



FRT-2

Título del Proyecto de Investigación a que corresponde el Reporte Técnico:

Modelo de distribución de conocimiento a través de redes sociales como apoyo a la toma de decisiones.

TÍTULO DEL REPORTE TÉCNICO

Modelo de distribución de conocimiento a través de redes sociales como apoyo a la toma de decisiones.

Área del conocimiento al que pertenece

Ciencia de la Computación e Inteligencia Artificial

Autores del reporte técnico: (Apellidos, Nombre(s))

Martínez Silva José de Jesús.

Dr. Rodas-Osollo Jorge.

Dra. Olmos Sanchez Karla.

## **Modelo de distribución de conocimiento a través de redes sociales como apoyo a la toma de decisiones.**

### **Resumen del reporte técnico en español:**

Dentro de la resolución de problemas donde se trabaja con conocimiento del dominio de estructura informal como lo son los proyectos de desarrollo de innovación. El conocimiento requerido suele ser heterogéneo, estar distribuido o requerir de especialistas altamente capacitados a pesar de que dicho conocimiento se encuentre de una manera formal. En el conocimiento formal podemos encontrar una serie de metadatos con los cuales se relacionan autores y otros documentos a través de referencias, con dichas relaciones se puede formar una red social y al mismo tiempo formar un modelado por grafos de conocimiento usando cada publicación o unidad de conocimiento como un nodo y las referencias como enlaces. Realizando un análisis que permita hacer explícitas a las redes de conocimiento que en conjunto y de forma implícita poseen conocimiento que apoye a quienes participan en el proceso de generación de productos innovadores. Este estudio presenta una solución cognitiva para el análisis, procesamiento y modelación de una base de conocimiento como herramienta de apoyo para al desarrollo de proyectos de innovación.

**Resumen del reporte técnico en inglés:**

Within the resolution of problems where it works with knowledge of the domain of informal structure such as innovation development projects the knowledge required is usually heterogeneous, distributed or require highly trained specialists despite the fact that such knowledge could be found in a formal way. In the formal knowledge we can find a series of metadata with which authors and other documents are related through references, with these relationships can be formed a social network and can formed a modeling of knowledge distribution by graphs using each publication or unit of knowledge as a node and references as links. Performing an analysis that allows make explicit knowledge networks that together and implicitly have knowledge that supports those who participate in the process of generating innovative products. This study presents a cognitive solution for the analysis, processing and modeling of a knowledge base as a support tool for the development of innovation projects.

**Palabras clave (Al menos 3):** Redes de Conocimiento, Toma de decisiones, redes sociales.

**Usuarios potenciales (del proyecto de investigación):**

Empresas generadoras de innovación, Estudiantes, investigadores, entidades que requieran de grandes cantidades de conocimiento formal que sobrepasen sus capacidades de análisis para la resolución de problemas.

---

---

**Reconocimientos (agradecimientos a la institución que financió, estudiantes que colaboraron, instituciones que apoyaron a la realización del proyecto, etc.):**

---

---

## INTRODUCCIÓN

En la actualidad las empresas de desarrollo de innovación que atienden las necesidades de productos innovadores que demandan las empresas trabajan con dominios de estructura informal. Para obtener el conocimiento necesario se requiere de entidades proveedoras de conocimiento que en conjunto forman redes de trabajo implícitas. Esta compleja tarea exige del análisis, procesamiento y modelación de conocimiento que con el paso del tiempo crece su nivel de complejidad debido al crecimiento exponencial de conocimiento disponible y la integración de distintas tecnologías como aplicación para el desarrollo innovador de nuevas soluciones. El conocimiento necesario puede ser tácito, conocimiento explícito, conocimiento heterogéneo e inclusive en ocasiones de la generación de nuevo conocimiento. En particular el conocimiento heterogéneo y el explícito pueden encontrarse distribuidos debido a las distintas fuentes de conocimiento y al gran número de investigadores que distribuyen con publicaciones a través de internet o distintos medios de manera internacional.

La cantidad de conocimiento formal va en aumento en la actualidad de una manera casi exponencial gracias a un acelerado desarrollo tecnológico, la globalización y otros factores que han impulsado el desarrollo de investigaciones y publicaciones. Dichas publicaciones formales se encuentran distribuidas en distintos repositorios en internet. Analizar una fuente de conocimiento muy vasta y tan distribuida para el desarrollo de investigaciones o resolución de problemas puede acarrear costos elevados de tiempo, recurso humano y monetarios, que de no tenerse pondrían en riesgo la calidad del resultado del proyecto.

