Github專案倉庫

<https://github.com/shinkia1985/->

# 

# **財政部電子發票 API 測試工具**

歡迎 PR 財政部的 API 清單

前端：Angular 4.0

* 後端：ASP.NET NetCore 2.0 - ASP.NET Web Api
* Live Demo：<https://exfast.me/einvoiceApiTest>
* SourceCode：<https://github.com/shuangrain/EinvoiceApiTest>
* Blog：<https://blog.exfast.me/2017/12/sideproject-ministry-of-finance-e-invoicing-api-testing-tools/>

## **操作說明**

* 環境變數
  1. 簽章 (Signature)：將 Request 參數按照名稱排序並以 HMAC-SHA1 加密後加在 Query 後方傳送至財政部
  2. 測試環境：將資料傳送至[財政部測試環境](https://wwwtest.einvoice.nat.gov.tw/)
  3. Client 模式：以另開新頁的方式呼叫財政部 API
* 選擇 Api  
    
  要測試的 API
* 財政部加密用 Key  
    
  財政部提供的 HMAC-SHA1 加密用 AppKey
* Api 位置  
    
  呼叫的 API 位置
* Request  
    
  傳送給財政部的資料

## **使用方法**

1. 開啟 /src/appsettings.json
2. 將財政部提供的 AppKey 與 AppID 填上去
3. 用 Visual Studio 2017 開啟並執行即可 !!

## **Api 新增**

開啟 /src/Json/ApiList.json 且依照下面的參數新增 Api

#### **參數介紹**

| **參數名稱** | **參數說明** |
| --- | --- |
| TypeName | Api 類型 (依此欄位排序) |
| ApiName | Api 名稱 |
| ApiUrl | Api 位置 |
| Param | Api 需傳送的參數範本 (參數 timeStamp, expTimeStamp, uuid, appID 會被自動取代) |

OCR識別

## **光學字元辨識運作方式**

OCR 軟體或引擎的運作會透過一組步驟進行。

**1.影像分析，** 掃描器會讀取文件，並將其轉變為二進位資料。OCR 軟體將檢測掃描的檔案，並將亮色區域分類為背景，將深色區域分類為文字。

**2.分析前作業，** OCR 技術會透過一些不同的技術，使影像更加完善：

* 平滑處理文字影像的邊緣，以及消除數位影像污點。
* 透過傾斜掃描文件，修正在掃描期間發生的對齊問題。
* 多語言 OCR 技術的字跡辨識
* 整理影像中的線條與方塊。

**3.辨識文字，** OCR 技術會使用特徵擷取與圖樣比對來處理文字：

* 特徵擷取會將語言分解成幾個部分，例如封閉環、線條、線條方向與線條交點。從這裡開始，它將會使用這些部分來搜尋最符合或是最接近的結果。
* 圖樣比對則是指被稱為「字符」的字元影像完成分離，並與已經存檔的類似字符相互比較。只有在已存入的字符中有類似於所加入字符的尺寸與字型時，圖樣辨識作業才會進行。當影像掃描來源的文件字元為已知字型時，這種方法的發揮效果最好。

**4.處理後作業，** 當內容完成分析之後，系統會將擷取的文字資料轉變為電腦檔案。特定 OCR 軟體可以製作掃描文件前後版本的已註解的 PDF。如果 OCR 未能辨識文字，請務必檢查掃描是否為高品質、光線是否充足，且掃描內容並未傾斜。

## **OCR 技術的類型**

資料科學家將根據其應用範圍與使用，區分不同類型的 OCR 軟體類型。以下是幾個範例：

* 簡易光學字元辨識軟體會儲存不同的文字和字型影像圖樣為範本。這個軟體會使用圖樣比對演算法，找出文字影像之間的差異。它將在內部資料庫中逐字元進行分析。系統逐字詞複製文字就是光學字詞辨識。這項技術不可能擷取到每種字型和手寫樣式，因為這兩個要素都有無限數量，所以這個解決方案有其限制。
* 智慧字元辨識 (ICR) 軟體是現代 OCR 技術的其中一部分。ICR 會以人類方式進行文字閱讀。使用機器學習軟體時，機器可以如訓練成如人類一般的行為。被稱為神經網路的機器學習系統能研究文字，並重複處理影像。它會搜尋多種影像層面，例如，線條、曲線、環圈和交點，最後合併不同資料層面的結果，得到最後結論。
* 智慧字詞辨識技術的運作規則與 ICR 相同，但是這些技術會研究整個字詞，而不用先將影像修改成字元。
* 光學標記辨識會尋找文件中的浮水印、標誌和其他文字符號。

