

Repositorio Semántico de Investigadores del Ecuador

Fernando Baculima, José Luis Cullcay, Freddy Sumba, Xavier Sumba,
Mauricio Espinoza, Víctor Saquicela.
Departamento de Ciencias de la Computacion, Universidad de Cuenca, Ecuador

*Sexta Conferencia de Directores de Tecnología de Información, TICAL 2016
Buenos Aires, Septiembre, 2016*



Contenido

- Introducción
- Trabajos Relacionados
- Arquitectura
- Resultados
- Conclusiones

Introducción

Motivación

● Problemas

- empresas privadas y organizaciones públicas
 - Identificar intereses comunes entre investigadores
 - Ej. gobierno ecuatoriano → talento humano
 - buscar oportunidades de colaboración con equipos de investigación existentes
- nivel académico
 - Información de autores y publicaciones dispersa
 - DSpace de cada IES
 - DBLP, Scopus, Scholar, Academics

Situación actual

- ❑ Información fragmentada
- ❑ Procesos de búsqueda sintácticas
 - autor, área de conocimiento, etc.
- ❑ Modelos heterogéneos de acceso y publicación datos
 - APIs, Web, Bases de datos

-Publicaciones "Mauricio Espinoza"
-Filtrar manualmente los resultados
-Identificar otros investigadores en áreas similares



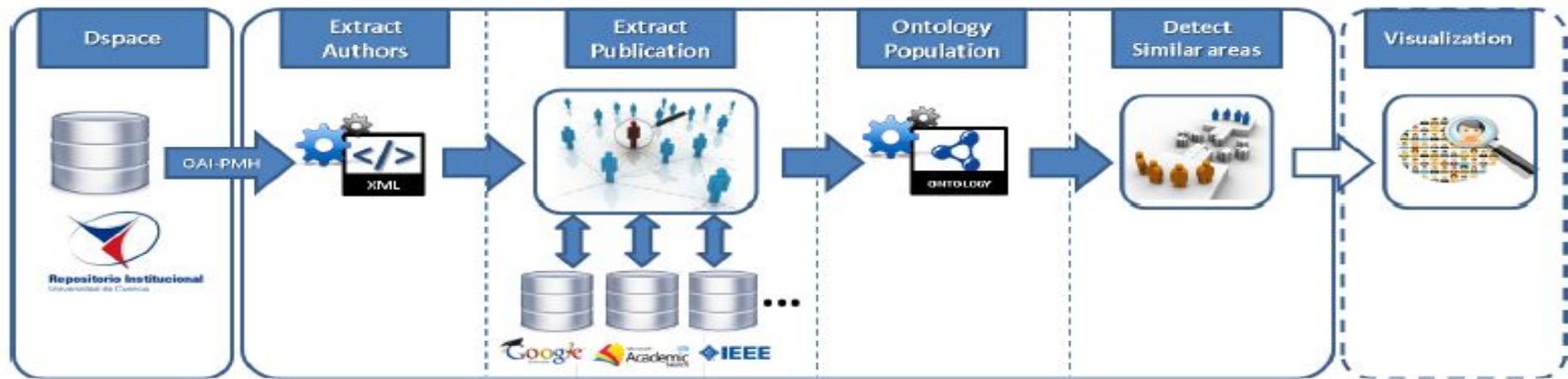
Trabajos relacionados

- ❖ Atanassova et al.
- ❖ Buter et al.
- ❖ Calero et al.
- ❖ Bergstrom et al.
- ❖ *U-papers.....*
 - Sin soporte a la identificación de áreas similares de conocimiento

- ❖ Mayoría de los trabajos
 - enfocadas en redes de colaboración explícitas
 - Por ejemplo, a través de referencias y citas de un determinado artículo

Arquitectura

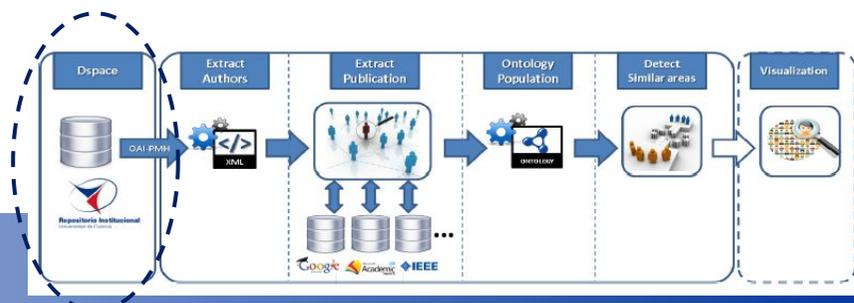
Componentes de la Arquitectura



<http://redi.cedia.org.ec>

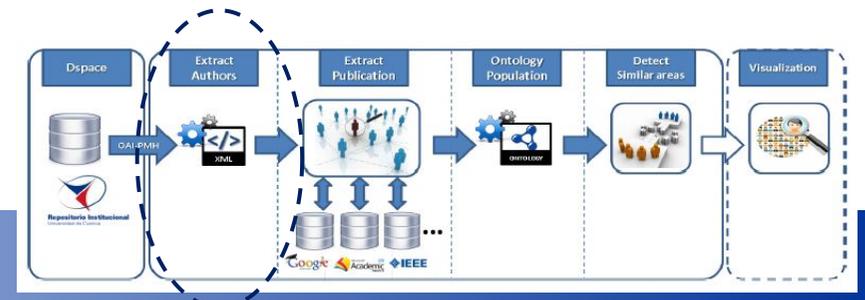
Fuentes de Datos de Autores (DSPACE)

- ❑ En DSPACE se encuentran Autores
 - No todos los autores en las IES son investigadores
- ❑ Extracción a través de OAI-PMH
 - proveedores de servicios cosechar metadatos de los proveedores de datos
- ❑ “Aplicación de Linked Open Data a Repositorios CEDIA”
 - Transformar información de DSPACE a RDF
 - Normas de linked data



