

PROYECTO DEL CURSO: INTEGRACIÓN DE FUENTES BIBLIOGRÁFICAS

Integración e Interoperabilidad

Curso 2020-21

Objetivo

- Desarrollar una aplicación que permita la búsqueda federada en diferentes recursos de información bibliográfica
 - Apariencia de recurso único
 - Empleando técnicas de integración estudiadas en clase
 - Empleando arquitectura orientada a servicios
 - Explotando tecnologías diversas

Enunciado

- Se desea realizar un integrador de información bibliográfica, que permita a un usuario realizar búsquedas por diferentes criterios (autor, título, fecha, palabras clave), de manera transparente, en un conjunto de repositorios.

Fuentes de datos



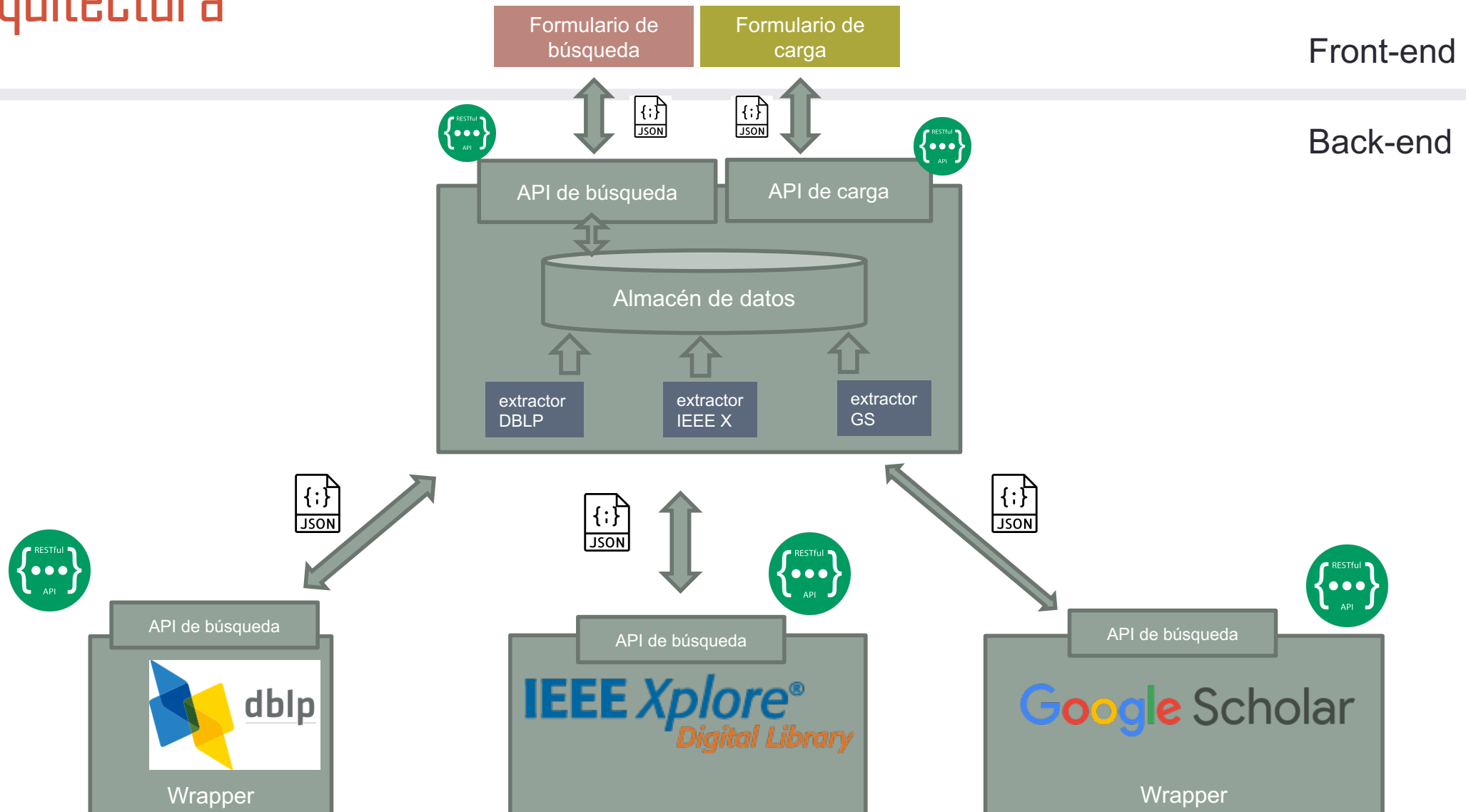
Colección de referencias de informática. Disponible via Web, vía API REST y como archivo XML.



Biblioteca digital del Institute of Electric and Electronics Engineers. Accesible vía API REST.

Google Scholar Buscador especializado en publicaciones académicas. Accesible vía Web.

Arquitectura



Se os facilitará...

- Vídeos de apoyo:
 - Construcción de APIs RESTful y creación de un cliente RESTful
 - Extracción de datos de fuentes semiestructuradas mediante Selenium
 - Manipulación de documentos XML mediante código JAVA
- Esquema global con el que crear el almacén de datos
- Acceso a herramientas de apoyo (Altova)
- Acceso a fuentes de datos:
 - Archivo XML con los datos de DBLP
 - URL de la API de IEEE Xplore: <https://developer.ieee.org>
 - Registrarse -> API Key
 - URL de Google Scholar: <https://scholar.google.com>
 - Construir *screen scrapper* con Selenium

Se os pide:

1. Diseñar y crear el almacén de datos basándoos en el esquema global
2. Diseñar e implementar el formulario y la API de búsqueda
3. Diseñar e implementar el formulario y la API de carga
4. Inferir los esquemas de las fuentes de datos
5. Definir los *mappings* semánticos entre los esquemas de las fuentes y el esquema global
6. Implementar los extractores de cada una de las fuentes
7. Diseñar e implementar las API REST que envuelvan DBLP y Google Scholar

Organización del trabajo

- Grupos de 4 personas
- Organizados en canales privados Teams
- Reuniones grupales durante las sesiones de prácticas
- Foco en los aspectos de ingeniería de datos: *mappings*, etc.
- No soporte a aspectos básicos de programación
- Lenguaje de programación y entorno de desarrollo de vuestra elección (se recomienda Java/Eclipse)

Hitos del proyecto

SEMANA	F. INICIO	PRÁCTICAS	Entregas
1	14/09		
2	21/09		
3	28/09		
4	05/10	(S) Enunciado del proyecto. Discusión sobre objetivos y alcance	
5	12/10	(SG) Elaboración esquemas fuentes y mappings	
6	19/10		
7	26/10		
8	02/11	(SG) Implementación extractores 1 (DBLP)	
9	09/11	(SG) Implementación extractores 2 (Scholar)	
10	23/11	(SG) Implementación extractores 3 (IEEE)	
11	30/11		Entrega 1: extractores en marcha
12	07/12	(SG) Implementación API (1)	
13	14/12	(SG) Implementación API (2)	
14	22/12	(SG) Implementación API (3)	
15	11/01		Entrega 2: Demo final/evaluación

(S): SÍNCRONA
(SG): SÍNCRONA POR GRUPOS