

Diseño de una Plataforma de Colaboración para investigación Científica basada en e-Science 3.0

Design of a collaboration platform for scientific research based on e-Science 3.0

Fernando Barraza¹, Hugo Ordoñez¹, Emilia Segovia¹

¹ Grupo de Investigación LIDIS. Universidad de San Buenaventura, Cali, Colombia

fbarraza@usbcali.edu.co , haordonez@usbcali.edu.co , ersegovia@usbcali.edu.co

RESUMEN. El concepto de e-Science 3.0 se ha venido acuñando en los últimos años como el uso de nuevas herramientas web inteligentes por parte de los investigadores para compartir, colaborar y publicar más fácilmente sus resultados. El grupo de investigación LIDIS ha diseñado una plataforma web con el propósito de apoyar a los investigadores en la formulación de proyectos de investigación utilizando conceptos de e-Science 3.0, mediante un conjunto de herramientas donde se aplican diversas técnicas y tecnologías basadas en minería de datos, inteligencia colectiva, computación en la nube, web semántica, aprendizaje automático y sistemas de recomendación.

ABSTRACT. The concept of e-Science 3.0 has been coining in recent years as the use of new intelligent web tools by researchers to share, collaborate and publish easily their results. LIDIS research group has designed a web platform in order to support researchers in developing research projects using concepts of e-Science 3.0, using a set of tools which various techniques and technologies based on data mining, collective intelligence, cloud computing, semantic web, machine learning and recommender systems, are applied.

PALABRAS CLAVE: Arquitectura de Software, Minería de Datos, Aprendizaje Automático, Web Semántica.

KEYWORDS: Software Architecture, Data Mining, Machine Learning, Semantic Web.

1. Introducción

Actualmente nuevas herramientas disponibles en internet brindan la posibilidad a los investigadores acceder a muchas más fuentes de información que en el pasado. Estas fuentes proveen información relacionada con las temáticas de los proyectos de investigación, adición de aspectos propios de la gestión de dichos proyectos (Kardefelt-Winther, 2014). Lo anterior ha derivado en el aumento significativo en la cantidad de información disponible, lo cual aumenta significativamente el tiempo de análisis que poseen los investigadores (Gerpott y Thomas, 2014). Por lo tanto surge la necesidad de implementar mecanismos que faciliten la recopilación, almacenamiento, procesamiento, análisis y recuperación de la información relevante existente en bases de datos especializadas tales como: journals, portales, librerías digitales, foros, redes sociales de científicos y sitios especializados de discusión, etc. Las cuales cuentan con cierto nivel de confianza sobre su calidad (T. V. B., 2013).

Por otra parte, a pesar de contar con más información persisten limitantes para que los investigadores establezcan relaciones efectivas de colaboración que permitan acceder de forma más precisa a trabajos e iniciativas de sus colegas. Estas limitantes reducen la oportunidad para trabajar en equipos multidisciplinarios y de forma interinstitucional [López y San-Juan, 2013]. En ese sentido e-Science 3.0 se presenta como una herramienta para potencializar la investigación (Hung, 2010). Esto dadas sus características tecnológicas avanzadas que en consonancia con las tecnologías de la Web 3.0 aportan al trabajo colaborativo, permitiendo hacer uso de las ventajas que actualmente aporta la Web.

Con base a lo anterior, en el presente artículo se presenta, el diseño de una plataforma basada en los conceptos y tecnologías de e-Science 3.0 la cual pretende apoyar el trabajo de los investigadores de forma colaborativa en las tareas propias del quehacer investigativo. La plataforma cuenta con un repositorio el cual indexa varias fuentes de información con resultados de investigación científica, tales como: ScienDirect, IEEE Xplore Digital Library, ACM Digital Library, dblp: DBLP Bibliography, entre otras.

El resto del documento está organizado de la siguiente manera, en la sección 2 es descrita conceptualmente la plataforma e-Science 3.0, la sección 3 describe el rol del modelo de colaboración en la plataforma, en la sección 4 la Arquitectura de la plataforma y finalmente en la sección 5 las conclusiones y el trabajo a futuro.

