 (3.18) ru4
  Type: User+Leet        Length 3  Entropy  6.728 (2.03) lp3
    Calculation Time 0.78ms
Score: 1
This is a zhuyin password. Add a few words to make it longer and stronger.
```

（圖四十一）更改強度的 zxcvbn 輸出結果

以下為更改前與更改後的比較結果：

項目	強密碼（未更改）	強密碼（經更改）
分數（滿分**4**分）	4分	4分
猜測數（**log10**）	24.0651	24.0651
密碼複雜度	79.9421	79.9421

（圖四十二）更改強度的 zxcvbn 實測強密碼比較結果

項目	注音文密碼（未更改）	注音文密碼（經更改）
分數（滿分**4**分）	3.1分	2.1分
猜測數（**log10**）	9.1623	9.1623
密碼複雜度	30.6424	30.6424

（圖四十三）更改強度的 zxcvbn 實測注音文密碼比較結果

由上兩表可以知道，在經過更改的強度演算法在注音文密碼會有作用，而在強密碼的方面，因為其取 log 後的猜測數大於 20，所以並不會影響，因此，我們可以得到這一個更改程式的方法既可以讓注音文顯示的強度降低，達到其警告使用者所設置的密碼的不安全性，也不會讓強密碼顯示的強度降低，失去密碼偵測器的目的。

四、 總測試

為更了解此演算法在注音文密碼的作用為何，我們從前面的頻率分析中所得的注音文密碼，選取約 25 組密碼進行分析，以下為在加入注音比對表以及更改強度的 zxcvbn 實測這些密碼的比較結果：

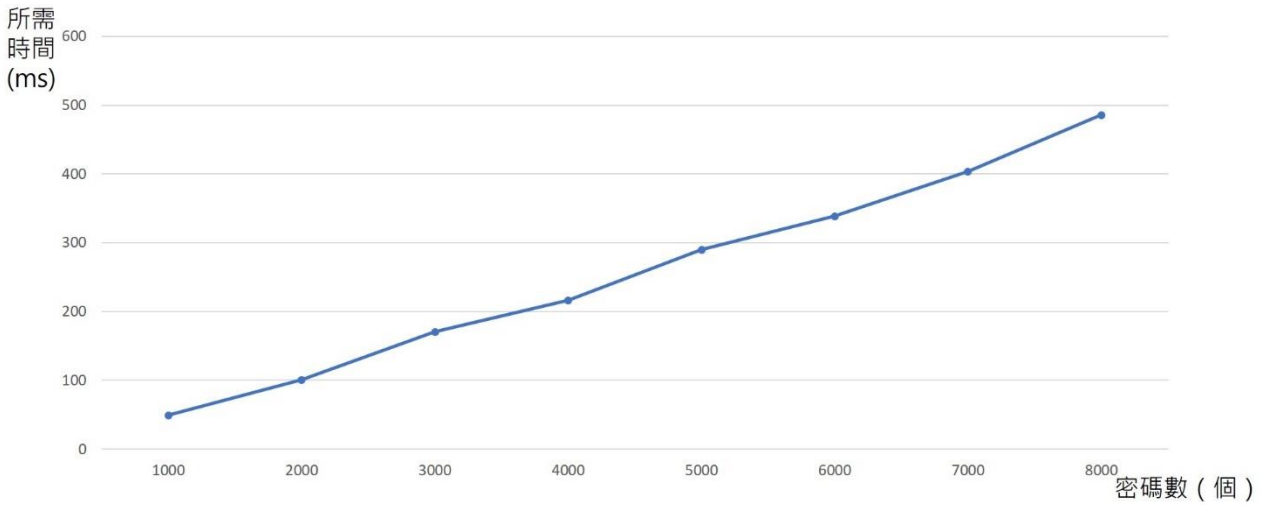
項目	注音文密碼（未更改）	注音文密碼（經更改）
分數（滿分4分）	3.72分	1.4分
猜測數（log10）	12.35	8.55
密碼複雜度	41.01	28.339

（圖四十四）加入注音比對表及更改強度實測注音文密碼前後比較

由以上表格我們可以知道，在未加入注音比對表及更改強度的注音文密碼的分數為 3.72 分，屬於中強密碼等級，而在加入注音比對表及更改強度後的分數為 1.4 分，屬於中弱等級。我們推測可能是因為許多人在設定密碼時，僅會將一個在腦海中想到的詞語輸入，而在此演算法中，僅僅是詞語的注音文密碼不會得到高分，分數才會低。

五、 效能分析

（一） 準確率評比公式

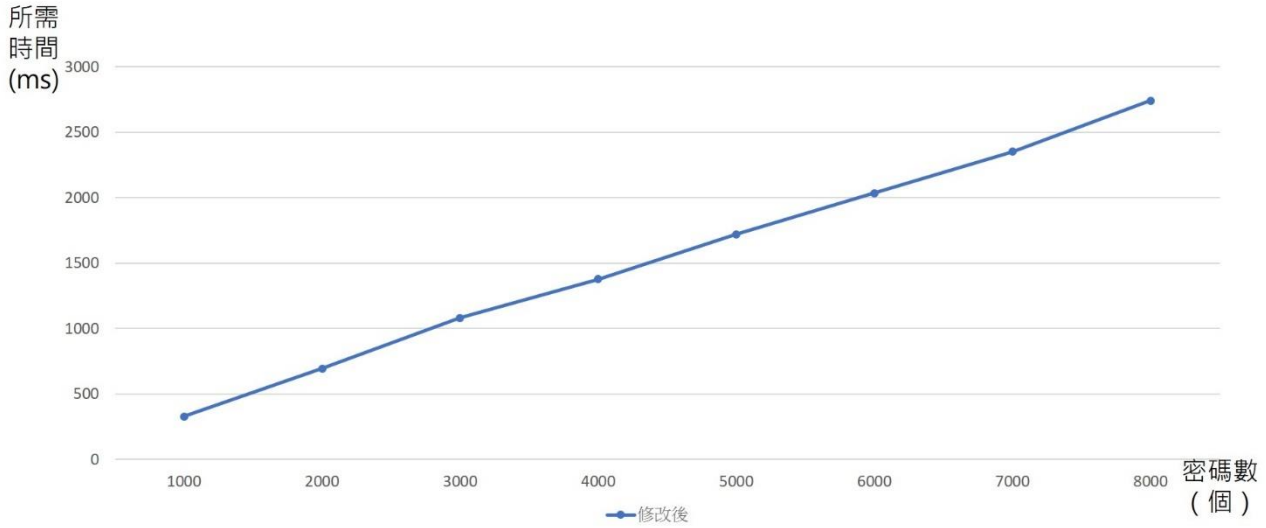


（圖四十五）不同密碼數所費時間

由圖三十七我們可以知道時間關係為線性關係。雖然在程式中有使用三

層迴圈，但因為真正與注音文密碼的判斷有關係的只有一層迴圈，因此，我們可以知道準確率評比程式的時間複雜度為 O(n)。

（二） 加入注音比對表的 zxcvbn 程式



（圖四十六）zxcvbn 加入注音表所費時間比較圖

由圖三十八我們可以知道，加入注音表的 zxcvbn 圖形為線性圖形。又因為我們只有加入注音比對表，並沒有改動其中的程式，而原先的程式的時間複雜度為 O(n)，所以，也可以知道改動後的程式的程式複雜度為 O(n)。雖然加入注音表的 zxcvbn 程式比未修改的程式慢，但時間都沒有超過 3 秒。可以說是十分的快速。

肆、 結論與應用

一、 總結與應用

（一） 我們有發現大家設定注音文密碼的方法可能的規則如下：

1. 直接將注音打出來，例如：我的密碼直接轉換為：ji32k7au4a83
2. 將注音文與英文混雜，例如：管理 admin 轉換為：ej03xu3admin

