

01MBID Fundamentos de la tecnología Big Data



viu

Universidad
Internacional
de Valencia

Sesión 6

De:



Planeta Formación y Universidades

> Agenda

- **Dudas**
- **Temas 6 y 7**

> Agenda

- **Dudas**
- **Temas 6 y 7**

> Dudas



> Agenda

- Dudas
- **Temas 6 y 7**

> Seguimos de la sesión anterior

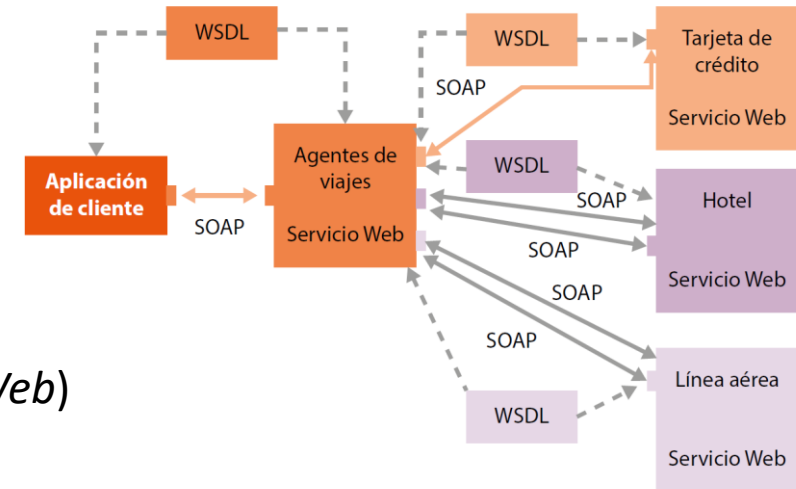
* PowerData.es



API, Servicios Web y Crawlers

Servicios Web

Los servicios web son un conjunto de aplicaciones o de tecnologías con capacidad para interoperar en la Web y con el fin de proporcionar servicios.



SOAP (***Simple*** Object Access Protocol)

WSDL (*Lenguaje de Descripción de Servicios Web*)



API, Servicios Web y Crawlers

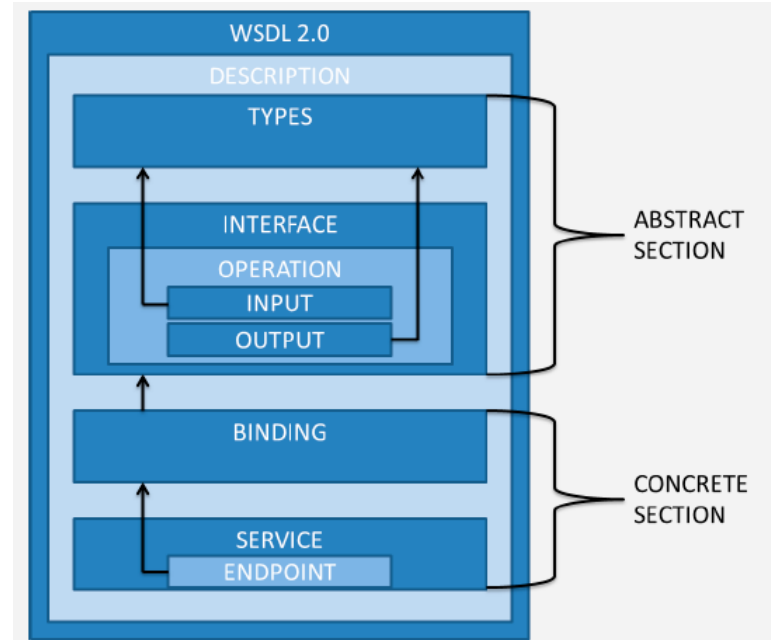
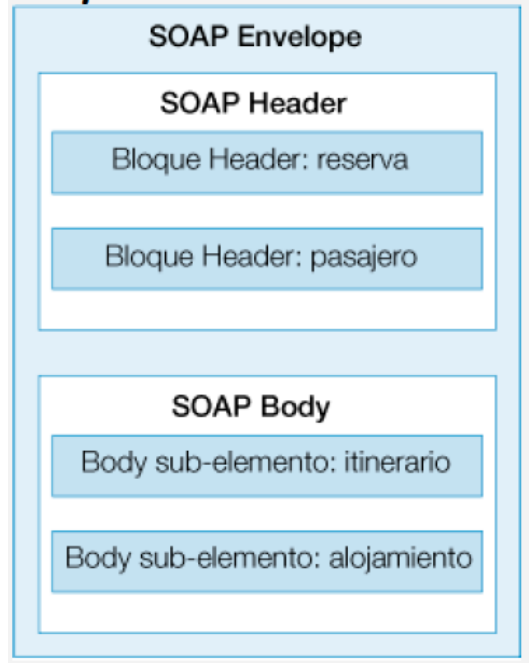
Servicios Web





