附件1-1：系統概述文件

編號：（主辦單位填寫）

專題名稱：EZRater

校名與科系：國立中央大學 資訊管理學系

團員成員：蘇勁安 何忻 陳怡均 羅曼禎 周嘉城

1. 前言

近年來，由於網路快速建設，行動裝置普及，人們可以無時無刻上網瀏覽資訊，網路成了多數人選擇接受資訊的主要渠道，在資訊繁多的狀態下，人腦無法處理大量資訊，因此越是誇大的文章，越能吸引人查看，網路上的資訊變得濫情、煽動，即使是長期在傳統媒體業的公司轉移到新媒體上，為了點擊率也逐漸屏棄媒體素養，提供強刺激性、帶有筆者情感的內容。同時網路的蓬勃發展也帶來社群媒體的使用者大幅增加，網紅成了新興職業，許多人寧願相信所謂的網紅撰寫的內容，對多數人來說網紅是所謂的素人，應該會帶給人更親民的內容，但網紅多數未經專業訓練，文章很容易變得濫情，無意中攝取大量無用資訊，造成腦內負荷。

因此EZRater旨在解決上述困擾，提供使用者最輕便簡潔的資訊，使人快速掌握文章要點。由於近期網路購物興盛，多數人在購物前會先找尋相關留言或是心得來補充產品資訊，進而決定要不要購買，因此EZRater從科技產品下手，也是目前多數人在購買前會先找尋資料的產品類別，提供使用者不帶情感的關鍵詞，降低使用者的搜尋時間及腦力消耗，利用最少資訊獲得最大效益。

1. 創意描述

網路文章充斥著無用資訊，人們在狹小的螢幕上閱讀就足以使人費神，再加上資訊氾濫，人腦處理過量資訊後就很難再專注，進而影響做事效率，因此我們透過資料抓取，大量資料前處理，AI協助辨別詞性、詞特徵，去除煩冗的文字，提供使用者2、3個主要形容詞，再對詞作評分，使使用者可以在幾秒內了解產品特性。

1. 系統功能簡介

* 定時動態爬蟲

透過設計的爬蟲技術，能夠避開網站限制，將產品文章內容截取至資料庫，再對其進行處理。也因科技產品日新月異，網路文章也是不斷更新，因此透過定時爬蟲來得到最新的內容，即時更新產品狀態。

