

1 1 0 全國大專校院數位人文
大數據學生競賽

網路霸凌守門員-打造更安全的網路世界

踏實組

團隊成員

張培莉、吳貫綸、黃慧蓮

吳東儒、鄭雅綿、林汪鴻宇

指導老師

蔡芸琇 老師

中 華 民 國 111 年 1 月 22 日

目錄

壹、前言.....	3
一、創作動機與目的	3
二、作品介紹.....	3
貳、正文.....	4
一、作品使用手冊.....	4
二、開發流程與技術架構	11
參、市場分析.....	16
肆、團隊成員與分工說明	17
伍、未來展望.....	18
陸、參考資料.....	18

壹、前言

一、創作動機與目的

隨著科技進步，全球進入網路資訊時代，任何年齡層的人皆可以透過網路與朋友聊天、購物、遠距教學或遠距上班、瀏覽新聞及經營社群媒體等，網路帶給了我們方便及豐富的生活，而看似所有人都是得利者的網路時代，卻存在著一群網路霸凌的受害者。

網路霸凌是指透過網路為媒介，散播不雅圖片、影片、不實訊息及辱罵的言語，使這些訊息流傳於同儕或一般大眾，使受害者感到害怕，達到中傷他人目的（吳明隆、陳明珠，2012；張淑亭，2015）。根據兒福聯盟「2021 兒少網路社交焦慮與網路霸凌經驗調查」，有 8 成兒少認為網路霸凌情形嚴重，逾 2 成曾經歷網路霸凌，等於平均每 5 名兒少就有超過 1 名是網路霸凌受害者，其中被同學跟朋友網路霸凌比例最高。

為減少網路霸凌發生，本團隊嘗試站在受害者角度，將網路霸凌分為兩個階段，預防及求助階段。在預防階段，本團隊開發留言守門員(Chrome 應用程式)，在使用者上網時，可主動過濾掉惡意攻擊的字詞，避免上網者接收到文字攻擊；在求助階段，本團隊開發網路霸凌糾察員(LINE BOT)，受害者透過此官方帳號，可以了解如何替自己發聲，包含蒐集加害者犯罪證據及提告流程。除了上述功能外，官方帳號也提供網路霸凌相關資訊，供民眾認識網路霸凌及其案例，希望民眾能了解網路霸凌，避免成為加害者。

二、作品介紹

網路霸凌的形式與種類隨著科技發展日新月異，網路霸凌加害者因好玩、忌妒等因素，使用文字、圖片或影片等媒介，在網路上長期對受害者進行侮辱、抹黑等攻擊使其心理受創。令人難過的是，目前仍沒有一套完整的系統可以預防網路霸凌的發生，而本團隊站在受害者角度針對預防方法深入探討，發現受害者感到害怕及不舒服的是內容，於是開發出留言守門員。

留言守門員是一款 Chrome 瀏覽器的擴充套件，具有文字過濾功能，內建黑名單詞庫，可過濾網頁中令人感到不舒服的字詞，給網路使用者友善的上網環境，適合用於教育單位，本團隊將此應用的目標對象設定於國高中小學生，各學校單位可在校園內電腦安裝此瀏覽器擴充套件，當目標對象於電腦課教學瀏覽網頁時，可避免具侮辱性的惡意字詞攻擊，保護青少年的身心健康。因為網路霸凌的形式及變化遠超過可預防的速度，本團隊認為留言守門員不能解決所有網路霸凌問題，同時開發 LINE BOT 網路霸凌糾察員，適合各年齡層使

用，只需要加入網路霸凌糾察員的 LINE 官方帳號即可使用，功能主要分成兩個面向，第一個面向是認識網路霸凌，使用者可以透過與網路霸凌糾察員互動，了解什麼是網路霸凌、網路霸凌的過去案例以及網路霸凌會面臨的相關罰則。第二個面向是網路霸凌的救濟方法，使用者可以了解網路霸凌提告的流程，捍衛自身權益。此外，使用者也可以使用網路霸凌行為衡量的功能，向網路霸凌糾察員說明自身遇到的情境，而網路霸凌糾察員會判斷該情境適用於何種法條或是提告的成功率。本團隊期望透過留言守門員與網路霸凌糾察員的兩款應用，讓多數人認識網路霸凌以及減少網路霸凌所帶來的傷害。

