

BENEMERITA UNIVERSIDAD
AUTONOMA DE PUEBLA

Facultad de Ciencias de la Computación

Análisis de una Representación de Textos Usando *Lattices*

Tesis que presenta:

Liset Fraguela Cuesta

Para obtener el grado de Maestro en Ciencias de la Computación.

Asesor: Dr. Héctor Jiménez Salazar

Abril 2004

Índice general

1. Presentación del trabajo	5
2. Conocimientos básicos	7
2.1. Aproximación Superior e Inferior	9
2.2. Ejemplo	10
3. Aplicación en un Diccionario Onomasiológico	13
4. Pruebas	17
4.1. Aplicación directa del método	23
4.1.1. Aproximación Inferior	24
4.1.2. Aproximación Superior	27
4.2. Usando un diccionario de sinónimos	28
4.2.1. Aproximación Inferior	30
4.2.2. Aproximación Superior	33
4.3. Usando un <i>corpus</i> del área de Física	34
4.3.1. Aproximación Inferior	36
4.3.2. Aproximación Superior	37
5. Análisis de los resultados obtenidos	39
Conclusiones	43
Bibliografía	46
A.	49
A.1. Fragmento del glosario de una definición por vocablo	49
A.2. Fragmento del glosario de tres definiciones por vocablo	51
A.3. Fragmento del Diccionario de Sinónimos (truncado)	53

A.4. Fragmento del <i>corpus</i> (truncado)	55
A.5. Ejemplos de consultas dadas por estudiantes de ingeniería. . .	57
A.6. Obtención de A-Inf y A-Sup de algunas consultas del apéndice	
A.5	60

Capítulo 1

Presentación del trabajo

El propósito de cualquier Sistema de Recuperación de Información (SRI), es dar respuesta a una consulta, y proveer al usuario de textos que satisfagan sus necesidades de información. Esta necesidad se ha extendido y magnificado con el uso de buscadores en Internet. Aún las necesidades de información son múltiples y variadas. Por ejemplo, a menudo los usuarios necesitan diccionarios para buscar un vocablo o palabra que ha escapado a su memoria, aunque ellos recuerdan elementos de la definición del vocablo. Cualquier diccionario usado en este sentido, donde la búsqueda no es a través del vocablo sino por su descripción, es un diccionario onomasiológico (G.Sierra / J. McNaught[11]). Tal diccionario puede ser considerado un SRI, ya que éste provee al usuario de información que satisface sus necesidades.

Con este trabajo se pretende hacer un análisis del uso de *lattices* en la representación de textos para ser utilizada en la Recuperación de Información (RI). Se presentará una aplicación novedosa de los conceptos de aproximaciones en el *lattice* de conceptos propuesta por Keyun Hu et. al.[1]. Usando la idea similar de la teoría de agrupamiento rugoso y las propiedades únicas del *lattice* de conceptos, las aproximaciones superior e inferior de cualquier conjunto de objetos o atributos pueden ser halladas explotando el *inf-(sup)-irreducible* de elementos en un *lattice* de conceptos.

Los conceptos formales son parejas (*Objetos-Atributos*) con propiedades atractivas que permiten algunas aplicaciones como la recuperación de información. Con estas parejas se construyen *lattices* o retículas, donde se mantiene un orden parcial entre las parejas: hacia arriba están los más generales (sup), hacia abajo los más particulares (inf).

Se pretende ubicar un nuevo concepto en la *lattice* haciendo algunas

aproximaciones que tienen base en los conjuntos rugosos. Así, dado un conjunto de características X (atributos), su aproximación inferior es una colección de conjuntos de objetos cuyos atributos contienen dichas características. En cuanto a su aproximación superior, ésta es una colección de objetos cuya unión de sus atributos está contenida en las X . Así, se aproxima con base en características de los conceptos en la *lattice* al nuevo concepto para decir, esencialmente, a cuáles conceptos se parece.

Los resultados de las aproximaciones superior e inferior serán evaluados con respecto a precisión (*precision*) y evocación (*recall*), cuyas definiciones (adaptadas de Rijsbergen [8]) vendrán explicadas en el capítulo 2. En este capítulo también se expondrán algunas nociones sobre *lattices* de conceptos, agrupamiento rugoso y relaciones de orden, entre otras. Se realizará una breve exposición del planteamiento propuesto por Keyun Hu et. al.[1]. Además, se expone un ejemplo sencillo (B.Ganter / R.Wille[4]) para explicar de forma más amena el método utilizado.

En el capítulo 3 es aplicado el método expuesto en el capítulo anterior, donde dadas las características de un concepto de Física se obtiene(n) el(los) vocablo(s) más adecuado(s). Se usaron dos glosarios, uno en el cual para cada vocablo había una definición; y en el segundo, a cada vocablo le correspondían alrededor de 3 definiciones.

En el capítulo 4 se realizaron diferentes pruebas, en las cuales se hicieron búsquedas de cinco vocablos diferentes. En la primera prueba las palabras que componen nuestro glosario, así como las de la consulta, fueron preprocesados para trabajar con seudoraíces. En la segunda prueba se utilizó un diccionario de sinónimos para representar cada palabra; en la consulta y en las definiciones de los vocablos. La tercera prueba se utiliza un *corpus* del área de Física, específicamente de Mecánica y Cinemática (Página de Internet: Física con Ordenador [5]), con el cual, después de ser preprocesada, se expanden tanto los glosarios como las consultas. En este capítulo las pruebas son descritas en cuanto a datos de entrada, comandos utilizados para su realización, resultados obtenidos y comentarios acerca de éstos.

En el capítulo 5 se analizan los resultados de las pruebas realizadas en el capítulo anterior.

Al final se presentan las conclusiones a las que se llegó en este trabajo.

Se presenta en un apéndice (A) fragmentos de los glosarios, el diccionario de sinónimos, el *corpus*, además de consultas dadas por estudiantes de ingeniería y los resultados de aproximación superior e inferior de algunas de estas consultas.

Capítulo 2

Conocimientos básicos

En esta sección se expondrán las nociones básicas necesarias del *lattice* de conceptos y conjuntos rugosos brevemente, con base en Keyun Hu et. al.[1].

Primero comenzaremos con algunas definiciones básicas de relaciones de orden en un conjunto, que en nuestra aplicación será considerado finito.

Definición 1: Sea P un conjunto. Una relación binaria R definida en P es un subconjunto de $P \times P$. Así, $x, y \in P$ se dicen relacionados por R , xRy , si $(x, y) \in R$.

Definición 2: Sea P un conjunto, R una relación en P y x, y, z cualesquiera elementos de P . R es una relación:

- Reflexiva si xRx
- Simétrica si xRy implica yRx
- Transitiva si xRy y yRz implica xRz
- Antisimétrica si xRy y yRx implica $x = y$

Definición 3: Sea P un conjunto, R es una relación de orden parcial de P si es reflexiva, antisimétrica y transitiva.

Definición 4: Sea P un conjunto y S un subconjunto de P .

- Decimos que un elemento $x \in P$ es *cota superior* de S si $y \leq x, \quad \forall y \in S$. Similarmente, un elemento $x \in P$ es *cota inferior* de S si $x \leq y, \quad \forall y \in S$.

- El conjunto de todas las cotas superiores de $S \subset P$ lo denotamos con S^u , esto es, $S^u = \{x \in P \mid y \leq x \forall y \in S\}$. Análogamente, $S^l = \{x \in P \mid x \leq y \forall y \in S\}$, es el conjunto de cotas inferiores.
- $a \in S$ es el elemento *máximo* (\top) de S , si existe, cuando para todo $x \in S$, $x \leq a$. Dualmente tenemos el *mínimo* (\perp).
- Para $S \subset P$, también definimos, si existe, la *cota mínima superior* o $sup(S)$, como el mínimo de S^u y la *cota máxima inferior* o $inf(S)$, como el máximo de S^l . Particularmente para dos elementos $x, y \in P$, escribimos $sup(x, y)$ mediante $x \vee y$, e $inf(x, y)$ mediante $x \wedge y$.

Definición 5: Sea P un conjunto ordenado. Si $x \wedge y$ y $x \vee y$ existen para toda pareja $x, y \in P$, llamamos a P *lattice* (o *retícula*).

A continuación se expondrán algunas nociones sobre los *lattices* de conceptos.

Un contexto (G, M, φ) es un conjunto de G objetos, un conjunto de M rasgos y una correspondencia de G en M , φ . Se definen así las parejas (A, B) , donde $A \in G$ es llamado extento del concepto, $B \in M$ es llamado intento del concepto. Las parejas (A, B) deben cumplir:

$$A = B' = \{g \in G \mid \varphi(g) = m, \forall m \in B\} \quad (2.1)$$

$$B = A' = \{m \in M \mid \varphi(g) = m, \forall g \in A\} \quad (2.2)$$

En un contexto puede definirse una relación de orden tal que para cada pareja (A_1, B_1) y (A_2, B_2) existe el sup e inf , es decir, es un *lattice*, L .

El extento no es más que el conjunto de todos los objetos que satisfacen las características del intento. De la Tabla 2.1, de la sección 2.2 se podría dar el siguiente ejemplo de extento: $\{1, 2, 3, 5, 6\}$ y estos serían todos los objetos que tienen en común el rasgo b o intento $\{b\}$.

Una relación de orden parcial puede ser construida sobre nodos del *lattice* de conceptos. Dados (A_1, B_1) y (A_2, B_2) decimos que $(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2$, el orden precedente significa que (A_1, B_1) es *más particular* que (A_2, B_2) . El diagrama de Hasse del *lattice* puede ser generado usando esta relación de orden parcial: si $(A_1, B_1) \leq (A_2, B_2)$ y no hay otro nodo (A_3, B_3) , tal que $(A_1, B_1) \leq (A_3, B_3) \leq (A_2, B_2)$, hay una arista de (A_1, B_1) a (A_2, B_2) .

En la teoría de conjuntos rugosos, el sistema de información juega un rol similar como contexto en el *lattice* de conceptos. Un sistema de información es

un par ordenado $S = (G, M)$, donde G es un conjunto finito no-vacío llamado el universo, M es un conjunto de atributos finito no-vacío. Los elementos del universo son llamados objetos.

La aproximación rugosa de $A \subset G$ con respecto a $B \subset M$ es la pareja $(B_{\bullet}(A), B^{\bullet}(A))$, donde $B^{\bullet}(A)$ es la aproximación superior de A , la cual consiste en todos los objetos posiblemente pertenecientes a A ; y $B_{\bullet}(A)$ es la aproximación inferior, que consiste en todos los objetos definitivamente pertenecientes a A . En los sucesivos hablaremos genéricamente de aproximaciones para referirnos a algunas de las anteriores.

El *lattice* de conceptos incluye todos los conceptos del contexto, y los integra en un concepto virtual de jerarquía mientras que el agrupamiento rugoso provee interesante mecanismo de aproximación.

La evaluación de los resultados de las aproximaciones superior e inferior con respecto a precisión y evocación se definen a continuación (C.J. van Rijsbergen [8]).

Definición 6: La precisión (P) viene dada por las siguientes relaciones:

$$P = R/NRes \quad (2.3)$$

donde,

R : es la cantidad de respuestas correctas obtenidas por la aproximación.

$NRes$: es la cantidad de resultados obtenidos por la aproximación.

Definición 7: La evocación o *recall* (E) viene dada por la relación:

$$E = R/RC \quad (2.4)$$

donde,

RC : es la cantidad de respuestas correctas para una consulta (ver tabla 4.4).

2.1. Aproximación Superior e Inferior

Aplicando las nociones básicas de la sección anterior y agregando las siguientes definiciones y teoremas de la teoría de *lattices*, llegaremos a la aproximación propuesta.

Definición 8: Un elemento x ($x \neq \top$) es *inf-irreducible* en un *lattice* L si para cada $y, z \in L$, $x = y \wedge z$ implica $x = y$ o $x = z$; dualmente, un

elemento x ($x \neq \perp$) es *sup-irreducible* en un *lattice* L si para cualesquier $y, z \in L$, $x = y \vee z$ implica $x = y$ o $x = z$.

