

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：94111270

※申請日期：94.6.24

※IPC分類：G06F17/30, G06F17/32

一、發明名稱：(中文/英文)

專利文件語意結構之建立方法

Patent Document Content Construction Method

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

國立臺灣科技大學

National Taiwan University of Science and Technology

代表人：(中文/英文) 陳希舜

住居所或營業所地址：(中文/英文)

10672 台北市基隆路四段 43 號

NO. 43, SEC. 4, KEELUNG RD., TAIPEI CITY

國 稷：(中文/英文) 中華民國 R.O.C.

三、發明人：(共 4 人)

姓 名：(中文/英文)

- 1.蘇豐文/ SOO, VONWUN
- 2.林士能/ LIN, SHIHENG
- 3.楊世堯/ YANG, SHIHYAO
- 4.林斯寅/ LIN, SZUYIN

國 稷：(中文/英文)

- 1.中華民國 R.O.C.
- 2.中華民國 R.O.C.
- 3.中華民國 R.O.C.
- 4.中華民國 R.O.C.

四、聲明事項：

主張專利法第二十二條第二項第一款或第二款規定之事實，其事實發生日期為： 年 月 日。

申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

有主張專利法第二十七條第一項國際優先權：

1.

2.

無主張專利法第二十七條第一項國際優先權：

主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

主張專利法第三十條生物材料：

須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

九、發明說明：

【發明所屬之技術領域】

本發明是有關於一種文字結構擷取方法，且特別是有關於一種專利文件的文字結構擷取方法。

【先前技術】

過去當開發一新技術時，為了避免侵權，通常必須閱讀並比對數十甚至數百篇的專利文件。然而，專利文件係以文字型態表示，因此進行比對時，需採用全人工方式進行判讀。其中大部分專利文件屬性相似，但卻必須花費大量時間來分析鑑定，耗費很多寶貴人力，缺乏一套有效率的機制。因此，如何設計一種新的方法，能夠自動擷取專利文件之語意結構以及自動進行相似度比對，是業界非常需要的。

【發明內容】

因此本發明的目的就是在提供一種專利文件語意結構建立方法，可以將專利文件的專利申請範圍(Claims)做自動化分析與結構擷取。

本發明的另一目的就是在提供一種專利文件語意結構建立方法，可以整合專業領域詞庫，並將專業知識轉化成可分享可再利用的標準化知識。

本發明的又一目的就是在提供一種專利文件語意結構建立方法，可以幫助知識之擷取與檢索，提供更精準的專業資訊。

根據本發明之上述目的，提出一種專利文件語意結構建立方法。依照本發明一較佳實施例，此方法包括下列步驟。建立一領域之一詞彙詞庫，其中此詞彙詞庫包含此領域中之複數個專業詞彙。這些專業詞彙形成一階層式架構。對一專利之一申請專利範圍進行一語意語法註記，以分辨此申請專利範圍中之專業詞彙、停用字、一般詞彙以及標點符號。利用詞彙詞庫建立該申請專利範圍之一結構關係(structural relation)。此結構關係包括申請專利範圍中之專業詞彙、一般詞彙以及申請專利範圍中之專業詞彙互相間的三元關係。

本發明至少具有下列優點，其中每一實施例可以具有一個或多個優點。本發明的專利文件語意結構建立方法可以將專利文件的專利申請範圍(Claims)做自動化分析與結構擷取。本發明的專利文件語意結構建立方法可以整合專業領域詞庫與領域知識，並將專業知識轉化成可分享可再利用的標準化本體知識供再利用。本發明的專利文件語意結構建立方法可以幫助知識之擷取與檢索，提供更精準的專業資訊。

【實施方式】

本發明提出了一個新的專利文件語意結構擷取系統。此系統可以自動分析專利文件的語意結構，並且將語意結構擷取出來，最後再透過圖形化介面顯示專利文件的語意結構。本發明主要描述內容即為“將專利文件轉換成為以領域知識(domain knowledge)為基礎的機讀式(machine

readable)語意架構(semantic structure)”。

由於專利申請範圍是專利的權利範圍依據與發明最大權利的判定準則，所以深度了解專利申請範圍的內容是最有價值的部份。要讓電腦可以自動剖析專利聲請範圍的語意內容，並且讓人快速了解其內容，必須先克服四個問題。第一、要了解專利內容所描述的專業領域詞彙。第二、要了解專利文件的法律用語和寫作規則。第三、為了讓電腦理解專利內容，須將專利聲請範圍轉換成為機讀式語意架構。第四、為了讓人快速了解專利內容，可以將冗長的專利聲請範圍轉換成為簡潔易懂的圖型化表示法。

在本發明中提出了幾項方法來解決這些困難；藉由專業領域詞彙的建立，系統在剖析特定領域的專利文獻時，就可以得知哪些詞彙是專有名詞，以及專有名詞的涵義。接著透過註記過程，系統便可以知道專利聲請範圍中的每一個字詞的語法/語意資訊。至此可以解決第一個問題。而專利文獻的法律用語和專利申請範圍的寫作規則，過去只能藉由人來閱讀法律書籍。本發明以分析歸納來取得，並將得到的規則，轉為正規表示式(regular expression)來擷取資訊並構成語意架構。至此可以解決第二、三個問題。最後再把語意架構轉換為圖形，解決第四個問題。

因此，在專利文件語意結構擷取流程圖中(第 2 圖)，首先將專利文件由美國專利局(USPTO)擷取出來(步驟 202)，存到資料庫。接著利用辭典編輯工具(步驟 204、206)，讓專家進行半自動專業詞典的建構(thesaurus construction)(步驟 208)。辭典可以用來幫助理解專利文件的特有詞彙及

語意註記，並且也是相似度演算法的參考依據之一。建立好專業領域辭典，系統就可以進行語法語意的註記(Semantic/Syntactic Annotation) (步驟 212)。由於正規表示式的內容是經由人類分析專利的法律用語和寫作規則建立而成，所以系統將運用正規表示式來擷取專利文件的語意資訊。取得語意資訊後，系統將會轉換成 OWL 格式的語意結構(步驟 218)，並且將語意結構以圖形化的方式呈現給使用者(步驟 220、222)。使用者藉由圖形化介面可以修正正規表示式擷取有誤的部份(步驟 228)，修正完成之後，再以 OWL 的格式回存到資料庫中(步驟 224)，即完成專利文件語意結構擷取。

本實施例以化學機械研磨領域的專利文件為實例來說明。第 3 圖繪示專利申請範圍維護工具的一個例子。首先從美國專利商標局(United Status Patent and Trademark Office, USPTO)提供的專利資料中，將化學機械研磨領域的專利文件挑選出來(步驟 202)，並且透過專利申請範圍維護工具(第 3 圖)，將每一篇專利文件的專利申請範圍擷取出來，存放在我們的資料庫中(步驟 204)。

本發明會描述下列內容：

- (1) .擷取專利文件語意結構的重要性和難處。
- (2) .專業領域辭典建立。
- (3) .專利文件語法/語意註記。
- (4) .正規表示式擷取語意結構。
- (5) .圖形化呈現專利文件語意結構。

電腦要能夠理解語意的重要性：

傳統的資訊擷取(Information Retrieval)領域，大多都是利用關鍵字詞來代表一篇文章的內容，並未針對文章的語意結構進行理解，大多只停留在語法結構的分析，再利用統計的方式來做文章相似度的比對。但是用關鍵字具有下列缺點：

1. 無法準確的表達語意。
2. 會找到額外不相關的垃圾資訊(garbage information)。

若是想要進一步突破現有資訊擷取之藩籬，勢必要讓電腦能夠更精確與深入的分析文章的內容，進而達到語意層面上的理解，將非結構性的文章段落，轉成結構性的資訊，將得運用到下列的資訊技術：

1. 專業領域辭典的建立。
2. 語法語意的註記。
3. 運用自然語言處理技術辨識文章的章法結構。
4. 將取得的結構性資訊轉為機讀式架構。

第 1 圖繪示一化學機械研磨的基本架構。本發明將以一關於化學機械研磨之專利當作實施例。化學機械研磨(Chemical Mechanical Polishing/Planarization)化學機械研磨法：為一種全面性平坦化技術，利用化學蝕刻與機械磨削兩者相互作用下，將凸出的沈積層，加以去除的一種平坦化技術。除化學機械研磨研磨盤(Polishing plate)102 轉動外，化學機械研磨研磨頭(polishing head)104 亦同時進行自轉，並依特定之軌跡運動，以達到最佳之研磨效果。此外，

研磨頭 104 經利用真空吸盤夾持晶圓 106，會導致晶圓 106 變形。故真空壓力亦影響研磨平坦度。故化學機械研磨研磨頭 104 需進行運動軌跡控制、轉速控制、真空壓力控制等。依據第 1 圖系統製造與設計分類原則，可經由電腦輔助工程(CAE)分析，並由專利分析所得之功能定義，決定所需之致動器、控制目標、控制策略以及控制器。

專利描述的內容一般分為兩類：

- (1). 方法類：專利的內容描述主要為方法或流程的陳述。
- (2). 結構類：專利的內容描述主要為元件結構的陳述。

專利文件中申請專利範圍(Claim)敘述的特性：

- (1)和一般文件不同，專利申請範圍句子很長。
- (2)有附屬項和獨立項的專利申請範圍，附屬項會加入獨立項中的限制條件。
- (3)用字有限定，且法律保護範圍不同：例如 Comprising、Consisting of。

專業詞彙詞庫的建構 (Thesaurus Construction)

人類在文件中想要表達專業知識時，一般都會使用領域的專有名詞來表達特定觀念，並且詳加描述這些觀念之間的關係，來表達所要描述的對象，而專利文件即是如此構成的。第 4A 圖繪示詞彙辭庫(thesaurus)之編碼原則之一範例。第 4B 圖繪示專業詞彙詞庫建構流程之一範例；第 5 圖繪示詞彙詞庫編輯工具之一範例。如果我們希望電腦可以理解一份專利文件，進而達成機讀式(machine readable)的目標，首要的工作就是由電腦將專業領域中的專業詞彙

