

丹青

中英日文文件辨識系統

4.5 版



S/N:107450-02-01-W-C-070502-01

版權

版權所有：力新國際科技股份有限公司

初版日期：2002年8月

本書版權為力新國際科技股份有限公司所有，未取得書面授權，不得將本書內容以任何形式複製、翻印，或以電子檔案方式儲存、運輸。

程式光碟片中的範例圖檔，僅供個人展示、製作及簡報之用，不得用於商業拷貝、銷售或流傳等其他用途。

力新國際科技股份有限公司保留隨時更新本書內容之權利。

商標

IBM PC是International Business Machines Corporation 的註冊商標；MS Windows 98/Me/2000/XP是Microsoft Corporation 的註冊商標；Pentium、MMX是Intel Corporation的註冊商標。

本書提及的產品名稱皆為其所屬公司的註冊或未註冊商標。

www.newsoftinc.com

www.newsoft.com.tw

www.newsoft.co.jp

de.newsoft.eu.com

newsoft.net.cn

目錄

第一章

介紹.....	1
丹青的功能與特色	1
系統需求.....	4
硬體需求	4
軟體需求	4
系統安裝.....	4

第二章

基本概念.....	6
操作流程圖.....	6
認識丹青介面	8
原稿影像模式：	8
全頁圖文模式：	12
文稿編輯模式：	14
改變螢幕配置	17
縮放影像顯示比例	17
頁面移動及訊息查詢.....	18
使用線上說明	18
設定系統設定值.....	19

第三章

輸入影像.....	21
-----------	----

掃描與開啓	21
影像處理.....	22
旋轉文件影像	22
清除雜點及補漏白	24
切除	24
反白功能	24

第四章

文件辨識.....	25
設定辨識字集	25
設定辨識區域	26
設定版面格式	26
欄位設定	27
排列設定	27
表格設定	27
資料設定	27
設定版面範本	28
版面分析.....	29
自動分析版面	29
手動設定版面	29
改變辨識順序	29
儲存版面	30
辨識文件.....	30
選擇校對詞庫	31
設定校對詞庫	31
使用詞庫校對	32

自動辨識文件	33
自動辨識	33
設定自動範本	34

第五章

校對文稿	36
放棄辨識	36
校對文稿	36
再辨識	38
擦拭雜點再辨識	38
修補影像再辨識	39
文字切割再辨識	39
文字合併再辨識	39
文字行切割再辨識	40
文字行合併再辨識	40
區塊再辨識	40
區塊結合再辨識	41
區塊分開再辨識	41
學習新字	42
刪除學習字	42
學習字庫設定	42

第六章

輸出檔案	44
儲存辨識前的影像文件	44
儲存辨識後的圖文和表格	45

儲存常用的版面格式.....	46
列印.....	47
傳送.....	47

第七章

圖文辨識範例.....	50
辨識含有圖形及文字的文件.....	50

第八章

英文辨識範例.....	54
辨識英文文件.....	54

第九章

表格辨識範例.....	58
辨識一般表格影像.....	58
辨識暗線表格影像.....	62

第十章

自動辨識範例.....	67
自動辨識文件.....	67

附錄 A

用語說明.....	71
-----------	----

附錄 B

功能表.....	73
----------	----

附錄 C

功能圖示列及下拉式選單	78
-------------------	----

附錄 D

編輯工具箱圖示	80
---------------	----

附錄 E

掃描的建議	82
如何改善辨識品質	82
圖例一：掃描解析度建議	83
圖例二：標準	84
圖例三：太濃	84
圖例四：太淡	85

第一章

介紹

丹青中英日文文件辨識系統能讓您輕鬆且快速地將影像檔轉換成可編輯的文字檔案。其所能辨識之內容包括繁體中文、簡體中文、日文、英文、阿拉伯數字及含表格的文件。辨識後的文字檔案所佔的記憶體空間遠較未辨識前的影像檔案小。在經過校對後即可儲存成 TXT、RTF、DOC、XLS、SLK、CSV 等檔案，並且可以在一般的文書處理軟體中被開啓和編輯。您還可以將檔案存成 HTML 格式，透過網路瀏覽器直接開啓。

丹青的功能與特色

➤ 自動辨識

只要一個按鈕，便可自動分析、辨識、校對影像文件，並可轉換成可編輯的文字檔案。

➤ 高辨識速率

在一般的個人電腦上(Pentium III 667)，丹青系統每秒鐘所能辨識的中文字高達 150 個。

➤ 可辨識中文繁體、簡體、日文及純英文文件

丹青採用多辨識引擎架構，可依據您的需要，辨識中文繁體、簡體、日文及純英文文件。

➤ 可處理多頁文稿

可一次處理多達 50 頁的文稿辨識。

➤ 可辨識黑白、彩色之文件

無論是黑白或彩色之文件，皆能獲致極佳的辨識效果。

➤ 多字體辨識及重現

能夠辨識多種印刷字體，如明體、黑體、仿宋體、楷書、圓體、隸書等，並能在辨識後還原成原稿的字體。

➤ 原文重現編輯環境

辨識結果依照原文件的版面格式呈現，方便您校對編輯並節省重新排版的時間。

➤ 自動圖文分離及檔案再編輯

自動分離影像上的圖形和文字，並可將辨識後的結果儲存成 TXT、DOC、RTF 等文書檔案，在一般文書處理器如 Word 中再編輯。

➤ 表格辨識功能

丹青能夠辨識各種表格影像，並可將結果儲存成 XLS、SLK、CSV 等檔案在 Excel 中做進一步的處理，或將檔案儲存成 RTF 格式，在諸如 Word 一般的文書排版軟體內重現原影像上的表格原貌。

➤ 可儲存多種檔案格式

辨識後的結果可儲存成不同之檔案格式，如 TXT、RTF、DOC、XLS、SLK、CSV、HTML 等，方便您做不同的應用與處理。

➤ 直接傳送

可將辨識結果直接傳送至您所指定的應用軟體中再處理，例如可設定辨識結果自動儲存成 HTML 檔案並直接傳送至網路瀏覽器中。

➤ 自動偵測影像傾斜角度

自動偵測影像傾斜角度，並提供旋轉影像之功能。

➤ 自動校對

利用內建的常用詞庫自動校對辨識出的文字，並標示出辨識時所碰到的疑問字，節省您校對的時間。

➤ 候選字功能

提供字形上相似，或語意上前後相連之候選字，可供您輕鬆地更正辨識錯誤的文字。

➤ 學習新字功能

提供學習新字的功能，可將較易辨識錯誤的字元輸入到學習資料庫中，於下一次辨識時使用，以提高辨識之正確率。

➤ 資料交換

可經由中文視窗環境的剪貼簿輸入影像檔或是輸出圖文檔案，而與其它視窗應用軟體交換資料。

➤ 中文橫直排及單多欄辨識

自動分析橫排、直排、橫直排並存以及單欄或多欄的影像文件。若輸入的影像版面過於複雜，您也可以自己設定文字影像的格式，以利於系統做正確的分割及辨識。

➤ 存取版面

提供版面存取功能，方便您將常用的版面格式儲存成丹青的版面檔案 (*.tpl)，並可套用在新輸入的文件影像上，以節省版面分析的時間。

➤ 文稿修改

提供圖文對照(也就是原文字影像與辨識出的文字相互對照)的文稿校對功能，讓您能夠利用鍵盤、候選字或個人詞庫等更正辨識錯誤的文字。

➤ 再辨識功能

提供合/分字、合/分行、合/分區塊、變更區塊屬性再辨識功能，讓您能夠修正錯誤的辨識結果，以便利校對文稿。

系統需求

硬體需求

- ◆ Pentium II (含) 以上的 IBM 電腦或其他相容電腦
- ◆ 64MB 以上的主記憶體(建議 128MB 以上)
- ◆ 螢幕解析度 800×600 Hi-Color
- ◆ 磁碟空間 300MB 以上
- ◆ CD-ROM Drive
- ◆ 支援 TWAIN 介面的掃描器（請參考掃描器的使用手冊）

軟體需求

- ◆ Microsoft Windows 98/Me/2000/XP

注意： Windows 98/Me 不支援[多國語系使用者介面設定]功能。

系統安裝

1. 將丹青的安裝光碟片放入光碟機中。
2. 依照螢幕上的指示完成安裝程序。

注意: 在 Windows 2000/XP 環境下安裝丹青系統前，請先將作業環境回復至電腦預設的語言介面，可避免在安裝過程發生錯誤。

安裝完成後，點選[開始]功能表中的[程式集]，選擇[丹青中英日文文件辨識系統>丹青中英日文文件辨識系統]，開始執行本軟體。您也可以桌面的丹青圖示上快速按兩下，直接進入丹青系統。

注意: 在安裝完成後，[丹青中英日文文件辨識系統]的群組選項中會包含以下選項:

[多國語系使用者介面設定]：您可使用此功能切換丹青系統至下列的設定：繁體中文、簡體中文、日文、自動偵測。請點選[開始]功能表中的[程式集]，選擇[丹青中英日文文件辨識系統>多國語系使用者介面設定]，設定語言後，開啓丹青程式，程式將會切換至您所選擇的語言使用介面。(此功能須電腦作業環境可支援多國語系)。

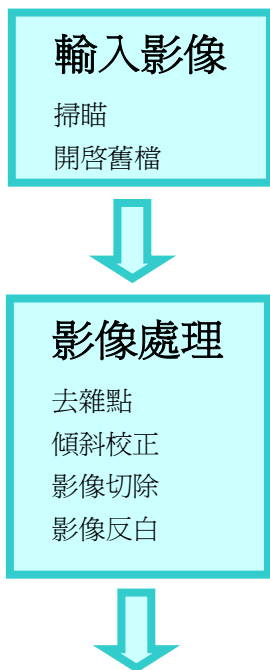
[解除安裝丹青中英日文文件辨識系統]：您可以執行這個程式，完整地移除丹青系統。

第二章

基本概念

本章爲您描述使用丹青中英日文文件辨識系統時，必備之基本概念及操作技巧，如：瞭解丹青系統操作流程及操作介面、改變螢幕配置、縮放影像顯示比例、查看檔案資訊、在不同頁面間移動、設定系統設定值，以及使用線上說明等等。

操作流程圖



辨識文件

自動辨識
圖文辨識
表格辨識
英文辨識



校對文稿

字詞校對
再辨識
學習新字



輸出檔案

儲存圖檔
儲存文字檔
列印
傳送

認識丹青介面

丹青系統包含三種操作介面模式：

原稿影像模式：

在原稿影像模式中，您可以完成辨識前的所有準備工作，如透過掃瞄器或磁碟機輸入影像，使用編輯工具修飾影像，以及最重要的辨識項目設定（如設定辨識語言、文字排列的方式和是否含有表格等）。



原稿影像模式

功能表列

功能表列分類列出所有可供您使用之指令，如開啓、儲存、編輯或辨識檔案等等。

檔案(F) 編輯(E) 查看(V) 格式(I) 分析辨識(R) 文件(D) 說明(H)

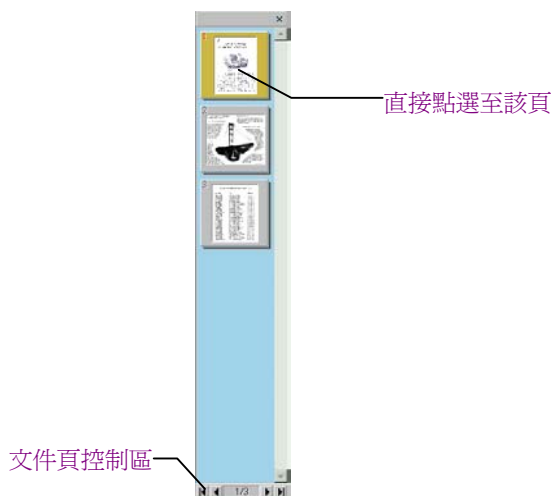
若您要開啓一個功能表，您只要在功能表上按一下滑鼠左鍵或是按一下該功能表上提示的英文字並同時按住[Alt] 鍵即可開啓。

同樣地，如果您要執行某個指令，您只要將游標移到該指令上然後按一下滑鼠左鍵或是按一下該指令提示的英文字即可。

若您不想執行任何指令而要關閉已開啓的功能表時，您只要在該功能表外部任何一處按一下滑鼠左鍵或按[Esc] 鍵即可。

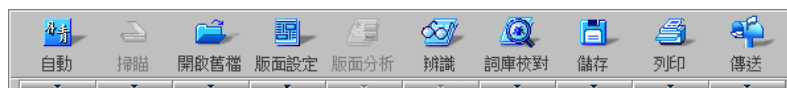
縮影視窗&文件頁控制區

縮影視窗將已開啓之檔案以縮圖方式顯示，可供您直接選取。您也可以利用下方之文件頁控制區的按鍵，移動至您想要的頁面。



功能圖示列

功能圖示列上的圖示可以讓您快速且輕鬆地執行各種功能表指令；每個功能圖示列下方都有一個下拉式的指令選單，可提供您更多的選擇。



按此處叫出下拉式選單

工作區

您可在工作區中執行影像處理的工作，例如轉正傾斜之文件影像，清除雜點，切除不需辨識之部分，使得辨識結果更令人滿意。

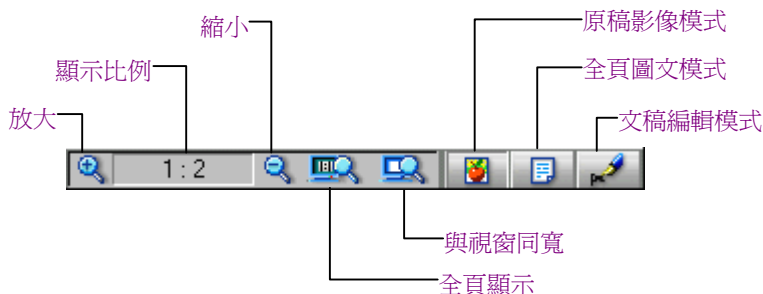
訊息狀態列

訊息狀態列顯示目前游標所在位置的 X 座標和 Y 座標、游標所在物件的相關資訊、文字輸入模式，以及目前所選擇的辨識字集，可按下訊息狀態列中的右方選擇欲使用的辨識字集。



