

# Semántica en la Web

Claudio Gutierrez  
DCC, Universidad de Chile

Taller de la Web, CIW, USACH 2009

# Agenda

1. Definiciones previas
  - Semántica, Web, Computador, Humano
2. Computadores y humanos
3. Web Computadores => protocolos
4. Web Documentos => links
5. Web Datos => semantica
6. Web Humanos => lenguaje

# Semántica: definición

**Semántico, ca.** (del gr. σημαντικός).

1. adj. Perteneciente o relativo a la significación de las **palabras**.
2. f. Estudio del **significado** de los signos lingüísticos y de sus combinaciones

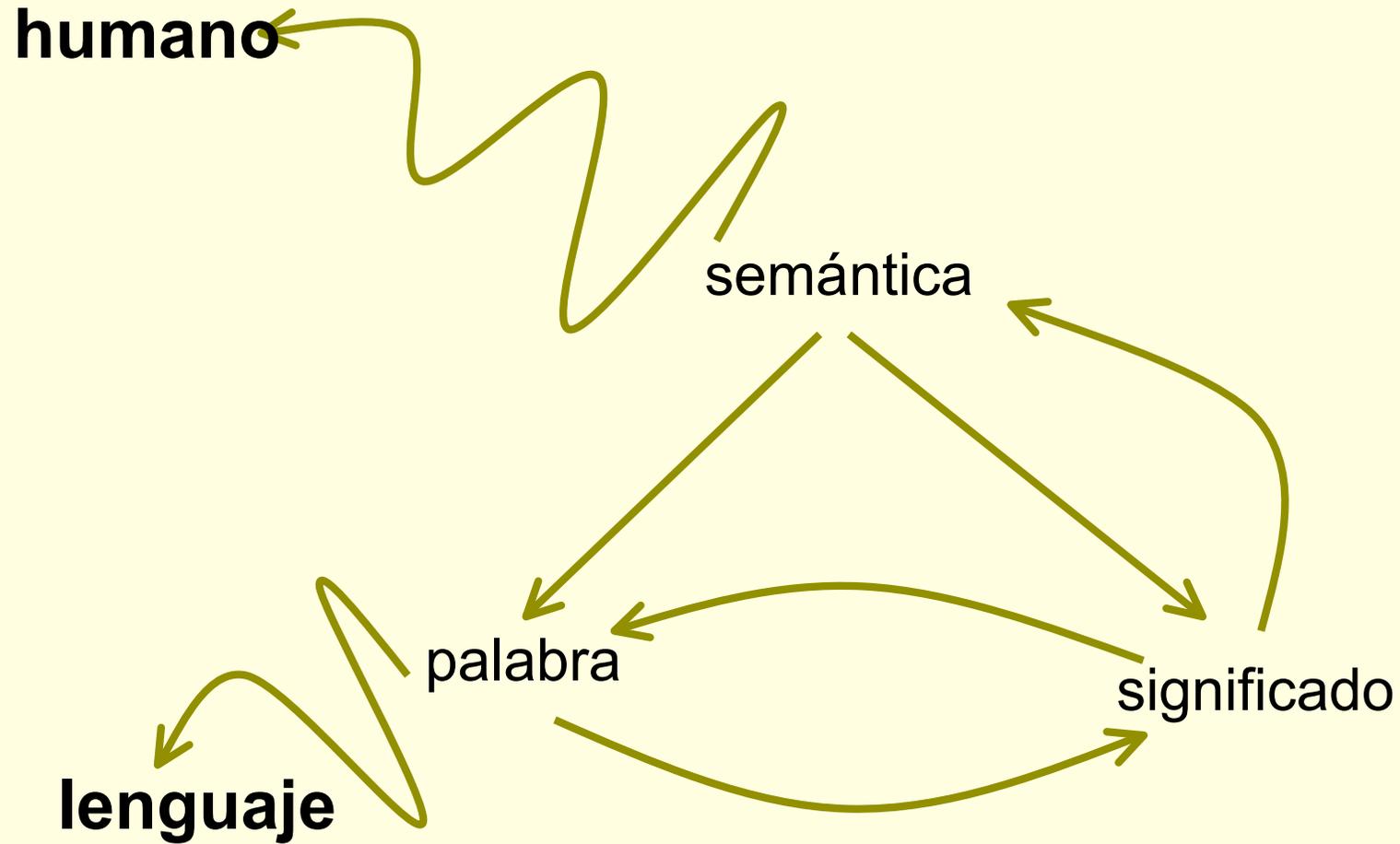
# Nos estará web... la RAE?

**palabra:** segmento del discurso unificado  
habitualmente por el acento y el significado

**significado:**

1. Significación o sentido de una **palabra** o de una frase.
1. Contenido **semántico** de cualquier tipo de sign

# El grafo de la RAE



# Computadores (RAE)

Máquinas dotadas de una **memoria** de **gran capacidad** y de métodos de tratamiento de la información, capaz de resolver problemas matemáticos y lógicos mediante la utilización **automática** de programas.