(domain terminology)擷取出來，並且經由領域專家的判別(第 4A 圖、第 4B 圖)，透過詞彙詞庫編輯工具(第 5 圖)，將專業詞彙編輯成專業知識之詞彙詞庫。當領域專家完成編輯，就可以得到階層式架構的詞彙詞庫。在這個階層式架構中，每一個階層以及每一各詞彙都會有一個編碼，此編碼是賦予專業的詞彙分類的觀念，屬於同一階層同一群的詞彙，就表示它們都是同樣類型的東西，而機器便可以藉由編碼猜出語意內涵，得知此詞彙是哪一種材料、元件、還是工具...等等。

舉例來說，“rotating speed”這兩個字應該結合在一起當成機械領域的特定詞彙，將之拆開為“rotating”和“speed”兩個字，皆不能正確的表示所要表達的概念。如第 4B 圖所示，“rotating speed”在辭典中的語意編碼(semantic code)為“B1:2:2:1:1”，而“Rotational Speed”在辭典中的語意編碼為“B1:2:2:1”，所以電腦將可以判別“rotating speed”為“Rotational Speed”之下的一個特定概念。

專業詞彙詞庫的建構流程如第 4B 圖所示。首先將選定的化學機械研磨專利申請範圍經由專業辭典搜尋者(Domain Terminology Finder)將專利申請範圍中可能的專業詞彙挑選出來，其中專業辭典搜尋者是經由我們所設計的一些自然語言處理規則所構成，藉由統計的方式，統計專利文件宣告中，常常會出現的一字詞、二字詞...五字詞，挑選出可能是專業詞彙的專有名詞、專業領域會出現的多字詞(multiword terms)、單字詞.singleton words)。接著領域專家把正確的專業詞彙由系統建議的專業詞彙列表中挑選

出來，並且將它歸類到所屬的階層中，如此，便完成專業詞彙詞庫的建構。如上述的詞彙詞庫建構流程，我們可以得到合乎標準的詞彙詞庫(步驟 208)。

而專業辭典搜尋者是藉由統計的方式，統計相同領域的專利文件宣告中，常常會出現的一字詞、二字詞...五字詞。

詞彙詞庫(Thesaurus)編碼原則：

- 需要有 UID (root UID=000)
- 需要知道是 concept 或 instance
- 需要知道該 node 在 thesaurus 的深度
- 需要知道 parent node
- 編碼方式 : (001->999)(0|1)(00-99)(001->999)

系統目前所擁有的專業領域辭典：

在本發明中有三個專業領域辭典，分別為一.機械元件辭典：收集化學機械研磨領域中的機械元件詞。二.單位辭典：收集化學機械研磨領域的單位詞彙。 3. 屬性辭典：收集化學機械研磨領域的參數詞彙。

第 6 圖繪示一晶圓和一研磨墊的關係。第 7 圖繪示第 6 圖之晶圓和研磨墊之間的三元關係(triple relation)。化學機械研磨機具內研磨墊(Polishing Pad)602 和晶圓(Wafer)604 的關係為研磨(Polish)605，如此就可以很清楚的描述“研磨墊研磨晶圓”的關係。當機器處理專利文件時，藉由詞彙詞庫的支援，就能夠明白專利申請範圍中所提到的元件，元件之間所具有的關係，以及元件具備的相關屬性。

語意/語法註記

請同時參照第 2 圖，接下來進行語意/語法註記 (Semantic/Syntactic Annotation)(步驟 212)。為了讓機器可以處理專利文件，電腦必須先分析專利文件中的專業字詞的語意，以及每各字的語法資訊(例如字的詞性)，以利近一步做專利結構的擷取。第 8 圖繪示語意/語法加註流程圖的一個範例。在語意/語法註記流程中，首先進行專利申請範圍的斷詞，以單字詞為基本單位，進行詞性(POS)的語法註記。本實施例採用史丹佛大學所開發的 JavaNLP parser 來表記詞性，此剖析器(parser)會自動判斷每一個輸入的句子的句型結構，將之一一剖析成為詞組結構，並且運用機率和統計的方式，針對詞組中的每一個字給予一個特定的可能的詞性。第 9 圖繪示 JavaNLP 所產生的一範例剖析樹 (parsing tree)。第 9 圖所展示的，就是 JavaNLP parser 分析：“A polishing pad comprising: a first layer; a second layer; a hole formed in the polishing pad, the hole having: a first section in the first layer of the polishing pad.”所生成的詞性剖析樹。

系統將分成四個部份來進行語意註記：

(1) 專業詞彙註記：標註專利申請範圍中的專業詞彙。由專業辭典搜尋者 (Domain Thesaurus Tagger) 來達成，藉由專業詞彙詞庫的支援，將標註的號碼與專業詞彙詞庫比對，可以得知該專業詞彙的語意。

(2) 停用字(Stop Word)註記：標註專利申請範圍中的

停用字。如：“the”、“a”...等等，由停用字標記者（Stop Word Tagger）達成。

(3)一般詞彙註記：標註專利申請範圍中的一般生活詞彙之動詞。由詞網標記者(WordNet Tagger)來達成，藉由詞網（WordNet）的支援，可以標註得知該動詞的語意。

(4)標點符號註記：標註專利申請範圍中的標點符號，由標點符號標記者（Punctuation Tagger）來達成。第 10 圖為語意/語法註記範例。第 10 圖係針對某一條專利申請範圍進行語意/語法註記的結果。

接下來描述正規表示式擷取語意結構(如第 2 圖之步驟 214、216)。由於專利申請範圍是專利的權利範圍依據與發明最大權利的判定準則，所以深度了解專利申請範圍的內容是最有價值的部份。而專利申請範圍的撰寫方式有幾種法律格式，所以很適合用電腦來擷取其語意內容。然而正由於專利申請範圍在法律上重要的地位，專利撰寫的風格也常因為法院的判決時有所變，加上用詞遣字也較冷僻，文法規則和一般的文章也不同，使得專利申請範圍的內容不易閱讀，連帶加深了電腦剖析上的困難。接下來將說明如何利用正規表示式來擷取專利聲請範圍的語意結構。

正規表示式就是一種用來描述文字字串(text string)的模板(template or pattern)。它由一些字母以及一些具有特殊涵義的超字元(meta-characters)所構成這個模版，可以用來

擷取或是描述符合這種模版的文字字串。簡單的說，正規表示式是定義語言的語言。

在 1956 年數學家 Stephen Kleene 建構了一套數學符號系統-正規集 (regular sets)。很快的它就被應用在計算機領域中的編譯器的 scanner 與 lexical analysis。所以正規表示式是源於自動理論與正規語言理論。正規表示式是由對應的字串集合所定義，此集合稱為“由正規表示式所產生的語言”，並且可以被表達為 $L(r)$ 。

第 11 圖繪示一正規表示式的超字元 (meta-characters) 功能。運算優先權： $*$ > and > or

舉例來說：

$$L(a|b^*) = \{a, \varepsilon, b, bb, bbb, bbbb, \dots\}$$

$$L((a|b)^*) = \{\varepsilon, a, b, aa, ab, ba, bb, \dots\}$$

第 12 圖繪示用來擷取專利語意結構的八類正規表示式。本發明根據專利聲請範圍的撰寫方式，定義了八類正規表示式來擷取專利文獻的語意結構。

1. 一般類 (Common)：

此分類項主要的目的，就是設定一些基本常用的正規表示式，供其他類的正規表示式使用。第 13A 圖繪示此項分類之正規表示式以及其解釋。

2. 專利申請範圍類 (Claim)：

此分類項主要的目的，就是從整篇專利文件鑑別出專利

申請範圍，並且自動將每一項申請範圍切開，接著判別每一項專利申請範圍的類型是獨立項或是附屬項，並且判斷此專利申請範圍的內容類型是描寫一發明之裝置或機械結構，或是描寫一發明之方法或程序，或是屬於其他類型 (other type) 的內容描述。第 13B 圖繪示此項分類之正規表示式。解釋：在專利文件中，專利申請範圍通常有一個固定的寫法和開頭。第 13C 圖繪示一些固定的寫法和開頭的範例。例如範例 1 為美國專利 6,544,104 專利申請範圍 1，專利申請範圍的開頭寫作模式都是這樣，所以可以用來斷專利申請範圍。範例 2，表示一般附屬項都會有這樣的關鍵字，可以用來判別是獨立或是附屬項。範例 3 為美國專利 6,569,004 專利申請範圍 1，表示一般方法類型的專利申請範圍都有這樣的關鍵字，可以用來判斷專利申請範圍的內容型態。在本實施例中，只有處理結構類型的專利申請範圍。

3. 元件類 (Component)：

此分類項主要的目的，就是用來擷取專利申請範圍所描述的元件。第 13D 圖繪示用來擷取元件的來源。用來擷取元件的來源有兩種方式，第一種是藉由詞性分析(步驟 1302)，判斷該詞組是否為元件。另一種來源乃是由專業領域詞庫而來(步驟 1304)，只要是辭庫裡面的詞彙都是元件。第 13E 圖繪示正規表示式類別的執行順序。

第 13F 圖繪示 component(x) 之正規表示式。解釋：在專利申請範圍撰寫風格中，藉由分析語法資訊，可以發現其實元件大多都具有固定的詞性，並且大多都會在元件

前面加上一個冠詞。其中冠詞 said 是專利文件特有的寫作方式。本發明可以評估利用詞性擷取元件的涵蓋率和準確率。第 13G 圖繪示在美國專利 6,273,800 找出的元件之範例。範例 1 為美國專利 6,273,800 專利申請範圍 1，運用 regexComponent1 裡面所定義的詞性組合，可以找出如第 13G 圖的元件。

