

# 應用網路爬蟲技術於網路食品平台資訊之初探

郭岳翰 王兆儀

衛生福利部食品藥物管理署食藥戰情中心

## 摘要

在後疫情時代網購商業型態興起，為強化網路通路之食安管理效能，透過運用網路爬蟲技術及食品雲鏈結挖掘高關注之食品與販售業者資訊，以利針對於網路潮流所帶來的食品營業新型態進行把關，防範於未然。鑑於不同網路平台對販售業者及消費者提供了不同的服務項目，填報資訊方面也不盡相同，本報告探討電商平台、美食外送平台及部落格三種類型平台之資訊概況、公開資料表格探討及擷取資料倉儲之設計。隨著資料量日益擴大，除考量人力、時間、成本或效益等經濟因素，更不可忽視其可能帶來的風險，其中法令遵循就是關鍵議題之一。倘若未來對系統性規劃及網路擷取資料等相關應用有需求，需先行評估是否有觸犯法律之疑慮，以落實風險管理。

**關鍵詞：**爬蟲、網路平台、食品安全管理

## 前言

在後疫情時代網購商業型態興起，除了線下常見食品販售銷路，如超商、購物中心、雜貨店等，民眾亦可於網路上獲得比實體通路更多更廣之資訊及商品，且購買容易及方便，因此網路的食安管理勢必納入管控食安風險的一環中。而在眾多可能造成食安風險的網路資訊中，尤以網路電商平台能影響的範圍更甚，因其提供消費者低成本的網路店面平台，可輕易的進行商品交易；除了網路提供的賣場平台之外，部落格也是民眾日常獲得食品資訊來源之一，部落客常提供前往餐廳之美食心得，或接受業者業務配合請求，於文章內容中介紹業者之食品及聯絡購買方式，而這些文章敘述往往對消費者產生莫大的影響力，相對於電商平台只顯示產品資訊，經由知名部落客描述經驗的

產品往往較吸引消費者衝動購買；另一須關注的食品販售平台為美食外送平台，此平台提供民眾在家隨時隨地瀏覽及購買居所附近之餐廳業者及市場業者提供之食品選項，在後疫情時代中，提供民眾不需出門也能解決每日飲食所需之便利性。

隨著網路逐漸成為日常必需品，民眾可隨時隨地瀏覽網頁，也因此無法忽視網路賣場的商機，而為了突顯差異化，網路賣場會使用不同的規範在網路中販賣商品，而這往往難以控管，無法得知每個商家的正確資訊可能會造成潛在的風險，尤以食品為電商中常見的販賣物品種類之一，無法得知食品流向可能造成食品安全上的疑慮，可能致使消費者暴露在高風險的食品安全當中。

為強化網路通路之食安管理效能，透過運用網路爬蟲技術及食品雲鏈結挖掘高關注之食

品與販售業者資訊，以對於網路潮流所帶來的食品營業新型態把關，防範於未然。另網路爬蟲資料擷取具有一定限制，主要因爬蟲係模擬人的操作模式，進而獲取網頁數據的一種程序。我們生活中幾乎每天都在應用爬蟲，如在Google中搜索到的內容幾乎都是爬蟲採集下來的，所以網路爬蟲作為一門技術，技術本身是不違法的，爬蟲本身不被法律禁止，可以採集對大眾、所有人公開的「公開資訊」，但用途須合理；惟如該網站有聲明禁止爬蟲採集或者轉載商業化時，網路爬蟲行為就具備一定的法律風險。爬蟲所帶來風險主要為違反網站意願，例如網站採取反爬措施後，強行突破其反爬措施、爬蟲干擾了被訪問網站的正常運營(如搶門票等)；此外在網路上的東西，無論是文字還是圖片，都受到法律的保護(如著作權法、刑法、公平交易法及個人資料保護法)，因此如果沒有經過同意，是不可以任意使用其資料的。一般人容易忽視爬蟲程式擷取的資訊是受到法律保護類型的數據或資訊。隨著資料經濟等崛起，透過工具擷取及運用相關資料或資訊，除考量人力、時間、成本或效益等經濟因素，更不可忽視其可能帶來的風險，其中法令遵循就是關鍵議題之一。倘若未來需系統性規劃及網路擷取資料等相關應用，需先行評估是否有觸犯法律之疑慮，另為避免後續造成不必要的困擾或損失，除研發、企劃等人員外，建議也應有法律專業人士加入，從專案開始前就應參與，以落實風險管理<sup>(1)</sup>。

