

搜 LAW: 裁判書搜尋系統



一、簡介

相較於外語法律資料庫，現行中文法律資料庫發展遲滯，最大宗使用之法律案件查詢系統仍為司法院裁判書查詢系統，然而，司法院裁判書查詢系統為匹配搜尋，並依照判決日期排序顯示所有包含查詢詞的判決書，導致重要的裁決被淹沒在眾多搜尋結果中；本專題之目標即為解決上述痛點。

二、特色

引入相關性搜尋

搜LAW賦予案件相關性進行排序，讓使用者能優先看到重要案件，減少瀏覽無用資料的時間。

相關性主要考量三個因子

- 查詢字出現次數:** 依循一般文件檢索，查詢字出現的次數越高，該案件被認為與查詢字閱相關。
- 查詢字位置分佈:** 從法律意義角度考量，段落查詢字密度越高，表示其相關法律概念被集中討論。
- 所屬法院層級:** 審級較高的法院可事實審或法律審下級法院的裁判，且最高法院的裁決具指標性。

重要段落預覽

不同於司法院裁判書系統和Lawsnote僅呈現查詢字首次或最後出現的段落，搜 LAW 能判斷查詢字在整篇判決中的分布狀況，找出密度最高的段落呈現，讓使用者在點開案件完整內文瀏覽之前就能知道是否為目標案件。

1 - 臺灣高等法院105年度 訴字第48號 判決 2017-06-06 侵權行為損害賠償

④又被告協紡企業社另以系爭鑑定意見未斟酌原告於系爭刑事案件103年11月16日警詢時稱：為閃避系爭汽車再與系爭貨車發生系爭車禍等情，系爭鑑定意見認原告就系爭車禍並無肇事原因，並不可採云云。然系爭車禍並非於原告閃避系爭貨車之際立即發生，已如上述，且原告於當日警詢中就系爭車禍發生經過即稱：「一開始我沿樹林區中正路往板橋方向直行，行經事故地點前，我一直行駛在靠路邊，但是（右）前方有一輛黑色自小客車（按

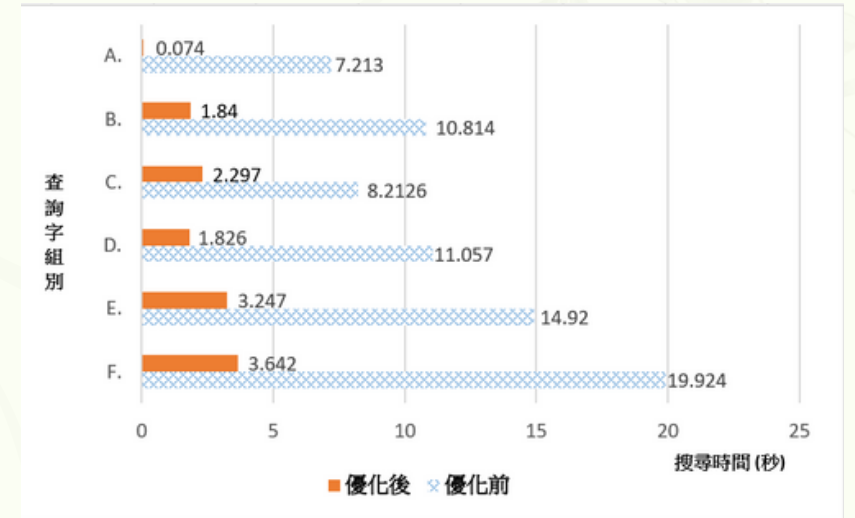
搜尋"汽車/10車禍&板橋"的結果列表與重要段落預覽
(至少一句符合汽車10個字內出現車禍，並且至少出現一次板橋)

創新語法設計

除了常見的布林搜尋語法，搜 LAW 是中文法律資料庫中，引入兩種語法「/S」、「/number」的首例，分別限制查詢字之間的文句位置、間隔距離，讓使用者能更具體描述目標案件具有的特質。

