



Construyendo un Sistema de Indexación y Búsqueda de Recursos Georreferenciados

Arturo Beltran, Laura Díaz, Joaquín Huerta



Motivación



Objetivo







YAHOO!



d m o z open directory project



WebCrawler®



altavista™
SEARCH SOFTWARE

YAHOO!®

Google






GeoNetwork
Opensource



➔ Metodología 

➔ Sistema de indexación y búsqueda de recursos

- + Recopilación mediante aplicaciones de tipo *crawler*
 - + Descripción mediante la generación automática de metadatos
 - + Indexación en base a sus características
 - + Publicación
 - + Permitir el descubrimiento a través de mecanismos de búsqueda
- 



➔ <http://nutch.apache.org>

+ *Crawler*

+ Proyecto Apache



+ Basado en Lucene

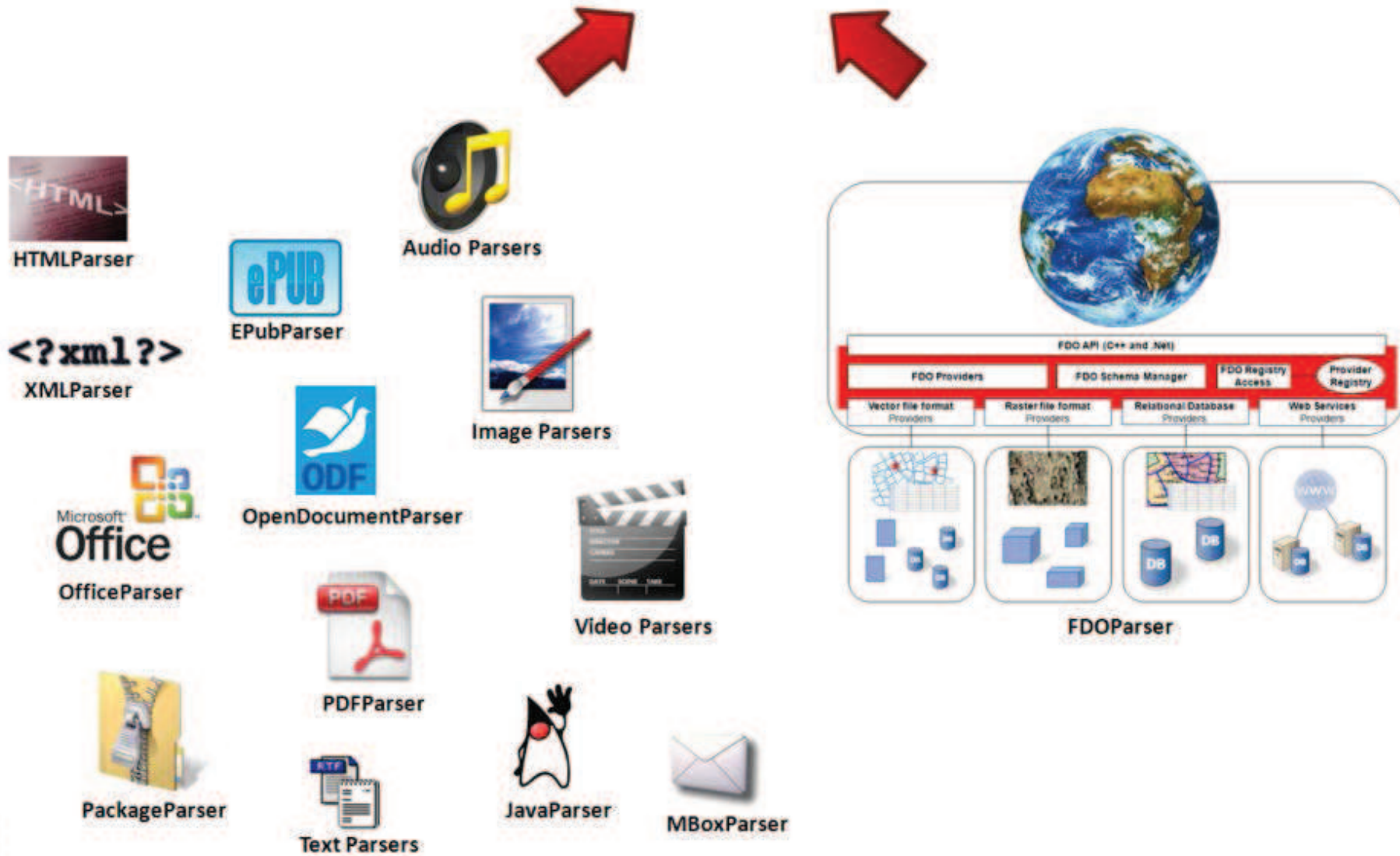


+ Java



+ Extensible (filtrado, indexación, procesamiento)

