

Minería de Grafos

Jose Aguilar

CEMISID, Escuela de Sistemas

Facultad de Ingeniería

Universidad de Los Andes

Mérida, Venezuela

Grafos

Un grafo **G** es un par ordenado de un conjunto de vértices **V** y un conjunto de aristas **E**

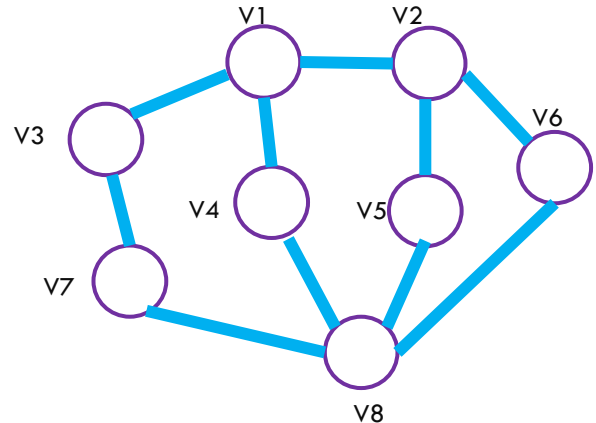
$$G = (V, E)$$

Par ordenado:

$(a, b) \neq (b, a)$ si $a \neq b$

Par No ordenado:

$\{a, b\} = \{b, a\}$

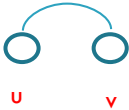


Grafos

Aristas:



Dirigido
(u, v)
(v, u)

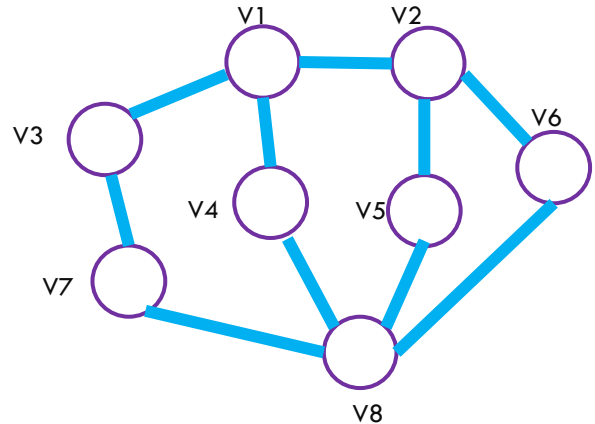


No Dirigido
{u,v}

Bucle



Varias
Aristas



$V = \{v1, v2, v3, v4, v5, v6, v7, v8\}$

$E = \{\{v1, v2\}, \{v1, v3\}, \{v1, v4\}, \{v2, v5\}, \{v2, v6\}, \{v3, v7\}, \{v4, v8\}, \{v5, v8\}, \{v6, v8\}, \{v7, v8\}\}$

Grafos

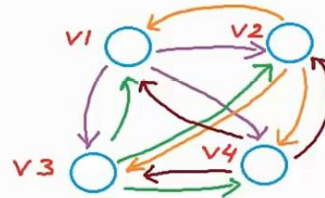
Número de Aristas:

if $|V| = n$

then,

$0 \leq |E| \leq n(n-1)$, if directed

$0 \leq |E| \leq \frac{n(n-1)}{2}$, if undirected



if $|V| = 10$, $|E| \leq 90$

if $|V| = 100$, $|E| \leq 9900$

Modelando Datos con Grafos...

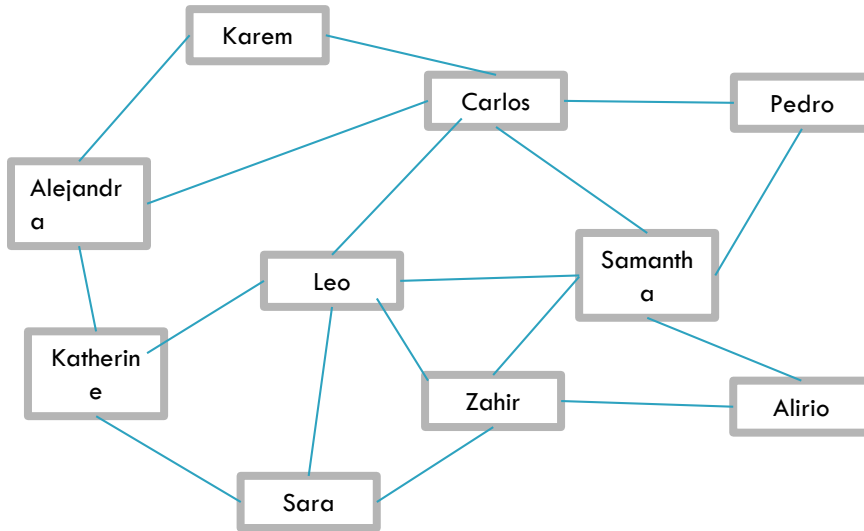
Los grafos son adecuados para la captura de las relaciones arbitrarias entre los diversos elementos.

<u>Instancia</u>		<u>Grafo</u>
Elemento	↔	Vertice
Atributos Elemento	↔	Etiquetas Vertices
Relaciones	↔	Arcos
Tipo de relaciones	↔	Etiquetas arcos

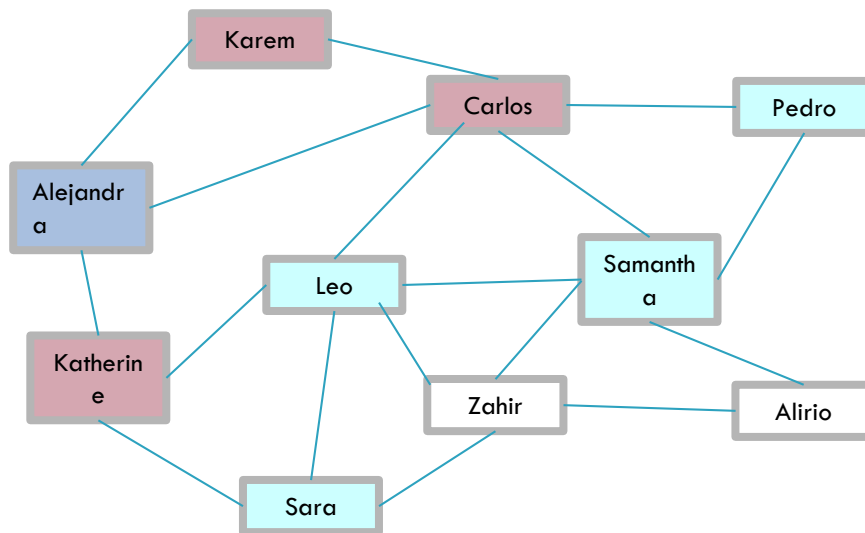
Proporcionan una enorme flexibilidad para el modelado de los datos, ya que permiten al modelador decidir cuáles son el tipo de relaciones a modelar

Grafos

Red Social
FACEBOOK



Grafos



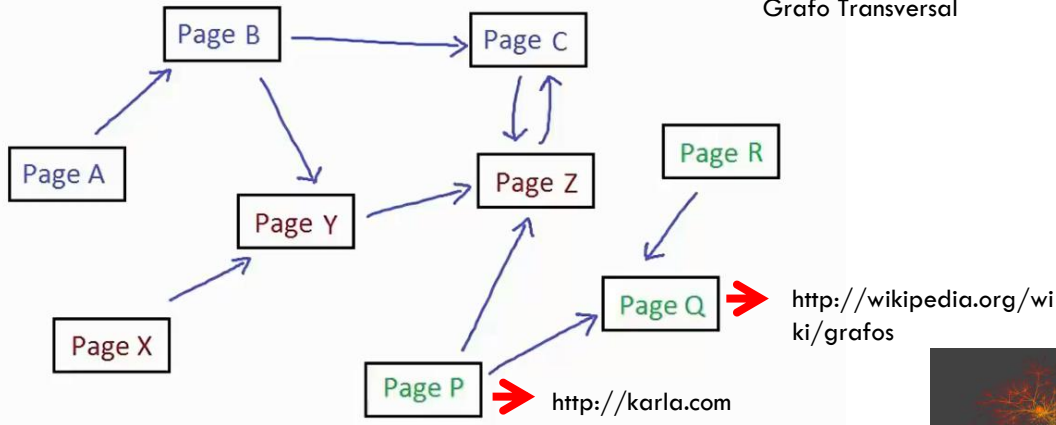
Red Social
FACEBOOK

Para Sugerir un amigo a ALEJANDRA hay que encontrar todos los nodos que tengan longitud del camino igual a 2.

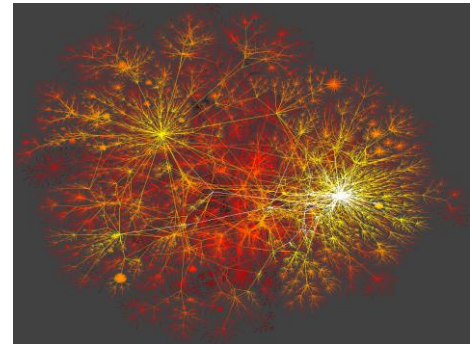
World Wide Web

Web – Crawling

Grafo Transversal

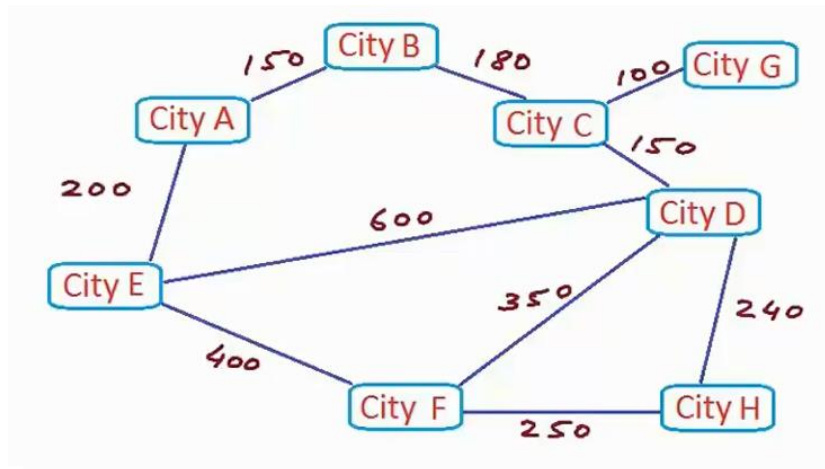


Internet



Grafos

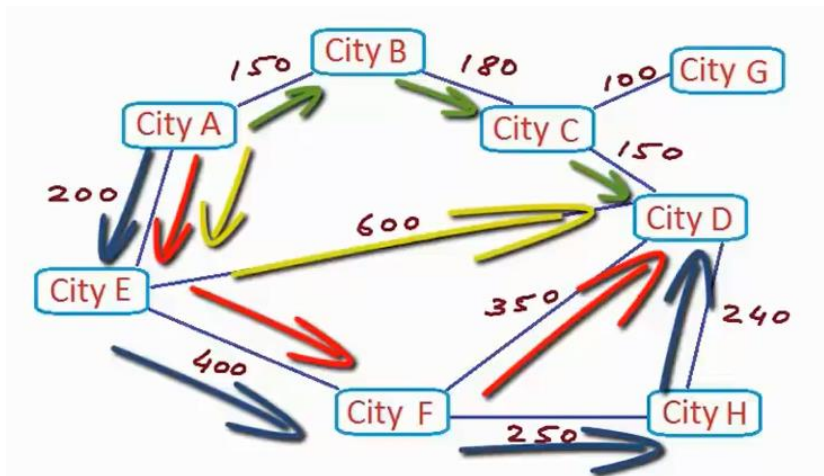
Grafos con Pesos VS Grafos sin Pesos



Red de Carreteras Inter urbanas

Grafos

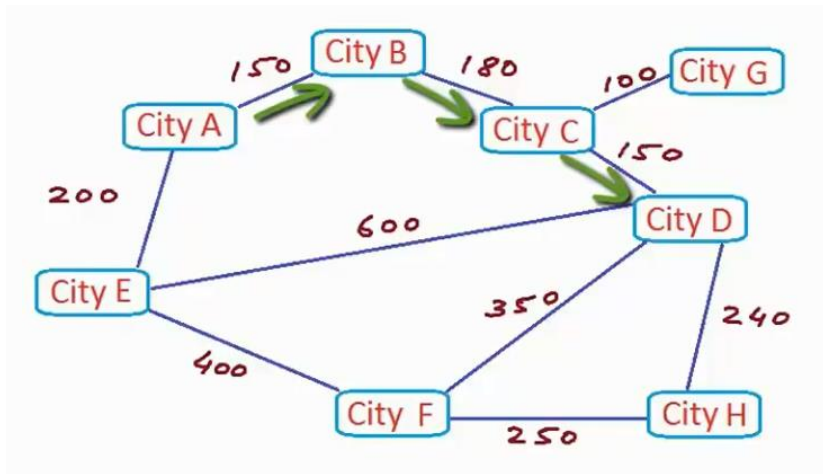
Grafos con Pesos VS Grafos sin Pesos



Red de Carreteras Inter urbanas

Grafos

Grafos con Pesos VS Grafos sin Pesos



Red de Carreteras Inter urbanas

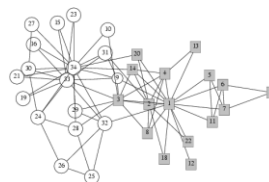
Redes en el mundo real

12

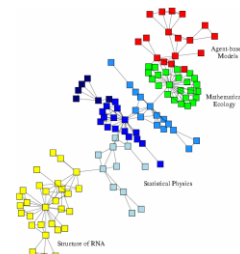
- **Redes de información:**
 - World Wide Web: hyperlinks
 - Redes de citación
 - Redes de Noticias y Blogs
- **Redes sociales**
 - Organizativas
 - Comunicativas
 - Colaborativas
 - Contactos sexuales
- **Redes tecnológicas:**
 - Energéticas
 - Transporte (aéreo, carreteras, fluviales,...)
 - Telefónicas
 - Internet
 - Sistemas Autónomos



Redes de amistad



Karate club network

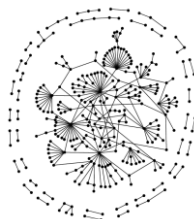


Redes de colaboración

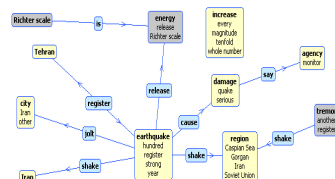
Redes en el mundo real

13

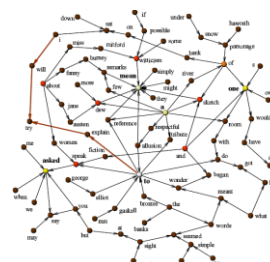
- Redes biológicas
 - ▣ Metabólicas
 - ▣ Cadenas alimenticias
 - ▣ Neuronales
 - ▣ Regulación Genética
- Redes de lenguaje
 - ▣ Semánticas
 - ▣ Lingüísticas
- Redes de software
- ...



Interacciones
entre las
proteínas de
la levadura

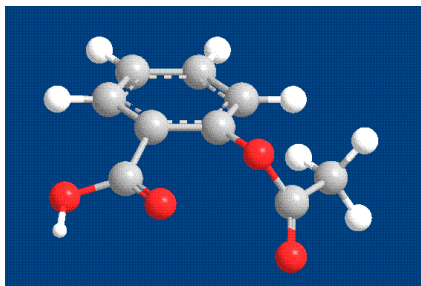


Red semántica

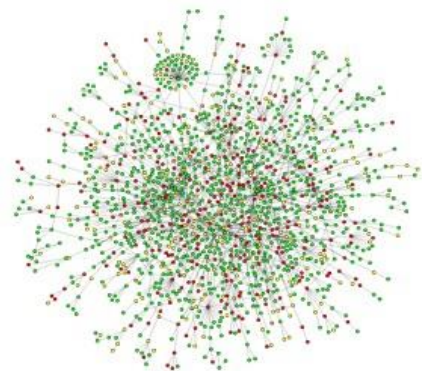


Red Lingüística

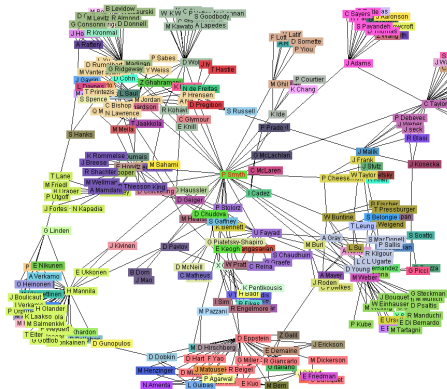
Grafos



Aspirina



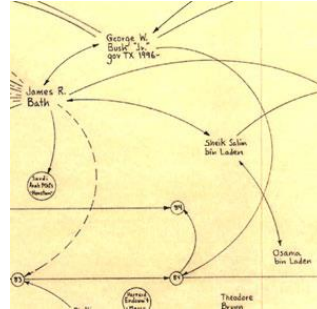
Red de interacción de proteína levadura



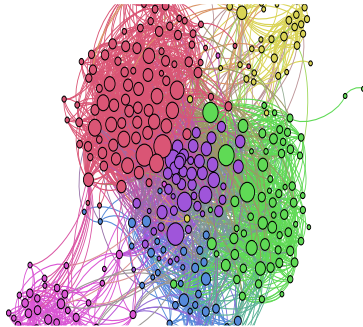
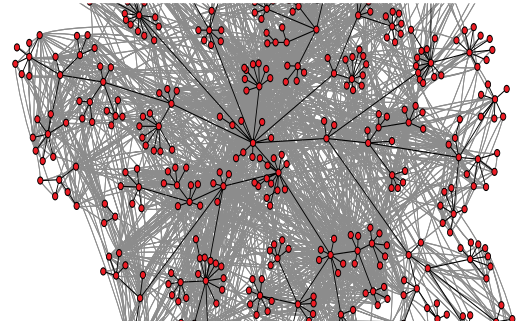
Red Co-autores de libros

Grafos

Mark Lombardi: rastreo y Mapeo fiascos financieros globales en los años 1980 a partir de fuentes públicas, como los artículos de noticias.



Relación de empleados de una organización



Facebook de alguien

Los colores separan componentes fuertemente conectados de la red.

Los arcos negros denotan estructura organizacional y los grises son interacciones por correo electrónico.

Red Social



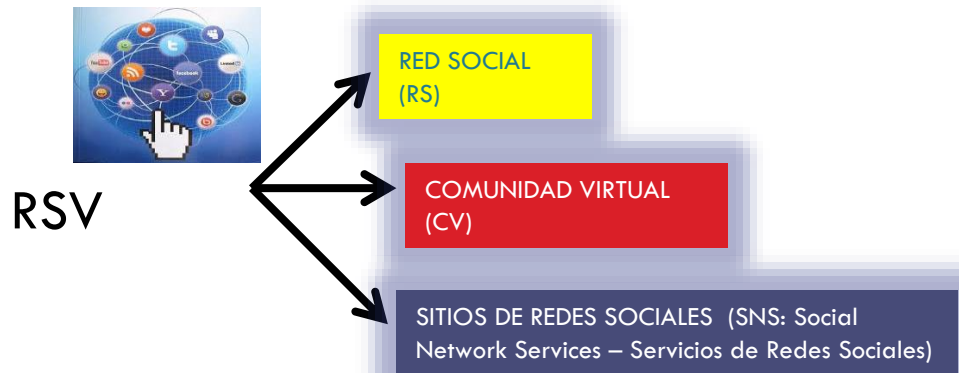
- El **análisis de redes sociales** estudia esta estructura social aplicando la teoría de grafos.
- Se analiza:
 - Si existen estructuras de comunidades ocultas
 - La difusión o las opiniones.
 - La influencia del todo en las partes y viceversa.
 - La difusión de nuevas ideas y prácticas (teoría de difusión de innovaciones).
 - El efecto producido por la acción selectiva de los individuos en la red
 - Grafos de colaboración para ilustrar buenas (amistad, alianza, citas) y malas (odio, ira) relaciones entre los seres humanos.

RED SOCIAL VIRTUAL

WEB 2.0 Y 3.0



- ❑ **Facebook** es la red social más utilizada del mundo
- ❑ **Twitter:** red social de microblogging.
- ❑ **LinkedIn** red de usuarios profesionales, y
- ❑ **Youtube** red de alojamiento de vídeos.
- ❑ **Google+**, apuesta de Google por las redes sociales.
- ❑ **Instagram**, red de intercambio de imágenes



Herramientas

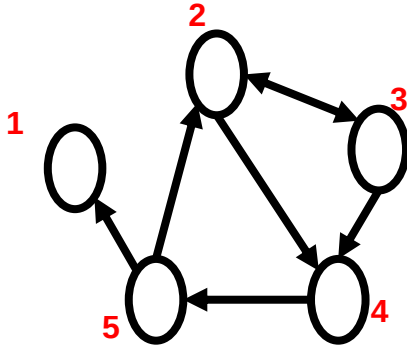
- **Gephi** (visualization and basic network metrics)
- **NetLogo** (modeling network dynamics)
- **Pajek**: amplia funcionalidad basada en menús, incluyendo muchas, muchas métricas de red y manipulaciones
 - ▣ pero ... no extensible
- **Guess**: extensibles, herramientas de secuencias de comandos de análisis exploratorio de datos, pero la selección más limitada de métodos incorporados en comparación con Pajek
- **NetLogo**: plataforma general agente basado en la simulación con el apoyo de modelado excelente red
 - ▣ muchos de los demos en este curso fueron construidos con NetLogo
- **IGRAPH**: utilizado en la versión de nivel de doctorado. bibliotecas se puede acceder a través de R o Python. Rutinas escalan a millones de nodos. (for programming assignments)

Elementos de un grafo

- Dirigido
 - ▣ $A \rightarrow B$
 - A le gusta B, A le dio un regalo a B, A es hijo de B
- No dirigido
 - ▣ $A \leftrightarrow B$ o $A - B$
 - A y B se gustan, son semejantes
 - ▣ Peso (frecuencia de comunicación)
 - ▣ ranking (mejor amigo, segundo mejor amigo...)
 - ▣ tipo (amigo, pariente, co-trabajador)

Representación de los datos

Matriz de adyacencia



$A =$

0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	0	0	0	1
1	1	0	0	0

Lista de adyacencia

■ Todos los vecinos de cada nodo

- 1:
- 2: 3 4
- 3: 2 4
- 4: 5
- 5: 1 2

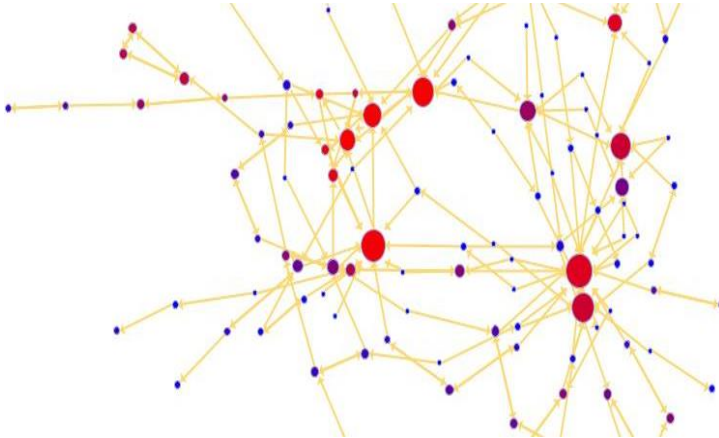
Lista de arcos

- 2, 3
- 2, 4
- 3, 2
- 3, 4
- 4, 5
- 5, 2
- 5, 1

■ Más fácil para redes

- Grandes
- Dispersas

Métricas



¿Cuál es el nodo con más arcos?

Métricas de redes

Cada métrica de red da respuesta a las siguientes preguntas:

➤ pregunta: ¿Quién es más central?

1) METRICA DE RED: centralidad

a) Centralidad de grado (degree centrality).

1) Indegree o grado de entrada

2) Outdegree o grado de salida

b) Centralidad de cercanía (closeness centrality).

c) Centralidad de intermediación (Betweenness centrality).

➤ pregunta: ¿Todo está conectado?

2) METRICA DE RED: los componentes conectados

- Componentes fuertemente conectados:

- Componentes Débilmente conectados:

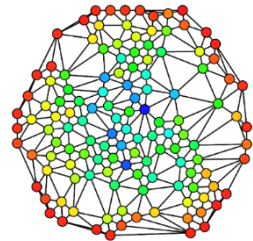
3) METRICA DE RED: tamaño de componente gigante(giant component)

➤ pregunta: ¿A qué distancia están las cosas?

4) METRICA DE RED: rutas más cortas

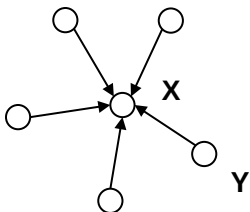
➤ pregunta: ¿Cómo densa son?

5) METRICA DE RED: densidad grafo

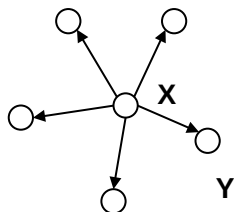


Métricas: Centralidad

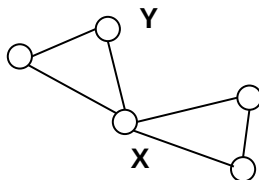
Medidas posibles de un vértice en un grafo, que determina su importancia relativa dentro de éste



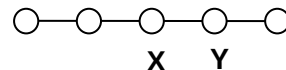
indegree



outdegree



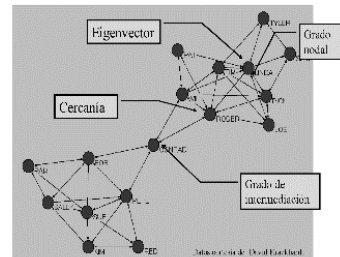
Betweenness
(intermediación)



Closeness
(cercanía)

Centralidad de vector propio
(eigenvector centrality)

Cuatro Aspectos de la Centralidad



Métricas: Propiedades de los nodos de la Red

Conexiones

indegree

cuantos arcos están dirigidos al nodo



indegree=3

$$\sum_{i=1}^n A_{ij}$$

outdegree

arcos que salen del nodo



outdegree=2

$$\sum_{j=1}^n A_{ij}$$

degree (in or out)

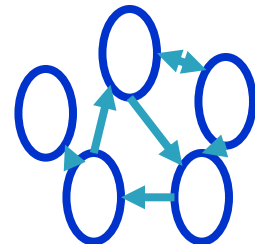
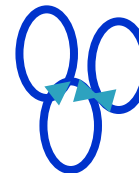
todos los arcos del nodo, entrada y salida



degree=5

Degree sequence: Lista ordenada de los grados de cada nodo

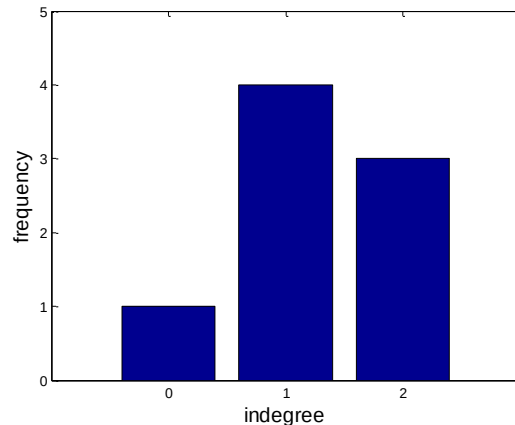
- In-degree sequence:
 - [2, 2, 2, 1, 1, 1, 1, 0]
- Out-degree sequence:
 - [2, 2, 2, 2, 1, 1, 1, 0]
- (undirected) degree sequence:
 - [3, 3, 3, 2, 2, 1, 1, 1]



Métricas: Propiedades de los nodos de la Red

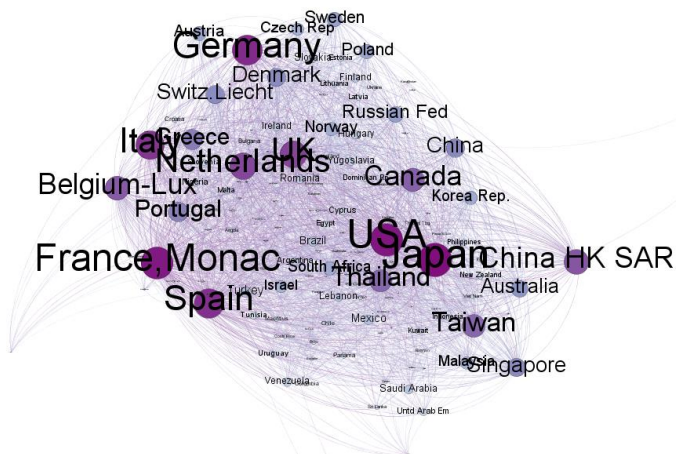
- **Degree distribution:** La frecuencia con la que ocurre cada grado

- In-degree distribution:
 - [(2,3) (1,4) (0,1)]
- Out-degree distribution:
 - [(2,4) (1,3) (0,1)]
- (undirected) distribution:
 - [(3,3) (2,2) (1,3)]

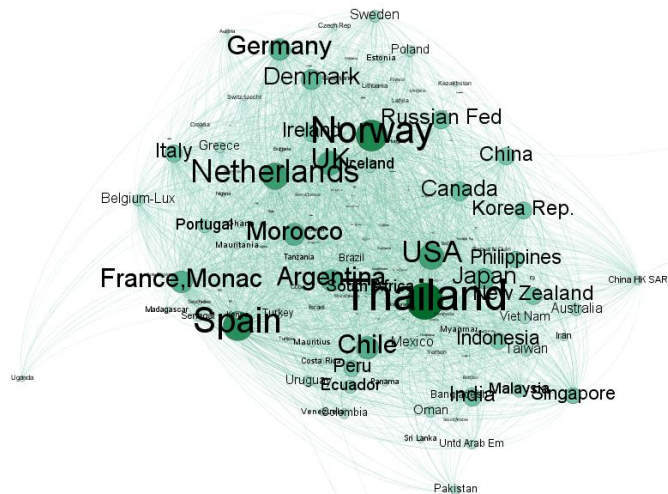


Métricas

InDegree



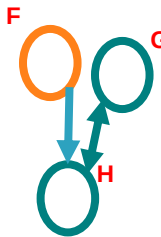
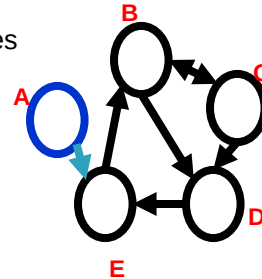
OutDegree



Métricas: Componentes conectados

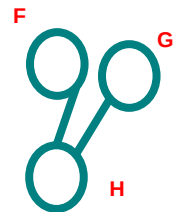
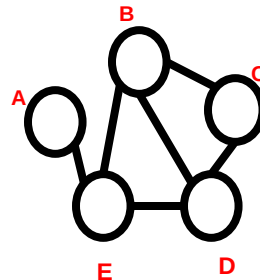
Componentes fuertemente conectados:

- Cada nodo dentro del componente se puede llegar desde cualquier otro nodo en el componente siguiendo los enlaces dirigidos
- Componentes fuertemente conectados
 - B C D E
 - G H



Componentes Débilmente conectados: cada nodo se puede llegar desde cualquier otro nodo siguiendo enlaces en cualquier dirección

- Componentes débilmente conectados
 - A B C D E
 - G H F



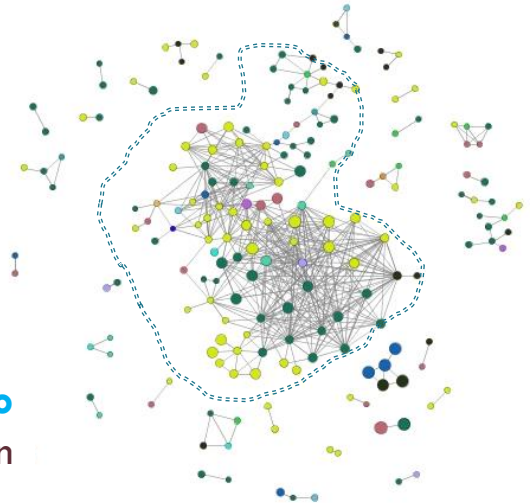
En las **redes no dirigidas** se habla simplemente de "**componentes conectados**"

Métricas: Componentes conectados

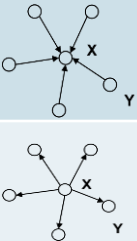
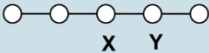
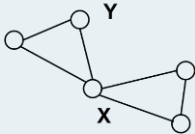
- Si el componente más grande ocupa una región significativa de la red o grafo, es llamado **giant component**

El componente gigante, consiste en un **grupo de nodos enlazados entre si**, y que agrupan a la **mayoría de los nodos** de la red.