4. 參考類 (Reference) :

此分類項主要的目的，就是用來建立元件之間的參考連結，並且建立專利申請範圍附屬項以及獨立項之間的連結。由於在專利申請範圍撰寫的法定格式中，第一次描述的元件前面都會加上冠詞 a 或是 an，第二次(含以上)又再度描述的時候，便會加上 the 或是 said 來加以區別，並且明確的指名是所指的是哪一個元件。在元件類這一正規表示式執行時，會把所有的元件擷取出來，但是沒有建立元件之間的參考；而參考類這一類正規表示式，便是要自動將第二次(含以上)描述的元件建立參考到第一次描述的元件。如此可以簡化資訊的複雜度，便於人類分析閱讀。

實際作法中，系統會找尋第二次(含以上)被描述的元件，如果此元件出現在獨立項，系統將會在同樣的專利申請範圍中找尋第一次被描述的地方在哪裡；如果此元件是出現在附屬項中，系統會先利用第二條正規表示式來判斷這一項附屬項是參考到那一項獨立項，找到之後，再運用相同的方式建立參考索引。第 13H 圖繪示此類(reference)正規表示式類別的執行順序。第 13I 圖繪示此類(reference)之正規表示式。

解釋：例如範例 1 為美國專利 6,273,800 專利申請範圍 1，“polishing pad (Component_Token_1)”為第一次出現在專利申請範圍的研磨墊元件，而“polishing pad (Component_Token_6)”為第二次出現在專利申請範圍中的研磨墊元件，系統便會自動建立關聯表，裡面載明“Component_token_6”等於“Component_token_1”。而在範例 2 中，雖然“apparatus (Component_Token_23)”是在第二項專利申請範圍中被描述，但是系統還是能夠藉由正規表示式自動判斷，得知“Component_token_23”其實就是指第一項專利申請範圍中“Component_token_0”。

請參照第 13J 圖，第 13J 圖繪示在美國專利 6,273,800 找出的元件之範例。

5. 屬性類 (Attribute)：

此分類項主要的目的，就是要擷取專利申請範圍中元件的屬性描述。在此分類項中共有七個小項目，依序為屬性名稱(property)、指派關係(assignment)、值(value)、範圍(range)、單位(unit)、單位值(unitvalue)、屬性值(propertvalue)。

屬性名稱代表系統所要擷取的屬性名稱。第 13K 圖繪示正規表示法的定義範例。第 13L 圖繪示化學機械研磨之參數範例。因為化學機械研磨的參數眾多，在本實施例中元件的屬性分析將著重於化學機械研磨製程中各項製程監控參數、研磨時兩研磨面間的接觸形態及研磨劑(slurry)之流體狀況進行深入的探討，以利未來化學機械研磨製程專利之參數相似度比對。

“指派關係”表示屬性和屬性值之間的關係，關係可能為大於(greater than)、等於(equal to)、小於(less than)...等等情況。值代表屬性值，有可能是正負整數，正負小數，或是用英文描述的“one”、“two”、“three”等序詞。範圍用來抓取一數值範圍；“單位”代表屬性的單位，目前也是由人收及建立到資料庫，而“單位值”整合了“值”、“範圍”、“單位”，用來表示某值或是某範圍的數值及其單位；最後是“屬性值”，整合了“屬性名稱”、“指派關係”、“單位值”，用來表示某屬性在某單位的情況下和值的關係為何。用三元關係來表示可以定義成：PropertyValue(Property(x),Assignment(y),Valueunit(z))。

第 13M 圖繪示屬性類(attribute)之正規表示式類別的執行順序。第 13N 圖繪示屬性類(attribute)之正規表示式之範例。如第 13N 圖所示，屬性類這一分類項具有這七個成員。依據此一正規表示式，系統可以辨認波長(wavelength)為“屬性名稱”，“of”為“指派關係”，“190”和“350”為“值”並且因此構成“範圍”，奈米(nanometer)為“單位”，並和“範圍”構成“單位值”，最後在統合成為“屬性值”。經過了正規表示式的資訊擷取過程，系統可以在美國專利 6,454,634, 專利申請範圍 19 中，擷取出一屬性，並且得到如下的表示式：

PropertyValue(Property(wavelength),Assignment(of),ValueUnit(Range(Value(190),-,Value(3500)),-,Unit(nanometers)))。

第 13O 圖繪示由美國專利 6,454,634 專利申請範圍 19 中擷取屬性得到的表示式。

6. 功能類 (Functionality) :

此分類項主要的目的，就是用來擷取專利申請範圍中元件的功能性描述。在專利申請範圍中，元件時常會有功能性的描述，以便更清楚的界定此元件在此發明中的功用，及元件的法律權利範圍。第 13P 圖繪示此類(functionality)之正規表示式類別的執行順序。第 13Q 圖繪示此類(functionality)之正規表示式之範例。

解釋：例如在範例 1 為美國專利 6,517,425 專利申請範圍 1 中，系統可以根據正規表示式擷取出研磨墊元件，其功能性描述為“研磨一個表面”(polishing a surface)。第 13R 圖繪示根據正規表示式擷取出 polishing pad 元件之示意圖。

7. 從屬 (Contain) :

此分類項主要的目的，就是用來擷取專利申請範圍中，元件之間的從屬關係(part-of relation)，並且運用此關係將相關的兩個元件關聯起來，組合成一個三元關係。三元關係的形式定義為：Contain (Component(x), ContainVerb(m), Component(y))。專利中常用的從屬關係主要有五種，依序為“comprising”、“consisting of”、“essentially consisting of”、“including”、“having”。

第 13S 圖繪示此類(contain)之正規表示式類別的執行順序。第 13T 圖繪示此類(contain)之正規表示式之範例。解釋：例如在範例 1 為美國專利 6,517,425，專利申請範圍 1 中，系統可以根據正規表示式擷取出兩個三元關係：

1. Contain (polishing pad, comprising, lower resilient

portion)

2. Contain (polishing pad, comprising, upper polishing portion)

第 13U 圖繪示此類(contain)關係的示意圖。第 13V 圖繪示根據正規表示式擷取出 polishing pad 元件之示意圖。

8. 空間關係類 (Spatial)：

此分類項主要的目的，就是用來擷取專利申請範圍中，元件之間的空間關係(spatial relation)，並且運用此關係將相關的兩個元件關聯起來，組合成一個三元關係。三元關係的形式定義為：Spatial (Component(x), SpatialTerm(m), Component(y))。其中，具有空間關係的詞主要有介係詞和動詞兩種。在介係詞方面所擷取的詞有：“in”、“on”、“at”、“onto”、“opposite”、“surrounding”。動詞方面有：“position”、“bond”、“attach”、“coplanar”、“reflect”、“isolate”、“interpose”、“adhere”、“form”。

第 13W 圖繪示此類(spatial relation)之正規表示式類別的執行順序。第 13X 圖繪示此類(spatial relation)之正規表示式。解釋：例如在範例 1 為美國專利 6,273,800 專利申請範圍 1 中，系統可以根據正規表示式擷取出兩個三元關係：

1. Spatial (second surface, opposite, first surface)
2. Spatial (platen, attached, second surface of the support pad)

第 13Y 圖繪示此類(spatial relation)關係的示意圖。第 13Z

圖繪示根據正規表示式擷取出 polishing pad 元件之示意圖。

經過了以上八類正規表示式的資訊擷取，專利申請範圍的半結構化資料便可以被經由系統轉換成為結構化的資訊，並且以 XML 和 OWL 的格式來呈現。

以下舉了一個完整的範例來探討專利申請範圍語意結構的擷取過程。

專利申請範圍語意結構擷取範例：

當經過了語意/語法註記之後，專利申請範圍的每個字詞都已經保有語意/語法資訊，接著要進行專利申請範圍的結構擷取(如第 2 圖之步驟 218、220、222)。第 14 圖繪示專利申請範圍之元件結構圖之一範例。所謂擷取專利申請範圍的結構的意思，即我們將運用正規表示式，將專利申請範圍內所提到的元件、元件之間的關係、以及元件所具有的屬性自動擷取出來，並且表現成圖像化的關係架構(如第 14 圖)。在本實施例中稱此架構圖為語意結構圖(semantic graph)。專利申請範圍分為獨立項(independent claim)與附屬項(dependent claim)，對於這兩種型態的專利申請範圍之間的依存關係，系統也將會自動的進行參考連結。一張語意結構圖由一獨立項與其附屬項構成，若是一篇專利有多項獨立項，系統也將會自動建立多張語意結構圖。由於完整的語意結構圖像較為龐大，為了方便說明，在本實施例中將只用美國專利案號 6,524,176 的第一項獨立項來做範例說明。

第 15 圖繪示美國專利案號 6524176 的一個申請專利範圍。以美國專利 6,524,176 為例，第 15 圖為此專利的第

一條專利申請範圍，此項專利申請範圍為獨立項。其內容說明的是化學機械研磨中的研磨墊(polishing pad)結構，研磨墊含有(comprising)第一層(first layer)、第二層(second layer)、孔(hole)三個元件；而且這個孔元件又包含(having)第一區(first section)與第二區(second section)；而有一個塞子元件(plug)嵌在(embedded in)這個孔裡面，塞子包含上半部(upper portion)與下半部(lower portion)；其中塞子的上半部安插在(fit into)孔的第一區，塞子的下半部安插在孔的第二區。。第 16 圖繪示塞子和 孔的結構圖之一範例。第 17 圖繪示研磨墊的兩個層與實際的顯微照片對照圖。

