

Repositorio Semántico de Investigadores del Ecuador

Fernando Baculima^a, Jose Luis Cullcay^a, Freddy Sumba^a, Xavier Sumba^a, Mauricio Espinoza^a, Victor Saquicela^a.

^a Departamento de Ciencias de la Computacion, Universidad de Cuenca, Av. 12 de Abril y Av. Loja, Cuenca, Ecuador
fernando.baculima@ucuenca.edu.ec, jose.cullcay@ucuenca.edu.ec,
xavier.sumba93@ucuenca.ec, freddy.sumbao@ucuenca.ec, mauricio.espinoza@ucuenca.edu.ec,
victor.saquicela@ucuenca.edu.ec

Resumen. En los últimos años, el número de investigadores en Ecuador que publican sus trabajos en diferentes líneas de investigación ha aumentado considerablemente. La identificación de áreas similares de investigación se ha convertido en una necesidad para generar redes de investigadores interesados en la creación de nuevos proyectos en torno a las Instituciones de Educación Superior (IES). En este artículo se presenta el Repositorio Ecuatoriano de Investigadores (REDI), el cual acoge datos de investigadores ecuatorianos y sus respectivas publicaciones con el objetivo de identificar y visualizar investigadores que trabajen en áreas de conocimiento similares, entre otros.

Palabras Clave: Repositorio Semántico, Investigadores, Redes de Colaboración, Cluster, Web Semántica.

1 Introducción

En particular, para las empresas privadas y organizaciones públicas, representa una preocupación importante el identificar los intereses comunes entre sus investigadores. Un ejemplo de esto, es el caso del gobierno ecuatoriano, el cual en los últimos años ha venido promoviendo un programa masivo de educación en el extranjero dirigido a mejorar el talento humano que ha permitido al país contar con un número creciente de investigadores, los cuales a su retorno tienen que buscar oportunidades de colaboración con equipos de investigación existentes. Debido a que el enfoque en la investigación académica es reciente en Ecuador, aún no existen herramientas que asistan a los investigadores en esa búsqueda.

Actualmente, cada IES mantiene un repositorio digital con los datos de sus investigadores y algunas de sus publicaciones. Sin embargo, cuando se presenta la necesidad de proponer proyectos entre varias IES surgen las interrogantes: ¿Trabajan otras IES en un área de conocimiento similar a la mía?, ¿Quiénes son los investigadores que trabajan en mi área de conocimiento? Ahora mismo la respuesta a estas preguntas se realiza de forma manual, es decir, las redes para los proyectos se generan en base a amistades o consultas a otros investigadores. Dicha respuesta generó la necesidad de crear un repositorio nacional con datos de investigadores en donde se pueda identificar investigadores de áreas comunes de conocimiento a través de herramientas informáticas.

Para solventar este problema se creó el Repositorio Ecuatoriano de Investigadores (REDI)¹, cuyo propósito es extraer datos desde los repositorios de publicaciones como Scopus², Google Scholar³, Microsoft Academics⁴, etc., integrar los datos provenientes desde las diferentes fuentes y brindar información sobre autores, publicaciones y áreas de conocimiento de las Instituciones de Educación Superior con la implementación de tecnologías semánticas⁵ y de aprendizaje automático⁶. El principal objetivo de REDI es la detección de investigadores y áreas similares de conocimiento que motiven a un proceso de investigación colaborativo. Además, se implementaron modelos de visualización que facilitan al usuario la navegación entre diferentes entidades y autores que trabajan en áreas de conocimiento de distinta índole.

En la sección 2 se presenta una serie de trabajos relacionados que muestran la situación actual respecto a repositorios de investigadores y sus modelos de explotación de información, en la sección 3 se describe un escenario que muestra el proceso actual para la búsqueda de información relacionada a un investigador ecuatoriano, la sección 4 describe la arquitectura del repositorio ecuatoriano de investigadores, en la sección 5 se muestra el resultado de este trabajo y, finalmente, la sección 6 describe las conclusiones y trabajos futuros.