El componente gigante aparece en casi todas las redes sociales



Métricas principales de la centralidad

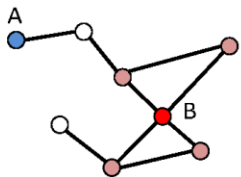
Métricas de centralidad		
a) Centralidad de grado (<i>degree centrality</i>): qué tantas conexiones directas tiene una unidad con otras unidades. Una unidad con alta centralidad de grado sirve como “conector” o “hub” de la red.	Indegree o grado de entrada outdegree o grado de salida	
b) Centralidad de cercanía (<i>closeness centrality</i>): que tan “cerca” se encuentra una unidad de la red de las otras, considerando tanto conexiones directas como indirectas. una unidad con alta centralidad de cercanía puede interactuar fácilmente con otras unidades, tiene la visibilidad del comportamiento de la red en su conjunto, y puede influir en ella.	closeness o cercanía	
c) Centralidad de intermediación (<i>Betweenness centrality</i>): índice de en qué tantas rutas más cortas entre 2 unidades cualesquiera de la red se encuentra una unidad dada. Estas unidades tienen el control del flujo de información dentro de la red.	betweenness o intermediación	

Métricas: Centralidad

MEDIDAS LOCALES DE CENTRALIDAD

Centralidad de grado

No dirigida

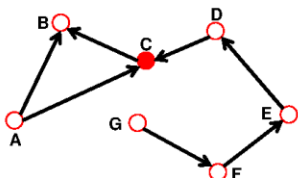


Centralidad de grado de un actor (C_D): número de enlaces que lo conectan con otros

$$C_D(A) = k_A = 1 \quad C_D(B) = k_B = 4$$

$C_D(i)$ se define en $\{0, g-1\}$, siendo g el número de nodos de la componente conexa

Dirigida



En redes dirigidas, se define el **Prestigio de entrada (*in-degree*)**, denominado **Soporte**, y el **Prestigio de salida (*out-degree*)**, denominado **Influencia**:

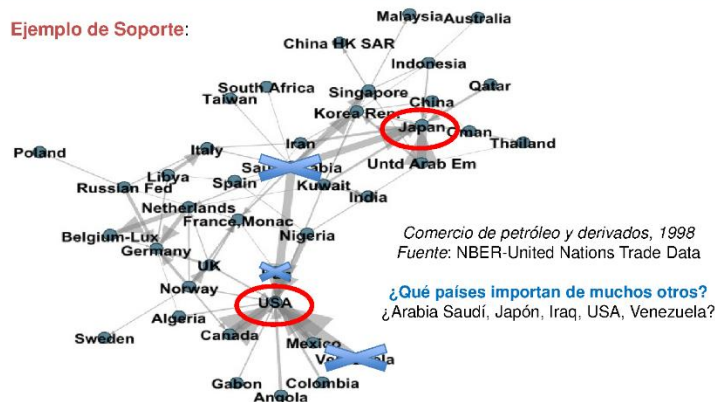
$$P_D^{in}(C) = k_C^{in} = 2 \quad P_D^{out}(C) = k_C^{out} = 1$$

Ambos se definen en $\{0, g-1\}$

MEDIDAS LOCALES DE CENTRALIDAD

Soporte

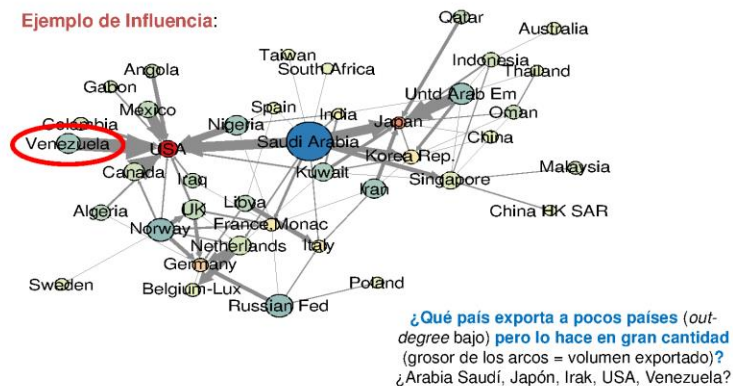
Ejemplo de Soporte:



MEDIDAS LOCALES DE CENTRALIDAD

Influencia

Ejemplo de Influencia:



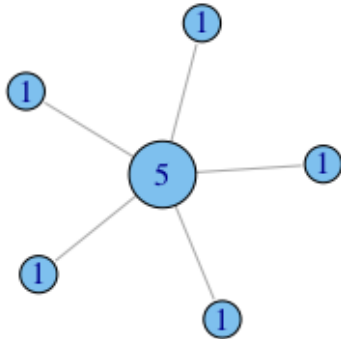
Métricas: Centralidad

$$C'_D(p_k) = \frac{\sum_{i=1}^n a(p_i, p_k)}{n-1}$$

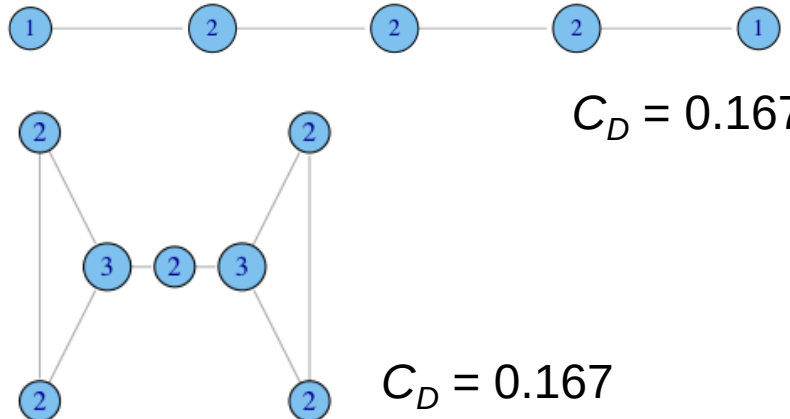
$$C_D = \frac{\sum_{i=1}^n [C_D(n^*) \cdot C_D(i)]}{[(N-1)(N-2)]}$$

Máximo valor de conexiones posibles en la red

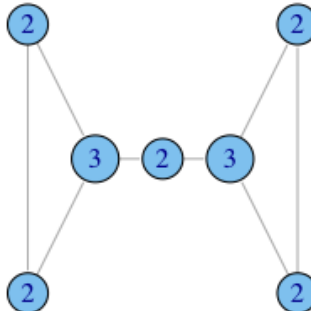
Formula de centralidad general de Freeman's



$$C_D = 1.0$$

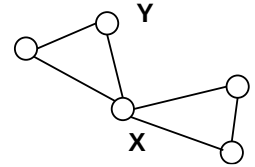


$$C_D = 0.167$$



$$C_D = 0.167$$

Métricas: betweenness



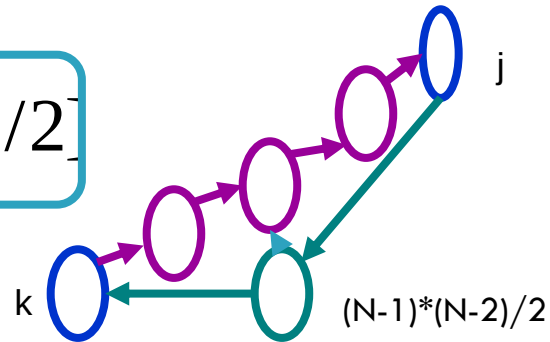
La **centralidad de intermediación** ve al nodo con una posición favorable en la medida que el nodo está situado entre los caminos entre otros pares de actores en la red.

$$C_B(i) = \sum_{j < k} g_{jk}(i) / g_{jk}$$

Pares de vértices posibles excluyendo el del mismo nodo

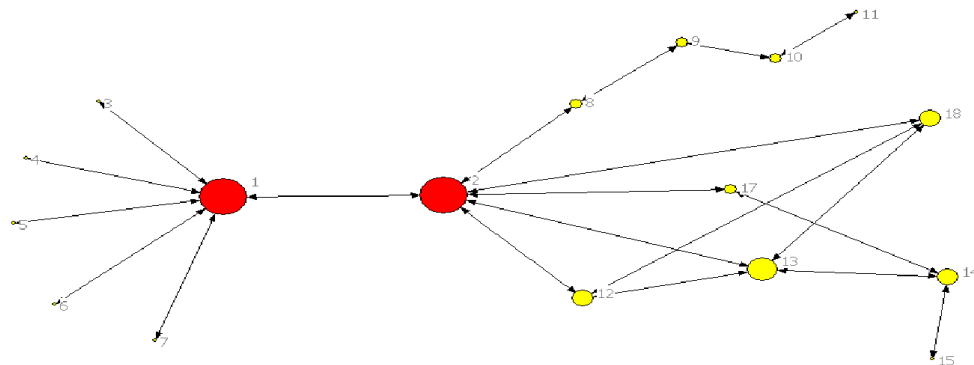
Donde g_{jk} = numero de caminos cortos que conectan jk
 $g_{jk}(i)$ = numero de caminos cortos en los que el nodo i se encuentra.

$$C'_B(i) = C_B(i) / [(n-1)(n-2)/2]$$



MEDIDAS LOCALES DE CENTRALIDAD

Grado vs. Intermediación (1)

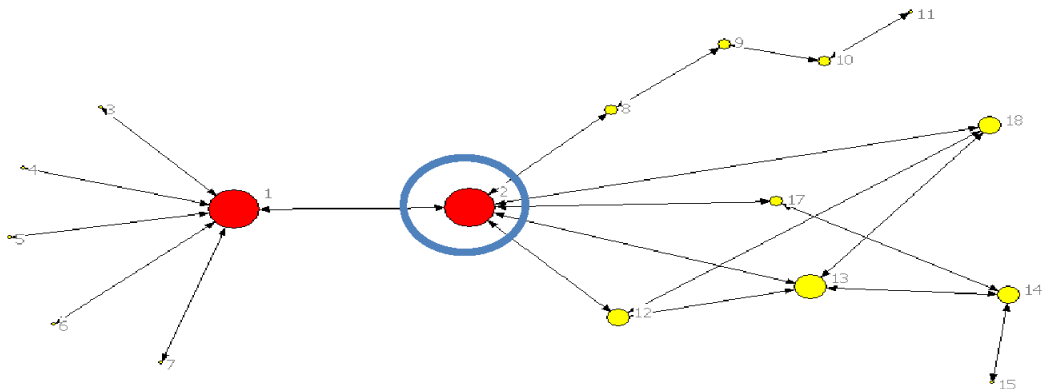


El grado de los nodos en rojo (1 y 2) es el mismo (6) pero, evidentemente, no son igual de importantes

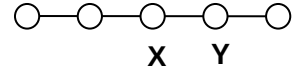
¿Por qué? Por la debilidad de esta medida, que solo toma en cuenta los vínculos inmediatos (a un nivel local) dejando de lado los vínculos indirectos (a nivel global)

MEDIDAS LOCALES DE CENTRALIDAD

Grado vs. Intermediación (2)



Nombre	Grado (C_D) Max = 17	Intermediación (C_B) Max = $17 \cdot 16 / 2 = 136$
1	6	70.00
2	6	96.50

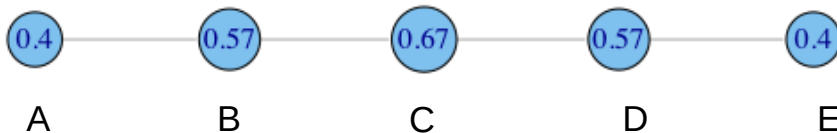


Métricas: closeness

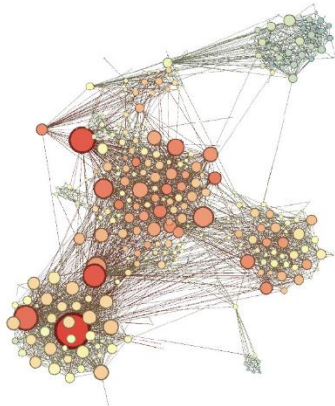
Distancia promedio del camino mas corto entre un nodo a todos los nodos.

Closeness Centrality:
$$C_c(i) = \frac{1}{\sum_{j=1}^N d(i, j)}$$

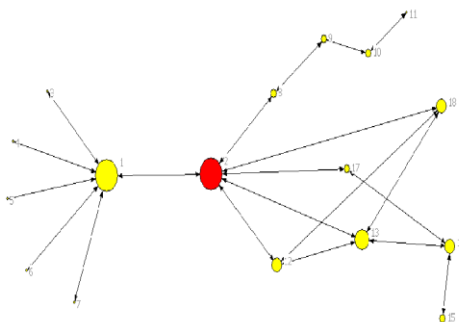
Normalized Closeness Centrality
$$C'_c(i) = (C_c(i)) / (N - 1)$$



$$C'_c(A) = \frac{1}{\sum_{j=1}^N d(A, j)} = \frac{1}{1+2+3+4} = \frac{1}{10} = 0.1$$



Red Personal de Contactos de Facebook de Oscar Cordón: el tamaño de los nodos indica el **grado** y el color la **cercanía** (más **azul**, menor valor; más **rojo**, mayor valor)



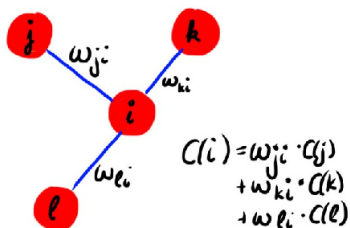
Actor	Lejanía	CercaníaN
2	34.000	50.000
1	40.000	42.500
13	42.000	40.476
17	44.000	38.636
8	44.000	38.636
12	45.000	37.778
18	45.000	37.778
14	52.000	32.692
6	56.000	30.357
5	56.000	30.357
7	56.000	30.357
3	56.000	30.357
4	56.000	30.357
9	56.000	30.357
15	66.000	25.758
10	70.000	24.286
16	82.000	20.732
11	86.000	19.767

MEDIDAS LOCALES DE CENTRALIDAD

Centralidad de vector propio (1)

La **Centralidad de vector propio** se basa en que la centralidad de un nodo concreto depende de cómo de centrales sean sus vecinos (**prominencia**)

La idea básica es que el poder y el status de un actor (**ego**) se define recursivamente a partir del poder y el status de sus vecinos (**alters**)



w_{ij} (a_{ij}) corresponde a la entrada de la matriz de adyacencia. Puede ser binaria $\{0,1\}$ o un peso numérico

La medida es válida para redes dirigidas (**Prestigio de rango**) y no dirigidas

Es una versión más elaborada de la Centralidad de grado al asumir que no todas las conexiones tienen la misma importancia. No se tiene en cuenta la cantidad sino la calidad de las mismas

Métricas: Eigenvector centrality

Bonacich

$$c_i(b) = \alpha (a + \sum_j b c_j) A_{ji}$$

$$c(b) = \alpha (I - bA)^{-1} A \mathbf{1}$$

- α constante de normalización
- β determina importancia de la centralidad de los vecinos
- A matriz de adjacencia (puede tener pesos)
- I matriz de identidad
- $\mathbf{1}$ matriz de unos.

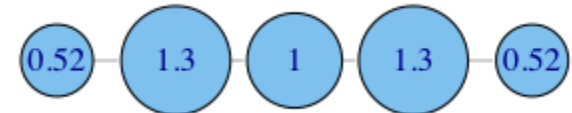
Si $\beta > 0$, los nodos tienen mayor centralidad cuando tienen enlaces con nodos con alta centralidad

$\beta = .25$



Si $\beta < 0$, los nodos tienen mayor centralidad cuando tienen enlaces a nodos no tan centrales

$\beta = -.25$



Otra medida local de centralidad basada en distancias es la **Centralidad de excentricidad (C_E)**. Se define como la inversa de la **excentricidad** (la máxima distancia geodésica) entre un actor y cualquier otro actor de la red:

$$C_E(i) = \frac{1}{\max_{j \in V(G)/i} d(i, j)}$$

$$C'_E(i) = C_E(i) / g - 1$$

Los actores con un mayor valor de excentricidad se denominan **actores periféricos**, los de menor valor forman el **centro de la red**

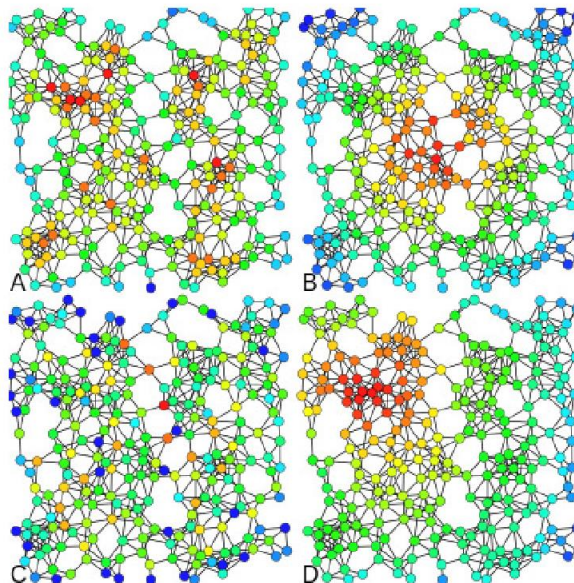
MEDIDAS LOCALES DE CENTRALIDAD

Comparativa

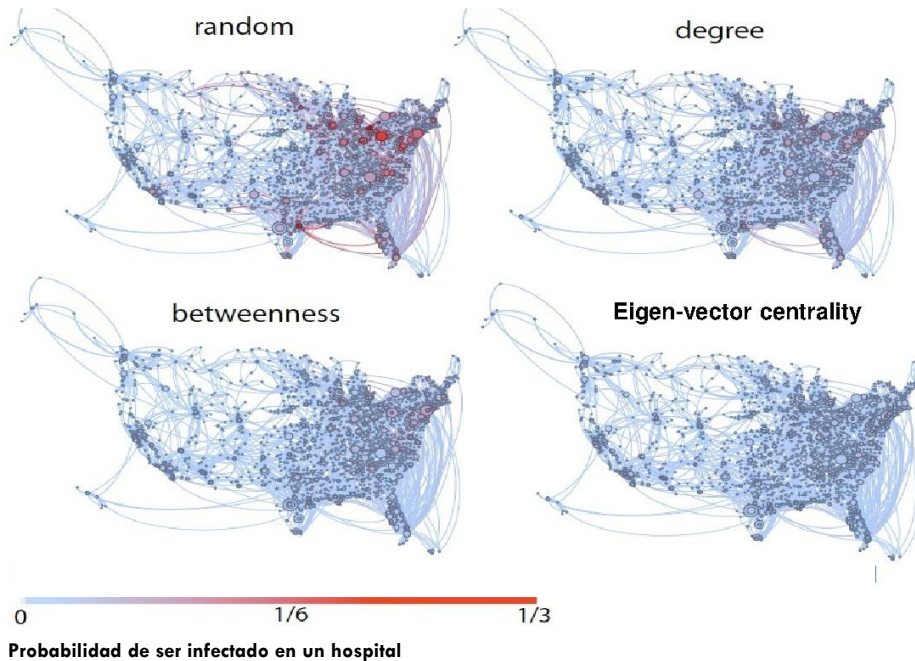
A) centralidad de grado; B) cercanía; C) intermediación; D) centralidad de vector propio

azul = menor valor

rojo = mayor valor



Red de transferencias de pacientes hospitalarios
Presupuesto para estrategias para evitar propagación de infecciones

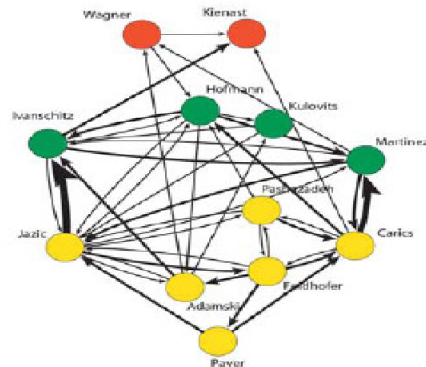
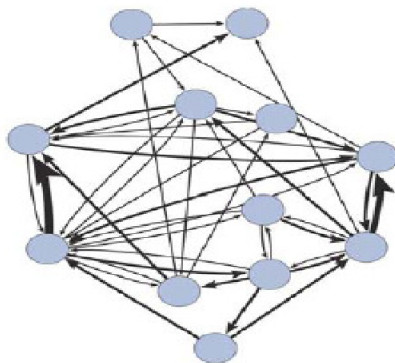


APLICACIONES

Análisis de juego en equipos de fútbol (1)

Descripción de un equipo de futbol de Viena llamado Rapid durante un juego realizado en diciembre del 2003

Describe el comportamiento en el Segundo tiempo



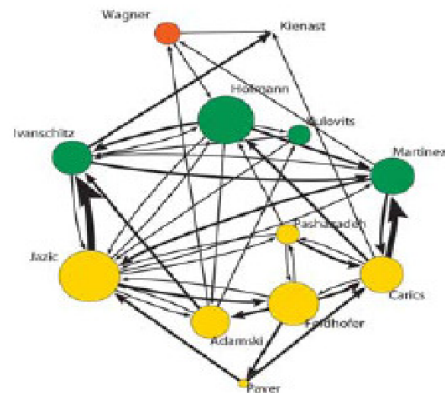
Grafo que describe los pases entre los jugadores

APLICACIONES

Análisis de juego en equipos de fútbol (2)

Posibles factores de análisis:

- ¿Qué jugador ha iniciado más pases (grado ponderado de salida)? **Jazic**
- ¿Qué jugador ha recibido más pases (grado ponderado de entrada)? **Jazic**
- ¿Quién ha controlado el juego del Rapid (centralidad)? **Jazic** y **Hoffman**
- ¿Qué jugadores han estado implicados en jugadas con el mayor número de pases (camino)? **Jazic**, **Hofmann**, **Feldhofer**, **Martinez** y **Carics**
- ¿Quién ha jugado con quién y quién no (análisis de los enlaces)? **Ni un solo pase de Ivanschitz a Wagner**
- ¿Qué grupos de jugadores han compuesto la columna vertebral del equipo (análisis de triadas)? **Por ejemplo, Feldhofer-Carics-Pashazadeh**
- ¿Qué jugadores han tenido un rol similar (análisis de enlaces)? **Por ejemplo, Ivanschitz / Martinez**



APLICACIONES

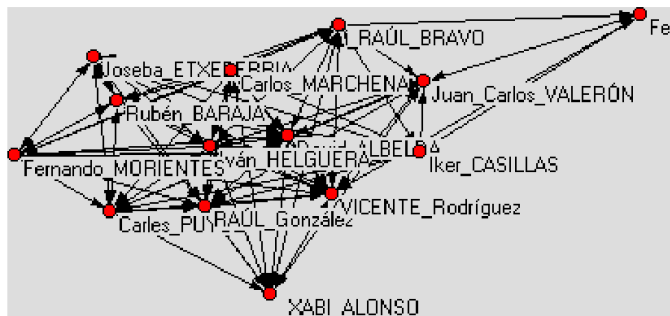
Análisis de juego en equipos de fútbol (4)

J.J. Merelo. Redes contra redes: el fútbol es así. <http://atalaya.blogalia.com/historias/19642>

Euro-copa Portugal 2004

Extraer las estadística de pases

España 1 – Rusia 0



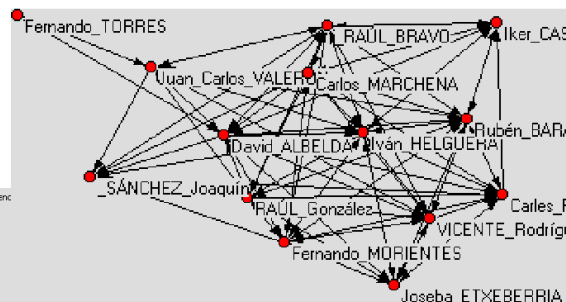
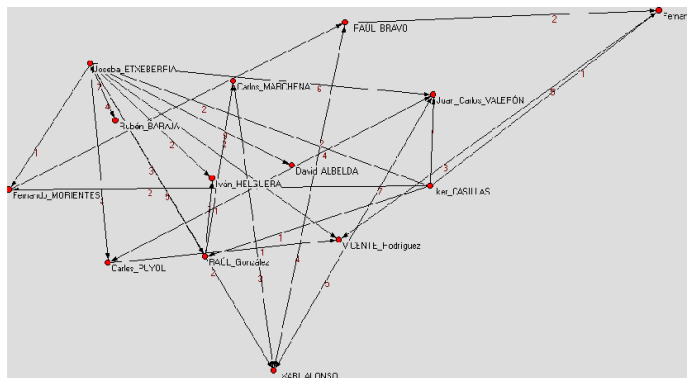
Una red donde, curiosamente, el jugador que tiene más centralidad es Iker Casillas, cuando debería ser un medio como Baraja

APLICACIONES

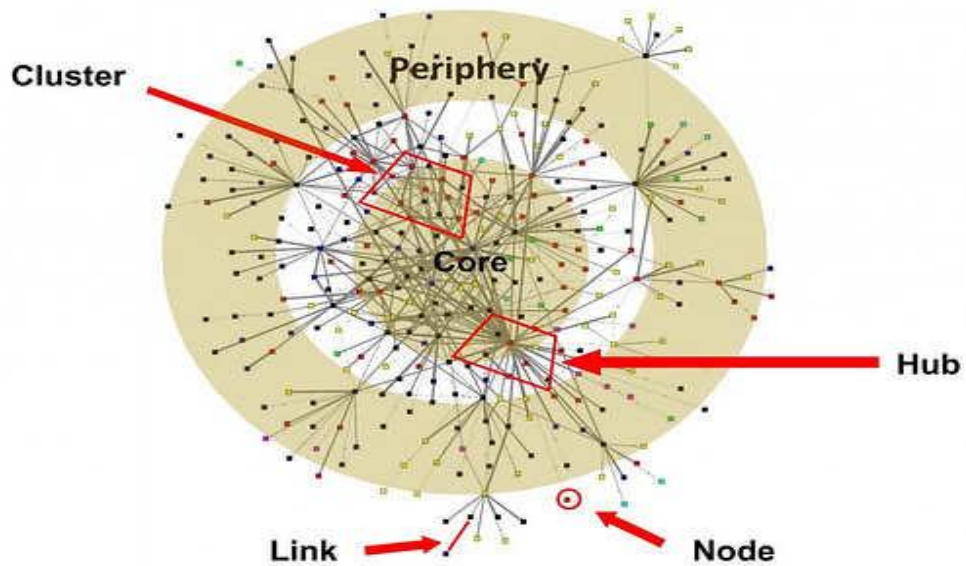
Análisis de juego en equipos de fútbol (5)

La situación no cambió mucho en el segundo partido (Grecia 1 – España 1) salvo que, en este caso, Albelda, Baraja y Helguera organizaron un poco más el juego. Y Fernando Torres a su bola, claro