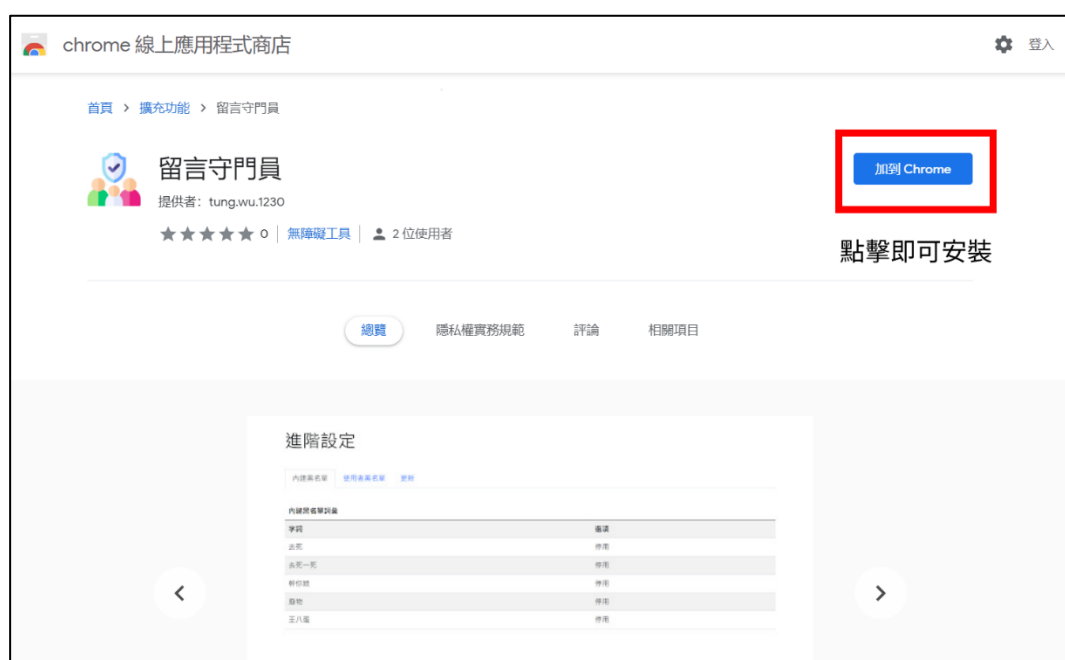
貳、正文

一、 作品使用手冊

1 留言守門員

1.1 安裝擴充功能

使用 Chrome 瀏覽器或 Edge 瀏覽器，前往 [Chrome 線上應用程式商店](#) 搜尋 [留言守門員](#)，並點選加到 Chrome 按鈕即可安裝，如圖[1]。



圖[1]，留言守門員 Chrome 應用程式商店。

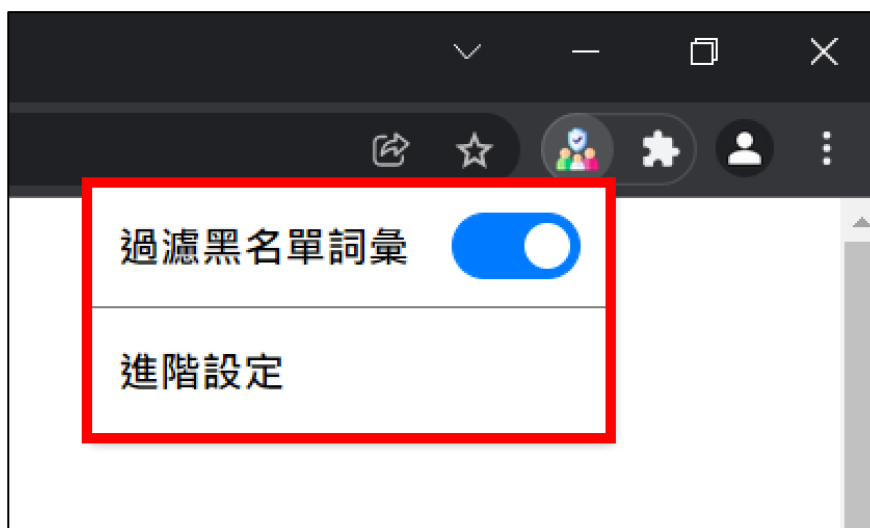
若成功安裝，擴充功能列表將出現留言守門員的圖示，並跳出感謝安裝的頁面，如圖[2]，在此畫面使用者可以了解套件運作方式以及如何使用。



圖[2]，感謝安裝頁面。

1.2 開始使用留言守門員

首次安裝後，過濾黑名單詞彙功能為開啟狀態，如圖[3]，當任何包含黑名單詞彙的網頁被載入時，黑名單詞彙的將被星符號取代，如圖[4]，透過遮蔽不當字詞的方法，可避免使用者直接閱讀到不當字詞。若網頁內容顯示異常或是欲關閉過濾功能，可點選擴充功能列表裡的留言守門員彈出視窗進行關閉，如圖[3]。



