

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 461 562**

51 Int. Cl.:

**G06F 17/30** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **31.01.2006 E 06709460 (7)**

97 Fecha y número de publicación de la concesión europea: **16.04.2014 EP 1856633**

54 Título: **Procedimiento para el tratamiento de datos en vistas a la extracción de los atributos principales de un producto**

30 Prioridad:

**07.03.2005 FR 0550596**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**20.05.2014**

73 Titular/es:

**SIGHT'UP (100.0%)  
104 ROUTE DU RAPOUX  
38690 TORCHEFELON, FR**

72 Inventor/es:

**BIGALET, FRANCK**

74 Agente/Representante:

**ISERN JARA, Jorge**

**ES 2 461 562 T3**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

**DESCRIPCIÓN**

Procedimiento para el tratamiento de datos en vistas a la extracción de los atributos principales de un producto

5 Sector técnico

La presente invención, se refiere a un procedimiento para el tratamiento de datos representativos de un producto, en vistas a su acceso mediante motores de búsqueda (buscadores), con una pertinencia acrecentada.

10 La invención, encuentra aplicación, de una forma particular, aunque no de una forma limitativa, en el tratamiento de datos representativos de los productos puestos a la venta, mediante sitios comerciales, en la red de Internet. Mediante "producto", se entenderá, por supuesto, no únicamente los productos fabricados, sino, igualmente, los servicios que puedan ser objeto de una oferta comercial y en donde, la compra, pueda igualmente realizarse por vía telemática.

15 En la parte que sigue de la descripción, la invención, se describirá en relación con los ejemplos ligados a esta aplicación específica, sin encontrarse no obstante limitada.

20 Técnicas correspondientes al arte anterior

El documento "Extracting salient keywords in a document that belong to a specific context", - "Palabras clave destacadas en un documento que pertenece a un contexto específico" -, RESEARCH DISCLOSURE, KENNET MASON PUBLICATIONS WESTBOURNE, GB, volumen 465, nº 13, Enero del 2003 (2003 -01), describe un procedimiento para extraer palabras clave principales de un documento, utilizando glosarios específicos para contextos diferentes.

30 Hoy en día, los sitios comerciales que proponen la venta de productos y de servicios, en la red de Internet, encuentran un interés en cuanto al hecho de relevar sus ofertas, mediante motores de búsqueda especializados, puestos a disposición mediante sitios de referencia. Estos motores de búsqueda o buscadores, especializados, proponen, al comprador, diferentes categorías de productos y servicios, y permiten, al comprador potencial, el efectuar una búsqueda, a partir de las palabras clave correspondientes a las características del producto.

35 Con objeto de realizar este cometido, el sitio de referencia, recolecta los datos representativos de los productos comercializados por los diferentes sitios comerciales, y determina las características principales que podrán ser objeto de búsqueda mediante la palabra clave.

40 Se concibe el hecho de que, este tratamiento, es una operación relativamente compleja, debido a que, ésta, depende eminentemente de la lengua empleada para la descripción del producto, del vocabulario relativo a una categoría de artículos, así como de la sintaxis empleada.

45 Así, de este modo, hoy en día, la extracción de los atributos o características de un producto, se realizan mediante un análisis manual y sistemático del conjunto de ofertas. Este tratamiento, permite caracterizar cada producto mediante los atributos, permitiendo la clasificación de estos productos, en diferentes categorías. Los diferentes productos, se catalogan, a continuación, en las bases de datos que permiten su acceso mediante la selección de los diferentes campos de la base de datos.

50 De una forma más precisa, algunos sitios comerciales que explotan bases de datos particulares, permiten afinar la elección del comprador, autorizando la elección de los criterios opcionales dependientes de los criterios ya recolectados.

55 Se concibe el hecho de que, estas técnicas de extracción de las características, son particularmente fastidiosos, puesto que, éstas, necesitan un análisis individual y manual de cada uno de los productos cuya oferta se expone, por parte del sitio comercial. Adicionalmente, además, la clasificación de los diferentes productos, en las bases de datos, no es técnicamente realmente satisfactoria. En efecto, la estructuración de las ofertas en las bases de datos, se adapta, de una forma relativa, a los productos de carácter técnico, como por ejemplo, en el sector informático, debido al hecho de que, cada producto presenta unas características, en un número limitado, las cuales pueden declinarse o no reconocerse, en una elección relativamente limitada.