2. Trabajos relacionados

Con la irrupción de las tecnologías web y el modelo de computación en la nube se ha aumentado la disposición hacia compartir recursos e información en proyectos de investigación, motivo por el cual han aparecido algunas plataformas tecnológicas en la nube, algunas de uso libre y otras de pago que brindan diferentes niveles de integración de datos entre los recursos y herramientas disponibles en internet (Narock y Fox, 2012). En el contexto colombiano cabe resaltar la iniciativa en ese sentido del Centro Virtual de Estudios en Altas Energías (CEVALE2) (Townend, Xu y Austin, 2012).

Sin embargo la atención desde los primeros años se han concentrado principalmente en temas de infraestructura y servicios de procesamiento de datos (De Roure y Goble, 2010) utilizados en áreas como bioinformática, computación, astronomía, física y medicina entre otros, donde las aplicaciones van desde la utilización de sistemas de workflow científicos hasta herramientas para gestionar la publicación de resultados en la nube. Un ejemplo es e-SC (Hiden et al., 2012), una plataforma que está desplegada en la nube y que permite a los científicos acceder a datos, procesarlos y generar resultados utilizando simplemente un navegador de internet común y corriente.

En el contexto iberoamericano cabe destacar el estudio realizado por Arcila, Piñuel y Calderín (2013) donde a través de una encuesta en línea y un posterior panel de expertos se pudieron obtener resultados sobre las actitudes de los investigadores hacia la e-Science, pero tal vez lo más importante, la forma como en particular en nuestra región los investigadores ven internet en relación con los hábitos de trabajo colaborativo y la dinámica para compartir el conocimiento generado. Aunque en los resultados del estudio anteriormente citado



se identificó una actitud positiva hacia la e-Science por parte de los investigadores también se vislumbró un bajo conocimiento o uso frecuente de herramientas con tecnologías avanzadas para trabajar en proyectos colaborativos. Igualmente se identificó una percepción generalizada (mas del 50% de encuestados) que las herramientas básicas como el correo electrónico y el software de videoconferencia son suficientes para conducir un trabajo de investigación entre colegas, pero más aún y un poco preocupante es que mas del 60% de los investigadores iberoamericanos nunca han presentado un proyecto con pares de otras instituciones diferentes a la suya, aunque hubiesen tenido alguna participación en comunidades virtuales y hubiesen expresado un alto grado de satisfacción en ellas.

La idea es entonces basados en la oportunidad detectada anteriormente, ofrecer una plataforma e-Science que integra diferentes técnicas de descubrimiento de información de tal forma que los investigadores podrán conocer la relación de sus trabajos y áreas de interés con trabajos más allá de lo que ellos normalmente consideran como su rango de acción y colaboración. Estas técnicas son ampliamente utilizadas en sitios web de uso masivo como NetFlix, Youtube, Facebook, entre otros. Al final se busca que los investigadores participantes de la plataforma de e-Science tengan una herramienta para mejorar su nivel de producción científica.

3. Diseño conceptual de la plataforma e-Science 3.0

Actualmente se tienen diferentes iniciativas para que la información y resultados de los proyectos de investigación estén disponibles a un público más extenso a través de comunidades y plataformas como ResearchGate, Mendeley, LinkedIn, entre otras (Chunli, 2012), las cuales claramente un enfoque colaborativo y que además permiten la interoperabilidad con bases de datos especializadas en resúmenes de artículos, páginas de journals en línea, portales científicos y muchos otros de los recursos disponibles en la red (Stapleton, Helms-Park y Radia, 2006; Sun et al., 2012).

Ante lo anterior nos planteamos la siguiente pregunta: Como gestionar la gran cantidad de información disponible, pero almacenada en diferentes formatos de tal forma que apoye la colaboración entre investigadores en sus proyectos de investigación.

Como respuesta a este interrogante se plantea el diseño de la plataforma presentada en este artículo (ver figura 1). La plataforma está compuesta por una serie de módulos que permiten el trabajo colaborativo de los investigadores, un conjunto de herramientas web que permiten el acceso a la información recolectada en la plataforma, la cual puede ser almacenada y gestionada en la nube. A continuación se describe cada uno de los módulos y sus componentes.

