中華民國第58屆中小學科學展覽會作品說明書

高級中等學校組 電腦與資訊學科

052510

反切密碼-風斬槍靴滅世嘛喇

學校名稱:臺中市立臺中第一高級中等學校

作者:

高一 林堅壬

高一 賴彥呈

高一 柯品安

指導老師:

林奇鋒

關鍵詞:密碼、自然語言、反切

摘要

我們當初的理念是是希望能將傳統的加密方式融合現代的密碼學,達成更高的安全性, 後來我們發現能結合語言學的部分,使得加密變得更有變化。

我們的研究利用「反切」這種古代的拼音方式來設計一套加密規則,並利用它替特定的字串加解密。我們並沒有試著挑戰一般的加解密,而是在原本的基礎上對內文動手腳,也就是說最終可能依然得依賴一般的加密方式,但在此之前我們已將內文經過加密成為密文,使得即使破解了一般的加密,結果依然是一串無法理解的密文,像是第二道保險一般,使文字在原有的加密基礎上再多一層保障。

壹、研究動機

在現今的社會中,資訊在我們生活中占了很大一部分,因此資訊安全的重要性在這個情況之下也變得更為顯著。

在發現這個問題之後,我們便想著有沒有方法能夠在密碼學的方面上做出一些新的改變,而在尋找方法的過程中,我們試著向古人借鏡,便發現了反切音這種形式。

反切是一種音韻學的發展分支之一,原是一種注音方式,分為反切上字及反切下字,發展到後期有人將反切音作為傳遞密文的方式,以此為構想,我們由**分**到**ム**的聲母,一、**メ**、**山**的介母,**Y**到**儿**的韻母以及音調,由一字分成二字,將聲母及介母分成上半部,韻母以及音調組成下半部。

例如:「心」,各隨機選擇一個上半部同樣為「T一」以及下半部同樣為「**ウ**」的字,像「小奔」即為「心」的反切。

歷史上最早將反切應用在密碼上的是明朝名將戚繼光將軍,他將此應用在軍事通訊上。 他是將中文密文藏在詩句中。戚繼光專門編了兩首詩歌,作為「密碼本」,一首是:「柳 邊求氣低,波他爭日時。鶯蒙語出喜,打掌與君知」;另一首是:「春花香,秋山開,嘉賓歡 歌須金杯,孤燈光輝燒銀缸。之東郊,過西橋,雞聲催初天,奇梅歪遮溝。」

兩首詩歌是反切碼全部秘密所在。取前一首中的前 20 個字的聲母,依次分別編號 1-20;取後一首 36 字韻母,順序編號 1-36。再將當時福州方言字音的八種聲調,也按順序編上號碼 1-8,形成完整的「反切碼」體系。其使用方法是:如送回的情報上的密碼有一串是 5-25-2,對照聲母編號 5 是「低」字,韻母歌編號 25 是「西」字,兩字的聲母和韻母合到一起了是 di,對照聲調是 2,就可以切射出「敵」字。

戚繼光將軍的方法以現代知識來說便是利用由詩詞組成的「資料庫」以及數字組成的「密文」來達到加密的效果,於是我們結合反切音和反切碼的原理制定了一套規則,而有了現在這個研究。

貳、研究目的

- 一、找出一套加解密規則,能適用大多數的字音
- 二、能較完整地利用反切的原理進行中文的加解密
- 三、希望能從語言學的角度達成加密的效果,證實古人的智慧也能應用在先進的 科技上。

參、研究設備與器材

一、開發環境:

- (—) Microsoft Visual Studio Community 2017
- (二) Microsoft Office Access 2016
- (三) Microsoft .Net Framework 4.5.2

二、電腦設備:

- (→) CPU: Intel Core i7 7700 3.60GHz
- (二)記憶體:16.0GB
- (三)硬碟:SSD: ADATA 512GB、HDD: Seagate 2TB 7200RPM

肆、研究過程與方法

一、流程



- 研究密碼特性
- 研究反切特性

(二)尋找 資料來源並 製作資料庫

- 採用新酷音輸入法
- 將其依需求製作成資料庫



- 設定注音分割規則
- 訂定轉換規則

(四)編寫程式並除錯、修改

- 分工分別完成加密部分和解密部分
- 經過除錯測試發現規則缺陷並修改

(五)測試 不同數據並 統整

- 整合加、解密部分為一測試程式
- 測試多筆資料並統整數據

研究過程

二、

一開始決定以反切音作為研究對象後,首先先查詢反切音相關的資料,了解它的規則以及特性,以掌握研究的方向。而在這個過程中發現了戚繼光將軍的「反切碼」這個應用方法, 反切碼的原理使我們聯想到當今「資料庫」的概念,而有了利用資料庫相互比對的方法加密 中文的構想,並開始實際嘗試。

既然是以資料庫為中心的構想,開始之後第一個遇到的問題自然就是資料庫了,無論是字音的判斷或是加密、解密都需要資料庫才能達成,於是我們便開始尋找資料庫。在尋找的過程中,我們以開源的語詞庫為目標為原則尋找,以利研究的方便,最後便找到了新酷音輸入法的語詞庫作為我們這次研究的資料庫。

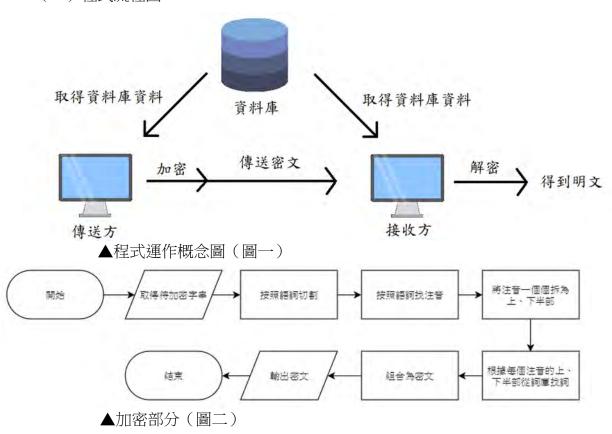
有了資料庫後,便是要訂定轉換規則了,一開始我們先從簡單的方法開始嘗試,也就是「逐字隨機加密」,但在做了幾筆測試之後,發現總會有幾個字轉換出來的的字不一樣,在調查那幾個字之後,才發現是中文特性——字多音造成的問題,所以我們就繼續再想新的加密規則。

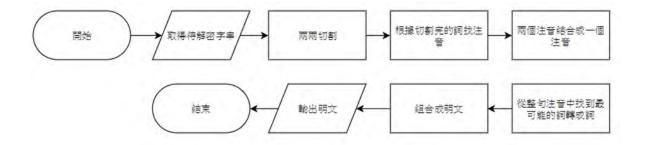
而在發現逐字隨機加密的問題之前我們就有想著是否能讓加密出來的文字具有一點文學的意義,而能夠讓人看不出那串文字是密文,就像是戚繼光將軍所用的詩句一般,能夠混淆 試圖破解的人。所以我們在想新的方法時就往這個方向去思考,最後得出了「以詞加密」這 個加密規則。這個加密規則雖然不能使整串密文具有連貫的意義,但每兩個字都是具有意義 的詞而能夠達到類似的效果。

因為撰寫程式時,加密與解密是由兩個人分別負責的,所以最後需要一個整合的過程使整個程式更符合邏輯,且也能夠將兩個部分結合並撰寫大量測試的程式。在完成之後經過一 些值錯的過程以及測試及統計的過程後便終於完成了這次的研究。

三、研究說明

(一)程式流程圖





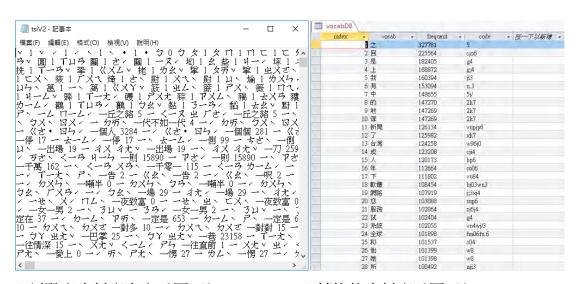
▲解密部分(圖三)

(二)資料庫處理:

在找資料庫時,我們希望它是開源的,以便做任何的研究而不會受太多限制, 所以找到了新酷音輸入法的字詞庫。

字詞庫的內容有字、詞以及他們的使用頻率及讀音,在判斷讀音以及加密實都是不可或缺的資料。

但是它的字詞庫是 TXT 檔,不利於資料搜尋的效率,且注音符號在程式中處理 相對英文字不便,所以我們將這個 TXT 檔做簡單的轉換後匯入至 Access 中成為一個 新資料庫。



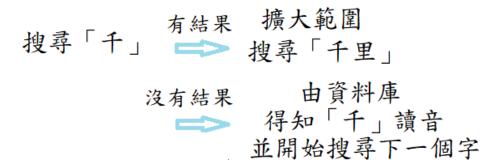
▲新酷音資料庫內容(圖四)

▲轉換後資料庫(圖五)

(三)判斷語句字音:

我們判斷字音的方法是以詞切割語句,而實際的方法如下圖:

以「千里共嬋娟」為例:



▲判斷字音流程解說圖(圖六)

如此循環數次後便可得到「千里共嬋娟」的正確讀音。

(四)注音加密規則:

▲注音分部解說圖(圖七)

如上圖所示,我們將一個字的注音分成上半部和下半部。上半部由聲母和介母組成,而下半部由韻母與聲調組成,以便加密時做使用。

而加密的規則我們想了兩種:

