(19)中華民國智慧財產局
(12)發明說明書公開本 (11)公開編號：TW 201250505 A1
(43)公開日： 中華民國 101 (2012) 年 12 月 16 日

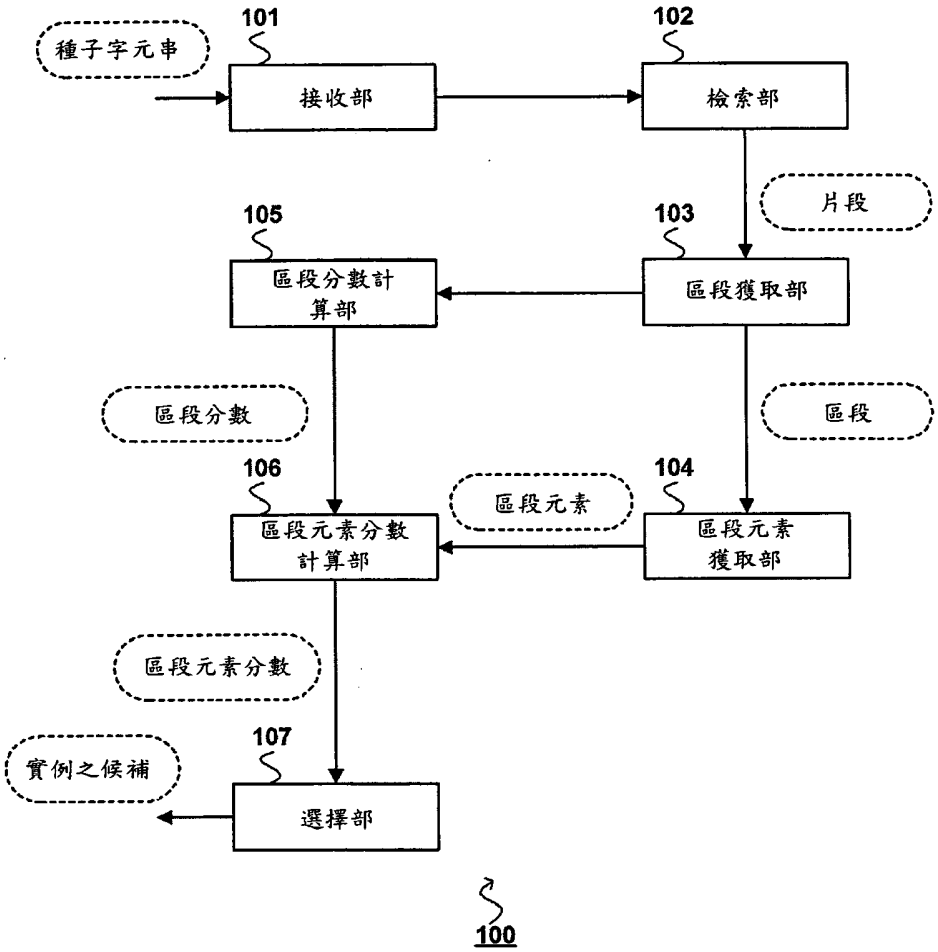
(21)申請案號：101106600 (22)申請日： 中華民國 101 (2012) 年 02 月 29 日
(51)Int. Cl.： *G06F17/30 (2006.01)*
(30)優先權：2011/03/04 日本 2011-048124
(71)申請人：樂天股份有限公司 (日本) RAKUTEN, INC. (JP)
日本
(72)發明人：萩原正人 HAGIWARA, MASATO (JP)
(74)代理人：陳長文
申請實體審查：有 申請專利範圍項數：7 項 圖式數：13 共 51 頁

(54)名稱

集合擴張處理裝置、集合擴張處理方法、程式、及非暫時性記錄媒體

(57)摘要

接收部(101)接收種子字元串。檢索部(102)獲取包含種子字元串之文件之片段。區段獲取部(103)以區段分隔字元串分隔該片段而獲取區段。區段元素獲取部(104)以區段元素分隔字元串分隔區段而獲取區段元素。區段分數計算部(105)根據區段元素之長度之標準偏差，而計算區段之區段分數。區段元素分數計算部(106)根據種子字元串之位置與區段元素之位置之距離、及區段分數，而計算區段元素之區段元素分數。選擇部(107)基於區段元素分數，自區段元素中選擇任意一個作為種子字元串之擴張集合中所含之實例之候補。



100：集合擴張處理裝置

101：接收部

102：檢索部

103：區段獲取部

104：區段元素獲取部

105：區段分數計算部

106：區段元素分數計算部

107：選擇部

(19)中華民國智慧財產局
(12)發明說明書公開本 (11)公開編號：TW 201250505 A1
(43)公開日： 中華民國 101 (2012) 年 12 月 16 日

(21)申請案號：101106600 (22)申請日： 中華民國 101 (2012) 年 02 月 29 日
(51)Int. Cl.： *G06F17/30 (2006.01)*
(30)優先權：2011/03/04 日本 2011-048124
(71)申請人：樂天股份有限公司 (日本) RAKUTEN, INC. (JP)
日本
(72)發明人：萩原正人 HAGIWARA, MASATO (JP)
(74)代理人：陳長文
申請實體審查：有 申請專利範圍項數：7 項 圖式數：13 共 51 頁

(54)名稱

集合擴張處理裝置、集合擴張處理方法、程式、及非暫時性記錄媒體

(57)摘要

接收部(101)接收種子字元串。檢索部(102)獲取包含種子字元串之文件之片段。區段獲取部(103)以區段分隔字元串分隔該片段而獲取區段。區段元素獲取部(104)以區段元素分隔字元串分隔區段而獲取區段元素。區段分數計算部(105)根據區段元素之長度之標準偏差，而計算區段之區段分數。區段元素分數計算部(106)根據種子字元串之位置與區段元素之位置之距離、及區段分數，而計算區段元素之區段元素分數。選擇部(107)基於區段元素分數，自區段元素中選擇任意一個作為種子字元串之擴張集合中所含之實例之候補。

六、發明說明：

【發明所屬之技術領域】

本發明係關於集合擴張處理裝置、集合擴張處理方法、程式、及非暫時性(non-transitory：非暫時性)記錄媒體，尤其係關於包含於意義上相同之類別之字之獲取者。

【先前技術】

網上購物中，購物場所中處理之商品按類別分開並提示予使用者。例如，專利文獻1中，刊登商品之頁中，揭示有顯示商品之類別「家電商品」、「書籍」、「電腦」等之資訊收發系統。使用者藉由自該等之類別中選擇希望購入之商品之類別，可容易地縮小商品之範圍。

另一方面，爲了系統化建構並維持人名、地名、或商品名等之固有表現需龐大之費用。因此，正在積極研究利用計算機自動獲取固有表現之意義上之關係性之自動獲取技術。例如，非專利文獻1中，揭示有自分隔寫法之字中擷取意義上之詞彙類別之算法(稱為「g-Espresso算法」)。又，非專利文獻2中，揭示有自非分隔寫法之字中擷取意義上之詞彙類別之算法(稱為「g-Monaka算法」)。

先前技術文獻

專利文獻

專利文獻1：日本特開2009-48226號公報

非專利文獻

非專利文獻1：小町守(Mamoru Komachi)、工藤拓(Taku Kudo)、新保仁(Masahi Shimbo)、松本裕治(Yuji

Matsumoto)，「以濃縮型引導演算法基於圖表之語義漂移之分析(Graph-based analysis of semantic drift in espresso-like bootst rapping algorithms)。」EMNLP 2008之會議記錄中，第1011-1020頁，2008。

非專利文獻2：萩原正人、小川泰弘、外山勝彦，「基於圖表核函數之來自非分隔寫法之意義上之詞彙類別之擷取」，語言處理學會第15次年度大會演講論文集，第697-700頁，2009年

【發明內容】

發明所欲解決之問題

如上所述之購物場所中，由於每天都有新商品上市，故手動進行商品之類別之記錄工作會來不及，即使為大量之使用者檢索之商品，仍有未設置有其商品所屬之類別之情形。然而，對商家而言，每次新商品上市即調查應該記錄之類別一事負擔很大，且存在希望其自動選擇應該記錄之類別之候補之要求。

本發明為解決如上所述之問題者，且目的在於提供一種適合於選擇屬於意義上相同之類別之字之候補之集合擴張處理裝置、集合擴張處理方法、程式、及非暫時性記錄媒體。

解決問題之技術手段

本發明之第1觀點之集合擴張處理裝置，其特徵為包含：

接收部，其係接收種子字元串；

檢索部，其係檢索包含上述接收之種子字元串之文件，而獲取該檢索之文件之片段；

區段獲取部，其係以特定之區段分隔字元串分隔上述獲取之片段，而獲取包含將出現於上述接收之種子字元串之前後之字元串、及將該種子字元串按出現順序排列之字元串之區段；

區段元素獲取部，其係以特定之區段元素分隔字元串分隔上述獲取之區段之各者，而獲取區段元素；

區段分數(Segment Score)計算部，其係根據於該區段中出現之區段元素之各者之長度之離差或標準偏差，而計算上述獲取之區段之各者之區段分數；

區段元素分數計算部，其根據上述接收之種子字元串於上述獲取之區段中出現之位置與該區段元素於該區段中出現之位置之距離、及針對該區段計算之區段分數，而計算該區段之各者中所含之區段元素之各者之區段元素分數；及

選擇部，其係根據針對上述獲取之區段元素之各者所計算之區段元素分數，自該區段元素中選擇任意一個作為將包含上述接收之種子字元串之集合擴張而成之擴張集合中所含之實例之候補。

又，上述觀點之集合擴張處理裝置，其中進而包含：

擷取部，其係由使用上述實例之候補進行檢索而獲取之片段，生成包含上述擷取之實例之候補之n元語法之相連詞模型(N-gram)，並根據該相連詞模型之上述接收之種子

字元串之前後之語境與該實例之候補之前後之語境，計算該種子字元串與該實例之候補之相似度，且根據該相似度，自該實例之候補中，擷取將包含該種子字元串之集合擴張而成之擴張集合中所應含有之實例。

又，上述觀點之集合擴張處理裝置，其中：

針對上述獲取之區段之各者，若於該區段中出現之區段元素之各者之長度之標準偏差超過特定之臨限值之情形，上述區段分數及上述區段元素分數成為該區段中所含之區段元素不會被上述選擇部選擇作為上述實例之候補之值。

又，上述觀點之集合擴張處理裝置，其中：

於上述獲取之區段之各者中出現之區段元素之各者之區段元素分數，相對於上述接收之種子字元串於該區段中出現之位置與該區段元素於該區段中出現之位置之最短距離呈指數衰減。

本發明之第2觀點之集合擴張處理方法，其特徵在於其係具備接收部、檢索部、區段獲取部、區段元素獲取部、區段分數計算部、區段元素分數計算部、及選擇部之集合擴張處理裝置所執行者，且包含：

接收步驟，由上述接收部接收種子字元串；

檢索步驟，由上述檢索部檢索包含上述接收之種子字元串之文件，而獲取該檢索之文件之片段；

區段獲取步驟，由上述區段獲取部以特定之區段分隔字元串分隔上述獲取之片段，而獲取包含將出現於上述接收之種子字元串之前後之字元串、及將該種子字元串按出現

順序排列之字元串之區段；

區段元素獲取步驟，由上述區段元素獲取部以特定之區段元素分隔字元串分隔上述獲取之區段之各者，而獲取區段元素；

區段分數計算步驟，由上述區段分數計算部根據於該區段中出現之區段元素之各者之長度之離差或標準偏差，而計算上述獲取之區段之各者之區段分數；

區段元素分數計算步驟，由上述區段元素分數計算部根據該區段中上述接收之種子字元串出現之位置與該區段中該區段元素出現之位置之距離及針對該區段計算之區段分數計算包含於上述獲取之區段之各者之區段元素之各自之區段元素分數；及