API, Servicios Web y Crawlers

Servicios Web





API, Servicios Web y Crawlers

Servicios Web

Servicios REST

Representational State Transfer

- ☐ Cualquier interfaz entre sistemas que use HTTP para obtener datos o generar operaciones sobre esos datos
 - Todos los formatos posibles, como XML y JSON.
- ☐ Alternativa en auge frente a protocolos como SOAP, que tienen gran capacidad pero también mucha complejidad.

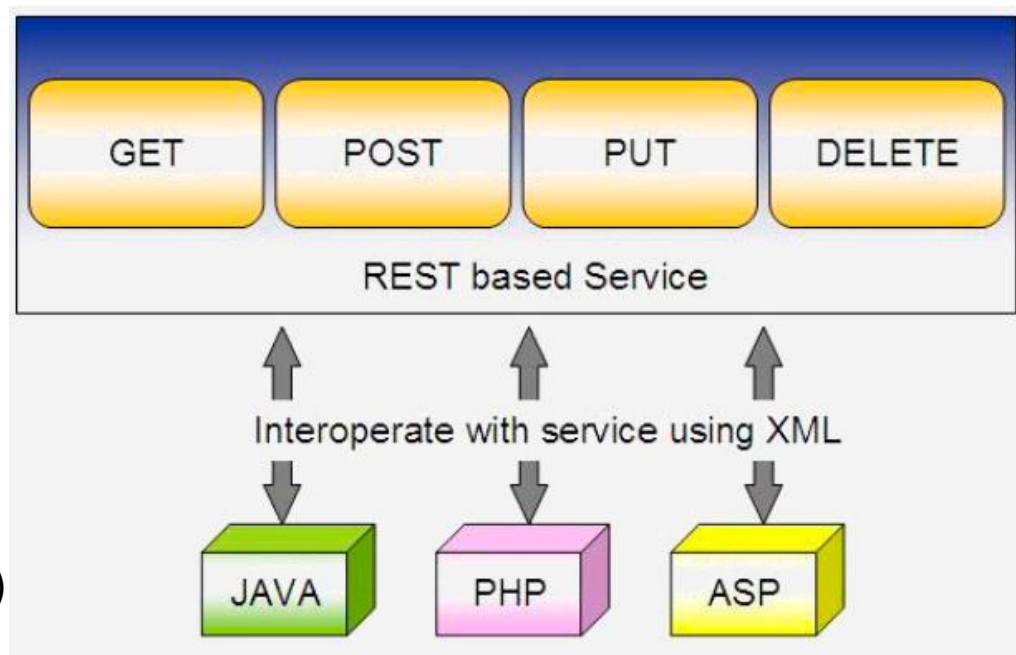


API, Servicios Web y Crawlers

Servicios Web

Servicios REST

- ☐ Cliente/Servidor
- ☐ Sin estado
- ☐ Interfaz uniforme
- ☐ URI (*Uniform Resource Identifier*)





