國立台灣海洋大學資訊工程學系專題報告

不雅文字自動改寫系統

**作者**

00157035 蘇暉博 e-mail:00157035@ntou.edu.tw

00157139 黃振傑 e-mail: kentak812@yahoo.com.tw

報告編號: NTOUCSE104-015-01

指導教授：林川傑 博士

中華民國 104年 xx 月 xx 日

Project Report, NTOU CSE

Indelicate text rephrasing system

**Author**

00157035 蘇暉博 e-mail:00157035@ntou.edu.tw

00157139 黃振傑 e-mail: kentak812@yahoo.com.tw

Project Report Number: NTOUCSE104-015-01

Advisor：Dr. 林川傑

20xx / xx / xx

目錄

[一、 緒論 4](#_Toc431734358)

[1.1. 動機 4](#_Toc431734359)

[二、 定義及資料準備 5](#_Toc431734360)

[2.1. 不雅詞彙的定義 5](#_Toc431734361)

[2.2. 不雅詞彙範圍 5](#_Toc431734362)

[2.3. 資料蒐集及整理 5](#_Toc431734363)

[2.3.1. 由PTT蒐集資料 5](#_Toc431734364)

[2.3.2. Twitter 9](#_Toc431734365)

[三、 不雅詞彙規則改寫 10](#_Toc431734366)

[3.1. 取代規則 10](#_Toc431734367)

[3.2. 取代結果 17](#_Toc431734368)

[四、 實驗 17](#_Toc431734369)

[4.1. 評估 17](#_Toc431734370)

[4.2. 討論 19](#_Toc431734371)

[五、 線上系統 20](#_Toc431734372)

[5.1. 網頁版不雅文字自動改寫系統 20](#_Toc431734373)

[5.2. FB應用程式版不雅文字自動改寫系統 21](#_Toc431734374)

[六、 結論 22](#_Toc431734375)

# 緒論

## 動機

有鑑於時下網路言論自由，在發表自身看法、立場或是抒發個人心境時，往往不經意地採用帶有侮辱性(insult)的詞彙，或者是其他與性(sex)有關，有間接或直接造成騷擾他人、侮辱他人疑虞的詞彙。因為文字有的妨礙觀讀，理解文意，有的甚或妨礙到他人名譽，造成公然侮辱而不自知，所以我們出現了用其他有意義文字，來取代這些文字的想法。

目前就我們所熟知的相關系統，對於不雅文字的處理方式，大抵是只針對不雅文字所進行的遮罩，例如：「你在幹麻？」被取代為「你在\*麻？」。在或者是，將不雅字直接改變成其他與所要表達並無關聯的字，例如：「幹勒！」被取代為「我不該說髒話的，對不起對不起！」

對不雅文字直接採用遮罩的方式固然簡單而有效，但亦會影響讀者在觀讀之時，對原文的辨別上有困難，因而產生困惑，或因文中所要表達的文意不明而被混淆。

故本專題的目的在於研究一套系統，可以針對我們所定義的不雅文字，以自動化、程式化的方式做有效的改寫為有意義，且不帶有侮辱性或是與性有關的不雅詞彙，以期最大限度地達到不影響文章原來所表達的語意及語氣，又能通順地表達出文章內容為最終目的。

# 定義及資料準備

## 不雅詞彙的定義

本專題欲處理的不雅詞彙定義為，帶有侮辱性(insult)的詞彙，或者是其他與性(sex)有關而會間接或直接造成騷擾他人、侮辱他人疑虞的詞彙。這些詞彙在本專題又稱為不雅文字(Indelicate text)。

## 不雅詞彙範圍

首先於2.1節處，先行定義一些不雅詞彙如下所列，做為此研究初步定義的不雅詞彙的範圍，並以之為「種子」，凡任意詞彙中有包含「種子」並符合定義的即成為我們要處理的對象，以此擴大定義範圍及涉及的不雅詞彙。

* 有「**性交**」意涵的字：「幹」、「操」的同音字等
* 有「**性提示**」、「**性意涵**」的字：「婊子、破麻、賤貨、騷貨、賤婊、賤人、賤女人、淫娃、懶趴、覽趴、懶覺、機掰、機巴、雞掰、機歪、洨、臭鮑、屄、屌」
* 有「**侮辱性**」意涵的字：「賽你、去死、吃屎、甲賽、三小、他媽的、你媽的、你他媽的」