## 運用科技偵測網路販售食品平台

本研究初步以Python套件BeautifulSoup及Selenium爬蟲技術針對不同網路平台，所公開之圖片及商品、店家文字段落等相關可能隱藏業者資訊之區塊(如商品描述區、商品規格區、賣場描述及文章內容等)進行資料擷取並儲存，其中包括電商平台：Shopee蝦皮購物、

PChome線上購物、momo購物網、露天拍賣、Yahoo奇摩購物中心、博客來、Rakuten台灣樂天、生活市集、Pcone松果購物、東森購物；美食外送平台：foodpanda、Uber-Eats；部落客：愛食記、痞客邦、隨意窩三大類型共17個網站資料，針對網路平台擷取資料及限制於下節說明。

## 研析各平台資訊以利鎖定高關注議題

本研究針對電商平台、美食外送平台及部落格資料分析。不同網路平台對販售業者及消費者提供了不同的服務項目，例如像momo購物網、PChome線上購物等由中心統一控管之網路平台，消費者無法得知販賣者太多資訊，平台也不提供評價資訊，而是以良好的售後服務及維護商品高品質來取代，使消費者信任而安心購買；而在蝦皮、露天拍賣等平台則是提供評價、星級等方式，使消費者能查看其他已購買過的消費者的購買經驗及平台使用心得，來決定是否於該平台購物，因此每個平台能擷取的資料不一，唯一共通點即是販售商品皆有文字敘述。

以下針對網路平台之電商平台、美食外送平台及部落格平台資訊概況研析。

### 一、電商平台之資訊概況分析

透過電商平台之商品資訊概況分析，將十大網購平台商品資訊(表一)，以平台有無預設商品資訊之表格作為初步判斷商品資訊是否完整的指標，發現有預設商品資訊表格之網購平台，因其刊登之資訊有固定填寫欄位，整體填報資訊較無預設商品資訊表格之網購平台完整。而無預設商品資訊表格之網購平台，普遍在商品或網頁中提及其市招名稱、描述商品之文字段落及圖片，以下將透過區分有無預設商品資訊表格，及利用不同資料擷取模式進行網路爬蟲，分析不同平台間業者填報及業者於不

表一、十大電商平台之商品資訊分析表

平台	網站有無預設商品資訊表格 <sup>a</sup>	無預設商品資訊表格之資訊提供形式		
		商品名稱	文字段落	圖片
Shopee蝦皮購物	X	O	O	-
露天拍賣	X	O	-	O
PChome線上購物	O	-	-	-
momo購物網	O	-	-	-
Yahoo奇摩購物中心	O	-	-	-
博客來	O	-	-	-
生活市集	X	O	-	O
東森購物	O	-	-	-
Rakuten台灣樂天	O	-	-	-
Pccone松果購物	X	O	O	O

<sup>a</sup> 十大網購平台之商品資訊評估：部分平台內資訊填寫位置不一致，以多數業者填寫之欄位為主。

同平台重複之情形。

(一)電商平台資訊

依據2017年12月25日衛生福利部(下稱衛

福部)公告之「以通訊交易方式訂定之食品或餐飲服務定型化契約應記載及不得記載事項」法規<sup>(2)</sup>，其中應記載事項包含「企業經營者資訊」、「定型化契約解釋原則」、「商品資訊」、「付款方式說明」、「契約履行及確認機制」、「商品交付地、交付日期及交付方式」、「商品訂購數量上限」、「受領物之檢視義務」、「消費爭議之處理」及「訴訟管轄」，本次分析僅就「企業經營者資訊」及「商品資訊」探討。以販賣食品而言，平台的資訊可分為「食品商品資訊」及「食品業者資訊」(表二)，「食品商品資訊」包含品名、內容物名稱及淨量、容量或數量、食品添加物、原產地、有效日期及交易總價等資訊；「食品業者資訊」則包含業者名稱、地址、電話、食品業者登錄字號等。

