

Problemas de Minería: minería de datos, minería semántica, minería de texto

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

Situación actual



- La revolución digital ha permitido que la captura de datos sea fácil, y su almacenamiento tenga un costo casi nulo.

Enormes cantidades de datos son recogidas y almacenadas en BD en la vida diaria.

- **Resultado:** Para analizar estas enormes cantidades de datos, las herramientas tradicionales de gestión de datos y las herramientas estadísticas no son adecuadas.

Situación actual



- Los datos por sí solos no producen beneficio directo.

Su verdadero valor consiste en poder extraer información útil para la toma de decisiones

- Tradicionalmente se analizaban datos con la ayuda de técnicas estadísticas (resumiendo y generando informes) o validando modelos sugeridos manualmente por los expertos.

Situación actual



- Estos procesos son irrealizables a medida que aumenta el tamaño de los datos.
- Bases de datos con un nº de registros del orden de 10^9 y 10^3 de dimensión, son fenómenos relativamente comunes.
- La tecnología informática puede automatizar este proceso.

Minería de datos

Definiciones iniciales



- **Datos:** hechos o medidas que describen características de objetos, eventos o personas, es la materia prima de la que se obtendrá la información.
- **Información:** Datos analizados y presentados en forma adecuada, de interés para un observador en un momento determinado.
- **Conocimiento:** información procesada para emitir juicios que llevan a conclusiones.
- **Meta Conocimiento:** Reglas que permiten obtener conocimiento.



Principales características de la información

- **Significado (semántica):**

¿Qué quiere decir?

Del significado extraído de una información, cada individuo evalúa las consecuencias posibles y adecúa sus actitudes y acciones de manera acorde a las consecuencias previsibles que se deducen del significado de la información.

- **Importancia (relativa al receptor):**

¿Trata sobre alguna cuestión importante?

La importancia de la información para un receptor, se referirá a en qué grado cambia la actitud o la conducta de los individuos.



Principales características de la información

- **Vigencia (en la dimensión espacio-tiempo):**

¿Es actual o desfasada?

En la práctica la vigencia de una información es difícil de evaluar, ya que en general acceder a una información no permite conocer de inmediato si dicha información tiene o no vigencia.

.

- **Validez (relativa al emisor):**

¿El emisor es fiable? ¿Puede entregar información no válida?

- **Valor (receptor):**

¿Cuán útil resulta para el destinatario la información?

.

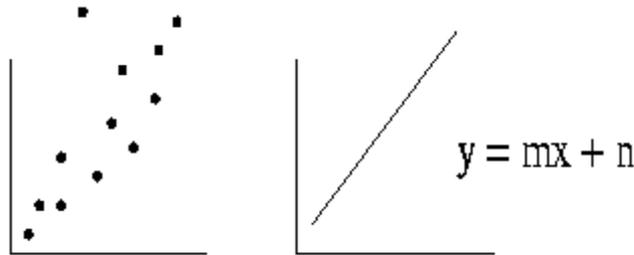
Definiciones iniciales



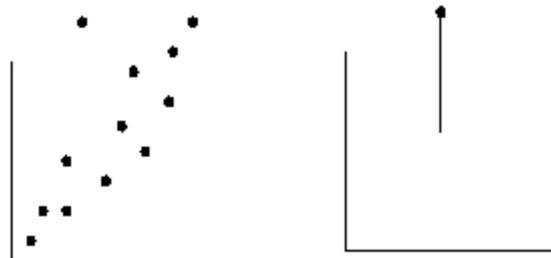
Conocimiento: Modelo vs. Patrón

Hand, Mannila y Smyth

Modelo: Habla de todo el conjunto de datos



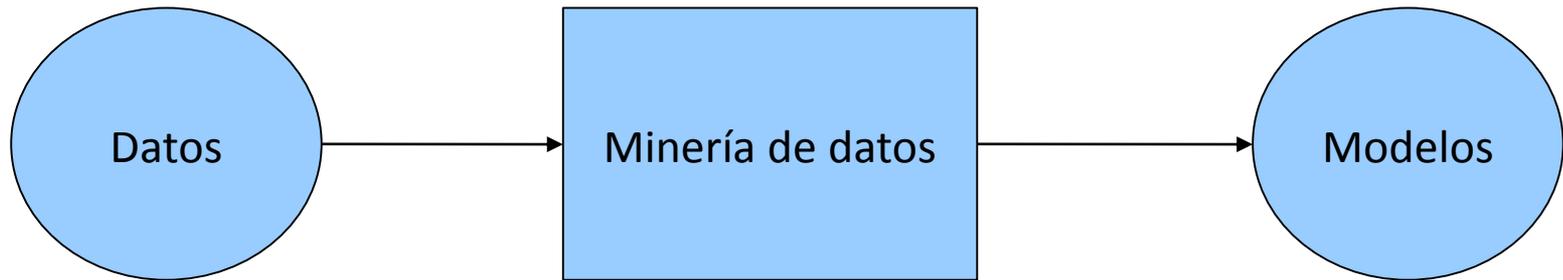
Patrón: Habla de una región particular de datos.



Datos y Modelos => Conocimiento

- Los datos se obtienen de:
 - Bases de datos (relacionales, espaciales, temporales, documentales, multimedia, etc)
 - World Wide Web
- **Modelos descriptivos:** identifican patrones que explican o resumen los datos:
 - Reglas de asociación
 - Clustering:
- **Modelos predictivos:** estiman valores de variables de interés (a predecir) a partir de valores de otras variables
 - Regresión
 - Clasificación

Una visión simplificada de la minería de datos



- Los “modelos” son el producto de la minería de datos...
- ...y dan soporte a las estrategias de decisión que se tomen



¿Qué es la Minería de Datos?

La minería de datos tiene como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten hacia la toma de decisión.

Minería de datos es la exploración y análisis de grandes cantidades de datos con el objeto de encontrar patrones y reglas significativas (conocimiento)



¿Qué es la Minería de Datos?

Análisis de grandes volúmenes de datos para encontrar relaciones no triviales, y para resumirlos de manera que sean entendibles y útiles.

Hand, Mannila y Smyth

Extracción de patrones y modelos interesantes, potencialmente útiles y datos en base de datos de gran tamaño.

Hand



¿Qué es la Minería de Datos?

- Data Mining (Minería de datos)
 - Proceso de extraer conocimiento util y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos (Witten and Frank, 2000)

- Knowledge Discovery in Databases – KDD (Descubrimiento de Conocimiento en Bases de Datos)
 - Proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles, a partir de los datos (Fayyad y col. 1996)



Minería de datos

- Proceso de utilizar datos “crudos” para inferir importantes relaciones entre ellos
- Colección de técnicas poderosas para analizar grandes volúmenes de datos
- No existe un solo enfoque para minería de datos sino un conjunto de técnicas que se pueden utilizar de manera independiente o en combinación

Lo que no es la minería de datos



Búsqueda de un número de teléfono en el directorio telefónico

Usar un motor de búsqueda en Internet para obtener información acerca de "Amazon"

La fuerza bruta crujido de datos a granel

La presentación de los datos de diferentes maneras

Una tarea intensiva de base de datos

Qué si es Data Mining?



- Una palabra de moda para una clase de técnicas que encuentran patrones en los datos
- Un conjunto de técnicas que se encuentran las relaciones que anteriormente no han sido descubiertos

La minería de datos es un campo multidisciplinario



Estadística vs Minería de datos

	Estadística	Mineria de datos
Construcción de modelos	Ceñido a premisas y teoremas	Mayor libertad en la construcción, interpretable
Búsqueda	Test de la razón de la verosimilitud	Metaheurísticos
Transparencia	Más complicados de interpretar	Más claros y sencillos
Validación	No	Sí

Minería de Datos vs. Estadística

Además, en minería de datos tenemos:

- 1.- Mayor volumen de datos (diferencia fundamental)
- 2.- Mayor complejidad de los datos (miles de atributos o dimensiones).
- 4.- Énfasis está en buscar modelos más que en verificar un modelo fijo.



OLAP vs Minería de datos

OLAP	Minería de datos
¿Cuál es la proporción media de accidentes entre fumadores y no fumadores?	¿Cuál es la mejor predicción para accidentes?
¿Cuál es la factura telefónica media de mis clientes y de los que han dejado la compañía?	¿Dejara X la compañía? ¿Qué factores afectan a los abandonados?
¿Cuánto es la compra media diaria de tarjetas robadas y legítimas?	¿Cuáles son los patrones de compra asociados con el fraude de tarjetas?

OLAP proporciona una muy buena vista de lo que está sucediendo, pero no se puede predecir lo que va a suceder en el futuro o por qué está sucediendo

¿Qué es la Minería de Datos?

- **Métodos Descriptivos**

Encontrar patrones interpretable que describen los datos.

- **Métodos de Predicción**

Utilizar algunas variables para predecir los valores desconocidos o futuros de otras variables.

¿Qué es la Minería de Datos?

La minería de datos se puede dividir en:

- **Minería de datos predictiva (mdp):** usa primordialmente técnicas estadísticas.
- **Minería de datos para descubrimiento de conocimiento (mddc):** usa principalmente técnicas de inteligencia artificial.

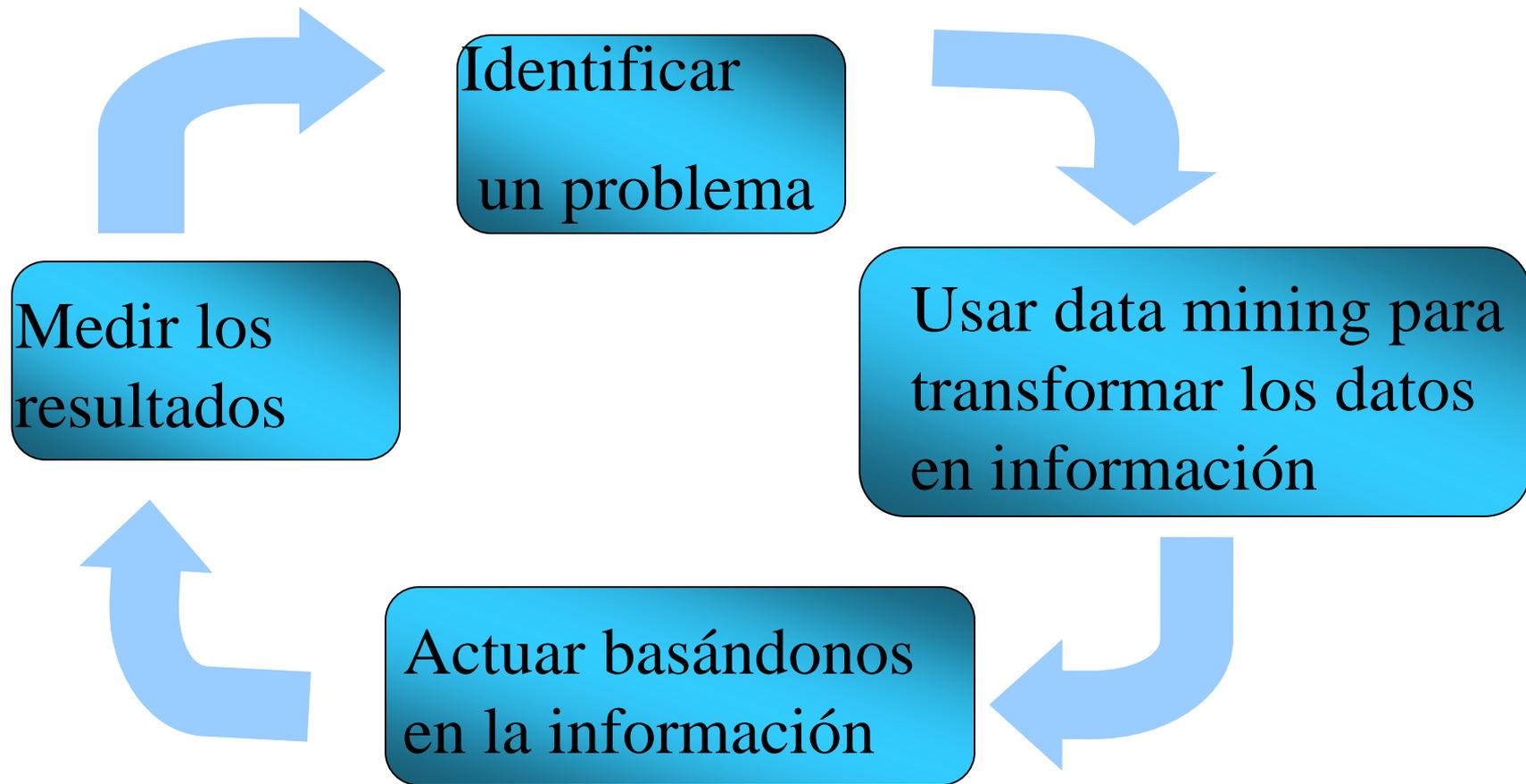
Tipos de aplicaciones de la minería de datos

- Clasificación [predictivo]
- Clustering [descriptivo]
- Descubrimiento Regla Asociación [descriptivo]
 - Análisis de dependencia de datos
 - correlación y causalidad
- Descubrimiento Patrones Secuenciales [descriptivo]
 - Análisis de series de tiempo, asociaciones secuenciales
- Regresión [predictivo]
- Tendencia y Desviación de detección [predictivo]
- Filtros Colaborativos [predictivo]
- Resumir
- Descripción de Conceptos-
 - Descripción de características
 - descripción discriminante

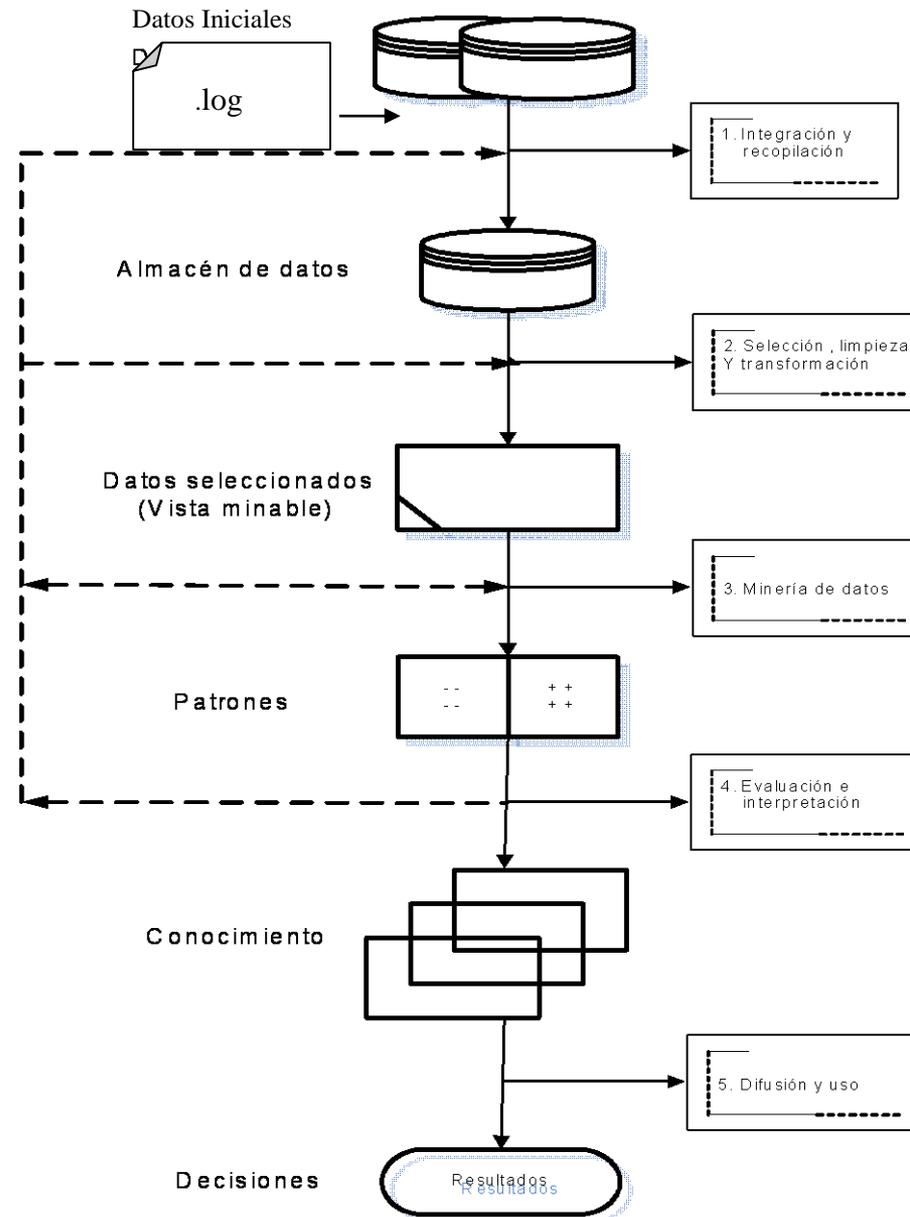
Importante

- La promesa de Data Mining es encontrar los patrones
- Simplemente el hallazgo de los patrones no es suficiente
- Debemos ser capaces de entender los patrones, responder a ellos, actuar sobre ellos,

El ciclo de data mining



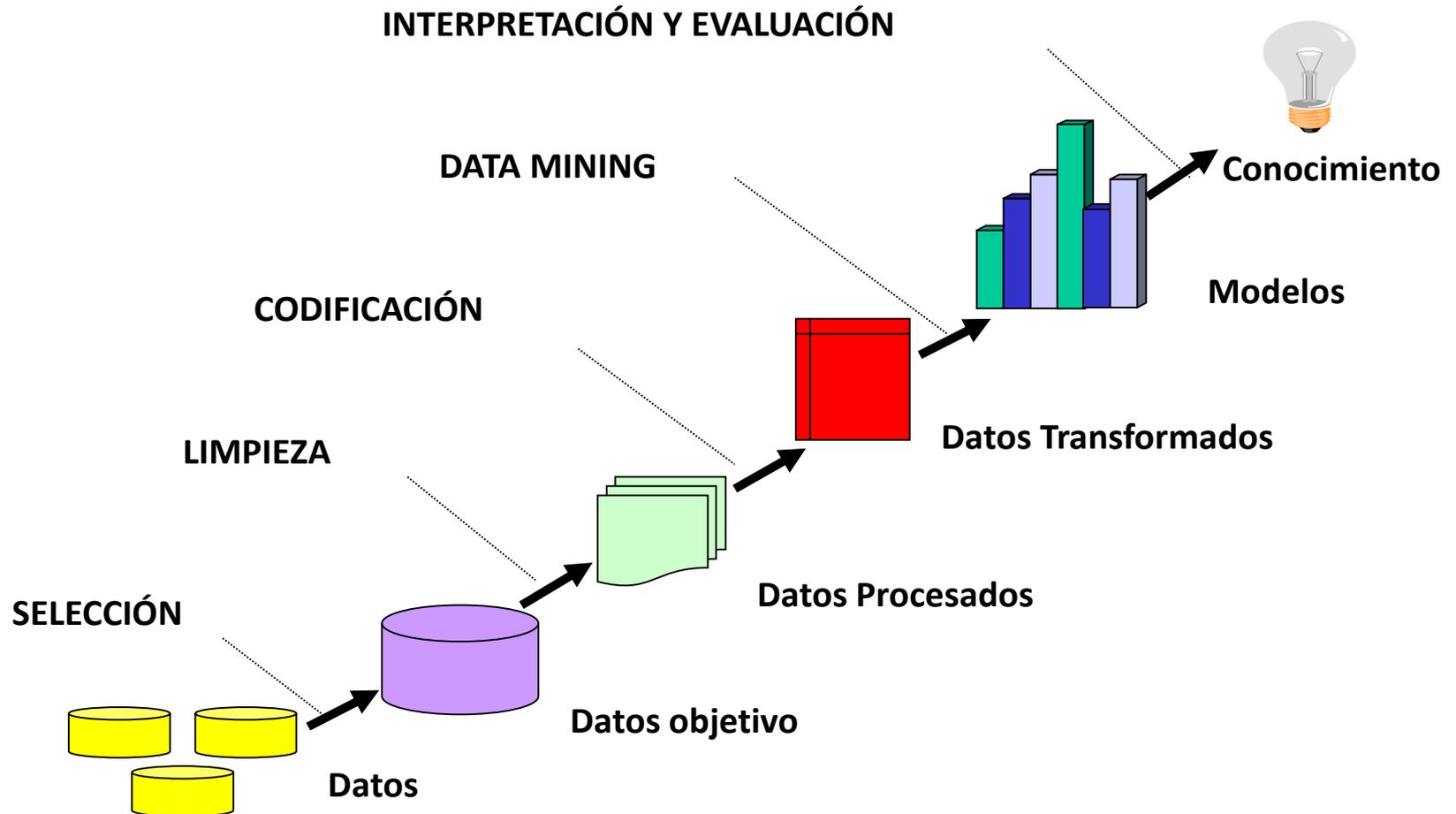
Esquema básico: Minería de datos



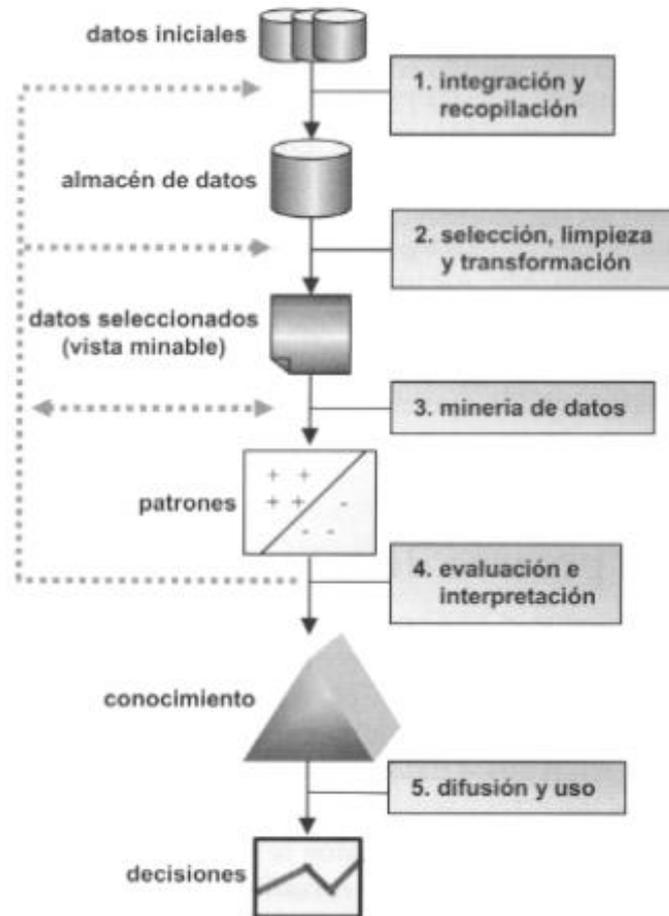
KDD – Knowledge Discovery in Databases

- La minería de datos se encuadra dentro de un proceso mucho mayor conocido como KDD (Knowledge Discovery from Databases)
- Fases del proceso iterativo e interactivo
 1. Integración y recopilación de datos
 2. Selección, limpieza y transformación
 3. Minería de datos
 4. Evaluación e interpretación
 5. Difusión y uso

El Proceso de KDD



KDD



Fase 1

- Integración y recopilación de datos
 - Decisiones estratégicas basadas en el análisis, la planificación y la predicción: datos en varios departamentos
 - Cada fuente de datos distintos formatos de registro, diferentes grados de agregación, diferentes claves primarias,
 - Integración de múltiples bases de datos: almacenes de datos (data warehousing)
 - Almacén de datos aconsejable cuando el volumen de información es grande. No estrictamente necesario (archivos de texto, hojas de cálculo, ...)

Fase 2

- Selección, limpieza y transformación
 - Calidad del conocimiento descubierto depende (además del algoritmo de minería) de la calidad de los datos analizados
 - Presencia de datos que no se ajustan al comportamiento general de los datos (*outliers*)
 - Presencia de datos perdidos (*missing values*)
 - Selección de variables relevantes (*feature subset selection*)
 - Selección de casos aleatoria en bases de datos de tamaño ingente
 - Construcción de nuevas variables que faciliten el proceso de minería de datos
 - Discretización de variables continuas

Fase 3

- Minería de datos
 - Modelos descriptivos
 - Reglas de asociación
 - Clustering
 - Modelos predictivos
 - Regresión: regresión lineal, regresión logística
 - Clasificación supervisada: clasificadores Bayesianos, redes neuronales, árboles de clasificación, inducción de reglas, K-NN, combinación de clasificadores

Fase 4

- Evaluación e interpretación
 - Técnicas de evaluación: validación simple (training + test), validación cruzada con *k-fold*, bootstrap
 - Reglas de asociación: cobertura (soporte), confianza
 - Clustering: variabilidad intra y entre
 - Regresión: error cuadrático medio
 - Clasificación supervisada: porcentaje de bien clasificados, matriz de confusión, análisis ROC
 - Modelos precisos, comprensibles (inteligibles) e interesantes (útiles y novedosos)

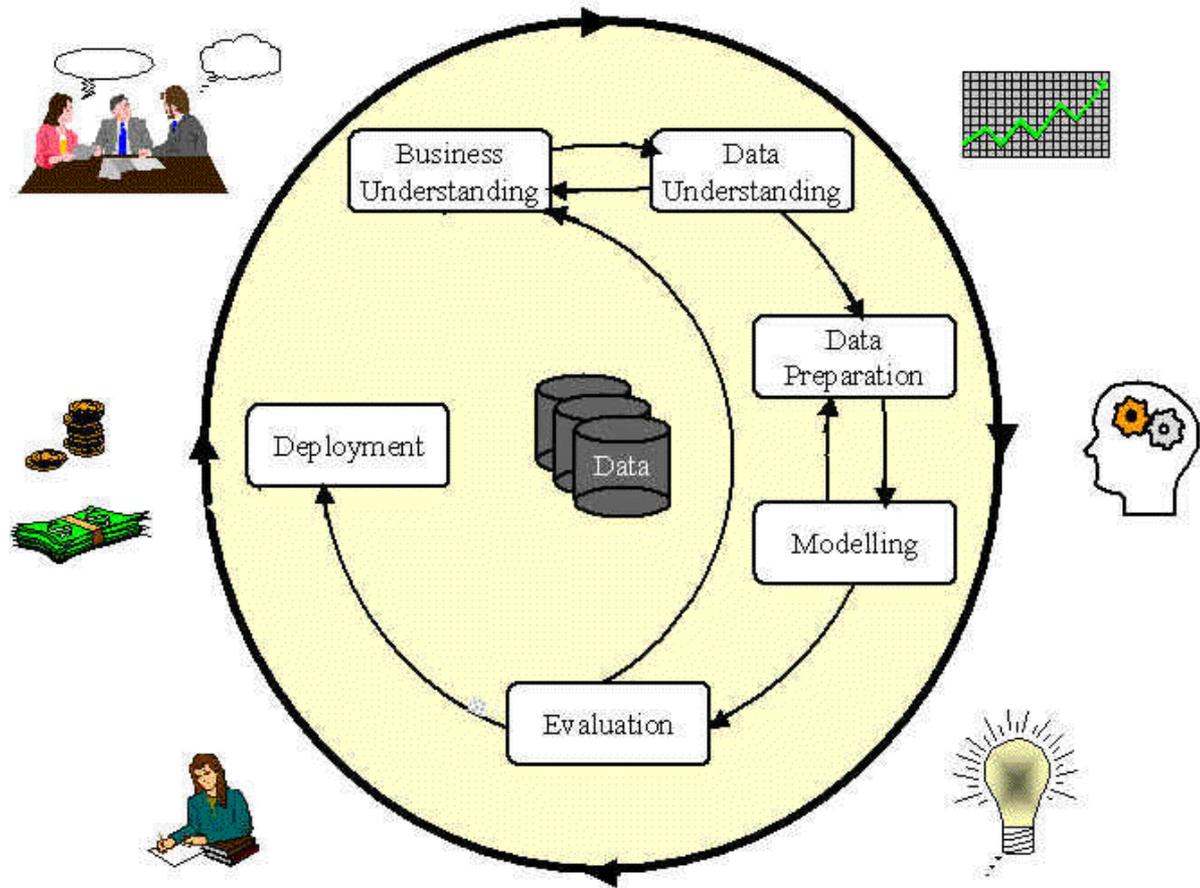
Fase 5

- Difusión y uso
 - Difusión: necesario distribuir, comunicar a los posibles usuarios, integrarlo en el *know-how* de la organización
 - Medir la evolución del modelo a lo largo del tiempo (patrones tipo pueden cambiar)
 - Modelo debe cada cierto tiempo de ser:
 - Reevaluado
 - Reentrenado
 - Reconstruido

CRISP-DM

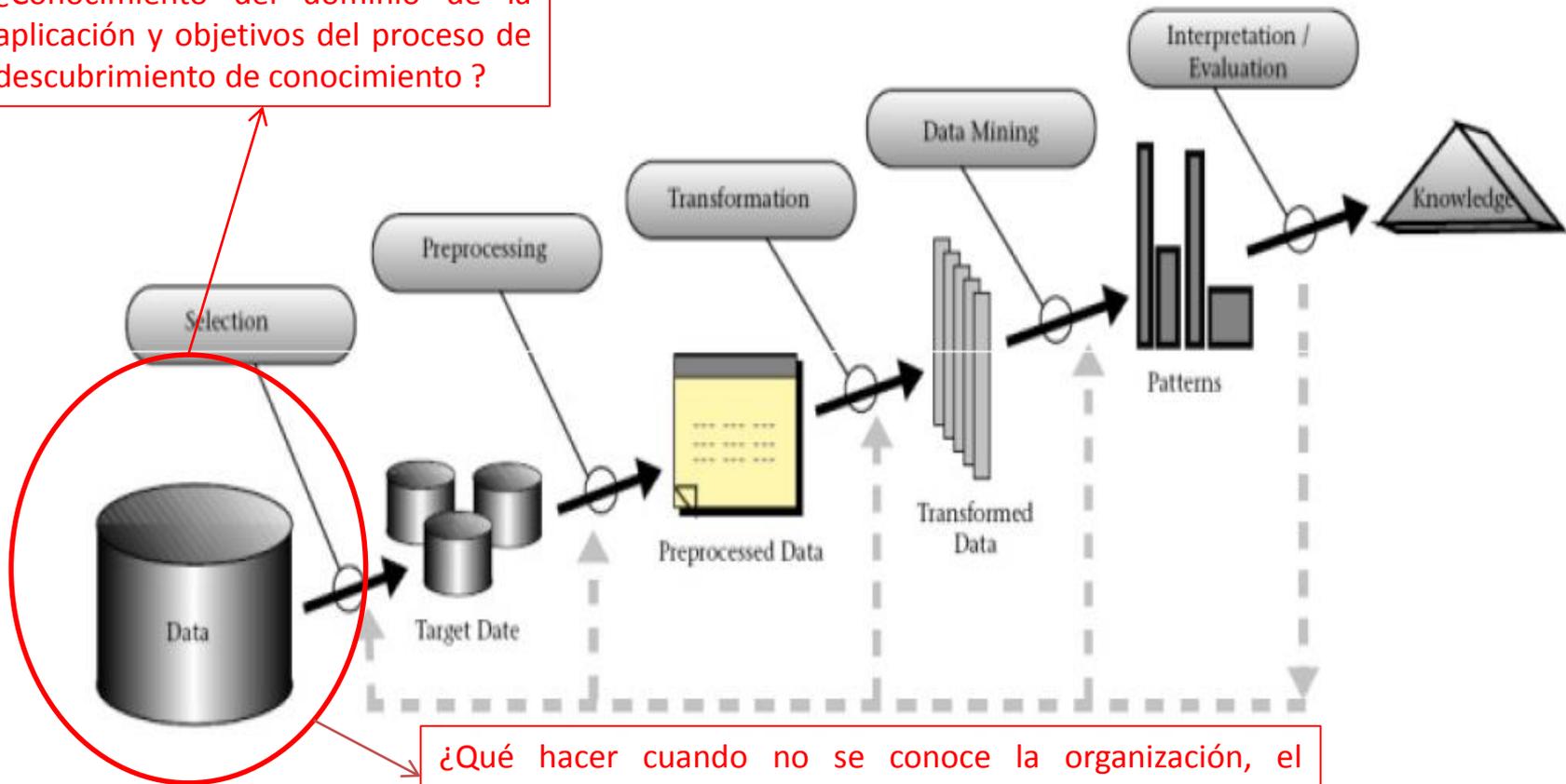
- El proceso de extracción de datos debe ser fiable y repetible por personas con pocos conocimientos de minería de datos
- CRISP-DM proporciona un marco uniforme para
 - directores
 - documentación de la experiencia
- CRISP-DM es flexible para tener en cuenta las diferencias
Diferentes problemas de negocio
Datos diferentes

Fases en CRISP-DM



Problemas

¿Conocimiento del dominio de la aplicación y objetivos del proceso de descubrimiento de conocimiento?



¿Qué hacer cuando no se conoce la organización, el problema, o los procesos a estudiar?

Ubicua e Invisible Minería de datos

- **Ubicuo Data Mining**

- La minería de datos se utiliza en todas partes, por ejemplo, las compras en línea

- **Invisible Minería de datos**

- Invisible: Las funciones de minería de datos se construyen en las operaciones de la vida diaria
- Ex. Google search: Los usuarios pueden no ser conscientes de que están examinando los resultados devueltos por los datos
- La minería de datos invisible es muy conveniente ?
- Minería Invisible debe considerar la eficiencia y la escalabilidad, la interacción del usuario, la incorporación de conocimientos y técnicas de visualización, la búsqueda de patrones interesantes, en tiempo real, ...

- **Trabajo adicional:**

- La integración de la minería de datos en tecnologías empresariales y científicos existentes para proporcionar herramientas de minería de datos específicos del dominio

Privacidad, Seguridad e Impactos Sociales de Data Mining

- Muchas aplicaciones de minería de datos no tocan datos personales
Por ejemplo, la meteorología, la astronomía, la geografía, la geología, la biología, y otros datos científicos y de la ingeniería
- Muchos estudios DM desarrollan algoritmos escalables para encontrar patrones generales o estadísticamente significativas, sin tocar los individuos
- El verdadero problema de privacidad: el acceso sin restricciones a los registros de individuo, la información especialmente sensible a la privacidad
- Método de preservación de privacidad de minería de datos
 - Cifrado: por ejemplo, firmas ciegas, encriptación biométrica y bases de datos anónimos (información personal se cifra y se almacena en diferentes lugares)

Métodos de preservación de privacidad de minería de datos

- preservación de Privacidad (sensible a la privacidad):
 - La obtención de los resultados de minería válidos sin dar a conocer los valores de los datos sensibles subyacentes
- métodos de minería de datos para preservación de Privacidad :
 - La asignación al azar (por ejemplo, perturbación): añadir ruido a los datos con el fin de enmascarar algunos valores de atributo de los registros
 - K-anonimato y l-diversidad: Alterar registros individuales de manera que no se pueden identificar de forma única
 - k-anonimato: Los mapas registrados guardados en por lo menos otros k registros
 - l-diversidad: la aplicación de la diversidad dentro del grupo de los valores sensibles
 - Preservación Distribuido privacidad: Los datos particionados y distribuidos, ya sea horizontal, vertical o una combinación de ambos
 - Modificar los datos o resultados de la minería, por ejemplo, ocultando algunas reglas de asociación o distorsionar ligeramente algunos modelos de clasificación

Minería en otras clases de Datos

- Minería de Datos Espaciales
- Minería espacio-temporal y Objectivos en movimiento
- Minería Cyber-físico de datos del sistema: salud, control de tráfico aéreo, simulación de inundaciones
- Minería de datos multimedia
- Minería de datos de texto
- Minería de datos Web
- Minería de datos Streams

Minería de Secuencia de Datos

- Buscar Similitud en serie temporal de datos
- Regresión y Análisis de Tendencias en series temporales de datos
- Minería Patrón Secuencial en secuencias simbólicas
- Clasificación de Secuencia
- Alineación de secuencias biológicas

Minería de Gráfos y Redes

- Minería de Patrón de Gráfo
- Modelado estadístico de Redes
- Agrupación y clasificación de grafos y redes homogéneas
- Agrupación, clasificación de las Redes heterogéneos
- Descubrimiento, clases, y Predicción de Enlace en Redes de Información
- Búsqueda de Similitud en Redes de Información:
- Evolución de las redes de información social

Minería avanzada

- Lazo autónomo de Analítica de Aprendizaje y Académica
- Servicios de Analítica de Aprendizaje Sociales
 - Minería Web, big data
- Ontologías Emergentes
 - Minería Ontológica
- Sistema de recomendación inteligente
 - Basado en mecanismos de razonamiento
- Sistema de reconocimiento emocional
 - Crónicas
- Análisis del proceso de aprendizaje
 - Serious Games, Aprendizaje Conectivista

Framework para Kaas

Mucha minería semántica

Abridged Declaration of Independence

A Declaration By the Representatives of the United States of America, in General Congress Assembled.