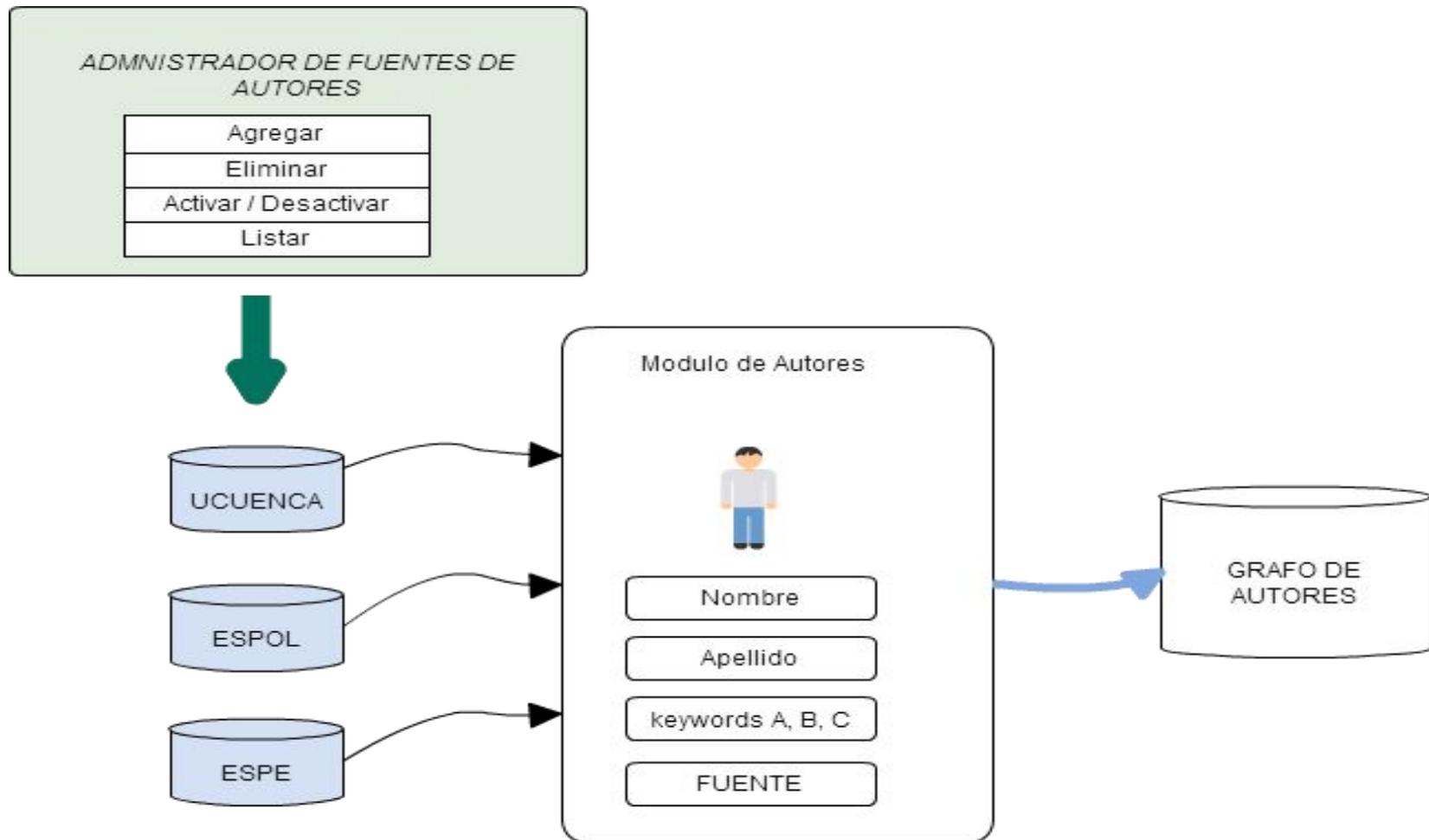
Extracción de Autores (i)

- ❖ Modelo único información de los autores
 - Nombre, Apellido, Keywords, etc
 - FOAF : <http://xmlns.com/foaf/spec/>
 - ontología legible para las máquinas
 - describe a las personas, sus actividades y sus relaciones con otras personas y objetos



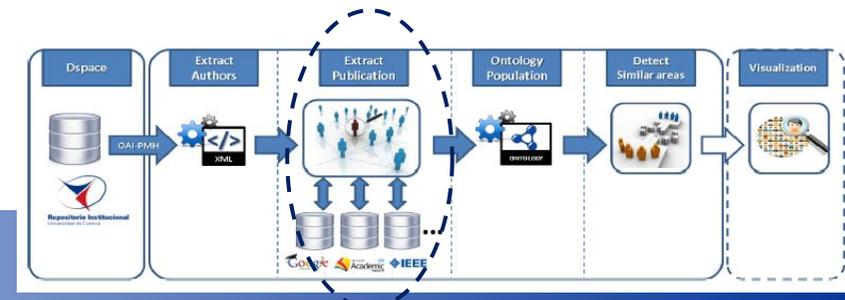
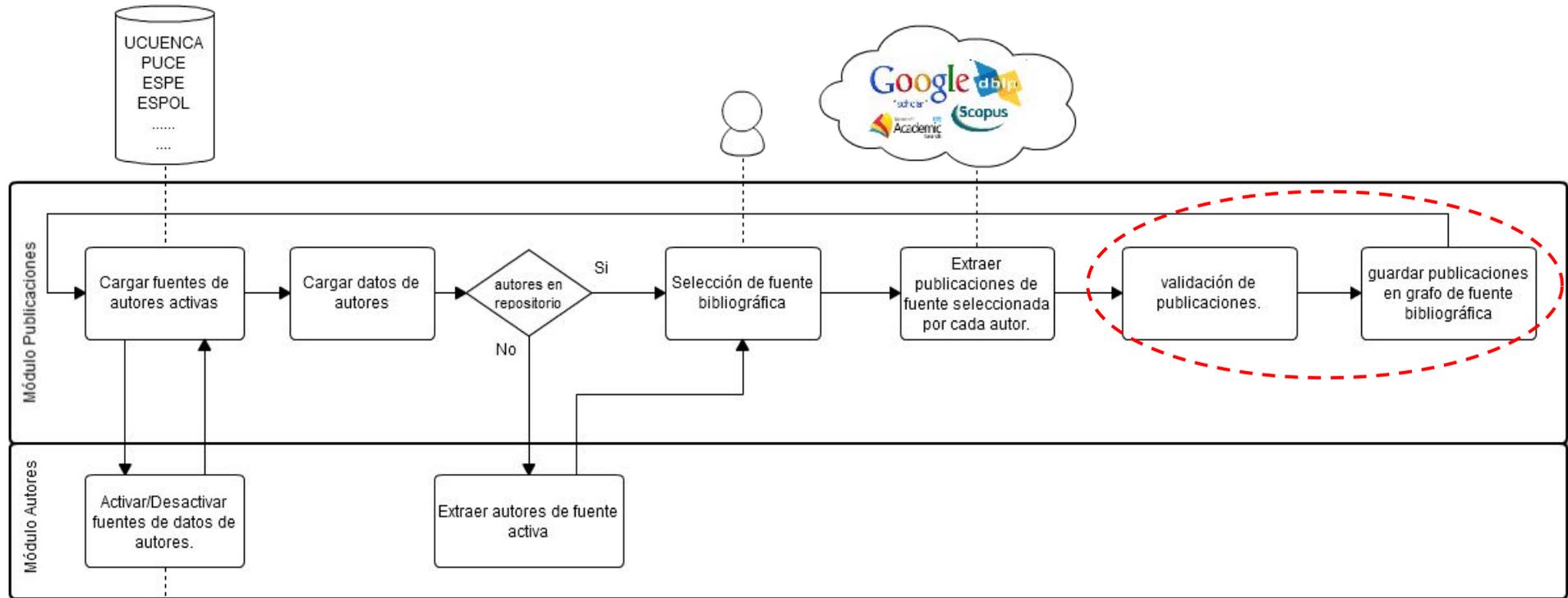
Extracción de Autores (ii)

