



Más allá de la digitalización: El rol del acceso a información lexicográfica en Internet

Contenido

- I. Introducción
- II. Acceso a datos lexicográficos:
 - Perspectiva del usuario
 - Perspectiva del investigador
- III. Acerca de (Web) APIs para datos lexicográficos:
 - ¿Qué es una (Web) API?
 - C-SALT APIs for Sanskrit Dictionaries
 - VedaWeb
 - Kosh: APIs para diccionarios
- IV. Preguntas y respuestas

I Introducción : El término “digitalización”

- El **escaneo** de imágenes es un paso fundamental en la digitalización de un diccionario.
- *Pero* mientras el **texto** contenido en las imágenes no se haya **extraído** y **estructurado**, no se puede hablar de digitalización propiamente tal.

I Introducción: Más allá

- Digitalización = Retrodigitalización
- Diccionario o información lexicográfica *nacida* digital, no es parte de este concepto.
- El foco de esta charla es el acceso a información lexicográfica, ya sea producto de la digitalización de diccionarios impresos o de origen digital.

II Acceso a info. lexicográfica: Usuarios

- Aplicaciones web (desktop)
- Aplicaciones móviles (específicas para smartphones)
- Desde otras aplicaciones (smartphones, eReaders, desktop)

II Acceso a info. lexicográfica: Investigadores

- Aplicaciones web (desktop)
- Repositorios de datos (GitHub, o portales especializados)
- (Web) Application Programming Interfaces (APIs)

III ¿Qué es una (Web) API?

“A server-side web API is a programmatic interface consisting of one or more publicly exposed endpoints to a defined request–response message system, typically expressed in JSON or XML, which is exposed via the web—most commonly by means of an HTTP-based web server” (Fuente: https://en.wikipedia.org/wiki/Web_API)

III C-SALT APIs for Sanskrit Dictionaries I

- Las C-SALT (*Cologne South Asian Languages and Texts*) APIs for Sanskrit Dictionaries están siendo desarrolladas como parte del proyecto VedaWeb.
- El objetivo de las APIs es ofrecer los diccionarios más importantes del portal [Cologne Digital Sanskrit Dictionaries](#) (36 diccionarios), la fuente lexicográfica en sánscrito más importante a nivel mundial.