2 Trabajos Relacionados

Existen varios trabajos previos en el ámbito de la explotación de información de autores y sus respectivas publicaciones. A continuación se describirán los esfuerzos presentados por algunos de estos trabajos con el fin de determinar el estado actual sobre los repositorios de investigadores y las herramientas que permiten visualizar este tipo de información.

Atanassova et al. [1] presentan un sistema que permite recuperar información de publicaciones científicas. La información recuperada es anotada semánticamente y el sistema permite filtrar resultados de acuerdo a facetas semánticas. Las anotaciones son realizadas mediante un *Rule-Based*⁷ que identifica pistas lingüísticas específicas en una ontología lingüística. En este trabajo los autores presentan esfuerzos para la anotación semántica y explotación de las publicaciones, no presentan un enfoque para autores ni la identificación de redes de colaboración entre autores.

Buter et al. [2] describen tres estrategias de búsqueda para identificar áreas de investigación convergentes utilizando datos extraídos desde las publicaciones y citaciones de WoS⁸, incluyendo las ciencias sociales y humanitarias. Este trabajo se limita a identificar áreas de investigación convergentes y recolectar estadísticas sobre las áreas encontradas, pero los autores no presentan propuestas para proporcionar una

¹ <http://redi.cedia.org.ec>

² <https://www.elsevier.com/solutions/scopus>

³ <https://scholar.google.com/>

⁴ <http://academic.research.microsoft.com/>

⁵ <https://www.w3.org/standards/semanticweb/>

⁶ <http://www.cs.bris.ac.uk/~flach/mlbook/materials/mlbook-beamer.pdf>

⁷ <http://www.aaai.org/Papers/Buchanan/Buchanan04.pdf>

⁸ <http://webofscience.com>

fácil navegabilidad a través de los resultados que han generado ni herramientas para asistir a los investigadores que buscan potenciales colaboradores en cada área.

Una propuesta para identificar grupos de investigación en un campo de investigación en particular es presentada por Calero et al. [3]. Con una combinación de técnicas de mapeo bibliométrico y análisis de redes, los autores han identificado y clasificado *clusters* de autores para representar los grupos de investigación. En este caso, utilizan una base de datos que contiene publicaciones, autores y sus afiliaciones con las IES. Sobre esta base de datos, buscan los términos sintácticamente relacionados con Nanotecnología, lo cual restringe su utilidad en contextos donde no se cuenta con toda esa información, además de requerir un mantenimiento manual de la base de datos de publicaciones. Adicionalmente, no se presenta una forma de visualización de los resultados para los usuarios.

Bergstrom et al. [4] describen una herramienta llamada *PaperCube*⁹. Esta herramienta permite visualizar redes de publicaciones y autores relacionados por sus citas. Los modelos de visualización para publicaciones permiten descubrir información acerca de un tema en particular, mientras que los modelos de visualización de autores permiten descubrir redes de autores basándose en sus colaboraciones y en las citas de sus publicaciones. *PaperCube* facilita la visualización de tales redes basándose únicamente en las citas y en los coautores de la publicación, pero no descubre posibles relaciones con otros trabajos que no hayan sido explícitamente citados. Necesariamente las relaciones están plasmadas sobre los documentos.

U-papers es un sistema desarrollado en la Universidad de Chile [5]. Se trata de un repositorio de artículos científicos cuyo objetivo, descrito por Chacón [6], es compartir publicaciones científicas desarrolladas por investigadores de la Universidad de Chile. El autor menciona que eventualmente el sistema ira evolucionando para acoger a investigadores de otras universidades. *U-papers* presenta 3 principales servicios: 1) Que un usuario cualquiera, inclusive de otro país, pueda acceder a *U-papers* y encontrar la publicación deseada. 2) Que el proceso que realizan los docentes actualmente, relativos a la declaración de nuevas publicaciones, sea un proceso automatizado. 3) Mantener la mayor cantidad de archivos asociados a las publicaciones. Este trabajo resulta ser bastante útil de cara a un usuario que desea buscar publicaciones relacionadas con cualquier temática en un entorno específico, en este caso artículos de la Universidad de Chile, sin embargo, no es posible identificar a investigadores que estén trabajando en un área de conocimiento en particular.