相關OCR軟體

### **1.Adobe Acrobat Pro**

Acrobat Pro 可為您提供所需的所有 OCR 工具，讓您能夠簡化工作流程並確保文件管理效率。使用 Pro 版的 Acrobat 時，您可使用所有基本 OCR 功能以及對文件加上注釋和留下回饋意見，比較兩份文件的選項，掃描表格的特殊工具，以及其他功能。在掃描完成後幾秒鐘，文件便可在您的電腦上進行調整。Acrobat OCR 可完美搭配免費 Adobe Scan 應用程式使用，讓您掃描文件，並且轉換成 PDF。藉由 Adobe OCR 工具的協助，文字將自動進行辨識，而且您還可以視需要進行調整。

### **2.OmniPage Ultimate**

這個軟體以高轉換準確性聞名業界。OmniPage Ultimate 可建立自訂工作流程，讓文件自動以正確格式傳送到正確位置

### **3.Abbyy FineReader**

Abbyy FineReader 提供在轉換紙本文件至數位文件時的所有必要工具。這個軟體可以辨識文字，並將結果轉換為 PDF、不同的 Microsoft Office 格式與其他格式。您可以比較文件、加上註解、注釋和其他標示。這套軟體可以批次轉換大量文件，並可處理許多輸出格式及 192 種不同語言。

### **4.Readiris**

Readiris 可支援許多不同的檔案格式，以及在文件中加入簽章與安全性保護，還有注釋、浮水印與註解等功能。

### **5. Rossum**

這套 OCR 解決方案可用於掃描發票上的重要資訊，並匯出至您使用的任何程式。這套軟體使用 AI 來掃描文件上的重要資訊，而非使用範本格式進行掃描。這個做法很有用，因為不同的發票可能有不同的資訊呈現格式。

## **真實世界的 OCR 應用**

有許多種 OCR 軟體和工具可因應不同日常的個人和專業需求。

* 版面分析軟體會將掃描的文件區分成適合使用 OCR 的不同區段。
* OCR 引擎會識別字元。
* 可在其他軟體中加入 OCR 能力的軟體開發套件。範例包括文件影像處理管理系統、記錄管理解決方案，或是表格處理應用程式。
* 在一或多個 OCR 引擎上的圖形化介面。

## 

## **OCR 如何協助簡化流程及提高效率**

人力、時間和成本都能透過 OCR 軟體和技術大幅降低。將靜態文件轉換成數位、智慧型、可搜尋的文件時，可讓資訊搜尋更簡單，因此利於公司為客戶與員工提供更棒的經驗。將 OCR 技術整合至企業模式之後，您可以：

* 轉換手寫文件成可編輯數位文件。
* 掃描表格(手寫內容)，以供自動檢閱、驗證、分析與編輯。節省花在手動處理文件和資料輸入的時間。
* 尋找重要文件時可採用快速搜尋資料庫中的詞組，而且不需要手動翻閱實體檔案。
* 資料獲得安全保障與集中管理，因為數位文件發生遺失或遭竊的機會較低。
* 獲取最準確且最新的資訊。

透過 OCR 技術，方便性與存取性都大為提高

**※ORC技術部分，比較偏向ICR智慧字元辨別，讓AI學習，可以訓練成如人類一般的行為。能研究文字，並重複處理影像。**

### **QR碼的原理**

### QR碼（Quick Response Code）是日本Denso Wave公司於1994年開發的二維條碼識別碼，為重視高速讀取的「進化形條碼」。尤其是在手機鏡頭可以讀取之後，快速普及開來 [3]。

### 稱它「進化形」的原因，來自幾個特色：比起條碼，它所含的資訊量大增， 而且髒汙、破損就算達到30％，數據也能修復，而且它的標示只需要條碼的十分之一面積。

### 以前條碼處理的資訊量最多20行，相較之下QR碼因有二維方向的條碼可做變化，容量極大，最大資料容量是7089字（只含數字時）。數據修復稱為「錯誤訂正功能」，因QR碼內特殊圖案設計和辨識定位標記有容錯能力，QR碼圖形如果有破損，仍然可以被機器讀取內容。製作數據的人，也可以選擇可修復的標籤。

