

# 專題與研究能力

## 目錄

一、聞到有先後：基於多源文本的段落生成以及資料加值系統實作 CHRONONEWS: MULTI-SOURCE TEXTUAL PARAGRAPH GENERATION AND VALUE-ADDED SYSTEM IMPLEMENTATION. ....	2
二、PROPOSAL : ENHANCING SEARCH ACCURACY FOR UNIVERSITY REGULATIONS .....	5
三、成大夏漢民太空中心研究助理證明.....	9

# 國立成功大學

## 專題報告

聞到有先後：基於多源文本的段落生成  
以及資料加值系統實作

ChronoNews: Multi-Source Textual  
Paragraph Generation  
and Value-Added System Implementation

專題生：鄭宇辰、陳冠廷  
指導教授：高宏宇 教授

中華民國一十二年五月

# 聞到有先後：基於多源文本的段落生成以及資料加值系統實作

## ChronoNews: Multi-Source Textual Paragraph Generation and Value-Added System Implementation.

指導教授：高宏宇

專題成員：陳冠廷、鄭宇辰

開發工具：ChatGPT，Python

測試環境：Google Colab

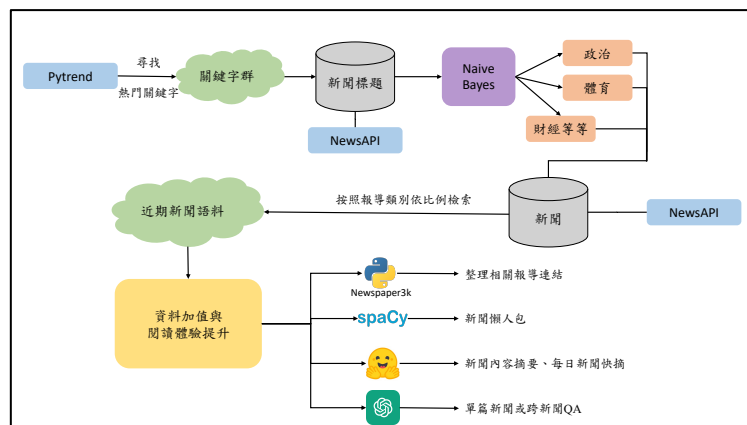
### 一、簡介：

近年來，台灣人越來越難從新聞中獲得國際新聞。我們希望開發一個整合平台，帮助大家更輕鬆的接觸國外的新聞資訊。

我們以 Python 作為主要開發語言，透過 PyTrend、NewsAPI、Newspaper3k 等套件擷取每日熱門關鍵字及其近日相關新聞（First Layer）。使用 Kaggle 資料集訓練 NLTK 的 Multinomial Naive Bayes Classifier；將相關新聞標題簡單分類，再根據類別給予不同比例來源的新聞（Second Layer）。如，關鍵字出現某好萊塢演員，近期報導大多針對其社會事件官司，則選用自“buzzfeed”，“ign”等娛樂平台的新聞會比來自“BBC”，“CNN”的新聞還少。

我們以 Second Layer 的文本製作 corpus，以 SpaCy 識別 corpus 的命名實體。根據《懶人包超文本敘事設計分析：順序與結構元素的討論》（陳雅惠，2018）所列舉的懶人包結構樣式，我們將人／組織的行為（plot）透過 Hugging Face, OpenAI 提供的套件及相關工具重新生成標題，並且針對某熱門關鍵字的 corpus 進行摘要。最後將摘要及其翻譯（使用 Googletrans 套件）並陳。

透過 ChatGPT 協助製作呈現網站，以 Flask 作為後端，html、css、js 作為開發語言，具有響應式網站的設計，讓使用者在不同裝置上都有良好的體驗。



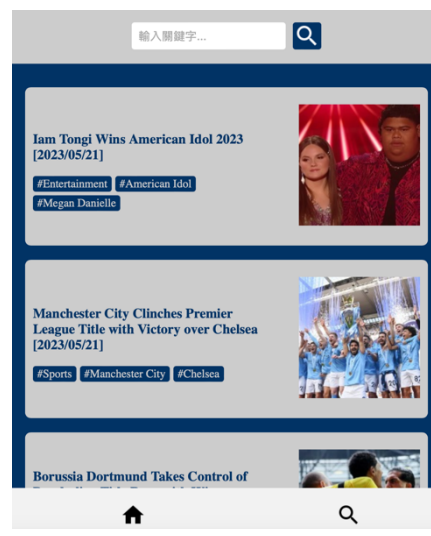
圖一：系統架構

## 二、測試結果：

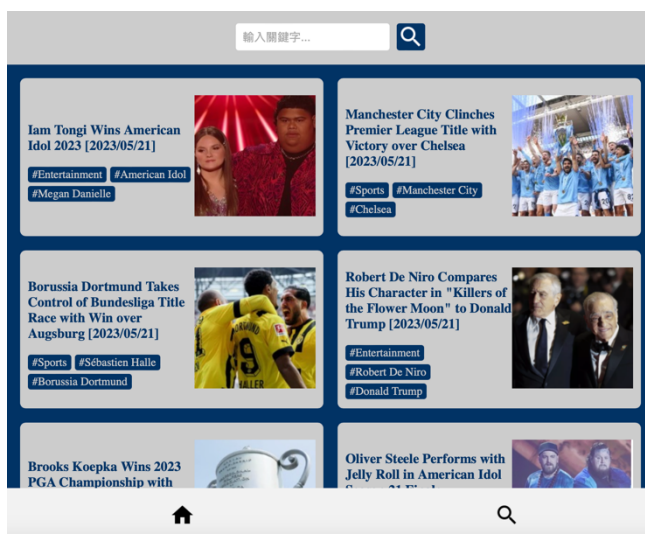
- 首頁的新聞類別能夠正確顯示，並且透過圖形大小可視化，使用者可以清楚看出今日熱門關鍵字在各類別的比例大致如何。
- 搜尋頁的標題以及 tag 出自於 SpaCy、Hugging Face 以及 OpenAI 套件。
- Multinomial Naive Bayes Classifier 可以透過 First Layer 的標題正確判斷新聞內容應該歸屬於哪一個類別。
- 網站介面可以符合各種比例的視窗、裝置，讓使用者獲得最舒適的操作體驗。
- 歡迎大家來現場看看我們的懶人包內容。



