



Aplicaciones

Jose Aguilar
CEMISID, Escuela de Sistemas
Facultad de Ingeniería
Universidad de Los Andes
Mérida, Venezuela

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

se basa en definir los **documentos relevantes** según el grado de similitud entre la consulta del usuario y los documentos recuperados.

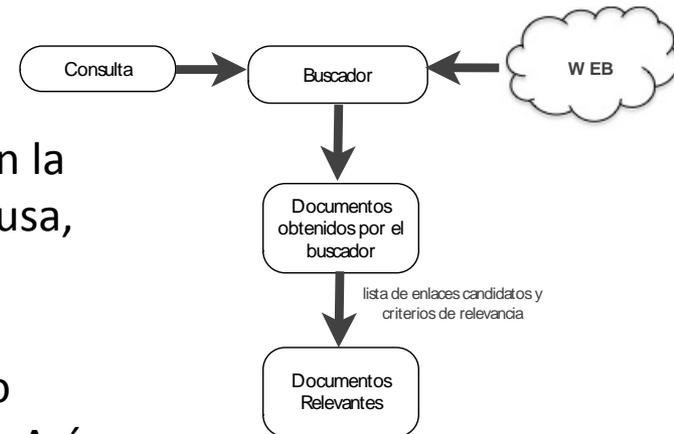
Para determinar esa similitud se usa ndistintos modelos en la literatura: el booleano, el vectorial, los basados en lógica difusa, redes neuronales, o redes bayesianos.

El servicio propuesto utiliza el modelo vectorial. Este modelo representa la consulta y los documentos mediante vectores . Así, un vocabulario de tamaño t definirá un espacio t -dimensional tal que un documento d_j es representado por un vector

$$d_j = (w_{1j}, w_{2j} \dots w_{tj})$$

y paralelamente una consulta q es representado como un vector

$$q = (w_{1q}, w_{2q} \dots w_{tq})$$



Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

- **Equiparación parcial**, capacidad del sistema para ordenar los resultados, basado en el grado de similaridad entre cada documento recuperado y la consulta.
- **Ponderación de los términos de los documentos y de los términos de la consulta**, el cual consiste en dar un valor real a los términos que reflejen su importancia en el documento y en la consulta.

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Para cumplir con los principios mencionados, se deben realizar los siguientes procesos:

- 1. Análisis de frecuencia**, el cual consiste en la contabilización del número de ocurrencias de los términos que se encuentra en la consulta y en los documentos recuperado,
- 2. Obtención de los pesos TF-IDF**, que consiste en el cálculo de la importancia de un término para discriminar y representar al documento y/o colección de documentos. Para hallar esos pesos se debe previamente definir la frecuencia inversa (IDF), la cual es calculada como:

$$IDF(\text{termino}) = \log_2\left(\frac{N}{DF}\right)$$

$$TF - IDF = TF(\text{termino}) * IDF(\text{termino})$$

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Donde:

N = número total de documentos de la colección

DF = número de documentos en los que aparece el término

TF = Frecuencia de aparición del término (n) en el documento

La frecuencia, como los pesos, son usados para el cálculo de la similaridad. Para medir la similitud entre un documento d_j y una consulta q , se va emplear la siguiente formula

$$\text{sim}(d_j, q) = \cos \theta = \frac{d_j \times q}{|d_j| \times |q|} = \frac{\sum_1^t w_{ij} \times w_{iq}}{\sqrt{\sum_1^t w_{ij}^2} \times \sqrt{\sum_1^t w_{iq}^2}}$$

donde $|d_j|$ y $|q|$ son las normas de los vectores que caracterizan al documento y a la consulta respectivamente

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Otra medida de interés es la Precisión, que se define como la proporción de los documentos recuperados que son relevantes.

Para el cálculo se utiliza la siguiente fórmula:

$$\text{Precisión} = \frac{\text{cantidad de documentos relevantes recuperados}}{\text{cantidad de documentos recuperados}}$$

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Supongamos que un usuario realiza la siguiente consulta en lenguaje natural en el buscador google:

“Universidad de Los Andes de Mérida”

Después de pasar por un proceso de interpretación de la consulta se obtiene la siguiente consulta booleana:

(“Universidad de los Andes” and Mérida and Venezuela) or (ULA Mérida and Venezuela) or (“Universidad de los Andes” and “Núcleo Mérida” and Mérida and Venezuela) or (ULA and “Núcleo Mérida” and Mérida and Venezuela),

Esta consulta se realiza en el buscador Google, y se obtiene el conjunto de enlaces a documentos

[http://es.wikipedia.org/wiki/Universidad_de_Los_Andes_\(Venezuela\)](http://es.wikipedia.org/wiki/Universidad_de_Los_Andes_(Venezuela))
<https://www.facebook.com/ula.venezuela>
<https://www.facebook.com/pages/Facultad-de-Ingenier%C3%ADa-ULA-Venezuela-Sitio-Oficial/25808485420578>
http://www2.ula.ve/andes/images/stories/inf_gestion_cap_1.pdf
<http://lama.adm.ula.ve/ptderecho/>
http://www2.ula.ve/andes/images/stories/pd/ula_cap1.pdf
<http://www.venezuelaaola.com/2014/02/decanos-de-ula-merida-estudian-suspender-clases/>
<http://www.venezuelaaola.com/2014/>
<http://www.elideshare.net/MANUELLIT0R>
<http://www.noticias24.com/venezuela/noticia/221873/suspender-las-actividades-academicas-indefiniidamente-en-la-universidad-de-los-andes/>
http://ar.ask.com/web?z=ula+de+merida+venezuela&acro=998&sem&tid=1332&corco=uf_8&fr=1&ad=sem&an=google_s&my=8&kwf=ula%20de%20merida%20venezuela&resq=3393209898&pla=8&mob=8&sc=sk&id=8&op=211&wss=40510943168&agid=9114244968&date=20131204
<http://www.mofl.gov.ve/Noticias/6134>
<http://www.actualidadyente.com/noticias-de-merida-venezuela/32-academicas/12888-hoy-la-ula-cumple-229-anos-de-fundada>
<http://www.actualidadyente.com/noticias-de-merida-venezuela/48-informacion-general-merida/72148-ula-crea-observatorio-de-derechos-humanos>
<http://www.aluniversa.com/nacional-y-politica/740710-decanos-de-ula-merida-podran-suspender-clases-por-violencia>
<http://eluniversario.net/ula-suspendidas-actividades-academicas-indefiniidamente-y-administrativas-hasta-el-lunes-17-de-febrero-en-el-nucleo-de-merida/>
<http://canalnoticia.com/ve/index.php/noticias-venezuela/item/26049-decanos-de-ula-merida-podran-suspender-clases-por-violencia>
<http://carlosramosnivas.com/2013/07/24/tesoros-de-merida-universidad-de-los-andes-primera-universidad-republicana-de-latinoamerica/>
http://www.ajcib.org.ve/pselo.php?did=51316-49162007000400019&script=sci_artext
<http://www.redajc.org/pdf/795/79504709.pdf>
<http://ppd.org.ve/portal/>
<http://www.almomento360.com/portal/estudiantes-de-la-ula-merida-protestan-en-contra-de-la-ineguridad/>
http://www.linkedin.com/search?orig=TSEO_SNA&firstName=Judi&lastName=Vega&M_Gevn%3A0&tr=TSEO_SN
<http://meridamusical.wordpress.com/artistas/agociaciones/boque-de-santa-rosa/>
<http://veasnoticias.com/universidad-de-los-andes-suspender-actividades/>
<http://www.minci.gob.ve/2012/11/entregan-financiamiento-a-investigadores-peil-de-merida/>
<http://noticiasvenezuela.info/2013/12/01/nu-sarueba-maestria-en-ciencias-de-la-actividad-fisica-y-los-deportes-para-la-ula-merida/>
<http://www.sagrasspica.com/31054-venezuela-rector-ula-rechaza-ingreso-grupos-armados-marlene-suspension-clases>
<http://informe21.com/universidad-los-andes>
<http://www.iberamerica.net/venezuela/prensa-generalista/noticias24.com/20140213/noticia.html?d=QG0812>
http://maribelsuarezmancha.blogspot.com/2014_02_09_archive.html
<http://temendenezco.com/ve/mediante-encuesta-evaluaran-la-situacion-para-un-possible-reinicio-de-actividades-en-la-ula/>

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Con los resultados obtenidos de la Web, se extraen los enlaces que contienen la página y se guarda en una lista.

Se utilizando los siguientes criterios para la selección de los enlaces candidatos:

- Documentos: html
- Lenguaje de los documentos: español
- No tomar en cuenta los siguientes enlaces:
- Los enlaces que comience con https, debido a que la información no se puede procesar por que están usando el protocolo seguro de transferencia de hipertexto, es decir (entidades bancarias o cualquier tipo de servicio que requiera envié de datos personales o claves).
- Los enlaces que terminen con alguna de las siguientes extensiones: .pdf, .jpg, etc.
- Los enlaces que direccionn a aplicaciones tales como slideshare, youtube, [linkedin](#), etc.
- Los enlaces que direcciona a Motores de búsqueda, por ejemplo [www.ask.com](#), etc.

Al aplicar ese filtro se eliminan los enlaces de la lista anterior que no cumplen con esa condición

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Al obtener la lista de los enlaces candidatos, se realiza el proceso de obtención de los documentos relevantes.

Este consiste en un procesamiento simple, con los siguientes pasos:

1. Se forma el vector de consulta, para lo cual se descompone la consulta en cuatro subconsultas, que se muestran a continuación:

q_1 = “Universidad de Los Andes” and Mérida and Venezuela

q_2 = “ULA” and Mérida and Venezuela

q_3 = “Universidad de Los Andes” and “Núcleo Mérida” and Mérida and Venezuela

q_4 = “ULA” and “Núcleo Mérida” and Mérida and Venezuela

Esas son las posibles combinaciones de las palabras existentes en la consulta original del usuario,

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

2. Luego se obtiene el **vector de consulta** .

Consulta	Universidad de Los Andes	Mérida	Venezuela	ULA	Núcleo Mérida
q_1	1	1	1	0	0
q_2	0	1	1	1	0
q_3	1	1	1	0	1
q_4	0	1	1	1	1

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

3. Ahora se trabaja con los documentos a los que enlaza la lista de los enlaces recuperados, para obtener los relevantes

- a) Hacer un primer filtrado, usando los criterios definidos por el usuario sobre las características básicas que debe tener un documento (tipo, lenguaje, etc.)
- b) Se procede a generar la matriz de frecuencia, para ello se recogen las apariciones de cada termino de la consulta en el documento que se esta procesando,

Enlace Doc	Universidad de Los Andes	Mérida	Venezuela	ULA	Núcleo Mérida
<u>http://es.wikipedia.org/wiki/Universidad de Los Andes (Venezuela)</u>	34	55	134	109	8
<u>http://llama.adm.ula.ve/pderecho/</u>	2	5	2	83	2
<u>http://www.venezuelaaldia.com/2014/02/decanos-de-ula-merida-estudian-suspender-clases/</u>	1	20	18	22	2
<u>www.venezuelaaldia.com/tag/ula/</u>	31	40	262	162	2
<u>http://www.noticias24.com/venezuela/noticia/221873/suspenden-las-actividades-academicas-indefinidamente-en-la-universidad-de-los-andes/</u>	7	9	5	4	7

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

4. Se procede a calcular la frecuencia inversa de cada documento para cada uno de los términos (la matriz DF-IDF).
5. Luego se calcula las similitudes existentes entre los distintos enlaces y el vector Q de la pregunta