圖[3]，留言守門員彈出視窗。

2016/07/01 10:05

〔即時新聞／綜合報導〕時常有人因為罵人挨告，賠償金額從0到上百萬均有，有網友自行挑選相關新聞後，整理出「台灣罵人價目表」，雖然無法窮盡所有判例，但也讓人莞爾。

一般而言，一個字的國罵最常出現，但可能會被界定為台語發語詞，因此無罪。至於更具羞辱性的像是「**」、「**」等詞彙，則會讓當事人名譽受損，獲得賠償。有趣的是，在這份「台灣罵人價目表」中罰款金額最高的，並非一般用來罵人的語句，而是作家馮光遠形容馬英九與金溥聰的「特殊性關係」，當時金溥聰怒告馮加重誹謗求償200萬元。高等法院判馮光遠須賠100萬元並在4大報頭版登報道歉1天。全案可上訴。

網友整理出的「台灣罵人價目表」，依價格高低排列如下：

罵「X」只是台語發語詞 無罪

罵「更年期到了」判賠2千

罵「**」判賠3千

罵「敗類」判賠3千

罵「米蟲」判賠6千

罵「***」判賠1萬

罵「不要臉髒東西」判賠2萬

罵「你去**啦」判賠2.5萬

罵「**就是矯情」判賠5萬元

圖[4]，留言守門員過濾示意圖。來源：[自由時報](#)

1.3 進階設定

點選留言守門員彈出視窗裡的進階設定可進入黑名單詞庫管理頁面，如圖[5]，使用者可透過此頁面查看內建黑名單與使用者黑名單。內建黑名單可進行操作包含啟用、停用字詞，若將字詞進行停用，則網頁載入時將不會過濾被停用的字詞。若使用者想增加過濾字詞，則可以在管理使用者黑名單之頁面進行新增與刪除。最後一項功能是更新，若本團隊有新增更多不當字詞，使用者可透過此功能獲取更多的內建黑名單字詞。