API, Servicios Web y Crawlers
Servicios Web

Servicios REST vs SOAP



Consider "Martin Lawrence" as your data

SOAP



REST





API, Servicios Web y Crawlers

Crawlers

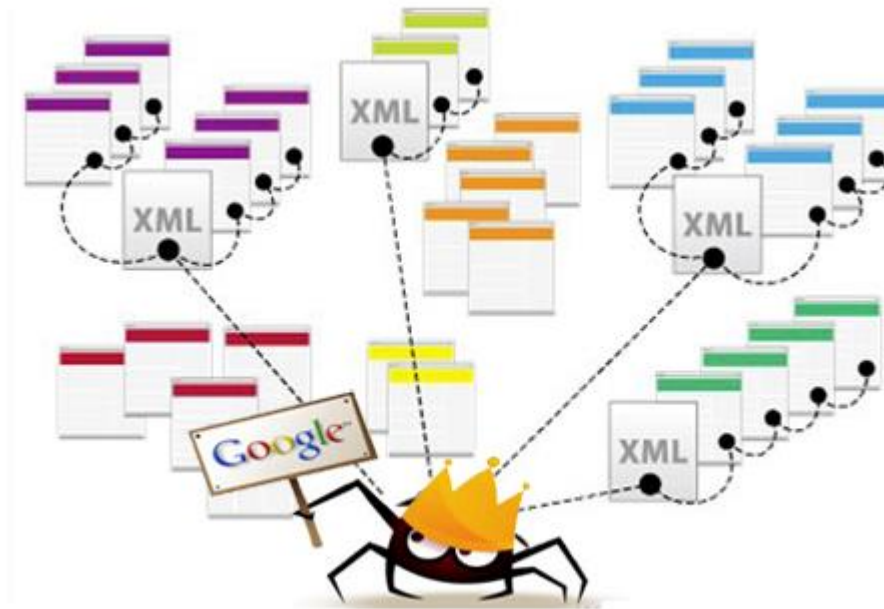
Traducción del inglés, araña de la web. Software que se encarga de recorrer los enlaces de las páginas web de forma sistemática y automática.





API, Servicios Web y Crawlers

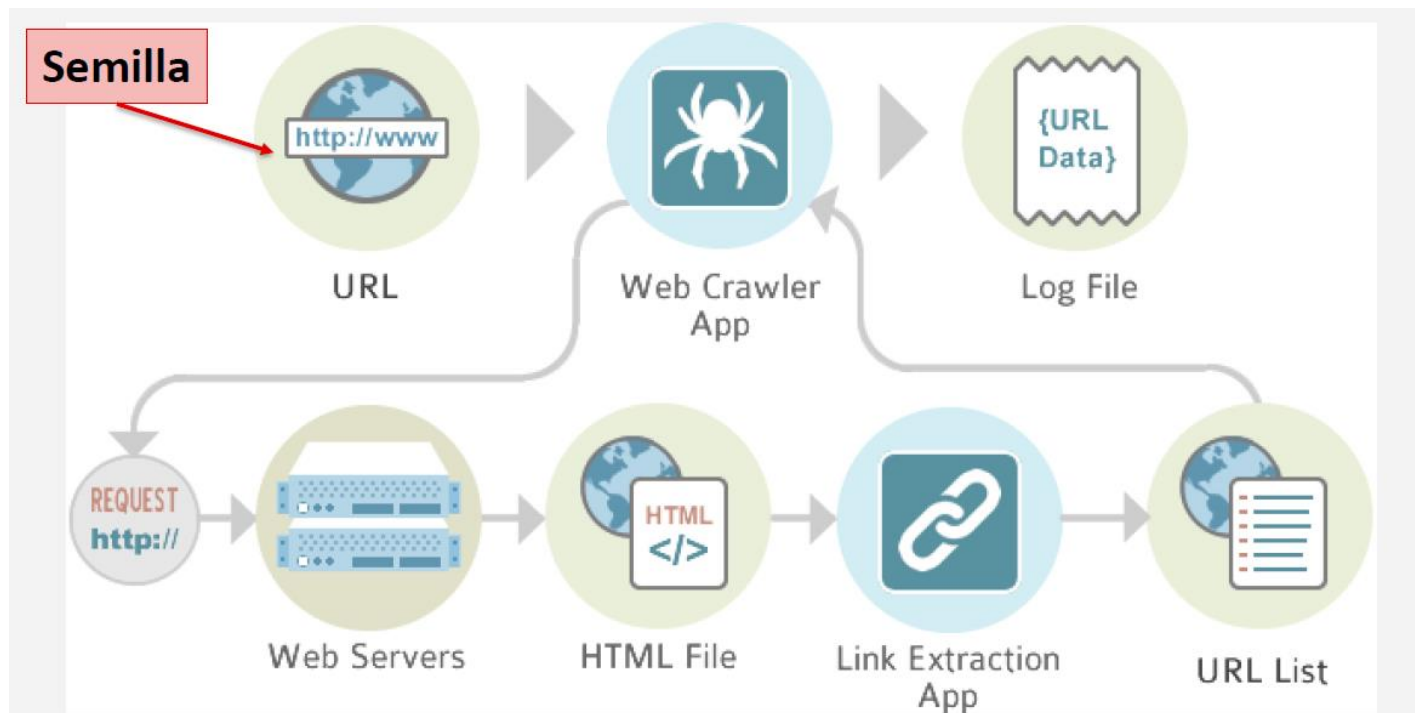
Crawlers





API, Servicios Web y Crawlers

Crawlers



> **Métodos de Integración y Motores de Búsqueda**

- 1) ¿Qué es integración de datos?
- 2) Extracción, Transformación y Carga
- 3) Otros aspectos de la integración de datos
- 4) API, Servicios Web y Crawlers
- 5) **Motores de búsqueda**