細部探討10大網購平台填寫產品或業者資訊之情形，「食品商品資訊」部分，有預設商品資訊表格之6家網購平台中，多數

表二、十大電商平台提供之商品(食品)資訊表

平台 <sup>ab</sup>	電商平台提供之商品(食品)資訊評估										
	有無預設商品資訊表格	食品商品資訊						食品業者資訊			
	品名	內容物名稱及淨量、容量或數量	食品添加物	原產地	有效日期	交易總價	名稱	地址	電話	食登字號	
Shopee蝦皮購物	X	O	O	X	O	X	O	X	X	X	X
露天拍賣	X	O	O	X	X	X	O	X	X	X	X
PChome線上購物	O	O	O	X	O	O	O	O	O	O	O
momo購物網	O	O	O	O	O	O	O	O	O	O	O
Yahoo奇摩購物中心	O	O	O	X	O	O	O	O	O	X	O
博客來	O	O	O	X	O	O	O	O	O	O	O
生活市集	X	O	O	X	O	O	O	X	X	X	X
東森購物	O	O	O	O	O	O	O	O	O	O	O
Rakuten台灣樂天	O	O	O	X	X	X	O	O	O	O	X
Pccone松果購物	X	O	O	X	X	X	O	X	X	X	X

<sup>a</sup> 部分平台內資訊填寫位置不一致，以多數業者填寫之欄位為主；shopee蝦皮購物、露天拍賣

<sup>b</sup> 平台僅提供該刊登業者之帳號，故後續不納入刊登者帳號進行業者資訊評估。

平台之預設表格皆有填寫上述食品商品資訊，其中「Rakuten台灣樂天」於原產地及有效日期之欄位仍有資訊缺漏情形；而在無預設商品資訊表格之4家網購平台中，除「生活市集」之商品資訊相對較為完整外，其餘3家皆有部分商品資訊未填寫之情形。「食品業者資訊」部分，在有預設商品資訊表格之6家網購平台中「Rakuten台灣樂天」無提供「食品業者登錄字號」(表格中簡稱:食登字號)，及「Yahoo奇摩購物中心」無提供「電話」等部分資訊缺漏；而無預設商品資訊表格之4家網購平台，皆無特定欄位提供食品業者相關資訊。綜整而言「食品商品資訊」方面，Shopee蝦皮購物、露天拍賣、Pcone松果購物及Rakuten台灣樂天等平台之產品資訊揭露較不完整，而「食品業者資訊」方面Shopee蝦皮購物、露天拍賣、生活市集及Pcone松果購物之平台設計過於仰賴食品業者自行登打，其資訊揭露亦不完整，倘發生食品事件時，不易對問題商品追蹤溯源，保護消費者力度相對較低。

## (二)電商平台資料倉儲設計

為利網路食品平台資料擷取後進行業者資訊比對，不同網路平台針對販售業者及消費者提供不同的服務方式，如知名電商平台momo及PChome等，皆是由後台統一控管之平台，消費者無法直接於網站上得知販售業者詳細資訊，也多不提供消費者評價等資訊，而是轉以完善售後服務及維護商品品質來取代，使消費者信任而安心購買；蝦皮、露天拍賣平台，其提供消費者評價、商品評分等方式，使消費者能透過其他消費者之購買評價及使用心得來選擇商品，因此針對各平台擷取資料皆不一致，唯一共通點為販售商品皆有文字敘述，故於設計擷取各平台之欄位需分別設