Definición 9: Dado el contexto (G, M, φ) y correspondiente *lattice* de conceptos L , se define $gR = \{m \in M \mid \varphi(g) = m\}$ para $g \in G$.

Dado lo anterior, se tiene la siguiente definición.

Definición 10: Una relación binaria J en G es definida como: $g_1 J g_2$ si y sólo si $g_1 R \subseteq g_2 R$, donde $g_1, g_2 \in G$.

Claramente, J es reflexivo, anti-simétrico y transitivo. De esta manera, J es una relación de orden parcial en G . Denotamos las cotas superiores de g como la clase de orden parcial, $[g]$; específicamente, $[g] = \{g^* \in G : g J g^*\}$.

Definición 11: Sea SI el conjunto de todos los elementos *sup-irreducibles* en L .

P es el conjunto de todos los pares $([g], [g]')$. En Keyun Hu et. al.[1] se prueba que $P = SI$

Similarmente, dado $Rm = \{g \in G \mid \varphi(g) = m\}$, para un contexto (G, M, φ) , se tiene:

Definición 12: Una relación binaria K sobre M es definida como: $m_1 K m_2$ si y solo si $Rm_1 \subseteq Rm_2$, donde $m_1, m_2 \in M$.

Denotamos la clase parcial de m como $[m]$, específicamente, $[m] = \{m^* \in M : m K m^*\}$.

Sea II , el conjunto de todos los elementos *inf-irreducibles* de L .

Q es el conjunto de todos los pares $([m]', [m])$. Igualmente, en Keyun Hu et. al.[1] se prueba que $Q = II$.

Dados $B \subseteq M$, la aproximación inferior (A-Inf) y la aproximación superior (A-Sup) de B con respecto a L son:

$$B_\bullet = \cap \{B \subseteq B^* \mid (B^*, B) \in II\} \quad (2.5)$$

y

$$B^\bullet = \cup \{B^* \subseteq B \mid (B^*, B) \in SI\} \quad (2.6)$$

Las mejores aproximaciones inferiores y superiores están dadas por (B'_\bullet, B_\bullet) y $(B^{\bullet'}, B^\bullet)$.

2.2. Ejemplo

Lo anteriormente expuesto será explicado a continuación mediante un ejemplo tomado de B.Ganter /R.Wille[4]. En este ejemplo tenemos los ob-

jetos, $G = \{1, 2, 3, 4, 5, 6, 7, 8\}$ y los atributos, $M = \{a, b, c, d, e, f, g, h, i\}$, como se muestra en la Tabla 2.1.

φ		a	b	c	d	e	f	g	h	i
1	Sanguijuela	1	1	0	0	0	0	1	0	0
2	Pez	1	1	0	0	0	0	1	1	0
3	Rana	1	1	1	0	0	0	1	1	0
4	Perro	1	0	1	0	0	0	1	1	1
5	Yerba mala	1	1	0	1	0	1	0	0	0
6	bejuco	1	1	1	1	0	1	0	0	0
7	Frijol	1	0	1	1	1	0	0	0	0
8	Maíz	1	0	1	1	0	1	0	0	0

Tabla 2.1 Correspondencia φ de G en M para el ejemplo.

Las características representan:

a=necesita agua, b=vive en agua, c=vive en tierra

d=necesita clorofila, e=dos vainas para semillas

f=una vaina para semillas, g=puede moverse por los alrededores

h=tiene extremidades, i=amamanta a sus crías.

Las clases de orden parcial de objetos y atributos pueden ser computados como se muestra en las tablas 2.2 y 2.3.

Clase	Elementos	Elem. del SI
[1]	{1, 2, 3}	({1, 2, 3}, {a, b, g})
[2]	{2, 3}	({2, 3}, {a, b, g, h})
[3]	{3}	({3}, {a, b, c, g, h})
[4]	{4}	({4}, {a, c, g, h, i})
[5]	{5, 6}	({5, 6}, {a, b, d, f})
[6]	{6}	({6}, {a, b, c, d, f})
[7]	{7}	({7}, {a, c, d, e})
[8]	{6, 8}	({6, 8}, {a, c, d, f})

Figura 2.2 Elementos del SI

Clase	Elementos	Elem. del II
$[a]$	$\{a\}$	$(\{1, 2, 3, 4, 5, 6, 7, 8\}, \{a\})$
$[b]$	$\{a, b\}$	$(\{1, 2, 3, 5, 6\}, \{a, b\})$
$[c]$	$\{a, c\}$	$(\{3, 4, 6, 7, 8\}, \{a, c\})$
$[d]$	$\{a, d\}$	$(\{5, 6, 7, 8\}, \{a, d\})$
$[e]$	$\{a, c, d, e\}$	$(\{7\}, \{a, c, d, e\})$
$[f]$	$\{a, d, f\}$	$(\{5, 6, 8\}, \{a, d, f\})$
$[g]$	$\{a, g\}$	$(\{1, 2, 3, 4\}, \{a, g\})$
$[h]$	$\{a, g, h\}$	$(\{2, 3, 4\}, \{a, g, h\})$
$[i]$	$\{a, c, g, h, i\}$	$(\{4\}, \{a, c, g, h, i\})$

Figura 2.3 Elementos del II

Como aplicación planteamos el siguiente problema: se desea encontrar la mejor aproximación en el *lattice* dado el conjunto de atributos $B = \{g, h\}$. Debe entonces calcularse $(B_{\bullet}$ y $B^{\bullet})$, es decir, cuáles objetos aproximan a B .

La aproximación inferior se obtiene primeramente dando el conjunto de objetos que cuentan con el atributo g y con el h e intersectando el resultado de ambos, ecuación (2.5). El atributo g está como característica de los objetos 1, 2, 3 y 4; y h lo tienen los objetos 2, 3 y 4. Siendo su aproximación inferior el conjunto de objetos $\{2, 3, 4\}$.

La aproximación superior es obtenida uniendo los objetos en los que su conjunto de atributos esté contenido en la consulta dada (B), ecuación (2.6). En este caso no hay aproximación superior (\emptyset).

Por otra parte podría determinarse las aproximaciones superior e inferior dado un conjunto de objetos, A , A^{\bullet} y A_{\bullet} , respectivamente. Aunque ésta forma no la utilizaremos en nuestra aplicación (pues nos interesan las búsquedas en diccionarios onomasiológicos: dados los atributos obtener los objetos).

Por ejemplo, el conjunto $A = \{2, 4\}$, sus aproximaciones superior e inferior son computadas como:

$$\begin{aligned}
 A^{\bullet} &= \cap \{A \subseteq A^* \mid (A^*, A^*) \in II\} = & (2.7) \\
 &= \cap \{\{1, 2, 3, 4, 5, 6, 7, 8\}, \{1, 2, 3, 4\}, \{2, 3, 4\}\} = \{2, 3, 4\}
 \end{aligned}$$

$$\begin{aligned}
 A_{\bullet} &= \cup \{A^* \subseteq A \mid (A^*, A^*) \in SI\} = & (2.8) \\
 &= \cup \{4\} = \{4\}
 \end{aligned}$$

Capítulo 3

Aplicación en un Diccionario Onomasiológico

El problema es que, dada la descripción de algún concepto obtener las palabras clave cuya definición se aproxime a la descripción dada. Nuestra propuesta consiste en utilizar el método anterior y aplicarlo en Sistemas de Recuperación de Información. En este caso se realizó la prueba con un listado de conceptos (glosario) de la Física¹, obtenido de los trabajos que desarrolla G.Sierra /J.McNaught[2].

Un glosario está formado por una lista de registros, cada uno está compuesto por el nombre del concepto o vocablo (D), que sería la entrada del diccionario, y a continuación el conjunto de palabras clave (T), que describen dicho vocablo. El vocabulario utilizado estará formado por las palabras clave que describen a todos los vocablos del glosario, eliminando de éste las palabras cerradas.

Las clases parciales pertenecientes a los elementos de II se obtienen de una tabla inversa donde se representan por cada palabra clave de la definición, T_i , todos los vocablos donde ésta aparece. Las clases que pertenecen a los SI , por otra parte, se obtienen de una tabla donde aparece cada vocablo D_j representado por las palabras clave que componen su definición.

Partiendo de estas dos tablas se realizan las aproximaciones superior e inferior dada una consulta. Para la aproximación superior se utiliza la tabla del SI y para la aproximación inferior la tabla del II .

Estas aproximaciones se obtienen realizando las siguientes operaciones:

¹Recurso amablemente brindado por el Grupo de Ingeniería Lingüística de la UNAM.

Aproximación Inferior por atributo o palabra clave (T_{\bullet}).

$$T_{\bullet} = \cap \{T \subseteq T^* \mid (T^{*'}, T^*) \in II\} \quad (3.1)$$

Más detalladamente se realiza lo siguiente. En primer lugar hay que partir de una tabla inversa (II), que no es otra cosa que un listado de palabras clave -las cuales conforman el vocabulario del *corpus*- y asociados a cada una de estas palabras los vocablos en cuya definición ocurren dichas palabras clave. Luego, dada una consulta, que a su vez debe tener las mismas características que el texto (es decir, truncado, sin palabras repetidas y sin palabras cerradas), para cada palabra clave que compone dicha consulta se determina en qué definiciones de vocablos aparece. El próximo paso será traslapar el resultado anterior y la Aproximación Inferior (ecuación 3.1) estará dada por aquellos vocablos que contengan en su descripción a las palabras clave de la consulta.

Aproximación Superior por atributo o palabra clave (T^{\bullet}).

$$T^{\bullet} = \cup \{T^* \subseteq T \mid (T^{*'}, T^*) \in SI\} \quad (3.2)$$

Donde,

- T son las palabras clave que conforman la consulta,
- T^* son las palabras clave que pertenecen al *inf-irreducible*,
- $T^{*'}$ son aquellos vocablos que contienen las palabras clave T^* ,

A continuación se explicará con más detalle el método empleado para obtener dicha aproximación. Primeramente se parte de una tabla de vocablos (SI) representado con sus respectivas definiciones, donde el texto se encuentra truncado, sin palabras repetidas y sin palabras cerradas. La aproximación superior (ecuación 3.2) es la unión del conjunto de vocablos en el que las palabras clave de su definición están contenidas en la consulta.

Teniendo en cuenta que en muy pocos casos se van cumplir las contenciones exactas en ambas aproximaciones, se introduce el criterio de *mayoría*. Éste consiste en aceptar como válido un vocablo donde las palabras clave que lo definen aparezcan en la mayoría de los casos (la mitad del total más uno) en la consulta, para la Aproximación Superior. En el caso de la Aproximación Inferior, ésta se refiere a que la mayoría de las palabras clave de la consulta estén contenidas en la definición de un vocablo.

Puede suceder que no se utilicen en la consulta las palabras clave que aparecen en las definiciones de los vocablos sino algún sinónimo de éstas, y es por ello que en una de las pruebas se hará la representación de las palabras clave, tanto de las que componen las definiciones como de la consulta, con un diccionario de sinónimos.²

En la tercera prueba se utilizará un *corpus* del área de la Física para expandir, tanto los glosarios como las consultas, con el fin de aumentar la evocación sin que se afecte mucho la precisión de los resultados obtenidos en las pruebas anteriores. Dicha expansión es realizada de la siguiente manera: se buscan las palabras clave de las definiciones, así como las de la consulta, en las entradas del *corpus*, si éstos coinciden se le agrega a la palabra clave de la definición, o de la consulta, el texto relacionado con ellos que aparece en el *corpus*. Se espera que la expansión mediante el *corpus* mejore los resultados de A-Inf y A-Sup ya que el usuario puede buscar un concepto por temas relacionados a éste, por ejemplo, en Física la *fuerza* está relacionada con la aceleración de un cuerpo (masa) y a su vez la *aceleración* está relacionada con cambio de velocidad con el tiempo; entonces, masa, velocidad, aceleración, son temas relacionados con la *fuerza*.

