



**Benemérita Universidad
Autónoma de Puebla**

Facultad de Ciencias de la Computación

**“Un Modelo de Representación de Textos
basado en Sintagmas”**

Tesis de Maestría

Para obtener el grado de

Maestro en Ciencias de la Computación

Presenta:

Miguel Rodríguez Hernández

Asesor:

Dr. Héctor Jiménez Salazar

Puebla, Pue.

Julio 2004

Dedicatoria

Dedico esta Tesis a mis Padres: Carlos Rodríguez Loeza y Ernestina Hernández S. por darme el don de la vida y a mis hermanos por su apoyo moral y económico, en especial a mi hermana Laura Rodríguez Hernández por su apoyo invaluable en todo momento. A mi esposa Luz María Ramírez Gaspar por su amor, apoyo y paciencia, a mis hijas Luz del Carmen y Maria del Rayo por ser el motivo que me impulsa a seguir adelante.

Agradecimientos

Quiero dar las gracias a mis maestros: M.C. David Pinto Avendaño y Dra. Darnes Vilariño Ayala por su apoyo y compartirme sus conocimientos, en especial a mi asesor Dr. Hector Jiménez Salazar por su tiempo, apoyo, paciencia y conocimientos que hicieron posible el logro de este trabajo. También agradezco el apoyo de mis amigos Hilario Salazar Martínez y David Alanis Urquieta, quienes me brindaron su amistad sincera y apoyo incondicional. Agradezco a Dios por la existencia y su infinita misericordia.

Índice general

Dedicatoria	II
Agradecimientos	II
Índice de figuras	VI
Lista de tablas	VII
1. Introducción	1
2. Recuperación de Información y Representación	3
2.1. El modelo Vectorial	5
2.2. El Modelo Booleano	6
2.3. El Modelo de Conjuntos Difusos	7
2.3.1. Conceptos de la Teoría de Conjuntos Difusos	8
2.3.2. Recuperación de Información empleando Conjuntos Difusos	9
2.4. Evaluación de un SRI	10
2.5. Lingüística de <i>corpus</i>	13
2.5.1. Asociación de Términos a partir de un <i>corpus</i>	14
2.5.2. Asociación de Términos mediante Análisis Lingüístico	15
2.6. Creación Automática de Tesoros	16
2.6.1. Palabras relacionadas	17
2.6.2. Verbos comunmente asociados	17
2.6.3. Expresiones comunes	17
2.6.4. Familias de palabras	17

<i>ÍNDICE GENERAL</i>	IV
2.6.5. Algunos desarrollos en la FCC.	18
2.7. Estructura Conceptual y Comunicación.	18
2.8. Grado de Pertenencia a un Dominio	21
2.8.1. Pertenencia a un Dominio	21
2.8.2. Representación de un Dominio	22
2.8.3. Pertenencia de Frases a un Dominio	22
2.9. Obtención del Extracto de un Texto	23
2.9.1. Uso de las <i>Relaciones de Sentido</i> para la Extracción de Oraciones	24
2.9.2. Representación de Textos utilizando <i>Relaciones de Sentido</i>	24
2.10. Reducción de Términos índice	25
3. Representación de documentos	27
3.1. Representación de documentos utilizando <i>relaciones de sentido</i>	27
3.1.1. Ejemplo	29
3.2. Consultas al sistema	33
3.2.1. Ejemplo	34
4. Pruebas	37
4.1. Consultas dadas por Estudiantes	38
4.1.1. Relaciones extraídas de un <i>Corpus</i>	40
4.1.2. Relaciones de Sinonimia	41
4.2. Consultas del Glosario	42
4.3. Comparación de Modelos	43
5. Conclusiones	45
Bibliografía	47
Apendices	49
A. Documentos de Prueba	49
A.1. Fragmento del <i>corpus</i> utilizado en la prueba	49
A.2. Fragmento de la colección de documentos	50

A.3. Fragmento de la colección de documentos representados utilizando un <i>corpus</i>	51
A.4. Fragmento de la colección representada con relaciones de sinonimia	54
A.5. Fragmento de la colección representada con relaciones de hiponimia	55
A.6. Fragmento de la colección representada con relaciones de hiperonimia	61
A.7. Fragmento de la colección representada con toda las relaciones .	64
A.8. Un Fragmento de Relaciones de Sinonimia	72

Índice de figuras

2.1. Diagrama básico de un SRI	5
--	---

Lista de tablas

3.1. Representación de los índices de la colección \mathcal{D}	31
3.2. Diccionario inverso de la colección representada	35
4.1. Diferentes tipos de pruebas realizadas	38
4.2. Consultas preprocesadas para el concepto Tierra	39
4.3. Resultados de las consultas de Estudiantes de Ingeniería	40
4.4. Resultados de las consultas de Estudiantes de Ingeniería utilizando relaciones de sinonimia	41
4.5. Desempeño promedio del sistema usando diversas relaciones . .	42
4.6. Modelo vectorial <i>vs</i> sintagmas (16 consultas de estudiantes) . .	43

Capítulo 1

Introducción

Hoy en día la demanda por información sigue creciendo al grado que la dependencia de las organizaciones e individuos de este recurso es crucial. Internet se ha convertido en el principal proveedor de información y por lo tanto, juega un papel estratégico en la toma de decisiones. Sin embargo, es deseable por todos los usuarios tener acceso a la información que satisface mejor sus necesidades. Así, la construcción de buenos buscadores en esta abrumadora cantidad de textos, imágenes y otro tipo de medios se ha convertido en un reto ineludible.

Estamos frente al problema de realizar búsquedas de información con base en los conceptos que contiene un texto y no en las palabras.

En este trabajo se experimentó con una representación de textos alternativa a la tradicional, la cual, tomó en cuenta el contenido semántico de los textos con el propósito de ser utilizada en la recuperación de información.

La idea principal en este trabajo es el "sentido de un sintagma", el cual se obtiene al combinar las propiedades de los conceptos asociados a los términos que ocurren en el sintagma. Se implantó un sistema que representa un documento considerando los sintagmas que contienen los términos índices del documento. Las consultas se representan de manera semejante a la representación de un documento y la búsqueda se realiza en forma semejante a la planteada por el modelo booleano de recuperación de información.

A continuación se describe de manera sintética el contenido de los capítulos que integran este trabajo.

En el capítulo 2 se describen los trabajos realizados alrededor de las *relaciones de sentido*, que conforman la base sobre la que el presente fue realizado.

En el capítulo 3 se plantea el método de representación de documentos con *relaciones de sentido* extraídas de un *corpus* y de EuroWordNet(EWN). Se ejemplifica el mismo con la finalidad de precisar las ideas que sustentan el método. Adicionalmente, se plantea el problema de la consulta al sistema y se presenta un ejemplo.

En el capítulo 4 se describen las pruebas realizadas al sistema con diferentes conjuntos de prueba y diccionarios de *relaciones de sentido*. Además se realiza una prueba adicional, confrontando el modelo basado en sintagmas con el modelo vectorial

Al final se presentan las conclusiones de este trabajo seguidas de los apéndices con parte de datos y resultados tomados de las pruebas.

Capítulo 2

Recuperación de Información y Representación

En este capítulo se hará una revisión de los trabajos que fundamentan la propuesta de una representación alternativa para sistemas de Recuperación de Información. A continuación se describe en forma breve la parte central de cada trabajo.

El trabajo de G. Ruge [11] incide en el uso de los términos de asociación de primer y segundo orden, con el fin de hacer consultas a un SRI.

En el caso del trabajo de Grefenstte [7] se proponen e implementan varias técnicas para crear un tesoro automáticamente a partir de una colección textual (*Corpus*). Esta técnicas son llamadas de conocimiento pobre, ya que no se requiere conocimiento del dominio para crear el tesoro.

Así mismo, se mencionan algunos resultados obtenidos del Laboratorio de Recuperación de Información de la Facultad de Ciencias de la Computación (FCC): herramientas y recursos.

En el trabajo de Josefina García Fajardo[6] se plantea que la formación del sentido de un sintagma puede verse como una combinación de las propiedades de los conceptos asociados a los términos que ocurren en un sintagma. Ella analiza diferentes casos que permiten plantear, en el marco de un modelo de discurso, como se lleva a cabo la formación del sentido de un sintagma.

CAPÍTULO 2. RECUPERACIÓN DE INFORMACIÓN Y REPRESENTACIÓN⁴

La estimación del grado de pertenencia de una palabra, así como el grado de pertenencia de una frase nominal simple se realizó en el trabajo de Jiménez Salazar H.[8]. El estudio realizado mostró buenas perspectivas de aplicación de estos métodos en la clasificación de frases.

En el trabajo de Reyes A. Berenice y Moyotl H. Edgar [4] se describe la identificación de las palabras más representativas de un texto con base en el punto de transición, es decir, la frontera entre las palabras de alta y baja frecuencia, derivado de la Ley de Zipf.

La presentación de un enfoque basado en *corpus* para la obtención automática del extracto de un texto, basándose en las *relaciones de sentido* de una palabra se realiza en Salazar M. H.[12].

La recuperación de información tiene sus orígenes en las bibliotecas y centros de documentación en los que se requerían búsquedas bibliográficas de libros y artículos de revista. El objetivo principal de cualquier centro de documentación es satisfacer las necesidades reales y potenciales de información de todos los usuarios, proporcionándoles la información veraz, pertinente, justo a tiempo y al menor costo.

Los usuarios que consultan un sistema de recuperación de información (SRI) para buscar información deben traducir su necesidad informativa en una consulta adecuada al sistema de recuperación. Esto supone utilizar un conjunto de términos que expresen semánticamente su necesidad.

El objetivo de la recuperación de información puede plantearse como sigue: Dada una necesidad de información y un conjunto de documentos, se requiere ordenar los documentos de mayor a menor relevancia para esa necesidad y presentarlos al usuario. Para ilustrar el problema nos apoyamos en la figura 2.1([adaptada de Rijsbergen,1979])[3].

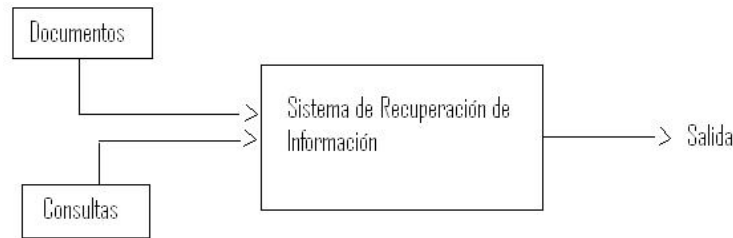


Figura 2.1: Diagrama básico de un SRI

En esta figura se ha supuesto que el SRI es una caja negra que acepta documentos y consultas y con base en la colección de documentos obtiene una salida, que es el conjunto de documentos que satisfacen la consulta.

El principal problema en este sistema es obtener representaciones homogéneas de elementos y consultas y procesar convenientemente esas representaciones para obtener la salida.

La investigación en recuperación de información busca diseñar sistemas que acepten consultas en lenguaje natural y proporcionen documentos adecuados a tales consultas, ordenados según algún criterio del sistema, de acuerdo con las características de los documentos y a las necesidades informativas expresadas por el usuario en su consulta [Belkin y Croft 1987], [13]

Aunque existen varios modelos para la recuperación de información, a continuación se describen únicamente los más importantes para este proyecto. Varios de los aspectos sobre recuperación de información tratados en este trabajo toman como referencia básica a [Baeza-Yates] [2]

2.1. El modelo Vectorial

Este es uno de los modelos más conocidos y difundidos para los sistemas de recuperación de información. En este modelo un documento D_j puede considerarse como un vector representado por sus términos índice. Los términos índice pueden estar ponderados de acuerdo a su importancia, esto es,

$$D_i = (d_{i1}, d_{i2}, \dots, d_{im}), \quad (2.1)$$

donde d_{ik} representa el peso del término k -ésimo en el documento i .

Por otro lado tendremos la frecuencia de un término k en el documento i , denotada por tf_{ik} y la frecuencia de un término en la colección de documentos denotada por df_k , esto es el número de documentos que hacen referencia al término k . Para asignar el peso de un término k se utiliza:

$$idf_k = \log_2(M) - \log_2(df_k) + 1 \quad (2.2)$$

donde, M el número de documentos. Finalmente el peso estará dado por la siguiente fórmula:

$$d_{ik} = tf_{ik} \cdot idf_k \quad (2.3)$$

Dados los vectores de dos documentos, D_i y D_j , es posible medir el coeficiente de similitud entre ellos, $sim(D_i, D_j)$. El modo más común para determinar la similitud entre dos documentos, representados mediante el modelo vectorial, es calculando el coseno del ángulo formado por ambos vectores:

$$sim(D_i, D_j) = \frac{\sum_{k=1}^m d_{ik} \cdot d_{jk}}{\sqrt{\sum_{k=1}^m d_{ik}^2 \cdot \sum_{k=1}^m d_{jk}^2}} \quad (2.4)$$

En esta ecuación se incluye la normalización de los vectores, para no privilegiar a documentos largos frente a otros documentos menos extensos.

2.2. El Modelo Booleano

El modelo booleano es un modelo simple de recuperación de información basado en la teoría de conjuntos y álgebra Booleana. Ya que el concepto de conjunto es bastante intuitivo, el modelo booleano provee un marco de referencia, el cual es fácil de comprender por parte del usuario de un sistema de recuperación de información. En este modelo, las consultas se especifican como expresiones booleanas las cuales contienen una semántica precisa.

Dada su inherente simplicidad y formalismo puro, el modelo Booleano recibió gran atención en el pasado y fue adoptado por muchos de los primeros

sistemas bibliográficos comerciales.

El modelo Booleano considera que los términos índice están presentes o ausentes en un documento. Como resultado, se asume que los pesos de los términos índice (w_{ij}) son todos binario, es decir, $w_{ij} \in \{0, 1\}$. Esto es, podemos ver un documento representado en el modelo Booleano como un vector de unos y ceros, es decir, si existe un 1 en la entrada del vector, entonces el término correspondiente se encuentra en el documento, en caso contrario no se encuentra y tendrá un valor de 0 en la entrada del vector. Una consulta q se compone de términos índice unidos mediante tres conectivos: *y*, *o* y *no*.

Para conocer la similitud entre un documento d_j representado mediante el modelo Booleano y una consulta q ($sim(d_j, q)$), también representada, realizamos lo siguiente:

$$sim(d_j, q) = \begin{cases} 1 & \text{si } d_j \text{ es relevante a } q, \\ 0 & \text{si } d_j \text{ NO es relevante a } q \end{cases} \quad (2.5)$$

Si la $sim(d_j, q) = 1$ entonces el modelo booleano predice que el documento d_j es relevante a la consulta q (o puede no serlo). En otro caso, la predicción es que el documento no es relevante. La principal ventaja de este modelo es su simplicidad; pero también presenta desventajas, una de ellas es que un empatamiento exacto puede producir gran cantidad de documentos o muy pocos.

2.3. El Modelo de Conjuntos Difusos

Representar documentos y consultas mediante conjuntos de palabras clave da como resultado descripciones, las cuales están solo relacionadas parcialmente al contenido semántico. Como resultado, la correspondencia o empatamiento de un documento con los términos de una consulta es aproximada o ambigua. Esto se puede modelar considerando que cada término de la consulta define un conjunto difuso y que cada documento tiene un grado de pertenencia (generalmente menor que 1) en este conjunto.

2.3.1. Conceptos de la Teoría de Conjuntos Difusos

La teoría de conjuntos difusos trata de la representación de clases cuyos límites no están bien definidos. La idea principal es asociar una función de pertenencia a los elementos de cada conjunto. Esta función toma valores en el intervalo $[0, 1]$, con un valor 0 para la no pertenencia al conjunto y un valor de 1 que corresponde a la pertenencia total.

Los valores de pertenencia entre 0 y 1 indican elementos marginales de la clase. Así, la pertenencia en un conjunto difuso es una noción gradual en lugar de abrupta.

Definición :Un subconjunto difuso A del universo U se caracteriza por una función de pertenencia $\mu_A : U \rightarrow [0, 1]$, la cual, asocia a cada elemento u de U un número $\mu_A(u)$ en el intervalo $[0, 1]$.