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident; that all men are created equal and independent; that from that equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying it's foundation on such principles and organizing it's power in such form, as to them shall seem most likely to effect their safety and happiness. Prudence indeed will dictate that governments long established should not be changed for light and transient causes: and accordingly all experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies; and such is now the necessity which constrains them to expunge their former systems of government. the history of his present majesty is a history of unremitting injuries and usurpations, among which no one fact stands single or solitary to contradict the uniform tenor of the rest, all of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith yet unsullied by falsehood.

How many “big”, “medium”, and “small” words are used?

Histograma de longitud de palabras

- Mucho (amarillo)= 10 + letras
- Medio (rojo)= 5 a 9 letras
- Poco (azul)= 2 a 4 letras
- Morado= 1 letra

ANNOUNCED DECLARATION OF INDEPENDENCE
A Declaration By the Representatives of the United States of America, in General Congress Assembled.

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident; that all men are created equal and independent; that from that equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying it's foundation on such principles and organizing it's power in such form, as to them shall seem most likely to effect their safety and happiness. Prudence indeed will

dictate that governments long established should not be changed for light and transient causes; and accordingly all experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies; and such is now the necessity which constrains them to expunge their former systems of government. the history of his present majesty is a history of unremitting injuries and usurpations, among which no one fact stands single or solitary to contradict the uniform tenor of the rest, all of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith yet unsullied by falsehood.

Histograma de longitud de palabras

Abridged Declaration of Independence

Mapa 1
204 palabras

A Declaration By the Representatives of the United States of America, in General Congress Assembled.

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident; that all men are created equal and independent; that from that equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying it's foundation on such principles and organizing it's power in such form, as to them shall seem most likely to effect their safety and happiness. Prudence indeed will

- Amarillo, 17
- Rojo, 17
- Azul, 107
- Morado, 3

dictate that governments long established should not be changed for light and transient causes: and accordingly all experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies; and such is now the necessity which constrains them to expunge their former systems of government. the history of his present majesty is a history of unremitting injuries and usurpations, among which no one fact stands single or solitary to contradict the uniform tenor of the rest, all of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith yet unsullied by falsehood.

- Amarillo, 20
- Rojo, 71
- Azul, 93
- Morado, 6

Mapa 2
190 palabras

Histograma de longitud de palabras

Mapa 1

A Declaration By the Representatives of the United States of America, in General Congress Assembled.

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident, that all men are created equal and independent, that from that equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed, that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying it's foundation on such principles and organizing it's power in such form, as to them shall seem most likely to effect their safety and happiness. Prudence indeed will

Amarillo, 17
Rojo, 17
Azul, 107
Morado, 3

Mapa 2

dictate that governments long established should not be changed for light and transient causes: and accordingly all experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies, and such is now the necessity which constrains them to expunge their former systems of government. the history of his present majesty is a history of unremitting injuries and usurpations, among which no one fact stands single or solitary to contradict the uniform tenor of the rest, all of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith yet unswerving by falsehood.

Amarillo, 20
Rojo, 71
Azul, 93
Morado, 6

“Shuffle step”

Reducción

• Amarillo, 37

• Rojo, 148

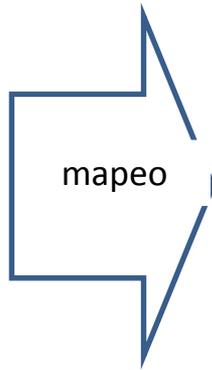
• Azul, 200

• Morado, 9

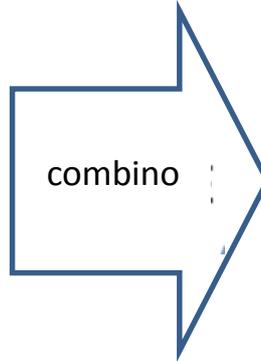


Simple análisis redes sociales: Contar amigos

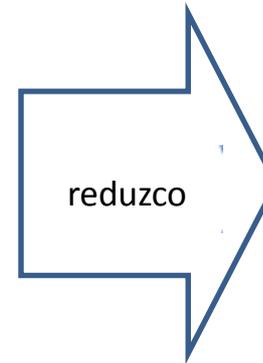
Entrada
Jim,Sue
Sue,Jim
Lin,Joe
Joe,Lin
Jim,Kai
Kai,Jim
Jim,Lin
Lin,Jim



Jim,1
Sue,1
Lin,1
Joe,1
Jim,1
Kai,1
Jim,1
Lin,1



Jim,(1,1,1)
Lin,(1,1)
Soe,(1)
Joe,(1)
Kai,(1)



Salida
Jim, 3
Lin,2
Soe,1
Joe,1
Kai,1

Minería de Texto

DM



Dato estructurado

HomeLoan (
Loanee: Frank Rizzo
Lender: MWF
Agency: Lake View
Amount: \$200,000
Term: 15 years
)

Multimedia



Loans(\$200K,[map],...)

texto libre

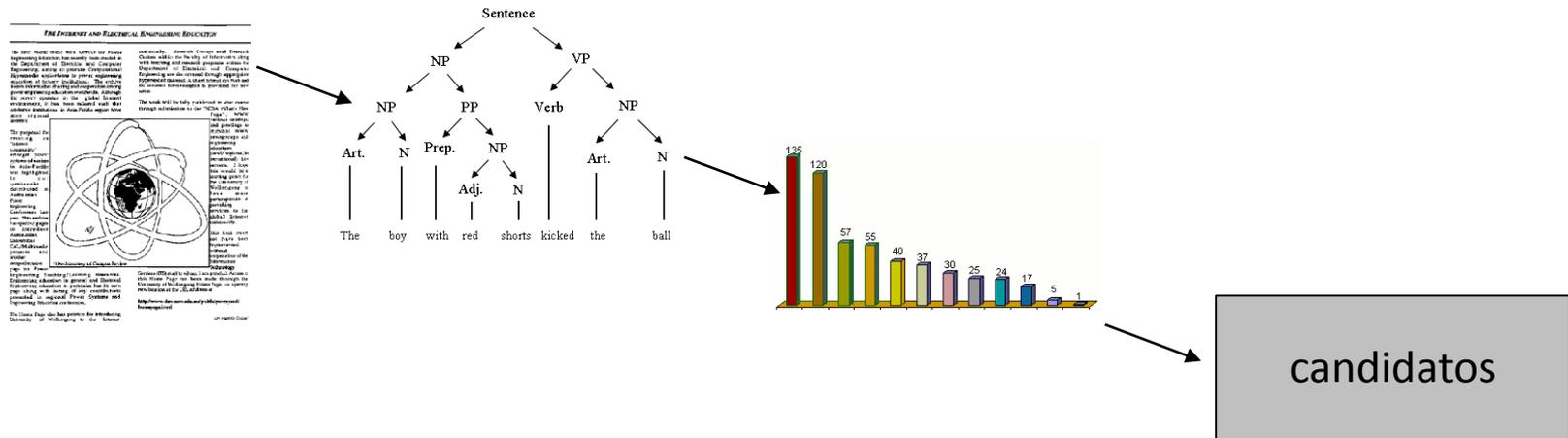
Frank Rizzo bought his home from Lake View Real Estate in 1992.
He paid \$200,000 under a 15-year loan from MW Financial.

Hypertexto

[Frank Rizzo](#)
Bought
[this home](#)
from [Lake View Real Estate](#)
In **1992**.
...

Minería de Textos

- Desarrollo y explotación de corpus lingüísticos.
- Reconocimiento de patrones lingüísticos.
- Explotación de métodos y recursos estadísticos.



Etapas de la minería de texto

1. **Selección de documentos** implica la identificación y recuperación de los documentos potencialmente relevantes de un conjunto grande (por ejemplo, Internet).
2. **Pre-tratamiento documento** incluya la limpieza y la preparación de los documentos, por ejemplo, eliminación de información extraña, corrección de errores, la normalización ortográfica, tokenización, etiquetado, etc.
3. **Procesamiento de documentos** consiste principalmente en la extracción de información. Para la Web Semántica es extracción de metadatos

Técnicas de la minería de texto

1. Selección y filtrado de documentos

técnicas de Recuperación de Información (RI)

1. Pre-procesamiento de documentos

técnicas de PLN

1. Procesamiento de documentos

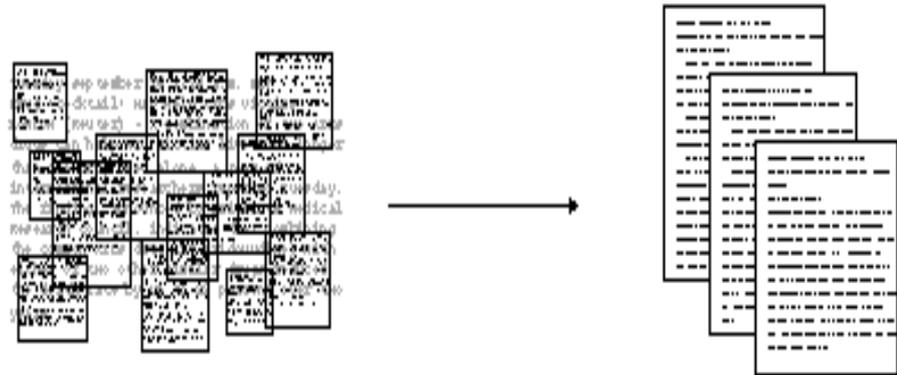
técnicas PNL/estadísticas/IA

Extracción de metadatos

- **Extracción de metadatos explícita** involucra información que describe el documento, como la información que contiene la cabecera de los documentos HTML (títulos, resúmenes, autores, fecha de creación, etc.)
- **Extracción de metadatos implícita** implica información semántica que se deduce, es decir, información endógena como los nombres de las entidades y las relaciones contenidas en el texto. Esto implica esencialmente las técnicas de extracción de información, a menudo con la ayuda de una ontología.

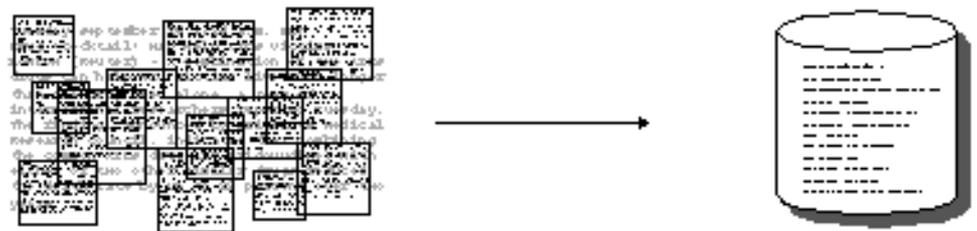
Extracción Inform. (EI) no es Recuperación de Inform. (RI)

RI busca información a partir de grandes colecciones de textos (por lo general, la Web) en respuesta a palabras clave o consultas específicas



Se analizan los documentos recuperados

EI saca hechos e información estructurada de los contenidos de las grandes colecciones de textos



Se analizan los hechos

Minería Semántica: Perspectivas Actuales

Jose Aguilar

CEMISID,
Dpto. de Computación
Facultad de Ingeniería

Minería Semántica



Determinar relaciones semánticas

- Minería de Datos Semánticos (*Semantic Data Mining*)
- Minería de la Web Semántica (*Semantic Web Mining*)
- Minería Ontológica (*Ontology Mining*).
- Minería de Texto

Análisis de las redes de aprendizaje
Análisis del contenido de aprendizaje
Análisis del contexto de aprendizaje

La Minería de Datos es un área bastante madura en las Ciencias Computacionales, cuyo principal objetivo es la extracción de conocimiento,.



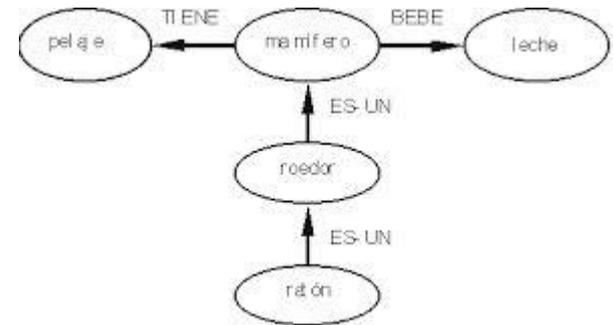
La Minería de Datos ha requerido ser enriquecido estos últimos años, debido a la necesidad de incorporar **contenido semántico**.

Minería Semántica

Minería Semántica



	A	B	C	D	E
1	NOMBRES	CARGO	TELEFONOS	LOCALIDAD	SUELDO
2	Daniela Cárdenas	Chef	3166294789-2574986	ENGATIVA	\$ 1.700.000
3	Gabriela Reyes	Subchef	327459836-4354822	SAN CRISTOBAL	\$ 110.000
4	Carmen Vanegas	Enologo	3154689857-2157458	KENEDDY	\$ 950.000
5	Cristina Pomras	Chef Pastelera	3146874953-6874235	BOSA	\$ 130.000
6	Liliana Cruz	Chef Panadera	3201478951-7451825	SUBA	\$ 1.500.000
7	Paola Cristancho	Soucier	3157489514-4785126	CHAPINERO	\$ 800.000
8	Camila Davalos	Cajera	3214875961-7584621	TEUSQUILLO	\$ 700.000
9	Lina Bohorquez	Mesera	3012574816-2245783	CANDELARIA	\$ 600.000
10	Pamela Carrasco	Mesera	3157485912-2485796	CANDELARIA	\$ 600.000
11	Lorena Valencia	Mesera	3204578963-2487512	ENGATIVA	\$ 600.000
12	Jairo Arevalo	Parquesidero	3002157459-2861459	BOSA	\$ 489.500
13				TOTAL	\$ 8.199.500



Minería Semántica

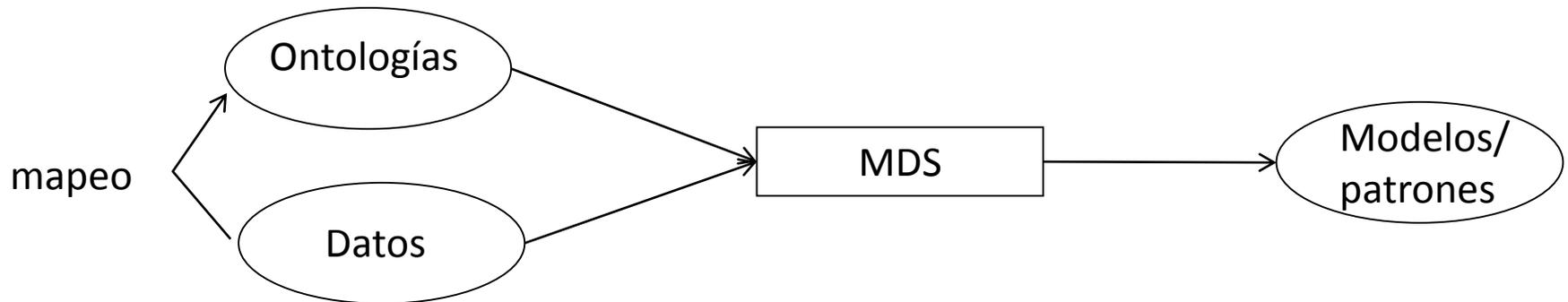


La minería semántica se encarga de extraer conocimiento semántico desde **diferentes fuentes semánticas**,

- Páginas web,
- Contenido sin estructura en la web,
- Contenido estructurado en la web,
- Grafos anotados,
- Ontologías,
- Tabla de Datos, entre otros

Minería de Datos Semánticos (MDS)

- Desafíos de la Minería de Datos (MD): incorporar conocimiento de un dominio desde los datos.
Minar recursos anotados semánticamente, con ontologías utilizadas como conocimientos de fondo
- Añadir contenido semántico a/desde los datos usando técnicas de MD para la extracción de conocimiento (en este caso, la fuente es contenido semántico).



Minería de Datos Semánticos

- El proceso de MDS se da en dos pasos,
 1. Enriquecimiento semántico,
 2. Aplicación de técnicas de MD como tal.
- En el primer paso se usan ontologías, o cualquier contenido semántico, y se realiza un mapeo con la data que se va a trabajar, almacenada normalmente en bases de datos.
- En el segundo paso se aplican técnicas de MD para buscar patrones, relaciones, y en general, cualquier operación que explote el enriquecimiento semántico.

conocimiento

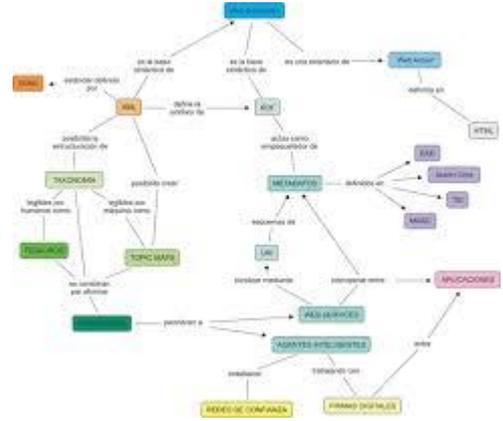
Minería de Datos Semánticos

Dado: tabla de datos de transacciones, bases de datos relacionales, documentos de texto, páginas Web, ... una o más ontologías de dominio

	A	B	C	D	E
1	NOMBRES	CARGO	TELEFONOS	LOCALIDAD	SUELDO
2	Daniela Cárdenas	Chef	3166294789-2574986	ENGATIVA	\$ 1.700.000
3	Gabriela Reyes	Subchef	327459836-4354822	SAN CRISTOBAL	\$ 110.000
4	Carmen Vanegas	Enologo	3154689857-2157458	KENEDDY	\$ 950.000
5	Cristina Porras	Chef Pastellera	3146874953-6874235	BOSA	\$ 130.000
6	Liliana Cruz	Chef Panadera	3201478951-7451825	SUBA	\$ 1.500.000
7	Paola Crisitancho	Soucier	3157489614-4785126	CHAPINERO	\$ 800.000
8	Camila Davalos	Cajera	3214875961-7584621	TEUSQUILLO	\$ 700.000
9	Lina Bohorquez	Mesera	3012574818-2245783	CANDELARIA	\$ 600.000
10	Pamela Carrasco	Mesera	3157485912-2485796	CANDELARIA	\$ 600.000
11	Lorena Valencia	Mesera	3204578963-2487512	ENGATIVA	\$ 600.000
12	Jairo Arevalo	Parquedero	3002157459-2861459	BOSA	\$ 489.500
13			TOTAL		\$ 8.199.500



Encontrar: un modelo de clasificación, un conjunto de patrones



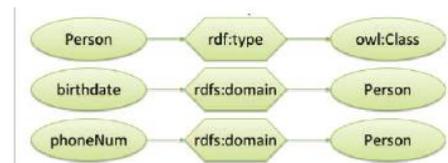
Minería de Datos Semánticos

- **Actual escenario de la MDS:** Minería de datos **empíricos** con ontologías como conocimiento de fondo
 - Abundantes datos empíricos,
 - Escaso conocimiento de fondo
- **Futuro escenario de MDS:**
 - Volumen creciente de ontologías y colecciones de datos semánticamente anotados
 - más de 6 billones de tripletas RDF
 - más de 200 millones de enlaces

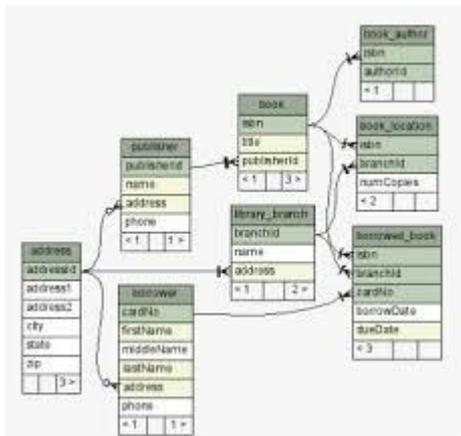
Definición relacional



Ontología



Minería de Datos Semánticos



Paso 1

1. Definir características Relacionales
2. construir una tabla proposicional

	f1	f2	f3	f4	f5	f6	fn
g1	1	0	0	1	1	1	0	0	1	0	1
g2	0	1	1	0	1	1	0	0	0	1	0
g3	0	1	1	1	0	0	1	1	0	0	1
g4	1	1	1	0	1	1	0	0	1	1	0
g5	1	1	1	0	0	1	0	1	1	0	1
g1	0	0	1	1	0	0	0	1	0	0	1
g2	1	1	0	0	1	1	0	1	0	1	1
g3	0	0	0	0	1	0	0	1	1	1	0
g4	1	0	1	1	1	0	1	0	0	1	0

	f1	f2	f3	f4	f5	f6	fn
g1	1	0	0	1	1	1	0	0	1	0	1
g2	0	1	1	0	1	1	0	0	0	1	0
g3	0	1	1	1	0	0	1	1	0	0	1
g4	1	1	1	0	1	1	0	0	1	1	0
g5	1	1	1	0	0	1	0	1	1	0	1
g1	0	0	1	1	0	0	0	1	0	0	1
g2	1	1	0	0	1	1	0	1	0	1	1
g3	0	0	0	0	1	0	0	1	1	1	0
g4	1	0	1	1	1	0	1	0	0	1	0

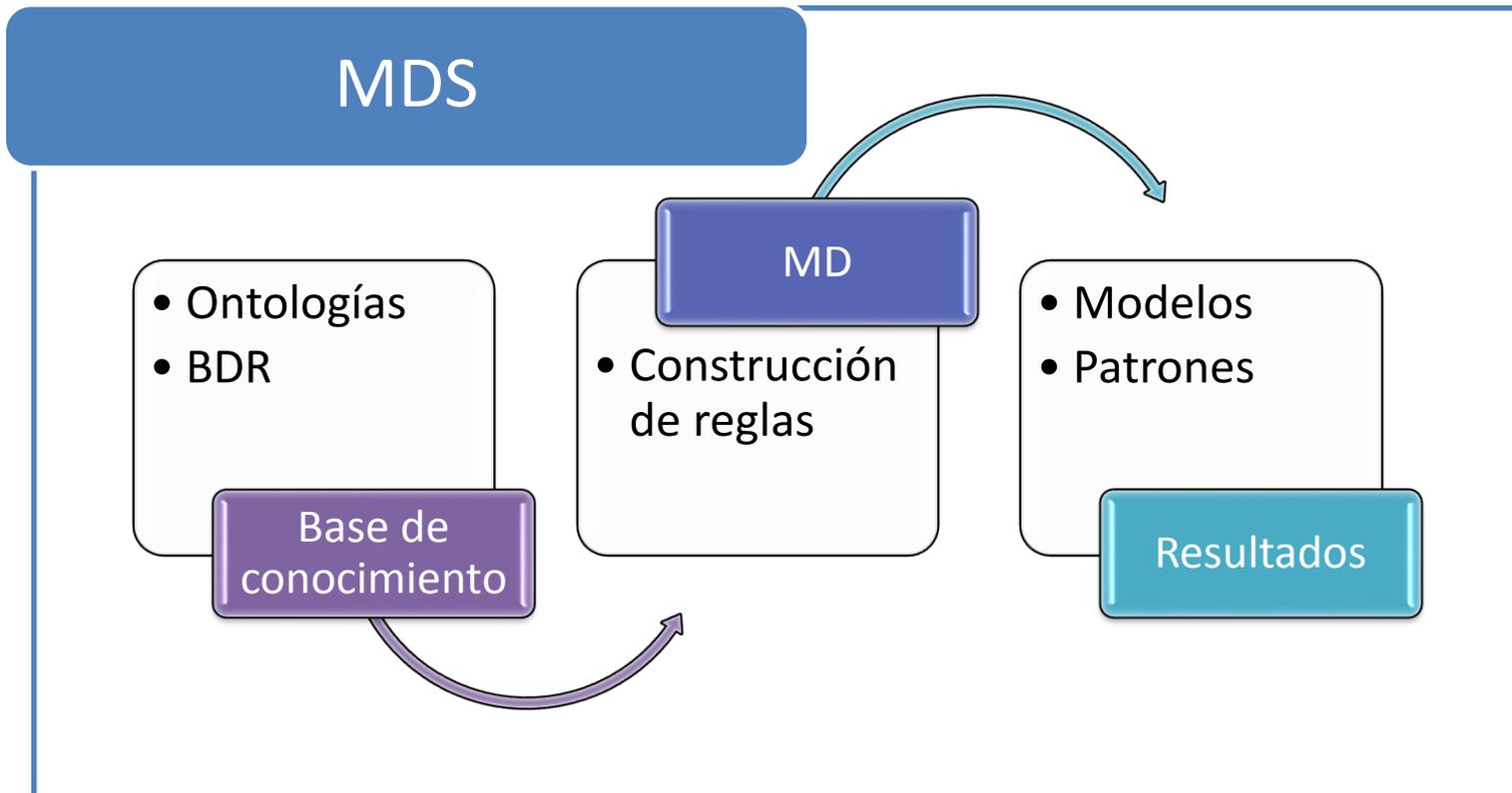
Paso 2

Descubrir reglas

```

If sepal width < 2.55 and petal length < 4.95 and
  petal width < 1.55 then Iris versicolor
If petal length ≥ 2.45 and petal length < 4.95 and
  petal width < 1.55 then Iris versicolor
If sepal length ≥ 6.55 and petal length < 5.05 then Iris versicolor
If sepal width < 2.75 and petal width < 1.65 and
  sepal length < 6.05 then Iris versicolor
If sepal length ≥ 5.85 and sepal length < 5.95 and
  petal length < 4.85 then Iris versicolor
If petal length ≥ 5.15 then Iris virginica
If petal width ≥ 1.85 then Iris virginica
If petal width ≥ 1.75 and sepal width < 3.05 then Iris virginica
If petal length ≥ 4.95 and petal width < 1.55 then Iris virginica
    
```

Recapitulando



SDM no solo es para construir reglas, otras formas de conocimiento se pueden generar (grupos, etc.)

Minería de la Web Semántica (MWS)

- Es la integración de dos áreas de conocimiento,
 - **Web Semántica (Semantic Web)**
 - **Minería en la Web (Web Mining)**

La **Web Semántica** es usada para darle significado a los datos que se encuentran en la Web.

La **Minería en la Web** se usa para extraer patrones de comportamiento en la Web.

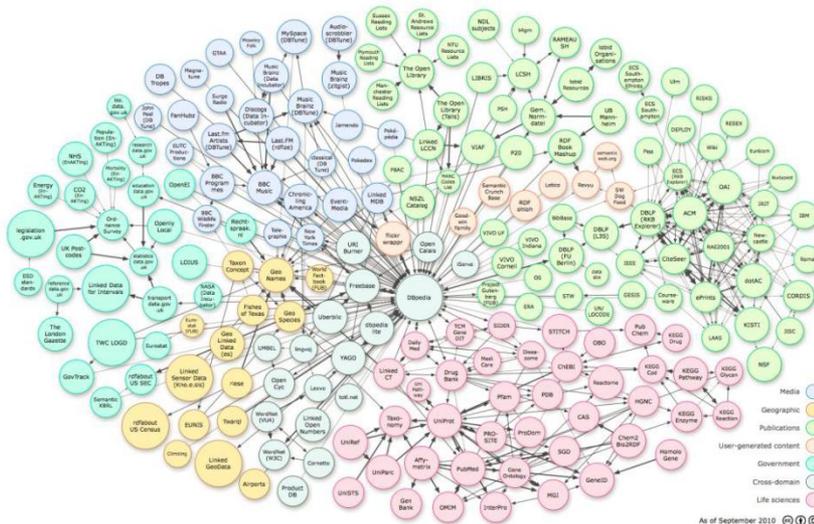
Minería de la Web Semántica

Cambio de paradigma de la minería de datos a la minería de conocimiento

- Minería de la Web Semántica: Minería del conocimiento codificado en ontologías de dominio,

Dos tipos de recursos semánticos

- Ontologías de Dominio
- Ontologías del flujo de trabajo de la minería de datos

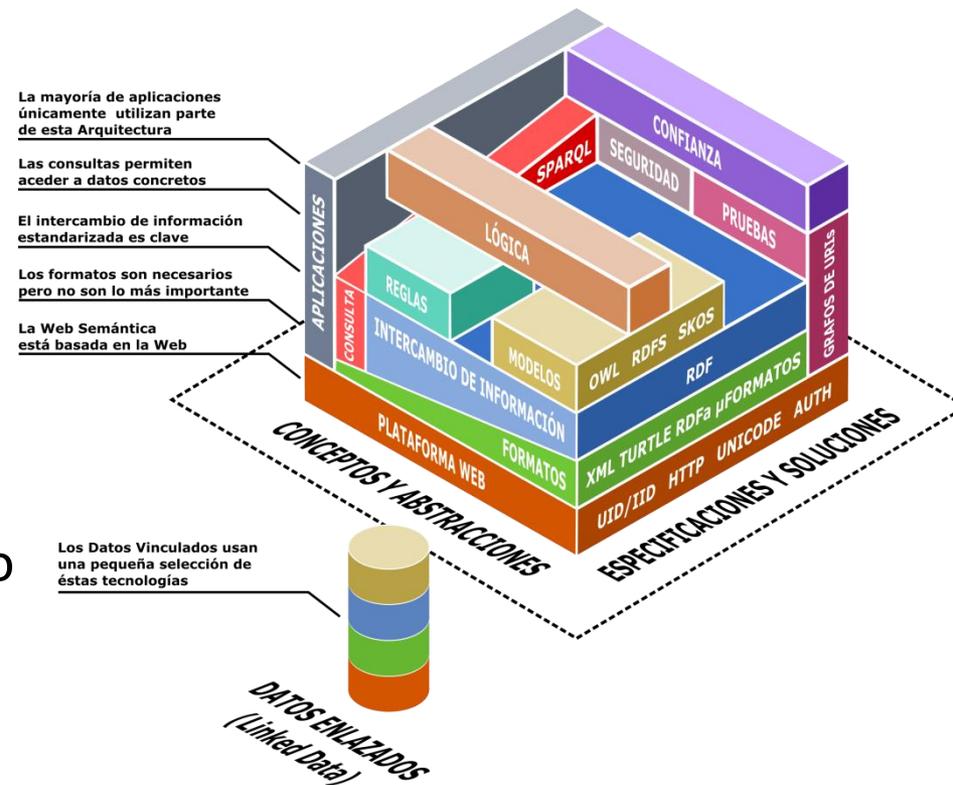


Minería de la Web Semántica

La diferencia de MWS con MDS es el propósito y lo que se está minando.

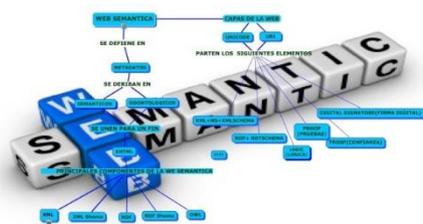
MWS mina datos de la Web, y los resultados son usados en la Web.

- La web semántica es expresada en formatos como OWL, RDF, XML,
- Son los recursos que van a ser minados para extraer conocimiento de la web semántica



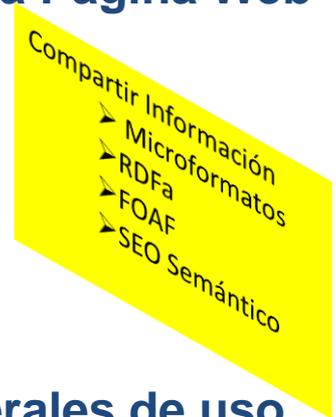
Existen varios tipos de **Minería Web** que se pueden aplicar en la MWS:

- El contenido de la web,
- La estructura de la web
- El uso que se hace de la web.



Resultados de la Búsqueda
Contenido de la Página Web

Enlaces



Patrones generales de uso
Patrones personales de acceso

Web mining

Minería del contenido de la Web

Es el descubrimiento de información útil desde los contenidos textuales y gráficos de los documentos Web, y tiene sus orígenes en el procesamiento del lenguaje natural y en la recuperación de la información.

Minería de la estructura de la Web

Es el proceso de descubrir el modelo subyacente a la estructura de enlaces de la Web y analiza, fundamentalmente, la topología de los hipervínculos (con o sin descripción de los enlaces)

Minería del uso de la Web

Es la aplicación de técnicas de minería de datos para descubrir patrones de acceso (o hábitos) desde los sitios Web. El principal objetivo es entender y servir mejor las necesidades de las aplicaciones basadas en Web.

Minería de Web Semántica

El minado de contenido, es una forma de *Text Mining*, que se aplica al contenido en la Web.

Por ejemplo, identificar en una página términos similares.

El minado de la estructura estudia el esqueleto que forman los enlaces entre las páginas de la Web, se mina un conjunto de enlaces.

El minado del uso de la web, se enfoca en minar un historial de uso de usuarios

Por ejemplo, consultas que hacen en una página, movimientos que los usuarios hacen entre páginas, etc.

Minería de la Web Semántica

Es usual que en el minado del uso, se usen las otras dos técnicas (de estructura y de contenido), haciendo al minado de uso la más completa forma de **Minería Web** y la más usada

Minería Web

Microformatos

XFN (XHTML Friends Network)

XFN quick reference

relationship category	XFN values
friendship (at most one):	friend acquaintance contact
physical:	met
professional:	co-worker colleague
geographical (at most one):	co-resident neighbor
family (at most one):	child parent sibling spouse kin
romantic:	muse crush date sweetheart
identity:	me

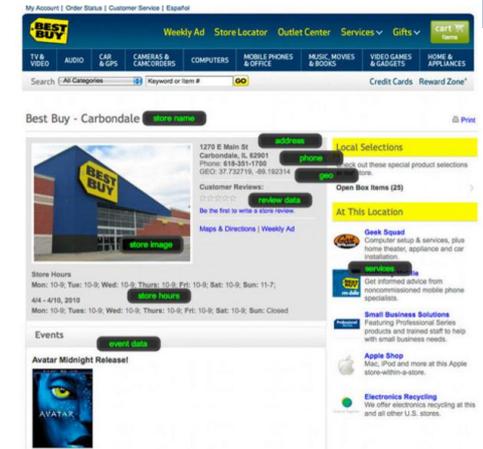
hCard
hcalendar

FOAF

```
<?xml version="1.0" standalone="yes"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1">
  <foaf:Person>
    <foaf:name>
      Taniana Josefina Rodríguez de Paredes
    </foaf:name>
    <foaf:mbox rdf:resource="mailto:taniana@ula.ve/">
    <foaf:knows>
      <foaf:Person>
        <foaf:name> Jose Aguilar </foaf:name>
        <foaf:mbox rdf:resource="mailto:aguilar@ula.ve/">
      </foaf:Person>
    </foaf:Knows>
  </foaf:Person>
</rdf:RDF>
```

Social Data Analytics
Social Network Analytics
Linked Data

SEO Semántico



Best Buy employees entered information into the blogs every day, using online forms that output RDFa. Myers told us that the use of RDFa makes "human input from our store employees more visible on the Web."

Best Buy is using Good Relations, a Semantic Web vocabulary for e-commerce that describes product, price, and company data.

RDFa

Knowledge Graph



Minería Ontológica (MO)

Actualmente, con el gran crecimiento en las cantidades de ontologías disponibles sobre un dominio de conocimiento dado, ha llevado a la MO a explorar técnicas que puedan extraer conocimiento adicional de un conjunto de ontologías, para lograr un dominio de conocimiento más amplio.