3 Escenario

Actualmente, la mayoría de los documentos que son frutos del trabajo de investigadores son indexados en repositorios de publicaciones (RDP) como: Scopus¹⁰, Microsoft Academics¹¹, Google Scholar¹², etc. Esto genera que gran cantidad de información se encuentre fragmentada, pues algunos RDP tienen información que

⁹ <http://papercube.peterbergstrom.com/>

¹⁰ <https://www.elsevier.com/solutions/scopus>

¹¹ <http://academic.research.microsoft.com/>

¹² <https://scholar.google.com/>

otros no tienen. Es por esto que los esfuerzos de los usuarios que tratan de buscar información sobre algún investigador o sobre alguna área de conocimiento en particular son altos, principalmente en términos de tiempo.

Según análisis previos realizados y la revisión de trabajos relacionados, no se ha detectado la existencia de un repositorio dedicado a la identificación de áreas similares de conocimiento. La mayoría de los trabajos presentan propuestas o soluciones enfocadas únicamente en las redes de colaboración creadas a través de las referencias y citas de un determinado artículo, es decir, estas relaciones necesariamente deben estar plasmadas en los documentos.

Por otra parte, en Ecuador, actualmente no existe un repositorio dedicado a la integración de datos de investigadores y sus publicaciones; y mucho menos una plataforma que permita la visualización de esta información, así como la posibilidad de visualizar las áreas de conocimiento similares. Solamente existen repositorios individuales en las IES que contienen información de los autores de tesis de pregrado, posgrado y algunas pocas publicaciones, por lo que estos repositorios se enfocan en trabajos de estudiantes y no en investigadores.

A continuación, se describe un escenario común para la búsqueda de información de un investigador ecuatoriano. A un usuario que conoce el nombre y las áreas de conocimiento del investigador ecuatoriano llamado “Mauricio Espinoza”, le interesa obtener información de las publicaciones de este investigador, para lo cual, desde su ordenador accede al sitio Google Scholar, digita el nombre del investigador sobre los campos de búsqueda, resultado de esto se obtiene una gran cantidad de información en los cuales se listan las publicaciones de todos los investigadores llamados “Mauricio Espinoza” sin importar el área de conocimiento ni el país al que pertenece el investigador. El exceso de información confunde al usuario sobre las publicaciones que pertenecen al investigador de su interés, y este se ve en la necesidad de filtrar manualmente los resultados, tomándole una gran cantidad de tiempo. Posteriormente, el usuario logra seleccionar las publicaciones del investigador de su interés. Ahora, surge la necesidad de identificar a otros investigadores ecuatorianos que trabajen en áreas de conocimiento similares a las del investigador “Mauricio Espinoza”, pero, Google Scholar no provee visualizaciones que permitan descubrir e identificar de una manera fácil, rápida e intuitiva a los investigadores que trabajan en el área de conocimiento de interés, lo que obliga al usuario a utilizar otros RDP, sin tener la certeza de que obtendrá los datos y visualizaciones que permitan identificar investigadores que trabajen en áreas de conocimiento similares.

Con el objetivo de mejorar la experiencia del usuario y dadas las condiciones en este contexto, en este trabajo se presenta al REDI, que integra en un solo repositorio la información referente a áreas de conocimiento y publicaciones de los investigadores ecuatorianos. Esta plataforma ofrece varias visualizaciones amigables e intuitivas para el usuario. Además, REDI propone la integración de publicaciones que se encuentran dispersas en los diferentes RDP. Estas soluciones permitirán ahorrar tiempo y esfuerzo cuando se desee descubrir e identificar a investigadores que trabajan en áreas de conocimiento similares.

4 Arquitectura

En esta sección se describe la arquitectura general del REDI, el cual permite extraer información de investigadores ecuatorianos, sus publicaciones, detectar áreas similares de conocimiento y generar modelos de visualización para el usuario final. En la figura 1 se muestra los cinco componentes o procesos que conforman la arquitectura de REDI y que se detallan a continuación.