Las operaciones más utilizadas en los conjuntos difusos son: el complemento de un conjunto difuso, la unión de dos o más conjuntos difusos y la intersección de dos o más conjuntos difusos.

Definición: Sea U el conjunto universo, A y B dos subconjuntos difusos de U , y \bar{A} el complemento de A relativo a U . También, sea u un elemento de U . Entonces,

$$\mu_{\bar{A}}(u) = 1 - \mu_A(u) \quad (2.6)$$

$$\mu_{A \cup B} = \max(\mu_A(u), \mu_B(u)) \quad (2.7)$$

$$\mu_{A \cap B} = \min(\mu_A(u), \mu_B(u)) \quad (2.8)$$

$$(2.9)$$

La teoría de conjuntos difusos es útil para representar imprecisión o ambigüedad y ha sido aplicada a varios dominios.

2.3.2. Recuperación de Información empleando Conjuntos Difusos

Un enfoque adicional para modelar los procesos de recuperación de información es adoptar un *tesoro*, el cual define las relaciones entre los términos y consiste de una lista precompilada de palabras importantes en un dominio de conocimiento dado; para cada palabra en esta lista, se encuentra un conjunto de palabras relacionadas; en su variación más común, son derivadas de la relación de sinonimia.

La idea básica es expandir el conjunto de términos índice de la consulta con términos relacionados (obtenidos del tesoro) tal que los documentos relevantes adicionales pueden ser recuperados mediante la consulta del usuario. También, se puede usar un tesoro para modelar el problema de la recuperación de información en términos de conjuntos difusos como a continuación se explica. Se puede construir un tesoro mediante la definición de una matriz \vec{c} de correlación término a término (llamada también matriz de conexión de palabras clave) cuyas filas y columnas están asociadas a los términos índice de la colección de documentos. En esta matriz \vec{c} , se puede definir un factor $c_{i,l}$ de correlación normalizado entre dos términos k_i y k_l mediante:

$$c_{i,l} = \frac{n_{i,l}}{n_i + n_l - n_{i,l}} \quad (2.10)$$

donde n_i es el número de documentos que contienen el término k_i , n_l es el número de documentos que contienen el término k_l y $n_{i,l}$ es el número de documentos que contienen ambos términos.

Se puede utilizar la matriz \vec{c} de correlación de términos para definir un conjunto difuso asociado a cada término índice k_i . En este conjunto difuso un documento d_j tiene un grado de pertenencia $\mu_{i,j}$ calculado como:

$$\mu_{i,j} = 1 - \prod_{k_l \in d_j} (1 - c_{i,l}) \quad (2.11)$$

lo cual calcula la suma algebraica de todos los términos en el documento d_j . Expliquemos esto con más detalle. Un documento d_j pertenece al conjunto difuso asociado al término k_i si sus propios términos están relacionados al índice

k_i , mientras exista al menos un término índice k_i de d_j , el cual está fuertemente relacionado al índice k_i , entonces $\mu_{i,j} \sim 1$ y el índice k_i es un buen índice difuso para el documento d_j . En el caso que todos los términos índice de d_j se encuentren débilmente relacionados a k_i , el índice k_i no es un buen índice difuso para d_j $\mu_{i,j} \sim 0$. La adopción de una suma algebraica de todos los términos en el documento d_j permite una transición suave para los valores del factor $\mu_{i,j}$.

El usuario establece su necesidad de información proporcionando una consulta semejante a una consulta Booleana. Así, como sucede con el modelo booleano, esta consulta se convierte a su forma normal disyuntiva. Por ejemplo, la consulta $[q = k_a \wedge (k_b \vee \neg k_c)]$ se puede escribir en forma normal disyuntiva como $[q_{dnf} = (1, 1, 1) \vee (1, 1, 0) \vee (1, 0, 0)]$, donde cada uno de los componentes es un vector ponderado con valores binarios asociados a la tupla (k_a, k_b, k_c) . Estos vectores con pesos binarios son componentes conjuntivos de q_{dnf} . Sea cc_i una referencia al i -ésimo componente conjuntivo. Entonces

$$q_{dnf} = cc_1 \vee cc_2 \vee \dots \vee cc_p$$

donde p es el número de componentes conjuntivos de q_{dnf} .

El procedimiento para calcular los documentos relevantes a una consulta es análogo al procedimiento adoptado por el modelo Booleano clásico. La diferencia es que aquí se trata con conjuntos difusos, en vez de conjuntos Booleanos.

2.4. Evaluación de un SRI

La calidad de un sistema de recuperación de información está determinada por los resultados comparativos entre documentos recuperados y documentos relevantes para una consulta dada. Existen otras medidas de calidad que se refieren a velocidad de procesamiento y espacio de almacenamiento, pero en este trabajo se concentra la atención en la evaluación de la recuperación.

La evaluación de un SRI se realiza utilizando una colección de pruebas perfectamente caracterizada. Esta colección consiste en un conjunto de documentos y un conjunto de consultas. Para cada una de las consultas, (consultas

supervizadas) varios especialistas han seleccionado los documentos pertinentes de la colección. Esos documentos serán los documentos relevantes para la consulta dada. La evaluación de un sistema de recuperación consiste entonces en comparar, para cada una de las consultas de la colección de pruebas, los documentos que el sistema ha obtenido; esto es, los documentos recuperados y los documentos marcados como relevantes para esa consulta (consultas supervizadas).

El método más habitual para medir la calidad de un sistema de recuperación de información es la utilización de gráficas de precisión-evocación. Las medidas de precisión y evocación se basan en a , el número de documentos recuperados y b , el número de documentos relevantes; c es la intersección entre a y b .

Así, la precisión determina cuántos documentos son relevantes:

$$P = \frac{|c|}{|a|}$$

y la evocación determina cuántos documentos relevantes se han recuperado[13].

$$E = \frac{|c|}{|b|}$$

Para realizar el cálculo de la precisión por niveles consideremos lo siguiente:

Primero se define una consulta supervisada Q como una pareja donde su primer componente es una expresión y el segundo un subconjunto de \mathcal{D} : $(q, r) \in \mathcal{V}^+ \times 2^{\mathcal{D}}$. En seguida se define la gradación de una respuesta.

Sea $S = \langle \mathcal{D}, \text{Rep}, \mathcal{E} \rangle$ un SRI, (q, r) una consulta supervisada con $\#r = nr$ y supongamos que la gradación de $E(q)$ es $[D_{q_i}]_{1 \leq i \leq n_q}$. Dado el primer elemento de esta lista, digamos D_{q_j} , que ocurre en r , referimos al nivel de evocación 1. A continuación se dan las pautas para la evaluación por niveles.

Para el nivel de evocación 1 ($100 \times (1/nr) \%$), tenemos $P = 1/j$ y $E = 1/n_r$. Para el nivel de evocación 2 ($100 \times (2/n_r) \%$) $P = 2/j$ y $E = 2/n_r$, donde j es la posición en $[D_{q_i}]_i$ del segundo documento relevante ($\in r$), etc.

Para N_q consultas $(q_i, r_i)_i$: $P_i(l)$ es la precisión en el nivel l de la consulta i y la precisión para el nivel l

$$\bar{P}(l) = \frac{1}{N_q} \sum_{i=1}^{N_q} P_i(l) \quad (2.12)$$

Los niveles de evocación estándar, 0 %, 10 %, ... , 100 %, para una consulta q_i , corresponden a tomar como niveles de evocación 1 el 10 % de los documentos de r . Los valores de precisión se interpolan mediante:

$$P(l_j) = P(l)_{l_j \leq l \leq l_{j+1}} \quad (2.13)$$

Asimismo, se define el índice F_1 , que combina P y E, como:

$$F_1 = \frac{2 \times P \times E}{P + E} \quad (2.14)$$

Este valor permite medir el desempeño global de un SRI mediante consultas supervisadas.

2.5. Lingüística de *corpus* y Asociación Automática de Términos

Un *thesaurus* (Tesoro) en el campo de la Información, es una colección ordenada de conceptos que sirve para indexar documentos y recuperarlos. Un aspecto importante es, no solamente definir los términos sino también las relaciones entre ellos. Tales relaciones son: sinonimia (contenedor - receptáculo); términos más generales o hiperónimos (contenedor - tanque), hipónimos (tanque - contenedor), la relación parte-de (carro - tanque), antonimia (aceleración - desaceleración) y compatibilidad (carro - manejar). Se introduce el concepto de asociación de términos para estas relaciones diferentes. El tesoro tiene diferentes funciones en el campo de Información y Documentación. Durante la producción de documentos pueden ser utilizados para normalizar el vocabulario utilizado en el documento. Pueden ser utilizados también para construir una representación normalizada de los documentos para propósitos de recuperación de información. En este caso los términos índices son únicamente términos del tesoro y no todas las palabras que aparecen en el documento. Normalmente un indexado basado en tesoro se realiza manualmente, es decir, seleccionando los términos apropiados del tesoro que caracteriza el contenido de un documento. Sin embargo, Fuhr (1996)[11] sugiere un enfoque automático, es decir, elaborar una función para mapear conjuntos de palabras de los documentos hacia los términos del tesoro. La desventaja de un sistema de recuperación basado en términos que no usan indexado basado en tesoros, es que los documentos relevantes contienen otros términos que la consulta no encuentra. Por *ejemplo* para la consulta original PALABRA, AGRUPAMIENTO, TERMINO, ASOCIACIÓN los siguientes términos se podrían encontrar en los títulos de los documentos relevantes por ejemplo: Experimentos sobre la Determinación de las Relaciones Entre Términos, Una Taxonomía de los Sustantivos y Verbos del Inglés

Ejemplo de variación o asociación de términos

TERMINO, PALABRA CLAVE, DESCRIPTOR
PALABRA, SUSTANTIVO, ADJETIVO, VERBO
CO-OCURRENCIA, INTERCAMBIABLE, PRIMER ORDEN, SEGUNDO ORDEN
AGRUPAMIENTO, GRUPO, EXPANSIÓN, CLASE
SIMILITUD, SINONIMIA, ANTONIMIA, HIPONIMIA, HIPERONIMIA,
ASOCIACIÓN, RELACIÓN, VÍNCULO,

Tales asociaciones de términos se pueden usar en dos formas para ampliar el conjunto de documentos relevantes recuperables. En primer lugar, la asociación de términos se pueden presentar al usuario quien selecciona términos relevantes para sus necesidades de información. Harman(1988)[11] encontró que apoyar al usuario con diferentes relaciones de términos es fructífero. En segundo lugar, la consulta original del usuario se puede expandir automáticamente mediante los términos asociados.

Es importante mencionar que la construcción y mantenimiento de un tesoro requiere mucho esfuerzo. Éste es un problema porque los tesoros son dependientes del dominio construcción; debido a que cada dominio necesita su propio tesoro.

2.5.1. Asociación de Términos a partir de un *corpus*

Los enfoques de *corpus* pueden ser aplicados a cualquier base de datos textual. En este caso, el texto de la base de datos puede funcionar como el *corpus* necesario. Se asume que la utilización de términos en un *corpus* caracteriza su significado. El enfoque de asociación de términos mediante análisis estadístico consiste en extraer una colección de términos únicos del *corpus* y extraer los contextos para cada término único del *corpus*, esto es, dado un término x del *corpus* se extrae la oración donde x ocurre. A estos contextos se les llama asociación de términos de primer orden. Se supone que son semánticamente compatibles como "carro" y "llanta". Para los términos del *corpus* con gran cantidad de asociaciones de términos de primer orden se les llama asociaciones de términos de segundo orden, los cuales, se supone que son intercambiables

en el contexto y por lo tanto semánticamente similares.

En los experimentos que se realizaron en [11] con asociaciones de términos de primer orden y asociaciones de términos de segundo orden no se observó mejoría en el rendimiento de recuperación mediante expansión de consulta.

2.5.2. Asociación de Términos mediante Análisis Lingüístico Basado en *corpus*.

Por medio de un análisis estadístico se puede determinar si un término se refiere a algún otro en una oración o sentencia. Estas relaciones sintácticas de términos constituyen los contextos de los términos con poco significado - consisten de solamente un término. Las relaciones sintácticas de los términos extraídas de un *corpus* se pueden ver como asociaciones de primer orden.

La comparación de términos con base en estas relaciones sintácticas conduce a las asociaciones de segundo orden basadas en la lingüística. Contrastando las estadísticas de primer orden, las asociaciones de primer orden basadas en lingüística son semánticamente compatibles, por lo tanto, se supondría que las asociaciones de segundo orden basadas en la lingüística resultaran más similares semánticamente que aquellas basadas en estadísticas.

Ruge [11] explica que con los experimentos en *corpora* se pueden repetir los experimentos psicológicos.

2.6. Construcción Automática de Tesoros a partir de Texto Plano

Partiendo del hecho de que las unidades lingüísticas están relacionadas en un campo o área del conocimiento, los tesoros nos proporcionan la idea de qué objetos son importantes a qué campos o áreas. Esta premisa sugiere la creación de un tesoro; pero no resulta fácil tal creación debido a que se requiere mucho esfuerzo para su creación manual y mantenimiento. En esta sección se analiza cómo se crea automáticamente un tesoro a partir de texto plano usando técnicas de conocimientos pobre; esto es, sin información semántica.

Un problema que surge frecuentemente en los sistemas de recuperación de información es tener términos alternativos para expresar una consulta, por tanto, un tesoro sería un gran apoyo para resolver este problema. Existen algunos problemas en la construcción y actualización de tesoros, por lo que se han examinado técnicas para acelerar la creación manual de tesoros o de crearlos y actualizarlos automáticamente. Estas tareas pueden ser apoyadas mediante el uso de las computadoras para procesar colecciones de texto de gran tamaño.

Un método automático para la construcción manual de tesoros es el uso de los contextos de las palabras, en la forma en que Phillips(1985)[7] utilizó una ventana textual de cinco palabras antes y cinco palabras después de cada palabra, como contexto de una palabra. En [11] se muestra que el uso de frases nominales se pueden tomar como contexto para organizar palabras similares. De algunas técnicas mencionadas anteriormente se puede afirmar que la construcción de un tesoro puede ser automática o al menos acelerada.

Se les llama técnicas de conocimiento pobre porque no se requiere conocimiento del dominio para su uso. A continuación se describen las técnicas para crear un tesoro automáticamente.

2.6.1. Palabras relacionadas

Se extraen las palabras relacionadas mediante el concepto de vecinos cercanos que se define a continuación. Son vecinos cercanos $word_1$ y $word_2$ si $word_1$ es una de las N palabras más similares a $word_2$ y $word_2$ es una de las N palabras más similares a $word_1$. El valor de N es arbitrario.

2.6.2. Verbos comunmente asociados

En [7] la lista de verbos se obtiene automáticamente a partir de un tesoro de medicina que es usado como parte de los atributos y mediante los cuales se calcula una similitud entre nominales.

2.6.3. Expresiones comunes

Las expresiones comunes que se consideraron fueron expresiones de dos palabras. La frecuencia de tales términos de dos palabras en un *corpus* es más baja que aquellas palabras individuales. Aquí se sugiere un enfoque que muestra que contextos más amplios producen mejores resultados para palabras poco frecuentes.

2.6.4. Familias de palabras

Esta sección muestra el resultado de un proceso que encuentra familias de palabras usando como contexto un *corpus*. Grefenstette[7] encontró que al utilizar como contexto de cada palabra los números de documentos en que ésta aparece en variantes de grupos de palabras morfológicas se obtendrán buenos resultados. El mismo fenómeno aparece cuando los números de oraciones se utilizan como contextos, pero a un menor grado. Las variaciones morfológicas son numerosas y a menudo dependientes del dominio.

Los resultados obtenidos en [7] demuestran que es posible la creación automática de tesoros a partir de texto plano, de manera similar como se hace manualmente.

2.6.5. Algunos desarrollos en la FCC.

En este apartado se describe una herramienta morfosintáctica que se ha trabajado en el Laboratorio de Recuperación de Información de la FCC[1], esto es un tesoro. Se trata de un recurso de información documental de suma importancia para llevar a cabo, por ejemplo, la expansión de consultas y mejorar la evocación de los sistemas de Recuperación de Información (SRI).

