

Servicio de nomenclátor con el motor de búsqueda Solr.

Caso práctico en la provincia de Lugo.

Diego Alberto Arias Prado (diegoalberto.arias@usc.es)

Manuel Pérez Gómez (manuel.perez.gomez@usc.es)

Francisco Velayos Pardo (f.velayos@deputacionlugo.org)

Rafael Crecente Maseda (rafael.crecente@usc.es)

Laboratorio de Territorio (LaboraTe), G.I.-1934-TB, Universidad de Santiago de Compostela

<http://laborate.usc.es/>

Diputación Provincial de Lugo

<http://www.deputacionlugo.org/>



masterra



LaboraTe
G.I.-1934-TB

A historical map of the Iberian Peninsula, specifically focusing on Galicia and the Cantabrian region. The map is titled 'OCEANO CANTABRICO' at the top and 'GALICIA' in the upper left. It shows geographical features like rivers, coastlines, and various towns. The map is overlaid with a grid of latitude and longitude lines. The text 'OCCIDENTAL' is visible on the left edge, and 'PROVINCIA DE LUGO' is visible on the right edge. The map is the background for the text.

Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

- **Un servicio de nomenclátor se define como uno que admite como entrada el nombre de un fenómeno y devuelve su ubicación geográfica mediante unas coordenadas.**
- **Propuesta de estándar de nomenclátor: WFS-G.**

A historical map of the Iberian Peninsula, specifically focusing on Galicia and the Cantabrian region. The map is titled 'OCEANO CANTABRICO' at the top. A box in the upper left corner is labeled 'GALICIA' and contains text. The map shows geographical features like rivers, mountains, and coastlines. The text 'OCCIDENTAL' is on the left edge, and 'PROVINCIA DE LUGO' is on the right edge. The map is overlaid with a semi-transparent text box containing the main title and bullet points.

Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

- **Objetivo: crear un servicio de nomenclátor que devuelva la ubicación de las entidades singulares de la provincia de Lugo.**
- **En Galicia – y también en Lugo - la población vive en un gran número de entidades singulares, muchas de ellas de carácter diseminado.**

Servicio de nomenclátor con Solr.

Caso práctico provincia de Lugo.

	Lugo	Salamanca
población	355.195 (~0,76% total España)	352.986
extensión (km²)	9.859 (~1,95% total España)	12.349
entidades singulares	9.776 (~15% total España)	987
entidades diseminadas	8.596 (~21% total España)	712
carreteras prov. (km)	~4.200	~2.415

fuentes de datos:

- *Instituto Nacional de Estadística*
- *Encuesta de Infraestructura y Equipamientos Locales de la provincia de Lugo, año 2011*
- *Diputación Provincial de Salamanca, Plan Provincial Carreteras 2006 - 2012*

A historical map of the Iberian Peninsula, specifically focusing on Galicia and the Cantabrian region. The map is titled 'OCEANO CANTABRICO' at the top and 'GALICIA' in a box on the left. It shows geographical features like rivers, mountains, and coastlines. The text 'OCCIDENTAL' is on the left edge, 'OCCIANO' is on the left side, and 'ASTURIAS' and 'REYNO' are visible on the map. The map is aged and has a yellowish tint.

Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

- **Necesidad de un servicio de nomenclátor.**
 - **Es uno de los servicios básicos que cualquier Infraestructura de Datos Espaciales debería implementar, de acuerdo al Consejo Superior Geográfico.**
 - **En el caso de Galicia, la búsqueda por nombre de entidades singulares puede ser especialmente difícil, debido a lo elevado de su número.**

The background of the slide is a historical map of Galicia and the Cantabrian Ocean. The map shows the coastline of Galicia, with various towns and rivers labeled. The text 'OCEANO CANTABRICO' is visible at the top, and 'GALICIA' is in the top left corner. The map is detailed with geographical features and text in a historical script.

Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

¿Qué es un motor de búsqueda *full-text*?