（二） 可能的取材來源為：

1. 常用詞語，例如：筆記本→1u3ru41p3、密碼→au4a83
2. 職位，例如：管理→admin 或是 ej03xu3
3. 自己的名字

（三） 注音文的占比大約在密碼總數的 0.3%

（四） 在注音文偵測程式的方面，我們的程式可以將注音文密碼所顯示的強度降低，並加上警語，但強密碼的顯示強度依然不變

（五） 因為注音是一個有規律的輸入法，建議大家還是不要使用注音文密碼。

二、 未來展望

（一） 希望之後可以寫出一個程式，推測注音文密碼的可能性為何，以供他人設定密碼參考。

（二） 因為本研究探討中文輸入法為注音輸入法，但中文輸入法仍有倉頡輸入法以及無蝦米輸入法等，若能將這些輸入法一併考量，使用者將會有更大的一分保障。除了中文以外，其他拼音語言如日文、韓文也皆可變為一種應用。

（三） 可以將人工轉換的部分撰寫程式以減低錯誤率。

伍、 參考文獻及其他

一、 Ding Wang and Ping Wang, *Peking University* ; Debiao He, *Wuhan University* ; Yuan Tian, *University of Virginia* .2019 “Birthday, Name and Bifacial-security: Understanding Passwords of Chinese Web Users” In Proceedings of the 28th UNENIX Security Symposium, pages 1537-1554

二、 Xavier de Carné de Carnavalet and Mohammad Mannan, *Concordia Institute for Information Systems Engineering* .2014 “From *Very Weak to Very Strong* : Analyzing Password-Strength Meters”

三、 Daniel Lowe Wheeler, *Dropbox Inc.* 2016“zxcvbn: Low-Budget Password Strength Estimation” In Proceedings of the 25th UNENIX Security Symposium, pages 157-173

四、 Alexa Internet.(2019).Top Sites in Taiwan. Retrieved from https://www.alexa.com/topsites/countries/TW

五、 Alexa Internet.(2019).The top 500 sites on the web. Retrieved from https://www.alexa.com/topsites

六、 ';--have i been pwned?. Pwned Passwords. Retrieved from

https://haveibeenpwned.com/Passwords

七、 TeamsID Business Password Manager.The Top 50 Worst Passwords of 2018.Retrieved from https://www.teamsid.com/100-worst-passwords/

（50~100）,https://www.teamsid.com/100-worst-passwords-top-50/