計，以利後續應用之完整性。針對電商平台之公開資訊及相關可能隱藏業者資訊區塊，進行完整擷取並於資料庫中儲存，下表三為電商平台欄位設計之資料表。

表三、電商平台資料表欄位設計

資料來源	欄位	內容範例
店家	賣場識別碼	S+2碼來源網站代碼+7碼流水號 S010000001
店家	來源網站	O皮購物
店家	來源網址	https://shOOec.tw/t5OO12?categoryId=66&itemId=18OOO
店家	賣場描述區文字	你好~ 農曆年前訂單截止為OOO，也就是2月9號早上10點後的訂單，2月16日依序出貨。 我們是桃園八OOO場準備邁向40年的老店...
店家	賣場描述區圖片(多筆)	二進位資料(Binary)
店家	商家名稱	長OO鬆店
店家	賣場評價	5.0
店家	賣場評價則數	8,299
店家	聊聊表現	96%
店家	粉絲數	609
店家	資料擷取日期	YYYYMMDD
店家	食品業者登錄字號	A-0000000-00000-0
店家	地址	臺O市大O區敦OO路二段100號OO樓
店家	電話	02-2OO0-0000
店家	統一編號	27OOO489
店家	品牌	長OO鬆

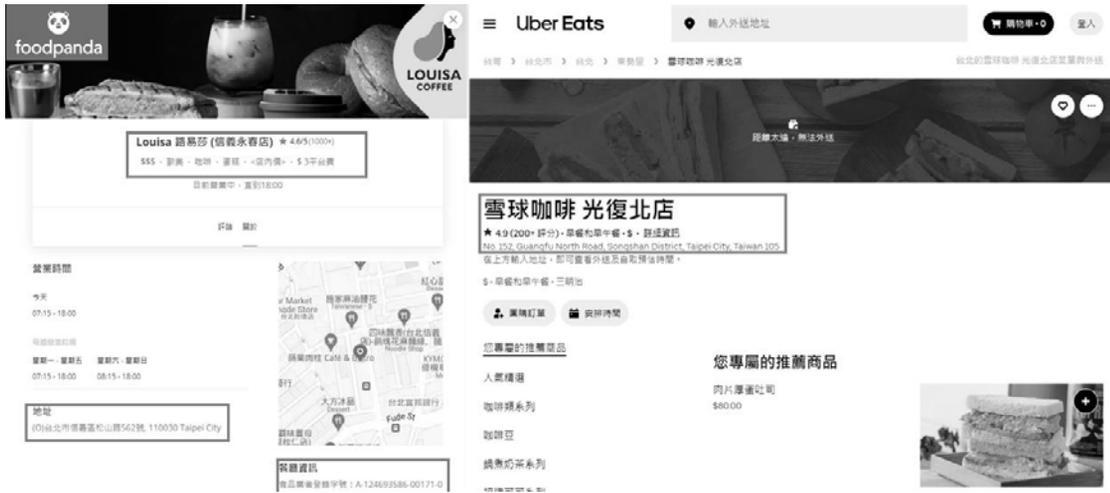
## 二、美食外送平台之資訊概況分析

為因應衛福部2019年4月26日公告修正「應申請登錄使得營業之食品業者類別、規模及實施日期」<sup>(2)</sup>，食品物流業自2019年7月1日起應辦理食品業者登錄使得營業，「美食外送平台業」如有參與或提供食品運送服務等

表四、美食外送平台資訊分析表

平台	固定格式	無預設商品資訊表格之資訊提供形式						
		店家名稱	店家地址 <sup>a</sup>	評論數	評分	餐飲分類	營業時間	非登字號 <sup>a</sup>
foodpanda	○	○	○	○	○	○	○	▲
Uber-Eats	○	○	●	○	○	○	○	▲

<sup>a</sup> ●有英文地址；▲平台之資訊為業者自行登入，並非所有業者皆有提供非登字號。



圖一、美食外送平台網站資訊

行為，即屬食品安全衛生管理法規範之「食品物流業」，應依法辦理登錄，並符合該法之相關規定，本次就foodpanda及Uber-Eats兩家美食外送平台之公開資料研析發現，其格式大致相似，惟主要差別為中英文地址，如表四、圖一。

美食外送平台資料擷取差異不大，主要為店家地址有中英差別(如表五)，故於設計擷取平台之欄位皆需完整保留原始資訊，以利後續需進行英文轉譯及分析。針對美食外送平台之公開資訊及相關可能隱藏業者資訊區塊，進行完整擷取並於資料庫中儲存，下表五為美食外送平台欄位設計的資料表。