Enlace Doc	Universidad de Los Andes	Mérida	Venezuela	ULA o (ULA)	Núcleo Mérida
http://es.wikipedia.org/wiki/Universidad_de_Los_Andes_(Venezuela)	5,6817395184	14,964374999	25,814440444	15,480071531	2,3964822549
http://llama.adm.ula.ve/pderecho/	0,3342199717	1,3603977272	0,3852901559	11,787577404	0,5991205637
http://www.venezuelaaldia.com/2014/02/decanos-de-ula-merida-estudian-suspender-clases/	0,1671099858	5,4415909087	3,467611403	3,1244181072	0,5991205637
www.venezuelaaldia.com/tag/ula/	5,1804095609	10,883181817	50,473010421	23,007078789	0,5991205637
http://www.noticias24.com/venezuela/noticia/221873/suspenden-las-actividades-academicas-indefinidamente-en-la-universidad-de-los-andes/	1,1697699008	2,4487159089	0,9632253897	0,5680760195	2,096921973

6. Finalmente, se aplica el método del coseno para obtener el orden de relevancia de los documentos,

VECTOR DE CONSULTA TF-DFI

Consulta	Universidad de Los Andes	Mérida	Venezuela	ULA	Núcleo Mérida
q_1	0,167	0,272	0,193	0	0
q_2	0	0,272	0,193	0,142	0
q_3	0,167	0,272	0,193	0	0,299
q_4	0	0,272	0,193	0,142	0,299

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Filtro	Yahoo	Google
<i>Enlaces pdf</i>	<i>9</i>	<i>23</i>
<i>Enlaces con autenticación de acceso</i>	<i>23</i>	<i>24</i>
<i>Total de enlaces relevantes candidatos en el primer filtro</i>	<i>141</i>	<i>126</i>

RESULTADOS DEL PRIMER FILTRADO

archivos pdf y con autenticación de acceso

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Filtro	Yahoo	Google
<i>Documento no html o el lenguaje es diferente español</i>	47	57
<i>Enlaces inactivos</i>	14	5
<i>Total de enlaces relevantes candidatos en el segundo filtro</i>	47	64

resultados del segundo filtrado

criterios tales como tipo de documento html, el lenguaje español y enlaces inactivos

resultados de la medida de precisión

Filtro	Yahoo	Google
<i>Precisión</i>	0,2716	0,3699

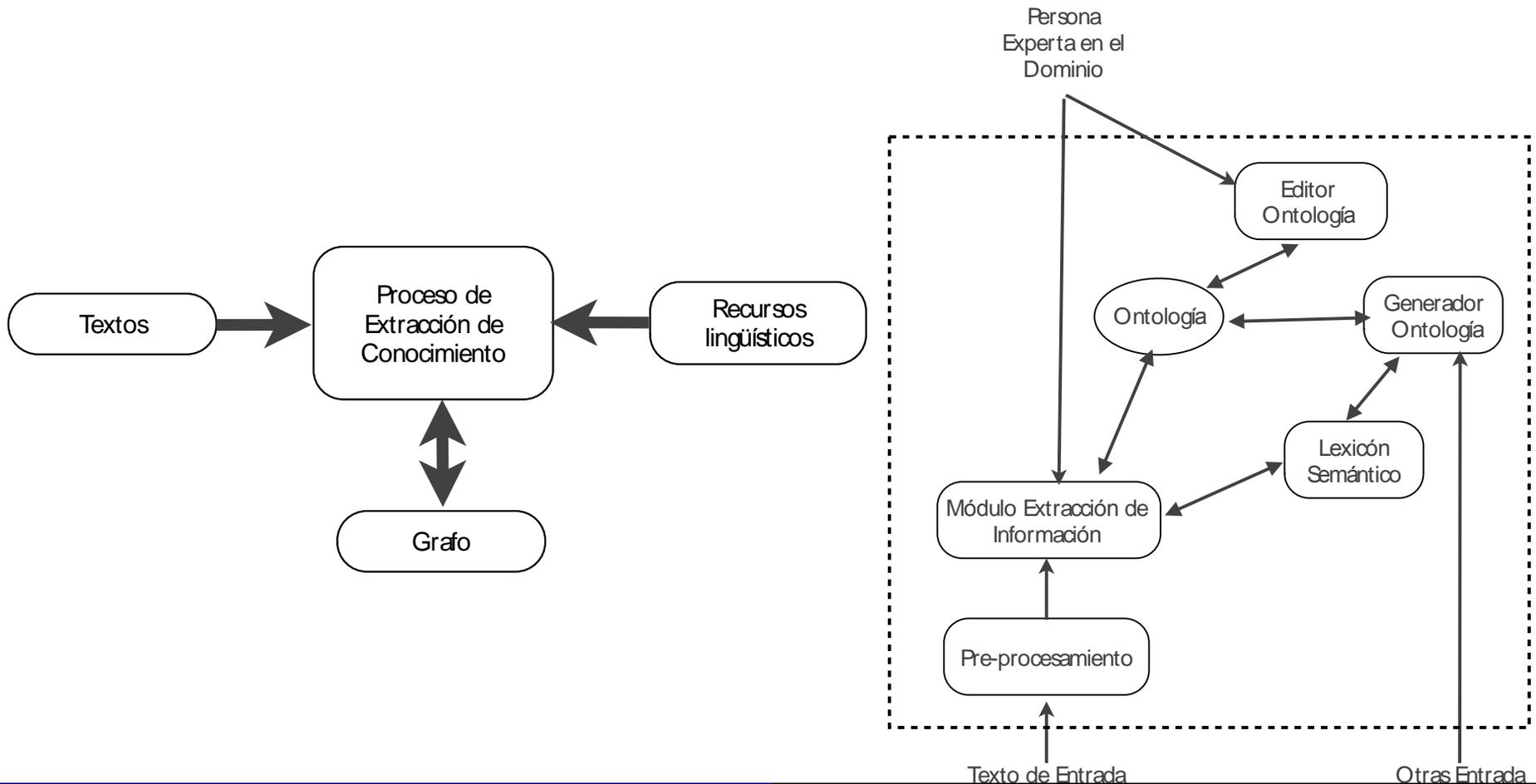
Nuestro sistema da todos los documentos como relevantes (precisión 100%).

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

valor de precisión para diferentes consultas sobre yahoo y google

Consultas Aleatorias	Yahoo	Google
<i>Consulta 1</i>	<i>0,2716</i>	<i>0,3699</i>
<i>Consulta 2</i>	<i>0,299</i>	<i>0,323</i>
<i>Consulta 3</i>	<i>0,253</i>	<i>0,312</i>
<i>Consulta 4</i>	<i>0,2612</i>	<i>0,3419</i>

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web



Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Lematizador

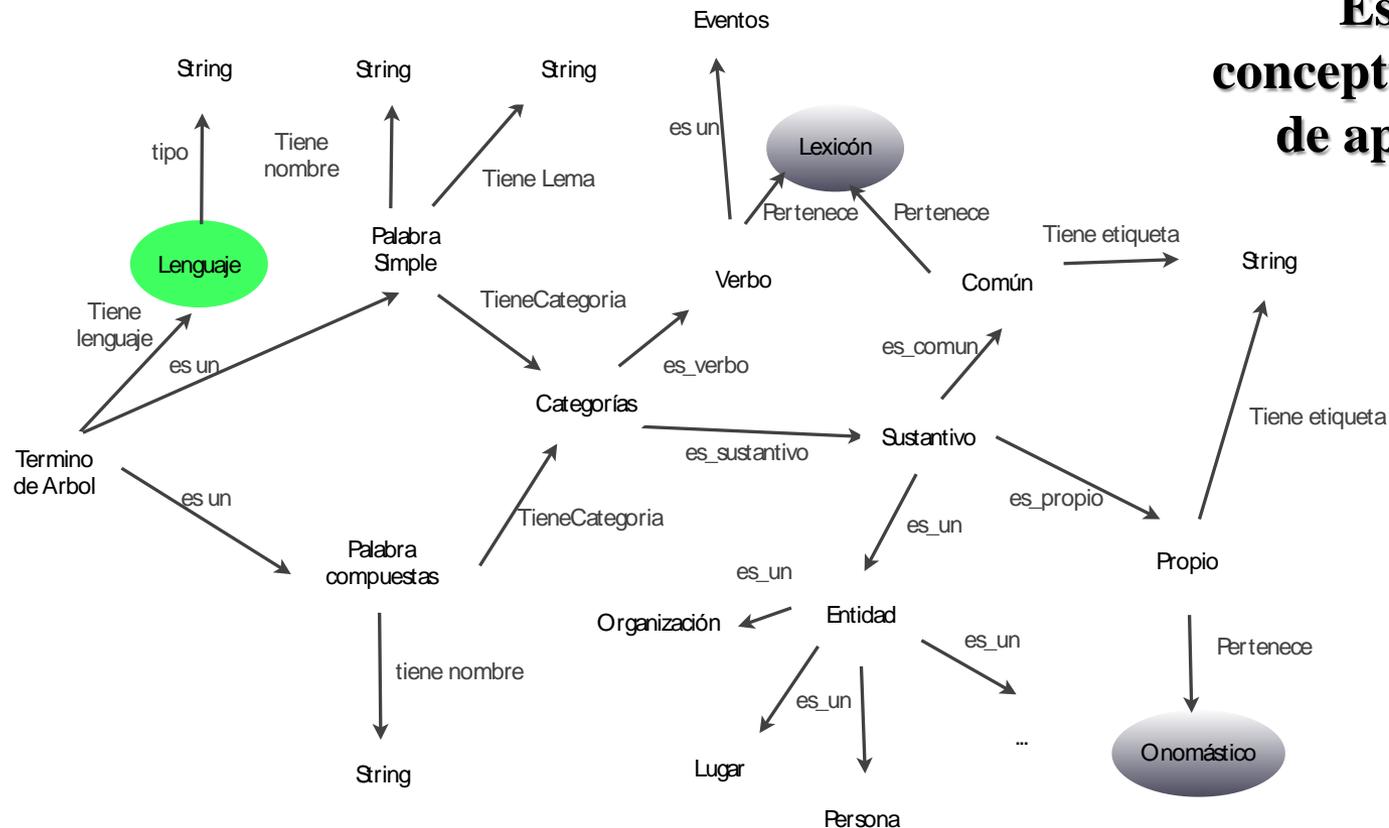
Es una parte del procesamiento lingüístico que determina el lema (forma canónica) de una palabra o token.

Recursos Lingüísticos

- **Lexicón:** El lexicón para un lenguaje (en este caso, el español) es la colección de palabras validas de un lenguaje, que son indexadas desde el lexema de la palabra, y describe todos sus posibles usos [1][2].
- **Onomasticon:** En el onomasticon se describen los nombres propios utilizados para nombrar personas, organizaciones, lugares y otros, según un dominio específico.
- **Corpus y diccionarios on-line:** Se utilizan corpus que son una colección de textos en lenguaje natural, y diccionarios electrónicos en formato legible, desde donde se permite extraer la información morfológica de las palabras.

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

El grafo de aprendizaje esta compuesto por un conjuntos de axiomas básicos para inferir nuevo conocimiento.