<Apache Tika/>



Apache

Solr



➔ <http://lucene.apache.org/solr>

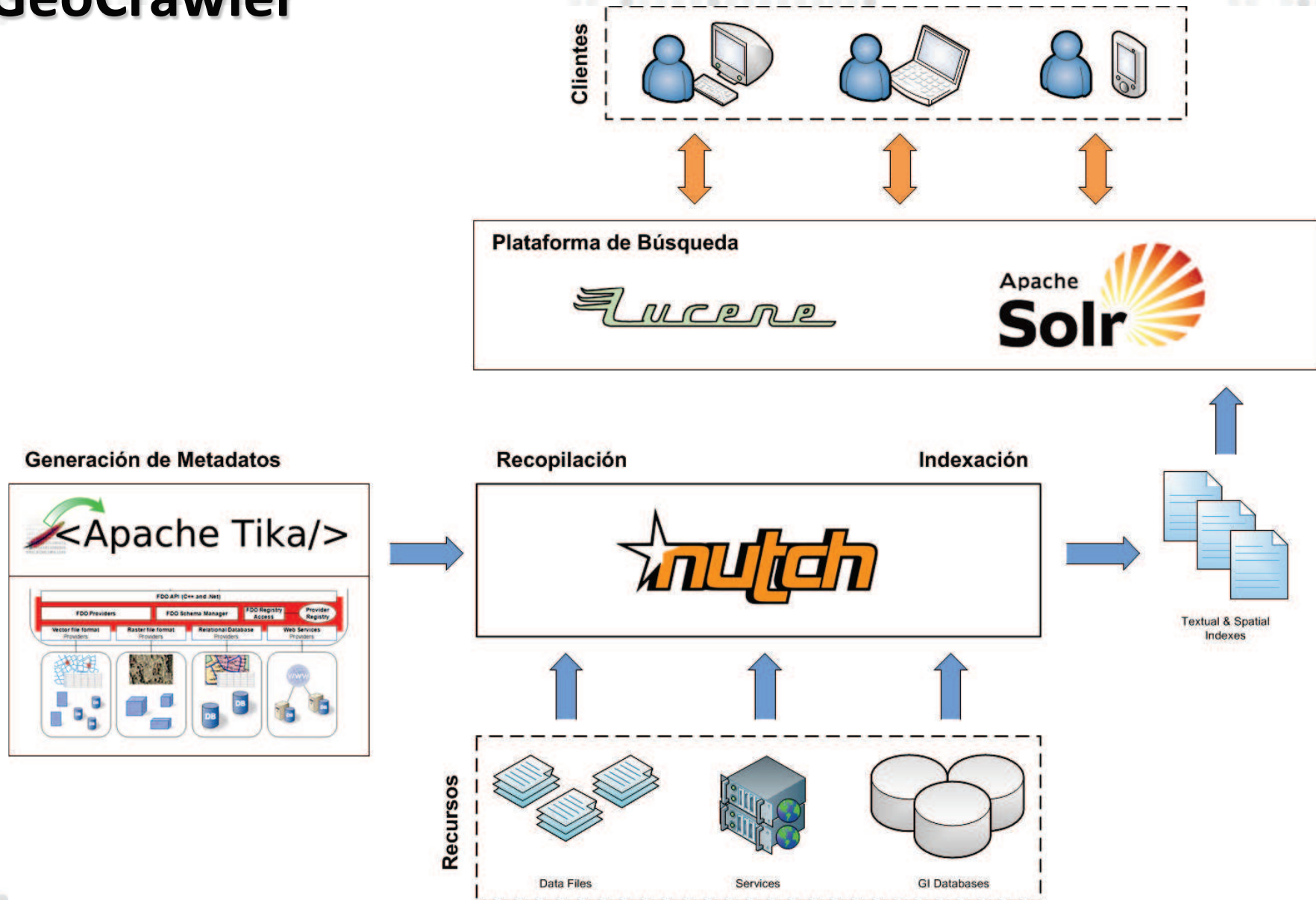
- + Plataforma de búsqueda
- + Proyecto Apache
- + Basado en Lucene
- + Java
- + Ejecución en Tomcat
- + Interfaces REST (HTTP/XML y JSON)
- + Capacidades de búsqueda avanzadas
- + Eficiente, escalable y extensible



Lucene



GeoCrawler





Entorno

→ Video 1

Estructura Directorios

- ExampleData
 - test
 - dir
 - hospitales.shp
 - cuatro.txt
 - tres.txt
 - tres.old
 - world_borders.shp
 - not.valid
 - manzanas.shp
 - uno.txt
 - testdir
 - test2
 - dir2
 - cuatro2.txt
 - hola.dat
 - tres2.txt
 - origen.shp
 - paradas_TSP.shp
 - uno2.txt
 - file.test