## 資料蒐集及整理

資料蒐集的方式係根據2.2節所定義之不雅詞彙為基準，再採用人工搜尋和程式擷取所需之研究資料的這兩種方式來完成蒐集，詳細方式會在本節之兩小節作介紹。

### 由PTT蒐集資料

PTT是一個免費、開放、自由且擁有大量使用者的網路論壇，裡面更有各種的看板可提供各類議題的討論。

因本專題需要大量的不雅詞彙可供我們分析、觀察之用，考慮到PTT在這一點上十分符合本專題之需求，故我們採用Google做為搜尋引擎，並以PTT作為本專題主要資料的來源處。

【Step1】圖一為使用「幹你」作為關鍵字並於Google搜尋，其中「site:ptt.cc」為google snippets。



圖一、Google搜尋

【Step2】於Google搜尋後的會產生大量搜尋結果，考慮到若以手動的方式擷取文章內容並無效率，故改以手動的方式取得搜尋後的頁面原始碼。因為原始碼的部分即有文章所在的網址，此時再透過程式將網址給擷取出來即可。

下例為完成搜尋後，於頁面空白處點下右鍵檢視頁面原始碼，再將取得的網頁原始碼以純文字檔的方式儲存為檔案(編碼為utf-8)，供後續步驟所用。

搜尋結果可設定為1頁包含100則，本專題的做法為每個關鍵字約莫取前2~3頁，即200至300則的搜尋結果

<a href="**https://www.ptt.cc/man/Hate/D9DC/D5D2/D839/D858/DCE3/  
M.1277619431.A.D16.html**" > </a></li></ul></div></div></div><span class="st">作者: joebryant (幹的好!兒子) 看板: Hate 標題: [超幹] 女友和別人上床時間: Thu Apr 5 01:44:56 2007 劈腿是吧<em>幹你</em>娘公的劈腿攬叫爛光光母的走在路上被輪姦到靠背</span>

<a href="**https://www.ptt.cc/bbs/HatePolitics/M.1438939402.A.E1F.html**" >5 篇文章 - ‎1 位作者</div><span class="st">標題Re: [黑特] <em>幹你</em>媽的柯建銘. 時間Fri Aug 7 17:23:19 2015. ※ 引述《sa080691 (<wbr>帕森)》之銘言： : 幹，柯建銘我真他媽超不爽你: 在新竹當立委的時後亂搞，搞&nbsp;...</span>

<a href="**https://www.ptt.cc/bbs/HatePolitics/M.1438351365.A.7EA.html**" >[黑特] 我幹你娘機掰挺吳部長新課綱到底 - 批踢踢實業坊</a></h3><div class="s"><div>2015年7月31日 - 15 篇文章 - ‎11 位作者</div><span class="st">幹破你娘老雞掰反課綱的你娘機掰是在反啥小我幹破你娘機掰一群屁孩你他媽的有啥資格反課綱連老師都有被民進黨煽動的<em>幹你</em>老母有夠悲哀當道&nbsp;...</span>

【Step3】此範例中為取得搜尋後的總共兩百筆結果的網頁原始碼，再透過程式將原始碼中所包含之網址輸出成純文字檔 (不處理網址中包含Gossiping**、**sex、LGBT\_SEX的部分)。僅列舉部分網址。

https://www.ptt.cc/man/Hate/D9DC/D5D2/D839/D858/DCE3/M.1277619431.A.D16.html

https://www.ptt.cc/bbs/PuzzleDragon/M.1437528058.A.148.html

https://www.ptt.cc/bbs/Hate/M.1437168194.A.50A.html

https://www.ptt.cc/bbs/HatePolitics/M.1438939402.A.E1F.html https://www.ptt.cc/bbs/KanColle/M.1439425825.A.9BF.html

https://www.ptt.cc/bbs/HatePolitics/M.1438351365.A.7EA.html

【Step4】在取得大量文章所在的網址後，需要透過程式來抓取網址所對應到的網頁內容，並擷取出主要的文章部分。

透過程式抓取網頁內容的所有文字(包含屬於網頁的標記語言，非文章內容的部分)並輸出為純文字檔，在使用另一程式擷取出主要的文章(從「作者」至「--」結尾)。此範例為未擷取前的網頁內容片段。