選擇步驟，由上述選擇部根據針對上述獲取之區段元素之各者所計算之區段元素分數，自該區段元素中選擇任意一個作為將包含上述接收之種子字元串之集合擴張而成之擴張集合中所含之實例之候補。

本發明之第3觀點之程式，其特徵在於使電腦作為下述各部發揮功能：

接收部，其係接收種子字元串；

檢索部，其係檢索包含上述接收之種子字元串之文件，而獲取該檢索之文件之片段；

區段獲取部，其係以特定之區段分隔字元串分隔上述獲取之片段，而獲取包含將出現於上述接收之種子字元串之前後之字元串、及將該種子字元串按出現順序排列之字元

串之區段；

區段元素獲取部，其係以特定之區段元素分隔字元串分隔上述獲取之區段之各者，而獲取區段元素；

區段分數計算部，其係根據於該區段中出現之區段元素之各者之長度之離差或標準偏差，而計算上述獲取之區段之各者之區段分數；

區段元素分數計算部，其根據上述接收之種子字元串於上述獲取之區段中出現之位置與該區段元素於該區段中出現之位置之距離、及針對該區段計算之區段分數，而計算該區段之各者中所含之區段元素之各者之區段元素分數；及

選擇部，其係根據針對上述獲取之區段元素之各者所計算之區段元素分數，自該區段元素中選擇任意一個作為將包含上述接收之種子字元串之集合擴張而成之擴張集合中所含之實例之候補。

本發明之第4觀點之記錄程式之非暫時性之電腦可讀取之記錄媒體，其特徵在於使電腦作為下述各部發揮功能：

接收部，其係接收種子字元串；

檢索部，其係檢索包含上述接收之種子字元串之文件，而獲取該檢索之文件之片段；

區段獲取部，其係以特定之區段分隔字元串分隔上述獲取之片段，而獲取包含將出現於上述接收之種子字元串之前後之字元串、及將該種子字元串按出現順序排列之字元串之區段；

區段元素獲取部，其係以特定之區段元素分隔字元串分隔上述獲取之區段之各者，而獲取區段元素；

區段分數計算部，其係根據於該區段中出現之區段元素之各者之長度之離差或標準偏差，而計算上述獲取之區段之各者之區段分數；

區段元素分數計算部，其根據上述接收之種子字元串於上述獲取之區段中出現之位置與該區段元素於該區段中出現之位置之距離、及針對該區段計算之區段分數，而計算該區段之各者中所含之區段元素之各者之區段元素分數；及

選擇部，其係根據針對上述獲取之區段元素之各者所計算之區段元素分數，自該區段元素中選擇任意一個作為將包含上述接收之種子字元串之集合擴張而成之擴張集合中所含之實例之候補。

上述程式與執行程式之電腦相獨立，且可經由電腦通信網進行散發、販賣。又，上述記錄媒體可與電腦相獨立進行散發、販賣。

發明之效果

根據本發明，可提供適合於選擇屬於意義上相同之類別之字之候補之集合擴張處理裝置、集合擴張處理方法、程式、及非暫時性記錄媒體。

【實施方式】

本發明之實施形態之集合擴張處理裝置100，如圖1所示，連接於購物伺服器200。購物伺服器200連接於網路

300。於網路300中，連接有使用者操作之複數個終端裝置401、402~40n。購物伺服器200經由網路300在終端裝置401~40n中提示登錄於購物伺服器200中之商品之資訊，且自終端裝置401~40n接收商品之訂購。一般而言，登錄於購物伺服器200中之商品根據商品之種類區分類別，並提示至終端裝置401~40n之使用者。集合擴張處理裝置100為就購物伺服器200中經營之商品進行集合擴張處理，並提示商品之類別之候補者。

此處，所謂「集合擴張」，即供給少數之正確設定作為種子，獲取屬於與種子意義上相同之類別之字語之集合之作業。例如，將廚房用品之「中華炒鍋」、「壓力鍋」作為種子之情形，屬於意義上相同之類別之字語為「砂鍋」、「雪平鍋」、及「塔吉鍋」等。即，集合擴張處理裝置100，若供給「中華炒鍋」、「壓力鍋」作為種子，則作為屬於與其等相同之類別「鍋」之字語，獲取「砂鍋」、「雪平鍋」或「塔吉鍋」等。

以下，就實現本發明之實施形態之集合擴張處理裝置100之典型資訊處理裝置500進行說明。

(1. 資訊處理裝置之概要構成)

如圖2所示，資訊處理裝置500具備CPU(Central Processing Unit：中央處理器)501、ROM(Read only Memory：唯讀記憶體)502、RAM(Random Access Memory：隨機存取記憶體)503、NIC(Network Interface Card：網卡)504、圖像處理部505、聲音處理部506、DVD-ROM(Digital Versatile Disc

ROM：唯讀型數位多功能光碟機)光碟機507、介面508、外部記憶體509、控制器510、監視器511與揚聲器512。

CPU 501控制資訊處理裝置500整體之動作，並交換與各構成元素連接之控制信號或資料。

ROM 502中，記錄有電源接通後執行之IPL(Initial Program Loader：初始程式載入器)，藉由執行該IPL，在RAM 503中讀出特定之程式並開始藉由CPU 501而進行之該程式之執行。又，ROM 502中，記錄有資訊處理裝置500整體之動作控制所需要之操作系統之程式或各種之資料。

RAM 503為用以暫時存儲資料或程式者，且保持有自DVD-ROM讀出之程式或資料、其他及通信所需之資料等。

NIC 504為用以將資訊處理裝置500連接於網路300等之電腦通信網者，且係藉由按照構成LAN(Local Area Network：區域網路)時使用之10BASE-T/100BASE-T規格者、用以使用電話線路連接於網路之類比數據機、ISDN(Integrated Services Digital Network：整合服務數位網路)數據機、ADSL(Asymmetric Digital Subscriber Line:非對稱數位用戶線)數據機、或使用有線電視線路連接於網路之寬頻數據機等與進行該等與CPU 501之中介之介面(未圖示)構成。

圖像處理部505，藉由CPU 501或圖像處理部505具備之圖像運算處理器(未圖示)對自DVD-ROM等讀出之資料進行加工處理後，將其記錄於圖像處理部505具備之圖框記

憶體(未圖示)。記錄於圖框記憶體中之圖像資訊，在特定之同步時機轉換為視頻信號，並輸出至監視器511。藉此，可實現各種之網頁顯示。

聲音處理部506，將自DVD-ROM等讀出之聲音資料轉換為類比聲音信號，並自連接於其之揚聲器512輸出。又，在CPU 501之控制下，在資訊處理裝置500進行之處理之進行中生成應該產生之聲音，並使與其對應之聲音自揚聲器512輸出。

於安裝於DVD-ROM光碟機507中之DVD-ROM中，例如，存儲有用以實現實施形態之集合擴張處理裝置100之程式。藉由CPU 501之控制，DVD-ROM光碟機507進行相對安裝於其中之DVD-ROM之讀出處理，且讀出所需之程式或資料，該等係暫時存儲於RAM 503等中。

介面508中，可裝卸地連接有外部記憶體509、控制器510、監視器511及揚聲器512。

外部記憶體509中，可重寫地存儲有關於使用者之個人資訊之資料等。

控制器510接收在資訊處理裝置500之各種之設定時等進行之操作輸入。資訊處理裝置500之使用者，藉由經由控制器510進行指示輸入，可適宜地將該等之資料記錄於外部記憶體509中。

監視器511將藉由圖像處理部505輸出之資料提示予資訊處理裝置500之使用者。

揚聲器512將藉由聲音處理部506輸出之聲音資料提示予

處理裝置500之使用者。

另外，資訊處理裝置500，使用硬碟等之大容量外部存儲裝置，且可以發揮與安裝於ROM 502、RAM 503、外部記憶體509、DVD-ROM光碟機507之DVD-ROM等相同之功能之方式構成。

以下，就上述資訊處理裝置500中實現之實施形態之集合擴張處理裝置100之概要構成，參照圖1至13進行說明。藉由接通資訊處理裝置500之電源，執行作為實施形態之該集合擴張處理裝置100發揮功能之程式，且實現實施形態之集合擴張處理裝置100。

(2.實施形態1之集合擴張處理裝置之概要構成)

實施形態1之集合擴張處理裝置100為選擇包含於擴張包含種子字元串之集合之擴張集合之實例之候補者。

本實施形態之集合擴張處理裝置100如圖3所示，係包含接收部101、檢索部102、區段獲取部103、區段元素獲取部104、區段分數計算部105、區段元素分數計算部106與選擇部107。

以下，集合擴張處理裝置100，以作為屬於廚房用品之鍋之類別之字進行適當之字(實例)之候補之提示之情形為例進行說明。

接收部101接收種子字元串。所謂種子字元串，例如，為包含於屬於「鍋」之類別之字之集合之正確之字(「中華炒鍋」或「壓力鍋」等)。例如，如圖4所示，使用者在WEB網頁之檢索引擎之檢索欄601中，將用空間分隔使全

部之種子字元串連結者作為查詢輸入，並按壓檢索按鈕602。該情形，接收部101將輸入至檢索欄601之「中華炒鍋」及「壓力鍋」作為種子字元串接收。另，檢索引擎之種類為任意。

本實施形態中，CPU 501及控制器510合作作用，並作為接收部101發揮功能。

檢索部102檢索包含接收之種子字元串之文件，並獲取片段。此處，所謂片段，為使用WEB網頁之檢索引擎時，包含作為檢索結果顯示之查詢之文本部份。檢索部102相對WEB網頁之檢索引擎，將用空間分隔使全部之種子字元串連結者作為查詢輸入，從而獲取檢索結果之，例如，前300件之片段之表。例如，檢索部102，將「中華炒鍋 壓力鍋」作為查詢使用檢索引擎進行WEB網頁之檢索，獲取包含得到之種子字元串「中華炒鍋」及「壓力鍋」之圖4之片段1、2、3~300(未圖示)。另，檢索部102不限於如上所述利用外部裝置獲取文件，亦可在內部具備檢索功能。例如，檢索部102亦可設為使用Web檢索API獲取片段。

本實施形態中，檢索部102與CPU 501及NIC 504合作作用，作為檢索部102發揮功能。

區段獲取部103，藉由用特定之區段分隔字元串分隔獲取之片段，獲取包含將出現於種子字元串之前後之字元串與種子字元串按出現順序排列之字元串之區段。片段爲了在包含有檢索字之網頁中如何使用該檢索字對使用者而言一目了然，用特定之分隔字元串分隔為一般情況。例如，

將特定之區段分隔字元串作為「...」。例如，區段獲取部103藉由Unicode(萬國碼)之NFKC使獲取之片段1、2、3~300正規化，且統一為小寫，並藉由區段分隔字元串「...」分割複數之字元串。且，區段獲取部103剔除分割之字元串中重複之字元串，將剩餘之字元串作為區段而獲取。藉由將獲取之片段統一為小寫，例如，可對應型號等之字元串未以大寫·小寫統一之情形。圖5中，區段獲取部103顯示自片段1獲取之區段1-1~1-3。