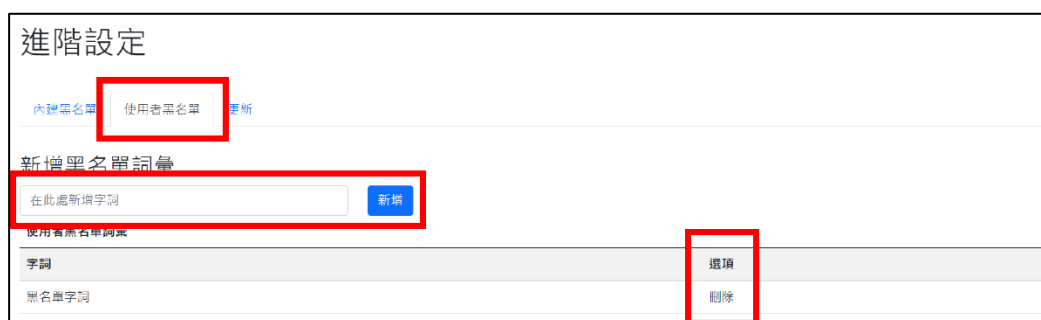
進階設定

內建黑名單 使用者黑名單 更新

內建黑名單詞彙

字詞	選項
下賤	停用
人渣	停用
他媽的	停用
你他媽	停用
你媽	停用
你媽的	停用

圖[5]，進階設定管理頁面。



圖[6]，管理使用者黑名單。



圖[7]，更新官方資料庫。

2 網路霸凌糾察員

2.1 加入網路霸凌糾察員 LINE 官方帳號

首次加入網路霸凌糾察員官方帳號的使用者，可以閱讀使用教學，糾察員提供四大主要功能及兩個附加功能供使用者選擇。

主要功能分別為網路霸凌相關條款介紹、過去提告案例（相關新聞）、提告流程介紹、網路霸凌行為衡量，次要功能為留言守門員套件介紹及本團隊理念設計。



圖[8]，圖文選單。

2.2 功能介紹

➤ 網路霸凌相關條款介紹

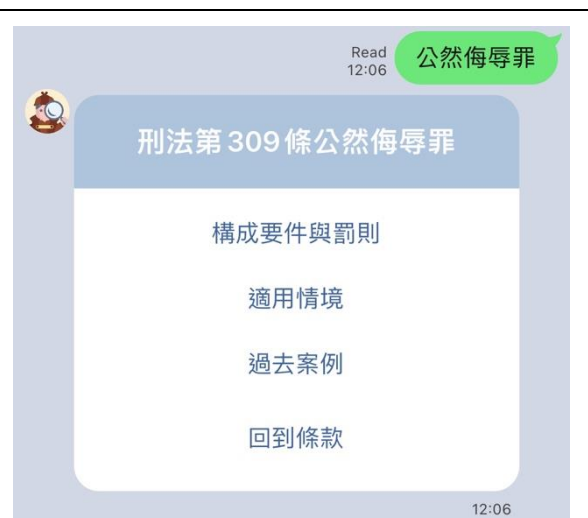
點選下方選單展開選單列表，並點選**網路霸凌相關條款**按鈕即可使用該功能。

選擇任一法條後，該按鈕提供三大功能供使用者選擇：

1. 構成要件與罰則
2. 適用情境
3. 過去案例



圖[9]，網路霸凌相關條款功能列。

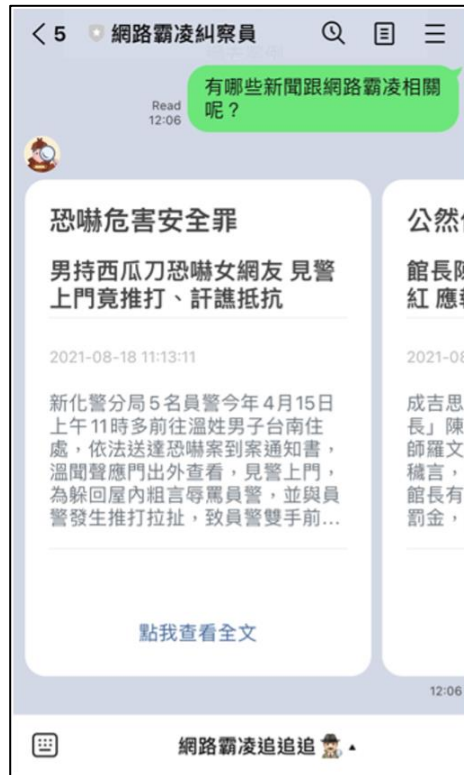


圖[10]，點擊法條後的功能列。

➤ 過去提告案例（相關新聞）

點選下方選單展開選單列表，並點選**相關新聞**按鈕即可使用該功能。

糾察員將顯示最新網路霸凌之相關法條新聞 10 筆。



圖[11]，網路霸凌相關新聞頁面。

➤ 想提告怎麼辦

點選下方選單展開選單列表，點選**想提告怎麼辦**即可使用該功能。