<div style="border: #999 1px solid; padding: 10px; background-color: #eee;"><pre>作者: joebryant (幹的好!兒子) 看板: Hate

標題: [超幹] 女友和別人上床

時間: Thu Apr 5 01:44:56 2007

劈腿是吧

幹你娘公的劈腿攬叫爛光光

母的走在路上被輪姦到靠背靠幕

他媽劈腿不得好死

下輩子當圓規 幹你娘讓妳他媽劈個夠本

--

【Step5】將蒐集而來的大量文章經過程式處理後，變成方便觀察的格式。此時再以人工觀察的方式大量觀察這些資料，最終便歸納出本系統所欲處理的不雅詞彙範圍。詳細範圍及改寫規則請參照第三章。

所有蒐集來的網頁內文以程式刪除多餘的空白、空行或重複句子後，彙整到同一個檔案內，再將所有句子中包含2.1節所定義之不雅詞彙(以行為單位，行中包含不雅詞彙即輸出該行，行首為對應的關鍵字)輸出至純文字檔。以此做觀察歸納的來源。

幹 幹你娘公的劈腿攬叫爛光光

幹 下輩子當圓規 幹你娘讓妳他媽劈個夠本

幹 幹，柯建銘我真他媽超不爽你

屌 怕你酸不贏！Damian Lillard神回覆屌嗆白目網友

懶覺 突然, 她想起之前有聽大人說過睡懶覺

機掰 我第一次這麼機掰一個老師過...以前罵罵就算了，今天真的...

### Twitter

由於Twitter做為網路上流行的社群網站之一，而其亦內建有關鍵字搜尋功能，故將Twitter納入本實驗的第二個蒐集資料的來源。

Twitter與Google最大的迥異在於，Twitter不能藉由擷取網頁原始碼一次取得大量的資料，必須以純人工的方式一則一則複製貼文，十分費時，無法符合本實驗大量分析資料之需求。故將之列為次要的資料來源，僅做為參考及歸納規則之用。

【Step1】開啟並登入Twitter帳號後於搜尋欄位中分別輸入2.1節所定義之不雅詞彙做為關鍵字搜尋後得到包含關鍵字的相關貼文(此處使用「幹你」作為關鍵字)



圖二、Twitter搜尋

【Step2】將所有蒐集來的Twitter貼文，依當初搜尋時的關鍵字，以手動的方式複製並存放到不同的純文字檔中，做為後面觀察、分析取代規則的資料來源。

於Twitter關鍵字搜尋後，出現的前約20至30則的貼文，擷取其含有不雅詞彙的句子，並自行分存於檔名與關鍵字相應的txt檔中。Twitter搜尋結果中部份推文中帶有簡體字或者日文中的漢字，故以編碼為utf-8進行儲存，避免資料流失。

幹你娘這三小wwwwwwwwwwwwwww

幹你娘勒一小時騎八點八公里的運動才一百大卡喔

難怪大家搶著要去賣淫幹你娘

幹你娘在跟我靠背一句試試看

幹 你連這都不會？

幹...好勵志...（跪哭）

幹，耍什麼白痴，抗旱鐵牌咧，何不乾脆生一個抗幅射鐵牌，這樣馬英九最愛的核四就可以繼續蓋了。

底下有一則留言很好笑：「這些小傢伙的其中一員幹掉了奧薩瑪賓拉登！？」

你這個婊子

婊子们 谢谢你给我们哥哥涨粉了

俺就是婊子，老娘这次不立牌坊了，你们还想咋的？！

回頭再想想鬼燈的冷撤裡的那隻狠兔子，就覺得她不過是個性格超爛的臭婊子w

# 不雅詞彙規則改寫

此節所列的規則產生方式係由第二節所蒐集來的資料，透過觀察、歸納後並討論出來的取代方式，為人工觀察後產生的結果。

## 取代規則

此處僅簡單列舉實驗中使用之取代規則，並未窮舉所有取代方式。例句中的上句加框字為被取代字，下句之加框字為取代字，若同句內有出現一則以上取代規則，則以該例句之規則為主，其餘出現規則以底線標記之。