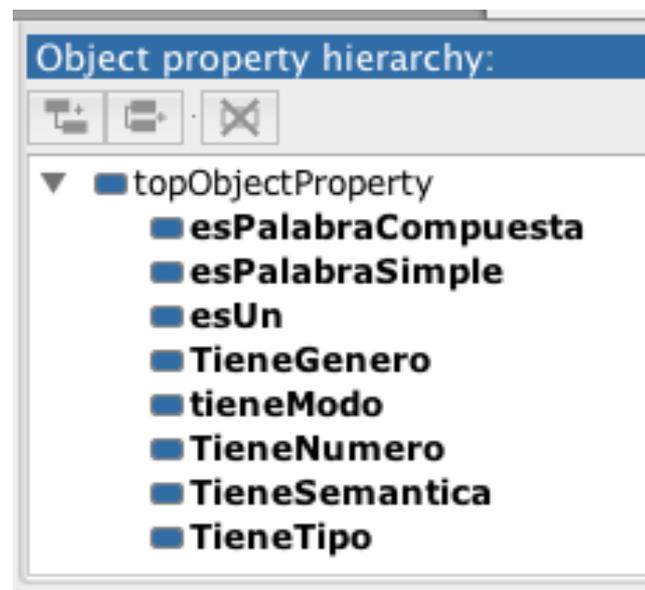


Esquema conceptual del grafo de aprendizaje

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

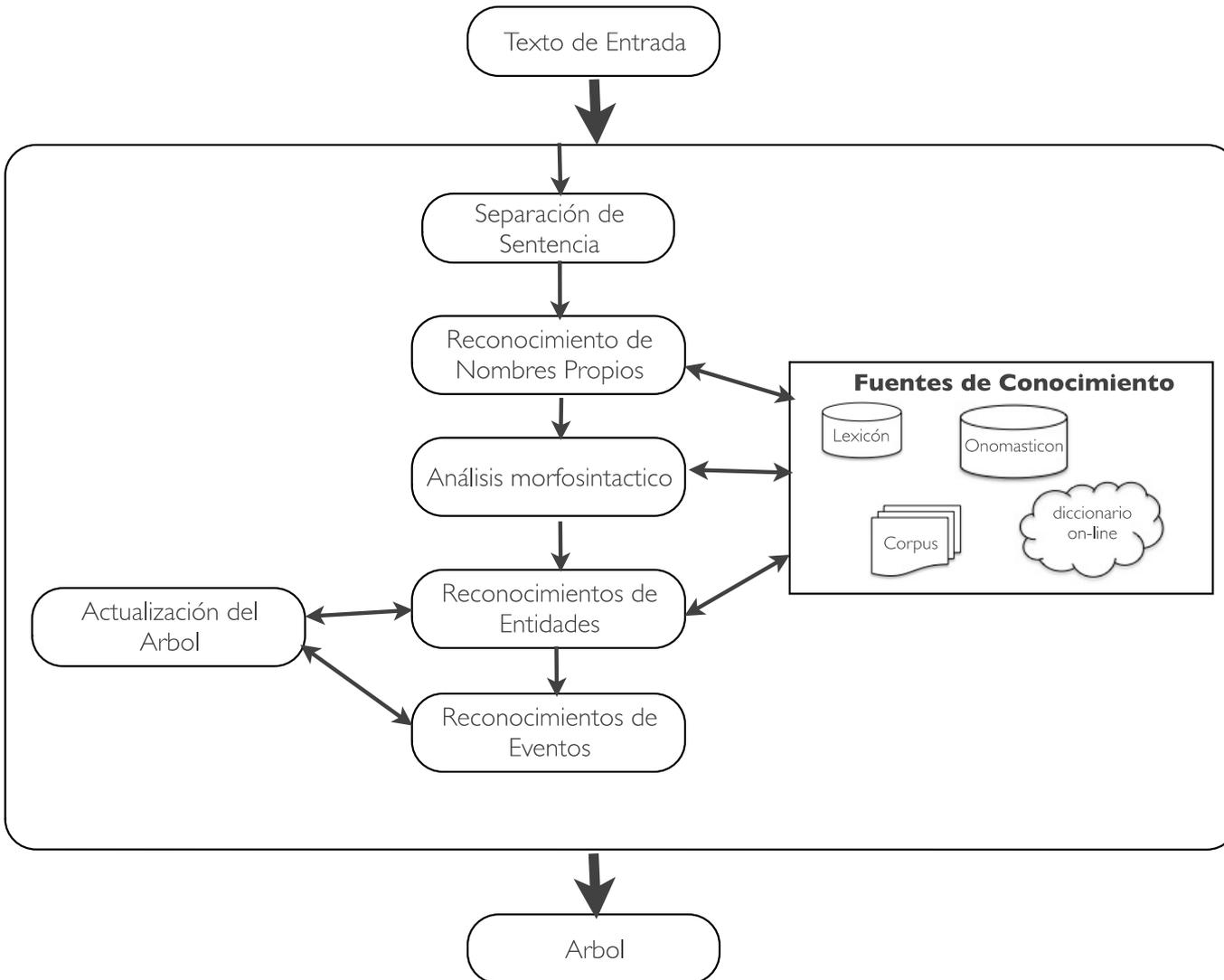


Las clases definidas en el grafo de aprendizaje



Las relaciones definidas en el grafo de aprendizaje

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web



Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Algunas posibles aplicaciones del servicio

- Creación lexicones electrónicos para los sustantivos y verbos.
- Detección de relaciones *es_un* (superclase-subclase), el cual se basa en la semejanza de los conceptos (uno de los conceptos incluye al otro)
- Creación automática de la base terminológica de un dominio

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

ejemplo 1: analizar sistios web

Procesamiento de texto no estructurado

Introducir el Texto

La Universidad de Los Andes es una universidad pública y autónoma ubicada en los andes venezolanos con su sede principal y rectorado en la ciudad de Mérida; fundada por el clero como casa de estudios el 29 de marzo de 1785, elevada luego a seminario y finalmente reconocida como Universidad el 21 de septiembre de 1810 bajo decreto expedido por la Junta Gubernativa de la provincia de la Corona de España.

Es una de las principales universidades de Venezuela por la cantidad de estudiantes que alberga, por su nivel académico y por sus aportes en investigación que han contribuido al estudio y desarrollo de las ciencias. La universidad tiene como propósito fortalecer la formación integral iniciada en los ciclos de educación primaria y secundaria, además de formar equipos profesionales y técnicos necesarios para el desarrollo y progreso de Venezuela.

La universidad está conformada por 11 facultades repartidas en el Núcleo Mérida (ubicado en la ciudad de Mérida), 3 núcleos autónomos localizados en las ciudades de San Cristóbal, Trujillo y El Vígía, dos extensiones universitarias con estudios de pregrado, postgrado y actualización profesional en Tovar y en Valera, extensiones de actualización profesional en las ciudades de Barinas, Guanare, Barquisimeto, Maracaibo, Caracas, entre otras, y diversas instalaciones universitarias dentro del territorio nacional como estaciones experimentales, haciendas de producción agrícolas, reservas naturales para el desarrollo de la fauna y flora y laboratorios de investigación.

Separar Sentencia

Texto no estructurado

Reconocimiento de Entidades y Relaciones Candidatas

Entidades Candidatas

Universidad de Los Andes	universidad	andes	sede
ciudad	Mérida	universidades	Venezuela
cantidad	estudiantes	nivel	aportes
investigación	estudio	desarrollo	ciencias
propósito	formación	ciclos	educación
equipos	progreso	facultades	Núcleo
núcleos	ciudades	San Cristóbal	Trujillo
El Vígía	extensiones	estudios	pregrado
actualización	Tovar	Valera	Barinas
Guanare	Barquisimeto	Maracaibo	Caracas
instalaciones	territorio	estaciones	haciendas
producción	reservas	fauna	flora
laboratorios			

Relaciones Candidatas

es	Es	alberga	han
contribuido	tiene	fortalecer	formar
está			

Actualizar el Arbol

Universidad de Los Andes
Desarrollado por Tatiana Rodríguez y Inca Anillar

Entidades y Relaciones Candidatas

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

ejemplo 1

	Total	Suma Frecuencia	Promedio	Máxima Frecuencia
Entidades Candidatas	251	903	7.19	38
Relaciones Candidatas	121	347	5,74	37

	Frecuencia	Peso	Porcentaje de Relevancia
Consejo Directivo	38	36,83	96,92
tesis	25	28,78	75,73
jurado	21	25,76	67,80
programa	21	25,76	67,80
tutor	21	25,76	67,80
Aspirante	18	23,29	61,28
doctorado	17	22,42	58,99
investigación	17	22,42	58,99
Plan de Formación	17	22,42	58,99
miembro	15	20,59	54,19
profesor	15	20,59	54,19
examen	14	19,64	51,68
créditos	12	17,64	46,42
caso	11	16,58	43,64
Area	10	15,49	40,76
estudiante	10	15,49	40,76
Estudios	10	15,49	40,76
grado	10	15,49	40,76
lapso	10	15,49	40,76
Programa de doctorado	10	15,49	40,76
artículo	9	14,35	37,77
comisión	9	14,35	37,77
conocimiento	9	14,35	37,77
doctoral	9	14,35	37,77
Facultad de Ingeniería	9	14,35	37,77
informe	9	14,35	37,77
postgrado	9	14,35	37,77
publicación	9	14,35	37,77
Año	8	13,17	34,65
candidatura	8	13,17	34,65
consejo	8	13,17	34,65
doctor	8	13,17	34,65
grupo	8	13,17	34,65
mes	8	13,17	34,65
reglamento	8	13,17	34,65
investigador	7	11,93	31,39
nivel	7	11,93	31,39
tipo	7	11,93	31,39
Admisión	6	10,63	27,96
criterios	6	10,63	27,96
curso	6	10,63	27,96
país	6	10,63	27,96
Programa de doctorado en ciencias aplicadas	6	10,63	27,96
actividades	5	9,25	24,34

Entidades relevantes con el criterio de peso $\geq 10,06$

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Ejemplo 2: Construir ontologías

En los textos analizados del Doctorado se encontraron las siguientes sentencias que tiene el verbo 'ser'

El **doctorando** es inmerso totalmente en la dinámica del grupo de investigación al cual pertenece su **tutor** y sigue los lineamientos previamente establecidos por éste en el **plan de formación**.

cualquier **investigador** cualificado que sea miembro de un grupo de investigación consolidado de la Universidad de Los Andes es, potencialmente **un tutor** del programa. Si el **Plan de Formación** no *es aceptado* por la Comisión de Admisión, el aspirante con su tutor podrán modificarlo y someterlo una vez más a la consideración de la Comisión, en un lapso de un mes.

Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

ejemplo 2

Por cada sentencia candidata se realiza el análisis morfosintáctico, generando el árbol sintáctico.

- En la sentencia 1 no se puede establecer ninguna relación con la entidades.
- En la sentencia 2 se establece la relación investigador es un tutor, por lo tanto
- En la sentencia 3 no se puede establecer la relación debido a que el verbo “es aceptado”

Por lo tanto, se puede obtener



Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

Description: LexiconElectronico		Description: LexiconElectronico	
Members +			
◆ ADMISIÓN	?	◆ INVESTIGACIÓN	?
◆ APROBADO	?	◆ JURADO	?
◆ ARTÍCULO	?	◆ LAPSO	?
◆ ASPIRANTE	?	◆ MIEMBROS	?
◆ AÑOS	?	◆ NIVEL	?
◆ CASO	?	◆ NOMBRAR	?
◆ CO-TUTOR	?	◆ PARTE	?
◆ COMISIÓN	?	◆ PAÍS	?
◆ CONOCIMIENTO	?	◆ PODRÁ	?
◆ CRITERIOS	?	◆ PODRÁN	?
◆ CRÉDITOS	?	◆ POSEER	?
◆ CUMPLIR	?	◆ PRESENTAR	?
◆ DEBE	?	◆ PROFESORES	?
◆ DEBEN	?	◆ PROGRAMA	?
◆ DEBERÁ	?	◆ REGLAMENTO	?
◆ DEBERÁN	?	◆ SEA	?
◆ DESARROLLAR	?	◆ SER	?
◆ DOCTOR	?	◆ SERÁ	?
◆ DOCTORADO	?	◆ SERÁN	?
◆ ELABORAR	?	◆ SIGUIENDO	?
◆ ES	?	◆ TENDRÁ	?
◆ ESTAR	?	◆ TENDRÁN	?
◆ ESTARÁ	?	◆ TESIS	?
◆ ESTÁN	?	◆ TIPO	?
◆ FORMAR	?	◆ TUTOR	?
◆ GRADO	?	◆ ÁREA	?
◆ HA	?		
◆ HABER	?		

