

【11】證書號數：I667581

【45】公告日：中華民國 108 (2019) 年 08 月 01 日

【51】Int. Cl. : G06F17/30 (2006.01) H04L29/12 (2006.01)  
H04L12/917 (2013.01)

發明

全 6 頁

【54】名稱：網址分類方法及其電腦程式產品

【21】申請案號：105140768 【22】申請日：中華民國 105 (2016) 年 12 月 09 日

【11】公開編號：201822030 【43】公開日期：中華民國 107 (2018) 年 06 月 16 日

【72】發明人：楊富丞 (TW)；呂栢頤 (TW)

【71】申請人：中華電信股份有限公司  
桃園市楊梅區電研路 99 號

【74】代理人：葉璟宗；卓俊傑

【56】參考文獻：

US 6691163B1

US 6721747B2

US 7299274B2

US 2012/0323898A1

審查人員：莊榮昌

【57】申請專利範圍

1. 一種網址分類方法，包含：在目標網址可取得網址描述內容條件下，則擷取一個或複數個網址描述資訊，並分析各該網址描述資訊與多個類別的每一類別之一個或多個類別關鍵字之間的關聯度以獲得該目標網址對應每個該些類別的一機率值，並將該目標網址分類至該些類別中該機率值最大的一第一類別，其中各該網址描述資訊包括一單詞。
2. 如請求項 1 所述之網址分類方法，更在該在目標網址未能取得該網址描述內容條件下，則由該目標網址萃取一個或多個網址關鍵字，並將該網址關鍵字與一個或多個訓練集網址關鍵字進行比對，以將該目標網址分類至具關聯性之該訓練集網址關鍵字所對應的該類別內。
3. 如請求項 2 所述之網址分類方法，係依據該網址關鍵字與該訓練集網址關鍵字其吻合字串之數量以分析其關聯性。
4. 如請求項 3 所述之網址分類方法，係依據該吻合字串所座落之網址層級以分析其關聯度。
5. 如請求項 4 所述之網址分類方法，係依據該吻合字串之數量以及所座落之網址層級進行權重計算，以分析其關聯度。
6. 如請求項 1 所述之網址分類方法，係分析各該網址描述資訊與該類別關鍵字間的機率分佈，以計算該關聯度。
7. 如請求項 6 所述之網址分類方法，係基於潛在狄利克雷分配模型分析該機率分佈。
8. 如請求項 7 所述之網址分類方法，其中該潛在狄利克雷分配模型係採用 Gibbs sampling 推估關聯度。
9. 如請求項 1 所述之網址分類方法，更對目標網址之關鍵字進行翻譯，以進行分類作業。
10. 一種網址分類電腦程式產品，當電腦裝置載入並執行該電腦程式產品，可完成如請求項 1 至 9 任一項所述之方法。

圖式簡單說明

圖 1 為本案實施範例之建立分類模式流程圖。

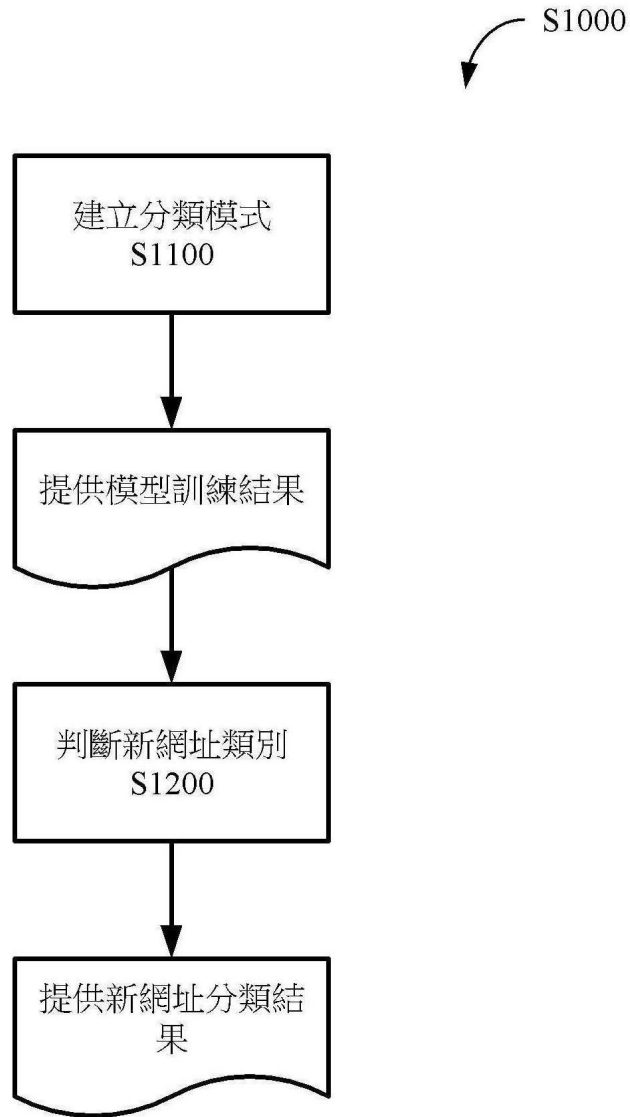
(2)

