



Generación Dinámica de Redes de Colaboración en México y América Latina (GReCo)

Reporte final

Proyecto financiado por la Corporación Universitaria para el Desarrollo de Internet CUDI

Investigadores responsables:

Dr. J. Alfredo Sánchez Huitrón
Universidad de las Américas- Puebla (UDLAP) , COORDINADOR

Dr. Juan Carlos Lavariega Jarquín
Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM)

Ing. Arianna Becerril García
Universidad Autónoma del Estado de México (UAEM)

Becarios:

UDLAP:

Federico Hernández Bolaños (Coordinador Técnico)

Efrén Hernández Cosco

Alfredo Ramos Lezama

ITESM:

Eric Santiago Balam Sabido

Alejandro Delgado Cárdenas

José Luis Madrid Marquez

UAEM:

Josué Isaac Legorreta Pérez,

UDLAP • ITESM • UAEM



GReCo: Generación Dinámica de Redes de Colaboración

Resumen

Este documento presenta una síntesis de las actividades realizadas y resultados obtenidos en el proyecto “Generación Dinámica de Redes de Colaboración en México y América Latina”, al que nos referimos de forma abreviada como el Proyecto GReCo. El proyecto fue auspiciado por la Corporación Universitaria para el Desarrollo de Internet (CUDI) con el objetivo central de aprovechar colecciones de documentos existentes para facilitar la detección y visualización de redes de colaboración entre investigadores e instituciones del país y la región.

El proyecto se basó en tres grandes áreas de acción, lideradas respectivamente por cada una de las instituciones participantes: recolección y normalización de datos (UAEM), diseño e implementación de mecanismos de inferencia (ITESM), y diseño e implementación de interfaces de visualización y consulta (UDLAP).

Se describen las actividades y resultados en cada área: metodología para recolección y normalización de metadatos para las colecciones disponibles de tesis y revistas electrónicas; adaptación de algoritmos para inferencia y representación de redes de colaboración a diferentes niveles, y consulta y visualización de redes integrando conceptos como nodos y enlaces, campos de estrellas, lentes de ojo de pescado y ontologías interactivas.

El reporte se ha organizado en los siguientes apartados: presentación del proyecto GReCo, resultados alcanzados, actividades, documentación generada, referencias y un directorio de participantes.

Adicionalmente a este reporte general del proyecto, se encuentran disponibles los reportes técnicos presentados por cada institución participante (UDLAP, ITESM y UAEM).

GReCo: Generación Dinámica de Redes de Colaboración

1. Presentación

El presente proyecto se enfoca a generar redes que reflejan la afinidad temática y autoral entre documentos y permitir su visualización y búsqueda para detectar dinámicamente y oportunidades de colaboración entre autores e instituciones.

Tomando como base las colecciones disponibles en la Red Abierta de Bibliotecas Digitales (RABiD, <http://www.rabid.org.mx>), entre las que se cuentan revistas especializadas y de divulgación, documentos digitalizados, así como tesis de licenciatura y posgrado, GReCo aprovecha mecanismos existentes basados en OAI-PMH para las colecciones participantes para recolección de sus metadatos.

Las estrategias centrales para lograr los objetivos del proyecto se han basado en la construcción de índices y aplicación de técnicas de inferencia para medir la proximidad semántica entre documentos y sus metadatos (previamente recolectados y normalizados), y generar una representación interna de las correspondientes redes entre autores y temas. Utilizando esta representación, se ofrecen varios servicios de acceso para explorar las redes, incluyendo dos clases de interfaces de visualización de redes y grupos resultantes, así como respuestas a preguntas específicas acerca de oportunidades potenciales de colaboración.

El proyecto inicia con las colecciones que las instituciones participantes ofrecen a través de RABiD, pero la metodología puede aplicarse a otros repositorios a nivel CUDI y a nivel latinoamericano que permitan detectar textual y visualmente oportunidades de colaboración y apoyar la generación de propuestas y proyectos conjuntos.

2. Resultados obtenidos

2.1 Arquitectura

GReCo se ha basado en la arquitectura propuesta inicialmente, la cual se ilustra en la Figura 1. Un aspecto clave de la arquitectura es la disponibilidad de servidores de metadatos asociados a las colecciones participantes. En el caso de las colecciones de UAEM, UDLAP e ITESM, los servidores de metadatos cumplen con el Protocolo de Cosecha de Metadatos de la Iniciativa de Archivos Abiertos (OAI-PMH por sus siglas en inglés). Cabe señalar que en la construcción de estos servidores se ha aprovechado el software vOAI desarrollado por UDLAP y disponible públicamente a través de la Red Abierta de Bibliotecas Digitales (RABiD, <http://www.rabid.org.mx>).