Motores de búsqueda



Recuperación de la Información

- ☐ Representación
- ☐ Almacenamiento
- ☐ Organización
- ☐ Acceso



Motores de búsqueda



Recuperación de la Información: Áreas

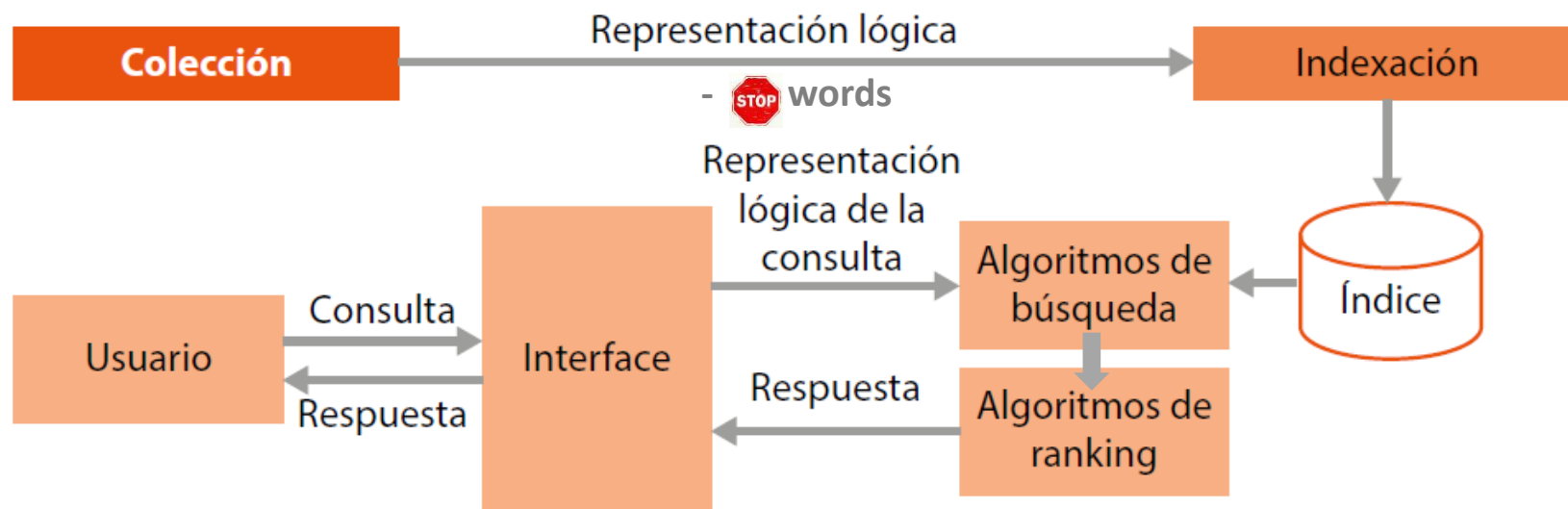
- ☐ Estructuras de datos
- ☐ Métodos de indexación
- ☐ Sistemas Distribuidos
- ☐ Algoritmos de compresión
- ☐ Algoritmos de ranking
- ☐ Optimización en búsquedas
- ☐ Data Profiling (e.g. estadísticas)
- ☐ ...



* Salton, G. Y Mc Gill, M.J. "Introduction to Modern Information Retrieval". New York. Mc Graw-Hill Computer Series. 1983.