藉由正規表示式，電腦將可以一步一步的對專利申請範圍進行剖析。首先截取出專利申請範圍中的元件(由正規表示式:元件類來達成)，例如範例中的研磨墊、孔、第一層、第二層、第一區、第二區、塞子、上半部、下半部。接著系統會建立元件之間的參考關係(由正規表示式參考類來達成)。例如在專利申請範圍的撰寫中，如果元件是第一次被描述，就會在元件的前面加上冠詞 a 或是 an，之後的描述，不管是同一項申請範圍中，或是在他項中描述，皆會具體的指明其原始出處是在哪裡，並且會在元件前面加上 the 或是 said，以利消除文件語意的歧異性。建立完參考關係後，系統將擷取各元件在專利申請範圍內所描述的屬性及其屬性值(由正規表示式屬性類來達成)。屬性將會記錄屬性名稱、屬性值、屬性的單位。另外專利申請範圍若有元件的功能性描述，系統也會將之結取記錄下來(由正規表示式功能類來達成)。最後，系統再把元件之間的關係擷取出

來，並且自動在元件之間建立關聯。此次所擷取的關係，包含有空間關係的詞(由正規表示式空間關係類來達成)，例如範例中的嵌、安插...等詞，以及從屬關係的詞(由正規表示式從屬關係類來達成)，如範例中的包含、含有。

在專利申請範圍之語意結構圖中，一對元件的關係稱為三元關係(triple)，三元關係中以兩個元件和它們之間的關係為基本單位。第 18 圖繪示此專利申請範圍之語意結構圖。由第 18 圖可知，語意結構圖是由許多的三元結構所構成。

利用正規表示式擷取所得的資訊，系統會自動的將之轉換成為 XML 與 OWL 機讀式檔案(如第 2 圖中步驟 218)。

由於系統含有專業詞彙詞庫，而且會將詞庫的階層式架構轉換成為本體知識，所以如果元件在語意註記的階段，具有專業詞彙詞庫的註記，便可以知道這各元件是屬於類別(class)或是某一類別的實體(instance)。對於沒有專業詞彙詞庫註記的元件，系統統一將之歸屬為 Component 這各類別的實體。另外，元件之間的關係，也將有一個規範。

圖形化呈現專利文件語意結構

當系統藉由正規表示式取得了語意資訊後，雖然以 OWL 的格式來表達，但是對於人類來說，機讀式的檔案還是很難立刻知道閱讀的。第 18 圖繪示專利申請範圍之元件結構圖之範例。藉由正規表示法與字詞單元(Tokens)的定義，電腦將可以一步一步的以元件為核心，利用具有空間

關係的字詞，把專利申請範圍所提到的元件關係的結構擷取出來，並且把元件的屬性也都擷取出來。在專利申請範圍之元件結構圖中，我們將整個結構圖稱為 (Structure Graph)，一對元件的關係稱為三元關係，三元關係中以元件 (Component) 為單位，每一個元件記錄著專利申請範圍提到的屬性。因此，若是將 OWL 的檔案轉換成為圖形化的表達方式，使用者就可以由圖立刻得知 claim 的內容為何，更可藉由圖形化的介面，立刻對照語意結構圖與專利申請範圍的文字，以便立即掌握專利的關鍵資訊。當使用者發現正規表示式擷取有誤時，也可以運用圖形化介面，直接更新語意結構圖，系統將直接修正 OWL 檔案，回存到資料庫中。

本發明至少具有下列優點，其中每一實施例可以具有一個或多個優點。本發明的專利文件語意結構建立方法可以將專利文件的專利申請範圍 (Claims) 做自動化分析與結構擷取。本發明的專利文件語意結構建立方法可以幫助知識之擷取與檢索，提供更精準的專業資訊。

雖然本發明已以一較佳實施例揭露如上，然其並非用以限定本發明，任何熟習此技藝者，在不脫離本發明之精神和範圍內，當可作各種之更動與潤飾，因此本發明之保護範圍當視後附之申請專利範圍所界定者為準。

【圖式簡單說明】

為讓本發明之上述和其他目的、特徵、優點與實施例

能更明顯易懂，所附圖式之詳細說明如下：

第 1 圖繪示一化學機械研磨的基本架構；

第 2 圖繪示本發明之專利文件語意結構建立方法之系統架構圖；

第 3 圖繪示專利申請範圍維護工具的一個例子；

第 4A 圖繪示詞彙辭庫(thesaurus)之編碼原則之一範例；

第 4B 圖繪示專業詞彙詞庫建構流程之一範例；

第 5 圖繪示詞彙詞庫編輯工具之一範例；

第 6 圖繪示一晶圓和一研磨墊的關係；

第 7 圖繪示第 6 圖之晶圓和研磨墊之間的三元關係；

第 8 圖繪示語意/語法加註流程圖的一個範例；

第 9 圖繪示 JavaNLP 所產生的一範例剖析樹(parsing tree)；

第 10 圖繪示語意/語法加註流程圖的一個範例；

第 11 圖繪示一正規表示式的超字元(meta-characters)功能；

第 12 圖繪示用來擷取專利語意結構的八類正規表示式；

第 13A 圖繪示一般類之正規表示式以及其解釋；

第 13B 圖繪示專利申請範圍類之正規表示式；

第 13C 圖繪示一些固定的寫法和開頭的範例；

第 13D 圖繪示用來擷取元件的來源；

第 13E 圖繪示元件類之正規表示式類別的執行順序；

第 13F 圖繪示 component(x)之正規表示式式；

第 13G 圖繪示在美國專利 6,273,800 找出的元件之範例；

第 13H 圖繪示參考類正規表示式類別的執行順序；

第 13I 圖繪示 參考類之正規表示式；

第 13J 圖繪示在美國專利 6,273,800 找出的元件之範例；

第 13K 圖繪示正規表示法的定義範例；

第 13L 圖繪示化學機械研磨之參數範例；

第 13M 圖繪示屬性類之正規表示式類別的執行順序；

第 13N 圖繪示屬性類之正規表示式之範例；

第 13O 圖繪示由美國專利 6,454,634 專利申請範圍 19 中擷取屬性得到的表示式；

第 13P 圖繪示功能描述類之正規表示式類別的執行順序；

第 13Q 圖繪示功能描述類之正規表示式之範例；

第 13R 圖繪示根據正規表示式擷取出研磨墊元件之示意圖；

第 13S 圖繪示從屬關係類之正規表示式類別的執行順序；

第 13T 圖繪示從屬關係類之正規表示式之範例；

第 13U 圖繪示從屬關係類關係的示意圖；

第 13V 圖繪示根據正規表示式擷取出研磨墊元件之示意圖；

第 13W 圖繪示空間關係類之正規表示式類別的執行順

序；

第 13X 圖繪示空間關係類之正規表示式；

第 13Y 圖繪示此類(spatial relation)關係的示意圖；

第 13Z 圖繪示根據正規表示式擷取出研磨墊元件之示意圖；

第 14 圖繪示專利申請範圍之元件結構圖之一範例；

第 15 圖繪示美國專利案號 6524176 的一個申請專利範圍；

第 16 圖繪示塞子和 孔 的結構圖之一範例；

第 17 圖繪示研磨墊的兩個層與實際的顯微照片對照圖； 以及

第 18 圖繪示此專利申請範圍之語意結構圖。

【主要元件符號說明】

102：研磨盤 104：研磨頭

106：晶圓

602：研磨墊 604：晶圓

605：研磨

五、中文發明摘要

專利文件語意結構之建立方法

一種專利文件語意結構之建立方法，包括下列步驟。建立一領域之一詞彙詞庫，包含此領域中之複數個專業詞彙。對一專利之一申請專利範圍進行一語意語法註記，以分辨此申請專利範圍中之專業詞彙、停用字、一般詞彙以及標點符號。利用定義之正規表示式，包括一般類、專利申請範圍類、元件類、參考類、屬性類、功能描述類、從屬關係類和空間關係類，來擷取該申請專利範圍之語意結構，以建立該申請專利範圍之一結構關係。此結構關係包括申請專利範圍中之專業詞彙、一般詞彙以及申請專利範圍中之專業詞彙互相間的三元關係。

六、英文發明摘要

Patent Document Content Construction Method

A patent document content construction method is described. The method includes the following steps. A domain-specific thesaurus including a plurality of domain-specific terms is constructed. A semantic/syntactic annotation is performed on a claim of a patent to identify domain-specific terms, stop words, general terms, and punctuations. A defined regular sets are used to classify the words in a claim to build a structural relation of the claim. The defined sets include Common, Claim, Component, Reference, Attribute, Functionality, Contain, Sapatial. The structural relation includes the domain-specific terms, the general terms, and the triple relations of the domain-specific terms in the claim.

