

## Qué es la Web Semántica?

---

Grupo de Concepción de Sistemas de Información

## Qué es la Web Semántica?

---

- *“La Web Semántica no es una web separada sino una extensión de la actual, en la cual la información tiene un significado bien definido, habilitando a las computadoras y las personas a trabajar en cooperación”.*
  - (T. Berners-Lee, J. Hendler., O. Lassila. The Semantic Web. Scientific American Mayo 2001)
- Por lo que conviene repasar algunas características de la Web Actual.

## Principios de la Web

- **Apertura:** Cualquier persona u organización puede utilizarla como proveedor o consumidor. Esto tiene en cuenta:
  - **Accesibilidad:** se puede acceder remotamente.
  - **No propiedad:** La Web no es propiedad de nadie.
  - **Control Consensuado:** La estructura de la Web es controlada y manejada por un a organización abierta (W3C) que tiene procedimientos bien definidos para la toma de decisiones de forma consensuada.
  - **Usable:** El uso de la infraestructura es simple e irrestricto.
- **Interoperabilidad:** La Web es independiente del hardware y software. Sólo hay que usar los protocolos adecuados.
- **Autoría y Edición descentralizada:** El contenido puede aparecer, ser modificado o removido de una forma descentralizada, sin control central.
  - (AAA:Anyone can say Anything about Any topic)

## Funcionalidades Básicas de la Web (Arquitectura Funcional)

- **Funcionalidades**
  - Una mecanismo de referencia global: URL's, URN's o URI'S
  - Un nivel de transporte: HTTP.
  - Una interfase independiente de la plataforma: HTML.
- **Consecuencia:**
  - Un conjunto de funcionalidades simples que construyen un mecanismo totalmente escalable, con posibilidades de crecimiento casi infinitas.

## Los Problemas de la Web (1)

---

- El **acceso a los datos** es limitado:
  - **Los documentos estan indexados en texto.**
    - Paris es:
      - Ciudad: Francia, EEUU, Canadá
      - Personas: Una figura de la Guerra de Troya, Paris Hilton o una hija de Michael Jackson.
  - **La búsqueda es a “lo que calza mejor”.**
    - La decisión de la pertinencia e integración de lo relevante queda a cargo del usuario.
  - **Los datos no están accesibles.**
    - Los datos están en las páginas, lo que complica su procesamiento. Solución parcial: Scrapping.

## Los Problemas de la Web (2)

---

- La Web está **Orientada a personas y no hay mucha ayuda para ellas.**
  - La Web puede ser vista como una colección de páginas que deben ser interpretadas necesariamente por personas.
- Una solución:
  - Habilitar a las computadoras a extraer, e integrar información para los usuarios.

## Qué es Semántica?

---

- Wikipedia:
  - El término **semántica** (del griego *semantikos*, "lo que tiene significado") se refiere a los aspectos del significado, sentido o interpretación de signos lingüísticos como símbolos, palabras, expresiones o representaciones formales.
- Se encara desde diferentes disciplinas:
  - Lingüística.
  - Psicología.
  - Matemáticas.
  - **Informática.**

## Los Problemas de la Semántica

---

- Diferentes significados para las referencias textuales:
  - Paris es:...
- El mismo significado para referencias diferentes:
  - Cesta, canasta
  - auto, automóvil, carro, coche.
- Diferentes significados para las frases:
  - "Se murió la perra de mi suegra".
- Significados implícitos.
  - "Ud. tiene menos palabra que un mudo !".

## Semántica en la Web?

---

- Hay que introducirla de forma que los problemas anteriores sean tratables.
- De forma que los nodos sean procesables por máquinas, orientándose a:
  - Facilitar la creación y manipulación de datos a través de aplicaciones.
  - Ayudar a los usuarios a crear información a partir de los datos.
- Todo esto, sin violar los principios de La Web (AAA).

## Pero... Que es la Web Semántica?

---

- Si alguien busca el manual de una cámara de fotos determinada:
  - Quiere el manual como documento?
  - Quiere información sobre ese modelo de cámara?
- “...*isn't the documents which are actually interesting, it is the things they are about!*” (TBL).
- Es la *Web de Cosas* (**Web de Datos**).