²Ofrecido amablemente por el Grupo de Ingeniería Lingüística de la UNAM, G.Sierra /J.McNaught[2].

Capítulo 4

Pruebas

En este capítulo se expondrá cómo fueron realizadas las diferentes pruebas de los programas que se utilizaron para obtener las aproximaciones superior e inferior.

Para estas pruebas se tuvo en cuenta los siguientes datos de entrada, primeramente contamos con dos glosarios del área de Física¹, uno de ellos tiene una definición para cada vocablo y el otro tres (en los apéndices A.1 y A.2 se encuentra una selección de los glosarios de una y tres definiciones, respectivamente). Otro dato de entrada es la consulta, la cual primeramente se buscó que diera un resultado evidente. Las consultas iniciales fueron definiciones de vocablos que aparecían en los glosarios para así conocer con exactitud que resultados obtendríamos. También fue utilizado un diccionario de sinónimos, y un *corpus* de la misma área que los glosarios. En cuanto al diccionario de sinónimos y el *corpus*, éstos sólo se usaron en una prueba cada uno, mientras que los glosarios y las consultas se utilizaron en todas las pruebas, con el fin de poder comparar los resultados de A-sup y A-Inf utilizando los diferentes métodos (ver tabla 4.3).

Las pruebas se llevaron a cabo en diferentes grados de complejidad, como explicaremos a continuación:

1. Programas individuales.
 - La eliminación de palabras cerradas de los glosarios se efectúa por:

```
./palcerr lispalcerr.txt glosario1.txt > glosar1.txt  
./palcerr lispalcerr.txt glosario3.txt > glosar3.txt
```

¹G.Sierra /J.McNaught[2].

donde:

palcerr: programa que elimina las palabras cerradas (por ejemplo, "la", "el", "del", "esta", "este", etc.) de las definiciones de los glosarios.

lispalcerr.txt: Listado de palabras cerradas.

glosario1.txt: Glosario con una definición por cada vocablo.

glosar1.txt: Glosario con una definición por cada vocablo sin palabras cerradas.

glosario3.txt: Glosario con tres definiciones por vocablo.

glosar3.txt: Glosario con tres definiciones por vocablo sin palabras cerradas.

- La eliminación de palabras repetidas por entrada en los glosarios se obtiene mediante los comandos:

```
./palrep glosar1.txt > glosa1.txt
```

```
./palrep glosar3.txt > glosa3.txt
```

donde,

palrep: programa que elimina las palabras repetidas por cada entrada de los glosarios

glosa1.txt: Glosario con una definición por vocablo, sin palabras cerradas ni repetidas.

glosa3.txt: Glosario con tres definiciones por vocablo, sin palabras cerradas ni repetidas.

- El truncamiento consiste en obtener las seudoraíces de las palabras claves que componen tanto los glosarios como las consultas y sustituir éstas seudoraíces por las palabras clave. Con este paso se termina de obtener los *SI* para una y tres definiciones por vocablo utilizados en la prueba del método directo de la sección 4.1.2. El truncador usado está disponible en el Laboratorio de Recuperación de Información de la FCC de la BUAP ². El truncamiento de glosarios y consultas se realiza ejecutando:

```
./trunc glosa1.txt > SI1.txt
```

```
./trunc glosa3.txt > SI3.txt
```

```
./trunc Consulta.txt > C.txt
```

²Se trata de un truncador para palabras del español basado en el algoritmo de Prter (R.Aceves et. al. [10])

donde,

trunc: programa que trunca las palabras de los glosarios y las consultas.

SI1.txt: glosario con una definición por vocablo cuyas palabras clave de estas definiciones están truncadas, sin palabras repetidas y sin palabras cerradas.

SI3.txt: glosario con tres definiciones por vocablo cuyas palabras clave de estas definiciones están truncadas, sin palabras repetidas y sin palabras cerradas.

Consulta.txt: Archivo de consultas (son cinco consultas diferentes).

C.txt: Archivo de consultas truncadas (ver Tabla 4.4).

- Comandos para la obtención del vocabulario:

```
./vocab SI1.txt > voc1.txt
```

```
./vocab SI3.txt > voc3.txt
```

donde,

vocab: programa que obtiene las palabras clave de las definiciones de todo el glosario.

voc1.txt: Vocabulario con las palabras clave de las definiciones de *SI1.txt*.

voc3.txt: Vocabulario con las palabras clave de las definiciones de *SI3.txt*.

- Los comandos para obtener la representación de glosarios y consultas mediante el diccionario de sinónimos son:

```
./equiv-c sinon.txt C.txt > Ceq.txt
```

```
./equiv-t sinon.txt SI1.txt > SI1eq.txt
```

```
./equiv-t sinon.txt SI3.txt > SI3eq.txt
```

donde,

equiv-c: programa que sustituye cada palabra de *C.txt* por el identificador de la clase de sinónimos a la que pertenece. En caso de no encontrar dicho sinónimo se mantiene la palabra clave que aparece en la consulta.

equiv-t: programa que sustituye cada palabra de los *SI1.txt* y *SI3.txt* por su sinónimo representativo.

sinon.txt: Diccionario de sinónimos, que consta de sinónimo representante con un listado de sinónimos (clase).

Ceq.txt: Igual que *C.txt* pero sus palabras clave son sinónimos tomadas de *sinon.txt*.

SI1eq.txt: Igual que *SI1.txt* pero sus palabras clave son sinónimos tomadas de *sinon.txt*.

SI3eq.txt: Igual que *SI3.txt* pero sus palabras clave son sinónimos tomadas de *sinon.txt*.

- Con los siguientes comandos se efectúa la expansión de glosarios y consultas mediante un *corpus*:

```
./expcorpus C.txt corpus.txt > C-corpus.txt
```

```
./expcorpus SI1.txt corpus.txt > SI1-corpus.txt
```

```
./expcorpus SI3.txt corpus.txt > SI3-corpus.txt
```

donde,

expcorpus: programa que expande *C.txt*, *SI1.txt* y *SI3.txt* mediante un *corpus*. Busca cada palabra clave de lo que se desea expandir en la entrada del *corpus* y se expande a dicha palabra clave *T*, con sus relaciones de sentido, H.Jiménez /G.Morales[6], (las palabras clave que constituyan las oraciones donde ocurra *T* en el *corpus*).

corpus.txt: *corpus* del área de la Física (Mecánica y Cinemática).

C-corpus.txt: Es el *C.txt* pero expandido mediante el *corpus*.

SI1-corpus.txt: Es el *SI1.txt* pero expandido mediante el *corpus*.

SI3-corpus.txt: Es el *SI3.txt* pero expandido mediante el *corpus*.

- La obtención del *II* se obtiene ejecutando:

```
./terdoc SI1.txt > II1.txt
```

```
./terdoc SI3.txt > II3.txt
```

```
./terdoc SI1eq.txt > II1eq.txt
```

```
./terdoc SI3eq.txt > II3eq.txt
```

```
./terdoc SI1-corpus.txt > II1-corpus.txt
```

```
./terdoc SI3-corpus.txt > II3-corpus.txt
```

donde,

terdoc: programa que obtiene una tabla donde cada renglón indica, por cada palabra clave, la lista de vocablos cuya definición conduce a la palabra clave.

II1.txt: Son los elementos del *II* utilizados para obtener la A-Inf en el método directo; para una definición por vocablo.

II3.txt: Son los elementos del *II* utilizados para obtener la A-Inf en el método directo; para tres definiciones por vocablo.

II1eq.txt: Son los elementos del *II* utilizados para obtener la A-Inf usando un diccionario de sinónimos; para una definición por vocablo.

II3eq.txt: Son los elementos del *II* utilizados para obtener la A-Inf usando un diccionario de sinónimos; para tres definiciones por vocablo.

II1-corpus.txt: Son los elementos del *II* utilizados para obtener la A-Inf usando un *corpus*; para una definición por vocablo.

II3-corpus.txt: Son los elementos del *II* utilizados para obtener la A-Inf usando un *corpus*; para tres definiciones por vocablo.

2. Grupos de programas.

- Para obtener las A-Inf se ejecutó:

`./a-inf $1 $2 $3` cuya definición es:

`./relac $1 $2 | ./inters - > $3`

donde,

`relac`: obtiene una lista de vocablos para cada palabra clave de la consulta.

`inters`: hace la intersección de los resultados obtenidos en `relac`.

Los parámetros *\$1*, *\$2* y *\$3* vienen en la tabla 4.1.

Método	Glos.	\$1	\$2	\$3
Directo	1	<i>C.txt</i>	<i>II1.txt</i>	<i>aprox-inf1.txt</i>
Directo	3	<i>C.txt</i>	<i>II3.txt</i>	<i>aprox-inf3.txt</i>
Sinónimo	1	<i>Ceq.txt</i>	<i>II1eq.txt</i>	<i>aprox-inf1eq.txt</i>
Sinónimo	3	<i>Ceq.txt</i>	<i>II3eq.txt</i>	<i>aprox-inf3eq.txt</i>
<i>corpus</i>	1	<i>C-corpus.txt</i>	<i>II1-corpus.txt</i>	<i>aprox-inf1-corpus.txt</i>
<i>corpus</i>	3	<i>C-corpus.txt</i>	<i>II3-corpus.txt</i>	<i>aprox-inf3-corpus.txt</i>

Tabla 4.1 Valores de los parámetros *\$1*, *\$2* y *\$3* utilizados para obtener las A-Inf.

- Para obtener las A-Sup se ejecutó:

`./a-sup $1 $2 > $3`

donde,

`a-sup`: obtiene la unión de todos aquellos vocablos en los cuales la mayoría de las palabras clave de su definición estén contenidas en la consulta.

En la tabla 4.2 se presentan los valores de los parámetros $\$1$, $\$2$ y $\$3$.

Método	Glos.	$\$1$	$\$2$	$\$3$
Directo	1	<i>C.txt</i>	<i>SI1.txt</i>	<i>aprox-sup1.txt</i>
Directo	3	<i>C.txt</i>	<i>SI3.txt</i>	<i>aprox-sup3.txt</i>
Sinónimo	1	<i>Ceq.txt</i>	<i>SI1eq.txt</i>	<i>aprox-sup1eq.txt</i>
Sinónimo	3	<i>Ceq.txt</i>	<i>SI3eq.txt</i>	<i>aprox-sup3eq.txt</i>
<i>corpus</i>	1	<i>C-corpus.txt</i>	<i>SI1-corpus.txt</i>	<i>aprox-sup1-corpus.txt</i>
<i>corpus</i>	3	<i>C-corpus.txt</i>	<i>SI3-corpus.txt</i>	<i>aprox-sup3-corpus.txt</i>

Tabla 4.2 Valores de los parámetros $\$1$, $\$2$ y $\$3$ utilizados para obtener las A-Sup.

Las pruebas de grupos de programas anteriores (obtención de las A-Inf y las A-Sup) se explican más detalladamente en las siguientes secciones de este capítulo. En la tabla 4.3 se muestran específicamente en qué secciones podemos encontrar dichas pruebas en dependencia del método empleado.

Método	Aprox. Inferior	Aprox. Superior
<i>Directo</i>	Sección 4.1.1	Sección 4.1.2
<i>Sinónimos</i>	Sección 4.2.1	Sección 4.2.2
<i>Corpus</i>	Sección 4.3.1	Sección 4.3.2

Tabla 4.3 Pruebas realizadas y su ubicación en este trabajo.

A continuación se muestran los resultados de aplicar los diferentes métodos explicados, partiendo de dos glosarios, las consultas, un diccionario de sinónimos y un *corpus* y utilizando los programas descritos anteriormente. Cada uno de los métodos se organizó en cuatro partes: descripción, comando, resultado y comentario.