En [1], el objetivo fue el refinamiento de técnicas para la construcción automática de *tesoros* en dominios específicos.

Para la representación semántica de los conceptos de un dominio, se partió del hecho de que la técnica basada en ventana, también conocida como conocimiento pobre, no requiere ninguna información semántica previa y depende de la frecuencia de la ocurrencia de los términos en cada contexto para realizar la representación.

En este trabajo [1] se realiza la extracción de palabras semánticamente relacionadas tomando como contexto la oración que contiene el vocablo. Esta técnica es una variante de la técnica de ventana propuesta por Grefenstette y cae dentro de la categoría de técnicas de conocimiento pobre. Posteriormente se realizó el preprocesado del *corpus* y una vez hecho esto, se procedió a calcular la similitud entre los contextos de las palabras elegidas; una de ellas constituirá el vocablo en la entrada del *tesoro*.

Los resultados obtenidos en [1] fueron alentadores si se considera que la mayoría de la investigación acerca del procesamiento de texto está enfocada al inglés y no al español.

2.7. Estructura Conceptual y Comunicación.

Se hace una propuesta para considerar cómo es que se llegan a comprender los conceptos desde el punto de vista de las interpretaciones de las acciones lingüísticas. La definición del universo de discurso - universo en el que se contextualiza una interpretación - abarca tres componentes:

1. Uno en el que se va construyendo la interpretación del discurso a partir

de la información gramatical y conceptual(I.D)

2. A través del cual recogemos la representación perceptual de la situación del habla, (R.P)
3. Corresponde a nuestra representación de estados, hechos, situaciones particulares de experiencias previas al acto lingüístico (R.E.P)

El enfoque se centra en la función de los conceptos en distintos contextos de acción lingüística. El lenguaje natural presenta características como la siguientes:

1. En ocasiones parece que a un objeto le viene bien un nombre , desde un punto de vista; y que tomando en cuenta otro punto de vista, no le queda bien el mismo nombre.
2. La combinación de dos significados en una estructura sintáctica puede hacerse a partir de unos de sus componentes, aunque esos conceptos tengan otros componentes que parezcan incompatibles.

Los componentes que conforman un concepto son parcialmente compartidos por una misma unidad lingüística.

Con base en los hechos antes mencionados, se puede decir que el concepto (k_1) que corresponde a un término (t_1) de la lengua es un conjunto de intersecciones de propiedades ($A_i A_j, \dots$) y que la expresión de dicho término desencadena un sentido a partir de dos propiedades, por lo menos, que se intersecten en el conjunto que conforma el concepto del término:

$$t_1 \rightarrow A_i A_j \quad A_i A_j \in k_1 \quad (2.15)$$

Textualmente García Fajardo J.[6] comenta:

”Con esta caracterización, en una estructura sintáctica pueden combinarse dos conceptos a pesar de que parezcan incompatibles, porque la amalgama no tiene que ocurrir con todos los componentes de cada concepto, pues para cada término será suficiente con dos propiedades que se intersectan en el conjunto que constituye su

concepto. Este principio permite combinar todos los significados lexicos que puedan ocurrir en una estructura sintagmáticamente bien formada y obtener así un número indefinido de significados sintagmáticos, resultados de esas combinaciones. Esta creatividad combinatoria permite hablar de hechos que no corresponden a una realidad del presente.”

Además, textualmente establece:

”Dos conceptos distintos pueden compartir algunas propiedades. Para que una propiedad forme parte de un conjunto conceptual determinado, se requiere que se interseque con la intersección de otras dos propiedades de dicho conjunto. Sean A, B, C tres conceptos. Si $A \cap B \neq \emptyset$ y el resultado se interseca con C , y es distinto del vacío, entonces, la intersección entre los tres conjuntos constituye un conjunto conceptual. Siguiendo el mismo principio, si una nueva propiedad D se interseca con $B \cap C$, las cuatro propiedades constituyen un mismo conjunto conceptual, y así sucesivamente. Ahora, si $D \cap E$, E otra propiedad, pero E no se interseca con otras propiedades del concepto al que pertenecen las intersecciones de A, B y C , resulta entonces que $D \cap E$ forma otro concepto, uno que comparte la propiedad D con el anterior.”

Se puede concluir que un concepto que corresponde a un término t se forma de la intersección de, por lo menos, dos propiedades que dan sentido al término t .

2.8. Grado de Pertenencia a un dominio y Métodos de Clasificación

El contexto (entendido como la vecindad de una palabra, o la estructura que contiene una frase) juega un papel importante en la clasificación y desambigüación. El principal problema consiste en clasificar una entidad lingüística en un dominio. En [8] se estudia el grado de pertenencia de un término a un dominio y posteriormente se analiza también el grado de pertenencia de una frase a un dominio.

A continuación se analizan los métodos para obtener el grado de pertenencia de una palabra y una frase nominal.

2.8.1. Pertenencia a un Dominio

Un dominio incluye el conjunto de términos, es decir, palabras o frases que aparecen en un lenguaje. Se precisa sobre la noción de pertenencia a dominios y representación de una palabra.

Representación de una palabra. Dado un vocabulario V , un diccionario $D : V \times \mathbb{N} \rightarrow M$, donde $M \subset V^*$, $(x, i) \mapsto D(x, i) = z$, asocia a cada palabra x y acepción i la definición z , conviniendo en que para cada $x \in V$ existe $n_x \in \mathbb{N}$ tal que $\forall n > n_x, D(x, n) = nil$. El diccionario inverso asociado a D es $D_{-1}(x) = \{(y, i) \in V \times \mathbb{N} | \exists i \in \mathbb{N} : x \text{ ocurre en } D(y, i)\}$, todas las definiciones donde ocurre el término x . El contexto de una palabra x respecto a D se define como:

$$C(x) = \bigcup_{(y,i) \in D_{-1}(x)} D(y, i) \quad (2.16)$$

2.8.2. Representación de un Dominio

La definición de la representación de un dominio se define a continuación en forma procedural:

1. Sea B_0 un conjunto de palabras básicas en el dominio D
2. Enriquecer B_0 mediante *inducción léxica*: B
3. Sea B' la unión de $\mathcal{C}_D(w)$, para cada $w \in B$.

Donde B_0 es un conjunto inicial de palabras del dominio Dom . Se obtiene B' que es el conjunto de palabras representantes del dominio.

Sea **Dom** un dominio y sea w una palabra. Se denota mediante $pert(w, \mathbf{Dom})$ el *grado de pertenencia* de w a un dominio **Dom**. Se procede a comparar dos enfoques con el objetivo de calcular $pert(w, \mathbf{Dom})$:

Utilizando una métrica. El **Dom** y x son establecidos en un espacio métrico y se decide si x se encuentra en la vecindad de **Dom**

Utilizando conjuntos rugosos. Se asocia a **Dom** una función $\mu_{Dom} : \mathcal{V} \rightarrow [0, 1]$ la cual es la función de pertenencia en realidad.

En la comparación entre ambos enfoques resultó que el enfoque métrico fue mejor que el enfoque rugoso pero cuando se utilizaron dos clases de términos disjuntas, la "incluida" y la "externa" entonces el enfoque métrico resultó menos eficiente que el enfoque rugoso.

2.8.3. Pertenencia de Frases a un Dominio

Para considerar términos determinados mediante frases, no sólo palabras, se utiliza la noción de *sentido de una frase* [6]. Se asume intuitivamente que el sentido de un sintagma está determinado por la combinación de propiedades de las palabras que aparecen en el sintagma. Se aproxima como sigue:

$$\Psi : \mathcal{V} \rightarrow \mathcal{V}^*, w \mapsto \Psi(w) = \left(\bigcup_{i \in n} D(w, i) \right) \cup \mathcal{C}(w) \quad (2.17)$$

donde $\Psi(w)$ consiste de todas las palabras que aparecen en las definiciones de w y en el contexto de w , con respecto al diccionario D . Para cualesquiera dos

palabras w_1, w_2 se tiene $\Upsilon(w_1, w_2) = \Psi(w_1) \cap \Psi(w_2)$. Luego para un sintagma arbitrario se define en una forma iterativa:

$$\xi(w_1, \dots, w_n, w) = \begin{cases} \xi(w_1, \dots, w_n) \cup \bigcup_{i=1}^n \Upsilon(w_i, w) & \text{si } \Psi(w) \cap \xi(w_1, \dots, w_n) \neq \emptyset, \\ 0 & \text{en otro caso} \end{cases} \quad (2.18)$$

donde $\xi(w_1, w_2) = \Upsilon(w_1, w_2)$. Ahora, para un dominio B , sea B' su representación. Por lo tanto para cualquier sintagma w_1, \dots, w_n se define el grado de pertenencia del sintagma $w_1 \dots, w_n$ a un dominio B como

$$\mu_B(w_1 \dots w_n) = \frac{\#(B' \cap \xi(w_1, \dots, w_n))}{\#\xi(w_1, \dots, w_n)}. \quad (2.19)$$

El concepto asociado a un sintagma se determina mediante las características comunes de los conceptos asociados a las palabras en el sintagma de la frase. Se menciona también que en el caso de frases largas, la intersección de los componentes tiende a ser vacía.

En [8] se practicaron dos métodos para calcular el grado de pertenencia de palabras y frases nominales al dominio de discurso: uno se basó en conjuntos rugosos, mientras que el segundo en una métrica. El primero provee mejores resultados pero depende de criterios extra para seleccionar adecuados valores del umbral.

2.9. Obtención del Extracto de un Texto mediante un Enfoque basado en *Corpus*

La idea principal del trabajo[12] consiste en obtener automáticamente un extracto de un texto, es decir, la oraciones más significativas de un documento. Esto se logra mediante las siguientes fases o etapas que se comentan someramente aquí.

Uno de los puntos importantes para la obtención del extracto es el uso de las *relaciones de sentido*.

Para lograr obtener el extracto se propone un método sencillo que toma en cuenta el contenido semántico de los textos. Este método se apoya en la

idea de las *relaciones de sentido* unidas a una palabra, más bien, se usa una aproximación a las relaciones de sentido debido a que se está utilizando únicamente un *corpus* como recurso lingüístico. En la literatura han existido trabajos que analizan las *relaciones de sentido* de una palabra tal como los términos de primer orden[11](términos que co-ocurren en el contexto de una palabra) y algunas aplicaciones muy exitosas como la construcción de tesoros.

2.9.1. Uso de las *Relaciones de Sentido* para la Extracción de Oraciones

El trabajo de [12] refiere a D. Marcu quien propuso un método para comparar el resumen de un artículo con algunas oraciones del texto fuente utilizando una función de similitud. Aquí, se sigue el método de Marcu en parte; las principales diferencias son que el método de Marcu no utiliza un *corpus* y que éste inicia con las parejas resumen-texto. En este trabajo se utilizó el contexto de una palabra w extraída de un *corpus* para representar a w , específicamente las oraciones como contextos. Este enfoque es simple: primero se representa cada una de las oraciones con los términos de primer orden de sus componentes (palabras). Posteriormente se calcula la similitud entre una oración y el texto completo excepto por z (el complemento de z). A mayor similitud entre una oración z y su complemento, más representativa es z del texto.

Para obtener el extracto de un texto automáticamente, el sistema desarrollado en [12] ejecuta tres pasos, preprocesamiento, representación del texto y la extracción de texto.

2.9.2. Representación de Textos utilizando *Relaciones de Sentido*

El vocabulario de un texto es el conjunto de palabras que aparecen en ese texto. La representación de una palabra w basada en sentidos considera un *corpus* C_1 que provee los contextos de w . Para cada palabra del vocabulario V , se asocian todos los contextos del *corpus* donde la palabra ocurre. De esta forma V se

puede representar como:

$$V = \{(x, y) | x, y \text{ ocurren en } S, \text{ y } S \text{ está en } C_1\}. \quad (2.20)$$

Consideremos una oración como una tupla, $S = (x_1, \dots, x_k)$, (sin repeticiones), así, dado el texto $T_1 = (S_1, S_2, \dots, S_n)$ su representación es:

$$T_2 = (Rep(S_1, V), \dots, Rep(S_n, V)). \quad (2.21)$$

donde

$$Rep(S, V) = (Rep(x_1, V), \dots, Rep(x_k, V)). \quad (2.22)$$

y

$$Rep(x, V) = \{y | (x, y) \in V\}. \quad (2.23)$$

donde k es el número de palabras del texto a ser procesado. El extracto de un texto se determina calculando la similitud - una simplificación de la función de Jackard- entre las oraciones S_i y su representación \overline{S}_i . En los experimentos se utilizó un *corpus* para determinar las relaciones de sentido de los términos y de 96 documentos de variados temas, además de 10 documentos de prueba para analizar los extractos resultantes.

Después de evaluar los resultados, los autores concluyeron que es conveniente realizar más experimentos, pues el método tiene buenas perspectivas .

2.10. Reducción de Términos Índice usando el Punto de Transición

En esta sección se analiza la reducción de términos índice usando el punto de transición. En este trabajo se identificaron los términos más representativos de un texto con base en el punto de transición, es decir, la frontera entre la palabras de alta y baja frecuencia derivado de la Ley de Zipf de frecuencias bajas. Un texto se representa así por los términos alrededor del punto de transición. Los resultados obtenidos mostraron que con un número reducido de palabras representativas se puede obtener un desempeño en representación de documentos comparable al método vectorial.

En concreto, este trabajo[4], analiza la forma de reducir los términos índice de un documento para su representación y con base en algunas investigaciones, se supone que las palabras con frecuencia media tienen el mayor contenido semántico del texto donde aparecen. Con base en el análisis de los términos de baja frecuencia se determina el punto de transición como:

$$n = \frac{\sqrt{1 + 8I_1} - 1}{2} \quad (2.24)$$

donde I_1 es el número de términos de frecuencia 1.

Una vez que se obtiene el punto de transición, deben seleccionarse términos alrededor de éste para conformar el conjunto de palabras que representan al texto. Los experimentos que se realizaron indican que tomando una banda de frecuencia del 25 % alrededor del punto de transición se obtienen buenos resultados.

En base a la obtención del punto de transición se puede representar un texto con un número reducido de términos índice. Esto facilita el manejo de análisis de documentos y gracias a esta representación se puede contribuir con la categorización de textos, que consiste en la asignación automática de documentos a categorías predefinidas. Esto ayuda a usuarios a acceder a gran cantidad de textos disponibles en Internet y en sus organizaciones.

La aplicación desarrollada en [4], propuso determinar si un texto pertenece a un dominio particular, esto es un subproblema de categorización de textos. Los resultados obtenidos en este trabajo fueron equiparables al método vectorial clásico.

Capítulo 3

Representación de documentos

En el capítulo precedente se han revisado varios trabajos que inciden en la noción de “asociación de términos”. Esta idea, como se vió en [11] está vinculada con las áreas de Psicología y Lingüística. Más aún, a través de este concepto es posible hacer un planteamiento para explotar la semántica de los términos que aparecen en los textos. Por ejemplo, se ha visto que este concepto es útil en la construcción de tesoros [7], idea que efectivamente se ha comprobado en nuestro ámbito dentro del Laboratorio de Recuperación de Información de la FCC [1]. Los términos de asociación de primer orden que son los contextos que rodean a una palabra en un texto constituyen el punto de partida para representar la palabra en cuestión, y otros experimentos demuestran un grado de efectividad no despreciable, incluso competitiva con métodos tradicionales. Además el consumo de recursos no es dispendioso, ya que únicamente es necesario un *corpus*; que entre más grande y balanceado se esperan mejores resultados. Así, resulta atractivo investigar esta representación que es más semántica que la tradicional. Se han de realizado diversos experimentos que constatan el valor de la representación.

3.1. Representación de documentos utilizando *relaciones de sentido*

Entre los experimentos realizados con esta representación podemos citar al que clasificó frases nominales que ocurren en las respuestas a preguntas abiertas [8]. En este caso, se representaron las respuestas usando como *corpus* un diccionario

del español. Tal trabajo se sustenta en la propuesta que hace J. García [6] sobre el sentido de un sintagma (ecuación 2.18). La propuesta se basa en la combinación que los seres humanos hacemos de las propiedades asociadas a los términos, de ahí la fórmula que se emplea para representar un sintagma. Otro trabajo en el que subyace la representación basada en términos de primer orden es el efectuado para obtener el extracto de un texto [12].