Resumen: gran capacidad, automatización

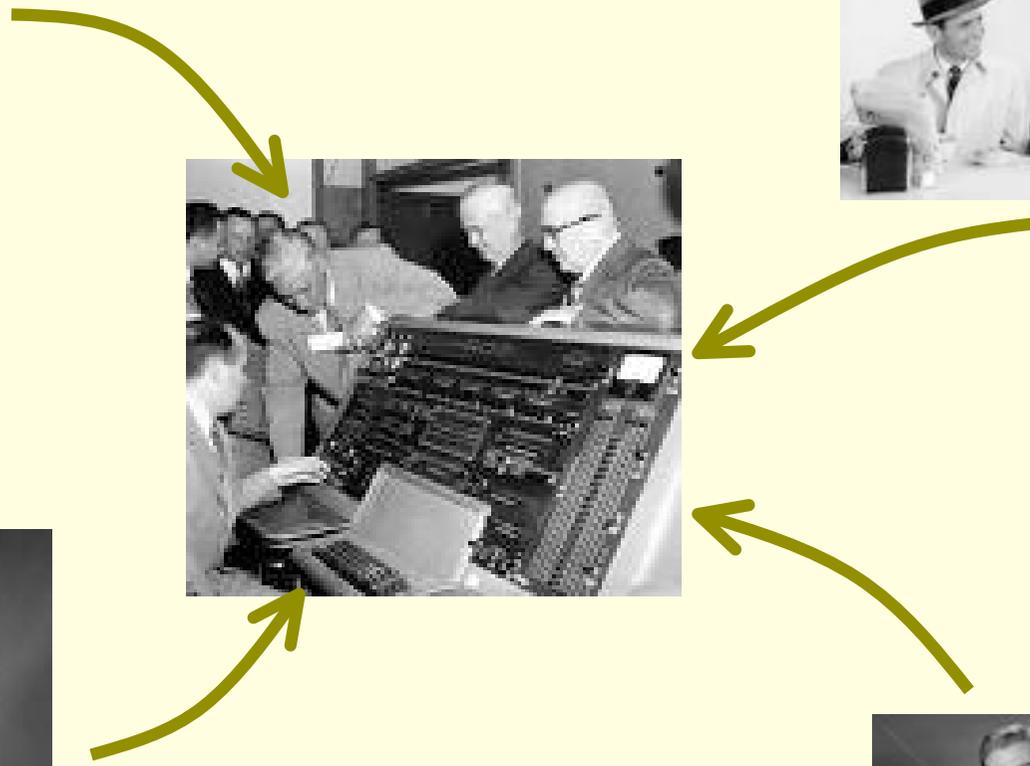
# Web (RAE)

- Conjunto de elementos organizados para determinado fin
- Conjunto de ordenadores o de equipos informáticos conectados entre sí que pueden intercambiar **información**.

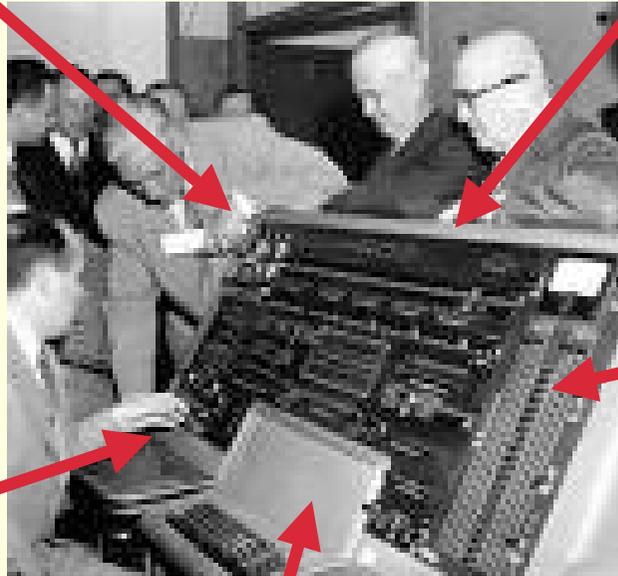
(También: “Capote de sayal”... ???)

*iiiiNOOOOOOO!!!!*

# Prehistoria: Uno para muchos



# El computador como centro



*(Nota: sólo hombres...)*

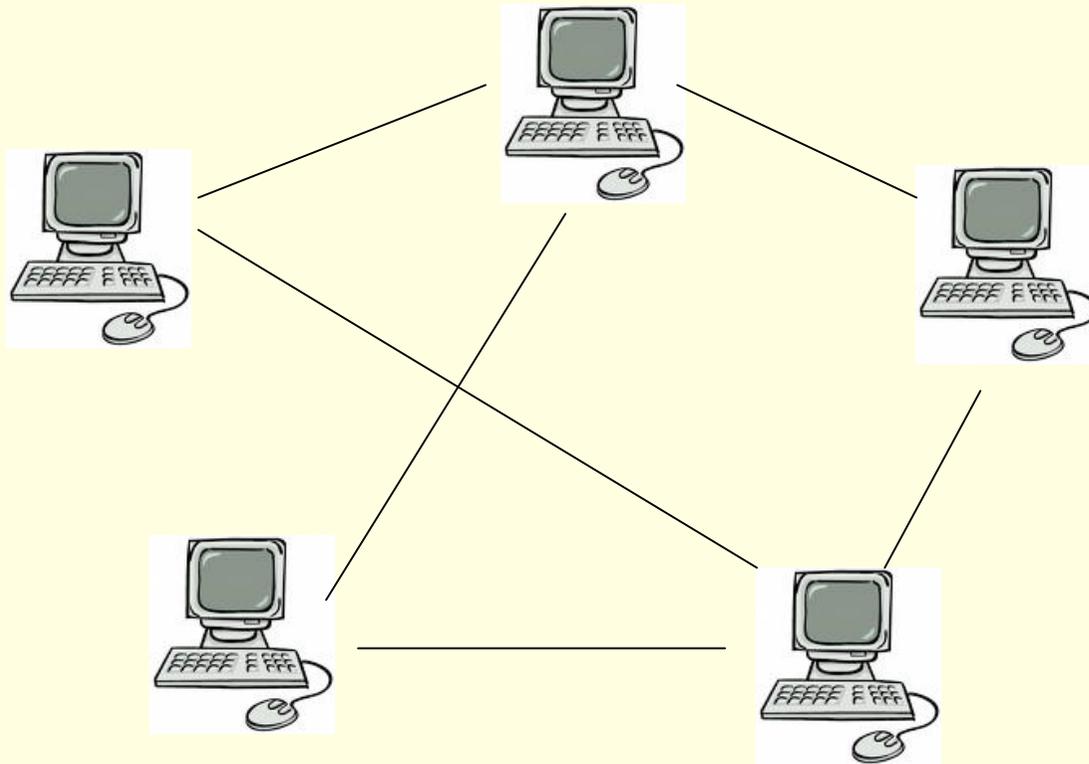
# Futuro: muchos para una...