- Solr – basado en Lucene - es un sistema que hace búsquedas de tipo texto completo (*full-text*).
- En una búsqueda de texto completo, el motor de búsqueda examina todas las palabras en todos los documentos almacenados.
- Este paradigma supone una ruptura con los sistemas de bases de datos relacionales, en los que la información se encuentra muy estructurada.

A historical map of Galicia and the Cantabrian Ocean. The map shows the coastline of Galicia, with various towns and rivers labeled. The text 'OCEANO CANTABRICO' is visible at the top, and 'GALICIA' is in the top left corner. The map is detailed with geographical features and text in a historical script.

Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

- **¿Por qué un motor de búsqueda *full-text*?**
 - **Solución tradicional:** almacenar la información en una base de datos relacional con capacidades espaciales.
 - **Sin embargo:**
 - Esta tecnología ni es buena ni es rápida haciendo búsquedas de texto.
 - **Nuestra propuesta:** utilizar el motor de búsqueda *full-text* Solr, basado en la tecnología Lucene.

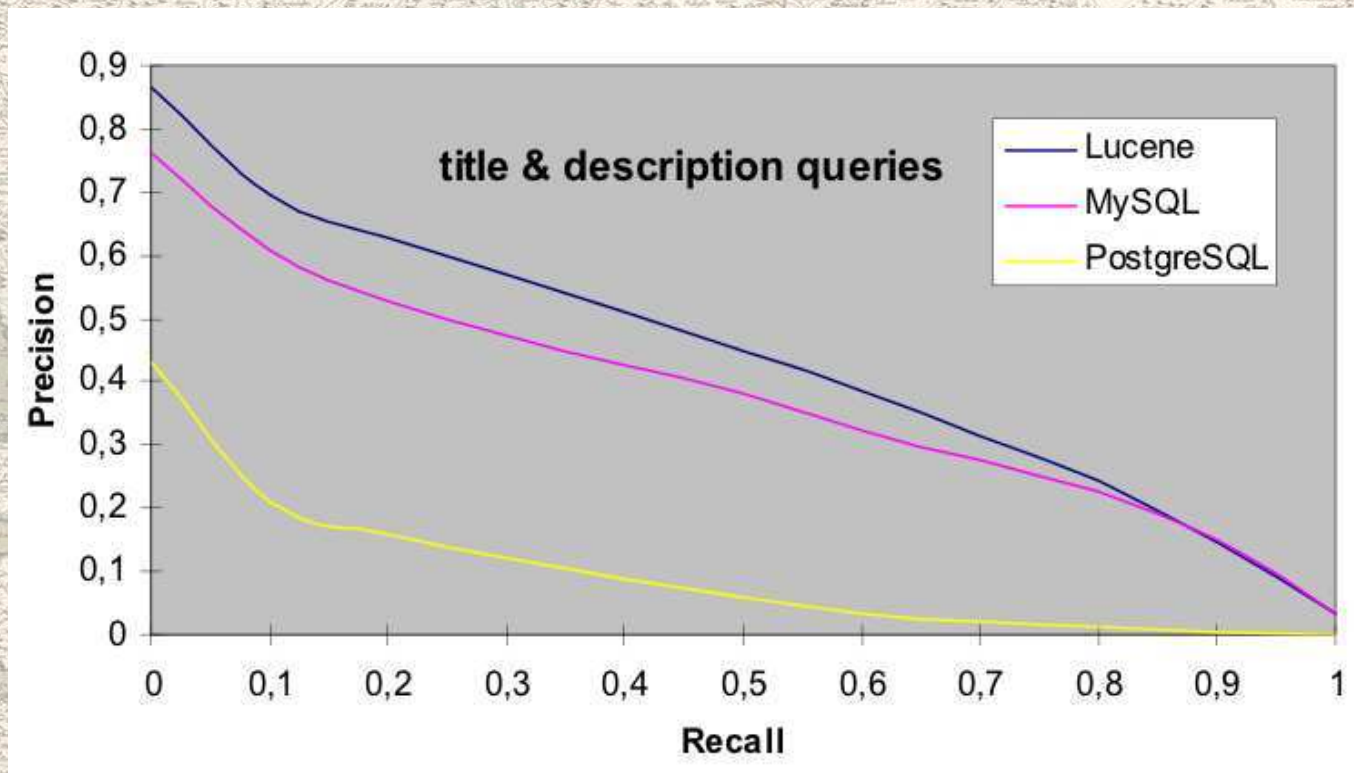
Servicio de nomenclátor con Solr.