Una de las principales fuentes de conocimiento son los artículos científicos, medio que cuenta con una estructura formal que puede ser fácilmente consultada: título, autores, referencias, resúmenes, palabras clave, por mencionar los más importantes. De acuerdo con datos obtenidos del portal SciELO.org [1],[2] (Scientific Electronic Library Online por sus siglas en inglés), una de las hemerotecas virtuales más importantes en américa latina. Los accesos a

archivos ya sean en html, pdf o, abstractos ha ido en aumento incremental tal y como se muestra en la figura 1.

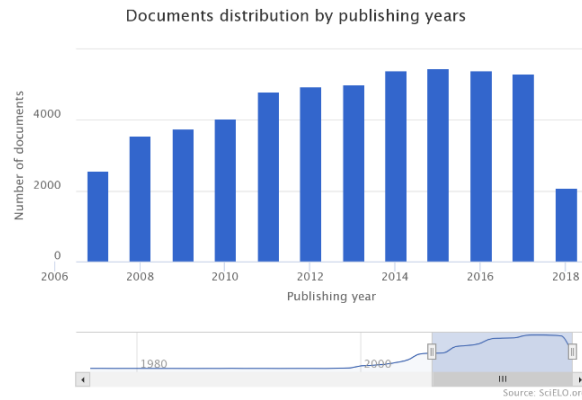


Fig. 1. Archivos registrados en SciELO.org por año.

## PLANTEAMIENTO

El proceso de generación de productos innovadores involucra todo un ejercicio de resolver problemas, satisfacer necesidades o atender dificultades que exigen utilizar conocimiento nuevo o muy actualizado y altamente especializado que puede ser tácito o explícito. Además, dicho conocimiento suele encontrarse disperso: el tácito geográficamente y el explícito en distintas fuentes en internet por medio de publicaciones y repositorios. Para potenciar el proceso de generación es conveniente identificar la red o conjunto de redes que poseen el conocimiento para atender las dificultades implícitas en el mismo proceso; lo cual, es una ardua tarea que demanda una gran destreza y una buena inversión de tiempo por parte de analistas de información para la localización del conocimiento especializado necesario para proveer el producto pretendido.

## Objetivo General.

Brindar soporte al desarrollo de proyectos de innovación mediante el uso de una solución cognitiva para el manejo de la convergencia del conocimiento en dominios de estructura informal utilizando artículos científicos.

## Objetivos específicos.

- Identificar las necesidades de gestión de conocimiento en las empresas de desarrollo de innovación.
- Desarrollar una aplicación cognitiva para el apoyo en procesamiento de conocimiento en proyectos de innovación.
- Verificar el beneficio y viabilidad del uso de una herramienta de apoyo en un ambiente productivo.

El crecimiento de información a causa de avances tecnológicos que se incrementa de forma exponencial año tras año demanda la implementación rápida de soluciones cognitivas que potencien su análisis. Las organizaciones que realizan procesos de generación de innovación realizan análisis exhaustivos de información para localizar el conocimiento requerido. El desarrollo de una herramienta cognitiva que pueda localizar el conocimiento aporta beneficios de tipo técnico a proyectos ayudando en la reducción de tiempo; tiempo que puede significar costo monetario o ventajas competitivas en las organizaciones.

## Antecedentes

### Convergencia Tecnológica

La convergencia tecnológica, la cual nos ayuda entender cómo es que se va dando el nacimiento de nuevos campos de estudio tecnológico gracias a que elementos discretos o heterogéneos ya sean tecnologías o distintas industrias se fusionan dando lugar a un nuevo campo de estudio, por ejemplo: La en la robótica, esta rama tecnológica emerge a través de la unión de otras tecnologías como la ingeniería mecánica, sistemas de control, electrónica y software [3]. Al mismo tiempo esta tecnología podría formar parte de las ramas que convergen en otra tecnología.

Un método reciente para la medición de la relación tecnológica es el método de “Module-based mining methodology” presentado por Kose, T., & Sakata, I. (2018) [4] aplicado en el campo de la robótica. Este método busca identificar la convergencia tecnológica de una manera más precisa utilizando un método de modulación de Newman y aplicando un análisis a las redes de referencias.

### Análisis de Redes de Referencias

El análisis de redes de referencias es una forma de minado de tecnología también llamado “Tech mining” el cual está compuesto por una combinación de metodologías de investigación científica. Esta técnica también ha sido utilizada para determinar o detectar nuevas tecnologías y para rastrear tecnologías emergentes (Tu and Seng, 2012) [6] incluyendo el análisis de redes de referencias.

[4] Toshiro Kose y Ichiro Sakata usan el análisis de redes de referencias para categorizar las publicaciones, factorizar los sectores superpuestos en un simple sector y calcular y evaluar la convergencia de tecnologías.

## Marco teórico

### Social Network Clustering.

La clusterización por “caminata aleatoria” es un proceso que puede encontrar comunidades en una red, en otras palabras, cuando se usa el método de “caminata aleatoria”, se escanean los nodos en algunos pasos; comenzando con un nodo inicial y, en función de un proceso aleatorio, progresa a los nodos vecinos.

