

# 字級及掃描解析度與中文光學文字辨識系統辨識率關係之研究——以蒙恬認識王專業版為例

指導教授：陳昌郎 老師

研究學生：施軍宇 辛炳宏

黃英銓 羅子堯

## 摘要

現今這資訊爆炸的社會中，早期記錄保存文字資料的方式已不再適合，尤其在現今講求效率、速度及精確的工作品質下，我們會透過中文光學文字辨識系統將紙本資訊轉為文字檔案，將文字作數位歸檔、編排及整理，除了可防止印刷品老化流失甚至失傳，還可將文字資訊作更有彈性的運用，例如書籍翻新、網路傳播，甚至強化學術界的資訊交流機制。

本研究主要的目的在於研究市售的專業版中文辨識軟體，對於一般書籍中文字辨識能力，本研究採用蒙恬認識王專業版，對常見內文字進行辨識能力測試，並探討其『辨識能力』、『掃描解析度』、『字級大小』以及與『掃描解析度與字級乘積大小』的關係，藉此瞭解最佳辨識效果的字級與掃描解析度組合，例如當辨識的文件是 12 級字時，以 300DPI 的解析度掃描後再進行辨識，即可得到最好的辨識率，如此一來可讓我們在選擇掃描解析度時有客觀的依據，也可節省文稿掃描及檔案存取時間。

在早期的研究中，可以得知中文光學文字辨識系統在於字級、掃描解析度和辨識結果有著顯著的相關聯性，因此本研究將同一字型，不同字級影像文件加以辨識，取得各字級最好之辨識率，這樣一來當我們拿到有需辨識的文件時，便可以直接對文件加以判斷，得到最好的辨識效果。

## 壹、緒論

### 一、問題陳述

從人類開始會思考及記錄的那一刻起，圖與文字便存在於我們日常生活中，藉著圖與文字的流傳及對於歷史文化的傳承，刻畫出一部人類文明的發展史。為了延續文化的傳承、新聞的傳播、知識的傳遞及

學術性文章等，將印刷品上的文字及其他載體上的文字，轉換成一種可供辨識的文字類型，是未來不可缺的趨勢。

現今這資訊爆炸的社會中，早期記錄保存文字資料的方式已不在適合，尤其在現今講求效率、速度及精確的工作品質下，我們所需要的是一套能夠減少人力資源、提高整體效率、快速記錄儲存、降低錯誤率的光學