4.1. Aplicación directa del método

Descripción:

Esta primera prueba fue realizada con el contexto de conceptos del área de la Física. Para esto se usaron dos glosarios, uno en el cual para cada vocablo había una definición; y en el segundo, a cada vocablo le correspondían como máximo 3 definiciones. Las palabras que componen nuestro glosario se encuentran, como se ha explicado, truncadas de las que restaron al ser eliminadas las cerradas, los signos de puntuación y las palabras repetidas del texto original.

Las consultas utilizadas versaron sobre nociones de Mecánica y Cinemática dadas por profesores de Ingeniería, estudiantes de computación y, otras, extraídas del libro de L.Landau /E.Lifshitz[3], las cuales se presentan en la Tabla 4.4.

Para visualizar mejor el criterio de mayoría utilizado en este método emplearemos las expresiones de las ecuaciones 4.1 y 4.2 en los resultados de la A-Inf y A-Sup, respectivamente.

$$M_{inf} = Np/NC \quad (4.1)$$

Donde,

M_{inf} : evaluación de mayoría en la A-Inf.

Np : cantidad de palabras clave de la consulta que se encuentran en la definición de un vocablo.

NC : la cantidad de palabras de la consulta.

$$M_{sup} = Np/ND \quad (4.2)$$

Donde,

M_{sup} : evaluación de mayoría en la A-Sup.

ND : la cantidad de palabras clave de la definición del vocablo.

	Consulta (truncadas)	Resp. Correcta (RC)
C1	tercer planet sistem sol	<i>Tierra</i>
C2	choqu conserv cantidad movimient lineal energi	<i>Choque elástico</i>
C3	aplic fuerz provoc desplaz mism direccion	<i>Energía mecánica</i>
C4	movimient sistem describ line rect	<i>Movimiento lineal Movimiento rectilíneo (Rectilíneo)</i>
C5	cambi veloc tiemp	<i>Aceleración (Aceleración angular y Aceleración centrípeta)</i>

Tabla 4.4 Consultas utilizadas en la primera prueba.

Entre las respuestas correctas de la Consulta C5 se encuentran entre paréntesis los hipónimos de *Aceleración*, mientras que la consulta C4 tiene tres posibles respuestas correctas, dos de ellas son sinónimos y la que se encuentra entre paréntesis es un hiperónimo de las anteriores.

4.1.1. Aproximación Inferior

Comando:

1. La A-Inf, usando el glosario de una definición por vocablo, se obtiene ejecutando:

```
./relac C.txt II1.txt | ./inters - > aprox-inf1.txt
```

2. La A-Inf, usando el glosario de tres definiciones por vocablo, se obtiene ejecutando:

```
./relac C.txt II3.txt | ./inters - > aprox-inf3.txt
```

donde,

relac: Mencionado en las pruebas de grupos de programas de este capítulo.

C.txt: archivo de las Consultas (C1,...,C5). Ver figura 4.

inters: Mencionado en las pruebas de grupos de programas de este capítulo.

III.txt y *II3.txt*: se refiere a las clases parciales pertenecientes al *II*; para glosarios con una definición por vocablo y para tres definiciones respectivamente.

aprox-inf1.txt y *aprox-inf3.txt*: resultado de la aproximación inferior; para glosarios con una definición por vocablo y para tres definiciones respectivamente.

Resultados:

Los resultados obtenidos en esta prueba se muestran a continuación. En la Tabla 4.5 aparece la aproximación inferior utilizando el glosario que cuenta con una definición por cada vocablo y la Tabla 4.6 con el glosario de tres definiciones por vocablo. Cada uno de estos resultados fueron obtenidos utilizando el comando antedicho para las consultas mostradas en la figura 4.4.

Consulta	Aprox. Inf. (Np/NC)
C1	Tierra (4/4) y Sol (3/4)
C2	No tiene
C3	Energía mecánica (4/6)
C4	Movimiento lineal (4/5) y Sistema de referencia inercial (3/5)
C5	Aceleración (3/3), Choque (3/3), Aceleración angular (3/3), Aceleración centrípeta (2/3), Estado de movimiento (2/3), Dilatación del tiempo (2/3), Choque elástico (2/3), Choque inelástico (2/3), Velocidad (2/3), Velocidad instantánea (2/3), Teoría electromagnética de la luz (2/3), Teoría de la relatividad (2/3), Movovimiento uniformemente retardado (2/3), Momento de torsión (2/3) y Marea (2/3)

Tabla 4.5 Resultados de la A-Inf con el método directo, usando el glosario de una definición por cada vocablo.

Las respuestas correctas fueron señaladas en itálica para resaltarlas con respecto al resto de los resultados de la A-Inf.

Comentarios:

En el caso especial de la consulta C2, la cual no tiene resultado en la Tabla 4.5 y en la 4.6, se puede decir que en ocasiones el usuario no utiliza una descripción como las que aparecen en el glosario y, por lo tanto, no se encuentra la respuesta deseada. Esta consulta en particular fue obtenida en el libro L.Landau /E.Lifshitz[3].

Con respecto a los resultados obtenidos en las Tablas 4.5 y 4.6 se puede decir que la evocación aumentó en algunos casos (particularmente en la A-Inf de la consulta C5) y que la precisión se mantuvo igual. Teniendo en cuenta lo anterior, y con el objetivo de buscar que la precisión y la evocación mejoraran se introdujo un diccionario de sinónimos para representar las palabras clave de las consultas y de las definiciones.

Consulta	Aprox. Inf. (Np/NC)
C1	<i>Tierra</i> (4/4) y <i>Sol</i> (3/4)
C2	No tiene
C3	<i>Energía mecánica</i> (4/6)
C4	<i>Movimiento lineal</i> (4/5) Sistema de referencia inercial (3/5)
C5	<i>Aceleración</i> (3/3), <i>Choque</i> (3/3), <i>Aceleración angular</i> (3/3), <i>Aceleración centrípeta</i> (2/3), Estado de movimiento (2/3), <i>Dilatación del tiempo</i> (2/3), <i>Choque elástico</i> (2/3), <i>Choque inelástico</i> (2/3), <i>Velocidad</i> (2/3), <i>Velocidad instantánea</i> (2/3), <i>Teoría electromagnética de la luz</i> (2/3), <i>Teoría de la relatividad</i> (2/3), <i>Movimiento uniformemente retardado</i> (2/3), <i>Momento de torsión</i> (2/3), <i>Marea</i> (2/3) <i>Energía cinética de rotación</i> (2/3) <i>Día solar</i> (2/3), <i>Oscilación</i> (2/3) y <i>Frecuencia</i> (2/3)

Tabla 4.6 Resultados de la A-Inf con el método directo, usando el glosario de tres definición por cada vocablo.

4.1.2. Aproximación Superior

Comando:

1. La A-Sup, usando el glosario de una definición por vocablo, se obtiene ejecutando:

```
./a-sup C.txt SI1.txt > aprox-sup1.txt
```

2. La A-Sup, usando el glosario de una definición por vocablo, se obtiene ejecutando:

```
./a-sup C.txt SI3.txt > aprox-sup3.txt
```

donde,

C.txt: archivo de las Consultas (C1,...,C5). Ver Tabla 4.4.

a-sup: Mencionado en las pruebas de grupos de programas de este capítulo.

SI1.txt y *SI3.txt*: clases parciales correspondientes al *SI*, para glosarios con una definición por vocablo y para tres definiciones, respectivamente.

aprox-sup1.txt y *aprox-sup3.txt*: resultado de la aproximación superior, para glosarios con una definición por vocablo y para tres definiciones, respectivamente.

Resultados:

Los resultados obtenidos en esta prueba se muestran a continuación. En la Tabla 4.7 aparecen las aproximación superior utilizando el glosario que cuenta con una definición por cada vocablo y la Tabla 4.8 con el glosario de tres definiciones por vocablo.

Comentarios:

Los resultados obtenidos en esta prueba con respecto a las Aproximación Superiores (A-Sup) fueron los esperados. Para las tres primeras consultas no hubo resultados puesto que es difícil que las descripciones de los vocablos estuvieran contenidas en la consulta.

Consulta	Aprox. Sup. (N_p/ND)
C1	No tiene
C2	No tiene
C3	No tiene
C4	<i>Movimiento lineal (4/7)</i> <i>y rectilíneo (2/3)</i>
C5	<i>Aceleración (3/4) y</i> <i>Aceleración angular (3/5)</i>

Tabla 4.7 Resultados de la A-Sup de la primera prueba usando el glosario de una definición por cada vocablo.

Consulta	Aprox. Sup.
C1	No tiene
C2	No tiene
C3	No tiene
C4	<i>Movimiento lineal (4/7)</i> <i>y rectilíneo (2/3)</i>
C5	<i>Aceleración (3/4) y</i> <i>Aceleración angular (3/5)</i>

Tabla 4.8 Resultados de la primera prueba usando el glosario de tres definición por cada vocablo.

4.2. Usando un diccionario de sinónimos

Descripción:

En esta prueba se utilizó un diccionario de sinónimos³ (41555 Kb) para representar las palabras clave de la consulta y de las definiciones mediante una palabra "equivalente". El diccionario de sinónimos es una lista de registros, cada uno está compuesto por el sinónimo representante (un total de 523) y a continuación el listado de sinónimos de dicho vocablo (en el apéndice A.3 se muestra una parte del diccionario de sinónimos utilizado). La sustitución de las palabras clave, tanto de la consulta como de las definiciones, por su sinónimo representante (en caso que lo tenga) se lleva a cabo de la siguiente forma:

³Se trata de una parte de un diccionario de sinónimos, amablemente proporcionado por el Grupo de Ingeniería Lingüística de la UNAM.

1. Buscar la palabra clave de la consulta o de la definición dentro de las palabras en el listado de sinónimos.
2. Si se encuentra la palabra clave en el listado de sinónimos entonces la palabra clave se sustituye por el sinónimo representante del diccionario de sinónimos.
3. Si no se encuentra la palabra clave en dicho listado de sinónimos se mantiene ésta, ya que es posible que una palabra no tenga sinónimos.

El diccionario original fue preprocesado uniendo todos los sinónimos, que aparecían en éste, para formar la lista. Hay que considerar que los sinónimos de una palabra pueden estar referidos a diferentes contextos, por ejemplo, la palabra *cuerpo* puede tener siete contextos diferentes, como se muestra a continuación:

- cuerpo: tronco
- cuerpo: cadáver
- cuerpo: espesor, grosor, grueso y volumen
- cuerpo: tamaño y grandor
- cuerpo: consistencia y densidad
- cuerpo: soma
- cuerpo: tomo, volumen, totalmente, enteramente, del todo, completamente, espesar, concretarse, materializarse y realizarse

Después de preprocesado el ejemplo anterior nos queda *cuerpo* como vocablo representante y a continuación el listado de todos sus sinónimos. Es decir:

cuerpo: tronco, cadáver, espesor, grosor, grueso, volumen, tamaño, grandor, consistencia, densidad, soma, tomo, totalmente, enteramente, del todo, completamente, espesar, concretarse, materializarse y realizarse.

Es por ello que en algunas consultas no resultan "legibles". Teniendo en cuenta esta representación las consultas quedaron como se muestra en la Tabla 4.9.

4.2.1. Aproximación Inferior

Comando:

1. La A-Inf, usando el glosario de una definición por vocablo, se obtiene ejecutando:

```
./relac Ceq.txt II1eq.txt | ./inters - > aprox-inf1eq.txt
```

2. La A-Inf, usando el glosario de tres definiciones por vocablo, se obtiene ejecutando:

```
./relac Ceq.txt II3eq.txt | ./inters - > aprox-inf3eq.txt
```

donde,

relac: Mencionado en las pruebas de grupos de programas de este capítulo.

Ceq.txt: archivo de las Consultas representadas con sinónimos (*C1eq*,...,*C5eq*). Ver Tabla 4.9.

inters: Mencionado en las pruebas de grupos de programas de este capítulo.

II1eq.txt y *II3eq.txt*: se refiere a las clases parciales pertenecientes al *II* representando las palabras clave con su sinónimo equivalente, para glosarios de una definición por concepto y tres definiciones, respectivamente.

aprox-inf-eq1.txt y *aprox-inf-eq3.txt*: resultado de la aproximación inferior, para glosarios de una definición por concepto y tres definiciones, respectivamente.