圖 2 為本案實施範例之網址分類模型建立流程圖。

圖 3 為網址描述擷取與前處理之步驟流程圖。

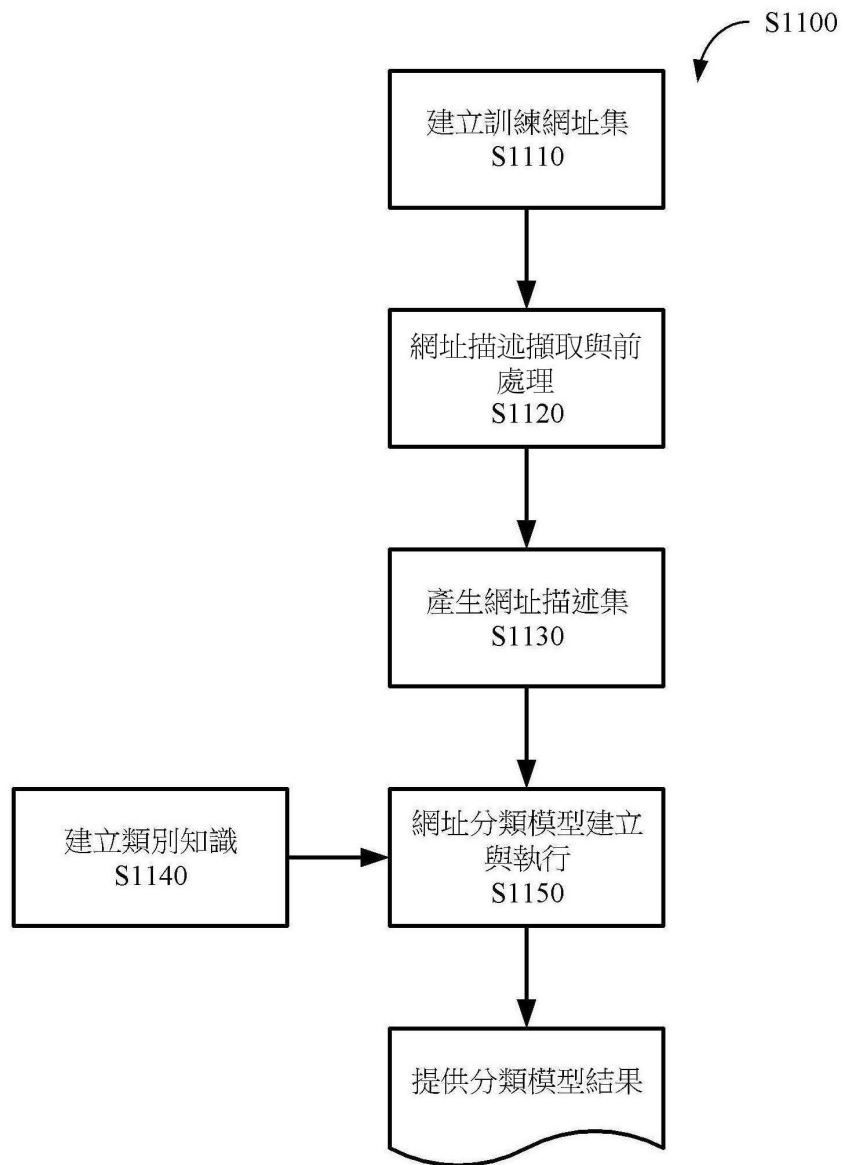
圖 4 為本案網址分類模型建立與執行之步驟流程圖。

圖 5 為本案判斷新網址類別並產生新網址分類結果之步驟流程圖。



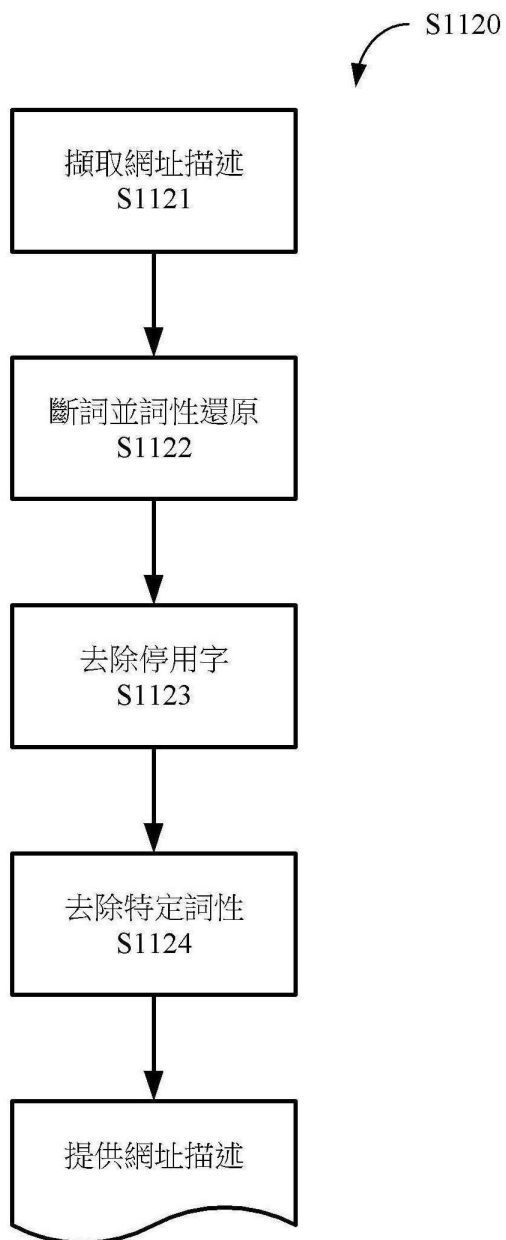
【圖1】

(3)