60 Como contraste de ello, esta clasificación en bases de datos, no es susceptible de poderse adaptar a los productos que pertenezcan a sectores en donde, la oferta, sea extremadamente variada y difusa. Se puede citar, de una forma particular y a título de ejemplo, el sector del vestir. La estructuración en bases de datos que estereotipan, de algún modo, las características potenciales de un producto, no convierten a este mecanismo en adaptable a los mercados en gran evolución.

Por otra parte, se ha constatado el hecho de que, esta forma de extracción de las características de un producto, presenta deficiencias, en términos de pertinencia.

5 Así, de este modo, en la siguiente descripción: "couverture 75 x 100, en laine polaire uni, livrée dans un sac PVC 100% poliéster", - "manta de 75 x 100 en lana polar uniforme (de un solo color), suministrada en un saco de PVC 100% de poliéster" -, los términos "cobertura" y "saco", corresponden, ambos, a una definición potencial del tipo de artículos. El producto de esta forma descrito, se considera, por lo tanto, mediante las técnicas actuales, como siendo, a la vez, una "manta", o un "saco". Este resultado, no es por supuesto pertinente, ya que, el artículo de este modo descrito, es una manta y no un saco, el cual constituye un accesorio de la venta. Se obtiene el mismo tipo de razonamiento, mediante la descripción de artículos tales como "vestido negro con un cinturón rojo", en la cual, los procedimientos actuales, determinan como palabras clave, los términos "cinturón" y "rojo", los cuales corresponden a características de un accesorio del artículo principal, la cual es un "vestido negro".

15 Dicho de otro modo, las técnicas existentes, presentan una deficiencia en términos de pertinencia.

Un objeto de la invención, es el de permitir la mejora de la pertinencia de indexación, procediendo a tratar, de la forma más apropiada, los casos de las figuras en donde, la descripción de un producto, comprende varias palabras, las cuales pertenecen a un mismo tipo de atributo, y en el cual, únicamente una de estas palabras, es representativa de una característica importante del producto.

20 Otro objeto de la invención, es la de proporcionar un procedimiento que pueda adaptarse de una forma sencilla a las diferentes lenguas que conciernen a la descripción de los productos.

#### Exposición de la invención

25 La presente invención, se refiere, por lo tanto, a un procedimiento para la determinación de las características principales de un producto definido por una descripción que combina una pluralidad de palabras.

Este procedimiento, combina una pluralidad de etapas sucesivas.

30 En primer lugar, para cada palabra de la descripción, se determina el hecho de si, ésta, pertenece a un glosario predeterminado. Adicionalmente, además, para cada una de estas palabras que pertenecen a un glosario, se efectúan las siguientes operaciones.

35 En una primera etapa, se asignan, a esta palabra, una pluralidad de codificaciones. Estas codificaciones, se elaboran en función del glosario predeterminado, al cual pertenece la palabra considerada, así como, eventualmente y dado el caso, los glosarios a los cuales pertenecen las otras palabras de la descripción.

40 En una segunda etapa, se procede a analizar, mediante un categorizador bi-clase, el conjunto de estas codificaciones elaboradas para la palabra en cuestión, de forma que se determine el hecho de si la palabra considerada, es un atributo principal del producto.

45 Así, de este modo, la extracción de los atributos o características de un producto, se realiza mediante comparación de las palabras de su descripción, con los glosarios. Estos glosarios, son conjuntos de palabras relativos al tipo de características del producto. Cada categoría de artículos, posee, por lo tanto, unas características particulares. Así, de este modo, a título de ejemplo, en la categoría de los artículos de vestir, existen diferentes glosarios que reagrupan cada uno de los tipos de ropa, el color, la materia, u otras características específicas para un artículo de vestir.