### QR碼被運用在各種不同的領域， 除了官方網站的網址、簡單取得特賣資訊等促銷用途之外，也應用在電子票卡或機場的登機證。在社群媒體上，也提供以QR碼與粉絲或朋友共享訊息的服務。

### 最近，也能看到與插圖或照片組合、設計性高的QR碼，任何人都能製作它，各位不妨來挑戰看看。

### **附註**

### [1] 台灣所使用的商品條碼為EAN碼，全名為European Article Number（歐洲商品條碼），台灣在1985年加入EAN會員，現在買東西結帳時，服務人員掃瞄商品上的條碼就是EAN條碼。台灣國碼為471。

### [2] 本書封底也印有ISBN碼，不妨參考看看。台灣的ISBN條碼第二段為957或986。

### [3] Denso Wave公司雖然保有QR碼的專利，但並未行使權利，而是將格式公開，讓任何人都能使用它。因此現在全世界不用花成本便能使用這種編碼。

### 

### QR Code：Quick Response Code，快速響應矩陣圖碼

QR Code是Quick Response Code的縮寫，因為發明者希望QR Code可以快速解碼內容，故以「Quick Response(快速反應)」命名。QR Code的發明可追溯至1994年的日本，由汽車零件業者Denso Wave公司發明，用來追蹤生產線上的貨物，並且讓條碼能容納日文漢字、假名訊息。

與一維條碼最大的差異在於，二維條碼能夠承載更多資料，而QR Code能儲存的資訊種類囊括了數字、字母、二進制(位元組)、以及日文、漢字。可以這麼說，一維條碼代表的資訊以「數字編號」為多，號碼用來識別商品，但QR Code可以讓日文、漢字資料直接記錄在條碼裡，也就是讓條碼代表一句話、或是一串網址。

### QR Code標準在1999年發佈 應用包含物流、驗票、網站登錄

日本的QR Code標準在1999年公佈，而其對應的ISO國際標準(ISO/IEC 18004)則在2000年獲得。ISO(國際標準化組織)認定QR Code為國際規格，其規格公開，雖由Denso Wave公司持有專利，但不會被執行。

而除了能記錄日文與漢字之外，QR Code的辨識速度快、360度全方位都能被識別、容錯能力高、佔位面積更小，讓它被廣泛運用在產品跟蹤、物品識別、文件管理等方面。除此之外個人名片、網站登錄等地方也常常見到QR Code的蹤跡。

不過，QR Code並非任何情況下都能夠辨識，比如說QR Code過度擠壓變形、印刷不夠清晰、或是外框沒有足夠的白邊的話，就會造成識讀困難。雖然如此，和傳統一維條碼相較，QR Code可以360度掃描辨識，容錯能力更是讓它在視覺造型上有了更多的變化。

### 有40種尺寸與4種容錯級別 最多能容納7089字元

先不論微型QR Code等特殊變形，一般常見的QR Code共有40種不同的尺寸版本，版本依據儲存密度與結構區分，每一種QR Code的組成結構、能夠儲存的資訊量多寡都不完全相同。QR Code的尺寸版本以「 Version」標示，如版本1，也就是 Version 1，QR Code尺寸為21x21位元， Version 2的尺寸為25x25位元，Version 40則為177x177位元，Version越高，QR Code的尺寸越大，能夠承載的資訊量也就越多。

是否有看過融入標誌的QR Code呢？「容錯」又稱「糾錯」，是指當某部分資訊缺失的時候，其它部分會自動填補，讓整體資訊依然可以完整辨識，容忍錯誤發生。簡單來說，就是當QR Code條碼缺了一角，或是沒有辨識完全，資訊依然可以被掃瞄出來，這就是QR Code容錯能力的表現，是一維條碼辦不到的。

QR Code的4個容錯等級：

* L(低)：可修正7%的字碼。
* M(中)：可修正15%的字碼。
* Q(中高)：可修正25%的字碼。
* H(高)：可修正30%的字碼。

容錯能力越高，同等級的QR Code能夠容量的資訊量就越少，這是因為將原本能夠儲存資料的地方，拿去做容錯備份，於是一般儲存資料的空間被壓縮，能夠儲存的資料也就變少。由以上訊息可以得知，資訊容量最大的QR Code版本為Version 40、容錯能力為L等級，此版本最多可以容納7089個數字，或是4296個字母，或1800個漢字。一般的一維條碼最多只能輸入20個資訊量。



▲QR Code由無數個黑白小格子組成，每一個格子就是一個位元(bits)。由上圖可知，組成一個純數字(Numeric only)需要3又1/3個位元，也就是3又1/3個小格子，至於要組成一個ISO 8859-1編碼的二進制字元，則需要8 bits，也就是8個小格子。以資訊容量最大的Version 40、容錯能力為L等級的QR Code來說，純數字最多能夠容納7089個，也就是能夠表示7089個0至9的數字。

**※透過QR Code掃描技術可快速將發票號碼快速複製或上傳，但還是需訓練AI只需擷取所需要的數字資訊部分**