Cabe mencionar que hay otras propuestas realizadas por el grupo de RI de la FCC. Por ejemplo, la descrita en [4], la cual propone una forma alternativa de elección de índices.

Los trabajos antedichos nos orientan a realizar una prueba de la representación con términos de primer orden y el sentido de un sintagma para conocer su efectividad en la recuperación de información.

La propuesta que se hace para representar un documento es tomar cada oración que contenga índices y hacer su representación con la idea expresada por “el sentido de un sintagma”. Por tanto, requeriremos de un *corpus* que proporcione los términos de primer orden. Veamos a detalle el planteamiento de la propuesta y al final un ejemplo.

Estamos interesados en representar documentos mediante las *relaciones de sentido* de los índices de documentos. Este problema está motivado por la necesidad de explorar representaciones que tomen en cuenta el significado de los términos que contienen los documentos. Para resolver el problema de la representación de documentos con relaciones de sentido partimos de aproximar las *relaciones de sentido* de una palabra x con las palabras que ocurran en los contextos de x . Tomemos como contexto de una palabra las oraciones donde ocurra la palabra. Para lograr lo anterior emplearemos un *corpus* C . Denotemos con C' el *corpus* C preprocesado. Definimos entonces las palabras y del contexto de x según la relación xVy , donde V es:

$$V = \{(x, y) | x, y \text{ ocurren en } S, \text{ con } S \in C'\} \quad (3.1)$$

Debido a que los contextos son oraciones, definimos una oración como una tupla:

$$S = (t_1, t_2, \dots, t_k)$$

donde t_1, t_2, \dots, t_k son términos índices de un documento D

Para continuar con la solución del problema en cuestión debemos representar en principio a los términos índice de las oraciones con base en sus contextos obtenidos mediante (ecuación 3.1). Por lo tanto, nuestra representación \bar{t} para un término t sería:

$$\bar{t} = \{y | (t, y) \in V\} \quad (3.2)$$

Para el problema que tratamos de resolver, un documento D está compuesto de oraciones S_{ij} , donde i representa al documento al cual pertenece la oración y j representa el número de oración en el documento. Por ejemplo, podemos ver al documento D_1 como:

$$D_1 = (S_{11}, S_{12}, \dots, S_{1n})$$

De esta forma la j -ésima oración del i -ésimo documento se denotará como:

$$S_{ij} = (t_{ij1}, \dots, t_{ijn})$$

Una vez obtenida la representación de los términos índices que integran las oraciones de un documento (mediante la ecuación 3.2), podemos obtener la representación $\overline{S_{ij}}$ de una oración S_{ij} , realizando la intersección de todas las combinaciones de los términos representados y uniendo todas estas intersecciones. Esto se logra mediante la siguiente expresión:

$$\overline{S_{ij}} = \bigcup \overline{t_{ijk}} \cap \overline{t_{ijl}}, \quad k \neq l \quad (3.3)$$

Una vez que se obtienen las representaciones de todas las oraciones de un documento (mediante 3.3), se debe realizar la unión de cada una de las oraciones representadas para obtener así la representación de un documento mediante:

$$\overline{D_i} = \bigcup \overline{S_{ij}} \quad (3.4)$$

Con esto se tendría por completo la representación de un documento D .

3.1.1. Ejemplo

El ejemplo siguiente ayuda a precisar la ideas expuestas sobre la representación de documentos.

A continuación se muestran las oraciones de C'

1. *accion Magnitud define producto energia absorbida durante proceso duracion mismo*
2. *aceleracion razon cambio velocidad cambio tiempo*

En este caso, el vocabulario V_c del corpus C' es:

$V_c = \{ \text{cambio, velocidad, tiempo, razon, proceso, mismo, Magnitud, energia, durante, duracion, define, accion, absorbida, aceleracion, producto} \}$

Para el ejemplo para $i = 2$ tenemos la colección de documentos $\mathcal{D} = \{D_1, D_2\}$ donde, después de eliminar las palabras cerradas cada documento es:

$$D_1 = ((\text{accion producto energia tiempo})) \quad (3.5)$$

$$D_2 = ((\text{aceleracion Cociente incremento velocidad tiempo tienden cero})) \quad (3.6)$$

El vocabulario nos sirve, como se ha dicho, para construir los contextos de cada palabra de C' . A partir del cálculo de \bar{t} se determinaron los contextos para cada palabra del vocabulario, $V_{\mathcal{D}}$, de la colección \mathcal{D} . La tabla 3.1 muestra a x y a \bar{x} para cada $x \in V_{\mathcal{D}}$. En la tabla aparecen solamente las palabras que pueden representarse *i.e* que ocurren en V_c vocabulario:

x	\bar{x}
accion	absorbida define duracion energia Magnitud mismo proceso accion durante producto
producto	absorbida define duracion energia Magnitud mismo proceso accion durante producto
energia	absorbida define duracion energia Magnitud mismo proceso accion durante producto
tiempo	cambio razon tiempo velocidad aceleracion
aceleracion	cambio razon tiempo velocidad aceleracion
velocidad	cambio razon tiempo velocidad aceleracion
absorbida	absorbida define duracion energia Magnitud mismo proceso accion durante producto
define	absorbida define duracion energia Magnitud mismo proceso accion durante producto
duracion	absorbida define duracion energia Magnitud mismo proceso accion durante producto
razon	cambio razon tiempo velocidad aceleracion
cambio	cambio razon tiempo velocidad aceleracion
Magnitud	absorbida define duracion energia Magnitud mismo proceso accion durante producto
mismo	absorbida define duracion energia Magnitud mismo proceso accion durante producto
proceso	absorbida define duracion energia Magnitud mismo proceso accion durante producto
durante	absorbida define duracion energia Magnitud mismo proceso accion durante producto

Tabla 3.1: Representación de los índices de la colección \mathcal{D}

Representemos ahora cada uno de los documentos usando la representación de los términos que se obtuvo con la ec. 3.1. Así, usando ahora la ecuación 3.3 obtenemos:

$$\begin{aligned}
\overline{S_{11}} &= (\overline{accion} \cap \overline{producto}) \cup (\overline{accion} \cap \overline{energia}) \cup (\overline{accion} \cap \overline{tiempo}) \cup \\
&\quad (\overline{producto} \cap \overline{energia}) \cup (\overline{producto} \cap \overline{tiempo}) \cup \\
&\quad (\overline{energia} \cap \overline{tiempo}) \\
&= (\text{absorbida define duracion energia Magnitud mismo proceso accion durante producto}) \\
&\quad \cup (\text{absorbida define duracion energia Magnitud mismo proceso accion durante producto}) \\
&\quad \cup (\emptyset) \\
&\quad \cup (\text{absorbida define duracion energia Magnitud mismo proceso accion durante producto}) \\
&\quad \cup (\emptyset) \\
&\quad \cup (\emptyset) \\
\overline{S_{11}} &= (\text{absorbida define duracion energia Magnitud mismo proceso accion durante producto}) \quad (3.7)
\end{aligned}$$

Para el documento D_1 solamente se tiene S_{11} . Posteriormente, para el segundo documento:

$$\begin{aligned}
\overline{S_{21}} &= (\overline{aceleracion} \cap \overline{Cociente}) \cup (\overline{aceleracion} \cap \overline{incremento}) \cup (\overline{aceleracion} \cap \overline{velocidad}) \cup \\
&\quad (\overline{aceleracion} \cap \overline{tiempo}) \cup (\overline{aceleracion} \cap \overline{tienden}) \\
&\quad \cup (\overline{aceleracion} \cup \overline{cer0}) \\
&\quad \cup (\overline{Cociente} \cap \overline{incremento}) \cup (\overline{Cociente} \cap \overline{velocidad}) \cup (\overline{Cociente} \cap \overline{tiempo}) \\
&\quad \cup (\overline{Cociente} \cap \overline{tienden}) \cup \\
&\quad (\overline{Cociente} \cap \overline{cer0}) \cup (\overline{incremento} \cap \overline{velocidad}) \\
&\quad \cup (\overline{incremento} \cap \overline{tiempo}) \cup (\overline{incremento} \cap \overline{tienden}) \cup (\overline{incremento} \cap \overline{cer0}) \\
&\quad \cup (\overline{velocidad} \cap \overline{tiempo}) \cup (\overline{velocidad} \cap \overline{tienden}) \cup (\overline{velocidad} \cap \overline{cer0}) \cup \\
&\quad (\overline{tiempo} \cap \overline{tienden}) \cup (\overline{tiempo} \cap \overline{cer0}) \cup (\overline{tienden} \cap \overline{cer0}) \\
&= (\emptyset) \cup (\emptyset) \cup (\text{cambio razon tiempo velocidad aceleracion}) \\
&\quad \cup (\emptyset) \cup (\emptyset) \cup (\emptyset) \cup (\emptyset) \cup (\emptyset) \cup (\emptyset) \cup (\emptyset) \cup (\emptyset) \\
&= (\text{cambio razon tiempo velocidad aceleracion}) \cup (\emptyset) \cup (\emptyset) \cup (\emptyset) \cup (\emptyset) \cup (\emptyset) \\
\overline{S_{21}} &= (\text{cambio razon tiempo velocidad aceleracion}) \quad (3.8)
\end{aligned}$$

Como $D_1 = (S_{11})$ y $\overline{D_1} = \bigcup \overline{S_{1j}}$ tenemos que $\overline{D_1} = \{ \text{absorbida define duracion energia Magnitud mismo proceso accion durante producto} \}$ y para $D_2 = (S_{21})$ siguiendo el mismo procedimiento tenemos que $\overline{D_2} = \{ \text{cambio razon tiempo velocidad celeracion} \}$.

3.2. Consultas al sistema

Ahora se trata de realizar una consulta q al sistema para obtener los documentos más relevantes. Lo anterior deberá hacer uso de una representación que considere el significado de las consultas, al igual que se hizo con los documentos.

Para realizar una consulta $q = (q_1, q_2, \dots, q_k)$ al sistema, primero debemos representar los términos q_1, q_2, \dots, q_k de q , usando $\overline{q_1}, \overline{q_2}, \dots, \overline{q_k}$. Esto es equivalente a representar una oración usando la ecuación (3.9):

$$\overline{q} = \bigcup (\overline{q_i} \cap \overline{q_j}), \quad i \neq j \quad (3.9)$$

Además de la representación de q , consideramos apropiado utilizar un archivo inverso I . Para cada entrada, I se compone de un término del vocabulario V'_D y toda la lista de documentos en los que este término ocurre. Así, para cada término que se encuentra en V' asociamos los documentos en los cuales este término ocurre:

$$t_i \mapsto r_i = (d_{i1}, \dots, d_{im})$$

donde d_{ik} es el identificador de un documento D tal que $t_i \in \overline{D}$.

El archivo inverso es entonces, $I = ((t_i, r_i))_{t_i \in V'_D}$. Una vez obtenido el archivo inverso I podemos proceder a obtener los documentos más relevantes a la consulta q mediante la intersección de la lista de documentos en los cuales los términos de \overline{q} ocurren. La idea para implantar esta búsqueda booleana es obtener los documentos que tengan, de la consulta representada, más relaciones de sentido. Esto significa ordenar de mayor a menor los documentos extraídos de I según el número de términos que de \overline{q} tengan en su representación.

Asociamos al documento D_i la puntuación :

$$s_i = |\overline{D_i} \cap \bar{q}| \quad (3.10)$$

Así, la respuesta, B , a la consulta q sería:

$$B(q) = (D_{1'}, D_{2'}, \dots, D_{k'})$$

donde $1', 2', \dots, k'$ representan el ordenamiento descendiente de $1, \dots, k$ según s_1, \dots, s_k .

3.2.1. Ejemplo

El objetivo del siguiente ejemplo es hacer posible la comprensión del proceso de ejecución de consultas al sistema desarrollado en este trabajo. Para realizar una consulta a nuestro sistema, primero debemos obtener el vocabulario V'_D de la colección de documentos representados. En nuestro caso resultó:

$V'_D = (\textit{velocidad tiempo, razon, producto, proceso, mismo, Magnitud energia, durante, duracion, define, cambio, aceleracion, accion, absorbida})$

El siguiente paso es generar el alchivo inverso I , que contiene el identificador de los términos de V'_D en la colección de documentos. La tabla 3.2 muestra el diccionario inverso obtenido.

Términos	Documentos
absorbida	1
define	1
cambio	2
duracion	1
energia	1
razon	2
Magnitud	1
accion	1
mismo	1
proceso	1
durante	1
tiempo	2
producto	1
velocidad	2
aceleracion	2

Tabla 3.2: Diccionario inverso de la colección representada

Con la creación del archivo inverso I , se pueden ejecutar las consultas al sistema de recuperación de información con *relaciones de sentido*. Es necesario representar las consultas, antes de que sean ejecutadas por el sistema. Se realizaron dos consultas q_1 y q_2 para este ejemplo. Donde $q_1 = \{ \text{La acción es la fuerza que un cuerpo ejerce sobre un segundo cuerpo} \}$ y para $q_2 = \{ \text{Incremento de la velocidad en una unidad de tiempo} \}$. Se eliminaron palabras cerradas de q_1 y q_2 antes de realizar su representación. Las consultas realizadas resultaron $B(q_1) = \emptyset$, dado que $\overline{q_1} = \emptyset$ y para $B(q_2) = (\text{razon cambio tiempo velocidad aceleracion})$. El proceso para obtener los documentos más relevantes con base en la consulta q_2 es tomar cada término de $\overline{q_2}$ y buscarlo en I , registrando el número de documentos en el que el término se encuentra, y así para los restantes términos de $\overline{q_2}$. Una vez terminado con todos los términos de la consulta se procede a obtener la frecuencia del número de documentos, ordenándolos en orden descendente. El resultado de ejecutar la consulta q_2 ,

recuperó el documento número 2 solamente. El número de referencia a los términos que representan al documento 2 fue de 5.

Capítulo 4

Pruebas

En este capítulo se muestran las pruebas realizadas al sistema de recuperación de información que implementa un nuevo modelo para la representación de documentos.

La mayor parte de las pruebas se basan en la información contenida en un glosario de Física con 255 vocablos, y tres definiciones por vocablo.

El sistema considera, como se explicó en el capítulo 3, una colección de documentos y una fuente que proporciona relaciones de sentido. Asimismo, con fines del cálculo del desempeño del sistema, es necesario contar con preguntas supervisadas.

En todas las pruebas se trabajó con una colección de documentos provista por el glosario, específicamente, cada documento fue visto como una definición del vocablo; tomando la primera definición de las tres.

En el caso de la fuente que proporcionó las relaciones de sentido, en este caso se tomó para algunas pruebas un *corpus* también provisto por el glosario; la segunda definición. En otras pruebas se empleó EWN; solamente las relaciones de hiponimia, hiperonimia y sinonimia para sustantivos.

Finalmente una colección de consultas se formó con la tercera definición de cada vocablo del glosario.

Todos los términos de la colección de documentos, el *corpus*, las consultas y las relaciones de EWN fueron truncadas.

Se muestra en las secciones siguientes pruebas con los recursos anteriores, asimismo una prueba con el modelo vectorial. En resumen, la tabla 4.1 muestra las pruebas realizadas y las secciones en las que se describen los resultados.

Tabla 4.1: Diferentes tipos de pruebas realizadas

Prueba	Recurso	Sección	Consulta
<i>Corpus</i>	<i>Corpus</i>	4.1.1	Est. de Ing.
Sinonimos	EWN	4.1.2	Est. de Ing.
Comparación de Relaciones	EWN y <i>Corpus</i>	4.2	Glosario de Física
Comparación de Modelos	Sintagmas y Vectorial	4.3	Est. de Ing.