Casi el 90% de los pases fueron los mismos. Esta es la diferencia de las dos redes (sin Joaquín porque es un nodo nuevo)



Es curioso ver también que la "autoridad" de la red es Vicente, un extremo. Lo lógico sería que las autoridades fueran los delanteros, pero Morientes y Raúl se hallan ahí perdidos, en la maraña de la red



Métricas: Comunidades

☐ **Mutualidad**

- ☐ Cada miembro conoce a todos los miembros

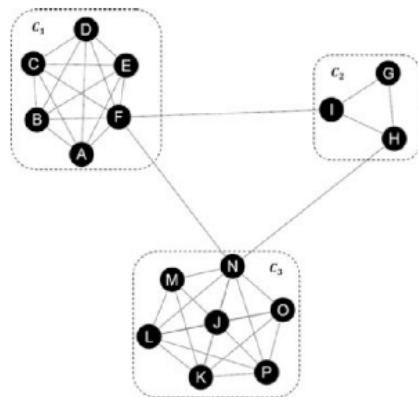
☐ **Frecuencia**

- ☐ Cada miembro conoce al menos k miembros del grupo

☐ **Cercanía**

- ☐ Los miembros están separados por máximo de n saltos

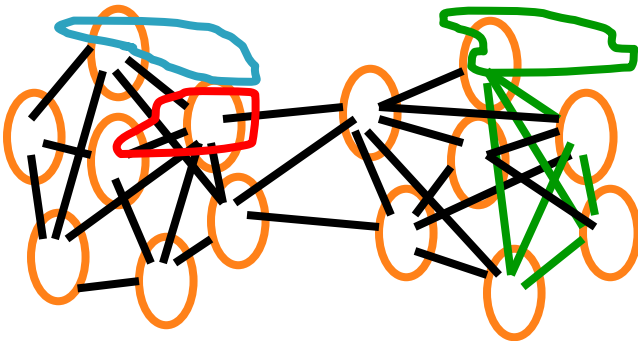
Métricas: Comunidades



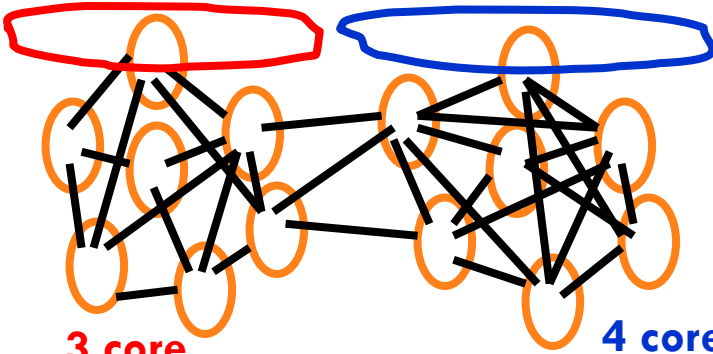
- En esta red, hay **tres comunidades**: C_1 , C_2 y C_3
- Cada comunidad está formada por un grafo completo (un **clique**) de tamaño variable ($C_1 = K_6$, $C_2 = K_3$ y $C_3 = K_7$)
- La densidad de enlaces entre las comunidades es muy baja. Los pocos enlaces que existen son **puentes**

Cliques y K-core

Cada miembro del grupo posee un link a todos los miembros del grupo



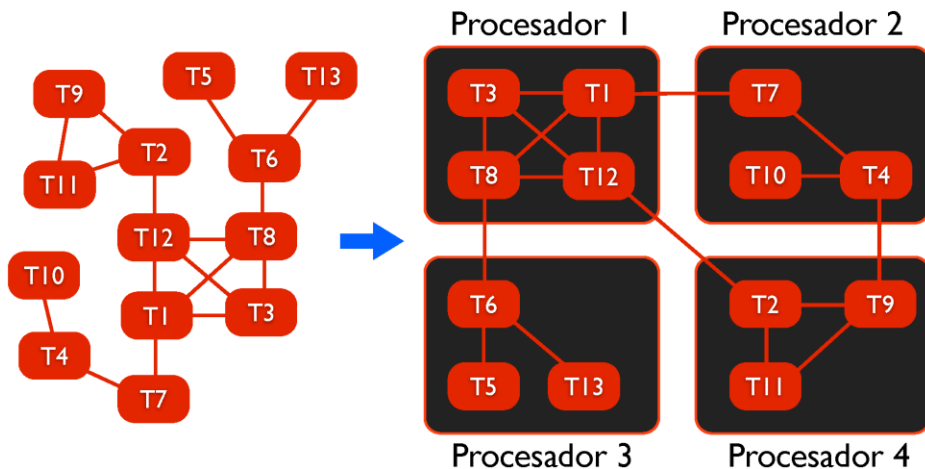
- ☐ Si se pierde un enlace, deja de ser un clique
- ☐ No es interesante que todos estén conectados con todos
- ☐ No hay medidas de centralidad dentro de un clique



Cada miembro del grupo está conectado con k otros miembros del grupo

Computación en Paralelo

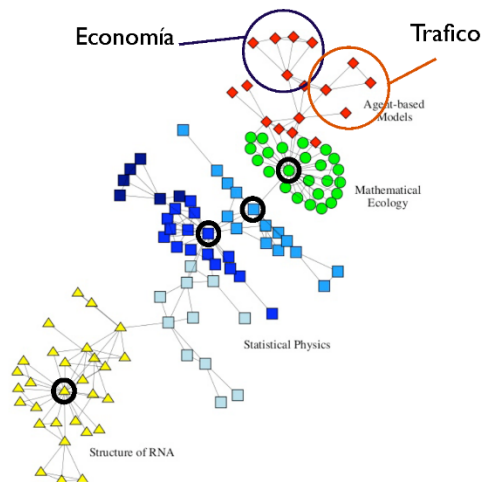
(Particionado de Grafos)



Distribución mas eficiente de tareas en un conjunto de procesadores

Redes de colaboración científica

- ◆ Modelos basados en agentes para estudiar problemas de economía y flujo de tráfico
 - Modelos matemáticos en ecología
 - Física estadística
 - ▲ Estructura del ARN
- Formación de comunidades en torno a la metodología
 - El centro de las comunidades corresponde al jefe del grupo de investigación.



Red de colaboración de científicos del Instituto de Santa Fe en Nuevo México

M. Girvan and M. E. J. Newman (2002). "Community structure in social and biological networks". Proc. Natl. Acad. Sci. USA 99 (12): 7821–7826. doi:10.1073/pnas.122653799. PMC 122977. PMID 12060727.

Métricas: Comunidades

Algoritmos de detección

- Edge Betweenness Method, M. Girvan and M. E. Newman (GN)
- Fast greedy modularity optimization, A. Clauset, M. E. Newman, and C. Moore (Clauset et al.)
- Exhaustive modularity optimization via simulated annealing, R. Guimerá, M. Sales-Pardo (Sim ann.)
- Multi-Level Aggregation Method based on modularity, V. D. Blondel, J.-L. Guillaume, R. Lambiotte (Blondel et al.)
- Divisive algorithm based on the edge-clustering coefficient, F. Radicchi, C. Castellano, F. Cecconi (Radicchi et al.)
- Clique Percolation Method for finding communities, G. Palla, I. Derenyi, I. Farkas (Cfinder)
- Graph clustering by flow simulation, S. van Dongen (MCL)

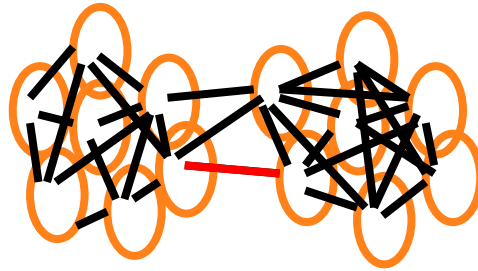
Betweenness clustering

Entrada: Grafo inicial $G(E,V)$

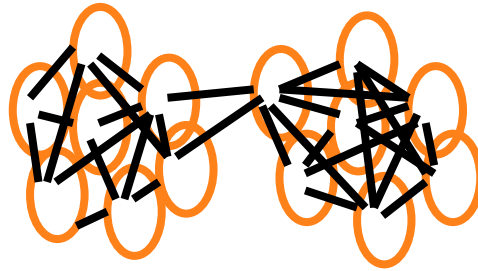
Inicio:

1. Calcular la intermediación para todas las aristas en la red.
2. Generar un nuevo grafo al remover las aristas con un valor alto de intermediación.
3. Recalcular la intermediación para todas las aristas del nuevo grafo.
4. Repetir el paso 2 hasta que todas las aristas tengan un nivel bajo de intermediación.

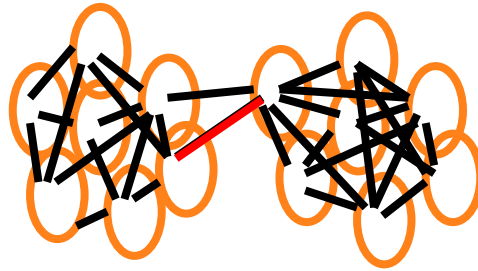
Betweenness clustering



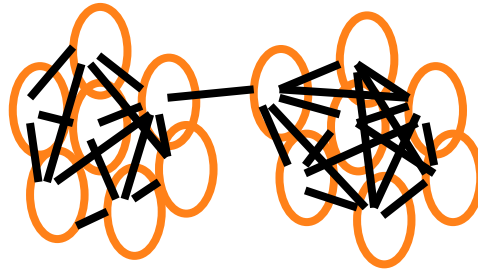
Betweenness clustering



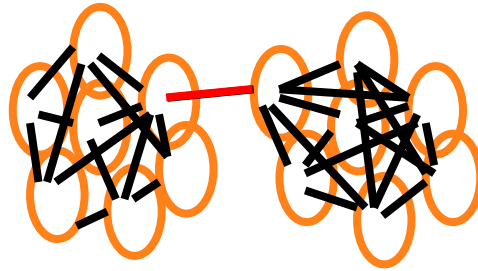
Betweenness clustering



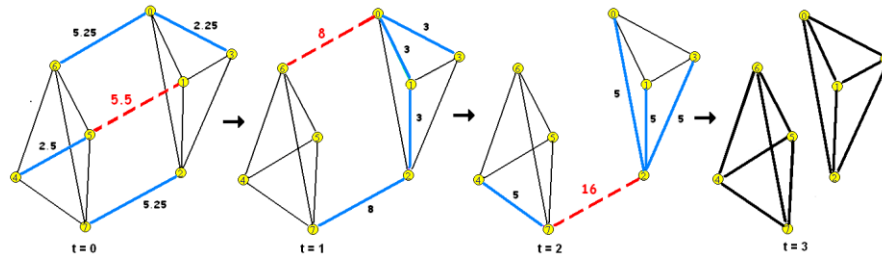
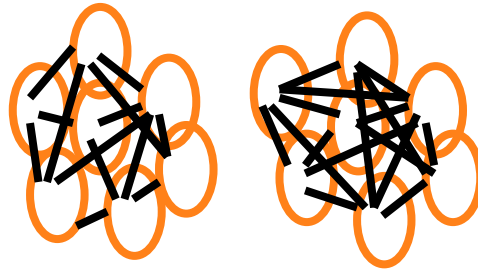
Betweenness clustering



Betweenness clustering



Betweenness clustering



Modularidad

- Métrica diseñada para dividir la red en módulos, clusters, comunidades.
- Es usada para maximizar métodos para hallar comunidades
- Una red con alta modularidad significa que:
 - Es muy densa entre nodos de una misma comunidad
 - Pero con conexiones dispersas entre nodos de comunidades distintas

Modularidad

$$Q = \frac{1}{2m} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w)$$

1 Si v y w están en la misma comunidad, 0 si no

Matriz de adyacencia

Probabilidad de un arco entre dos nodos, es proporcional a sus grados

Detección basada en la Modularidad

Heurísticas utilizadas para la optimización de la modularidad:

- Recocido Simulado (Guimera and Amaral)
- Optimización extrema (J. Duch and A. Arenas)
- Algoritmos voraces (Clauset et al.)
- Reformulación de la modularidad en términos de las propiedades espectrales de la red. (Newman)

Las dos ultimas heurísticas han resultado efectivas sin embargo poseen un problema inherente al concepto de modularidad llamado *limite de resolución*.

Modularity and community structure in networks, M. E. J. Newman, Proc. Natl. Acad. Sci. USA 103, 8577–8582 (2006).

Wikipedia, 2011. Modularity (networks). [en línea] Disponible en: <[http://en.wikipedia.org/wiki/Modularity_\(networks\)](http://en.wikipedia.org/wiki/Modularity_(networks))>
[Consultado el 17 Noviembre 2011].

MEDIDAS GLOBALES

Existen varias medidas globales en SNA. La mayoría son las mismas empleadas para analizar cualquier otro tipo de red, que ya hemos estudiado:

1. **Diámetro y Radio**
2. **Distancia media**
3. **Grado Medio**
4. **Densidad**
5. **Coeficiente de Clustering Global**
6. **Reciprocidad**

Diámetro ($d_{\max}$): longitud del camino mínimo más largo de la red

En redes grandes, se puede determinar con el algoritmo de búsqueda primero en anchura

Equivale al valor máximo de excentricidad para todos los nodos de la red:

$$E(i) = \max_{j \in V(G) / i} d(i, j) \quad d_{\max} = \max \{E(i) : i \in V(G)\}$$

En el contexto del SNA, esta métrica da una **idea de la proximidad entre pares de actores en la red**, indicando cómo de lejos están en el peor de los casos

Las redes más dispersas suelen tener un mayor diámetro que las más densas al existir menos caminos entre cada par de nodos

Radio (r): Valor mínimo de excentricidad para toda la red:

$$r = \min \{E(i) : i \in V(G)\}$$

► **Densidad:** Actividad global de la red

► En redes no dirigidas:

$$\Delta = L[g(g-1)/2]$$

► En redes dirigidas:

$$\Delta = L / g(g-1)$$

L, número de enlaces
g, número de nodos

Distancia media ($\langle d \rangle$) para un grafo dirigido:

$$\langle d \rangle = \frac{1}{2L_{\max}} \sum_{i, j \neq i} d_{ij} \quad d_{ij} \text{ es la distancia geodésica entre los nodos } i \text{ y } j$$

En un **grafo no dirigido** $d_{ij} = d_{ji}$. De este modo, sólo es necesario contar la longitud de los caminos una vez:

$$\langle d \rangle = \frac{1}{L_{\max}} \sum_{i, j > i} d_{ij}$$

La medida da una idea de cómo de lejos están los distintos actores en promedio. En SNA representa la **eficiencia del flujo de información en la red**

Medida de red : Densidad de una red no dirigida

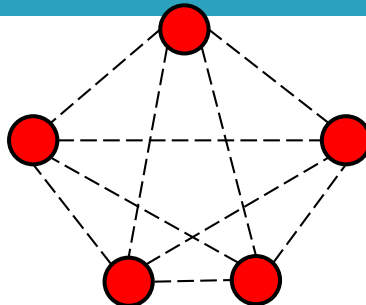
En el siguiente grafo:

Las conexiones posibles son:

10 conexiones posibles en el grafo.

Pos.= posibles

Eff . = efectivas



Medidas de red
Densidad de una red no dirigida

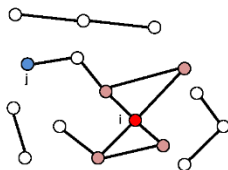
$$d = \frac{n^{\circ} \text{ effective edges}}{n^{\circ} \text{ possible edges}}$$

Eff.=0 Pos.=10 d=0	Eff.=2 Pos.=10 d=0.2	Eff.=4 Pos.=10 d=0.4	Eff.=8 Pos.=10 d=0.8	Eff.=10 Pos.=10 d=1

MEDIDAS GLOBALES

Grado medio

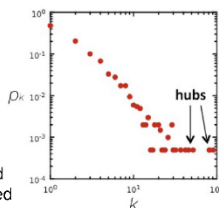
No dirigida



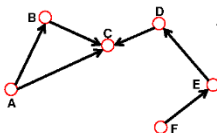
$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i$$

$$\langle k \rangle = \frac{2L}{N}$$

N – número de nodos de la red
L – número de enlaces de la red



Dirigida



$$\langle k^{in} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{in}, \quad \langle k^{out} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

$$\langle k \rangle = \frac{L}{N}$$

MEDIDAS GLOBALES

Densidad y Coeficiente de Clustering

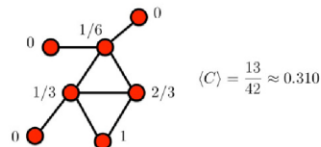
La **densidad D** = L/L_{max} mide el grado de conectividad de la red social a nivel global

El coeficiente de clustering C_i mide la **densidad local**, el ratio de vecinos de un nodo que están conectados entre sí

El **coeficiente de clustering medio** ($\langle C \rangle$) es una **medida global** que indica la probabilidad de que dos vecinos de un nodo de la red escogido aleatoriamente estén conectados entre sí:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$$



Las redes sociales reales son redes de mundos pequeños (*small worlds*) y suelen tener valores de $\langle C \rangle$ altos. Eso implica que la **transitividad** entre nodos aparece más y con más fuerza, incrementando la probabilidad de que se formen **cliques**

MEDIDAS GLOBALES

Reciprocidad

La **reciprocidad (R)** en un **grafo dirigido** mide la tendencia de pares de actores a tener conexiones mutuas entre ellos

La forma más habitual de calcularla es como un ratio entre el número de conexiones (*díadas*) mutuas (*#mut*) y el número de conexiones totales en la red, las mutuas y las asimétricas (*#asim*):

$$R = \frac{\#mut}{\#mut + \#asim}, \quad R \in [0,1]$$

Indica la probabilidad de que dos actores de la red apunten el uno al otro

Una *díada asimétrica* es un par de actores que presentan un arco en una u otra dirección pero no en ambas. Una *díada mutua* incluye los arcos en las dos direcciones

APLICACIONES

Aprendizaje Colaborativo por Ordenador (1)

Aprendizaje Colaborativo por Ordenador: énfasis en las relaciones entre los actores en un curso on-line en BSCW

Relación entre el profesor y los alumnos en un curso, así como entre los propios alumnos de distintos grupos

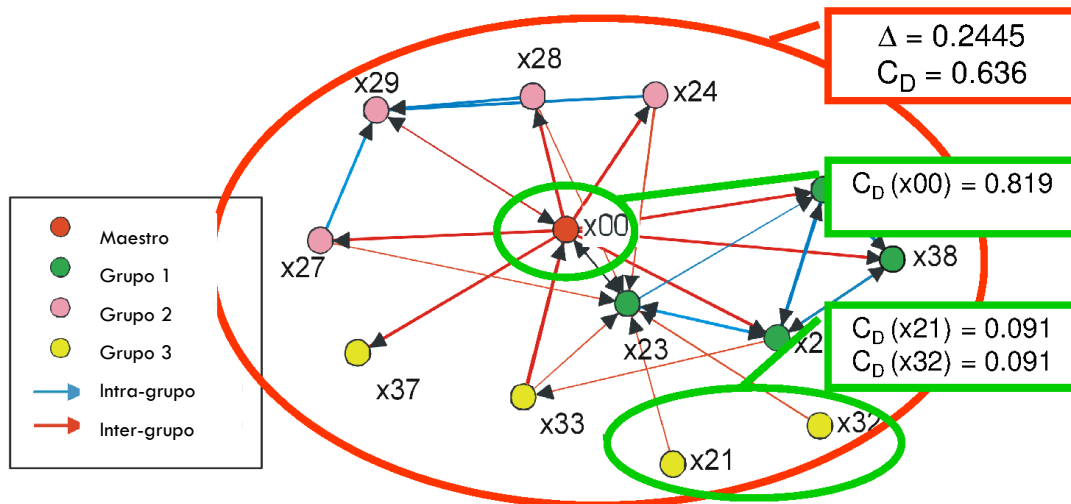
Red Social con distintas variantes: trimodal egocéntrica (tres grupos de alumnos-profesor) y trimodal, unimodal completa (todos los alumnos con todos) y trimodal completa (miembros de un grupo con los de otros)

Pregunta global: ¿Cómo ayudar a los profesores a monitorizar aspectos colaborativos de aprendizaje mediante la tecnología?

Análisis con dos medidas globales (Densidad de la red Δ y Centralización de grado C_D) y dos medidas locales (Centralidades de grado y de cercanía)

APLICACIONES

Aprendizaje Colaborativo por Ordenador (2)

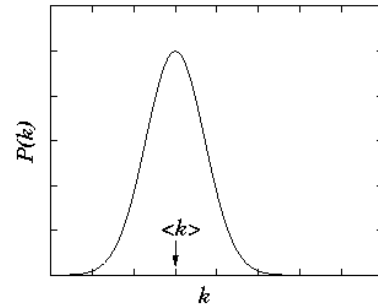


Grafos aleatorios

En Matemáticas se denomina **grafo aleatorio** a un **grafo** que es **generado** por algún **tipo de proceso aleatorio**.

- Nodos conectados al azar
- Número de aristas incidentes en cada nodo se distribuye Poisson

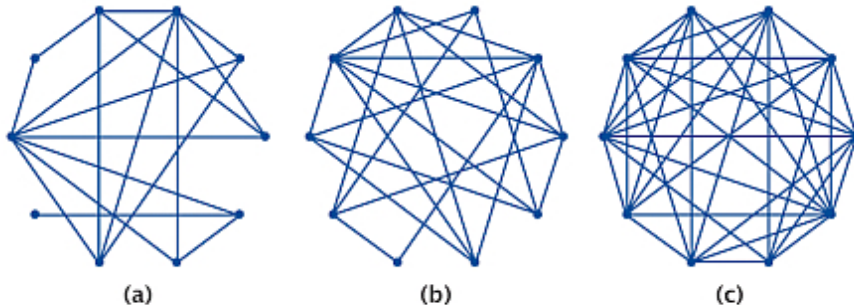
Distribución de Poisson



Grafos Aleatorios

Iniciador de la Teoría de Redes Complejas: Erdos-Renyi (50's).

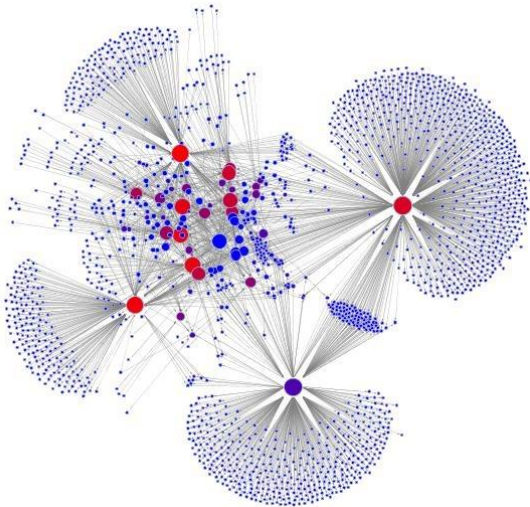
- Construye la red enlazando nodos elegidos al azar según determinada probabilidad



Grafos Aleatorios con distinta probabilidad de conexión entre pares de nodos (A) $p=0.3$, (B) $p=0.5$ y (C) $p=0.8$

En este **modelo Erdős–Rényi** se tiene que **un nuevo nodo se enlaza con igual probabilidad con el resto de la red,**

Power-law



- ❑ Nodos aparecen con el tiempo (growth model)
- ❑ Nodos prefieren unirse a nodos populares (preferential model)
- ❑ Nodos viejos Mueren
- ❑ Algunos nodos son mas sociable
- ❑ Las amistades pueden desaparecer

Propagación de epidemias

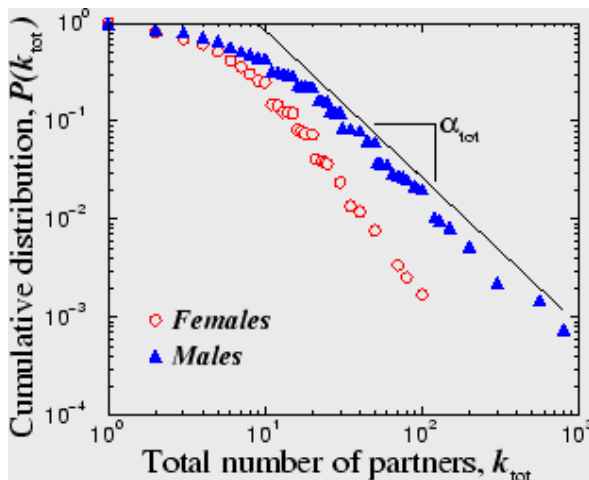
El papel de los centros en las epidemias

- En una red del tipo power-law, un virus puede persistir por más baja sea su capacidad de infección
- Muchas redes del mundo real hacen exhibir power-leyes:
 - contactos sexuales
 - redes de correo electrónico

Propagación de epidemias: **RED DE CONTACTOS SEXUALES**

Nodos: individuos

Links: relaciones sexuales



↓

HAY UNOS POCOS NODOS
CON MAYOR PROBABILIDAD
DE CONTAGIAR QUE OTROS
(HUBS)

↓

ESTRATEGIAS DE PREVENCIÓN DE
EPIDEMIAS

Difusión de información en las redes

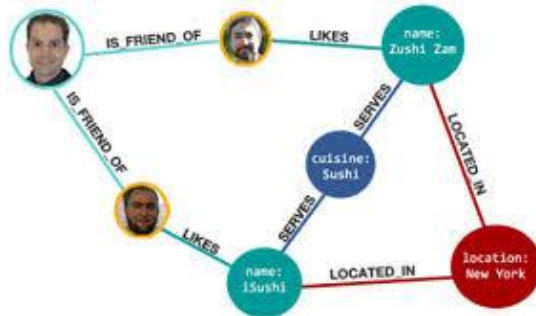
- **Factores que influyen en la difusión de información**
 - **estructura de la red:** la que se conectan los nodos?
 - **fuerza de los lazos:** qué tan fuerte son las conexiones?

- **Estudios en la difusión de la información:**
 - **Granovetter:** la fuerza de los lazos débiles
 - **J-P Onnela et al:** fuerza de los lazos intermedios
 - **Kossinets et al:** fuerza de los lazos de backbone
 - **Davis:** enclavamientos de mesa y adopción de practices

Aplicaciones

- Epidemiología (propagación de virus).
- Tolerancia frente a ataques (deliberados).
- Procesos de optimización (publicidad).
- . . .

BASES DE DATOS ORIENTADAS A GRAFOS (BDOG)

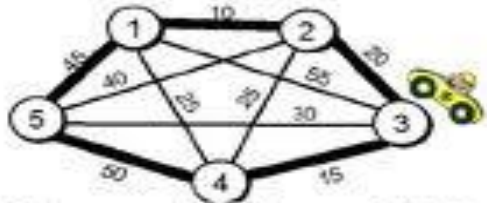


- **BDOG:**
- Representan la información como nodos de un grafo y sus relaciones con las aristas del mismo.
- Una BDOG debe estar absolutamente normalizada.
-

VENTAJAS DE UNA BDOG



Grafo que representa las redes de comunicación



Grafo que representa 5 ciudades con sus distancias.

VENTAJAS

- Consultas más amplias y no demarcadas por tablas
- No hay que definir un número determinado de atributos
- Los registros también son de longitud variable
- Se puede recorrer directamente la base de datos de forma jerárquica

MOTORES DE MODELAMIENTO GRAFICO DE UNA BDOG

[AllegroGraph](#) - Escalable y de alto rendimiento.

[Bigdata](#) - RDF/base de datos orientada a grafo.

[CloudGraph](#) - .NET usa tanto los grafos como clave/valor para almacenar los datos.

[Cytoscape](#) - Bioinformática

[DEX/Sparksee](#) - De alto rendimiento, permite escalar billones de objetos. Comercializada por [Sparsity Technologies](#).

[Filament](#)

[GraphBase](#)

Graphd, backend de [Freebase](#)

[Horton](#)

[HyperGraphDB](#) - Base de datos opensource basada en la idea de hipergrafo.

[InfiniteGraph](#)

[InfoGrid](#) - Open Source

[Neo4j](#) - Open Source.

[OrientDB](#) - Base de datos orientada a grafos y documental.

[OQGRAPH](#)

[sones GraphDB](#)

[VertexDB](#)

[Virtuoso Universal Server](#)

[R2DF](#)

Minería de Grafos

Objetivo: Desarrollar algoritmos para extraer y analizar grafos.

- Búsqueda de patrones en ellos
- Búsqueda de grupos de grafos similares (clustering)
- Construcción de modelos de predicción para las grafos (clasificación)

- Aplicaciones
 - ▣ descubrimiento motivo estructural
 - ▣ reconocimiento de proteínas
 - ▣ ingeniería inversa en VLSI
 - ▣ Mucho más ...

Por qué Minería de Grafos?

□ Los grafos son ubicuos

- ▣ Compuestos químicos (quimio-informática)
- ▣ Estructuras de las proteínas, las vías/redes biológicas (Bioinformática)
- ▣ Flujo de programas, flujo de tráfico, flujo de trabajo
- ▣ bases de datos XML, Web, de redes sociales

□ Grafos es un modelo general

- ▣ Árboles, secuencias, lazos, etc.