ejemplo 2

Lexicón electrónico de
sustantivos y verbos

Reglas de Asociación

Reglas que implican relaciones

Sombreado: parado (standing)

No sombreado: acostado (lying)

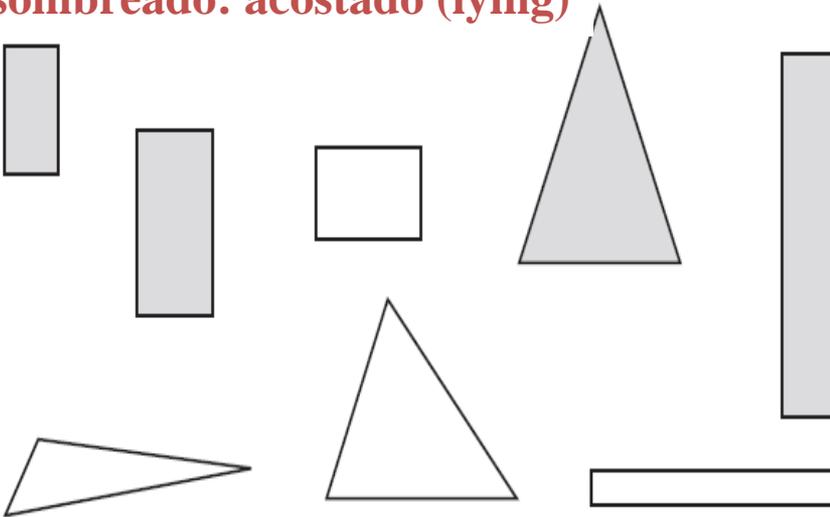


Tabla con datos de entrenamiento

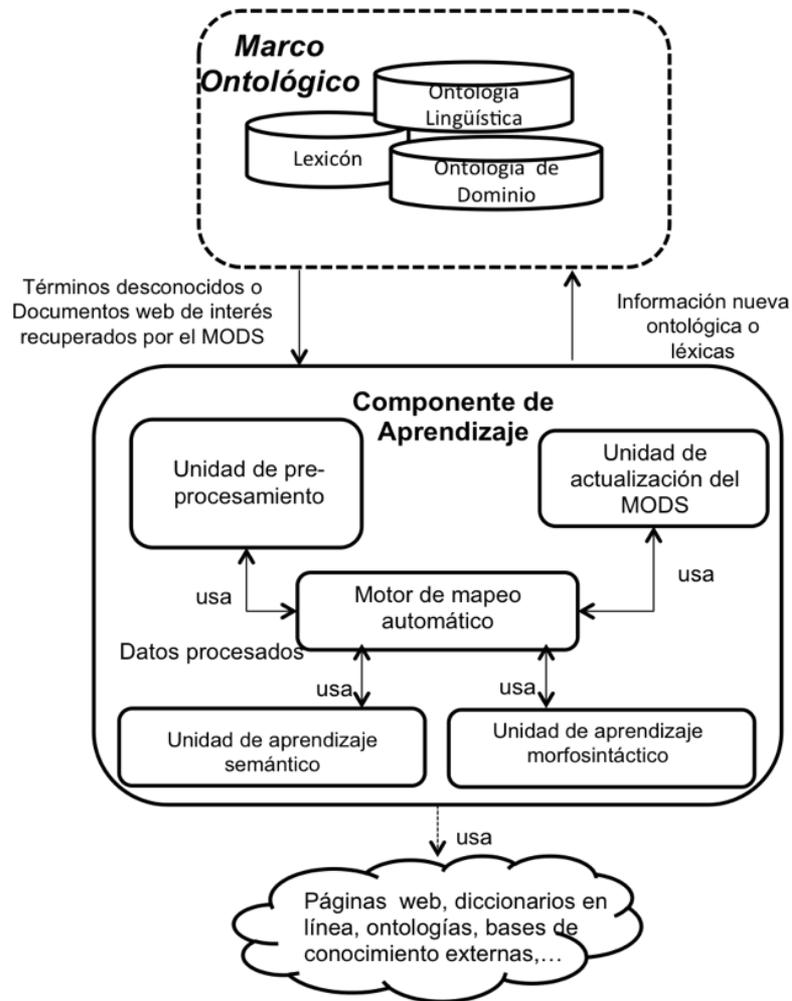
Width	Height	Sides	Class
2	4	4	standing
3	6	4	standing
4	3	4	lying
7	8	3	standing
7	6	3	lying
2	9	4	standing
9	1	4	lying
10	2	3	lying

Reglas



if width ≥ 3.5 and height < 7.0 then lying
if height ≥ 3.5 then standing

APRENDIZAJE ONTOLÓGICO

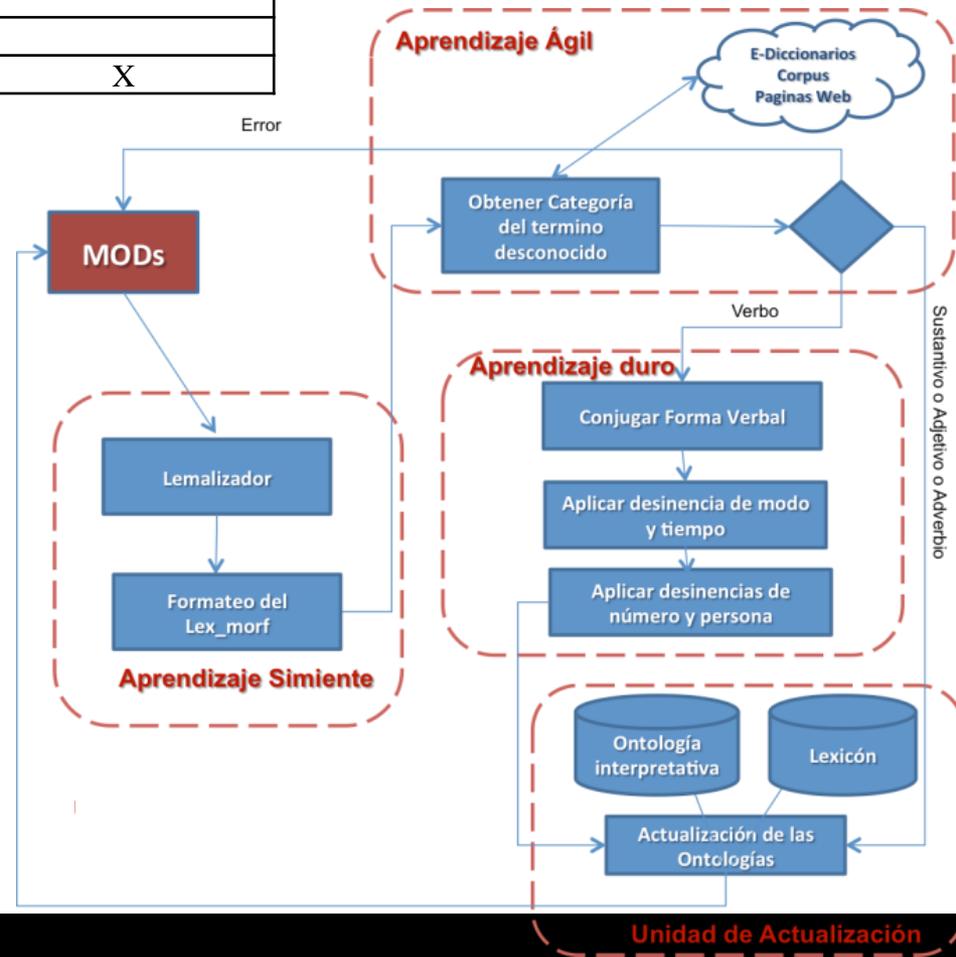


Aprendizaje morfosintáctico

Categoría	Genero	Número	Tipo	Modo	Tiempo	Persona
Sustantivo	X	X	X			
Adjetivo	X	X	X			
Adverbio			X			
Verbo				X	X	X

Requerimientos de aprendizaje de información morfosintáctica

Macro Algoritmo del aprendizaje morfosintáctica



Aprendizaje morfosintáctico



Aprendizaje de un sustantivo: Para la entrada:

lex_mor(cama, Desconocido, null, null, null, null, null, null, null, null, null).

Se pasa por:

1. El aprendizaje simiente y determina cual es su forma canónica.
2. El aprendizaje ágil, el cual determina que es un sustantivo y determina su estructura gramatical, dando como resultado lo siguiente: lex_mor('cama', 'sustantivo', 'NULL', 'femenino', 'singular', 'NULL', 'NULL', 'NULL', 'NULL', 'NULL', 'NULL', 'var(sustantivo)')
3. Finalmente, se actualiza el lexicón.

Aprendizaje morfosintáctico

Aprendizaje de verbo: Para la entrada

lex_mor(compra, Desconocido, NULL, NULL, NULL, NULL, NULL, NULL, NULL, NULL)

```
lex_mor(compro,Desconocido,NULL,NULL,NULL,NULL,NULL,NULL,NULL)
```

Procesando el componente del Aprendizaje Ontológico

Palabra encontrada

```
lex_mor('comprar','verbo','transitivo','NULL','NULL','NULL','NULL','NULL','NULL','NULL')
```

```
lex_mor('compro','verbo','transitivo','NULL','NULL','Indicativo','primera persona','presente','NULL','NULL','NULL')
lex_mor('compraba','verbo','transitivo','NULL','NULL','Indicativo','primera persona','imperfecto','NULL','NULL','NULL')
lex_mor('compré','verbo','transitivo','NULL','NULL','Indicativo','primera persona','preterito','NULL','NULL','NULL')
lex_mor('compraré','verbo','transitivo','NULL','NULL','Indicativo','primera persona','futuro','NULL','NULL','NULL')
lex_mor('compraría','verbo','transitivo','NULL','NULL','Indicativo','primera persona','condicional','NULL','NULL','NULL')
```

```
lex_mor('compras','verbo','transitivo','NULL','NULL','Indicativo','segunda persona','presente','NULL','NULL','NULL')
lex_mor('comprabas','verbo','transitivo','NULL','NULL','Indicativo','segunda persona','imperfecto','NULL','NULL','NULL')
lex_mor('compraste','verbo','transitivo','NULL','NULL','Indicativo','segunda persona','preterito','NULL','NULL','NULL')
lex_mor('comprarás','verbo','transitivo','NULL','NULL','Indicativo','segunda persona','futuro','NULL','NULL','NULL')
lex_mor('comprarías','verbo','transitivo','NULL','NULL','Indicativo','segunda persona','condicional','NULL','NULL','NULL')
```

```
lex_mor('compra','verbo','transitivo','NULL','NULL','Indicativo','tercera persona','presente','NULL','NULL','NULL')
lex_mor('compraba','verbo','transitivo','NULL','NULL','Indicativo','tercera persona','imperfecto','NULL','NULL','NULL')
lex_mor('compró','verbo','transitivo','NULL','NULL','Indicativo','tercera persona','preterito','NULL','NULL','NULL')
lex_mor('compará','verbo','transitivo','NULL','NULL','Indicativo','tercera persona','futuro','NULL','NULL','NULL')
lex_mor('compararía','verbo','transitivo','NULL','NULL','Indicativo','tercera persona','condicional','NULL','NULL','NULL')
```

Se pasa por los siguientes módulos:

1. Aprendizaje simiente, el cual determina el infinitivo del verbo (es a su vez su forma canónica).
2. Luego pasa por el aprendizaje ágil, el cual determina que es un verbo, y descubre que tipo de verbo es.
3. Aprendizaje duro, debido a que es un verbo. da la forma de conjugación de acuerdo al tipo de verbo (regular o irregular), dando como salida la siguiente información.