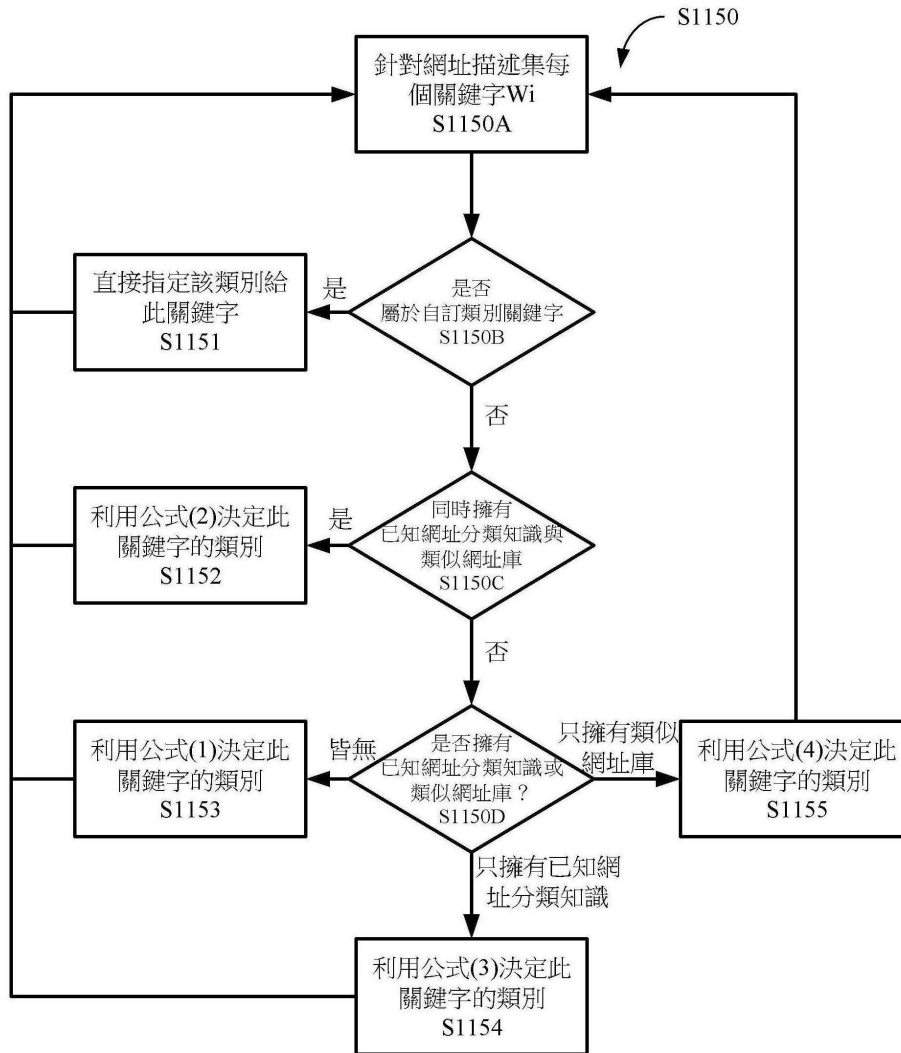
【圖2】

(4)