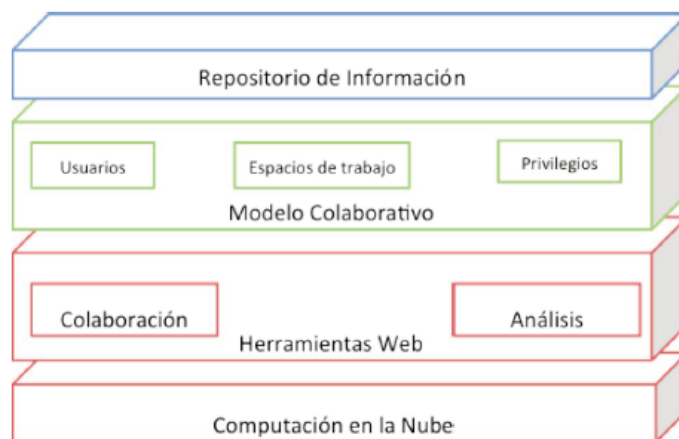


Figura 1. Diseño conceptual de la plataforma colaborativa basada en e-Science 3.0.

Repositorio de Información: Este módulo resuelve la persistencia de la información que es obtenida por la plataforma a través de diferentes herramientas como crawlers, web services, etc. El repositorio soporta datos estructurados y no estructurados.

Modelo de Colaboración: Este módulo se enfoca en resolver la forma en que los investigadores pueden compartir y colaborar con sus colegas a través de la plataforma pues son ellos quienes con acciones explícitas o tácitas dan valor a los datos obtenidos y procesados. El modelo de colaboración resuelve entonces inquietudes sobre quien, como, donde y cuando la información es obtenida, procesada, clasificada y valorada. Aunque en estudios anteriores se ha probado como la influencia de la colaboración está asociada a la productividad en la investigación [4], los resultados preliminares al responder estas preguntas apuntan a que no existe un patrón predeterminado en la forma como la relación conocimiento-investigación es compartida y esta responde a diferentes factores de diversa índole como el tipo de investigación, las temáticas y áreas de conocimiento, la afiliación de los investigadores, el contexto sociodemográfico de los proyectos que producen resultados de investigación, la presencia de patentes involucradas en dichos resultados, entre otros factores (Hung, 2010).

En el modelo de colaboración es entonces donde se incorporan los diferentes elementos de un entorno colaborativo como son la definición de colaboradores o usuarios, los grupos o redes de colaboración llamados también espacios de trabajo, y los aspectos de permisos de acceso sobre los datos publicados definido esto último como privilegios.

Debido a que no existen patrones comunes únicos de colaboración, el diseño contempla un modelo inicial que se puede extender (basados en repositorios de citas y resúmenes de artículos, bases de datos especializadas, públicas y privadas, redes sociales, sitios web científicos, etc.) por lo cual se ha considerado la aplicación de tecnologías de la web semántica para tal efecto (Narock y Fox, 2012).

Herramientas Web: El objetivo de este modulo es potencializar por parte de los investigadores su participación en la formulación de proyectos de alta complejidad, propiciar las dinámicas para compartir resultados de investigación haciendo más visibles sus trabajos entre sus colegas. El diseño contempla un conjunto de herramientas básicas típicas de un entorno colaborativo en línea como son blogs, wikis, etc. Sin embargo existe otro grupo de herramientas consideras de análisis en la cual se aplican modelos y técnicas de las áreas de la computación y las tecnologías de la información como son: minería de datos, máquinas de aprendizaje, sistemas de recomendación, web semántica, cuyo objetivo es el brindar información más clara y clasificada a los investigadores (Townend, Xu y Austin, 2012).

Computación en la nube: El último elemento en el diseño conceptual de la plataforma es el modelo de distribución de trabajo computacional a utilizar para resolver el aspecto de la velocidad de procesamiento de los datos. Con respecto a este aspecto, desde el año 2001 en la primera reunión europea de e-Science (All Hands Meeting, UK), se ha evolucionado desde la discusión del uso de la computación en grid hacia la computación en la nube. La motivación para esto son las ventajas que en disponibilidad de recursos que el modelo de computación en la nube aporta de forma elástica a las necesidades de almacenamiento y desempeño (T. V. B., 2013).