1. 「幹」+ {代名詞} + {名詞} →「真是可惡」  
   {標點符號} +「幹」+ {代名詞} + {名詞} → {標點符號} +「真是可惡」

**例如：**幹你娘 看轉播也看到快吐血  
**→** 真是可惡 看轉播也看到快吐血

本規則的**「幹」**可替換為：**「賽」、「操」、「肏」**

1. 「幹」前後加上任意{標點符號} →「可惡」前後加上任意{標點符號}

**例如：**幹！被老師罵了之後還跟我講一堆屁話！明明是我在幫你擦屁股！！！   
**→** 可惡！被老師罵了之後還跟我講一堆屁話！明明是我在幫你擦屁股！！！

1. {形容副詞} +「幹」→ {形容副詞} +「怒」

**例如：**剛剛遇到4所以來發這篇 ...... 各位還有遇過什麼很幹的事嗎  
**→** 剛剛遇到4所以來發這篇 ...... 各位還有遇過什麼很怒的事嗎

1. 「幹」出現在句首 → 「可惡」  
   複數個「幹」 → 「可惡」

例如：幹 最雞掰的是  
→ 可惡 最機車的是

1. 「賤婊」→「壞女人」

例如：賤婊會跟賤婊conversation   
→ 壞女人會跟壞女人conversation

本規則的**「賤婊」**可替換為：**「婊子」、「破麻」、「賤婊子」、「淫蕩」、「淫娃」、「賤貨」、「賤女人」**

1. 「賤人」→「壞人」

例如：賤人 就是 一個 使勁兒 的 傷害 別人，然後 自己 在哪兒 吶喊 疼痛 的人。  
→ 壞人 就是 一個 使勁兒 的 傷害 別人，然後 自己 在哪兒 吶喊 疼痛 的人。

1. 「機掰」→「機車」  
   「機機歪歪」→「機車」

例如：我要跟他借車還擺機掰臉!!!! 摔我車沒貼錢就算了，換我要跟他借車就在那邊機掰洨!!!  
→ 我要跟他借車還擺機車臉!!!! 摔我車沒貼錢就算了，換我要跟他借車就在那邊機車什麼!!!