50 El procedimiento puesto en ejecución mediante la presente invención, necesita el empleo de los glosarios, los cuales resultan de un inventario manual, los más exhaustivo que sea posible, de los términos relativos a la calidad del glosario considerado, en una muestra representativa lo más reducida posible.

55 De una forma global, el conjunto de las descripciones de los productos, se somete a una comparación con estos diferentes glosarios. En el caso de una concordancia de un término de la descripción, con una de las palabras de uno de los glosarios, el término de esta forma identificado, se considera, entonces, como definiendo un atributo potencial, correspondiente al glosario al cual éste pertenece.

60 Dicho de otro modo, la invención, consiste en proceder a efectuar un tratamiento, sobre cada descripción, el cual no se limita a la simple búsqueda en concordancia con los términos de un glosario, sino que, adicionalmente, además, éste efectúa un conjunto de operaciones destinadas a eliminar los casos no pertinentes de pertenencia a un glosario. De una forma más precisa, a cada uno de los términos susceptibles de constituir un atributo esencial del producto, puesto que éstos pertenecen a un glosario, se asocia un cierto número de codificaciones. Estas codificaciones, se elaboran en función de los otros términos de la descripción que pertenecen a glosarios. Así, de

este modo, es posible detectar los casos de figuras, en donde, la pertenencia a un glosario, no significa sistemáticamente, para una palabra, el que ésta represente una característica principal del producto.

A título de ejemplo, la presencia, en la descripción, de dos palabras susceptibles de representar un tipo de artículo, es un riesgo de error, en la determinación del atributo "tipo" de artículo. Así, de este modo, en la expresión "Vestido negro con un cinturón rojo", los términos "Vestido" y "cinturón" son, ambos, representativos del tipo de ropa, aunque únicamente el término "vestido", corresponda al tipo de artículo real en cuestión. La determinación de la posición, en el interior de la descripción de cada una de las palabras relativas a un tipo de artículo, puede permitir discriminar el caso de figuras, y afectar al atributo "tipo", el término que es efectivamente pertinente, y en la especie, la palabra "vestido".

El análisis de las diferentes codificaciones asignadas a una palabra determinada, permite así, de este modo, mediante técnicas de autoaprendizaje, el confirmar el hecho de que, la palabra en cuestión, se trata efectivamente de una característica del producto, para el tipo de glosario al cual ésta pertenece. Y a la inversa, éste permite eliminar los casos, en donde, ésta, corresponda a una característica de un accesorio del artículo, o bien, otra vez, a una característica auxiliar.

En la práctica, el tratamiento en conformidad con la invención, puede pasar, de una forma ventajosa, por una primera etapa consistente en sustituir, en la descripción, las palabras pretendientes a un glosario, por el nombre del citado glosario, de tal forma que se elaboren las codificaciones características, mediante el análisis de esta descripción, después de la sustitución.

En otros términos, el tratamiento en conformidad con la invención, puede realizarse sobre una descripción, en la cual, los términos pertenecientes a un glosario, se han reemplazado por el nombre del mismo glosario. En esta descripción sustituida, los términos característicos, ya no se encuentran presentes, y su pertinencia, se determina, por lo tanto, en función de la posición con relación a los otros términos de la descripción, los cuales, en sí mismos, puedan haberse sustituido.

Con objeto de atender al nivel más elevado de prestaciones, puede ser ventajoso el hecho de generar codificaciones suplementarias. Así, de este modo, para cada una de las palabras pertenecientes a un glosario determinado, se puede asignar, a la citada palabra, una pluralidad de codificaciones complementarias, las cuales se elaboran en función del glosario predeterminado, al cual pertenece la palabra considerada y, así como otras palabras de la descripción.

Estas codificaciones complementarias, se elaboran, por lo tanto, en función de las otras palabras de la descripción, y no únicamente en función de los glosarios a los cuales pertenecen estas palabras. Estas codificaciones complementarias, se efectúan, por lo tanto, teniendo en cuenta la descripción, antes de las operaciones de sustitución mencionadas anteriormente, arriba.