□ Diversidad de grafos

- ▣ Dirigidos vs. no dirigidos, etiquetados vs. no etiquetados (arcos y vértices), ponderados, con ángulos y geometrías (topológicos en 2-D/3-D)

□ La complejidad de los algoritmos: muchos problemas son de alta complejidad

Minería de Patrones de Grafos

Minería subgrafo frecuentes

- **Encontrar subgrafos frecuentes dentro de un grafo**
 - ▣ SUBDUE (DOMINAR)
- **Encontrar (sub)grafos frecuentes en un conjunto de grafos**
 - ▣ *Support* (frecuencia de ocurrencia) no inferior a un umbral mínimo
 - ▣ Enfoques basado en Apriori
 - ▣ Enfoques de crecimiento del patrón (Pattern-growt)
- **Aplicaciones de la minería de patrones de grafos**
 - ▣ Minería de estructuras bioquímicas, de flujos de programas, de estructuras XML y comunidades de la Web
 - ▣ Construcción de sistemas de clasificación, agrupación, compresión, comparación y análisis de correlación para grafos

Enfoques de Minería subgrafo frecuentes

□ Enfoques basados en Apriori

- AGM: Inokuchi, et al. (PKDD'00)
- FSG: Kuramochi and Karypis (ICDM'01)
- PATH: Vanetik and Gudes (ICDM'02, ICDM'04)
- FFSM: Huan, et al. (ICDM'03)

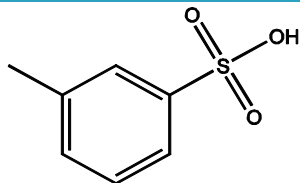
□ Enfoques de crecimiento del patrón (Pattern-growt)

- MoFa, Borgelt and Berthold (ICDM'02)
- gSpan: Yan and Han (ICDM'02)
- Gaston: Nijssen and Kok (KDD'04)

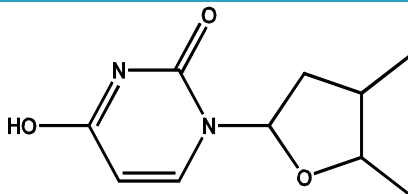
□ Minería de patrones cercanos

- CLOSEGRAPH: Yan & Han (KDD'03)

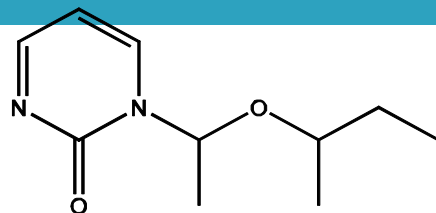
Descubrir subgrafos repetitivos (patrones) en Compuestos químicos



(A)



(B)



(C)

El soporte de un patrón dado, P, es :

$$sup_P = \sum_{G' \in D_P} S(P, G') / |D|$$

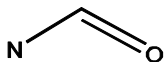
D colección de grafos

D_P representa el subconjunto de grafos en D que contienen un subgrafo aproximadamente isomorfo a P, dado el umbral de similitud τ .

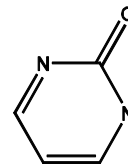
Un patrón P es frecuente si su soporte es mayor o igual que un umbral de frecuencia σ dado por el usuario.

FRECUENTES PATRONES (MIN SOPORTE ES >2)

(1)

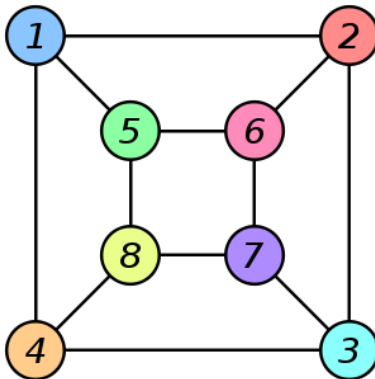
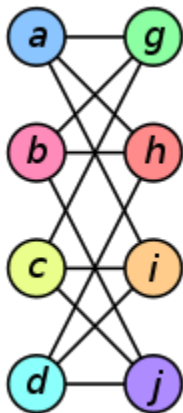


(2)



Descubrir subgrafos repetitivos (patrones)

Isomorfismo entre dos grafos G y H es una biyección F entre los conjuntos de sus vértices $F: V(G) \rightarrow V(H)$ que preserva la relación de adyacencia.



$F(a)=1$
 $F(b)=6$
 $F(c)=8$
 $F(d)=3$
 $F(g)=5$
 $F(h)=2$
 $F(i)=4$
 $F(j)=7$

Medidas de similitud basadas en los patrones de grafos

▣ Medidas de similitud basado en características

- Cada grafo se representa como un vector de características
- Vector de distancia

▣ Medida de similitud basada en la Estructura

- Subgrafo común máximo
- Grafo edita distancia: inserción, supresión, y re-etiquetado

▣ Subgrafos frecuentes y discriminativos son características de indexación de alta calidad

Datos enlazados o vinculados (Linked Data)

Datos enlazados



- Linked Data y Web de Datos
- RDF y RDFS
- Vocabularios y Ontologías en Linked Data
- Consultas RDF con SPARQL
- Aplicaciones con Linked Data

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 Linked Data?

Problema en la recuperación de la información

Text: "Pluto"



Entity Mapping
Disambiguation

Pluto

a Disney cartoon character

Pluto

a Roman god

Pluto

a song by Björk

HMS Pluto

a ship

...

Pluto

a dwarf planet

- Ambigüedad del lenguaje natural
- Diferentes palabras / expresiones para el mismo concepto (Sinónimos, metáforas)

Por qué Linked Data?

Problema en la recuperación
de la información



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

Para una comunicación exitosa:

- La información debe ser transmitida correctamente (**Sintaxis**)
- El significado (**semántica**) de la información transmitida debe interpretarse correctamente (= **comprensión**)

La **comprensión** depende de:

- El contexto tanto del remitente como del receptor

El **contexto** del emisor y del receptor depende de:

- La experiencia (**conocimiento del mundo**) tanto del remitente como del receptor

Representación del conocimiento

- How do I represent the following fact:
"Pluto has been discovered in 1930"?



UML instance

```
<a href="http://en.wikipedia.org/wiki/Pluto">  
  Pluto  
</a> has been discovered in 1930.
```

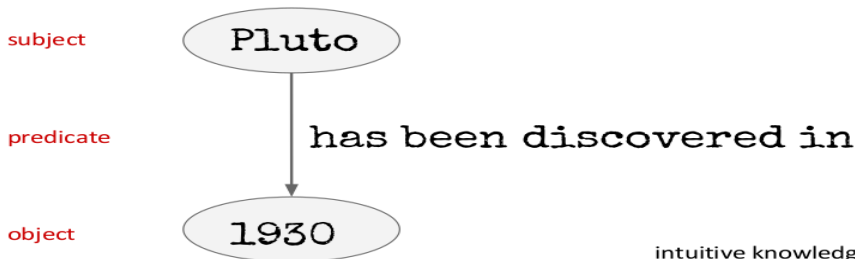
HTML

```
<planet name = "Pluto" discovered="1930" />
```

XML



- How do I represent the following fact:
"Pluto has been discovered in 1930" in an intuitive way?



intuitive knowledge representation with a **directed graph**

Representación del conocimiento

- **RDF Statements (RDF-Triple):**

Subject + Property + Object / Value

URI

URI

URI / Literal

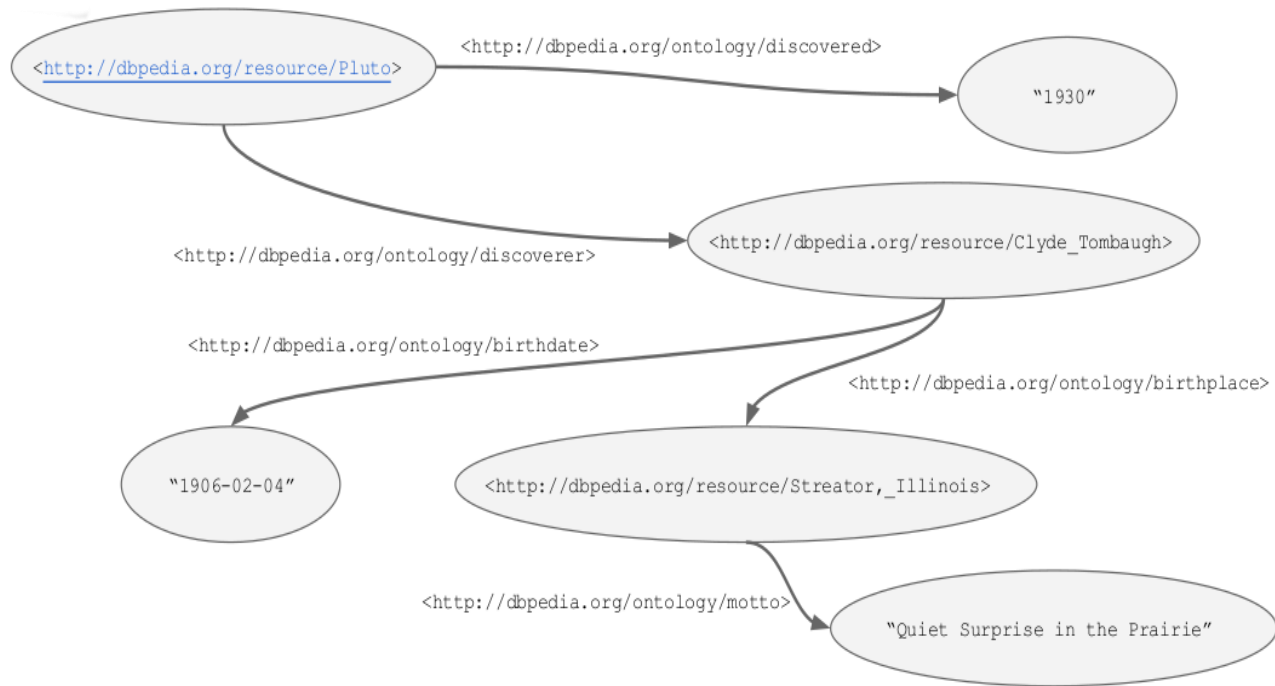
RDF Building Blocks

N-Triples Serialization

`<http://dbpedia.org/resource/Pluto>` `<http://dbpedia.org/ontology/discovered>` `"1930"` .



Representación del conocimiento



Representación del conocimiento

<http://dbpedia.org/resource/Pluto> <http://dbpedia.org/ontology/discovered> "1930" .
<http://dbpedia.org/resource/Pluto> <http://dbpedia.org/ontology/discoverer> [http://dbpedia.org/resource/Clyde Tombaugh](http://dbpedia.org/resource/Clyde_Tombaugh) .
<http://dbpedia.org/resource/Pluto> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology/CelestialBody> .
<http://dbpedia.org/resource/Pluto> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://schema.org/place> .
... ..
[http://dbpedia.org/resource/Clyde Tombaugh](http://dbpedia.org/resource/Clyde_Tombaugh) <http://dbpedia.org/ontology/birthdate> "1906-02-04" .
[http://dbpedia.org/resource/Clyde Tombaugh](http://dbpedia.org/resource/Clyde_Tombaugh) <http://dbpedia.org/ontology/birthplace> [http://dbpedia.org/resource/Streator, Illinois](http://dbpedia.org/resource/Streator,Illinois) .
... ..
[http://dbpedia.org/resource/Streator, Illinois](http://dbpedia.org/resource/Streator,Illinois) <http://dbpedia.org/ontology/motto> "Quiet Surprise in the Prairie" .
[http://dbpedia.org/resource/Streator, Illinois](http://dbpedia.org/resource/Streator,Illinois) http://www.w3.org/2003/01/geo/wgs84_pos#lat "41.120834"^^xsd:float .
[http://dbpedia.org/resource/Streator, Illinois](http://dbpedia.org/resource/Streator,Illinois) http://www.w3.org/2003/01/geo/wgs84_pos#long "-88.835281"^^xsd:float .
... ..

Subject	Property	Object
---------	----------	--------

RDF Triples

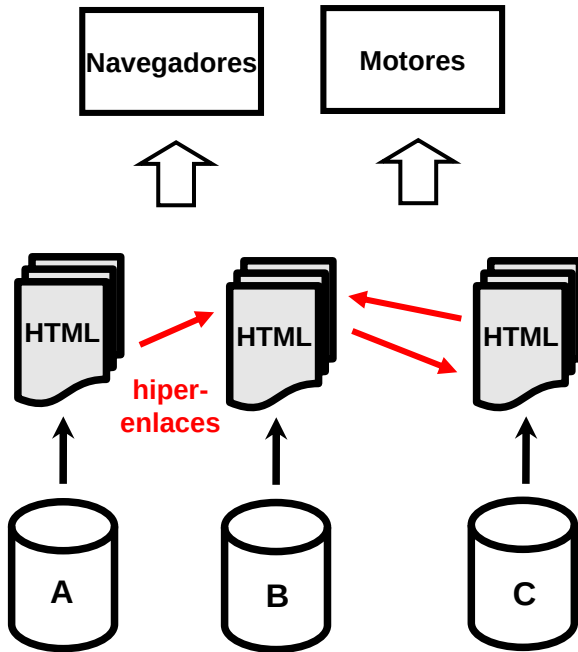
— Individuos
(Entidades)

— Clases

— Literales / Valores

— Vocabularios /
Ontologías

Clásica Web



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

Problema

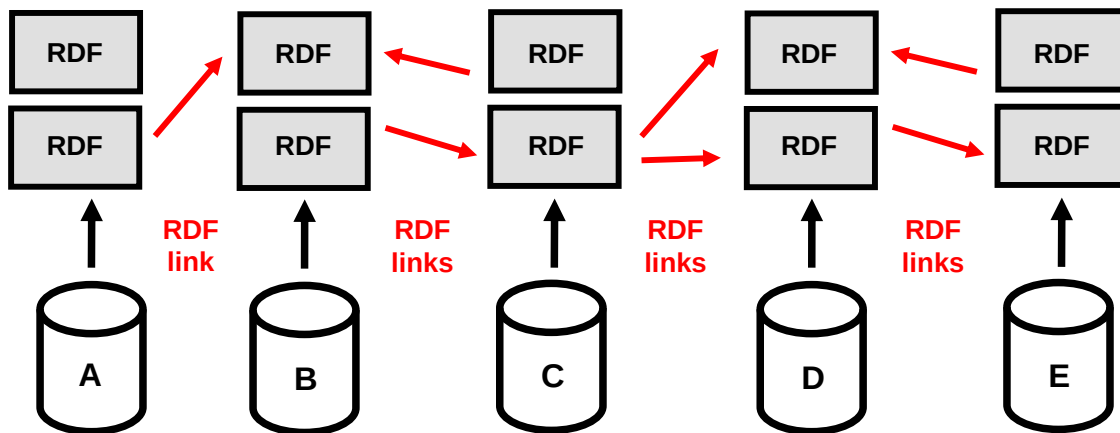
Difícil hacer cosas inteligentes con él.

Solución

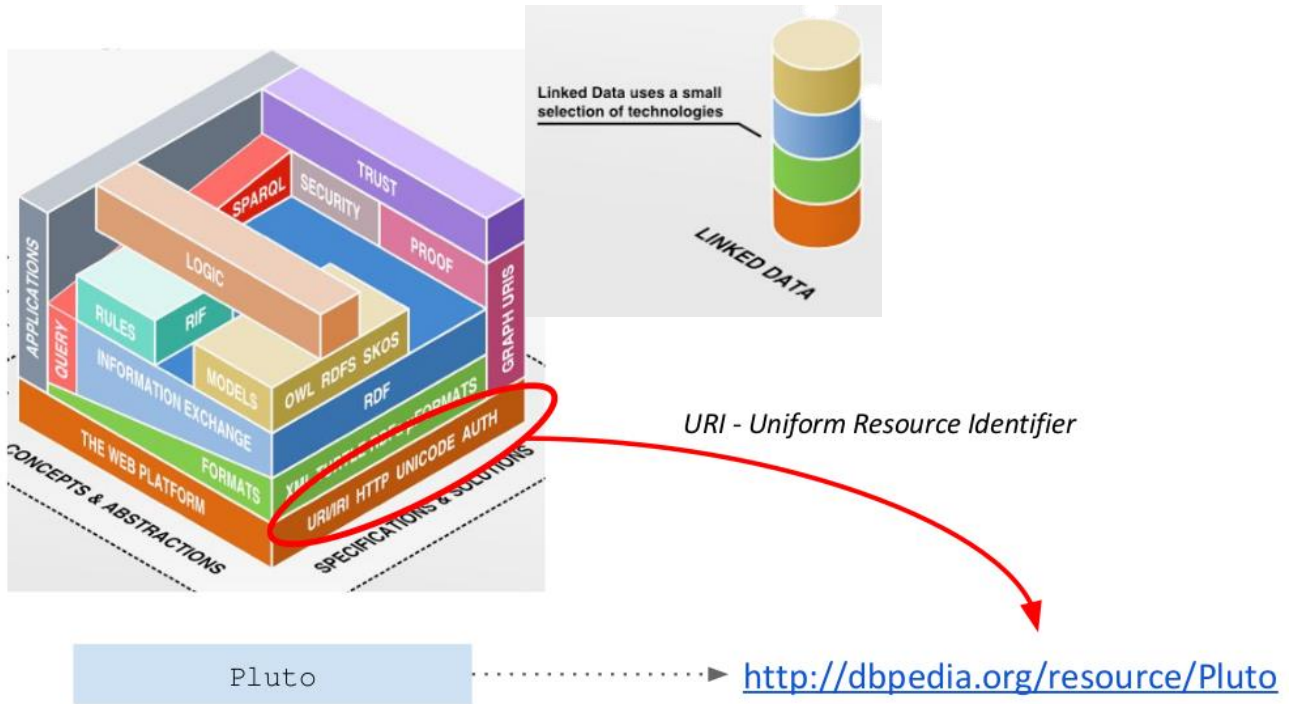
Aumentar la estructura del contenido de la Web.

Datos enlazados (Linked Data)

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



Tecnología de la Web Semántica (URI)



Tecnología de la Web Semántica (URI)

<http://en.wikipedia.org/wiki/Pluto>



Wikipedia: The free encyclopedia

Pluto

From Wikipedia, the free encyclopedia

This article is about the dwarf planet. For other uses, see Pluto (disambiguation).

Pluto (minor planet designation: **134340 Pluto**) is a dwarf planet in the Kuiper belt, a ring of bodies beyond Neptune. It was the first Kuiper belt object to be discovered, the largest and second most massive known dwarf planet in the Solar System and the tenth-largest and tenth-most massive known object directly orbiting the Sun. It is the largest known trans-Neptunian object by volume but is less massive than Eris, a dwarf planet in the scattered disc. Like other Kuiper belt objects, Pluto is primarily made of ice and rock^[a] and is relatively small—about one-sixth the mass of the Moon and one-third its volume. It has a moderately eccentric and inclined orbit during which it can be from 30 to 49 astronomical units or AU (4.4–7.3 billion km) from the Sun. This means that Pluto periodically comes closer to the Sun than Neptune, but a stable orbit—intersection with Neptune presents them from colliding. In 2014, Pluto was 30.8 AU from the Sun. Light from the Sun takes about 5.5 hours to reach Pluto at its average distance (39.4 AU).^[b]

Pluto was discovered in 1930 by Clyde Tombaugh, and was originally considered the ninth planet from the Sun. After 1992, its status as a planet led into question because the discovery of several objects of similar size in the Kuiper belt. In 2005, Eris, which is 27% more massive than Pluto, was discovered, which led the International Astronomical Union (IAU) to define the term "planet" formally for the first time the following year.^[c] This definition excluded Pluto and reclassified it as a member of the "dwarf planet" category (and specifically as a plutoid).^[d] Some astronomers think that Pluto, as well as the other dwarf planets, should be considered planets.^[e] Pluto has five known moons: Charon (the largest, with a diameter just over half that of Pluto), Styx, Nix, Kerberos, and Hydra.^[f] Pluto and Charon are sometimes considered a binary system because the barycenter of their orbits does not lie within either body.^[g] The IAU has not formalized a definition for binary dwarf planets, and Charon is officially classified as a moon of Pluto.^[h]

On 14 July 2015, the *New Horizons* spacecraft became the first spacecraft to fly by Pluto.^{[i][j][k]} During its brief flyby, *New Horizons* made detailed measurements and observations of Pluto and its moons.^[l]

Contents [hide]

- 1 History
- 1.1 Discovery
- 1.2 Name
- 1.3 Project X discovered
- 1.4 Classification
- 1.4.1 IAU classification
- 2 Orbit and rotation
- 2.1 Relationship with Neptune
- 2.1.1 Other factors
- 2.2 Rotation
- 2.3 Daylight
- 2.4 Quaoz satellite
- 3 Surface
- 3.1 Surface
- 3.2 Internal structure
- 4 Moons and rings
- 4.1 Moons
- 4.2 Satellites
- 7 Origin
- 8 Observation and exploration
- 8.1 Observation
- 8.2 Exploration
- 9 Gallery
- 10 See also

Pluto

High-resolution image composite of Pluto seen by the *New Horizons* spacecraft on 14 July 2015 from a distance of 43,500 kilometers

Discovery

Discovered by	Clyde W. Tombaugh
Discovery date	18 February 1930
Designation	134340 Pluto
MPC designation	4871 Jutic ^[a]
Provisional name	Pluto
Named after	Dwarf planet
Minor planet category	Trans-Neptunian object
	Plutoid
	Kuiper belt object
	Plutino
	Plutonian

Adjectives

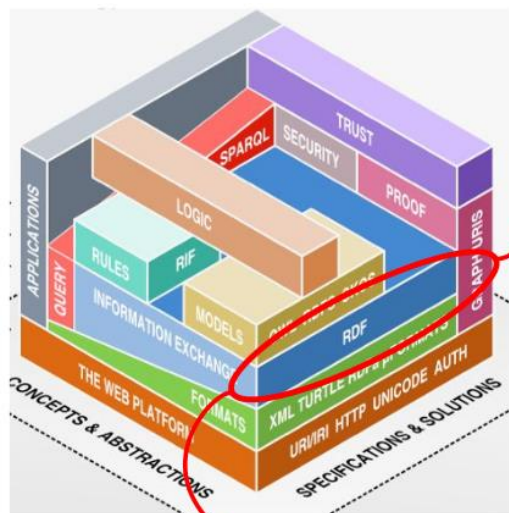
Orbital characteristics^[b]

Epoch	J2000
Aphelion	49,309 AU
Perihelion	31,917,000,000 km
	20,800 AU
	14,427,000,000 km
	21 September 1989 ^[c]
Semi-major axis	30,487 AU
	31,974,000,000 km
Eccentricity	0.14949
Orbital period	248.34 years ^[d]
	90,879 d ^[e]
	164.18 AU ³ /Plutonian solar day ^[f]
Spinodal period	98.575 days ^[g]
Average rotation	4.57 days ^[h]



<http://dbpedia.org/resource/Pluto>

Tecnología de la Web Semántica (RDF)



<http://dbpedia.org/resource/Pluto>

```
:Pluto rdf:type dbo:Planet .  
:Pluto foaf:name "Pluto"@en .  
:Pluto dbo:discoverer :Clyde_Tombaugh .  
:Pluto dbo:discovered "1930-02-18"^^xsd:date .  
:Clyde_Tombaugh rdf:type dbo:Person .  
:Clyde_Tombaugh dbo:birthdate "1906-02-04"^^xsd:date .  
...
```

RDF Resource Description Framework



RDF Triple

:Pluto

RDF Subject

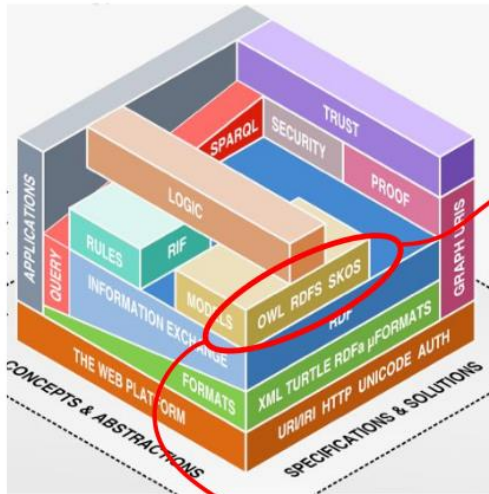
rdf:type

RDF Property

dbo:Planet .

RDF Object

Tecnología de la Web Semántica (RDFs)



<http://dbpedia.org/ontology/Planet>

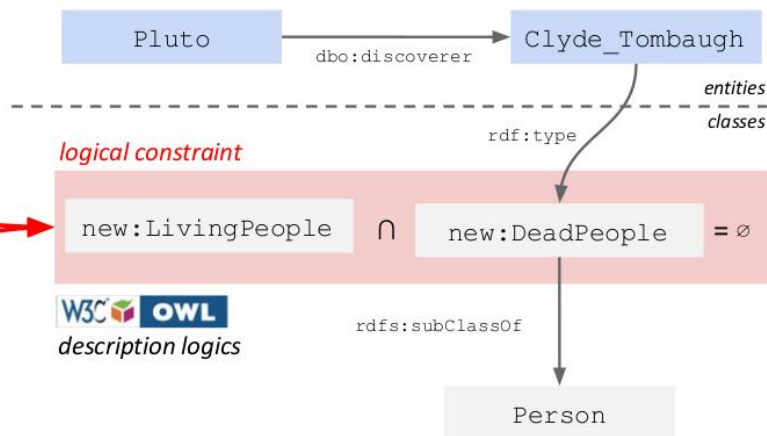
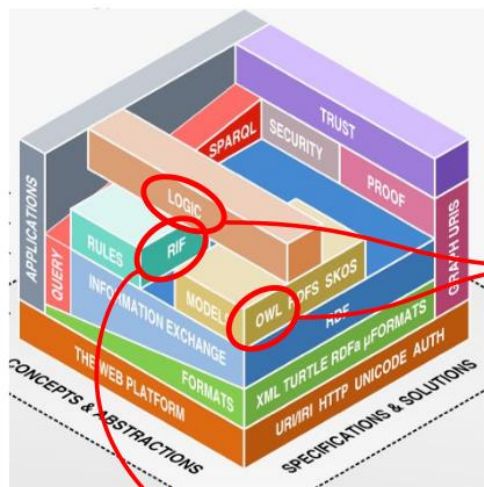
```
dbo:Planet rdf:type owl:class .
dbo:Planet rdfs:subClassOf dbo:CelestialBody .
dbo:discovered rdf:type rdf:Property .
dbo:discovered rdfs:domain owl:Thing .
dbo:discovered rdfs:range xsd:date .
dbo:discoverer rdf:type rdf:Property .
dbo:discoverer rdfs:domain owl:Thing .
dbo:discoverer rdfs:range dbo:Person .
...
```

W3C RDFS

RDF Schema



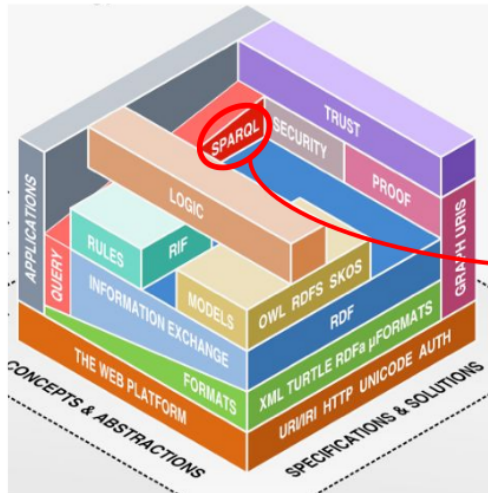
Tecnología de la Web Semántica (OWL)



+ logical rules

$$\forall x. \exists y. \text{deathDate}(x, y) \wedge \text{Person}(x) \wedge \text{Date}(y) \rightarrow \text{DeadPeople}(x)$$

Tecnología de la Web Semántica (SPARQL)



Look for all **space missions in the Solar System** which have become a **satellite** of their target

PREFIX dcterms: <<http://purl.org/dc/terms/>>

PREFIX skos: <<http://www.w3.org/2004/02/skos/core#>>

PREFIX dbp: <<http://dbpedia.org/property/>>

PREFIX dbc: <<http://dbpedia.org/resource/Category:>>

SELECT distinct ?s ?o

FROM <<http://dbpedia.org/>>

WHERE{

?s dcterms:subject/skos:broader*

dbc:Discovery_and_exploration_of_the_Solar_System ;

dbp:satelliteOf ?o .

}

Reglas para Datos enlazados (Linked Data)

- Se debe permitir:
 - ▣ Seguir 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.

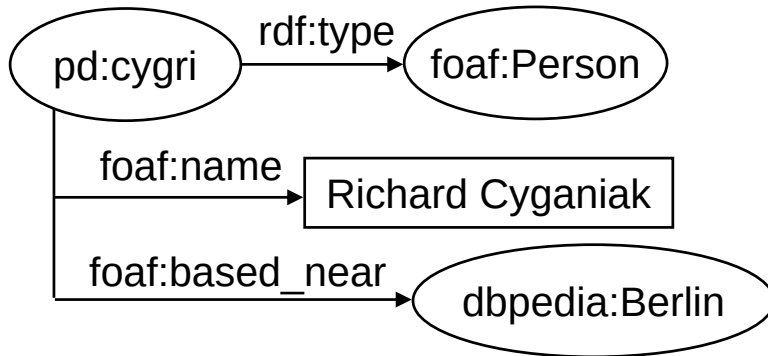
Reglas para Datos enlazados (Linked Data)

1. Utilizar **URIs** como nombre para identificar las cosas.
2. Aprovechar el **HTTP de la URIs**, para que la gente pueda **buscar** esos nombres.
3. Proporcionar **información útil**, cuando alguien busca un URI, usando los **estándares** (RDF, SPARQL) .
4. Incluir **enlaces a otros URIs**, para que puedan descubrir más cosas.