4. Arquitectura de la plataforma

La arquitectura de software que soporta la construcción de la plataforma está definida siguiendo el estilo de arquitectura de pizarra (Lyberis et al., 2014). Este estilo permite el desarrollo de soluciones que no tienen un patrón determinístico, si no que responden a la aplicación de diferentes soluciones parciales que en forma colaborativa que construyen una solución final que no es necesariamente la más óptima (Capilla et al., 2014). La anterior caracterización del estilo de arquitectura de pizarra coincide en muchos aspectos presente en el diseño conceptual de la plataforma como son la multiplicidad de fuentes de información que aportan a una solución, la multiplicidad de herramientas con diferentes técnicas y modelos computacionales aplicadas al descubrimiento de información, entendiendo esto último como nuevo conocimiento generado entre científicos

que determinan claramente las heurísticas para establecer relaciones de confianza entre colegas y veracidad de los datos.

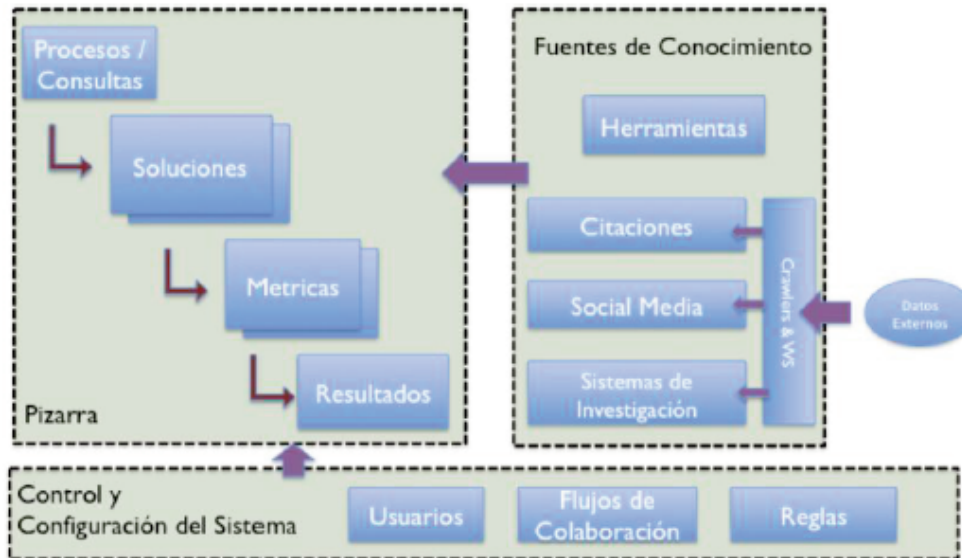


Figura 2. Arquitectura de Software de la plataforma e-Science 3.0

La arquitectura de la plataforma cuenta con los siguientes componentes principales:

Pizarra (blackboard): Representa el espacio de soluciones posibles de colaboración, las fuentes de conocimiento que están compuestas por las herramientas de adquisición de información (así como de otros sistemas de información).

El espacio de soluciones es un área de tránsito o de soluciones parciales en la resolución de una consulta, la cual se define como una posible relación de colaboración entre los usuarios de la plataforma. Este espacio de soluciones recibe principalmente todos los datos obtenidos y filtrados de las diferentes fuentes de información existentes en la plataforma. El componente observa una estructura jerárquica que permite descomponer en niveles cada elemento que participa en la solución final. El componente de procesos y consultas (queries) ofrece el más alto nivel de solución a una oportunidad o necesidad de colaboración. En este nivel se definen cuales consultas y procesos son las más apropiadas para entregar una respuesta. En el siguiente nivel se abordan las instancias de posibles soluciones obtenidas a través de los procesos y consultas ejecutados en el nivel anterior. En el tercer nivel, el de métricas, se valoran las calificaciones de las instancias de las soluciones y en el último nivel se entregan los datos resultantes de cada herramienta o fuente de información obtenida.