❖ Métodos de actualización controlados

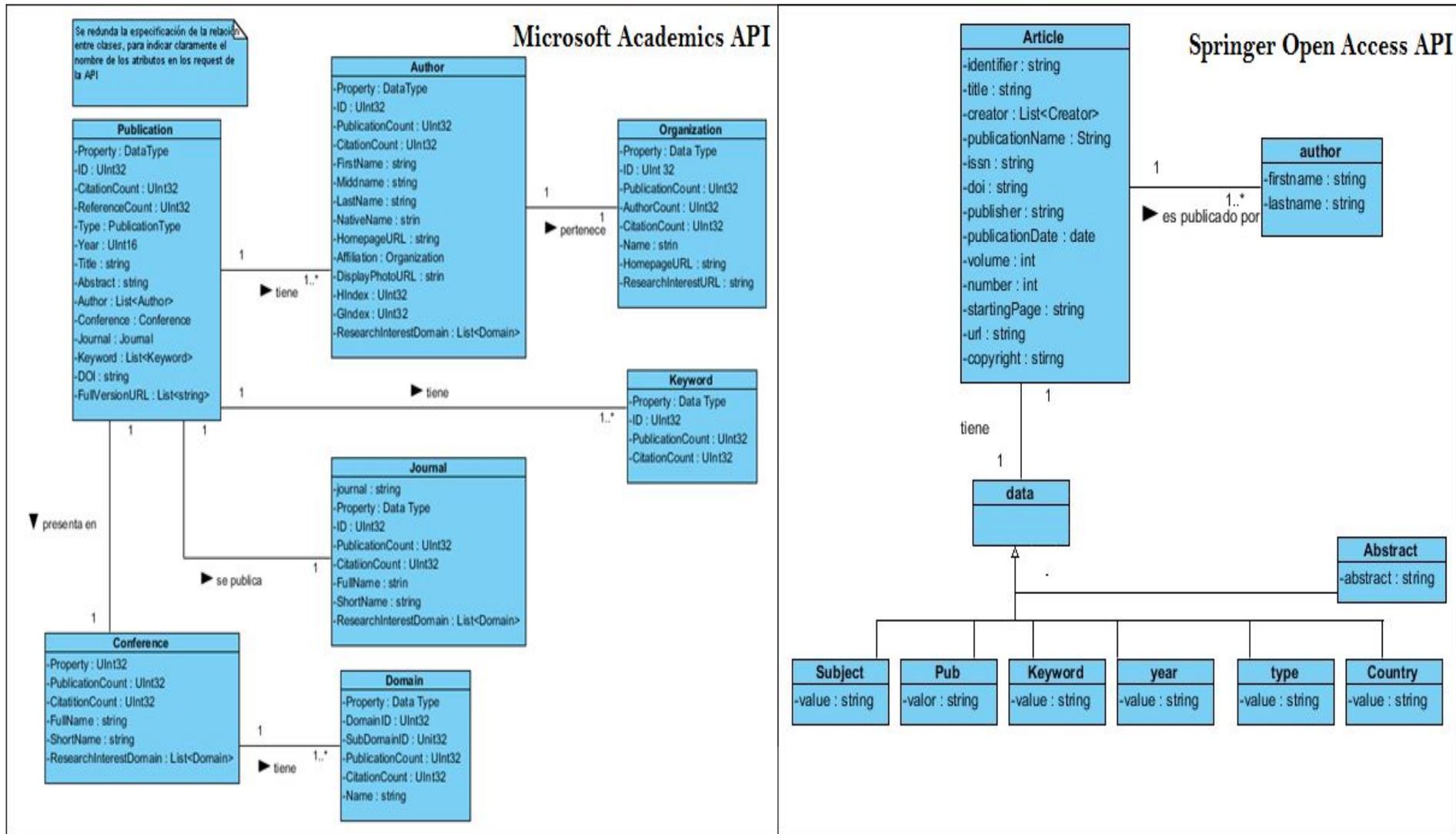


Extracción de Publicaciones (i)

- ❑ Se busca publicaciones por cada autor
 - SCOPUS, DBLP, etc.