另，區段分隔字元串並不限於「...」之字元串。檢索部102使用之檢索引擎或Web檢索API提示之片段，例如，用「---」或「##」之字元串分隔之情形，將區段分隔字元串作為「---」或「##」之字元串。又，獲取區段之技術並不限於使用區段分隔字元串獲取區段之技術。根據使用之檢索引擎或Web檢索API提示之片段，適當獲取區段。例如，一個片段未被「...」之記號分隔而提示之情形，將該片段作為一個區段。又，預先，以條列書寫等提示相當於片段內之區段之部份之情形，將相當於條列書寫之一列之部份作為一個區段。

本實施形態中，CPU 501作為區段獲取部103發揮功能。

區段元素獲取部104，藉由將獲取之各個區段用特定之區段元素分隔字元串分隔，獲取區段元素。例如，所謂特定之區段元素分隔字元串，為標點符號或記號(「、」、「，」、「。」、「！」、「〔」、「〕」等)，藉由該等之區段元素分隔字元串分隔區段，從而獲取區段元素。

例如，區段元素獲取部104，若用區段元素分隔字元串分隔圖5之區段1-1、1-2、1-3，則獲取圖6之區段元素群1-1P(區段元素P_i(i=1~5))、1-2P(區段元素P_i(i=1~12))、1-3P(區段元素P_i(i=1~5))。

本實施形態中，CPU 501作為區段元素獲取部104發揮功能。

區段分數計算部105，根據出現於該區段之區段元素之各自之長度之離差或標準偏差計算獲取之區段之各自之區段分數。此處，就獲取之各個區段，該區段中出現之區段元素之各自之長度之標準偏差超過特定之臨限值之情形，使區段分數及後述之區段元素分數設為包含於該區段之區段元素不會作為實例之候補被選擇部107選擇之值。本實施形態中，雖用Unicode(萬國碼)之字數定義區段元素之長度，但並不限定於此。例如，作為區段元素之長度，可採用其他之字元碼之位元組數。

例如，如圖5所示，雖區段1-1、1-3中包含一般詞句，但區段1-2中不包含一般詞句。且，區段1-1、1-3中所含之區段元素之長度之差，大於區段1-2中所含之區段元素之長度之差。即，包含一般詞句之區段，一般而言，與不包含一般詞句之區段相比，具有區段中所含之各區段元素之長度不一致之傾向。且，包含一般詞句之區段中，由於不包含屬於與種子字元串相同意義範圍之實例之情形較多，故作為獲取實例之候補之區段並不適當。因此，以下將區段元素之長度之標準偏差超過特定之臨限值之區段，自所

要獲取實例之候補之區段中剔除。

本實施形態中，將特定之臨限值設為5.00。又，區段分數計算部105在區段元素之長度之標準偏差未達5.00之情形時，將標準偏差之值本身作為區段分數，而當標準偏差為5.00以上之情形時，將區段分數設為5.00。

圖7中顯示區段分數計算部105所計算之區段分數。圖7之表中，關聯記載有將種子字元串作為查詢而獲取之「片段701a」、片段701a中所含之「區段702a」、區段702a中所含之「區段元素703a」、區段元素703a之「長度704a」、長度704a之「標準偏差705a」、根據標準偏差705a而計算之「區段分數706a」、及藉由後述之區段元素分數計算部106計算之「區段元素分數707a」。

例如，如圖7之704a所示，區段分數計算部105求得區段1-1中所含之區段元素P_i(i=1~5)、區段1-2中所含之區段元素P_i(i=1~12)、及區段1-3中所含之區段元素P_i(i=1~5)之長度。且，區段分數計算部105，如圖7之705a所示，求得區段1-1中所含之區段元素P_i(i=1~5)之長度之標準偏差為「5.89」、區段1-2中所含之區段元素P_i(i=1~12)之長度之標準偏差為「1.34」、區段1-3中所含之區段元素P_i(i=1~5)之長度之標準偏差為「5.27」。因此，區段分數計算部105，如圖7之706a所示，求得區段1-1之區段分數為「5.00」，區段1-2之區段分數為「1.34」，區段1-3之區段分數為「5.00」。

本實施形態中，CPU 501作為區段分數計算部105發揮功

能。

區段元素分數計算部106根據該區段中接收之種子字元串出現之位置與該區段中該區段元素出現之位置之距離及針對該區段計算之區段分數計算包含於獲取之各個區段之區段元素之各自之區段元素分數。

例如，如上所述，區段元素之各自之長度之標準偏差超過特定之臨限值之情形，使區段元素分數為區段元素作為實例之候補不被選擇部107選擇之值。例如，區段元素分數計算部106，區段分數為「5.00」之情形，使區段元素分數為「0」。另一方面，區段分數為未達「5.00」之情形，區段元素分數計算部106根據區段中接收之種子字元串出現之位置與該區段中該區段元素出現之位置之距離計算區段元素分數。此處，所謂區段中種子字元串出現之位置s_j(j：種子字元串之數)、及區段中區段元素出現之位置p_i，如圖6所示，為區段中按出現順序排列區段元素時之區段內中之出現順位，所謂距離，為位置s_j與位置p_i之出現順位之差。即，若種子字元串為「中華炒鍋」及「壓力鍋」，則區段1-2中種子字元串「壓力鍋」(P_4)出現之位置s_1為第「4」號，種子字元串「中華炒鍋」(P_8)出現之位置s_2為第「8」號。又，區段1-2中區段元素「親子鍋」(P_5)出現之位置p_5為第「5」號，種子字元串「中華炒鍋」(P_8)與區段元素「親子鍋」(P_5)之距離為3。

且，區段元素分數計算部106，根據區段中種子字元串出現之位置s_j與區段中區段元素出現之位置p_i基於以下之

式(數1)計算區段元素分數S_i。根據該式(數1)，隨著與最近之種子字元串之距離以指數衰減之分數為各區段元素之區段元素分數。本實施形態中α=0.8。在圖7之區段元素分數707a中顯示計算結果。

[數1]

$$S_i = \max_j \exp(-\alpha |p_i - s_j|)$$

上述中，雖求得隨著與最近之種子字元串之距離以指數衰減之分數，但分數之求得方法中可有各種變化。例如，存在複數種子字元串之情形下，分別求各種子字元串與區段元素之距離，亦可將隨著求得之距離之平均值以線形衰減之分數作為各區段元素之區段元素分數。

以上，雖記載了區段內出現種子字元串之情形之一例，但出現種子字元串之相似字之情形亦可同樣進行計算。具體而言，將「中華炒鍋」及「壓力鍋」作為種子字元串之情形下，若檢索部中除種子字元串之外用種子字元串之相似字進行檢索，則可得到包含有稱為「中華炒guo」或「壓力guo」之種子字元串之相似字之片段。如此之情形下，區段元素分數計算部106中，藉由使用眾所周知之漢字假名文字轉換程式等，可將種子字元串之相似字作為種子字元串同樣地進行處理。如此，即使為種子字元串之相似字出現於區段內之情形，仍可根據數1計算區段元素分數S_i。

本實施形態中，CPU 501作為區段元素分數計算部106發

揮功能。

選擇部107根據針對獲取之各個區段元素計算之區段元素分數，自該區段元素選擇任意一個作為包含於擴張包含接收之種子字元串之集合之擴張集合之實例之候補。此處，所謂擴張集合，係實施集合擴張處理後獲取之集合，且係包含於與種子字元串意義上相同之類別之字之集合。例如，選擇部107，自實例之候補剔除區段元素分數之值為未達「0.10」之區段元素，選擇剩餘之區段元素作為實例之候補。即，選擇部107，由於自區段1-1、1-3獲取之區段元素之區段元素分數全部為「0」(圖7)，故將自區段1-1、1-3獲取之區段元素從候補中剔除。且，選擇部107，如圖8所示，自區段1-2獲取之區段元素中，將區段元素分數之值為未達「0.10」之「製義大利麵機」、「其他」、及「進而價格」之區段元素剔除，並將剩餘之區段元素作為包含於與「中華炒鍋」及「壓力鍋」意義上相同之類別之實例之候補進行選擇。另，本實施形態中，以一個片段為例，就選擇實例之候補之技術進行說明，實際上，自多數之片段獲取區段元素從而求出區段元素分數，選擇實例之候補。該情形，於相同區段元素，可自不同之片段分別求出區段元素分數。尤其，包含於與種子字元串意義上相同之類別之區段元素，由於考慮到包含於複數之片段中之情形較多，故可獲取複數之區段元素分數之可能性較高。因此，獲取複數之區段元素分數之情形，將該等之和或最大值作為該區段元素之區段元素分數之值。藉由如此處

理，可選擇更適當之實例之候補。

本實施形態中，CPU 501作為選擇部107發揮功能。

(3.實施形態1之集合擴張處理裝置之動作)

其次，就本實施形態之集合擴張處理裝置100之各部進行之動作，使用圖9之流程圖進行說明。若電源進入集合擴張處理裝置100且進行特定之操作，則CPU 501開始圖9之流程圖中所示之集合擴張處理。

首先，接收部101接收種子字元串(步驟S101)。例如，接收部101，如圖4所示，將作為查詢輸入至WEB頁之檢索引擎之檢索欄601之「中華炒鍋」及「壓力鍋」作為種子字元串接收。

其次，檢索部102檢索包含接收之種子字元串之文件，並獲取片段(步驟S102)。例如，檢索部102將種子字元串中「中華炒鍋」及「壓力鍋」作為查詢進行檢索，如圖4所示，從而獲取檢索結果之前300件之片段1、2、3~300。另，雖檢索部102獲取之片段之數為任意，但藉由獲取大約100件以上之片段，可選擇更適當之實例之候補。

其次，區段獲取部103藉由將檢索部102獲取之片段用區段分隔字元串分隔獲取區段(步驟S103)。例如，區段獲取部103用區段分隔字元串「...」分隔片段1、2、3~300，從而獲取區段。例如，區段獲取部103自片段1，如圖5所示，獲取區段1-1~1-3。

若獲取區段(步驟S103)，則區段元素獲取部104藉由將該區段用特定之區段元素分隔字元串分隔獲取區段元素

(步驟S104)。例如，將區段1-1~1-3用區段元素分隔字元串(「、」、「，」、「。」、「！」、「〔」、「〕」等)分隔，從而獲取圖6之區段元素(區段元素群1-1P、1-2P、1-3P)。

若獲取區段元素(步驟S104)，則區段分數計算部105，根據區段包含之區段元素之長度之標準偏差計算該區段之各自之區段分數(步驟S105)。例如，區段分數計算部105，於區段元素之長度之標準偏差為未達5.00之情形，將標準偏差之值本身作為區段分數，區段元素之長度之標準偏差為5.00以上之情形，將區段分數設為5.00。即，區段分數計算部105求得標準偏差為「5.89」之區段1-1之區段分數為「5.00」，標準偏差為「1.34」之區段1-2之區段分數為「1.34」，標準偏差為「5.27」之區段1-3之區段分數為「5.00」。

其次，區段元素分數計算部106，根據區段中接收之種子字元串出現之位置與該區段中該區段元素出現之位置之距離及針對該區段計算之區段分數計算區段元素之區段元素分數(步驟S106)。例如，區段元素分數計算部106，於區段分數為「5.00」之情形，使區段元素分數為「0」，區段分數為未達「5.00」之情形，根據使用區段中種子字元串出現之位置與區段元素出現之位置之距離之式(數1)，計算區段元素分數707a(圖7)。

且，選擇部107，根據針對獲取之區段元素計算之區段元素分數，選擇屬於與種子字元串意義上相同之類別之實

例之候補(步驟S107)。例如，選擇部107，如圖8所示，選擇區段元素分數之值為「0.10」以上之區段元素作為實例之候補。

根據本實施形態，由於「親子鍋」或「塔吉鍋」為包含於與種子字元串之「中華炒鍋」或「壓力鍋」相同之「鍋」之類別之用語，故可選擇屬於意義上相同之類別之字之候補。

(4.實施形態2之集合擴張處理裝置之概要構成)