表五、美食外送平台資料表欄位設計

資料來源	表格欄位	範例
店家	主鍵	1
店家	店家名稱	麥OO國OO號店
店家	食品業者登錄字號	A-2000005-0000001-8
店家	地址	臺O市OO區長O路100-2號O樓, Taipei, 104
店家	類別	早餐和早午餐
店家	評分	4.9
店家	評分筆數	495
商品	主鍵	1
商品	商品名稱	經典OOO牛O起OO堡
商品	商品敘述	自製手OO肉 加上OO司配 上爽口美生 ----- 原價：225

### 三、部落格之資訊概況分析

部落格之定位是以主題性為出發點，並以此為中心吸引相同族群來重複瀏覽，最後達成讓消費者透過關鍵字聯想到常看的部落格；另部分消費者也可直接搜尋自己想要的關鍵字來找尋符合自己需求的餐廳，或搜尋餐廳名稱直接看該餐廳的食記及網友評價，來滿足需求。故如何透過觀察消費者行為規劃，未來資訊擷取成為主要目標。針對部落格之公開資訊及相關可能隱藏業者資訊區塊，進行完整擷取並於資料庫中儲存，下表六與圖二為部落格欄位設計的資料表。

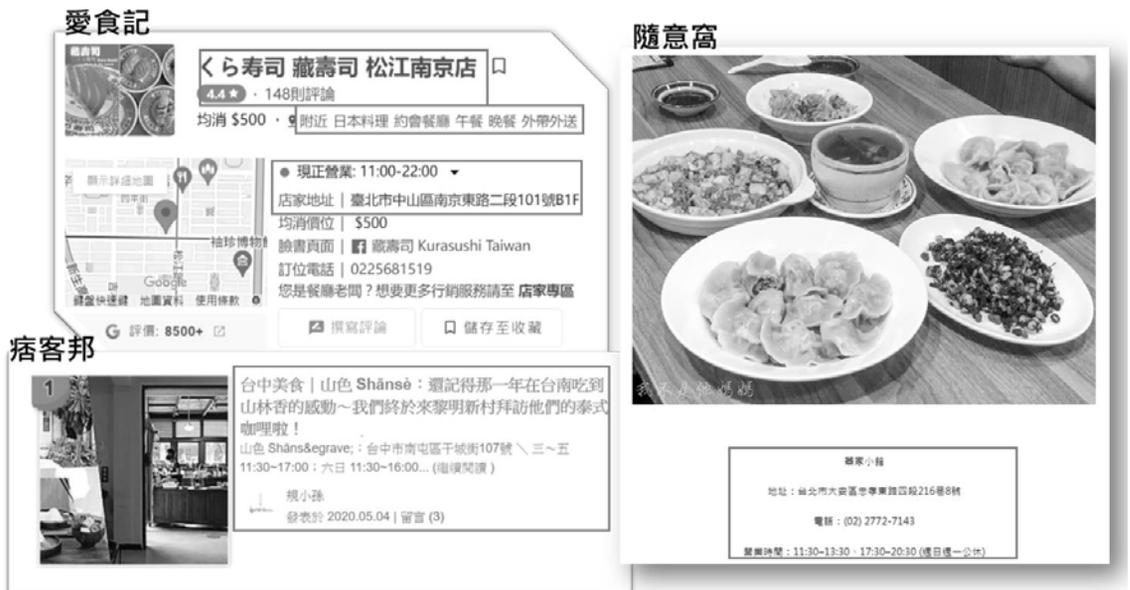
部落格資訊主要針對部落客所撰寫與推薦內容之公開完整資訊擷取並納入資料庫中儲存，下表七為部落格欄位設計的資料表。

### 建立資料庫挖掘新興議題，拓廣風險 監測範圍

針對上述不同平台類型中不同的產品及業者資訊區塊進行分析，將其設計為三種資料表，分別對應到電商平台、美食外送平台及部落格，並針對各平台欄位擷取之資料以完整資訊進行資料結構化及欄位規整，建立爬蟲資

表六、部落格之資訊分析表

平台	固定格式	無預設商品資訊表格之資訊提供形式					
		店家名稱	店家地址	評論數	評分	文章分類	營業時間
愛食記	O	O	O	O	O	O	O
痞客邦	O	O	O	O	X	X	▲
隨意窩	X	O	O	O	X	X	▲