* 資料處理

將文章從一塊原石變成價值高的黃金就需仰賴資料處理，因此我們設計了一套中文資料處理流程，去除標點符號、中文使用者常用的詞 （e.g. ‘的’）

* AI判斷關鍵詞

利用為科技產品產生的AI對文章判斷，擷取關鍵詞，以達到不帶個人偏頗的判斷，提供客觀關鍵字。

* 產品過濾

系統提供快速的過濾系統，使使用者能夠輕易且迅速找到自己所需資訊，降低使用者需花費的時間。

1. 系統特色

* 簡潔易懂

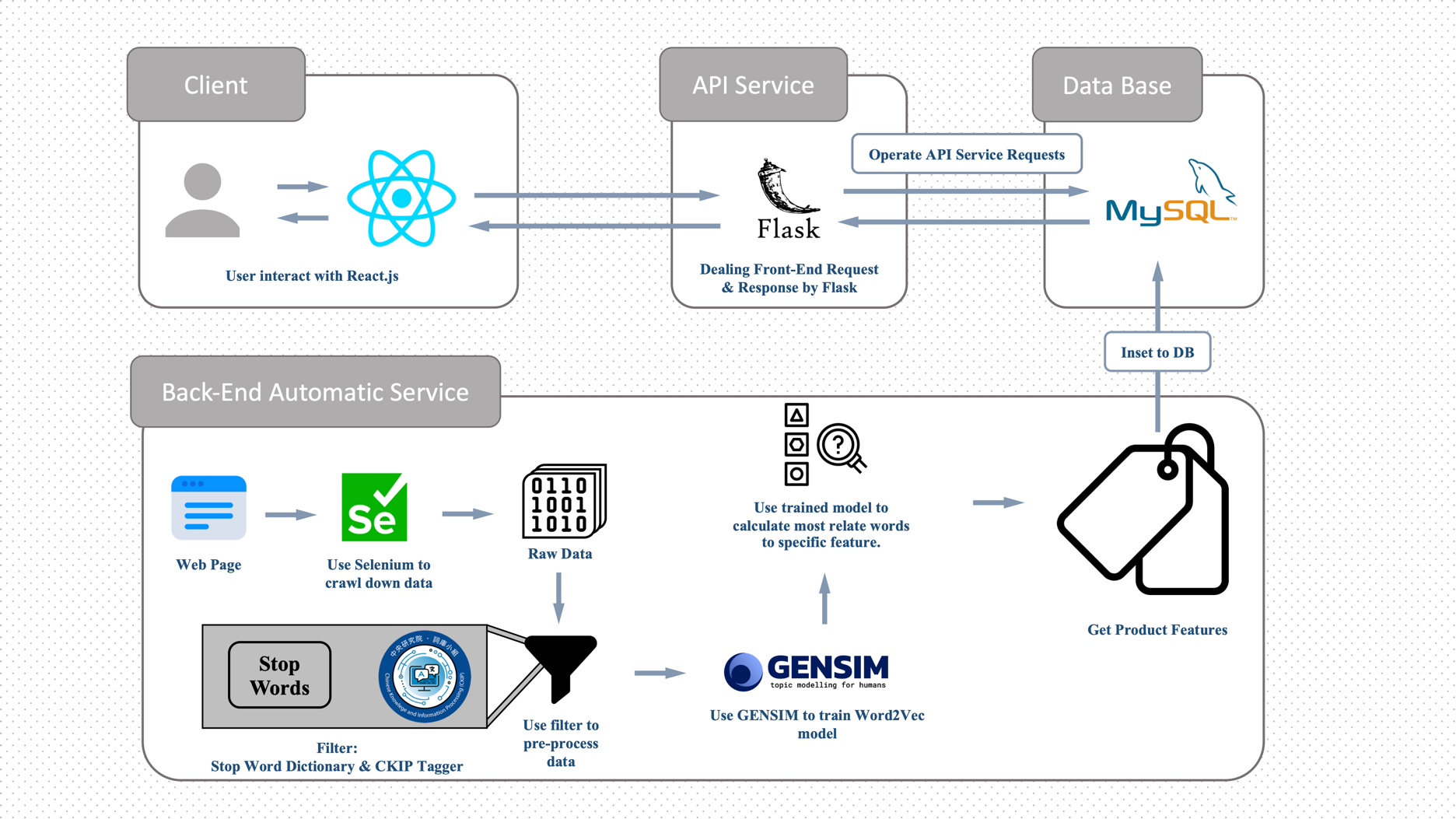
系統旨在提供最少資訊獲得最大效益，因此在網頁呈現介面上下功夫，提供不分男女老少皆可快速上手之網頁內容，且能快速掌握資訊，節省時間。

* 客製模型

透過全為科技產品的文章訓練AI，得出適用於科技產品的模型，進而增進判斷精確度。

* 定時更新

科技產品生命週期短暫，評論文章日新月異，因此系統定時更新能夠提供即時最新的資訊給使用者，進而達成提供最少資訊創造最大價值的效益。

1. 系統開發工具與技術
2. React.js: 提供使用者良好的戶動界面。
3. Flask: 高效率與資料庫進行C/R/U/D互動。
4. MySQL: 建立關連式資料庫，達到系統需過濾資料等關聯需求。
5. Selenium: 動態爬蟲工具，模擬人為操作，有效避開網頁限制抓取資料。
6. CKIP Tagger: 中研院切詞工具，提供精確的切詞，以利系統找出關鍵詞。
7. GENSIM: Word2Vec工具，協助計算詞相似度，以求出產品關鍵詞。
8. 系統使用對象

對科技產品有需求且在找尋相關資訊以利協助決定購買之人

1. 系統使用環境

適用於具有瀏覽器可上網之裝置

1. 結語

此系統旨在針對新聞、社群媒體對3C產品的報導，擷取其形容此產品的關鍵詞，透過AI運算，對關鍵詞進行情緒分析，提供報導對產品的評論。