三、關鍵技術

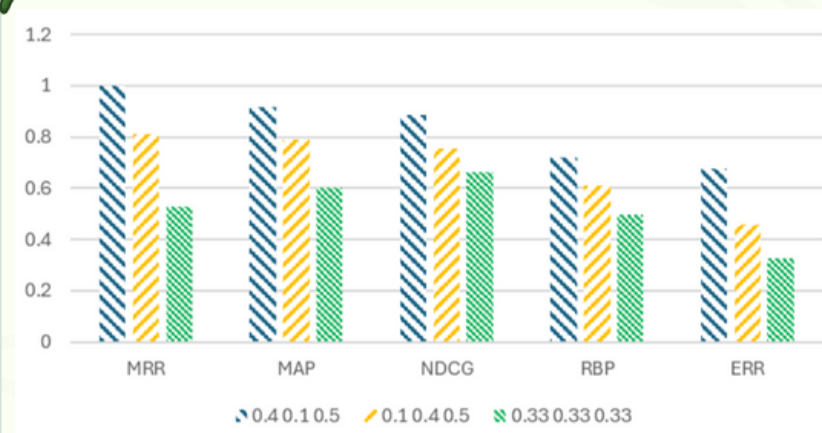
- 1.資料垂直表示法
- 2.常用欄位索引
- 3.建立實體化視圖
- 4.減少服務端跟資料庫的互動次數



四、參數實驗與驗證

參數

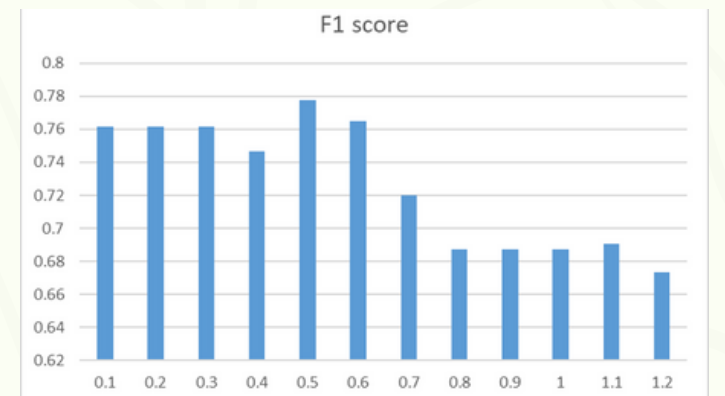
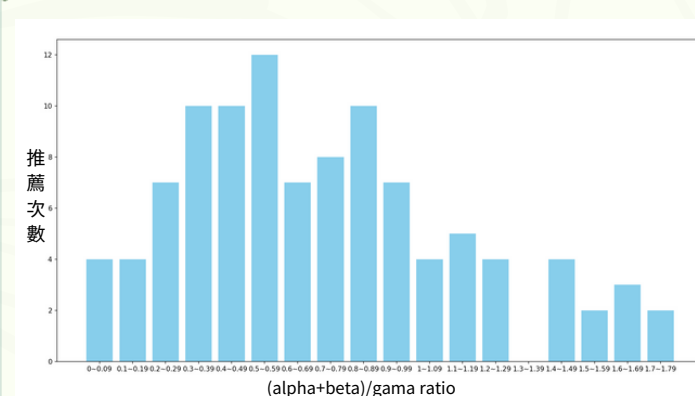
(查詢字出現的次數、重要段落內查詢字的密度、裁決法院層級)



透過三組參數的評估
得出因子重要性：
法院層級>次數>密度

注意到密度跟次數在概念上有雷同之處，而此觀察延伸出一新問題：「次數」跟「法院層級」在什麼樣的比例下，可以產生更好的排序結果？為了探究這個問題，於是從兩方面著手：線性迴歸和參數掃描分析。

實驗驗證



在數據不充裕的情況下，線性迴歸推薦之參數變化很大：改以其相關統計值呈現，多組回歸結果認同「法院層級」的權重應至少佔比0.5，對於最佳佔比卻沒有明顯結論。

以參數掃描分析應證線性迴歸結論：得出當「次數」與「法院層級」的權重比值在0.5時表現最好（結果一致）。

系統	相關性搜尋	重要段落預覽	搜尋語法支援/S、/N	資料庫語系	搜尋語意規則	其他
司法院裁判書系統				中文	無優先度	法律名詞解釋
Lawsnote	✓	✓		中文	無優先度	特定身分搜尋
Westlaw	✓	✓	✓	英語	符號有優先順序	West Key Number System
LexisNexis	✓	✓	✓	英語	符號有優先順序	三種模式：圖像、詳細、互動圖表
搜LAW	✓	✓	✓	中文	無優先度	可自動化處理批次新增案件