lex_mor('comprar', 'verbo', 'transitivo', 'NULL', 'NULL', 'NULL', 'NULL', 'NULL', 'NULL', 'NULL')

lex_mor('compro', 'verbo', 'transitivo', 'NULL', 'NULL', 'Indicativo', 'primera persona', 'presente', 'NULL', 'NULL', 'NULL')

lex_mor('compraba', 'verbo', 'transitivo', 'NULL', 'NULL', 'Indicativo', 'primera persona', 'imperfecto', 'NULL', 'NULL', 'NULL')

4. Finalmente, actualizar el lexicón

Aprendizaje morfosintáctico

Módulo Análisis Morfológico

Consulta: Universidad de Los Andes

Resultado del Análisis Morfológico

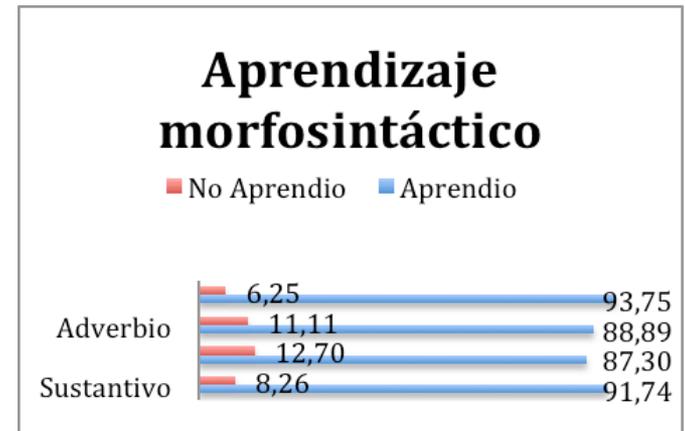
```
lex_mor(Universidad,sustantivo,NULL,femenino,singular, NULL,NULL,NULL,NULL,var(sustantivo))
lex_mor(de,preposicion,NULL,NULL,NULL,NULL,NULL,var(preposicion))
lex_mor(de,verbo,verbo_monosilabo,NULL,singular, Imperativo,presente,NULL,NULL,tercera
persona,var(verbo))
lex_mor(de,verbo,verbo_monosilabo,NULL,singular, Subjetivo,presente,NULL,NULL,primera
persona,var(verbo))
lex_mor(de,verbo,verbo_monosilabo,NULL,singular, Subjetivo,presente,NULL,NULL,tercera
persona,var(verbo))
lex_mor(de,sustantivo,Común,Femenino,Singular, NULL,NULL,NULL,NULL,var(sustantivo))
lex_mor(Los,sustantivo,NULL,masculino,singular, NULL,NULL,NULL,NULL,var(sustantivo))
lex_mor(Andes,sustantivo,NULL,masculino,singular, NULL,NULL,NULL,NULL,var(sustantivo))
lex_mor(Andes,verbo,intransitivo,NULL,singular, Subjetivo,presente,NULL,NULL,segunda
persona,var(verbo))
```

Componente léxico	Categoría	Tipo	Genero	Número	Modo	Persona	Tiempo
Universidad	sustantivo	NULL	femenino	singular			
de	preposicion						
de	verbo	verbo_monosilabo	NULL	singular	Imperativo	tercera persona	presente
de	verbo	verbo_monosilabo	NULL	singular	Subjetivo	primera persona	presente
de	verbo	verbo_monosilabo	NULL	singular	Subjetivo	tercera persona	presente
de	sustantivo	Común	Femenino	Singular			
Los	sustantivo	NULL	masculino	singular			
Andes	sustantivo	NULL	masculino	singular			
Andes	verbo	intransitivo	NULL	singular	Subjetivo	segunda persona	presente

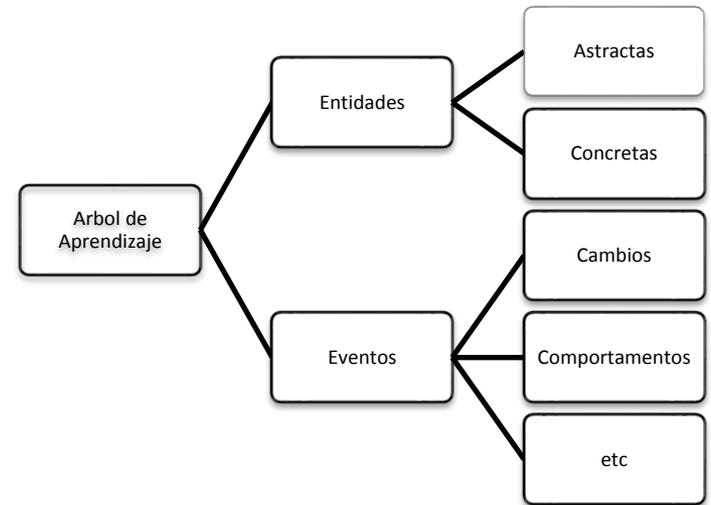
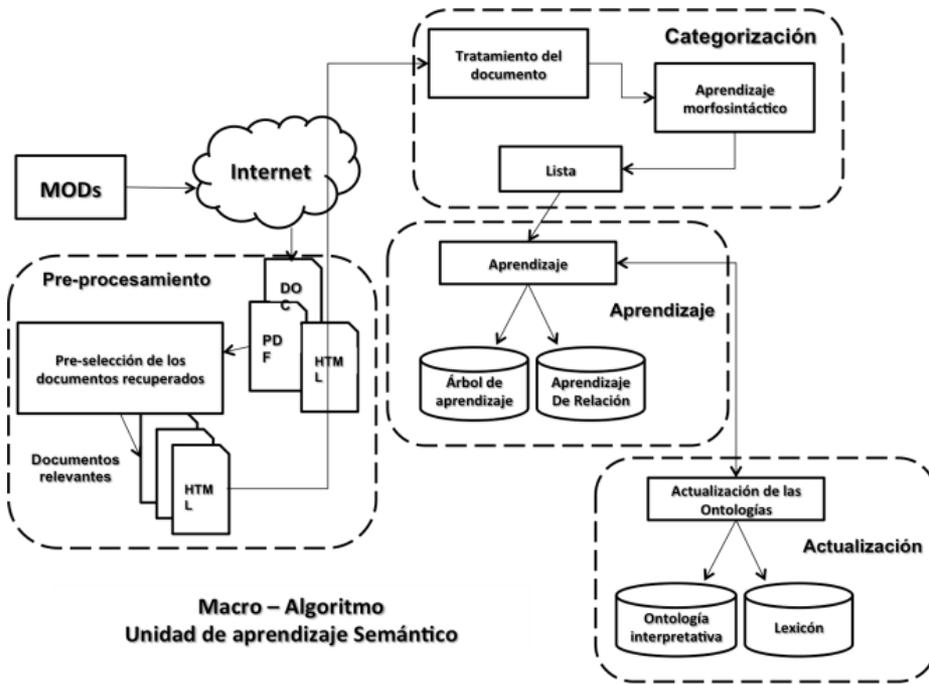
Término	Aprendió	No Aprendió	Total
Sustantivo	500	45	545
Adjetivo	220	32	252
Adverbio	80	10	90
Verbo	75	5	80

Resultado del análisis léxico-morfológico de los componentes léxicos: Universidad, de, Los, Andes.

Aprendizaje morfo-sintáctico



APRENDIZAJE SEMÁNTICO PARA EL MODS



Árbol de Aprendizaje

Macro Algoritmo de la Unidad de Aprendizaje Semántico

APRENDIZAJE SEMÁNTICO PARA EL MODS

Supongamos que un usuario realiza la siguiente consulta en lenguaje natural en Google:

“Universidad de Los Andes de Mérida”

Posible consulta

**“Universidad de los Andes” and Mérida and Venezuela)
or (ULA Mérida and Venezuela) or (“Universidad de los
Andes”
and “Núcleo Mérida” and Mérida and Venezuela) or (ULA
and “Núcleo Mérida” and Mérida and Venezuela),**

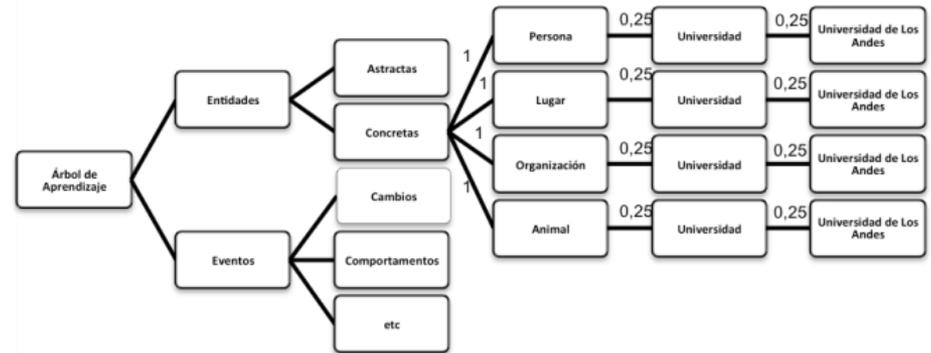
[http://es.wikipedia.org/wiki/Universidad_de_Los_Andes_\(Venezuela\)](http://es.wikipedia.org/wiki/Universidad_de_Los_Andes_(Venezuela))
<https://www.facebook.com/ula.venezuela>
<https://www.facebook.com/pages/Facultad-de-Ingenier%C3%ADa-ULA-Venezuela-Sitio-Oficial/258084854230578>
http://www2.ula.ve/plandes/mages/stories/inf_gestion_cap_ii.pdf
<http://lama.adm.ula.ve/pderecho/>
http://www2.ula.ve/plandes/mages/stories/pdula_capii.pdf
<http://www.venezuelaalcia.com/2014/02/decanos-de-ula-merida-estudian-suspender-clases/>
<http://www.venezuelaalcia.com/tag/ula/>
<http://www.slideshare.net/MANUELLIT0R>
<http://www.noticias24.com/venezuela/noticia/221873/suspenden-las-actividades-academicas-indefinidamente-en-la-universidad-de-los-andes/>
http://ar.ask.com/web?q=ula+de+merida+venezuela&qsrc=999&l=sem&siteid=1532&qenc=utf-8&lfr=1&ad=semA&an=google_s&mtty=b&kwd=ula%20de%20merida%20venezuela&net-p&cre=3393026980&pla=&mob=&sou=s&aid=&adp=211&kwid=60510843168&agid=9114244968&date=20131204
<http://www.mcti.gob.ve/Noticias/16134>
<http://www.actualidadygente.com/noticias-de-merida-venezuela/32-academicas/12888-hoy-la-ula-cumple-229-anos-de-fundada>
<http://www.actualidadygente.com/noticias-de-merida-venezuela/48-informacion-general/merida/12146-ula-crea-observatorio-de-derechos-humanos>
<http://www.eluniversal.com/nacional-y-politica/140210/decanos-de-ula-merida-podran-suspender-clases-por-violencia>
<http://eluniversitario.net/ula-suspenden-actividades-academicas-indefinidamente-y-administrativas-hasta-el-lunes-17-de-febrero-en-el-nucleo-de-merida/>
<http://canal5demercado.com.ve/index.php/noticias-venezuela/item/36049-decanos-de-ula-merida-podran-suspender-clases-por-violencia>
<http://carlosarasmosrivias.com/2013/07/24/tesoros-de-merida-universidad-de-los-andes-primera-universidad-republicana-de-latinoamerica/>
http://www.scieilo.org.ve/scieilo.php?pid=S1316-49102007000400019&script=sci_artext
<http://www.redalyc.org/pdf/705/70504702.pdf>
<http://ppd.org.ve/portal/>
<http://www.alimento360.com/portal/estudiantes-de-la-ula-merida-protestan-en-contra-de-la-inseguridad/>
http://www.linkedin.com/vsearch?o?orig=TSEO_S&firstName=Judith&lastName=Vega&f_G=ve%3A0&rk=TSEO_SN
<http://meridamusical.wordpress.com/artistas/agrupaciones/vozes-de-santa-rosa/>
<http://veusunoticias.com/universidad-de-los-andes-suspende-actividades/>
<http://www.minci.gob.ve/2012/11/entregan-financiamiento-a-investigadores-pai-de-merida/>
<http://noticiasvenezuela.info/2013/12/cnu-aprueba-maestria-en-ciencias-de-la-actividad-fisica-y-los-deportes-para-la-ula-merida/>
<http://www.lagranepoca.com/31054-venezuela-rector-ula-rechaza-ingreso-grupos-armados-mantiene-suspension-clases>
<http://informe21.com/universidad-los-andes>
<http://www.iberamerica.net/venezuela/prensa-generalista/noticias24.com/20140213/noticia.html?tid=QG0812>
http://marbelsuarezmancha.blogspot.com/2014_02_09_archive.html
<http://elmeridenezgo.com.ve/mediante-encuesta-evaluaran-la-situacion-para-un-posible-reinicio-de-actividades-en-la-ula/>