該按鈕提供三大功能供使用者選擇：

1. 網路霸凌是什麼
2. 提告的事前準備
3. 如何進行提告（自訴、告訴）

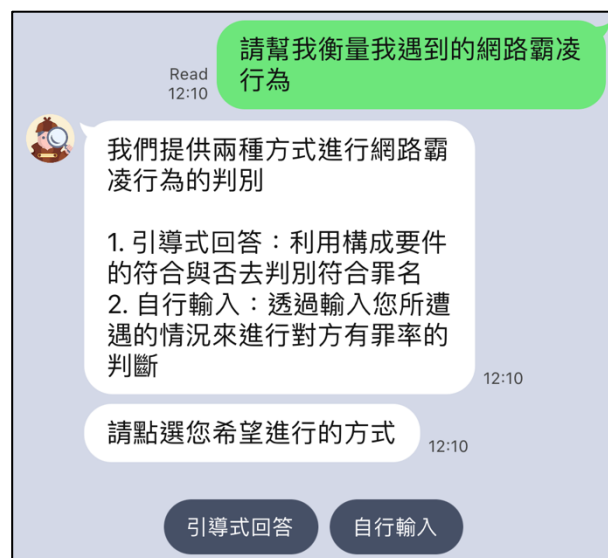


圖[12]，提告流程功能列。

➤ 網路霸凌行為衡量

點選下方選單展開選單列表，點選**網路霸凌行為衡量**即可使用該功能。

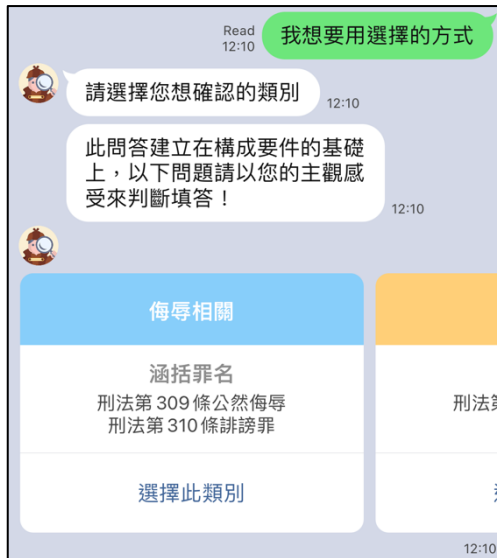
該按鈕提供兩功能供使用者選擇：



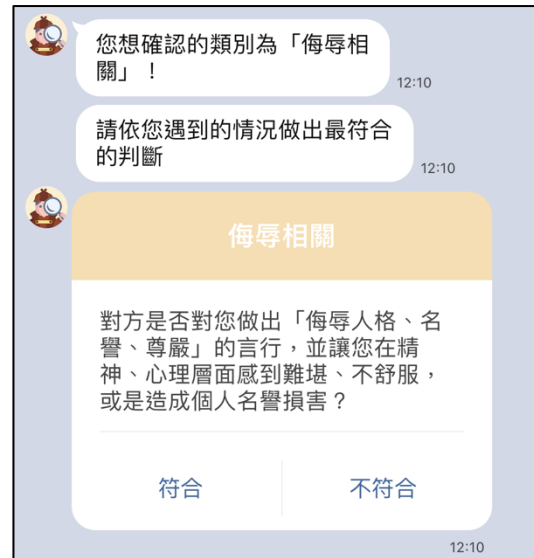
圖[13]，網路霸凌行為衡量提供兩種方式進行。

1. 問答模式：

使用者可先選擇想確認的條款類別，再透過基於構成要件的選項答題模式，系統將依所選情境推敲出符合該情形之構成要件的法條。



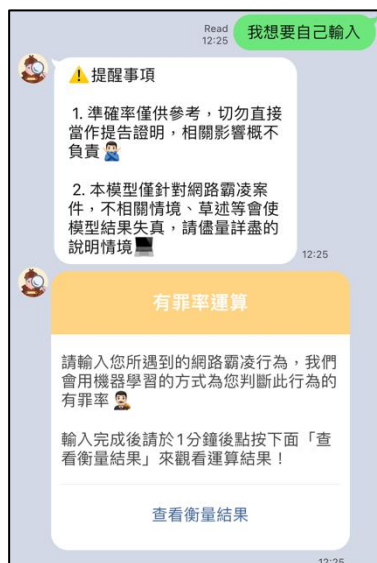
圖[14]，引導式回答介面-1。



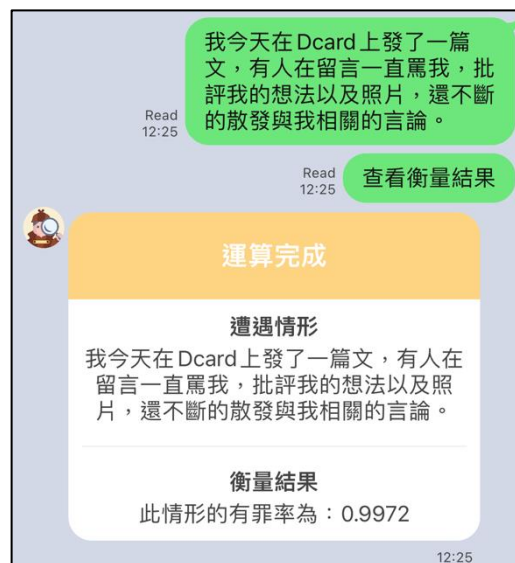
圖[15]，引導式回答介面-2。

2. 短文輸入模式：

使用者可輸入一段網路霸凌事件發生始末，系統將依該情況分析後顯示符合提告的可能勝訴率。



圖[16]，自行輸入介面-1。



圖[17]，自行輸入介面-2。

二、開發流程與技術架構

1 留言守門員

1.1 系統架構

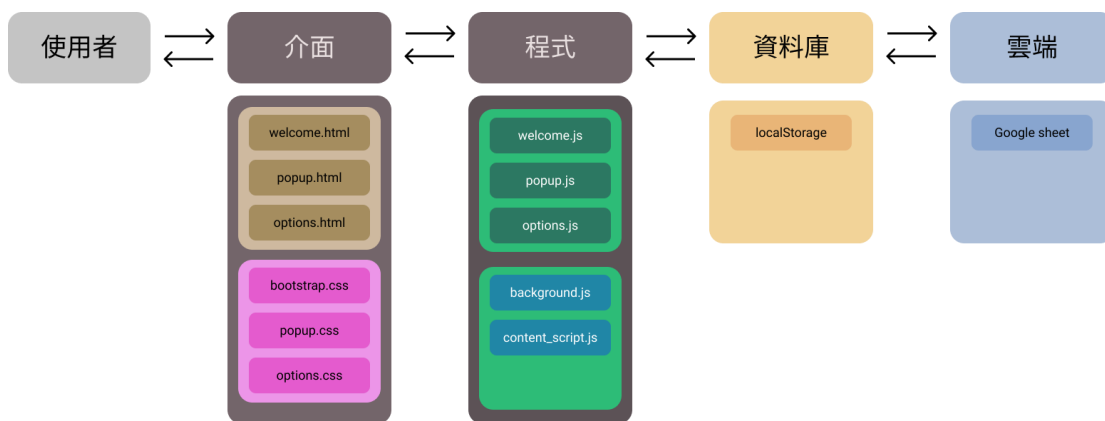
留言守門員是 Chrome 瀏覽器的擴充功能，基於 Manifest V3 版本所開發，其架構如下圖[18]所示，主要由四大部分組成，分別為介面、程式、資料庫及雲端。