En la práctica, pueden emplearse múltiples codificaciones, más o menos complejas. Estas codificaciones, pueden combinarse entre ellas, según una lógica específica de la invención.

En efecto, el procedimiento, puede comportar una etapa de selección de un número limitado de codificaciones, elegidas de entre un grupo de codificaciones potenciales. En efecto, puede ser útil el proceder a no retener más que ciertas codificaciones particulares, los cuales permitan discriminar ciertos casos más pertinentes.

En la práctica, la selección de estas diferentes codificaciones, puede depender del glosario al cual pertenece la palabra para la cual se busca la adaptación de las codificaciones. En otros términos, se retendrán ciertas codificaciones particulares, para tratar las palabras que pertenezcan a un glosario específico. Estas codificaciones, pueden ser diferentes, según el glosario considerado, pero también, según la categoría del producto, o bien, también, la lengua de la descripción.

En la práctica, existe una casi infinidad de codificaciones posibles, entre las cuales, algunas de ellas, se detallan en la parte que sigue de esta descripción.

Así, de este modo, ciertas codificaciones, pueden tener en consideración las palabras situadas antes o después de la palabra para la cual se elabora la codificación. De una forma más precisa, una primera codificación, puede consistir en identificar las palabras o los glosarios a los cuales pertenecen las palabras que se encuentran hasta en un número de emplazamientos determinados, antes o después de la palabra para la cual se ha elaborado la codificación. Dicho de otro modo, se observa cuáles son las palabras próximas a un término que pertenece a un glosario, para sacar conclusiones sobre el grado de pertinencia del término considerado, en calidad de característica del artículo.

Una codificación más sofisticada, puede consistir en combinar la codificación anteriormente mencionada, arriba, simultáneamente, para las palabras que se encuentren antes y después del término estudiado.

Otro tipo de codificación, puede consistir en enumerar el número de palabras que pertenecen, en la descripción, al mismo glosario que aquél al que pertenece la palabra a codificar, después de la deduplicación.

5 En este caso, esta codificación, permite detectar el hecho de si, varias palabras de la descripción, corresponden al mismo tipo de característica, como, por ejemplo, a un color o a una forma. En este caso, puede resultar útil la combinación con otra codificación.

10 Otro tipo de codificación, puede consistir en asignar, a una palabra dada, una codificación correspondiente al número de coincidencias o casos de la palabra dada en la descripción del producto. Dicho de otro modo, esta codificación, permite detectar el hecho de si, un mismo término, interviene varias veces en la descripción, lo cual puede significar que ésta es, mucho más probablemente, una de las características principales del artículo.

15 En otro tipo de codificación, puede asignarse, a una palabra dada, la posición en la descripción de esta palabra, con relación a las otras palabras pertenecientes al mismo glosario, pero que son diferentes. Dicho de otro modo, en una descripción en la cual aparecen dos términos que pertenecen al mismo glosario, estos dos términos diferentes, no serán por lo tanto objeto de asignación de la misma codificación, lo cual permite discriminar diferentes casos de figuras. Es también posible el diferenciar el caso, en donde, un término, corresponde a una característica principal del artículo, o a un accesorio o auxiliar.

20 A título de ejemplo, en la descripción, “vestido negro con un cinturón rojo”, los términos “negro” y “rojo”, que pertenecen al mismo glosario de colores, no recibirán la misma codificación.

25 Otro tipo de codificación asignada a una palabra dada, puede consistir en identificar las palabras (o los glosarios a los cuales pertenecen las palabras), situadas entre la palabra considerada, y la primera palabra que pertenece al mismo glosario, situada después de la palabra dada, en la descripción.

30 Dicho de otro modo, en el caso en donde, varias palabras pertenecen al mismo glosario, esta codificación, es sensible a los términos que se encuentran situados entre estas dos palabras que pertenecen al mismo glosario. De una forma más precisa, esta codificación, permite detectar el hecho de si, entre estas dos palabras que pertenecen al mismo glosario, se encuentran presentes términos de lenguaje no característicos para el producto, o por el contrario, que correspondan a una característica de otra naturaleza.