本規則的**「機掰」**可替換為**：「雞巴」、「機歪」、「機八」、「雞八」  
、「雞歪」  
「機機歪歪」**可替換為：**「雞雞歪歪」**

1. 「歸」+「覽趴會」→「滿」+「肚子氣」  
   「覽趴會」→「一肚子氣」

**例如：**你們一定都有「幹，這樣會起火吧？」的疑問，所以我每次尻的時候都很生氣，這也是歸覽趴會的由來  
**→** 你們一定都有「可惡，這樣會起火吧？」的疑問，所以我每次尻的時候都很生氣，這也是滿肚子氣的由來

本規則的**「歸」**可替換為：**「規」、「龜」**  
**「覽趴會」**可替換為：**「覽趴火」、「懶趴會」、「懶趴火」、「攬趴會」  
、「攬趴火」**

1. 「懶趴」→「那話兒」

**例如：**少跟我靠北 老子只是懶的鳥你 醜懶趴只會在那邊豪洨  
**→** 少跟我靠北 老子只是懶的鳥你 醜那話兒只會在那邊唬人

本規則的**「懶趴」**可替換為：**「覽趴」、「懶覺」、「懶較」、「覽覺」、  
「覽較」**

1. 「去死」→「去撞牆」

**例如：**特別愛說「齁唷~~」「咦」「去死啦」  
**→** 特別愛說「齁唷~~」「咦」「去撞牆啦」

本規則的「去死」可替換為：「去吃屎」、「去死一死」

1. {動詞} +「三小」→ {動詞} +「什麼」

**例如：**標題[福龍] 179話，李淳風你到底在衝三小...  
**→** 標題[福龍] 179話，李淳風你到底在衝什麼...

1. {介係詞} +「豪洨、唬洨」→ {介係詞} +「瞎扯」

例如：不貼圖人家以為我在豪洨，幹，超像炳中哥啊wwwww #不一樣的美男  
→ 不貼圖人家以為我在瞎扯，可惡，超像炳中哥啊wwwww #不一樣的美男

本規則的**「豪洨」**可替換為：**「唬洨」、「虎洨」、「毫洨」**

1. 「豪洨」→「唬人」

**例如：**妳他媽的這種幹話是想唬洨誰啊？  
**→** 妳他媽的這種X話是想唬人誰啊？

本規則的**「豪洨」**可替換為：**「唬洨」、「虎洨」、「毫洨」**

1. 「甲洨」→「撞牆」

**例如：**這樣還好意思跟人家說你經理？ 甲洨厚啊～  
**→** 這樣還好意思跟人家說你經理？ 撞牆厚啊～

本規則的**「甲洨」**可替換為：**「假洨」、「呷洨」**

1. 「三洨」→「什麼」

**例如：**[翻譯] PewDiePie - 你他媽的我到底在玩三洨  
**→** [翻譯] PewDiePie - 你他媽的我到底在玩什麼

本規則的**「三洨」**可替換為：**「撒洨」、「殺洨」、「啥洨」、「沙洨」**

1. 「的洨」+ {標點符號}→「的精液」+ {標點符號}  
   「吃洨」 →「吃精液」

**例如：**你要吃我的洨？  
**→** 你要吃我的精液？

1. 「魯洨」→「糾纏」  
   「盧洨洨」→「煩人精」

**例如：**常聽到有人說...你不要魯洨...  
**→** 常聽到有人說...你不要糾纏...

本規則的**「魯洨」**可替換為：**「盧洨」、「嚕洨」**  
本規則的**「盧洨洨」**可替換為：**「魯洨洨」、「嚕洨洨」**

1. 其它不符合上面規則之「洨」→「小」

**例如：**洨明跟小華在聊天  
**→** 小明跟小華在聊天

1. 「牛屄」→「厲害」

**例如：**看中國台灣同胞整天在那吵ikea 的發音就覺得特牛逼  
**→** 看中國台灣同胞整天在那吵ikea 的發音就覺得特厲害

本規則的**「牛屄」**可替換為：**「牛逼」**

1. 「傻屄」→「傻子」

**例如：**買不起還非要建，你不是傻屄還有誰是傻屄？  
**→** 買不起還非要建，你不是傻子還有誰是傻子？

本規則的**「傻屄」**可替換為：**「傻逼」**

1. 「臭屄」→「臭下體」

**例如：**屌你老媽子,臭屄  
**→** 理你老媽子,臭下體

本規則的**「臭屄」**可替換為：**「臭逼」**

1. 其它不符合上面規則之「屄」→「女生下體」

**例如：**她說沒有,因為她沒有屄毛  
**→** 她說沒有,因為她沒有女生下體毛

1. {形容副詞} +「屌」→{形容副詞} +「厲害」

例如：我是蠻屌的  
→ 我是蠻厲害的

1. 「屌打」→「打爆」

例如：看完格雷的五十道陰影 只有對這首歌有印象結果是以前的神曲 REMIX成情慾版 XDDDDD 屌打其他歌曲  
→ 看完格雷的五十道陰影 只有對這首歌有印象結果是以前的神曲 REMIX成情慾版 XDDDDD 打爆其他歌曲

1. 「屌」+ {代名詞} →「理」+ {代名詞}

例如：人家根本就不會屌你 去分配只會被其他兵吐口水而已  
→ 人家根本就不會理你 去分配只會被其他兵吐口水而已

1. 其它不符合上面規則之「屌」→「那話兒」

例如：後來知道屌這個字的原義後就再也不講了  
→ 後來知道那話兒這個字的原義後就再也不講了

1. 正常使用下的「幹」→ 「=」(意指不取代)

例如：(也沒拍照幹嘛的 我好傻==)  
→ (也沒拍照幹嘛的 我好傻==)