Fuentes de Conocimiento: Con respecto al componente de fuentes de conocimiento, la arquitectura está concebida de tal manera que cada fuente sea independiente una de otra, esto siguiendo claramente el patrón de arquitectura adoptado. Cada fuente de información tiene representada su implementación en un componente independiente (crawlers, web services, citations databases, etc.). La dinámica propia de estos componentes es obtener información externa a la plataforma, filtrarla dependiendo de su tipo o formato y posteriormente entregarla bajo demanda de los procesos que se disparan desde la pizarra en un modelo de comunicación publicador-subscriptor. Para esto se puede apoyar en el componente denominada Herramientas el cual es explicado de forma amplia más adelante.

Control y configuración del sistema: El componente de configuración y control del sistema implementa el modelo de colaboración de la plataforma. Se basa en tres subcomponentes principales: Los usuarios (normalmente investigadores), los flujos de trabajo predefinidos que permiten definir como interconectar e intercalar

procesos de consulta sobre el espacio de soluciones y las reglas, las cuales definen aspectos sobre privilegios y permisos de los usuarios sobre los datos almacenados en la plataforma.

Herramientas de la plataforma: Las herramientas disponibles en la plataforma implementan los algoritmos que resuelven las consultas sobre posibles relaciones de colaboración entre los investigadores, estas construyen un modelo de pizarra a partir de soluciones parciales entregadas por parte de cada herramienta. Existen en el diseño herramientas preconcebidas basadas en sistemas de recomendación, minería de datos, aprendizaje de máquina y algoritmos de análisis predictivo.

Herramienta	Técnicas o tecnologías utilizadas
Recomendador de Proyectos	Motor de Recomendación
Observatorio de Convocatorias	Minería de Datos
Observatorio de llamados a trabajos	Motor de Recomendación
Generador de Iniciativas de proyectos	Minería de Datos
Compositor de flujos de colaboración	Maquinas de aprendizaje

Tabla 1. Herramientas inicialmente concebidas en la plataforma.

Recomendador de proyectos de investigación. Esta herramienta utiliza tecnologías de motores de recomendación para encontrar similitudes entre proyectos en los cuales participan diferentes investigadores. Dicha recomendación se base en dos criterios, el primero son los puntajes de clasificación dados sobre la relevancia de los proyectos con respecto a un proyecto en particular y el segundo es la similitud en la definición del marco lógico entre los mismos proyectos. Esto se conoce como un filtro colaborativo doble aplicado en un sistema de recomendación y que se puede observar en sistemas en línea como yahoo.com, amazon.com y netflix.com (Hiden et al., 2012).

Observatorio de convocatorias para financiación de proyectos. La idea es poder suministrar información a los investigadores sobre posibles fuentes de financiamiento disponibles para sus investigaciones pero no solamente proveyendo una lista de dichas fuentes si no una clasificación de la experiencia de los investigadores frente a ellas, indicando su pertinencia, facilidad de acceso, ejecución y otros aspectos de especial relevancia para los proyectos de investigación.

Observatorio de llamados a trabajos de investigación, donde se aplicarán el mismo concepto de usar criterios ampliados de búsqueda que brinden una mayor experiencia de uso al investigador.

Generado de iniciativas de proyectos de investigación el cual mediante aplicación de técnicas de minería de datos permite sobre una caracterización de los perfiles de investigadores, autores de artículos, libros, blogs, etc. establecer posible relaciones de agrupación entre ellos y otros interesados en resultados de investigación.

Es posible que en casos particulares, una solución se resuelva a partir de una sola fuente de datos, pero lo más común es que se requiera de una colaboración de diferentes herramientas accediendo múltiples fuentes de datos, por lo cual en el diseño se ha incorporado una herramienta basada en aprendizaje de máquina que permita descubrir nuevos patrones de colaboración basados en las reglas iniciales definidas en el modelo. En el diseño de la plataforma hemos llamado dicha herramienta como el compositor de flujos de colaboración.

4. Conclusiones y trabajos futuros

Se ha diseñado una plataforma de colaboración para investigadores siguiendo principios de e-Science 3.0. El diseño de la plataforma contempla la composición de un grupo de herramientas colaborativas basadas en diferentes técnicas y modelos computacionales que permiten mejorar la experiencia de los investigadores frente



a plataformas con herramientas tradicionales de colaboración.