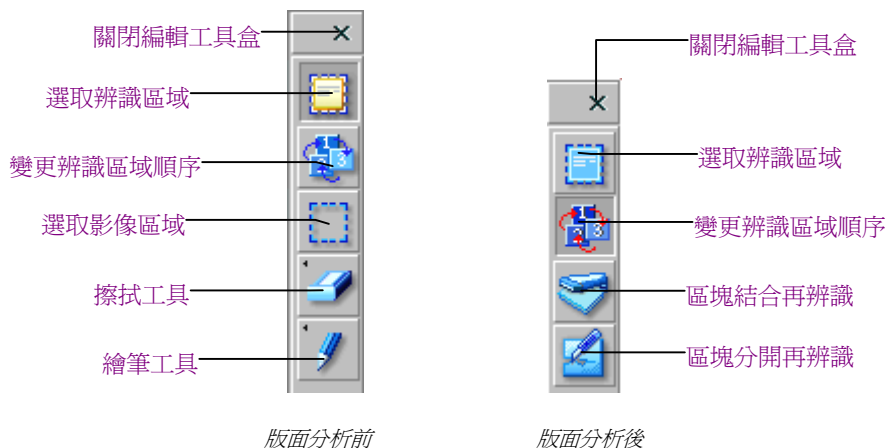
查看工具盒

查看工具盒可讓您縮小或放大影像的顯示比例，並可選擇原稿影像模式、全頁圖文模式及文稿編輯模式。



編輯工具箱

您可以利用所提供之編輯工具，處理辨識過程中各個階段的文件，例如編輯掃描進來之文件影像稿、更動文件之區塊設定、校正辨識後之文稿等等。編輯工具箱也會隨著辨識階段的不同提供您不同之編輯工具。以下為版面分析前後原稿影像視窗所分別顯示的編輯工具箱。



原稿影像視窗在版面分析前後所分別顯示的編輯工具箱



全頁圖文模式：

[全頁圖文模式]可供您查看辨識後的文件全貌，並調整與版面相關的設定，如合併/分割區塊、更改區塊屬性、調整區塊之辨識順序等；您也可以在此模式中直接校正辨識錯的字。




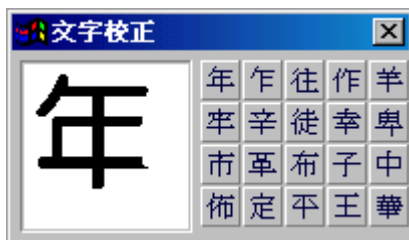
全頁圖文模式

工作區


工作區內顯示辨識後的文件全貌。若您按下文字校對工具，工具區內會出現一些藍底黃字之字樣，其代表的是系統在辨識後所顯示的疑問字；若您按下變更辨識區塊順序工具，各區塊會以紅線框出，並標示出區塊之辨識順序編號。

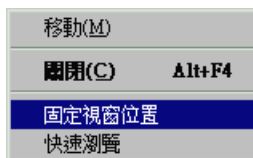
文字校正視窗

當您發現系統辨識錯誤的字時，可選擇文字校對工具，並在該錯誤字上按一下，工具區內將會出現如下的視窗：



文字校正視窗

您可以按一下，在選單中設定文字校正視窗屬性：



文字校正視窗之選單

- | | |
|---------------|--------------------------------------------------------------|
| 移動 | 選擇以鍵盤之上下左右鍵移動文字校正視窗的位置。 |
| 關閉 | 關閉文字校正視窗。 |
| 固定視窗位置 | 在您用滑鼠拖曳文字校正視窗至合適的位置後，可選擇此項以固定視窗位置；若您沒有選擇此項，視窗位置則依您所點選的文字而移動。 |
| 快速瀏覽 | 選擇此項，視窗內的文字將依游標移動而變化，可供您快速瀏覽；若您沒有選擇此項，視窗則會在您點選某文字時才出現該字之候選字。 |

編輯工具箱

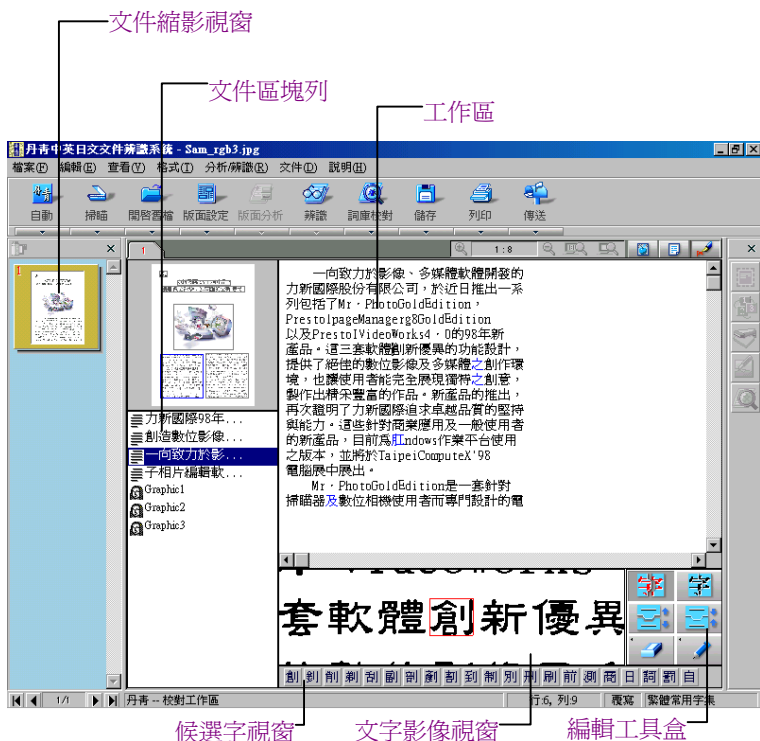
編輯工具箱提供您在全頁圖文模式中，所需的校對編輯工具。



全頁圖文模式中的編輯工具箱

文稿編輯模式：

在[文稿編輯模式]中最重要的工作就是核對系統辨識後的文稿。[文稿編輯模式]可供您查看辨識後的文字，並可針對文字進行校正的工作。丹青系統提供多種校對文稿的功能，如候選字、分/合字再辨識、分/合行再辨識、校對詞庫等，您也可以直接輸入正確的字來進行校對。



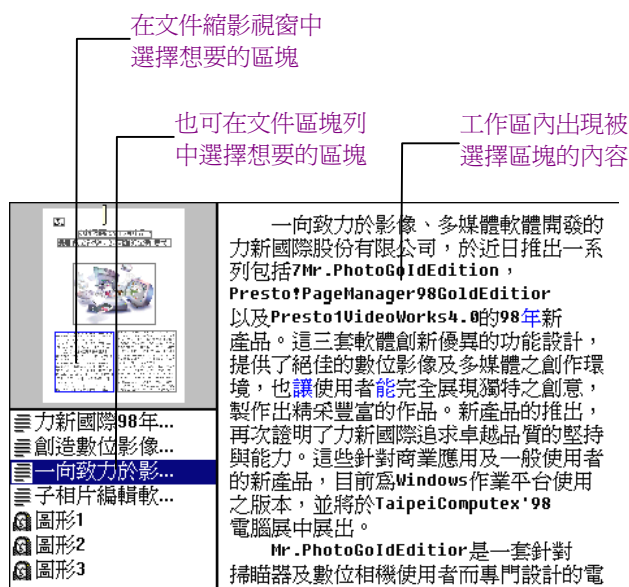
文稿編輯模式

工作區

工作區內顯示辨識後的文字；藍色字樣代表的是系統在辨識時遭遇到的疑問字。

文件縮影視窗&文件區塊列

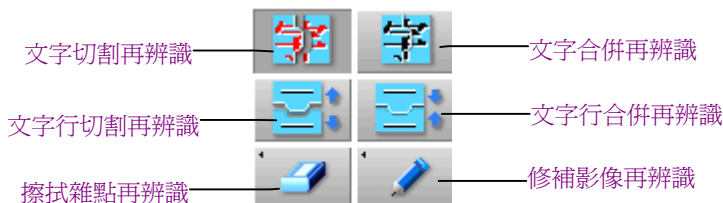
當您將游標移至文件縮影視窗時，您所點選的區塊將會顯示在工作區中。您也可以文件區塊列中選擇想要的區塊；該區塊之內容將會直接顯示於工作區中。



文件縮影視窗&文件區塊列

編輯工具盒

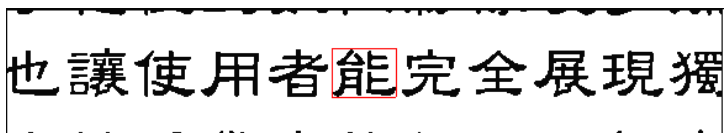
編輯工具盒提供您在文稿編輯模式中，所需的校對編輯工具。



文稿編輯模式中的編輯工具盒

文字影像視窗

文字影像視窗能放大顯示游標所指的文字之原圖影像，方便您校對與編輯文稿。



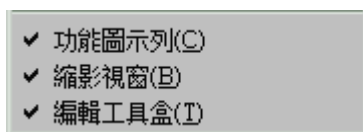
候選字視窗

您可在候選字視窗中點選正確之文字，以取代系統辨識時所誤認的字。



改變螢幕配置

部分螢幕上的視窗可依需要而顯示或隱藏，您可以選擇[查看]功能表底下的[功能圖示列]、[縮影視窗]、或[編輯工具盒]指令來改變螢幕的配置。



縮放影像顯示比例

視窗內的影像可被放大或縮小至您所需要的顯示比例，使您在編輯時更加得心應手。您可以選擇[查看]功能表底下的[與視窗同寬]、[全頁顯示]、或[實際大小]指令，或是[縮小顯示]、[放大顯示]、[縮小]、[放大]等指令來改變影像顯示的大小。

頁面移動及訊息查詢

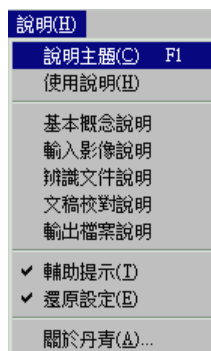
當您同時編輯數頁文稿時，您可以選擇[文件]功能表底下的指令迅速地跳至特定的頁面，並可選擇[本頁訊息]指令來查詢相關影像之大小及解析度。



影像文件資訊對話盒

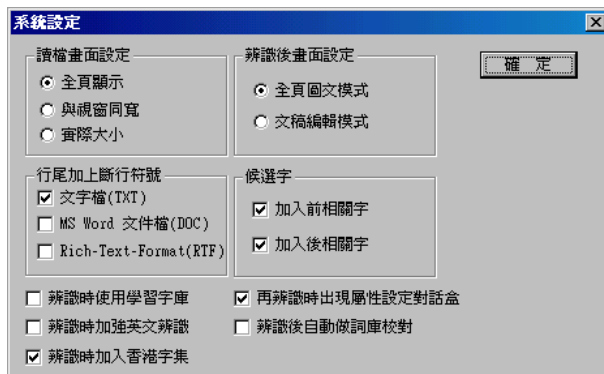
使用線上說明

若您需要快速取得相關指令或功能之說明與介紹，請按[說明]選單下的指令，即可獲得您所需的資訊。



設定系統設定值

您可選擇[檔案>系統設定]指令來設定丹青系統的預設值，[系統設定]對話盒如下：



[系統設定]對話盒

讀檔畫面設定

選擇最適合您的畫面顯示比例，如[全頁顯示]、[與視窗同寬]、[實際大小]等

辨識後畫面設定

選擇辨識後螢幕切換至[全頁圖文模式]或[文稿編輯模式]

行尾加上斷行記號

選擇辨識後需要在每行行尾加上斷行符號的檔案格式

候選字

選擇候選字的依據標準為前相關字或後相關字

前相關字:以被選字之前的字為依據，在詞意上相關的字。以「中英文」一詞為例，若英為被選字，則以「中」為依據的詞意相關字有：中心、中華、中國、中央等

後相關字:以被選字之後的字為依據，在詞意上相關的字。以「中英文」一詞為例，若英為被選字，則以「文」為依據的詞意相關字有：俄文、歐文、古文、公文等

辨識時使用學習字庫	在辨識時使用學習字庫，可提高辨識的正確率
辨識時加強英文辨識	在辨識中英文夾雜的文件時，加強對英文的辨識
辨識時加入香港字集	加入香港流通文件中特殊字的辨識字集
再辨識時出現屬性設定對話盒	在執行再辨識功能之前，出現屬性設定對話盒，可更改相關之設定
辨識時自動做詞庫校對	辨識時以內建的常用詞庫自動校對文稿

第三章

輸入影像

所有要辨識的影像檔案都必須先輸入到丹青系統的影像處理視窗，才能做更進一步的辨識處理。因此，在辨識之前，您必須先知道如何將影像檔案輸入到丹青系統。

掃描與開啓

大部分待辨識的影像檔案都經由掃描過程而取得，您可在丹青系統內先設定掃描器，並將掃描後的影像檔案直接輸入。您也可以直接開啓磁碟內的影像檔案進行辨識。

若要設定掃描器：

1. 選擇[檔案>掃描器設定]。
2. 在[掃描器設定]對話盒中選擇您的掃描器驅動程式之後，按[確定]。

若要輸入影像：

丹青系統可辨識以 TIF(G3, G4, PackBits)，PCX，BMP 以及 JPG 檔案格式所儲存的影像。

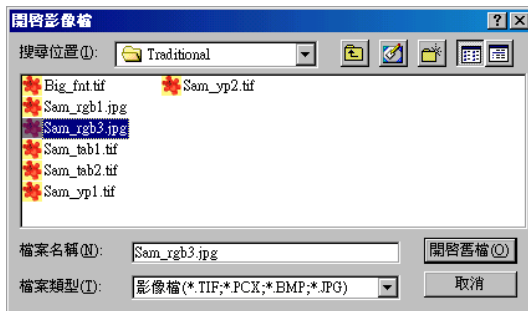
注意： 若一個 TIF 檔案內含有多頁影像，丹青只讀取第一頁的影像。

您可自下列選擇任何一種方式將檔案輸入：

- ◆ 選擇功能圖示列上的[掃描]圖示或[檔案>掃描文件]指令。

系統會開啓掃描介面，直接從掃描器輸入影像。(關於使用掃描介面，請參考掃描器使用手冊。)