Una de las características más comunes de estas redes se llama estructura de la comunidad, que representa grupos conectados (clusters) donde deben existir muchos enlaces dentro de cada grupo y pocos entre los grupos. Los grupos resultantes son una fracción de individuos que tienen características similares o están conectados a través de relaciones. Los grupos en las redes sociales se corresponden con las relaciones sociales y se utilizan para comprender la estructura de datos, como las estructuras de la organización, la colaboración científica y las relaciones en las redes de telecomunicaciones.

El objetivo de la detección de la comunidad de gráficos es la identificación de módulos y su estructura jerárquica mediante el uso de la información codificada en la topología de gráficos [10].

### Modularidad de redes sociales.

Dentro del análisis de redes sociales complejas existen indicadores que nos ayudan a identificar que tan buena es la partición dentro de una red social. Uno de los estudios más completos lo muestra Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. [11], en donde se muestra como medir la sensibilidad de varios métodos de modularidad, los clasifican dependiendo su alcance y se comparan evaluando el costo computacional y su desempeño dependiendo en las redes que son aplicados. De manera individual el set de algoritmos presentado por Newman, M. E., & Girvan, M. [12] muestra un proceso iterativo que entrega un métrico objetivo con el cual se pueden identificar el mejor número de comunidades en las cuales debe de estar dividida una red social, demostrando ser altamente efectivo con costo de computo moderado de acuerdo con Danon, L. et all (2005) [11].

## METODOLOGÍA

### Herramienta.

Para poder apoyar a las empresas generadoras de productos de innovación desarrollamos una herramienta cognitiva que nos apoyara en la identificación de redes de conocimiento, realizar los clusters en base a la convergencia del conocimiento y al mismo tiempo poder ofrecer una utilería de búsqueda de artículos científicos usando las redes de conocimiento dentro de la base de conocimiento para brindar soporte durante la toma de decisiones. La herramienta cognitiva que hemos desarrollado cuenta con la base de datos de aminer.org en su versión DBLP-Citation-network V4 [18].

### Redes de Conocimiento

Con el alto crecimiento de publicaciones científicas y su correcta documentación, en la actualidad encontramos millones de artículos disponibles en librerías electrónicas con acceso al público en general, los cuales individualmente cada uno referencian una lista de artículos que



representan el conocimiento base usado durante su desarrollo. De esa manera se han ido creando enlaces entre todas las publicaciones realizadas, algunos autores ganando muchas más conexiones gracias a la importancia o novedad del tema y algunos otros quedando un poco pobres de enlaces, con estas características podemos representar el conocimiento utilizando nodos y enlaces que formaran una red, esto es a lo que llamamos redes de conocimiento.

Título	Referenciado	Referencias
Sistemas en tiempo real y embebidos.	3	2
Programación distribuida y multiprocesador.	6	2

Tabla 1. Ejemplo de número de referencias y número veces que es referenciado un artículo. (Base de datos Aminer v2).

Visualizando la red de conocimiento resultante se podría observar que artículos se encuentran cerca o más lejos de la nube de conocimiento con mayor interés o donde se está desarrollando con mayor fuerza el tema y así poder discriminar datos individuales que a pesar de su número de relaciones no se encuentra tan cerca de aquellos que si son parte primordial de la nube de conocimiento en el tema que se está investigando.

### Nodos y Relaciones.

Para identificar los nodos y relaciones existentes entre artículos científicos utilizamos la definición de tipos de citas tal y como las nombra (Kose T. & Sakata I. 2018) [4]. En donde las "citas directas" son las que se encuentran dentro de las referencias de cada artículo, las "uniones bibliográficas" son aquellos documentos X, Y que tienen una cita directa a un documento A, entre ellos existe una unión bibliográfica, por ultimo tenemos las "Co-Referencias", así llamadas las relación que existe entre los listados de documentos referenciados S y T en un documento inicial A tal y como se muestra en la figura 1.

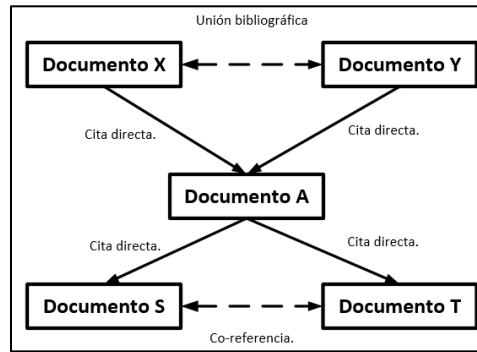


Fig. 1 Definición de tipos de citas y relaciones.