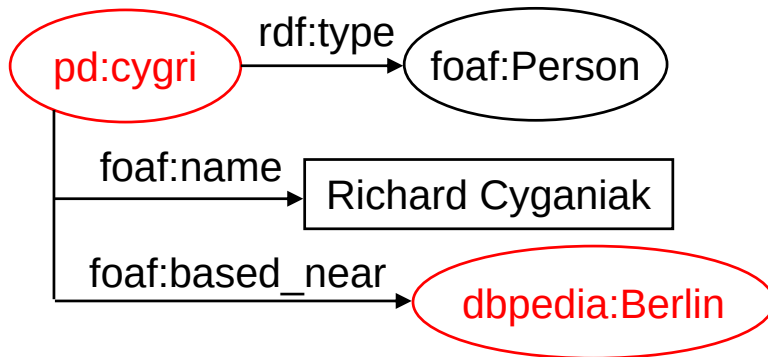
Propiedades de la Web de datos enlazados

- 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

Modelo RDF



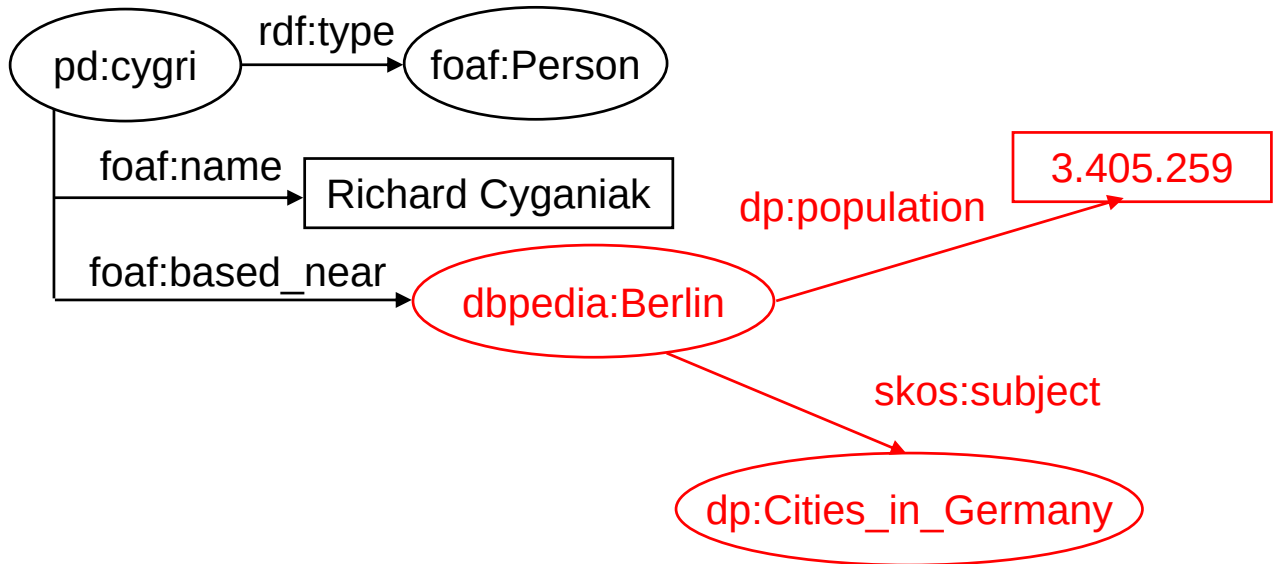
Datos son identificados con URIs http



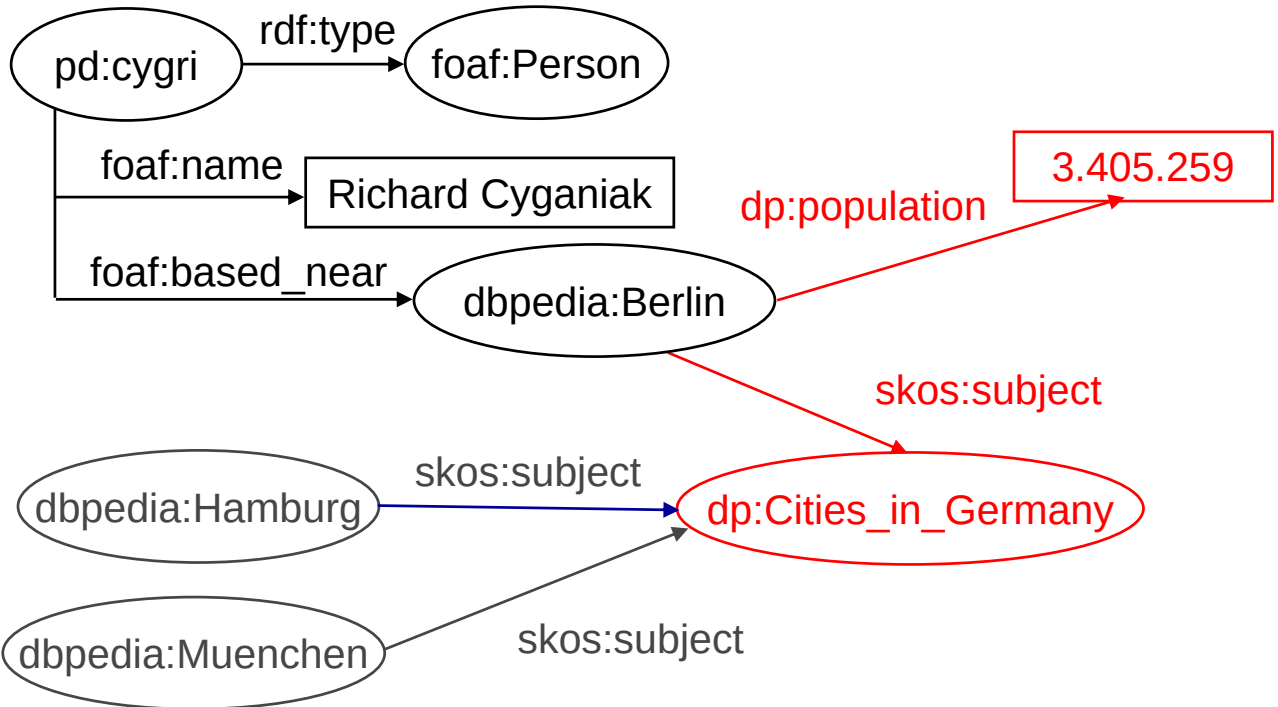
pd:cygri = <http://richard.cyganiak.de/foaf.rdf#cygri>

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

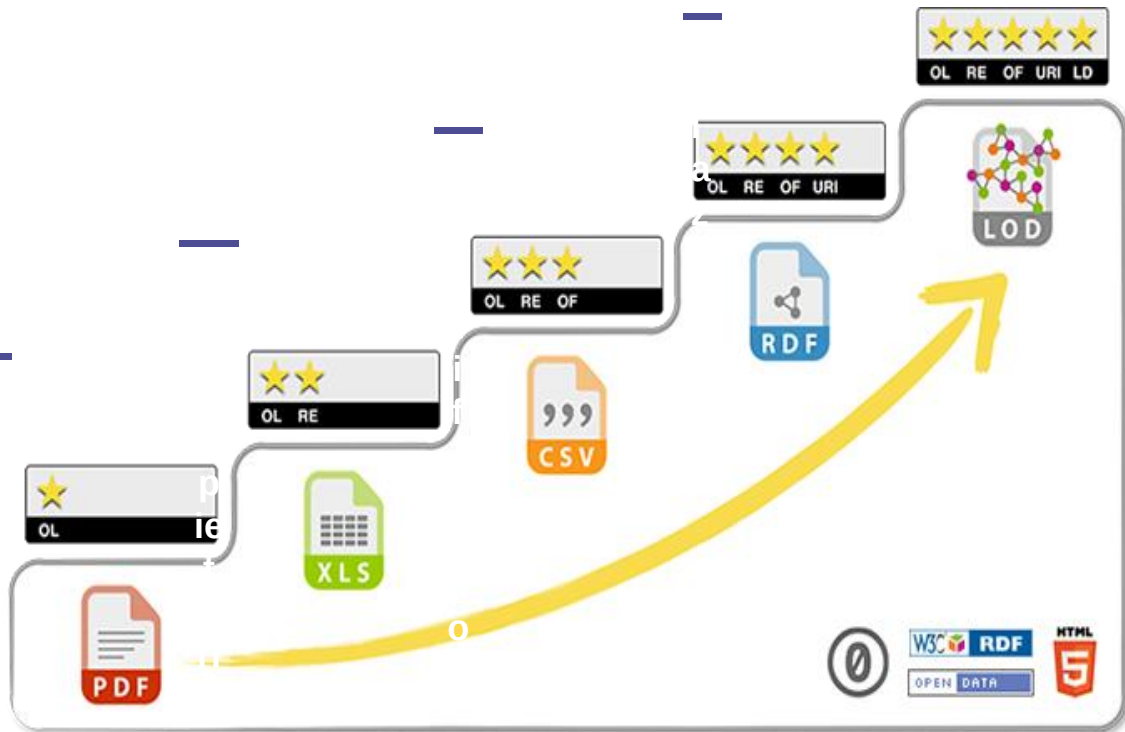
URIs en la Web



Enlazando URIs en la Web



Linked Open Data



Beneficios de utilizar Linked Data



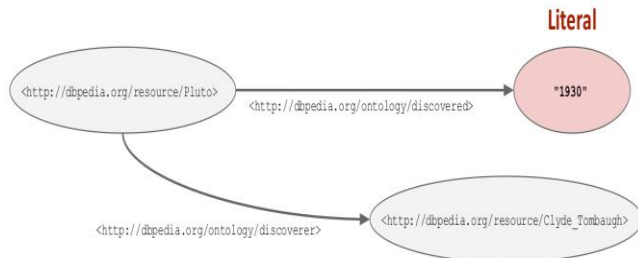
The screenshot shows the BBC Music website. At the top, there's a navigation bar with the BBC logo, a 'Sign in' button, and links for News, Sport, Weather, Shop, Earth, Travel, and More. A search bar is on the right. Below this is a 'MUSIC' section with a 'Search Music' bar and links for Tracks, Performances, Playlists, Artists, and More. A 'My Music' button is on the right. The main content area features a large image of a woman with long, dark hair, possibly a DJ, with her hair flying. Below this image is a blue banner with white text listing benefits of using Linked Data. At the bottom, there are two music player widgets. The left one shows 'Shape Of You' by Ed Sheeran, and the right one shows 'Blood Money' by Protoje. Both have play buttons and a plus sign to add to a playlist.

- La información se agrega dinámicamente a partir de datos externos disponibles públicamente (Wikipedia, MusicBrainz, Last.FM, ...)
- Sin Screen Scraping
- No requiere API especializada
- Todos los datos están disponibles como estándar abierto
- Acceso a datos a través de una simple petición HTTP
- Los datos siempre están actualizados sin interacción manual

Popular on Radio 2

Popular on Radio 3

RDF y RDFS: Literales y tipos de datos



- Typed literals can be expressed via XML Schema datatypes

- Namespace for typed literals:

<http://www.w3.org/2001/XMLSchema#>

- Examples:

"Semantics"^^<<http://www.w3.org/2001/XMLSchema#string>>

"1161.00"^^<<http://www.w3.org/2001/XMLSchema#float>>

"2015-08-02"^^<<http://www.w3.org/2001/XMLSchema#date>>

- Language Tags denote the (natural) language of the text:

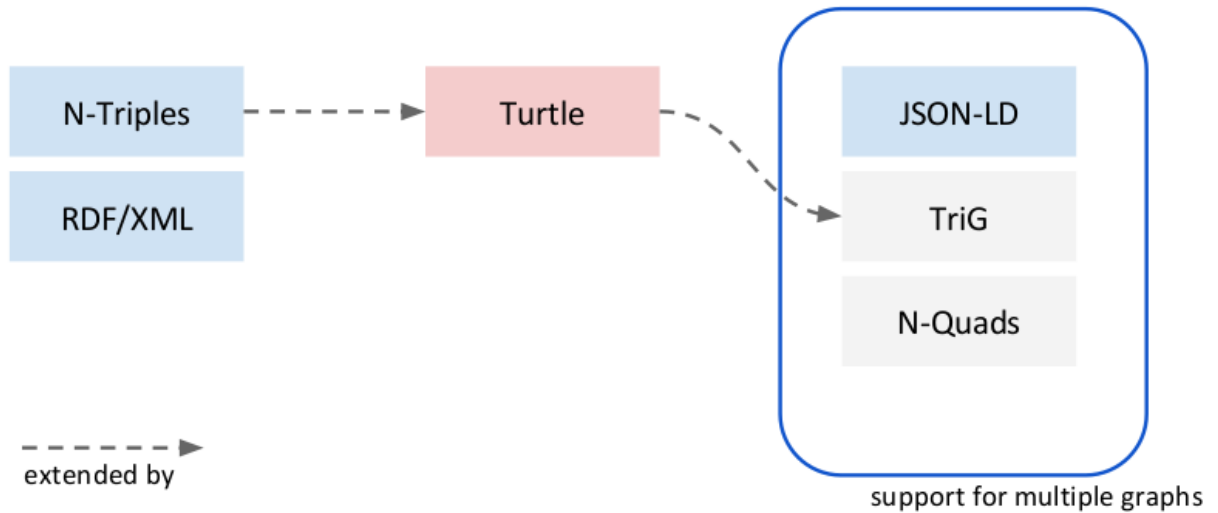
- Example:

"Semantik"@de , "Semantics"@en

	xsd:string	Character strings
Core types	xsd:boolean	true, false
	xsd:decimal	Arbitrary-precision decimal numbers
	xsd:integer	Arbitrary-size integer numbers
IEEE floating-point numbers	xsd:double	64-bit floating point numbers incl. ±Inf, ±0, NaN
	xsd:float	32-bit floating point numbers incl. ±Inf, ±0, NaN
	xsd:date	Dates (yyyy-mm-dd) with or without timezone
Time and date	xsd:time	Times (hh:mm:ss.sss...) with or without timezone
	xsd:dateTime	Date and time with or without timezone
	xsd:dateTimeStamp	Date and time with required timezone
Recurring and partial dates	xsd:gYear	Gregorian calendar year
	xsd:gMonth	Gregorian calendar month
	xsd:gDay	Gregorian calendar day of the month
	xsd:gYearMonth	Gregorian calendar year and month
	xsd:gMonthDay	Gregorian calendar month and day
	xsd:duration	Duration of time
	xsd:yearMonthDuration	Duration of time (months and years only)
Limited-range integer numbers	xsd:dayTimeDuration	Duration of time (days, hours, minutes, seconds only)
	xsd:byte	-128...+127 (8 bit)
	xsd:short	-32768...+32767 (16 bit)
	xsd:int	-2147483648...+2147483647 (32 bit)
	xsd:long	-9223372036854775808...+9223372036854775807 (64 bit)
	xsd:unsignedByte	0...255 (8 bit)
	xsd:unsignedShort	0...65535 (16 bit)
	xsd:unsignedInt	0...4294967295 (32 bit)
	xsd:unsignedLong	0...18446744073709551615 (64 bit)
	xsd:positiveInteger	Integer numbers >0
Encoded binary data	xsd:negativeInteger	Integer numbers <0
	xsd:nonNegativeInteger	Integer numbers ≥0
	xsd:nonPositiveInteger	Integer numbers ≤0
	xsd:nonZeroInteger	Integer numbers ≠0
Miscellaneous XSD types	xsd:hexBinary	Hex-encoded binary data
	xsd:base64Binary	Base64-encoded binary data
	xsd:anyURI	Absolute or relative URIs and IRIs
	xsd:language	Language tags per [BCP47]
	xsd:normalizedString	Whitespace-normalized strings
	xsd:token	Tokenized strings
	xsd:NMTOKEN	XML NMTOKENs
	xsd:Name	XML Names
	xsd:NCName	XML NCNames

RDF y RDFS: Serializaciones

- RDF comes with several different **serialization formats**:
 - N-Triples, RDF/XML, JSON, Turtle, TriG, N-Quads, RDFa, ...



RDF y RDFS: Serializaciones

- **RDF/XML Notation**

S

P

O

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ns0="http://dbpedia.org/ontology/">
  <rdf:Description rdf:about="http://dbpedia.org/resource/Pluto">
    <ns0:discovered>1930</ns0:discovered>
    <ns0:discoverer rdf:resource="http://dbpedia.org/resource/Clyde_Tombaugh"/>
  </rdf:Description>
</rdf:RDF>
```

- **N-Triples Notation**

- **URIs/IRIs** in angle brackets
- **Literals** in quotation marks
- Triple ends with a **period**

```
<http://dbpedia.org/resource/Pluto> <http://dbpedia.org/ontology/discovered> "1930" .
<http://dbpedia.org/resource/Pluto> <http://dbpedia.org/ontology/discoverer>
<http://dbpedia.org/resource/Clyde_Tombaugh> .
```

RDF y RDFS: Reificación

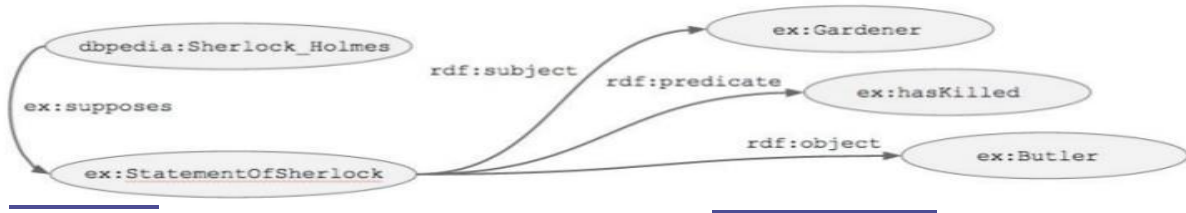
Sherlock Holmes supposes that the gardener has killed the butler

- **Part 1:** the gardener has killed the butler

`ex:Gardener ex:hasKilled ex:butler .`

- **Part 2:** Sherlock Holmes supposes...

`dbpedia:Sherlock_Holmes ex:supposes ????`



```
@prefix dbpedia: <http://dbpedia.org/resource/> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix ex: <http://example.org/CrimeStories#> .
```

```
dbpedia:SherlockHolmes ex:supposes ex:StatementOfSherlock .  
ex:StatementOfSherlock a rdf:Statement ;  
  rdf:subject      ex:Gardener ;  
  rdf:predicate    ex:hasKilled ;  
  rdf:object       ex:Butler .
```

RDFS

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
@prefix rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix :    <http://example.org/Space#> .
```

```
:Planet rdf:type      rdfs:Class ;  
        rdfs:subClassOf :CelestialBody .  
:Satellite rdf:type    rdfs:Class ;  
        rdfs:subClassOf :CelestialBody .  
:ArtificialSatellite rdf:type    rdfs:Class ;  
        rdfs:subClassOf :Satellite .
```

Class Definitions

```
:satelliteOf rdf:type  rdf:Property ;  
             rdfs:domain :CelestialBody .  
             rdfs:range  :CelestialBody .
```

Property Definitions

```
:Earth  rdf:type :Planet .  
:Moon   rdf:type :Satellite ;  
        :satelliteOf :Earth .  
:Sputnik1 rdf:type :ArtificialSatellite ;  
        :satelliteOf :Earth ;  
        rdfs:label   "Sputnik 1"@en ;  
        rdfs:comment "the first artificial Earth satellite in 1957" .
```

Instance Definitions

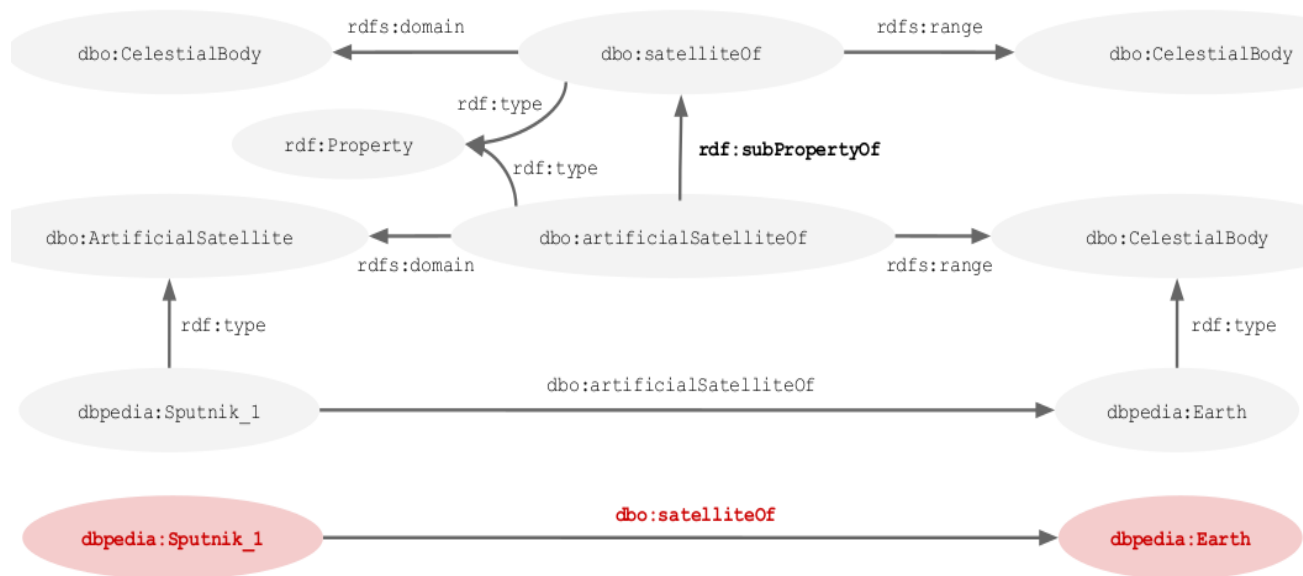
RDFS: Semántica



- Deduction of **new facts** from a **class hierarchy**.

$$\forall i, c_1, c_2: T(i, \text{rdf:type}, c_1) \wedge T(c_1, \text{rdfs:subClassOf}, c_2) \rightarrow T(i, \text{rdf:type}, c_2)$$

RDFS: Semántica



- Deduction of **new facts** from **subproperty** relationships

$$\forall i_1, i_2, p_1, p_2: T(i_1, p_1, i_2) \wedge T(p_1, \text{rdfs:subPropertyOf}, p_2) \rightarrow T(i_1, p_2, i_2)$$

Anotaciones Semánticas en la Web

En principio hay tres maneras de integrar datos estructurados con anotaciones semánticas explícitas dentro de documentos HTML



Microformatos (μ Format)



RDFa



HTML5 Microdatos (incluso schema.org)

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

Expresando Relaciones

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

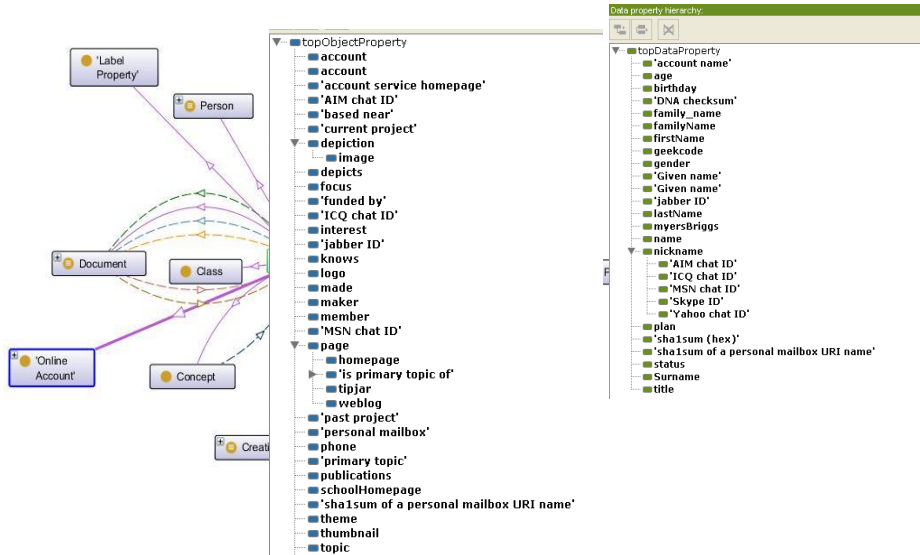
```
<?xml version="1.0" standalone="yes"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <foaf:Person>
    <foaf:name>
      Jose Aguilar
    </foaf:name>
    <foaf:mbox rdf:resource="mailto:aguilar@ula.ve"/>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

Indicando Datos Externos

Para navegar por FOAF es por medio del esquema RDFS y su propiedad **<rdfs:seeAlso>**

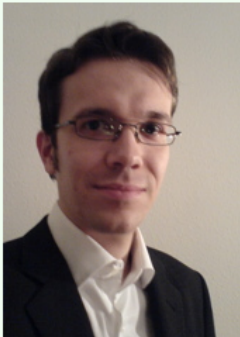
```
<?xml version="1.0" standalone="yes"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <foaf:Person>
    <foaf:name>
      Jose Aguilar
    </foaf:name>
    <foaf:mbox rdf:resource="mailto:aguilar@ula.ve"/>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



Richard Cyganiak

URI:

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

Berlin

URI:

Property	Value	Sources
population	3398888	G2
type	http://dbpedia.org/City 	G2
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.	G2
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.	G2
label	Berlin	G2
sameAs	http://sws.geonames.org/2950159/ 	G2
subject	http://dbpedia.org/resource/category/Berlin 	G2
subject	http://dbpedia.org/resource/category/Capitals_in_Europe 	G2
subject	http://dbpedia.org/resource/category/Cities_in_Germany 	G2
subject	http://dbpedia.org/resource/category/German_state_capitals 	G2
subject	http://dbpedia.org/resource/category/Host_cities_of_the_Summer_Olympic_Games 	G2
subject	http://dbpedia.org/resource/category/States_of_Germany 	G2
sourceURL	Berlin 	G1
depiction		G2
page	http://en.wikipedia.org/wiki/Berlin 	G2
is birthplace of	Adolf von Baeyer 	G2
is birthplace of	http://dbpedia.org/resource/person/Albert_Speer_%28the_younger%29 	G2

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.