Caso práctico provincia de Lugo.

- ¿Por qué un motor de búsqueda *full-text*?
 - Posibilidad de hacer búsquedas complejas.
 - Clasificación de los resultados.
 - Buen soporte para múltiples idiomas.
 - *Stop words*.
 - *Stemming*: reducción de palabras a su raíz.
 - Escalabilidad, gracias a la replicación eficiente de buscadores.
 - Posibilidad de hacer búsquedas distribuidas.
 - Paginación de resultados.

Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

- ¿Por qué un motor de búsqueda *full-text*?

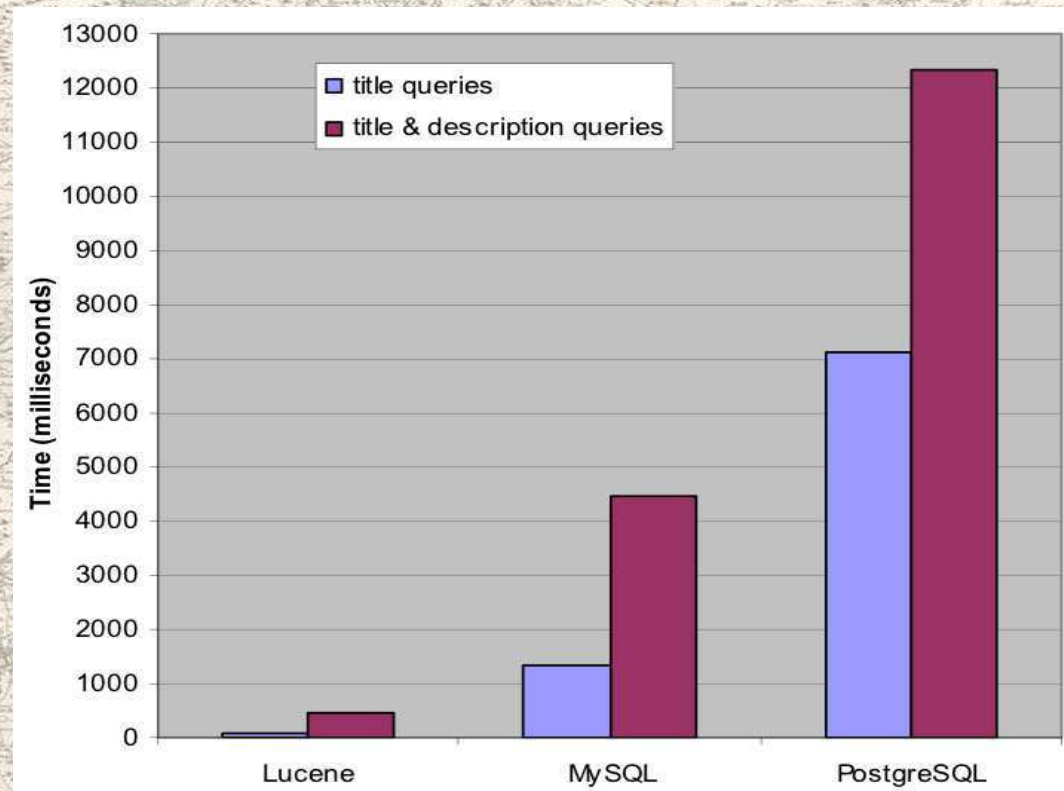


Lucene/Solr da resultados más precisos.

referencia: Arslan A., Yilmazel O.; 2008; "A comparison of Relational Databases and information retrieval libraries on Turkish text retrieval"; International Conference on Natural Language Processing and Knowledge Engineering

Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

- ¿Por qué un motor de búsqueda *full-text*?