# El humano como centro



# Web 0: computadores interconectados



# Web 0 (Internet)

- Cable: grosor, aislamiento, seguridad, ...
- Lenguaje: TCP/IP
  - protocolos de bajo nivel
  - Intercambio de paquetes
- Arquitecturas de red
- Características:
  - Semántica inexistente
  - ¿Dónde está el humano?

# Web 0 y BD: mientras...

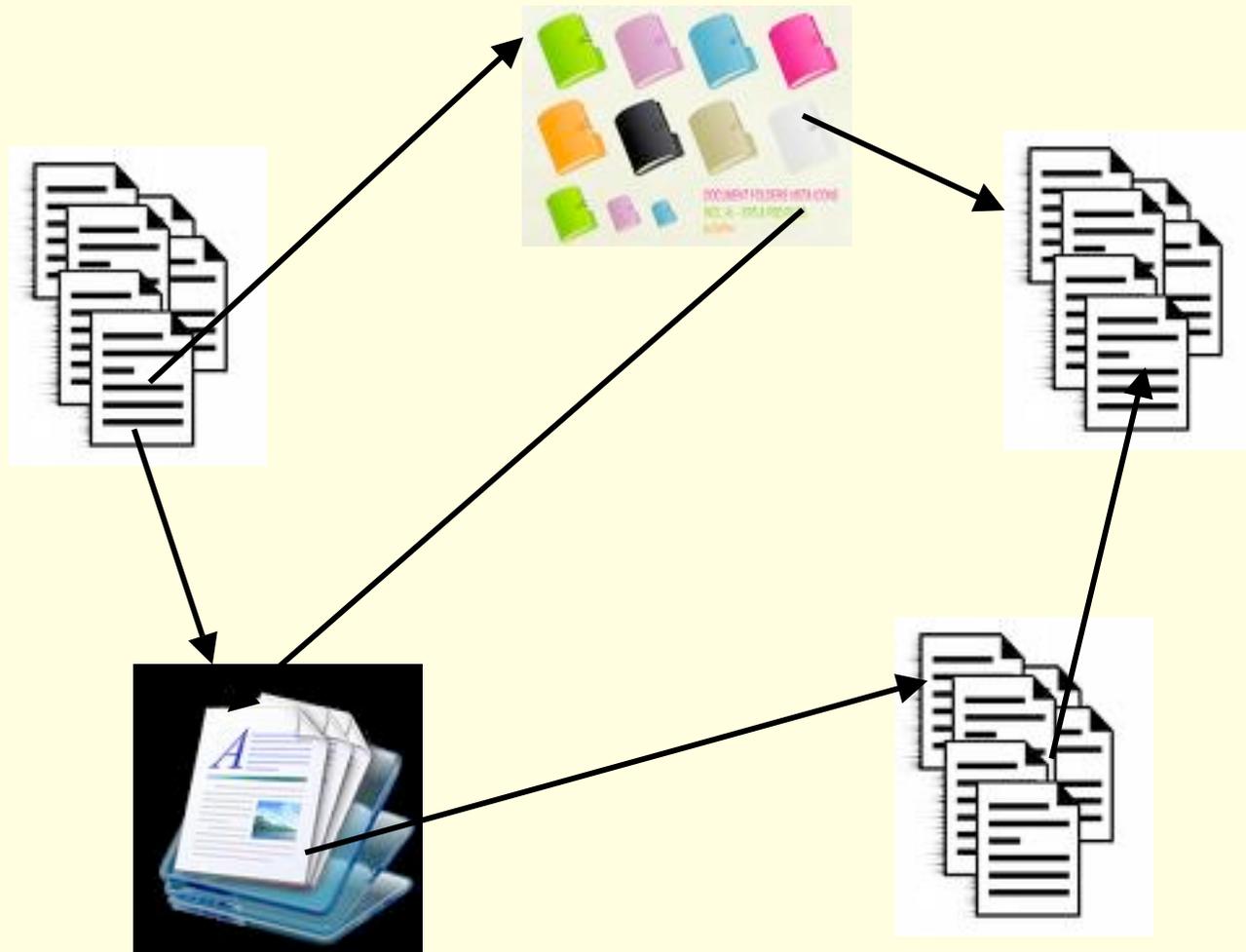
Mientras la Web se desarrollaba...  
expertos en BD diseñaban:

- Bases datos transaccionales
- Bases datos distribuídas

Pero esencialmente:



# Web Doc: documentos interconectados



# Web Doc: Tecnologías

- Documento = “Bolsa” de palabras
- Links entre bolsas
- Palabras y links como el centro del universo
- Lenguajes y protocolos:
  - HTTP y HTML

# Web Doc: Tecnologías

- “Semántica”: técnicas de IR  
pobres pero honradas...  
pero sobre todo MUY eficientes:
  - A) Análisis de palabras y textos
  - B) Análisis de links (¡redes!)
- Minería de Datos
  - Procesamiento de grandes volúmenes
  - Escasa semántica



# Web Doc: Tecnologías

- Peak del modelamiento de documentos
- Estándar: XML
- Lenguaje de Consulta: Xquery
- Bases de Datos
  
- Gran tema: Intercambio de información

# Web Doc y BD

Mientras la Web seguía su carrera...