介面部分使用技術以 HTML 與 CSS 為主，部分使用 Bootstrap 框架。首先透過 HTML 打造出介面骨幹，並將程式功能綁定於介面按鈕，再透過 CSS 美化介面，提供使用者簡單可操作的畫面。

程式部分使用 JavaScript 及 Chrome.* APIs 組成。JavaScript 負責邏輯判斷以及 DOM(Document Object Model)處理，Chrome.* APIs 與 Chrome 瀏覽器進行交互，負責取得使用者瀏覽的頁面以及儲存黑名單詞庫等資料。

資料庫由 Chrome.* APIs 中的 chrome.storage 所控制，透過程式與資料庫互動，可以儲存、讀取、更新及刪除本應用中的黑名單詞庫。

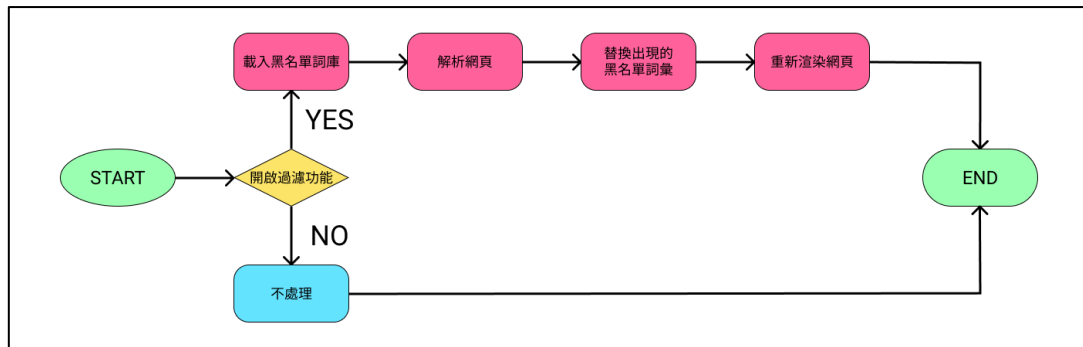
雲端使用 Google Sheets 為線上的黑名單詞庫，透過程式發出請求至雲端試算表，取得最新版本的黑名單字詞後寫入資料庫，留言守門員完成更新黑名單詞庫。



圖[18]，留言守門員-系統架構圖。

1.2 運作流程

留言守門員運作流程如圖[19]所示，每次瀏覽器載入頁面，程式首先會判斷過濾功能是否開啟，預設為開啟狀態，使用者可透過彈出視窗關閉過濾功能。若過濾功能開啟，程式將會載入黑名單詞庫，再解析網頁內容，比對網頁中特定標籤(<p>、、<a>)底下的文字是否包含黑名單字詞，將其中的黑名單字詞替換為星符號(*)並重新渲染內容，直到比對完所有特定標籤底下的文字。



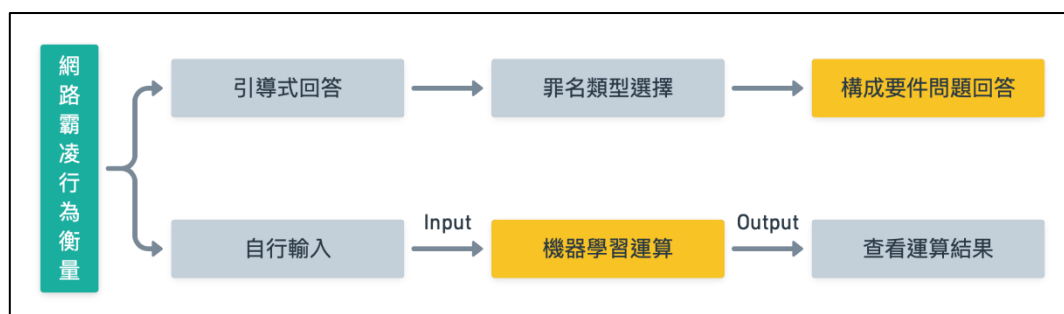
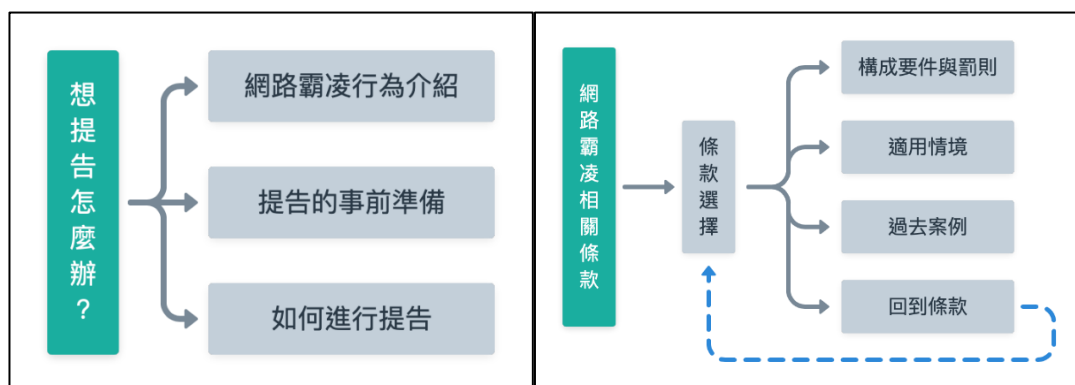
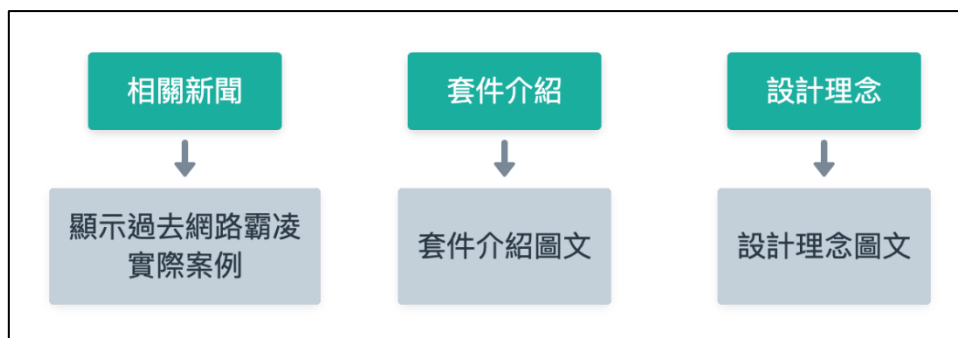
圖[19]，過濾功能運作流程圖。