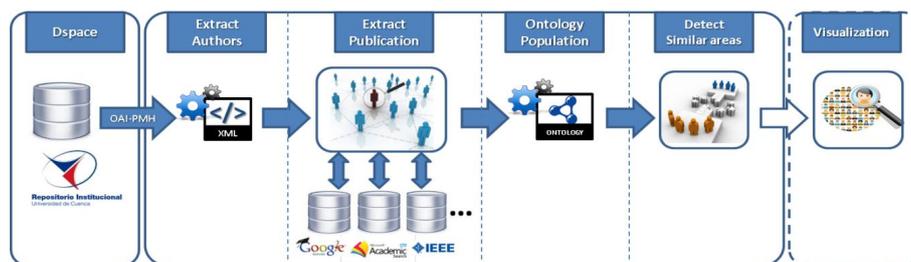


Fig. 1. Arquitectura del REDI.

4.1 Extracción de autores. Este proceso consiste en extraer datos de autores de trabajos académicos e investigativos que se encuentran en los repositorios digitales DSPACE¹³ de las IES. Esto se realiza mediante un servicio que extrae, transforma a formato RDF¹⁴ y almacena la información utilizando el modelo ontológico¹⁵ de datos de personas (FOAF¹⁶).

4.2 Extracción de publicaciones. Este proceso consiste en obtener información desde los RDP acerca de las publicaciones de los autores recuperados en el proceso anterior. Al igual que el componente anterior, se han implementado servicios que se conectan a las APIs y permiten el acceso a los datos de cada uno de los RDP. Si se ha encontrado posibles trabajos de un autor de DSPACE en los RDP, entonces se valida dicha relación para determinar que se trata de trabajos de ese autor. Luego se procede a almacenar los datos de las publicaciones relacionadas a los autores utilizando el modelo ontológico bibliográfico (BIBO¹⁷).

4.3 Población de la Ontología. La información extraída en los componentes anteriores se guarda en grafos insertados en un Triple Store¹⁸. En principio, los datos de autores se almacenan en un grafo de autores, mientras que los datos de publicaciones se almacenan en diferentes grafos, según cada RDP del cual provienen las publicaciones. En este punto, existe un proceso de integración de la información, ilustrado en la figura 2, que valida ciertos aspectos de la coherencia de la información

¹³ <http://www.dspace.org/>

¹⁴ <https://www.w3.org/RDF/>

¹⁵ <https://www.w3.org/standards/semanticweb/ontology.html>

¹⁶ <http://xmlns.com/foaf/spec/>

¹⁷ <http://bibliontology.com/>

¹⁸ https://www.w3.org/2001/sw/wiki/Category:Triple_Store

y la consolida en un grafo central. La población del grafo central está basado en el modelo ontológico mostrado en la figura 3.