35 Así, de este modo, en la expresión “azul o rojo o verde”, los términos azul, rojo, verde, corresponden a características potenciales de color, entre las cuales se encuentra una conjunción de coordinación, significando el hecho de que, los términos, se encuentran sobre el mismo plano de igualdad, en términos de pertinencia.

40 De forma inversa, en la expresión “un vestido negro con un cinturón rojo”, los términos “vestido” y “cinturón”, pertenecen al mismo glosario del tipo de artículo. La presencia de los términos “negro”, la cual es, en sí misma, una característica potencial del color y del término “o” son representativos de un cierto caso de figura, en donde se puede suponer el hecho de que, el segundo término representativo de las características del artículo, es decir, “cinturón”, corresponde a un accesorio del artículo principal, es decir, “vestido”.

45 En la práctica, pueden imaginarse múltiples codificaciones, conservando, al mismo tiempo, el principio de la invención.

En la práctica, el procedimiento, necesita una etapa de aprendizaje inicial, mediante el tratamiento manual de un cierto número de ejemplos, que permitan programar el categorizador bi-clase.

#### 50 Descripción resumida de las figuras

La forma de realizar la invención, así como las ventajas que se desprenden de ella, resaltarán, de una forma evidente, a raíz de la descripción del modo de realización que se facilita abajo, a continuación, y con el apoyo de la figura anexa, en la cual:

55 La figura 1, la cual comporta una tabla que ilustra el tratamiento en conformidad con la invención, de una descripción de producto.

#### 60 Forma de realizar la invención

Tal y como se ha mencionado ya, anteriormente, arriba, la presente invención, se refiere a un procedimiento para el tratamiento de datos que describen a productos y, de una forma más general, a toda oferta comercial, con la finalidad de extraer los atributos principales y permitir, así, de este modo, una búsqueda pertinente.

La figura 1, ilustra un ejemplo simplificado del tratamiento de una descripción de tamaño reducido, pero, ni que decir tiene, que puede aplicarse el mismo principio, a descripciones de tamaño mucho más importante.

5 Así, de este modo, en una primera etapa, la descripción, se analiza con el fin de determinar el hecho de si, cada una de estas palabras, pertenece a glosarios predeterminados. Esta descripción, puede seguir un tratamiento previo, por mediación de un lematizador, consistente en sustituir las palabras susceptibles de declinarse, en género y en número, mediante su forma en masculino y en singular.

10 Cada una de estas palabras, se compara, con el contenido de diferentes glosarios, de tal forma que, en el ejemplo de la figura 1, los términos “vestido” y “cinturón”, se identifican como perteneciendo al glosario “ARTÍCULO”, definido como el conjunto de artículos de vestir.

15 Del mismo modo, los términos “rojo” y “negro”, se identifican como perteneciendo al glosario “COLOR”, que reagrupa los colores de productos de vestir.

En un primer tiempo, se procede a realizar una descripción sustituida, reemplazando cada una de las palabras pertenecientes a un glosario, por el nombre del glosario correspondiente, de tal modo que, la frase “el vestido rojo con un cinturón negro”, se sustituya por la descripción siguiente “el artículo color con un artículo color”

20 A continuación, las codificaciones características, se elaboran para cada uno de los términos pertenecientes a un glosario. Tal y como ya se ha mencionado anteriormente, arriba, pueden construirse una infinidad de codificaciones y, la descripción que sigue, no retendrá más que un número limitado, por cuestiones de simplificación.

25 Así, de este modo, un primer tipo de codificación, consiste en definir los términos situados antes y después de la palabra considerada.

30 De una forma más precisa, puede revelarse como ventajoso, en la categoría de los artículos de vestir, en lengua francesa, el hecho de identificar los términos que pueden situarse hasta cuatro posiciones antes de la palabra estudiada, tomando en consideración el hecho de que, la palabra, puede situarse en el inicio de la descripción, y ser precedida por un número de palabras inferiores a cuatro. Las líneas referenciadas  $Z_{4\text{antes}}$ , en la tabla de la figura 1, recuperan, para cada uno de los términos pertenecientes a un glosario, las cuatro codificaciones correspondientes.