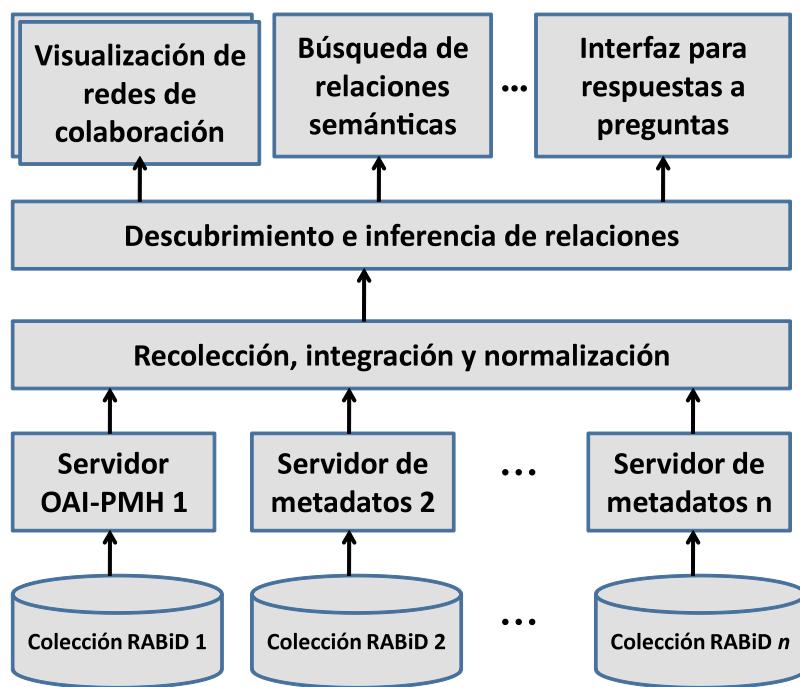


Figura 1. Arquitectura general de GReCo.

2.2 Recolección, integración y normalización

Para la recolección e integración de los metadatos obtenidos de las diversas colecciones mediante sus servidores OAI-PMH, se construyó una herramienta que produjo un formato de datos homogéneo para su posterior procesamiento por los mecanismos de inferencia. Dada la diversidad de formatos manejados por las fuentes (por ejemplo, sólo la

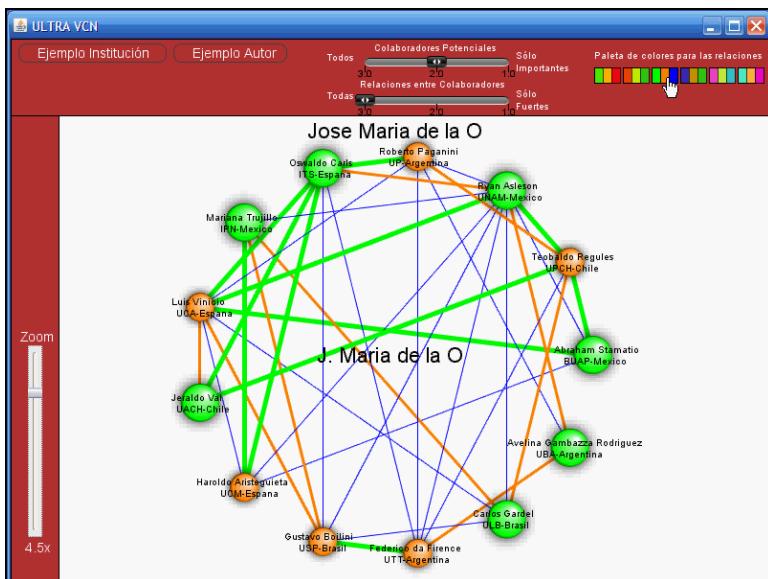
colección Redalyc comprende más de 600 revistas distintas con casi 140,000 artículos), fue necesario un gran esfuerzo para unificar su uso de los atributos del estándar Dublin Core, y así producir una base de metadatos que pudiera utilizarse por los otros componentes de la arquitectura. En este proceso se evaluaron también alternativas para un manejador de bases de datos común para el proyecto, optándose finalmente por eXist, una base de datos XML nativa disponible en código abierto.

2.3 Inferencia de redes de colaboración

Con respecto a los mecanismos de inferencia, se analizaron diversas aproximaciones al problema de determinar afinidades entre documentos, particularmente usando sus metadatos para producir agrupaciones que relacionan documentos, autores, instituciones y países. Se produjo una representación usando XML para cada uno de los niveles de interés (países, instituciones y autores) con el fin de facilitar la consulta y visualización. Una contribución clave del proyecto es la aplicación del algoritmo de agrupación conocido como FIHC para la construcción de ontologías ligeras y la inferencia de redes de colaboración.