1、逐字隨機加密

隨機加密解說圖(圖八)

假設某字經過字音判斷後得知讀音為「坐メム`」,先將它分成上下兩部分,分別是「坐メ」以及「ム`」,接著以隨機的方式將「坐メ」的部份加上下半部並將「ム

`_ 部份加上上半部後最後各自轉回中文字便形成密文。

優點:因為處理方法單純所以速度相對快速。

缺點:因破音字問題導致錯誤率相對高。

2、以詞加密



▲以詞加密(圖九)

假設某字經過字音判斷後得知讀音為「**出以以以**),先將它分成上下兩部分,分別是「**出以**」以及「**ム**)」,接著在資料庫中以**SQ** L 查詢首字上半部同為「**出以**」且尾字下半部同為「**ム**)」的兩字的詞,便為此字轉換為反切音後的密文。

優點:詞的同音的機率低所以在字音方面能夠達到接近 100%正確率,且密文的內容有文字上的意義可以混淆試圖破解的人而達到更好的加密效果。

缺點:處理複雜導致速度較慢。

伍、研究結果

我們在做測試時,除了針對不同字數作了測試,也對不同類型的字詞作了測試,以下是我們作的測資以及結果。

明文	密文	明文	密文
約會	預約回饋	你好不好	年底很好補助很好
你好	年底很好	茉莉花茶	沒錯領域回家長達
作業	坐坐一頁	可口可樂	可可可否可可垃圾
美國	魔鬼規模	反切密碼	發展契約名字魔法
倫敦	論文鍍金	不鏽鋼菜刀	不足銷售剛剛此外蛋糕
歐盟	歐洲美容	阿拉伯半島	二八喇叭本國本站代表
美術館	魔鬼碩士觀點	你有沒有吃飯	年底也有買回也有成績發現
礦泉水	款項去年說給	紅燒牛肉湯麵	歡迎商標念頭然後台商名片
腳踏車	減少台大唱歌	今天不想去上班	結婚天天補助香港事項榜單
五月天	位址漁業天天	你想吃牛肉麵嗎	年底香港成績念頭然後名片媽媽
出師表	傳輸社區表表	布宜諾斯艾利斯	補助以及農作似乎安泰領域似乎
太平洋	他在平衡銀行		

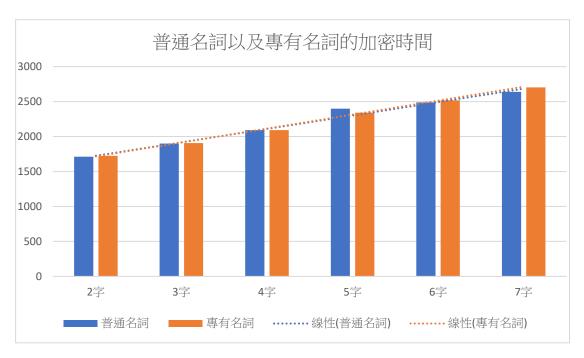
▲測試的測資及結果(圖十)

我們測資中的語詞內容分為兩類,普通名詞以及專有名詞,上圖中以黃色註記的測資即 為專有名詞。

	加密平均秒數	解密平均秒數	總花費平均秒數	正確率
2字	1717.333333	1306.833333	3024.666667	10
3字	1903.5	1310.5	3214.666667	10
字	2092.25	1298	3391	10
				10
字	2370.5	2361	4732.5	
字	2503.5	2688.5	5193	10
7字	2660.333333	2860	5521	10
	普通字詞	解密平均秒數	總花費平均秒數	正確率
字	1711.666667	1298,666667	3010.666667	10
			3207.333333	
字	1900.333333	1306.333333		10
字	2091.5	1301.5	3394	10
字	2399	2483	4883	10
字	2489	2779	5269	10
字	2639	3000.5	5640	10
	專有名詞	解密平均秒數	總花費平均秒數	正確率
VC-2-7	Part Control of the C			
字	1723	1315	3038.666667	. 10
字	1906.666667	1314.666667	3222	10
字	2093	1294.5	3388	10
字	2342	2239	4582	10
字	2518	2598	5117	10
字	2703	2579	5283	10
7.				
	加密時間標準差	解密時間標準差	總花費時間標準差	正確率標準差
字	47.45666667	39.75	76.32833333	
字	23.865	21.26833333	31.73	T
字	33.7225	23.0425	44.82	T T
5字	83.505	78.995	144.005	
字	37.095	28.135	45.38	
7字	113.6066667	124.0233333	225.6666667	
	加密時間標準差	解密時間標準差	總花費時間標準差	正確率標準差
2字	70.78666667	52.64333333	114.0666667	
3字	23,49	23.42666667	33,10666667	
4字	20.16	20.73	29.79	
5字	134.59	113.66	232	
6字	47,285	25.355	59.85	
7字	100.71	110.34	202.795	
	加密時間標準差	解密時間標準差	總花費時間標準差	正確率標準差
2字	24.12666667	26.85666667	38,59	**************************************
	and the second s	*		*
3字	24.24	19.11	30.35333333	
4字	47.285	25.355	59.85	
5字	32.42	44.33	56.01	
6字	35.52	27.25	46.14	
7字	139.4	151.39	271.41	
	加密時間相對標準差	解密時間相對標準差	總花費時間相對標準差	正確率相對標準差
2字	0.027666667	0.03045	0.025266667	0
3字	0.012533333	0.016233333	0.009866667	. 0
1字	0.0138	0.017725	0.0132	0
5字	0.03495	0.0328	0.02985	0
6字	0.013475	0.01045	0.00875	0
7字	0.042266667	0.0437	0.040733333	0
	加密時間相對標準差	解密時間相對標準	總花費時間相對標準差	正確率相對標準差
2字	0.041333333	0.040466667	0.0378333333	正唯守相利徐华定
	P 1955 N. S.			
3字	0.012366667	0.017933333	0.0103	. 0
1字	0.01005	0.0159	0.00875	0
5字	0.0561	0.0458	0.0475	0
5字	0.0156	0.0104	0.0085	. 0
7字	0.03765	0.0362	0.0354	0
	加密時間相對標準差	解密時間相對標準	總花費時間相對標準差	正確率相對標準差
2字	0.014	0.020433333	0.0127	0
3字	0.0127	0.014533333	0.009433333	0
4字	0.01755	0.01955	0.01765	, 0
5字	0.01735	0.0198	0.0122	0
	0.0138			
	0.0114	0.0105	0.009	0
6字 7字	0.0515	0.0587	0.0514	0

▲測驗結果統計數據(圖十一)

統計數據包含了加密、解密以及總時間在不同字數及字詞類型下的平均、標準差以及相 對標準差。

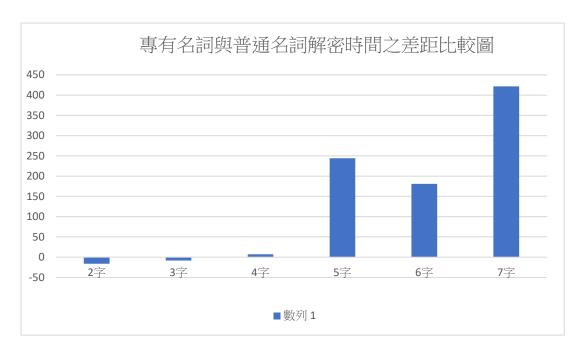


▲普通名詞以及專有名詞的加密時間比較(圖十二)

從(圖十二)可以觀察出加密的所需時間與加密字數成正相關,但普通名詞與專有名詞間的差距時高時低,看不出一定的相對關係。

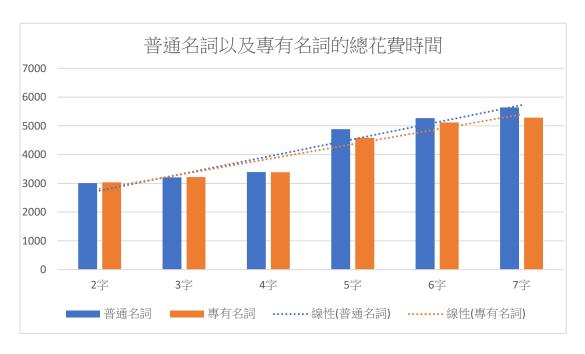


▲普通名詞以及專有名詞的解密時間比較(圖十三)



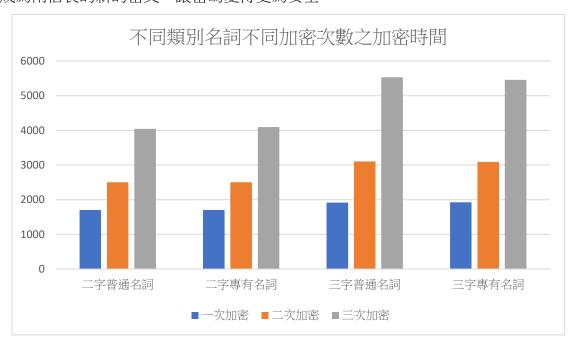
▲專有名詞與普通名詞解密時間之差距比較圖(圖十四)

從(圖十三)可以觀察出解密的時間在四字以下相差不大,但是一旦大於五字就迅速增高,原因是因為四字以下時所需判斷的詞數多為一個而一旦多於四字就要判斷兩個或更多詞,所以會花更多時間。而由(圖十四)可知解密時普通名詞與專有名詞的時間差距也是在四字以下時差距不大,但在五字以上時差距變得明顯,原因是因為專有名詞在由字音轉換為中文字時能以用更少次迴圈完成且自數越多節省的時間越多。