1.3 開發工具

程式語言、套件、工具	說明	版本
HTML	超文本標記語言，用於呈現網頁元素。	HTML5
CSS	階層式樣式表，用於美化網頁元素。	CSS 3
JavaScript	用於邏輯處理、與瀏覽器互動之程式語言。	ECMAScript 2019
Bootstrap	網頁開發前端框架，用於美化頁面。	v 5.1.3
Chrome Extension API	Chrome 提供 extension 許多特殊用途的 API。	Manifest V3
Google Sheet	Google 試算表，用於儲存官方黑名單字詞資料庫使用。	

2 網路霸凌糾察員

2.1 流程圖



2.2 開發工具

程式語言、套件、工具	說明	版本
Python	程式語言，用於匯入統整各種專案套件。	3.7.4
Flask	是一個使用 Python 編寫的輕量級 Web 應用框架，用於建立 Line Bot 的 Server 產生 Webhook，並回應 Line Server 的請求。	1.1.1
LINE BOT API	Line 官方 API，讓 data 可於 bot server 及 LINE Platform 之間傳遞。	1.19.0

Pygsheets	為 Google 自有的 API，可將 Python 連接上 Google Sheets。	2.0.5
Google Sheet	Google 試算表，用於儲存相關法條新聞及輸入情境。	
Pytorch	是一個開源的機器學習庫，用於處理網路霸凌行為衡量使用者自行輸入情境的文字處理。	1.10.1
Scikit-learn	用於 Python 的自由軟體機器學習庫。使用其中的 Logistic Regression 機器學習模型。	0.21.3
Transformers	用於自然語言處理(NLP)最先進的預訓練模型，使用其中的 BertTokenizerFast、AutoModelForMaskedLM、AutoModelForCausalLM 及 AutoModelTokenClassification 套件。	3.4.0
Microsoft Azure	微軟的公用雲端服務 (Public Cloud Service) 平台，透過此平台服務將機器人部署至雲端。	

參、市場分析

一、產品優勢

留言守門員(Chrome 擴充套件)與網路霸凌糾察員(LINE BOT)皆屬於免費應用，任何網路使用者都可透過網路安裝擴充套件或加入網路霸凌糾察員好友，兩項服務皆無排他性，兩項服務也提供簡易的教學，讓無經驗的使用者也可以透過教學引導，輕鬆地使用服務，具有親和力(Accessibility)。

目前國內市場上尚無過濾中文不當字詞的工具，而本團隊開發留言守門員致力於打造友善的上網環境，預計應用於教育領域，學校電腦可安裝此瀏覽器擴充套件，保護每一位學童瀏覽網頁的內容。更進一步推廣到家庭，家長也可以安裝此瀏覽器擴充套件，不用時時刻刻陪在兒童身邊，可以放心地讓兒童上網。

網路霸凌稽查員也是一款具教育性質的服務，本團隊蒐集網路資訊、整合有關網路霸凌的法條以及案例，目的在於提供使用者正確的資訊，讓使用者透過與網路霸凌稽查員的 LINE BOT 對話，認識網路霸凌、了解網路霸凌的求助管道。本團隊期望使用本服務的使用者，獲得正確的觀念後，也可以向身邊的親朋好友傳達，藉由多數人努力，減少網路霸凌事件的發生。