4.1. Consultas dadas por Estudiantes

Se trata de 16 consultas que son descripciones del concepto "Tierra"¹. La tabla 4.2 muestra estas consultas truncadas. Cada una de estas consultas fue introducida al sistema, utilizando el *corpus* y relaciones de sinonimia provenientes de EWN como apoyo a la representación por sintagmas. Las tablas 4.3 y 4.4 muestran los resultados obtenidos para las dos formas de representación (*corpus* y relaciones de sinonimia), usando la ecuación 3.3.

¹Recurso amablemente proporcionado por la M.C. Liset Fraguela Cuesta[5]

Tabla 4.2: Consultas preprocesadas para el concepto Tierra

Id	Consulta
q_1	tercer planet sistem sol vid diferent demas
q_2	mas efecto fuerz atraccion cuerpos
q_3	terc planet cercan Sol tard 24 hor gir propi eje rotacion cuenta divers cap recubr encuentr divers ecosistem encuentr divers tip biodivers
q_4	planet lug habit viv
q_5	planet habit Hombre unic vida
q_6	porcion material sol abund planet
q_7	convertir dia muramos
q_8	punt escencial lodo
q_9	planet sistem sol vivimos
q_{10}	primer cap cortez terrestr conoc tiempo
q_{11}	planet ocup terc lug cuatn posicion sistem sol Cuenta condicion propiedad favorec evol reproduccion vid organism
q_{12}	element natural capaz vida
q_{13}	nuev planet integr sistem solar
q_{14}	terc planet sistem sol satelit natural llam Luna
q_{15}	ecosistem conform plant animal element agua
q_{16}	terc enlac circuit

4.1.1. Relaciones extraídas de un *Corpus*

La tabla 4.3 muestra los resultados de las consultas a nuestro SRI. Algunas consultas fueron realizadas por algunos estudiantes de la Facultad de Ingeniería Mecánica de la BUAP. Asimismo, Id, #D, Pos, P, E y F_1 representan el Identificador de consulta, el Número de documentos recuperados, la Posición del documento correcto recuperado, la Precisión, la Evocación y el índice que permite medir el desempeño global del SRI respectivamente.

Tabla 4.3: Resultados de las consultas de Estudiantes de Ingeniería

Id	#D	Pos	P	E	F_1
q_1	235	17	0.00425	1	0.00847
q_2	236	200	0.00423	1	0.00843
q_3	235	132	0.00425	1	0.00847
q_4	137	99	0.00729	1	0.01449
q_5	167	No hubo	0	0	0
q_6	222	13	0.0045045	1	0.00896
q_7	0	No hubo	0	0	0
q_8	0	No hubo	0	0	0
q_9	234	15	0.00427	1	0.00851
q_{10}	255	75	0.00392	1	0.0078
q_{11}	235	68	0.00425	1	0.00847
q_{12}	0	No hubo	0	0	0
q_{13}	227	13	0.00440	1	0.00877
q_{14}	235	38	0.00425	1	0.00847
q_{15}	0	No hubo	0	0	0
q_{16}	0	No hubo	0	0	0
Promedios	151.125	41.875	0.00285	0.625	0.00568

4.1.2. Relaciones de Sinonimia

La tabla 4.4 muestra los resultados obtenidos al realizar las consultas de la tabla 4.2 utilizando relaciones de sinonimia para la expansión tanto de las consultas como de la colección de documentos.

Tabla 4.4: Resultados de las consultas de Estudiantes de Ingeniería utilizando relaciones de sinonimia

C	#D	Pos	P	E	F_1
q_1	38	3	0.02631	1	0.01739
q_2	53	No hubo	0	0	0
q_3	27	No hubo	0	0	0
q_4	11	2	0.09090	1	0.08695
q_5	9	2	0.11111	1	0.10526
q_6	9	4	0.11111	1	0.05405
q_7	0	No hubo	0	0	0
q_8	25	No hubo	0	0	0
q_9	35	2	0.02857	1	0.02816
q_{10}	15	No hubo	0	0	0
q_{11}	52	8	0.01923	1	0.00479
q_{12}	3	No hubo	0	0	0
q_{13}	33	10	0.03030	1	0.00604
q_{14}	43	2	0.02325	1	0.02298
q_{15}	1	No hubo	0	0	0
q_{16}	0	No hubo	0	0	0
Promedios	22.125	2.0625	0.0275	0.5	0.02035

Se ve claramente que el uso de relaciones de sinonimia mejora en un factor de 4 al empleo de un *corpus* como fuente que provee las *relaciones de sentido*.

4.2. Consultas del Glosario

Para las 255 consultas se realizaron las pruebas con relaciones de sinonimia(SINON), hiponimia(HIPO), hiperonimia(HIPER) y los tres tipos de relaciones contenidas en un sólo archivo(TODAS) para la expansión de consulta, y de la colección de documentos. Los resultados promediados se muestran en la tabla 4.5.

Tabla 4.5: Desempeño promedio del sistema usando diversas relaciones

Relación	#D	Pos	P	E	F_1
SINON	48.07843	7.050980	0.050846	0.560784	0.037055
HIPER	132.6352	26.77254	0.01768	0.73725	0.015286
HIPO	113.7803	23.60392	0.023224	0.694117	0.018432
TODAS	180.1921	45.2	0.014089	0.835294	0.012144
<i>CORPUS</i>	219.6117	78.32156	0.0065	0.898039	0.011090

Como puede observarse hay una mejora sustancial usando como soporte para la representación de sintagmas solamente extrayendo las relaciones de sinonimia (SINON). Aunque se nota, también mejor desempeño al usar relaciones de hiponimia (HIPO) con respecto a todas (TODAS) y a las de hiperonimia (HIPER). Es importante notar que el empleo de un *corpus* para proveer *relaciones de sentido* tiene el desempeño más bajo.

La alta precisión en el nivel de evocación uno, al usar sinónimos, puede observarse en la columna etiquetada como "Pos".

4.3. Comparación de Modelos

Se realizó una prueba adicional, confrontando el modelo basado en sintagmas con el modelo vectorial. La tabla 4.6 presenta resultados promedios de las consultas realizadas con los dos modelos de recuperación de información. En ambos modelos se consideró el uso de sinónimos, y como se aprecia en el modelo vectorial el uso de sinónimos no provee mejora.

Tabla 4.6: Modelo vectorial *vs* sintagmas (16 consultas de estudiantes)

Modelo	Precisión	Evocación	F1
Vectorial	0.021	0.5	0.041
Vectorial sinónimos	0.021	0.5	0.041
Sintagmas <i>corpus</i>	0.002	0.6	0.005
Sintagmas sinónimos	0.024	0.5	0.045

* * *

En resumen, se tiene que 1) el uso de sinónimos es indispensable, 2) el modelo basado en sintagmas muestra un comportamiento ligeramente mejor al vectorial; lo cual debe comprobarse por significancia estadística.

En el caso del modelo basado en sintagmas, cuando se realiza la representación de los términos usando un *corpus* vemos que se introduce mucho "ruido": solamente 57% de los términos pudieron representarse con el corpus. Como se espera, el empleo de sinónimos conlleva a una mejora, y al realizar la equivalencia entre términos el desempeño es casi el mismo al del modelo vectorial. Es importante aclarar que la versión de EWN utilizada es pobre (sólo el 51% de las palabras del glosario aparecen en él).

Las anteriores observaciones deben aún comprobarse con mayor exhaustividad. Así, está por comprobarse, por ejemplo, si eligiendo términos de los contextos del *corpus* y filtrándolos con la medida información mutua hay una mejora. De igual forma, es necesario corroborar los resultados obtenidos con otras colecciones de textos.

En los apéndices se agregaron fragmentos de los documentos de prueba, de la colección, del *corpus*, de la colección representada, de las representaciones utilizando relaciones de sinonimia (SINON), hiponimia (HIPO), hiperonimia (HIPER) y todas (TODAS), en los cuales se muestra la forma en que se encuentran arreglados los documentos, con el fin de mostrar los datos utilizados.

Capítulo 5

Conclusiones

En el presente trabajo se experimentó con una propuesta de representación semántica de textos con el fin de ser utilizada en la recuperación de información. Se partió de la idea del "sentido de un sintagma" y se implantó un sistema que realiza la representación de un documento considerando los sintagmas que contienen los términos índice del texto (aquellos términos que resultan al final del preprocesamiento del texto). Además, las consultas se representan de manera semejante a los documentos y la búsqueda se efectúa como en el modelo booleano de recuperación de información.

Las pruebas se apoyaron en tres colecciones, cada una consistió de 255 definiciones del área de física. La primera correspondió a la colección de documentos, la segunda al *corpus* y la tercera a la colección de consultas supervisadas. También se efectuaron 16 consultas truncadas para el concepto Tierra, obtenidas de estudiantes de la Facultad de Ingeniería Mecánica de la BUAP. El sistema se apoya en un *corpus* y relaciones de sinonimia (extraídas de EuroWordNet).

Los resultados obtenidos muestran, en todos los casos, un desempeño absoluto bajo, debido principalmente al material empleado en el experimento. El número de elementos de la colección es de 255 y la mayoría de éstos está compuesto de solamente una oración. Con respecto al número de documentos se utilizó un conjunto grande de consultas, cuyas características hacen difícil la recuperación exacta: por el origen coloquial de estas expresiones, el vocabulario de ellas es variado y, aunque se refieren al dominio, el tipo de lenguaje es informal

y diferente al de los documentos. Aún así, consideramos importante realizar el experimento con este material. Todas estas observaciones afectan a ambos modelos; sintagmas y vectorial.

El uso de relaciones de sinonimia en la representación de textos mejoró el desempeño con respecto a cualquier otra (hiponimia, hiperonimia, todas y *corpus*). Asimismo, el modelo basado en sintagmas supera ligeramente el desempeño al modelo vectorial

Pueden realizarse otras pruebas como la combinación de relaciones de sinonimia con hiperonimia, sinonimia con hiponimia e hiperonimia con hiponimia, provista por una fuente más rica y aplicando la representación en diversos puntos del procesamiento dentro del Sistema de Recuperación de Información. Asimismo, experimentar con un *corpus* de mayor tamaño.

Bibliografía

- [1] Aceves, R., Castilla Y.: *Resultados del Laboratorio de Recuperación de Información de la FCC*. Facultad de Ciencias de la Computación. BUAP
- [2] Baeza-Yates Ricardo: *Modern Information Retrieval*, Addison Wesley. 1999
- [3] C.J. van Rijsbergen: *Information Retrieval*. University of Glasgow, pp. 114-117. Second Edition, 1999.
- [4] Reyes A. Berenice, Moyotl H. Edgar: *Reducción de términos índice usando el punto de transición*. Facultad de Ciencias de la Computación. BUAP
- [5] Fraguela Cuesta L.: *Análisis de una Representación de Textos usando Lattices*. Tesis de Maestría en Ciencias de la Computación. BUAP. 2004
- [6] García Fajardo Josefina: *Estructura conceptual y comunicación* , El Colegio de México 1995.
- [7] Grefenstette Gregory: *Automatic Thesaurus Generation from Raw Text using Knowledge-Poor Techniques*, Rank Xerox Research Centre.
- [8] Jiménez Salazr H., Morales Luna G.: *Domain membership and clasificación methods* FCC. BUAP, CINVESTAV-IPN
- [9] Robbins Arnold. *Effective awk programming*, O'Reilly, Third Edition, 2001
- [10] Rodríguez Hernández M.: *Representación de Documentos utilizando Relaciones de Sentidos*. Tesis de Licenciatura. Facultad de Ciencias de la Computación. BUAP. Junio 2004.

- [11] Ruge Gerda: *Combining corpus linguistics and human memory models for automatic term association*, AI Group. Institut für Informatik, München. 1999
- [12] Salazar Martínez H.; Jimenez Salazar H. & Pinto A. D.: *Extraction: a corpus-based Approach*. Facultad de Ciencias de la de la Computación. BUAP
- [13] Zazo Rodríguez Angel: *Recuperación de información utilizando el modelo vectorial*, Participación en el taller CLEE-2001. Informe Técnico.

Apéndice A

Documentos de Prueba

A.1. Fragmento del *corpus* utilizado en la prueba

En este fragmento se muestran las definiciones de los vocablos que conforman al *corpus*. Recordemos nuevamente que se considera un documento por definición de vocablo

- 1 Magnitud defin product energi absorb proces duracion mism
- 2 cuerp ejerc fuerz accion efectu primer fuerz igual contrari reaccion
- 3 razon cambi cambi veloc cambi tiemp
- 4 Razón cambi veloc angul cambi tiemp
- 5 cuerp muev trayectori curv mism rapidez veloc constant cambi direccion veloc cambi instant ac
- 6 orbit planet punt alej Sol
- 7 Medida dimension vertical
- 8 intens fuerz violenci disminu suaviz
- 9 Espacio recorr cuerp posicion extrem
- 10 elong maxim distanci posicion extrem señal
- 11 Relación semej parec cos distint
- 12 form angul
- 13 Unidad med superficial equivalent decametr cuadr
- 14 Astrónomo grieg nac Samos primer afrim Tierra gir alrededor Sol
- 15 Matemático cientif inventor nac Siarcusa Sicilia Estableció principi flotacion cuarp nombr principi palanc consider inici mecan matemat logr calcul are n

- 16 Instrumento determin constant gravit mid desplaz pequen esfer mas m unid liger varill suspend punt hil delg produc acerc gr
- 17 Bala menor calibr ordinari fusil
- 18 astronom danes nac Knudstrup observ Brahe minuci realiz registr result trascendental importanci cientif posterior Kepler
- 19 Movimiento cuerp camp gravitatori influenci gravedad
- 20 Forma energi deb agit molecul constituy sustanci
- 21 Aspecto diferencian porcion mism cos conjunt mism clas cos porcion conjunt med cont

A.2. Fragmento de la colección de documentos

En esta sección se muestra el fragmento de la colección de documentos que se utilizaron para la prueba del sistema.

- 1 product energi tiemp
- 2 Ley cuerp ejerc fuerz segund cuerp transmit primer fuerz colineal igual magnitud sent opuest
- 3 Cociente increment veloc tiemp tiend cer
- 4 veloc angul objet muev trayectori curv ejempl circul mid divid angul recorr objet tiemp tard hac veloc angul angul recorr uni
- 5 mism acel normal general usa punt muev describ circunferent acel dirig centr mism
- 6 Punto orbit planet dist Sol
- 7 Elevación cuerp superfici tierr
- 8 Efecto amortigu
- 9 movimient oscilatori espaci recorr cuerp posicion extrem angul movimient circular distanci movimient rectiline
- 10 valor maxim alcanz cicl magnitud vari period
- 11 Relación semej cos distint
- 12 Relativo angul
- 13 Superficie comprend perimetr
- 14 Astrónomo grieg primer manten Tierra gir alrededor Sol
- 15 Matemático inventor grieg especial famos descubrir relacion superfici volum

esfer cilindr circunscrit Formuló principi hidrodinam conoc principi Arquímedes
16 Experimento cab Cavendish determin constant gravit Empleó balanz torsion
med pequeñisim fuerz attract esferill plom
17 Bala fusil pequeñ
18 Astrónomo danes pioner desarroll instrument medicion astronom
19 Movimiento cuerp propuls camp gravitatori
20 Sensación experiment recib direct indirect radiacion sol aproxim fueg etc
21 Porción cos ciert magnitud pes numer

A.3. Fragmento de la colección de documentos representados utilizando un *corpus*

El fragmento de esta sección muestra la colección de documentos representada mediante *relaciones de sentido*.