此處的正常使用的定義係指出現於《同義詞詞林》中，有「幹」字但並不符合本專題定義的詞彙者。部分列舉如下：「幹嘛」、「幹練」、「精幹」、「樹幹」等。

1. 分類為「其他」類列舉如下：  
   「的豪洨」→「的誇張」  
   「眼洨」 → 「眼殘」  
   「擦洨」 → 「理會」  
   「怪洨」 → 「怪咖」  
   「睡懶覺」 → 「睡覺」

例如：這下好了 跟一個怪洨住就算了  
→ 這下好了 跟一個怪咖住就算了

1. 「他媽的」→ 「X」  
   「你媽的」→ 「你的」

例如：幹~來到菜市場還他媽的一直嫌髒!裝高貴~雖然不會殺價!  
→ 可惡~來到菜市場還 一直嫌髒!裝高貴~雖然不會殺價!

## 取代結果

為了觀察取代後的結果是否能達到本專題之研究目的：完成取代後是否能符合原本文章的文意，故依3.1節所訂定的取代規則，另行撰寫程式處理經由第二節的方式所蒐集來的資料。完成取代後，再輸出成方便觀察之樣式的純文字檔，以供第四章實驗所用，詳細如以下範例。

ID=03785

你天天講手機不會累，拎杯耳朵都快長繭了，幹！

你天天講手機不會累，拎杯耳朵都快長繭了，可惡！(，幹！->，可惡！)

ID=04354

同學B：我屌你！(你自生自滅吧！)

同學B：我理你！(你自生自滅吧！)(屌你->理你)

ID=04362

同：你去死啦！

同：你去撞牆啦！(去死->去撞牆)

ID=04404

和幹訓隊比，一般中新訓真的是天堂。

和幹訓隊比，一般中新訓真的是天堂。(幹訓->=)

ID=04433

哎唷哎唷這個屌

哎唷哎唷這個厲害(這個屌->這個厲害)

ID=04630

在那個賤人帶著她那叫做B的美麗小女嬰回來之前都是孤獨的。

在那個壞人帶著她那叫做B的美麗小女嬰回來之前都是孤獨的。(賤人->壞人)

ID=04674

大家不該把好人當成聖人當成傻屄，應該更加主動，不是嗎?

大家不該把好人當成聖人當成傻子，應該更加主動，不是嗎?(傻屄->傻子)

ID=06027

幹自己條件很好喔 媽的我是你男友的話我都不要了

可惡自己條件很好喔 媽的我是你男友的話我都不要了(幹->可惡)

# 實驗

## 評估

本實驗欲對3.2節的取代結果進行評估。實驗中會有某少部分規則所適用的句子稀少的問題，其原因是3.2節少部分規則本身並不夠詳盡，再或者是蒐集資料時因PTT於某一類規則之資料量稀少等因素所致。

因為本實驗係採用人工觀察的方式進行評估，故此實驗僅對每一種規則分類底下的資料做部分擷取：最多取至100組句子，若不足100組句子就取其所有句子

評估時則是針對所取的100組(或少於100組)句子其中的每一組(包含取代前、後的句子及所用的取代規則)進行對比：在句子被取代後是否能符合被取代前的句子之文意為判別方法，來決定此組是取代成功或者是取代失敗。

【評估過程】將2.2節所蒐集而來的共9557句資料，依3.1節的取代規則加以分類，共可分出29類，。每一分類內皆有若干與規則相符的句子。取得分類後首先以程式亂數排序之，再各自將句子輸出到與規則相符的純文字檔內，並以3.2節所示之方式呈現，再用人工觀察的方式，以Y或N(Y是正確、N是失敗)標註出取代完成後正確與否。最後將標註完的結果整理為表格。詳細請見表一。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **case** | **Y** | **N** | **accurate** | **case** | **Y** | **N** | **accurate** | |
| 1 | 85 | 15 | 0.85 | 16 | 39 | 5 | 0.89 |
| 2 | 91 | 9 | 0.91 | 17 | 98 | 2 | 0.98 |
| 3 | 98 | 0 | 1 | 18 | 78 | 22 | 0.78 |
| 4 | 76 | 24 | 0.76 | 19 | 56 | 44 | 0.56 |
| 5 | 98 | 2 | 0.98 | 20 | 99 | 1 | 0.99 |
| 6 | 97 | 3 | 0.97 | 21 | 30 | 11 | 0.73 |
| 7 | 70 | 30 | 0.7 | 22 | 47 | 53 | 0.47 |
| 8 | 93 | 5 | 0.95 | 23 | 98 | 2 | 0.98 |
| 9 | 81 | 19 | 0.81 | 24 | 93 | 1 | 0.99 |
| 10 | 88 | 12 | 0.88 | 25 | 30 | 14 | 0.68 |
| 11 | 2 | 0 | 1 | 26 | 68 | 32 | 0.68 |
| 12 | 12 | 0 | 1 | 27 | 93 | 7 | 0.93 |
| 13 | 96 | 4 | 0.96 | 28 | 10 | 0 | 1 |
| 14 | 42 | 4 | 0.91 | 29 | 86 | 14 | 0.86 |
| 15 | 90 | 10 | 0.9 | **total** | 2044 | 345 | 0.86 |

