

1、身份驗證API規格

1.1 註冊帳號（限開博）

1.2 使用者登入

登入到中研院數位文化中心OCR系統進行身份驗證，以驗證登入帳號是否取得授權使用系統的權限。登入成功後token的有效時間最長為24小時，逾時則失效。

■ 網址：https://ocr.ascdc.tw/web_api/auth.php

■ 呼叫方式：POST

■ 傳入參數

參數名稱	類型	描述說明
account	String	傳入登入者的帳號
password	String	傳入登入者的密碼

■ 回傳資料格式：JSON

回傳訊息中各項資料項目說明如下。

參數名稱	類型	描述說明
status	Number	API回傳狀態值 200表示登入成功 101表示資料庫連線失敗 102表示帳號或密碼錯誤 108表示發生不可預期的錯誤
message	String	額外除錯訊息(有錯誤才會回傳)
token	String	JWT (內含使用者資訊)
uid	Number	使用者uid
account	String	使用者帳號
name	String	使用者姓名
oid	Number	所屬機構
iat	timestamp	token建立時間
exp	timestamp	token有效時間，每次登入都會給一組新的(有效期限1天)

回傳範例：

```
{
  "access_token": "eyJ0NiJ9.eyJpc3M ...",
  "status": 200
}
```

2、影像辨識API規格

2.1. 新增我的藏書

新增一本藏書。

- 網址：https://ocr.ascdc.tw/web_api/create_book.php

- 呼叫方式：POST

- 傳入參數

參數名稱	類型	描述說明
token	String	傳入登入認證金鑰
title	String	傳入書名
author	String	傳入作者
public	Number	傳入開放程度,1表示公開,0表示私人
orientation	Number	傳入文字方向,1表示直書, 0表示橫書, 2表示自動偵測
傳入範例： { "token": "16543fac07fc8d58", "title": "古今圖書集成", "author": "陳夢雷", "public": 0, "orientation": 2 }		

- 回傳資料格式：JSON

回傳訊息中各項資料項目說明如下。

參數名稱	類型	描述說明
status	String	回傳狀態值 200表示新增藏書成功 101表示資料庫連線失敗 102 表示新增藏書失敗 JWT驗證token是否有效、是否被竄改，若回傳108且搭配 message 'Expired token'，可重新登入取得token 108表示發生不可預期的錯誤
bookid	Number	回傳藏書的編號，以作為之後上傳書頁之用
message	String	額外除錯訊息
{ "status": 200, " bookid ": 495 }		

2.2. 上傳影像檔到伺服器

上傳欲辨識的影像檔到數位文化中心的OCR伺服器，上傳書頁，一次上傳一張，可上傳之檔案格式包含：JPG、PNG、PDF 及 ZIP，檔案名稱僅接受半形之大小寫英文、數字、減號 - 及底線 _。

■ 網址：https://ocr.ascdc.tw/web_api/upload.php

■ 呼叫方式：POST

■ 傳入參數

參數名稱	類型	描述說明
token	String	傳入登入認證金鑰
page	File	傳入上傳的影像檔
bookid	Number	傳入藏書的編號
block_order	String	TBRL: 上到下右到左 (右到左的直書, 預設) TBLR: 上到下左到右 (左到右的直書) RLTB: 右到左上到下 (右到左的橫書) LRTB: 左到右上到下 (左到右的橫書)
language	Number (可選)	1: 中文 (預設) 5: 舊字體日文
orientation	Number (可選)	0: 橫書 1: 直書 2: 自動 (預設)
correction	Number (可選)	0: 史 1: 道 2: 佛 3: 醫藥 4: 現代辭典 -1: 不使用 (預設)
pages_per_img	Number (可選)	1: 1 頁 (預設) 2: 2 頁
remove_margin	Number (可選)	-1: 不進行額外前處理 (預設) 1: 有掃描時的黑色背景 2: 有掃描時的白色背景
has_mark	Boolean (可選)	有標點符號 (預設 false)
has_alphabet	Boolean (可選)	有橫式英文 (預設 false) orientation 為橫書才會生效
remove_anno	Boolean (可選)	移除註記符號 (預設 false)
is_inverted	Boolean (可選)	黑底白字 (預設 false)