1 realiz mism igual ley duracion total encuentr proces recorr teori veloc
absorb fenomen Magnitud universal movimient product unidad oscilatori llam
fuerz = vibracion defin line energi trabaj direct distanci cambi bas important
proporcional punt form sol propag decrec potencial Einstein cuerp

2 torsion tendenci movimient Propiedad Aumento llam plan tiemp varill
cuadr Fuerza horizontal camin * presion ejempl interval pequeñ 2 equilibri
existent = descrit potencial G I extrem M parec trabaj teori atom V Tierra
constitu Inmovilidad l m sosten magnitud q s v w circundant superfici desarroll
densidad moment tridimensional product inercial accion modul Sistema
exterior suscept lug posicion sent Llámase fictici luz disminu Contracción
muev ley $E_k=1/2$ vuelt analiz radial somet unid proporcional energi presenci
Hamburgo gravitatori comet respect rozamient propi element comprim particul
efect Límite reduc sig sum sal disting geometr sep Iw^2 inerci pose distint
mov hall tiend gir resist orbit kilogram conoc segment central vibratori
rapidez determin magnet instant U1 rectiline cambi caracteriz ortogonal
debil influenci perpendicul seccion origin component vibrant oscilatori opac

A.3. Fragmento de la colección de documentos representados utilizando un corpus⁵²

normal intermedi nac ac coorden rotacion hech eje aplic planet propag mv
aport atraccion produc consider Gravitación independient realiz Astro violenci
impetu derech curv mencion Fuerzas permanec dinam infinitesimal relacion
simbol actuant calcul Espacio recorr siguint Unive radi sistem segu estudi
interaccion Relación reaccion ondas cos circul volum actu paralel asteroid
localiz Tercera fundamental $\frac{Mm}{r^2}$ semejant grad present fisic correspond grup
ocasion gran electromagnet supon camp centr acerc propiedad angul encuentr
Vector ocurr contact anterior gra ondulatori sirv sujet pes trayectori Conjunto
rect deb principal punt ded dad resort multipl line fricción tres trat intervencion
intens mecan contrari Ciencia V1 V2 equival electr incorpor espaci necesari
Fuerte gravedad igual opuest 1905 especial Ingeniero virtud primer Fricción
tamañ Luz aleman observ estad invers corrient estir estat totatl atraid uniform
Fn unidad metr original Newton Realidad indic defin molecul constituyent
heterogene segund hipotet cocient recib dich rad/s rap obtien Einstein analog
asoci cte ray Universo fij engendr adecu Barra fas mism alrededor impel
resbal expresion cosen acel geometri dirig period fuerz Varilla exist conjunt
desplaz repos oposicion maxim sol cinet atravies complet efectu ciert ocup man
condicion mas Punto respond extern extension fenomen conic respons multiplic
lumin oscilant total opon objet Movimiento irregul KX atra vist elemental
universal deform aspect apunt bord Cambio especific cuerp larg Parte part
Desviación oscil giratori delimit amplitud vectorial Representación marc direct
cort cont veloc clas dimension restitu continu distanci Internacional repentín
transmit ausenci relat vari comunic porces suaviz cantidad elec mutu grand
direccion referent aproxim Cantidad ces concentr constant elast form ejerc
Número Incapacidad forz mantien utiliz expres resultant revolu Agregación
desliz modern pulg situ manifest vez Suma

3 mism gir cer dad igual ley fisic total med encuentr correspondient cort
1905 gravedad medi aument teori perd veloc nocion propon posicion consider
choqu universal movimient alrededor razon alter exist muev porces coeficient
propiedad Tierra mo llam presenci fuerz nombr Albert part luz transcurr planet
contradiccion energi recib interval resuelv dinam angul determin invari cambi
bas important instantane grand repos caracteriz proporcional repentín son

propag instant rapidez inelast esp galilean infinit tiemp larg movil onda restit
Sistema Razón Einstein espaci cuerp constant

4 realiz mism gir aplic cer dad ro igual engendr lumin elast conoc naturaleza
multipl ley fisic centr imaginari desempeñ duracion total horizontal vibrant
objet deb med Supongamos equivalent mas encuentr satelit ondulatori
invarianci correspondient cort caen recorr eter 1905 gravedad cuatr medi mi/s
traves 1887 triangul Polígono relacion rap aument cocient Línea teori radial
Llámase vuel POSTULADO trayectori contribuy perd propon veloc nocion
rect TEORÍA opuest posicion asoci etc equilibri precision acort meridian
barr arbitrari ray Iw^2 radiacion desigual Cantidad disting grad impetu 000
choqu cos horizont alrededor perpendicul movimient libr razon result tiend
alter exist product alfa are negati LONGITUD muev superfici Conjunto
porces matemat complet año coeficient interferometr propiedad direccion
Tierra contigu Componente camp 0 mo inercial distint presenci llam conserv
propuest camin fuerz Técnicamente emple Movimiento nombr Teoría rotacional
orbit epoc elemental mv Albert señal 90 particul infinitesimal part lleg
curv Lado vibracion respect = clas defin areol Paralelogramo luz transcurr
suel sep isotrop maquin coordin refier tierra original describ line Considera
component contradiccion planet energi recib sig mayor I electromagnet trabaj
prop aisl interval aproxim resuelv Predice menor dinam papel descubri cinet
Tipo una transformacion inerci tangent distanci vari invers coordin angul
sistem variacion Vector produc uni determin resultant invari descrit cambi bas
Michelson CONTRACCIÓN important Consideremos trat instantane orden
grand Sol paralel aceler inclin pal fij vector vaci ident repos rad/s cuadr
descripcion caracteriz proporcional tangencial valor funcion repentin punt marc
espacio sol form tren referent mecan lad 186 Asume pas excentr parabol
rectangul propag cantidad son seccion magnitud tard tres instant rapidez
invariant $E_k=1/2$ unic Figura rectiline inelast circulo esp $e=0$ galilean nul
longitud signif diseñ m medicion suc conic infinit plan tiemp actu torc Parte
libre esfer Espacio consecuent movil onda larg q acerc adecu moment simetri
constitu v restitudo cosen extrem w potencial ac Sistema GENERAL influenci
RELATIVIDAD Einstein uniform expresion espaci Razón Afirma eje cuerp tom

A.4. Fragmento de la colección representada con relaciones de sinonimia

- 1 product energi tiemp
- 2 Ley cuerp ejerc fuerz segund cuerp transmit primer fuerz colineal igual magnitud sent opuest
- 3 Cociente increment veloc tiemp tiend cer
- 4 veloc angul objet muev trayectori curv ejempl circul circulatori mid divid angul recorr objet tiemp tard hac veloc angul angul recorr uni
- 5 mism acel aceler aceler normal general usa punt muev describ circunferent acel aceler aceler dirig centr centrist mism
- 6 Punto orbit planet dist Sol
- 7 Elevación cuerp superfici tierr
- 8 Efecto amortigu
- 9 movimient oscillatori espaci recorr cuerp posicion extrem radical angul movimient circular distanci movimient rectiline
- 10 valor maxim alcanz cicl magnitud vari period
- 11 Relación semej cos distint
- 12 Relativo angul
- 13 Superficie comprend perimetr
- 14 Astrónomo grieg primer manten Tierra gir rotador alrededor Sol
- 15 Matemático inventor grieg especial famos descubri relacion superfici volum esfer cilindr circunscrit Formuló principi hidrodinam conoc principi Arquímedes
- 16 Experimento cab Cavendish determin constant gravit gravitatori Empleó balanz torsion med pequeñisim fuerz atract esferill plom
- 17 Bala fusil pequeñ
- 18 Astrónomo danes pioner desarroll instrument medicion astronom
- 19 Movimiento cuerp propuls camp gravitatori
- 20 Sensación experiment recib direct indirect radiacion sol aproxim fueg etc
- 21 Porción cos ciert magnitud pes numer
- 22 triangul rectangul lad form formador angul rect
- 23 consider orig fundament efect

24 principi suces preced caus

25 Físico quimic ingles Analizó air atmosfer determin densidad medi glob
descubr composicion agu conoc propiedad hidrogen creador electrostat

26 Referido especial fuerz alej enajen centr centrist

27 atra dirig impel centr centrist

28 punt pas pasador resultant fuerz dirig centr centrist tierr cuent etiquet
Geometría cuerp import sumist homogene

29 Término equivalent efect centr centrist gravedad

30 colision aparec cuerp fuerz intens actu interval tiemp brev

A.5. Fragmento de la colección representada con relaciones de hiponimia

1 product energi fuerza_hidraul energia_radiant energia_potencial energia_cinet
energia_calorif electr vataj energia_atom energia_altern energia_de_activ actin
bizarri tiemp perfect imperfect futur pas aorist present perfect imperfect futur
pas aorist present

2 Ley cuerp particul cromosom cuerpo_viv amazon ejerc fuerz torsion
estres empuj propulsion magnet gravit enlac repulsion mec sangre_vital
coaccion presion energi impuls choqu crescend segund cuerp particul cromosom
cuerpo_viv amazon transmit primer fuerz torsion stres empuj propulsion
magnet gravit enlac repulsion mec sangre_vital coaccion presion energi impuls
choqu crescend colineal igual conciudadan sucesor substitut coetane asoci
magnitud extension cantidad balumb med amplitud med dimension sent opuest
3 Cociente increment amplif suplement aument aument veloc calm gradual
decel apresur cel ligerez pas tiemp perfect imperfect futur pas aorist present
perfect imperfect futur pas aorist present tiend cer parafin cerill cer

4 veloc calm gradual decel apresur cel ligerez pas angul acimut pendent oblicu
angulo_rect axil objet muev trayectori curv rulet crecient arco sinu helic arco
curv extradados catenari ejempl precedent excepcion circul transito_rod mid
divid angul acimut pendent oblicu angulo_rect axil recorr objet tiemp perfect
imperfect futur pas aorist present perfect imperfect futur pas aorist present

tard hac veloc calm gradual decel apresur cel ligerez pas angul acimut pendent
oblicu angulo_rect axil angul acimut pendent oblicu angulo_rect axil recorr uni
5 mism acel ciclotron normal general George.Washington Napoleón Lee
Ulysses.Grant Cromwell César usa punt foc cenit siti esquin posicion lug
polo_celestial engast periheli perige nad punto_geograf destin centr fuent afeli
apoge àpic omblig boca_del_estomag notici asunt canast carrer touchdown segur
dian muev describ circunferent acel ciclotron dirig centr sed centrocamp nucle
omblig corazon centro_de_gravedad baricentr dian entrañ mism

6 Punto orbit planet estrella_matutin asteroid pluton neptun uran saturn jupit
mart globo_terrestr venus mercuri dist Sol

7 Elevación cuerp particul cromosom cuerpo_viv amazon superfici fond
exterior interior car altur car palad paladar_dur interfaz banc vistavision
macadan pist afirm tarim perpendicular paviment paviment car dian tablero_de_ajedrez
delantal tierr caj cien aren barr moh marg marg humus gred arcill caolin subsuel
tierr moren Europa subcontinent arbol lengu istmo peninsul archipelag isla
tundr estep never brezal orient

8 Efecto amortigu

9 movimient corrient cimbre natural cubism futur hum sindical expresion
dad vanguardi sindical torcedur bol tiron gir encorvadur contragolp bambole
maremot nub choqu murmull espum mar_de_fond desplaz chilen maniobr
apresur agit desmadr tempestad alborot perturb golp bisagr tijeret planteamient
alcanc postracion cerradur vibracion inversion arranc bisagr zarande rotacion
gir reciproc codaz carrerill cabece transplant traduccion acel crol corrient
desliz prision_prevent balanc caid ascension anunci acos conduccion paracaïd
desplaz arremet locomocion adelant puerta_traser escap oscilatori espaci asient
parking asient parking recorr cuerp particul cromosom cuerpo_viv amazon
posicion extrem apoge limit extrem final limit angul acimut pendent oblicu
angulo_rect axil movimient corrient cimbre natural cubism futur hum sindical
expresion dad vanguardi sindical torcedur bol tiron gir encorvadur contragolp
bambole maremot nub choqu murmull espum mar_de_fond desplaz chilen
maniobr apresur agit desmadr tempestad alborot perturb golp bisagr tijeret
planteamient alcanc postracion cerradur vibracion inversion arranc bisagr
zarande rotacion gir reciproc codaz carrerill cabece transplant traduccion

acel crol corrient deslizz prision_prevent balanc caid ascension anunci acos
conduccion paracaidd desplaz arremet locomocion adelant puerta_traser escap
circular distanci altur alej distancia_entre_ej aproxim movimient corrient
cimbrenatural cubism futur hum sindical expresion dad vanguardi sindical
torcedur bol tiron gir encorvadur contragolp bambole maremot nub choqu
murmull espum mar_de_fond desplaz chilén maniobr apresur agit desmadr
tempestad alborot perturb golp bisagr tijeret planteamient alcanc postracion
cerradur vibracion inversion arranc bisagr zarande rotacion gir reciproc codaz
carrerill cabece transplant traduccion acel crol corrient deslizz prision_prevent
balanc caid ascension anunci acos conduccion paracaidd desplaz arremet
locomocion adelant puerta_traser escap rectiline

10 valor talon cotiz valor_nominal PIB insignificanci futil insistent garr signific
grandez preminent trascendent magnitud realc cost cost preci maxim alcanz
cicl tric scoot motociclet tandem biciclet magnitud extension cantidad balumb
med amplitud med dimension vari period cambri ordovic silurian devon
carbonifer permic triasic juras cretaz terciari cuaternari octav minut apoge
novici incub prohibicion luna_de_miel alist tiempos_de_guerr tiempos_de_paz
turno_de_noch turno_de_di guardi turn judicatur estadi periodo_lun arrend siest
tiemp interval juventud rein mitad set entrad periodo_glacial prehistori gen
era cachipoll instant moment rat sequi años tempor tempor canicul estacion
prorrog tiempo_reglamentari año_platon neolit mesolit paleolit edad_de_piedr
edad_del_hierr edad_de_hierr edad_del_bronc trimestr periodo_academ conden
plaz mes dia cuarto_de_sigl medio_sigl sigl dec olimpi bimestr semestr
año año año moment fin_de_seman quincen seman noch noch noch dia
dia ultimos_moment edad mediana_solteri madurez flor_de_la_v los_novent
los_ochent los_setent los_sesent los_cincuent cuarenten los_treint pubertad
adolescent juventud tiempos_del_colegi niñez infanci edad pas mileni existent
vid vid horas_extraordinari dia_util tiempo_complet tiempo_parcial seman
seman duracion moment tiempo_libr hor

11 Relación semej air comp contraimag cos distint

12 Relativo angul acimut pendent oblicu angulo_rect axil

13 Superficie comprend perimetr

14 Astrónomo grieg Pitágoras Ptolomeo Homero Hipócrates Euclides Epicuro

Aristóteles Esopo arcadian espartan corinti ateniens dori eoli primer manten Tierra gir tomador alrededor pieceri crisol habit ambient Sol

15 Matemático inventor grieg Pitágoras Ptolomeo Homero Hipócrates Euclides Epicuro Aristóteles Esopo arcadian espartan corinti ateniens dori eoli especial urolog radioterapeut radiograf loquer proctolog ortoped ocul oncolog obstetr neurolog inmunolog hematolog ginecolog geriatr otorrinolaringolog otolog patolog dermatolog callist cardiocircujan pediater anest arboricultor investigador_especializ oriental ocul meteorolog decor grafolog dactiloscop educacion dietist canon agreg famos personal astro inmortal descubr chism redescubr hallazg relacion avasall amistad asimil lio amor superfici fond exterior interior car altur car palad paladar_dur interfaz banc vistavision macadan pist afirm tarim perpendicular pavimento pavimento car dian tablero_de_ajedrez delantal volum capac esfer globul bola_de_naftalin glob cilindr circunscrit Formuló principi fundament hidrodinam magnetohidrodinam conoc principi fundament Arquímedes

16 Experimento cab estoperol cordon fiqu cordon_de_apertur cordon_de_apertur plom acoll acoll atadur filet bramant tendeder cuerda_de_trip cordon cuerda Cavendish determin diligent perseveranci diligent perseveranci constant coeficient parametr gravit Empleó balanz roman microbal balanza_electron torsion med pequeñez tamañ circunferent fuerz tall grad dimension probabil concentr termometri telemed gradacion hidrometri calorimetri calibr audiometri contramed pequeñisim fuerz torsion estres empuj propulsion magnet gravit enlac repulsion mec sangre_vital coaccion presion energi impuls choqu crescend atract atencion esferill plom bolin

17 Bala fusil carabin pequeñ

18 Astrónomo danes pioner Boone bosquiman desarroll crecimient desplieg elabor adelant instrument violin trompet trombon concert cancion sax gaiter pianist percusion organ obo tañedor_de_laud musico_de_jazz clavicordi arpist guitarr flaut clarinet chelist bajon acompañant medicion millas_por_hor astronom Ptolomeo Carl_Friedrich_Gauss Arthur_Stanley_Eddington cosmolog astrofis