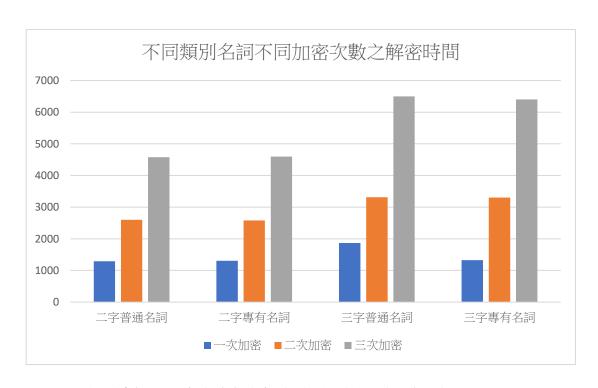


▲普通名詞以及專有名詞的總花費時間比較(圖十五)

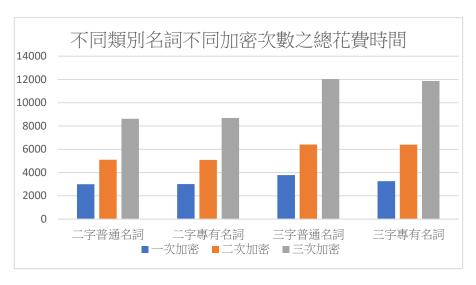
多次加密是另一個我們有嘗試的東西,原理很簡單,就是將加密後的密文在加密一次, 使他成為兩倍長的新的密文,讓密碼變得更為安全。



▲不同類別名詞不同加密次數之加密時間比較圖(圖十六)



▲不同類別名詞不同家密次數之解密時間比較圖(圖十七)



▲不同類別名詞不同加密次數之解密時間比較圖(圖十八)

從(圖十六)至(圖十八)可以觀察出無論是普通名詞還是專有名詞,不同加密次數所花的時間的差距大約成二次正相關,原因是因為加密一次是由一字轉為兩字,所以每加密一次,下一次加密的字數便會變為兩倍,也就需要花將近多兩倍的時間。

多次加密在有更安全的優點下,同時也有花費時間過多的影響,除了上面圖表中的二次 及三次以外,我們也有做更多次的測試,舉例來說加密僅僅兩個字五次就要花 30 秒,若要加 密七次甚至要花 2 分鐘,所以它的應用可能會受限。

陸、討論

經過實驗後我們發現我們已能確保字詞轉換的穩定性-也就是說每次轉換的結果能確保 一致-唯一現在的小缺陷便是受限於資料庫的豐富度和完整度,在轉換時有時會出現特定語 詞轉換錯誤,但這只要增加資料庫的語詞數便能得到充分的解決。

至於轉換時間偏長的問題我們認為這依然是可以解決的,由於時間因素,這次我們的程式幾乎都是使用.Net 內建的搜尋函數和資料結構,因此不能依據自己的需求調整。未來我們將按照我們的需求設計搜尋函數以及資料結構,再加上分散式運算,我們相信時間將能進一步壓縮。

關於成果我們發現經過大量測試之後,字數和所需時間呈現接近正比的關係,由此可見我們的規則是十分穩定的,不會因字數而有過度膨脹的時間複雜度。不過我們測試過許多不同的類別的語詞發現有些類別的詞受限於資料庫,正確率不高,例如詩詞類的語詞,同時我們發現越文言的詩詞越容易出錯,歸根究柢是資料庫沒有相關的詞的原因,我們也不可能無限擴充資料庫,於是我們希望之後能利用自然語言處理的方式結合機器學習,能讓斷詞更加省時一我們現在是利用類似「窮舉」的方式斷詞,自然耗時又易錯一並且讓我們不必因為語言習慣的改變而時時調整資料庫-程式將幫我們完成這一點。

延伸應用的部份除了前述的機器學習和自然語言之外,我們希望最終能讓它能加密一篇文章,當然,在這之前必須先解決效率和正確率的問題,同時我們也希望能融合雲端的技術,讓它更有實用性。另外,我們也能針對「語言」這個部份做延伸,我們目前比較成熟的想法是希望能發展出不同的方言的加密方式,例如廣東話之類的,當然目前尚處於構想階段,但我們相信既然能解決「普通話」的加密,廣東話也不是難事,只要掌握規則便行。

柒、結論

- 一、資料庫的設計將影響轉換的正確率
- 二、在資料庫被駭入時,可透過更改資料庫避免資料被竊取
- 三、轉換的字數與所需時間有一定的關係
- 四、字數與正確率沒有一定的關聯性
- 五、語詞類型會影響轉換速率
- 六、確認了反切音在密碼應用上的可能性

捌、參考資料及其他

- 一、參考資料
 - (一)字字珠璣中研院搜詞尋字造詞搜尋

http://words.sinica.edu.tw/sou/

(二)中研院現代漢語語料庫詞頻統計

http://elearning.ling.sinica.edu.tw/CWordfreq.html

(三)國科會數位博物館先導計畫-- 搜文解字

http://words.sinica.edu.tw/

(四)新酷音輸入法

http://chewing.im/

(五)臺灣博碩士論文知識加值系統:自由的博碩士學位論文全文資料庫-陳雅萍-《字彙》反切音系研究

https://ndltd.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi/login?o=dwebmge#XX

(六)反切 - 維基百科,自由的百科全書

https://zh.wikipedia.org/wiki/%E5%8F%8D%E5%88%87

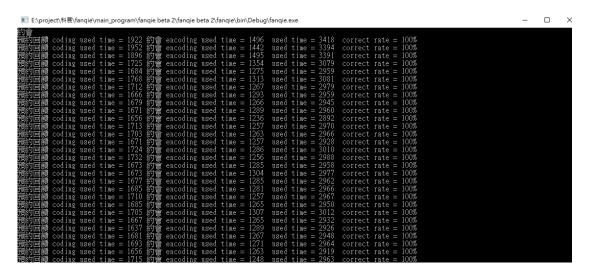
(七)密碼學-維基百科,自由的百科全書

https://zh.wikipedia.org/wiki/%E5%AF%86%E7%A0%81%E5%AD%A6

二、實驗數據

(一)普通名詞

二字



average coding time: 1704, average encoding time: 1290, average total time: 2995, average correct rate: 100%

SD of coding time: 102.32, SD of encoding time: 65.02, SD of total time: 154.87, SD of correct rate: 0.00

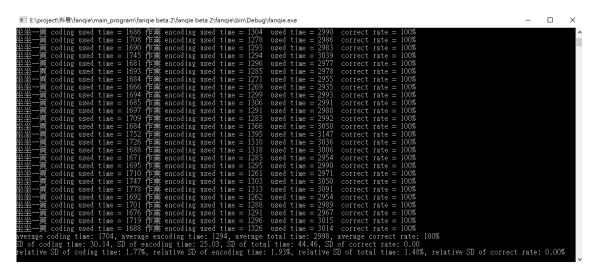
relative SD of coding time: 6.00%, relative SD of encoding time: 5.04%, relative SD of total time: 5.17%, relative SD of correct rate: 0.00%



average coding time: 1727, average encoding time: 1312, average total time: 3039, average correct rate: 100%

SD of coding time: 79.90, SD of encoding time: 67.88, SD of total time: 142.87, SD of correct rate: 0.00

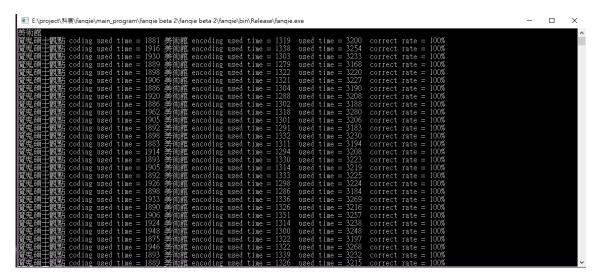
relative SD of coding time: 4.63%, relative SD of encoding time: 5.17%, relative SD of total time: 4.70%, relative SD of correct rate: 0.00%



average coding time: 1704, average encoding time: 1294, average total time: 2998, average correct rate: 100%

SD of coding time: 30.14, SD of encoding time: 25.03, SD of total time: 44.46, SD of correct rate: 0.00

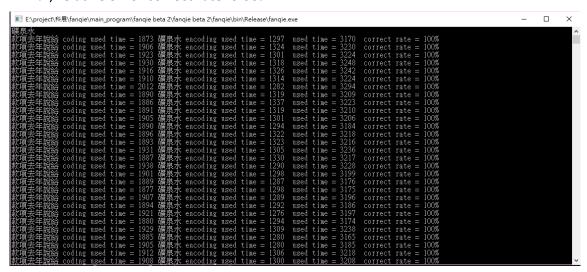
relative SD of coding time: 1.77%, relative SD of encoding time: 1.93%, relative SD of total time: 1.48%, relative SD of correct rate: 0.00%



average coding time: 1902, average encoding time: 1312, average total time: 3215, average correct rate: 100%

SD of coding time: 24.45, SD of encoding time: 23.73, SD of total time: 35.74, SD of correct rate: 0.00

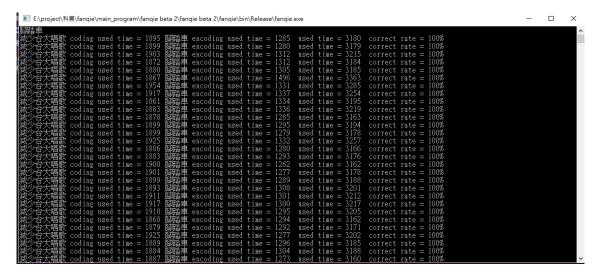
relative SD of coding time: 1.29%, relative SD of encoding time: 1.81%, relative SD of total time: 1.11%, relative SD of correct rate: 0.00%



average coding time: 1905, average encoding time: 1305, average total time: 3211, average correct rate: 100%