Motores de búsqueda

Recuperación de la Información: Arquitectura Básica



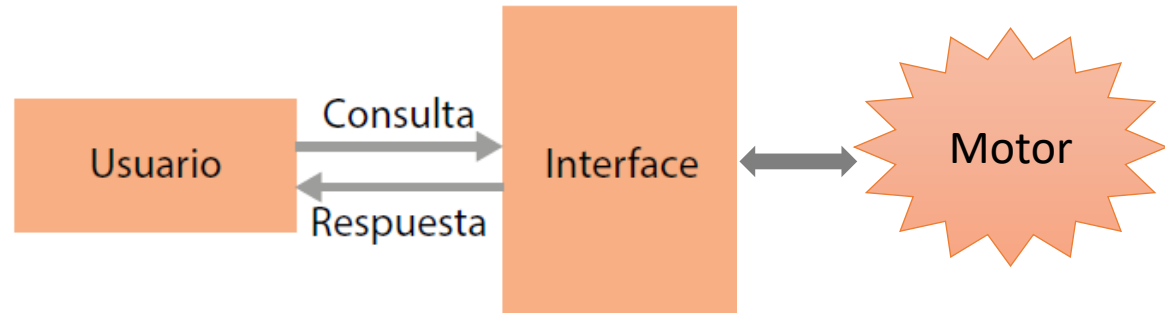
* Baeza-Yates & Ribeiro-Neto (1999).

Motores de búsqueda



Motor de Búsqueda o Buscador

Sistema informático encargado de realizar la búsqueda de información en la web. Generalmente utilizan crawlers para obtener los resultados.



Motores de búsqueda



Google: cómo funciona BERT, la mayor actualización del algoritmo del motor de búsqueda más usado en el mundo

<https://www.bbc.com/mundo/noticias-50223408>



Understanding searches better than ever before

<https://www.blog.google/products/search/search-language-understanding-bert/>



Motores de búsqueda



Understanding searches better than ever before

“**15 percent** of those queries are ones we **haven’t seen before**--so we’ve built ways to return results for queries we can’t **anticipate**.”

“we aren’t always quite sure about the best way to formulate a query. We might **not know** the **right words** to use, or how to **spell something**, because often times, we come to Search looking to learn --we don’t necessarily have the knowledge to begin with.”

Motores de búsqueda



Understanding searches better than ever before

“Search is about understanding language”

“With the latest advancements from our research team in the science of language understanding --made possible by machine learning-- we’re making a significant improvement to how we understand queries”

Motores de búsqueda



Understanding searches better than ever before

“Last year, we [introduced and open-sourced](#) a **neural network-based technique** for **natural language processing (NLP)** pre-training called **Bidirectional Encoder Representations from Transformers**, or as we call it--[BERT](#), for short. This technology enables anyone to train their own state-of-the-art question answering system.”

“models that process words in relation to all the other words in a sentence, rather than one-by-one in order.”

Motores de búsqueda



Understanding searches better than ever before

“BERT are so complex that they **push the limits** of what we can do using traditional **hardware**, so for the first time we’re using the latest [Cloud TPUs](#) to **serve search results and get you more relevant information quickly.**”

* **TPU** = Tensor Processing Unit

* **Cloud TPU** is the custom-designed machine learning ASIC that powers Google products like Translate, Photos, Search, Assistant, and Gmail

* **ASIC** = Application-Specific Integrated Circuit

Motores de búsqueda



Understanding searches better than ever before

“In fact, when it comes to **ranking results**, BERT will help **Search better understand** one in 10 searches in the U.S. in English, and we’ll bring this to more languages and locales over time.”

Motores de búsqueda



Understanding searches better than ever before

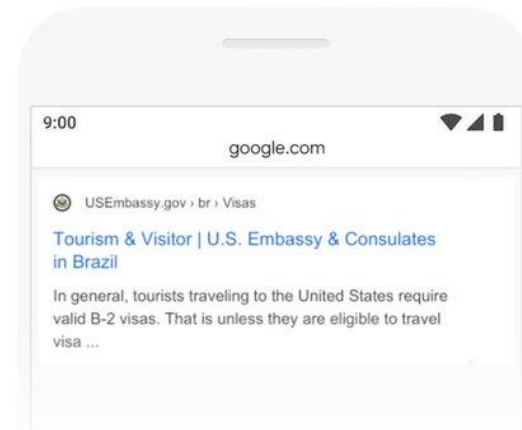


🔍 2019 brazil traveler to usa need a visa

BEFORE



AFTER



Motores de búsqueda



big data



Archie Query Form



Search for:

All



Enter keywords or phrases (Note: Searches metadata only by default. A search for 'smart grid' = 'smart AND grid')



Motores de búsqueda



<https://www.google.com/imghp?hl=es>



> Dudas



01MBID

Roger

roger.clotet@campusviu.es

Gracias



viu

Universidad
Internacional
de Valencia

universidadviu.com

De:



Planeta Formación y Universidades