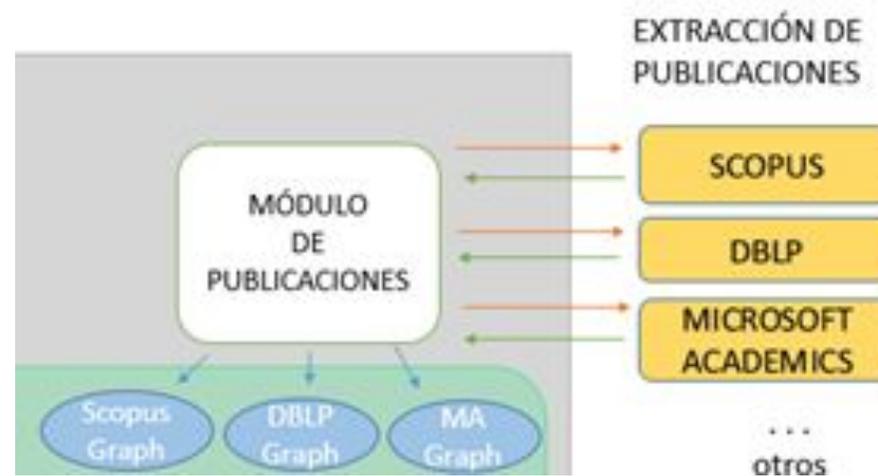


Extracción de Publicaciones (ii)



Extracción de Publicaciones (iii)

- Se almacena los datos en grafos separados por fuente de publicaciones

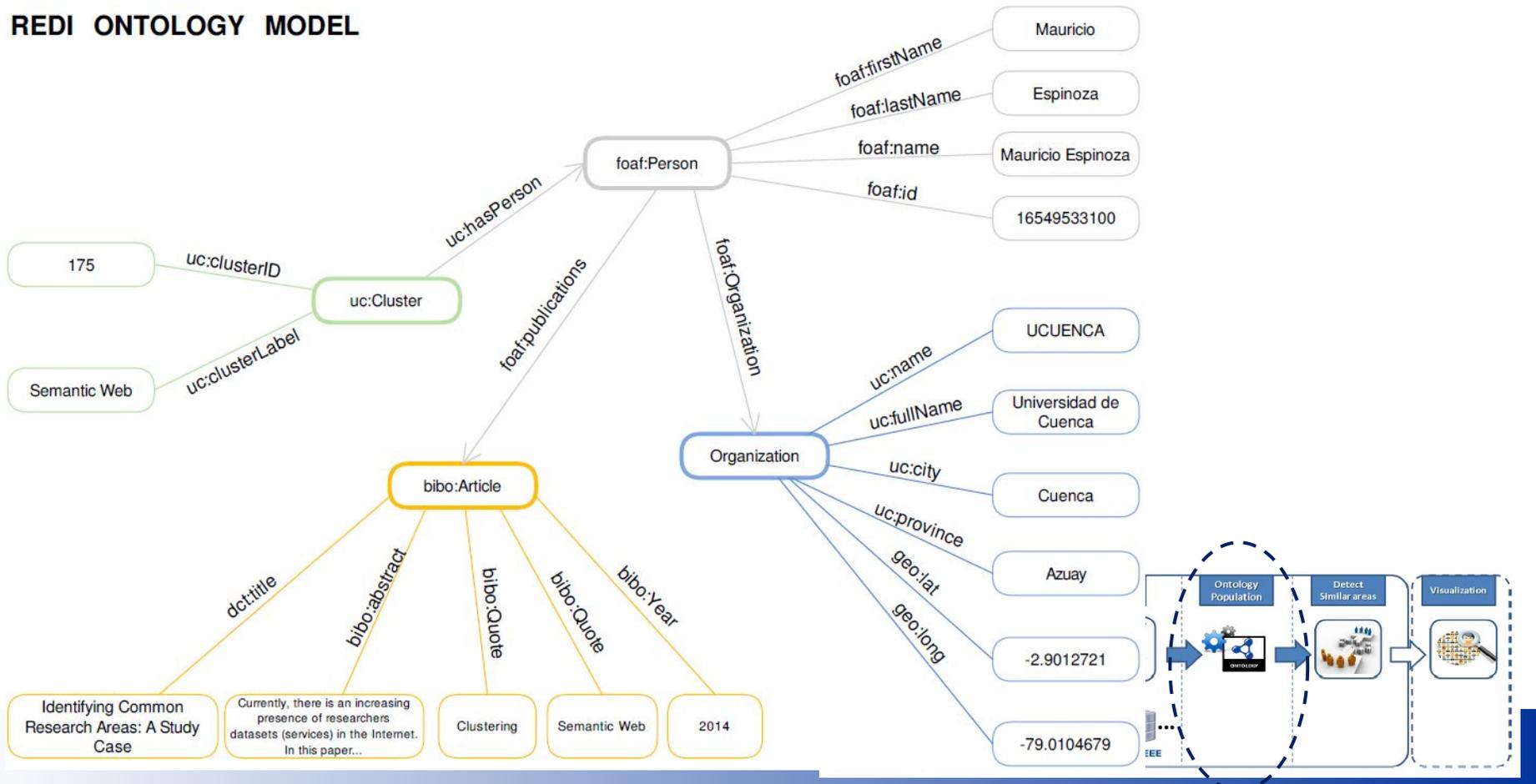


- modelo ontológico bibliográfico
- BIBO: <http://biblontology.com/>

Población de la Ontología (i)