15 recursos

11 válidos

6 txt

5 SHP

Recopilación, Generación de Metadatos e Indexación

 **Video 2**

Recursos Indexados

- ExampleData
 - test
 - dir
 - hospitales.shp
 - cuatro.txt
 - tres.txt
 - ~~tres.old~~
 - world_borders.shp
 - ~~not.valid~~
 - manzanas.shp
 - uno.txt
 - testdir
 - test2
 - dir2
 - cuatro2.txt
 - ~~hola.dat~~
 - tres2.txt
 - origen.shp
 - paradas_TSP.shp
 - uno2.txt
 - ~~file.test~~

11 recursos indexados

6 txt

5 SHP

~~4 recursos NO indexados~~

Metadatos Indexados

| Campo | Tipo | Descripción |
|-------------------------|--------|--|
| id | String | Identificador único del recurso (coincide con su URL) |
| host | URL | URL del host del recurso |
| site | String | Site del recurso |
| tstamp | Date | Fecha de indexación |
| anchor | String | Enlaces que contiene el recurso |
| type | String | <i>MIME Type</i> del recurso |
| contentLength | Long | Tamaño del archivo (si lo es) |
| date | Date | Fecha de creación del recurso |
| gc_url | URL | URL del recurso |
| gc_name | String | Nombre del recurso |
| gc_path | URL | Ruta de acceso del recurso |
| gc_type | String | Tipo de recurso |
| gc_lastModified | Date | Fecha de modificación |
| gc_length | Long | Tamaño |
| gc_description | Text | Descripción del recurso |
| gc_title | Text | Título del recurso |
| gc_keywords | Text | Palabras clave |
| gc_publisher | Text | Información acerca del autor del recurso |
| gc_schema | URL | Esquema según el tipo del recurso |
| gc_source | URL | Origen del recurso |
| gc_resourceName | String | Nombre del principal esquema interno del recurso |
| gc_resourceDescription | Text | Descripción del esquema interno del recurso |
| gc_resourceTitle | Text | Título del esquema interno del recurso |
| gc_resourceMinX | Double | Coordenada X_{min} del BBOX |
| gc_resourceMinY | Double | Coordenada Y_{min} del BBOX |
| gc_resourceMaxX | Double | Coordenada X_{max} del BBOX |
| gc_resourceMaxY | Double | Coordenada Y_{max} del BBOX |
| gc_resourceCRS | String | Nombre del sistema de coordenadas del recurso |
| gc_resourceGeometry | String | Tipo de geometría del recurso |
| gc_resourceNumRecords | Int | Número de registros del recurso |
| gc_resourceRestrictions | String | Restricciones del recurso |
| gc_resourceHints | String | Consejos del recurso |
| gc_attributeNames | String | Recopilación de los nombres de las <i>features</i> |
| gc_completeXML | String | XML completo generado por la plataforma de extracción de metadatos |

Plataforma de búsqueda

 **Video 3**

Clientes

The screenshot displays the GEOS web application interface. On the left side, there is a search panel with the following sections:

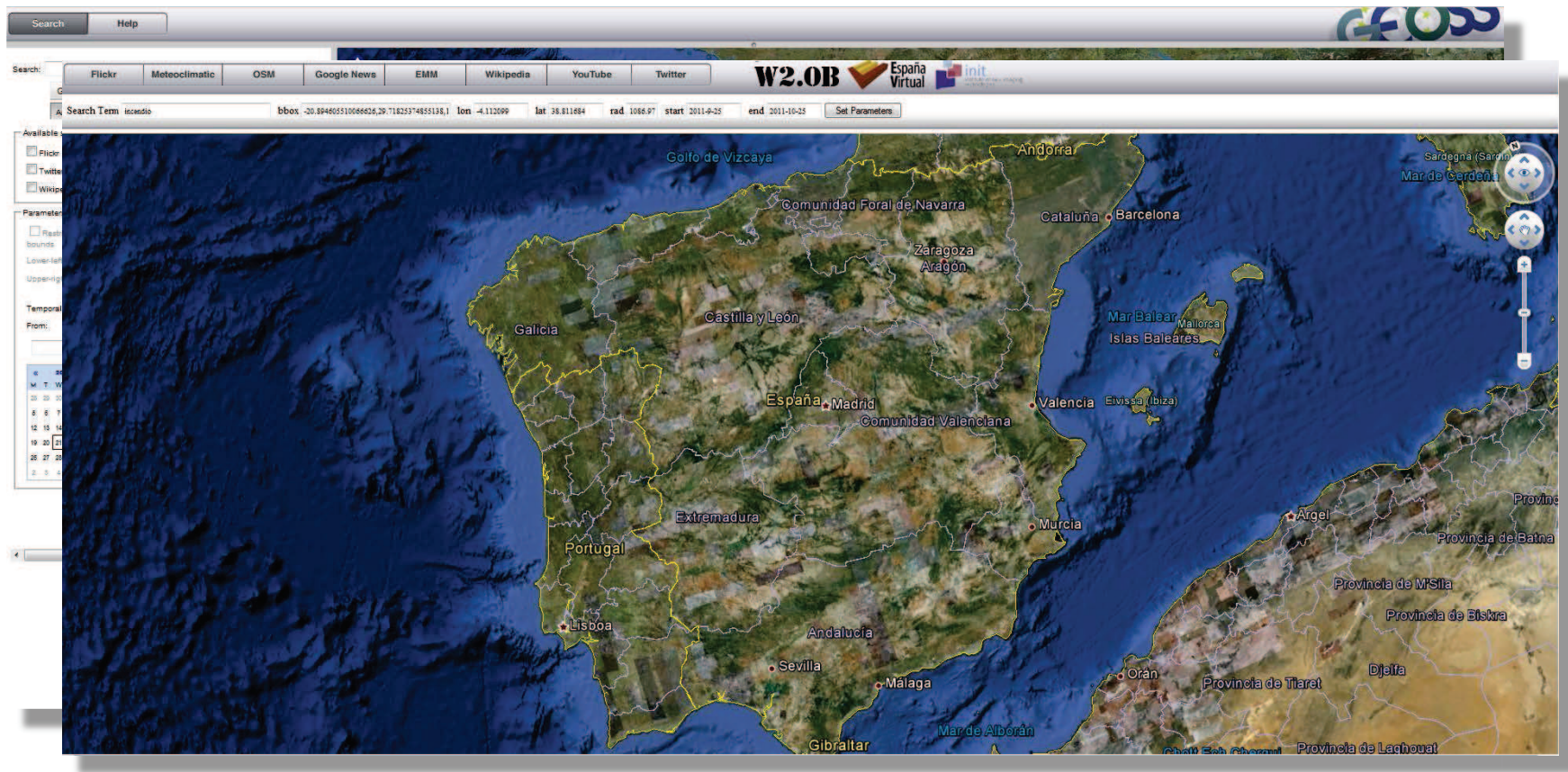
- Search:** Includes a search input field, a "Go!" button, and checkboxes for "Geotagged only" and "Add to results". Below this is an "Advanced Search" button.
- Available services:** A grid of checkboxes for various services: Flickr, YouTube, EMM, Twitter, Meteorclimatic, Google news, Wikipedia, and Eurogeos Broker.
- Parameters:** A checkbox for "Restrict search area to map bounds", followed by input fields for "Lower-left" and "Upper-right" coordinates.
- Temporal Search:** Fields for "From:" and "To:" dates, with two calendar pop-ups for December 2011 and December 2010.

The main area of the interface is a satellite map of Europe and the Mediterranean region. A legend in the top right corner, titled "Base Layer" and "Overlays", lists the following options:

- Base Layer: GEOS, Google map
- Overlays: Fires (entire season), Burned areas (entire season), Hot spots (entire season), FWI (today), Fires News, LAC (Disregarded)

At the bottom left of the map, there is a "Google" logo and the text "Imágenes ©2011 - Términos de uso".

Cientes



Cientes

The screenshot displays a web application interface for geospatial data analysis. The main window shows a satellite map of Madrid, Spain, with several overlays and search results. The interface includes a search bar at the top, a navigation pane on the left, and a main map area with various data layers and search results.

Search Results:

- Search Term:** iscedio
- Search Results:** Madrid (Madrid, Spain, ES).
- Search Term:** hotels
- Search Results:** The Westin Palace Hotel Madrid

Map Information:

- Altitude:** 20 km
- Lat:** 40.4500°
- Lon:** -3.7338°
- Elev:** 627 meters

Layers:

- Stars
- Atmosphere
- NASA Blue Marble Image
- Blue Marble (vNS) 2004
- RoadNet Landrail
- USDA NAIP
- USGS Urban Area Ortho
- Place Names
- World Map
- Scale bar
- Compass
- Terrain profile graph
- OpenSearch (hotels)

Twitter:

- Meteo Espagne - Madrid (11°C) #Espagne #Meteo**

The Westin Palace Hotel Madrid:

Famous hotel and always in the shadow of the Ritz in Madrid since it actually opened in 1912.

Conclusiones



Metodología



METADATOS

- + Pieza fundamental
- + Transparentes al usuario



Sistema de indexación y búsqueda de recursos

- + *Crawling* e indexación eficiente
- + Descripciones homogéneas
- + Potente interfaz de búsqueda



¿Preguntas?

Muchas gracias