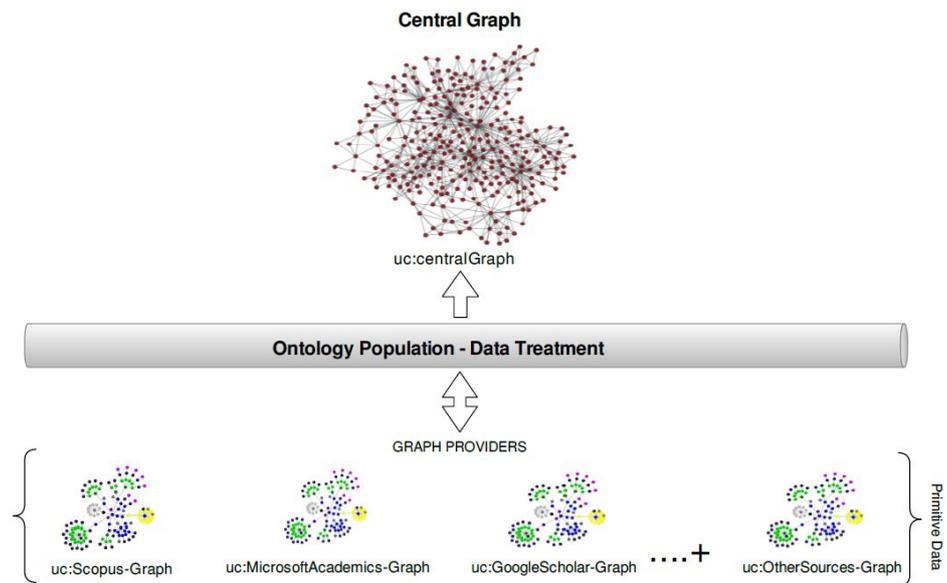


Fig. 2. Población de la ontología. Integración, validación y limpieza de información que permite generar el grafo central a partir del cual se explotará la información.

4.4 Detección de áreas similares de conocimiento. Realiza una detección de patrones que permite agrupar las entidades (investigadores y publicaciones) según los datos almacenados en el grafo central. En este proceso se ejecutan técnicas y algoritmos de minerías de datos, descritas en [9] y [10], que permiten detectar asociaciones entre los datos y agrupar entidades similares. En este punto, los grupos también reciben un nombre de acuerdo a parámetros que relacionan a los elementos de cada grupo como se describe en [11] y, finalmente, esta información es anotada semánticamente en el modelo ontológico.

REDI ONTOLOGY MODEL

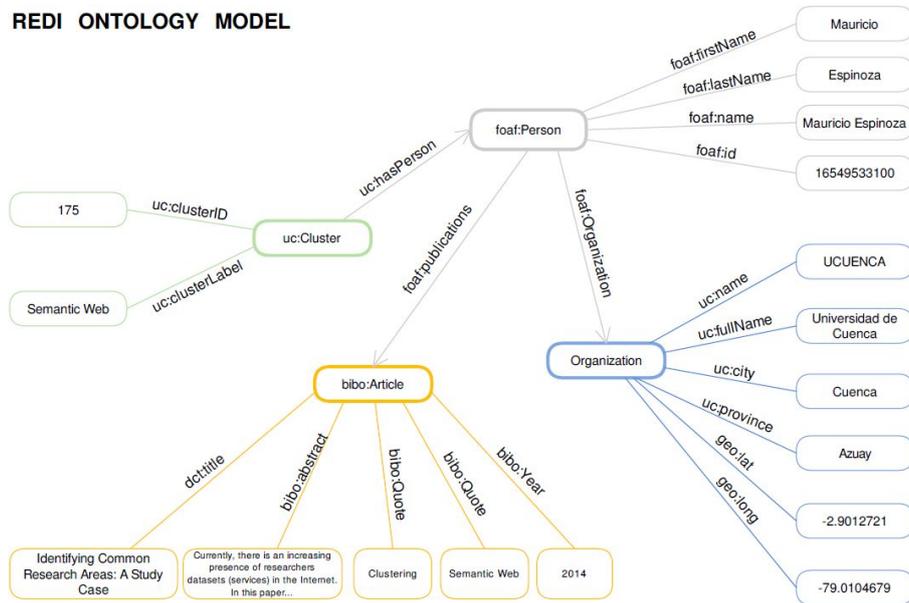


Fig. 3. Modelo Ontológico de REDI basado en FOAF y BIBO. Modelo ontológico representado con datos de una publicación del investigador “Mauricio Espinoza”.

4.5 Visualización. Finalmente se ha implementado una aplicación web que ofrece modelos de visualización que permiten descubrir la información almacenada sobre el repositorio. Estos modelos de visualización que permiten navegar y descubrir investigadores, publicaciones, autores relacionados y áreas de conocimiento de forma fácil e intuitiva, se construyeron utilizando la librería D3¹⁹. Esta librería permite crear vistas o modelos de visualización explotando grandes volúmenes de datos; el objetivo es enlazar las diferentes vistas creadas en una experiencia agradable e interactiva para el usuario final. De esta manera, la aplicación web permite navegar a través de una serie de rutas que posibilitan el descubrimiento de nueva información a través de las visualizaciones disponibles

REDI está construida sobre un LDP²⁰ llamado Apache Marmotta²¹, la cual fue extendida con los componentes descritos anteriormente en la arquitectura del sistema. De esta manera, se dispone de una plataforma de Linked Data que realice la extracción, integración de la información, además de la detección de los patrones y el manejo de consultas.