實施形態2之集合擴張處理裝置100，針對包含於擴張集合之實例之候補，藉由根據語境加以過濾，排除意義上無關之字。

本實施形態之集合擴張處理裝置100，如圖10所示，係包含接收部101、檢索部102、區段獲取部103、區段元素獲取部104、區段分數計算部105、區段元素分數計算部106、選擇部107及擷取部108。本實施形態之接收部101、檢索部102、區段獲取部103、區段元素獲取部104、區段分數計算部105、區段元素分數計算部106及選擇部107，具有與實施形態1相同之功能。以下，就具有不同功能之擷取部108進行說明。

首先，實例之候補當種子字元串之前後之語境與實例之候補之前後之語境相似，則認為與種子字元串意義上相似。因此，實施形態2之集合擴張處理裝置100，根據種子字元串之前後之語境與實例之候補之前後之語境求出種子字元串與實例之候補之相似度，並根據該相似度自實例之

候補之中擷取實例。藉此，可排除意義上無關之字。以下，集合擴張處理裝置100，根據用g-Monaka算法計算之相似度，將實例之候補分級，並擷取具有特定之值以上之相似度者作為實例。另，求相似度之技術並不限於g-Monaka算法。例如，亦可使用g-Espresso算法。

擷取部108根據藉由使用上述實例之候補進行檢索而獲取之片段，生成包含擷取之實例之候補之n元語法之相連詞模型。且，擷取部108根據該相連詞模型之接收之種子字元串之前後之語境與實例之候補之前後之語境計算該種子字元串與該實例之相似度，且根據該相似度，自該實例之候補，擷取應該包含於擴張包含該種子字元串之集合之擴張集合之實例。以下，詳細說明根據g-Monaka算法之相似度之計算技術。

擷取部108，將選擇部107選擇之各個實例之候補，對於WEB網頁之檢索引擎作為查詢輸入，從而獲取檢索結果之前300件之片段之表。且，擷取部108，對於獲取之片段，藉由Unicode(萬國碼)之NFKC正規化，並統一為小寫，去掉重複。又，剔除日語之比例極少而記號很多等，作為片段並不適當者。

其次，擷取部108，針對包含於剩餘之片段之集合之全部之文字n元語法，建構關聯矩陣M(u，v)。關聯矩陣M(u，v)用式(數2)表示。

[數2]

$$M(u,v)=\frac{pmi(u,v)}{\max pmi},\qquad pmi(u,v)=\log\frac{|u,v|}{|u,*||*,v|}$$

此處，｜u，v｜為n元語法u之後n元語法v連續之頻率，｜u，*｜、｜*，v｜分別為n元語法u、n元語法v本身之出現頻率。本實施形態中，｜u，v｜、｜u，*｜、｜*，v｜為將其等自身作為查詢檢索之情形之檢索結果數，pmi(u，v)使用獲取其等之檢索結果數之自然對數者。

其次，擷取部108生成將全部之n元語法之集合V作為節點集合，將M作為關聯矩陣表現之定向加權圖表(以下，稱為「相連詞模型」)G_M。在圖11中顯示生成之相連詞模型G_M之例。該圖表中，當n元語法u及n元語法v之各自之右側語境及左側語境相似，則可視為其等之意義相似。

此處，首先，對應引用分析技術之書誌結合之概念可思考n元語法u之右側語境與n元語法v之右側語境是否相似。所謂書誌結合，係指文獻x、y引用相同文獻。即，可對應n元語法u與n元語法v是否連接於相同n元語法思考書誌結合。另一方面，對應引用分析技術之共引用之概念可思考n元語法u之左側語境與n元語法v之左側語境是否相似。所謂共引用，係指文獻x、y被相同文獻引用。即，可對應思考n元語法u與n元語法v是否自相同n元語法連接。

因此，分別對應書誌結合矩陣及共引用矩陣求顯示n元語法u及n元語法v之右側語境及左側語境是否相似之相似度矩陣A_R、A_L。右側語境之相似度矩陣A_R及左側語境之相似度矩陣A_L可使用關聯矩陣M用式(數3)表示。

[數3]

$$A_R = \frac{1}{|V|^2} MM^T, \qquad A_L = \frac{1}{|V|^2} M^T M$$

擷取部108求針對全部之n元語法右側語境之相似度矩陣A_R、及左側語境之相似度矩陣A_L。

又，為了使n元語法u與n元語法v視為意義上相似，有必要右側語境及左側語境之兩者相似(以下，稱為「兩側近接制約」)。因此，擷取部108，如式(數4)所示，根據每個元素之加權一般化平均，求顯示n元語法u與n元語法v之相似度之相似度矩陣A。此處，m為調節該兩側近接制約之強度之參數，本實施形態中，m=0.1。

[數4]

$$A(i,j) = \sqrt[m]{\frac{1}{2}(A_R(i,j)^m + A_L(i,j)^m)}$$

且，擷取部108，使用該相似度矩陣A根據數5、數6之式求拉普拉斯核(Laplacian Kernel)$R_\beta(A)$。

[數5]

$$\tilde{R}_\beta(A) = \sum_{n=0}^{\infty} \beta^n (-\tilde{L})$$

[數6]

$$\tilde{L} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}, \qquad D(i,j) = \sum_j A(i,j)$$

$R_\beta(A)$之(i，j)元素與n元語法i與n元語法j之相似度對應。因此，擷取部108使用種子向量v_o(與種子字元串對應之元素為1，除此之外為0之矢量)，計算$R_\beta(A)v_o$並將計算之值作為相似度。β之值為任意，例如，為1.0-2。

例如，圖11之相連詞模型G_M中，「中華炒鍋」連接於「之」，且「烹飪夾」、「塔吉鍋」兩者都連接於「之」。又，連接於「中華炒鍋」之「之」雖連接於「塔吉鍋」，但未連接於「烹飪夾」。如此之情形下，「烹飪夾」相對於「中華炒鍋」之相似度$R_\beta(A)v_o$之值小於「塔吉鍋」相對於「中華炒鍋」之相似度$R_\beta(A)v_o$。

擷取部108，例如，將計算之相似度超過特定之值者作為實例擷取。例如，若如圖12所示求得相似度，並使特定之值為「0.10」，則擷取部108將「壓力鍋」、「中華炒鍋」、「親子鍋」、「塔吉鍋」、「伊賀陶器」作為實例擷取。

本實施形態中，CPU 501作為擷取部108發揮功能。

(5.實施形態2之集合擴張處理裝置之動作)

其次，就本實施形態之集合擴張處理裝置100之各部進行之動作使用圖13之流程圖進行說明。若電源進入集合擴張處理裝置100且進行特定之操作，則CPU 501開始圖13之流程圖中所示之集合擴張處理。另，圖13之流程圖中，附

有與圖9之流程圖相同之步驟號之步驟，進行與圖9之流程圖之處理相同之處理。因此，省略該等之說明。

若藉由選擇部107選擇實例之候補(步驟S107)，則擷取部108藉由使用實例之候補用檢索引擎檢索獲取片段(步驟S208)。例如，擷取部108將實例之候補作為查詢輸入至WEB網頁之檢索引擎，從而獲取檢索結果之前300件之片段之表。

其次，擷取部108自獲取之片段生成包含實例之候補之n元語法之相連詞模型(步驟S209)。例如，擷取部108，自300件之片段剔除不適當者，針對包含於剩餘之片段之集合之全部之文字之n元語法，求關聯矩陣M。且，如圖11所示，將全部之n元語法之集合V作為節點集合，生成將M(數2)作為關聯矩陣表現之相連詞模型G_M。

擷取部108根據相連詞模型之種子字元串之前後之語境與實例之候補之前後之語境，計算種子字元串與實例之候補之相似度(步驟S210)。例如，擷取部108根據式(數3)，求取右側語境之相似度矩陣A_R、及左側語境之相似度矩陣A_L，且如式(數4)所示，求取針對每個元素進行加權一般化平均之相似度矩陣A。進而，根據式(數5、6)，求取使用相似度矩陣A之拉普拉斯核$R_\beta(A)$，且乘以種子向量v_0，而求取相對於種子字元串之實例之候補之相似度。

擷取部108根據相似度而擷取實例(步驟S211)。例如，如圖12所示，擷取部108擷取所計算之相似度超過「0.10」者作為實例。又，或者擷取部108亦可自相似度

較高者擷取僅特定之個數。例如，實例之候補如圖12所示有9個之情形時，若將特定之個數設為4個，則擷取部108擷取相似度中前4個之「壓力鍋」、「中華炒鍋」、「親子鍋」、及「塔吉鍋」作為實例。

根據本實施形態，可排除意義上無關之詞語，且可對於視為包含於意義上相同之類別擷取更適當之用語。

另，實施形態1、2中，集合擴張處理裝置100雖顯示適用於生成購物場所之商品之類別之例，但並不限於此。例如，可應用於固有表現獲取或字典建構等。

本發明係基於2011年3月4日提出申請之日本專利申請案第2011-048124號。該案之說明書、申請專利範圍、圖式全體以引用的方式併入本文中。

產業上之可利用性

根據本發明，可提供適合於選擇屬於意義上相同之類別之字之候補之集合擴張處理裝置、集合擴張處理方法、程式、及非暫時性記錄媒體。

【圖式簡單說明】

圖1係顯示本發明之實施形態之集合擴張處理裝置與購物伺服器之關係之圖。

圖2係顯示實現本發明之實施形態之集合擴張處理裝置之典型之資訊處理裝置之概要構成之圖。

圖3係用以說明實施形態1之集合擴張處理裝置之概要構成之圖。

圖4係用以說明檢索之文件之圖。

圖5係用以說明區段之圖。

圖6係用以說明區段元素之圖。

圖7係用以說明區段分數及區段元素分數之圖。

圖8係用以說明選擇之實例之候補之圖。

圖9係用以說明實施形態1之集合擴張處理裝置之各部進行之集合擴張處理之流程圖。

圖10係用以說明實施形態2之集合擴張處理裝置之概要構成之圖。

圖11係用以說明相連詞模型之圖。

圖12係用以說明擷取之實例之圖。

圖13係用以說明實施形態2之集合擴張處理裝置之各部進行之集合擴張處理之流程圖。

【主要元件符號說明】

100	集合擴張處理裝置
101	接收部
102	檢索部
103	區段獲取部
104	區段元素獲取部
105	區段分數計算部
106	區段元素分數計算部
107	選擇部
108	擷取部
200	購物伺服器
300	網路

401	終端裝置
402	終端裝置
40n	終端裝置
500	資訊處理裝置
501	CPU
502	ROM
503	RAM
504	NIC
505	圖像處理部
506	聲音處理部
507	DVD-ROM光碟機
508	介面
509	外部記憶體
510	控制器
511	監視器
512	揚聲器
601	檢索欄
602	檢索按鈕

發明專利說明書

（本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫）

※ 申請案號：101106650

※ 申請日：101.2.29　　※IPC 分類：G06F 17/30 (2006.01)

一、發明名稱：（中文/英文）

集合擴張處理裝置、集合擴張處理方法、程式、及非暫時性記錄媒體

二、中文發明摘要：

接收部(101)接收種子字元串。檢索部(102)獲取包含種子字元串之文件之片段。區段獲取部(103)以區段分隔字元串分隔該片段而獲取區段。區段元素獲取部(104)以區段元素分隔字元串分隔區段而獲取區段元素。區段分數計算部(105)根據區段元素之長度之標準偏差，而計算區段之區段分數。區段元素分數計算部(106)根據種子字元串之位置與區段元素之位置之距離、及區段分數，而計算區段元素之區段元素分數。選擇部(107)基於區段元素分數，自區段元素中選擇任意一個作為種子字元串之擴張集合中所含之實例之候補。