十、申請專利範圍：

1. 一種專利文件語意結構建立方法，包含：

建立一領域之一詞彙詞庫，其中該詞彙詞庫包含該領域中之複數個專業詞彙，該些專業詞彙形成一階層式架構；

對一專利之一申請專利範圍進行一語意語法註記，以分辨該申請專利範圍中之專業詞彙、停用字、一般詞彙以及標點符號；以及

利用該詞彙詞庫建立該申請專利範圍之一結構關係 (structural relation)，該結構關係包含該申請專利範圍中之專業詞彙、一般詞彙以及該申請專利範圍中之專業詞彙互相間的三元關係。

2. 如申請專利範圍第 1 項所述之方法，更包含：

在該詞彙詞庫之階層式架構中，將相同類型的專業詞彙歸類為同一階層。

3. 如申請專利範圍第 1 項所述之方法，在進行該語意語法註記之前，先對該申請專利範圍進行詞性(POS)之語法註記。

4. 如申請專利範圍第 1 項所述之方法，在進行該語意語法註記之步驟中，更包含：

將該申請專利範圍之詞彙與該詞彙詞庫中的專業詞彙相比對，以決定該申請專利範圍之詞彙之語意。

5. 如申請專利範圍第 1 項所述之方法，其中該些停用字包含「a」以及「the」。

6. 如申請專利範圍第 1 項所述之方法，其中該申請專利範圍為一獨立項。

7. 如申請專利範圍第 1 項所述之方法，其中該申請專利範圍為一附屬項，且該方法一併對該附屬項以及該附屬項所附屬之獨立項進行該語意語法註記之步驟以及進行建立該結構關係之步驟。

8. 如申請專利範圍第 1 項所述之方法，更包含：以一結構圖(structure graph)顯示該申請專利範圍的該結構關係。

9. 如申請專利範圍第 8 項所述之方法，更包含：利用一正規表示法(regular expression)以及複數個字詞單元(tokens)的定義決定該結構圖。

10. 如申請專利範圍第 1 項所述之方法，更包含：利用一正規表示法(regular expression)對該申請專利範圍進行剖析(parsing)。

11. 如申請專利範圍第 10 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中之一元件(component)。

12. 如申請專利範圍第 10 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中元件間的參考連結(reference)。

13. 如申請專利範圍第 10 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中一元件的屬性(attribute)。

14. 如申請專利範圍第 10 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中一元件的功能性描述(functionality)。

15. 如申請專利範圍第 10 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中是否具有元件之間的從屬關係(part-of-relation)。

16. 如申請專利範圍第 10 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中是否具有元件之間的空間關係(spatial relation)。

17. 一種專利文件語意結構建立方法，包含：
對一專利之一申請專利範圍進行一語意語法註記，以

分辨該申請專利範圍中之專業詞彙、停用字、一般詞彙以及標點符號；以及

利用一詞彙詞庫建立該申請專利範圍之一結構關係 (structural relation)，該結構關係包含該申請專利範圍中之專業詞彙、一般詞彙以及該申請專利範圍中之專業詞彙互相間的三元關係。

● 18. 如申請專利範圍第 17 項所述之方法，其中該詞彙詞庫包含一領域中之複數個專業詞彙，該些專業詞彙形成一階層式架構。

● 19. 如申請專利範圍第 18 項所述之方法，更包含：在該詞彙詞庫之階層式架構中，將相同類型的專業詞彙歸類為同一階層。

● 20. 如申請專利範圍第 17 項所述之方法，在進行該語意語法註記之前，先對該申請專利範圍進行詞性(POS)之語法註記。

● 21. 如申請專利範圍第 17 項所述之方法，在進行該語意語法註記之步驟中，更包含：
將該申請專利範圍之詞彙與該詞彙詞庫中的專業詞彙相比對，以決定該申請專利範圍之詞彙之語意。

● 22. 如申請專利範圍第 17 項所述之方法，其中該些停

用字包含「a」以及「the」。

23. 如申請專利範圍第 17 項所述之方法，其中該申請專利範圍為一獨立項。

24. 如申請專利範圍第 17 項所述之方法，其中該申請專利範圍為一附屬項，且該方法一併對該附屬項以及該附屬項所附屬之獨立項進行該語意語法註記之步驟以及進行建立該結構關係之步驟。

25. 如申請專利範圍第 17 項所述之方法，更包含：以一結構圖(structure graph)顯示該申請專利範圍的該結構關係。

26. 如申請專利範圍第 25 項所述之方法，更包含：利用一正規表示法(regular expression)以及複數個字詞單元(tokens)的定義決定該結構圖。

27. 如申請專利範圍第 17 項所述之方法，更包含：利用一正規表示法(regular expression)對該申請專利範圍進行剖析(parsing)。

28. 如申請專利範圍第 27 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中之一元件(component)。

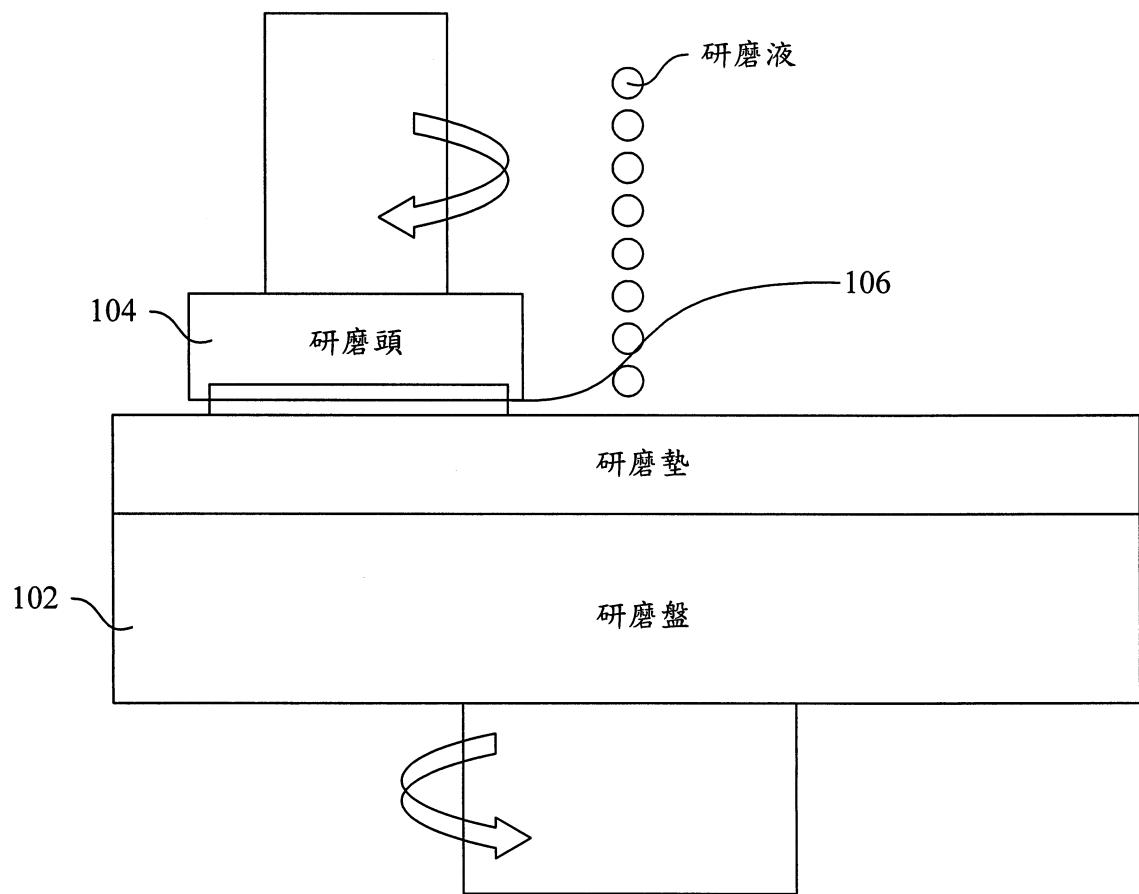
29. 如申請專利範圍第 27 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中元件間的參考連結 (reference)。

30. 如申請專利範圍第 27 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中一元件的屬性 (attribute)。

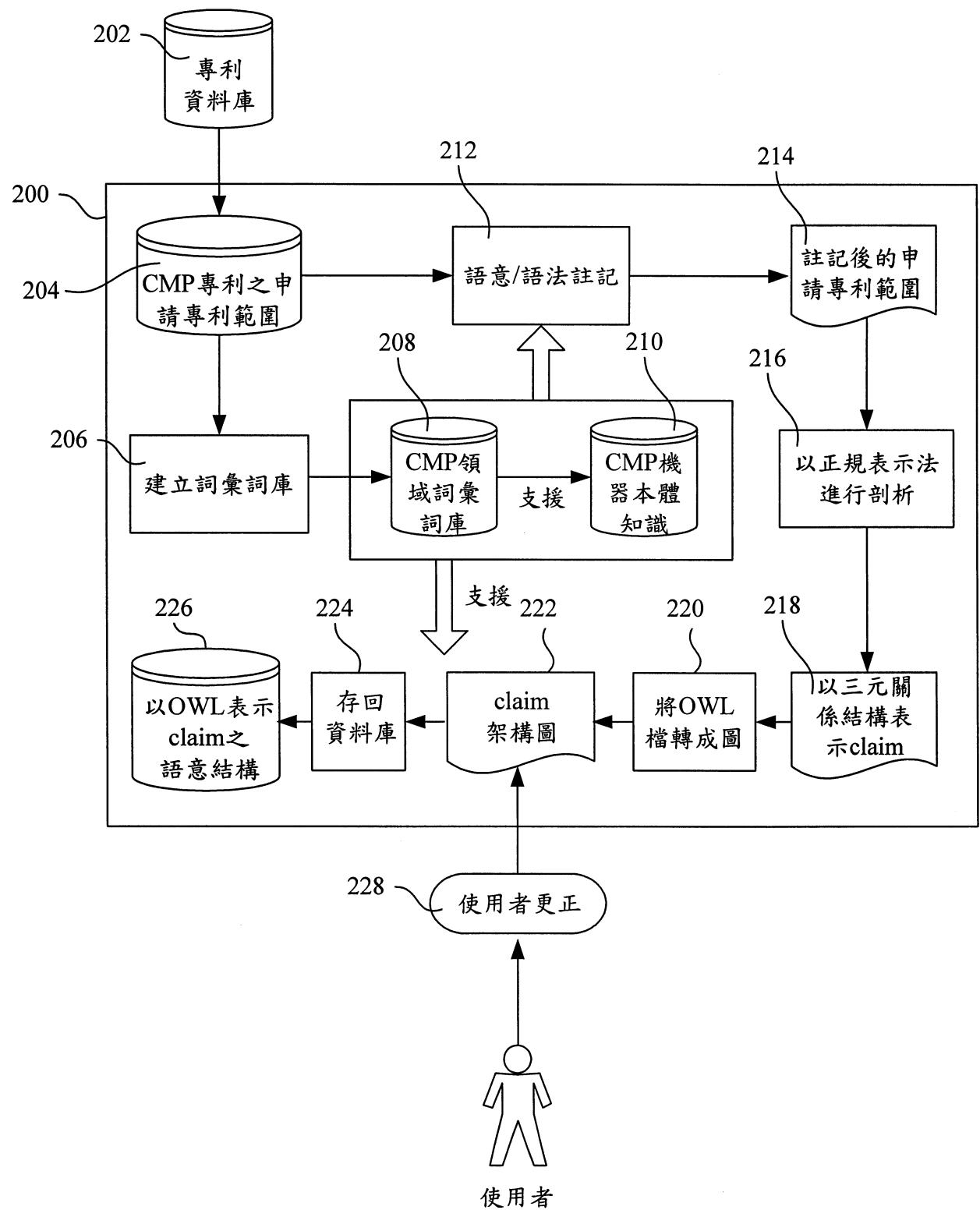
31. 如申請專利範圍第 27 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中一元件的功能性描述 (functionality)。