八、 維基百科。密碼強度。檢自：https://zh.wikipedia.org/wiki/%E5%AF%86%E7%A0%81%E5%BC%BA%E5%BA%A6。

九、 Github.Zxcvbn. Retrieved from https://github.com/dropbox/zxcvbn

十、 Github.Password data. Retrieved from https://github.com/robinske/password-data

十一、 Github Password data-2 Retrieved from https://github.com/danielmiessler/SecLists/tree/master/Passwords/Leaked-Databases

十二、 密碼強度檢測演算法分析及實現案例說明。檢自：https://www.itread01.com/content/1547107754.html

十三、 其他

〔註 1〕我們在找尋資料中，發現網路上有許多資料良莠不齊，因此最終選定 Alexa Internet 的資料作為參考，Alexa Internet 為一個亞馬遜的一個子公司，其專業是在網路的流量以及網站排名。

十四、 附錄

一百組密碼：

	1	2	3	4	5
1	vm6jo3xu4	123123123	aaaaaa	lol1234	123qwe123
2	ej03xu3admin	azerty	haha123	asdasd	lizard123
3	1u3ru41p3	123qwe	1234567890	1q2w3e4r	1qazxsw2
4	ji394su3	hallo123	Password1	acUn3t1x	lololol
5	ji3cp3gj94	lizardsquad	swag123	abcdefg	Lizard
6	ji32k7	qwerty123	lol123321	lol123456	boobs
7	au4a83	qwertyuiop	penis	qweqweqwe	lizardstresser
8	ji32k7au4a83	g00dPa$$w0rD	qwe123	swagswag	liverpool123
9	123456	nigger123	herpderp123	1q2w3e	faggot
10	lol123	lollol	123321	fuckyou123	1q2w3e4r5t
11	123456789	Password123	admin	qwert123	ed.fisica
12	12345	123qweasd	abcd1234	fuckoff	654321
13	test123	testtest	asdasdasd	qwerty1	asdfghjkl
14	password123	nigger	1qaz2wsx	1234qwer	football
15	123123	lizard	nascar48	niggers	12341234
16	password1	12345678	asdf1234	fucku	nick1234
17	abc123	azerty123	poop123	test1234	lolol
18	qwerty	yoloswag	pokemon	pass123	pussy
19	exedra345	1234567	azertyuiop	hahaha	aaabbbccc
20	penguinjoanne	lol12345	hello	blablabla	hola123

（圖四十六）一百組密碼

【評語】190004

- 本作品的原創性高，研究內容完整。

- 本作品的可擴充性高，有許多發展的空間

  （例如可加入理論分析，或討論其他語言和輸入法）。