三、英文發明摘要：

七、申請專利範圍：

1. 一種集合擴張處理裝置，其特徵為包含：

接收部，其係接收種子字元串；

檢索部，其係檢索包含上述接收之種子字元串之文件，而獲取該檢索之文件之片段；

區段獲取部，其係以特定之區段分隔字元串分隔上述獲取之片段，而獲取包含將出現於上述接收之種子字元串之前後之字元串、及將該種子字元串按出現順序排列之字元串之區段；

區段元素獲取部，其係以特定之區段元素分隔字元串分隔上述獲取之區段之各者，而獲取區段元素；

區段分數(Segment Score)計算部，其係根據於該區段中出現之區段元素之各者之長度之離差或標準偏差，而計算上述獲取之區段之各者之區段分數；

區段元素分數計算部，其根據上述接收之種子字元串於上述獲取之區段中出現之位置與該區段元素於該區段中出現之位置之距離、及針對該區段計算之區段分數，而計算該區段之各者中所含之區段元素之各者之區段元素分數；及

選擇部，其係根據針對上述獲取之區段元素之各者所計算之區段元素分數，自該區段元素中選擇任意一個作為將包含上述接收之種子字元串之集合擴張而成之擴張集合中所含之實例之候補。

2. 如請求項1之集合擴張處理裝置，其中進而包含：

擷取部，其係由使用上述實例之候補進行檢索而獲取之片段，生成包含上述擷取之實例之候補之n元語法之相連詞模型(N-gram)，並根據該相連詞模型之上述接收之種子字元串之前後之語境與該實例之候補之前後之語境，計算該種子字元串與該實例之候補之相似度，且根據該相似度，自該實例之候補中，擷取將包含該種子字元串之集合擴張而成之擴張集合中所應含有之實例。

3. 如請求項1或2之集合擴張處理裝置，其中

針對上述獲取之區段之各者，若於該區段中出現之區段元素之各者之長度之標準偏差超過特定之臨限值之情形，上述區段分數及上述區段元素分數成為該區段中所含之區段元素不會被上述選擇部選擇作為上述實例之候補之值。

4. 如請求項1之集合擴張處理裝置，其中

於上述獲取之區段之各者中出現之區段元素之各者之區段元素分數，相對於上述接收之種子字元串於該區段中出現之位置與該區段元素於該區段中出現之位置之最短距離呈指數衰減。

5. 一種集合擴張處理方法，其特徵在於其係具備接收部、檢索部、區段獲取部、區段元素獲取部、區段分數計算部、區段元素分數計算部、及選擇部之集合擴張處理裝置所執行者，且包含：

接收步驟，由上述接收部接收種子字元串；

檢索步驟，由上述檢索部檢索包含上述接收之種子字

元串之文件，而獲得該檢索之文件之片段；

區段獲取步驟，由上述區段獲取部以特定之區段分隔字元串分隔上述獲取之片段，而獲取包含將出現於上述接收之種子字元串之前後之字元串、及將該種子字元串按出現順序排列之字元串之區段；

區段元素獲取步驟，由上述區段元素獲取部以特定之區段元素分隔字元串分隔上述獲取之區段之各者，而獲取區段元素；

區段分數計算步驟，由上述區段分數計算部根據於該區段中出現之區段元素之各者之長度之離差或標準偏差，而計算上述獲取之區段之各者之區段分數；

區段元素分數計算步驟，由上述區段元素分數計算部根據該區段中上述接收之種子字元串出現之位置與該區段中該區段元素出現之位置之距離及針對該區段計算之區段分數計算包含於上述獲取之區段之各者之區段元素之各自之區段元素分數；及

選擇步驟，由上述選擇部根據針對上述獲取之區段元素之各者所計算之區段元素分數，自該區段元素中選擇任意一個作為將包含上述接收之種子字元串之集合擴張而成之擴張集合中所含之實例之候補。

6. 一種程式，其特徵在於使電腦作為下述各部發揮功能：

接收部，其係接收種子字元串；

檢索部，其係檢索包含上述接收之種子字元串之文件，而獲取該檢索之文件之片段；

區段獲取部，其係以特定之區段分隔字元串分隔上述獲取之片段，而獲取包含將出現於上述接收之種子字元串之前後之字元串、及將該種子字元串按出現順序排列之字元串之區段；

區段元素獲取部，其係以特定之區段元素分隔字元串分隔上述獲取之區段之各者，而獲取區段元素；

區段分數計算部，其係根據於該區段中出現之區段元素之各者之長度之離差或標準偏差，而計算上述獲取之區段之各者之區段分數；

區段元素分數計算部，其根據上述接收之種子字元串於上述獲取之區段中出現之位置與該區段元素於該區段中出現之位置之距離、及針對該區段計算之區段分數，而計算該區段之各者中所含之區段元素之各者之區段元素分數；及

選擇部，其係根據針對上述獲取之區段元素之各者所計算之區段元素分數，自該區段元素中選擇任意一個作為將包含上述接收之種子字元串之集合擴張而成之擴張集合中所含之實例之候補。

7. 一種記錄程式之非暫時性之電腦可讀取之記錄媒體，其特徵在於使電腦作為下述各部發揮功能：

接收部，其係接收種子字元串；

檢索部，其係檢索包含上述接收之種子字元串之文件，而獲取該檢索之文件之片段；

區段獲取部，其係以特定之區段分隔字元串分隔上述

獲取之片段，而獲取包含將出現於上述接收之種子字元串之前後之字元串、及將該種子字元串按出現順序排列之字元串之區段；

區段元素獲取部，其係以特定之區段元素分隔字元串分隔上述獲取之區段之各者，而獲取區段元素；

區段分數計算部，其係根據於該區段中出現之區段元素之各者之長度之離差或標準偏差，而計算上述獲取之區段之各者之區段分數；

區段元素分數計算部，其根據上述接收之種子字元串於上述獲取之區段中出現之位置與該區段元素於該區段中出現之位置之距離、及針對該區段計算之區段分數，而計算該區段之各者中所含之區段元素之各者之區段元素分數；及

選擇部，其係根據針對上述獲取之區段元素之各者所計算之區段元素分數，自該區段元素中選擇任意一個作為將包含上述接收之種子字元串之集合擴張而成之擴張集合中所含之實例之候補。