SD of coding time: 29.31, SD of encoding time: 18.39, SD of total time: 32.86, SD of correct rate: 0.00

relative SD of coding time: 1.54%, relative SD of encoding time: 1.41%, relative SD of total time: 1.02%, relative SD of correct rate: 0.00%

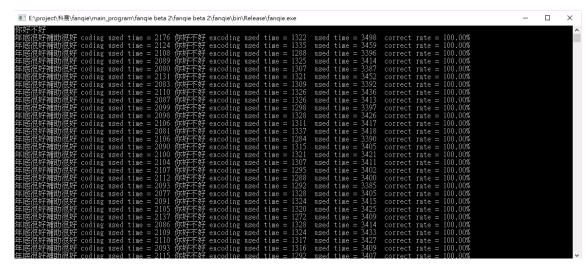


average coding time: 1894, average encoding time: 1302, average total time: 3196, average correct rate: 100%

SD of coding time: 16.71, SD of encoding time: 28.16, SD of total time: 30.72, SD of correct rate: 0.00

relative SD of coding time: 0.88%, relative SD of encoding time: 2.16%, relative SD of total time: 0.96%, relative SD of correct rate: 0.00%

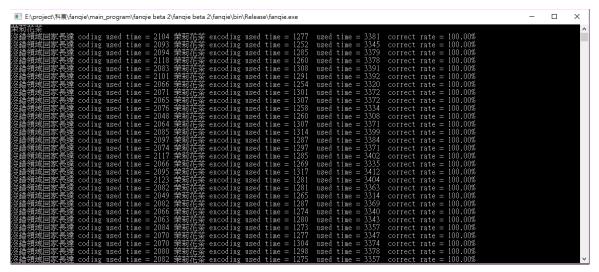
3、四字



average coding time: 2100, average encoding time: 1312, average total time: 3413, average correct rate: 100%

SD of coding time: 19.73, SD of encoding time: 21.02, SD of total time: 29.72, SD of correct rate: 0.00

relative SD of coding time: 0.94%, relative SD of encoding time: 1.60%, relative SD of total time: 0.87%, relative SD of correct rate: 0.00%

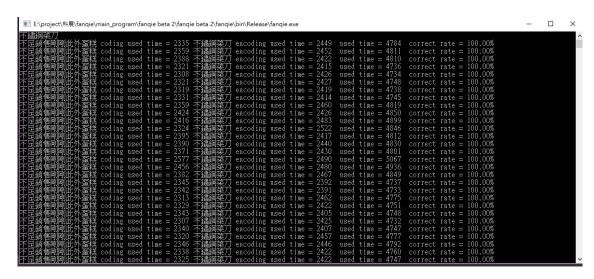


average coding time: 2083, average encoding time: 1291, average total time: 3375, average correct rate: 100%

SD of coding time: 22.39, SD of encoding time: 20.44, SD of total time: 29.86, SD of correct rate: 0.00

relative SD of coding time: 1.07%, relative SD of encoding time: 1.58%, relative SD of total time: 0.88%, relative SD of correct rate: 0.00%

4、五字



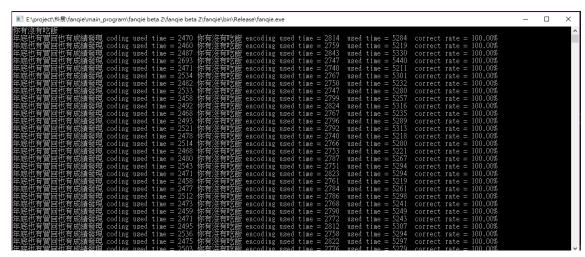
average coding time: 2399, average encoding time: 2483, average total time: 4883, average correct rate: 100%

SD of coding time: 134.59, SD of encoding time: 113.66, SD of total time: 232.00, SD of correct rate: 0.00

relative SD of coding time: 5.61%, relative SD of encoding time: 4.58%, relative SD of total time:

4.75%, relative SD of correct rate: 0.00%

5、六字

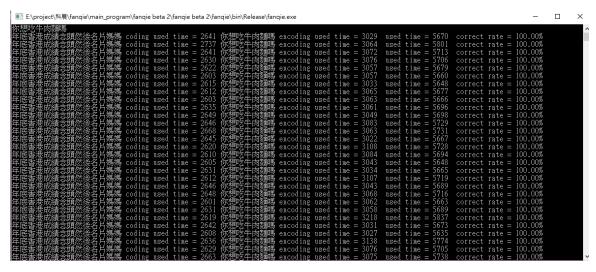


average coding time: 2489, average encoding time: 2779, average total time: 5269, average correct rate: 100%

SD of coding time: 38.67, SD of encoding time: 29.02, SD of total time: 44.62, SD of correct rate: 0.00

relative SD of coding time: 1.55%, relative SD of encoding time: 1.04%, relative SD of total time: 0.85%, relative SD of correct rate: 0.00%

6、七字



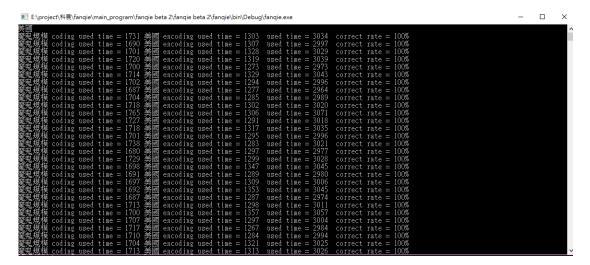
average coding time: 2756, average encoding time: 3199, average total time: 5956, average correct rate: 100%

SD of coding time: 133.23, SD of encoding time: 143.53, SD of total time: 271.12, SD of correct rate: 0.00

relative SD of coding time: 4.83%, relative SD of encoding time: 4.49%, relative SD of total time: 4.55%, relative SD of correct rate: 0.00%

(二)專有名詞

1、二字



average coding time: 1706, average encoding time: 1304, average total time: 3011, average correct rate: 100%

SD of coding time: 18.72, SD of encoding time: 23.60, SD of total time: 29.98, SD of correct rate: 0.00

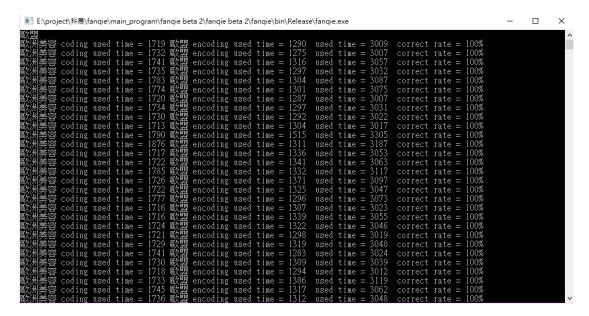
relative SD of coding time: 1.10%, relative SD of encoding time: 1.81%, relative SD of total time: 1.00%, relative SD of correct rate: 0.00%



average coding time: 1723, average encoding time: 1330, average total time: 3054, average correct rate: 100%

SD of coding time: 25.73, SD of encoding time: 24.56, SD of total time: 39.05, SD of correct rate: 0.00

relative SD of coding time: 1.49%, relative SD of encoding time: 1.85%, relative SD of total time: 1.28%, relative SD of correct rate: 0.00%

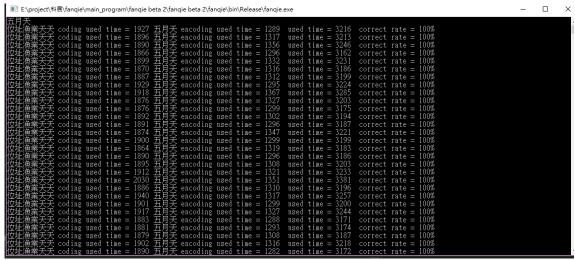


average coding time: 1740, average encoding time: 1311, average total time: 3051, average correct rate: 100%

SD of coding time: 27.93, SD of encoding time: 32.41, SD of total time: 46.74, SD of correct rate: 0.00

relative SD of coding time: 1.61%, relative SD of encoding time: 2.47%, relative SD of total time: 1.53%, relative SD of correct rate: 0.00%

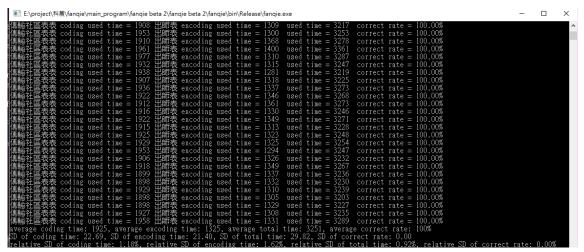




average coding time: 1898, average encoding time: 1309, average total time: 3207, average correct rate: 100%

SD of coding time: 25.33, SD of encoding time: 20.21, SD of total time: 32.59, SD of correct rate: 0.00

relative SD of coding time: 1.33%, relative SD of encoding time: 1.54%, relative SD of total time: 1.02%, relative SD of correct rate: 0.00%



average coding time: 1925, average encoding time: 1325, average total time: 3251, average correct rate: 100%

SD of coding time: 22.69, SD of encoding time: 21.40, SD of total time: 29.82, SD of correct rate: 0.00

relative SD of coding time: 1.18%, relative SD of encoding time: 1.62%, relative SD of total time: 0.92%, relative SD of correct rate: 0.00%