Resultados:

Los resultados de la aproximación inferior aparecen en las Tablas 4.10 y 4.11.

	Consulta equivalente (truncada)	Resp. Correcta (RC)
C1eq	tercer planet método sol	<i>Tierra</i>
C2eq	colision manten porción tiemp lineal potenci	<i>Choque elástico</i>
C3eq	entreg virtud produc mov mism sent	<i>Energía mecánica</i>
C4eq	tiemp método describ líne orifici	<i>Movimiento lineal</i> <i>Movimiento rectilíneo</i> <i>(Rectilíneo)</i>
C5eq	vuelto acel period	<i>Aceleración</i> <i>(Aceleración angular y</i> <i>Aceleración centrípeta)</i>

Tabla 4.9 Consultas utilizadas en el método de representación por sinónimos.

Consulta	Aprox. Inf. (Np/NC)
C1eq	<i>Tierra (4/4)</i> y <i>Sol (3/4)</i>
C2eq	No tiene (0/6)
C3eq	<i>Energía mecánica (4/6)</i>
C4eq	<i>Movimiento lineal (4/5)</i> y Sistema de referencia inercial (3/5)
C5eq	<i>Aceleración (3/3)</i> , <i>Aceleración angular (3/3)</i> <i>Aceleración centrípeta, (2/3)</i> <i>Choque (3/3)</i> , <i>Velocidad (2/3)</i> , <i>Marea (2/3)</i> , <i>Estado de movimiento (2/3)</i> , <i>Dilatación</i> <i>del tiempo (2/3)</i> , <i>Choque elástico (2/3)</i> , <i>Choque inelástico (2/3)</i> , <i>Velocidad</i> <i>instantánea (2/3)</i> , <i>Velocidad angular</i> <i>instantánea (2/3)</i> , <i>Teoría</i> <i>electromagnética de la luz (2/3)</i> ,

Tabla 4.10 Resultados de la A-Inf con la representación por sinónimos; usando el glosario de una definición por cada vocablo.

Consulta	Aprox. Inf. (Np/NC)
C5eq	Teoría de la relatividad (2/3), Movovimiento uniformemente retardado (2/3), Movovimiento uniformemente acelerado (2/3), Momento de torsión (2/3), Velocidad media (2/3), Experimento de Michelson y Morley (2/3), y Eventos simultáneos (2/3)

Tabla 4.10 Continuación.

Consulta	Aprox. Inf. (Np/NC)
C1eq	Tierra (4/4) y Sol (3/4)
C2eq	No tiene (0/6)
C3eq	<i>Energía mecánica</i> (4/6)
C4eq	<i>Movimiento lineal</i> (4/5) y Sistema de referencia inercial (3/5)
C5eq	<i>Aceleración</i> (3/3), <i>Aceleración angular</i> (3/3) <i>Aceleración centrípeta</i> , (2/3) Choque (3/3), Velocidad (2/3), Marea (2/3), Estado de movimiento (2/3), Dilatación del tiempo (2/3), Choque elástico (2/3), Choque inelástico (2/3), Velocidad instantánea (2/3), Velocidad angular instantánea (2/3), Teoría electromagnética de la luz (2/3), Teoría de la relatividad (2/3), Movimiento uniformemente retardado (2/3), Movimiento uniformemente acelerado (2/3), Momento de torsión (2/3), Velocidad media (2/3), Experimento de Michelson y Morley (2/3), Eventos simultáneos (2/3), Energía cinética de rotación (2/3), Día solar (2/3), Oscilación (2/3), Frecuencia (2/3), Velocidad angular (2/3), Velocidad tangencial (2/3) y Momento angular (2/3)

Tabla 4.11 Resultados de la A-Inf con representación por

sinónimos; usando el glosario de tres definiciones por cada vocablo.

Comentarios:

Los resultados de esta prueba son muy similares a los de la sección 4.1.1, pero la evocación y la precisión aumentaron. Ya que una misma consulta puede ser expresada de forma diferente por varios usuarios, y lo que difiere entre estas consultas es que se usan sinónimos para expresar la misma idea.

4.2.2. Aproximación Superior

Comando:

1. La A-Sup, usando el glosario de una definición por vocablo, se obtiene ejecutando:

```
./a-sup Ceq.txt SI1eq.txt > aprox-sup1eq.txt
```

2. La A-Sup, usando el glosario de una definición por vocablo, se obtiene ejecutando:

```
./a-sup Ceq.txt SI3eq.txt > aprox-sup3eq.txt
```

donde,

Ceq.txt: archivo de las Consultas representadas con sinónimos (C1eq,...,C5eq). Ver Tabla 4.9.

a-sup: Mencionado en las pruebas de grupos de programas de este capítulo.

SI1eq.txt y *SI3eq.txt*: clases parciales correspondientes al *SI* representando las palabras clave con su vocablo representante según el diccionario de sinónimos, para glosarios con una definición por vocablo, y tres definiciones, respectivamente.

aprox-sup-eq1.txt y *aprox-sup-eq3.txt*: resultados de la aproximación superior, para glosarios con una definición por vocablo, y tres definiciones, respectivamente.

Resultados:

Los resultados de la aproximación superior aparecen en las Tablas 4.12 y 4.13.

Consulta	Aprox. Sup. (N_p/ND)
C1eq	No tiene
C2eq	No tiene
C3eq	No tiene
C4eq	<i>Movimiento lineal (4/7)</i> <i>y rectilíneo (3/2)</i>
C5eq	<i>Aceleración (3/4) y</i> <i>Aceleración angular (3/5)</i>

Tabla 4.12 Resultados de la A-Sup con representación por sinónimos; usando el glosario de una definición por cada vocablo.

Consulta	Aprox. Sup. (N_p/ND)
C1eq	No tiene
C2eq	<i>Choque elástico (3/4)</i>
C3eq	Oscilador armónico simple (3/5)
C4eq	<i>Movimiento lineal (4/7)</i> <i>y rectilíneo (3/2)</i>
C5eq	<i>Aceleración (3/4) y</i> <i>Aceleración angular (3/5)</i>

Tabla 4.13 Resultados de la A-Sup con representación por sinónimos; usando el glosario de tres definiciones por cada vocablo.

Comentarios:

Para el caso de la consulta C2 de las Tablas 4.5, 4.6, 4.7 y 4.8, no se encontraron resultados en la A-Inf y la A-Sup, sin embargo en esta segunda prueba para el caso de la consulta C2eq se obtuvo en la A-Sup la respuesta esperada cuando se utilizó el glosario de tres definiciones por vocablo.

4.3. Usando un *corpus* del área de Física

Descripción:

Para esta prueba se utilizó un *corpus* del área de la Física, específicamente, Mecánica y Cinemática (tomados de la página de Internet: "Física con Ordenador" [5]) con el fin de expandir las consultas y los glosarios. El *corpus* consta de 20891 palabras y tiene un tamaño, después de ser procesado, de 115591 Kb (parte de este *corpus* procesado se muestra en el apéndice A.4). El *corpus* se divide en la entrada del *corpus* y a continuación las relaciones de sentido de dicha entrada.

La expansión de los glosarios y de las consultas (H.Salazar et. al.[9]) se realiza de la siguiente forma:

1. Buscar la palabra clave (del glosario o la consulta) en la entrada del *corpus*.
2. Si la palabra clave se encuentra en la entrada del *corpus* entonces se sustituye dicha palabra clave por el tema relacionado a ésta. Para tener una idea de cuánto se expanden las consultas usando el *corpus* se presenta la Tabla 4.14.
3. Si no se encuentra la palabra clave en la entradas del *corpus* entonces esta se mantiene como estaba.

En la cantidad de palabras clave de las consultas extendidas, al igual que la de los glosarios extendidos, no se consideran palabras repetidas.

Consulta original	Cantidad de palabras	Consulta expandida	Cantidad de palabras
C1	4	C1 – corpus	300
C2	6	C2 – corpus	617
C3	6	C3 – corpus	816
C4	5	C4 – corpus	645
C5	3	C5 – corpus	1051

Tabla 4.14 Cantidad de palabras de las consultas extendidas usando un *corpus* comparadas con la cantidad de palabras de las consultas originales.

Para esta prueba los resultados de A-Inf y la A-sup serán presentados teniendo en cuenta las relaciones para evaluar precisión, P (ecuación 2.3), y evocación, E (ecuación 2.4), definidas en el capítulo 2.

4.3.1. Aproximación Inferior

Comando:

1. La A-Inf, usando el glosario de una definición por vocablo, se obtiene ejecutando:

```
./relac C-corpus.txt II1-corpus.txt | ./inters - > aprox-inf1-corpus.txt
```

2. La A-Inf, usando el glosario de tres definiciones por vocablo, se obtiene ejecutando:

```
./relac C-corpus.txt II3-corpus.txt | ./inters - > aprox-inf3-corpus.txt
```

donde,

C-corpus.txt: Archivo de consultas expandidas mediante el *corpus* (C1-corpus...C5-corpus).

relac y *inters*: Mencionados en las pruebas de grupos de programas de este capítulo.

aprox-inf1-corpus.txt y *aprox-inf3-corpus.txt*: resultado de la aproximación inferior usando un *corpus*, para una definición por vocablo y para tres definiciones, respectivamente.

Resultados:

La evaluación de los resultados de la A-Inf se muestran en las Tablas 4.15 y 4.16.

Consulta	$E = R/RC$	$P = R/NRes$
C1 – corpus	1/1	1/301
C2 – corpus	0/1	0/287
C3 – corpus	1/1	1/287
C4 – corpus	3/3	3/287
C5 – corpus	3/3	3/287

Tabla 4.15 Evaluación de la A-Inf de la tercera prueba usando un *corpus* y glosario con una definición por cada vocablo.

Consulta	$E = R/RC$	$P = R/NRes$
C1 – corpus	1/1	1/806
C2 – corpus	0/1	0/642
C3 – corpus	1/1	1/642
C4 – corpus	3/3	3/642
C5 – corpus	3/3	3/642

Tabla 4.16 Evaluación de la A-Inf de la tercera prueba usando un *corpus* y glosario con tres definiciones por cada vocablo.

Comentarios:

Como se puede observar a partir de los resultados obtenidos en la A-Inf, para ambos glosarios, la evocación es muy buena ($E = 1$) pero la precisión es muy mala ($P \approx 0$). Estos resultados de E y P se deben a que la cantidad de respuestas que arroja el sistema es muy alta en relación a la cantidad de entradas de los glosarios de una y tres definiciones por vocablo (312 y 858 vocablos, respectivamente).

4.3.2. Aproximación Superior

Comando:

1. La A-Sup, usando el glosario de una definición por vocablo, se obtiene ejecutando:

```
./a-sup C-corpus.txt SI1-corpus.txt > aprox-sup1-corpus.txt
```

2. La A-Sup, usando el glosario de una definición por vocablo, se obtiene ejecutando:

```
./a-sup C-corpus.txt SI3-corpus.txt > aprox-sup3-corpus.txt
```

donde,

C-corpus.txt: archivo de las consultas expandidas usando el *corpus*.

a-sup: Mencionado en las pruebas de grupos de programas de este capítulo.

SI1-corpor.txt y *SI3-corpor.txt*: clases parciales correspondientes al *SI* extendidos mediante el *corpus*, para glosarios con una definición por vocablo y tres definiciones, respectivamente.

aprox-sup-corpor1.txt y *aprox-sup-corpor3.txt*: resultados de la aproximación superior, para glosarios con una definición por vocablo y tres definiciones, respectivamente.

Resultados:

Los resultados de la A-Sup de esta prueba se muestran en las Tablas 4.17 y 4.18.

Consulta	$E = R/RC$	$P = R/NRes$
C1 – corpus	0/1	0/7
C2 – corpus	0/1	0/200
C3 – corpus	1/1	1/301
C4 – corpus	3/3	3/219
C5 – corpus	3/3	3/298

Tabla 4.17 Evaluación de la A-Sup de la tercera prueba usando un *corpus* y glosario con una definición por cada vocablo.