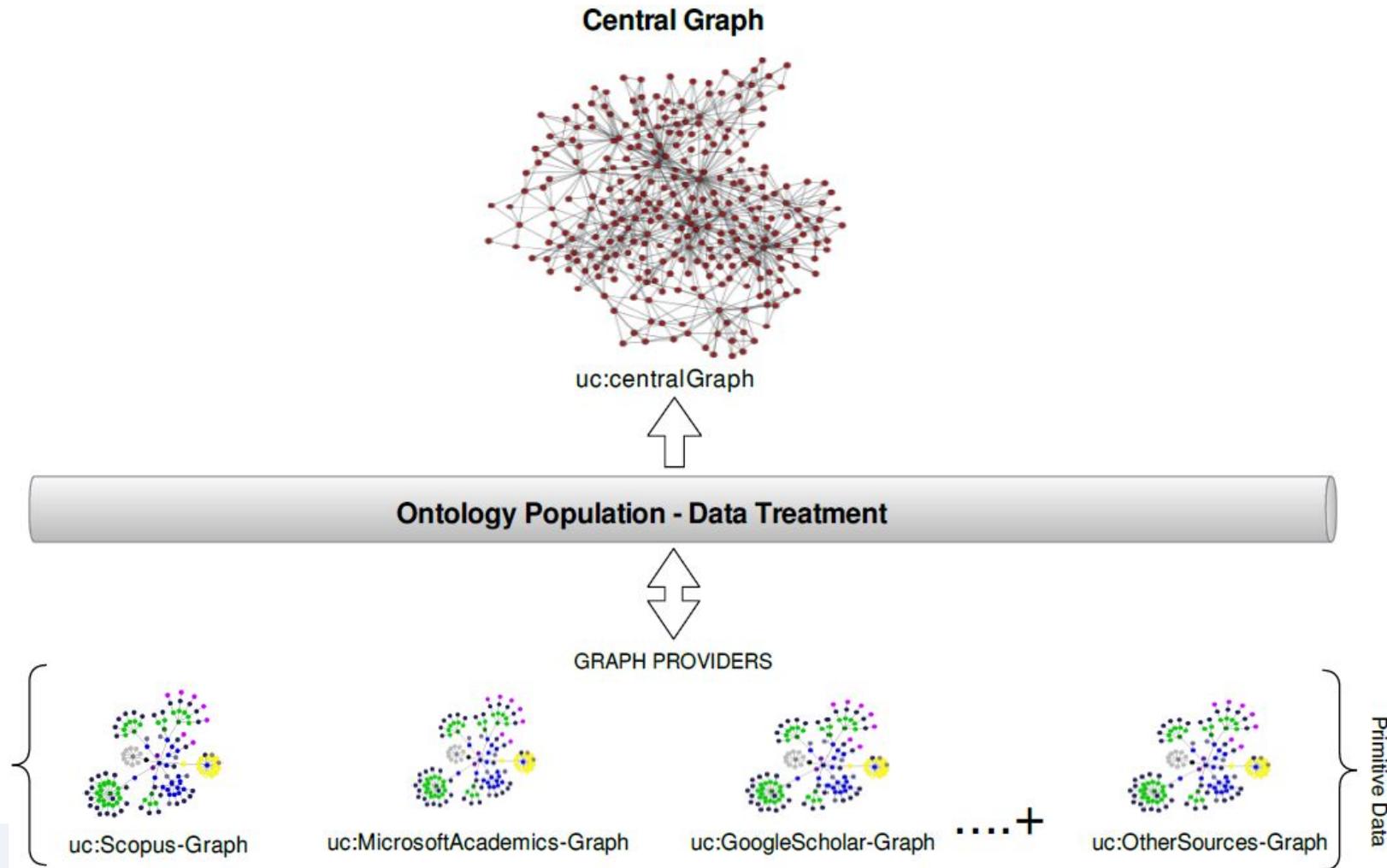
- ❑ Se almacena la información extraída en un Triple Store
 - Apache Marmotta
- ❑ Modelo ontológico común

REDI ONTOLOGY MODEL



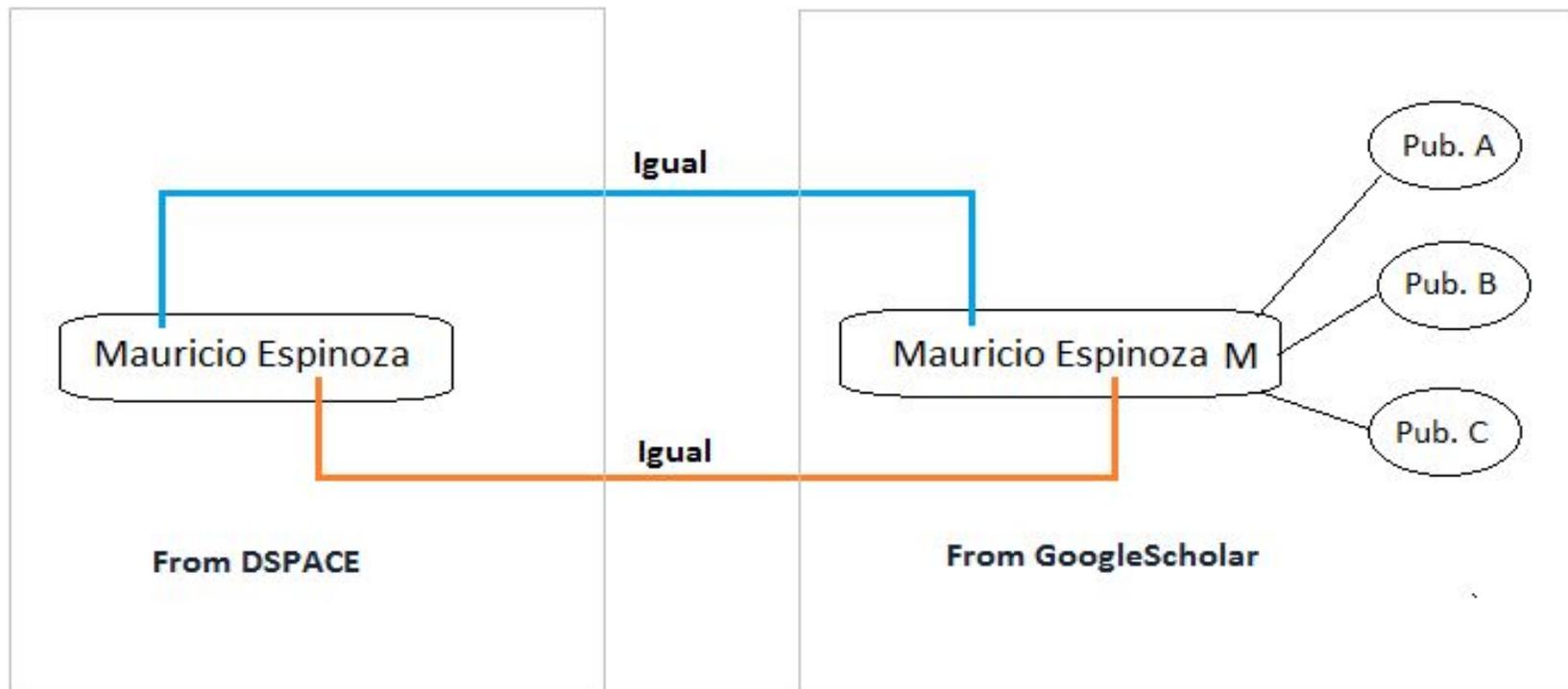
Población de la Ontología (ii)