■ 回傳資料格式：JSON

回傳訊息中各項資料項目說明如下。

參數名稱	類型	描述說明
status	Number	回傳狀態值 200表示上傳檔案成功 101表示資料庫連線失敗 102 表示上傳檔案失敗 JWT驗證token是否有效、是否被竄改，若回傳108且搭配 message 'Expired token'，可重新登入取得token 108表示發生不可預期的錯誤
queue_id	Number	回傳佇列唯一識別碼，以作為之後查詢辨識進度之用
<pre>{ "status": 200, "queue_id": 1666 }</pre>		

2.3. 查詢辨識進度

傳入檔案唯一識別碼查詢指定檔案的OCR辨識進度，或已辨識完成則回傳該檔案辨識出的文字內容資料

- 網址：https://ocr.ascdc.tw/web_api/queue.php

- 呼叫方式：POST

- 傳入參數

參數名稱	類型	描述說明
queue_id	Number	佇列識別碼
token	String	以Authorization: Bearer token方式夾帶

- 回傳資料格式：JSON Array

回傳訊息中各項資料項目說明如下。

參數名稱	類型	描述說明
status	Number	回傳狀態值 200表示成功 101表示資料庫連線失敗 102 表示查詢失敗或不存在

參數名稱	類型	描述說明
		103 表示等待處理中 108表示發生不可預期的錯誤
guids	Number	回傳該檔案唯一識別碼，作為之後查詢辨識文字內容資料用 若上傳PDF或ZIP會產生一個queue_id對應到多張書頁的狀況，因此回傳型態為陣列格式
<pre>{ "status": 200, "guids": [{ "guid": 1666 }, { "guid": 1667 }] }</pre>		

2.4. 取得辨識結果

傳入檔案唯一識別碼回傳該檔案辨識出的文字內容資料

■ 網址：https://ocr.ascdc.tw/web_api/query.php

■ 呼叫方式：POST

■ 傳入參數

參數名稱	類型	描述說明
guid	Number	傳入檔案唯一識別碼，此參數為URL的一部份
token	String	傳入登入認證金鑰 以Authorization: Bearer token方式夾帶

■ 回傳資料格式：JSON

回傳訊息中各項資料項目說明如下。

參數名稱	類型	描述說明
status	Number	回傳狀態值 200表示辨識成功 101表示資料庫連線失敗 102 表示沒有相關資料 JWT驗證token是否有效、是否被竄改，若回傳108且搭配message 'Expired token'，可重新登入取得token 108表示發生錯誤
result	Object	查詢結果
id	Number	此文字框的 id，每個文字框的 id 皆為唯一值
block_id	Number	該文字所屬的 block 之 id，每個 block 的 id 需要為唯一值
line_group_id	Number	該文字所屬的 line_group 之 id，每個 line_group 的 id 皆為唯一值
line_id	Number	該文字所屬的 line 之 id，每個 line 的 id 皆為唯一值
text	String	圖上的字元，經過校正後可能會和欄位 options 的候選字元不相同
vertical	Number	該文字所屬的 block 是屬於 直書還是 橫書，直書為 1，橫書為 0

參數名稱	類型	描述說明
x	Number	字左上角的 X 座標
y	Number	字左上角的 Y 座標
width	Number	字的寬度(像素)
height	Number	字的高度(像素)
options	JSONArray	該文字框的候選字列表，列表中的每一個項目皆有兩個值，第一個值是候選字元，第二個值是該候選字元的信心分數，例如： [["測", 1.0], ["則", 0.8], ...]
回傳範例：(請另見附檔：校正平台_對外API_回傳結果_規格說明.docx)		

2.5. 取得轉檔後的圖片

取得轉檔後的圖片

- 網址：https://ocr.ascdc.tw/web_api/get_image.php

- 呼叫方式：POST

- 傳入參數

參數名稱	類型	描述說明
guid	Number	傳入檔案唯一識別碼，此參數為URL的一部份
token	String	傳入登入認證金鑰 以Authorization: Bearer token方式夾帶

- 回傳資料格式：Content-Type: image/jpeg 圖檔