Las referencias directas son los enlaces principales para las uniones de este proyecto, y contendrán un valor y etiqueta “3”. Los enlaces por “Co-referencia” son los enlaces de segunda importancia los cuales contendrán un valor y etiqueta “2”. Por último, se usarán los enlaces tipo “Unión bibliográfica” con un valor y etiqueta “1”. Estas valoraciones serán utilizadas para la representación gráfica mostrando solo las uniones tipo 3 y 2. Los enlaces de tipo 1 serán mostrados en la representación gráfica solamente cuando un nodo o artículo científico aparece en los resultados, pero sin ningún enlace tipo 3 o 2. Esta categorización de enlaces entre artículos es de vital importancia para el desarrollo de esta solución cognitiva y generación del algoritmo del manejo de innovación.

### Detección de comunidades y modularidad.

Con el uso de la modularidad  $Q$  de Newman y Girvan se identificaran las comunidades en donde sus enlaces son más densos y en donde son muy débiles, para así generar clusters que delimitaran ciertos grupos de relaciones dándonos como resultado las redes de conocimiento. El tipo de relación usado ya sea referencias directas, uniones bibliográficas o co-referencias influyen directamente en el resultado, pero estas opciones serán seleccionables en el sistema.

$$Q = \sum (e_{ii} - a_i^2) = Tr(e) - \|e^2\|$$

Fig. 2 Formula de Modularidad  $Q$  Newman y Girvan.

## RESULTADOS

Usando la base de datos de aminer.org (DBLP-Citation-network V4) se realizaron una serie de pruebas en donde se generaron grafos con las relaciones registradas en las referencias directas. La cantidad de datos graficada fue limitada a 8000 nodos analizados como prueba.

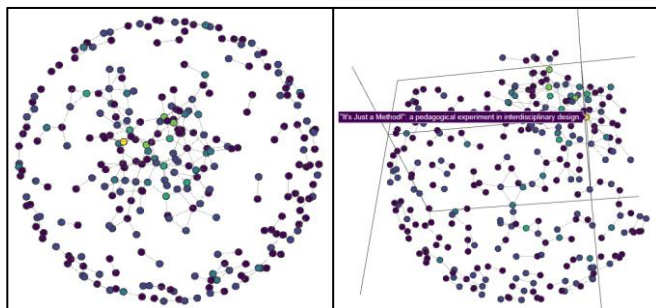


Fig. 3 Grafo de relaciones DBLP-Citation-network V4 limitada a 8000 nodos (vista 3D).

Para conocer el desempeño de la respuesta de la base de datos en conjunto con el proceso de análisis del software, se puso a prueba el sistema utilizando cantidades en específico de datos para su análisis en 30 ciclos repetitivos cada una, los resultados son los siguientes:

Tiempo promedio en seg.	Registros analizados
10.77487738	8000
13.45824571	10000
24.6751673	15000
36.59711232	20000
120.2087197	30000
266.6108944	40000

Fig. 4 Tabla resultado de pruebas realizadas.

## CONCLUSIONES

La limpieza de nodos, identificación de relaciones y clasificación por número de referencias fue posible gracias al uso de la herramienta descrita y nos muestra las capacidades para la implementación de los métodos de identificación de comunidades en redes sociales que se quieren utilizar, como lo es el método de modularidad Q descrito en esta investigación. En cuanto al tiempo de desempeño se encontró que, a partir de los 30 000 registros el tiempo comienza a mostrar una tendencia exponencial lo cual nos indica que las capacidades de computo de la maquina podrían estar llegando a su límite mermando el buen desempeño del proceso.

## REFERENCIAS (bibliografía)

- [1] SciELO, hemeroteca virtual conformada por una red de colecciones de revistas científicas en texto completo y de acceso abierto y gratuito: <http://www.scielo.org.mx/scielo.php#about>
- [2] Análisis de cantidad de publicaciones realizadas en los últimos 28 años: [https://analytics.scielo.org/w/publication/article?py\\_range=1990-2018](https://analytics.scielo.org/w/publication/article?py_range=1990-2018)
- [3] Kose, T., & Sakata, I. (2019). Identifying technology convergence in the field of ro-botics research. *Technological Forecasting and Social Change*, 146, 751-766.
- [4] Kose, T., & Sakata, I. (2018). Identifying technology convergence in the field of ro-botics research. *Technological Forecasting and Social Change*.
- [6] Tu, Y. N., & Seng, J. L. (2012). Indices of novelty for emerging topic detection. *Information processing & management*, 48(2), 303-325.
- [10] Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09), P09008.
- [11] Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09), P09008.
- [12] Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2), 026113.
- [13] Azizifard, N., Mahdavi, M., & Nasersharif, B. Modularity
- [14] Optimization for Clustering in Social Networks.
- [15] J. Duch, A. Arenas, "Community detection in complex networks using extremal optimization," *Physical review E* 72, 027104, 2005.
- [17] Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2), 026113.
- [18] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2008)*. Pp.990-998.