- ◆ 選擇功能圖示列上的[開啓舊檔]圖示或[檔案>開啓影像檔]指令。



[開啓影像檔]對話盒

在[開啓影像檔]對話盒中選擇一個影像檔案，之後按[開啓舊檔]鍵，將選擇的影像檔案輸入到丹青系統。


您也可以配合鍵盤上的[Shift]或[Ctrl]鍵，選擇多份檔案同時開啓。

影像處理

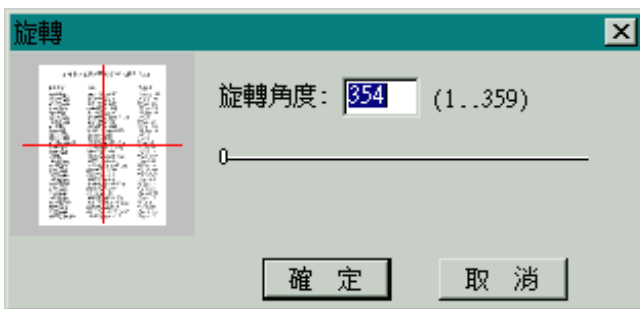
一張品質良好的文件影像，是獲致最佳辨識效果的關鍵因素。因此，在辨識前您應該檢視輸入的文件影像，並依情況做適當的影像處理，如轉傾斜之文件、除雜點、補漏白、切除不需辨識的部分、反白文件等，以提高辨識的正確率。

旋轉文件影像

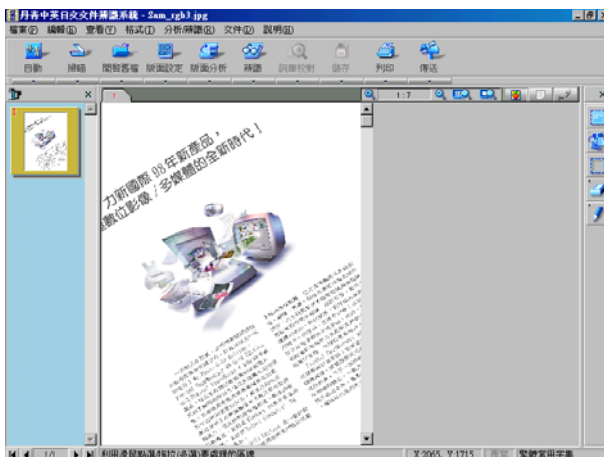
- 若影像傾斜角度小於 3 度，此為正常辨識可接受之範圍，您不需調整影像角度。
- 若影像傾斜角度為 90 度，可利用[編輯>旋轉>順(逆)時鐘旋轉 90 度]之指令，將文件影像轉正。
- 若影像傾斜角度為 180 度，可選擇[編輯>旋轉>旋轉 180 度]，將文件影像轉正。

若您要一次轉正多頁文稿，可按一下縮影視窗上方的圖示，在出現的選單中，選擇您要的指令，系統將會一次轉正目前在丹青系統中所開啓的全部影像。

- 若影像傾斜角度大於 3 度且小於 10 度，可利用[編輯>旋轉>任意角度旋轉]功能，系統會出現如下的對話盒，自動為您偵測文件影像應轉正的角度。





- 若影像的傾斜角度大於 10 度以上（如下圖），建議您重新掃描影像。




傾斜角度過大之文件影像

清除雜點及補漏白


若文件影像上有雜點（尤其是在文字區塊附近，與文字大小相近的雜點），可利用編輯工具盒上的[擦拭工具]將之去除；若影像上有漏白之部分，也可以[繪筆工具]補上，以提高正確率。

切除

若輸入的影像不需全部辨識，您可利用編輯工具盒上的[選擇影像區域]工具框選欲保留的區域，再選擇[編輯>切除]之指令，即可將不必要的部分切除。

反白功能

由於丹青系統無法辨識黑底白字的影像，若您輸入的正是此類影像，可利用[編輯>反白]之功能，將影像轉換成白底黑字之後，再進行辨識。

若您要反白多頁文稿，可按一下縮影視窗上方的圖示，在出現的選單中，選擇[全部反白]指令，系統將會一次反白目前在丹青系統中所開啓的全部影像。

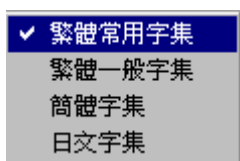
第四章

文件辨識

在執行文字辨識之前，您可以先做好一些辨識前的準備工作，如選擇辨識字集、框選辨識區域、設定版面格式、執行版面分析、指定校對詞庫等，使丹青系統在辨識時更快速而準確。此外，丹青也提供自動辨識文件之功能，從輸入以至辨識等各項流程皆能自動執行，讓您輕鬆地獲得想要的辨識結果。

設定辨識字集

當欲辨識文件裡包含中文文字時，您可以選擇[格式>設定辨識字集]，指定適合的辨識字集作為丹青系統辨識時的依據。




繁體常用字集	適合辨識一般白話文文件，如報紙、雜誌等。
繁體一般字集	適合辨識包含文言文的文件，如經文、古書、典籍等。選擇此項後的辨識速度會較使用[繁體常用字集]慢。
簡體字集	適合辨識簡體中文文件。
日文字集	適合辨識日文文件。

設定辨識區域

若您要辨識整份文件，在執行辨識之前並不需要設定辨識區域。若您只想辨識部分文件，則可先設定該部分為辨識區域，點選該區之後再執行版面分析、辨識等工作，系統將只辨識您所框選的部分。

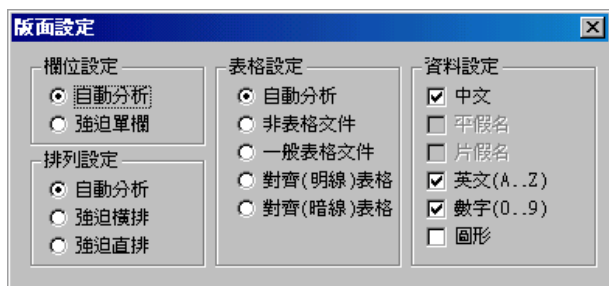
設定辨識區域的步驟如下：

1. 選擇編輯工具盒上的[設定辨識區塊]工具.
2. 拖曳滑鼠，框選欲辨識的影像區域即可。

若您要辨識數個辨識區域內的文字，可先分別設定各個欲辨識區域，之後再執行辨識，系統將會辨識所有設定的辨識區域。

設定版面格式

版面格式的設定主要在於設定欲辨識文件的屬性，包括文件的橫/直排、單/多欄、所使用的語言及表格相關的設定等等。選擇[格式>版面設定]指令之後，請依文件之內容選擇所需要的設定。



版面設定對話盒

欄位設定

欄位設定的預設值為[自動分析]，也就是讓系統自動偵測影像上文字部份的欄位格式。在下列情況下可強迫設定為單欄：

- ◆ 排版較稀疏的單欄文件（如條列式文件），請在文件辨識之前，選擇[強迫單欄]辨識影像上的文字。
- ◆ 影像上的文字部份為多欄位，但想存成單欄格式的文字檔案（如通訊錄），請在文件辨識之前，選擇[強迫單欄]辨識影像上的文字。

排列設定

排列設定的預設值為[自動分析]，也就是讓系統自動偵測影像文件的文字排列方式。若您的文字排列方式較為特殊，可依需要選擇[強迫直排]或[強迫橫排]，以獲得正確的版面分析結果。若您選擇[強迫直排]，那麼系統將以直排文字之順序辨識影像上的文字；若您選擇[強迫橫排]，那麼系統則以橫排文字之順序辨識文字。也就是說，辨識結果會因為您選擇不同的文字排列方式而不相同。

表格設定

表格設定的預設值為[自動分析]，也就是讓系統自動偵測影像文件為非表格文件或文件及其所屬型態。若您的文件不含表格，您可自己設定為[非表格文件]；若您的文件包含表格，則可依表格型態選擇[一般表格文件]、[對齊(明線)表格]或[對齊(暗線)表格]。

資料設定

可選擇影像文件所包含的內容屬性，如影像文件為中/英/中英混合文件、以及是否含有數字或圖形等。

設定版面範本

您可將經常需要設定的版面格式設定成版面範本，之後便可直接套用於相同類型的影像文件中，而不需一一重新設定。丹青系統保留十組版面範本，可供您直接使用，您也可以更改其中的設定資料。

若您要設定版面範本:

1. 選擇[格式>版面範本設定]。
2. 在出現的對話盒中，選擇欲設定的範本名稱，並設定相關的版面格式屬性，之後按[確定]即可。



版面範本設定對話盒

您也可以按一下[重新命名]，更改目前所選擇的範本名稱。


若您要套用版面範本:

1. 按一下功能圖示列的[版面設定]圖示下方的下拉式選單。
2. 在下拉式選單中選擇您想要的版面範本，即可將所選擇的範本直接套用於目前開啓的影像文件上。


版面分析

執行版面分析的目的在于將圖形與文字影像區域分離，分割出待辨識的區塊，並決定辨識區塊的順序，以利系統辨識。您可以讓系統自動執行版面分析，或是自己設定區塊及辨識順序；除此之外，您也能在執行版面分析之後，分別設定各個區塊的屬性並將版面儲存起來，當需要辨識相同版面的文件時，便可直接套用。

自動分析版面

按一下功圖示列上的[版面分析]，系統便會自動分離圖文並分割所有待辨識區域內的影像文字。在執行版面分析後，原影像文件上會出現系統所分析出的區塊框線。


手動設定版面

1. 選擇編輯工具盒的[選擇辨識區塊]工具.
2. 拖曳滑鼠，框選您要設成區塊的部分。被框出的部分會以黃色顯示。
3. 選擇[格式>設成區塊]，即可將所框選的部分設定成欲辨識的區塊。

改變辨識順序

在版面分析後所分割出的區塊都有其被辨識時的順序編號。若您選擇自動分析版面，系統會依照設定的文字橫/直排方式來決定辨識區塊的順序；若您自己設定區塊，辨識順序則依每個區塊設定之先後而定。當您按一下編輯工具盒的[變更辨識區塊順序]工具，每一個區塊的左上角會出現辨識時的順序編號。辨識順序的設定將會影響系統辨識後的文字內容排列方式。

變更辨識區塊順序的步驟如下：

1. 在編輯工具盒上選擇[變更辨識區塊順序]工具.
2. 在欲改變辨識順序的區塊上按住滑鼠，並拖曳滑鼠指標到另一個想改變成其辨識順序編號的區塊上；例如想改變區塊 3 成為區塊 2，只要將滑

鼠自編號 3 號的區塊拖曳到編號 2 號的區塊。您會發現，當您放開滑鼠指標的同時，辨識區域左上角的排列順序也跟著改變了。

您可以根據您的需要，設定各個區塊的辨識順序。

儲存版面

您可儲存經常使用的版面，於下次辨識時直接套用，可節省系統執行版面分析的時間。若您要儲存版面，選擇[格式>儲存版面]；若您要套用既有之版面，選擇[格式>開啓版面]，或是直接在[版面分析]圖示的下拉式選單中選擇。


辨識文件

在完成輸入影像檔案、設定辨識字體、設定辨識區域及變更辨識區塊順序等步驟之後，系統便可以根據您的設定開始辨識文件。

當完成辨識工作後，系統會自動進入[全頁圖文模式]或[文稿編輯模式]，讓您校對辨識後的文字檔案。您可以選擇[檔案>系統設定]，在[系統設定]對話盒中指定辨識後所出現的畫面顯示模式。

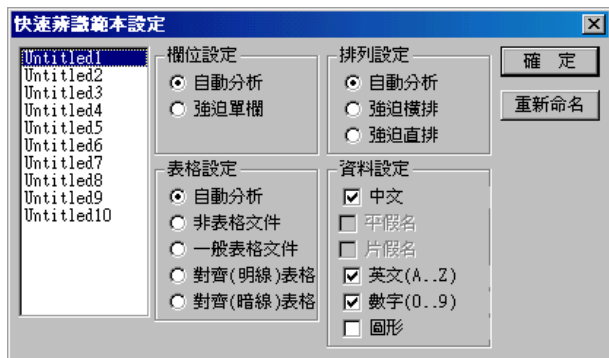
若您經常辨識某類文件，設定快速辨識範本將可為您省卻一一設定的工作，並能快速地呈現辨識後的結果。

若您要辨識文件：

按一下功能圖示列上的[辨識]圖示 。

若您要設定快速辨識範本：


1. 選擇[分析/辨識>快速辨識範本設定]。
2. 在出現的對話盒中，選擇欲設定的範本名稱，並設定相關的版面格式屬性，之後按[確定]即可。



快速辨識範本設定對話盒

您也可以按一下[重新命名]，更改目前所選擇的範本名稱。您所設定的範本將會出現於[辨識]圖示下方的下拉式選單中。

若您要套用快速辨識範本：

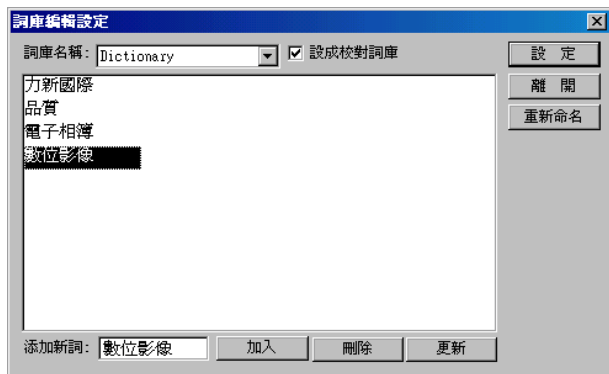
在[辨識]圖示  的下拉式選單中，選擇想要的快速辨識範本即可。

選擇校對詞庫

校對詞庫裡包含您常用的詞彙；在辨識的過程中系統將依您所選擇的校對詞庫執行辨識。因此，依需要設定不同種類的詞庫，在辨識不同種類的文件時，將更節省您在辨識及校對時所花費的時間。