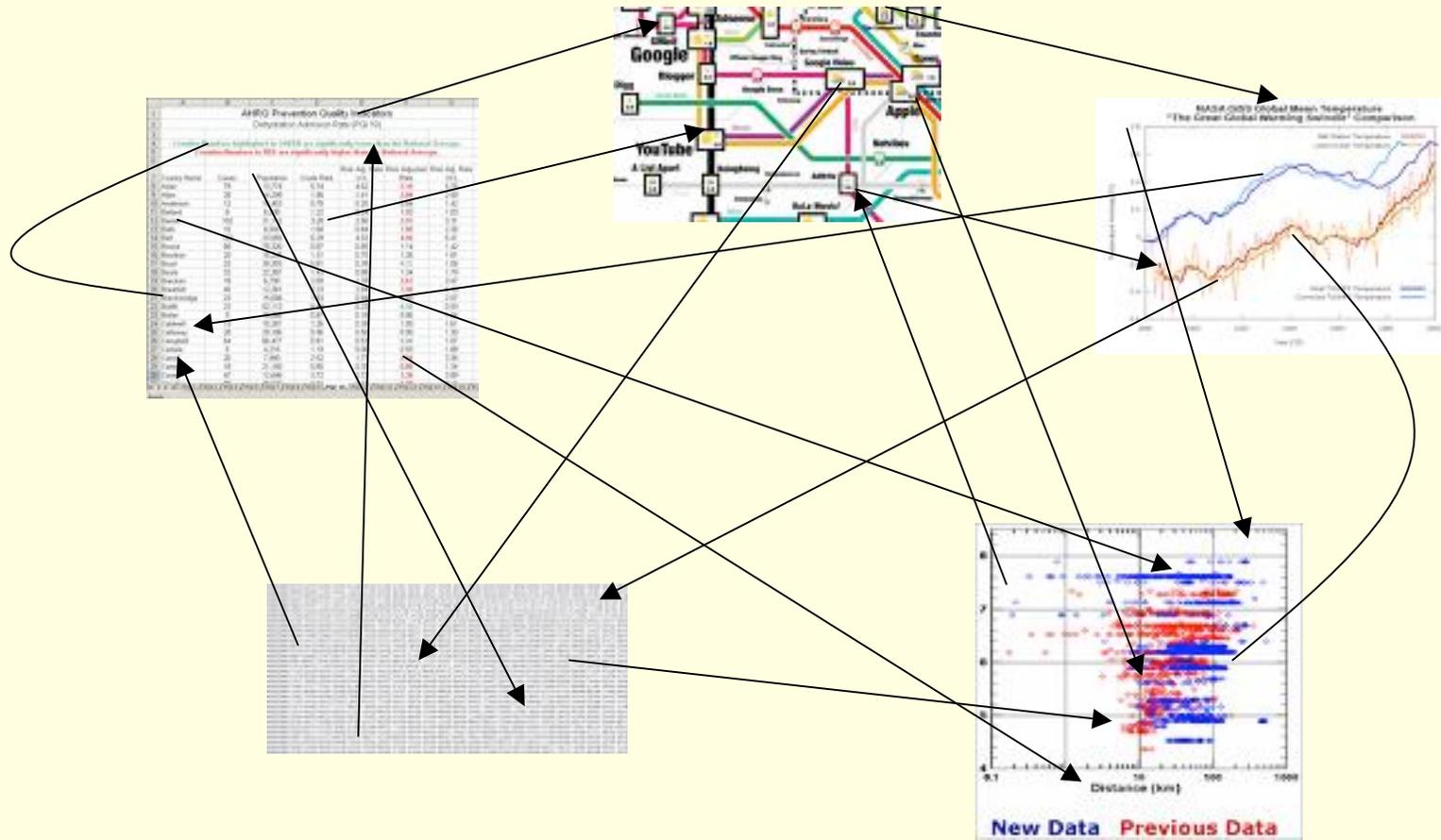
expertos en BD trabajando en:

- Bases datos clásicas
- Intercambio información con XML

¿Cuál es el principal problema de administración de datos hoy?