## Idea Básica: Interesan los Links

Lugares de Distribución

Tipo

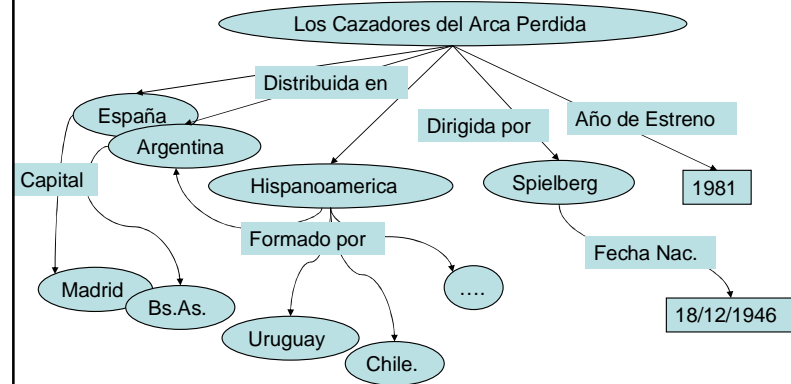
### Raiders of the Lost Ark

*Raiders of the lost ark* (conocida también como *Indiana Jones and the raiders of the lost ark*, en **España** y **Argentina** como *En busca del arca perdida* y en el resto de **Hispanoamérica** como *Los cazadores del arca perdida*) es una **película de aventuras** estadounidense dirigida por **Steven Spielberg**, estrenada en **1981** y producida por **George Lucas**. Es el primer filme en ser estrenado de la franquicia de *Indiana Jones* y el segundo en términos cronológicos de la serie. Obtuvo cinco premios Óscar, de un total de nueve nominaciones (que incluían la categoría de "mejor película").

El **Origen** fue escrito por **Lawrence Kasdan** y **George Lucas**. Fue dirigido por **Steven Spielberg**. El reparto estuvo integrado por **Harrison Ford**, **Karen Allen**, **Paul Freeman**, **John Rhys-Davies** y **Danholm Elliot**; el rodaje fue realizado en **Inlaterra** y **La Rochella**.



## Idea Básica: Tipear los links y los Recursos



## Ideas Básicas: Resumen

---

- Representar entidades, propiedades y relaciones entre las entidades:
  - La información interesante está en los links.
- Cada recurso se representa por un grafo.
- Cada arco representa alguna forma de relación entre recursos o datos.
- A cada recurso se lo puede clasificar según sus relaciones.
- Se pueden establecer relaciones entre esas clases ( o entre las relaciones !!!)
  - Una película se distribuye países o regiones. Si es una región, se distribuye en cada uno de los países de la región.
- Al tratar todo el grafo, se puede encontrar información que no se conocía explícitamente (inferencia)

## Cómo se Construye la Web Semántica ?

---

- Definir requerimientos.
- Definir arquitectura.
- Definir los elementos de la arquitectura.
- Estudiar las alternativas para cada elemento de la arquitectura.

## Requerimientos para una Arquitectura de la Web Semántica.

---

- Características desde el punto de vista de sistema informático:
  - **Representación simple de los datos.**
    - Análogo a los documentos HTML pero orientado a los datos.
  - **Sistema Distribuido:**
    - El sistema debe ser distribuido sin control centralizado, y facilitar la adopción y la escalabilidad.
    - Análogo a los sitios web, en donde los datos estén bajo el control de sus productores.
  - **Un mecanismo de referencia:**
    - Los datos deben estar inter-referenciados, permitiendo de esta forma el reuso de datos y definiciones de diferentes orígenes.

## Requerimientos para una Arquitectura de la WS.

---

- **Débilmente acoplados a través de lenguajes y protocolos comunes.**
  - Lenguajes estandarizados con mucha flexibilidad y posibilidad de especialización, pero sin que esto dificulte la comunicación.
  - Análogo al acoplamiento entre diferentes Web Servers mediante HTTP como transporte, y produciendo contenido HTML
- **Facilidad de publicación y consumo.**
  - Debe ser razonablemente fácil publicar y consumir datos en el sistema.
  - Análogo a el uso de HTML en la web, como mecanismo de publicación y consumo.