35 Se elabora el mismo tipo de codificación, para determinar los términos que se sitúan en una o dos posiciones después de la palabra considerada. Estas codificaciones, se reagrupan en las líneas  $Z_{2\text{después}}$  de la tabla de la figura 1.

40 La línea P de la tabla de la figura 1, ilustra otro tipo de codificación elaborado para cada uno de los temas pertenecientes a un glosario. De una forma más precisa, esta codificación, consiste en identificar la posición del término considerado, con relación a las otras palabras pertenecientes al mismo glosario, en la descripción. Así, de este modo, el término “vestido”, que pertenece al glosario “ARTÍCULO”, es el primer término de la descripción que pertenece a este glosario, siendo, el término “cinturón”, el segundo. Se encuentra el mismo orden, comparando el término “rojo”, con relación al término “negro”.

45 La línea X de la tabla de la figura 1, identifica otro tipo de codificación, consistente en enumerar el número de veces que la palabra considerada aparece en la descripción. En el ejemplo preciso de la tabla de la figura 1, cada uno de los términos pertenecientes a un glosario, no aparece más que una vez, en la descripción.

50 La línea N de la tabla de la figura 1, ilustra otro tipo de codificación, considerado como pertinente, únicamente para las palabras que pertenecen al glosario “ARTÍCULO”. Así, de este modo, esta codificación, no es representativa para las palabras de otro glosario y, de una forma más particular, para el glosario “COLOR”. Así, de este modo, en cuanto a lo concerniente a término “vestido”, que pertenece al glosario “ARTÍCULO”, un segundo término, a saber, el término “cinturón”, pertenece al mismo glosario “ARTÍCULO”, de tal forma que, la codificación N, adopta el valor 2, para la palabra “vestido”.

55 Lo mismo sucede, para todos los otros términos del glosario “ARTÍCULO” y, especialmente, para el término “cinturón”.

60 En la frase de aprendizaje, el ejemplo de la figura 1, se analiza para conducir a la conclusión de que, los términos “vestido” y “rojo”, deben retenerse, para definir los atributos de artículo y de color, mientras que, la inversa, los términos “cinturón” y “negro”, se desestimarán. Este ejemplo, participa, por lo tanto, en la programación del categorizador bi-clase, destinado a analizar el resultado de la codificación.

65 En el bien entendido, el ejemplo simplificado, con la finalidad de ilustraciones de la figura 1, puede tomar una forma más compleja, para corresponder a los casos de figura reales y, ello, funcionando mediante el mismo principio.

Se concibe, por lo tanto, el hecho de que, el procedimiento en conformidad con la presente invención, presenta múltiples ventajas, de entre las cuales:

- su adaptabilidad a diferentes lenguas, resultante de la capacidad del procedimiento, para analizar la estructura de la construcción de una descripción;
- 5 · la necesidad de un número muy reducido de ejemplos típicos, destinados al aprendizaje, en comparación con los procedimientos clásicos conocidos en la actualidad;
- una pertinencia muy grande, con una eliminación casi completa del ruido, ya que, la precisión, se encuentra cercana al 100%.