# Web Datos: datos interconectados

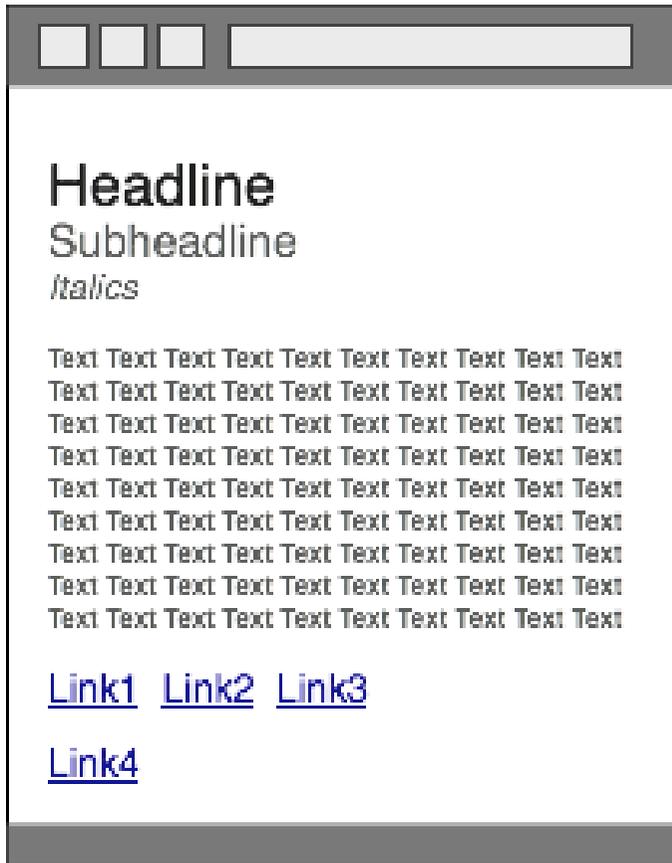


# Web Datos y BD

- Predicados (vínculo semántico) como el centro del universo
- Lenguajes:
  - **RDF** (Sujeto-predicado-objeto)
  - **SPARQL**: Lenguajes para procesar grafos
  - **Linked Data**: Protocolo para publicar datos
- **Temas nuevos**: GRANDES volúmenes  
grandes grafos (redes)

# Los grandes problemas

- Cómo **publicar**: formatos, metadatos, temporalidad, persistencia, privacidad
- Cómo **linkear**: protocolos, escalabilidad, mantenibilidad, semántica(!!!)
- Cómo **consultar / navegar**: distribución, relaciones, ranqueo, deducción(!!!)
- Cómo **integrar**: todas las anteriores...