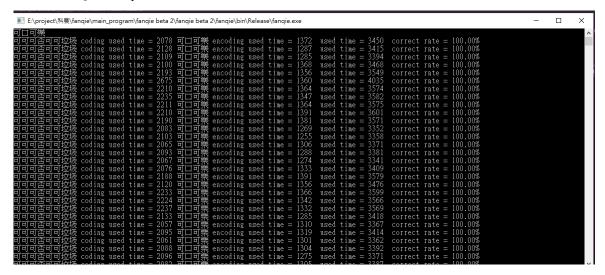


average coding time: 1897, average encoding time: 1310, average total time: 3208, average correct rate: 100%

SD of coding time: 24.70, SD of encoding time: 15.72, SD of total time: 28.65, SD of correct rate: 0.00

relative SD of coding time: 1.30%, relative SD of encoding time: 1.20%, relative SD of total time: 0.89%, relative SD of correct rate: 0.00%

3、四字



average coding time: 2104, average encoding time: 1300, average total time: 3404, average correct rate: 100%

SD of coding time: 72.39, SD of encoding time: 32.24, SD of total time: 92.09, SD of correct rate: 0.00

relative SD of coding time: 3.44%, relative SD of encoding time: 2.48%, relative SD of total time: 2.71%, relative SD of correct rate: 0.00%

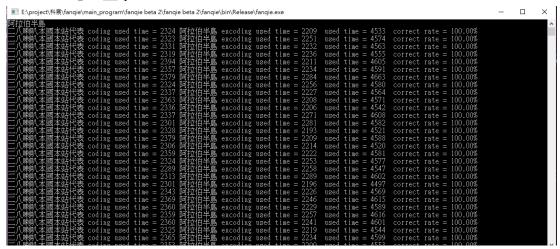


average coding time: 2082, average encoding time: 1289, average total time: 3372, average correct rate: 100%

SD of coding time: 22.18, SD of encoding time: 18.47, SD of total time: 27.61, SD of correct rate: 0.00

relative SD of coding time: 1.07%, relative SD of encoding time: 1.43%, relative SD of total time: 0.82%, relative SD of correct rate: 0.00%

4、五字

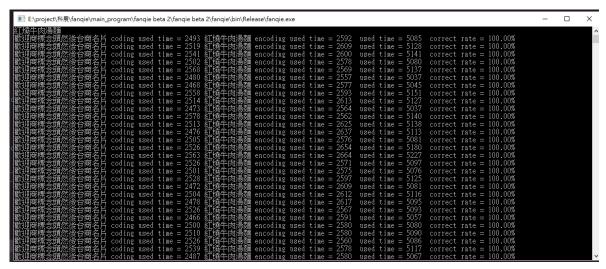


average coding time: 2342, average encoding time: 2239, average total time: 4582, average correct rate: 100%

SD of coding time: 32.42, SD of encoding time: 44.33, SD of total time: 56.01, SD of correct rate: 0.00

relative SD of coding time: 1.38%, relative SD of encoding time: 1.98%, relative SD of total time: 1.22%, relative SD of correct rate: 0.00%

5、六字

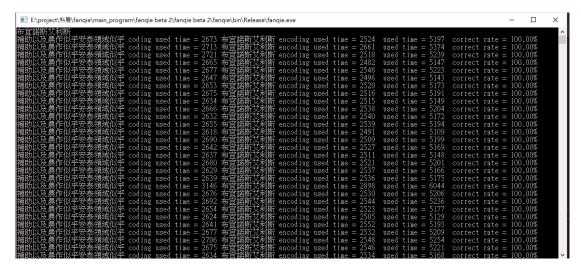


average coding time: 2518, average encoding time: 2598, average total time: 5117, average correct rate: 100%

SD of coding time: 35.52, SD of encoding time: 27.25, SD of total time: 46.14, SD of correct rate: 0.00

relative SD of coding time: 1.41%, relative SD of encoding time: 1.05%, relative SD of total time: 0.90%, relative SD of correct rate: 0.00%

6、七字



average coding time: 2703, average encoding time: 2579, average total time: 5283, average correct rate: 100%

SD of coding time: 139.14, SD of encoding time: 151.39, SD of total time: 271.41, SD of correct rate: 0.00

relative SD of coding time: 5.15%, relative SD of encoding time: 5.87%, relative SD of total time:

5.14%, relative SD of correct rate: 0.00%

【評語】052510

- 本研究將傳統的加密方式結合語言學,利用「反切」這種古 代拼音方式來設計一套加密規則,使得加密變得更有變化, 研究題目有其創新性。
- 利用注音進行加解密,饒富趣味,但加密與解密的唯一性有 待進一步探討。另加解密資料庫是否可為公開,或如有多個 資料庫,如何有效率的管理也須加以探討。
- 3. 建議針對加解密演算法時間複雜度進行分析,據以改善加解 密效能的問題。
- 4. 建議可對研究所提出方法與常見的加密方法做一比較,透過實驗驗證不同方法的優劣,以符合科學探究精神。

摘要

我們的研究是利用「反切」這種古代的拼音方式來設計一套加密規則,並利用它替特定的字串 加解密。

我們並沒有試著挑戰一般的加解密,而是在原本的基礎上對內文動手腳,也就是說最終可能依然得依賴一般的加密方式,但在此之前我們已將內文經過加密成為密文,使得即使破解了一般的加密,結果依然是一串無法理解的密文,像是第二道保險一般,使文字在原有的加密基礎上再多一層保障。

壹、研究動機

在現今的社會中,資訊在我們生活中占了很大一部分,因此資訊安全的重要性在這個情況之下也變得更為顯著。

在發現這個問題之後,我們便想著有沒有方法能夠在密碼學的方面上做出一些新的改變,而在尋找方法的過程中,我們試著向古人借鏡,便發現了反切音這種形式。

反切是一種音韻學的發展分支之一,原是一種注音方式,分為反切上字及反切下字,發展 到後期有人將反切音作為傳遞密文的方式

歷史上最早將反切應用在密碼上的是明朝名將戚繼光將軍,他將此應用在軍事通訊上。

他是將中文密文藏在詩句中。戚繼光專門編了兩首詩歌,作為「密碼本」:一首是:「柳邊求氣低,波他爭日時。鶯蒙語出喜,打掌與君知」;另一首是:「春花香,秋山開,嘉賓歡歌須金杯,孤燈光輝燒銀缸。之東郊,過西橋,雞聲催初天,奇梅歪遮溝。」

兩首詩歌是反切碼全部秘密所在。取前一首中的前20個字的聲母,依次分別編號1-20;取後一首36字韻母,順序編號1-36。再將當時福州方言字音的八種聲調,也按順序編上號碼1-8,形成完整的「反切碼」體系。其使用方法是:如送回的情報上的密碼有一串是5-25-2,對照聲母編號5是「低」字,韻母歌編號25是「西」字,兩字的聲母和韻母合到一起了是di,對照聲調是2,就可以切射出「敵」字。

貳、研究目的

- 一、找出一套加解密規則,能適用大多數的字音
- 二、能較完整地利用反切的原理進行中文的加解密
- 三、希望能從語言學的角度達成加密的效果 證實古人的智慧也能應用在先進的 科技上。

肆、研究過程與方法

一、流程圖

(一)確認主題查詢資料

- 研究密碼特性
- 研究反切特性

(二)尋找資 料來源並製作 資料庫

- 採用新酷音輸入法
- 將其依需求製作成資料庫

(三)訂定加密規則

- 設定注音分割規則
- 訂定轉換規則

(四)編寫程式並除錯、修改

- 分別完成加密和解密部分
- •除錯測試發現缺陷並修改

(五)測試不同數據並統整

- 整合加、解密部分
- 測試多筆資料並統整數據

多、研究設備與器材

- 一、開發環境:
- (一) Microsoft Visual Studio Community 2017
 - (_) Microsoft Office Access 2016
 - (\equiv) Microsoft .Net Framework 4.5.2
- 二、電腦設備:
- () CPU: Intel Core i7 7700 3.60GHz