設定校對詞庫

1. 選擇[分析/辨識>詞庫設定]。
2. 依您的需要設定下列選項：




詞庫編輯設定對話盒

重新命名	更換所選擇詞庫之名稱
加入	將您輸入於[添加新詞]文字框內的符號或字詞加至個人詞庫中
刪除	將您在詞庫中所選擇的字詞刪除
更新	更改您在詞庫中所選擇的字詞
設成校對詞庫	將所選擇的詞庫編輯設定成系統校對時使用的詞庫，您並可同時選擇多種詞庫為校對詞庫

使用詞庫校對

您可更換不同的詞庫，重新執行校對，使辨識結果更令人滿意。使用詞庫校對的步驟如下：

1. 選擇[分析/辨識>詞庫設定]，指定您要的校對詞庫，按[確定]即可。
2. 按一下功能圖示列上的[詞庫校對]圖示 ，系統將依據您所指定的詞庫再次辨識。

自動辨識文件

丹青提供自動辨識文件之功能，您可以隨著自動導引模式一一設定從輸入以至辨識等各項流程所必要之選項，之後由丹青為您執行自動辨識的工作。此外，您也能設定自動辨識範本，同時執行多份文件的自動辨識工作。

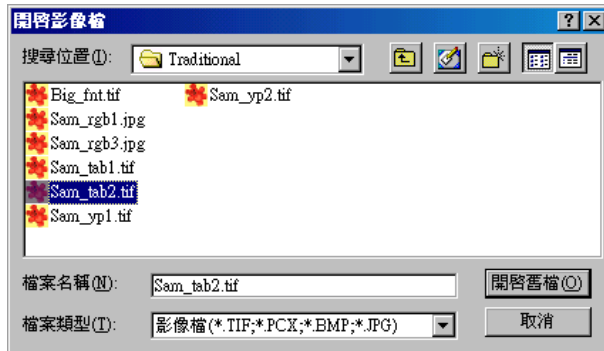
自動辨識



1. 按一下功能圖示列上的[自動]圖示。
2. 在出現的[自動導引辨識]對話盒中，選擇[開啓舊檔]。



3. 選擇您要開啓的影像文件。您可同時選擇開啓多份文件，這些文件的名稱會列在對話盒中。若您不需其中的部分文件，也可選擇[刪除]或[全部刪除]。



4. 選擇[下一步驟]，依次設定各個選項。



5. 重覆第四步驟的動作。當您完成所有的設定時，系統將開始執行辨識。

設定自動範本

您可事先設定自動辨識的屬性，使系統之自動辨識結果更符合您的需要。當您使用自動辨識功能時，系統將依據您所設定的自動範本執行辨識。

設定自動範本之步驟如下：

1. 選擇[檔案>自動範本設定]。
2. 在[自動範本設定]對話盒中，設定下列選項，之後按[確定]即可。



自動範本設定對話盒

重新命名

更換所選擇範本之名稱

新增頁

增加新的原稿影像檔案於範本中

刪除頁	刪除範本中所選擇的原稿影像檔案
全部刪除	刪除範本中全部的原稿影像檔案
欄位設定	設定範本或某個被選擇的影像檔案之欄位屬性
排列設定	設定範本或某個被選擇的影像檔案之文字橫直排
表格設定	設定範本或某個被選擇的影像檔案之表格性質
資料設定	設定範本或某個被選擇的影像檔所包含的資料內容
選擇版面範本	按一下[瀏覽]，可選擇合適的版面檔案(*.tpl)，並套用於範本中的所有檔案

第五章

校對文稿

當系統完成辨識後，您可在[全頁圖文模式]及[文稿編輯模式]中查看辨識結果。系統可顯示在辨識時遭遇到的疑問字，疑問字也許是原影像文字的模糊不清或是版面設定、分析有誤所導致，您可以依據實際狀況需要來校對文稿。以下則分別說明如何校對辨識後的檔案。


放棄辨識

若您在辨識後想放棄辨識結果並重新設定，請選擇[分析/辨識>放棄本頁辨識]或[分析/辨識>放棄全部辨識] 將辨識結果消除。

校對文稿

當辨識完成後，畫面會依您的系統設定值出現[全頁圖文模式]或[文稿編輯模式]。[全頁圖文模式]可供您觀看辨識後的文稿版面全貌，而[文稿編輯模式]則能分段顯示辨識後的結果。您可以在查看工具盒上選擇這兩種不同的顯示模式。

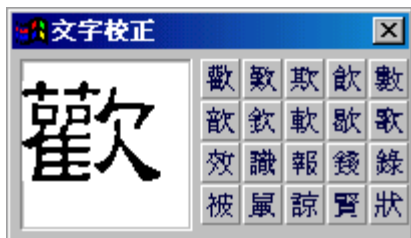
若您要在[全頁圖文模式]中校對文稿：

1. 按一下編輯工具盒上的[疑問字瀏覽]工具，文稿中的疑問字會以藍底黃字之字樣顯示。



藍底黃字之字樣顯示出疑問字

2. 使用[疑問字瀏覽]工具在第一個疑問字上點一下，並在出現的[文字校正]視窗中選擇正確的字。您所選擇的字將會取代指定的疑問字。



[文字校正]視窗

若在[文字校正視窗]中找不到您要的替代字，您也可以使用一般的鍵盤輸入法將文字輸入。

3. 同時按下鍵盤上的[Shift]及[F3]鍵，將游標移到下一個疑問字元，或使用[疑問字瀏覽]工具選擇任何辨識錯誤的字，繼續進行文字之校對即可。

若您要在[文稿編輯模式]中校對文稿：

1. 在文件瀏覽圖中直接點選欲校對的區塊，或是在文件區塊列中指定區塊，該段的辨識結果將出現於工作視窗內，且文稿中的疑問字會以藍色字樣顯示。
2. 用滑鼠點一下工作視窗內系統辨識錯誤的字，在[文字影像視窗]中，也同時會用紅線框出其對應的原字元影像。
3. 您可以在[文字影像視窗]下方的[候選字視窗]中，點選一個正確的候選字來替代辨識錯誤的字。您所選擇的候選字會替換掉工作區內滑鼠游標所在位置後的字元。

若在[候選字視窗]中找不到您要的替代字，請將游標移到工作區中辨識錯的文字左方，使用鍵盤輸入法直接輸入您需要的字。


4. 同時按下鍵盤上的[Shift]及[F3]鍵，將游標移到下一個疑問字元，或移動滑鼠選擇任何辨識錯誤的字，繼續進行文字之校對即可。

再辨識

某些影像文件可能無法使系統做出正確的分割，並因而造成辨識上的錯誤。您可以使用擦拭雜點、分/合字再辨識、分/合行再辨識與分/合區塊再辨識的功能，重新辨識錯誤結果。

擦拭雜點再辨識


去除影像上的雜點可以提高系統辨識的正確率。當系統已進入文稿編輯模式後，您可以使用[擦拭雜點再辨識]工具擦拭文字影像視窗中紅框內的字元。擦拭雜點再辨識的步驟如下：

1. 在工作區中選定系統辨識錯誤的字，文字影像視窗中也同時會出現以紅框框住的對應字元。
2. 選擇編輯工具箱中的[擦拭雜點再辨識]工具.
3. 拖曳滑鼠指標，將影像視窗內的雜點擦拭乾淨。之後在紅框外按一下或

按[Enter]鍵，系統會重新辨識該影像字元。您也可以按[Esc]鍵復原之前所擦拭掉的部分。


修補影像再辨識

增補影像上的漏白部分可以提高系統辨識的正確率。當系統已進入文稿編輯模式後，您可以使用[修補影像再辨識]工具增補文字影像視窗中紅框內的字元。修補影像再辨識的步驟如下：

1. 在工作區中選定系統辨識錯誤的字，文字影像視窗中也同時會出現以紅框框住的對應字元。
2. 選擇編輯工具箱中的[修補影像再辨識]工具.
3. 拖曳滑鼠指標，將影像視窗內的漏白部分補上。之後在紅框外按一下或按[Enter]鍵，系統會重新辨識該影像字元。您也可以按[Esc]鍵復原之前補漏白的部分。

文字切割再辨識

將相鄰二個或數個辨識錯的字元分開並予以重新辨識。使用[文字切割再辨識]工具的步驟如下：


1. 用滑鼠點一下工作區內辨識錯誤的字元，文字影像視窗中也同時會出現用紅線框出的對應字元影像。
2. 選擇編輯工具箱中的[文字切割再辨識]工具.
3. 在紅框中按住滑鼠左鍵，影像字元會被分割成二個部份。當您調整好紅線切割的位置時放開滑鼠，系統便會重新辨識被切割的兩個字元，同時，工作區內的辨識文字也會跟著更新。

文字合併再辨識

將相鄰二個或數個辨識錯的字元合併並予以重新辨識。使用[文字合併再辨識]工具的步驟如下：


1. 用滑鼠點一下工作區內辨識錯誤的字元，文字影像視窗中也同時會出現

用紅線框出的對應字元影像。

2. 選擇編輯工具盒中的[文字合併再辨識]工具 。
3. 在紅框中按住滑鼠左鍵，拖曳至欲合併的字後放開滑鼠按鍵，系統會重新辨識合併部份的字元，在此同時，工作區內的辨識文字也會跟著更新。


文字行切割再辨識

將因二行相連而辨識錯的文字分開並予以重新辨識。使用[文字行切割再辨識]工具的步驟如下：

1. 用滑鼠點一下工作區內辨識錯誤的文字，文字影像視窗中也同時會出現用紅線框出的對應部分。
2. 選擇編輯工具盒中的[文字行切割再辨識]工具 。
3. 在紅框中按一下滑鼠左鍵，相連文字行會被分割開並重新辨識，而工作區內的辨識文字也會跟著更新。



文字行合併再辨識

將被錯誤分割成二行之文字合併並予以重新辨識。使用[文字行合併再辨識]工具的步驟如下：

1. 用滑鼠點一下工作區內辨識錯誤的文字，文字影像視窗中也同時會出現用紅線框出的對應部分。
2. 選擇編輯工具盒中的[文字行合併再辨識]工具 。
3. 在紅框中按住滑鼠左鍵，拖曳至欲合併的文字行後放開滑鼠按鍵，系統會重新辨識合併部份的文字行，在此同時，工作區內的辨識文字也會跟著更新。



區塊再辨識

當您發現某區塊之版面分析錯誤，如文字之橫直排列錯誤或是中英語言設定錯誤時，您可以針對該區塊再次辨識。區塊再辨識的步驟如下：

1. 按一下查看工具盒上的[全頁圖文模式]工具。
2. 選擇編輯工具中的[設定辨識區域]工具，並在欲重新辨識的區塊上按一下滑鼠左鍵。被選擇的區塊會以黃色顯示。
3. 選擇[格式>文件格式設定]，重新設定該區塊的屬性。
4. 選擇[辨識/分析>區塊再辨識]，系統將重新辨識指定的區塊，並更新前次的辨識結果。



區塊結合再辨識

可合併被錯誤分割的區塊，再次辨識。

1. 按一下查看工具盒上的[全頁圖文模式]工具。
2. 選擇編輯工具中的[區塊結合再辨識]工具。
3. 按住滑鼠左鍵並拖曳滑鼠，將欲合併的區塊框住。
4. 放開滑鼠，系統將重新辨識合併的區塊，並更新前次的辨識結果。

區塊分開再辨識

可分割被錯誤合併的區塊，再次辨識。

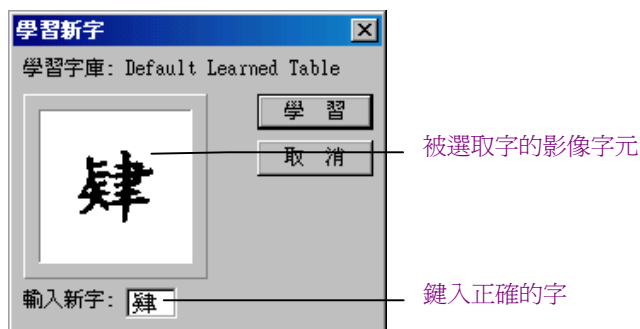
1. 按一下查看工具盒上的[全頁圖文模式]工具。
2. 選擇編輯工具中的[區塊分開再辨識]工具。
3. 按一下滑鼠左鍵並調整產生的紅線之位置(您也可托曳滑鼠直接拉出紅線)，將合併的區塊分開。
4. 放開滑鼠，系統將重新辨識被分割的區塊，並更新前次的辨識結果。

學習新字

當在校對文稿時，若系統經常辨錯某些文字，您可以使用[學習新字]的功能，將常辨識錯的字元輸入到學習資料庫中，留待以後辨識時使用。您也可依文件之性質，設定各種不同之學習字庫。

使用[學習新字]:

1. 選擇[分析/辨識>學習新字]指令，螢幕上會出現一個[新字學習]的對話盒。



在對話盒的上方為被選取字的影像字元(也就是在[文字影像視窗]中被紅框框選者)。

2. 在[新字]文字盒中輸入正確的字。
3. 按下[學習]鍵，將新字輸入到學習字庫中，並置換工作區內的錯誤字元。

刪除學習字

若您想刪除學習字庫中的字，請選擇[分析/辨識>刪除學習字]指令。在學習字庫中選擇欲刪除的字之後，按[刪除]即可。

學習字庫設定

[學習字庫設定]功能可讓您選擇特定的學習字庫做為系統之預設值，也可更改學習字庫的名稱。

1. 選擇[分析/辨識>學習字庫設定]指令。



2. 指定您想要的預設學習字庫，按[確定]。您也可選擇[重新命名]，更改學習字庫的名稱。

第六章

輸出檔案

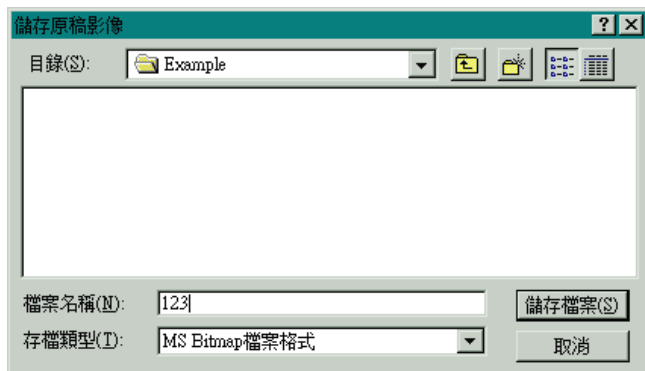
在丹青系統中您可儲存辨識前的影像文件、常用的版面格式、以及辨識後的圖文和表格，再加以使用及編輯。

儲存辨識前的影像文件

若您想保留這些在辨識前經掃描器或其他方式輸入的影像文件，可將之儲存成 BMP、TIFF、PCX、JPEG 等影像檔案格式，方便之後再辨識利用。

儲存影像文件：

1. 選擇[檔案>儲存本頁>儲存本頁原稿影像]指令。
2. 在[儲存原稿影像]對話盒中指定路徑、檔名及檔案格式，之後按[確定]即可。



[儲存原稿影像]對話盒

儲存辨識後的圖文和表格

丹青系統提供多種檔案格式，可儲存辨識後的圖文及表格。您可依需要選擇儲存本頁或儲存整份文件，並將之儲存成 TXT、DOC、RTF、XLS、SLK、CSV 等檔案格式，在小作家、Word、Excel 等文書處理器中編輯。此外，您還可以將檔案存成 HTML 格式，透過網路瀏覽器（如 Internet Explorer、Netscape Navigator 等）直接開啓。

若您要儲存辨識後的結果：

1. 若您要儲存本頁，選擇[檔案>儲存本頁>儲存本頁辨識結果]；若您要儲存整份文件，選擇[檔案>儲存文件辨識結果]。螢幕上將出現[儲存辨識結果]對話盒。



2. 指定路徑、輸入檔名，並選擇您要的存檔類型。

檔案格式	說明
TXT	純文字檔。選擇此種檔案格式，系統將只儲存檔案中的文字部分，而不會保留其中的圖形。
DOC	WIN Word 文書檔。選擇此種檔案格式，系統會將檔案儲存成 doc 檔；若檔案中包含圖形，系統會將圖形儲存成 JPEG 圖形檔並依序編號。當儲

	存多頁文件時，系統也會自動在各頁面之間加入分頁符號，以利區別。
RTF	可保留文字大小、版面位置及表格格線等資訊之檔案格式。若檔案中包含圖形，系統會將圖形儲存成 JPEG 圖形檔並依序編號。當儲存多頁文件時，系統也會自動在各頁面之間加入分頁符號，以利區別。
XLS	可在 Excel 試算表中開啓之檔案。
SLK	可在 Excel 試算表中開啓之檔案。
CSV	可在 Excel 試算表中開啓之檔案。
HTML	可透過網路瀏覽器開啓之檔案。若檔案中包含圖形，系統會將圖形儲存成 JPEG 圖形檔並依序編號。當儲存多頁文件時，系統也會自動在各頁面之間加入分頁符號，以利區別。

3. 按下[儲存檔案]鍵即可。

儲存常用的版面格式

若您經常需要使用某一種版面格式，可將此版面格式儲存成版面檔案 (*.TPL)，套用於辨識前的文件影像，節省您再次設定的時間並提高辨識之正確率。

儲存版面格式:

選擇[格式>儲存版面]指令，螢幕上將出現[儲存版面]對話盒。輸入檔名後，按[確定]即可。



【儲存版面】對話盒

若您要套用某個儲存的版面格式：

請選擇[格式>開啓版面]指令，在對話盒中選擇要套用的版面格式，按下[開啓舊檔]按鍵即可。

列印

丹青提供列印原稿及辨識結果之功能。若您需要列印辨識結果，可選擇[檔案>列印>列印辨識結果]；若需要列印原稿圖文，可選擇[檔案>列印>列印原稿影像]。

傳送

丹青系統提供直接傳送功能，可將文件影像檔或辨識結果傳送至電子郵件軟體或其他相關的應用軟體，如小作家、小畫家等。此外，丹青並保留 10 組傳送範本可讓您事先設定，在其中指定傳送的檔案格式及欲開啓的應用程式，之後便可直接套用範本，將檔案傳送至您想要的應用軟體中。

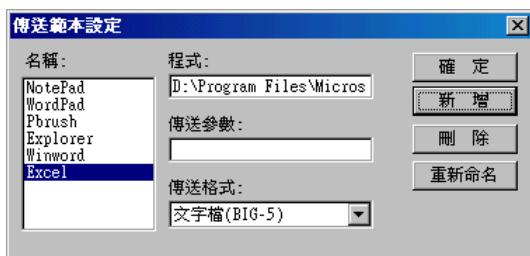
若您要將文件傳送成電子郵件：

按一下功能圖示列上的[傳送]圖示，便可將目前所開啓的檔案傳送至電子郵件軟體中。

注意： 您所安裝的電子郵件軟體必須支援 MAPI 的電子郵件系統，如 Exchange、Outlook Express 等，才能在丹青系統中直接傳送電子郵件。

若您要設定傳送之範本：

1. 選擇[檔案>傳送範本設定]。
2. 在[傳送範本設定]對話盒中設定所需之選項：



[傳送範本設定]對話盒

新加	加入新的傳送範本
刪除	刪除原有的傳送範本
重新命名	更換原有傳送範本之名稱
程式	輸入該應用程式之執行檔名
傳送參數	輸入該應用程式所需之參數
傳送格式	指定所傳出之檔案格式

若您要套用傳送之範本:

按一下功能圖示列[傳送]圖示  下方的下拉式選單 ，選擇您要套用的傳送範本即可。


注意: 在傳送的過程中，丹青系統會將檔案暫存於 *Pccrtemp* 的目錄裡，您可以定期刪除該目錄內的所有檔案，以節省您的電腦磁碟空間。

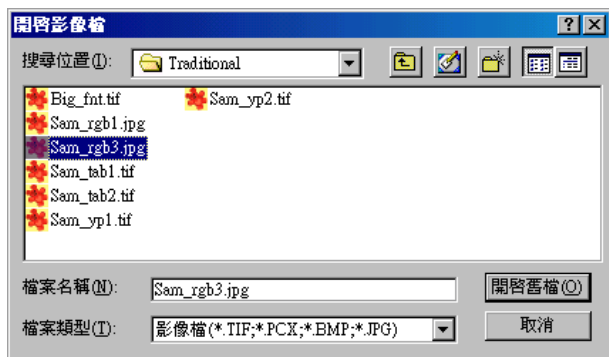
第七章

圖文辨識範例

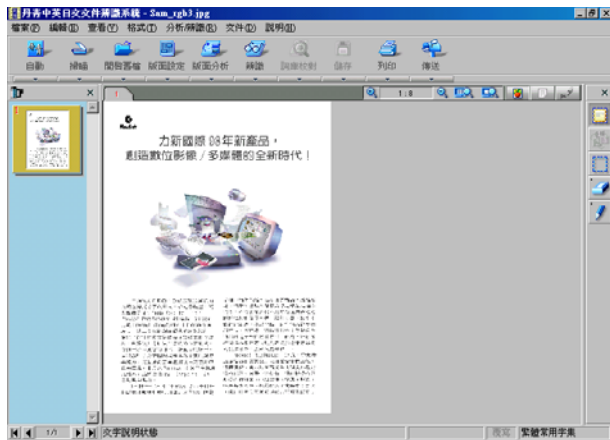
含有圖形及文字的文件是一般最常見的文件類型，本章將為您示範圖文文件的設定，辨識流程及其應用。

辨識含有圖形及文字的文件

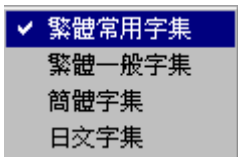
1. 按一下功能圖示列上的[開啓舊檔]圖示 。
2. 在出現的對話盒中，選擇丹青程式檔案夾中 **Example** 目錄裡的 **Sam_rgb3.jpg** 檔案。




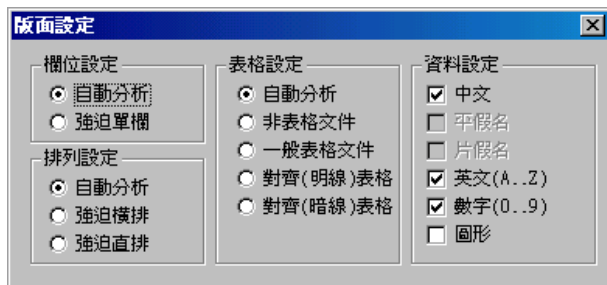
3. 按一下[開啓舊檔]，將檔案輸入。



4. 選擇[格式>設定辨識字集]，在次功能表中選擇[繁體常用字集]；或由訊息狀態列的右方選擇[繁體常用字集]。

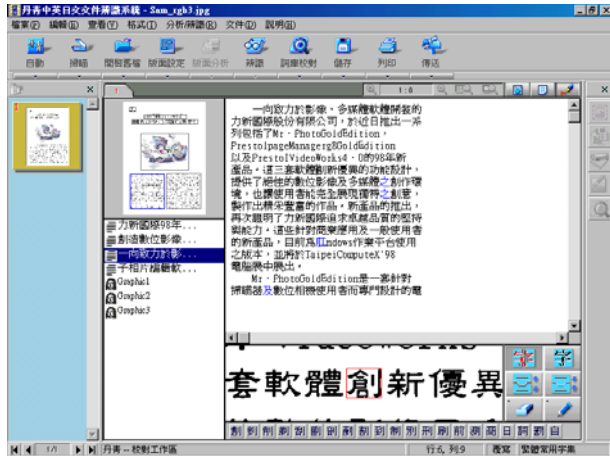


5. 按一下功能圖示列上的[版面設定]圖示 。
6. 在[版面設定]對話盒中，選擇[欄位設定－自動分析]、[排列設定－自動分析]、[表格設定－自動分析]、[資料設定－中文、英文、數字]。

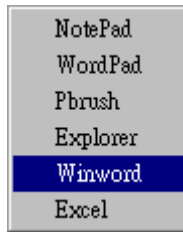


7. 按一下功能圖示列上的[辨識]圖示 。系統將開始為您辨識文件。
8. 選擇查看工具盒上的[全頁圖文模式]  及[文稿編輯模式]  工具查看辨

識結果，並依需要校對文稿。(校對文稿之詳細步驟請參考第 5 章)。



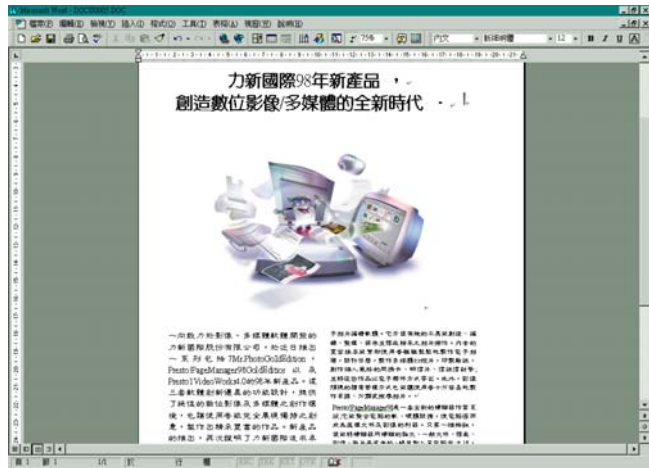
9. 在功能圖示列的[傳送]圖示  之下拉式選單中 ，選擇[WIN Word]。



注意： 您必須先在電腦中安裝 WIN Word 軟體，才能直接傳送辨識結果。在安裝之後，並且要先將之設定為傳送範本，才能在[傳送]圖示之下拉式選單中直接選取套用。關於傳送範本的設定，請參考第 6 章之說明。

注意： 您可以選擇[檔案>系統設定]，在對話盒中設定 WIN Word 文件檔加上斷行符號，便可保留文件原來的版面樣式。若您想要再編輯該文件，則不要選擇加上斷行符號，可方便文件的再編輯。