- Se limpia, valida e integra la información y se consolida en un grafo central



Población de la Ontología (iii)

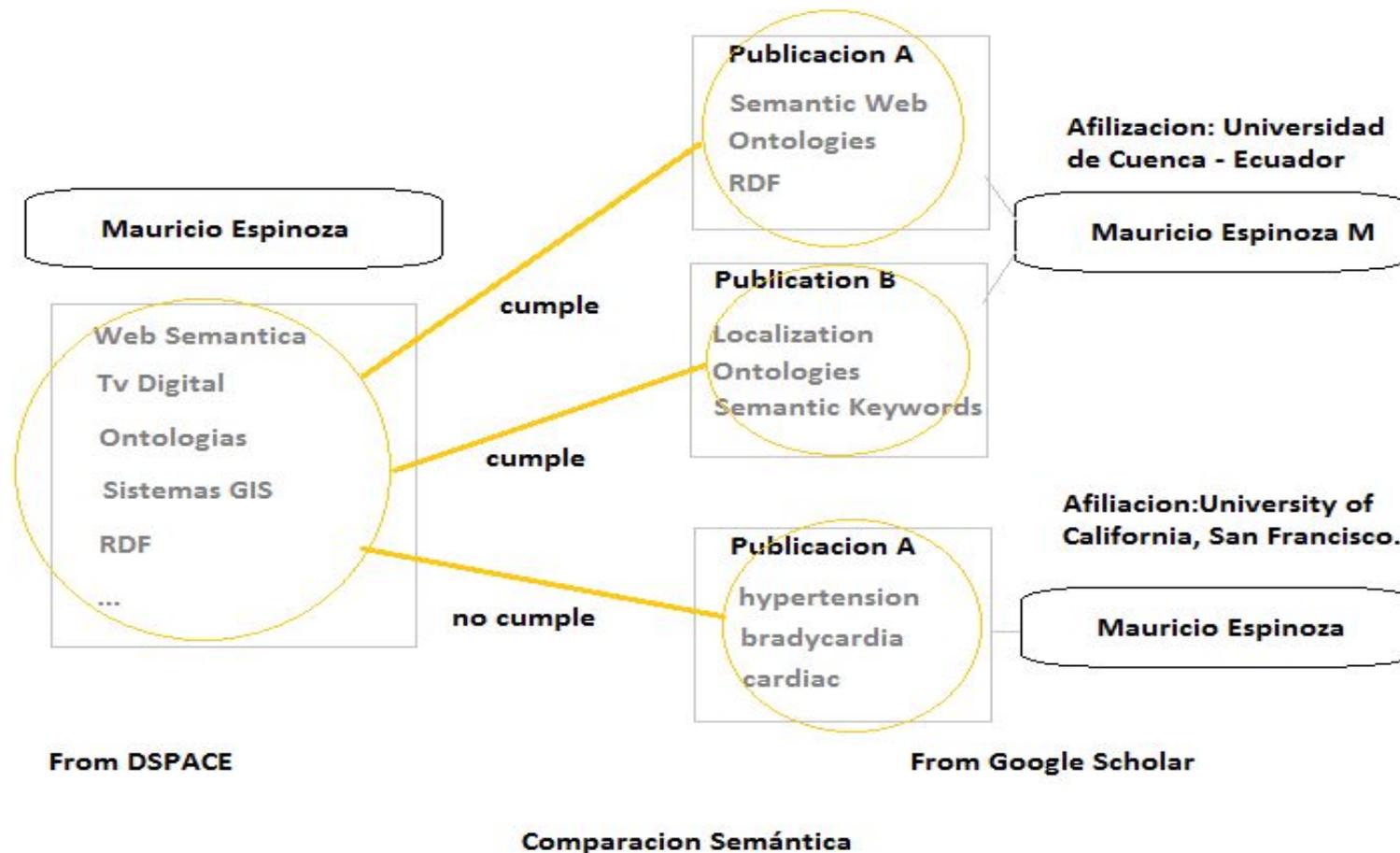
- ❖ Se ejecuta un proceso de desambiguación
 - Sintáctica y Semántica (Verificar si una publicación pertenece al autor)