二、SWOT 分析

1. Strength 優勢

- 留言守門員與網路霸凌糾察員皆屬於免費應用，皆有詳細使用教學。
- 留言守門員的黑名單字詞庫為可擴充的，未來若有其他侮辱他人字詞出現也不必擔心會有受限。

2. Weakness 劣勢

- 網路霸凌糾察員 - 網路霸凌行為衡量 此功能上，在模型訓練資料上之筆數因時間限制仍不足，可能造成輸出結果不夠準確。

3. Opportunity 機會

- 近年來，明星與青少年受網路霸凌影響，造成嚴重身心問題，最後選擇結束生命的血淋淋案例屢見不鮮，政府與民眾開始注意網路霸凌所帶來的負面影響，每一個網路使用者都應該重視網路霸凌的傷害以及避免成為網路霸凌的加害者。

4. Threat 威脅

- 在網路霸凌發生前，人們不會想杜絕霸凌環境，因此需多加強留言守門員的推廣。

- 當人遇見網路霸凌時，若是想求助於人，可能優先自行尋找網路，不會想直接使用本團隊設計之網路霸凌糾察員。

三、外部環境分析

(一) 網際網路成形

隨著時代推進，網際網路在全球已然成形，智慧型手機的出現大大降低了使用網際網路的技術和設備門檻。隨著手機用戶的增長，上網的比例也繼續增加，據 We Are Social 和 Hootsuite 的統計報告數據顯示，從 2021 年 1 月，全球有 46.6 億人使用網際網路。目前，全球網際網路普及率為 59.5%。由於使用人數的增長，網路霸凌事件的發生也趨近頻繁。

(二) 兒童網路社交使用時長增加

隨著上方所述之網際網路成形，兒童使用網路的年紀與時長正在改變。據兒童福利聯盟結果顯示，兒童在 2021 年的平均每週網路的使用時間為 42.7 小時，相較於 2020 年的 27.2 小時以及 2014 年的 12.9 小時，明顯高出許多，並且在裏頭顯示，有超過兩成的兒少曾遭受網路霸凌。

(三) 小結

由以上的分析結果得知，本團隊發展之留言守門員及網路霸凌糾察員具極大功效，除了留言守門員能夠阻擋網頁的霸凌文字，大幅降低網路霸凌的事件發生，若是自身真實遇見相關事件，也能根據網路霸凌糾察員的指示，透過裡頭的教學管道進行求助，打擊網路霸凌。因此，本團隊之作品適合推廣至社會，若是能進一步推展至校園環境，也將有效杜絕校園網路霸凌的事件。

肆、團隊成員與分工說明

系級	成員	負責項目
法律三 D	林汪鴻宇	法律相關背景知識整理、使用者需求設計
巨資三 A	吳貫綸	LINEBOT 製作、編寫作品構想書
巨資三 A	黃慧蓮	不雅字詞蒐集、判決書蒐集、影片製作及配音、圖文製作、Bert 模型製作
巨資三 A	吳東儒	Chrome 擴充套件設計及製作、編寫作品構想書
巨資三 A	鄭雅綿	不雅字詞蒐集、判決書蒐集、影片製作及剪輯、Bert 模型製作

巨資三 A	張培莉	LINEBOT 製作及設計使用者頁面、編寫作品構想書
-------	-----	----------------------------

伍、未來展望

（一）與教育相關單位合作

將此功能推廣至教育系統上，透過留言守門員與網路霸凌稽查員兩大設計，使學生擁有更好的上網環境，讓網路霸凌對於孩童的傷害降至最低，同時了解到網路霸凌的可怕之處，避免成為加害者。

（二）提高 LINE BOT 網路霸凌行為衡量功能 準確度

此功能由機器學習判斷有罪率，未來將持續增加模型訓練資料集，進而提高有罪率判斷之準確度。

（三）持續更新 留言守門員 資料庫詞彙

不雅詞彙將隨著時間的推進，在不同的世代有不同的用法，本團隊未來將定期透過網路爬蟲及 LINE BOT 使用者提供之不雅詞彙，持續更新網路不雅詞彙，讓留言守門員的資料庫維持在最新狀態，保持良好的過濾效果。

陸、參考資料

- [1] 紀惠英、郭依潔、范熾文（2021）。國中生網路霸凌行為及其相關因素之研究。學校行政，134 期（2021/07/16）。
- [2] 吳明隆、陳明珠（2012）。霸凌議題與校園霸凌策略。臺北市：五南。
- [3] 張淑亭（2015）。校園霸凌之初探。諮商與輔導，357，54-57。
- [4] 研究調查報告內容:研究調查報告:兒盟資源::兒福聯盟 (children.org.tw)