二、研究過程

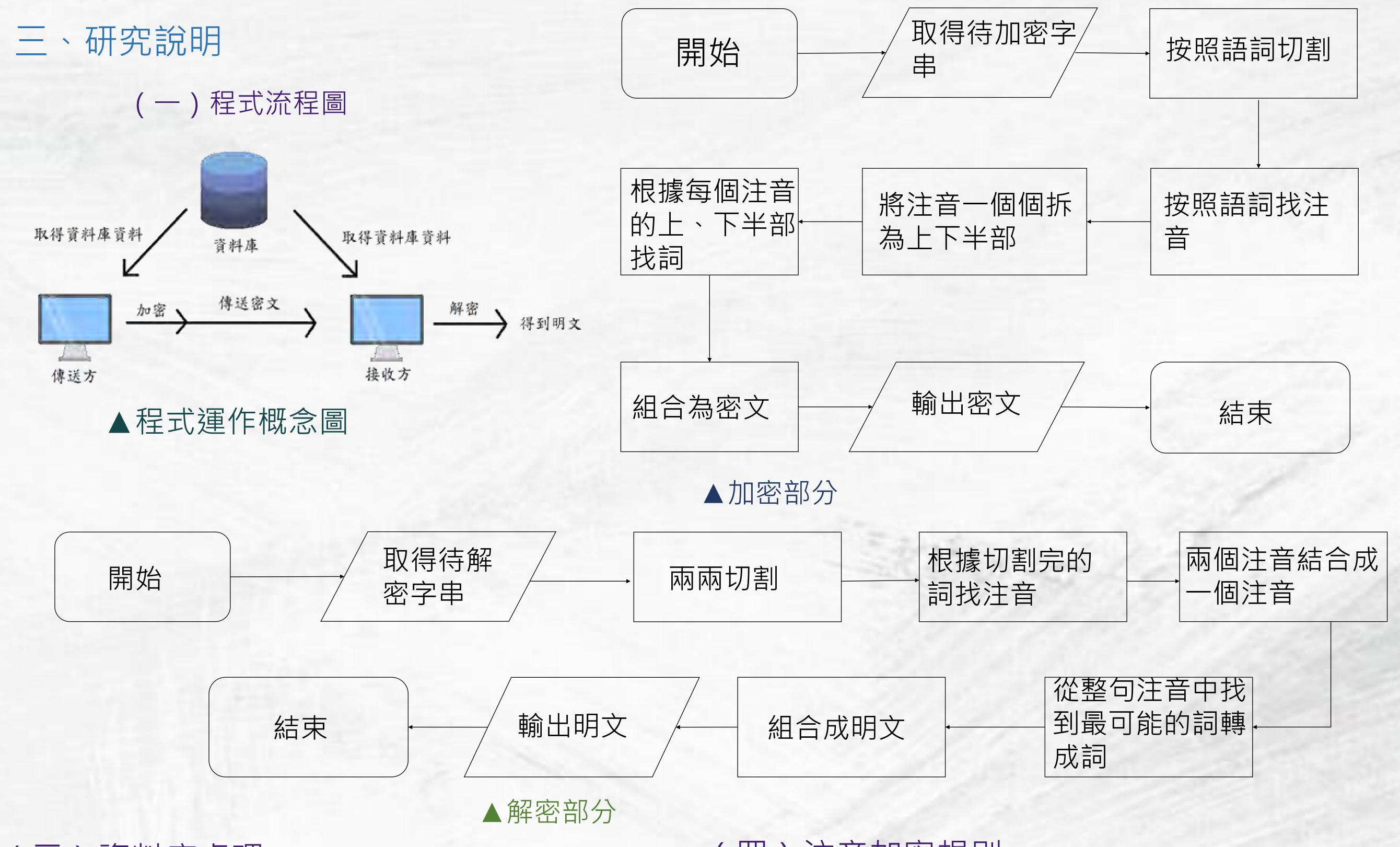
一開始決定以反切音作為研究對象後,首先 先查詢反切音相關的資料,了解它的規則以及特 性,以掌握研究的方向。

而在這個過程中發現了戚繼光將軍的「反切碼」這個應用方法,反切碼的原理使我們聯想到當今「資料庫」的概念,而有了利用資料庫相互比對的方法加密中文的構想,並開始實際嘗試。

在尋找資料庫的過程中,我們以開源的語詞庫為目標為原則尋找,以利研究的方便,最後便找到了新酷音輸入法的語詞庫作為我們這次研究的資料庫。

有了資料庫後,便是要訂定轉換規則了,一開始我們先從簡單的方法開始嘗試,也就是「逐字隨機加密」,但在做了幾筆測試之後,發現總會有幾個字轉換出來的的字不一樣,在調查那幾個字之後,才發現是中文特性——字多音造成的問題,所以我們就繼續再想新的加密規則。

而在發現逐字隨機加密的問題之前我們就有 想著是否能讓加密出來的文字具有一點文學的意 義,而能夠讓人看不出那串文字是密文,就像是 戚繼光將軍所用的詩句一般,能夠混淆試圖破解 的人。所以我們在想新的方法時就往這個方向去 思考,最後得出了「以詞加密」這個加密規則。 這個加密規則雖然不能使整串密文具有連貫的意 義,但每兩個字都是具有意義的詞而能夠達到類 似的效果。

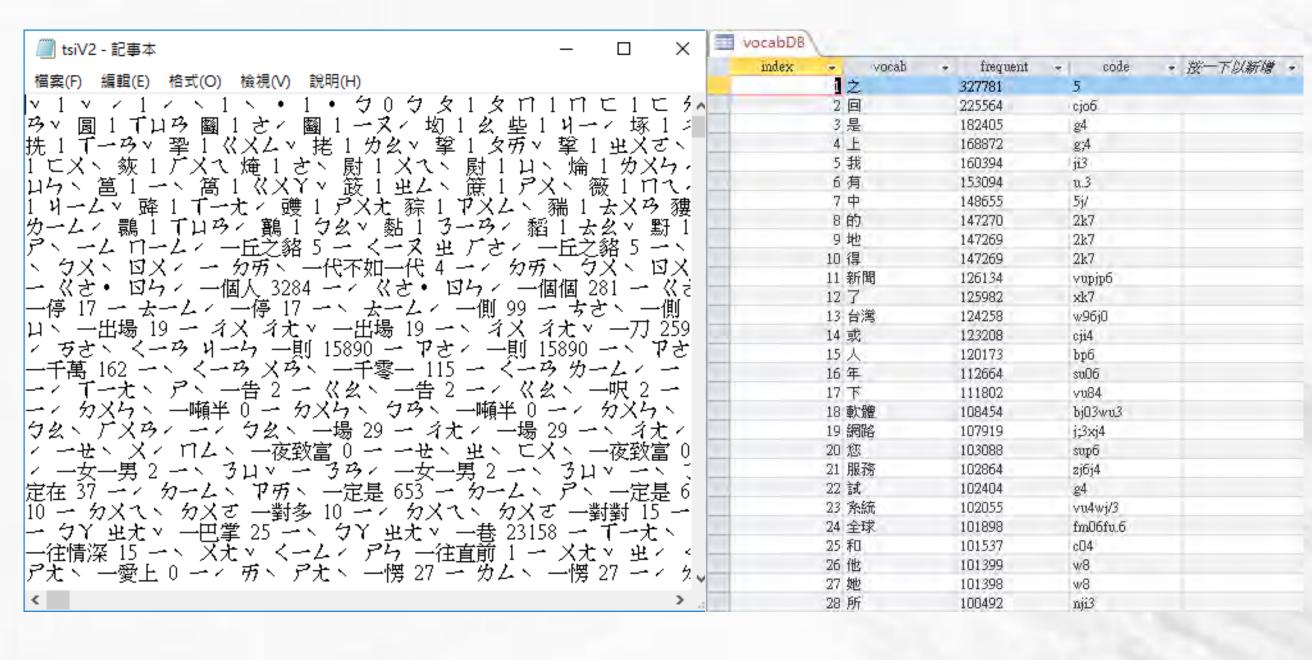


(二)資料庫處理

在找資料庫時,我們希望它是開源的,以便做任何的研究而不會受太多限制,所以找到了新酷音輸入法的字詞庫。

字詞庫的內容有字、詞以及他們的使用頻率及讀音,在判斷讀音以及加密實都是不可或缺的資料。

但是它的字詞庫是TXT檔,不利於資料搜尋的效率,且注音符號在程式中處理相對英文字不便,所以我們將這個TXT檔做簡單的轉換後匯入至Access中成為一個新資料庫。



▲新酷音資料庫內容

▲轉換後資料庫

(三)判斷語句字音

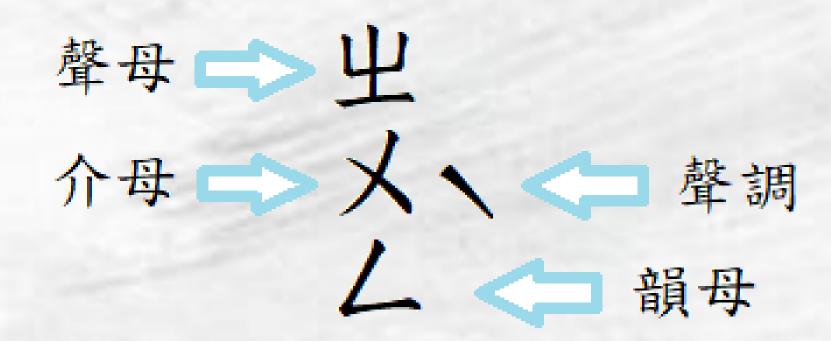
我們判斷字音的方法是以詞切割語句,而實際的方法如下圖

以「千里共嬋娟」為例:

搜尋「千」有結果擴大範圍搜尋「千里」

如此循環數次後便可得到「千里共嬋娟」的正確讀音。

(四)注音加密規則:



如上圖所示,我們將一個字的注音分成上半部和下半部。上半部由聲母和介母組成,而下半部由韻母與聲調組成,以便加密時做使用。

而加密的規則我們想了兩種:

1、逐字隨機加密

假設某字經過字音判斷後得知讀音為「坐 メム`」,先將它分成上下兩部分,分別是 「坐メ」以及「ム`」,接著以隨機的方式 將「坐メ」的部份加上下半部並將「ム`」 部份加上上半部後最後各自轉回中文字便形 成密文。

優點:因為處理方法單純所以速度相對快速。 缺點:因破音字問題導致錯誤率相對高。

2、以詞加密

出、今里 2 中用

假設某字經過字音判斷後得知讀音為「虫メム `」,先將它分成上下兩部分,分別是「虫メ」以 及「厶`」,接著在資料庫中以SQL查詢首字上 半部同為「虫メ」且尾字下半部同為「ㄥ`」的兩 字的詞,便為此字轉換為反切音後的密文。