Extracto de los documentos recuperados de consulta realizada por el MODS

APRENDIZAJE SEMÁNTICO PARA EL MODS

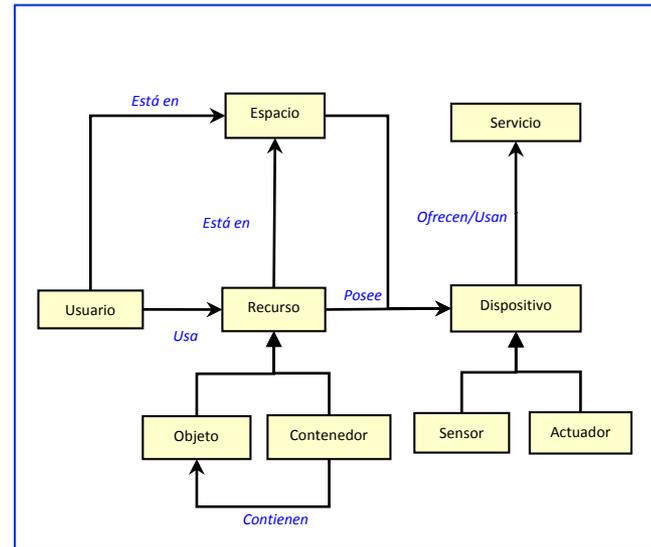
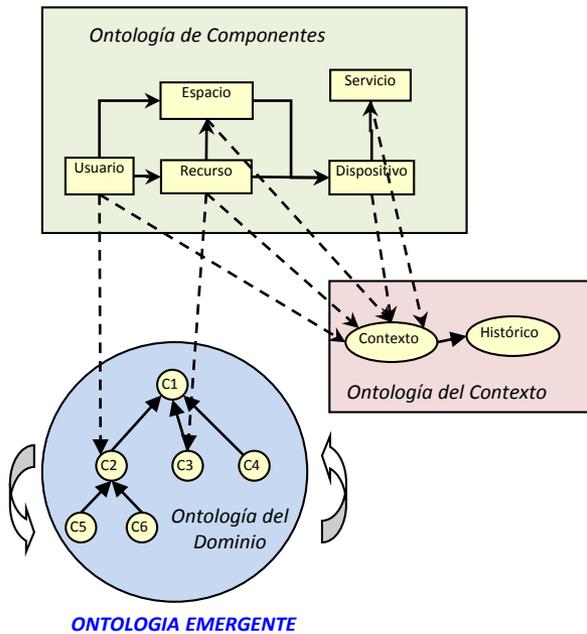
Término	Frecuencia	Categoría
Universidad	35	Sustantivo
Organización	13	Sustantivo
Dicta	7	Verbo
Mérida	4	Sustantivo
...

Ejemplo de la tabla de frecuencias



Ejemplo del árbol de aprendizaje

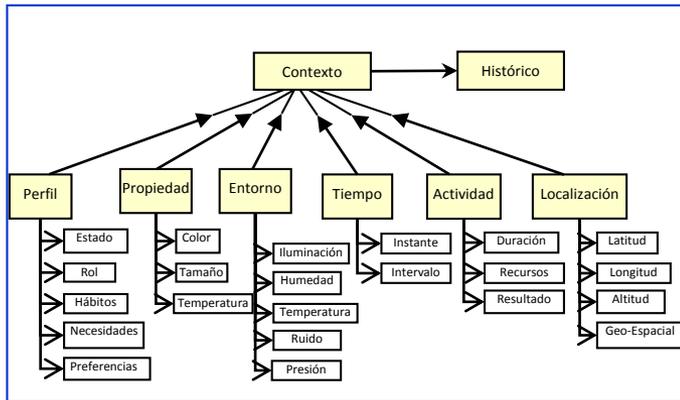
Ontología Emergente



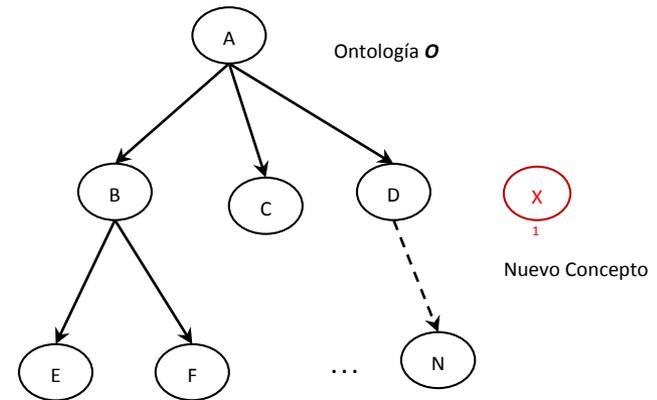
Ontología de Componentes de un Aml

Ontologías en un Aml

Ontología Emergente

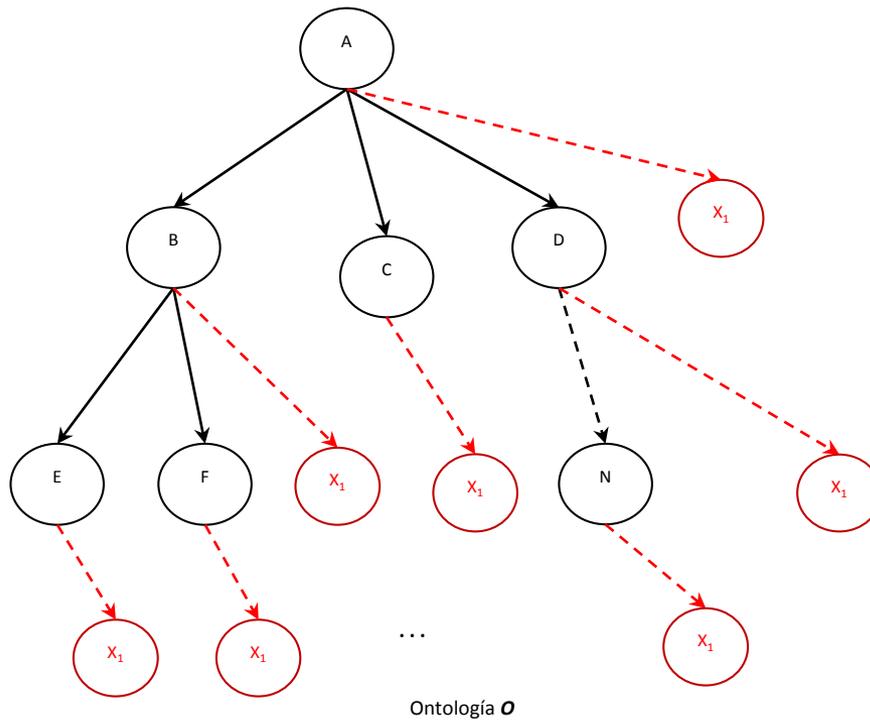


. Ontología de Contexto para un Aml



nuevo concepto que se debe incluir en una ontología

Ontología Emergente



Posibles posiciones de un nuevo concepto en la ontología

Ontología Emergente

Objeto	Propiedades		Ubicaciones			Tareas		
	Liq.	Frio	Nev	Desp	Lav	Preparar comida	Preparar bebidas	Limp
Pollo	0	0	1	0	0	1	0	0
Carne	0	0	1	0	0	1	0	0
Tomate	0	0	0	0	0	1	0	0
Arroz	0	0	0	1	0	1	0	0
Agua	1	1	0	0	0	0	1	0
Jugo	1	1	0	0	0	0	1	0
Lavaplatos Líquido	1	0	0	0	1	0	0	1
Refresco	1	1	0	0	0	0	1	0

Ontología Emergente

Coeficiente de similitud entre A y B usando el índice de *Jacckar*

$$S_{AB} = \frac{R}{R + T + V}$$

$$T = \sum_{j=1}^N X_{A_j} X_{B_j}$$

$$R = \sum_{j=1}^N X_{B_j} (1 - X_{A_j})$$

$$V = \sum_{j=1}^N X_{A_j} (1 - X_{B_j})$$

De manera general, la semejanza de un concepto X_i con un concepto C_j , para establecer la relación “es-un”, viene dada por:

$$S_{X_i C_j} = \frac{\sum_{C_k \in H} S_{X_i C_k}}{N}$$

Donde:

H : todas las clases derivadas de C_j (hijas)

N : número total de clases derivadas de C_j

Ontología Emergente

Si la hormiga está parada en un concepto R , podrá decidir colocar a X como sub-clase de C (que es hijo de R), en función de la semejanza entre X y C .

Probabilidad de transición:

$$P_{rs}^k = \frac{\gamma_{rs}^\alpha \cdot \eta_{rs}^\beta}{\sum_{u \in J_r^k} \gamma_{ru}^\alpha \cdot \eta_{ru}^\beta} \quad \text{Si } s \in J_r^k$$

η_{rs} : Cantidad de feromona.

γ_{rs} : Información Heurística.

J_r^k : Nodos aún no visitados por la hormiga k desde r .

α y β : Importancia de la información memorística (feromona) y heurística.