19 Movimiento cuerp particul cromosom cuerpo_viv amazon propuls camp competent campo_magnet campañ labranz interior pista_de_tenis aren camposant

patio_de_recre arrozal cespced chatarreri cortafueg campo_de_maiz campus
teologi artes ciencia_aplic arquitectur cienci alometri autodrom campo_de_golf
gravitatori

20 Sensación experiment tante recib salon direct indirect radiacion radiact
onda_electromagnet sol polv cristal plastic precipit tronc ovoid dad poliedr
anill concav convex diente_de engranaj aproxim adyacent propincu fueg
llamar fuego_de_artilleri fuego_supr fuego_prepar fuego_indirect fuego_direct
fuego_cruz descarga_cerr andan linea_de_fueg fuego_antiaere etc

21 Porción cos ciert magnitud extension cantidad balumb med amplitud
med dimension pes pesadez pisapapel contrapes plom contrapes carg
numer multiplic multipl divisor dividend raiz cub cuadr numero_complej
enter numero_natural cuot numero_atom bas cardinal ordinal aguil factor
sum constant decimal pagin coordin número_de_la_Seguridad_Social matricul
cuenta_bancari parvedad mayori preponderanci

22 triangul rectangul cuadr lad cabecer form condicion acronim abreviatur
entrad etim tem radical forma_singul forma_plural angul acimut pendent oblicu
angulo_rect axil rect vector diagonal diametr secant asintot perpendicul

23 consider orig fundament bas rudiment efect guarnicion caballeri

24 principi fundament suces truen fucilaz culebrin hundimient bonanz agit
fecund ataqu brot alegron fall cortocircuit pulsacion coyuntur acto chispaz fluj
bajam refluj pleam lat combin fusion son interrupcion malform bomb contagi
vibracion rep contact destin comienz comienz manifest materializ reaparicion
sal abort disfuncional migracion movimient cas deton incendio_forestal conflagr
fogat cham fogar cambi conclusion contingent episodi fenix dificultad experient
preced caus

25 Físico quimic Louis_Pasteur Faraday Dalton bioquim ingles jacobit
londinens anglosajon anglosajon anglosajon tac Analizó air hemipleji parapleji
esbeltez agil graci atmosfer lobreguez ciclon anticiclon masa_de_air determin
diligent perseveranci diligent perseveranci densidad medi medi median glob
globo_celest aerostat descubr chism redescubr hallazg composicion modul
recit cancion adagi nocturn mezcl poema_sinfon suit larg cadenci fantasi
tocat pastoral divertiment octav septet sextet quintet cuartet tri duo sol
intermezz intermezz adapt agu aguazal estrech bahi abra casc rap ester

vad baji lag las_profund mar corrient cistern pilet fos conoc propiedad
criterio_select magnitud disposicion actitud coloc asimetri simetri configur
lineal veloc duracion orden_temporal languidez fortalez propiedad_fis asexual
gener desluc vital fisiologi comport estatur flaquez gordur complexion crudez
insipidez dessabor acu ricur ricur amargor dulzur acedi salin picant cadenci
pian decrescend fuerz metal pit disonanci musical silenci ruid olor intens color
color soseri lumin claridad textur tient predisposicion predisposicion consistent
genotip fenotip conform estil edad herenci hidrogen creador peluquer reloj
vinicultor sastr trabajador_del_acer hilander camiser cosedor veler guarnicioner
productor perfum confeccion causant maquet fabricant joyer herrero artific
sombbrero cristaler fant bordador costurer artesano diseñador_de_mod confiter
zapater peletero coreograf droguer constructor cervecero cester artista arquitect
encajer electrostat

26 Referido especial urolog radioterapeut radiograf loquer proctolog ortoped
ocul oncolog obstetr neurolog inmunolog hematolog ginecolog geriatr otorrinolaringolog
otolog patolog dermatolog callist cardiocircujan pediater anest arboricultor
investigador_especializ oriental ocul meteorolog decor grafolog dactiloscop
educacion dietist canon agreg fuerz torsion estres empuj propulsion magnet
gravit enlac repulsion mec sangre_vital coaccion presion energi impuls choqu
crescend alej centr sed centrocamp nucle omblig corazon centro_de_gravedad
baricentr dian entrañ

27 atra dirig impel centr sed centrocamp nucle omblig corazon centro_de_gravedad
baricentr dian entrañ

28 punt foc cenit siti esquin posicion lug polo_celestial engast periheli perige
nad punto_geograf destin centr fuerz afeli apoge àpic omblig boca_del_estomag
notici asunt canast carrer touchdown segur dian pas pasa_de_Corintio
pasa_de_Esmirna resultant fuerz torsion estres empuj propulsion magnet gravit
enlac repulsion mec sangre_vital coaccion presion energi impuls choqu crescend
dirig centr sed centrocamp nucle omblig corazon centro_de_gravedad baricentr
dian entrañ tierra caj cien aren barr moh marg marg humus gred arcill caolin
subsuel tierra moren Europa subcontinent arbol lengua istmo peninsul archipiela
isla tundr estep never brezal orient cuent comentari consej anecdot inform
integral diferencial apreci conversion interpol extrapol Geometría cuerp particul

A.6. Fragmento de la colección representada con relaciones de hiperonimia 61

cromosom cuerpo_viv amazon import homogene

29 Término equivalent substitut homolog igual efect guarnicion caballeri
centr sed centrocamp nucle omblig corazon centro_de_gravedad baricentr dian
entrañ gravedad

30 colision trompaz choque_en_caden choqu aparec cuerp particul cromosom
cuerpo_viv amazon fuerz torsion estres empuj propulsion magnet gravit enlac
repulsion mec sangre_vital coaccion presion energi impuls choqu crescend intens
acomet maldad fuerz actu interval cicl etern latenci tiempo_intermedi interludi
interrupcion despues_del_mediodi am intercal tiemp perfect imperfect futur pas
aorist present perfect imperfect futur pas aorist present brev adehal

A.6. Fragmento de la colección representada con relaciones de hiperonimia

1 product fecund energi fenomeno_fis fuerz anim tiemp period categoria_gramatical
categoria_gramatical

2 Ley cuerp cos unidad cos ejerc fuerz fenomeno_fis agente_causal influenci
delanter delanter med intens fortalez fortalez propiedad agresion segund
lugartenient cuerp cos unidad cos transmit primer mec fuerz fenomeno_fis
agente_causal influenci delanter delanter med intens fortalez fortalez propiedad
agresion colineal igual alma equivalent magnitud proporcion valor propiedad
sent curs ton opuest

3 Cociente increment cantidad cantidad cantidad veloc razon propiedad tiemp
period categoria_gramatical categoria_gramatical tiend cer digit cantidad

4 veloc razon propiedad angul anguil anguiliform objet aparat muev trayectori
curv curv line seccion ejempl boletin circul acumul desplaz mid divid angul
anguil anguiliform recorr objet aparat tiemp period categoria_gramatical
categoria_gramatical tard dia hac veloc razon propiedad angul anguil
anguiliform angul anguil anguiliform recorr uni

5 mism acel razon movimient normal circunstanci general oficial_g usa punt
disc lug signo_de_puntu señal part component exit muev describ circunferent
longitud med acel razon movimient dirig centr are punt edifici mism

A.6. Fragmento de la colección representada con relaciones de hiperonimia 62

- 6 Punto orbit camin planet astro dist Sol divinidad_roman
- 7 Elevación cuerp cos unidad cos superfici cap limit artefact tierr tierr materi
cos lug
- 8 Efecto amortigu aparat
- 9 movimient estad grupo_social suces negoci cambi cambi oscillatori espaci
duracion are are caract recorr cuerp cos unidad cos posicion actitud extrem
extrem part grad angul anguil anguiliform movimient estad grupo_social
suces negoci cambi cambi circular distanci part disposicion movimient estad
grupo_social suces negoci cambi cambi rectiline
- 10 valor investig maxim cantidad limit aledañ alcanz cicl interval rep
vehiculo_de_rued magnitud proporcion valor propiedad vari period unidad_de_tiem
medicion tiemp distanci
- 11 Relación semej afin cos estado_de_las_c afirm abstraccion atribut artefact
entidad distint
- 12 Relativo angul anguil anguiliform
- 13 Superficie comprend perimetr line
- 14 Astrónomo grieg europe actividad_sexual primer mec manten chic Tierra
gir pagador alrededor region Sol divinidad_roman
- 15 Matemático inventor grieg europe actividad_sexual especial convoc famos
gerifalt descubr chivataz logr relacion estad estad pertinent consej superfici
cap limit artefact volum cantidad esfer form artefact cilindr circunscrit
Formuló principi hech concept hidrodinam hidraul conoc proceso_cognitivo_superior
principi hech concept Arquímedes
- 16 Experimento cab convex suboficial altur pedaz artefact conductor Cavendish
determin firmez firmez constant numer cantidad gravit fuerz Empleó balanz
instrumento_de_medicion torsion fuerz med unidad cantidad unidad cantidad
magnitud magnitud activ maniobr pequeñisim fuerz fenomeno_fis agente_causal
influenci delanter delanter med intens fortalez fortalez propiedad agresion
atract caracterist bellez esferill plom pes cab
- 17 Bala fusil arma_de_fueg pequeñ descendient
- 18 Astrónomo danes europe pioner coloniz desarroll proces mejor instrument
actor medicion dat astronom fisic
- 19 Movimiento cuerp cos unidad cos propuls camp estado_de_las_c estado_de_las_c

A.6. Fragmento de la colección representada con relaciones de hiperonimia 63

fenomeno_fis region lug lug region area_de_conoc panoram facet servici
gravitatori

20 Sensación experiment investig recib aposent direct administr orden indirect
implic radiacion fenomeno_fis logr sol composicion activ aviacion aproxim
version distanci afin conexion fueg combustion equip arremet etc chism

21 Porción cos estado_de_las_c afirm abstraccion atribut artefact entidad
ciert magnitud proporcion valor propiedad pes propiedad_fis artefact numer
cantidad_defin actuacion señal identif categoria_gramatical categoria_gramatical
cantidad

22 triangul poligon instrumento_de_percusion rectangul paralelogram lad part
form lozani palabr angul anguil anguiliform rect intestino_grues

23 consider orig fundament relacion principi explic filef efect dotacion

24 principi hech concept suces event preced caus fond justif

25 Físico quimic cientif ingles britan lengua_german Analizó air disfuncion
part part comport semej atmosfer condiciones_meteorolog determin firmez
firmez densidad solidez medi glob esfer represent aeronav descubr chivataz logr
composicion ajust music agu residu excrecion cos fluido_corporal servici conoc
proceso_cognitivo_superior propiedad atribut hidrogen gas element creador
alma electrostat

26 Referido especial convoc fuerz fenomeno_fis agente_causal influenci delanter
delanter med intens fortalez fortalez propiedad agresion alej aislamient desapeg
antipati distanci esquivaz centr are punt edifici

27 atra dirig impel centr are punt edifici

28 punt disc lug signo_de_puntu señal part component exit pas excedent
documento_judicial resultant fuerz fenomeno_fis agente_causal influenci delanter
delanter med intens fortalez fortalez propiedad agresion dirig centr are punt
edificio tierra tierra materi cos lug cuent deud chivataz factur pensamient
Geometría cuerp cos unidad cos import burgues homogene

29 Término equivalent cognicion efect dotacion centr are punt edifici gravedad

30 colision accident aparec cuerp cos unidad cos fuerz fenomeno_fis agente_causal
influenci delanter delanter med intens fortalez fortalez propiedad agresion
intens grad propiedad actu interval period tiempo period categoria_gramatical
categoria_gramatical brev abon

A.7. Fragmento de la colección representada con toda las relaciones

1 product fecund energi fenomeno_fis fuerz anim fuerza_hidraul energia_radiant energia_potencial energia_cinet energia_calorif electr vataj energia_atom energia_altern energia_de_activ actin bizarri tiemp period categoria_gramatical perfect imperfect futur pas aorist present

2 Ley cuerp cos unidad particul cromosom cuerpo_viv amazon ejerc fuerz fenomeno_fis agente_causal influenci delanter med intens fortalez propiedad agresion torsion estres empuj propulsion magnet gravit enlac repulsion mec sangre_vital coaccion presion energi impuls choqu crescend segund mec lugartenient cuerp cos unidad particul cromosom cuerpo_viv amazon transmit primer mec fuerz fenomeno_fis agente_causal influenci delanter med intens fortalez propiedad agresion torsion estres empuj propulsion magnet gravit enlac repulsion mec sangre_vital coaccion presion energi impuls choqu crescend colineal igual alma equivalent conciudadan sucesor substitut coetane asoci magnitud proporcion valor propiedad extension cantidad balumb med amplitud dimension sent protest curs ton opuest

3 Cociente increment cantidad amplif suplement aument veloc razon propiedad calm gradual decel apresur cel ligerez pas tiemp period categoria_gramatical perfect imperfect futur pas aorist present tiend cer materi lip digit cantidad parafin cerill

4 veloc razon propiedad calm gradual decel apresur cel ligerez pas angul coloc form anguil anguiliform acimut pendient oblicu angulo_rect axil objet enjuici aparat muev trayectori curv line seccion rulet crecient arco sinu helic extrados catenari ejempl dem admonicion inform boletin precedent excepcion circul redondez acumul desplaz transito_rod circulatori mid divid angul coloc form anguil anguiliform acimut pendient oblicu angulo_rect axil recorr objet enjuici aparat tiemp period categoria_gramatical perfect imperfect futur pas aorist present tard dia hac veloc razon propiedad calm gradual decel apresur cel ligerez pas angul coloc form anguil anguiliform acimut pendient oblicu angulo_rect axil angul coloc form anguil anguiliform acimut pendient oblicu

angulo_rect axil recorr uni

5 mism acel barr instrumento_cientif razon movimient ciclotron aceler normal circunstanci general ide cualidad oficial.g institut catol projim George_Washington Napoleón Lee Ulysses_Grant Cromwell César usa punt final grafem tej tabac adorn labor_de_costur disc lug signo_de_puntu señal part component exit punta_de_lanz punta_de_alfil pic punta_de_flech colill pav punto_de_medi punto_en_zigzag sobrehil lomill cadenet respunt embast punto_de_ganchill foc cenit siti esquin posicion polo_celestial engast periheli perige nad punto_geograf destin centr fuent afeli apoge àpic omblig boca_del_estomag notici asunt canast carrer touchdown segur dian muev describ circunferent longitud med acel barr instrumento_cientif razon movimient ciclotron aceler dirig centr are punt edifici sed centrocamp nucle omblig corazon centro_de_gravedad baricentr dian entrañ centrist mism

6 Punto orbit camin planet astro estrella_matutin asteroid pluton neptun uran saturn jupit mart globo_terrestr venus mercuri dist Sol divinidad_roman

7 Elevación cuerp cos unidad particul cromosom cuerpo_viv amazon superfici cap limit artefact fond exterior interior car altur palad paladar_dur interfaz banc vistavision macadan pist afirm tarim perpendicular pavement dian tablero_de_ajedrez delantal tierr materi cos lug caj cien aren barr moh marg humus gred arcill caolin subsuel moren Europa subcontinent arbol lengu istmo peninsul archipelag isla tundr estep never brezal orient