¹⁹ <https://d3js.org/>

²⁰ Linked Data Platform: <https://www.w3.org/TR/ldp/>

²¹ <http://marmotta.apache.org/>

5. Resultados

La implementación y puesta en marcha de REDI²² ha generado ciertos resultados que se describen a continuación. REDI presenta la información obtenida a través de su interfaz de usuario accesible desde un navegador web, e incluye diferentes modelos de visualización. En las siguientes subsecciones se presentan dos ejemplos de las visualizaciones disponibles para el usuario final que ilustran el potencial de REDI.

5.1 Visualización de Grupos por Área de Conocimiento

Con el objetivo de facilitar la búsqueda de posibles colaboradores para investigadores que trabajan en la misma área de conocimiento, se ha implementado esta vista que permite visualizar los grupos (*clusters*) de publicaciones y autores identificados por la plataforma. Considerando que un clúster es un conjunto de datos agrupados en base a ciertas características, mediante esta búsqueda se relacionan las publicaciones con los clústeres del área a la que pertenecen, es decir, permite visualizar los clústeres y la información acerca de las publicaciones. Esta visualización refleja los resultados obtenidos mediante el componente de detección de áreas similares que fue el encargado de agrupar los autores y publicaciones (Figura 4). Esta visualización es útil cuando un investigador quiere explorar los grupos más relevantes que han sido encontrados y que podrían ser de interés, pero no tiene un autor o publicación específica en mente.



Fig. 4. Visualización de grupos de investigadores clasificados por área de conocimiento. En esta vista se visualizan los grupos de investigadores que han sido detectados a partir de la ejecución de los algoritmos de *clustering*.

²² <http://redi.cedia.org.ec>

5.2 Visualización de investigadores

El propósito de esta visualización es facilitar la exploración inicial de la información recolectada por la plataforma sobre autores y sus publicaciones. Esta vista es especialmente útil cuando el usuario conoce a un autor y quiere explorar a sus coautores y a otros autores relacionados por las temáticas de sus publicaciones, los cuales pueden ser potenciales colaboradores. De este modo, el usuario obtiene información de investigadores, sus publicaciones y la relación que tienen con otros investigadores. El resultado es la visualización de un árbol que permite explorar investigadores y publicaciones, con nodos que representan dichas entidades (Figura 5).

La información que se presenta en el árbol proviene de tres componentes del sistema: extracción de autores, publicaciones, e integración de la información. Los nodos representan a investigadores y las ramas representan sus publicaciones y a los coautores. El componente de detección de áreas similares (descrito en la sección arquitectura) permite presentar un listado de enlaces a investigadores con intereses relacionados, que fueron vinculados mediante el proceso de agrupamiento. Estos enlaces facilitan una búsqueda rápida de posibles colaboradores.

El resto de visualizaciones también presentan de forma interactiva la información obtenida por los componentes del sistema. Otras funciones que se encuentran disponibles son: búsqueda por palabras clave, grupos por ubicación geográfica, listados desde una nube de palabras clave, y estadísticas. Cada una de ellas puede asistir a un investigador en la búsqueda de diferente información que pueda ser de utilidad. Con el objetivo de demostrar la utilidad de las visualizaciones presentadas, en la siguiente subsección se describe un escenario de búsqueda utilizando la aplicación REDI.

5.3 Escenario de búsqueda utilizando REDI

En la sección 3 se describió el escenario para la búsqueda del investigador ecuatoriano “Mauricio Espinoza” sobre los RDP. En esta sección se describe la búsqueda de este investigador ejecutada sobre REDI y a continuación se muestran los resultados obtenidos.