表一、取代規則正確率

## 討論

本實驗因為受到資料蒐集不齊全(文意不清)、取代規則不夠詳盡、遇到取代規則包含專有名詞等技術上的困難，或是取代結果正確與否的主觀想法的因素所影響，故可發現部分取代規則之正確率呈現較低的情形。需等待日後發現更符合原意之規則，或是另有其他較好的取代方式，方得以修正。

以下會加以說明正確或失敗的例子。

**正確：**

ID=06162 y

幹，死賤人，這麼怕冷滾回你家去喝你媽的奶啦！

可惡，死壞人，這麼怕冷滾回你家去喝你的奶啦！(幹，->可惡，) (賤人->壞人) (你媽的->你的)

**失敗：**

1. **資料蒐齊不齊全：**文意不清，導致取代失敗

ID=00837 n

: 聊天時雞巴咖啡

: 聊天時機車咖啡(雞巴->機車)

1. **取代規則不夠詳盡：「**幹0糧」應套用3.1節的規則1，但因未能窮舉規則1之所有情形而無法套用該規則，導致取代失敗

ID=05411 n

幹0糧，各位大朋友小碰友下午好

可惡0糧，各位大朋友小碰友下午好(幹->可惡)

1. **專有名詞：**牛逼洋行為一個店名，不應該取代

ID=04065 n

到台中的第一站，就是到知名漢堡店牛逼洋行開眼界，

到台中的第一站，就是到知名漢堡店厲害洋行開眼界，(牛逼->厲害)

1. **主觀因素：**

A認為(嚕洨->糾纏)可以

ID=00539 y

: 嚕洨的人很多 我們沒有辦法盡如人意

: 糾纏的人很多 我們沒有辦法盡如人意(嚕洨->糾纏)

B認為此處(嚕洨->糾纏)不通順，應為嚕洨->煩人

ID=00539 n

: 嚕洨的人很多 我們沒有辦法盡如人意

: 糾纏的人很多 我們沒有辦法盡如人意(嚕洨->糾纏)

# 線上系統

## 網頁版不雅文字自動改寫系統

在Input框中輸入要取代的字或文章，按下Submit按鈕遞交後，於下方Output框中顯示系統取代後的結果。



圖三、網頁版畫面

## FB應用程式版不雅文字自動改寫系統

於原本網頁版中新增了發布到塗鴉牆功能：圖三中Post To Facebook按鈕可將取代結果發布至個人塗鴉牆上(發布後的結果如圖四)。並設計了Facebook Social Plugins 可以按讚或是分享此網頁至Facebook，並提供Facebook留言板可留言交流之用。

圖四、FB塗鴉牆貼文截圖



# 結論

本專題從第四章的實驗中的數據可看出，雖有部分規則正確率較低仍有待改進，但大部分的規則正確率仍是較高的，意即本專題的系統若僅針對一般的文章做取代後，於前後文文意上的處理並無太大問題。 未來若有相關研究或是任何實際應用方面的需求，相信本專題的取代規則或是研究過程都能給予參考、幫助，甚至是直接套用本專題的系統，但是相關程式內容的改寫，規則取代正確率的提高，還有帶進一步的研究。