8 Efecto amortigu aparat

9 movimient estad grupo_social suces negoci cambi corrient cimbre natural cubism futur hum sindical expresion dad vanguardi torcedur bol tiron gir encorvadur contragolp bambole maremot nub choqu murmull espum mar_de_fond desplaz chilen maniobr apresur agit desmadr tempestad alborot perturb golp bisagr tijeret planteamient alcanc postracion cerradur vibracion inversion arranc zarande rotacion reciproc codaz carrerill cabece transplant traduccion acel crol desliz prision_prevent balanc caid ascension anunci acos conduccion paracaid arremet locomocion adelant puerta_traser escap oscillatori espaci duracion are caract asient parking recorr cuerp cos unidad particul cromosom cuerpo_viv amazon posicion punt actitud extrem ideologi radical parte_externa_del_cuerp part grad punta_del_ded ded garr pie miembr alet pinz

apoge limit final angul coloc form anguil anguiliform acimut pendient oblicu
angulo_rect axil movimient estad grupo_social suces negoci cambi corrient
cambre natural cubism futur hum sindical expresion dad vanguardi torcedur bol
tiron gir encorvadur contragolp bambole maremot nub choqu murmull espum
mar_de_fond desplaz chilén maniobr apresur agit desmadr tempestad alborot
perturb golp bisagr tijeret planteamient alcanc postracion cerradur vibracion
inversion arranc zarande rotacion reciproco codaz carrerill cabece transplant
traduccion acel crol desliz prision_prevent balanc caid ascension anunci acos
conduccion paracaíd arremet locomocion adelant puerta_traser escap circular
distanci part disposicion altur alej distancia_entre_ej aproxim movimient estad
grupo_social suces negoci cambi corrient cambre natural cubism futur hum
sindical expresion dad vanguardi torcedur bol tiron gir encorvadur contragolp
bambole maremot nub choqu murmull espum mar_de_fond desplaz chilén
maniobr apresur agit desmadr tempestad alborot perturb golp bisagr tijeret
planteamient alcanc postracion cerradur vibracion inversion arranc zarande
rotacion reciproco codaz carrerill cabece transplant traduccion acel crol desliz
prision_prevent balanc caid ascension anunci acos conduccion paracaíd arremet
locomocion adelant puerta_traser escap rectiline

10 valor duracion creenci cantidad importanci bravur investig talon cotiz
valor_nominal PIB insignificanci futil insistent garr signific grandez preminent
trascendent magnitud reale cost preci maxim cantidad limit aledañ alcanz
cicl interval rep vehiculo_de_rued tric scoot motociclet tandem biciclet
magnitud proporcion valor propiedad extension cantidad balumb med
amplitud dimension vari period duracion regul medio_de_comun comunicant
autor tiempo_geolog cantidad unidad_de_tiempp medicion tiemp distanci
ritm fotop present peridista_de_deport periodista_de_prensa_r corresponsal
articul cambric ordovic silurian devon carbonifer permic triasic juras cretaz
terciari cuaternari octav minut apoge novici incub prohibicion luna_de_miel
alist tiempos_de_guerr tiempos_de_paz turno_de_noch turno_de_di guardi turn
judicatur estadi periodo_lun arrend siest interval juventud rein mitad set entrad
periodo_glacial prehistori gen era cachipoll instant moment rat sequi años
tempor canicul estacion prorrog tiempo_reglamentari año_platon neolit mesolit
paleolit edad_de_piedr edad_del_hierr edad_de_hierr edad_del_bronc trimestr

periodo_academ conden plaz mes dia cuarto_de_sigl medio_sigl sigl dec olimpi
bimestr semestr año fin_de_seman quincen seman noch ultimos_moment edad
mediana_solteri madurez flor_de_la_v los_novent los_ochent los_setent los_sesent
los_cincuent cuarenten los_treint pubertad adolescent tiempos_del_colegi niñez
infanci pas mileni existent vid horas_extraordinari dia_util tiempo_complet
tiempo_parcial tiempo_libr hor

11 Relación semej afin air comp contraimag cos creacion estado_de_las_c afirm
abstraccion atribut artefact entidad distint

12 Relativo angul coloc form anguil anguiliform acimut pendent oblicu
angulo_rect axil

13 Superficie comprend perimetr line

14 Astrónomo grieg europe actividad_sexual Pitágoras Ptolomeo Homero
Hipócrates Euclides Epicuro Aristóteles Esopo arcadian espartan corinti
ateniens dori eoli primer mec manten chic Tierra gir institu cambi movimient
viraj rotacion pagador guiñ sorte desviacion piruet vuelte tomador rotador
alrededor region pieceri crisol habit ambient Sol divinidad_roman

15 Matemático inventor grieg europe actividad_sexual Pitágoras Ptolomeo
Homero Hipócrates Euclides Epicuro Aristóteles Esopo arcadian espartan
corinti ateniens dori eoli especial facult expert convoc urolog radioterapeut
radiograf loquer proctolog ortoped ocul oncolog obstetr neurolog inmunolog
hematolog ginecolog geriatr otorrinolaringolog otolog patolog dermatolog
callist cardiocircujan pediater anest arboricultor investigador_especializ oriental
meteorolog decor grafolog dactiloscop educacion dietist canon agreg famos
gerifalt personal astro inmortal descubri chivataz logr chism redescubri
hallazg relacion estad pertinent consej avasall amistad asimil lio amor
superfici cap limit artefact fond exterior interior car altur palad paladar_dur
interfaz banc vistavision macadan pist afirm tarim perpendicular paviment
dian tablero_de_ajedrez delantal volum cantidad capac esfer form artefact
globul bola_de_naftalin glob cilind circunscrit Formuló principi hech concept
fundament hidrodinam hidraul magnetohidrodinam conoc proceso_cognitivo_superior
principi hech concept fundament Arquímedes

16 Experimento cab convex suboficial altur pedaz artefact conductor
estoperol cordon fiq cordon_de_apertur plom acoll atadur filet bramant

tendeder cuerda_de_trip cuerda Cavendish determin filosof firmeza diligent
perseverancia constant numer cantidad coeficient parametr gravit fuerz
gravitatori Emplé balanz instrumento_de_medicion roman microbal balanza_electron
torsion fuerz med unidad cantidad magnitud activ maniobr pequenez
tamaño circunferent fuerz tall grad dimension probabil concentr termometri
telemed gradacion hidrometri calorimetri calibr audiometri contramed pequenísimo
fuerz fenomeno_fis agente_causal influenci delanter med intens fortalez
propiedad agresion torsion estres empuj propulsion magnet gravit enlac
repulsion mec sangre_vital coaccion presion energi impuls choqu crescend atract
caracterist bellez atencion esferill plom elemento_metal pes cab bolin

17 Bala fusil propulsion arma_de_fueg tirote cañone cañonaz descarg carabin
pequeñ descendient

18 Astrónomo danes europe pioner coloniz Boone bosquiman desarroll
proces mejor crecimient desplieg elabor adelant instrument aparat maner
actor instrumento_de_vient instrumento_de_cuerda instrumento_de_percusion
caja_de_mus birimba organo_electron celest organill empulguer estac son
instrumento_cientif rued astiller instrumento_opt instrumento_med instrumento_de_medicion
brujula_giroscop guillotín cam garrot cadals teodolit planchet camara_cl ferul
silla_electr instrumento_de_dibuj cauteri azoter analiz violin trompet trombon
concert cancion sax gaiter pianist percusion organ obo tañedor_de_laud
musico_de_jazz clavicordi arpist guitarr flaut clarinet chelist bajon acompañant
medicion dat millas_por_hor astronom fisic Ptolomeo Carl.Friedrich.Gauss
Arthur.Stanley.Eddington cosmolog astrofis

19 Movimiento cuerp cos unidad particul cromosom cuerpo_viv amazon
propuls camp estado_de_las_c fenomeno_fis region lug area_de_conoc panoram
facet servici competent campo_magnet campañ labranz interior pista_de_tenis
aren camposant patio_de_recre arrozal cespced chatarreri cortafueg campo_de_maiz
campus teologi artes ciencia_aplic arquitectur cienci alometri autodrom
campo_de_golf gravitatori

20 Sensación experiment investig tante recib aposent salon direct golp
administr orden indirect implic radiacion fenomeno_fis logr radiact onda_electromagnet
sol sustanci estad form coloid gerifalt estrell solf composicion activ aviacion
polv cristal plastic precipit tronco ovoid dad poliedr anill concav convex

diente_de_engranaj aproxim version distanci afin conexion adyacent propincu
fueg combustion equip arremet llamar fuego_de_artilleri fuego_supr fuego_prepar
fuego_indirect fuego_direct fuego_cruz descarga_cerr andan linea_de_fueg fuego_antiaere
etc chism

21 Porción cos creacion estado_de_las_c afirm abstraccion atribut artefact
entidad ciert magnitud proporcion valor propiedad extension cantidad balumb
med amplitud dimension pes propiedad_fis artefact pesadez pisapapel contrapes
plom carg numer cantidad_defin actuacion señal identif categoria_gramatical
cantidad multiplic multipl divisor dividend raiz cub cuadr numero_complej
enter numero_natural cuot numero_atom bas cardinal ordinal aguil factor
sum constant decimal pagin coordin número_de_la_Seguridad_Social matricul
cuenta_bancari parvedad mayori preponderanci

22 triangul poligon instrumento_de_percusion rectangul paralelogram cuadr lad
part cabecer form lozani palabr condicion acronim abreviatur entrad etim tem
radical forma_singul forma_plural formador angul coloc form anguil anguiliform
acimut pendent oblicu angulo_rect axil rect line intestino_grues vector diagonal
diametr secant asintot perpendicular

23 consider orig fundament armon relacion principi explic filet bas rudiment
efect artificial text impresion aparent dotacion autodetermin riend vet
jurisdicción control influenci dominio_absolut energi persu alquimi valenci
eficaci autor mordac guarnicion caballeri

24 principi hech concept fundament suces event truen fucilaz culebrin
hundimient bonanz agit fecund ataqu brot alegron fall cortocircuit pulsacion
coyuntur acto chispaz fluj bajam refluj pleam lat combin fusion son
interrupcion malform bomb contagi vibracion rep contact destin comienz
manifest materializ reaparicion sal abort disfuncional migracion movimient cas
deton incendio_forestal conflagr fogat cham fogar cambi conclusion contingent
episodi fenix dificultad experient preced caus fond justif

25 Físico quimic relacion_social ciencias_natural cientif fotoquim electroquim
quimica_fis quimica_inorgan quimica_organ Louis_Pasteur Faraday Dalton
bioquim ingles britan lengua_german jacobit londinens anglosajon tac
Analizó air disfuncion part comport semej hemipleji parapleji esbeltez agil
graci atmosfer gas part condiciones_meteorolog espacio_aere lobreguez ciclón

anticlón masa_de_air determin filosof firmez diligent perseveranci densidad
solidez medi transmision estadist calceteri median glob esfer represent aeronav
globo_celest aerostat descubr chivataz logr chism redescubr hallazg composicion
ajust music modul recit cancion adagi nocturn mezcl poema_sinfon suit
larg cadenci fantasi tocat pastoral divertiment octav septet sextet quintet
cuartet tri duo sol intermezz adapt agu residu excrecion cos fluido_corporal
servici aguazal estrech bahi abra casc rap ester vad baji lag las_profund mar
corrient cistern pilet fos conoc proceso_cognitivo_superior propiedad atribut
criterio_select magnitud disposicion actitud coloc asimetri simetri configur
lineal veloc duracion orden_temporal languidez fortalez propiedad_fis asexual
gener desluc vital fisiologi comport estatur flaquez gordur complexion crudez
insipidez dessabor acu ricur amargor dulzur acedi salin picant cadenci pian
decrescend fuerz metal pit disonanci musical silenci ruid olor intens color soseri
lumin claridad textur tient predisposicion consistent genotip fenotip conform
estil edad herenci hidrogen gas element creador alma peluquer relojer vinicultor
sastr trabajador_del_acer hilander camiser cosedor veler guarnicioner productor
perfum confeccion causant maquet fabricant joyer herrer artific sombrerer
cristaler fant bordador costurer artesan diseñador_de_mod confiter zapater
peleter coreograf droguer constructor cervecer cester artist architect encajer
electrostat

26 Referido especial facult expert convoc urolog radioterapeut radiograf
loquer proctolog ortoped ocul oncolog obstetr neurolog inmunolog hematolog
ginecolog geriatr otorrinolaringolog otolog patolog dermatolog callist cardiocircujan
pediatr anest arboricultor investigador_especializ oriental meteorolog decor
grafolog dactiloscop educacion dietist canon agreg fuerz fenomeno_fis agente_causal
influenci delanter med intens fortalez propiedad agresion torsion estres
empuj propulsion magnet gravit enlac repulsion mec sangre_vital coaccion
presion energi impuls choqu crescend alej aislamient desapeg antipati distanci
esquivez enajen centr are punt edifici sed centrocamp nucle omblig corazon
centro_de_gravedad baricentr dian entrañ centrist

27 atra dirig impel centr are punt edifici sed centrocamp nucle omblig corazon
centro_de_gravedad baricentr dian entrañ centrist

28 punt final grafem tej tabac adorn labor_de_costur disc lug signo_de_puntu

señal part component exit punta_de_lanz punta_de_alfil pic punta_de_flech
colill pav punto_de_medi punto_en_zigzag sobrehil lomill cadenet pespunt
embast punto_de_ganchill foc cenit siti esquin posicion polo_celestial engast
periheli perige nad punto_geograf destin centr fuent afeli apoge àpic omblig
boca_del_estomag notici asunt canast carrer touchdown segur dian pas cantidad
lug son veloc puent jug viaj locomocion decision realiz tiemp escrutini
period fruta_sec excedent documento_judicial coll col maniobr viejos_tiemp
tiempo_inmemorial antañ pasa_de_Corintio pasa_de_Esmirna pasador resultant
fuerz fenomeno_fis agente_causal influenci delanter med intens fortalez
propiedad agresion torsion estres empuj propulsion magnet gravit enlac
repulsion mec sangre_vital coaccion presion energi impuls choqu crescend dirig
centr are punt edifici sed centrocamp nucle omblig corazon centro_de_gravedad
baricentr dian entrañ centrist tierr materi cos lug caj cien aren barr moh marg
humus gred arcill caolin subsuel moren Europa subcontinent arbol lengu istmo
peninsul archipelag isla tundr estep never brezal orient cuent histori deud
chivataz factur pensamient comentari consej anecdot inform integral diferencial
apreci conversion interpol extrapol etiquet Geometría cuerp cos unidad particul
cromosom cuerpo_viv amazon import comercial burgues chiv sumist homogene
29 Término equivalent estatus paridad cognicion substitut homolog igual efect
artificial text impresion aparient dotacion autodetermin riend vet jurisdiccio
control influenci dominio_absolut energi persu alquimi valenci eficaci autor
mordac guarnicion caballeri centr are punt edifici sed centrocamp nucle omblig
corazon centro_de_gravedad baricentr dian entrañ centrist gravedad
30 colision accident trompaz choque_en_caden choqu aparec cuerp cos unidad
particul cromosom cuerpo_viv amazon fuerz fenomeno_fis agente_causal
influenci delanter med intens fortalez propiedad agresion torsion estres empuj
propulsion magnet gravit enlac repulsion mec sangre_vital coaccion presion
energi impuls choqu crescend intens grad propiedad acomet maldad fuerz
actu interval period cicl etern latenci tiempo_intermedi interludi interrupcion
despues_del_medioidi am intercal tiemp period categoria_gramatical perfect
imperfect futur pas aorist present brev abon adehal

A.8. Un Fragmento de Relaciones de Sinonimia

brom bromist
confirm ratificatori comprob
escision hend
dolor dolorens
dicion termin
imperfeccion tachador
alcantarill debilit
autoriz licenci
alumbr alumbr
achuchon despach empell
circul circulatori
agent justici
boxe embotell
comport actu
dificultad difficult
añagaz reclam
censur reproch blasfem
milici reserv
piedad apiad
arreat arrob arreat
lobo_de_m marin
atrev afrent
lasitud abrum
denigr denigratori
alboroz alegr bromist
recomend recomend recomend
derroc expuls arroj
conexion acopl
transmigr transmigratori
ambicion aspiratori
apreton conservatori metens aplast estruj

estiliz dar_estil
medicament medic
desfalc malvers
arbitri trazador
conclusion termin metens
firmez resolutori
fundamental integr
asedi sitiador
centr centrist
pro ventaj
pensamient discurr
omeg abon
aficion gustabl alegr
absorcion adquisitori
miramient precautori
substitut reserv reserv
aislamient aisl
sublim sublimatori sublimatori
descentraliz descentraliz
adivin resolutori adivinatori
sep apart aisl separ
cambi relev
discusion querell
cuent etiquet