El usuario final debe acceder a la página web de REDI²³ y, sobre el cuadro de búsqueda, digitar el nombre del investigador ecuatoriano de su interés, en este caso “Mauricio Espinoza”. A continuación REDI muestra un árbol con las publicaciones de este investigador, los temas o áreas de conocimiento y los investigadores que trabajan en áreas similares al investigador “Mauricio Espinoza”. Estos resultados se muestran en la figura 5.

²³ <http://redi.cedia.org.ec>

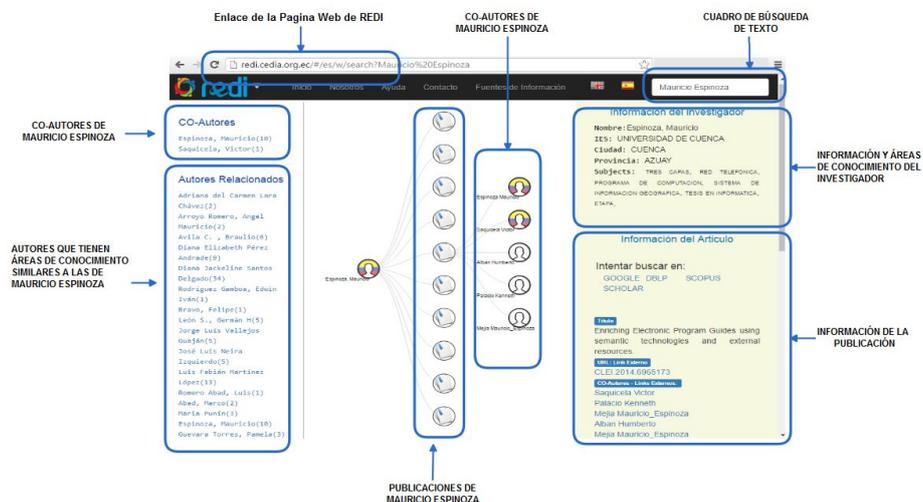


Fig. 5. Visualización de publicaciones e investigadores relacionados con el investigador ecuatoriano “Mauricio Espinoza”. REDI permite realizar la consulta y visualización de investigadores ecuatorianos con modelos de vistas de fácil navegabilidad y descubrimiento de información [8].

En caso de que exista más de un investigador con el mismo nombre, REDI muestra a todos los investigadores, junto a sus áreas de conocimiento, que tengan el nombre digitado por el usuario, para que este seleccione el investigador de su interés. REDI es una plataforma que está en constante actualización de datos. En la siguiente tabla se muestra la cantidad de información que hasta el momento se han cargado en la plataforma.

Tabla 1. Estado actual de los datos en REDI. Presenta los totales de las entidades que se encuentran almacenadas en la versión actual de la plataforma

Recurso	Valor
Autores	912
Publicaciones	3488
Keywords	4475
Grupos detectados	78
IES identificadas en el repositorio	16

Esta información puede ser consultada directamente desde la Plataforma Servidor REDI²⁴, o a través de la página Web en la sección de estadísticas²⁵.

²⁴ <http://190.15.141.85:8080/core/admin/dataview.html>

²⁵ <http://redi.cedia.org.ec/#/es/data/statistics>

6. Conclusiones y Trabajo Futuro

En este trabajo se presenta una aplicación que apoya a los investigadores en la búsqueda de colegas que se encuentran trabajando en un área de conocimiento similar y que pueden convertirse en potenciales colaboradores en futuros proyectos. Como resultado, la aplicación REDI se encarga de la detección de investigadores y áreas similares de conocimiento, los cuales son presentados mediante modelos de visualización que posibilitan la navegación entre publicaciones y autores que trabajan en distintas áreas.