**REIVINDICACIONES**

- 5 1.- Procedimiento para el tratamiento de datos, puestos en ejecución mediante ordenador, en vistas a la determinación de los atributos principales de un producto, para su indexación y su acceso mediante motores de búsqueda, con una pertinencia incrementada, definido mediante una descripción que incluye un pluralidad de palabras, en el cual:
- para cada palabra, se determina si ésta pertenece a un glosario predeterminado;
  - para cada palabra que pertenece a un glosario:
    - 10 · se asigna, a la citada palabra, una pluralidad de codificaciones elaborados en función del glosario predeterminado, al cual pertenece la citada palabra y los glosarios a los cuales pertenecen las otras palabras de descripción,
    - 15 · se analiza, con un categorizador bi-clase, el conjunto de codificaciones elaborados para la citada palabra, para determinar el hecho de si, la citada palabra, es un atributo principal del producto retenido, con el fin de mejorar la pertinencia de indexación.
- 20 2.- Procedimiento, según la reivindicación 1, el cual comporta una etapa consistente en sustituir, en la descripción, las palabras pertenecientes a un glosario, por el nombre del citado glosario y, a continuación, la elaboración de las codificaciones mediante el análisis de la descripción, después de la sustitución.
- 25 3.- Procedimiento, según la reivindicación 1, en el cual, para cada palabra que pertenece a uno o varios glosarios:
  - se asigna, a la citada palabra, una pluralidad de codificaciones complementarias, elaboradas en función del glosario predeterminado, al cual pertenece la citada palabra, y las otras palabras de la descripción,
  - se analiza, con un categorizador bi-clase, el conjunto de codificaciones complementarias elaboradas para la citada palabra, para determinar si la citada palabra, es un atributo principal del producto.
- 30 4.- Procedimiento, según la reivindicación 1, caracterizado por el hecho de que, éste, comporta una etapa de selección de un número restringido de codificaciones, entre un grupo de codificaciones principales.
- 35 5.- Procedimiento, según la reivindicación 4, en el cual, las codificaciones seleccionadas, dependen del glosario al cual pertenece la citada palabra.
- 6.- Procedimiento, según la reivindicación 1, caracterizado por el hecho de que, el categorizador bi-clase, se programa mediante autoaprendizaje.
- 40 7.- Procedimiento, según la reivindicación 1, caracterizado por el hecho de que, una codificación ( $Z_n$ ), asignada a una palabra dada, consiste en identificar las palabras o los glosarios a los cuales pertenecen las palabras que encuentran hasta n emplazamientos después o antes de la citada palabra dada, en la descripción.
- 45 8.- Procedimiento, según la reivindicación 1, caracterizado por el hecho de que, una codificación ( $Z_{nm}$ ), asignada a una palabra dada, consiste en identificar las palabras o los glosarios a los cuales pertenecen las palabras que encuentran hasta n emplazamientos después de la citada palabra dada, y las palabras o los glosarios a los cuales pertenecen las palabras que se encuentran hasta m emplazamientos, antes de la citada palabra dada, en la descripción.
- 50 9.- Procedimiento, según la reivindicación 1, caracterizado por el hecho de que, una codificación (A), asignada a una palabra dada, consiste en identificar las palabras o los glosarios a los cuales pertenecen las palabras que encuentran situadas entre la citada palabra dada, y la primera palabra que pertenece al mismo glosario, situada después de la palabra dada en la descripción.
- 55 10.- Procedimiento, según la reivindicación 1, caracterizado por el hecho de que, una codificación (X), asignada a una palabra dada, consiste en enumerar el número de casos de la palabra dada, en la descripción del producto.
- 11.- Procedimiento, según la reivindicación 1, caracterizado por el hecho de que, una codificación (N), asignada a una palabra dada, consiste en enumerar el número de palabras en la descripción que pertenece al mismo glosario que aquél al que pertenece la citada palabra dada, y que son diferentes de la citada palabra dada.
- 60 12.- Procedimiento, según la reivindicación 1, caracterizado por el hecho de que, una codificación (P), asignada a una palabra dada, consiste en identificar la posición, en la descripción de la citada palabra dada, con relación a las otra palabras que pertenecen al mismo glosario, en la descripción.

ES 2 461 562 T3

Descripción	EI	vestido	rojo	con	un	cinturón	negro
Descripción sustituida	EI	ARTÍCULO	COLOR	con	un	ARTÍCULO	COLOR
Codificaciones							
Z <sub>4antes</sub>		EI	EI			ARTÍCULO	COLOR
			ARTÍCULO			COLOR	con
						con	un
						un	ARTÍCULO
A <sub>4después</sub>		COLOR	con			COLOR	∅
		con	un				
P		1	1			2	2
X		1	1			1	1
N		2	No representativo			2	No representativo
Resultado		RETENIDO	RETENIDO			RECHAZADO	RECHAZADO
Atributos retenidos		VESTIDO	ROJO				

FIGURA 1