## Requerimientos para una Arquitectura de la WS: Lenguajes

---

1. Deben poder representar entidades individuales, las relaciones entre estas entidades y datos concretos sobre estas entidades.
  - Debe ser un Modelo de Datos, en el sentido que se le da en Base de Datos.
  - Ej: Los Cazadores del Arca Perdida, fue dirigida por Steven Spielberg. Se estrenó en 1981.
2. Esas entidades y propiedades deben ser intercambiables entre diferentes máquinas.
  - Con esto se logra que se puedan realizar combinaciones de datos (igual que en BD) independientemente del servidor que los provea.

## Requerimientos para una Arquitectura de la WS: Lenguajes

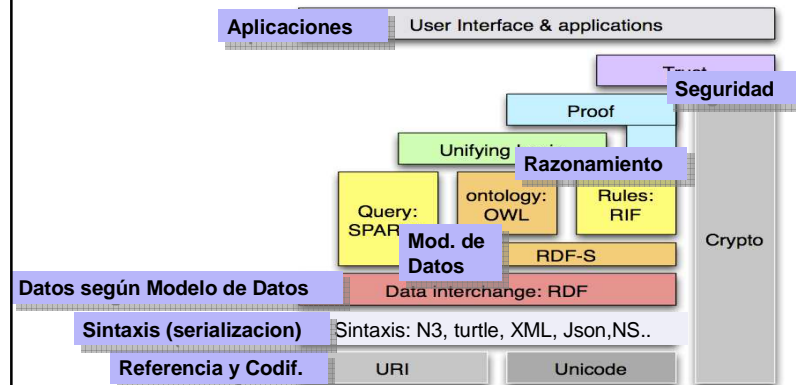
---

3. Las entidades deben poder ser referenciables incluso más allá de los límites de su nodo.
  - Permite diferenciar a Eduardo Fernandez – guitarrista - de Eduardo Fernandez – docente del InCo.
4. El modelo de datos debe ser expresivo pero a la vez, procesable por máquinas:
  - Una persona puede no ser capaz de comprender todos los datos de una entidad o todo un esquema completo. Ej: BBC.
5. El modelo requiere lenguajes de consulta y manipulación que permitan las selecciones y/o agregaciones de datos.

## Requerimientos para una Arquitectura de la WS: Lenguajes

6. Es deseable que el modelo soporte mecanismos de razonamiento.
  - Esto permite resolver casos complejos de modelado (ej: definición de cadenas de propiedades) y/o obtener datos que en realidad están implícitos. También facilita la integración.
7. Se necesitan protocolos de transporte tanto para los datos en sí mismos como para las consultas y resultados.
  - Ej: Debe estar claro que devuelve y cómo se hace una consulta
8. Ese transporte debe soportar seguridad
  - Encriptación, firma, etc.

## Una Arquitectura para la WS.



*Tim Berners-Lee. Artificial Intelligence and The Semantic Web. AAAI 2006. Boston*

## Bibliografía

---

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*.
- Berners-Lee, Tim. (2006). Artificial Intelligence and The Semantic Web. Presented at the The Twenty-First National Conference on Artificial Intelligence, Boston: AAAI Press. Retrieved from <http://www.w3.org/2006/Talks/0718-aaai-tbl/>
- Domingue, J., Fensel, D., & Hendler, J. A. (2011). Introduction to the Semantic Web Technologies. In J. Domingue, D. Fensel, & J. A. Hendler (Eds.), *Handbook of Semantic Web Technologies* (pp. 1–41). Springer Berlin Heidelberg. Retrieved from [http://dx.doi.org/10.1007/978-3-540-92913-0\\_1](http://dx.doi.org/10.1007/978-3-540-92913-0_1)
- Harth, A., Janik, M., & Staab, S. (2011). Semantic Web Architecture. In J. Domingue, D. Fensel, & J. A. Hendler (Eds.), *Handbook of Semantic Web Technologies* (pp. 43–75). Springer Berlin Heidelberg. Retrieved from [http://dx.doi.org/10.1007/978-3-540-92913-0\\_2](http://dx.doi.org/10.1007/978-3-540-92913-0_2)