Consulta	$E = R/RC$	$P = R/NRes$
C1 – corpus	0/1	0/32
C2 – corpus	0/1	0/603
C3 – corpus	1/1	1/829
C4 – corpus	3/3	3/652
C5 – corpus	3/3	3/822

Tabla 4.18 Evaluación de la A-Sup de la tercera prueba usando un *corpus* y glosario con tres definiciones por cada vocablo.

Comentarios:

Al igual que los resultados de la A-Inf de ésta prueba, la evocación es muy buena y la precisión muy mala. Los resultados de la A-Inf trajeron más resultados correctos que la A-Sup.

Capítulo 5

Análisis de los resultados obtenidos

Es importante recordar que en el capítulo 4 se analizaron algunas consultas que fueron propuestas a partir de opiniones de profesores de ingeniería y definiciones en libros de Física con la finalidad de controlar más el experimento, debido a que el glosario tiene descripciones académicas. Con el objetivo de obtener ejemplos de cómo los usuarios pueden hacer búsquedas en un diccionario onomasiológico, se pidió a 30 estudiantes de ingeniería que describieran con sus palabras los vocablos (respuestas correctas en la tabla 4.4). Esta recolección de descripciones es un buen ejemplo de la problemática real que puede enfrentar un sistema de consulta onomasiológica, pero no fueron consideradas en el capítulo anterior para facilitar el análisis y tener una idea del comportamiento del sistema. Aún así los resultados para estas consultas (apéndice A.5) se realizó y se muestran en el apéndice A.6.

Los resultados del capítulo 4 que se analizarán serán fundamentalmente los relacionados con el método directo y el método usando un diccionario de sinónimos. Al comparar los resultados de éstos se podría decir que el uso de un diccionario de sinónimos para la representación, tanto de las consultas como de los glosarios (secciones 4.2.1 y 4.2.2), aumenta levemente la precisión con respecto al método directo. Esta mejora se intuía antes de aplicar este segundo método, sencillamente porque una misma consulta puede ser expresada por varios usuarios de manera diferente, y esta diferencia es en parte por el uso de sinónimos. Por ejemplo, para la consulta C3, en nuestro caso, la expresamos de la siguiente manera: *aplicación de una fuerza que provoca desplazamiento en una misma dirección*, sin embargo otro usuario

podría expresar lo mismo diciendo: *Generación de una fuerza que produce un movimiento en el mismo sentido*. En ambos casos estaríamos buscando que nuestro sistema nos devolviera como respuesta correcta *Energía mecánica* o *Trabajo mecánico*.

Aunque se puede considerar que hubo una mejora entre los resultados del método de representación de palabras claves usando un diccionario de sinónimos con respecto al método directo, sus resultados de A-Sup y A-Inf son muy parecidos. Por ejemplo, A-sup aumentó la precisión utilizando el glosario de tres definiciones para cada vocablo en la consulta **C2eq** y la cantidad de resultados aumentó en las A-Inf de la consulta **C5eq**, al usar los dos glosarios y en la A-Sup de la consulta **C4eq**. Aparte de estos cambios, los demás resultados se mantuvieron igual a los del método anterior. Nos preguntamos, sin embargo, por qué no mejoraron aún más estos resultados. Un elemento que influye en tales resultados está relacionado con las entradas del diccionario de sinónimos utilizado. Otra limitación con la que contamos es que el diccionario de sinónimos utilizado en estas pruebas es un diccionario general y, por otra parte, no se considera el contexto que puede tener la palabra. Hay que tener en cuenta que partimos de dos glosarios donde como máximo contábamos con tres definiciones por vocablo. Sin embargo, siempre habrá más de tres formas de describir dichos vocablos, y ésta puede ser la razón de por qué la consulta **C2** sólo tiene respuesta en la A-Sup de método usando un diccionario de sinónimos en el glosario de tres definiciones por vocablo.

Las pruebas realizadas utilizando un *corpus* del área de la Física (Mecánica y Cinemática) para expandir tanto las consultas como los glosarios, en este caso se esperaba que la evocación aumentara pero sin afectar mucho la precisión. Sin embargo, la cantidad de respuestas aumenta exageradamente, al punto en que la cantidad de resultados de las A-Inf y las A-Sup son mucho más de la mitad de los vocablos que aparecen en los glosarios e, incluso así, hay casos en que no aparece la respuesta deseada. La precisión, por otra parte, disminuyó mucho. Al comparar la cantidad de resultados de la A-Inf y la A-sup con respecto a la cantidad de respuestas correctas en estos resultados se observa que aunque aparezca la respuesta correcta en los resultados de la A-Inf y la A-Sup estos pierden relevancia al ser la cantidad de respuestas excesivamente elevada (ver Tablas 5.1 y 5.2).

El fracaso de la tercera prueba se debe a lo siguiente: primeramente, el tema utilizado en los glosarios, así como el del *corpus*, es un tema muy limitado (solamente a un área de la Física) en el cual casi todos los vocablos están

relacionados a nociones tales como, movimiento, fuerzas, desplazamiento, velocidad, cambio, tiempo, masa, etc. Por lo que al hacer una expansión, donde se tomaban temas relacionados con una palabra clave dada, casi todos los vocablos tenían la misma descripción extendida. Por este motivo la cantidad de respuestas fue tan grande.

Con respecto al empeoramiento de la precisión en la tercera prueba, en las Tablas 5.1 y 5.2 se comparan sus resultados con respecto a las dos primeras pruebas utilizando la relación de precisión, $P = R/NRes$ (definida en el capítulo 2). Mientras el valor de P se acerque a 1, mayor será la precisión.

Consulta	Directo		Sinónimo		Corpus	
	A-Inf	A-Sup	A-Inf	A-Sup	A-Inf	A-Sup
C1	1/2	0/0	1/2	0/0	1/301	0/7
C2	0/0	0/0	0/0	0/0	0/287	0/200
C3	1/1	0/0	1/1	0/0	1/287	1/301
C4	1/2	2/2	1/2	2/2	3/287	3/219
C5	3/15	2/2	3/21	2/2	3/287	3/298

Tabla 5.1 Comparación de resultados de las A-Inf y A-Sup de los tres métodos en cuanto a precisión (mediante la relación $P = R/NRes$); usando glosario de una definición por vocablo.

Consulta	Directo		Sinónimo		Corpus	
	A-Inf	A-Sup	A-Inf	A-Sup	A-Inf	A-Sup
C1	1/2	0/0	1/2	0/0	1/806	0/32
C2	0/0	0/0	0/0	1/1	0/642	0/603
C3	1/1	0/0	1/1	0/1	1/642	1/829
C4	1/2	2/2	1/2	2/2	3/642	3/652
C5	3/19	2/2	3/27	2/2	3/642	3/822

Tabla 5.2 Comparación de resultados de las A-Inf y A-Sup de los tres métodos en cuanto a precisión (mediante la relación $P = R/NRes$); usando glosario de tres definiciones por vocablo.

En las tablas anteriores (Tablas 5.1 y 5.2) se observa que la relación $P \approx 0$ al utilizar el método de expansión mediante un *corpus*, por lo que la precisión es muy baja. En contraste, los dos primeros métodos (Directo y Sinónimo) tienen solamente un caso (para A-inf) con $P = 0$.

Conclusiones

En este trabajo se obtuvieron los vocablos cuya definición se aproxima a la descripción de un concepto dado por un usuario. Dichas aproximaciones, superior e inferior, se basan en la teoría de conceptos formales y los conjuntos rugosos. Las aproximaciones fueron obtenidas empleando diferentes métodos llamados Directo, Sinónimo y *corpus*.

A continuación compararemos los mejores resultados de A-Inf y A-Sup en cuanto a precisión y evocación, del método Directo y Sinónimo, con respecto a los siguientes aspectos:

1. Comparación de los resultados de las aproximaciones del método Directo, usando glosario de una definición con respecto al de tres definiciones por vocablo.

Para las consultas C1 hasta la C4 sus resultados de A-Inf y A-Sup son los mismos usando ambos glosarios. Sin embargo, para la consulta C5 su A-Inf usando el glosario de tres definiciones por vocablo aumenta la cantidad de respuestas y siguen apareciendo las mismas que al usar el glosario de una definición por vocablo. Es por ello que disminuye un poco la precisión al usar el glosario de tres definiciones por vocablo con respecto al de una definición por vocablo. Esta disminución de la precisión se observa en las Tablas 5.1 y 5.2 donde la relación $P = R/NRes$, para la consulta C5, disminuye al usar el glosario de tres definiciones.

2. Comparación de los resultados de las aproximaciones del método de representación por Sinónimos, usando glosario de una definición con respecto al de tres definiciones por vocablo.

Los resultados de A-Inf de esta prueba son iguales usando ambos glosarios excepto para la consulta C5. Para esta consulta los resultados de

A-Inf aumentan la cantidad de respuestas al usar el glosario de tres definiciones por vocablo, aunque con la misma cantidad de respuestas correctas que al usar el glosario de una definición por vocablo. Por lo tanto, disminuye la precisión para la consulta C5 usando el glosario de tres definiciones, igual que ocurrió al obtener las A-Inf del método anterior.

Con respecto a las A-Sup de esta prueba aumenta la precisión para la consulta C2 al usar el glosario de tres definiciones por vocablo con respecto al de una definición por vocablo. Mientras que para la consulta C3 aumenta solamente la cantidad de respuestas (ver Tabla 5.2).

3. Comparación de los resultados de A-Inf y A-Sup del método Directo.
Se obtienen resultados de A-Inf para cuatro consultas. En el caso de la consulta C2 (que no tiene resultados de A-Inf) se debe a que en la forma en que se describe al vocablo en dicha consulta, no aparece en los glosarios.
La A-Sup sólo se obtiene para dos consultas (C4 y C5). Estos resultados obtenidos en la A-Sup son más precisos que los de las A-Inf (ver figuras 13 y 14). El hecho de que solamente dos consultas tengan resultados de A-Sup se debe a que las palabras claves de la definición deben estar contenidas en la consulta y en ocasiones esta contención no se cumple.
4. Comparación de los resultados de A-Inf y A-Sup del método de Sinónimo.
Al igual que al utilizar el método Directo, se obtienen resultados de A-Inf para cuatro consultas, mientras que la A-Sup mejora la precisión al usar el glosario de tres definiciones por vocablo con respecto al de una definición por vocablo, como se explicó en el punto 2.
5. Comparación de resultados de A-Inf del método de Sinónimos con respecto al método Directo.
Mediante el método de Sinónimos para la A-Inf aumenta la cantidad de respuestas con respecto al método directo, sin embargo sigue obteniéndose las mismas respuestas correctas con ambos métodos. Se esperaba que también aumentara la precisión pero esto no ocurrió. Lo anterior se debe al uso de un diccionario de sinónimos no especializado y a que los glosarios como máximo cuentan con tres definiciones por

vocablo mientras que en realidad siempre hay más formas de describir dicho vocablo.

6. Comparación de resultados de A-Sup del método de Sinónimos con respecto al método Directo.

Los resultados de la A-sup mejoran en cuanto a precisión al aplicar el método de Sinónimos con respecto al método Directo. La explicación de esta mejora se comentó en el capítulo anterior.

Del análisis anterior se puede llegar a la conclusión de que el método de representación de palabras clave (de la consulta y los glosarios) mediante un diccionario de sinónimos, para la obtención de aproximaciones inferior y superior, es mejor que el método Directo.

A continuación se proponen varios puntos para mejorar, en trabajos futuros, los diferentes métodos empleados en este trabajo para la obtención de A-Inf y A-Sup; ya que hacer más eficientes los SRI actuales obtendríamos un mejor acceso y aprovechamiento de la información.