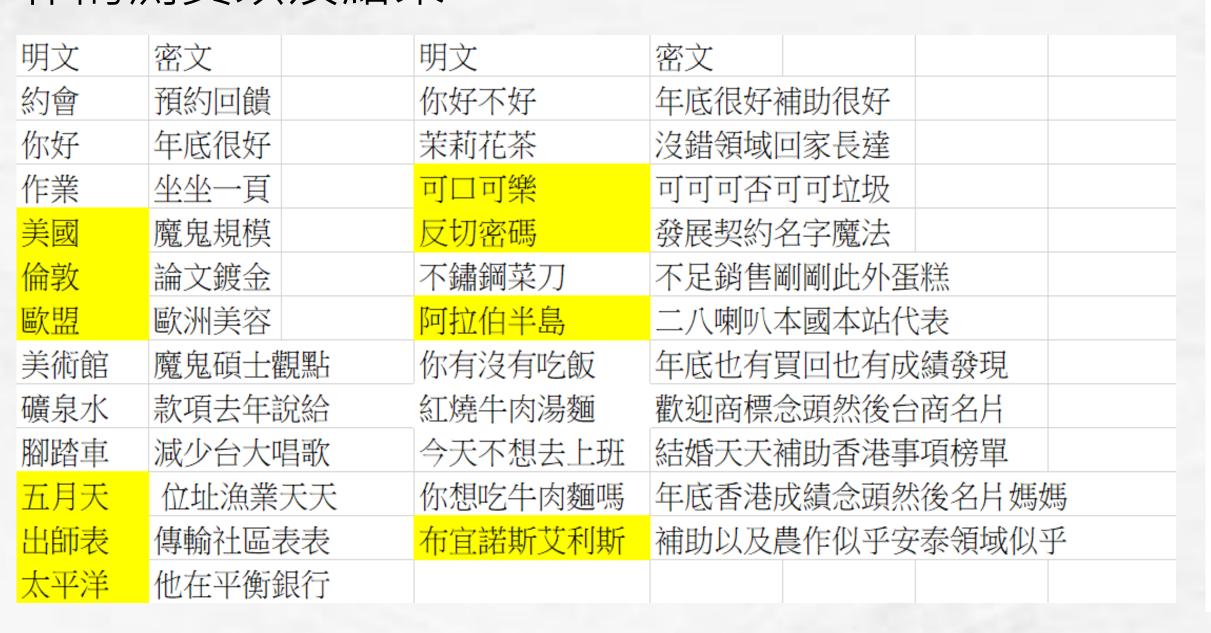
優點:詞的同音的機率低所以在字音方面能夠達到接近100%正確率,且密文的內容有文字上的意義可

以混淆試圖破解的人而達到更好的加密效果。

缺點:處理複雜導致速度較慢。

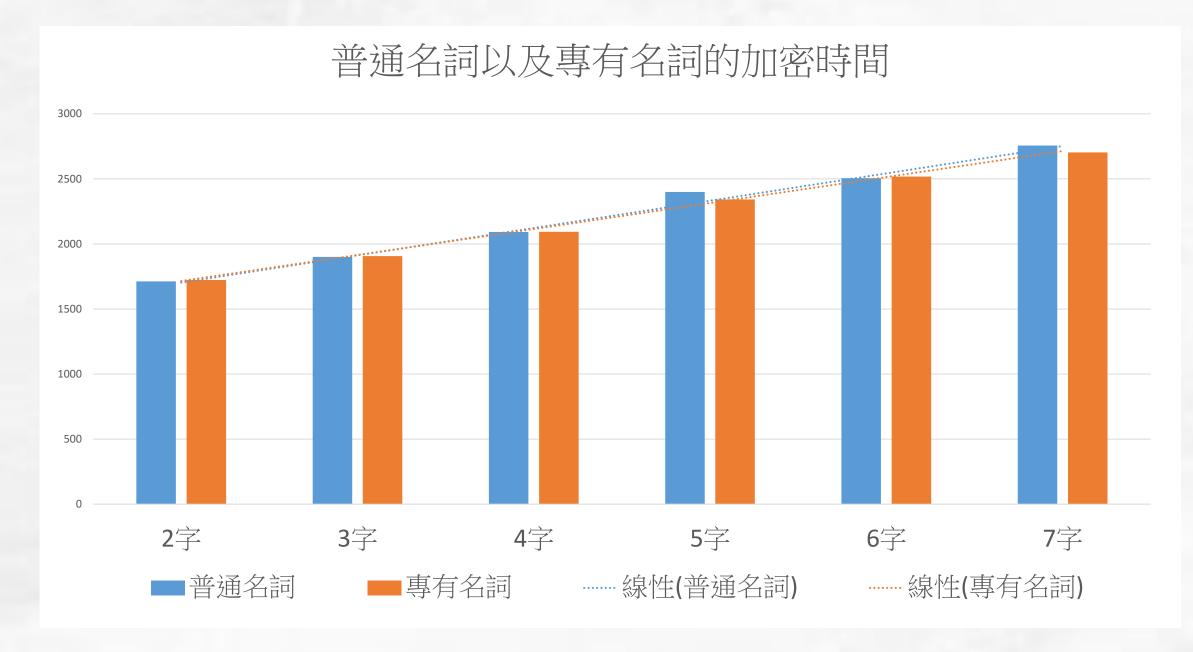
伍、研究結果

我們在做測試時,除了針對不同字數作了測 試,也對不同類型的字詞作了測試,以下是我們 作的測資以及結果。



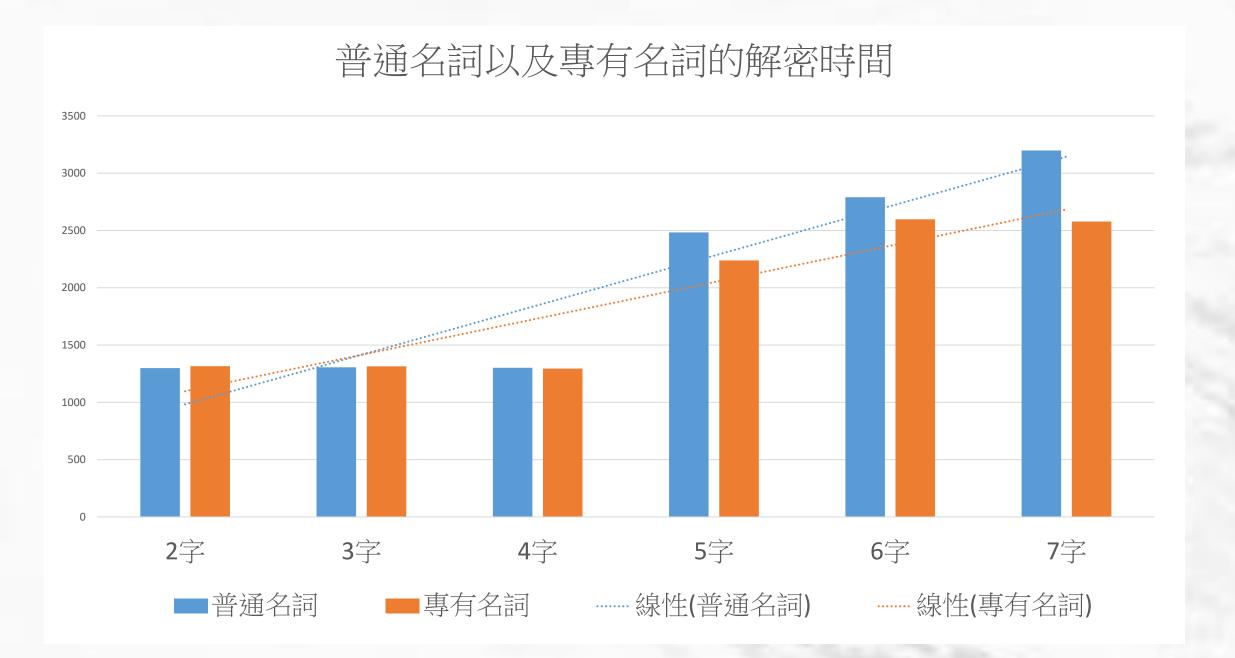
▲測試的測資及結果

我們測資中的語詞內容分為兩類,普通名詞以及 專有名詞,上圖中以黃色註記的測資即為專有名詞

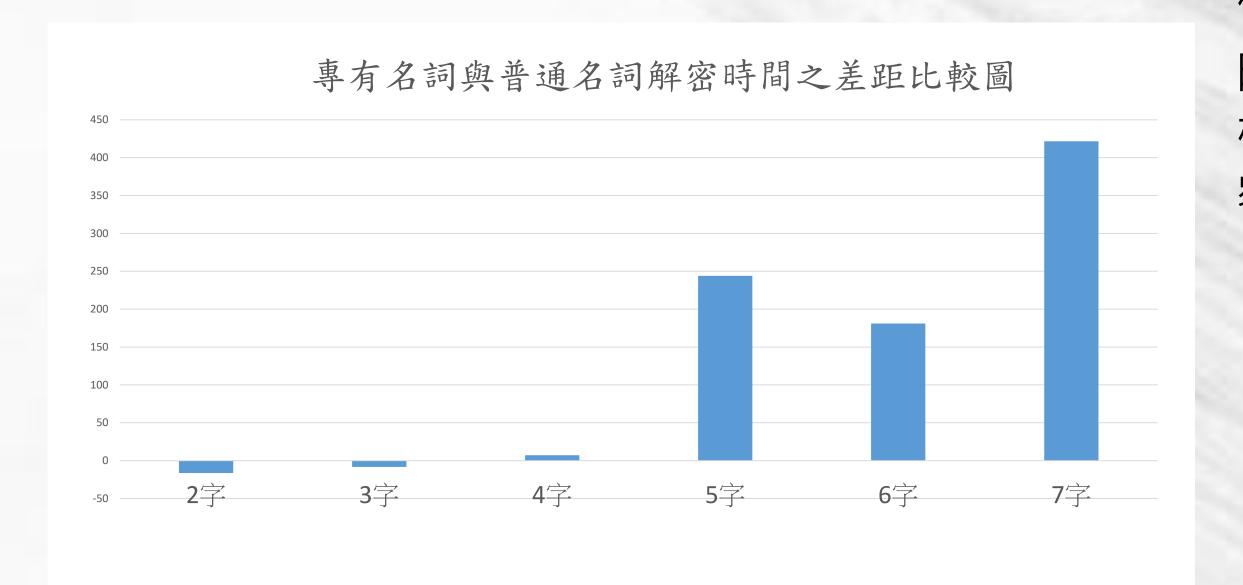


▲普通名詞以及專有名詞的加密時間比較(圖一)