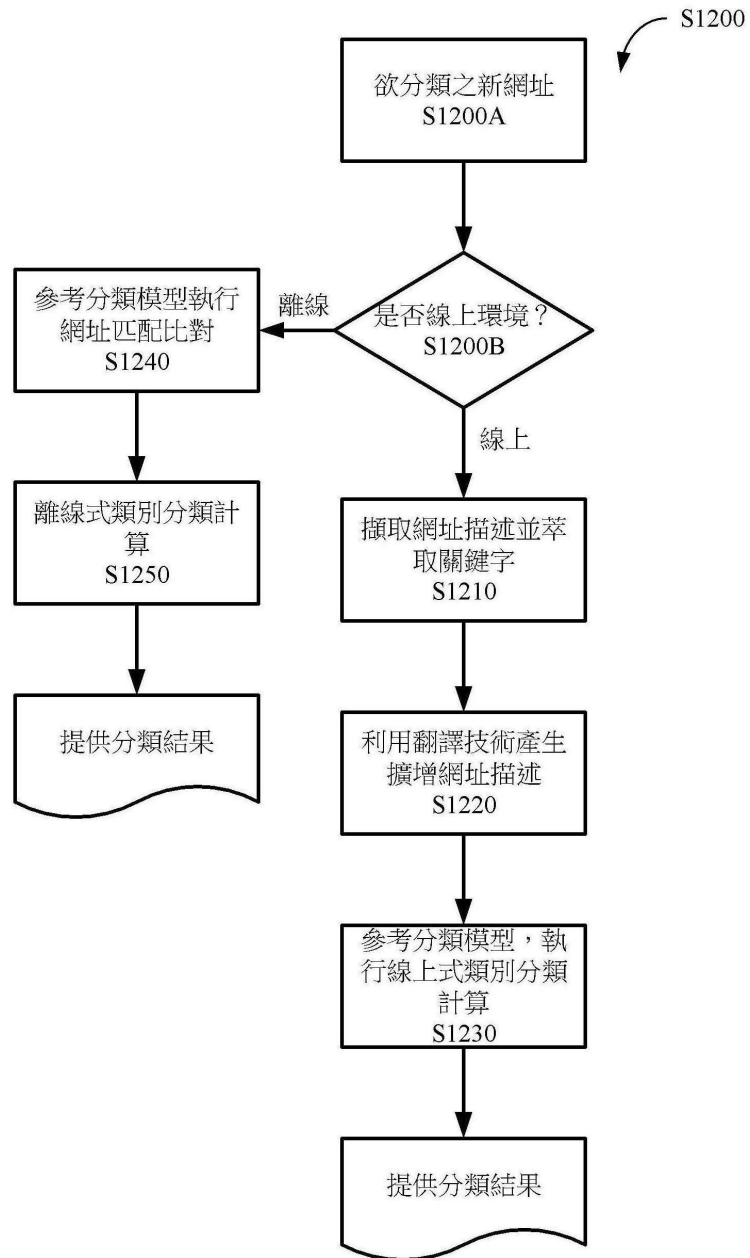
【圖3】

(5)



【圖4】

(6)



【圖5】