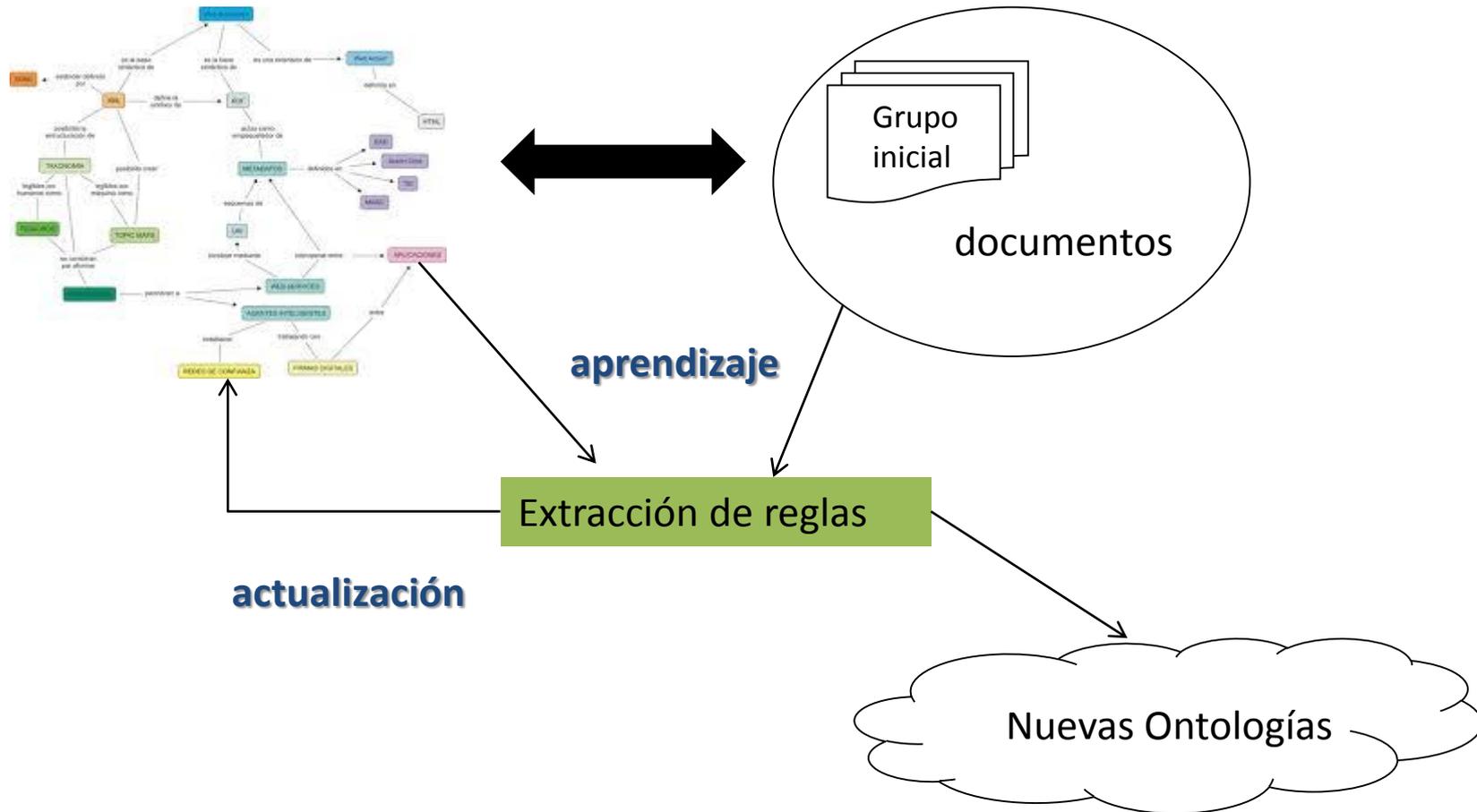
1. La extracción de: patrones de comportamiento, entre otras características,
2. Con la finalidad de construir o enriquecer ontologías.

- **Extracción de Reglas:** extrae reglas de un conjunto de ontologías.
- **Integración de Ontologías:** busca el vocabulario compartido entre varias ontologías.
- **Enlazado de Ontologías:** encuentra relaciones entre entidades de distintas ontologías.
- **Mezcla de Ontologías:** mezcla la información de varias ontologías con el fin de estandarizar conocimiento.
- **Alineación de Ontologías:** Identifica conceptos semejantes entre ontologías.

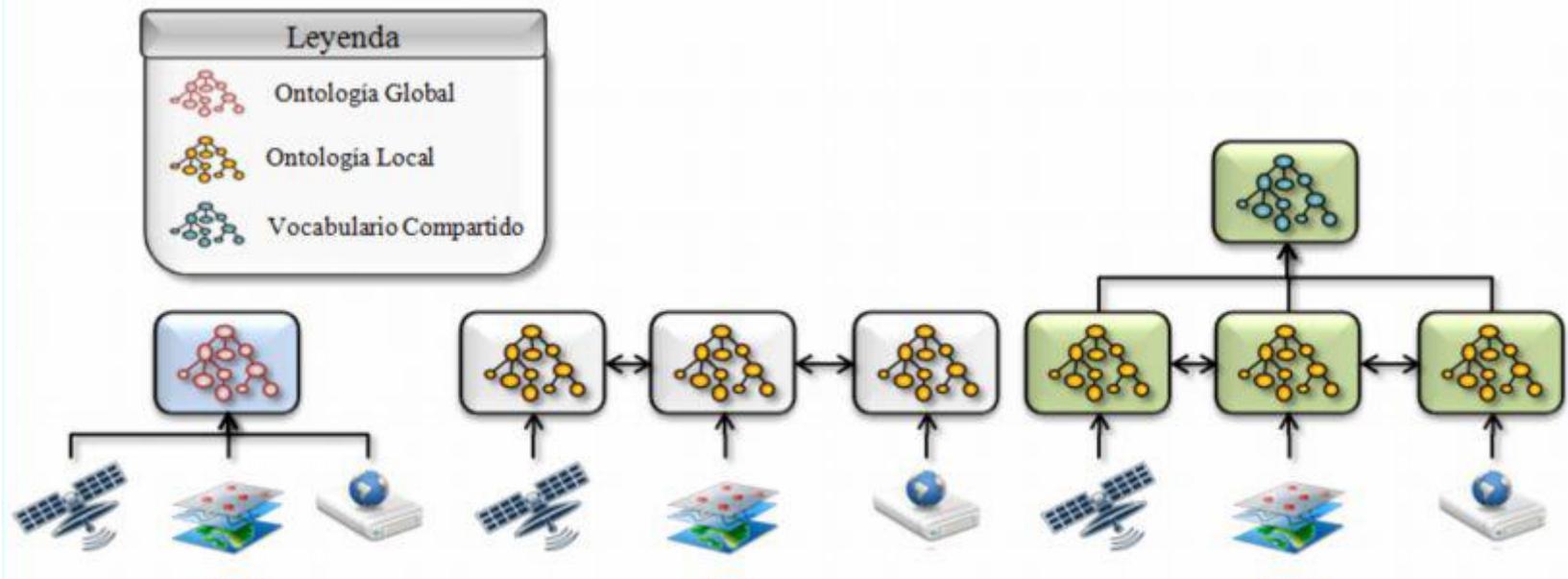
Ontologías Emergentes

Extracción de Reglas

Ontología



Integración de ontologías



Mecanismos de MO

Enlazado de ontologías: es el proceso de encontrar relaciones entre entidades que pertenecen a diferentes ontologías. Los resultados pueden ser usados para visualizar correspondencias, transformar una fuente en otra, crear un conjunto de relaciones o reglas entre las ontologías, generar consultas para extraer información desde las dos ontologías, etc.

Enlazado débil de Ontologías: es una correspondencia entre conceptos idénticos. En este caso, básicamente lo que se realiza es la intersección de las ontologías, a partir de la cual se podrían hacer inferencias específicas en cada ontología.

Enlazado Fuerte de Ontologías: Es realizado de manera semiautomático, con la ayuda de un experto del conocimiento global que se está enlazando, el cual puede definir nuevos conceptos, así como enlaces que relacionan conceptos de ontologías distintas, creando así una Meta-Ontología con partes de conocimiento de las ontologías enlazadas.

Todos los estudios se diferencian en cómo buscar el concepto mas similar

- Dadas dos ontologías A y B
- Para cada par de conceptos, uno de A y uno de B
 - Buscar similitud léxica
 - Buscar similitud semántica
- Calcular una matriz de similitud para todos los pares
- Escoger conceptos similares
- Si existen conceptos similares
 - copiar los conceptos similares
 - Copiar los hijos y sus relaciones de los conceptos similares
 - Detectar inconsistencias
 - Resolver inconsistencias
- Si no, no se puede realizar el enlazado

Enlazado débil

Enlazado débil

Todos los estudios se diferencian en cómo buscar el concepto más similar

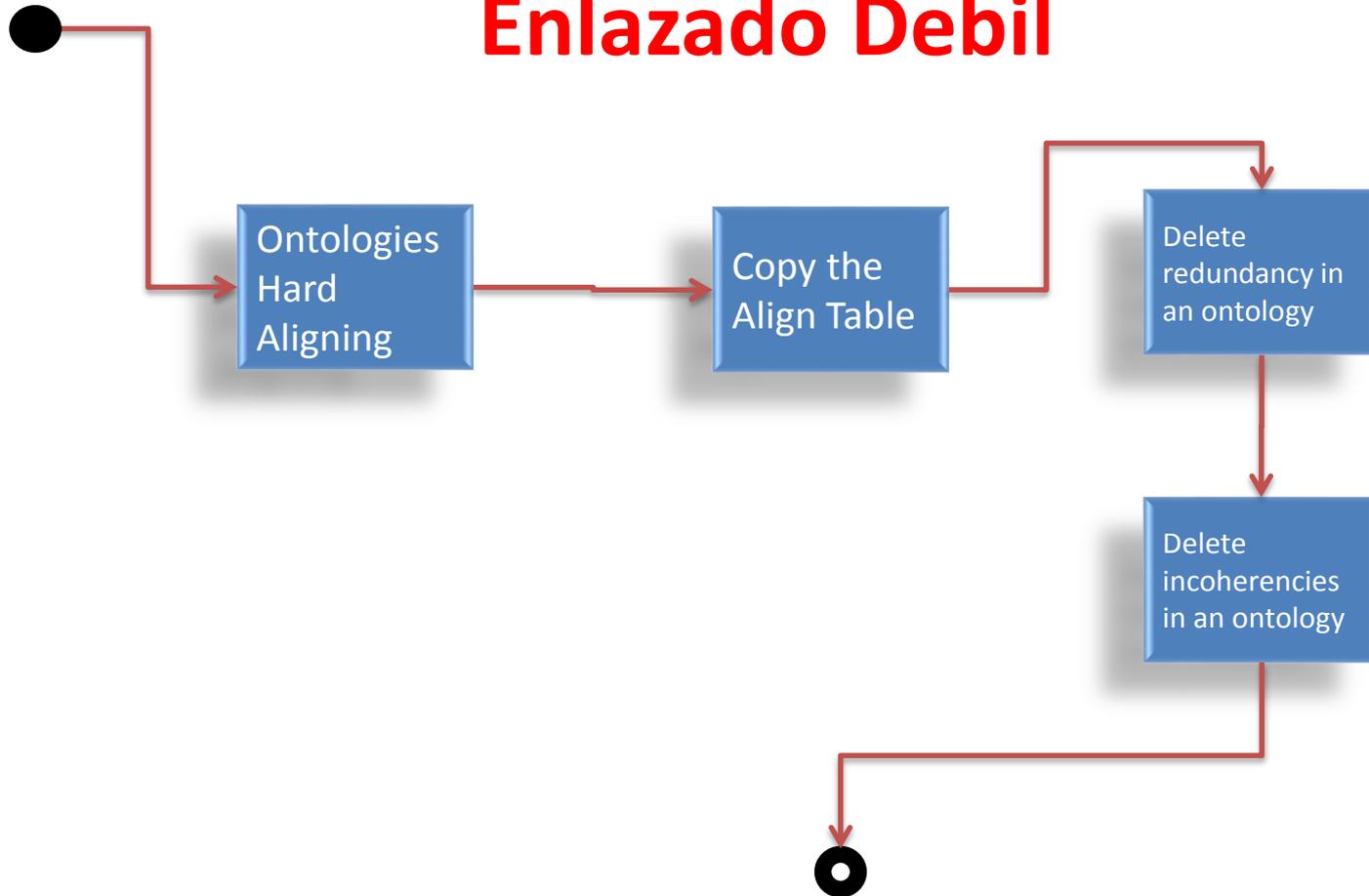
Enlazado débil

Dadas dos ontologías A y B
Para cada par de conceptos, uno de A y uno de B:
 Buscar similitud léxica
 Buscar similitud semántica
Calcular una matriz de similitud para todos los pares de conceptos
Escoger conceptos similares
Si existen conceptos similares
 copiar los conceptos similares
 Copiar los hijos y sus relaciones de los conceptos similares
 Detectar inconsistencias
 Resolver inconsistencias
Si no, no se puede realizar el enlazado

Se crea una matriz de pares de conceptos con su similitud

Enlazado débil

Enlazado Debil



Enlazado Fuerte

Dadas una tabla de conceptos idénticos T por un experto

Generar una nueva ontología C

Determinar qué conceptos de T se copiaran en C

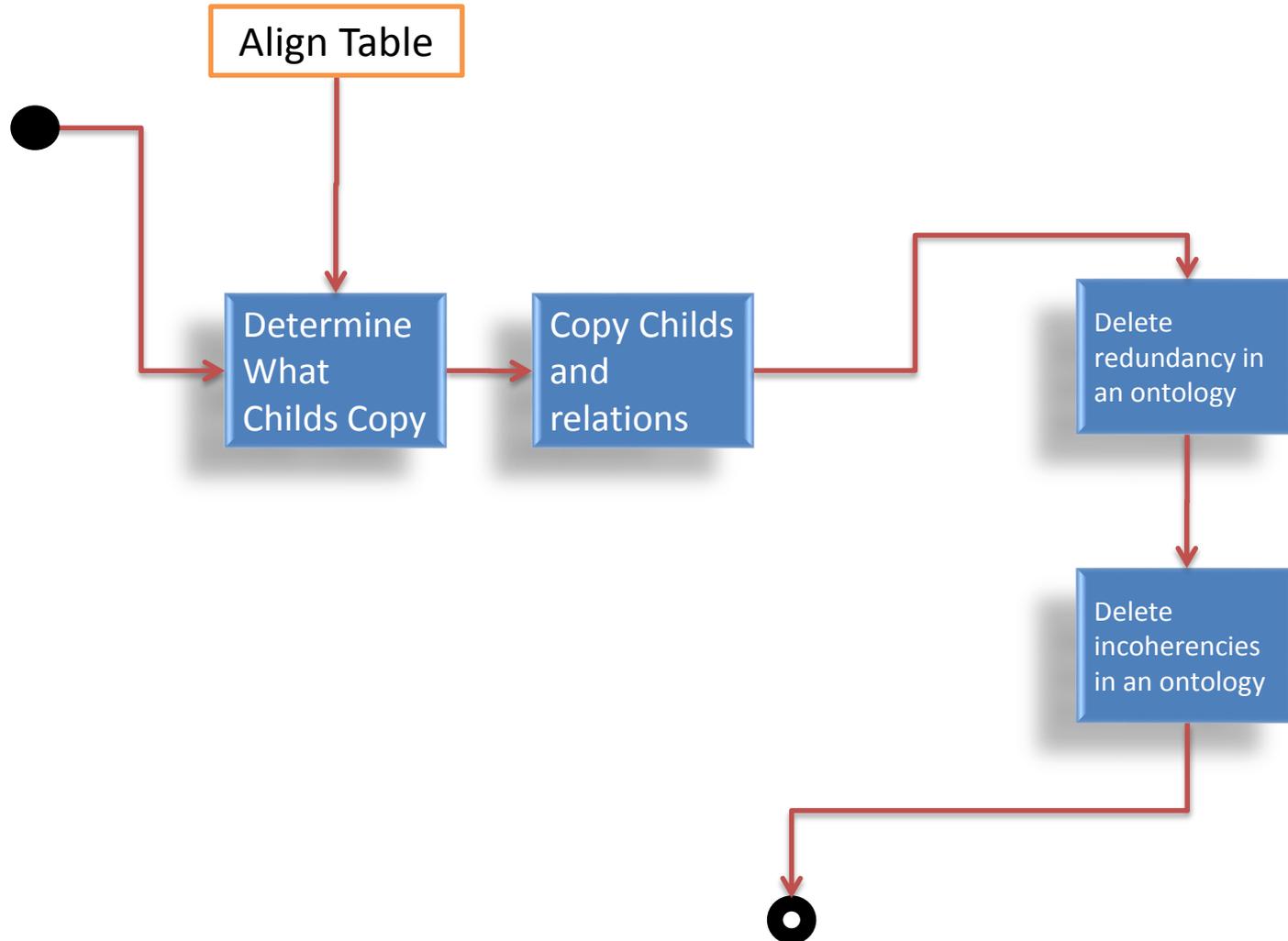
Copiar los conceptos escogidos en C

Para cada concepto escogido (Buscar en orden todos sus hijos)

Determinar qué concepto hijo se copiaran

Copiar el hijo en C

Enlazado Fuerte



Mecanismos de MO

Mezclado

Mezclado de ontologías: es el proceso donde varias ontologías dentro de un mismo dominio se unen para estandarizar el conocimiento, hacer crecer el conocimiento y tener el conocimiento total de manera local. Los mezcladores unen ontologías que manejan el mismo conocimiento, pero con diferente representaciones, o que poseen representaciones parciales de dicho conocimiento, tal que las ontologías pueden coincidir en ciertos conceptos y en otros no.

Mezcla Débil de Ontologías: se toma una ontología A, la copian como resultado C, y la van enriqueciendo con la otra B, comparando todos los conceptos de la ontología C (que son los mismos de A en este momento) con los de la ontología B, enriqueciendo los conceptos de C con sus conceptos semejantes de B. Dejando por fuera parte del conocimiento de B.

Mezcla Fuerte de Ontologías: Es una mezcla débil, pero incorporándole el conocimiento dejado por fuera de B, ya sea, por ejemplo, por una técnica de enlazado.

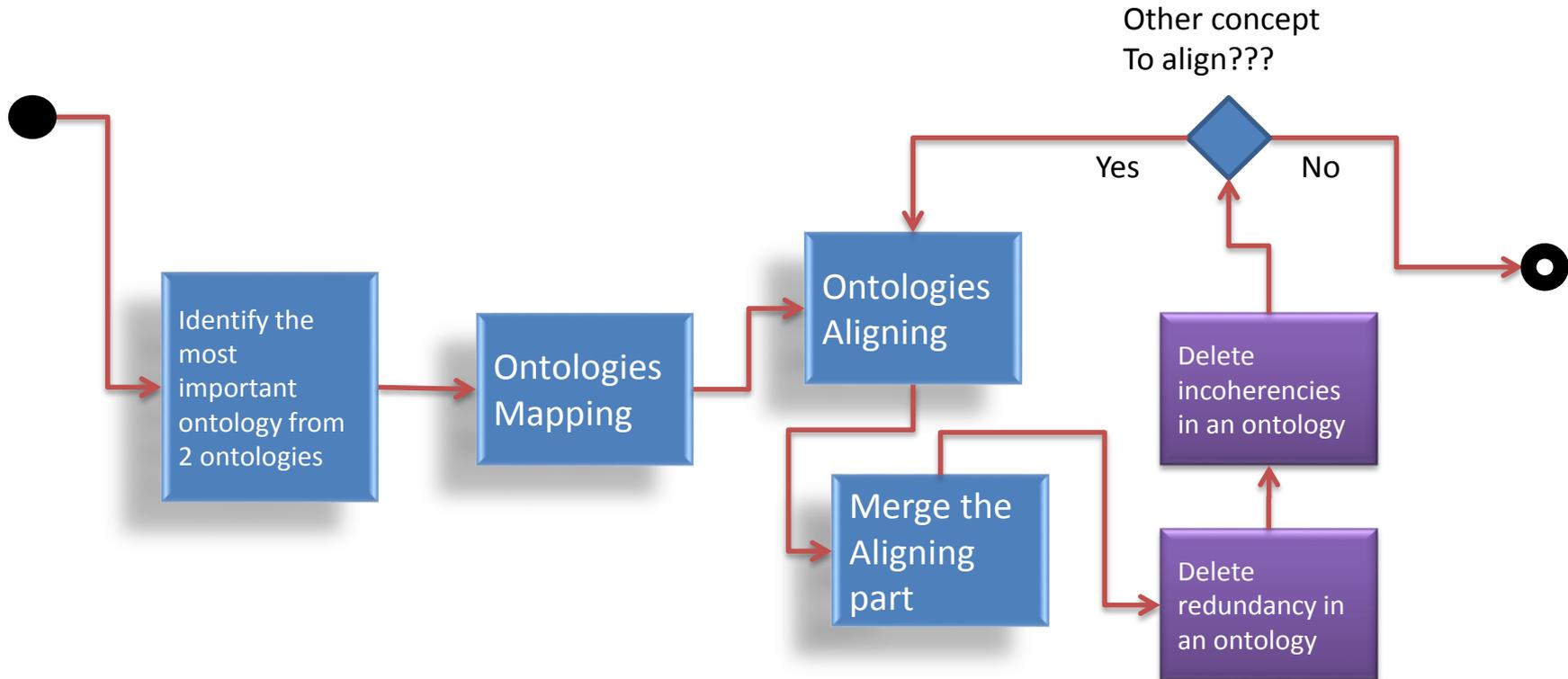
Macro-algoritmo para la Mezcla Débil de Ontologías

- **Copia** ontología de A a C.
- A partir del concepto raíz en C, se sigue un **recorrido en profundidad**:
 - **Buscar** en B cada concepto C_c de C (utilizando algún método (como COS) para **calcular el concepto más cercano** en otra ontología (CMS).
 - Si hay un **CMS** en B, se seleccionan las relaciones del CMS en B que se pueden añadir a C, así como los nuevos conceptos, de la siguiente manera :
 - Se **enriquecen las relaciones** de C_c que son sinónimas con las del CMS;
 - Se **añaden las relaciones** nuevas del CMS a C_c .
 - Se **copian los conceptos** que no tenga C que se encuentren en las nuevas relaciones.
 - Se **detectan inconsistencias** entre las relaciones de C_c y CMS
 - Se **mantienen sólo las relaciones que sean consistentes** en C_c
 - Si no se logra resolver las inconsistencias, **permanecen en C_c las relaciones originales** de A.
 - Se **copian los conceptos hijos** de CMS a C.
 - Si no hay un CMS, se toma el siguiente concepto C_c en profundidad

se diferencian en cómo buscar el concepto mas similar

Se dejan conceptos de B por fuera

Mezcla Débil



Mezcla Fuerte 1/2

Dadas dos ontologías A y B, se mezclaran en C

Identificar la ontología con mayor conocimiento entre A y B

Copio La ontología con mayor conocimiento en C

Para cada concepto c en la ontología con menos conocimiento

Buscar en C el concepto más similar (cms, escoger método)

Si c posee un concepto similar (cms) en C

Se enriquecen las relaciones del cms en C con las relaciones que sean sinónimas a las su concepto más similar c

Se crean nuevas relaciones en cms que no tenga de su concepto más similar c

Se copian los conceptos hijos de c en el cms de C

Detectar inconsistencias

Resolver inconsistencias

Mezcla Fuerte 2/2

Si no se encuentra un cms de c en C

- Identificar donde se puede enlazar

- Crear el enlace

- Crear el concepto c en C

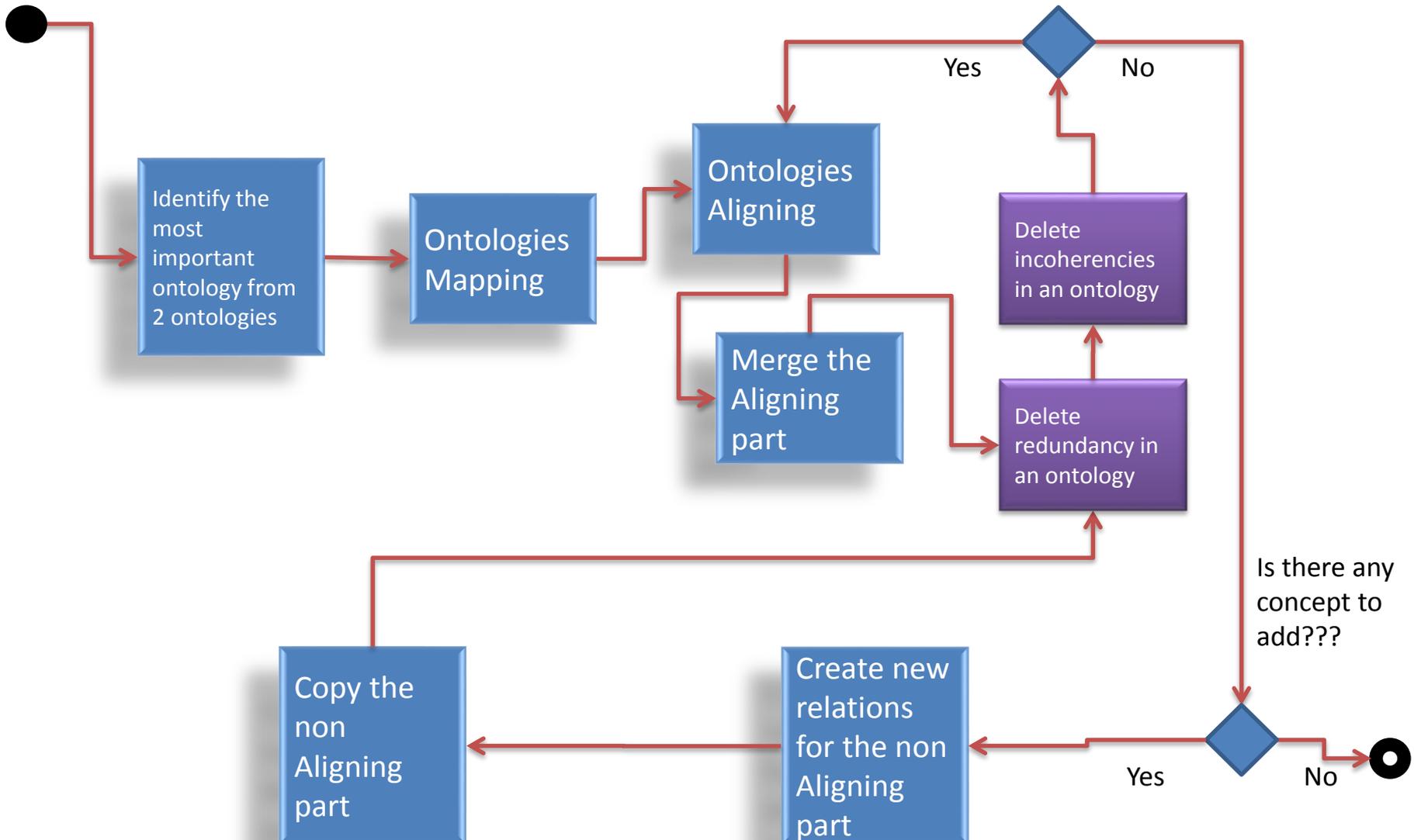
- Copiar todos los hijos de c

- Detectar inconsistencias

- Resolver inconsistencias

Pasar a buscar el siguiente concepto c en la ontología con menos conocimiento

Mezcla Fuerte



Alineación de ontologías

identificar conceptos de una ontología que sean semejantes en
las otras
ontologías

Distancia semántica entre cada par de conceptos en
ontologías distintas

**Existen varios métodos y herramientas para realizar la
alineación de ontologías**

Alineación de ontologías

Técnicas de alineación de ontologías

- Basado en similitud lingüística (*linguistic matching*)
- Basado en similitud de grafos (*graph matching*)

Alineación de ontologías

alineación

Esta compuesto por los siguientes elementos:

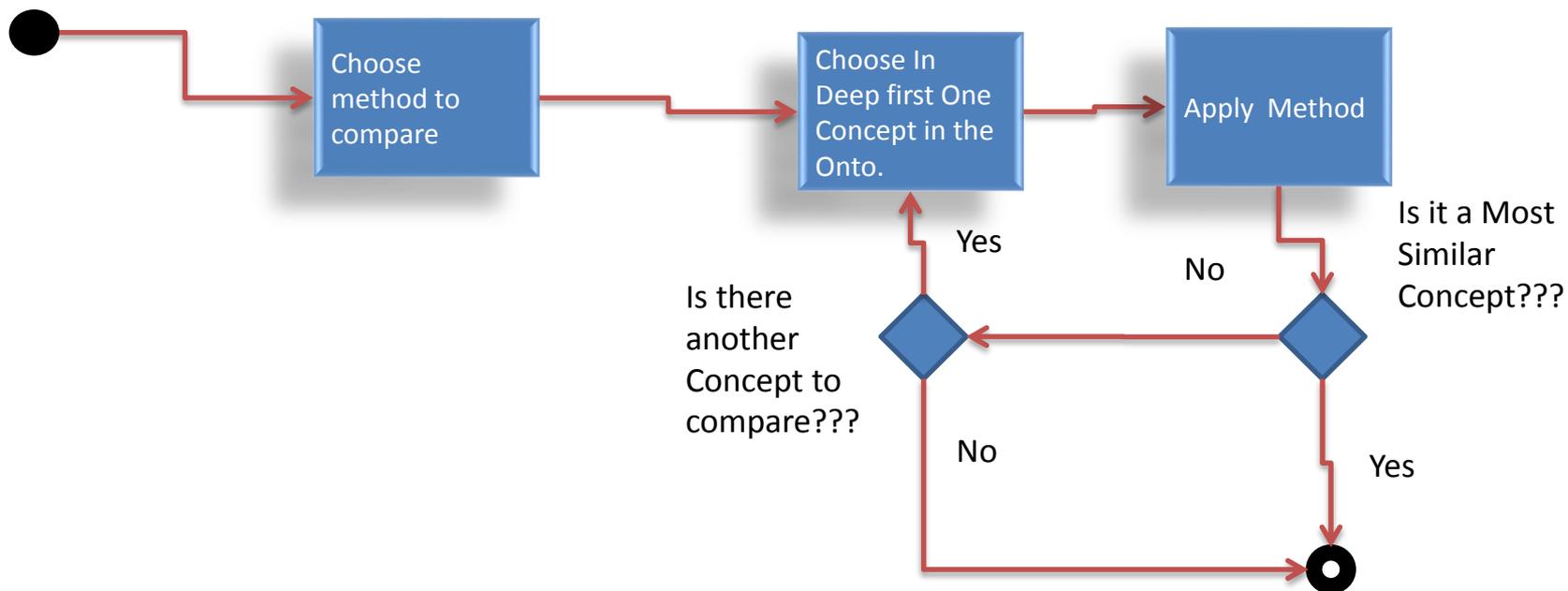
- dos ontologías $O1$ y $O2$,
- un conjunto p de parámetros,
- un conjunto r de recursos para la alineación, y
- una función f de alineación, que retorna un conjunto de correspondencias A'

La función f integra diversos recursos para encontrar correspondencias entre dos conceptos.

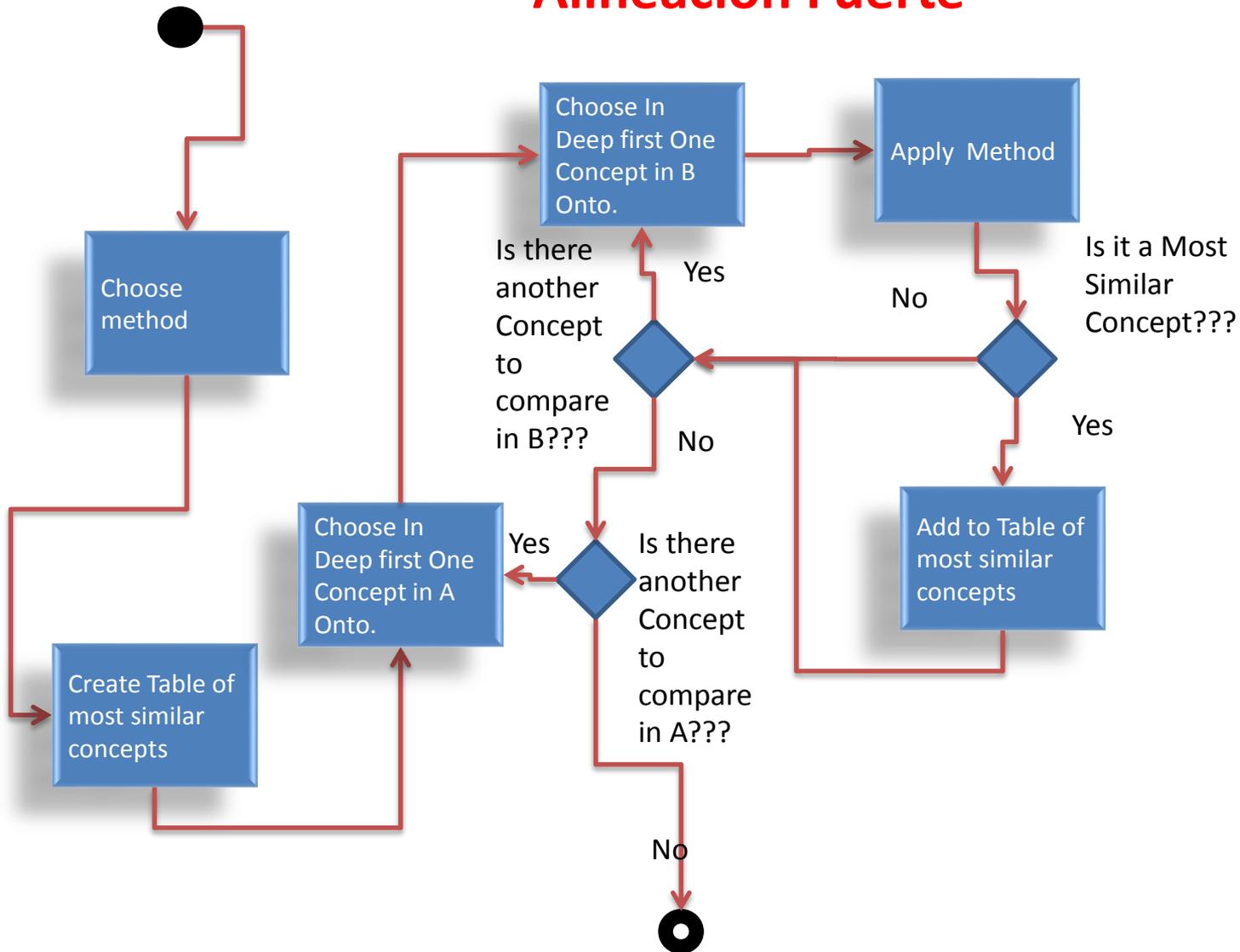
En cada $O1$ y $O2$ se analizan parte de sus elementos como: conceptos, propiedades de conceptos y jerarquía de conceptos.

- El conjunto p representa los requisitos para realizar la alineación; $p = \{\text{lenguaje de diseño OWL, número de elementos, vocabulario del idioma, no inferencias}\}$.
- El conjunto de recursos se refiere a los elementos empleados para obtener el conjunto de correspondencias $r = \{\text{conjunto medidas de similitud, algoritmo AdaBoost, algoritmo de clasificación K-Vecinos}\}$.
- El conjunto A' simboliza todas las correspondencias semánticas.

Alineación Débil



Alineación Fuerte



Métodos para calcular conceptos cercano

Suponen que CA sea un concepto o nodo en la ontología A y PA su predecesor. COM busca encontrar el concepto más parecido CB a CA en la ontología B, y PB (predecesor del concepto CB) a PA que aún no se ha encontrado.

Cuatro casos para calcular la similitud:

Caso A: El concepto CA coincide con CB en B y los predecesores PA y PB

Dados CA y PA, se busca en B por dos conceptos, CB y PB, de manera que la definición de PB coincide con la mayoría de las palabras que definen PA, y la mayoría de las palabras que definen CB coinciden con la definición de CA. En ese caso retorna CB (conocido como CMS).

Métodos para calcular conceptos cercano

Caso B: PA coincide con PB, pero no hay coincidencia entre CA y CB.

PB se encuentra, pero no CB.