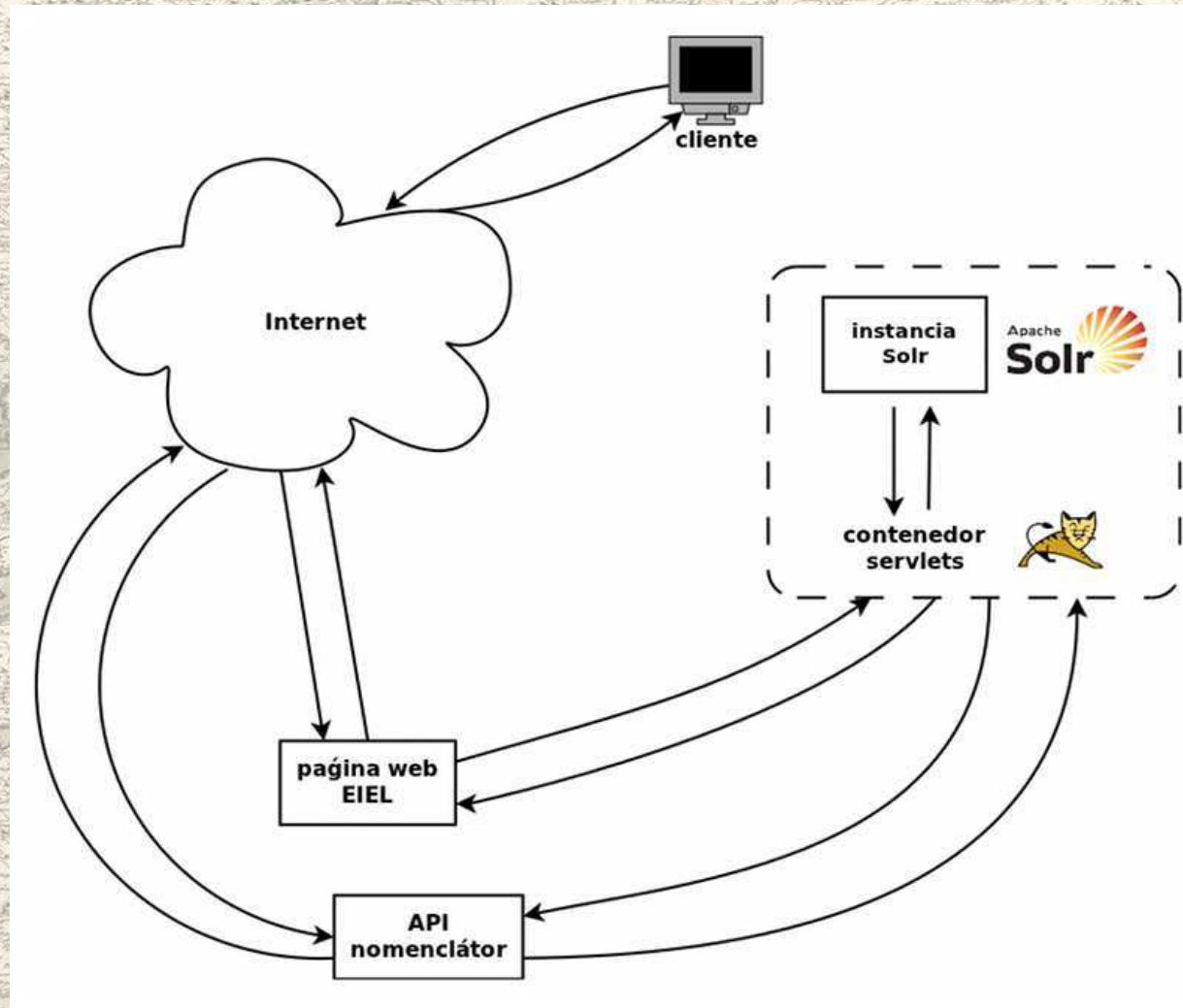
Con Lucene los tiempos de búsqueda son mucho menores.

referencia: Arslan A., Yilmazel O.; 2008; "A comparison of Relational Databases and information retrieval libraries on Turkish text retrieval"; International Conference on Natural Language Processing and Knowledge Engineering

Servicio de nomenclátor con Solr.