圖二、部落格網站資訊

表七、部落格資料表欄位設計

表格欄位	範例
文章識別碼	B+2碼來源網站代碼+7碼流水號 B01000001
來源網站	痞客邦
來源網址	<a href="https://ange00000y.tw/blog/post/50000001?utm_source=PIXNET&amp;utm_medium=Hashtag_article">https://ange00000y.tw/blog/post/50000001?utm_source=PIXNET&amp;utm_medium=Hashtag_article</a>
文章標題	秒殺00000餅，還來不及起00000賣光「合00000店」、超上00000
文章內容文字	喜歡吃傳統炭烤00000朋友應該知道著台00000餅店「合成00000店」 民生社區新中街必吃小吃、銅板美食,聞著炭烤香氣就知道到了…
文章內容圖片 (多筆)	二進位資料(Binary)
部落客名稱	梅O(A00000y)享樂日記
閱讀次數	36,362

料庫以利於後續進行業者比對，期能解決不同資料擷取時間之業者重複性、停歇業等相關問題，滾動式更新業者資訊，強化擷取時效及比對之效能。未來除食品外，可針對本署所管轄之藥品、醫療器材及化粧品，進行擷取，並透過相關法規規範，協助判斷如於網路販售之合法性或業者是否領有執照等議題，建立違規判斷機制，期能達成未來應用之完整性，以及對網路販售之風險把關。

### 未來展望

在現今網路世代，數據種類多樣化、數量不斷增加且產生速度越來越快，但這些隨之而來的巨量資料，如果能有效處理及應用，當前人力無法有效解決的問題或許將有機會迎刃而解。

鑒於現今，只要資訊在網頁上是公開可用的，任何人皆可應用網路爬蟲等相關技術來獲取所需的數據。網路爬蟲讓複雜的數據變得可以搜尋，於現行大數據時代中資訊變得更容易獲取，於此也推動了新的市場和商業機會，同時也伴隨著法律方面的擔憂及衍生了另外的

管理問題。未來爬蟲技術仍有很大的發展，如何在未有觸犯法律疑慮之情況下靈活應用，管理之手法亦有與時俱進之必要。

### 參考文獻

1. 陳宏志。2020。網路爬蟲排除協議初探資料擷取仍有法可循。[<https://www.netadmin.com.tw/netadmin/zh-tw/viewpoint/6AEF7E7ECAD64EC1B21BFC54D45D6C88>]。
2. 衛生福利部。2017。以通訊交易方式訂定之食品或餐飲服務定型化契約應記載及不得記載事項。106年12月25日衛授食字第1061303394號公告。[<https://www.ey.gov.tw/Page/DFB720D019CCCB0A/8acac077-aa47-4d3c-a48d-db2928641378>]。
3. 衛生福利部。2021。應申請登錄始得營業之食品業者類別、規模及實施日期。110年4月28日衛授食字第1101300738號公告。[<https://www.fda.gov.tw/tc/includes/GetFile.ashx?id=f637551173813066216>]。

# A Preliminary Study on Applying Web Crawling Technology in Online Food Platform Information.

YUE-HAN KUO AND CHAO-YI WANG

Decision Support Center, TFDA, MOHW

## ABSTRACT

The rise of quick turn e-commerce business model trends in the post-pandemic era, prompting the need to strengthen food safety management through online channels. This can be achieved by utilizing web crawling technology and food cloud integration to mine high-concern food product and vendor information to facilitate targeting on the internet. The objective is to proactively monitor and regulate the emerging food business models driven by online trends and takes precautions before they happen. In view of the fact that different online platforms offer distinct services to both vendors and consumers, and the required information reporting also vary. this report explores the information overview, investigates publicly available data tables, and outlines the design of a data extraction and storage system for three types of platforms: e-commerce platform, food delivery platform, and blogs. As the amount of data continues to expand, considerations extend beyond economic factors such as manpower, time, cost, and benefits, the potential risks it may bring cannot be ignored, among which legal compliance is one of the key concerns. For any future systematic planning and applications related to web data extraction, a preliminary evaluation of potential legal implications is crucial to ensure robust risk management practices.

Key words: web crawler, internet platform, food safety management