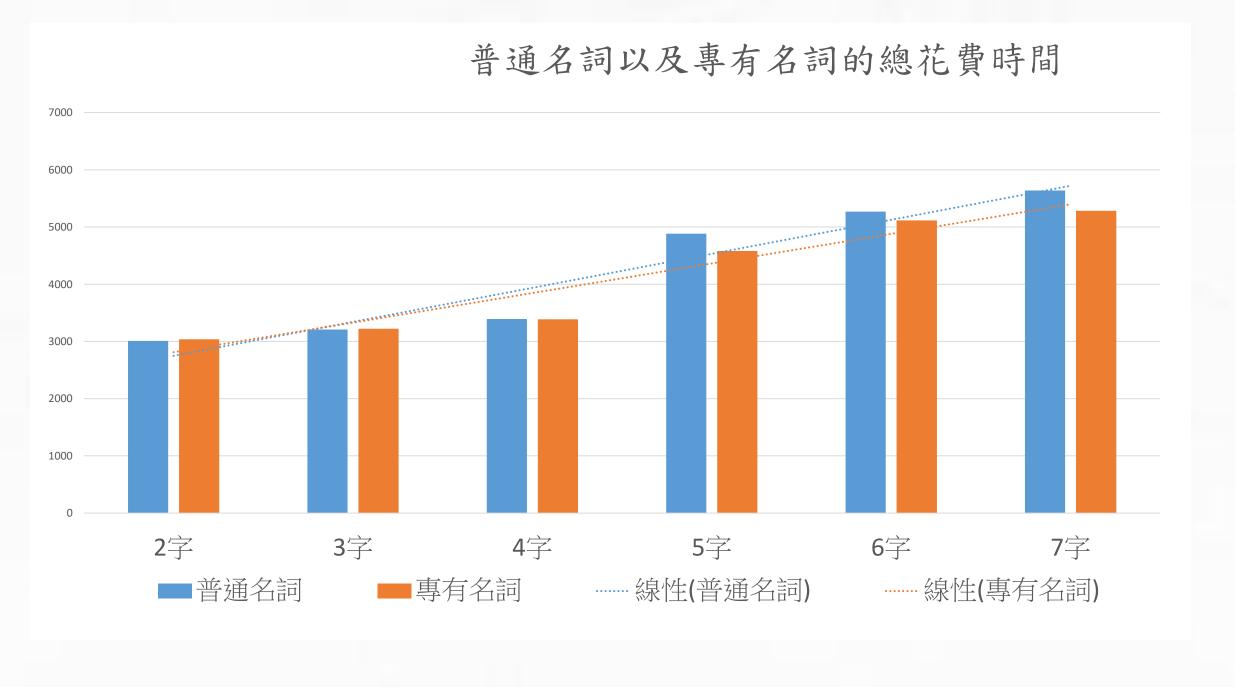
從(圖一)可以觀察出加密的所需時間與加密字 數成正相關,但普通名詞與專有名詞間的差距時高 時低,看不出一定的相對關係。



▲普通名詞以及專有名詞的解密時間比較(圖二)

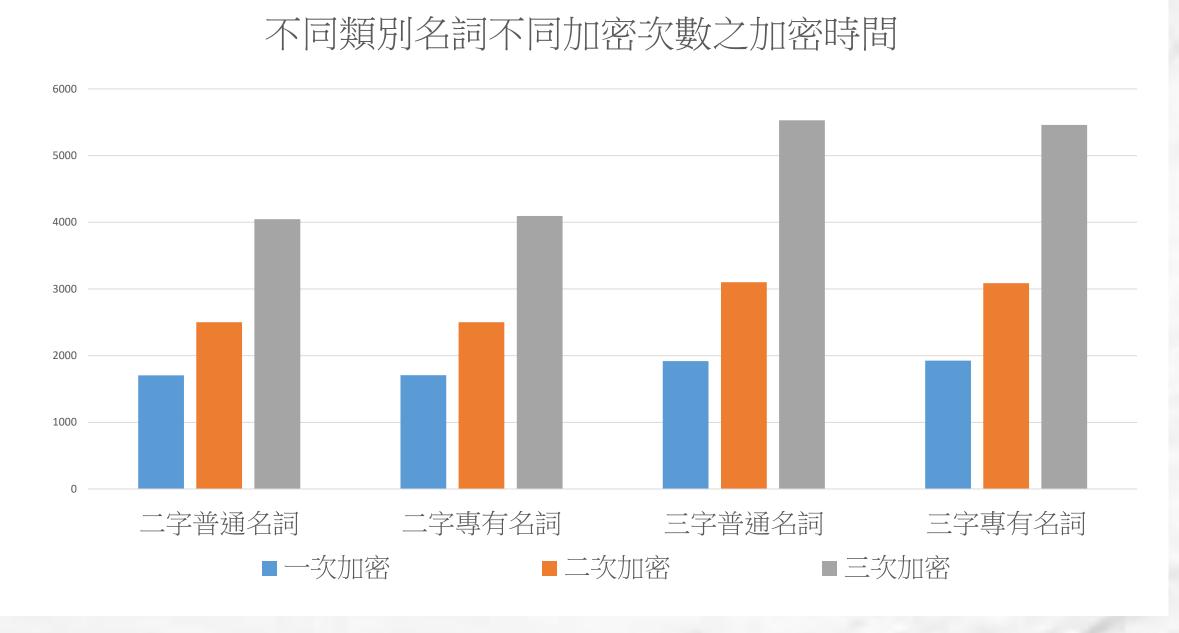


▲專有名詞與普通名詞解密時間之差距比較圖 (圖三)



▲普通名詞以及專有名詞的總花費時間比較(圖四)

多次加密是另一個我們有嘗試的東西,原理很 簡單,就是將加密後的密文在加密一次,使他成為 兩倍長的新的密文,讓密碼變得更為安全。



多次加密在有更安全的優點下,同時也有花費 時間過多的影響,除了上面圖表中的二次及三次以 我們也有做更多次的測試,舉例來說加密僅僅 兩個字五次就要花30秒,若要加密七次甚至要花2 分鐘,所以它的應用可能會受限。

陸、討論

至於轉換時間偏長的問題我們認為這依然是可 以解決的,由於時間因素,這次我們的程式幾乎都 是使用.Net內建的搜尋函數和資料結構,因此不能 依據自己的需求調整。未來我們將按照我們的需求 設計搜尋函數以及資料結構,再加上分散式運算, 我們相信時間將能進一步壓縮。

關於成果我們發現經過大量測試之後,字數和 所需時間呈現接近正比的關係,不會因字數而有過 度膨脹的時間複雜度。不過我們測試過許多不同的 類別的語詞發現有些類別的詞受限於資料庫,正確 率不高,例如詩詞類的語詞,同時我們發現越文言 的詩詞越容易出錯,歸根究柢是資料庫沒有相關的 **詞的原因,我們也不可能無限擴充資料庫,於是我** 們希望之後能利用自然語言處理的方式結合機器學 習,能讓斷詞更加省時一我們現在是利用類似「窮 舉」的方式斷詞,自然耗時又易錯一並且讓我們不 必因為語言習慣的改變而時時調整資料庫 - 程式將 幫我們完成這一點。

延伸應用的部份除了前述的機器學習和自然語 言之外,我們希望最終能讓它能加密一篇文章,當 然,在這之前必須先解決效率和正確率的問題 時我們也希望能融合雲端的技術,讓它更有實用性。 另外,我們也能針對「語言」這個部份做延伸 們目前比較成熟的想法是希望能發展出不同的方言 的加密方式,例如廣東話之類的,當然目前尚處於 構想階段,但我們相信既然能解決「普通話」的加 密,廣東話也不是難事,只要掌握規則便行。

- 一、資料庫的設計將影響轉換的正確率
- 二、在資料庫被駭入時,可透過更改資料庫避免資 料被竊取
- 轉換的字數與所需時間有一定的關係
- 四、字數與正確率沒有一定的關聯性
- 、語詞類型會影響轉換速率
- 六、確認了反切音在密碼應用上的可能性

捌、参考資料及其他

- (一)字字珠璣中研院搜詞尋字造詞搜尋
- http://words.sinica.edu.tw/sou/
- (二)中研院現代漢語語料庫詞頻統計

https://tinyurl.com/cssvz2p

(三)國科會數位博物館先導計畫--搜文解字

- http://words.sinica.edu.tw/ (四)新酷音輸入法
- http://chewing.im/
- 五)陳雅萍-《字彙》反切音系研究

https://tinyurl.com/y7rpm4q4