- En este caso, de forma recursiva con PA como parámetro, se pasa a confirmar que PB es un predecesor de CA.
- Si un primo de PB (PB') pasa a ser la raíz de la ontología (OBRoot), entonces el algoritmo termina sin éxito. Si eso no sucede, entonces CA se busca en B a través de cada hijo de PB (el hijo debe coincidir con la mayoría de sus propiedades de CA).
- Si el candidato CB tiene hijos, se verifica que coincidan con los hijos de CA. Si un CB' se encuentra con las propiedades esperadas de CA, el algoritmo termina devolviendo con éxito CB'. De lo contrario, COM intenta encontrar CB' entre los hermanos de PB.
- Si eso no sucede, se busca entre los nietos de PB. Si CB' no se encuentra, entonces el valor más cercano a la CA es un hijo (desconocido, no está presente) de PB, por lo tanto, COM devuelve "hijo de PB" (lo que significa que un hijo PB que no existe todavía en B es el concepto más similar a CA).

Métodos para calcular conceptos cercano

Caso C: CA coincide con CB, pero no hay ninguna coincidencia entre el PA y PB.

- Si CB se encuentra, pero no PB, entonces se comprueba si el abuelo de CB en B es similar a la PA, o si el bisabuelo de CB en B es similar a PA.
- Si este es el caso, entonces el concepto más similar de PA en B es el abuelo o bisabuelo de CB y se termina el algoritmo.
- Si no se encuentra, entonces se verifica si la mayoría de las relaciones y los valores de CA coinciden con los de CB y si la mayoría de los hijos CA coinciden con la mayoría de los hijos de CB.
- Si las propiedades y los hijos coinciden, entonces la respuesta es CB y el algoritmo termina, a pesar de que PB no se ha encontrado en B.
- Si sólo una parte de las propiedades y los hijos coincide entonces la respuesta es "Probablemente CB" y se termina el algoritmo.
- Si no hay propiedades ni los hijos son iguales, entonces la respuesta es "no existe" y el algoritmo concluye.

Métodos para calcular conceptos cercano

Caso D: CA no coincide con el CB y PA no coincide con PB.

Si CB no existe y tampoco PB, entonces la respuesta de la COM es "no existe " y termina el algoritmo.

Validación de ontologías

Se recibe como entrada dos ontologías y valida si se pueden utilizar en el proceso de alineación,

También adquiere información sobre las clases y sus respectivas jerarquías.

Las actividades a realizar son las siguientes:

1. Verificar el lenguaje de diseño de las dos ontologías de entrada.
2. Obtener la jerarquía de clases de ambas ontologías.
3. Extraer información de cada clase.
4. Verificar si las propiedades de las clases son adecuadas para realizar el proceso de alineación.

DetECCIÓN DE CORRESPONDENCIAS SEMÁNTICAS SEGÚN ZAGAL R.

Las medidas de similitud seleccionadas para el proceso de alineación se dividen en dos grupos:

Similitud en base a términos: Se enfoca en el nombre de las entidades en las ontologías, principalmente en el nombre de las clases.

Similitud semántica: Su alcance va más allá de los nombres de las entidades, se enfoca en los componentes que definen la semántica de una clase:

- **Similitud entre propiedades de clases:** Considera las coincidencias existentes entre las propiedades de dos clases.
- **Similitud entre superclases:** Se refiere al par de superclases con mayor similitud respecto a dos clases comparadas.

Similitud léxica

- La Distancia de Levenshtein o distancia de edición (edit distance), fue creada en 1965 por el científico ruso Vladimir Levenshtein.
- La idea consiste en determinar el número mínimo de operaciones requeridas para transformar una cadena de caracteres en otra,
- Estas operaciones son: inserción, eliminación o sustitución de un carácter.

Por ejemplo, la distancia de Levenshtein entre los términos "hotel" y "hostal" es de dos, porque se necesitan al menos dos operaciones elementales para cambiar un término en el otro término.

Similitud entre propiedades

- Es la verificación de si cada propiedad p_n del conjunto de propiedades P de una clase $C1$, coincide con otra propiedad p'_m del conjunto de propiedades de P' de otra clase $C2$.
- Para realizar dicha comparación entre cada propiedad, se utilizan las etiquetas de ambas propiedades (o nombres) como entrada a una medida de similitud léxica (por ejemplo, la distancia de Levenshtein)

Similitud entre Superclases

También conocida como similitud entre conceptos ascendentes, a partir de dos clases iniciales.

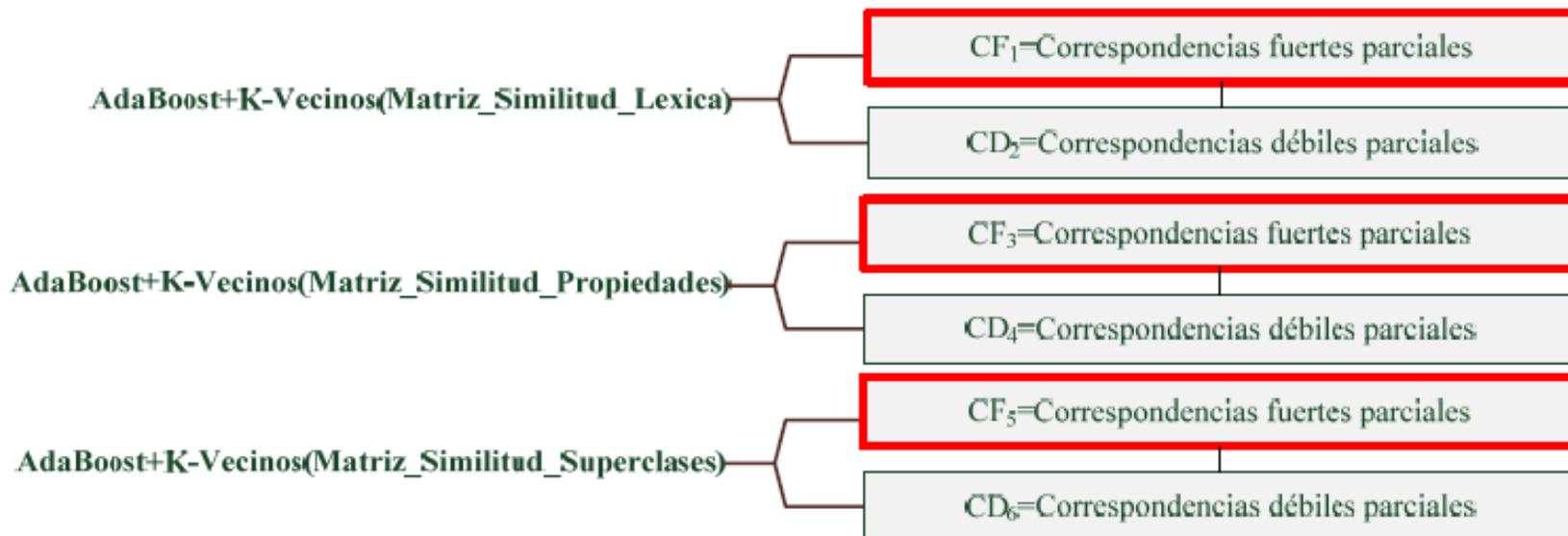
La similitud entre algún par de superclases puede influir para establecer una correspondencia entre dos clases.

Para calcular esta medida de similitud es necesario calcular previamente la similitud léxica y de propiedades de todas las parejas de clases.

El objetivo es comparar parejas de superclases y seleccionar la pareja de superclases con mayor similitud entre sí, a fin de encontrar de manera aproximada la pareja de superclases en donde coincidan o converjan las clases C1 y C2.

Correspondencias semánticas

- Para clasificar si hay o no una correspondencia fuerte, se usa el algoritmo general de AdaBoost mas el de KVecinos (AdaBoost+K-vecinos)
- Como entrada recibe valores de una matriz de similitud.
- La salida es la clasificación de estos valores en dos posibles clases: una clase que agrupa los valores de similitud más altos y otra clase que agrupa los valores de similitud más bajos.



Crterios para determinar la correspondencia semántica fuerte

Caso	“Correspondencia fuerte parcial” SIMILITUD LÉXICA	“Correspondencia fuerte parcial” SIMILITUD ENTRE PROPIEDADES	“Correspondencia fuerte parcial” SIMILITUD ENTRE SUPERCLASES	“Correspondencia semántica fuerte”
1	SI	SI	SI	SI
2	SI	SI	NO	SI
3	SI	NO	SI	NO
4	NO	SI	SI	SI
5	SI	NO	NO	NO
6	NO	SI	NO	NO
7	NO	NO	SI	NO
8	NO	NO	NO	NO

Conceptos que son inconsistentes

Una inconsistencia surge cuando dos afirmaciones diferentes toman valor en un concepto monovaluado

$V = a$ y por otro lado dice que $V = b$, donde a es diferente a b .

Hay una contradicción porque V no puede tomar mas de un valor (Como es el caso de las definiciones de conceptos).

Para los servicios de

- Detectar incoherencias
- Detectar redundancia

Conceptos que son redundantes

Una redundancia surge cuando la misma afirmación toma valor en varios conceptos.

$V = a$ y por otro lado dice que $C = a$. Hay una redundancia debido a que el valor (definición) de V es idéntico al de C .

Caso de dos conceptos con nombres distintos que se refieren a lo mismo.

Para los servicios de

- Detectar incoherencias
- Detectar redundancia

Datos enlazados o datos vinculados (Linked Data)

Método de publicación de datos estructurados para que se puedan interconectar

Se basa en tecnologías Web, tales como HTTP, FOAF, OWL, RDF y los URI, pero en vez de utilizarlos para páginas web, se extienden para compartir información de una manera que puede ser leída automáticamente por computadores.

Web de enlaces de información interconectadas

- [DBpedia](#) - conjunto de datos extraído de Wikipedia; contiene unos 3,4 millones de conceptos descritos por un millardo de tripletas (1000 millones), que incluyen resúmenes en once idiomas
- [Bibliografía DBLP](#) - información bibliográfica de artículos científicos, con información de 800.000 artículos, 400.000 autores y aproximadamente 15 millones de tripletas
- [riese](#) - datos estadísticos de 500 millones de europeos (el primer conjunto de datos enlazados en [XHTML+RDFa](#))

Por qué Linked Data?

- Muchas ontologías con **información similar en algunas de sus partes:**
 - Por ejemplo, Nombres, CI, Dirección, Número telefónico
- Esas partes comunes **podrían interconectarse, y juntar todos los datos desde múltiples ontologías en una gigante colección de datos, para ser consultada.**

Eso debería llevar a crear un **enjambre/araña de ontologías en el mundo**, y cada ontología sería un **nodo del gigante grafo.**

Por qué no una simple Ontología con los datos interconectados?

- Acceso a los Datos y tiempo de razonamiento seria enorme
- Las cargas de datos en tiempo real seria muy compleja y embotellar la red
- No existe actualmente computador que pudiera procesar esa cantidad de datos masivos

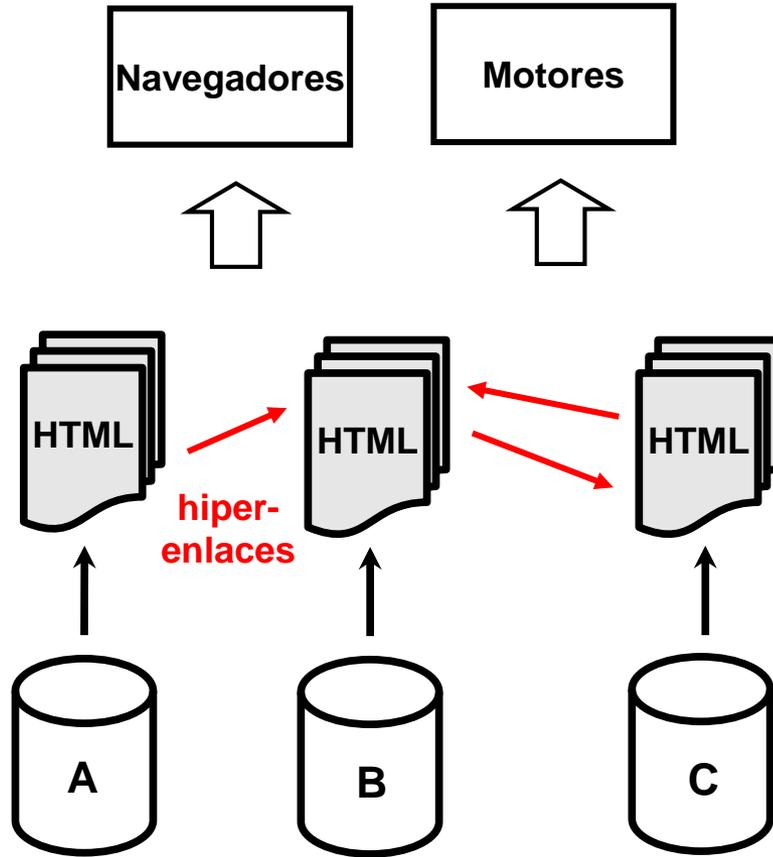
Alternativa...

- Las ontologías se “enlazarían” a través de las partes comunes (Nombre, Dirección, etc.)
- **Usuarios conocen las ontologías que ellos requieren consultar**
- Las consultas se hacen sobre las ontologías individuales
- **La parte común de una ontología se conecta con las partes similares de las otras**
- A partir de ese “enlazado”, se extrae localmente un subconjunto de esa global ontología compartida

Reglas para Datos enlazados (Linked Data)

- A partir de allí, se debe permitir:
 - Seguir esos enlaces
 - Combinar la información guardada en las ontologías
- Todos los datos (cosas) tienen un URI
- Ese URI es un válido URL
- Debe haber una pagina con ese URL, el cual contenga los datos representados por ese URI
- El URL nunca cambia
- Cuando alguien busca un URI, se provee información útil en RDF.
- Se incluyen instrucciones RDF que enlaza a otros URIs para descubrir cosas relacionadas.

Clásica Web



1. URLs como único identificador para recuperar información
2. HTML para compartir información
3. Híper-enlaces

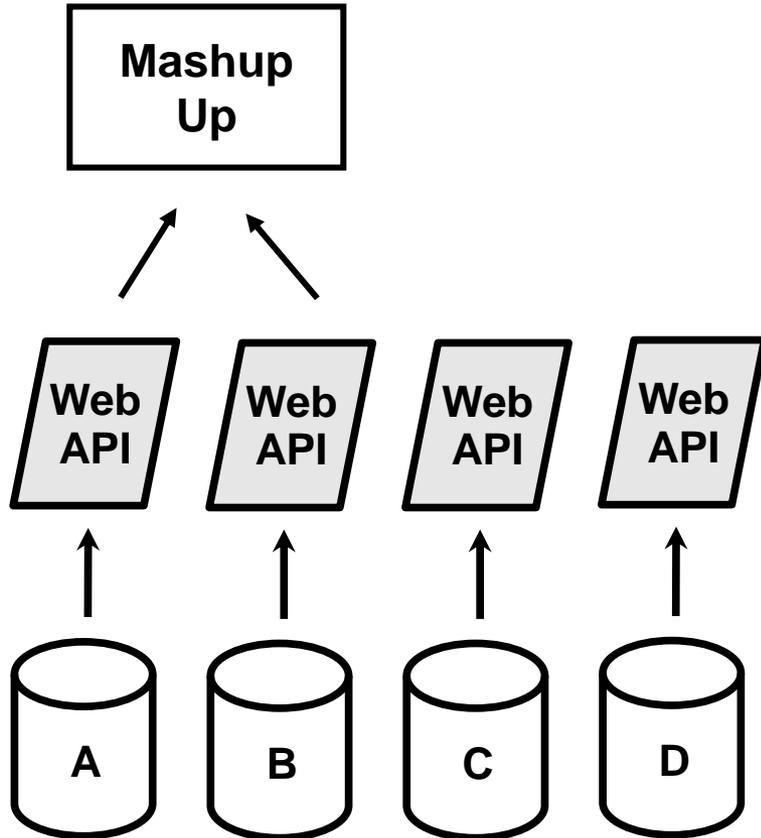
Problema

Difícil hacer cosas inteligentes con él.

Solución

Aumentar la estructura del contenido de la Web.

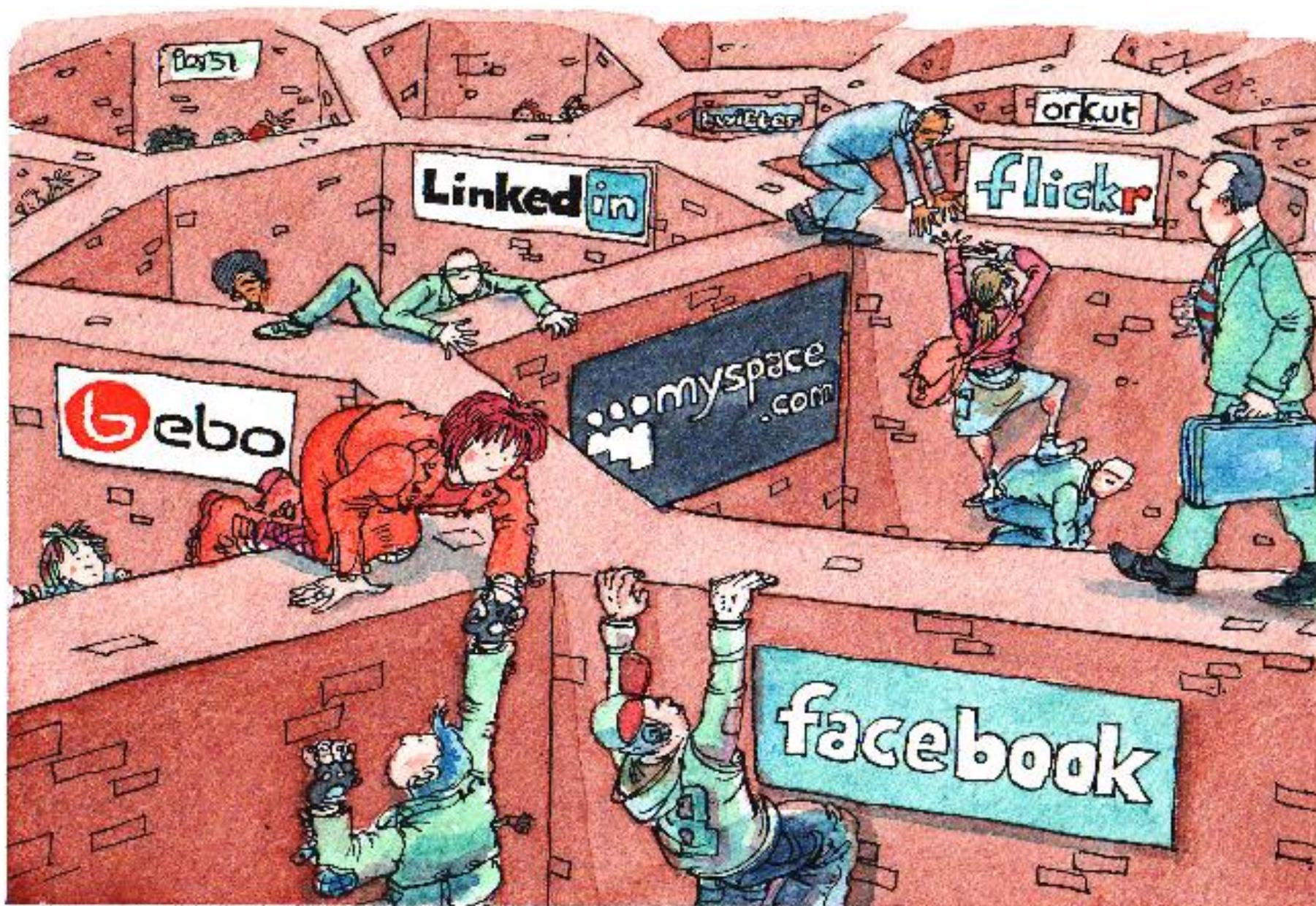
APIs y Mashups en la Web



1. APIs provee interfaz
2. Mashups se basan en conjunto fijos de fuentes de datos.
3. No se pueden establecer Híper-enlaces entre los objetos de datos.



APIs corta la Web en Valles



Qué es Mashup?

Es la combinación de datos relacionados para facilitar su uso

- Grafos
- Sitios Webs

Los datos están conectados, para describir conocimientos desde pequeños datos disjuntos ahora conectados.

Ejemplos de motores Mashups



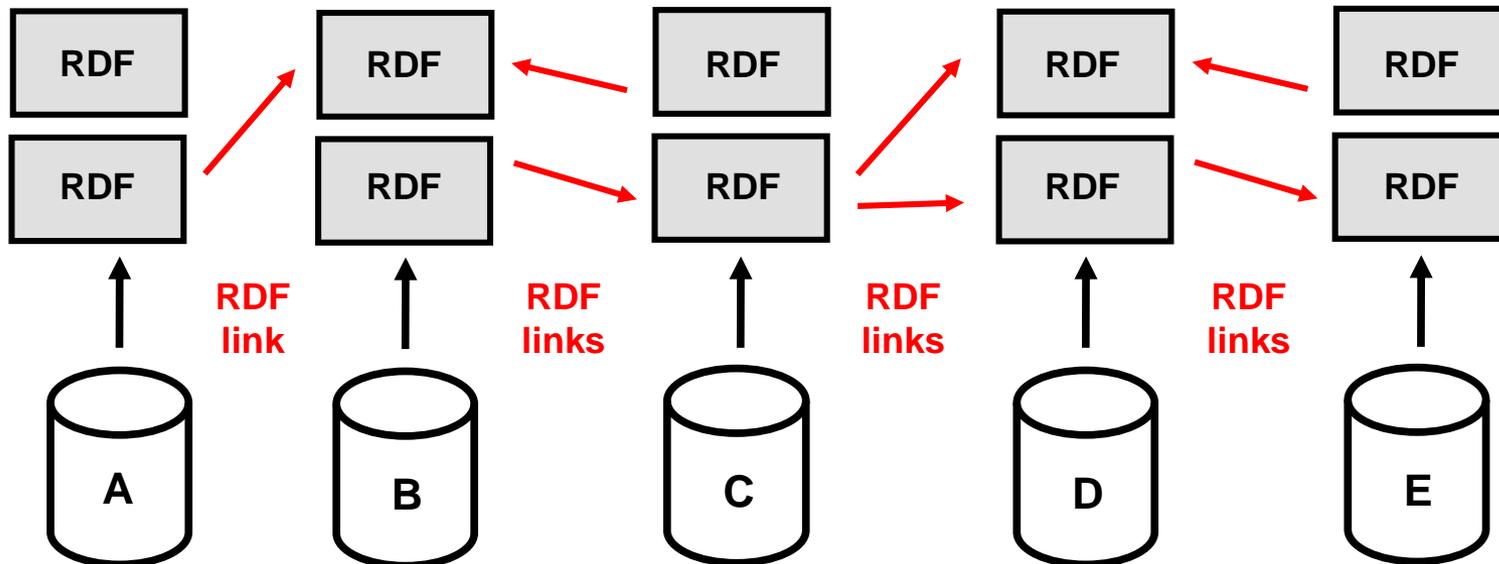
<http://www.jackbe.com/enterprise-mashup/>



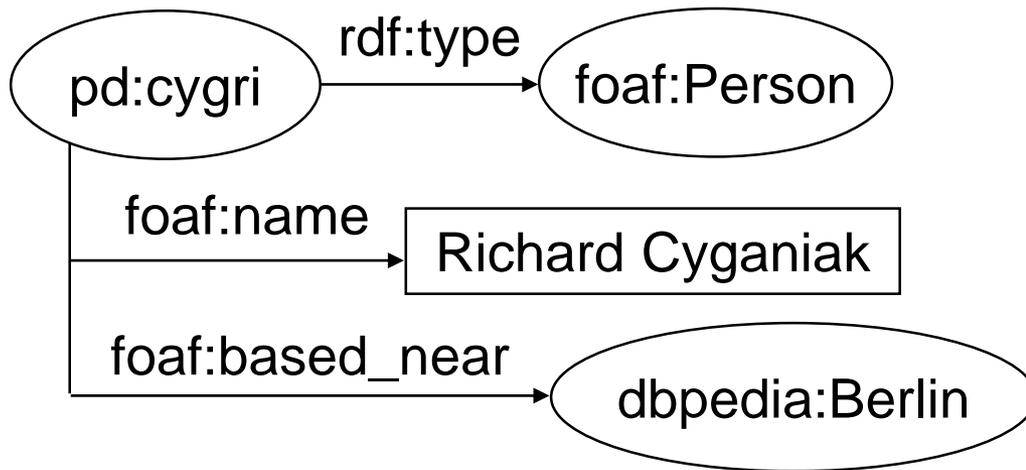
<http://www-01.ibm.com/software/lotus/products/mashups/>

Datos enlazados (Linked Data)

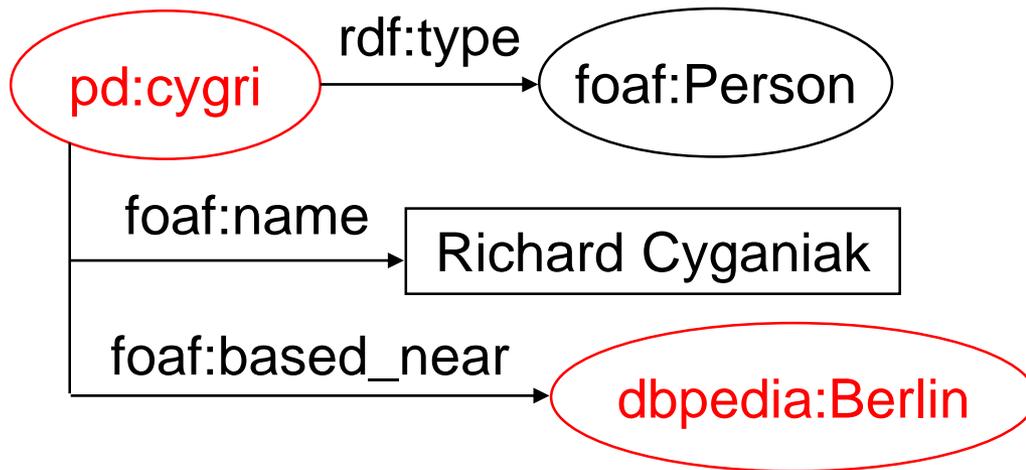
1. Publica datos estructurados en la Web,
2. Establece enlaces entre datos en diferentes Fuentes.



Modelo RDF



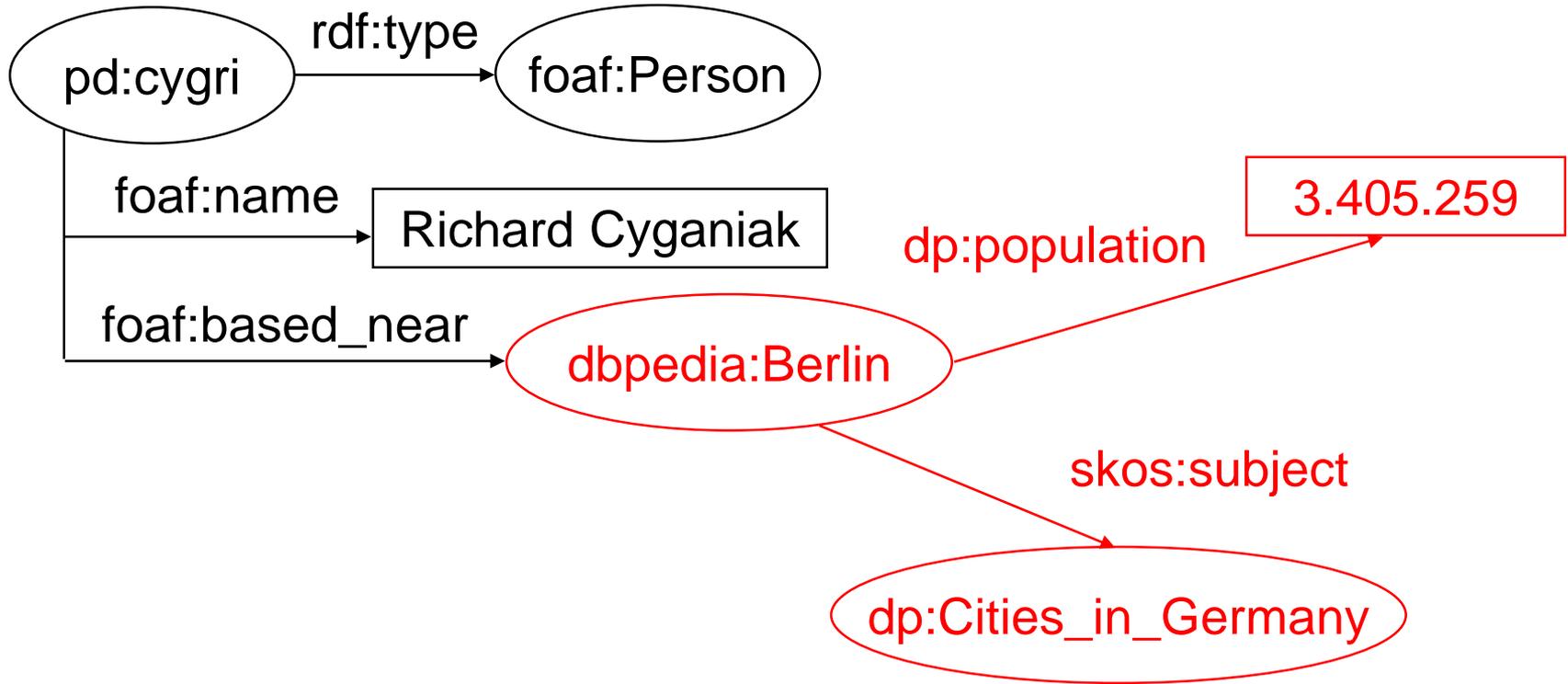
Datos son identificados con URIs http



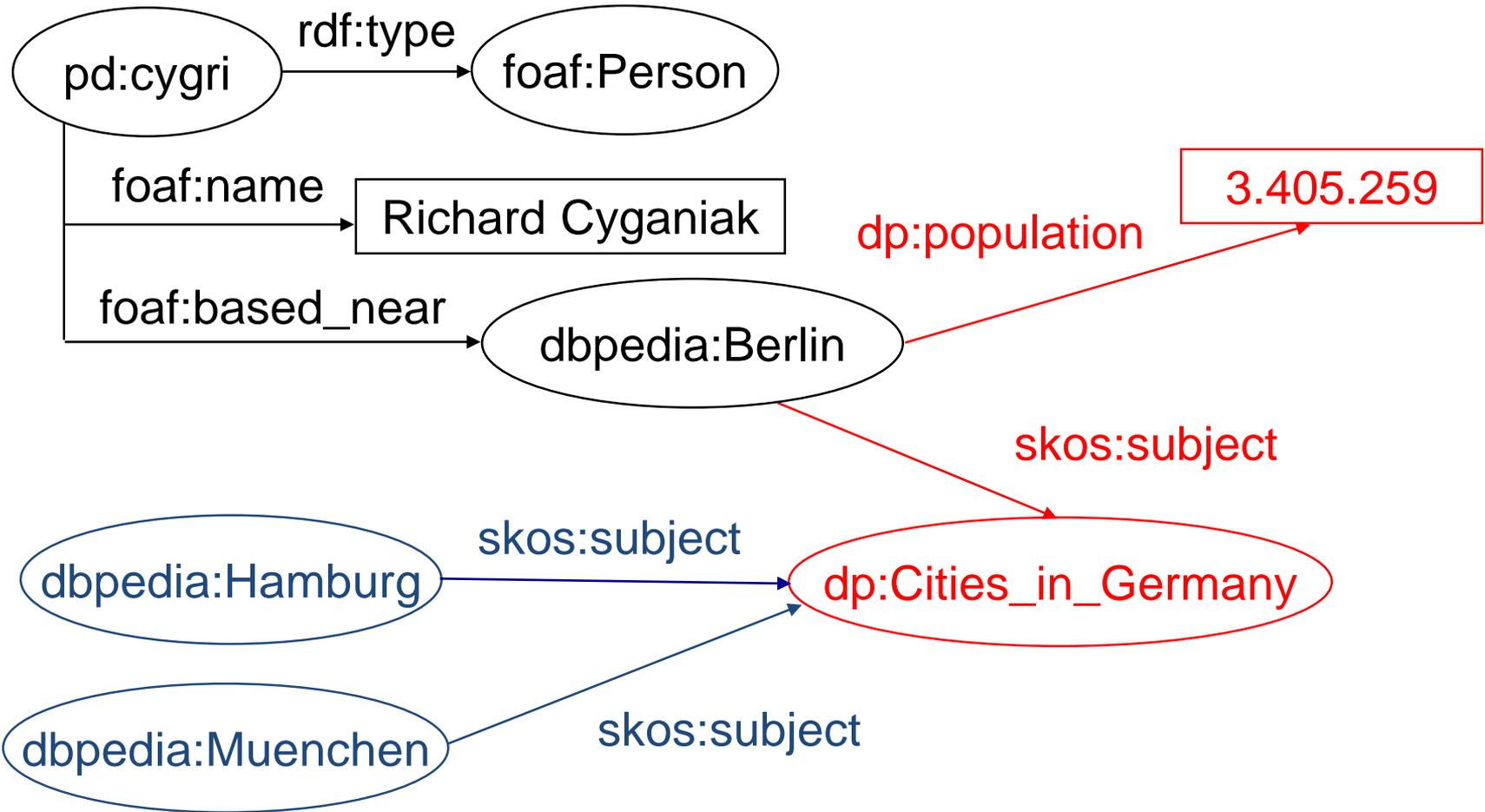
pd:cygri = <http://richard.cyganiak.de/foaf.rdf#cygri>

dbpedia:Berlin = <http://dbpedia.org/resource/Berlin>

URIs en la Web



Enlazando URIs en la Web



FOAF

- Permite crear paginas Web para describir personas, vínculos entre ellos, y cosas que hacen y crean.
- Es un vocabulario RDF (<http://xmlns.com/foaf/spec/>) que permite tener disponible información personal de forma sencilla y simplificada para que pueda ser procesada, compartida y reutilizada.

FOAF: Conceptos fundamentales

- **Concepto básico:** <foaf:Person>
- **Propiedades simples:** name, title, familyName, nick
- **Enlaces web:** depiction (foto), homepage, workplaceHomepage
- **Relaciones entre personas:** <foaf:knows>.