圖二：系統首頁



圖三：系統搜尋頁



圖四：系統搜尋頁（演示 RWD 效果）



[repo](#)

# 國立成功大學

## 研究提案

Enhancing Search Accuracy for University Regulations

學生：鄭宇辰

中華民國一十二年九月

# **Proposal : Enhancing Search Accuracy for University Regulations**

## **Title**

Enhancing Search Accuracy for University Regulations

## **Introduction**

In light of a recent incident where I sustained an injury on university property due to adverse weather conditions, I embarked on a search for relevant regulations governing such occurrences. It came to my attention that the existing regulation search system in place is notably deficient. Notably, I discovered that it only yields results when specific keywords like '傷' (injury) are entered, rendering it ineffective for queries involving related terms such as '致傷' (inflict injury).

This experience has served as a catalyst for me to propose the development of an easily accessible and efficient regulation search system tailored to the needs of both university staff and students at National Cheng Kung University (NCKU).

## **Statement**

The primary objective of this project is to enhance the precision and effectiveness of retrieving university regulations, thereby reducing the gap between university staff, students, and the regulatory framework.

## **Approach/ Methods**

To achieve the project's objectives of improving the accuracy of searching university regulations and enhancing accessibility for staff and students, the following approach and methods will be employed:

### 1. Data Collection:

- Gather existing university regulations and associated documents, ensuring a comprehensive and up-to-date database of regulations.

### 2. Named Entity Recognition:

- Implement NER algorithms to identify and extract key entities, such as regulation titles, dates, and departments, to enhance search relevance.

### 3. Keyword Expansion:

- Develop a keyword expansion mechanism to identify synonyms and related terms, allowing users to find regulations even when using variant keywords.

#### 4. Indexing and Search Engine:

- Create an efficient indexing system to organize and store regulations and metadata for fast retrieval.
- Implement a robust search engine that incorporates NLP and machine learning techniques to improve the relevance of search results.

#### 5. Machine Learning:

- Explore the use of machine learning models, such as topic modeling or classification algorithms, to categorize regulations into thematic areas for better organization and retrieval.

By implementing these approaches and methods, the project aims to provide a more accurate, efficient, and user-friendly system for accessing university regulations, ultimately bridging the gap between staff, students, and the regulatory framework.

### **Implications of Research**

The successful execution of this project is expected to have far-reaching implications and benefits for National Cheng Kung University (NCKU), its staff, students, and the broader academic community. These implications can be summarized as follows:

**1. Enhanced User Experience:** The implementation of an advanced and user-friendly regulation search system will significantly enhance the overall user experience for both staff and students. They will be able to quickly and efficiently access the information they need, thereby improving productivity and reducing frustration.

**2. Improved Accuracy and Relevance:** The use of natural language processing (NLP) techniques and advanced search algorithms will result in more accurate and relevant search results. This will ensure that users can find the regulations they need, even when using variant keywords or synonyms.

**3. Accessibility:** The user-friendly interface and enhanced search capabilities will promote inclusivity by making it easier for all users.

**4. Transparency and Compliance:** The project will contribute to greater transparency in regulatory compliance by ensuring that university regulations are easily accessible and understandable. This will assist in ensuring that all members of the university community are aware of and adhere to relevant rules and policies.



**5. Data-Driven Decision-Making:** The project will generate valuable data on user search behavior and regulatory needs. This data can inform data-driven decision-making processes within the university administration, leading to more informed policy adjustments and improvements.

In conclusion, this project has the potential to positively impact the university's operations, governance, and overall academic environment by improving the accessibility, accuracy, and user-friendliness of its regulatory framework. The implications outlined above reflect the project's broader significance for NCKU and its community.

## **Reference**

Nay, J. (2021). Natural Language Processing for Legal Texts. In D. Katz, R. Dolin, & M. Bommarito (Eds.), *Legal Informatics* (pp. 99-113). Cambridge: Cambridge University Press. doi:10.1017/9781316529683.011

## **Repo**



[NCKU\\_RQA](#)

# 國立成功大學

## 研究助理證明

### Gemini 立方衛星計畫

學生：鄭宇辰

中華民國一一二年九月

# 成大夏漢民太空中心研究助理證明



國立成功大學  
National Cheng Kung University

窮理致知 SEEK FOR TRUTH; TOIL FOR GOOD

## 夏漢民太空科技中心

### 參與 Gemini 立方衛星計畫證明書

學號(Student ID): F74096255

學系(所): 資訊工程學系

學生鄭宇辰 中華民國89年11月8日生

於本中心 擔任 Gemini 立方衛星計畫

指令與資料處理次系統(Command & Data Handling (C&DH)) 小組 組長

負責開發立方衛星飛行軟體(On Board Computer(OBC))

特此證明

中心主任

教授兼夏漢民太空  
科技中心主任 陳炳志



中華民國 112 年 9 月 27 日