32. 如申請專利範圍第 27 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中是否具有元件之間的從屬關係 (part-of-relation)。

33. 如申請專利範圍第 27 項所述之方法，其中該正規表示法包含鑑別該申請專利範圍中是否具有元件之間的空間關係 (spatial relation)。

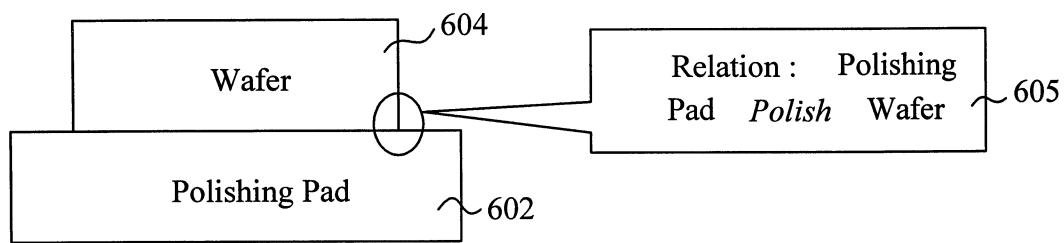


第 1 圖

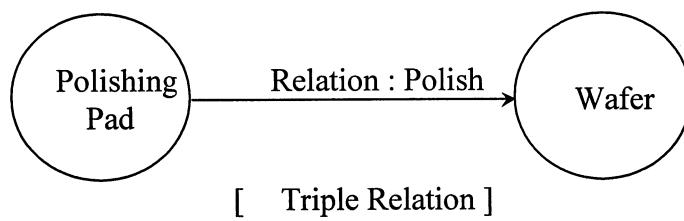


第 2 圖

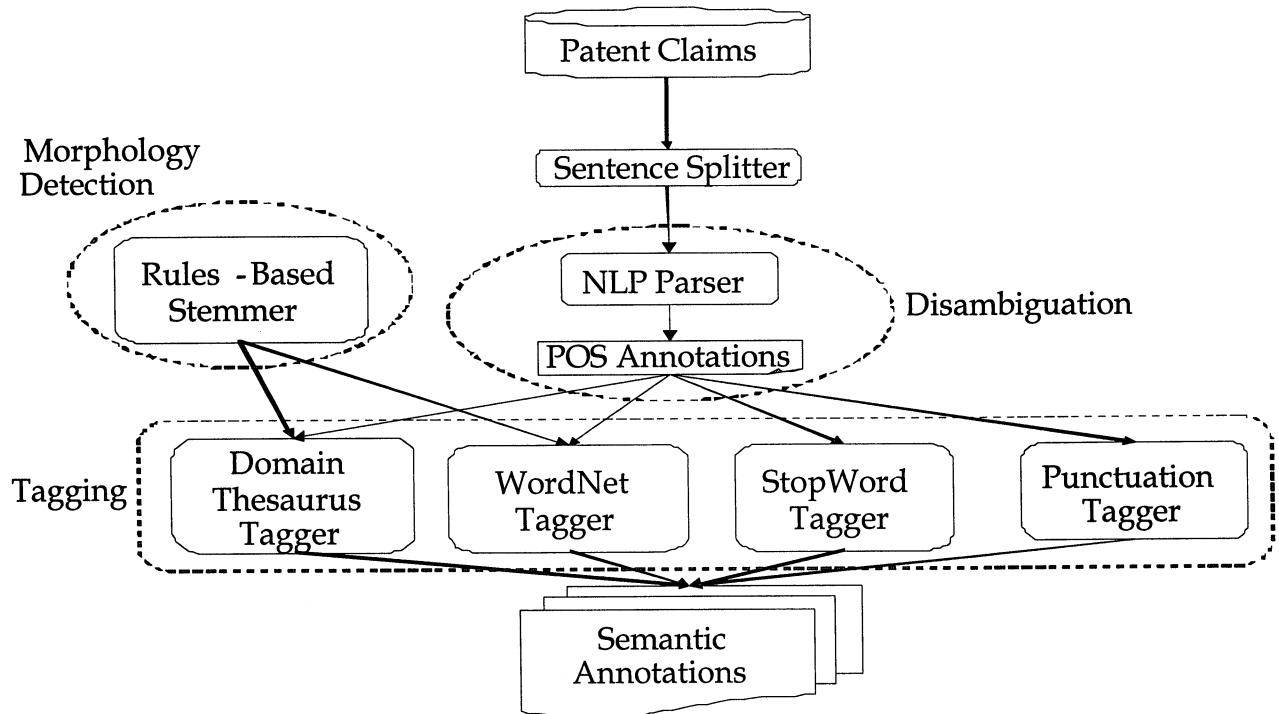
I267756



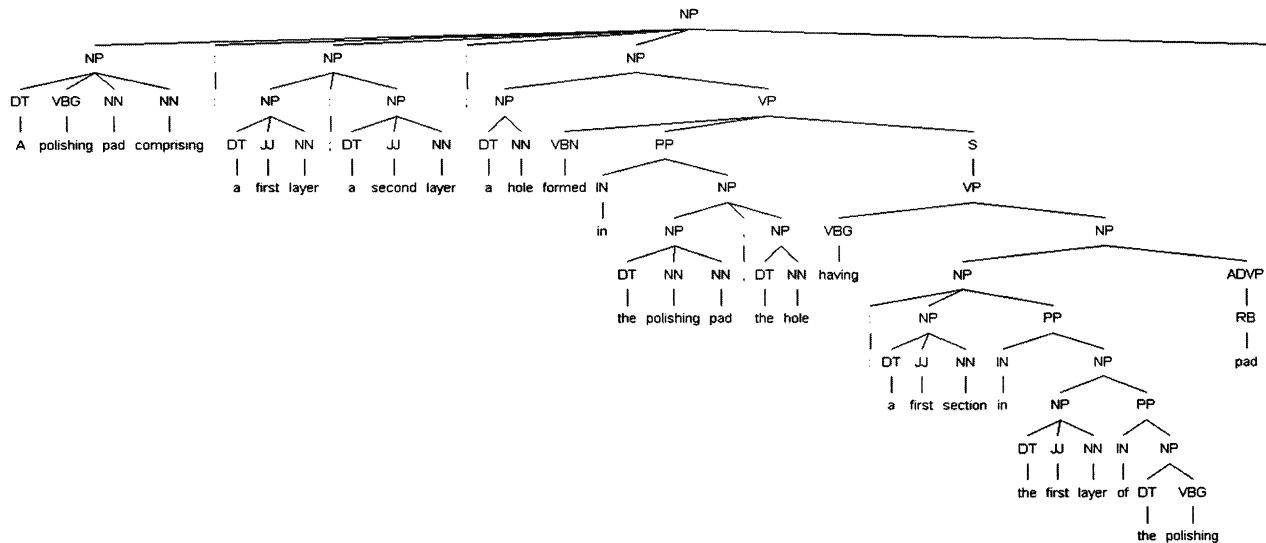
第 6 圖



第 7 圖



第 8 圖



第 9 圖

A semiconductor wafer polisher comprising: a rotatable wafer carrier having a wafer receiving surface for releasably retaining a semiconductor wafer;



```

<stopword _article id= '001">A </stopword_article> <mechanical
id= '1322"> semiconductor wafer polisher </mechanical>
<wordnet _verb id= "1796878"> comprising </wordnet_verb>
<punctuation id= '010300">: </punctuation> <stopword _article
id= '001">a </stopword_article> <mechanical id= '1342"> rotatable
wafer carrier </mechanical> <wordnet _verb
id= "1876679"> having </wordnet_verb> <stopword _article
id= '001">a </stopword_article> <mechanical id= '1354"> wafer
receiving surface </mechanical> <stopword _preposition
id= '048">for </stopword_preposition> <pos_RB> releasably </pos_RB>
<wordnet _verb id= "1842985"> retaining </wordnet_verb>
<stopword _article id= '001">a </stopword_article> <mechanical
id= '1342"> semiconductor wafer </mechanical><punctuation
id= '010100">; </punctuation>

```

第 10 圖

		token	explanation
1	quantifiers	*	X , zero or more times
2		+	X , one or more times
3		?	X , once or not at all
4		{ n,m}	X , at least n but not more than m times
5		{ n,}	X , at least n times
6		{ n}	X , exactly n times
7	Logical operators		Either X or Y
8		^	not X
9		()	X, as a capturing group
10		[]	Just one char in the []
11		-	region
12		.	Any character

第 11 圖

	類別	功能描述
1.	{ Common }	一些共通用的法則，供其他的正規表示式使用。 例如：“space”（表示文章中的空格）
2.	{ Claim }	用來判別專利聲請範圍的種類。例如：獨立項，附屬項...等
3.	{ Component }	用來擷取專利聲請範圍中所描述的元件與專業領域詞彙。例如：“first layer”，“polishing pad”
4.	{ Reference }	用來建立索引表，內容為專利聲請範圍的獨立項和附屬項之間的交互參考關係，以及元件之間交互索引的關連。例如：“The polishing pad of claim 1...”
5.	{ Attribute }	用來擷取專利聲請範圍中元件的屬性及其屬性值。 例如：“a wavelength of from about 190 to about 3500 nanometers”
6.	{ Functionality }	用來擷取專利聲請範圍中元件的功能性描述。例如：“a pad for polishing a surface...”
7.	{ Contain }	用來擷取專利聲請範圍中元件之間的從屬關係。例如：“A polishing pad comprising a hole..”
8.	{ Spatial }	用來擷取專利聲請範圍中元件之間的空間關係。 Ex. the hole is embedded in the polishing pad ...