八、圖式：

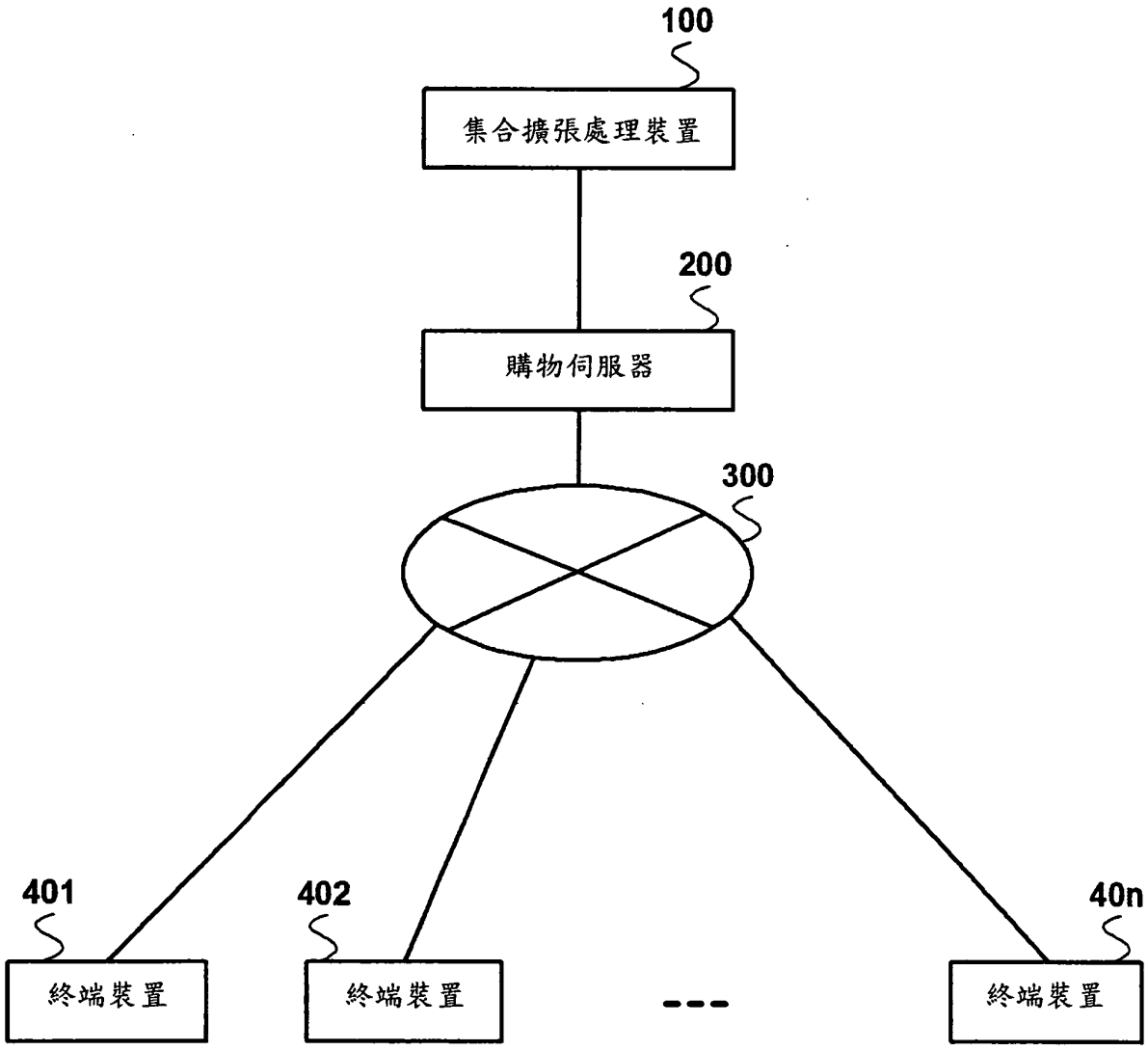


圖 1

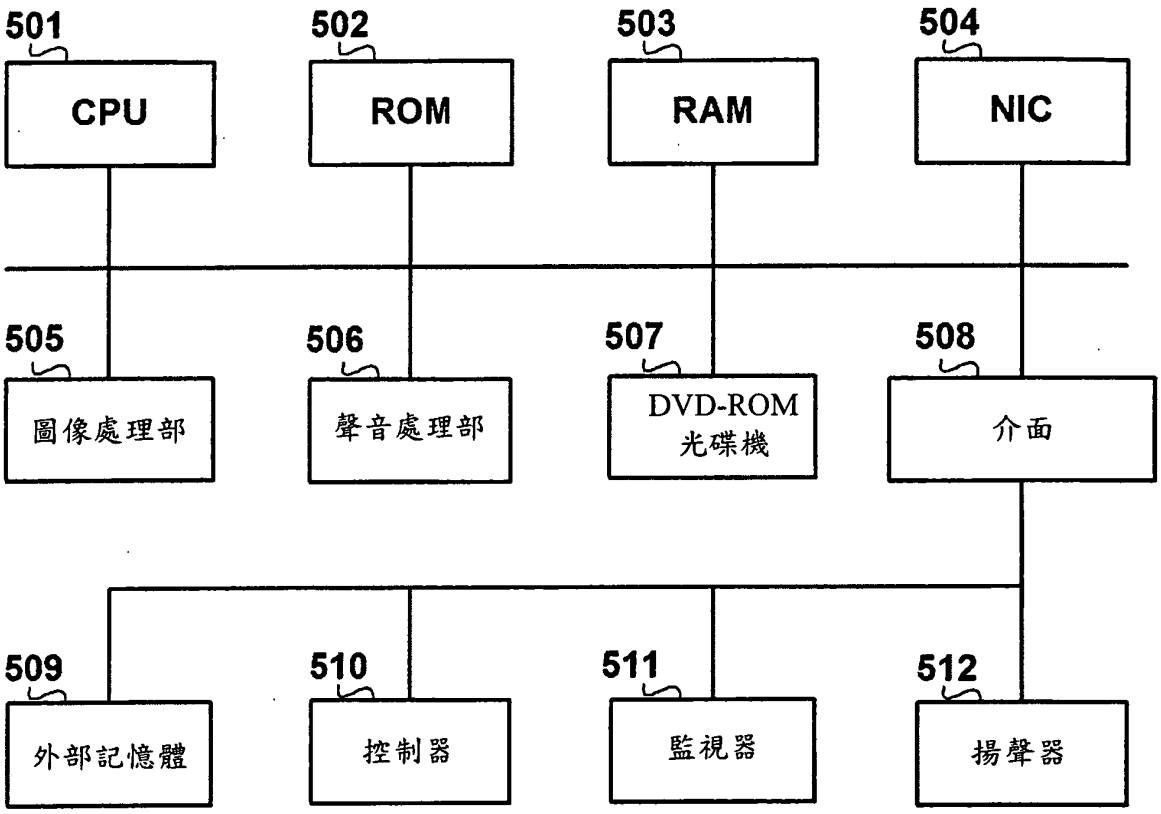

501
CPU
502
ROM
503
RAM
504
NIC
505
圖像處理部
506
聲音處理部
507
DVD-ROM
光碟機
508
介面
509
外部記憶體
510
控制器
511
監視器
512
揚聲器
500

圖 2

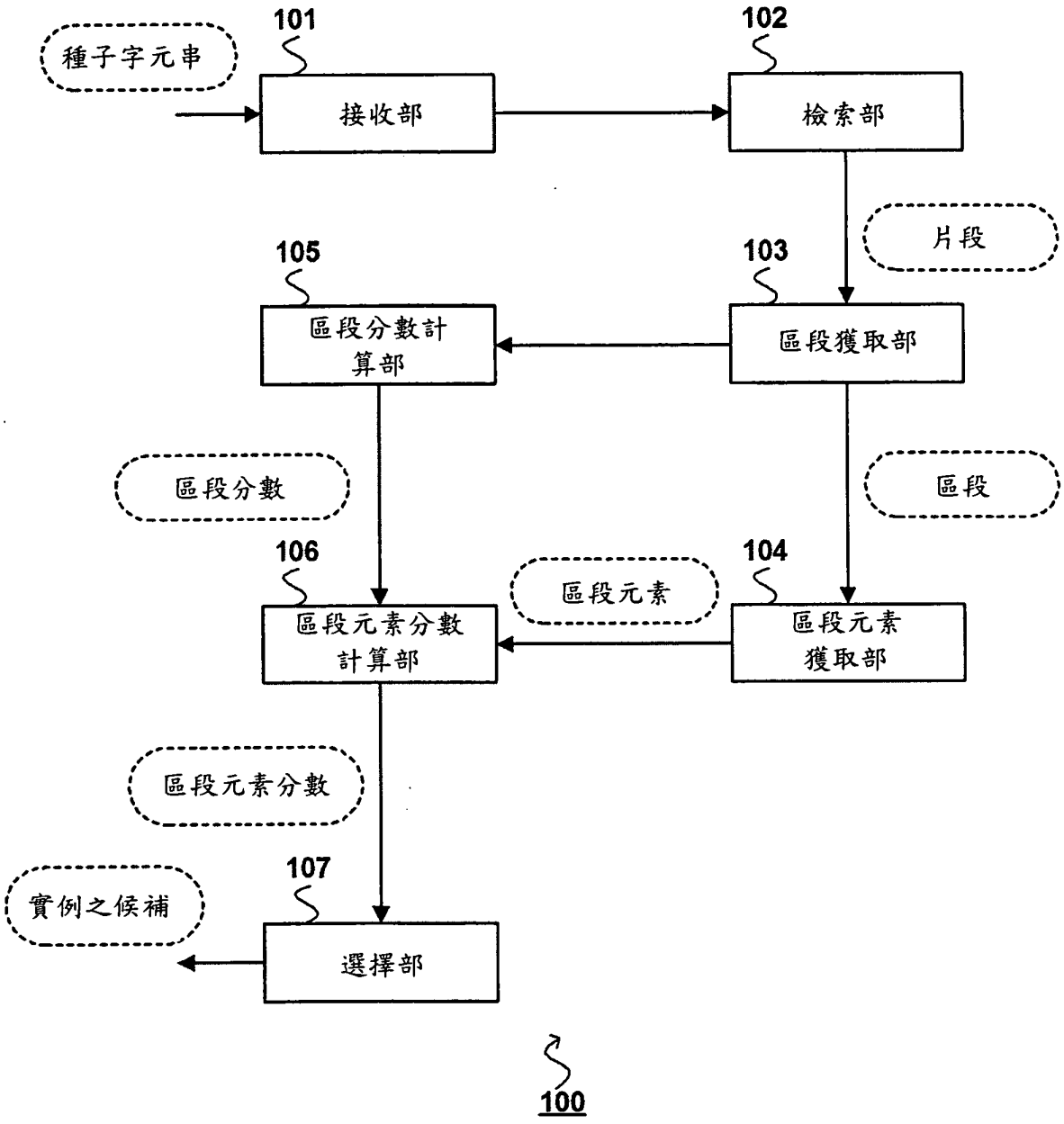

種子字元串
101
接收部
102
檢索部
片段
105
區段分數計算部
103
區段獲取部
區段分數
區段
106
區段元素分數計算部
區段元素
104
區段元素獲取部
區段元素分數
107
實例之候補
選擇部
100

圖 3

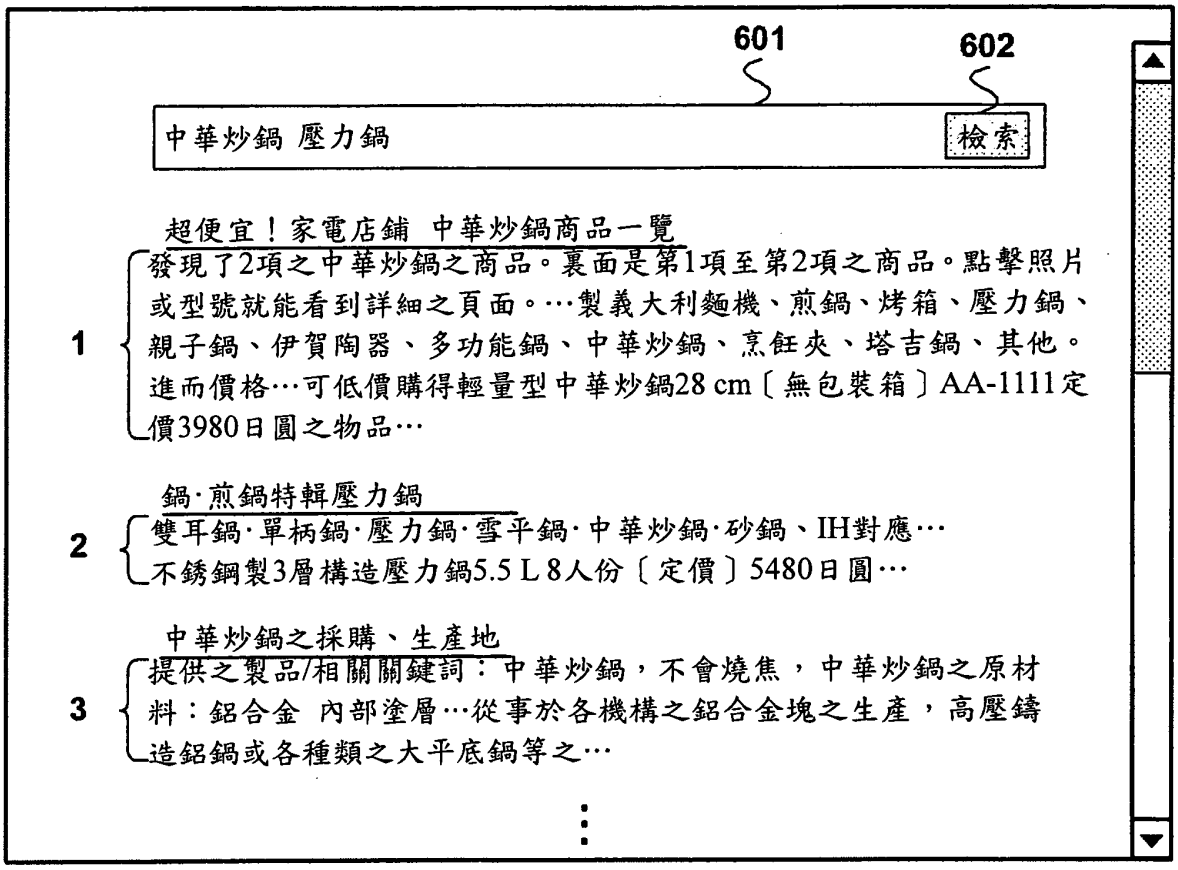

601
602
中華炒鍋 壓力鍋
檢索
超便宜！家電店鋪 中華炒鍋商品一覽
1
發現了2項之中華炒鍋之商品。裏面是第1項至第2項之商品。點擊照片或型號就能看到詳細之頁面。…製義大利麵機、煎鍋、烤箱、壓力鍋、親子鍋、伊賀陶器、多功能鍋、中華炒鍋、烹飪夾、塔吉鍋、其他。進而價格…可低價購得輕量型中華炒鍋28 cm〔無包裝箱〕AA-1111定價3980日圓之物品…
鍋·煎鍋特輯壓力鍋
2
雙耳鍋·單柄鍋·壓力鍋·雪平鍋·中華炒鍋·砂鍋、IH對應…
不銹鋼製3層構造壓力鍋5.5 L 8人份〔定價〕5480日圓…
中華炒鍋之採購、生產地
3
提供之製品/相關關鍵詞：中華炒鍋，不會燒焦，中華炒鍋之原材料：鋁合金 內部塗層…從事於各機構之鋁合金塊之生產，高壓鑄造鋁鍋或各種類之大平底鍋等之…