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

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)

Creador de XFN 1.1

Nombre	Jose Aguilar
URL	http://www.ing.ula.ve/~aguila ☐ 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
Construir Link Restablecer	

```
<a href="http://www.ing.ula.ve/~aguila/" rel="friend met co-worker colleague">Jose Aguilar</a>
```

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">
  <span class="fn">Sally Ride</span>
  <span class="n">
    <span class="honorific-prefix">Dr.</span>
    <span class="given-name">Sally</span>
    <abbr class="additional-name">K.</abbr>
    <span class="family-name">Ride</span>
    <span class="honorific-suffix">Ph.D.</span></span>,
  <span class="nickname">sallykride</span> (IRC)
  <div class="org">Sally Ride Science</div>
  
  <a class="url" href="http://sally.example.com">u</a>,
  <a class="email" href="mailto:sally@example.com">e</a>
  <div class="tel">+1.818.555.1212</div>
  <div class="adr">
    <div class="street-address">123 Main st.</div>
    <span class="locality">Los Angeles</span>,
    <abbr class="region" title="California">CA</abbr>,
    <span class="postal-code">91316</span>
    <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	Tatiana Rodriguez
middle name	
family name	Tatiana Josefa Rodrig
organization	Universidad de Los Andes
street	
city	Merida
state/province	Merida
postal code	9101
country name	Venezuela
phone	
email	tatiana@ula.ve
url	http://tatiana.novacorp.co/
photo url	
AIM screenname	
YIM screenname	
Jabber screenname	
Categories (comma separated)	

[Resettable](#) [Build HTML](#)

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

```
<div id="hcard-Tatiana-Rodriguez-Tatiana-Josefa-Rodriguez-de-Paredes"
class="vcard">
  <a class="url" href="http://tatiana.novacorp.co/">
    <span class="given-name">Tatiana Rodriguez</span>
    <span class="additional-name">Josefa</span>
    <span class="family-name">Rodriguez de Paredes</span>
  </a>
  <div class="org">Universidad de Los Andes</div>
  <a class="email" href="mailto:tatiana@ula.ve">tatiana@ula.ve</a>
  <span class="locality">Merida</span>
  </div>
```

preview

Tatiana Rodriguez Tatiana Josefa Rodriguez de Paredes
Universidad de Los Andes
Merida, Merida, 9101 Venezuela
The hCard came with the hCard.o-matic

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)
 - rrule, rdate
 - category, description
 - uid
 - geo (latitude, longitude)
 - attendee (partstat, role), contact, organizer
 - attach
 - status
- editors note: this list is incomplete (an incomplete list is better)

hCalendar Creator

hCalendar-o-matic

summary

Seminario 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> : <span class="summary">Seminario de Minería Semántica</span> at <span class="location">Mérida</span>
```

```
<p style="font-size: smaller;">This <a href="http://microformats.org/wiki/hcalendar">hCalendar event</a> brought to you by the <a href="http://microformats.org/code/hcalendar/creator">hCalendar Creator</a>.</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> : <span class="summary">Seminario de Minería Semántica
```

preview

November 17, 2013 7–8am : Seminario 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



```
<p vocab="http://xmlns.com/foaf/0.1/"
  prefix="ov: http://open.vocab.org/terms/"
  resource="#harald" typeof="Person">
  My name is
  <span property="name">Harald Sack</span>
  and you can give me a ring via
  <span property="phone">1-800-555-0527</span>.
  My favorite beverage is
  <span property="ov:preferredBeverage">Green Tea</span>
</p>
```

RDFa ejemplo

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

BEST BUY PRODUCTOS SERVICIOS OFERTAS Buscar Best Buy 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 11/1/15–12/31/15.
Conozca más >



ENVÍO GRATIS
Válido para compras de electrodomésticos grandes desde \$399. Incluye el seguro y recargo gratis.
Conozca más >

Electrodomésticos de cocina grandes >

Desde ruidosa 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 simplifiquen su vida y mejoren la apariencia de su mostrador.



Café, té y espresso Olas y cacerolas

My Account | Order Status | Customer Service | Español

BEST BUY Weekly Ad Store Locator Outlet Center Services > Gifts > Gift Me Yours

TV & VIDEO AUDIO CAM & CAM. RECORDERS COMPUTERS MOBILE PHONES & APPS MUSIC MOVIES & BOOKS VIDEO GAMES & APPS HOME & APPLIANCES

Search (All Categories) > Respond to Item # > Credit Cards Reward Zone >

Best Buy - Carbondale **store name** > **address** > **phone** > **local selections** > **open box items (28)** > **at this location** > **events** > **avatar midnight releases** > **electronics recycling**

1270 E Main St
Carbondale, IL 62901
Phone: 618-381-1700
Geo: 37.732719, -90.192314

Customer Reviews: 5.0 (2011)
Be the first to write a review.

Store Hours
Mon: 10-9; Tue: 10-9; Wed: 10-9; Thurs: 10-9; Fri: 10-9; Sat: 10-9; Sun: 11-7;
404 - 410, 2010
Mon: 10-9; Tues: 10-9; Wed: 10-9; Thurs: 10-9; Fri: 10-9; Sat: 10-9; Sun: Closed

Events
Avatar Midnight Release!

Electronics Recycling
Use offer electronics recycling at this and all other U.S. stores.

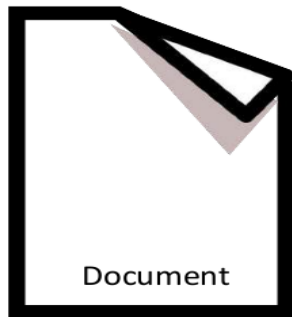
Best Buy employees entered information into the blogs every day, using online forms that output RDFa. Myers told us that the use of RDFa makes "human input from our store employees more visible on the Web."

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

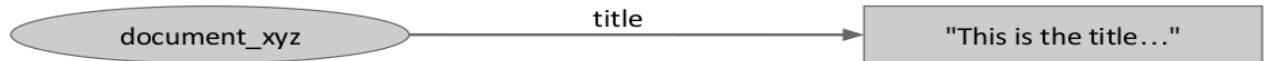
Vocabularios

Los vocabularios definen las **relaciones y términos** utilizados para **describir y representar** un área específica.

- The **Dublin Core Vocabulary** (Dublin Core Metadata Element Set)



- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

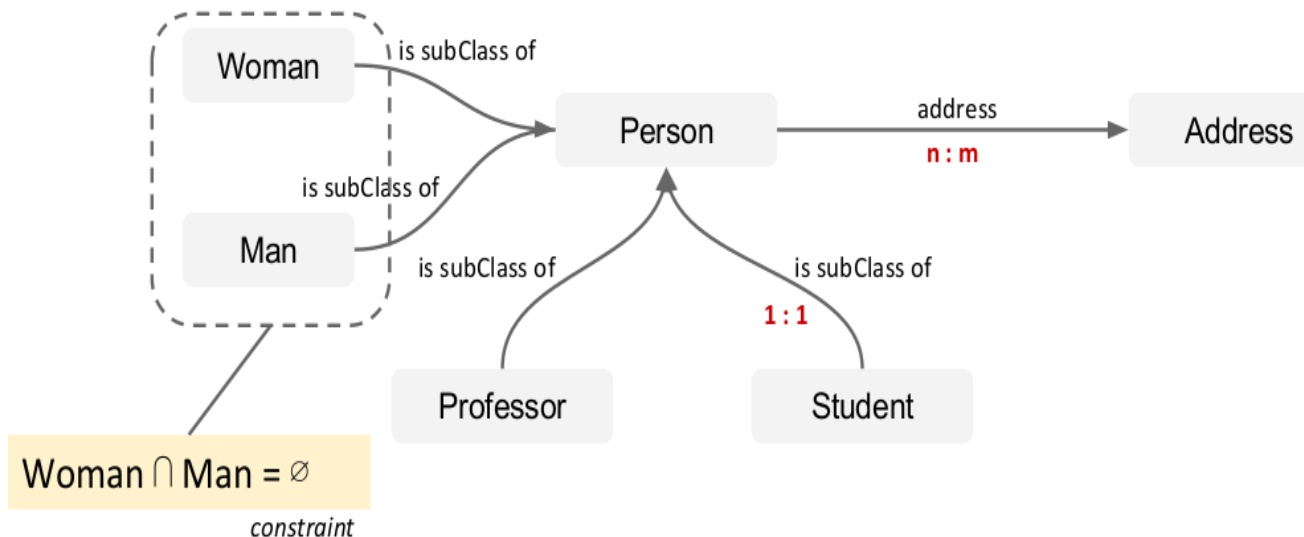


PREFIX dct: [<http://purl.org/dc/terms/>](http://purl.org/dc/terms/) .

[<http://example.org/document_xyz>](http://example.org/document_xyz) dct:title "This is the title..."@en .

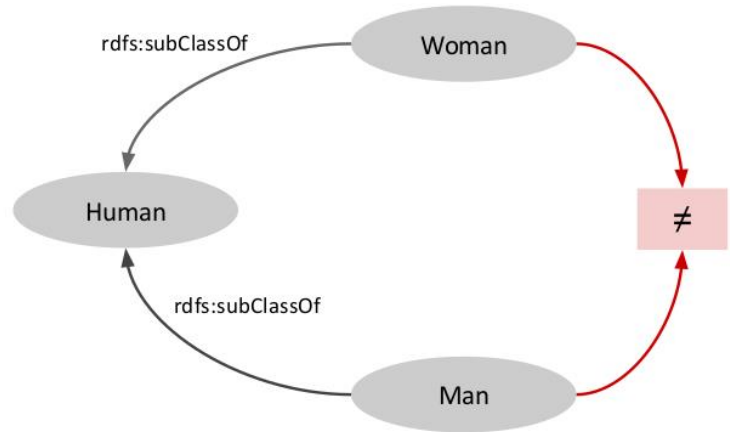
Ontologías

Una ontología es una especificación **explícita** y **formal** de una **conceptualización compartida**.



RDFs no es suficiente...

- Disjunctive Classes



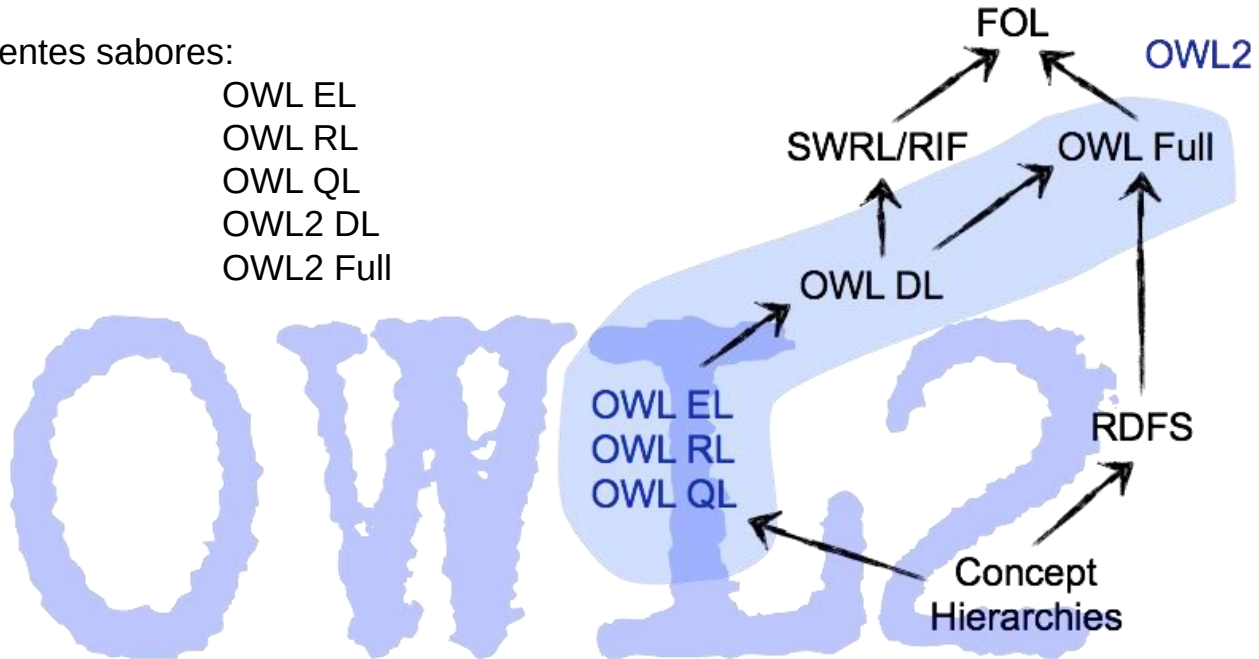
- RDFS Subclass relation cannot express disjunctive class (subclass) membership

OWL

Es un fragmento semántico de la lógica de primer orden (FOL)

Diferentes sabores:

- OWL EL
- OWL RL
- OWL QL
- OWL2 DL
- OWL2 Full



OWL

Class Expressions

- Class names A, B
- Conjunction $C \sqcap D$
- Disjunction $C \sqcup D$
- Negation $\neg C$
- Exist. property restriction $\exists R.C$
- Univ. property restriction $\forall R.C$
- Self $\exists S. \text{Self}$
- Greater-than $\geq n \ S.C$
- Less-than $\leq n \ S.C$
- Enumerated classes $\{a\}$

Properties

- Property names R, S, T
- Simple properties S, T
- Inverse properties R^{-}
- Universal property U

Tbox (Class axioms)

- Inclusion $C \sqsubseteq D$
- Equivalence $C \equiv D$

Rbox (Property Axioms)

- Inclusion $R_1 \sqsubseteq R_2$
- General Inclusion $R^{(-)}_1 \circ R^{(-)}_2 \circ \dots \circ R^{(-)}_n \sqsubseteq R$
- Transitivity
- Symmetry
- Reflexivity
- Irreflexivity
- Disjointness

Abox (Facts)

- Class membership $C(a)$
- Property relation $R(a, b)$
- Negated property relation $\neg S(a, b)$
- Equality $a=b$
- Inequality $a \neq b$

OWL

```
:Book a owl:Class .  
:Writer a owl:Class .  
:Novel a owl:Class ;  
    rdfs:subClassOf :Book .  
:Poet a owl:Class ;  
    rdfs:subClassOf :Writer .  
  
:Book owl:disjointWith :Writer .
```

Novel $\sqsubseteq$ Book
Poet $\sqsubseteq$ Writer
Book $\sqcap$ Writer $\sqsubseteq \perp$

Por inferencia, implica que **:Novel** y **:Poet** son también clases disjuntas

OWL y Ontologías en Linked Data

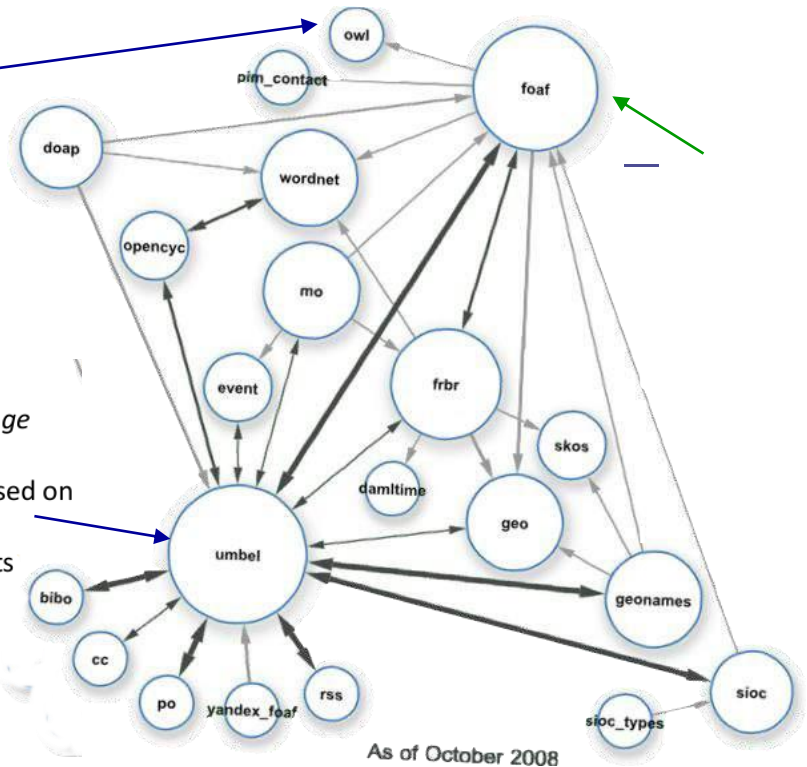
OWL

`owl:sameAs` connects identical individuals

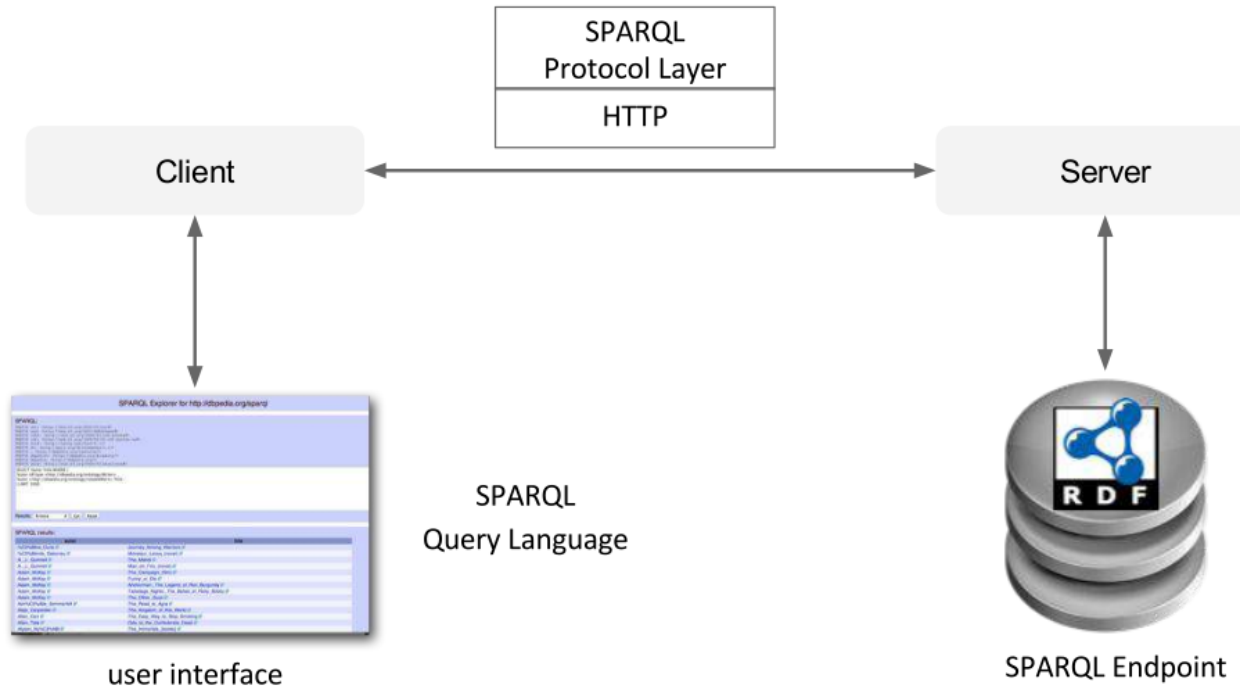
`owl:equivalentClass` connects equivalent classes

UMBEL

- „Upper Mapping and Binding Exchange Layer“
- Subset of OpenCyc as RDF Triples based on SKOS and OWL
- Upper Ontology with 28.000 concepts (`skos:Concept`)
- 46.000 Mappings into DBpedia, geonames, e.a.
(`owl:equivalentClass`,
`rdfs:subClassOf`)
- Links to more than 2m Wikipedia pages



SPARQL



SPARQL

- SPARQL **Variables** are bound to RDF terms
 - e.g. **?title, ?author, ?address**
- In the same way as in SQL,
a **Query for variables** is performed via **SELECT statement**
 - e.g. **SELECT ?title ?author ?published**
- A SELECT statement returns Query Results as a **table**

SPARQL Query

?title	?author	?published
1984	George Orwell	1948
Brave New World	Aldous Huxley	1932
Fahrenheit 451	Ray Bradbury	1953

SPARQL Result

SPARQL

Graph Pattern Matching

Triple Pattern

`?country` `dbo:capital` `?capital` .

RDF Graph

`dbpedia:Venezuela` `rdf:type` `dbo:Country` .

`dbpedia:Venezuela` `dbo:capital` `dbpedia:Caracas` .

`dbpedia:Venezuela` `dbprop:language` `"Spanish"` .

`dbpedia:Germany` `rdf:type` `dbo:Country` .

`dbpedia:Germany` `dbo:capital` `"Berlin"` .

`dbpedia:Germany` `dbp:language` `"German"` .

...

SPARQL: consulta

Buscar autores y los títulos notables de sus trabajos

```
PREFIX :      <http://dbpedia.org/resource/>
PREFIX rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:   <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo:    <http://dbpedia.org/ontology/>
```

specifies namespaces

```
SELECT ?author_name ?title
```

specifies output variables

```
FROM <http://dbpedia.org/>
```

specifies graph to be queried

```
WHERE {
```

```
  ?author rdf:type dbo:Writer .
  ?author rdfs:label ?author_name .
  ?author dbo:notableWork ?work .
  ?work rdfs:label ?title .
```

*specifies graph pattern
to be matched*

```
}
```

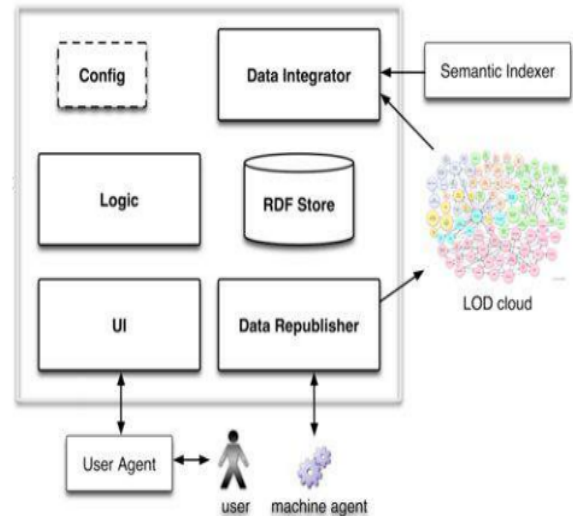
SPARQL: endpoint

- Native RDF Stores
 - Apache Jena TDB - <http://jena.apache.org>
 - AllegroGraph - <http://www.franz.com/agraph/allegrograph/>
 - GraphDB - <http://ontotext.com/products/ontotext-graphdb-owlim/>
 - Blazegraph - <https://www.blazegraph.com/>
- DBMS backed RDF Stores
 - Apache Jena SDB - <http://jena.apache.org>
 - Oracle Spatial and Graph: RDF Semantic Graph
<http://www.oracle.com/technetwork/database/options/spatialandgraph/overview/rdfsemantic-graph-1902016.html>
- Hybrid RDF Stores
 - Open Link Virtuoso - <http://virtuoso.openlinksw.com>
 - Sesame - <http://rdf4j.org/>
- W3C maintains a list of triple stores:
https://www.w3.org/2001/sw/wiki/Category:Triple_Store

Aplicaciones Web con Linked Data

- Required Components:

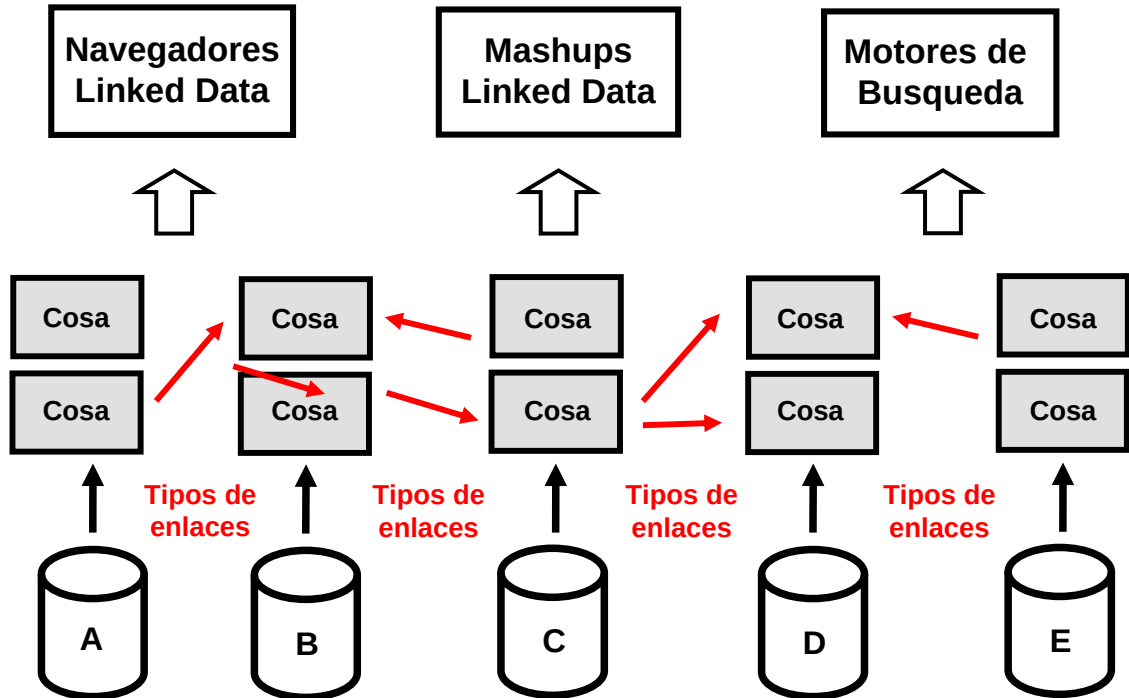
- **Local RDF Store**
 - caching of results
 - permanent storage
- **Logic** (Controller) and
- **User Interface** (=Business Logic)
 - (not LOD specific)
- **Data Integration component**
 - get data directly from LOD-Cloud or
 - via Semantic Indexer
- **Data Re-/Publishing component**
 - write back application dependent data into the Web of Data



Aplicaciones Web con Linked Data

- The easiest way is to make use of a suitable library:
 - SPARQL Javascript Library
http://www.thefigtrees.net/lee/blog/2006/04/sparql_calendar_demo_a_sparql.html
 - ARC for SPARQL (PHP)
<https://github.com/semsol/arc2/wiki>
 - dotNetRDF (C#)
<https://dotnetrdf.github.io/>
 - Jena/ARQ (Java)
<http://jena.apache.org/>
 - Sesame (Java)
<http://rdf4j.org/>
 - **SPARQL Wrapper (Python)**
<http://rdflib.github.io/sparqlwrapper/>

Aplicaciones



Aplicaciones Web con Linked Data

Extraer estructurada información desde Wikipedia

Amsterdam



The Keizersgracht at dusk

Location of Amsterdam

Coordinates:  52°22'23"N 4°53'32"E

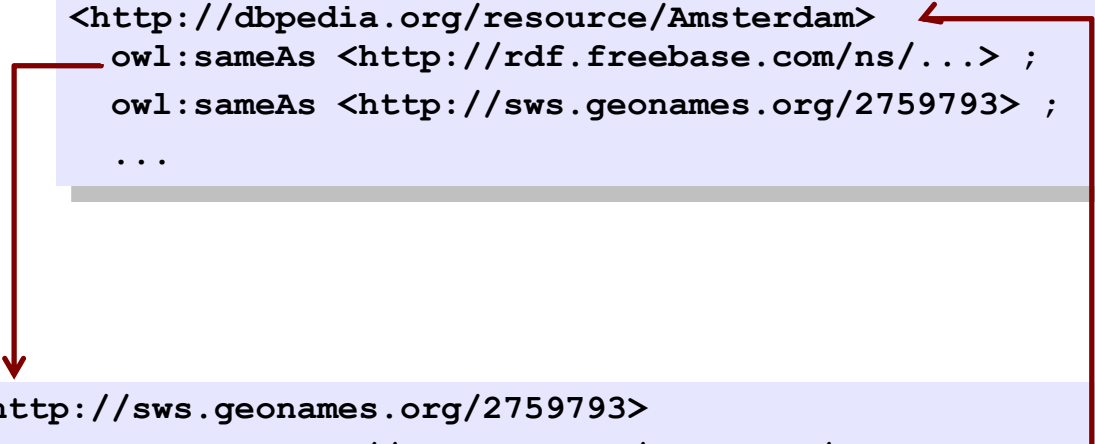
Country	Netherlands
Province	North Holland
Government	
- Type	Municipality
- Mayor	Job Cohen ^[1] (PvdA)
- Aldermen	Lodewijk Asscher Carolien Gehrels Tjeerd Herrema Maarten van Poelgeest Marijke Vos
- Secretary	Erik Gerritsen
Area ^{[2][3]}	
- City	219 km² (84.6 sq mi)
- Land	166 km² (64.1 sq mi)
- Water	53 km² (20.6 sq mi)
- Urban	1,003 km² (387.3 sq mi)
- Metro	1,815 km² (700.8 sq mi)
Elevation ^[4]	2 m (7 ft)
Population (1 October 2008) ^{[5][6]}	
- City	755,269
- Density	4,459/km² (11,548.8/sq mi)
- Urban	1,364,422
- Metro	2,158,372
- Demonym	Amsterdammer
Time zone	CET (UTC+1)
- Summer (DST)	CEST (UTC+2)
Postcodes	1011 – 1109
Area code(s)	020
Website: www.amsterdam.nl 	

```
@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

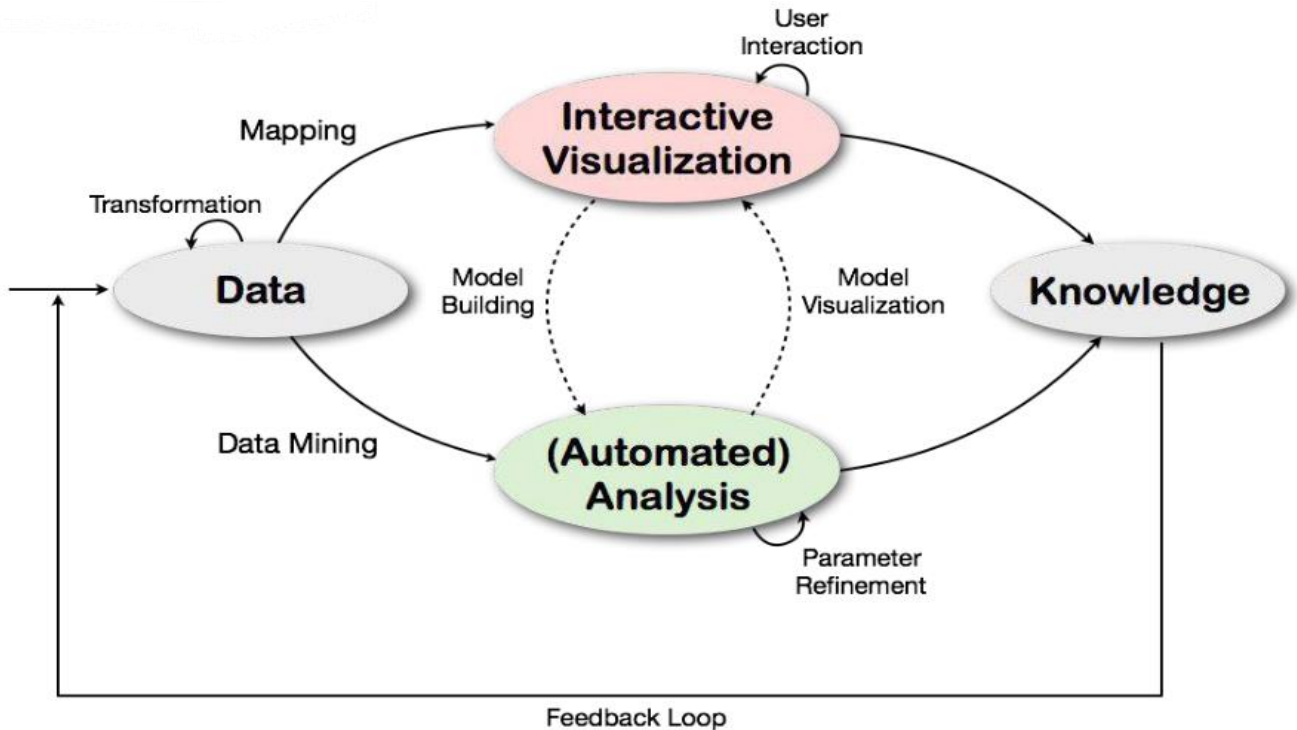
```
<http://dbpedia.org/resource/Amsterdam>  
  owl:sameAs <http://rdf.freebase.com/ns/...> ;  
  owl:sameAs <http://sws.geonames.org/2759793> ;  
  ...
```



```
<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...

Análisis en Linked Data



Análisis en Linked Data