III C-SALT APIs for Sanskrit Dictionaries II

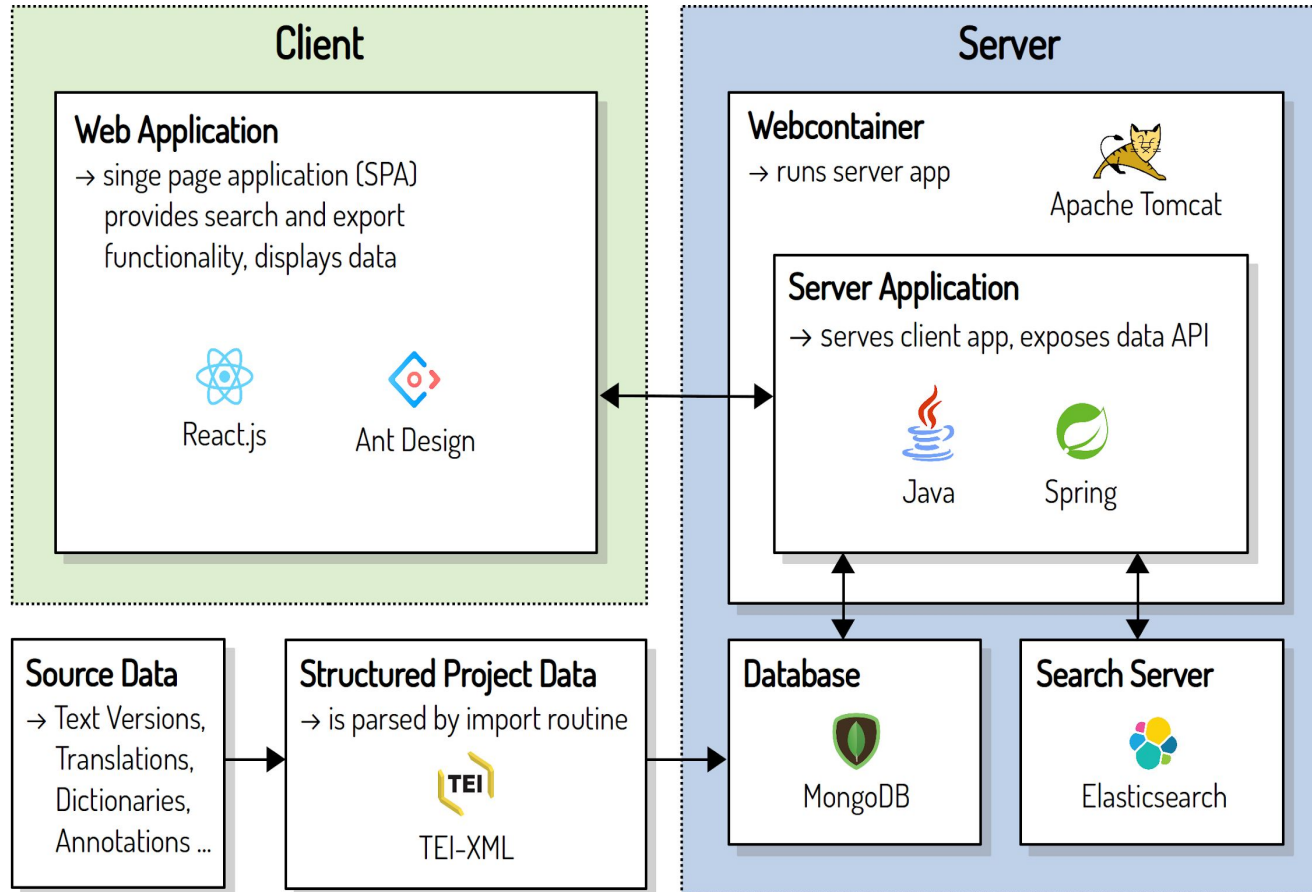


III VedaWeb I

- VedaWeb tiene como objetivo ofrecer una plataforma de investigación para textos védicos. El texto piloto es el Rigveda (RV).
- El RV es uno de los textos indoeuropeos más antiguos, siendo el más antiguo de los textos indoarios (fines segundo milenio a. C. aprox.).
- El RV contiene aprox. 160.000 palabras. La Ilíada y la Odisea juntas contienen aprox. 190.000 palabras.