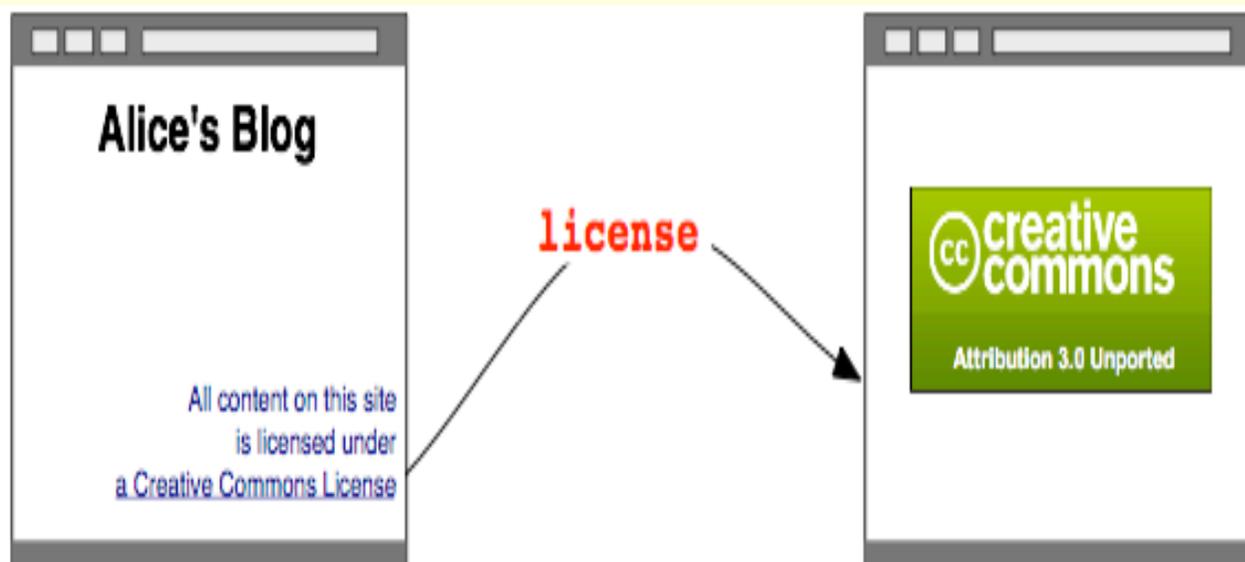


Lo que ven las  
máquinas hoy



Lo que ven  
los humanos

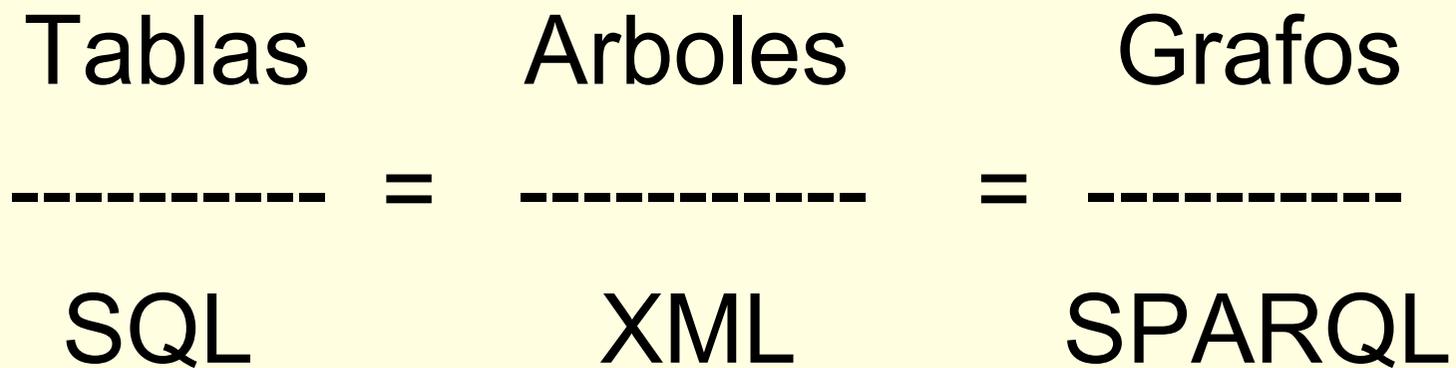
# Agregar semántica mínima



# RDF

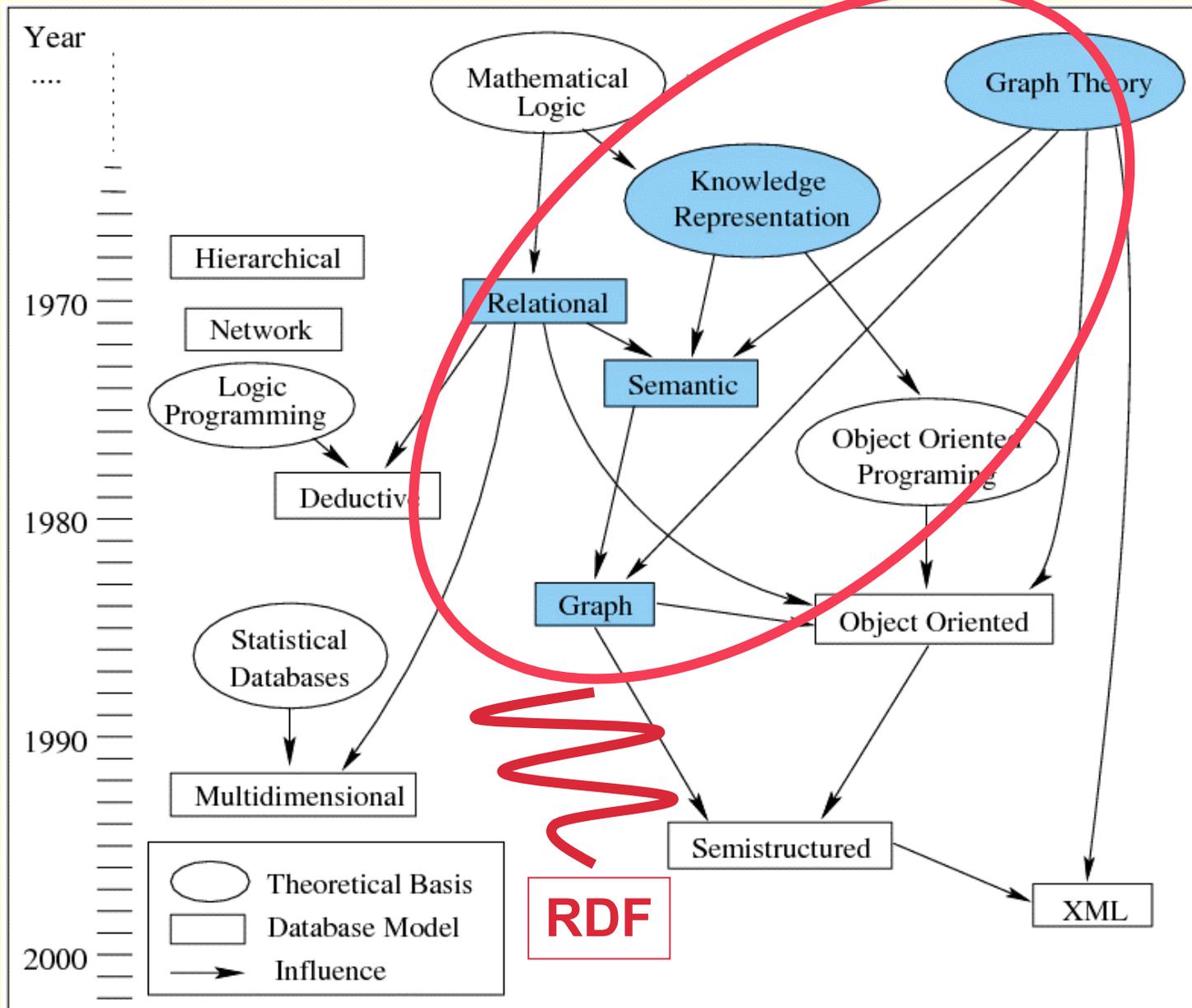
- RDF: Resource Description Framework  
recom: leer **RDF PRIMER**
- SPARQL: leng. de consulta para RDF  
recom: **escuchar a M. Arenas** en un rato más (y pregúntenle mucho...)
- RDFa: agregar semántica a los links
- recom: **MANOS a la OBRA!!**

# Consulta, navegación y transformación de datos



# BD Grafos

- Vuelven del mas allá...
- Tema HOT hoy día debido a la gran presencia de fuentes de datos de grafos
- Recomendación: **charlas**
  - Pablo Barceló: sobre posibles desarrollos
  - Renzo Angles, Bases Datos de Grafos

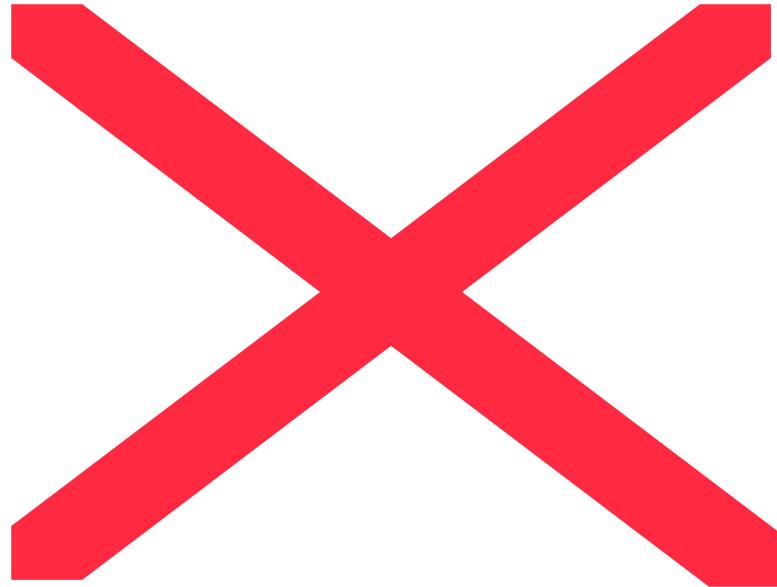


# Linked data: Publicar

- (en serio, no para uno mismo)
- Para quien?
- Buenas prácticas
- El tema de la escalabilidad
- El tema de la propiedad