Ejemplo

```
<?xml version="1.0" standalone="yes"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <foaf:Person>
    <foaf:name>
      Jose Aguilar
    </foaf:name>
    <foaf:mbox rdf:resource=mailto:aguilar@ula.ve/>
  </foaf:Person>
</rdf:RDF>
```

Indicando Datos Externos

Para navegar por FOAF es por medio del esquema RDFS y su propiedad

`<rdfs:seeAlso>`

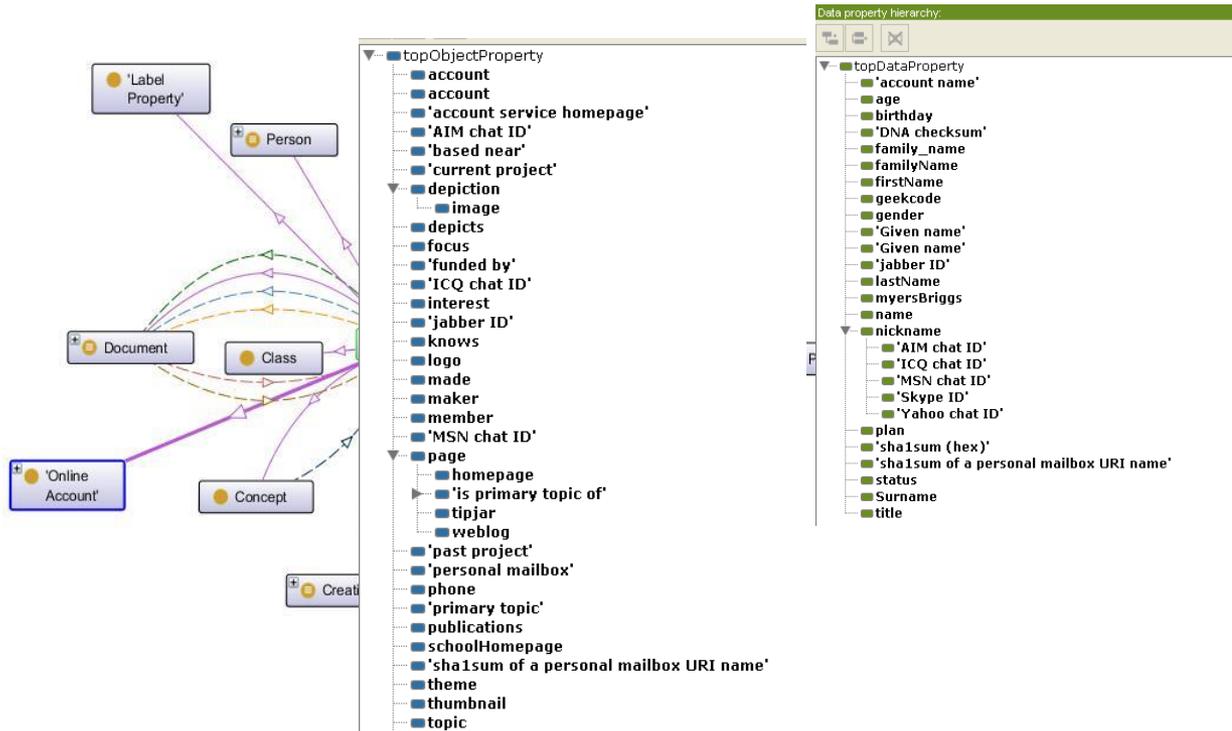
Expresando Relaciones

Para expresar que se conoce a alguien en FOAF es por medio de la propiedad **Knows**

```
<?xml version="1.0" standalone="yes"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <foaf:Person>
    <foaf:name>
      Jose Aguilar
    </foaf:name>
    <foaf:mbox rdf:resource=mailto:aguilar@ula.ve/>
    <foaf:knows>
      <foaf:Person>
        <foaf:name> Taniana Rodriguez</foaf:name>
        <foaf:mbox rdf:resource=mailto:taniana@ula.ve/>
        <rdfs:seeAlso rdf:resource="paginaWebTaniana/foaf.rdf" />
      </foaf:Person>
    </foaf:knows>
  </foaf:Person>
</rdf:RDF>
```

Especificación de FOAF



Generar un FOAF desde FOAF a Matic

- FOAF-a-matic que permite generar un FOAF desde FOAF a Matic
 - <http://www.foaf-a-matic.org/>
- ```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:foaf="http://xmlns.foaf.org/2000/01/01/foaf#" xmlns:vcard="http://www.w3.org/2006/10/vcard-rdf/1.0#" xmlns:adm="http://www.foaf-a-matic.org/2006/01/01/admin#">Personal
Información acerca de usted, y como pueden contactarlo.
foaf:title Titulo (Sr, Sra, Dr, etc)
foaf:name Nombre/s
foaf:surname Apellido
foaf:nickname Apodo
foaf:mbox Su Dirección de Email
foaf:homepage <foaf:homepage Su Homepage
foaf:photo <foaf:photo Su Foto
foaf:phone <foaf:phone Teléfono
Trabajo
Información acerca de su lugar de trabajo.
foaf:workHomepage <foaf:workHomepage Homepage del Trabajo
foaf:workDescription <foaf:workDescription Página describiendo su trabajo
Escuela
foaf:givenName <foaf:givenName ¿A que escuela fue?
foaf:familyName <foaf:familyName Homepage de su Escuela
Gente Que Conoce
foaf:workInfo <foaf:workInfo Informe a FOAF-a-matic acerca de la gente que conoce. Presione "Agregar Amigo" para agregar espacio para agregar más gente. Si su amigo ya tiene una descripción FOAF, entonces introduzca la
foaf:knows <foaf:knows Tambien".
foaf:knows <foaf:knows Amigo-- Nombre Email Ver Tambien
foaf:knows <foaf:knows Amigo-- Nombre Email Ver Tambien
foaf:knows <foaf:knows Amigo-- Nombre Email Ver Tambien
</foaf:knows <foaf:knows Agregar Amigo
</foaf:knows
Generate Results
</foaf:knows <foaf:knows Ahora que ha completado los detalles, usted es listo para convertirse a FOAF...
</rdf:RDF <foaf:knows Proteger direcciones de email de los spammers

```

# Richard Cyganiak

URI: 

| Property        | Value                                                                                                                       | Sources                                                                     |
|-----------------|-----------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| event           | ...                                                                                                                         | <a href="#">G2</a>                                                          |
| type            | <a href="http://xmlns.com/foaf/0.1/Person">http://xmlns.com/foaf/0.1/Person</a>                                             | <a href="#">G1</a> <a href="#">G2</a> <a href="#">G3</a> <a href="#">G4</a> |
| seeAlso         | <a href="http://richard.cyganiak.de/cygri.rdf">http://richard.cyganiak.de/cygri.rdf</a>                                     | <a href="#">G2</a>                                                          |
| seeAlso         | <a href="http://richard.cyganiak.de/foaf.rdf">http://richard.cyganiak.de/foaf.rdf</a>                                       | <a href="#">G3</a>                                                          |
| nearest airport | ...                                                                                                                         | <a href="#">G1</a>                                                          |
| phone           | tel:+49-175-5630408                                                                                                         | <a href="#">G1</a>                                                          |
| sameAs          | <a href="#">Richard Cyganiak</a>                                                                                            | <a href="#">G1</a>                                                          |
| based_near      | ...                                                                                                                         | <a href="#">G1</a>                                                          |
| based_near      | <a href="#">Berlin</a>                                                                                                      | <a href="#">G1</a>                                                          |
| based_near      | <a href="http://sws.geonames.org/2950159/">http://sws.geonames.org/2950159/</a>                                             | <a href="#">G1</a>                                                          |
| currentProject  | <a href="http://page.mi.fu-berlin.de/~cyganiak/foaf.rdf#StatCvs">http://page.mi.fu-berlin.de/~cyganiak/foaf.rdf#StatCvs</a> | <a href="#">G3</a>                                                          |
| currentProject  | <a href="http://www.wiwiss.fu-berlin.de/suhl/bizer#d2rq">http://www.wiwiss.fu-berlin.de/suhl/bizer#d2rq</a>                 | <a href="#">G3</a>                                                          |
| depiction       |                                           | <a href="#">G4</a>                                                          |
| gender          | male                                                                                                                        | <a href="#">G1</a>                                                          |
| holdsAccount    | <a href="#">cygri@delicio.us</a>                                                                                            | <a href="#">G1</a>                                                          |

# Berlin

URI: 

| Property         | Value                                                                                                                                                                                                                                          | Sources            |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| population       | 3398888                                                                                                                                                                                                                                        | <a href="#">G2</a> |
| type             | <a href="http://dbpedia.org/City">http://dbpedia.org/City</a> <a href="#">↗</a>                                                                                                                                                                | <a href="#">G2</a> |
| comment          | Berlin is the capital city and one of the sixteen Federal States of Germany. It is the country's largest city in area and population, and the second most populous city in the European Union.                                                 | <a href="#">G2</a> |
| comment          | Berlin ist die deutsche Bundeshauptstadt und als Stadtstaat ein eigenständiges Land der Bundesrepublik Deutschland. Berlin ist die bevölkerungsreichste und flächengrößte Stadt Deutschlands und nach Einwohnern die zweitgrößte Stadt der EU. | <a href="#">G2</a> |
| label            | Berlin                                                                                                                                                                                                                                         | <a href="#">G2</a> |
| sameAs           | <a href="http://sws.geonames.org/2950159/">http://sws.geonames.org/2950159/</a> <a href="#">↗</a>                                                                                                                                              | <a href="#">G2</a> |
| subject          | <a href="http://dbpedia.org/resource/category/Berlin">http://dbpedia.org/resource/category/Berlin</a> <a href="#">↗</a>                                                                                                                        | <a href="#">G2</a> |
| subject          | <a href="http://dbpedia.org/resource/category/Capitals_in_Europe">http://dbpedia.org/resource/category/Capitals_in_Europe</a> <a href="#">↗</a>                                                                                                | <a href="#">G2</a> |
| subject          | <a href="http://dbpedia.org/resource/category/Cities_in_Germany">http://dbpedia.org/resource/category/Cities_in_Germany</a> <a href="#">↗</a>                                                                                                  | <a href="#">G2</a> |
| subject          | <a href="http://dbpedia.org/resource/category/German_state_capitals">http://dbpedia.org/resource/category/German_state_capitals</a> <a href="#">↗</a>                                                                                          | <a href="#">G2</a> |
| subject          | <a href="http://dbpedia.org/resource/category/Host_cities_of_the_Summer_Olympic_Games">http://dbpedia.org/resource/category/Host_cities_of_the_Summer_Olympic_Games</a> <a href="#">↗</a>                                                      | <a href="#">G2</a> |
| subject          | <a href="http://dbpedia.org/resource/category/States_of_Germany">http://dbpedia.org/resource/category/States_of_Germany</a> <a href="#">↗</a>                                                                                                  | <a href="#">G2</a> |
| sourceURL        | <a href="#">Berlin</a> <a href="#">↗</a>                                                                                                                                                                                                       | <a href="#">G1</a> |
| depiction        |                                                                                                                                                             | <a href="#">G2</a> |
| page             | <a href="http://en.wikipedia.org/wiki/Berlin">http://en.wikipedia.org/wiki/Berlin</a> <a href="#">↗</a>                                                                                                                                        | <a href="#">G2</a> |
| is birthplace of | <a href="#">Adolf von Baeyer</a> <a href="#">↗</a>                                                                                                                                                                                             | <a href="#">G2</a> |
| is birthplace of | <a href="http://dbpedia.org/resource/person/Albert_Speer_%28the_younger%29">http://dbpedia.org/resource/person/Albert_Speer_%28the_younger%29</a> <a href="#">↗</a>                                                                            | <a href="#">G2</a> |

# Microformats

- Son XML tags que son incorporado dentro de las páginas Web para soportar declaraciones semánticas
  - Enriquecen sitios webs con atributos, con el fin de hacer una declaración semántica, para los agentes de software.
- ✓ XFN (**XHTML Friends Network**, <http://gmpg.org/xfn/>), representa relaciones de personas usando hyperlinks.

Por ejemplo:

- ✓ Supongamos que en la página Web de Taniana Rodríguez tiene un enlace a la página Web del Jose Aguilar
  - ❖ ¿Ellos son amigo?
  - ❖ ¿Ellos trabajan juntos?
  - ❖ ....
- ✓ Si se añade explícitamente en la página Web de Taniana
  - ❖ `<a href=ref="friend co-worker"> Jose Aguilar </a>`
  - ❖ indica que Taniana y Jose trabajan juntos (friend como co-worker están definido en XFN microformat)

*XFN quick reference*

| <i>relationship category</i> | <i>XFN values</i>                                                                                               |
|------------------------------|-----------------------------------------------------------------------------------------------------------------|
| friendship (at most one):    | <a href="#">friend</a> <a href="#">acquaintance</a> <a href="#">contact</a>                                     |
| physical:                    | <a href="#">met</a>                                                                                             |
| professional:                | <a href="#">co-worker</a> <a href="#">colleague</a>                                                             |
| geographical (at most one):  | <a href="#">co-resident</a> <a href="#">neighbor</a>                                                            |
| family (at most one):        | <a href="#">child</a> <a href="#">parent</a> <a href="#">sibling</a> <a href="#">spouse</a> <a href="#">kin</a> |
| romantic:                    | <a href="#">muse</a> <a href="#">crush</a> <a href="#">date</a> <a href="#">sweetheart</a>                      |
| identity:                    | <a href="#">me</a>                                                                                              |

## Creador de XFN 1.1

Nombre

URL   otra dirección web que me pertenece

amistad  contacto  conocido  amigo  ninguno

físico  conocido en persona

profesional  compañero de trabajo  colega

geográfico  compañero de vivienda  vecino  ninguno

familiar  hijo  padre  hermano  matrimonio  familiar  ninguno

romántico  musa  atracción  cita  amor

```
Jose Aguilar
```

# Microformats



hCard (<http://microformats.org/wiki/hcard>) es un microformato que permite marcar los datos de cualquier persona o entidad

## Properties

Common hCard properties (inside class vcard)

- fn - name, formatted/full. required
- n - structured name, container for:
  - honorific-prefix - e.g. Ms., Mr., Dr.
  - given-name - given (often first) name
  - additional-name - other/middle name
  - family-name - family (often last) name
  - honorific-suffix - e.g. Ph.D., Esq.
- nickname - nickname/alias, e.g. IRC nick
- org - company/organization
- photo - photo, icon, avatar
- url - home page for this contact
- email - email address
- tel - telephone number
- adr - structured address, container for:
  - street-address - street #+name, apt/ste
  - locality - city or village
  - region - state or province
  - postal-code - postal code, e.g. U.S. ZIP
  - country-name - country name
- bday - birthday. ISO date.
- category - for tagging contacts
- note - notes about the contact

```
<div class="vcard">
 Sally Ride
 (
 Dr.
 Sally
 <abbr class="additional-name">K.</abbr>
 Ride
 Ph.D.),
 sallykride (IRC)
 <div class="org">Sally Ride Science</div>

 w,
 e
 <div class="tel">+1.818.555.1212</div>
 <div class="adr">
 <div class="street-address">123 Main st.</div>
 Los Angeles,
 <abbr class="region" title="California">CA</abbr>,
 91316
 <div class="country-name">U.S.A.</div></div>
 <time class="bday">1951-05-26</time> birthday
 <div class="category">physicist</div>
 <div class="note">1st American woman in space.</div>
</div>
```

## hCard Creator

hCard-o-matic

given name	<input type="text" value="Taniana Rodriguez"/>
middle name	<input type="text"/>
family name	<input type="text" value="Taniana Josefina Rodrig"/>
organization	<input type="text" value="Universidad de Los Andes"/>
street	<input type="text"/>
city	<input type="text" value="Mérida"/>
state/province	<input type="text" value="Mérida"/>
postal code	<input type="text" value="5101"/>
country name	<input type="text" value="Venezuela"/>
phone	<input type="text"/>
email	<input type="text" value="taniana@ula.ve"/>
url	<input type="text" value="http://taniana.novacorp.co/"/>
photo url	<input type="text"/>
AIM screenname	<input type="text"/>
YIM screenname	<input type="text"/>
Jabber screenname	<input type="text"/>
Categories (comma separated)	<input type="text"/>

Warning - publishing your email address, phone number or instant messenger screenname on the web can open it up to abuse.

```
code
<div id="hcard-Taniana-Rodriguez-Taniana-Josefina-Rodriguez-de-Paredes"
class="vcard">
 <span class="given-
name">Taniana Rodriguez

Taniana Josefina Rodriguez de
Paredes
</div>
<div class="org">Universidad de Los Andes</div>
taniana@ula.ve
<div class="adr">
Mérida
```

### preview

Taniana Rodriguez Taniana Josefina Rodriguez de Paredes  
Universidad de Los Andes  
[taniana@ula.ve](mailto:taniana@ula.ve)  
Mérida , Mérida , 5101 Venezuela  
This hCard created with the hCard creator.

# Microformats

- ✓ hcalendar (<http://microformats.org/wiki/hcalendar>), es un estándar de microformat de la información de un evento, en formato iCalendar

## Property List

hCalendar properties (sub-properties in parentheses like this)

### Required:

- ▶ **dtstart** ([ISO date](#))
- ▶ **summary**

### Optional:

- ▶ location
- ▶ url
- ▶ dtend (ISO date), duration (ISO date duration)
- ▶ rdate, rrule
- ▶ category, description
- ▶ uid
- ▶ geo (latitude, longitude)
- ▶ attendee (partstat, role), contact, organizer
- ▶ attach
- ▶ status
- ▶ editor's note: this list is incomplete (an incomplete list is better)

## hCalendar Creator

hCalendar-o-matic

**summary**  
Seminarío de Minería Semántica

**location**  
Mérida

**url**  
http://

**start** November 17 2013 7 :  
**end** November 17 2013 8 :  
**TZ** none GMT hour(s) from

**description**

**tags**  
(comma separated)

Reset Build It!

### code

```
<div class="vevent" id="hcalendar-Seminario-de-Mineria-Semántica-"><time datetime="2013-11-17T07:00" class="dtstart">November 17, 2013 7</time>-<time datetime="2013-11-17T08:00" class="dtend">8am</time> : Seminarío de Minería Semántica at Mérida
```

```
<p style="font-size: smaller;">This hCalendar event brought to you by the hCalendar Creator.</p></div>
```

### compact code

```
<div class="vevent" id="hcalendar-Seminario-de-Mineria-Semántica-"><time datetime="2013-11-17T07:00" class="dtstart">November 17, 2013 7</time>-<time datetime="2013-11-17T08:00" class="dtend">8am</time> : Seminarío de Minería Semántica
```

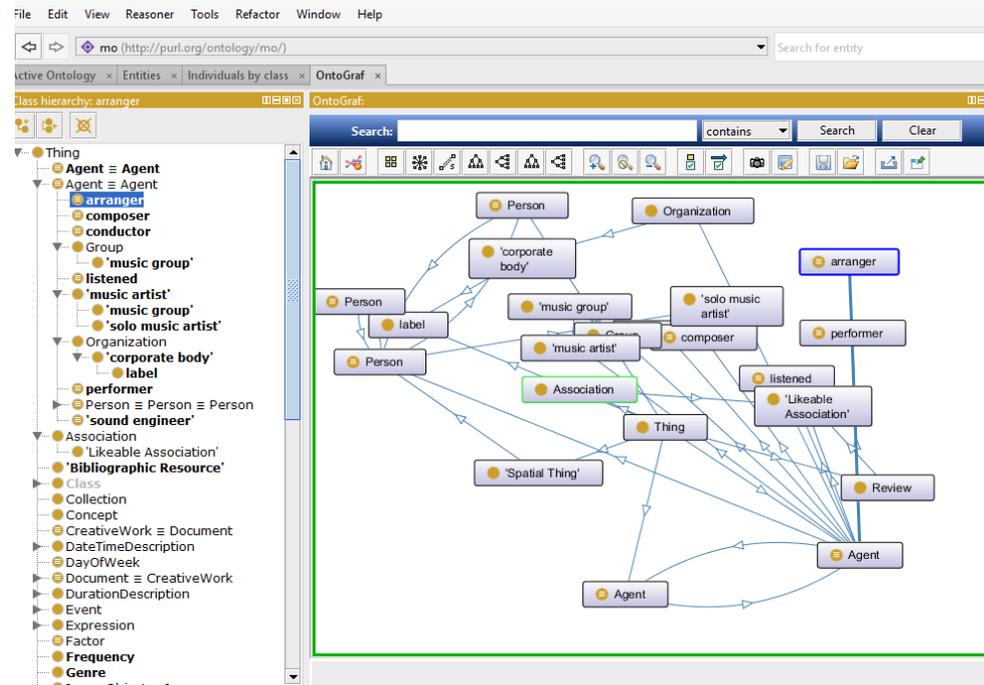
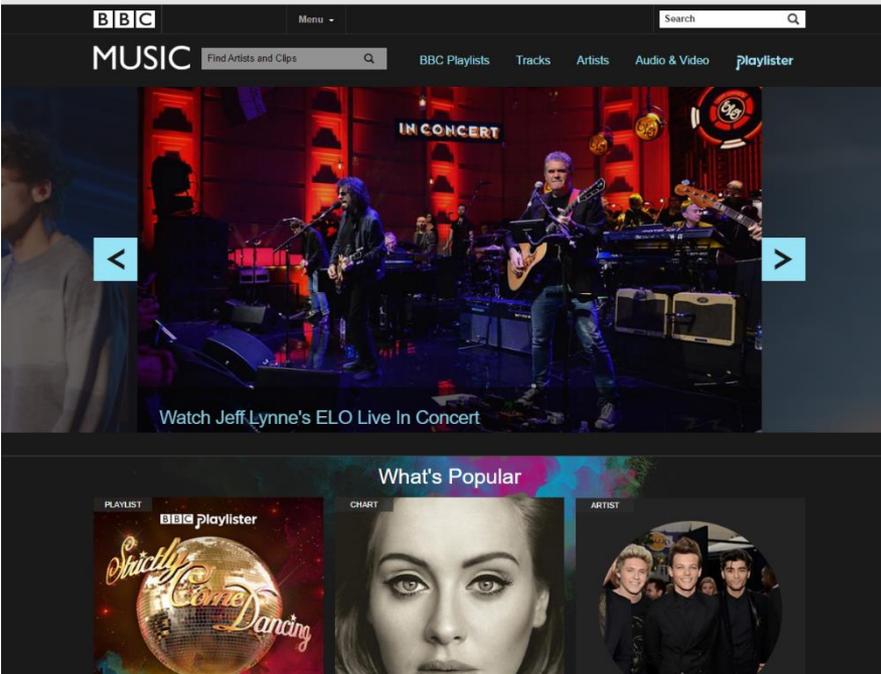
### preview

November 17, 2013 7–8am : Seminarío de Minería Semántica at Mérida

This [hCalendar event](#) brought to you by the [hCalendar Creator](#).

# RDFa

- RDFa añade semántica a las páginas Web
- Las máquinas encuentran y recuperan información de las páginas Web mas fácil usando RDFa
- RDFa define atributos para las palabras o frases que pueden ser tratadas como entidades semánticas



# RDFa ejemplo

Servicio experto. Precio inigualable. Aviso semanal Tarj. crédito Tarj. regalo Ideas de regalos Registro Pedidos Tiendas

**BEST BUY** PRODUCTOS SERVICIOS OFERTAS  Iniciar sesión Crear cuenta

**PRECIOS DE BLACK FRIDAY DISPONIBLES MEJORES OFERTAS EN ELECTRODOMÉSTICOS GRANDES: 25%-40% DE**

Best Buy

## Electrodomésticos

**BLACK FRIDAY**  
- PRECIOS DISPONIBLES -  
**25%-40% MENOS**  
OFERTAS EN ELECTRODOMÉSTICOS GRANDES  
Se aplican restricciones.  
Ver electrodomésticos grandes en oferta >

FINANCIACIÓN DE 18 MESES SOBRE COMPRAS DE \$479 O MÁS  
Oferta válida 11/1/15-12/26/15.  
[Conozca más >](#)

ENVÍO GRATIS  
Válido para compras de electrodomésticos grandes desde \$399. Incluye el acarreo y reciclaje gratis.  
[Conozca más >](#)

### Electrodomésticos de cocina grandes >

Desde nitida y clara hasta elegante y moderna, cree una apariencia de lujo en su cocina.

Refrigeradores

Lavavajillas

### Electrodomésticos de cocina >

Mejore sus habilidades culinarias con electrodomésticos que simplifican su vida y mejoran la apariencia de su mostrador.

Ollas y cacerolas

My Account | Order Status | Customer Service | Español

**BEST BUY** Weekly Ad Store Locator Outlet Center Services Gifts

TV & VIDEO AUDIO CAR & GPS CAMERAS & CAMCORDERS COMPUTERS MOBILE PHONES & OFFICE MUSIC, MOVIES & BOOKS VIDEO GAMES & GADGETS HOME & APPLIANCES

Search    Credit Cards Reward Zone®

Best Buy - Carbondale [store name](#)

1279 E Main St  
Carbondale, IL 62901  
Phone: 618-351-1709  
GEO: 37.732719, -89.192314

Customer Reviews: [review data](#)  
Be the first to write a store review

Maps & Directions | Weekly Ad

Store Hours  
Mon: 10-9; Tue: 10-9; Wed: 10-9; Thurs: 10-9; Fri: 10-9; Sat: 10-9; Sun: 11-7;  
414 - 476, 2010  
[store hours](#)

Mon: 10-9; Tues: 10-9; Wed: 10-9; Thurs: 10-9; Fri: 10-9; Sat: 10-9; Sun: Closed

### Local Selections

[address](#) [phone](#) [geo](#) [store image](#)

Check out these special product selections from this store.

### Open Box Items (25)

### At This Location

- Geek Squad**  
Computer setup & services, plus home theater, appliance and car installation.
- services**  
Get informed advice from noncommissioned mobile phone specialists.
- Small Business Solutions**  
Featuring Professional Series products and trained staff to help with small business needs.
- Apple Shop**  
Mac, iPod and more at this Apple store-within-a-store.
- Electronics Recycling**  
We offer electronics recycling at this and all other U.S. stores.

### Events

[event data](#)

Avatar Midnight Release!

Best Buy employees entered information into the blogs every day, using online forms that output RDFa. Myers told us that the use of RDFa makes "human input from our store employees more visible on the Web."

Best Buy is using [Good Relations](#), a Semantic Web vocabulary for e-commerce that describes product, price, and company data.

# Propiedades de la Web de datos enlazados

- Cualquiera puede publicar datos
- Entidades están conectadas por enlaces
  - Un global grafo de datos que expande las fuentes de datos, descubriendo nuevas.
- Datos se auto-describen
  - Si una aplicación encuentra un dato con vocabulario no familiar, la aplicación, usando el URIs que identifica los términos del vocabulario, puede encontrar los RDFS o OWL con su definición.
- Esta Web es Open Data

# SEO (Search Engine Optimization)

Según Wikipedia, el SEO es:

“es el proceso de mejorar la visibilidad de un sitio Web”

Entidades y tripletas: la base de la Web Semántica

- ya no son palabras claves, se trata ahora de entidades (personas, lugares, organización, eventos, objetos, etc.)
- Las entidades pueden tener múltiples relaciones con otras entidades.

The screenshot shows a Google search for "quien es el director de matrix". The search bar is at the top, and the results are displayed below. The first result is "The Matrix / Directores" with a sub-heading "Andy Wachowski" and a photo of Lana Wachowski. Below this, there are several search results from Wikipedia and other sources, including "The Matrix - Wikipedia, la enciclopedia libre", "Hermanos Wachowski - Wikipedia, la enciclopedia libre", "Larry Wachowski, el director de «The Matrix», se llama Lana ...", and "Larry Wachowski, director de 'Matrix', cambia de sexo - El ...".

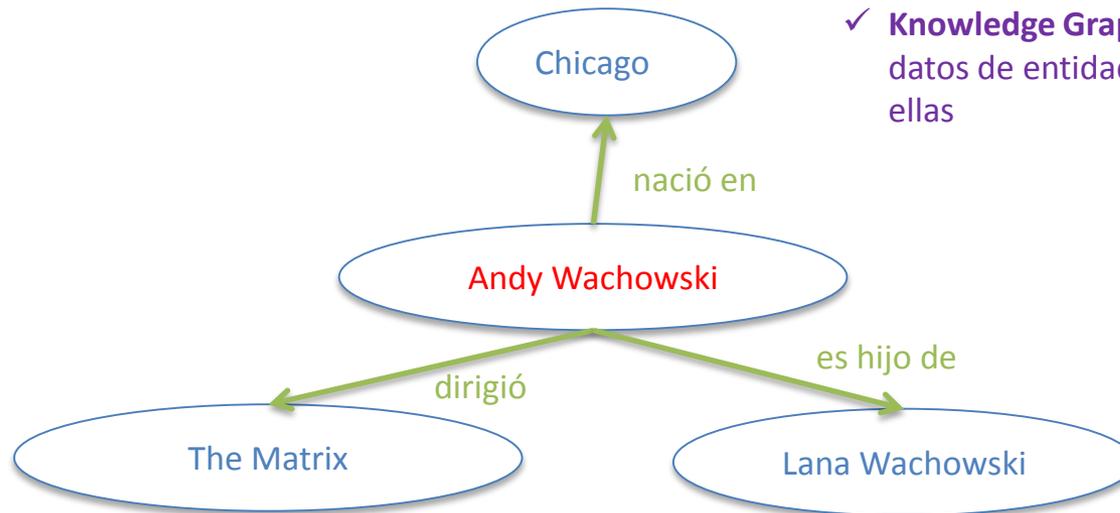
- **Andy Wachowski** nació en Chicago
- **Andy Wachowski** es hijo de Lana Wachowski
- **Andy Wachowski** dirigió The Matrix

- Información puede ser extraída de diferentes fuentes: Dbpedia, IMDB, Wikipedia etc.
- Basado en una representación del conocimiento  
**Sujeto + Predicado + Objeto**

El sujeto es la entidad que se esta describiendo,  
el predicado es que se esta describiendo del sujeto  
el objeto es el valor del predicado

# SEO Semántico

Andy Wachowski nació en Chicago  
Andy Wachowski es hijo de Lana Wachowski  
Andy Wachowski dirigió The Matrix

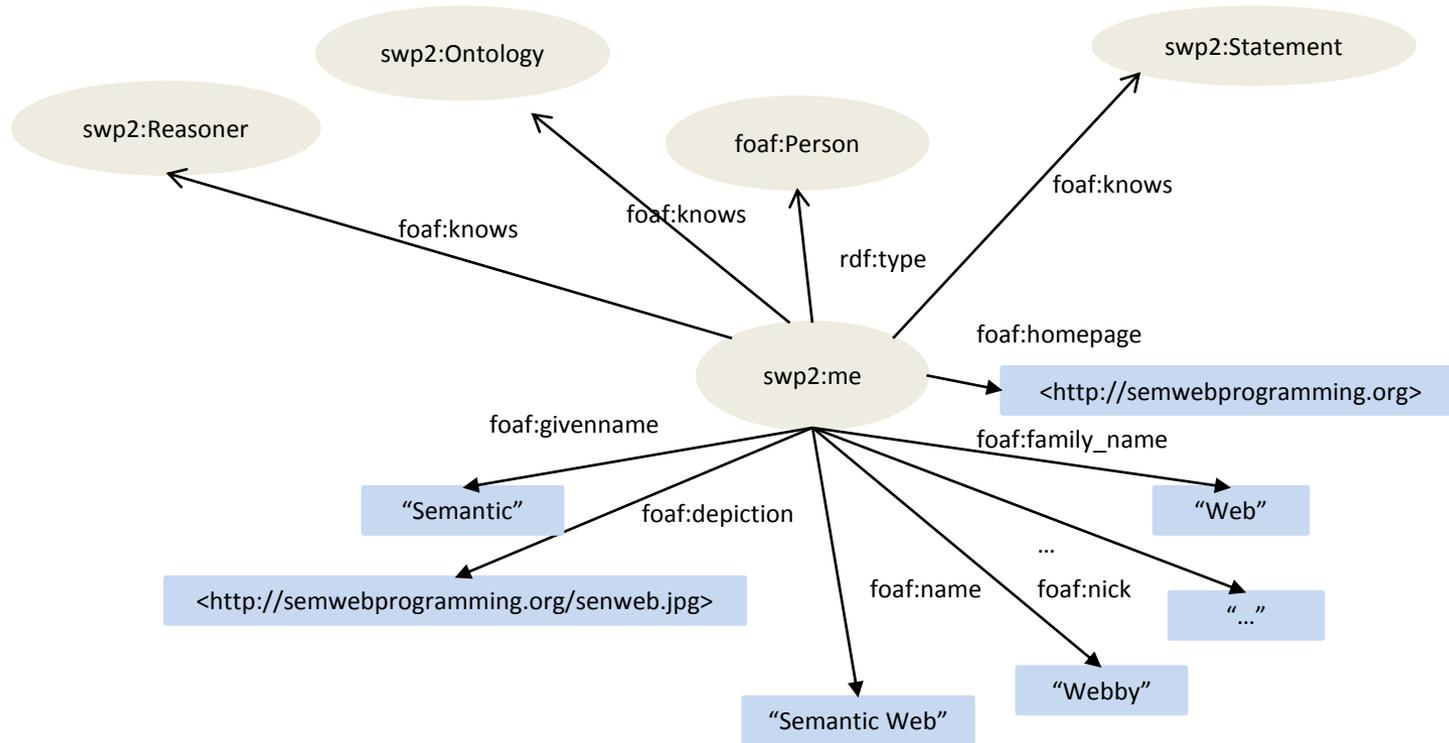


- ✓ Las tripletas se representa en grafos.
- ✓ **Knowledge Graph** -> es una base de datos de entidades y relaciones entre ellas

- ✓ SEO semántico tiene como objetivo de ayudar a los buscadores a entender exactamente de qué trata tus páginas.
- ✓ Para ello, sigue los siguientes pasos
  - Determinar las entidades correspondientes a la página.
  - Desambiguarlas directamente
  - Desambiguarlas indirectamente.