III VedaWeb II

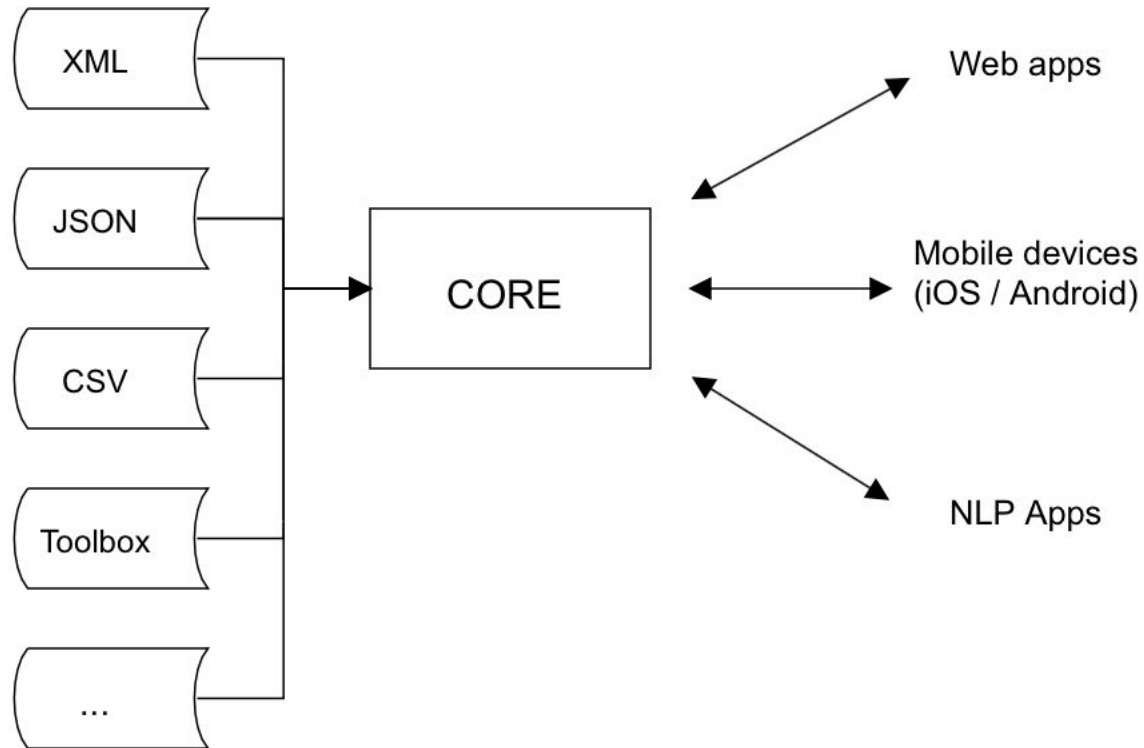
VedaWeb Application Architecture



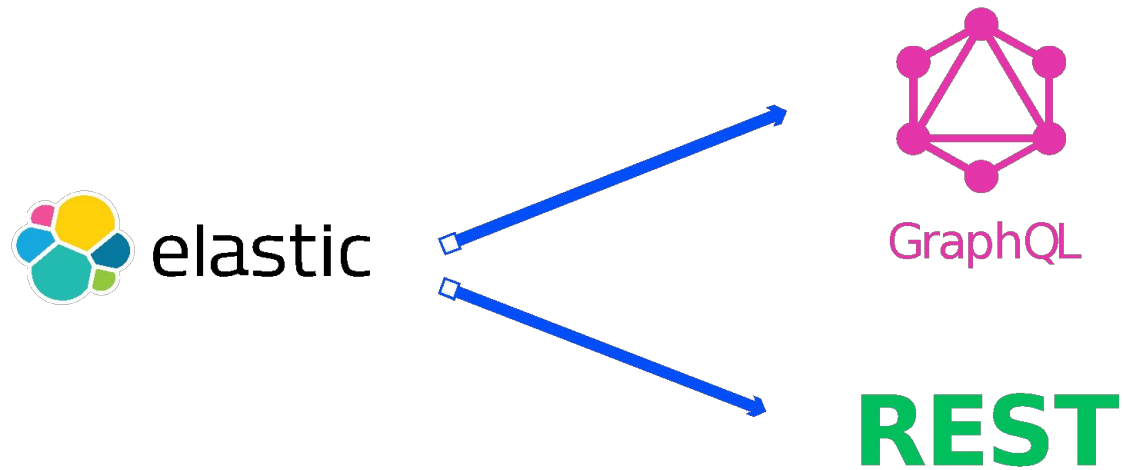
III - Kosh: APIs para diccionarios I

- ¿Cómo crear un framework genérico de APIs para diccionarios?
- Problema: información lexicográfica en diferentes formatos

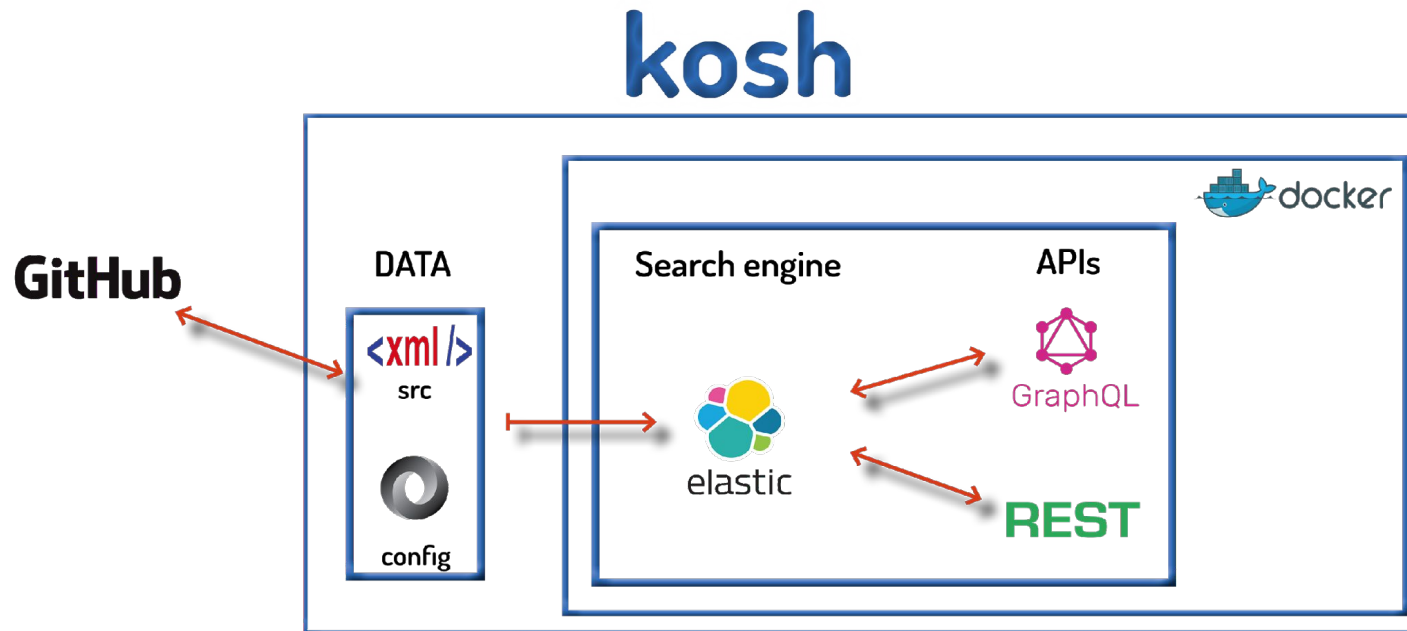
III - Kosh: APIs para diccionarios II



III - Kosh: APIs para diccionarios III



III - Kosh: APIs para diccionarios IV



IV – Preguntas y respuestas

¡Muchas Gracias!