Los resultados muestran que el proceso realizado en REDI comprende una mejora con respecto al proceso de búsqueda manual que se realiza en los RDP o incluso a los resultados de otros proyectos que tienen un menor alcance. Un aporte significativo ha sido la agrupación investigadores por áreas similares, la cual utiliza técnicas semánticas y de machine learning para crear conjuntos relacionados a pesar de que no exista una equivalencia sintáctica entre los metadatos de los autores y publicaciones. Otra contribución importante es la creación de vistas que muestren los resultados al usuario de forma interactiva. También es importante recalcar que Marmotta, la plataforma utilizada por REDI, se especializa en el manejo de Linked Data, lo que facilita la exportación de información en formatos recomendados por las especificaciones de World Wide Web Consortium (W3C²⁶) para modelos de metadatos.

Adicionalmente, Marmotta permite que los datos se actualicen mediante procesos que liberan al usuario de la carga de actualizar manualmente la información y a su vez ayuda a que exista un mejor resultado de detección de áreas similares y descubrimiento de información.

Como trabajos futuros, se plantea continuar mejorando la plataforma en diferentes aspectos. Uno de ellos es el mejoramiento de la búsqueda y asociación de publicaciones pertenecientes a los investigadores. Existen posibilidades de mejora para el algoritmo que se encarga de dicha asociación, especialmente para los casos en los que se cuentan con nombres abreviados o keywords limitadas en cantidad y significado.

Otro aspecto a mejorar involucra las técnicas de descubrimiento de nueva información a partir de los datos de los que dispone la plataforma actualmente. Debido a que la plataforma utiliza tecnologías de la Web Semántica, se podría emplear un razonador basado en ontologías para descubrir relaciones de interés entre las entidades consideradas.

Adicionalmente se plantea mejorar el agrupamiento (*clustering*) de investigadores y publicaciones mediante el análisis de más parámetros, como por ejemplo el título, resumen y otras secciones del documento, y así mejorar la calidad de los patrones encontrados. Además, es necesario continuar la mejora de las vistas disponibles para la visualización en la aplicación web, así como el tiempo de respuesta de algunas opciones.

Finalmente, se proyecta integrar dentro de la plataforma REDI a otras redes de investigadores latinoamericanos, lo cual serviría de apoyo a investigadores

²⁶ <http://www.w3c.es/>

interesados en promover la colaboración entre grupos de investigación de diferentes países.

Agradecimientos

Este trabajo forma parte del proyecto “Repositorio Semántico de Investigadores del Ecuador” financiado por CEDIA²⁷ y el Departamento de Ciencias de la Computación de la Universidad de Cuenca.

Referencias

1. Atanassova, Iana, Marc Bertin: The Semantic Publishing Challenge, Task 3 : In-use task, 11th European Semantic Web Conference (ESWC-2014).
2. Buter, R. K., Noyons, E. C. M., & Van Raan, A. F. J. (2010). Identification of converging research areas using publication and citation data. *Research Evaluation*, 19(1), 19-27 (2010).
3. Calero, Clara, Buter, Renald, Cabello Vald , Cecilia: How to identify research groups using publication analysis: an example in the field of nanotechnology. *Scientometrics* (2013).
4. Bergstrom, Peter , Darren C Atkinson: Augmenting the Exploration of Digital Libraries with Web-Based Visualizations. *IEEE* (2009).
5. A publications repository from the Computer Science Department of the University of Chile, <https://upapers.dcc.uchile.cl/index/>
6. Chacón, Candia, Ignacio, Renald: Desarrollo de un repositorio de artículos científicos. Tesis. (2012).
7. Saquicela, V., Bermeo, J., Espinoza, M., Palacio, K., Villazon-Terrazas B.: Identifying Common Research Areas: A Study Case. *ENC* (2014).
8. Repositorio Ecuatoriano de Investigadores, <http://redi.cedia.org.ec/#/es/>
9. Rokach, Lios, Maimon, Oded: *Data Mining and Knowledge Discovery Handbook. Clustering Methods*. Boston, MA (2005).
10. N. O. Andrews and E. A. Fox: *Recent developments in document clustering*. Virginia Tech. Department of Computer Science, 2007.
11. Lau, Jey Han and Grieser, Karl and Newman, David and Baldwin, Timothy: *Automatic Labelling of Topics Models*. Portland, Oregon (2011).

²⁷ <https://cedia.org.ec/>