10. 系統將會直接傳送辨識結果至 WIN Word 軟體中。在 Word 中選擇[檢視 > 整頁顯示]，即可瀏覽辨識後的文件全貌。




傳送至 WIN Word 軟體中的辨識結果

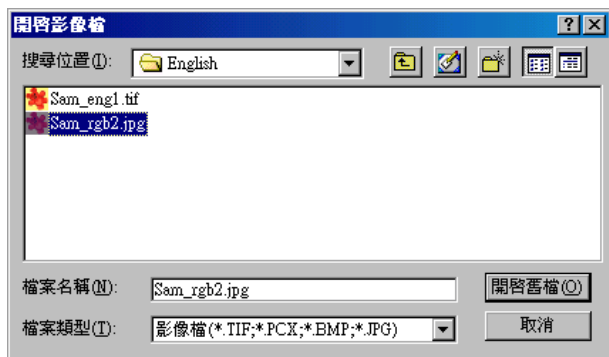
第八章

英文辨識範例

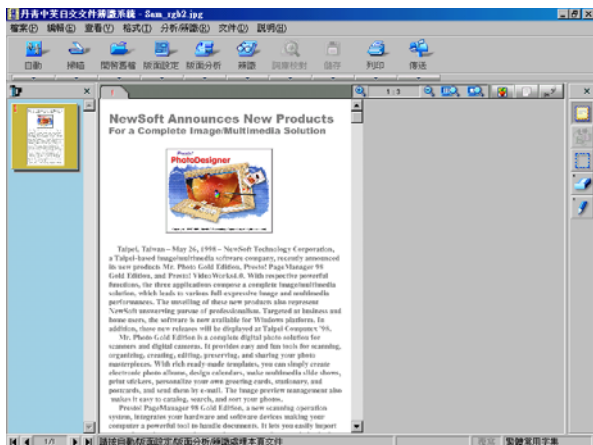
除了中文文件外，丹青也能為您辨識英文文件。以下將為您介紹英文文件的辨識步驟。


辨識英文文件

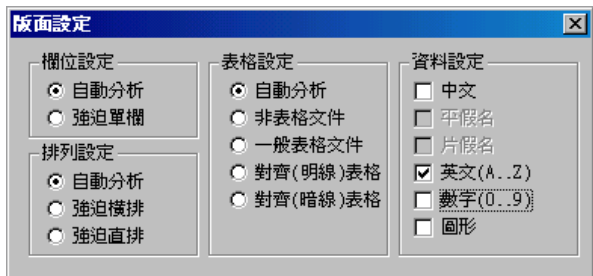
1. 按一下功能圖示列上的[開啓舊檔]圖示 。
2. 在出現的對話盒中，選擇丹青程式檔案夾中 **Example** 目錄裡的 **Sam_rgb2.jpg** 檔案。



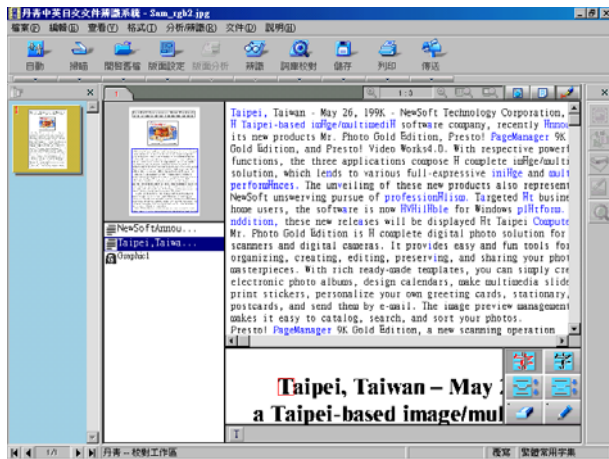
3. 按一下[開啓舊檔]，將檔案輸入。




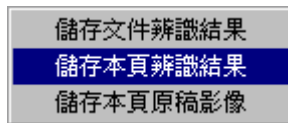
4. 按一下功能圖示列上的[版面設定]圖示。在[版面設定]對話盒中，選擇[欄位設定－自動分析]、[排列設定－自動分析]、[表格設定－自動分析]、[資料設定－英文]。



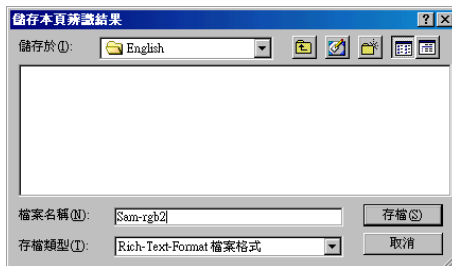
5. 按一下功能圖示列上的[辨識]圖示，系統將開始為您辨識文件。選擇查看工具盒上的[全頁圖文模式]及[文稿編輯模式]工具查看辨識結果，並依需要校對文稿。(校對文稿之詳細步驟請參考第 5 章)。



6. 校對完成後，您可在功能圖示列的[儲存]圖示  之下拉式選單中，選擇[儲存本頁辨識結果]。



7. 在儲存對話盒中，輸入檔案名稱 Sam-rgb2，並指定存檔類型為 RTF 檔案格式。



注意： 您可選擇[檔案>系統設定]，在對話盒中設定 RTF 文件檔加上斷行符號，便可保留文件原來的版面樣式。若您想要再編輯排版該文件，則不要選擇加上斷行符號。

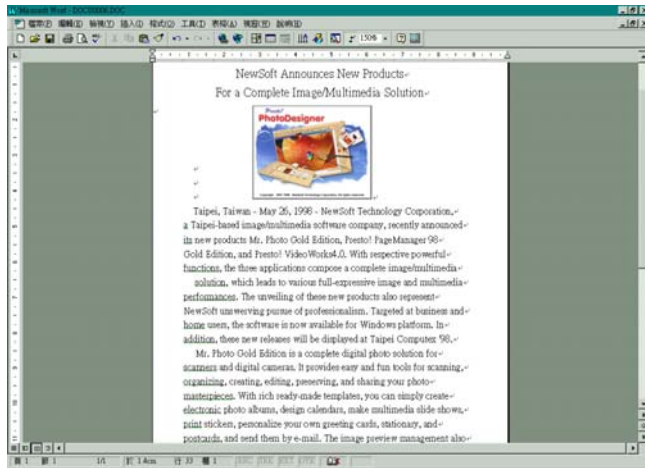
8. 您也可以直接傳送檔案至相關的軟體中。可在功能圖示列的[傳送]圖示



之下拉式選單中，選擇[WIN Word]。

注意： 您必須先在電腦中安裝 WIN Word 軟體，才能直接傳送辨識結果。在安裝之後，並且要先將之設定為傳送範本，才能在[傳送]圖示之下拉式選單中直接選取套用。關於傳送範本的設定，請參考第 6 章之說明。

9. 系統將會直接傳送辨識結果至 WIN Word 軟體中。在 WIN Word 中選擇[檢視>整頁顯示]，即可瀏覽辨識後的文件全貌。



在 WIN Word 中的辨識結果


注意： 若您不需要儲存圖形，也可選擇 txt 純文字檔案格式來儲存文件，所儲存的文件將只包含檔案中的文字內容。您可將該純文字檔傳送至小作家或計事本等文書軟體中。

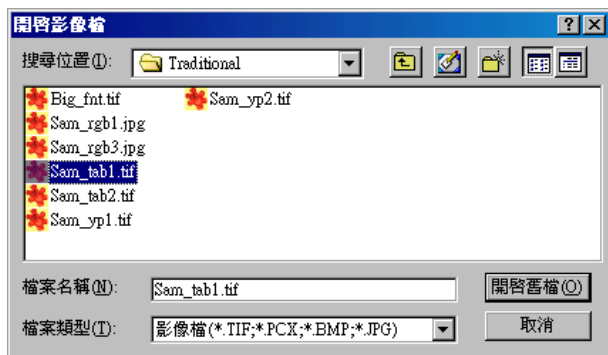
第九章

表格辨識範例

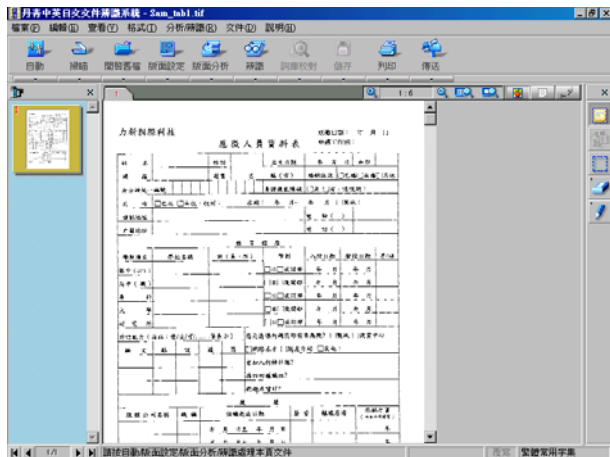
除了圖文並存及中英文文件的辨識外，丹青也能為您辨識各式各樣的表格，如公文、通訊錄、履歷表、成績單等。表格的形式及文字皆能保持原貌；您也能將表格的辨識結果傳送至 Word、Excel 等軟體中再進一步編輯處理。

辨識一般表格影像

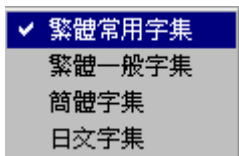
1. 按一下功能圖示列上的[開啓舊檔]圖示 。
2. 在出現的對話盒中，選擇丹青程式檔案夾中 Example 目錄裡的 Sam_TAB1.tif 檔案。




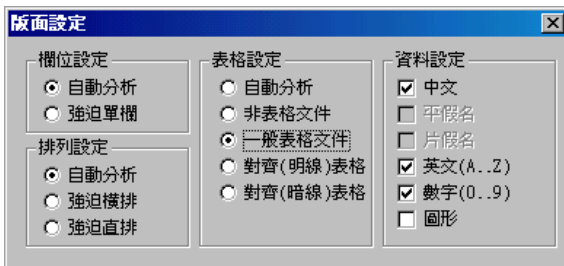
3. 按一下[開啓舊檔]，將檔案輸入。




- 選擇[格式>設定辨識字集]，在次功能表中選擇[繁體常用字集]；或由訊息狀態列的右方選擇[繁體常用字集]。

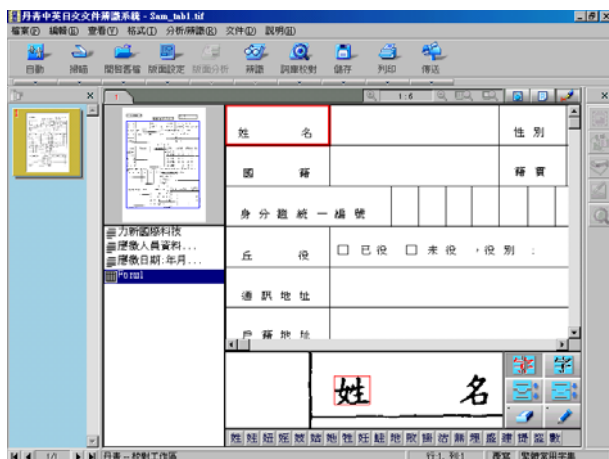


- 按一下功能圖示列上的[版面設定]圖示 。
- 在[版面設定]對話盒中，選擇[欄位設定—自動分析]、[排列設定—自動分析]、[表格設定—一般表格文件]、[資料設定—中文、英文、數字]。

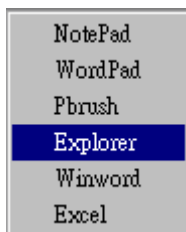


- 按一下功能圖示列上的[辨識]圖示 。系統將開始為您辨識文件。

8. 選擇查看工具盒上的[全頁圖文模式]  及[文稿編輯模式]  工具查看辨識結果，並依需要校對文稿。(校對文稿之詳細步驟請參考第 5 章)。

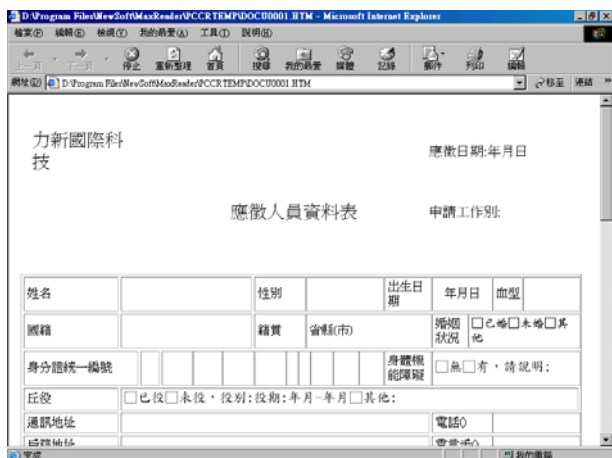


9. 您可以直接傳送檔案至相關的軟體中。可在功能圖示列的[傳送]圖示  之下拉式選單中，選擇[Explorer]。




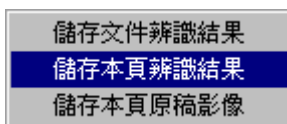
注意： 您必須先在電腦中安裝網路瀏覽器，才能直接傳送辨識結果。在安裝之後，並且要先將之設定為傳送範本，才能在[傳送]圖示之下拉式選單中直接選取套用。關於傳送範本的設定，請參考第 6 章之說明。

10. 系統將會直接傳送辨識結果至 Internet Explorer 網路瀏覽器中。



在 Internet Explorer 網路瀏覽器中的辨識結果

11. 您也可以選擇儲存檔案。可在功能圖示列的[儲存]圖示  之下拉式選單中，選擇[儲存本頁辨識結果]。




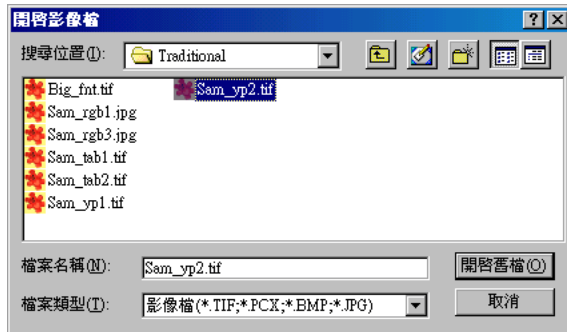
12. 在儲存對話盒中，輸入檔案名稱 Sam_TAB1，並指定存檔類型為 HTML 檔案格式。系統會將檔案儲存成 HTML 格式，並將檔案中的圖形依序編號儲存成*.JPEG 檔。



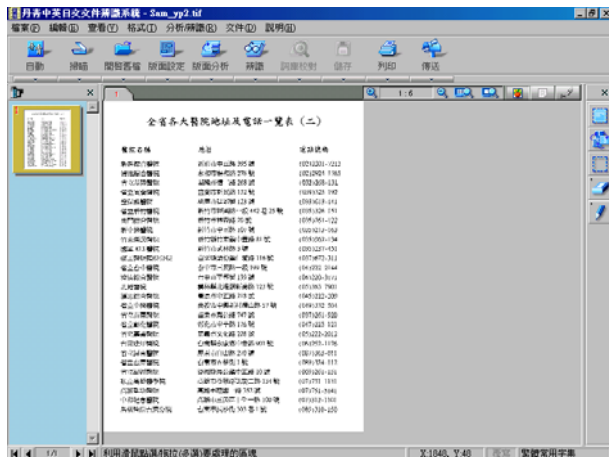
儲存辨識結果

辨識暗線表格影像

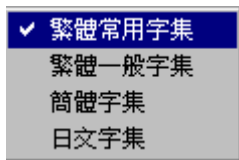
1. 按一下功能圖示列上的[開啓舊檔]圖示 。
2. 在出現的對話盒中，選擇丹青程式檔案夾中 **Example** 目錄裡的 **Sam_YP2.tif** 檔案。