```
PREFIX vrank:<http://purl.org/voc/vrank#>

SELECT ?author SAMPLE(?rank) as ?arank COUNT(?book) as ?bcount
FROM <http://dbpedia.org>
FROM <http://people.aifb.kit.edu/ath/#DBpedia_PageRank>
WHERE {
  ?author rdf:type dbo:Writer ;
          vrank:hasRank/vrank:rankValue ?rank .
  ?book dbo:author ?author .
} GROUP BY ?author
ORDER BY DESC(?arank)
```

author	arank	bcount
http://dbpedia.org/resource/William_Shakespeare	376.536	23
http://dbpedia.org/resource/Charles_Dickens	138.343	41
http://dbpedia.org/resource/Johann_Wolfgang_von_Goethe	123.192	7
http://dbpedia.org/resource/J. R. R. Tolkien	119.514	31
http://dbpedia.org/resource/Plutarch	104.906	1
http://dbpedia.org/resource/Voltaire	102.866	7
http://dbpedia.org/resource/Robert_Christgau	98.7989	1
http://dbpedia.org/resource/Mark_Twain	95.103	40
http://dbpedia.org/resource/Edgar_Allan_Poe	84.0183	44
http://dbpedia.org/resource/John_Milton	83.8589	2
http://dbpedia.org/resource/George_Orwell	78.8637	13
http://dbpedia.org/resource/Ovid	77.4266	1
http://dbpedia.org/resource/Leo_Tolstoy	75.5519	24
http://dbpedia.org/resource/Walter_Scott	74.7785	31
http://dbpedia.org/resource/T. S. Eliot	73.4492	4

Guardar
Resultados en
Formato cvs
Para análisis

-> authors2.csv

Análisis en Linked Data

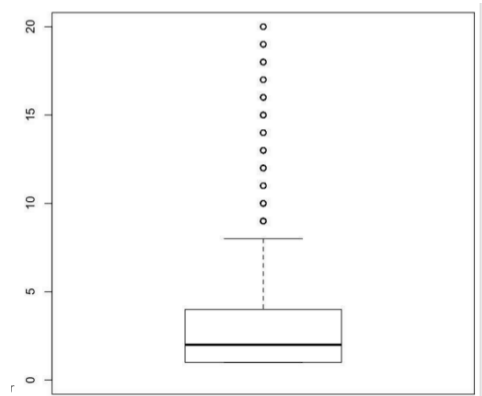
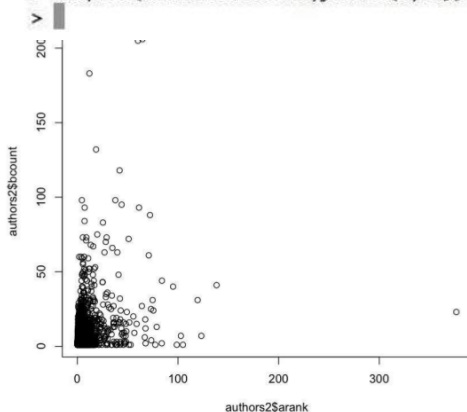
Ejemplo: numero de libros por autor

Hipotesis: los autores importantes escriben muchos libros

```
> authors2 <- read.csv("authors2.csv", header=TRUE)
> summary(authors2)
```

arank	bcount
Min. : 0.1548	Min. : 1.000
1st Qu.: 0.4566	1st Qu.: 1.000
Median : 0.9826	Median : 2.000
Mean : 3.0980	Mean : 4.238
3rd Qu.: 2.4136	3rd Qu.: 4.000
Max. : 376.5360	Max. : 206.000

```
> boxplot(authors2$bcount)
> boxplot(authors2$bcount,ylim=c(0,100))
> boxplot(authors2$bcount,ylim=c(0,20))
```



```
plot(authors2$arank, authors2$bcount)
```

```
cor(authors2$arank, authors2$bcount)
[1] 0.3560807
```

Consultas en Linked Data

Query Linked Data on the Web

Live in your browser, powered by Triple Pattern Fragments.



Choose datasources:

DBpedia 2016-04 x

Type or pick a query:

Award ceremonies in Dutch speaking countries ▼



```
SELECT DISTINCT ?artist ?common
WHERE {
  dbr:Neil_Armstrong dct:subject ?common.
  ?artist a dbo:Artist;
  dct:subject ?common.
}
```

Execute query

14 results in 1.9s

Query results:

?artist	http://dbpedia.org/resource/A._T._Mahmud
?common	http://dbpedia.org/resource/Category:1930_births

?artist	http://dbpedia.org/resource/Abbey_Lincoln
?common	http://dbpedia.org/resource/Category:1930_births

?artist	http://dbpedia.org/resource/Abdul_Rahim_Sarban
?common	http://dbpedia.org/resource/Category:1930_births

?artist	http://dbpedia.org/resource/Ada_Madsen
?common	http://dbpedia.org/resource/Category:1930_births

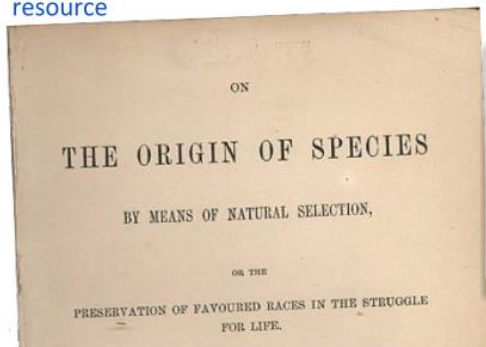
?artist	http://dbpedia.org/resource/Adel_Rootstein
?common	http://dbpedia.org/resource/Category:1930_births

?artist	http://dbpedia.org/resource/Adas_Cutler
----------------	---

Metadata y Anotación Semantica

```
<head profile="http://dublincore.org/documents/dcq-html/">
<title>Dublin Core</title>
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<meta name="DC.format" scheme="DCTERMS.IMT" content="text/html" />
<meta name="DC.type" scheme="DCTERMS.DCMItype" content="Text" />
<meta name="DC.publisher" content="Jimmy Wales" />
<meta name="DC.subject" content="Dublin Core Metadaten-Elemente. Anwendungen" />
<meta name="DC.creator" content="Björn G. Kulms" />
<meta name="DCTERMS.license" scheme="DCTERMS.URI" content="http://www.gnu.org/copyleft/fdl.html" />
<meta name="DCTERMS.rightsHolder" content="Wikimedia Foundation Inc." />
<meta name="DCTERMS.modified" scheme="DCTERMS.W3CDTF" content="2006-03-08" />
</head>
```

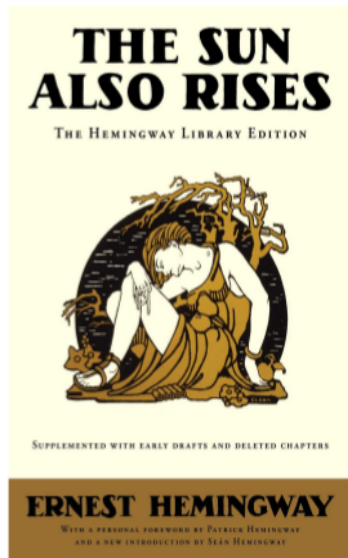
resource



On the Origin of Species, published on 24 November 1859, is a work of scientific literature by Charles Darwin which is considered to be the foundation of evolutionary biology. Its full title was *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. For the sixth edition of 1872, the short title was changed to *The Origin of Species*. Darwin's book introduced the scientific theory that populations evolve over the course of generations through a process of natural selection. It presented a body of evidence that the diversity of life arose by common descent through a branching pattern of evolution. Darwin included evidence that he had gathered on the *Beagle* expedition in the 1830s and his subsequent findings from research, correspondence, and experimentation.

unstructured metadata

Metadata y Anotación Semantica

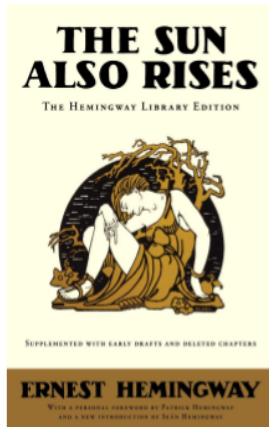


- The **semantics** of semantic metadata is explicitly and formally defined (Ontology) and therefore **machine understandable** (i.e. can be automatically read and correctly interpreted by a machine)

```
PREFIX dbr: <http://dbpedia.org/resource/> .  
PREFIX dbo: <http://dbpedia.org/ontology/> .
```

```
:9787532717071 dbo:author dbr:Ernest_Hemingway .
```

Metadata y Anotación Semantica



- **Formal Model:**

- An **annotation** A is a tuple (a_s, a_p, a_o, a_c) , where
 - a_s is the **subject** of the annotation (the annotated data)
 - a_o is the **object** of the annotation (the annotating data)
 - a_p is the **predicate** (the annotation relation) that defines the type of relationship between a_s and a_o and
 - a_c is the **context** in which the annotation is made.

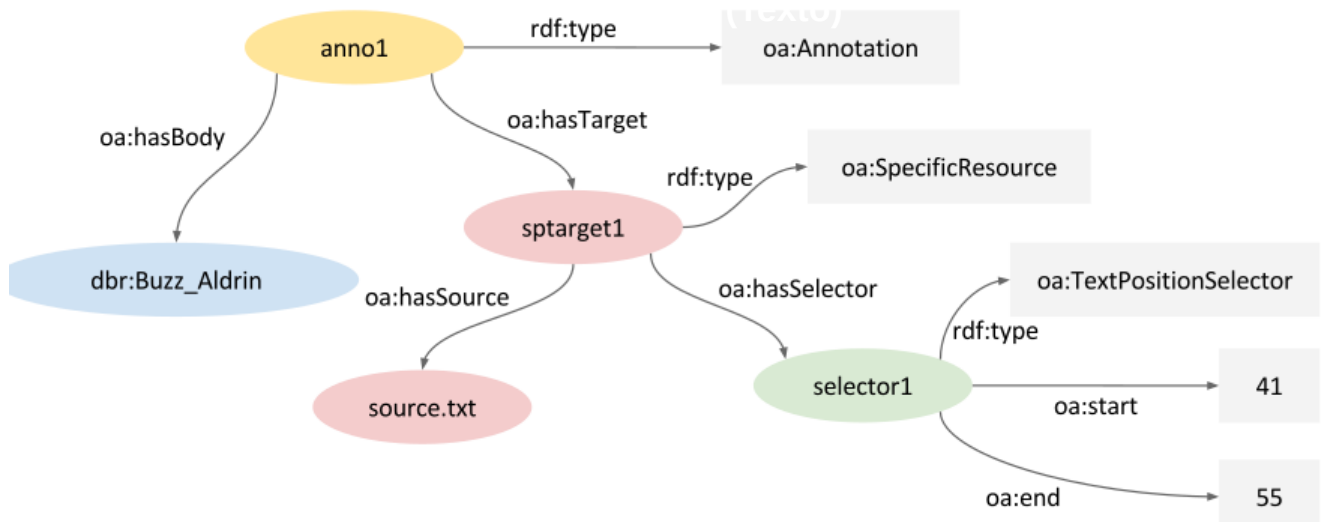
```
PREFIX dbr: <http://dbpedia.org/resource/> .  
PREFIX dbo: <http://dbpedia.org/ontology/> .  
  
.:9787532717071 dbo:author dbr:Ernest_Hemingway .
```

- created at Dec 11, 2013
- created by Lysander07
- ...

Metadata y Anotación Semantica

On July 16, 1969, Armstrong, along with **Edwin E. Aldrin, Jr.**, and Michael Collins, blasted off in the Apollo 11 vehicle toward the Moon.

Annotations

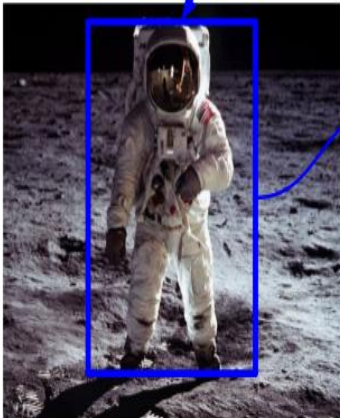


Metadata y Anotación Semántica

media fragment
<http://example.com/apollo11.ogv#t=10,20>

media fragment
<http://example.com/apollo11.ogv#t=20,30>

media fragment
<http://example.com/apollo11.ogv#t=30,44>



```
...
<div vocab="http://www.w3.org/ns/oa#"
  prefix="dctypes: http://purl.org/dc/dcmitype/
        foaf: http://xmlns.com/foaf/0.1/"
  typeof="Annotation"
  resource="#contentAnnotation-001">
  <div property="hasTarget"
    resource="http://example.com/apollo11.ogv#t=20,30&xywh=480,150,140,330"
    typeof="dctypes:video">
  </div>
  <div property="hasBody" typeof="SemanticTag">
    <a property="foaf:page" href="http://dbpedia.org/resource/Buzz_Aldrin">
      Buzz Aldrin
    </a>
  </div>
</div>
...
```

HTML with RDFa

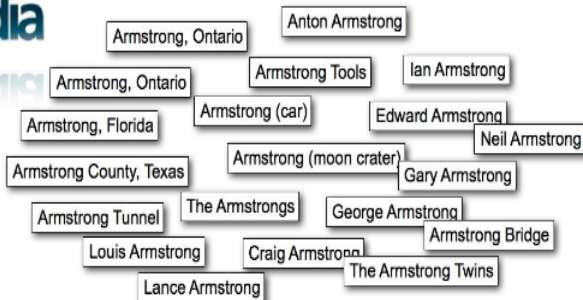
Named Entity Resolution



Named Entity Resolution

Armstrong landed the Eagle on the Moon

Determine all possible Entity Candidates



- linguistic analysis (POS tagging)
- normalization
- encoding and spelling
- special (language dependent) characters
- language dependent spellings
- abbreviations, acronyms
- type dependent spellings
- alternative names and synonyms
- fuzzy string mapping
- ...

Named Entity Resolution

Armstrong landed the Eagle on the Moon



Armstrong

448 entities

Eagle

95 entities

Moon

156 entities

George Armstrong Custer
Neil Armstrong
The Armstrong Twins
Armstrong, Florida
Craig Armstrong
Armstrong, Ontario
Armstrong (Moon Crater)
Armstrong Gun
Armstrong's Theorem
Louis Armstrong International Airport
Armstrong County, Texas
Joe Armstrong
Ian Armstrong
Armstrong Tunnel
Armstrong Automobile
Sir Thomas Armstrong
Louis Armstrong
Karen Armstrong
Armstrong (British Columbia)
Curtis Armstrong
Hilary Armstrong
William L. Armstrong

Eagle (Bird)
Eagle (heraldry)
USCGC Eagle
The Eagle (2011 film)
Eagle (comic)
Eagle (song)
The Eagle (newspaper)
War Eagle
Eagle (Moon Crater)
The Eagle (Pub)
Eagle TV
Eagle Falls (Washington)
Eagle (racehorse)
Armstrong Tunnel
John H. Eagle
Eagle (typeface)
Angela Eagle
Linda Eagle
James Philipp Eagle

Man on the Moon (film)
Moon (song)
Moon Son-Ri
C Moon
Moon 44
The Moon (Tarot card)
Man on the Moon (soundtrack)
Moon
Man on the Moon (musical)
Mr. Moon (song)
Moon (Band)
Moon OS
Moon 83
Lottie Moon
Edgar Moon
Gary Moon
William Moon
Francis Moon
Robert Charles Moon
Allan Moon
Fly me to the Moon (song)
Black Moon
Ban-Ki Moon

Named Entity Resolution

Select matching entities from all possible candidate entities:

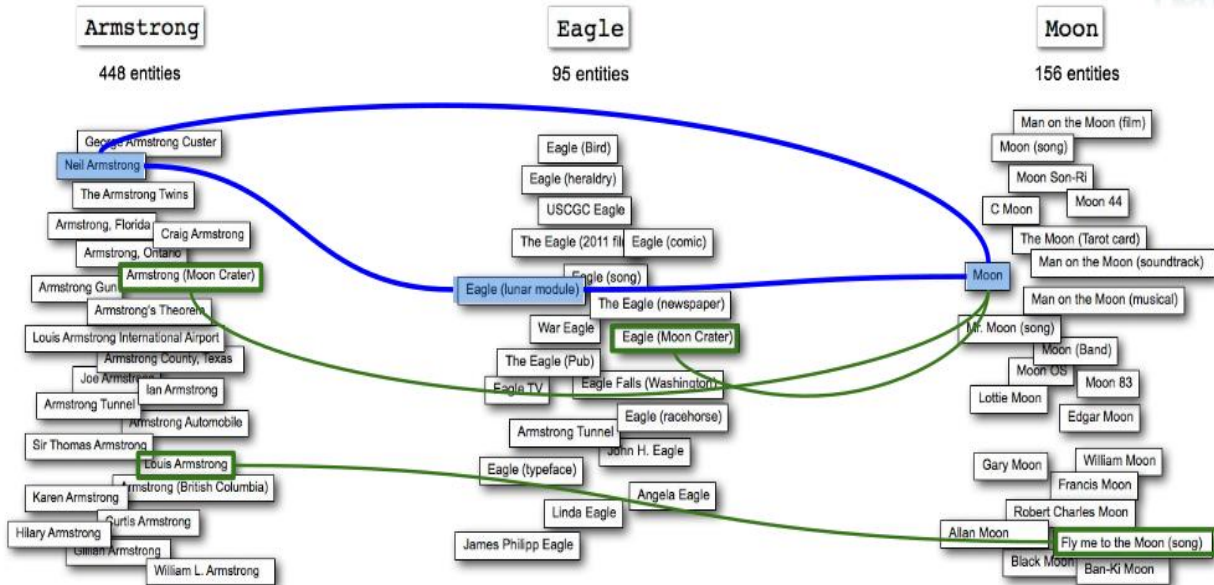
- Popularity based strategies
 - Linguistic strategies
 - Statistical strategies
 - Semantic based strategies
- 
- reference text corpus (wikipedia)
 - link graph (wikipedia)
 - semantic graph (dbpedia)

General Approach

1. Make an assumption
2. Do the strategies support or contradict your assumption
3. Make decision according to logical and probabilistic rules/constraints

Named Entity Resolution

Armstrong landed the Eagle on the Moon



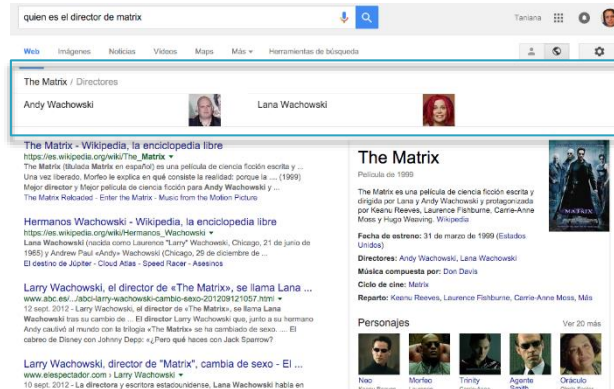
SEO (Search Engine Optimization)

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.



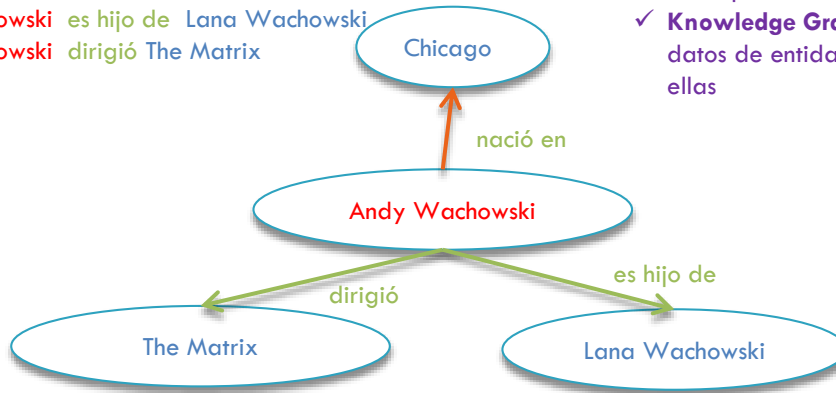
- Andy Wachowski nació en Chicago
- Andy Wachowski es hijo de Lana Wachowski
- Andy Wachowski dirigió The Matrix

- 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 datos de entidades y relaciones entre ellas

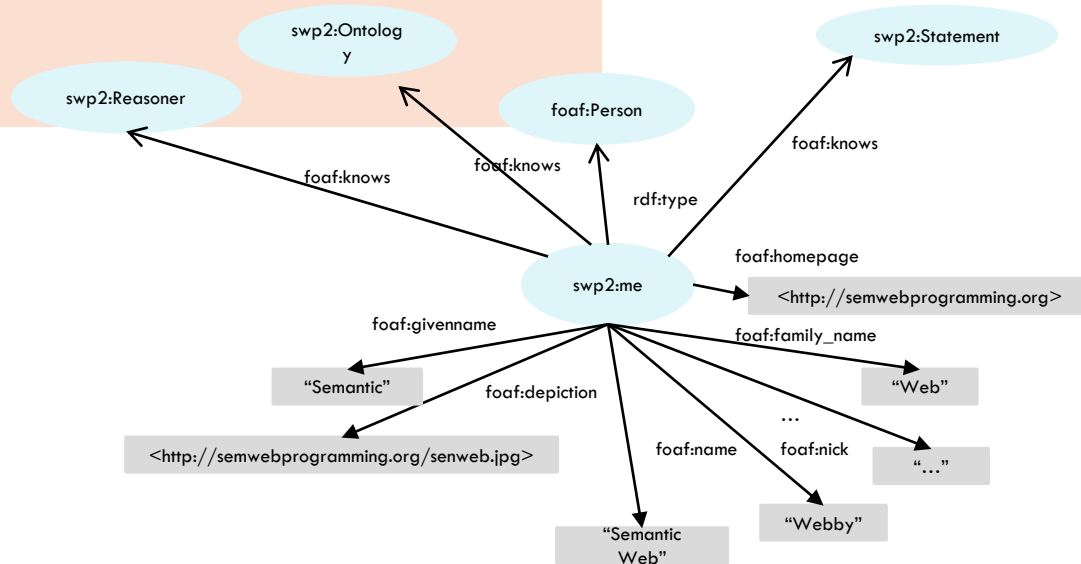
- ✓ SEO semántico tiene como objetivo de ayudar a los buscadores a entender exactamente de qué trata tus páginas.
- ✓ Para ello, sigue los siguientes pasos
 - Determinar las entidades correspondientes a la página.
 - Desambiguarlas directamente
 - Desambiguarlas indirectamente.

```

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> .

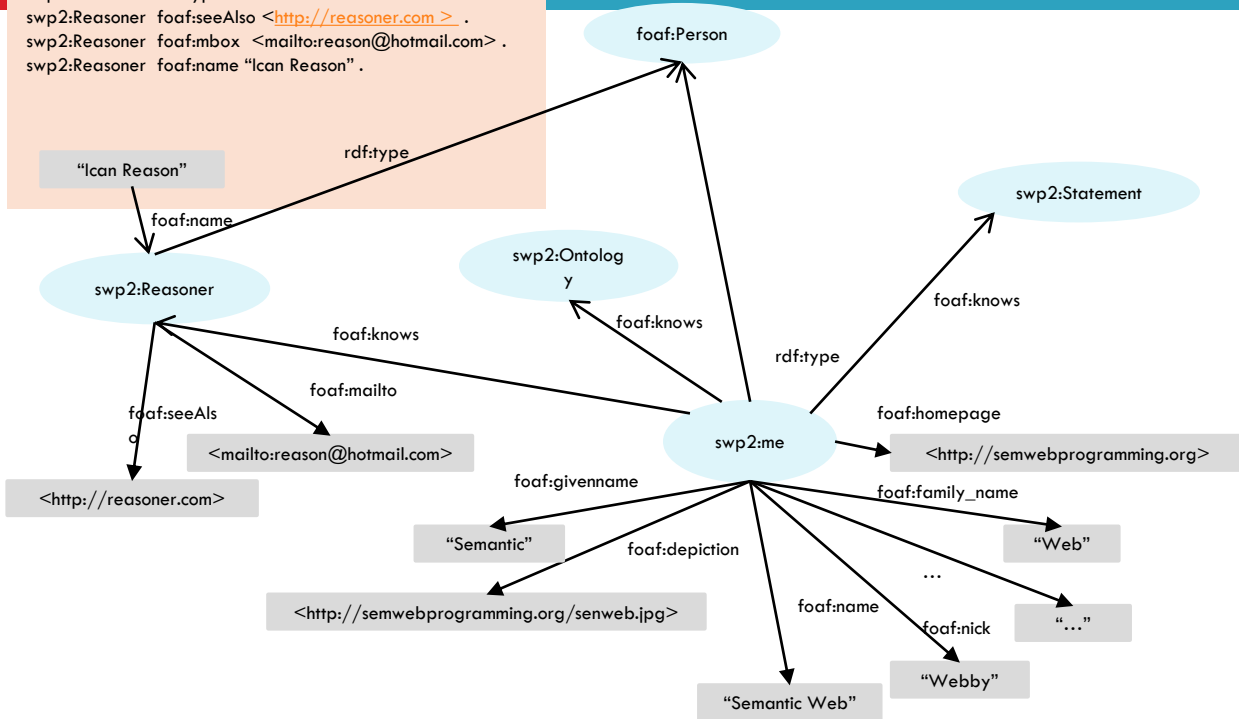
```

Ejemplo



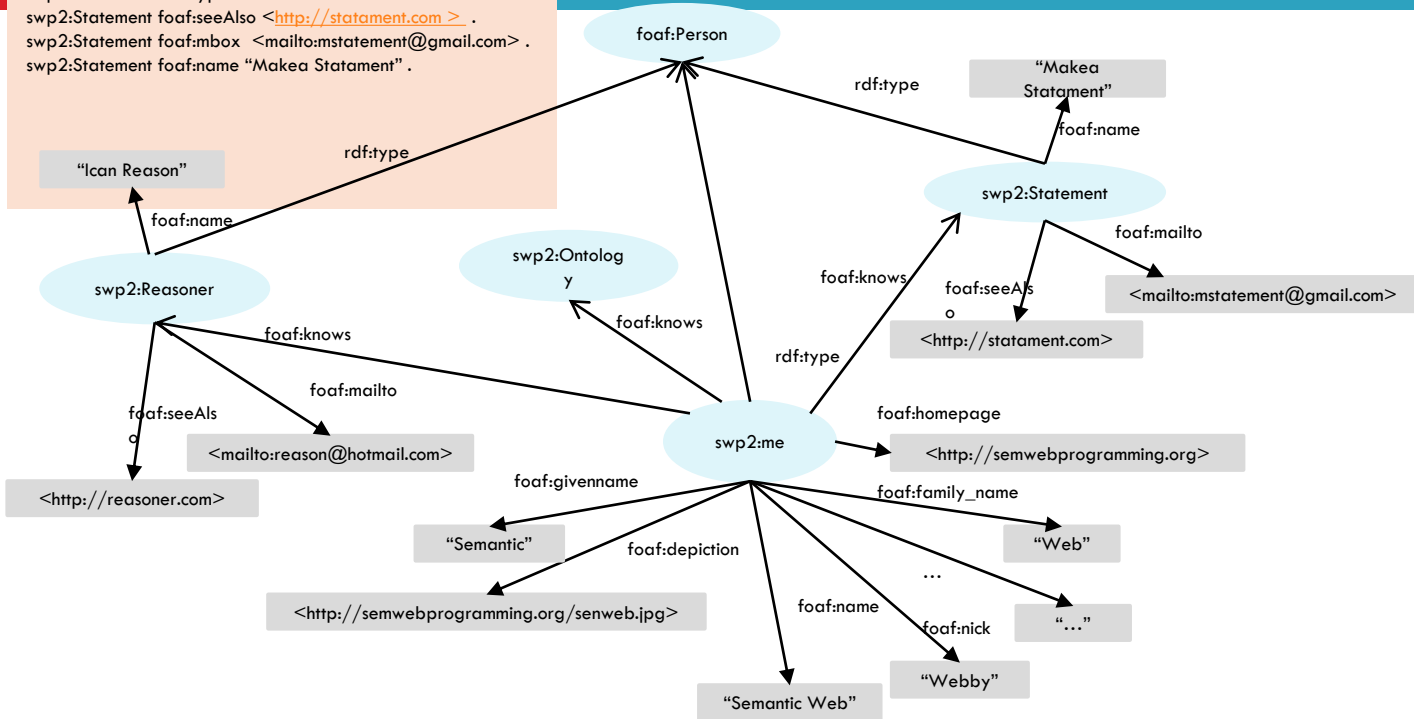
Ejemplo

```
swp2:Reasoner rdf:type foaf:Person .
swp2:Reasoner foaf:seeAlso <http://reasoner.com> .
swp2:Reasoner foaf:mbox <mailto:reason@hotmail.com> .
swp2:Reasoner foaf:name "Ican Reason" .
```



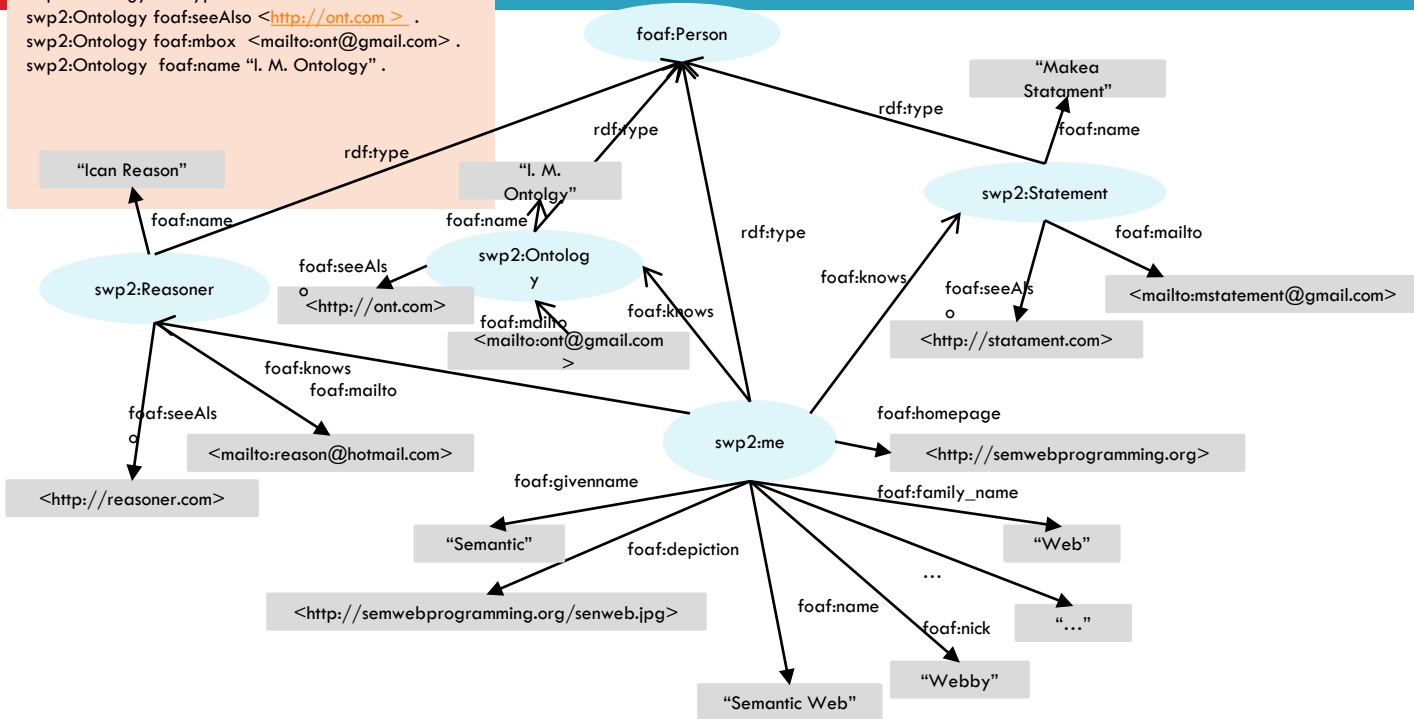
Continuación del Ejemplo