# Ejemplo

```
swp2:me rdf:type foaf:Person .
swp2:me foaf:depiction <http://semwebprogramming.org/senweb.jpg >.
swp2:me foaf:family_name "Web" .
swp2:me foaf:givenname "Semantic" .
swp2:me foaf:homepage <http://semwebprogramming.org >.
swp2:me foaf:knows "Reasoner" .
swp2:me foaf:knows "Statement" .
swp2:me foaf:knows "Ontology" .
swp2:me foaf:name "Semantic Web" .
swp2:me foaf:nick "Webby" .
swp2:me foaf:phone "<tel:410-679-8999>" .
swp2:me foaf:schoolInfoHomepage <http://www.web.edu > .
swp2:me foaf:title "Dr." .
swp2:me foaf:workInfoHomepage <http://semwebprogramming.com/dataweb.html > .
swp2:me foaf:workplaceHomepage <http://semwebprogramming.com > .
```

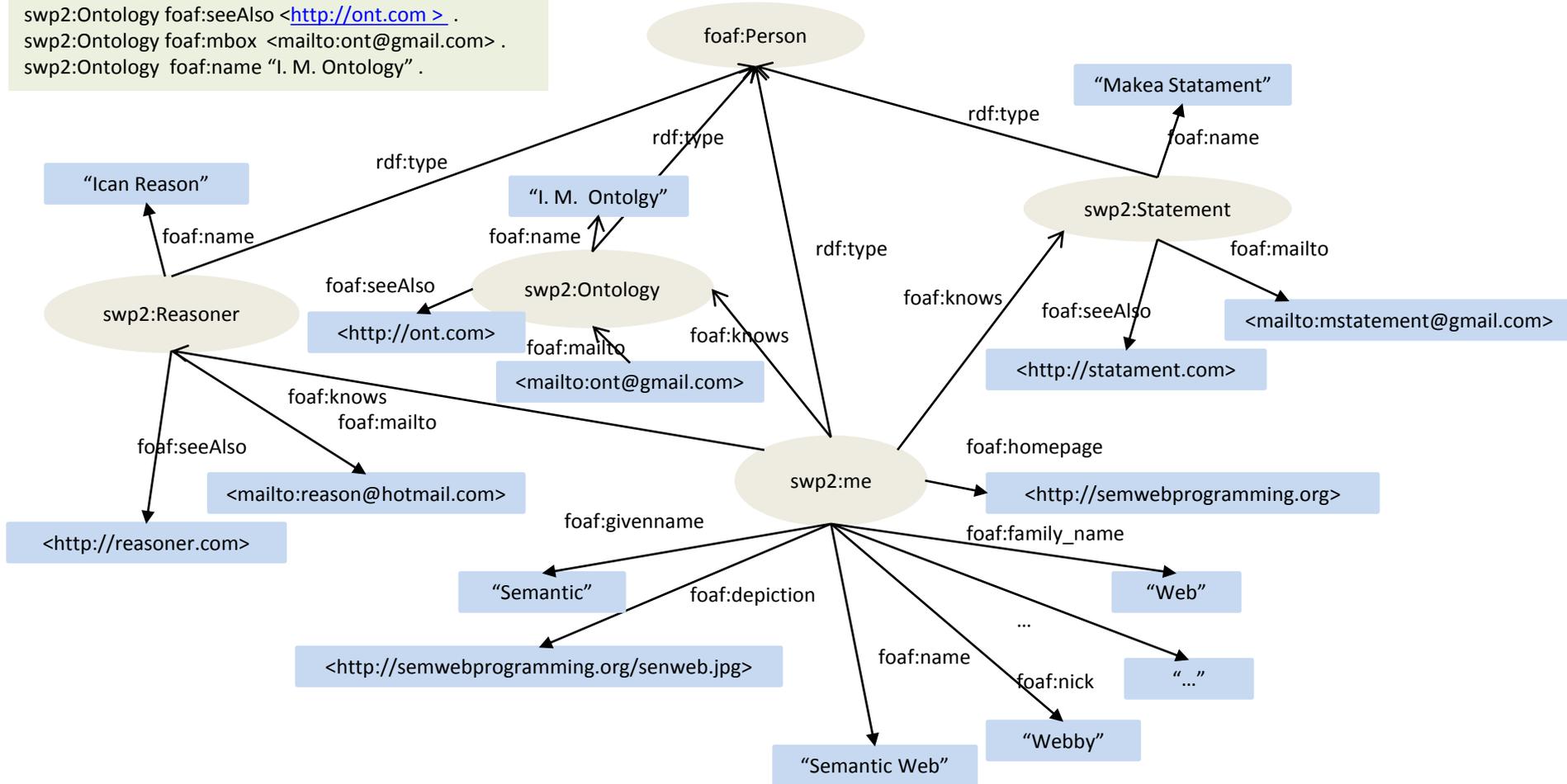






# Continuación del Ejemplo

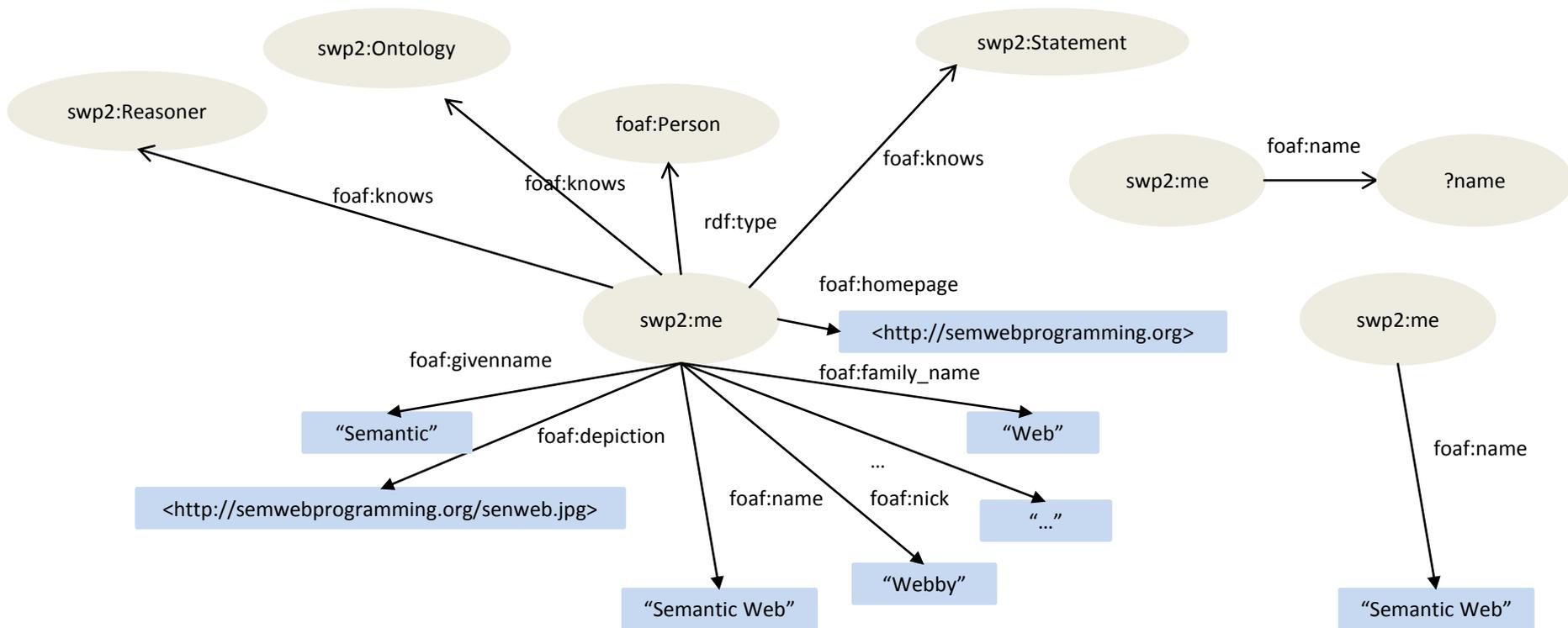
```
swp2:Ontology rdf:type foaf:Person .
swp2:Ontology foaf:seeAlso <http://ont.com> .
swp2:Ontology foaf:mbox <mailto:ont@gmail.com> .
swp2:Ontology foaf:name "I. M. Ontology" .
```



# Consulta

```
select DISTINCT ?name
where{
 swp2:me foaf:name ?name
}
```

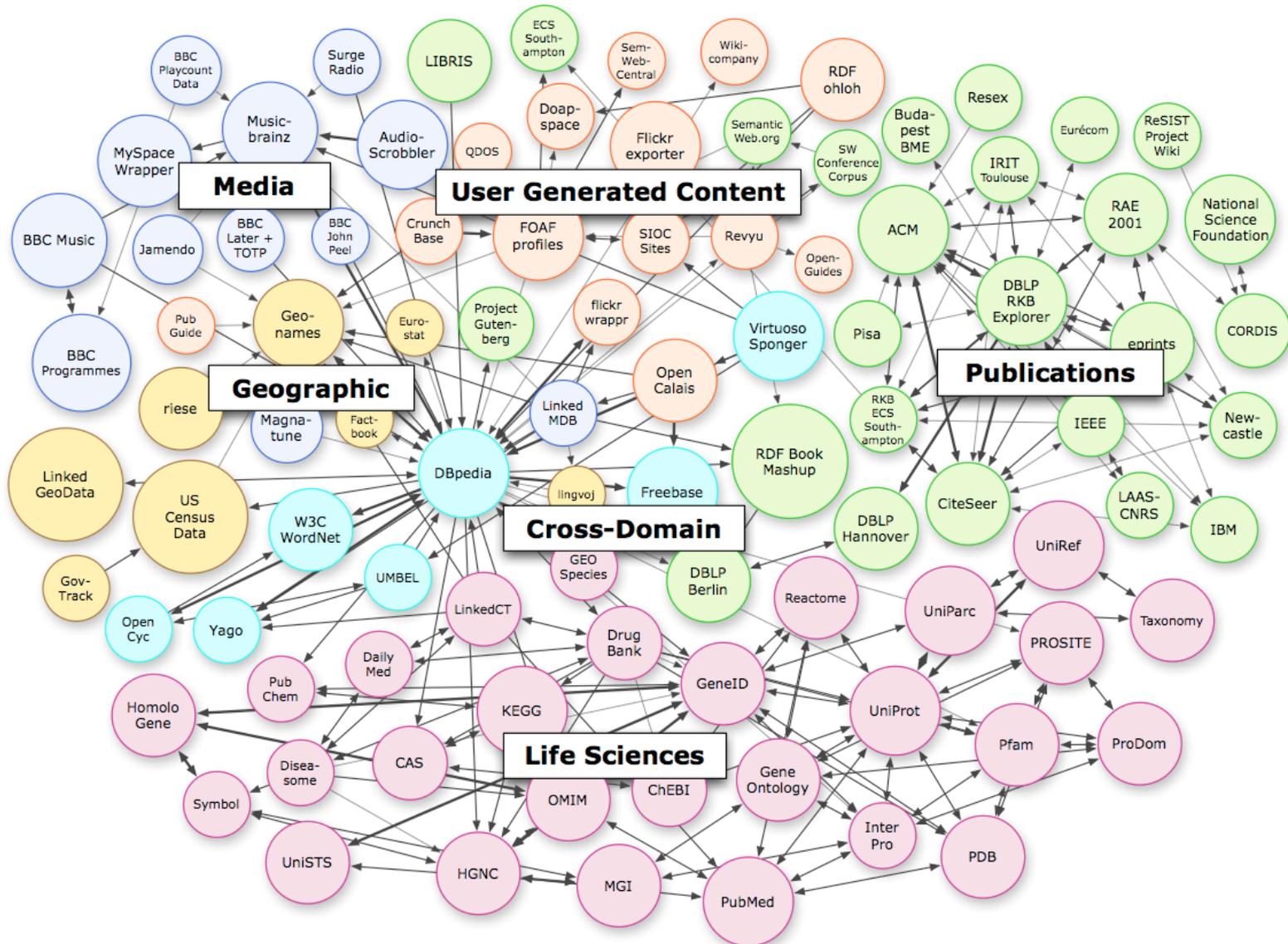
```
swp2:me rdf:type foaf:Person .
swp2:me foaf:depiction <http://semwebprogramming.org/senweb.jpg >.
swp2:me foaf:family_name "Web" .
swp2:me foaf:givenname "Semantic" .
swp2:me foaf:homepage <http://semwebprogramming.org >.
swp2:me foaf:knows "Reasoner" .
swp2:me foaf:knows "Statement" .
swp2:me foaf:knows "Ontology" .
swp2:me foaf:name "Semantic Web" .
swp2:me foaf:nick "Webby" .
swp2:me foaf:phone "<tel:410-679-8999>" .
swp2:me foaf:schoolInfoHomepage <http://www.web.edu >.
swp2:me foaf:title "Dr." .
swp2:me foaf:workInfoHomepage <http://semwebprogramming.com/dataweb.html> .
swp2:me foaf:workplaceHomepage <http://semwebprogramming.com > .
```







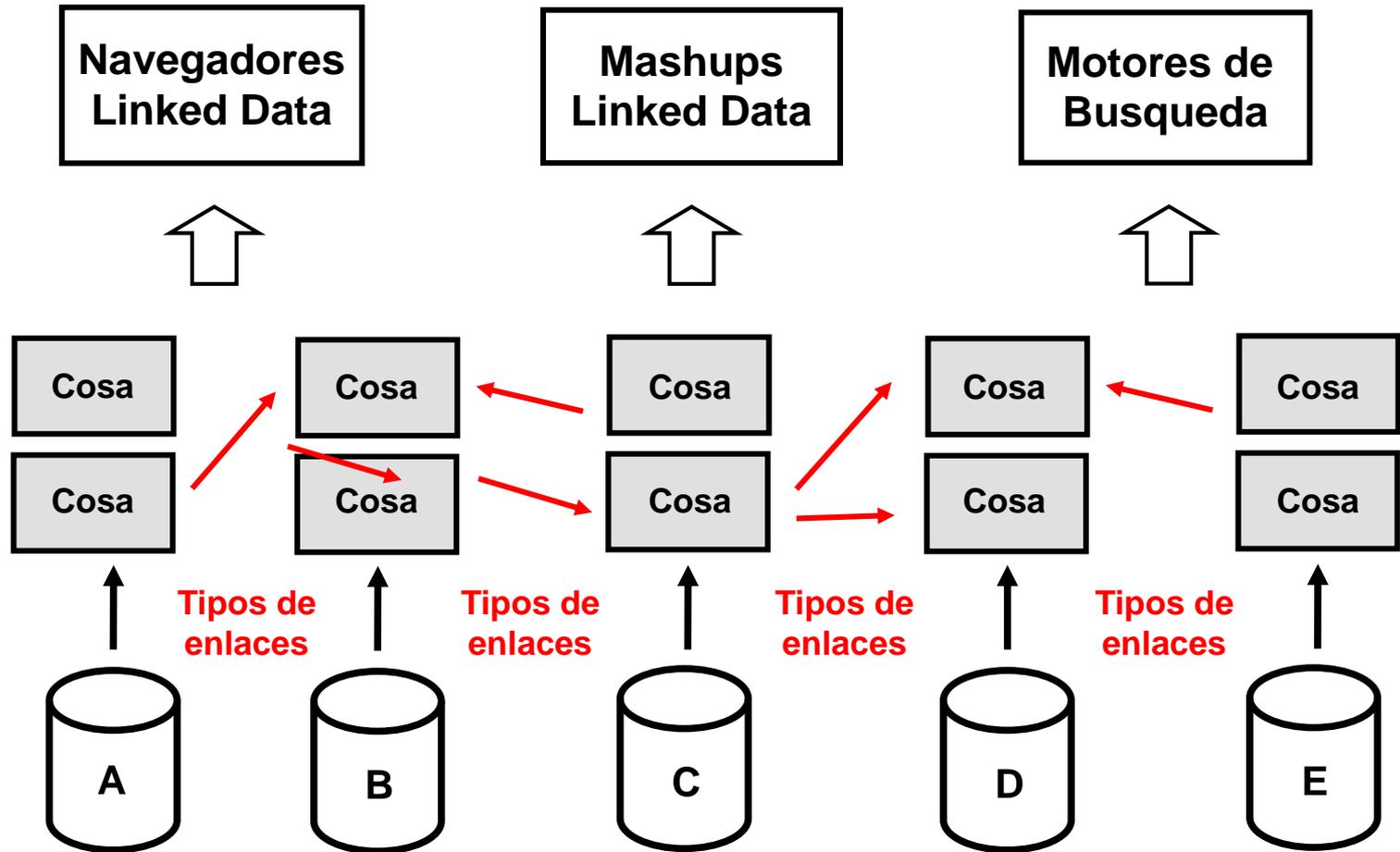
# Julio 2009



# Estadísticas de Julio 2009

<i>Dominio</i>	<i>No de Tripletas</i>	<i>%</i>	<i>No de enlaces</i>	<i>%</i>
Medios de comunicación	698.000.000	10,4%	1.238.000	0,8%
Publicaciones	212.000.000	3,2%	4.922.000	3,3%
Ciencias de la Vida	2.429.000.000	36,1%	133.199.000	89,4%
Datos Geográficos	3.097.000.000	46,0%	4.038.000	2,7%
Usuarios	76.000.000	1,1%	1.559.000	1,0%
Varios Dominios	214.000.000	3,2%	3.992.000	2,7%
<b>Total</b>	<b>6.726.000.000</b>		<b>148.948.000</b>	

# Aplicaciones



# Ejemplo Fuente de datos: DBpedia

## Amsterdam



The Keizersgracht at dusk

Location of Amsterdam

Coordinates:  52°22′23″N 4°53′32″E

<b>Country</b>	<b>Netherlands</b>
<b>Province</b>	<b>North Holland</b>
<b>Government</b>	
<span> </span> - <b>Type</b>	Municipality
<span> </span> - <b>Mayor</b>	Job Cohen <sup>[1]</sup> (PvdA)
<span> </span> - <b>Aldermen</b>	Lodewijk Asscher Carolien Gehrels Tjeerd Herrema Maarten van Poelgeest Marijke Vos
<span> </span> - <b>Secretary</b>	Erik Gerritsen
<b>Area</b> <sup>[2][3]</sup>	
<span> </span> - <b>City</b>	219 km <sup>2</sup> (84.6 sq mi)
<span> </span> - <b>Land</b>	166 km <sup>2</sup> (64.1 sq mi)
<span> </span> - <b>Water</b>	53 km <sup>2</sup> (20.5 sq mi)
<span> </span> - <b>Urban</b>	1,003 km <sup>2</sup> (387.3 sq mi)
<span> </span> - <b>Metro</b>	1,815 km <sup>2</sup> (700.8 sq mi)
<b>Elevation</b> <sup>[4]</sup>	2 m (7 ft)
<b>Population</b> (1 October 2008) <sup>[5][6]</sup>	
<span> </span> - <b>City</b>	755,269
<span> </span> - <b>Density</b>	4,459/km <sup>2</sup> (11,548.8/sq mi)
<span> </span> - <b>Urban</b>	1,364,422
<span> </span> - <b>Metro</b>	2,158,372
<span> </span> - <b>Demonym</b>	Amsterdammer
<b>Time zone</b>	CET (UTC+1)
<span> </span> - <b>Summer (DST)</b>	CEST (UTC+2)
<b>Postcodes</b>	1011 – 1109
<b>Area code(s)</b>	020

Website: [www.amsterdam.nl](http://www.amsterdam.nl) 

## Extraer estructurada información desde Wikipedia

```
@prefix dbpedia
<http://dbpedia.org/resource/>.
```

```
@prefix dbterm
<http://dbpedia.org/property/>.
```

```
dbpedia:Amsterdam
```

```
dbterm:officialName "Amsterdam" ;
```

```
dbterm:longd "4" ;
```

```
dbterm:longm "53" ;
```

```
dbterm:longs "32" ;
```

```
...
```

```
dbterm:leaderName dbpedia:Job_Cohen ;
```

```
...
```

```
dbterm:areaTotalKm "219" ;
```

```
...
```

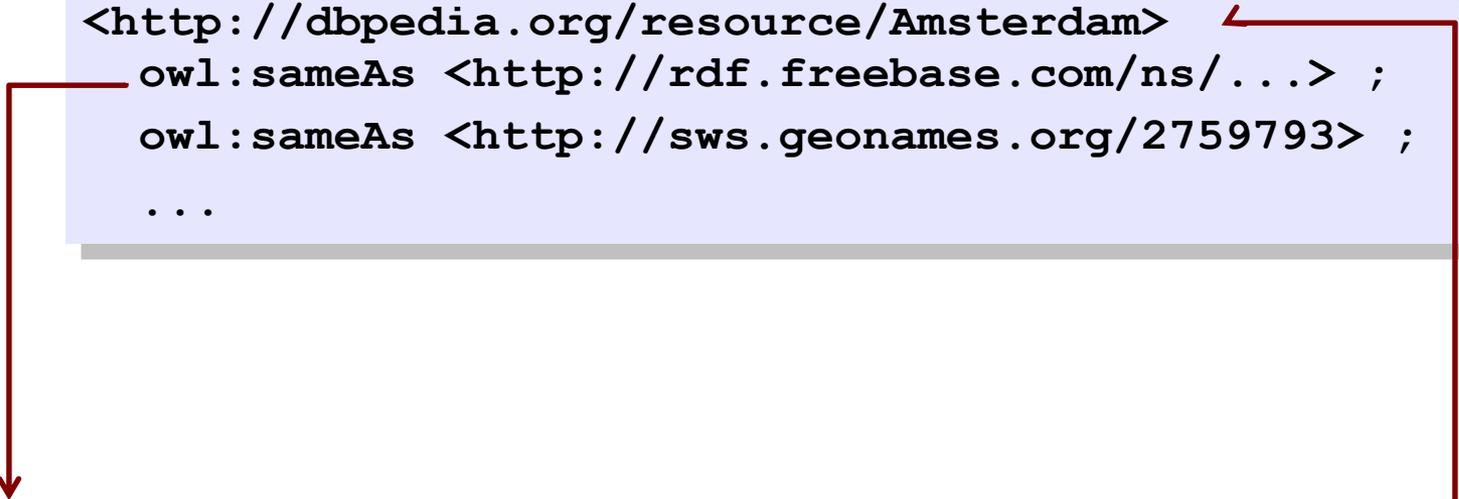
```
dbpedia:ABN_AMRO
```

```
dbterm:location dbpedia:Amsterdam ;
```

```
...
```

# Enlazado automático

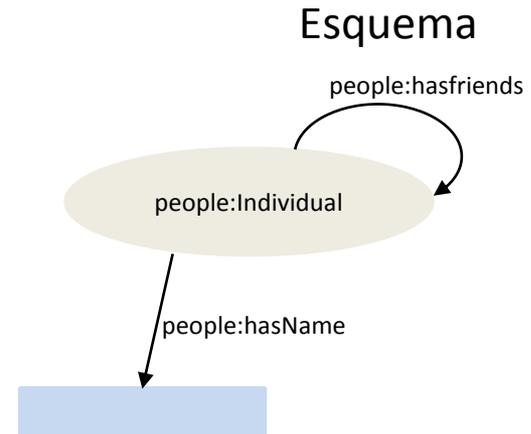
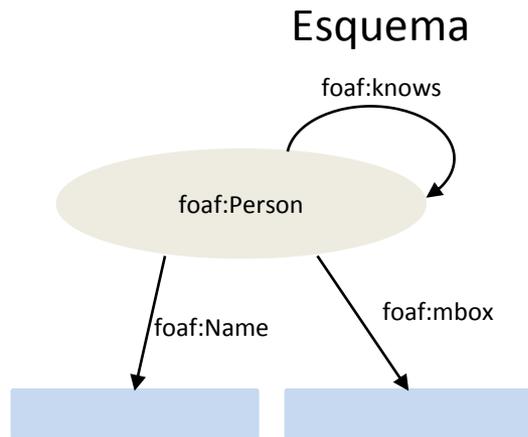
```
<http://dbpedia.org/resource/Amsterdam>
 owl:sameAs <http://rdf.freebase.com/ns/...> ;
 owl:sameAs <http://sws.geonames.org/2759793> ;
 ...
```



```
<http://sws.geonames.org/2759793>
 owl:sameAs <http://dbpedia.org/resource/Amsterdam>
 wgs84_pos:lat "52.3666667" ;
 wgs84_pos:long "4.8833333" ;
 geo:inCountry <http://www.geonames.org/countries/#NL> ;
 ...
```

**Computador puede saltar automáticamente desde una a otra...**

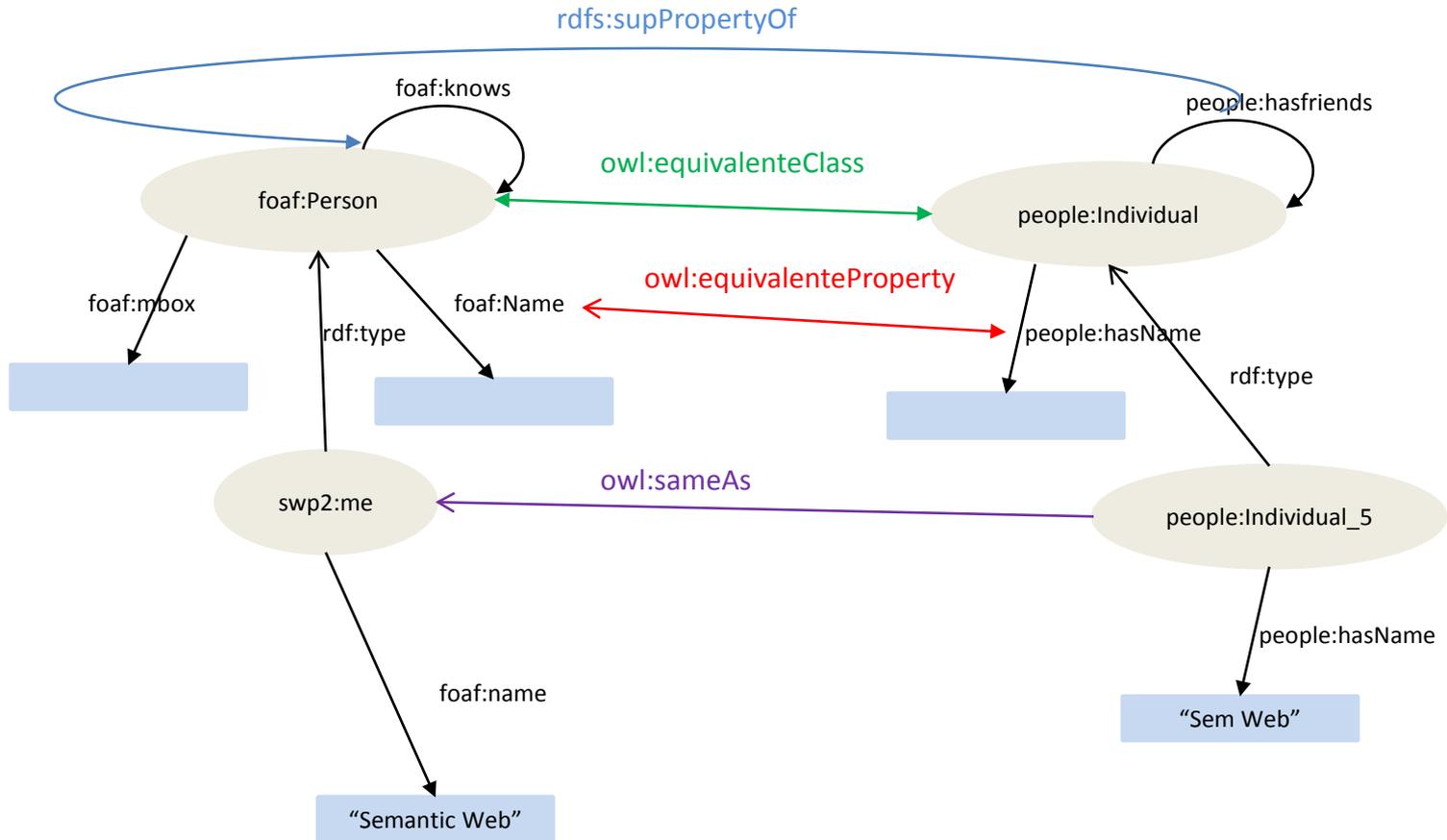
# Alineando



```
people:Individual owl:equivalentClass foaf:Person .
people:hasName owl:equivalentProperty foaf:Name .
people:hasfriends rdfs:subPropertyOf foaf:Knows .
```

# Alineando

```
people:Individual owl:equivalenteClass foaf:Person .
people:hasName owl:equivalentProperty foaf:Name .
people:hasfriends rdfs:subPropertyOf foaf:Knows .
```



## Servicio para la 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.

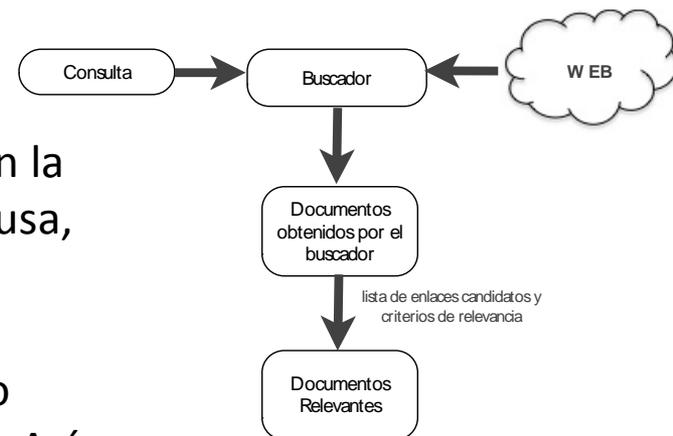
Para determinar esa similitud se usan distintos modelos en la literatura: el booleano, el vectorial, los basados en lógica difusa, redes neuronales, o redes bayesianas.

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

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



## Servicio para la 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.

## Servicio para la 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:

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

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

## Servicio para la 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

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

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

## Servicio para la 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:

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

## Servicio para la 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/ula_venezuela)  
<https://www.facebook.com/pages/FAcadem%20de%20Ingenier%C3%ADa-ULA-Venezuela-Sitio-Oficial/258084854230578>  
[http://www2.ula.ve/plandes/images/stories/inf\\_gestion\\_cap\\_1.pdf](http://www2.ula.ve/plandes/images/stories/inf_gestion_cap_1.pdf)  
<http://lama.adm.ula.ve/ptderecho/>  
[http://www2.ula.ve/plandes/images/stories/pdula\\_cap1.pdf](http://www2.ula.ve/plandes/images/stories/pdula_cap1.pdf)  
<http://www.venezuelaaola.com/2014/02/02/decanos-de-ula-merida-estudian-suspender-clases/>  
<http://www.slideshare.net/MANUELLIT0R>  
<http://www.noticias24.com/Venezuela/noticia/221873/suspenden-las-actividades-academicas-indefiniidamente-en-la-universidad-de-los-andes/>  
[http://ar.ask.com/web/?q=ula+de+merida+venezuela&asrc=998&i=sem&site=1332&serch=ul+&fr=1&ad=sem&an=google\\_s&my=bl&kwid=ula%20de%20merida%20venezuela&news&res=333030898&pla=4&mb=8&ou=sk&id=sk&op=011&w=40510943168&agid=5114244988&date=20131204](http://ar.ask.com/web/?q=ula+de+merida+venezuela&asrc=998&i=sem&site=1332&serch=ul+&fr=1&ad=sem&an=google_s&my=bl&kwid=ula%20de%20merida%20venezuela&news&res=333030898&pla=4&mb=8&ou=sk&id=sk&op=011&w=40510943168&agid=5114244988&date=20131204)  
<http://www.mof.gov.ve/Noticias/16134>  
<http://www.actualidadyante.com/noticias-de-merida-venezuela/32-academicas/2888-hoy-la-ula-cumple-229-anos-de-fundada>  
<http://www.universital.com/nacional-y-politica/149210/decanos-de-ula-merida-podran-suspender-clases-por-violencia>  
<http://eluniversitario.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://carlosramosnvas.com/2013/07/24/decanos-de-merida-universidad-de-los-andes-primera-universidad-republicana-de-latinamerica/>  
[http://www.scielo.org.ve/scielo.php?pid=S1516-4919\(2007\)000400019&script=sci\\_artext](http://www.scielo.org.ve/scielo.php?pid=S1516-4919(2007)000400019&script=sci_artext)  
<http://www.redhyc.org/pdf/735/70504795.pdf>  
<http://ppd.org.ve/portal/>  
<http://www.alnomeno360.com/portal/estudiantes-de-la-ula-merida-protestan-en-contra-de-la-ineguridad/>  
[http://www.linkedin.com/search/?orig=TSEO\\_SNA&firstName=Jordi&lastName=Veiga&Geo=V&trk=TSEO\\_SN](http://www.linkedin.com/search/?orig=TSEO_SNA&firstName=Jordi&lastName=Veiga&Geo=V&trk=TSEO_SN)  
<http://meridamunicipal.wordpress.com/articulos/agricultores-boyc-de-santa-rosa/>  
<http://vusunoticias.com/universidad-de-los-andes-suspende-actividades/>  
<http://www.minc.gov.ve/2012/11/entregan-financiamiento-a-investigadores-peil-de-merida/>  
<http://noticiasvenezuela.info/2013/12/01/una-gran-veba-maestria-en-ciencias-de-la-actividad-fisica-y-los-deportes-para-la-ula-merida/>  
<http://www.sagrados.com/31054-venezuela-rector-sila-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?id=QG0812>  
[http://marbelsuarezmancha.blogspot.com/2014\\_02\\_09\\_archive.html](http://marbelsuarezmancha.blogspot.com/2014_02_09_archive.html)  
<http://temendenzazo.com/ve/mediante-encuesta-evaluaran-la-situacion-para-un-possible-reinicio-de-actividades-en-la-ula/>

## Servicio para la 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 envió 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

## Servicio para la 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,

## Servicio para la 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

## Servicio para la 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

- Hacer un primer filtrado, usando los criterios definidos por el usuario sobre las características básicas que debe tener un documento (tipo, lenguaje, etc.)
- 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
<a href="http://es.wikipedia.org/wiki/Universidad_de_Los_Andes_(Venezuela)">http://es.wikipedia.org/wiki/Universidad de Los Andes (Venezuela)</a>	34	55	134	109	8
<a href="http://llama.adm.ula.ve/pderecho/">http://llama.adm.ula.ve/pderecho/</a>	2	5	2	83	2
<a href="http://www.venezuelaaldia.com/2014/02/decanos-de-ula-merida-estudian-suspender-clases/">http://www.venezuelaaldia.com/2014/02/decanos-de-ula-merida-estudian-suspender-clases/</a>	1	20	18	22	2
<a href="http://www.venezuelaaldia.com/tag/ula/">www.venezuelaaldia.com/tag/ula/</a>	31	40	262	162	2
<a href="http://www.noticias24.com/venezuela/noticia/221873/suspenden-las-actividades-academicas-indefinidamente-en-la-universidad-de-los-andes/">http://www.noticias24.com/venezuela/noticia/221873/suspenden-las-actividades-academicas-indefinidamente-en-la-universidad-de-los-andes/</a>	7	9	5	4	7

## Servicio para la Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

- Se procede a calcular la frecuencia inversa de cada documento para cada uno de los términos (la matriz DF-IDF).
- 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
<a href="http://es.wikipedia.org/wiki/Universidad_de_Los_Andes_(Venezuela)">http://es.wikipedia.org/wiki/Universidad_de_Los_Andes_(Venezuela)</a>	5,6817395184	14,964374999	25,814440444	15,480071531	2,3964822549
<a href="http://lama.adm.ula.ve/pderecho/">http://lama.adm.ula.ve/pderecho/</a>	0,3342199717	1,3603977272	0,3852901559	11,787577404	0,5991205637
<a href="http://www.venezuelaaldia.com/2014/02/decanos-de-ula-merida-estudian-suspender-clases/">http://www.venezuelaaldia.com/2014/02/decanos-de-ula-merida-estudian-suspender-clases/</a>	0,1671099858	5,4415909087	3,467611403	3,1244181072	0,5991205637
<a href="http://www.venezuelaaldia.com/tag/ula/">www.venezuelaaldia.com/tag/ula/</a>	5,1804095609	10,883181817	50,473010421	23,007078789	0,5991205637
<a href="http://www.noticias24.com/venezuela/noticia/221873/suspenden-las-actividades-academicas-indefinidamente-en-la-universidad-de-los-andes/">http://www.noticias24.com/venezuela/noticia/221873/suspenden-las-actividades-academicas-indefinidamente-en-la-universidad-de-los-andes/</a>	1,1697699008	2,4487159089	0,9632253897	0,5680760195	2,096921973

### 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

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

## Servicio para la Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

<b>Filtro</b>	<b>Yahoo</b>	<b>Google</b>
<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

## Servicio para la 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

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

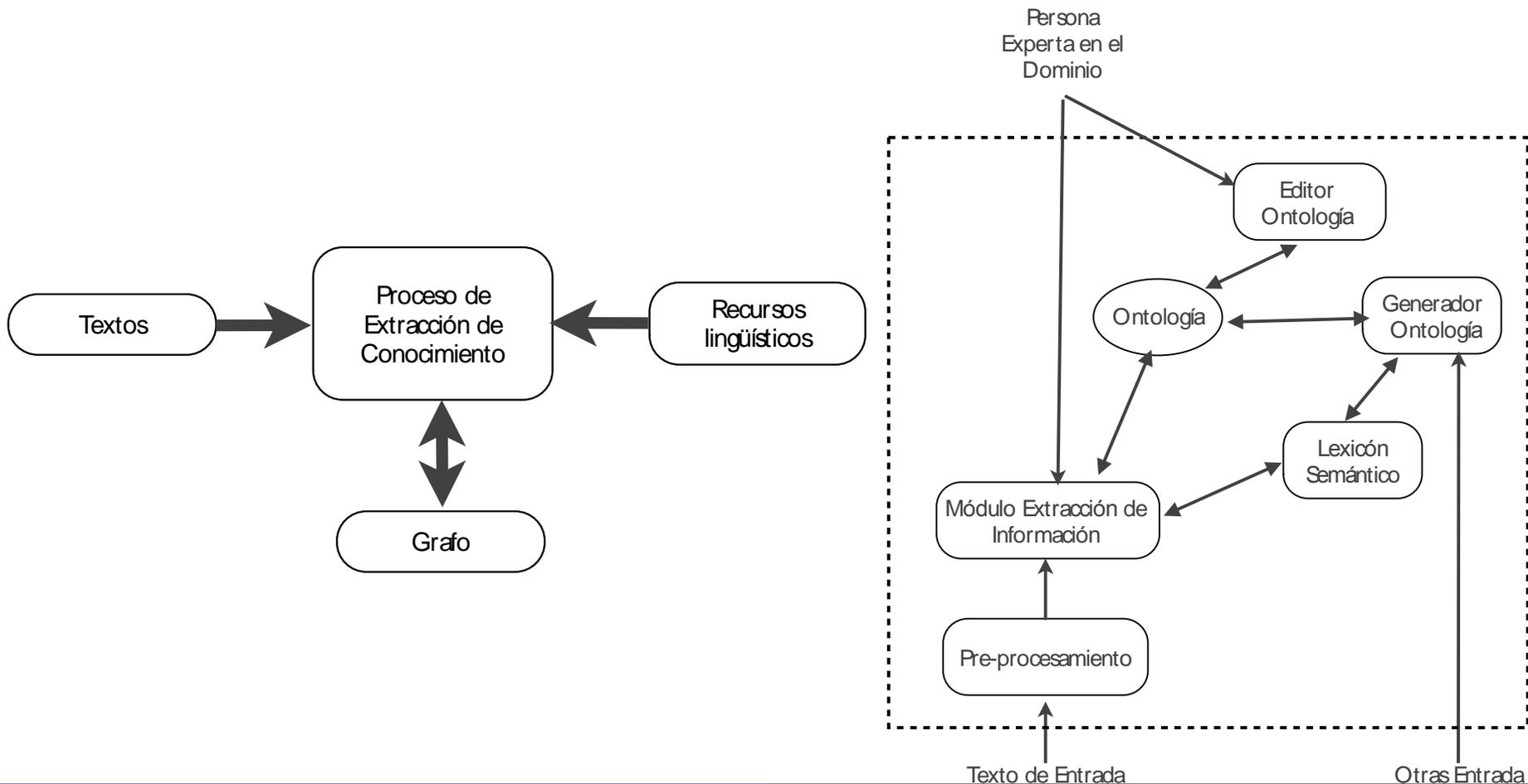
Filtro	Yahoo	Google
<i>Precisión</i>	0,2716	0,3699

## Servicio para la Extracción de Documentos Relevantes desde Documentos Recuperados en la Web

valor de precisión para diferentes consultas sobre yahoo y google

<b>Consultas Aleatorias</b>	<b>Yahoo</b>	<b>Google</b>
<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>

## Servicio de Extracción de conocimiento desde documentos no estructurados



## Servicio de Extracción de conocimiento desde documentos no estructurados

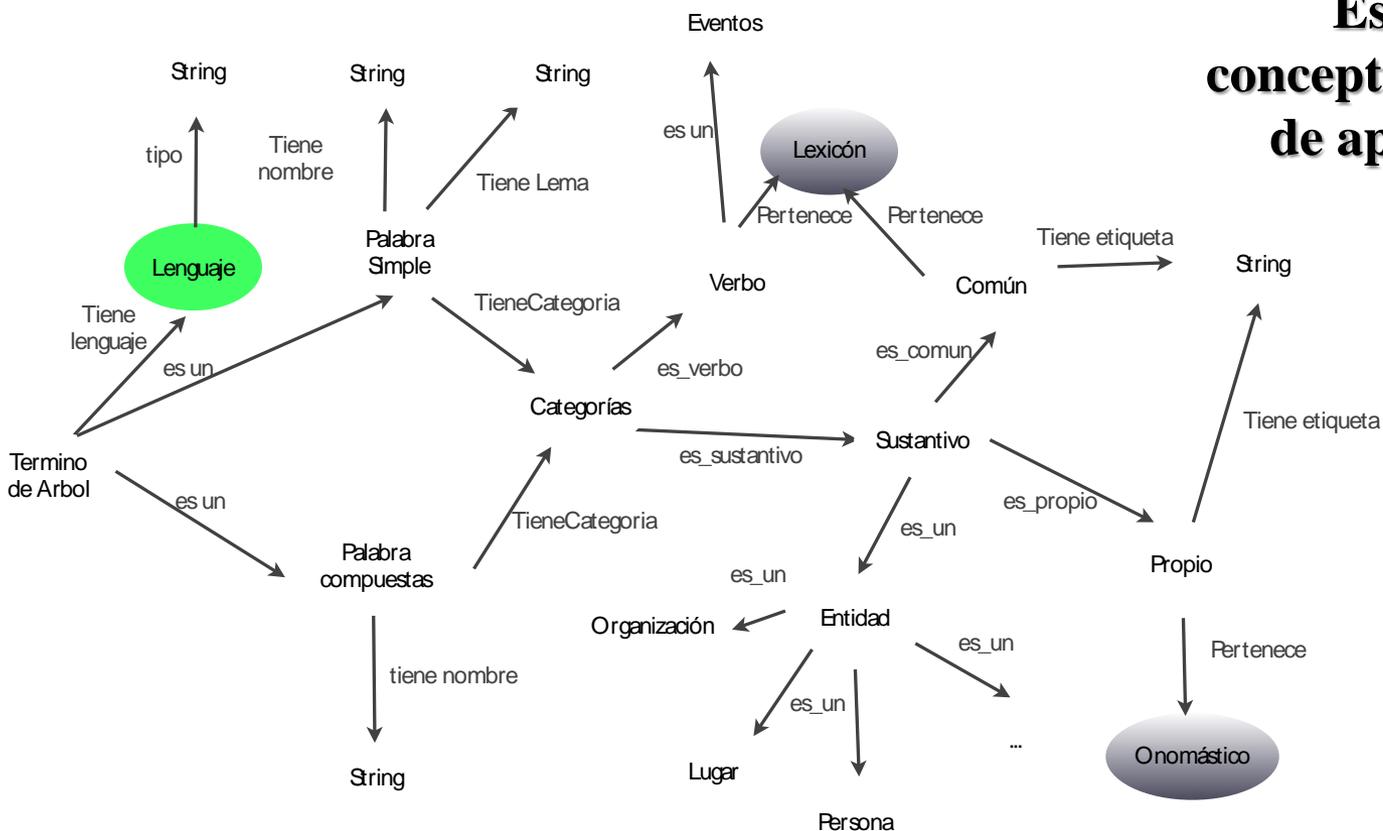
### *Lematizador*

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.

