

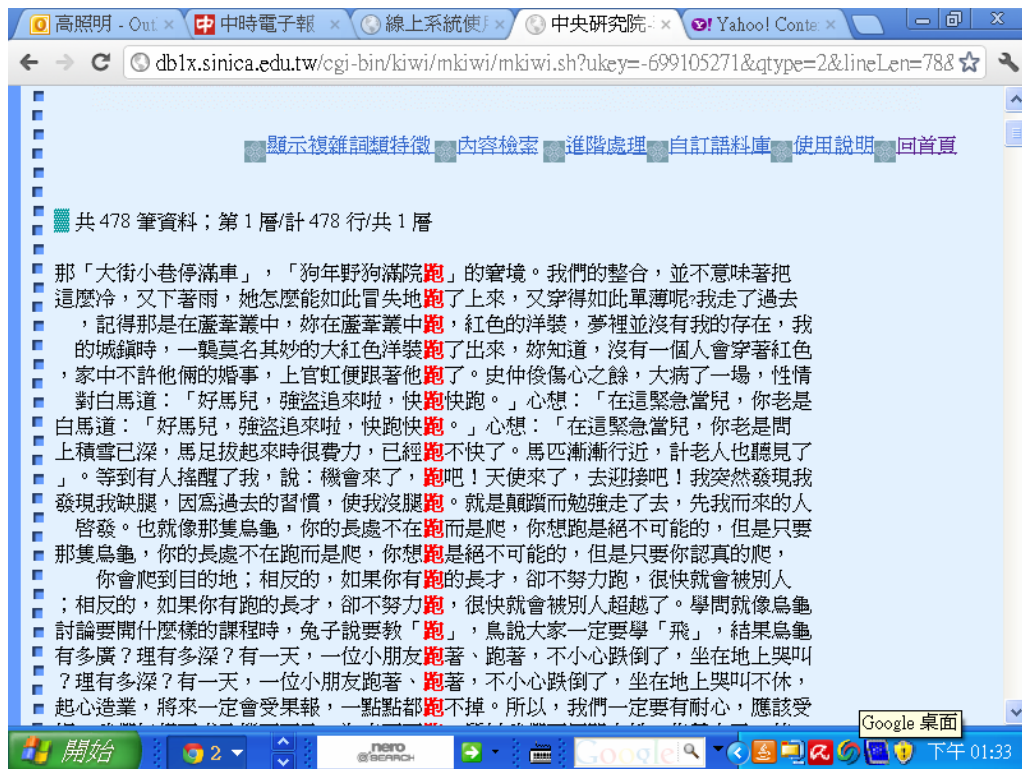
柒、中文語料庫的比較與分析

目前正體字的中文語料庫最具有代表性的兩個，分別是中研院詞詞知識庫小組的現代漢語平衡語料庫及 Sketch Engine。前者包含 1 千萬詞，後者則包含 4 億 5 千多萬詞，兩者除了語料的數量之外，最大的差別在於 Sketch Engine 從分詞到詞性標記到語法依存分析完全是程式自動處理，且沒有經過人工檢查，現代漢語平衡語料庫除了程式處理之外，則經過人工多次的檢查，正確率較高。至於檢索的方式，兩者都支援關鍵詞檢索，也可以搭配語詞性檢索，也都可以檢索關鍵詞左邊或右邊若干詞之內的搭配語。現代漢語平衡語料庫並支援各種重疊詞的檢索，以及依據文類來檢索。



圖十三 現代漢語平衡語料庫的檢索介面

<http://db1x.sinica.edu.tw/kiwi/mkiwi/>



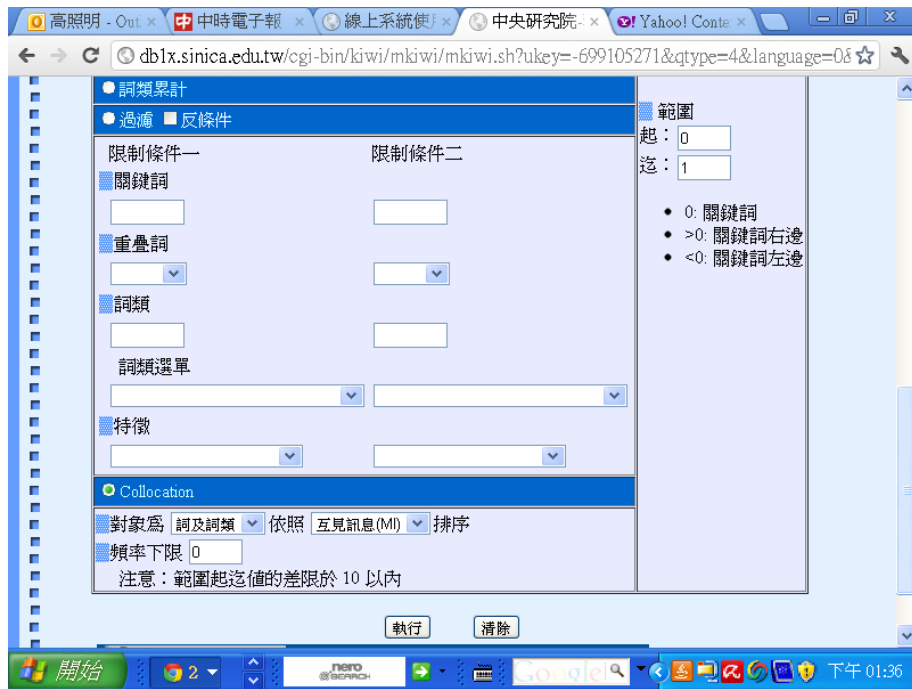
圖十四 現代漢語平衡語料庫檢索輸出結果

<http://db1x.sinica.edu.tw/kiwi/mkiwi/>



圖十五 現代漢語平衡語料庫檢索輸出結果(顯示詞性)

<http://db1x.sinica.edu.tw/kiwi/mkiwi/>



圖十六 現代漢語平衡語料庫進階檢索介面

<http://db1x.sinica.edu.tw/kiwi/mkiwi/>



圖十七 現代漢語平衡語料庫進階檢索功能中的搭配語的檢索

<http://db1x.sinica.edu.tw/kiwi/mkiwi/>



圖十八 Sketch Engine 詞彙速描介面

<http://www.sketchengine.co.uk/>



圖十九 Sketch Engine 詞彙速描輸出結果

<http://www.sketchengine.co.uk/>