圖 4

1-1

發現了2項之中華炒鍋之商品。裏面是第1項至第2項之商品。點擊照片或型號就能看到詳細之頁面。

1-2

製義大利麵機、煎鍋、烤箱、壓力鍋、親子鍋、伊賀陶器、多功能鍋、中華炒鍋、烹飪夾、塔吉鍋、其他。進而價格係

1-3

可低價購得輕量型中華炒鍋28 cm〔無包裝箱〕aa-1111定價3980日圓之物品

圖 5

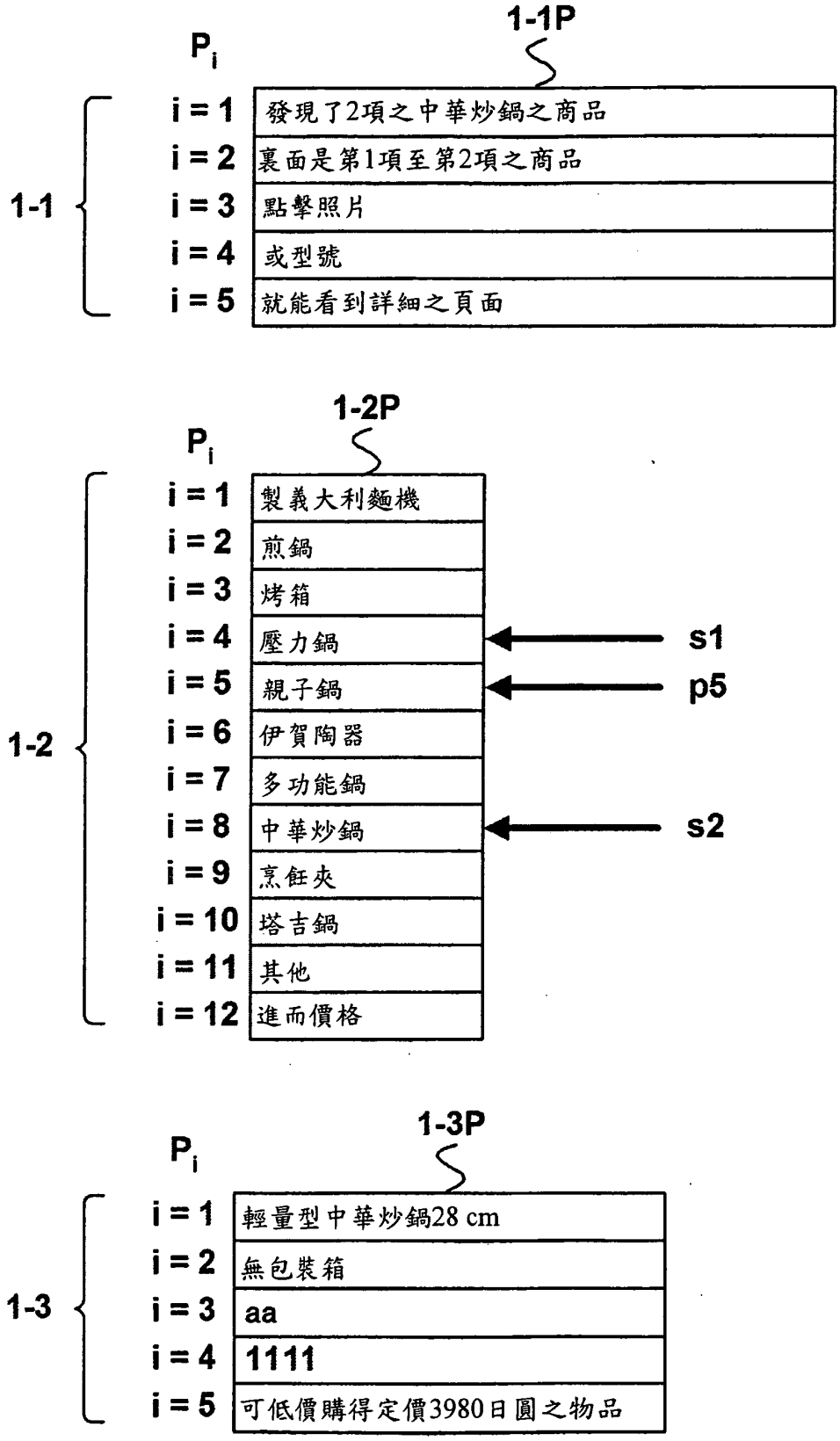

1-1P
Pi
1-1
i = 1
發現了2項之中華炒鍋之商品
i = 2
裏面是第1項至第2項之商品
i = 3
點擊照片
i = 4
或型號
i = 5
就能看到詳細之頁面
1-2P
Pi
1-2
i = 1
製義大利麵機
i = 2
煎鍋
i = 3
烤箱
i = 4
壓力鍋
s1
i = 5
親子鍋
p5
i = 6
伊賀陶器
i = 7
多功能鍋
i = 8
中華炒鍋
s2
i = 9
烹飪夾
i = 10
塔吉鍋
i = 11
其他
i = 12
進而價格
1-3P
Pi
1-3
i = 1
輕量型中華炒鍋28 cm
i = 2
無包裝箱
i = 3
aa
i = 4
1111
i = 5
可低價購得定價3980日圓之物品

圖 6

701a 片段	702a 區段	703a 區段元素	704a 長度	705a 標準偏差	706a 區段分數	707a 區段元素分數
1	1-1	P1	17	5.89	5.00	0
		P2	17			0
		P3	3			0
		P4	7			0
		P5	14			0
	1-2	P1	7	1.34	1.34	0.09
		P2	5			0.20
		P3	5			0.45
		P4	3			1.00
		P5	3			0.45
		P6	3			0.20
		P7	5			0.45
		P8	3			1.00
		P9	5			0.45
		P10	3			0.20
		P11	3			0.09
		P12	5			0.04
	1-3	P1	12	5.27	5.00	0
		:	:			:
		P5	14			0
2	:	:	:	:	:	:
3	:	:	:	:	:	:
:	:	:	:	:	:	:
300	:	:	:	:	:	:

圖 7

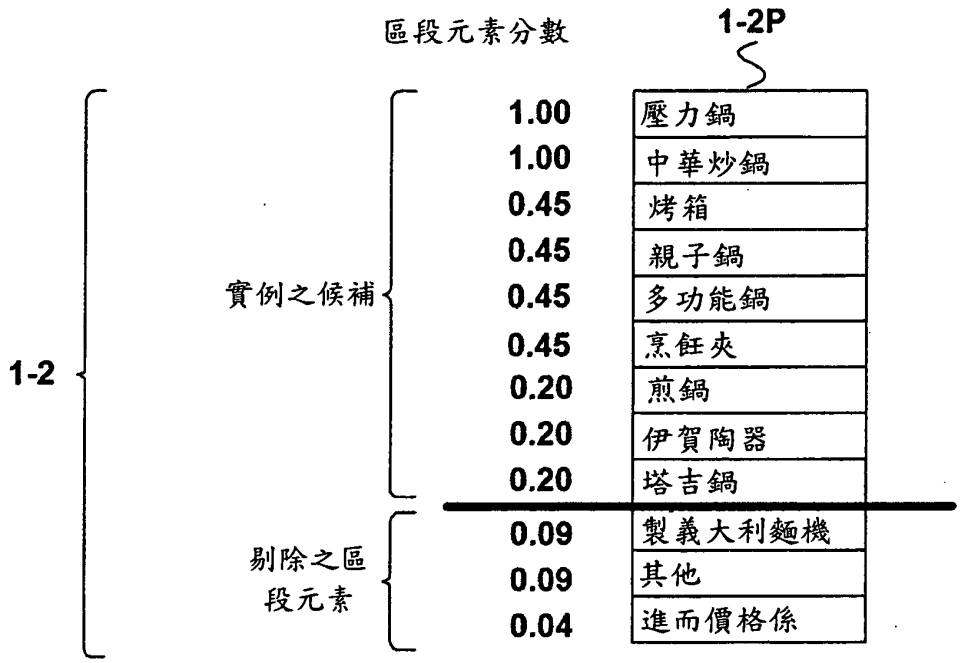

區段元素分數
1-2P
1-2
實例之候補
剔除之區段元素
1.00 壓力鍋
1.00 中華炒鍋
0.45 烤箱
0.45 親子鍋
0.45 多功能鍋
0.45 烹飪夾
0.20 煎鍋
0.20 伊賀陶器
0.20 塔吉鍋
0.09 製義大利麵機
0.09 其他
0.04 進而價格係

圖 8

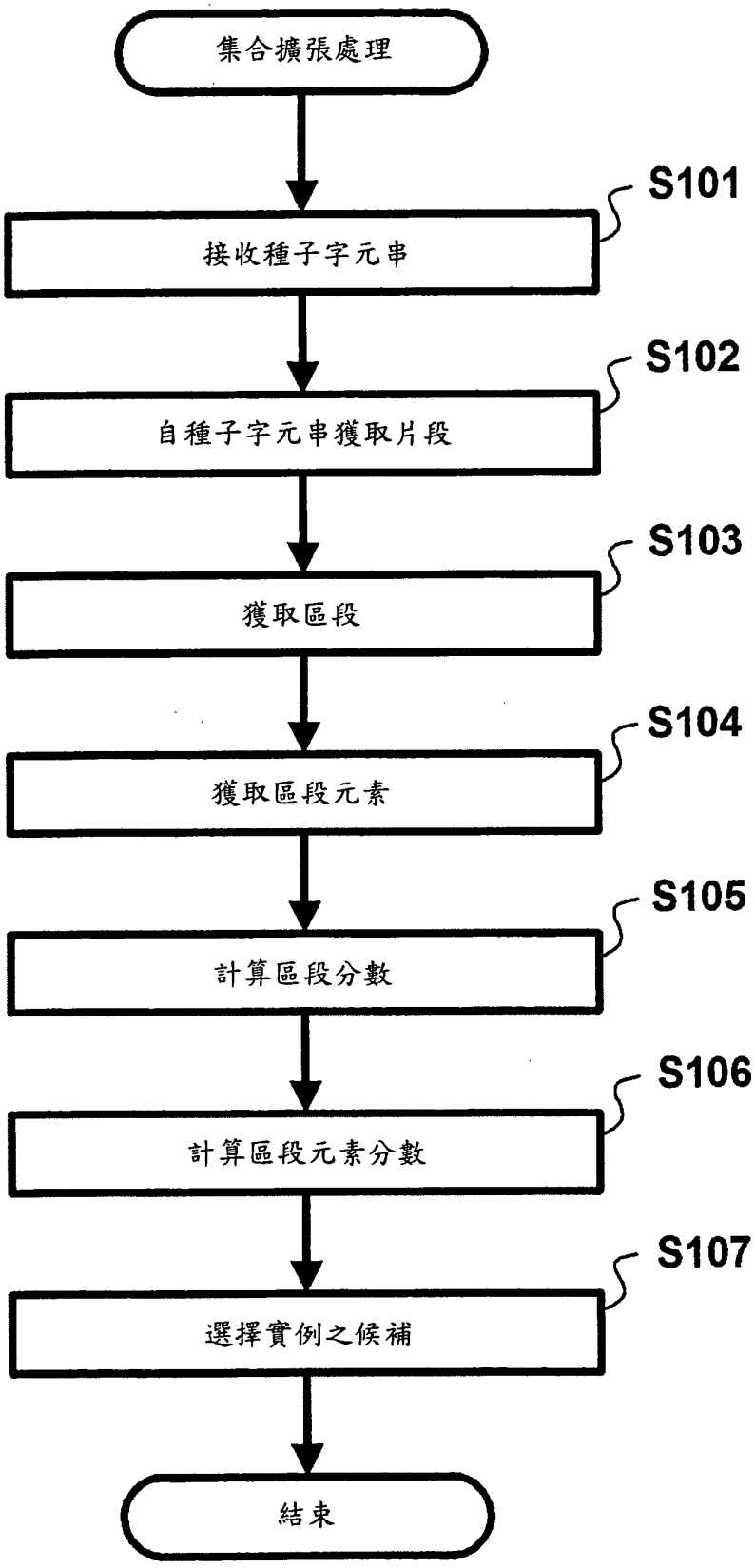

集合擴張處理
S101
接收種子字元串
S102
自種子字元串獲取片段
S103
獲取區段
S104
獲取區段元素
S105
計算區段分數
S106
計算區段元素分數
S107
選擇實例之候補
結束