- El método de representación de las consultas y de los glosarios utilizando un diccionario de sinónimos podría mejorarse. Si se utilizara un diccionario especializado tomando en cuenta los diferentes contextos que pueden estar referidos a una misma palabra clave, entonces se elegiría el sinónimo adecuado (desambiguación).
- Para los métodos Directo y usando Sinónimos si contáramos con glosarios más extensos en cuanto a formas de describir un vocablo determinado habría mayores posibilidades
- Dentro de las formas de describir un vocablo en el glosario sería recomendable que también se incluyeran algunas que no sea tan académicas. Por ejemplo, se puede describir la *Energía mecánica* o *Trabajo mecánico* como *cuando se empuja una cosa y esta se mueve*. Esta descripción puede ser una forma en que un usuario puede plantear su idea del vocablo que desea encontrar y es una prueba "más real" para las aplicaciones, tal como el experimento realizado en la UNAM, G.Sierra [7]. Otros ejemplos de cómo los usuarios pueden expresar una idea viene dado en el apéndice A.5 y los resultados de A-Inf y A-Sup de algunas de estas consultas se muestran en el apéndice A.6.

de encontrar la respuesta esperada.

- El método de expansión de consultas mediante un *corpus* fue desarrollado para enriquecer las consultas a los SRI. La prueba efectuada en este trabajo parte del hecho que mediante relaciones de sentido podría enriquecerse tanto la consulta como el glosario. La cantidad de respuestas aumentó demasiado porque no se controló la expansión. También influye el tamaño del *corpus*, su constitución u orientación (está más orientado a expresar matemáticamente los conceptos que a describirlos).

Para que éste método mejore se propone controlar la expansión agregando las relaciones de sentido más adecuadas para expandir.

Bibliografía

- [1] Keyun Hu, Yuefei Sui, Yuchang Lu, Ju Wang & Chunyi Shi: "Concept Approximation in Concept *Lattice*". PAKDD 2001, *LNAI* 2035, Springer Verlag, pp. 167-173 2001.
- [2] Gerardo Siera & John McNaught: "Natural Language System for Terminological Information Retrieval". CICALing 2003, *LNCS* 2588, Springer Verlag, pp. 541-552, 2003.
- [3] L.Landau & E.Lifshitz: *Curso Abreviado de Física Teórica, Mécanica y Electrodinámica*. Editorial Mir Moscú, 1979.
- [4] B.Ganter & R.Wille: *Formal Concept Analysis, Mathematical Foundations*. Berlin, Springer, 1999.
- [5] "Física con Ordenador". <http://www.sc.ehu.es/sbweb/fisica/default.htm>
- [6] Héctor Jiménez & G. Morales: "Grado de Pertenencia a un dominio y método de clasificación". *Computación y Sistemas*, Vol.5, N^o.4, pp. 288-295, 2002.
- [7] Gerardo Sierra: *Comunicación Personal*. Diciembre 2003.
- [8] C.J. van Rijsbergen: *Information Retrieval*. University of Glasgow, pp. 114 - 117. Second Edition, 1999.
- [9] H. Salazar, H. Jiménez & D. Pinto: "Text Extraction: a Corpus-based Approach". *Memorias del 30 Aniversario del PE en Computación de la BUAP*, ISBN-9688637114, pp 92-94, 2003.
- [10] R. Aceves, Y. De CasTilla, V. Cortés, H. Jiménez, C. Lucero, E. Mendoza, E. Mayotl, S. Paniagua, D. Pinto & B. Reyes: "Resultados del

Laboratorio de Recuperación de Información de la FCC". *Memorias del 30 Aniversario del PE en Computación de la BUAP*, ISBN-9688637114, pp 87-91, 2003.

- [11] G. Sierra & J. McNaught: "Design of an onomasiological search system: A concept-oriented tool for terminology". *Terminology*. Vol. 6 (1), 2000.

*

Apéndice A

A.1. Fragmento del glosario de una definición por vocablo

acción — magnitud que se define como producto de la energía absorbida durante un proceso por la duración del mismo

acción y reacción — cuando un cuerpo ejerce una fuerza sobre otro acción éste efectúa sobre el primero una fuerza igual y contraria reacción

aceleración — la razón de cambio del cambio de la velocidad al cambio del tiempo

aceleración angular — razón de cambio de la velocidad angular al cambio del tiempo

aceleración centrípeta — cuando un cuerpo se mueve en una trayectoria curva siempre con la misma rapidez su velocidad no es constante ya que va cambiando de dirección entonces la velocidad cambia a cada instante y por lo tanto hay una aceleración

afelio — en la órbita de un planeta el punto más alejado del Sol

altura — medida de la dimensión vertical de algo

amortiguado — con la intensidad la fuerza o la violencia disminuidas o suavizadas

amplitud — espacio recorrido por el cuerpo entre sus dos posiciones extremas

amplitud de una oscilación — es la elongación máxima la distancia entre las posiciones extremas de la señal

analogía — Relación de semejanza o parecido entre dos cosas distintas

angular — De forma de ángulo

área — Unidad de medida superficial equivalente al decámetro cuadrado

Aristarco (310-230 a JC) — astrónomo griego nacido en Samos el primero que afirmó que la Tierra giraba alrededor del Sol

Arquímedes (287 a 212? AJC) — Matemático científico e inventor nacido en Siarcusa Sicilia Estableció el principio de flotación de los cuerpos que hoy lleva su nombre y el principio de la palanca considerado como el inicio de la mecánica En matemáticas logró calcular el área de n

balanza de Cavendish — Instrumento para determinar la constante de gravitación en el cual se mide el desplazamiento de dos pequeñas esferas de masa m que están unidas con una ligera varilla suspendida por su punto medio de un hilo delgado producido por el acercamiento de dos g

balín — Bala de menor calibre que la ordinaria de fusil

Brahe Tycho (1546-1601) — astrónoma danés nacido en Knudstrup Las observaciones de Brahe minuciosamente realizadas y registradas resultaron de trascendental importancia para científicos posteriores como Kepler

caída libre — Movimiento de un cuerpo en un campo gravitatorio bajo la influencia de la gravedad

calor — Forma de energía debida a la agitación de las moléculas que constituyen una sustancia

cantidad — aspecto por el que se diferencian entre sí las porciones de la misma cosa o los conjuntos de la misma clase de cosas por el cual esas porciones o esos conjuntos se pueden medir o contar

cartesiano — Perteneciente o relativo al cartesianismo

cateto — Cada uno de los dos lados que forman el ángulo recto en el triángulo rectángulo

causa — Con relación a una cosa o suceso otra cosa u otro suceso que es el que produce aquellos

...

A.2. Fragmento del glosario de tres definiciones por vocablo

acción — Magnitud que se define como producto de la energía absorbida durante un proceso, por la duración del mismo.

acción — El producto de la energía y el tiempo.

acción — La acción es la fuerza que un cuerpo ejerce sobre un segundo cuerpo.

acción y reacción — Cuando un cuerpo ejerce una fuerza sobre otro (acción), éste efectúa sobre el primero una fuerza igual y contraria (reacción).

acción y reacción — Ley según la cual, cuando un cuerpo ejerce una fuerza sobre otro, el segundo cuerpo transmite al primero una fuerza colineal, de igual magnitud pero en sentido opuesto.

acción y reacción — Toda fuerza ejercida por A sobre B va aparejada con otra fuerza, igual y contraria, que B ejerce sobre A.

aceleración — La razón de cambio del cambio de la velocidad al cambio del tiempo

aceleración — Cociente de los incrementos de velocidad y tiempo cuando ambos tienden a cero

aceleración — Incremento de la velocidad en la unidad de tiempo

aceleración angular — Razón de cambio de la velocidad angular al cambio del tiempo.

aceleración angular — La velocidad angular de un objeto que se mueve en una trayectoria curva, como por ejemplo un círculo, se mide dividiendo el ángulo que recorre el objeto entre el tiempo que tarda en hacerlo. Así, la velocidad angular es el ángulo que se recorre en la uni

aceleración angular — Vector que mide la variación de la velocidad angular por unidad de tiempo en el movimiento circular

aceleración centrípeta — Cuando un cuerpo se mueve en una trayectoria curva, siempre con la misma rapidez, su velocidad no es constante, ya que va cambiando de dirección. Entonces la velocidad cambia a cada instante y por lo tanto hay una ac

aceleración centrípeta — Es lo mismo que aceleración normal,

pero por lo general se usa cuando un punto se mueve describiendo una circunferencia y por lo tanto su aceleración está dirigida hacia el centro de la misma.

aceleración centrípeta — Dirigida siempre hacia el centro, por lo que se le llama centrípeta (este término significa que busca el centro).

afelio — En la órbita de un planeta, el punto más alejado del Sol.

afelio — Punto que en la órbita de un planeta dista más del Sol

afelio — Punto más distante del Sol en la órbita de un planeta

altura — Medida de la dimensión vertical de algo

altura — Elevación que tiene un cuerpo sobre la superficie de la tierra

altura — Elevación o distancia de un cuerpo respecto a la tierra o a otra superficie de referencia

amortiguado — Con la intensidad, la fuerza o la violencia disminuidas o suavizadas

amortiguado — Efecto de amortiguar.

amortiguado — Un movimiento es amortiguado cuando se presenta una fuerza que causa que éste se atenúe.

amplitud — Espacio recorrido por el cuerpo entre sus dos posiciones extremas.

amplitud — En el movimiento oscilatorio, espacio recorrido por el cuerpo entre sus dos posiciones extremas. Es un ángulo en los movimientos circulares, y una distancia en los movimientos rectilíneos

amplitud — En una magnitud oscilatoria, valor máximo que alcanza

amplitud de una oscilación — Es la elongación máxima. La distancia entre las posiciones extremas de la señal

amplitud de una oscilación — El valor máximo que alcanza durante su ciclo una magnitud que varía periódicamente

amplitud de una oscilación — La de la cresta en relación con el nivel de señal constante

analogía — Relación de semejanza o parecido entre dos cosas distintas

analogía — Relación de semejanza entre cosas distintas

...

A.3. Fragmento del Diccionario de Sinónimos (truncado)

accion — acto hech actuacion combat bataill encuentr escaramuz
movimient gest ademan

acel — prontitud veloc rapidez prestez diligent activ cel precipit
pris apresur atolondr

aceler — veloz rap prest liger

actu — proced conduc port ejerc jug injustici iniqu ilegal ilicitud
arbitrariedad desafuer atropell legitim genuin autent verdader pur

agit — movimient trafag trajin inquietud intranquil conmocion
turbacion perturb ansi ansiedad congoj zozobr angusti tribul frenes
efervescent ardor exalt excit turbulent desord confusion alborot
revuelt motin

agu — linf gaseos aguardient acid orin pip pis mead dud vacil
mear

agud — delg puntiagud aguz afil fin sutil perspicaz discret ocu-
rrent gracios chistos ingeni viv penetrant estrident chirriant rechi-
nant profund oxiton

air — atmosfer vient aparient aspect port figur garb graci gallar-
di apostur climatiz insegur dudos march irse asemej semej parec
corr tir sal pas

aisl — solitari sol retir apart incomunic arrincon desatend olvid
posterg cartuj esporad ocasional excepcional suelt

alcanc — seguimient persecu alcanz

alcanz — logr consegu obten gan entend comprend penetr toc
tañ bast satisfac

aleman — german tudesc teuton

alrededor — aproxim

alta — admision ingres inscripcion incorpor

alter — cambi mud vari demud desarregl perturb inquiet trastorn
turb afect interes sobresalt estremec perd desconcert descompon
sal

altur — alto elev eminent cumbr altitud peralt excelent superior
altez ton excelent entramb

amortigu — apag debil mortecin

amplitud — extension anchur holgur

analogi — semej parec similitud
anterior — antecedent precedent previ
anticip — anterior prim años viej vetust antigu añej entrad
madur adult mayor gran machuch arcaic remot
aparat — instrument mec disposit artefact maquin pomp ostent
solemn ceremoni boat faust adorn atavi gal ornament ornat
aplic — adapt sobrepuest asidu esmer atencion estudi perseve-
ranci util utiliz utiliz
aplic — superpon sobrepon adapt destin adjudic dedic atribu
imput achac emple invert gast coloc trat esmer persever estudi
ingeniar atend
arbitrari — inmotiv caprich arbitral injust ilegal inicu despot
armon — armoni cadenci melod euritm musical eufon hiperton
aspect — aparient air car semblant cariz fisonomi plant presenci
pot fach pint catadur figur tall gir direccion curs vertient punt
fas period estad
astro — figur personal celebr
astronomi — uranografi cosmografi
anterior — retract
augment — acrecent increment crecimient adiccion sum añadidur
agreg adelant medr avanc alza sub elev encarec inflacion
ausenci — falt privacion carenci escasez
autor — causant creador inventor escritor actor
averigu — inquir indag investig busc enter sond ahond escud-
riñ escrit rebusc deduc desentrañ hall descubr sac
baj — descend abaj disminu mengu decrec dec caer abarat rebaj
depreci arri ape descabalg desmont humill abat
balanz — bascul roman
barr — band ray barrot pal liston pintalabi carmin
bas — asient apoy fundament cimient central sed alcali hidrox
busc — demand
caer — dec baj descend perd irse falt desaparec sucumb mor
perec extingu sent derrumb desplom ven percat record repar ad-
vert depend oscurec anohec
calcul — cont comput cuantific med calibr conjetur supon deduc
cre calibr

...