正規表示式(局部):

```
space = "(" + "\s" + idx + ")";
redundancy = "(.*)?"
```

解釋:

1. 一個空白字元
2. 任意長度的字元

第 13A 圖

Claim(x)

正規表示式(局部):

1. regexClaimSplit = space + "[0-9]*\\." + space + "[A-Z]"
2. regexClaimType = "(of|to|with|in)?"+ space + "claim" + space + "[0-9]* + "(,"+ space + "wherein)?"
3. regexContentType = "(A|An|The)" + space + "method" + space + "(of|for)?"

第 13B 圖

1. [][數字]*[.][][一個大寫字元]

(範例1) 1. A polishing pad for chemical mechanical polishing having, within a surface of the pad, at least.....

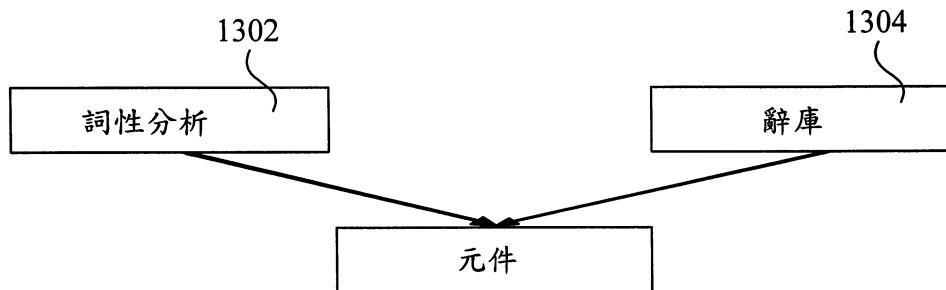
2. [(of|to|with|in)]?[][claim][][數字]*([,][][wherein])?

(範例2) 2. The polishing pad, as set forth in **claim 1**, wherein the...

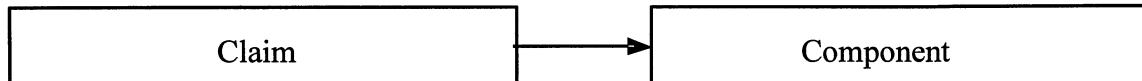
3. [(A|An|The)][][methid][][(of|for)]?

(範例3) 3. A **method** of joining a first section of...

第 13C 圖



第 13D 圖



第 13E 圖

Component(x)

正規表示式(局部):

```

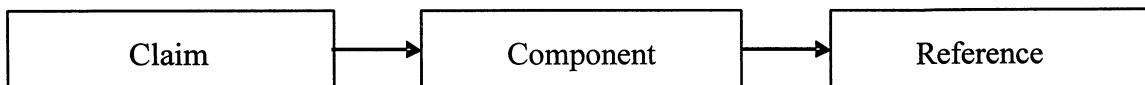
determin = "DT" + space
regexComponent1 = "(" + determin + "JJ" + space + "NN" + ")"|" +
"(" + determin + "VBG" + space + "NN" + ")"|" +
"(" + determin + "NN" + ")"|" +
"(" + determin + "JJ" + ")"...(略);
regexComponent2 = 由辭庫構成.
  
```

第 13F 圖

1.“ (a | an | the | said) (JJ NN)”or
“ (a | an | the | said) (V рG NN)”or
“ (a | an | the | said) (NN)”or
“ (a | an | the | said) (JJ)”or ...

(範例1) 1. An apparatus for supporting a polishing pad during planarization of a microelectronic substrate, comprising: an elongated at least partially compressible support pad having a first surface for engaging the polishing pad and a second surface opposite the first surface; and a platen attached to the second surface of the support pad, the platen being generally incompressible in a direction generally perpendicular to the second surface of the support pad, at least a portion of the platen and the support pad being movable with the polishing pad when the first surface of the support pad engages the polishing pad.

第 13G 圖



第 13H 圖

Reference

正規表示式:

```

regexReference_1 = "(" + "(The|the|said)" + space + regComponent + ")"
regexReference_2 = "(" + "(claim)" + space + "[0-9]+")"
  
```

第 13I 圖

1. (a|an) “component” ... (the|said) “component”

(範例1) 1. **An apparatus(Component_Token_0)** for supporting a **polishing**

pad(Component_Token_1) during planarization of a microelectronic

substrate(Component_Token_2), comprising:

an elongated(Component_Token_3) at least partially compressible support

pad(Component_Token_4) having a first surface(Component_Token_5) for engaging the

polishing pad(Component_Token_6) and a second surface(Component_Token_7)

opposite the first surface(Component_Token_8); ...(略)

2. the “component” of claim 1 ... (the|said) “component”;

(範例2) 2. **The apparatus(Component_Token_23) of claim 1** wherein the support

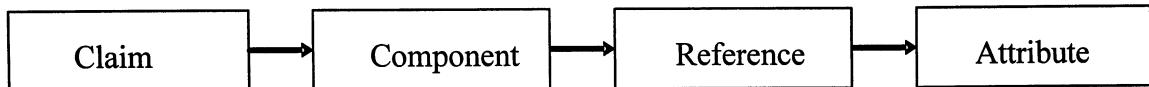
pad(Component_Token_24) and the platen(Component_Token_25) together form a

continuous loop(Component_Token_26) with the first surface(Component_Token_27) of
the support pad(Component_Token_28) facing outward.

第 13J 圖

項目	分類參數
一般參數	物理參數如：時間、速度、溫度、濃度、重量、面積、體積、力(強度)、應力、壓力、流量...等等。
CMP監控參數例	研磨粒子種類、研磨粒子大小、研磨粒子電化學特性、研磨液之介電常數、研磨液濃度、研磨液酸鹼值、晶圓轉速、轉盤速度、研磨時間、晶圓重量、晶圓面積、研磨墊特性。

第 13L 圖



第 13M 圖

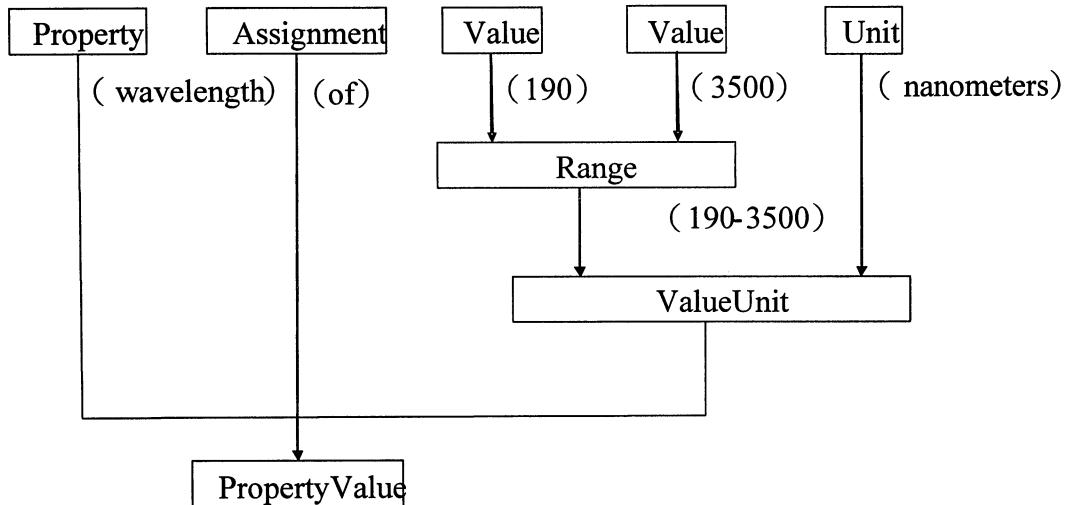
Attribute

正規表示式:

```

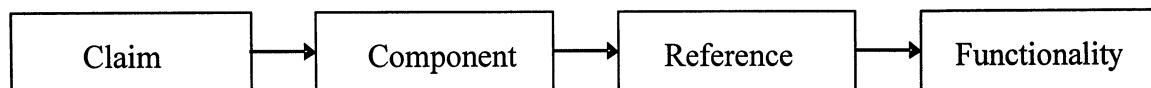
1. regexProperty = 從資料庫讀取
2. regexAssignment = 從資料庫讀取
3. regexValue = "(\d)+"+idx;
4. regexRange = "(from"+idx+"")+"(.*)?"+regValue+"(.*)?"+"(to|\-
)"+"(.*)?"+regValue;
5. regexUnit = 從資料庫讀取
6. regexValueUnit_1 = "(" + "("+regRange+")"+ " (" + space + ")?" +
" (" + regUnit+ "(" + space+ ")?" +"")" + ")"
    regexValueUnit_2="(" + "("+regValue+")" + "(" + space + ")?" +
" (" + regUnit+ "(" + space+ ")?" +"")" + ")";
7. regexPropertyValue = "("+regProperty + "(" + space + ")?" + ")" +
" (" + regAssignment+ "(" + space + ")?" + ")" + "
" (" + regValueUnit + ")" + "(" + space + ")?" ;
  
```

第 13N 圖



PropertyValue (Property (wavelength) ,
 Assignment (of) ,
 ValueUnit (Range (Value (190),Value (3500)) ,, Unit (nanometers)))

第 13O 圖



第 13P 圖

正規表示式(局部):