Actualización de la feromona

$$\gamma_{rs} = (1 - \rho) \cdot \gamma_{rs}$$

$$\gamma_{rs} = \gamma_{rs} + \Delta \gamma_{rs}$$

$$\Delta \gamma_{rs} = \sum_{k=1}^M \Delta \gamma_{rs}^k$$

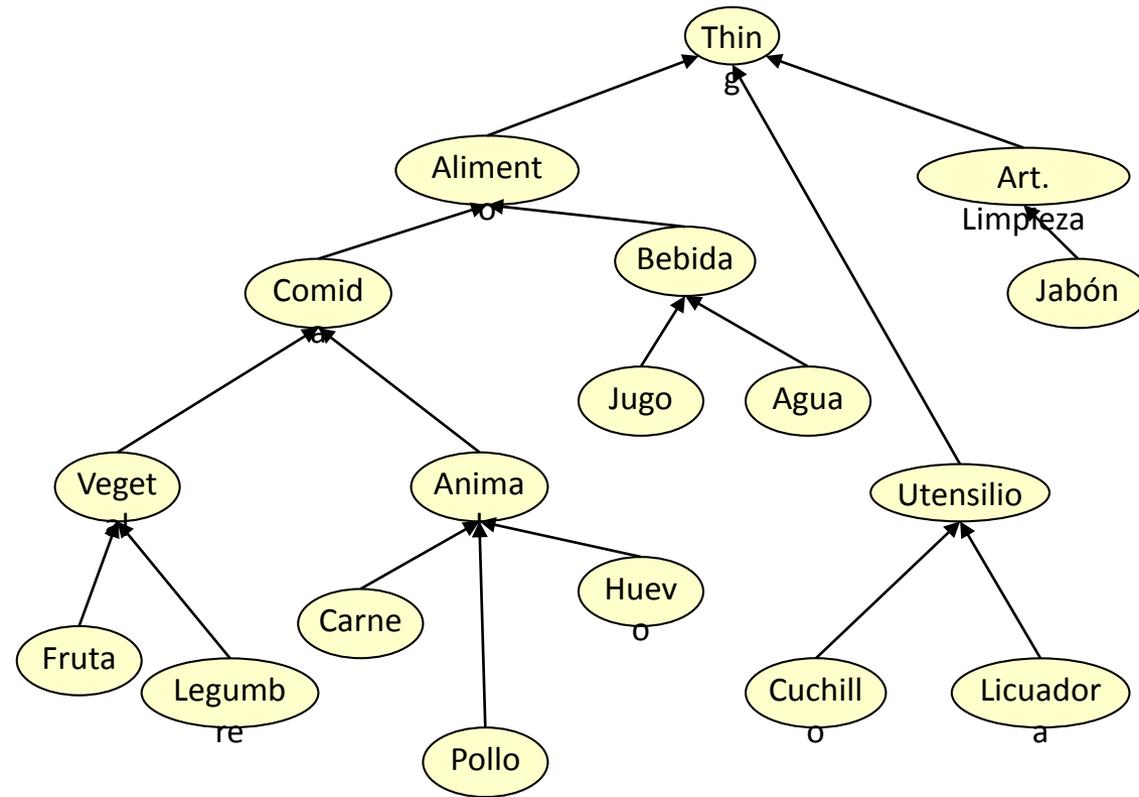
$$\Delta \gamma_{rs}^k = f(C(S_k))$$

S_k : concepto seleccionado por la hormiga k como mejor solución

Ontología Emergente

```
Repetir desde i=1 hasta i=Numero de Hormigas;
  Solución_Actual =“Thing”;
  Seguir=Verdadero;
  Repetir mientras (Seguir==Verdadero)
    Buscar hijos de Solución_actual();
    Si “Tiene hijos”
      Repetir desde j=1 hasta j=Numero de Hijos;
      Calcular Semejanza();
      Nueva_Solucion=Mayor_Semejanza();
      Si Nueva_Solucion==Solucion_Actual
        Seguir=Falso
      De lo Contrario
        Solucion_Actual=Nueva_Solucion
    Si “No Tiene hijos”
      Seguir=Falso
```

Ontología Emergente



Leche “es-un” Bebida (Alimento)
Aceite “es-un” Vegetal (Comida, Alimento)

Propiedades del Contexto		Nuevos Conceptos	
		Leche	Aceite
Características	Líquido	1	1
	Frío	1	0
Ubicaciones	Nevera	1	0
	Congelador	0	0
	Despensa	0	1
	Gabinete	0	0
Tareas	Preparar Comida	0	1
	Preparar Bebida	1	0
	Limpiar	0	0



Plataforma tecnológica de
Facebook

Facebook como estructura
social



Publicidad en Redes Sociales

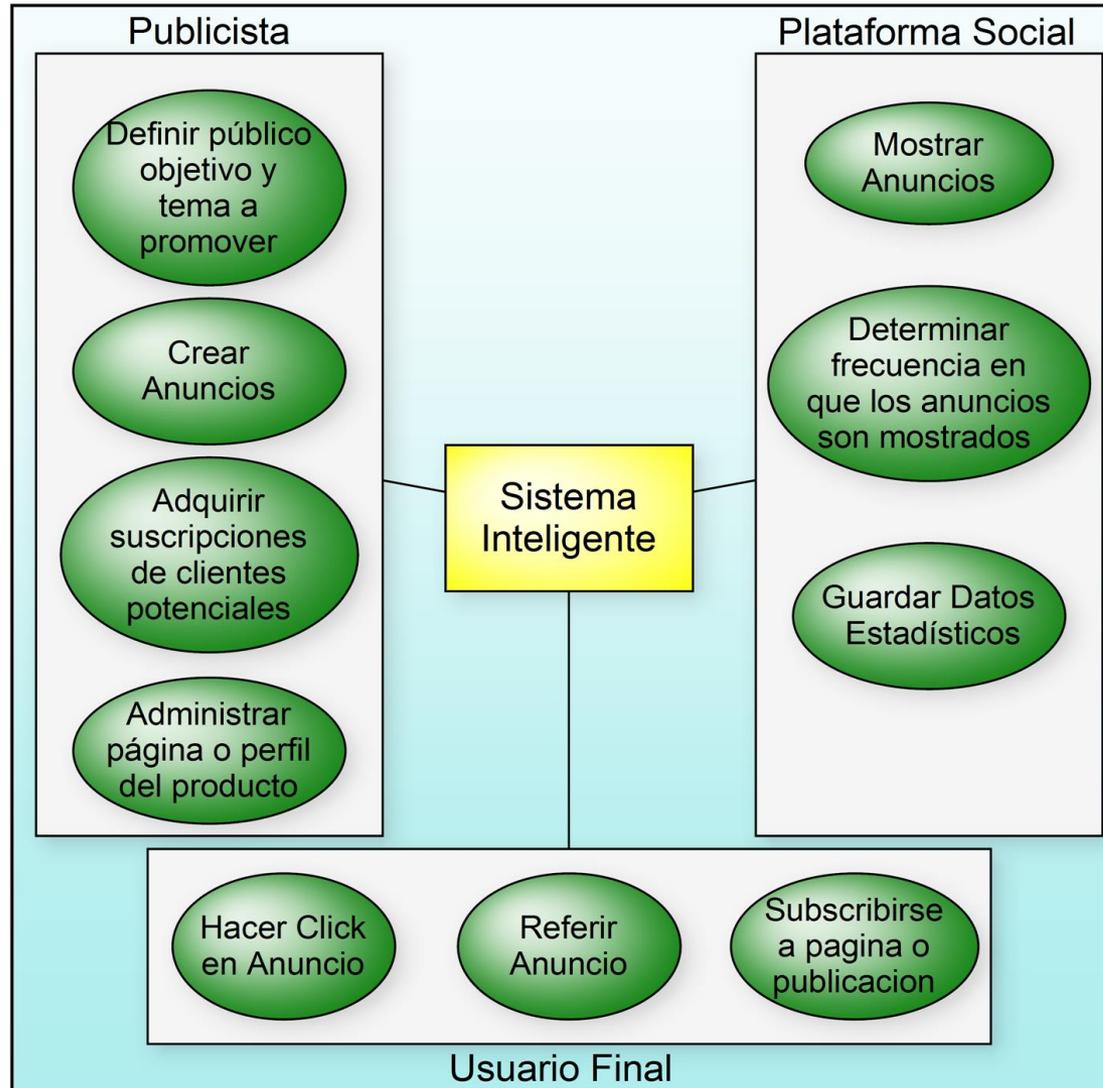
- Directa
 - Basada en la red de amigos
 - Colocada en páginas de redes sociales
- Indirecta

Componentes de la Publicidad en Redes Sociales

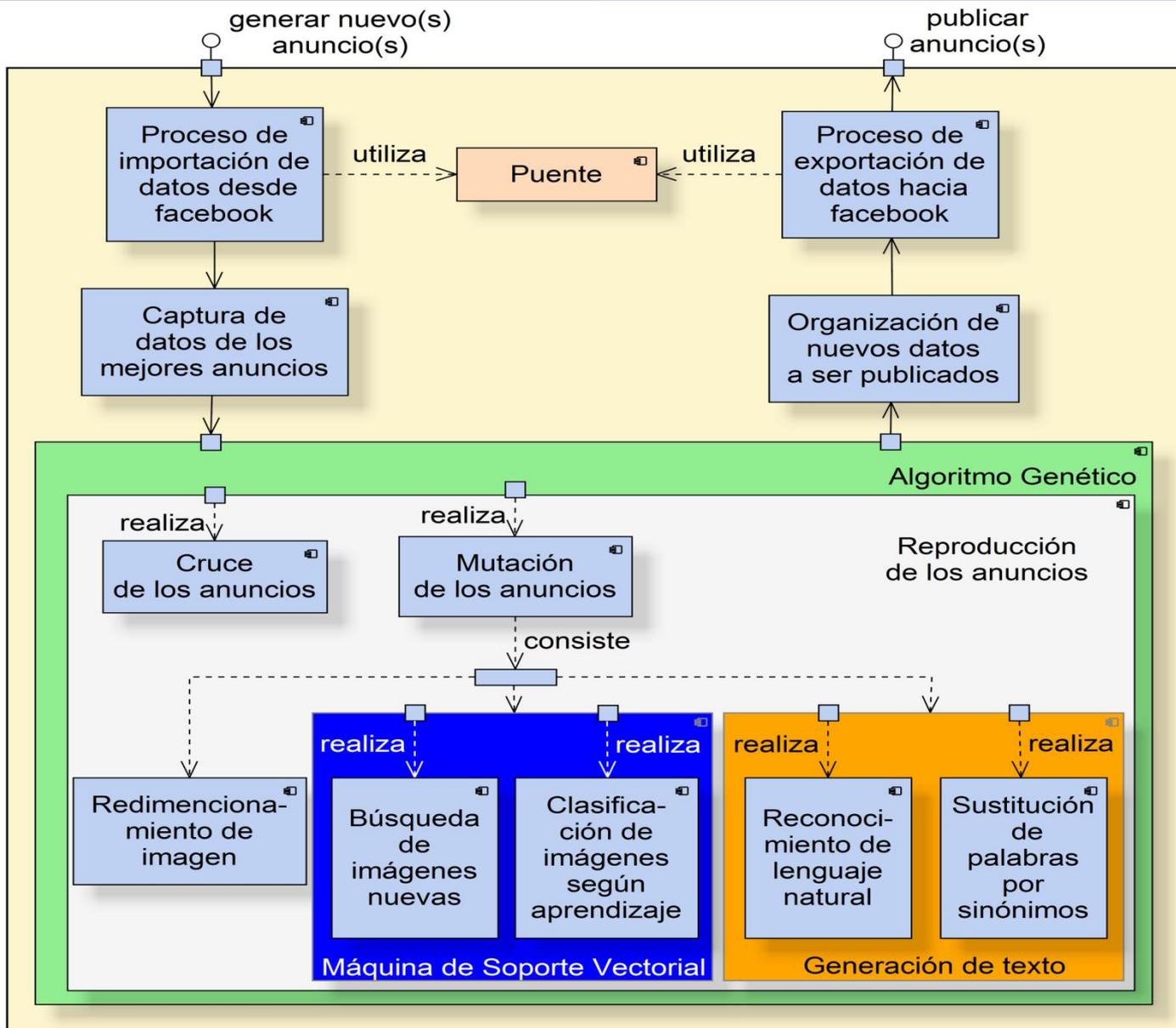
- Clicks
- Alcance
- Frecuencia
- Porcentaje de clicks del anuncio
- Interacciones
- Puja del anuncio