# Linked data: mucho más que publicar

- Se necesita consultar y navegar
- Se necesita enlazar, conectar, relacionar la información para darle valor
- Se necesita dar persistencia
- Resolver temas privacidad / seguridad
- Buenas prácticas / formas normales



Datos vinculados ("linkeados")



# Web 2.0?...

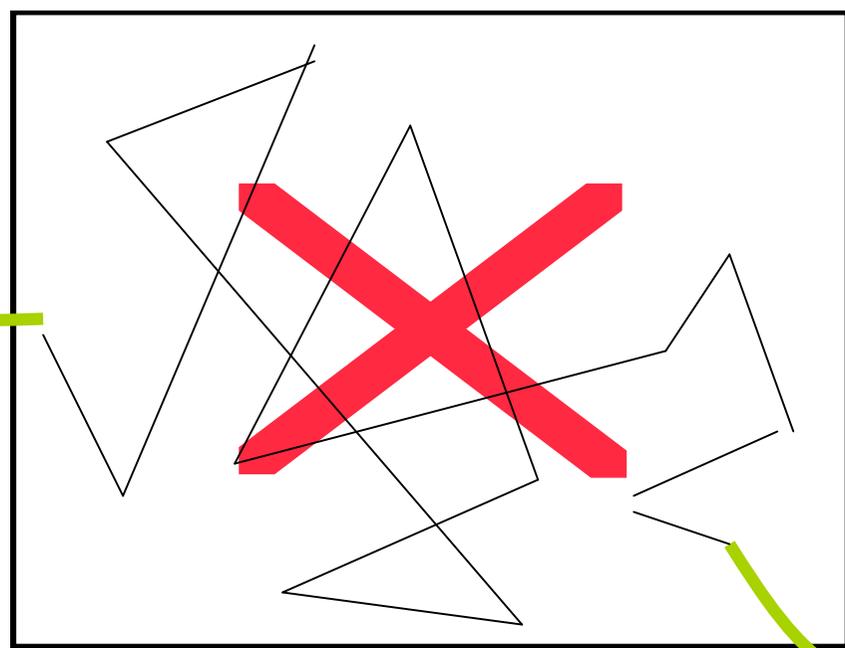
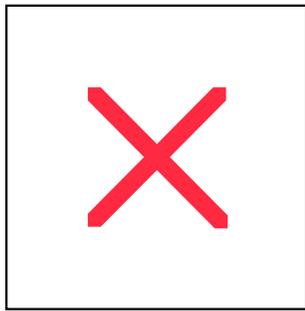
¿Keyword dropping?

¿Qué es?:

el “descubrimiento”  
que la red conecta  
humanos...

*(pero eso ya no lo  
sabíamos?...)*

# La web como pretexto para comunciar humanos



# Web “n”: humanos interconectados...

- Cómo publicar mi información, cómo hacerme “visible”
- Cómo usar la información que estará publicada para encontrar mis “intereses”
- Esencia: como quebrar la barrera entre producir y consumir información
- La vuelta a la “conversación”, pero ahora sin aburrirse con el vecino o el apoderado...