Caso práctico provincia de Lugo.

- Arquitectura del sistema



Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

galego  español

 inicio  WMS  WFS  EIEL  mapa web

IDEAL

Infraestructura de Datos Espaciales de la Administración Lucense

está usted en: inicio » nomenclátor

Inicio

visor de mapas

servicio WMS

servicio WFS

nomenclátor

metadatos 2009

Servicio de nomenclátor

El servicio de nomenclátor que ofrecemos permite la búsqueda de más de 9.400 entidades singulares distribuidas en toda la provincia de Lugo.

Para cada entidad singular encontrada se facilita la siguiente información: código INE, municipio, parroquia y coordenadas que permiten ubicarla geográficamente.

nombre entidad

buscar entidad en

» buscar

<http://www.idealugo.es/eiel2011/>

Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

 galego  español

 inicio  WMS  WFS  EIEL  mapa web

IDEAL

Infraestructura de Datos Espaciales de la Administración Lucense

está usted en: inicio » resultados de la búsqueda

resultados de la búsqueda de 'Vilar' en el municipio de Castroverde
4 entidades singulares encontradas

Inicio

visor de mapas

servicio WMS

servicio WFS

nomenciótor

metadatos 2009

acerca de la EIEL

EIEL 2010

EIEL 2009

EIEL 2008

EIEL 2005

EIEL 2000

EIEL 1995

Vilar

nombre en gallego

Vilar

nombre en castellano

Vilar

código INE

270110804

municipio

Castroverde

parroquia

Cellán de Calvos (San Salvador)

coordenadas WGS 84

lon = -7,31302 lat = 42,99913

(código EPSG 4326)

coordenadas ETRS 89 29N x = 637506,30515 y = 4762099,008689

(código EPSG 25829)

Vilar

nombre en gallego

Vilar

nombre en castellano

Vilar

código INE

270112029

municipio

Castroverde

parroquia

Montecubreiro (San Cibrao)

coordenadas WGS 84

lon = -7,316651 lat = 43,102871

(código EPSG 4326)

coordenadas ETRS 89 29N x = 636979,251139 y = 4773613,741043

(código EPSG 25829)

Servicio de nomenclátor con Solr.

Caso práctico provincia de Lugo.

- Información indexada.

```
<?xml version="1.0" encoding="utf-8" ?>
<add overwrite="true">
<doc>
<field name="id">95360</field>
<field name="ineCode">270010101</field>
<field name="spanishName">Abadín</field>
<field name="galicianName">Abadín</field>
<field name="centroid">SRID=25829;POINT(623678.046408781
4802438.68112146)</field>
<field name="envelope">SRID=25829;POLYGON((623314.279908025
4802146.92807464,623314.279745559
4802737.09325913,624049.428928126
4802737.09305599,624049.42909156 4802146.9278715,623314.279908025
4802146.92807464))</field>
<field name="xy">623678.046408781,4802438.68112146</field>
</doc>
```

...

A historical map of the Iberian Peninsula, specifically focusing on Galicia and the Cantabrian region. The map is titled 'OCEANO CANTABRICO' at the top and 'GALICIA' in a box on the left. It shows geographical features like rivers, mountains, and coastlines. The text 'OCCIDENTAL' is on the left edge and 'OCCIDENTAL' is on the right edge. The map is overlaid with a grid of latitude and longitude lines.

Servicio de nomenclátor con Solr. Caso práctico provincia de Lugo.

- **Líneas futuras de trabajo.**
 - **Hacer que el servicio cumpla con el estándar WFS-G.**
 - **Ampliar las búsquedas a otros accidentes geográficos.**
 - **Servicio de búsqueda de equipamientos e infraestructuras: saneamiento de aguas, alumbrado público, etc.**