```
swp2:Statement rdf:type foaf:Person .
swp2:Statement foaf:seeAlso <http://statement.com> .
swp2:Statement foaf:mbox <mailto:mstatement@gmail.com> .
swp2:Statement foaf:name "Makea Statment" .
```



Continuación del Ejemplo

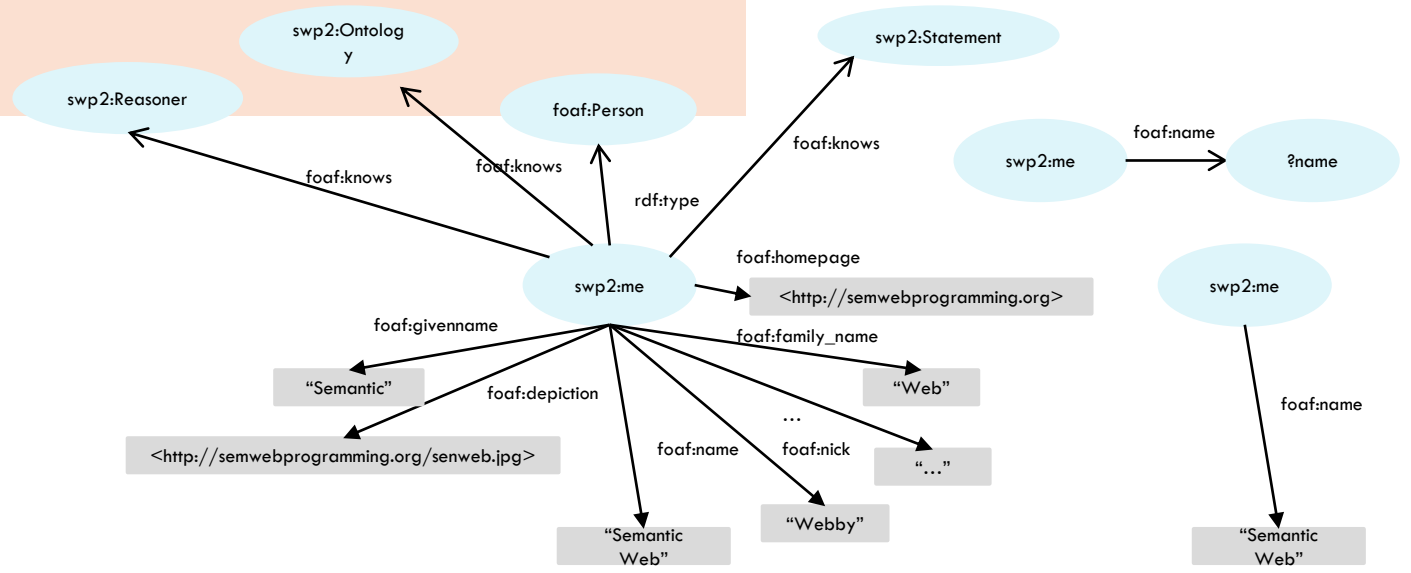
```
swp2:Ontology rdf:type foaf:Person .
swp2:Ontology foaf:seeAlso <http://ont.com> .
swp2:Ontology foaf:mbox <mailto:ont@gmail.com> .
swp2:Ontology foaf:name "I. M. Ontology" .
```



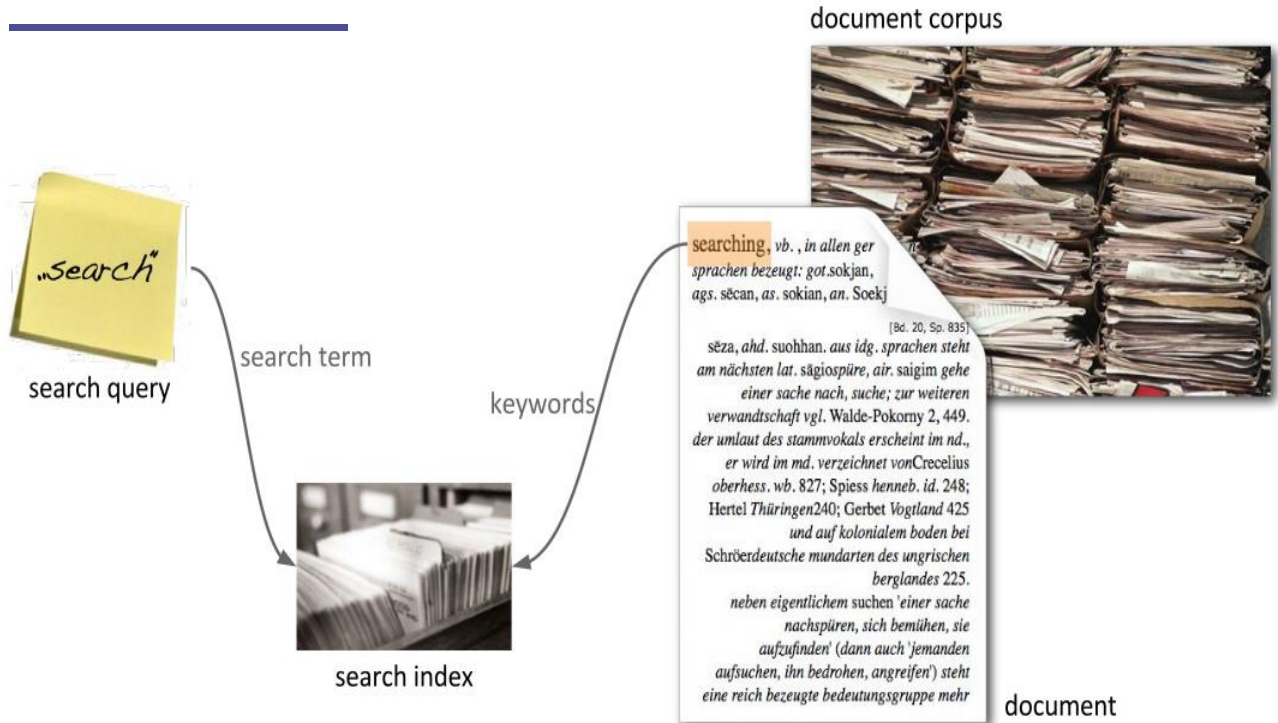

```
swp2:me rdf:type foaf:Person .
swp2:me foaf:depiction <http://semwebprogramming.org/senweb.jpg >.
swp2:me foaf:family_name "Web" .
swp2:me foaf:givenname "Semantic" .
swp2:me foaf:homepage <http://semwebprogramming.org >.
swp2:me foaf:knows "Reasoner" .
swp2:me foaf:knows "Statement" .
swp2:me foaf:knows "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 > .
```

Consulta

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



Búsqueda Semántica



Búsqueda Semántica

Search Query:

Armstrong on the Moon

Named Entity Resolution

dbr:Neil Armstrong

```
dbr: Moon
```

Indexing

The 2nd Man on the Moon

....
 Legendary astronaut **Buzz Aldrin** has revealed some captivating pieces of **Apollo 11** memorabilia on social media in the last few days.

dbr : Moon

```
dbo:Astronaut
```

```
dbr:Apollo_11
```

dbo:mission

dbr:Neil Armstrong

Entity-Based Query Matching

- simple entity matching
- similarity-based entity matching
- **relationship-based entity matching**
- ...

```

rdf:type

```

Named Entity Resolution

Búsqueda Semántica



Jules Verne (1828-1905)

On January 30, 1873, Jules Verne's famous novel "Around the World in 80 Days" (Le tour du monde en quatre-vingt jours) was published. Jules Verne's French valet, Jean Passepartout, and his friends at the London Reform Club. One day at the recent bank robbery, Phileas Fogg, but nobody knows his source of his fortune. One day at the London Reform Club, Phileas Fogg, and his newly employed French valet, Jean Passepartout, attempt to circumnavigate the world in 80 days on a £20,000 wager set by his friends at the London Reform Club.

birth year	1828
death year	1905
influenced by	Alberto Santos-Dumont
influenced	Alexandre Dumas
death place	Amiens
influenced	Antoine de Saint-Exupéry

The author of the novel "Around the World in 80 Days" is Jules Verne. He is a French writer, novelist, and playwright best known for his science fiction and adventure novels.

FOLLOW US



<http://refer.cx>

Tweets by @yovisto

yovisto.com

On Nov 4, 1847, the German composer and pianist of the Romantic era Franz Mendelssohn Bartholdy passed away. <http://blog.yovisto.com/file-mendelssohn>



Around the World in 80 Days

Permalink: <http://blog.yovisto.com/around-the-world-in-80-days/> Edit

Add Media

Paragraph



Jules Verne (1828-1905)

On January 30, 1873, Jules Verne's famous novel "Around the World in 80 Days" (Le tour du monde en quatre-vingt jours) was published by Pierre-Jules Hetzel in Paris, France. It is one of Jules Verne's most acclaimed stories, where Phileas Fogg of London and his newly employed French valet, Jean Passepartout, attempt to circumnavigate the world in 80 days on a £20,000 wager set by his friends at the London Reform Club.

Phileas Fogg lives at No. 7, Saville Row, Burlington Gardens. He's a rich bachelor, around 40 years of age, but nobody knows his source of his fortune. Overall, Phileas Fogg always lives life in clockwork precision. One day at the London Reform Club, Phileas Fogg, Flanagan, Fallentin, and Sullivan are talking about a recent bank robbery. This conversation leads to a wager on £20,000, where Fogg is quite sure he can travel around the world in eighty days, while Sullivan doesn't believe it can be done. Sullivan, Flanagan, and Fallentin think Fogg is not considering the unexpected. At the very same evening, Phileas Fogg is about to get on the train and start his expedition at 8:45 p.m. on Saturday

December 1872 together with his newly hired resourceful valet Jean Passepartout, who is caught unaware of the bet polishing shoes. Fogg tells Passepartout to pack only a few things, while everything else will be bought on the trip. The only

Publish

Status: P

Visibility:

Revision:

Publish

WP Statistics

Move to

Format

Stan

Gali

ima

Vide

Categories

All Categori

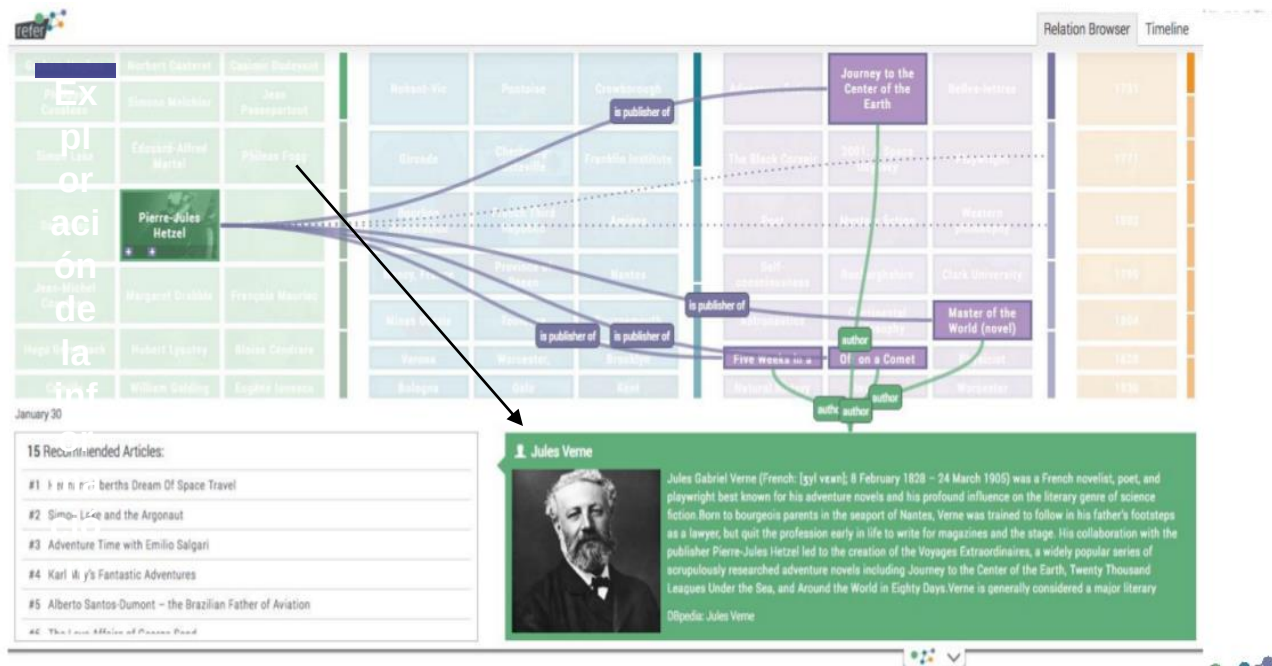
Búsqueda Semántica

es. Fogg tells Passepartout to pack only a few things, while everything else will be bought on the trip. The only will carry about is a carpet bag filled with £20,000.

ginning of the fabulous plot and from on we see how Phileas Fogg travels around the world and we witness adventures that he has together with his companion. While disembarking in Egypt, he is spotted by a dilligent detective named Fix, who has been dispatched in search of a bank robber. Because Fogg matches n of the bank robber, Fix mistakenly believes that he is the man he is looking for. To be able to secure a re, Fix goes on board the steamer crossing the Atlantic with Passepartout, but does not reveal his purpose. On the voyage, Fix gets th Passepartout, but does not reveal his purpose. Phileas and Passepartout and Fogg have Fix's true intentions, who wants to get Fogg.



Búsqueda Semántica



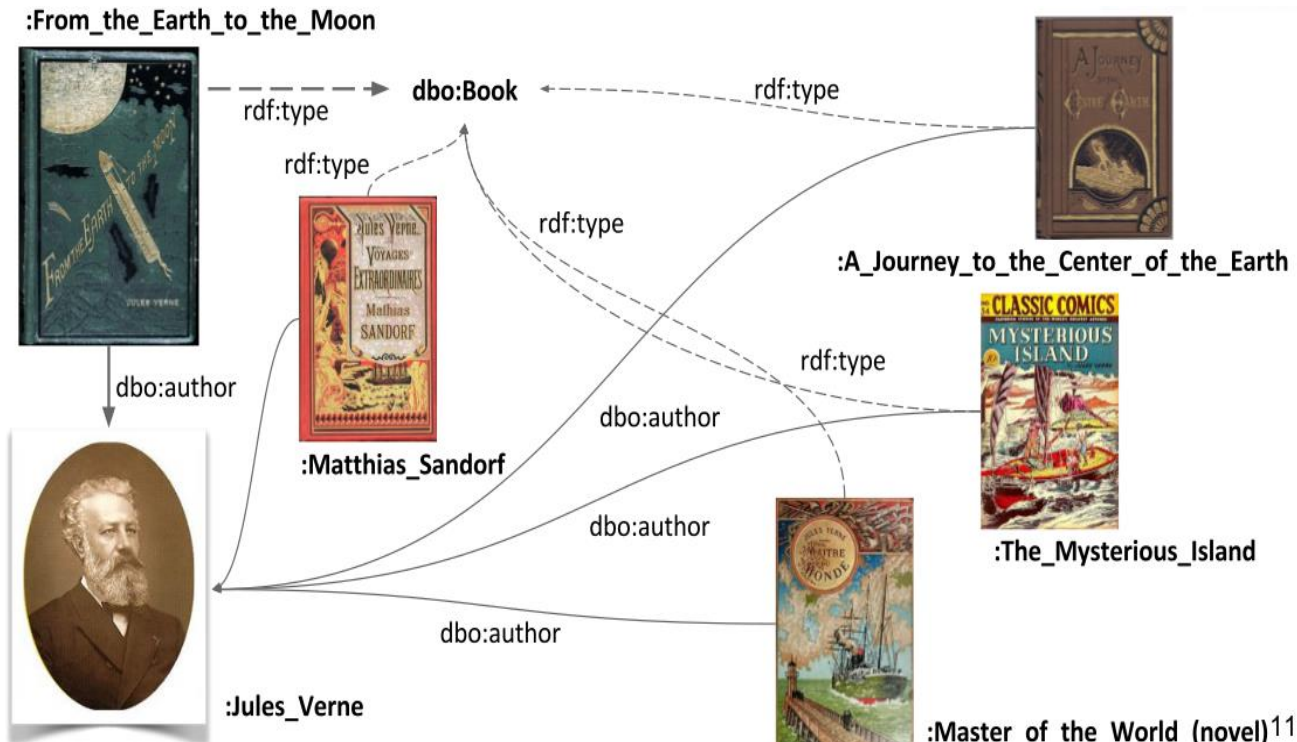
Sistemas Recomendadores



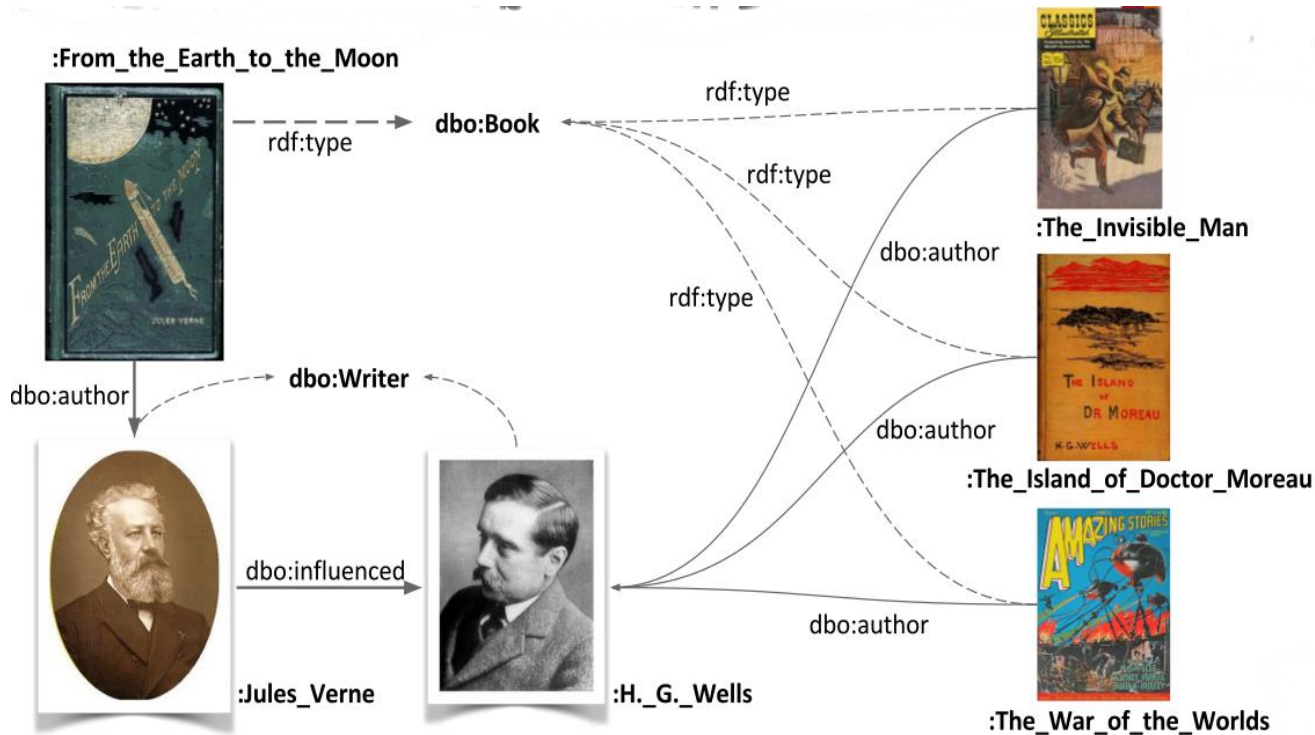
http://dbpedia.org/resource/From_the_Earth_to_the_Moon

 Browse using                              Formats                        	
About: From the Earth to the Moon	
An Entity of Type : work, from Named Graph : http://dbpedia.org , within Data Space : dbpedia.org	
From the Earth to the Moon (French: <i>De la terre à la lune</i>) is an 1865 novel by Jules Verne.	
Property	Value
dbp:abstract	- Von der Erde zum Mond ist ein Roman des französischen Autors Jules Verne. Der Roman wurde erstmals 1865 unter dem französischen Titel <i>De la Terre à la Lune</i> von dem Verleger Pierre-Jules Hetzel veröffentlicht. Die erste deutschsprachige Ausgabe erschien 1873 unter dem Titel <i>Von der Erde zum Mond</i>. Der englische Titel des Romans lautet <i>From the Earth to the Moon</i>. Es handelt sich um ein frühes Werk des Science-Fiction-Genres, das die Mondfahrt um etwa hundert Jahre vorwegnimmt. Allerdings geht es hier vor allem noch um die Vorbereitung des Abenteuers. Der Roman <i>Reise um den Mond</i> (<i>Autour de la Lune</i>) von 1870 setzte die Geschichte fort. ^[a] <i>From the Earth to the Moon</i> (French: <i>De la terre à la lune</i>) is an 1865 novel by Jules Verne. It tells the story of the Baltimore Gun Club, a post-American Civil War society of weapons enthusiasts, and their attempts to build an enormous sky-facing Columbiad space gun and launch three people—the Gun Club's president, his Philadelphia armor-making rival, and a French poet—in a projectile with the goal of a moon landing. The story is also notable in that Verne attempted to do some rough calculations as to the requirements for the cannon and, considering the comparative lack of any data on the subject at the time, some of his figures are surprisingly close to reality. However, his scenario turned out to be impractical for safe manned space travel since a much longer muzzle would have been required to reach escape velocity while limiting acceleration to survivable limits for the passengers. The character of Michel Ardan, the French member of the party in the novel, was inspired by the real-life photographer Félix Nadar. ^[a]
dbp:author	dbp:Jules_Verne
dbp:illustrator	dbp:Henri_de_Montaut
dbp:literaryGenre	dbp:Science_fiction
dbp:mediaType	dbp:Hardcover
dbp:publisher	dbp:Pierre-Jules_Hetzel
dbp:series	dbp:Voyages_extraordinaires

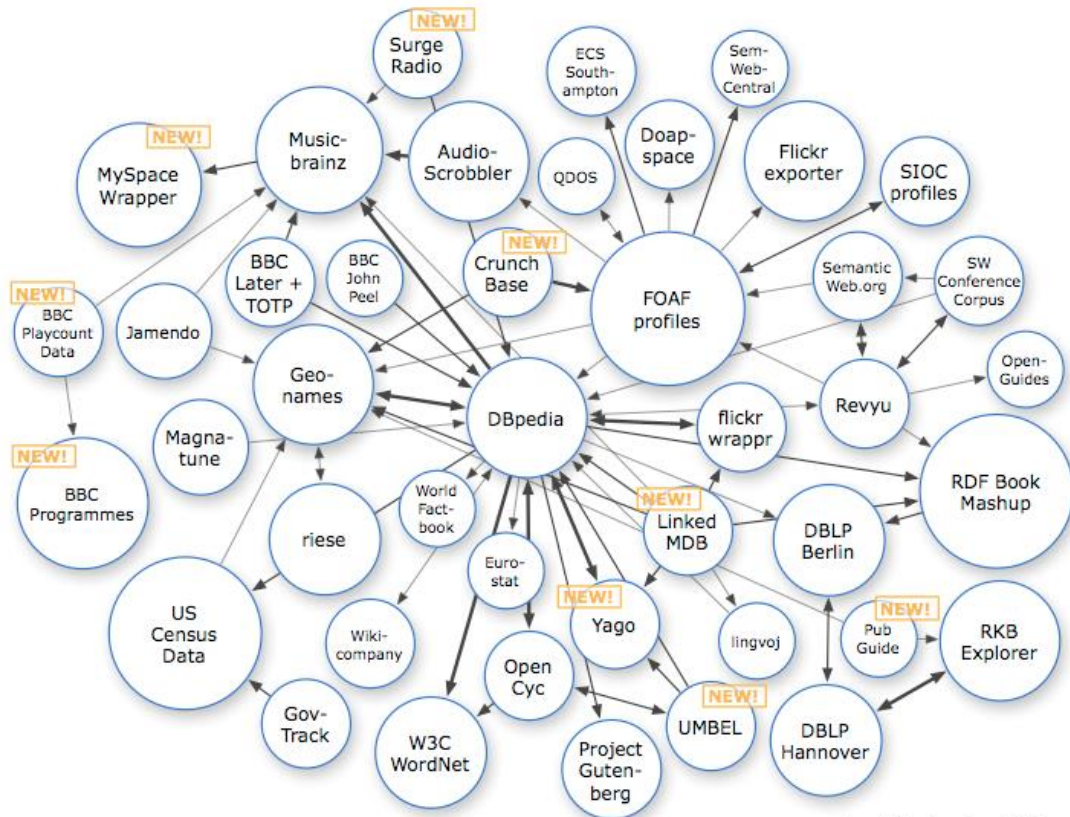
Sistemas Recomendadores



Sistemas Recomendadores

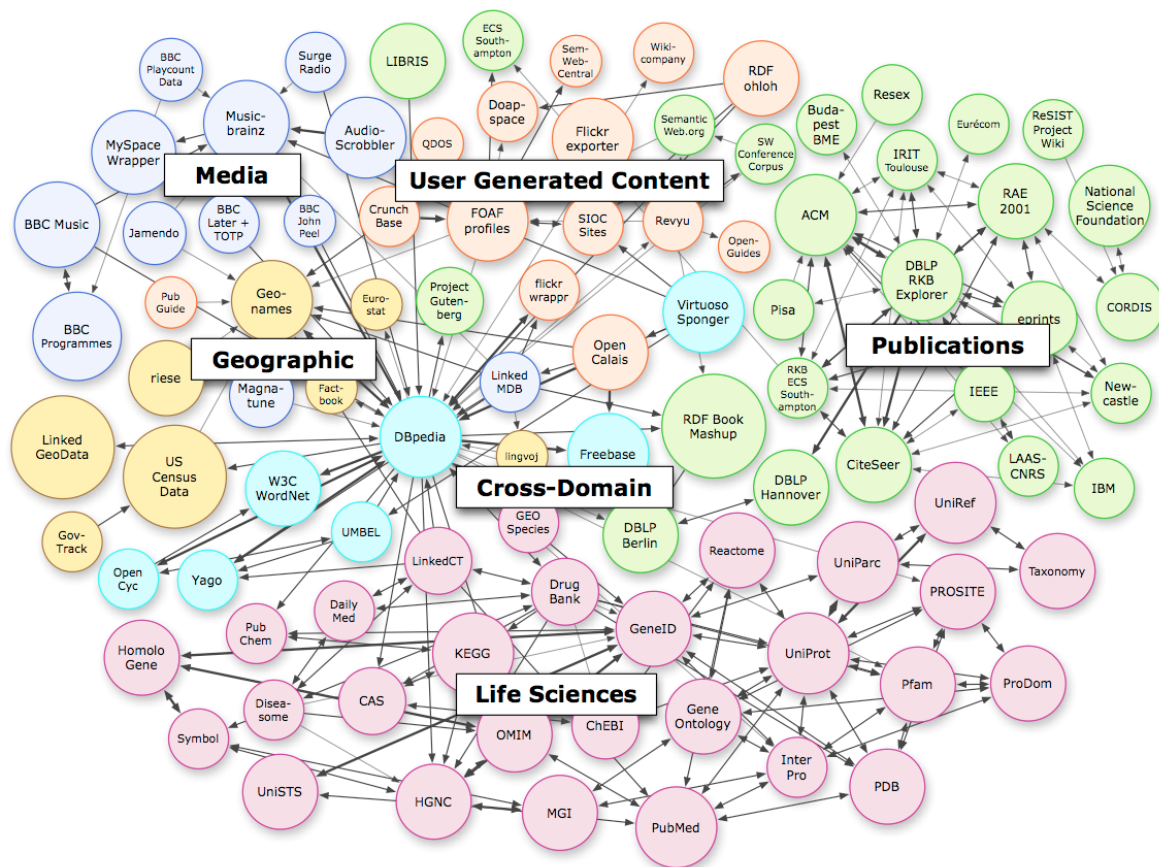


Septiembre 2008



As of September 2008

Julio 2014



Estadísticas de Julio 2014

<i>Dominio</i>	<i>No de Tripletas</i>	<i>%</i>	<i>No de enlaces</i>	<i>%</i>
Medios de comunicación	6.098.000.000	10,4%	10.238.000	0,8%
Publicaciones	2.012.000.000	3,2%	40.922.000	3,3%
Ciencias de la Vida	24.029.000.000	36,1%	1.033.199.000	89,4%
Datos Geográficos	30.097.000.000	46,0%	44.038.000	2,7%
Usuarios	976.000.000	1,1%	10.559.000	1,0%
Varios Dominios	2.014.000.000	3,2%	32.992.000	2,7%
<i>Total</i>	<i>>75.000.000.000</i>		<i>>1.000.000.000</i>	