圖 9

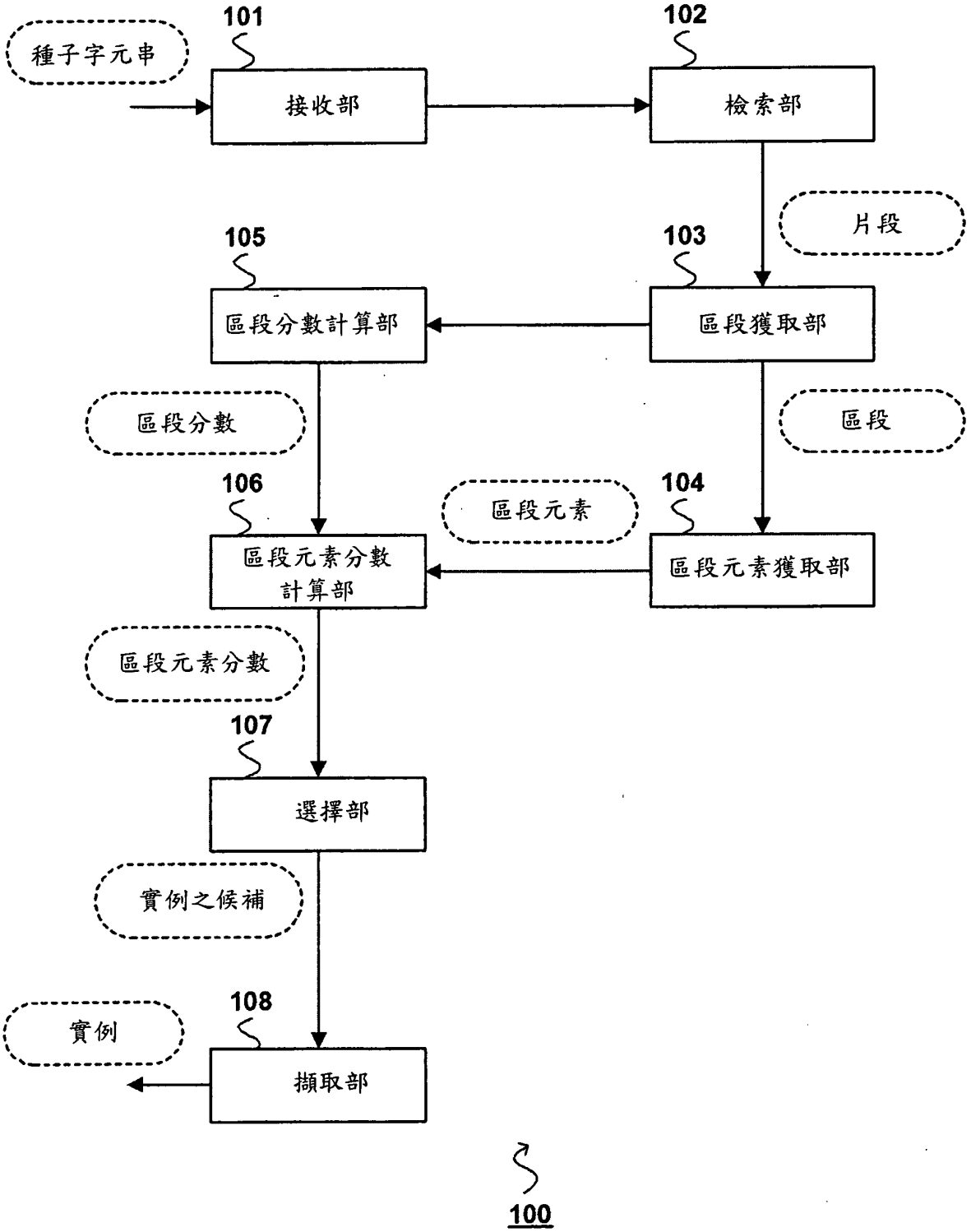

101
種子字元串
接收部
102
檢索部
片段
105
103
區段分數計算部
區段獲取部
區段分數
區段
106
區段元素
104
區段元素分數計算部
區段元素獲取部
區段元素分數
107
選擇部
實例之候補
108
實例
擷取部
100

圖 10

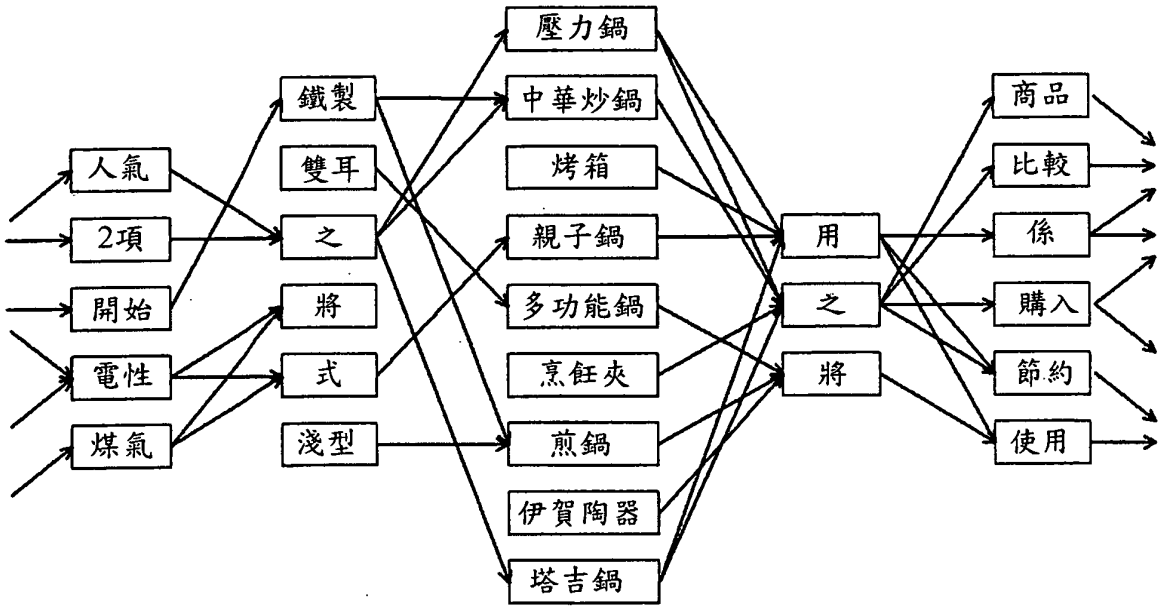

人氣
2項
開始
電性
煤氣
鐵製
雙耳
之
將
式
淺型
壓力鍋
中華炒鍋
烤箱
親子鍋
多功能鍋
烹飪夾
煎鍋
伊賀陶器
塔吉鍋
用
之
將
商品
比較
係
購入
節約
使用

圖 11

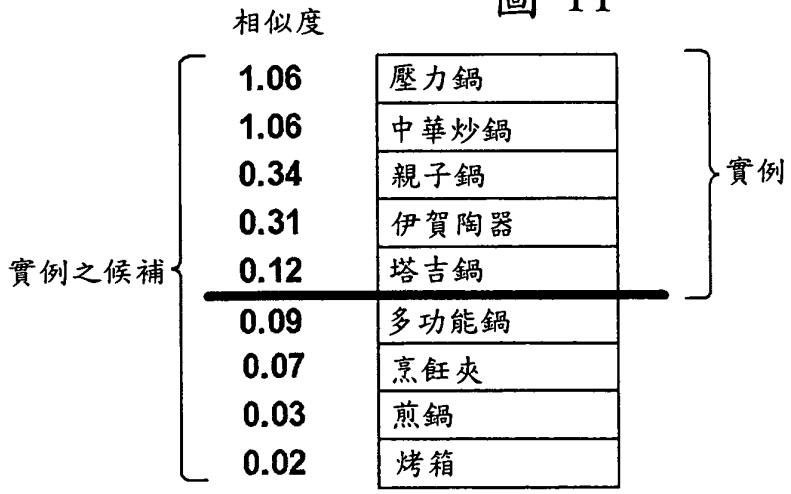

相似度
1.06
1.06
0.34
0.31
0.12
0.09
0.07
0.03
0.02
壓力鍋
中華炒鍋
親子鍋
伊賀陶器
塔吉鍋
多功能鍋
烹飪夾
煎鍋
烤箱
實例
實例之候補

圖 12

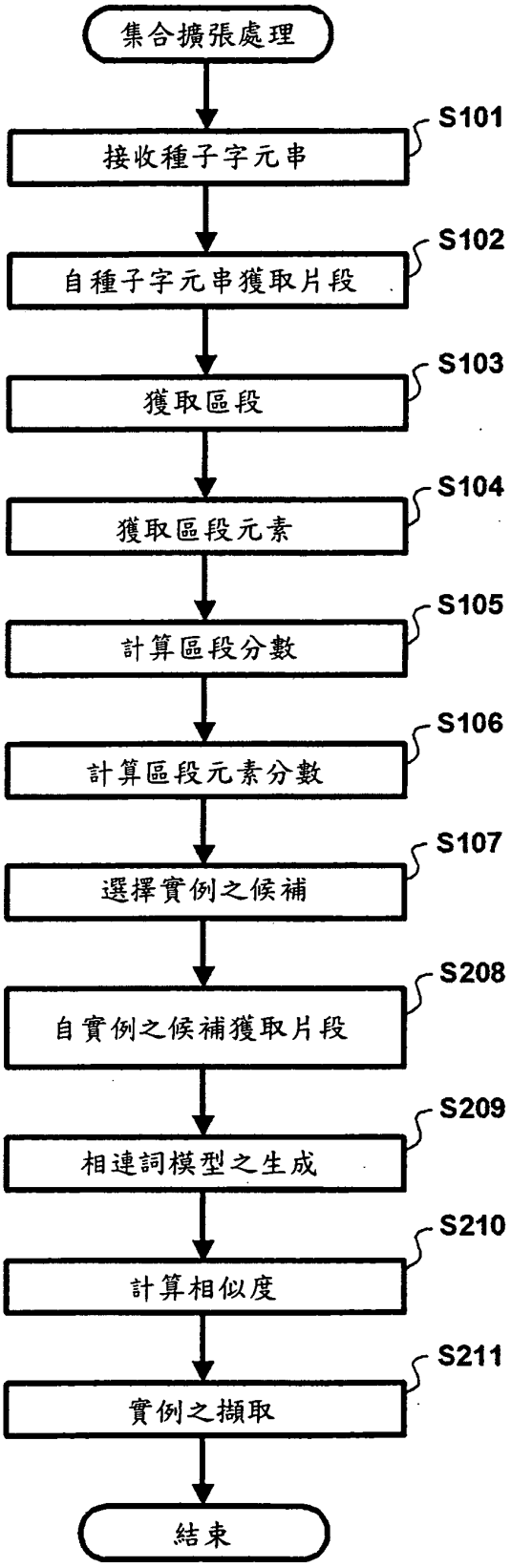

集合擴張處理
S101
接收種子字元串
S102
自種子字元串獲取片段
S103
獲取區段
S104
獲取區段元素
S105
計算區段分數
S106
計算區段元素分數
S107
選擇實例之候補
S208
自實例之候補獲取片段
S209
相連詞模型之生成
S210
計算相似度
S211
實例之擷取
結束

圖 13

四、指定代表圖：

(一)本案指定代表圖為：第（ 3 ）圖。

(二)本代表圖之元件符號簡單說明：

100	集合擴張處理裝置
101	接收部
102	檢索部
103	區段獲取部
104	區段元素獲取部
105	區段分數計算部
106	區段元素分數計算部
107	選擇部

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

(無)