## Servicio de Extracción de conocimiento desde documentos no estructurados



## Esquema conceptual del grafo de aprendizaje

## Servicio de Extracción de conocimiento desde documentos no estructurados

### Grafo de aprendizaje

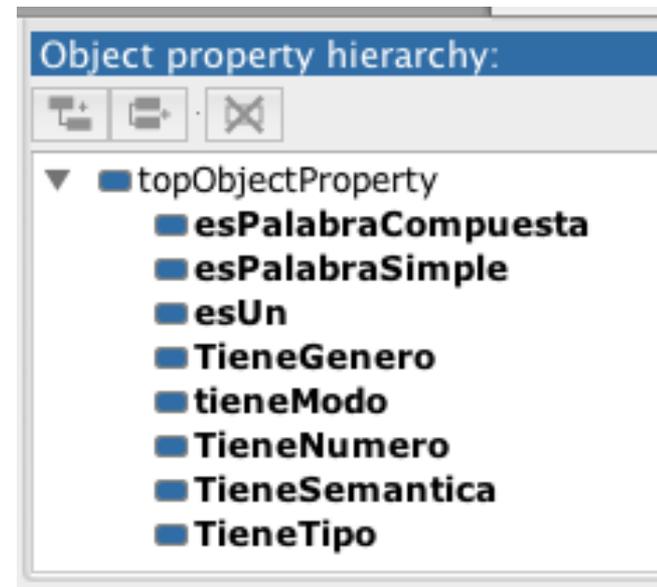
El grafo de aprendizaje esta compuesto por un conjuntos de axiomas básicos para inferir nuevo conocimiento.

Tiene un conjunto de reglas y se utiliza el motor de inferencia para generar conocimiento.

## Servicio de Extracción de conocimiento desde documentos no estructurados

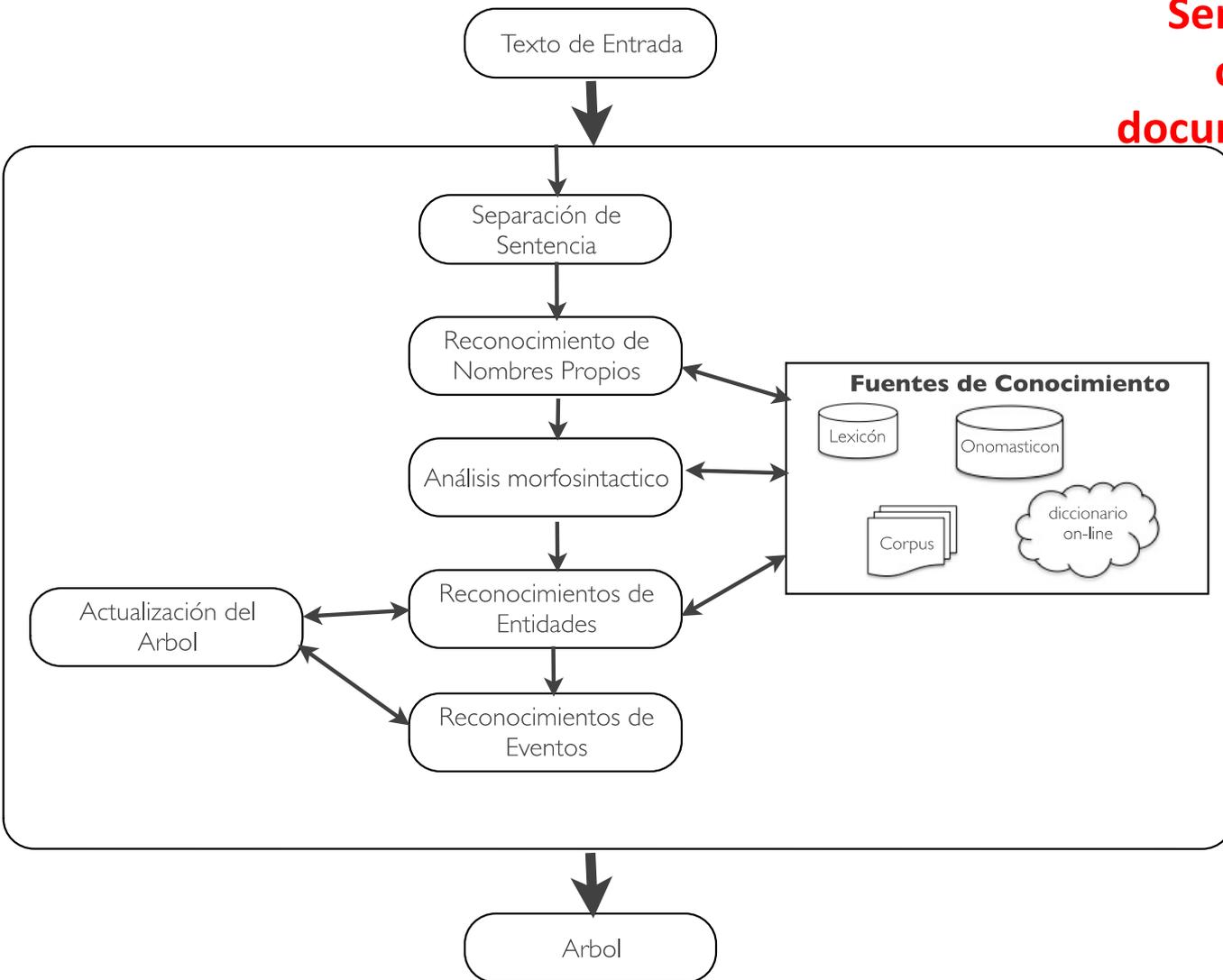


Las clases definidas en el árbol de aprendizaje



Las relaciones definidas en el grafo de aprendizaje

## Servicio de Extracción de conocimiento desde documentos no estructurados



## Servicio de Extracción de conocimiento desde documentos no estructurados

### *Algunas posibles aplicaciones del servicio*

- Creación de 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

# Servicio de Extracción de conocimiento desde documentos no estructurados (ejemplo 1)

## 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 Ince Anzures

## Entidades y Relaciones Candidatas

## Servicio de Extracción de conocimiento desde documentos no estructurados (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  $\geq 10,06$

## Servicio de Extracción de conocimiento desde documentos no estructurados (ejemplo 2)

**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.

## Servicio de Extracción de conocimiento desde documentos no estructurados (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



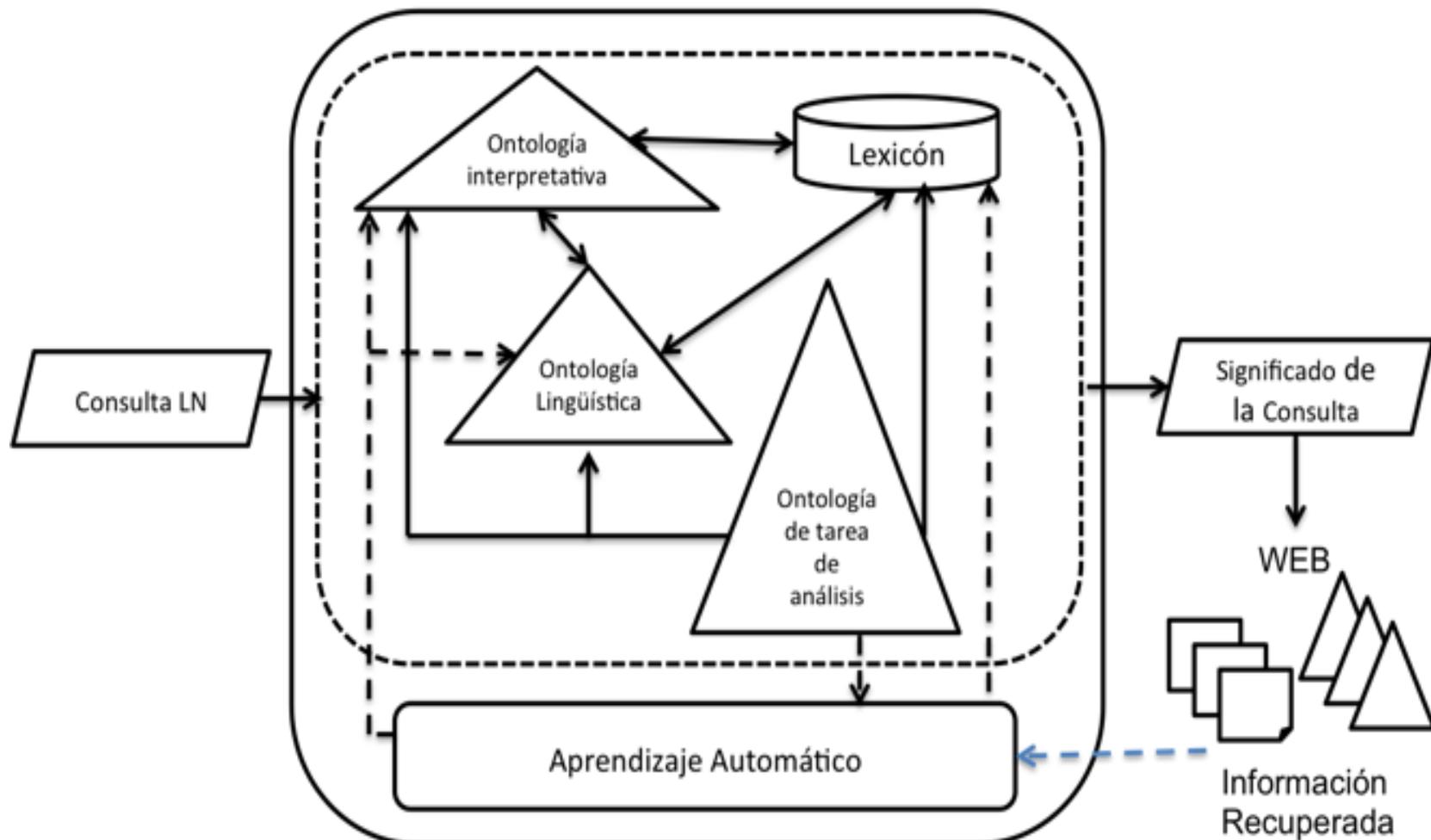
Description: LexiconElectronico	
Members +	
◆ ADMISIÓN	?
◆ APROBADO	?
◆ ARTÍCULO	?
◆ ASPIRANTE	?
◆ AÑOS	?
◆ CASO	?
◆ CO-TUTOR	?
◆ COMISIÓN	?
◆ CONOCIMIENTO	?
◆ CRITERIOS	?
◆ CRÉDITOS	?
◆ CUMPLIR	?
◆ DEBE	?
◆ DEBEN	?
◆ DEBERÁ	?
◆ DEBERÁN	?
◆ DESARROLLAR	?
◆ DOCTOR	?
◆ DOCTORADO	?
◆ ELABORAR	?
◆ ES	?
◆ ESTAR	?
◆ ESTARÁ	?
◆ ESTÁN	?
◆ FORMAR	?
◆ GRADO	?
◆ HA	?
◆ HABER	?

Description: LexiconElectronico	
◆ INVESTIGACIÓN	?
◆ JURADO	?
◆ LAPSO	?
◆ MIEMBROS	?
◆ NIVEL	?
◆ NOMBRAR	?
◆ PARTE	?
◆ PAÍS	?
◆ PODRÁ	?
◆ PODRÁN	?
◆ POSEER	?
◆ PRESENTAR	?
◆ PROFESORES	?
◆ PROGRAMA	?
◆ REGLAMENTO	?
◆ SEA	?
◆ SER	?
◆ SERÁ	?
◆ SERÁN	?
◆ SIGUIENDO	?
◆ TENDRÁ	?
◆ TENDRÁN	?
◆ TESIS	?
◆ TIPO	?
◆ TUTOR	?
◆ ÁREA	?

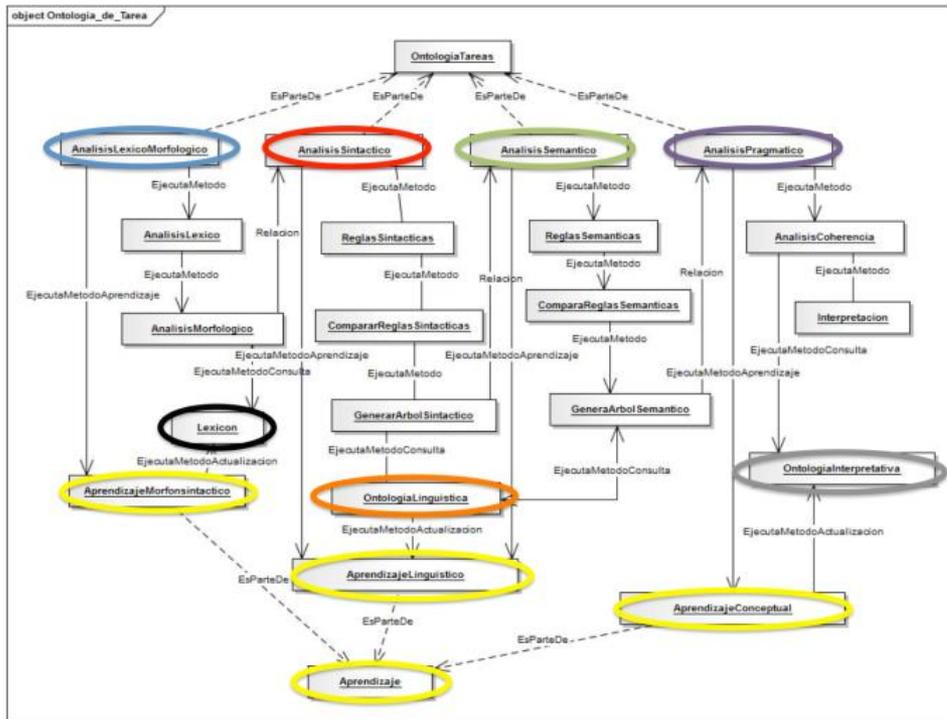
## Servicio de Extracción de conocimiento desde documentos no estructurados (ejemplo 2)

Lexicón electrónico de sustantivos y verbos

## Marco Ontológico Semántico para PLN



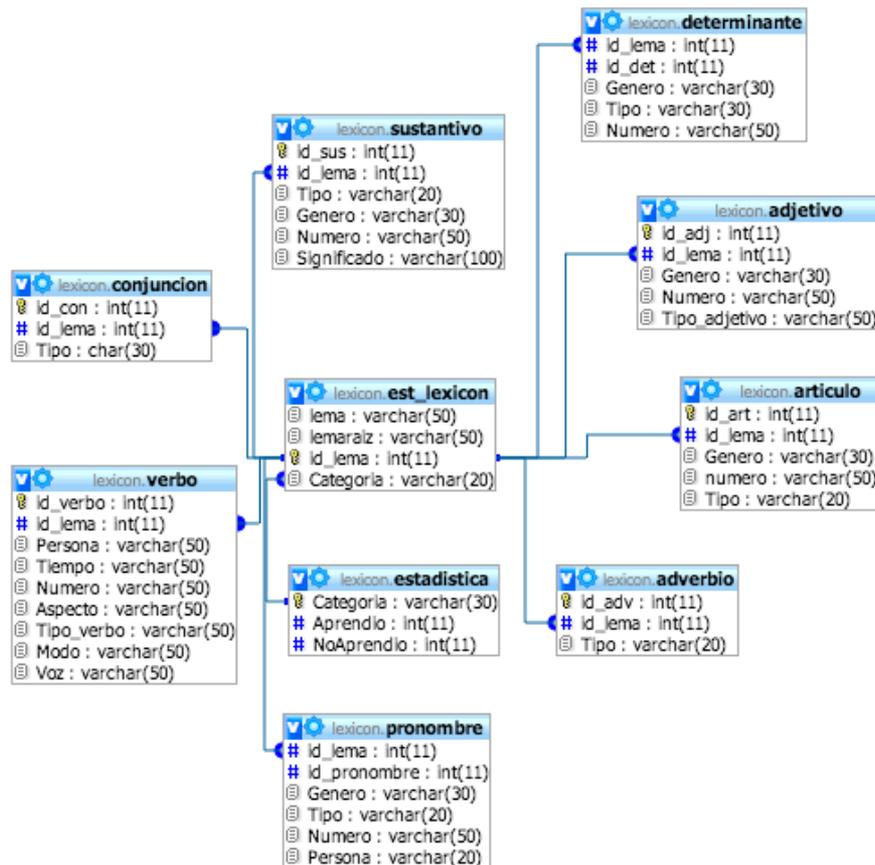
## Marco Ontológico Semántico para PLN



## Ontología de Tareas

*lex\_mor*(componente léxico, categoría, género, número, modo, tiempo, aspecto, voz, persona)

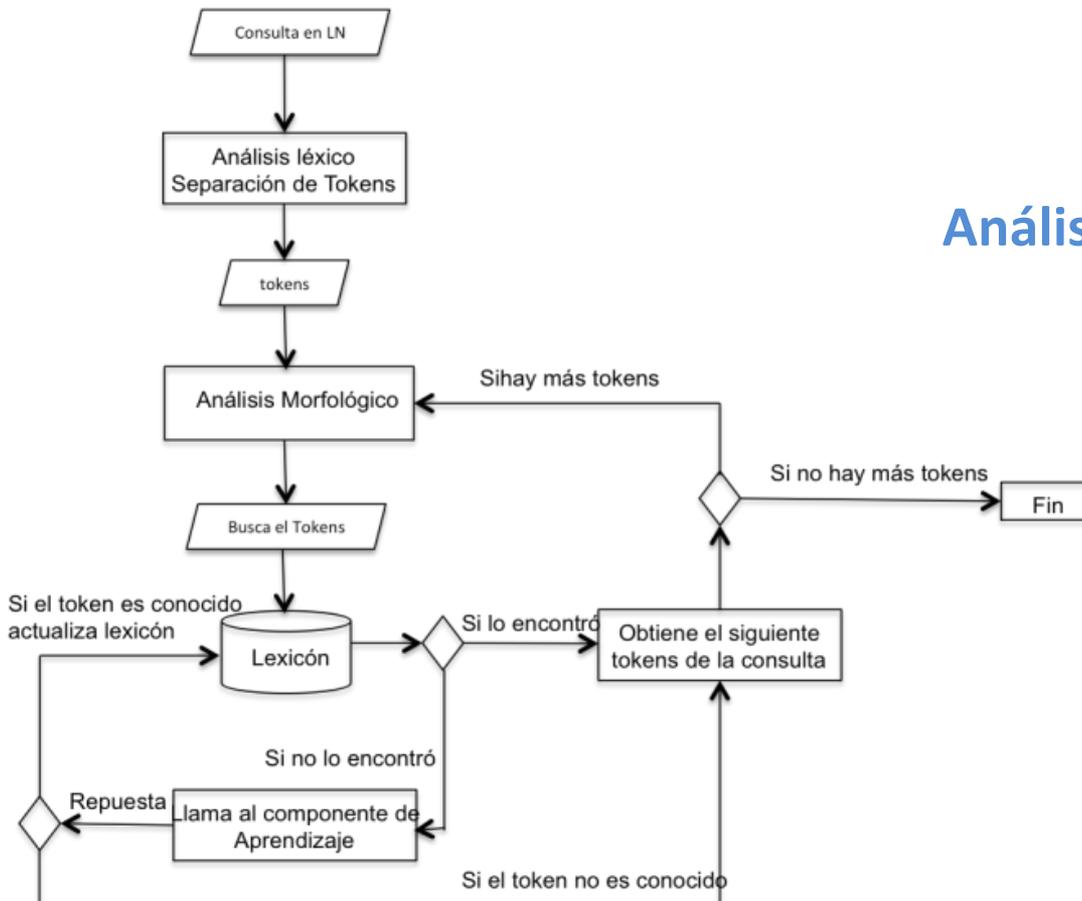
## Marco Ontológico Semántico para PLN



Base de datos  
del lexicón del MODS

## Marco Ontológico Semántico para PLN

### Análisis léxico-morfológico



# Marco Ontológico Semántico para PLN

## Ontología Interpretativa

Conceptos:

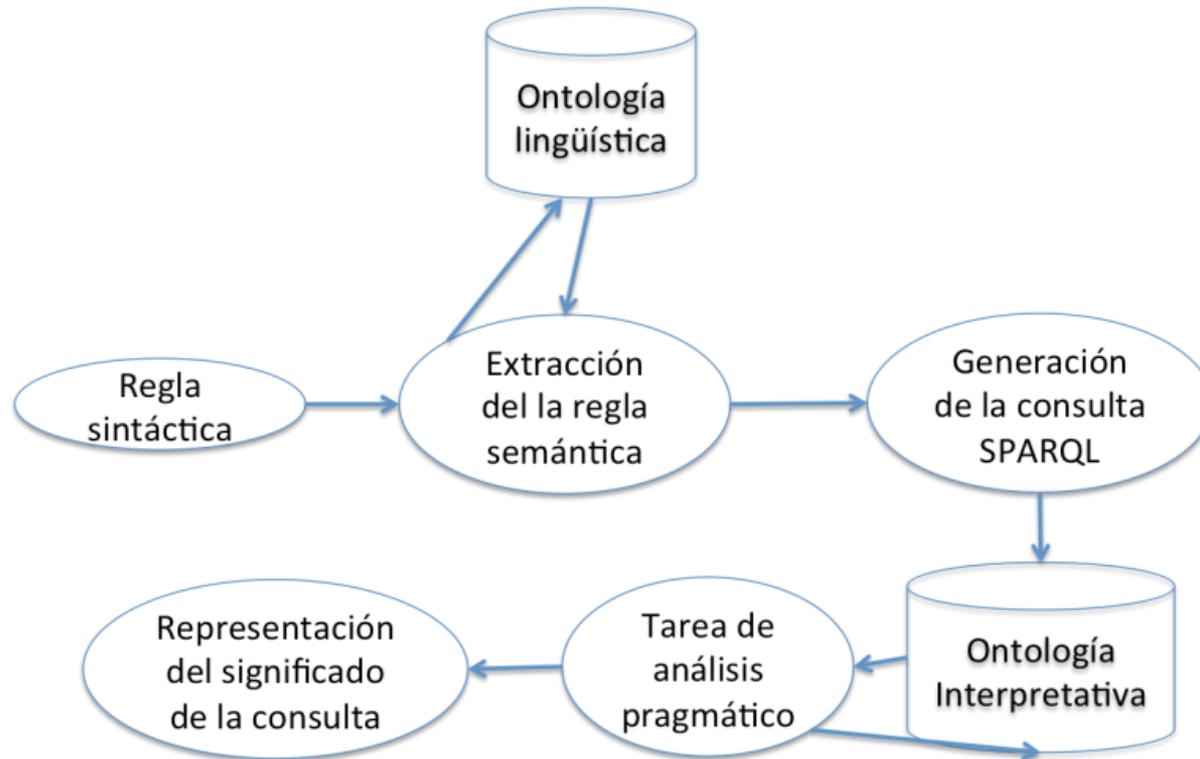
- **Entidades:** representan objetos físicos como abstractos (normalmente son los sustantivos, adjetivos y adverbios)
- **Eventos:** representan una acción (normalmente son verbos)
- **Relaciones:** indican las diferentes relaciones que puede existir entre los conceptos definidos previamente, o propiedades que pueden tener (normalmente son sinónimos, antónimo, parte de).

El concepto *Entidades* en el MODS puede ser:

- **Abstractos:** Pueden ser definiciones, teorías, etc.
- **Concretos:** Son objetos físicos, o que se pueden definir en algo específico.



## Marco Ontológico Semántico para PLN



### Uso de la Ontología Interpretativa

Marco Ontológico Dinámico Semántico  
MODS

**cemisid**  
centro de estudios en  
microelectrónica y sistemas distribuidos

Inicio Componentes del MODS Acerca de

Introduzca el texto a consultar

"tipos de mantenimientos" and (planta o industria) and (producción o product

Aproximadamente 81 resultados (0,11 segundos) Sort by: Relevance

con la tecnología de Google™ Búsqueda personalizada de Google

[Servicios - logo-maquinas](#)  
[www.mytsac.net/servicios.html](http://www.mytsac.net/servicios.html)  
En Máquinas y Tecnología SAC se examinan diseños, **productos**, ... de **producción**, disponibilidad de equipos, y la seguridad en la **planta industrial** en cada uno de los **tipos de mantenimientos** como son Preventivos, correctivos y predictivos.

[Industrial maintenance - Aula Fácil](#)  
[www.aulafacil.com/cursosenviados/Mantenimiento-Industrial.doc](http://www.aulafacil.com/cursosenviados/Mantenimiento-Industrial.doc)  
Formato de archivo: Microsoft Word  
**TIPOS DE MANTENIMIENTOS** .... Actualmente el mantenimiento busca aumentar y confiabilizar la **producción**; aparece el .... presente la construcción, diseño y modificaciones de la **planta industrial** como también debe tener a .... y productividad a la **industria**, los resultados se evalúan en cantidad y calidad de **producto**.

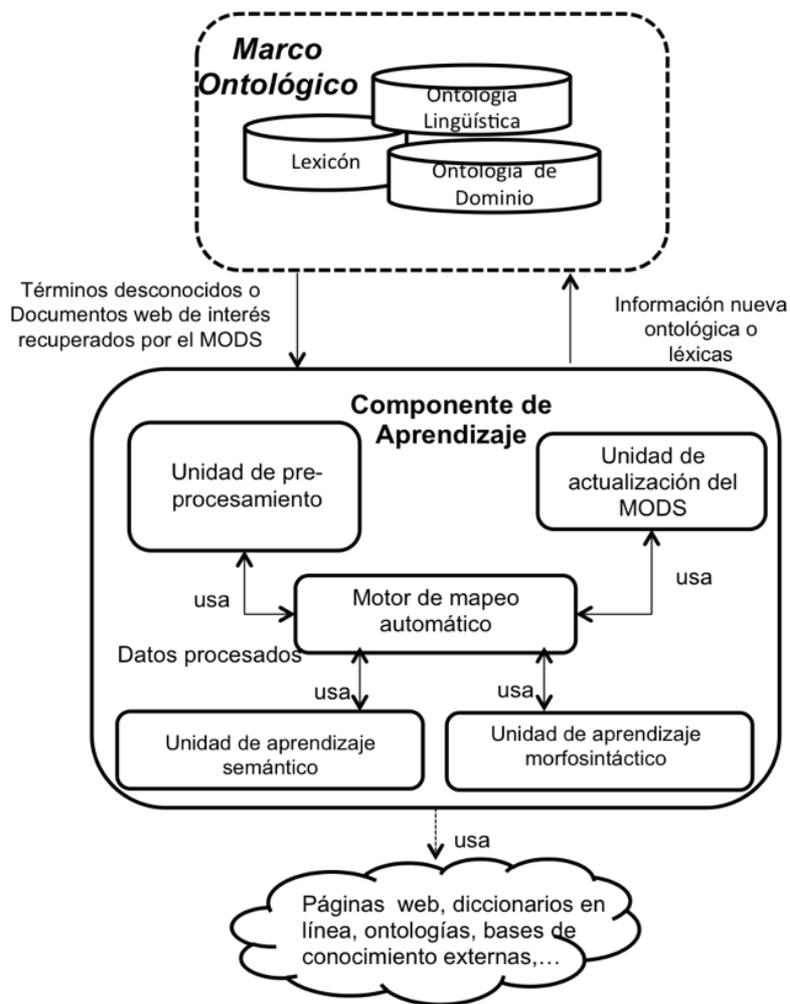
[Planificación de Paradas de Planta - sim ingeniería](#)  
[www.simingeneria.com.ar/.../Planificacion%20de%20paradas%20de%20planta\\_noviembre.htm](http://www.simingeneria.com.ar/.../Planificacion%20de%20paradas%20de%20planta_noviembre.htm)  
(El Mantenimiento en **Plantas** de Proceso). BUENOS AIRES ... Supervisores de Mantenimiento y **Produccion** u Operaciones. Responsables ... **Tipos de Mantenimientos** y características. 5. Distribucion ... Experiencia de 8 años en **Industria** Alimenticia como responsable de la elaboracion y control de calidad del **producto**.

## Marco Ontológico Semántico para PLN

### CONSULTA REALIZADA EN BUSCADORES USANDO MODS (P) Y SIN USARLO (SP)

CONSULTA	GOOGLE		YAHOO		ALTAVISTA	
Genéricas	SP	P	SP	P	SP	P
	1030	81	2050000	66700	181000	237
Especializada	SP	P	SP	P	SP	P
	70300	58	45600	32600	24300	70

## APRENDIZAJE ONTOLÓGICO

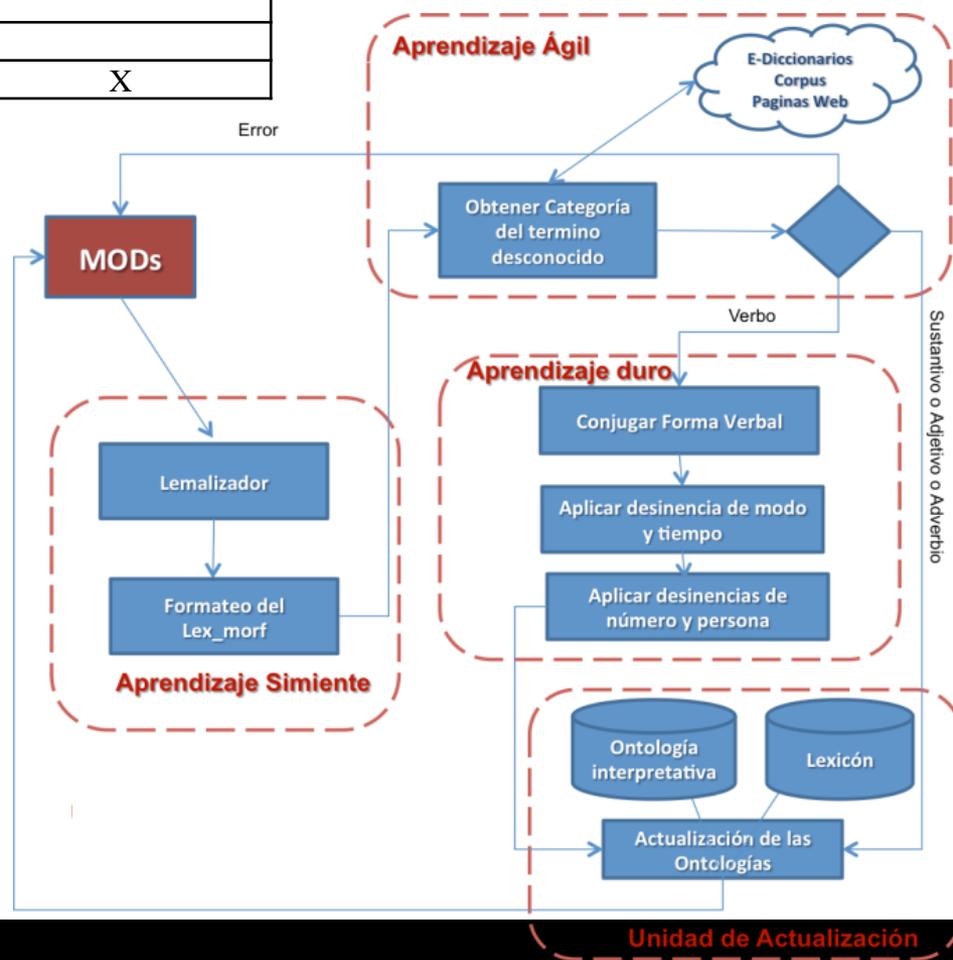


# Aprendizaje morfosintáctico

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



The screenshot shows the 'Marco Ontológico Dinámico Semántico' (MODS) interface. The header includes the logo for 'cemisid' (centro de estudios en microelectrónica y sistemas distribuidos). The main content area is titled 'Módulo Análisis Morfológico' and shows the results of a morphological analysis for the word 'cama'. The results include the function call `lex_mor(cama, Desconocido, NULL, NULL, NULL, NULL, NULL, NULL, NULL, NULL)`, the word found, the component found in the ontology, and the updated lexicon entry for 'cama'.