Sistema Adaptativo Inteligente



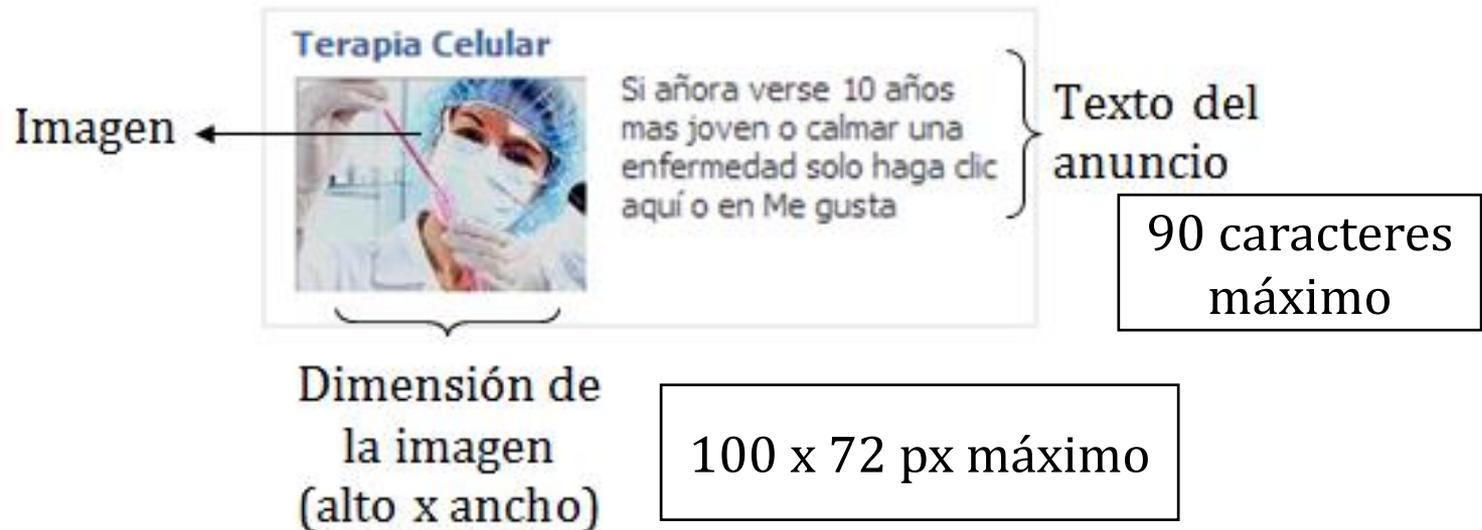
Arquitectura del Sistema



➤ Especificación Formal

✓ Algoritmo Genético

- Estructura del Individuo



- **Operadores Genéticos**

- ❖ *Cruce (contenido de imagen, dimensión de imagen, texto)*

Anuncio A

Células Madre Terapia Celular



Si desea verse 10 años mas joven o tratar una enfermedad solo haga clic aquí o en Me gusta

Anuncio B

Células Madre Terapia Celular



Si quiere verse 10 años mas joven o aliviar una enfermedad haga clic aquí o en Me gusta



Anuncio A

Células Madre Terapia Celular



Si desea verse 10 años mas joven o tratar una enfermedad solo haga clic aquí o en Me gusta

Anuncio B

Células Madre Terapia Celular

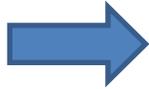
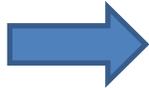
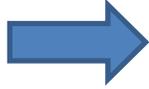


Si quiere verse 10 años mas joven o aliviar una enfermedad haga clic aquí o en Me gusta

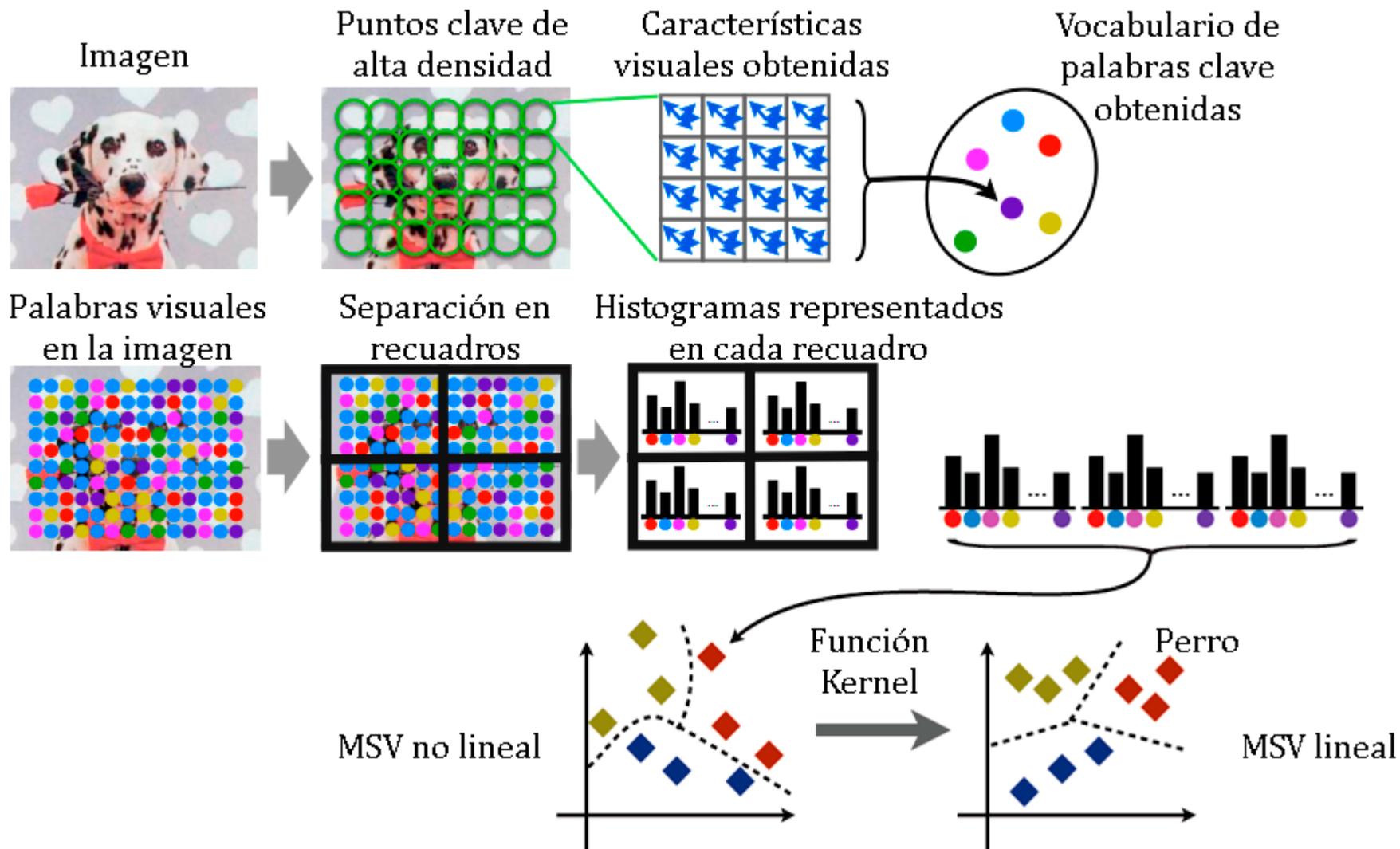
Cruce de imágenes entre dos anuncios

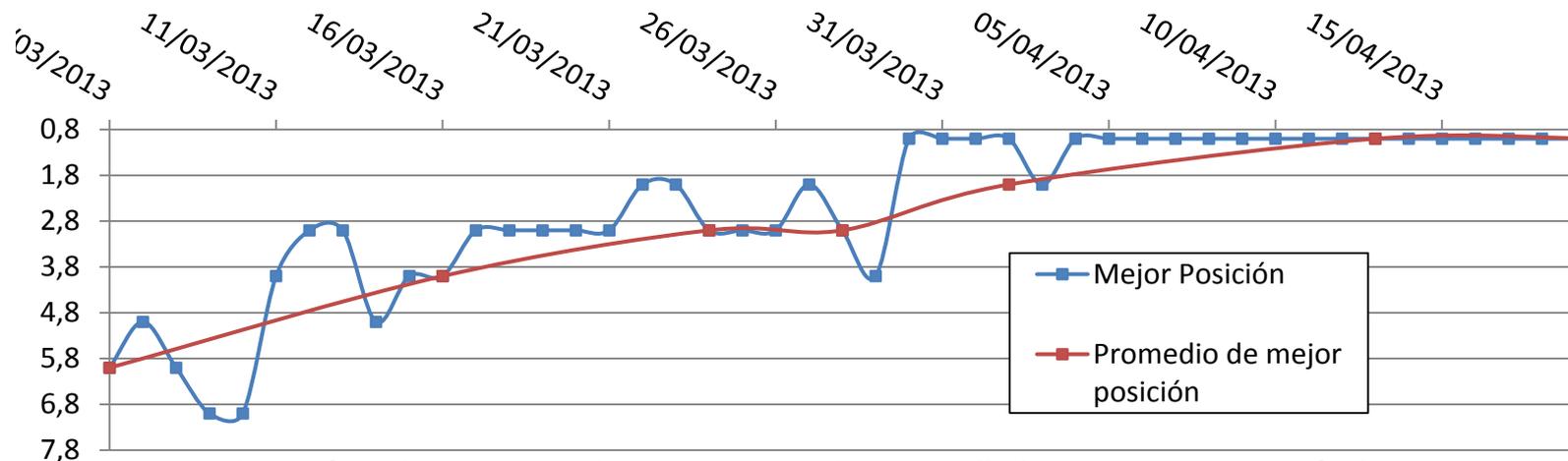
• Operadores Genéticos

❖ *Mutación*

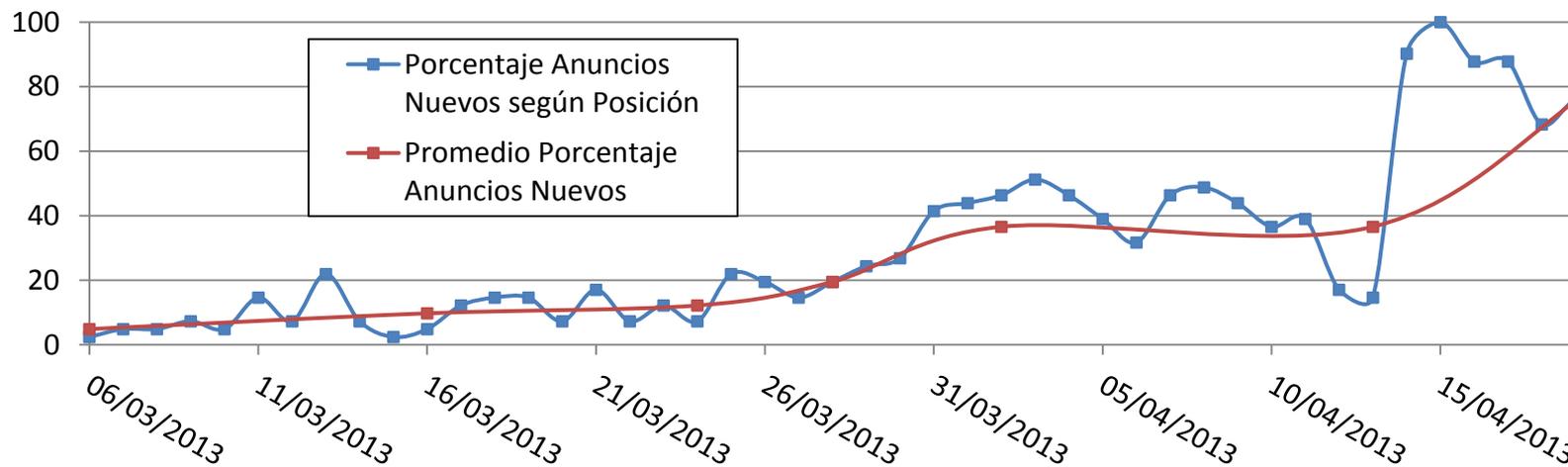
- Si se muta el texto  Generador de texto
- Si se muta la dimensión de la imagen  Se asigna un valor a la dimensión aleatoriamente
- Si se muta el contenido de la imagen  Máquina de Soporte Vectorial que escoge una imagen (nueva) adecuada

✓ Máquina de Soporte Vectorial





Evolución del caso de estudio “CMTC” (mejor posición)



Evolución del caso de estudio “CMTC”
 (porcentaje de anuncios nuevos entre los mejores 20)