Comparacion Sintactica de Nombre

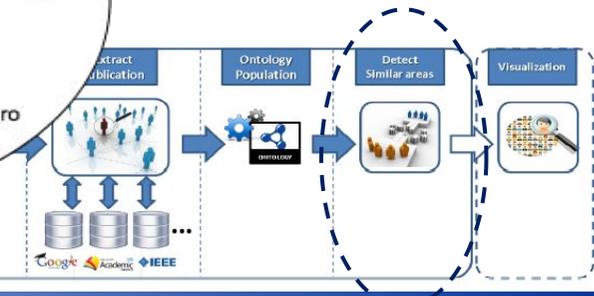
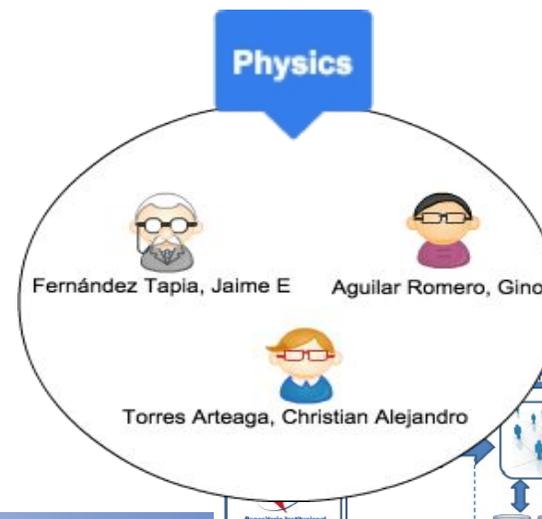
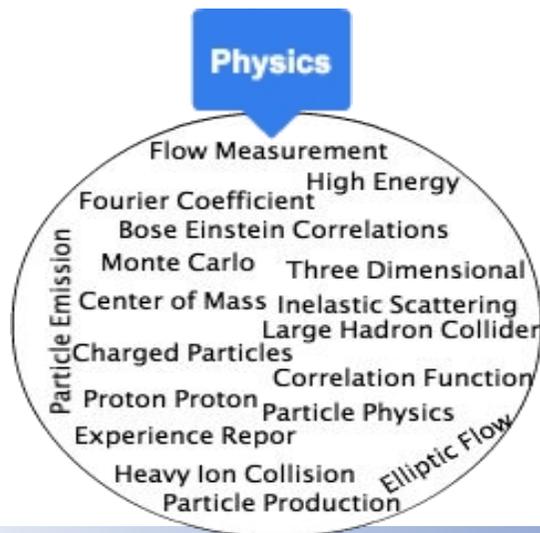
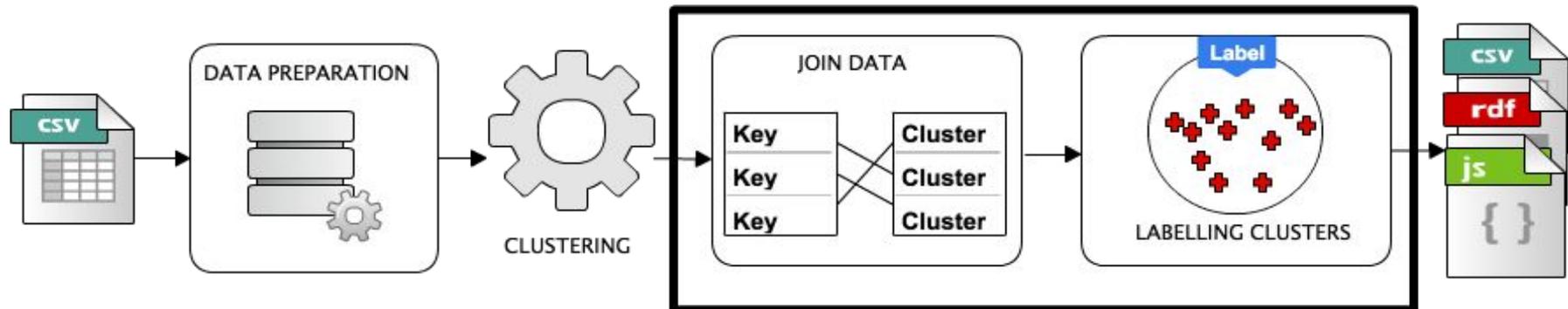
Población de la Ontología (iv)

- ❖ La desambiguación semántica compara los conceptos de las keywords



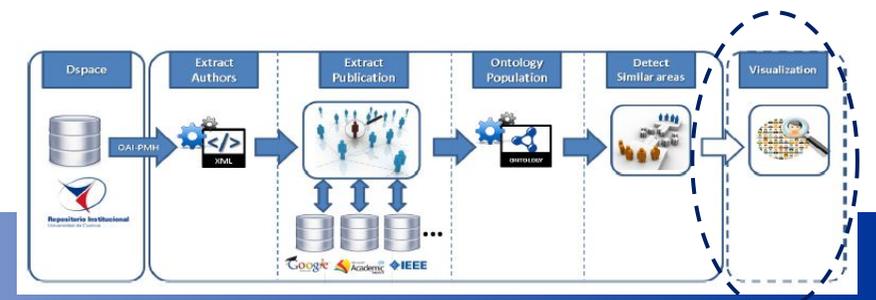
DetECCIÓN DE ÁREAS SIMILARES (i)

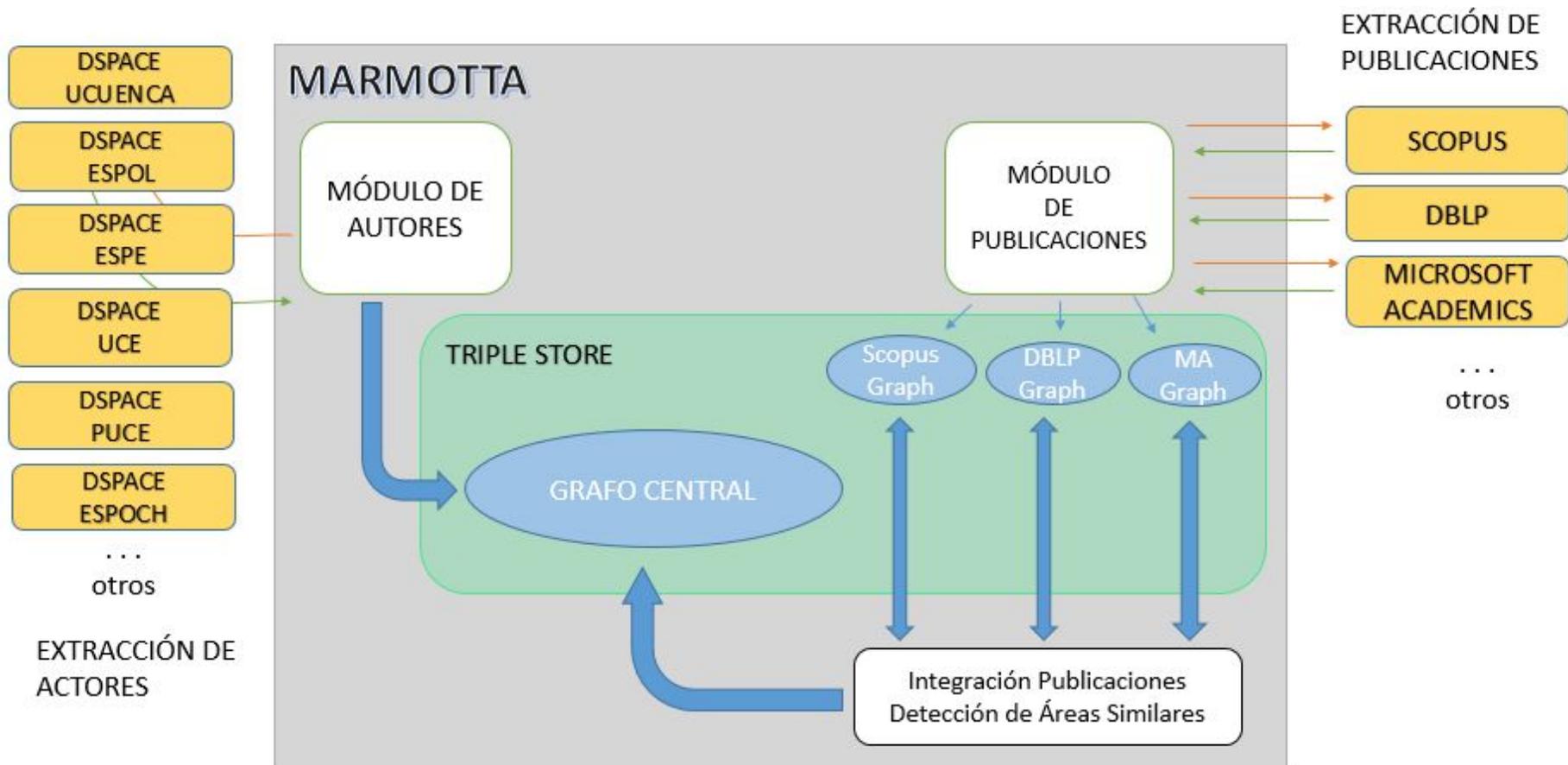
- ❖ Algoritmos de clustering: agrupar las entidades (investigadores y publicaciones) basado en keywords