3. 按一下[開啓舊檔]，將檔案輸入。

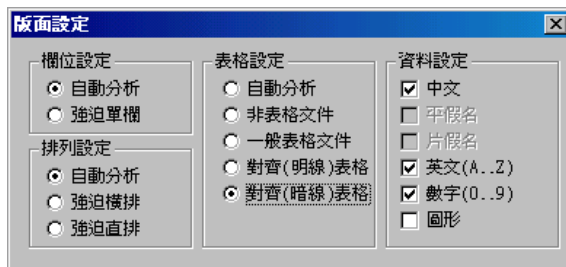



4. 選擇[格式>設定辨識字集]，在次功能表中選擇[繁體常用字集]；或由訊息狀態列的右方選擇[繁體常用字集]。

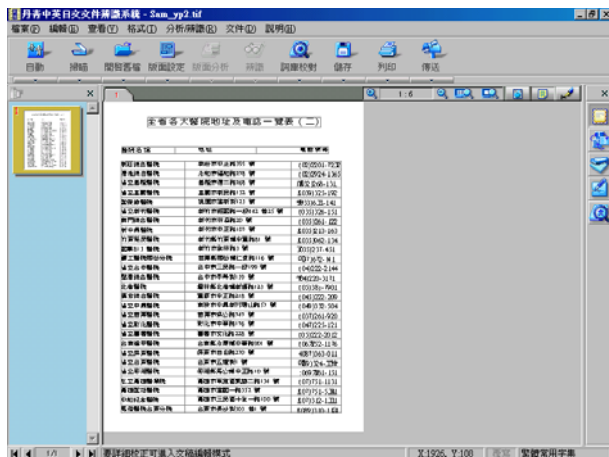


5. 按一下功能圖示列上的[版面設定]圖示 。

6. 在[版面設定]對話盒中，選擇[欄位設定－自動分析]、[排列設定－自動分析]、[表格設定－對齊(暗線)表格]、[資料設定－中文、英文、數字]。

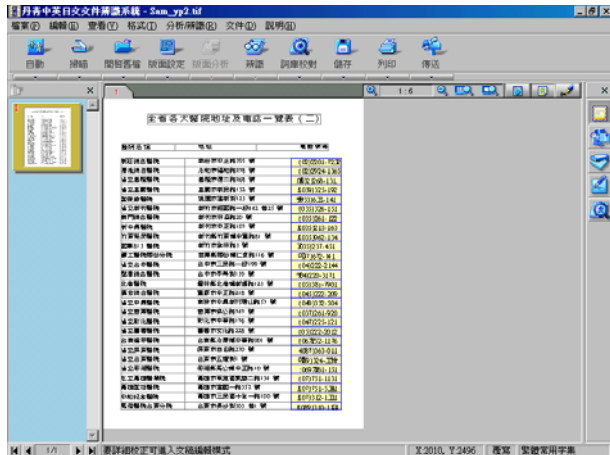


7. 按一下功能圖示列上的[辨識]圖示 。系統將開始為您辨識文件。

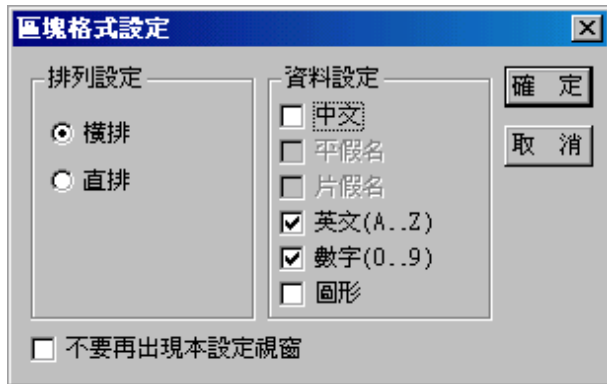


8. 選擇查看工具盒上的[全頁圖文模式]  及[文稿編輯模式]  工具查看辨識結果，並依需要校對文稿。(校對文稿之詳細步驟請參考第 5 章)。

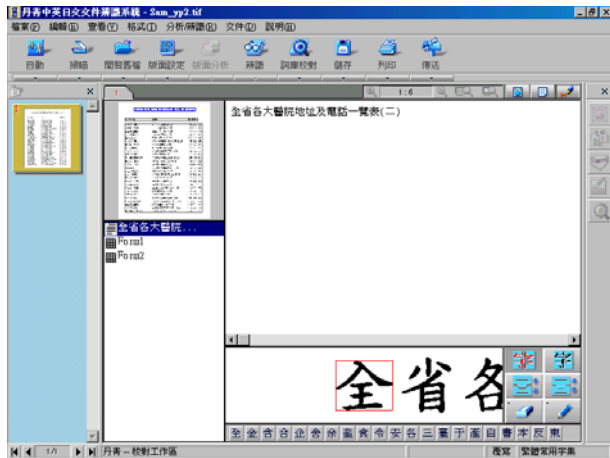
9. 由於電話號碼區塊無中文字，所以這裡重新設定區塊的屬性將會更提高辨識結果的正確性。請選擇編輯工具盒上的[選擇辨識區塊]工具，利用滑鼠圈選要重新辨識的區塊。被選擇的區塊會顯示黃色。



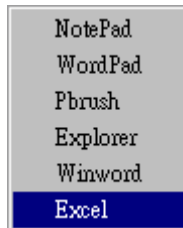
10. 選擇[分析/辨識>區塊再辨識]。在[資料設定]選項中，取消[中文]設定。按下[確定]，系統將開始為您重新辨識此區塊。



11. 選擇查看工具盒上的[全頁圖文模式]  及[文稿編輯模式]  工具查看辨識結果，並依需要校對文稿。(校對文稿之詳細步驟請參考第 5 章)。



12. 您可以直接傳送檔案至相關的軟體中。可在功能圖示列的[傳送]圖示之下拉式選單中，選擇[Excel]。




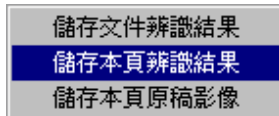
注意： 您必須先在電腦中安裝 **Excel** 試算表軟體，才能直接傳送辨識結果。在安裝之後，並且要先將之設定為傳送範本，才能在[傳送]圖示之下拉式選單中直接選取套用。關於傳送範本的設定，請參考第 6 章之說明。

13. 辨識結果將會直接傳送至 **Excel** 試算表軟體中。

醫院名稱	地址	電話號碼
新莊綜合醫院	新莊市中正路395號	(02)2201-7212
清池綜合醫院	永和市福和路278號	(02)2924-1365
省立基隆醫院	基隆市信二路268號	(03)268-131
省立宜蘭醫院	宜蘭市新民路152號	(039)325-192
鹿保綠醫院	桃園市建新街123號	(033)613-141
省立新竹醫院	新竹市經國路一段442巷25號	(035)326-151
南門綜合醫院	新竹市林森路32號	(035)261-122
新中興醫院	新竹市中正路107號	(035)213-163
竹東榮民醫院	新竹縣竹東鎮中豐路81號	(035)862-134
國華醫院	新竹市武林路3號	(035)237-451
鹽工醫院頭份分院	苗栗縣頭份鎮仁愛路116號	(037)672-311
省立台中醫院	台中市三民路一段199號	(04)222-2144
澄清綜合醫院	台中市平等路139號	(04)220-3171
北港醫院	雲林縣北港鎮新港路123號	(05)383-7901
鳳溪綜合醫院	豐原市中正路218號	(045)222-209
省立中興醫院	南投市中興新村環山路57號	(049)332-504

在 Excel 試算表軟體中的辨識結果

- 14.您也可以選擇儲存檔案。在功能圖示列的【儲存】圖示  之下拉式選單中選擇【儲存本頁辨識結果】。



- 15.在儲存對話盒中，輸入檔案名稱 Sam_YP2，並指定存檔類型為 EXCEL 檔案格式，即可儲存辨識結果。


將辨識結果儲存成 EXCEL 檔案格式

第十章

自動辨識範例

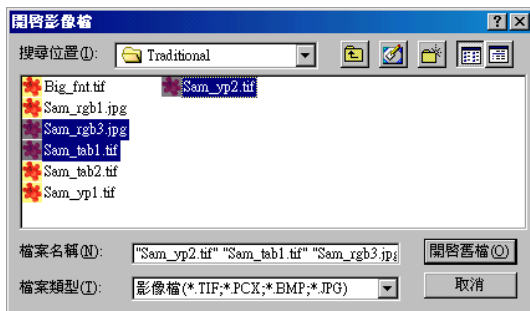
丹青系統提供方便快捷的自動辨識功能，從輸入以至辨識等各項流程皆能自動執行，讓您輕鬆地獲得高正確率的辨識結果。

自動辨識文件

1. 選擇[檔案>開新檔案]。
2. 按一下功能圖示列上的[自動]圖示 。
3. 在出現的對話盒中，選擇[開啓舊檔]。



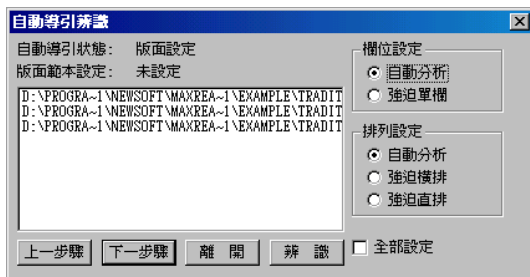
4. 開啓丹青程式檔案夾中 Example\Traditional 目錄，按住鍵盤上的[Ctrl]鍵並點選 Sam_RGB3.jpg、Sam_TAB1.tif 及 Sam_YP2.tif 等三個檔案。



5. 按下[開啓舊檔]，即可將欲辨識的文件讀入。

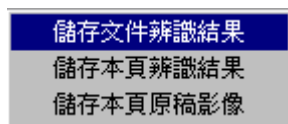


6. 選擇[下一步驟]。
7. 在出現的選項中，分別點選各個檔案，並一一設定為[欄位設定－自動分析]及[排列設定－自動分析]。

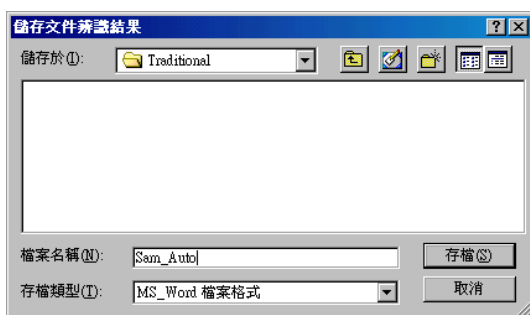


8. 選擇[下一步驟]。
9. 在[表格設定]的選項中，點選 Sam_RGB3 檔案名稱並設定為[非表格文件]；選擇 Sam_TAB1 檔案名稱並設定為[一般表格文件]；以及選擇 Sam_YP2 檔案名稱並設定為[對齊(暗線)表格]。

- 14.校對完成後，可在功能圖示列的[儲存]圖示之下拉式選單中選擇[儲存文件辨識結果]。



- 15.在儲存對話盒中，輸入檔案名稱爲 Sam_Auto，並指定存檔類型爲 Word 檔案格式。這三份文件將會被儲存在同一個檔案，並以換頁符號區分各個文件。



儲存文件辨識結果

用語說明

用語	說明
區塊屬性	文件版面及資料內容的特性，如文字的橫/直排、文字的語言、是否含有表格、單/多欄位等。
剪貼簿	文字及圖片的暫存區域。用以儲存[剪下來]和[貼下去]指令所處理的物件。
預設值	在程式中各種選項最初的設定。
解析度(DPI)	解析度的度量單位。印表機和掃描器的解析度是由每英吋所產生的點數來測量。DPI值愈高，解析度愈高。
記憶體	也稱為 RAM (隨機存取記憶體) 。這是電腦暫時存放資料的區域。您可以將記憶體中的內容複製到硬碟或磁片中永久保存。
下拉式選單	當您選取功能表列上的項目時所出現的指令表。
版面設定	文件資料內容的設定，如欄位設定、排列設定、表格設定、語言設定等。
版面分析	系統自動分析影像版面辨識區域、變更辨識區塊順序及設定文件屬性
區域	文件影像上欲辨識的矩形區域
區塊	版面分析後的矩形區塊
分字	將相鄰而辨識錯的字元分開
合字	將相鄰而辨識錯的字元合併
分行	將因二行相連而辨識錯的文字分開

合行	將被錯誤分割成二行之文字合併
分割區塊	可分割被錯誤合併的區塊
合併區塊	可合併被錯誤分割的區塊
候選字	與選取字字形上相似者或語意上前後相關者
學習新字	將常辨識錯的字元輸入到學習資料庫中，增強辨識正確

功能表

檔案	
開新檔案	開啓新文件，結束原文件。
自動	自動輸入、辨識及校對文件。
自動範本設定	設定自動功能的屬性。
開啓影像檔	開啓影像檔案。
儲存文件辨識結果	儲存整份文件之辨識結果，並以分頁符號區隔各頁面。
加入新頁	在文件末端新增數頁(由對話盒選取)。
刪除本頁	將本頁自文件中移除。
儲存本頁	以使用者輸入的名稱儲存本頁影像或文字。若您選擇以DOC、RTF、HTML格式儲存文字，系統會儲存檔案內的所有內容(圖形及文字)，並另將文件內所包含的圖形依序編號儲存。若您選擇儲存成純文字檔，則系統只會儲存檔案中的文字部分。
掃描文件	開啓掃描介面（請參閱掃描器的使用手冊）。
掃描器設定	選擇掃描器的來源。
列印	列印目前的檔案。
設定印表機	選擇印表機的來源。
傳送	將文件影像檔或辨識結果直接傳送至您事先設定的應用軟體中。
傳送範本設定	設定欲傳送檔案的應用軟體並及輸出之檔案格式。