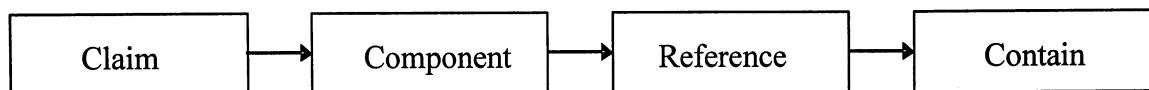
```
regexFunctionality = regComponent + space + "(for)?(.*)?"
```

第 13Q 圖

1. ... “component” (for ...) comprising|comprises:

(範例1)1. A polishing pad for polishing a surface, comprising:
 a first member defining a first polishing surface... (略)

第 13R 圖



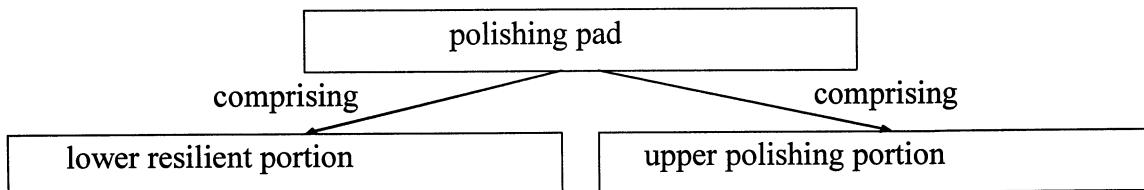
第 13S 圖

Comprising

正規表示式(局部):

```
regexComprising_1 = regComponent + space + "(for)?(.*)" +
"(comprising|comprises)" + ":" + space + "(" + "(*?)" + "(a|an|the)" + space +
regComponent + "(*?)" + "(;:)" + space + ")*" + "(" + "(*?)" + "(a|an|the)" +
space + regComponent + "(*?)" + ";" + space + "(and)" + space + ")*" + "(*?)" +
"(a|an|the)" + space + regComponent
```

第 13T 圖



第 13U 圖

1. ... “component” (for ...) comprising|comprises:

(a|an|the) “component” ...;

.....;

(a|an|the) “component” ...; and

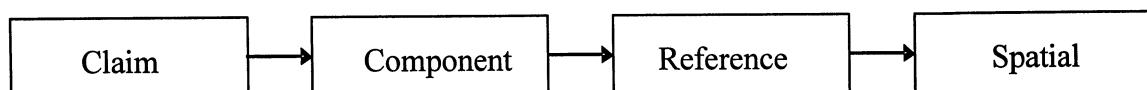
(a|an|the) “component”

(範例1)1. A **polishing pad** for polishing a surface, **comprising**:

a **first member** defining a first polishing surface and having a structurally degradable abrasive first material; and

a **surface abrasion** impeding second member defining a second polishing surface...(略)

第 13V 圖



第 13W 圖

Spatial

正規表示式(局部):

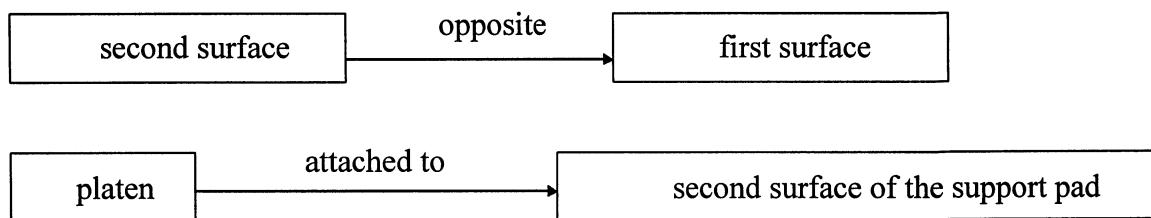
```

regexSpatial_1= "( regComponent + space +
"(in|on|at|onto|opposite|surrounding)" + space + regComponent )"

regexSpatial_2= (regComponent + space + (is|are|being) + space +
(positioned|bonds|bonded|bonding|attach|attached|attaches|attachable|
coplanar|reflected|reflect|isolated|interposed|interpose|adhere|adhered|
form|formed) + space + (to|between|in|on|opposite|with|by|from) +
space + regComponent)

```

第 13X 圖



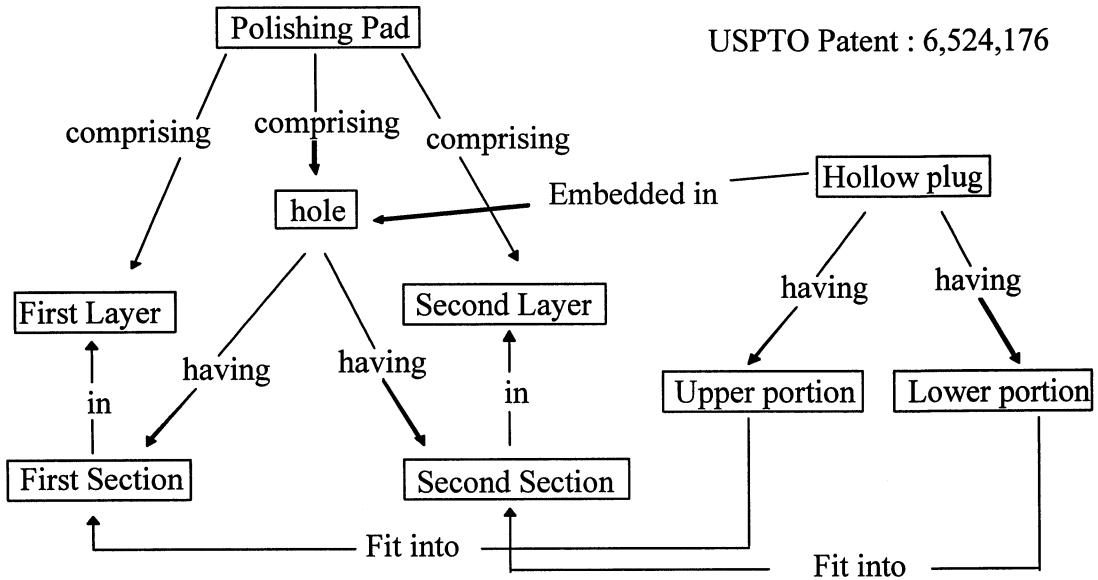
第 13Y 圖

... “component” ... (is or being) ... in (or other spatial term.) ... “component” ;

... “component” ... (is or being) ... attached (or other spatial term.) ... to ... “component” ;

(範例1)1. An apparatus for supporting a polishing pad during planarization of a microelectronic substrate, comprising: an elongated at least partially compressible support pad having a first surface for engaging the polishing pad and a **second surface opposite the first surface**; and a **platen attached to the second surface of the support pad**, ... (略)

第 13Z 圖



第 14 圖

What is claimed is:

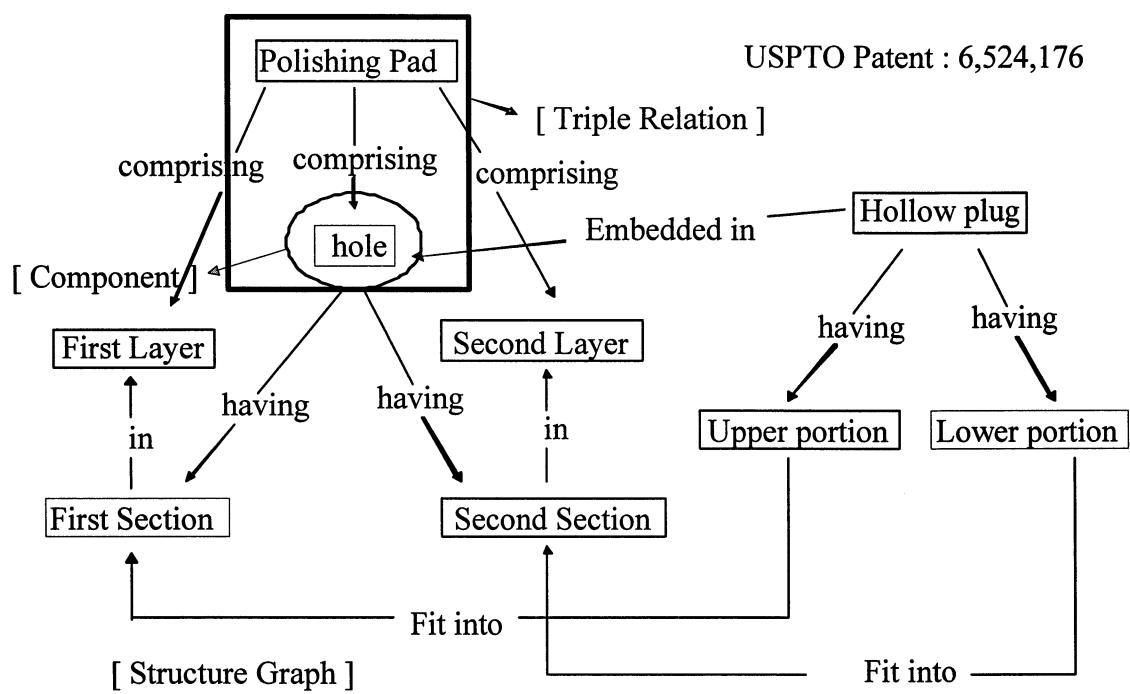
1. A polishing pad comprising:
 - a first layer;
 - a second layer;
 - a hole formed in the polishing pad, the hole having:
 - a first section in the first layer of the polishing pad; and
 - a second section in the second layer of the polishing pad; and
 - a hollow plug embedded in the hole;

wherein the hollow plug has an upper portion and a lower portion , and the upper portion of the hollow plug fits into the first section of the hole, and the lower portion of the hollow plug fits into the second section of the hole.

USPTO patent : 6524176 , Claim 1

第 15 圖

I267756



第 18 圖

七、指定代表圖：

(一)、本案指定代表圖為：第(2)圖

(二)、本案代表圖之元件符號簡單說明：

八、本案若有化學式時，請揭示最能顯示發明
特徵的化學式：