A.4. Fragmento del *corpus* (truncado)

acel — apliqu pis v cort air automóvil partícul 1014 x objet mpar aumentará valor grand + fren disminuirá acel m at aceleración encontr anteamient ocurr result ejercici permanec característ interval 6.25 emple carg v0 veloc s tiemp solución empl

aceler — atrapa anunci orig 45 5 v automóvil catódic tub improb relación x maner aut partícul + repos part pas televisor at cas m aceleración tocicl pong uniform gund electrón ectrón vez planteamient 5.25 retras ray motocicl s2 motociclet 3.00 región 10.0 0 úa aceler cambi loc arranc rapidez v0 15 emp 104m entra s constant t solución tas contin promedi alidadést tiemp

accept — estudiant eláre concurs prepar física explic especializ especial select ment recorr desarroll idóne creación partícul requ abiert investig actual dese foment permit curs matemát repos ambient distanci infinit ocurr ret clar determin aul realiz met cre referent accept problem analiz finit materi científic particip desempeñ grup físicos puert labor represent tiemp zon

acerc — aplic cer igual conoc aut alización partícul ganic grupal primer ap promedi aproximación inclus esper información objet med línea ecuacion resolución ón xi apoy recorr especif técnicas alg pequeñ aument cocient yésta aceleración pens veloc observ teorí vuelv obten vez todologi contien realment necessari movimient desarroll rep descripción personal complic necesit muev diferent 50 + concept complet palabr lug educación conflict estudi progr 2 azul tabl viaj curs disminuy roa part supondrá curv física instantáne defin luz precis coordin ejercici describ verbal problem gráfic resoluci sig útil expresión inicial trabaj desplaz noc afic interval aproxim términos físicos distanci tangent habil recolect buen aláre posterior determin explícit fácil cambi encontr dich ecuación trat ide orden tem análisis detall confront valor material punt ndient aprend capac form analiz +18 situación lluvi muestr d investigación tard rapidez posición ún 3t2 límite m convivent n tiemp per p 3.0 q larg lat tiv s t acerc moment v comprend sum brind x opinion pendent úmer z organiz eje tom resolv

acompañ — acompañ dibuj úres cuid estudi problem eg

acontec — visión física fundament rod juici form mund global

mod vid posesión conoc modern implic medi viv nec critic prec
 cis hombr acontec cons logr ltur acot x1 eláre x2 i integr nombr
 tegral describ x general tip real áre específic valor defin ecuación
 med azul aproxim element b obten sum curv recib ejempl acot
 figur eje límites límite deáre integral continu función f

activ — material activ disciplin propuest requ finaliz básica im-
 portant medi constanci trat clas import realiz encomend encarg
 cumpliment trabaj tiemp granéxit herramient consider lític actual
 al estudiant eláre model propi bas establec estructur útipl yópti-
 mos sistemát concurs prepar física lógica mundial aplicación inter-
 pret form geométric segur especializ especial select ment nivel
 técnica converg cinemát información preci desarroll idóne creación
 relación ident abiert investig moment reconoc concluy actual dese
 foment permit curs resolv pas análisis matemát maticesángul m
 ambient ocurr actualiz progresión pag ret clar determin secuenci
 esquem cálculo potenci enriquec trat result esfuerz aul acept met
 cre referent realiz consecuent viaj obtendrá analiz autoriz vuelv
 traslación gui problem materi científic particip plante tem método
 asum tad desempeñ grup bibliografí rode man buen intent puert
 labor represent aspect adopt fenómen dedicación g

actualiz — model bas establec estructur útipl través propi glob-
 aliz comprend mund enfrent lógica mundial aplicación interpret
 form geométric profesor global consult orient nivel técnica converg
 cinemát información generación recal maner llen aut relación
 ident revist best present modific moment reconoc concluy actual
 mencion red investig permit enigm educación juventud resolv pas
 análisis matemát maticesángul m uso ocument personalización
 progresión actualiz últim brind viv sencill clar recomendación se-
 cuenci esquem cálculo potenci enriquec trat mism internet aport
 result comend realiz consecuent viaj sell proces cambi incentiv
 problem vuelv traslación gui polítécn autoriz elángul sigl plante
 egres profesional científic pedagóg tem álisís fácilmente vertigin
 rode man bibliografí investigación aspect adopt fenómen public
 charl herramient pequen

...

A.5. Ejemplos de consultas dadas por estudiantes de ingeniería.

Las consultas se encuentran truncadas.

1. *Tierra*

- terc planet sistem sol , exist vid , diferent demas
- mas efectu fuerz atraccion cuerpos
- terc planet cercan Sol , tard 24 hor gir propi eje , rotacion .
Cuenta divers cap recubr . encuentr divers ecosistem encuentr divers tip biodivers .
- planet lug habit viv
- planet habit Hombre unic vida
- porcion material sol abund planeta
- convertir dia muramos
- punt esencial lodo
- planet sistem sol vivimos
- primer cap cortez terrestr , conoc , tiempo
- planet ocup terc lug cuatn posicion sistem sol . Cuenta condicion propiedad favorec evolu , reproduccion vid organismos
- element natural capaz vida
- nuev planet integr sistem solar
- terc planet sistem sol satelit natural llam Luna
- ecosistem conform plant , animal element agua
- terc enlac circuito

2. *Choque elástico*

- efectu cuerp gener reaccion
- relacion propiedad elast elementos
- rebote
- accion colision objet ultim elastico

- cuerp choc alej otro
- choqu cuerpos
- capac propiedad particul rebotar
- impact cuerp elasticidad
- capac cuerp reaccion encontr otro
- cuerp capac extend regres form original , coincid trayectoria
- accion realiz cuerp choc cuerp elast , intervien fuerz recib cuerp gener cuerp elastico
- cuerp rig contact choqu produc reaccion llam deformacion
- reaccion estir material
- cuerops encuentr , instant particul sufr cambi permanente

3. *Energía mecánica*

- realiz aplic fuerza
- fuerz efectu cuerp provoc reaccion
- fuerz desplaz lug , intervien elementos
- esfuerz cuerpo
- capac fuerz mec realiz accion segundo
- movimient realiz objet caus fuerz mism exterior
- Newton kilogram metr segund cuadrado
- fuerz movimient cuerp realiza
- capac cuerp realiz actividad
- esfuerz realiz ciert tip mecanismo
- movimient fuerz eherc cuerp import estado
- esfuerz realiz produc efectos
- accion realiz cuerpo
- accion realiz cuerp suministr calor tip energia
- esfuerz realiz gener movimient minim sea
- esfuerz realiz realiz tarea
- accion junt movimiento

A.5. EJEMPLOS DE CONSULTAS DADAS POR ESTUDIANTES DE INGENIERÍA.59

- tip energi produc esfuerz movimiento

4. *Movimiento lineal*

- desplaz superfici line recta
- deslize line accion
- cambi posicion line rect cuerpo
- recorr describ particul line recta
- trabaj realiz objet recorr line recta
- movimient particul desviarse
- form constant segu movimiento
- movimient particul derech , curvas
- movimient exist curvas
- desplaz line recta
- movimient desplaz constant mism direccion
- trayectori dimension varian
- cuerp sig trayectori rect desvie
- movimient trayect realiz sol sentido
- desplaz larg rect direccion sentido
- desplaz sol camino

5. *Aceleración*

- cambi subit velocidad
- cambi veloc respect distanci tiempo
- aument fuerz cuerp movimient , aplic mism direccion muev dich cuerpo
- aument increment velocidad
- capac vari veloc segund pasa
- increment veloc cuerp inici estatico
- 9.81 metr segund cuadrado
- cambi veloc respect tiempo
- necesit aument velocidad

- aument disminu veloc respect tiempo
- relacion exist metr segund cuadr
- cambi constant veloc variacion movimient cuerpo
- relacion exist trabaj aplic tiempo
- fuerz energi qu cuerp aument velocidad
- accion realiz cuerp intervien distanci tiempo
- segund deriv veloc respect tiempo
- veloc aplic cuerp

A.6. Obtención de A-Inf y A-Sup de algunas consultas del apéndice A.5

En este apéndice veremos resultados de A-Inf y A-Sup obtenidos a partir de algunas de las consultas del apéndice A.5. Las consultas utilizadas se muestran en la tabla A.1.

	Consulta	Cant. palabras
q1	terc planet sistem sol exist vid diferent demas	8
q2	mas efectu fuerz atraccion cuerpos	5
q3	terc planet cercan Sol tard 24 hor gir propi eje rotacion cuenta divers cap recubr encuentr ecosistem tip biodivers	19
q4	planet lug habit viv	4
q5	planet habit Hombre unic vida	5
q6	planet sistem sol vivimos	4
q7	primer cap cortez terrestr conoc tiempo	6
q8	planet ocup terc lug cuant posicion sistem sol cuenta condicion propiedad favorec evolu reproduccion vid organismos	16
q9	nuev planet integr sistem solar	5
q10	terc planet sistem sol satelit natural llam Luna	8

Tabla A.1 Selección de consultas dadas por estudiantes de ingeniería para obtener como resultado *Tierra*.

A.6. OBTENCIÓN DE A-INF Y A-SUP DE ALGUNAS CONSULTAS DEL APÉNDICE A.561

El método utilizado en estas pruebas es el Directo con el glosario de una definición por vocablo. En las tablas A.2 y A.3 se observan los resultados de A-Inf y A-Sup, respectivamente.

Consulta	Aprox. Inf. (Np/NC)
q1	No tiene
q2	Ley de la gravitación universal (3/5) y Fuerza de atracción gravitacional (3/5)
q3	No tiene
q4	No tiene
q5	No tiene
q6	Tierra (3/4) y Sol (3/4)
q7	No tiene
q8	No tiene
q9	no tiene
q10	No tiene

Tabla A.2 Resultados de la A-Inf.

Consulta	Aprox. Sup. (Np/ND)
q1	No tiene
q2	No tiene
q3	Movimiento rotacional (3/5)
q4	No tiene
q5	No tiene
q6	No tiene
q7	No tiene
q8	No tiene
q9	No tiene
q10	No tiene

Tabla A.3 Resultados de la A-Sup.

Con respecto a los resultados de las tablas A.2 y A.3 se puede observar que nuestro sistema no encuentra respuestas de A-Inf y A-Sup para la mayoría de las consultas expresadas por el usuario. Sólomente en el caso de la consulta q6 se obtiene la respuesta correcta en la A-Inf. Por lo que se puede observar la importancia de tener un glosario enriquecido con la mayor cantidad de definiciones posibles (académicas y no académicas), para expresar un mismo vocablo o concepto físico.