系統設定	設定丹青系統程式的執行方式。
結束	離開丹青中英日文文件辨識系統。

編輯	
復原	復原前一項執行的動作。
剪下	將選取的區域剪下來並存到電腦系統的剪貼簿上。
複製	複製選取的區域到剪貼簿上。
貼上	將剪貼簿上的影像資料貼在開啓的檔案上。
清除	刪除選取的區域。
全選	選擇工作區內所有的影像或文字。
切除	切除不需要的影像部分。
反白	調換影像的文字顏色與背景顏色。
旋轉	選擇順時針旋轉90度、旋轉180度、逆時針旋轉90度或任意角度旋轉等指令旋轉工作區內的文件影像。
尋找	尋找文稿中的某個特定字詞。
找下一個	尋找文稿中的下一個特定字詞。
取代	以指定的字詞取代某個特定字詞。
尋找第一個疑問字	尋找目前頁面中，丹青系統在辨識時所遭遇的第一個疑問字。
找下一個疑問字	尋找下一個疑問字。

查看	
與視窗同寬	將螢幕上的影像放大或縮小至與視窗同寬。
全頁顯示	將螢幕上的影像以全頁顯示。
實際大小	將螢幕上的影像以實際大小。

縮小顯示	以1/2到1/8倍的比例縮小顯示螢幕上的影像。
放大顯示	以2到8倍的比例放大顯示螢幕上的影像。
放大	放大顯示螢幕上的影像。
縮小	縮小顯示螢幕上的影像。
原稿影像模式	顯示輸入的原稿影像。
全頁圖文模式	顯示辨識後的圖文版面。
文稿編輯模式	顯示辨識後的文章編輯模式。
功能圖示列	顯示或隱藏功能圖示列。
縮影視窗	顯示或隱藏縮影視窗。
編輯工具盒	顯示或隱藏編輯工具盒。

格式	
版面設定	設定文件的資料屬性及其版面格式。
版面範本設定	設定常用文件之版面格式並可直接套用。
設定辨識字集	設定系統辨識時所使用的字集。
儲存版面	儲存目前工作區中的影像的版面格式資料，包括版面尺寸、設定的辨識區域及順序等。
開啓版面	套用已儲存的版面格式。
放棄分析結果	放棄目前的版面分析結果，讓使用者重新設定，再行分析。
設成區塊	設定框選部分為待辨識區塊。
字體設定	指定文字區塊的字體
字型大小設定	指定文字區塊的字型大小
輸入模式設定	設定文字輸入的模式為[插入]或[覆寫]。

分析/辨識

版面分析	分析影像上框選的辨識區域。
辨識	辨識分析後的文字。
詞庫校對	以指定的詞庫再辨識。
快速辨識範本設定	將常用的版面儲存起來，所儲存的範本會出現在辨識下拉式選單中，可直接套用執行辨識。
詞庫設定	設定個人常用的詞庫以作為詞庫校對之依據。
學習新字	讓丹青系統在做文字辨識的時候，參考學習字庫內的新字為辨識之根據。
刪除學習字	刪除學習字庫中的學習字。
學習字庫設定	指定學習字庫作為辨識之依據。
區塊再辨識	再次辨識所選擇的區塊。
放棄本頁辨識	放棄目前的分割、辨識結果，讓使用者重新設定，再行分割、辨識。
放棄全部辨識	放棄所有檔案的分割、辨識結果，讓使用者重新設定，再行分割、辨識。

文件

第一頁	顯示文件中的第一頁畫面。
最後頁	顯示文件中的第一頁畫面。
下一頁	顯示目前文件的下一頁畫面。
上一頁	顯示目前文件的上一頁畫面。
到第幾頁	顯示指定的頁面。
本頁訊息	選擇本頁的影像文件資訊。

說明

說明主題	選擇說明的內容或索引。
使用說明	顯示使用說明的使用方法。

輔助提示	顯示輔助的標題。
還原設定	還原至丹青系統程式的初始設定。
關於丹青	顯示丹青程式資訊、版本及版權。

附錄C

功能圖示列及下拉式選單

功能圖示列中包含了在辨識過程裡常用到的指令，其下拉式選單中可列出與該圖示相關之指令。您可以在各個圖示之下拉式選單中，按一下您要的指令，直接套用於文件上。

	自動 自動輸入、分析及辨識文件。您可選擇[檔案>自動範本設定]設定常用的自動範本；已設定的自動範本將會出現在下拉式選單中，可供您直接執行自動功能。
	掃描 從掃描器輸入影像。您可選擇[檔案>掃描器設定] 設定欲使用的掃描器；所有已安裝的掃描器將會出現在下拉式選單中。
	開啟舊檔 開啓欲辨識之影像檔案。在其下拉式選單中，會保留10組您最近開啓過的檔案。
	版面設定 設定文件的資料屬性及其版面格式。您可選擇[格式>版面範本設定]設定常用的版面格式，如英文文件、通訊錄、公文等；已設定的版面範本將會出現在下拉式選單中，可供您直接套用。
	版面分析 自動分析影像上所框選的辨識區域。您可選擇[格式>儲存版面]，將常用的版面儲存起來。在下拉式選單中，會保留10組您最近儲存的版面檔案。

	<p>辨識</p> <p>執行辨識指令。您可選擇[分析/辨識>快速辨識範本設定]，將常用的版面儲存起來；您所儲存的版面會出現在下拉式選單中，可供您直接套用執行辨識。</p>
	<p>詞庫校對</p> <p>以指定的詞庫再辨識。您可選擇[分析/辨識>詞庫設定]，將常用的詞庫儲存起來。系統保留10組詞庫供您設定。這10組詞庫會出現在下拉式選單中，您可直接選擇數個詞庫，再執行辨識。</p>
	<p>儲存</p> <p>以使用者輸入的名稱儲存本頁影像、文字或整份文件。您可選擇[儲存辨識結果]來儲存整份文件之辨識結果，[儲存文字檔]來儲存本頁之辨識結果，或是[儲存影像檔]來儲存原稿之影像。</p>
	<p>列印</p> <p>列印目前的檔案。您可選擇[檔案>印表機設定] 設定欲使用的印表機；所有已安裝的印表機將會出現在下拉式選單中。</p>
	<p>傳送</p> <p>將文件影像檔或辨識結果直接傳送至您所安裝的電子郵件軟體中。您可選擇[檔案>傳送範本設定] 設定欲傳送的應用軟體，如文書處理器或影像處理軟體等；您所設定的傳送範本會出現在下拉式選單中，可讓您直接點選將檔案傳至該軟體。</p>

編輯工具盒圖示

	放大觀看 放大顯示螢幕上的影像。
	縮小觀看 縮小顯示螢幕上的影像。
	全頁觀看 將螢幕上的影像以全頁顯示。
	與視窗同寬度觀看 將螢幕上的影像放大或縮小至與視窗同寬。
	原稿影像模式 顯示輸入的原稿影像。
	全頁圖文模式 顯示辨識後的圖文版面。
	文稿編輯模式 顯示辨識後的文稿編輯模式。
	設定辨識區塊工具 框選待辨識區域。使用滑鼠在影像上拖曳框選要做辨識的區域。
	變更辨識區塊順序工具 每一個框選出來的區塊都有一個辨識的序號，您可以決定它們的先後順序。 (請參考變更辨識區塊順序的使用說明。)

	選取影像區域工具 可使用滑鼠在影像上拖曳框選影像區域，之後再執行[切除]或[設定區塊]等功能。
	擦拭工具 清除文件上的雜點，以提高辨識效率。
	繪筆工具 添補影像文件漏白的部分。
	區塊結合再辨識工具 可合併被錯誤分割的區塊，再次辨識。
	區塊分開再辨識工具 可分割被錯誤合併的區塊，再次辨識。
	文字校對工具 顯示辨識時的疑問字。
	文字合併再辨識 將相鄰二個或數個辨識錯的字元合併並重新辨識。
	文字切割再辨識 將相鄰二個或數個辨識錯的字元分開並重新辨識。
	文字行合併再辨識 將被錯誤分割成二行之文字合併並予以重新辨識。
	文字行切割再辨識 將因二行相連而辨識錯的文字分開並予以重新辨識。

掃描的建議

當您掃描一份文件時，原稿品質的好與壞會直接影響到掃描的結果。而當您在掃描不同的文件影像時，所使用的解析度值也跟著不同。通常掃描一般文件（文字高度約為 3mm）時，建議您使用 400dpi 的解析度值。若字體稍小，掃描時則建議您提高解析度。

為了方便系統辨識，請您在使用系統辨識文件以前，用[擦拭工具]擦去影像上的雜點，以提高系統的辨識效率。

如何改善辨識品質

影響辨識品質的因素有三：影像品質、掃描時的解析度(dpi)、掃描時的明亮度。

- ◆ 影像品質：掃描時文件要放端正(誤差以三度為上限)、影像要清晰。
- ◆ 掃描時的解析度(dpi)：一般而言，影像檔內文字的大小在 40~50 Pixel 時，丹青會有最佳的表現，過與不及，皆會降低辨識品質。
- ◆ 以中國時報及聯合報本文字體的大小(3mm)為例，可用 400dpi 來掃描；較大的字體建議您使用 300dpi，詳細資料請參考圖例一[掃描解析度建議]。
- ◆ 掃描時的明亮度：太淡（造成斷線--請參考圖例四）或太濃（糊成一團--請參考圖例三）的影像都會降低系統的辨識率。因此在調整掃描的明亮度時，應注意不要讓筆劃簡單的字（如中、大、口）的筆劃斷掉，也不要讓筆劃複雜的字（如團、麗、籲）的內部糊成一團。

若兩者無法兼顧時應以保留筆劃簡單的字之橫向筆劃為優先。

圖例一：掃描解析度建議

建議值：400dpi

力新國際科技股份有限公司(10 point)

力新國際科技股份有限公司(11 point)

力新國際科技股份有限公司(12 point)

建議值：300dpi

力新國際科技股份有限公司(12 point)

力新國際科技股份有限公司(14 point)

力新國際科技股份有限公司(16 point)

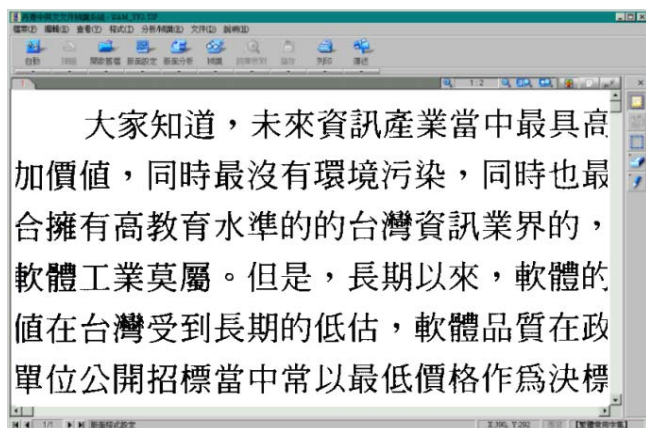
建議值：200dpi

力新國際科技股份有限公司(18 point)

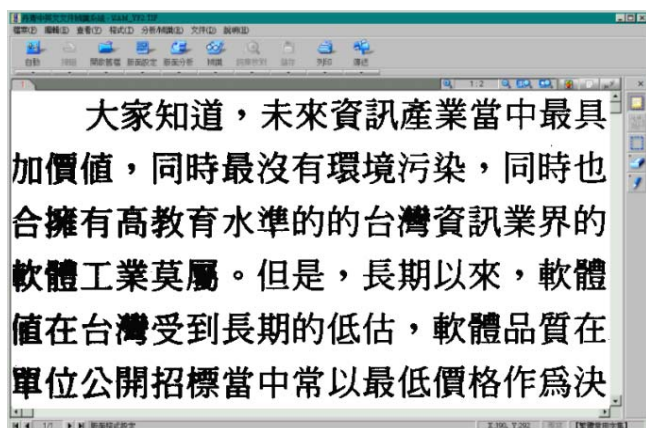
力新國際科技股份有限公司(20 point)

力新國際科技股份有限公司(22 point)

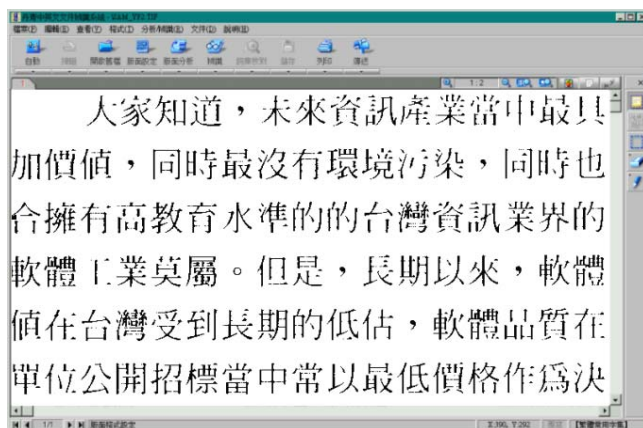
圖例二：標準



圖例三：太濃



圖例四：太淡



就實際例子而言，如果文件本身為白底黑字，則可利用掃描器驅動程式上的“自動明亮度調整”按鈕來為你設定掃描影像的明亮度值；若是文件為黃底黑字(如報紙類)，則可再增加亮度值，若是效果依舊不好，可再增加數個單位的值。

總體來說，當您覺得丹青系統的辨識品質未達到您的需求時，可試著調整影像的解析度與明亮度。其中明亮度是個經驗值，可能各家掃描器而不同。另外，試著給予辨識區塊屬性，如設定區塊屬性(有/無)漢字、設定區塊屬性(有/無)英文字母以及設定區塊屬性(有/無)數字等，也有助於系統的辨識。