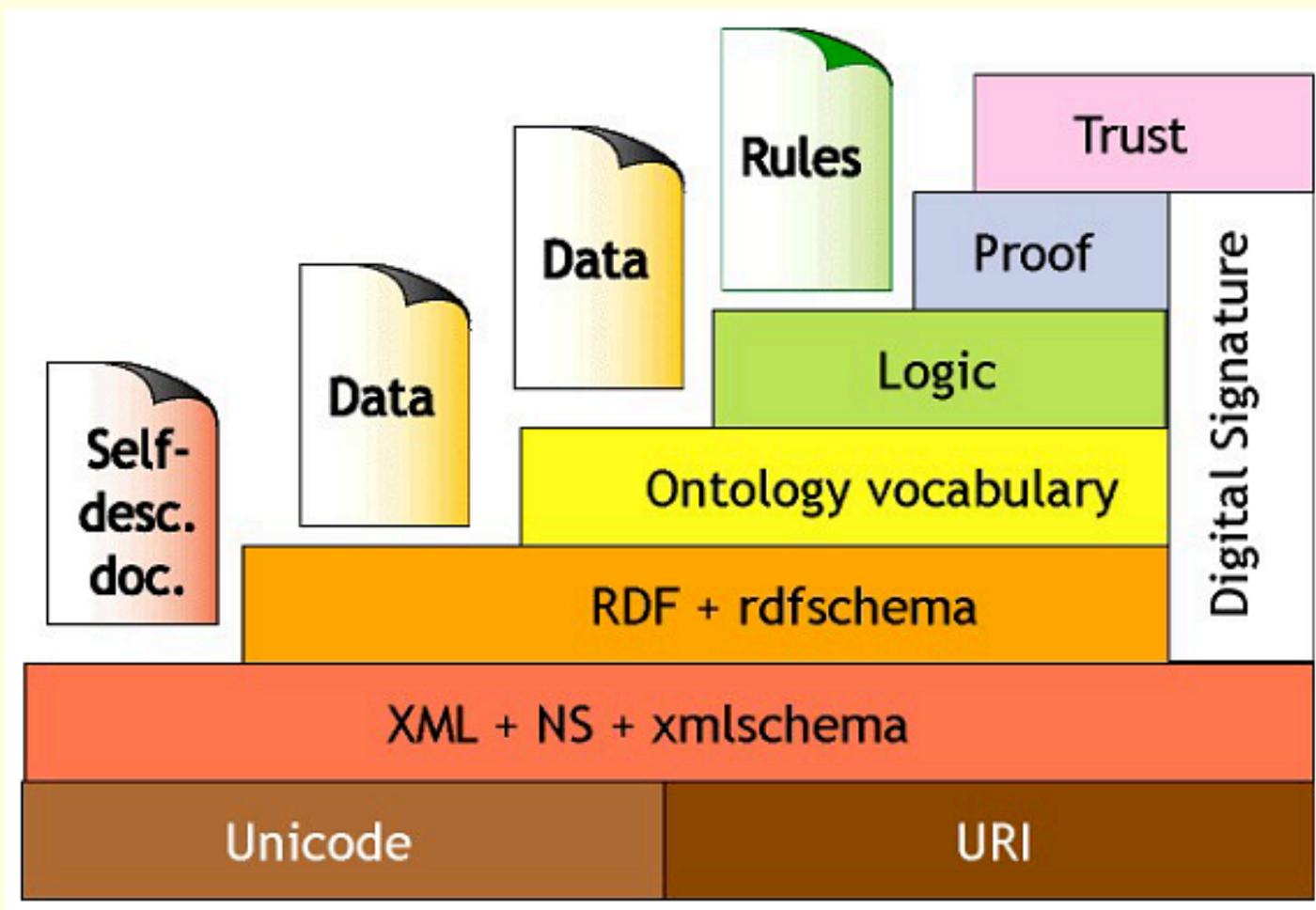
# Web humanos

- Cómo:
  - conversar con millones de humanos
  - escuchar (leer) a millones de humanos
  - hacer transacciones con millones de humanos

*Conditio sine qua non:*

**automatización de la semántica**

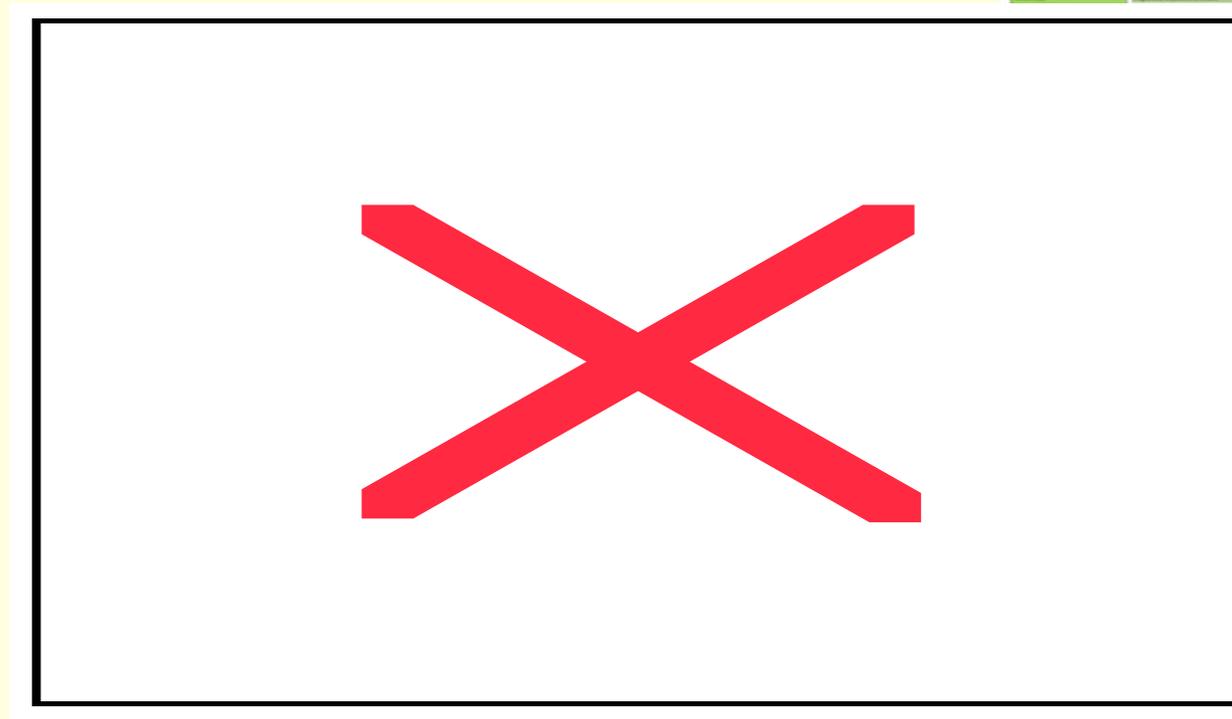
# Torre semántica



# Algunos ejemplos



The table is a complex grid with multiple columns and rows. The cells are color-coded: some are green, some are yellow, and some are white. The text within the cells is small and difficult to read, but it appears to be organized into sections or categories. The overall layout is dense and structured, resembling a data table or a detailed report.





Tim Berners-Lee:  
¡Datos sin procesar ahora!

Raw  
Data  
Now!