**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', '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,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,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))							
Lex_morf							
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

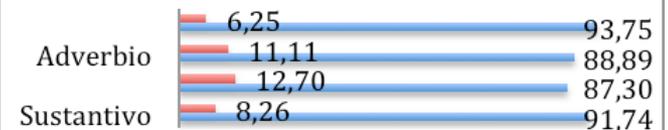
Término	Aprendió	No Aprendió	Total
Sustantivo	500	45	545
Adjetivo	220	32	252
Adverbio	80	10	90
Verbo	75	5	80

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

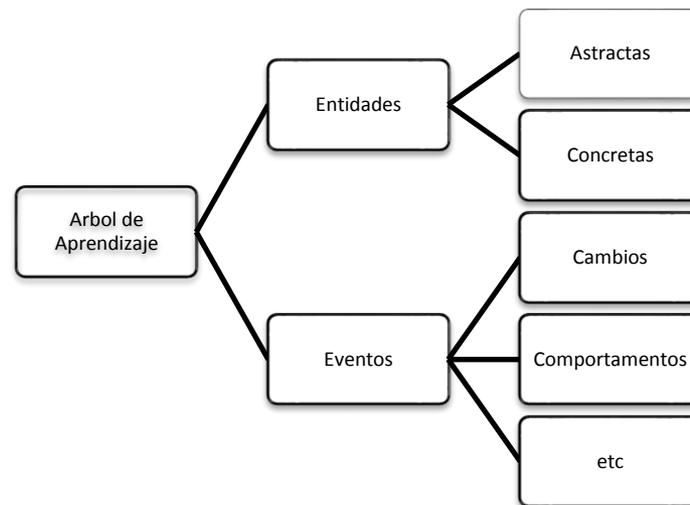
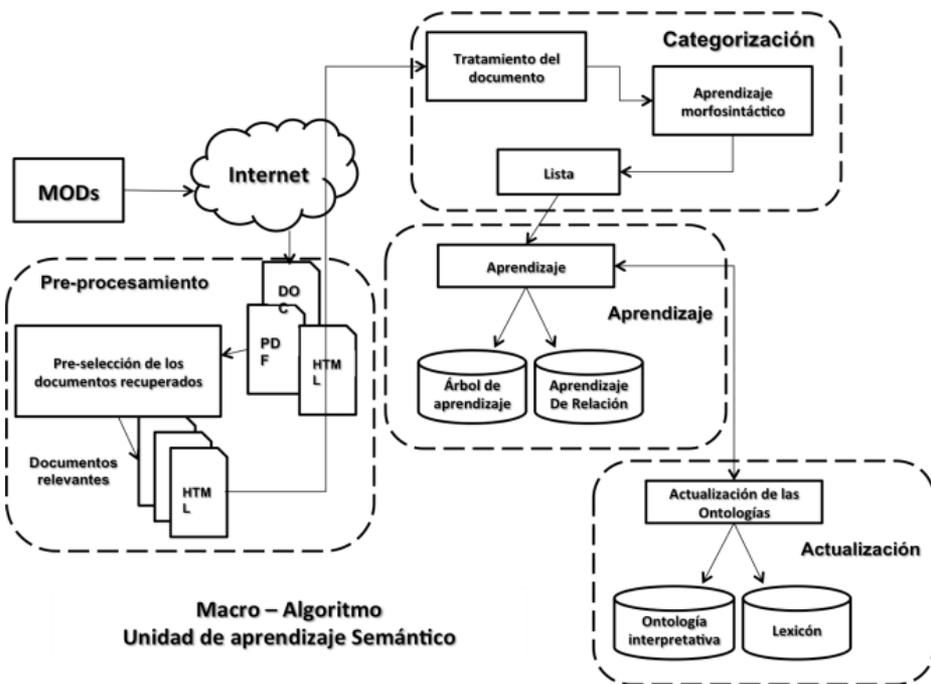
Aprendizaje morfo-sintáctico

### Aprendizaje morfosintáctico

■ No Aprendio ■ Aprendio



## 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),**

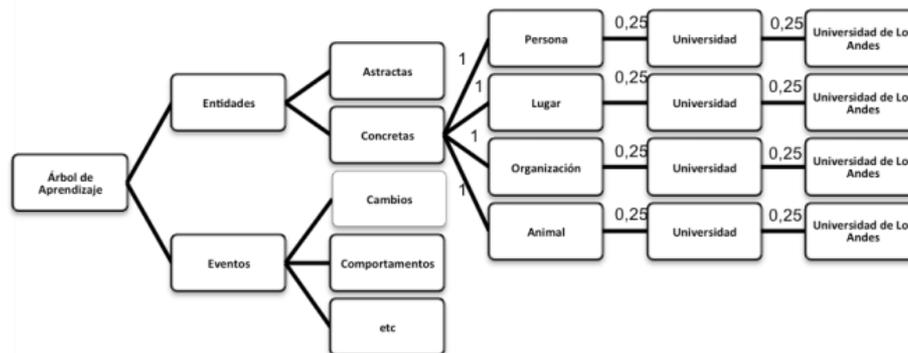
[http://es.wikipedia.org/wiki/Universidad\\_de\\_Los\\_Andes\\_\(Venezuela\)](http://es.wikipedia.org/wiki/Universidad_de_Los_Andes_(Venezuela))  
<https://www.facebook.com/ula.venezuela>  
<https://www.facebook.com/pages/Facultad-de-Ingenier%C3%ADa-ULA-Venezuela-Sitio-Oficial/258084854230578>  
[http://www2.ula.ve/plandes/images/stories/inf\\_gestion\\_cap\\_ii.pdf](http://www2.ula.ve/plandes/images/stories/inf_gestion_cap_ii.pdf)  
<http://lama.adm.ula.ve/pderecho/>  
[http://www2.ula.ve/plandes/images/stories/pdula\\_capii.pdf](http://www2.ula.ve/plandes/images/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=s&cre=33930269808&pla=&mob=&sou=s&aid=&adp=211&kwid=60510843168&agid=9114244968&date=20131204](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=s&cre=33930269808&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-aos-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://canal5noticia.com.ve/index.php/noticias-venezuela/item/36049-decanos-de-ula-merida-podran-suspender-clases-por-violencia>  
<http://carlosramosrivias.com/2013/07/24/tesoros-de-merida-universidad-de-los-andes-primer-universidad-republicana-de-latinamerica/>  
[http://www.scieio.org.ve/scieio.php?pid=S1316-49102007000400019&script=sci\\_artext](http://www.scieio.org.ve/scieio.php?pid=S1316-49102007000400019&script=sci_artext)  
<http://www.redalyc.org/pdf/705/70504702.pdf>  
<http://ppd.org.ve/porta/>  
<http://www.almomento360.com/porta/estudiantes-de-la-ula-merida-protestan-en-contra-de-la-inseguridad/>  
[http://www.linkedin.com/vsearch?orig=TSEO\\_SN&firstName=Judith&lastName=Vega&I\\_G=ve%3A0&rk=TSEO\\_SN](http://www.linkedin.com/vsearch?orig=TSEO_SN&firstName=Judith&lastName=Vega&I_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-pel-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=OQ0IB12>  
[http://marbelsuarezmancha.blogspot.com/2014\\_02\\_09\\_archive.html](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

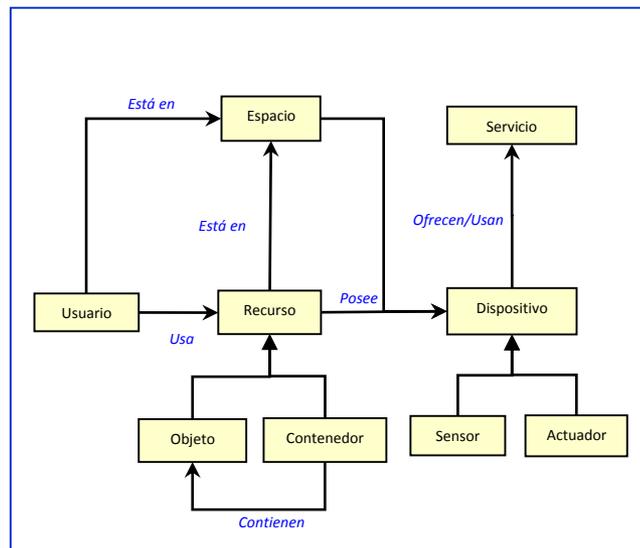
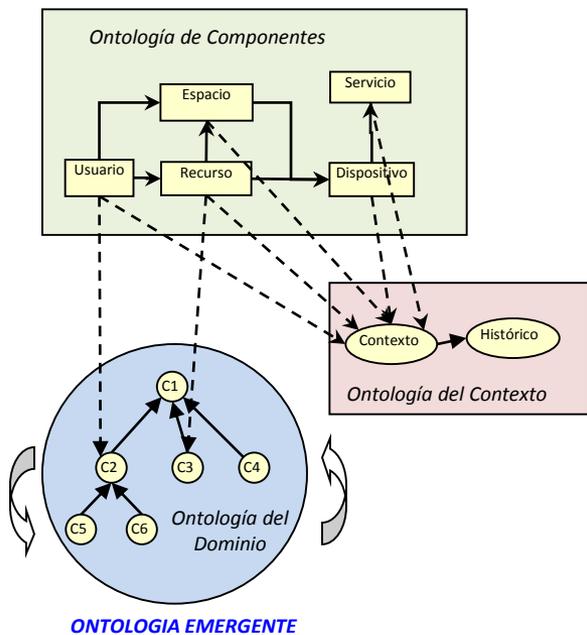
Término	Frecuencia	Categoría
Universidad	35	Sustantivo
Organización	13	Sustantivo
Dicta	7	Verbo
Mérida	4	Sustantivo
...	...	...

Ejemplo de la tabla de frecuencias



Ejemplo del árbol de aprendizaje

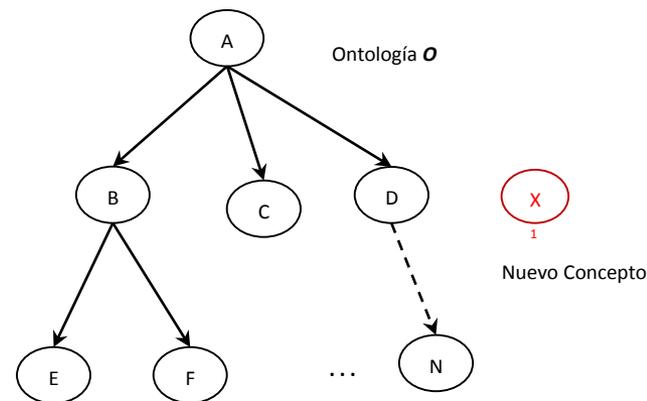
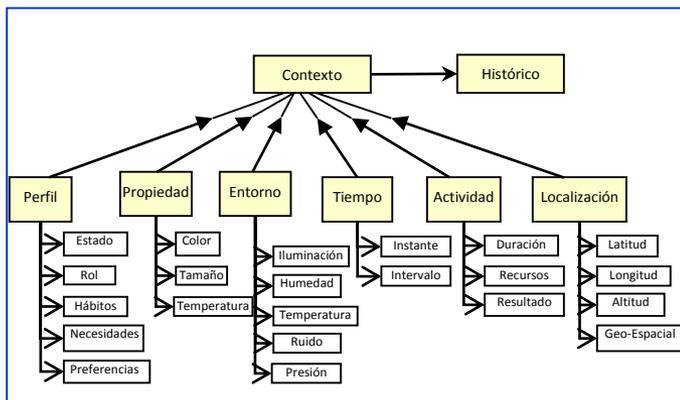
## Ontología Emergente



Ontología de Componentes de un AmI

Ontologías en un AmI

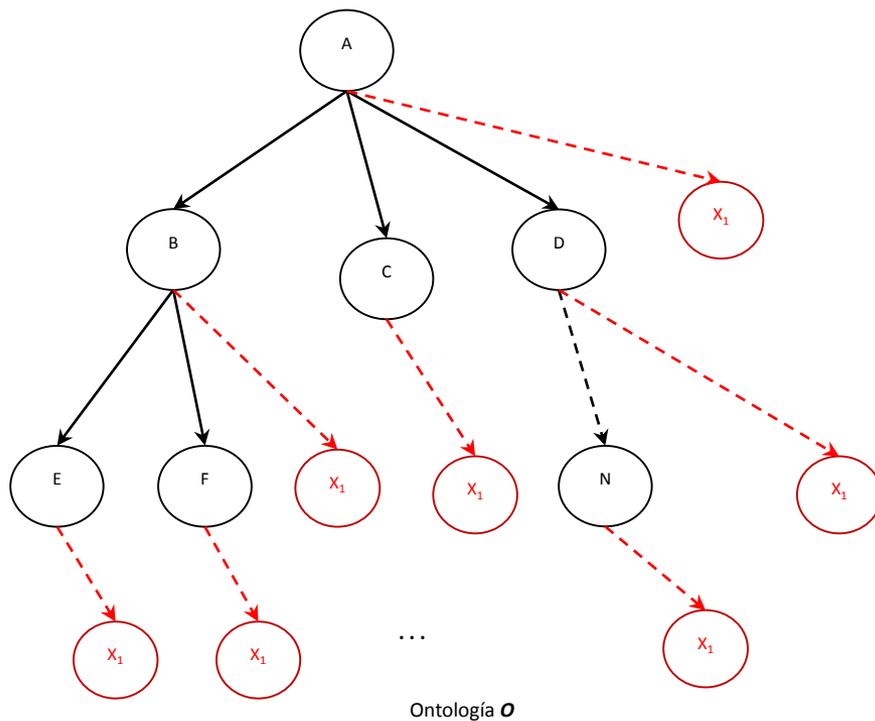
## Ontología Emergente



. 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
<b>Refresco</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>

## 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 subclase de  $C$  (que es hijo de  $R$ ), en función de la semejanza entre  $X$  y  $C$ .

**Probabilidad de transición:**

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

$\eta_{rs}$  : Cantidad de feromona.

$\gamma_{rs}$  : Información Heurística.

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

$\alpha$  y  $\beta$ : Importancia de la información memorística (feromona) y heurística.

**Actualización de la feromona**

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

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

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

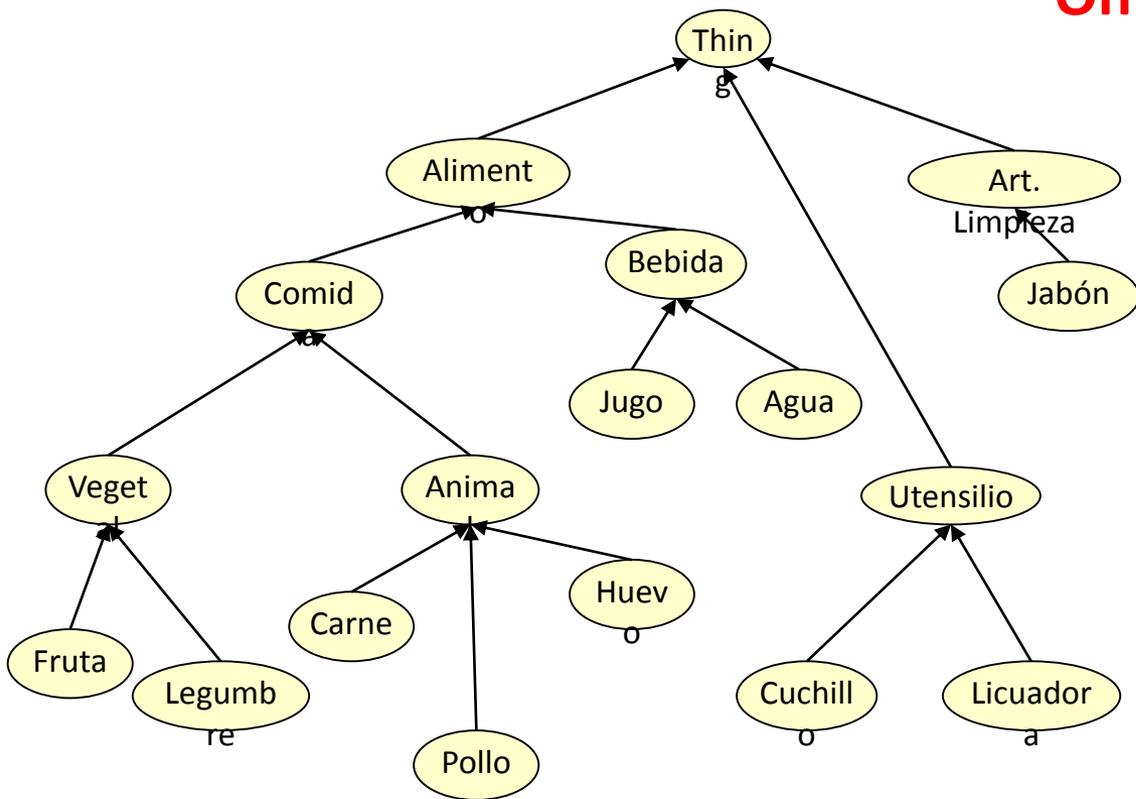
$$\Delta \gamma_{rs}^k = f(C(S_k))$$

$S_k$  : 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

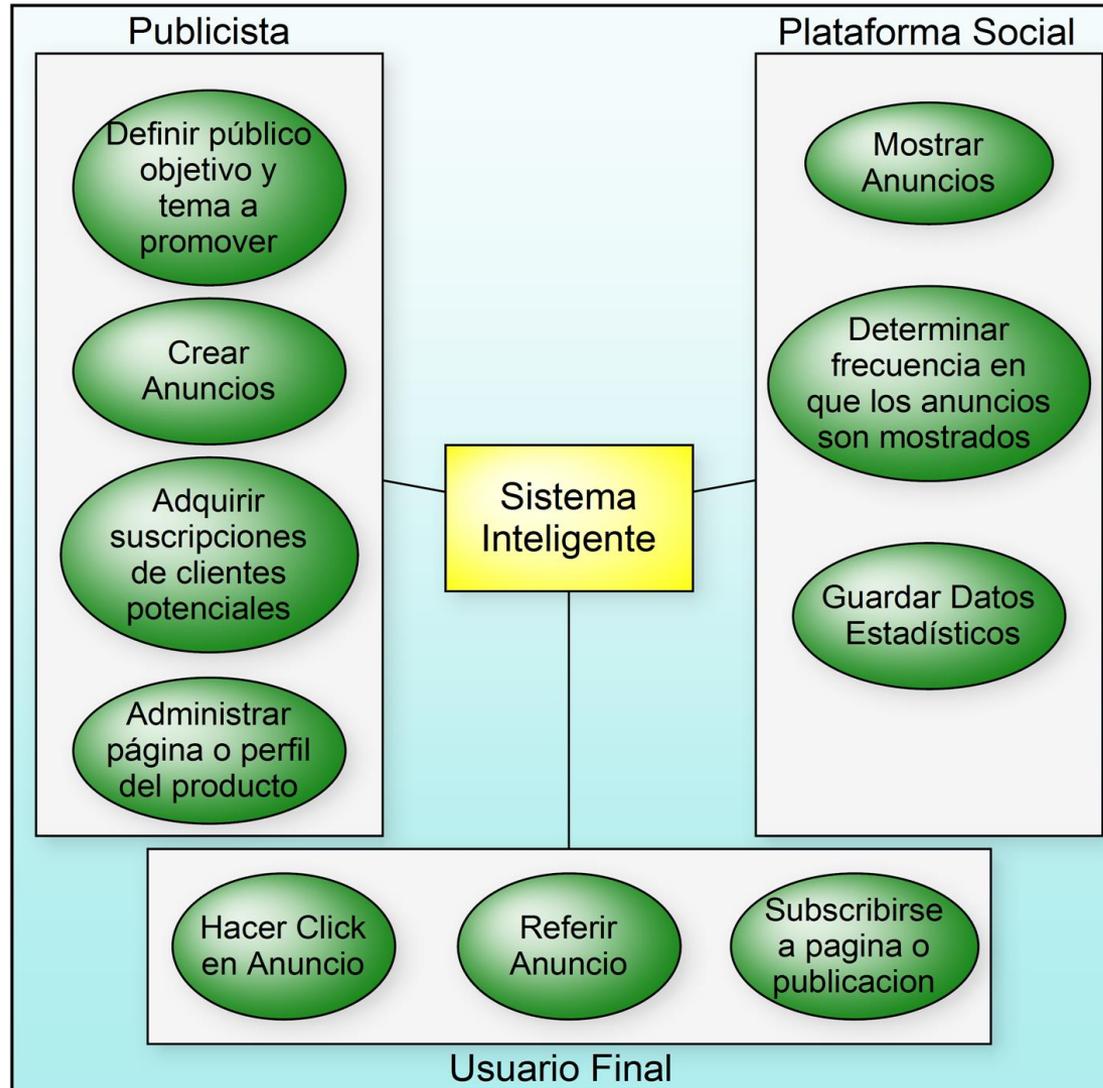
- Directa
  - Basada en la red de amigos
  - Colocada en páginas de redes sociales
- Indirecta

## Componentes de la Publicidad en Redes Sociales

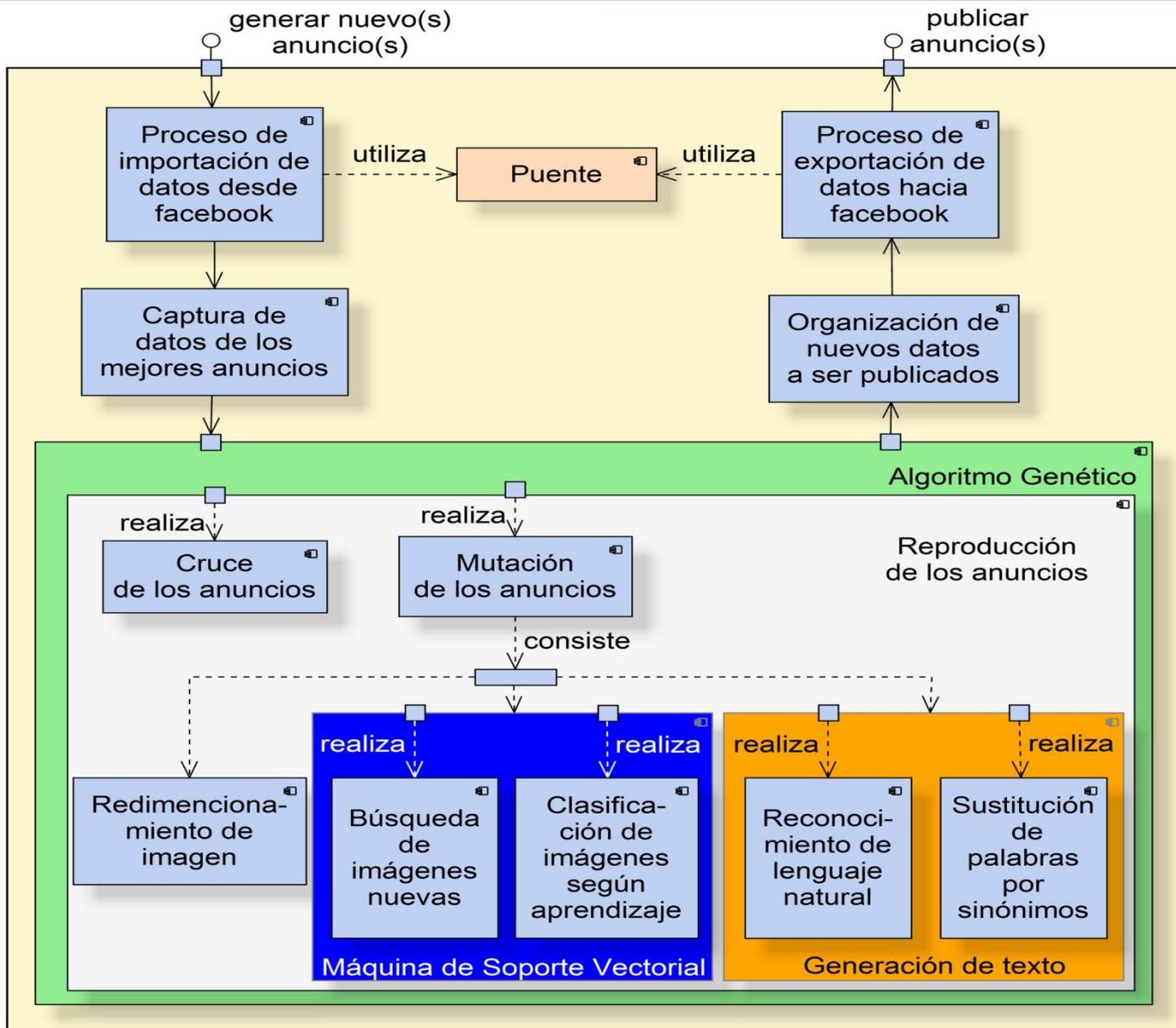
- Clicks
- Alcance
- Frecuencia
- Porcentaje de clicks del anuncio
- Interacciones
- Puja del anuncio



# Sistema Adaptativo Inteligente



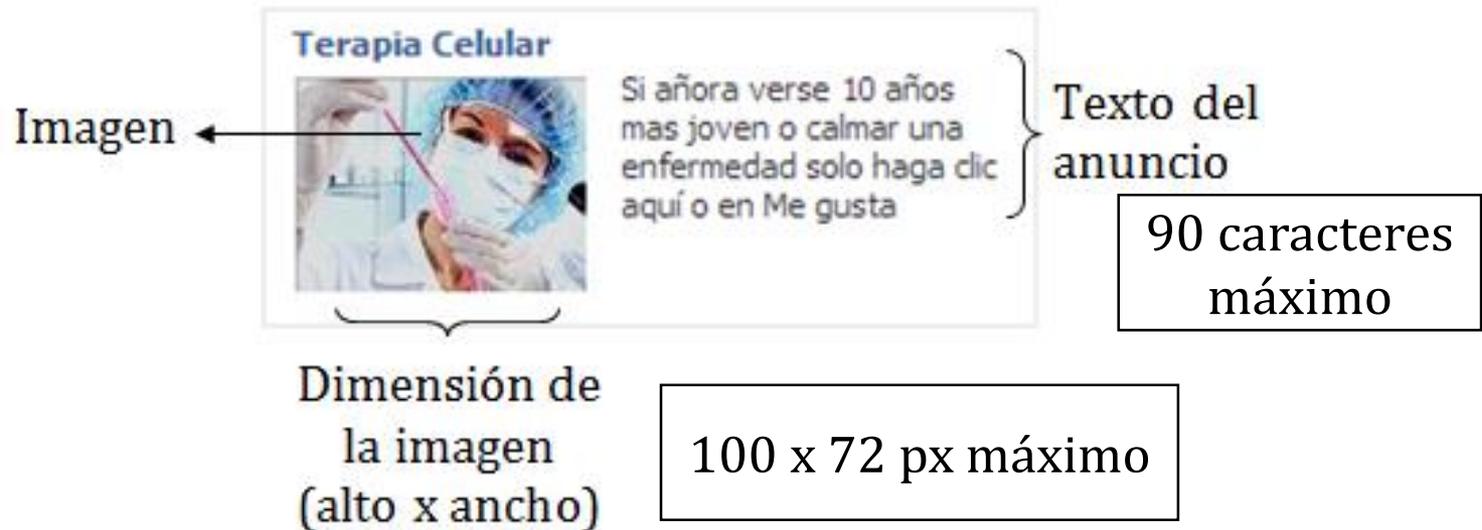
# Arquitectura del Sistema



## ➤ 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

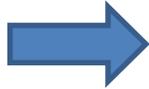
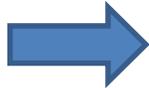


Si quiere verse 10 años mas joven o aliviar una enfermedad haga clic aquí o en Me gusta

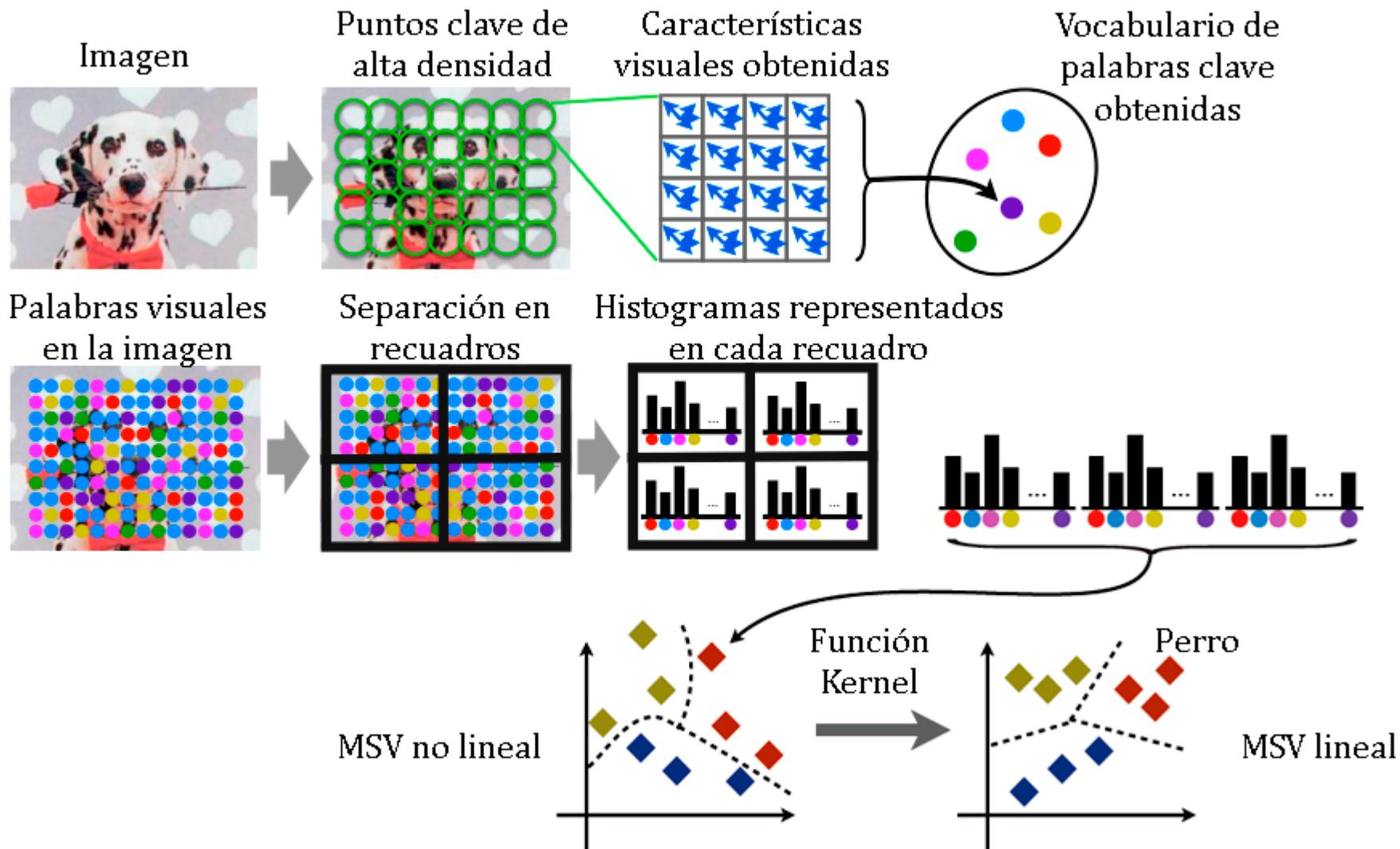
Cruce de imágenes entre dos anuncios

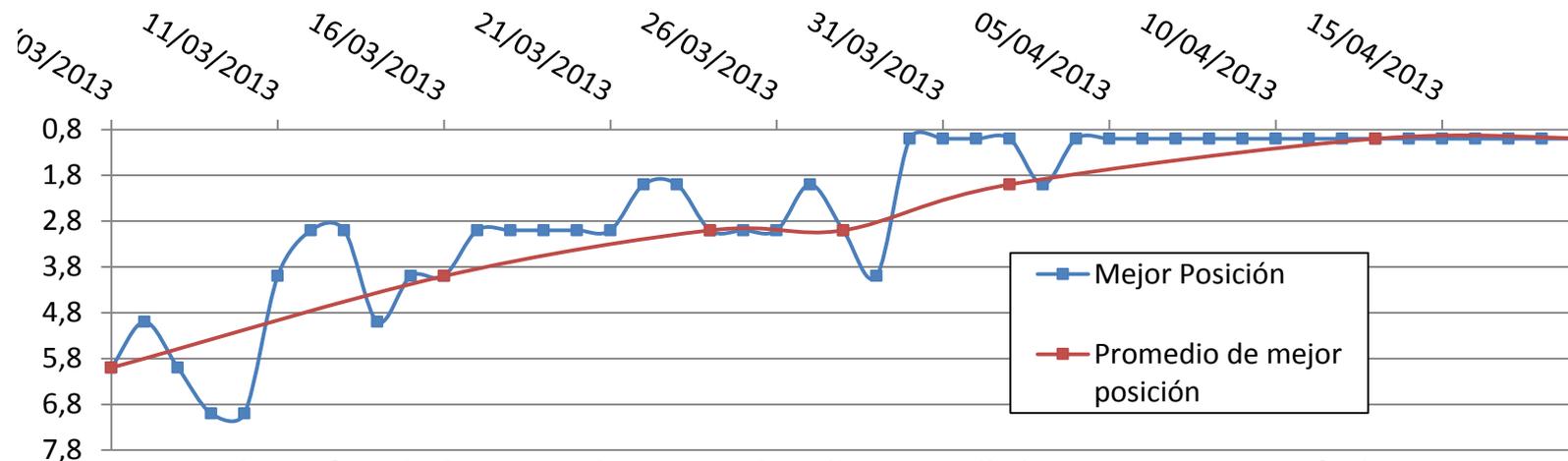
## • Operadores Genéticos

### ❖ *Mutación*

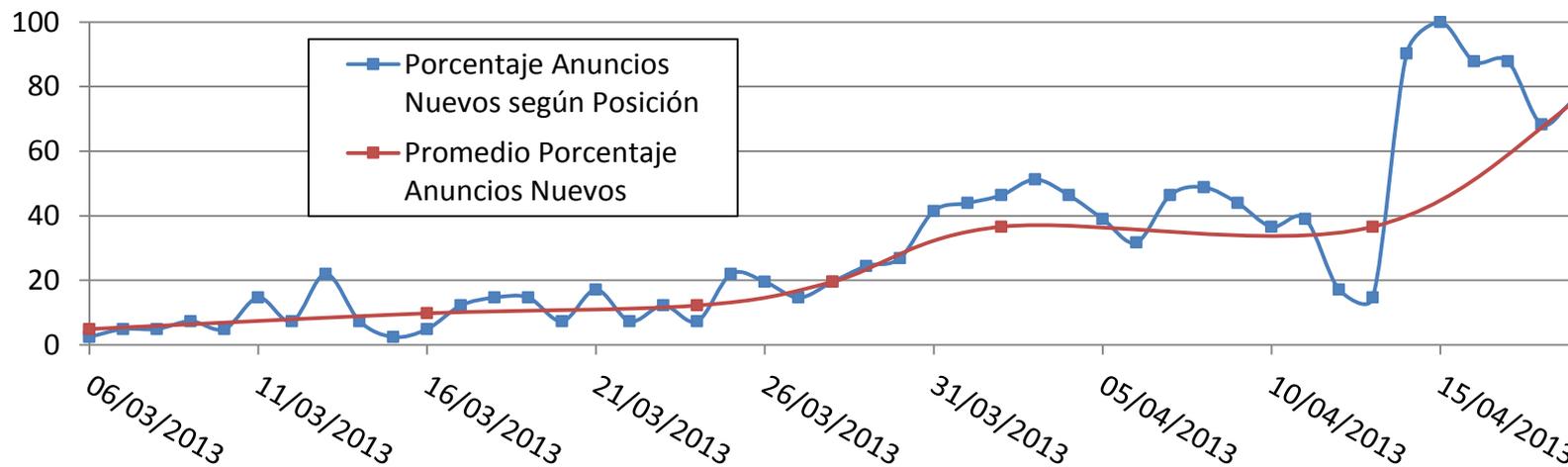
- Si se muta el texto  Generador de texto
- Si se muta la dimensión de la imagen  Se asigna un valor a la dimensión aleatoriamente
- Si se muta el contenido de la imagen  Máquina de Soporte Vectorial que escoge una imagen (nueva) adecuada

## ✓ 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)