Visualización de datos (i)

- ❑ Aplicación web que ofrece modelos de visualización
- ❑ Uso de la librería D3
 - <https://d3js.org/>
 - vistas o modelos de visualización explotando grandes volúmenes de datos





Resultados

Estado Actual Plataforma

<i>Recurso</i>	<i>Valor</i>
<i>Autores</i>	<i>912</i>
<i>Publicaciones</i>	<i>3488</i>
<i>Keywords</i>	<i>4475</i>
<i>Grupos detectados</i>	<i>78</i>
<i>IES identificadas</i>	<i>16</i>

Visualización de Grupos por Área de Conocimiento



Inicio

Nosotros

Ayuda

Contacto

Fuentes de Información



Buscar..

Agrupamiento de publicaciones

Agrupar por:

Área

superlattices

American Mathem

ers

Fuzzy Rules

Asymptotic Properties



Title: GeoLinked data and INSPIRE through an application case.

Keywords: "Coastal Area, Linked Data"

Author: "Saquicela, Victor"

Cluster: level

Flash Translation Laye

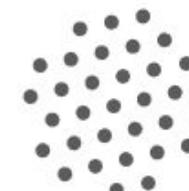
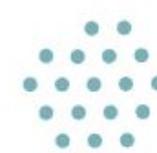
Lessons Learned

Commercial Off The She

3d structure

Computational Protein

Ensemble Monte Carlo



Visualización de investigadores

The image shows a screenshot of the REDI website search results for the name "Mauricio Espinoza". The browser address bar shows the URL: `redi.cedia.org.ec/#/es/w/search?Mauricio%20Espinoza`. The search results are annotated with several labels and arrows pointing to specific parts of the page:

- Enlace de la Pagina Web de REDI**: Points to the browser address bar.
- CO-AUTORES DE MAURICIO ESPINOZA**: Points to a vertical list of co-authors on the right side of the page.
- CUADRO DE BÚSQUEDA DE TEXTO**: Points to the search input field containing "Mauricio Espinoza".
- CO-AUTORES DE MAURICIO ESPINOZA**: Points to the "CO-Autores" section on the left, which lists Espinoza, Mauricio(10) and Saquicela, Victor(1).
- AUTORES QUE TIENEN ÁREAS DE CONOCIMIENTO SIMILARES A LAS DE MAURICIO ESPINOZA**: Points to the "Autores Relacionados" section on the left, which lists various other researchers and their publication counts.
- INFORMACIÓN Y ÁREAS DE CONOCIMIENTO DEL INVESTIGADOR**: Points to the "Información del Investigador" section on the right, which provides details about Mauricio Espinoza, including his affiliation (UNIVERSIDAD DE CUENCA), location (CUENCA, AZUAY), and research subjects.
- INFORMACIÓN DE LA PUBLICACIÓN**: Points to the "Información del Artículo" section on the right, which displays details about a specific publication, including the title, search engines (GOOGLE, DBLP, SCOPUS, SCHOLAR), and external links.
- PUBLICACIONES DE MAURICIO ESPINOZA**: Points to a central vertical list of publication icons, with lines connecting them to the co-authors and related authors.

Información del Investigador

Nombre: Espinoza, Mauricio
IES: UNIVERSIDAD DE CUENCA
Ciudad: CUENCA
Provincia: AZUAY
Subjects: TRES CAPAS, RED TELEFONICA, PROGRAMA DE COMPUTACION, SISTEMA DE INFORMACION GEOGRAFICA, TESIS EN INFORMATICA, ETAPA.

Información del Artículo

Intentar buscar en:
GOOGLE DBLP SCOPUS SCHOLAR

Título:
Enriching Electronic Program Guides using semantic technologies and external resources.

URL - Link Externo:
CLEI 2014.6965173

CO-Autores - Links Externos:
Saquicela Victor
Palacio Kenneth
Meja Mauricio_Espinoza
Alban Humberto
Meja Mauricio_Espinoza

Conclusiones

❑ Tecnológicas y científicas

- Aunar ventajas de iniciativas disjuntas
- Agrupación investigadores por áreas similares
- Marmotta: manejo de Linked Data y exportación a formatos recomendados por W3C

❑ Ámbito académico

- Dinamizar la creación de nuevas propuestas de investigación: Redes
- Mejorar la visibilidad: IES - investigador

Trabajos Futuros

- ❖ Mejorar visualización y tiempo de respuesta
- ❖ Implementación RDF Data Cube
- ❖ Clustering: Analizar más parámetros
 - título, abstract, etc.
 - Incrementar rendimiento y calidad de los patrones

Gracias!

mauricio.espinoza@ucuenca.edu.ec

victor.saquicela@ucuenca.edu.ec