La plataforma está diseñada para implementarse siguiendo un patrón de arquitectura de pizarra lo cual permite agregar por demanda nuevas fuentes de conocimiento. La plataforma igualmente soportar la creación de nuevos flujos de colaboración entre investigadores basando en un modelo semántico.

Inicialmente se ha implementado el sistema de recomendación de proyectos el cual actualmente se encuentra en una prueba piloto. En el plan de trabajo sigue la implementación de las demás herramientas, así como la inclusión en el diseño de otras nuevas, producto de la experiencia que se tenga con las inicialmente preconcebidas e implementadas.

Se espera que la plataforma e-Science 3.0 se convierta en un referente nacional y regional como espacio virtual de colaboración científica y de esta manera los investigadores participantes de la plataforma accedan con mayor interés, autonomía y nivel de confianza a las oportunidades de colaboración con otros colegas potencializando su trabajo y los resultados de los proyectos de investigación.

Cómo citar este artículo / How to cite this paper

Barraza, F., Ordoñez, H., & Segovia, E. (2016). Diseño de una Plataforma de Colaboración para investigación Científica basada en e-Science 3.0. *Campus Virtuales*, Vol. 5, Num. 1, pp. 92-98. Consultado el [dd/mm/aaaa] en www.revistacampusvirtuales.es

Referencias

- Arcila, C., Piñuel, J. L., & Calderín, M. (2013). La e-investigación de la Comunicación: actitudes, herramientas y prácticas en investigadores iberoamericanos. *Revista Comunicar*, (40), 111-118.
- Capilla, R., Bosch, J., Trinidad, P., Ruiz-Cortés, A., & Hinchey, M. (2014). An overview of Dynamic Software Product Line architectures and techniques: Observations from research and industry. *J. Syst. Softw.*, 91, 3-23.
- Chunli, L. (2012). On the Application of the Web-based Instructional Research in Teaching. *Energy Procedia*, 17, 723-727.
- De Roure, D., & Goble, C. (2010). Supporting e-Science Using Semantic Web Technologies – The Semantic Grid. *Semantic e-Science. Annals of Information Systems*, 11, 29-65.
- Gerpott, T. J., & Thomas, S. (2014). Empirical research on mobile Internet usage: A meta-analysis of the literature. *Telecomm. Policy*, 38(3), 291-310.
- Hiden, H., Woodman, S., Watson, P., & Cala, J. (2012). Developing cloud applications using the e-Science Central platform. *Philosophical Transactions of Royal Society A*, 2013 371, publicado 10 diciembre 2012.
- Hung, E. (2010). Los escenarios de colaboración académica desde el uso de las tic y escenarios virtuales. Caso CEVALE2. *Revista Digital RetoDigital*. Fundación Universidad del Norte. Colombia.
- Kardfelt-Winther, D. (2014). A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use. *Comput. Human Behav.*, 31, 351-354.
- López, R., & San-Juan, J. F. (2013). An e-Science project in Astrodynamics and Celestial Mechanics fields. *Comput. Phys. Commun.*, 184 (5), 1381-1386.
- Lyberis, S., Kalokerinos, G., Lygerakis, M., Papaefstathiou, V., Mavroidis, I., Katevenis, M., Pnevmatikatos, D., & Nikolopoulos, D. S. (2014). FPGA prototyping of emerging manycore architectures for parallel programming research using Formic boards. *J. Syst. Archit.*, 60(6), 481-493.
- Narock, T., & Fox, P. (2012). From science to e-Science to Semantic e-Science: A Heliophysics case study. *Comput. Geosci.*, 46, 248-254.
- Stapleton, P., Helms-Park, R., & Radia, P. (2006). The Web as a source of unconventional research materials in second language academic writing. *Internet High. Educ.*, 9(1), 63-75.
- Sun, M., Dou, H., Li, Q., & Yan, Z. (2012). Quality Estimation of Deep Web Data Sources for Data Fusion. *Procedia Eng.*, 29, 2347-2354.
- T. V. B. (2013). The future of science in the internet. *Portal Res*. 2012.
- Townend, P., Xu, J., & Austin, J. (2012). e-Science-towards the cloud: infrastructures, applications and research. *Philosophical Transactions of Royal Society A*, 2013 371, publicado 10 diciembre 2012.