2.4 Interfaces de consulta y visualización

En cuanto a las interfaces de acceso, se propusieron y se implementaron dos estrategias de visualización y un mecanismo de respuestas a preguntas específicas sobre colaboraciones potenciales. La primera estrategia de visualización está basada en la representación clásica de nodos y enlaces, pero se ha enriquecido mediante el uso de características gráficas (como color y volumen) que facilitan la identificación de los participantes de una red, así como con la introducción de filtros que permiten acotar rápidamente las redes de interés. La segunda estrategia integra de manera innovadora tres conceptos para ayudar en la detección visual de redes de colaboración: la representación de campos de estrellas para grandes volúmenes de información, el uso de múltiples lentes tipo “ojo de pescado” para examinar detalles de áreas de interés sin perder el contexto, y una representación gráfica de las ontologías en las que se organizan las entidades que participan en las redes de colaboración. Las interfaces de visualización resultantes se ilustran en la Figura 2.



(a)



(b)

Figura 2. Visualización (a) usando nodos y enlaces, y (b) usando campos de estrellas, ontologías y múltiples ojos de pescado.

3. Actividades realizadas

3.1 Videoconferencias, reuniones y visitas

En el transcurso del proyecto se llevaron a cabo videoconferencias mensuales para coordinar avances. Adicionalmente, se tuvieron reuniones presenciales en el contexto de las reuniones semestrales de CUDI en Puebla, Ixtapan de la Sal y Morelia. Asimismo, se llevaron a cabo visitas de los integrantes del proyecto a la Universidad de las Américas Puebla y

a colegas de la Universidad Politécnica de Puebla para tratar temas relacionados con los algoritmos de agrupación, el modelado de datos y evaluaciones preliminares de avances.

3.2 Presentaciones

Para compartir avances y recibir retroalimentación, se hicieron presentaciones formales del proyecto en sus diferentes etapas en la mesa de bibliotecas digitales de cada una de las reuniones semestrales de CUDI en Puebla, Ixtapan de la Sal y Morelia.

3.3 Cursos

En el contexto de la Reunión de Primavera 2009 de CUDI se ofreció el curso “Herramientas para gestión de colecciones de publicaciones electrónicas de acceso abierto”, orientado a promover la construcción de colecciones por instituciones y autores, las cuales alimentarán a proyectos como GReCo.

3.4 Otras actividades

La coordinación técnica instaló un portal del proyecto para difundir los resultados obtenidos y recibir retroalimentación. El portal está disponible en <http://ict.udlap.mx/greco>. También se configuró un wiki, disponible en el mismo portal o directamente en <http://ict.udlap.mx/wikis/greco>, para facilitar la documentación de las actividades y avances del proyecto.

4. Documentos Elaborados

En el transcurso de este proyecto se han elaborado los documentos listados a continuación:

Reporte Técnico Semestral GReCo UDLAP. Junio, 2009.
Reporte Técnico Semestral GReCo UAEM. Junio, 2009.
Reporte Técnico Semestral GReCo ITESM. Junio, 2009.
Reporte Técnico Final GReCo UAEM. Febrero 2010.
Reporte Técnico Final GReCo ITESM. Febrero 2010.
Reporte Técnico Final GReCo UDLAP. Febrero 2010.

Ramos Lezama, A., Sánchez, J. A. 2009. Visualization of Implicit Collaboration Networks using Multiple Fisheye Views on Ontological Starfields. Poster. Congreso Latinoamericano de Interacción Humano-Computadora (CLIHC 2009), Mérida, Yuc., Noviembre.

5. Referencias

GReCo

Portal del proyecto GReCo: <http://ict.udlap.mx/greco>

Wiki del proyecto GReCo: <http://ict.udlap.mx/wikis/greco>

ITESM: <http://www.itesm.mx>

UAEM: <http://www.uaemex.mx/>

Redalyc: <http://www.redalyc.org>

UDLAP: <http://www.udlap.mx>

Laboratorio de Tecnologías Interactivas y Cooperativas ICT

<http://ict.udlap.mx>

CUDI

Corporación Universitaria para el Desarrollo de Internet A.C.

<http://www.cudi.edu.mx>

Reunión de Primavera 2009 CUDI:

http://www.cudi.edu.mx/primavera_2009

Reunión de Otoño 2009 CUDI:

http://www.cudi.edu.mx/otono_2009

Reunión de Primavera 2010 CUDI:

http://www.cudi.edu.mx/primavera_2010

6. Directorio de Participantes

Universidad de las Américas Puebla (UDLAP)

Alfredo Sánchez Huitrón, investigador responsable, coordinador

Federico Hernández Bolaños (coordinador técnico)

Efrén Hernández Cosco, becario

Alfredo Ramos Lezama, becario

Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM)

Dr. Juan Carlos Lavariega Jarquín, investigador responsable

Eric Santiago Balam Sabido, becario

Alejandro Delgado Cárdenas, becario

José Luis Madrid Marquez, becario

Universidad Autónoma del Estado de México (UAEM)

Ing. Arianna Becerril García, investigador responsable

Josué Isaac Legorreta Pérez, becario