

## Avances sobre Reuso de Conocimiento en Foros de Discusión Técnicos

Gabriela Aranda, Nadina Martinez Carod, Alejandra Cechich,  
Pamela Faraci, Carina Noda, Mauro Sagripanti  
Grupo de Investigación en Ingeniería de Software del Comahue (GIISCo)  
<http://giisco.uncoma.edu.ar>  
Facultad de Informática. Universidad Nacional del Comahue  
Buenos Aires 1400, (8300) Neuquén  
Contacto: {gabriela.aranda, nadina.martinez, alejandra.cechich}@fi.uncoma.edu.ar

### Resumen

Los foros de discusión en Internet son utilizados por muchos técnicos informáticos para plantear sus dudas y pedir sugerencias para resolver algún problema particular. A partir de dicha pregunta, se abre un hilo de discusión (en el que suelen participar varios usuarios del foro), donde entre todos analizan el escenario, y por lo general, proponen una o más soluciones al problema en cuestión. Funcionando de esta manera, los foros de discusión se han convertido en plataformas colaborativas donde el conocimiento se explicita a la vez que se comparte.

Dado que existen muchos foros sobre las mismas temáticas (lenguajes de programación, aplicaciones específicas, etc.), es posible encontrar en la Web muchos hilos de discusión en diferentes foros que están relacionados al mismo problema. Cuando un técnico informático tiene un problema específico, suele utilizar un motor de búsqueda multi-propósito que le devuelve una lista extensa de páginas de varios tipos (blogs, foros, artículos, etc.), luego el técnico necesita navegar por varias páginas hasta descubrir cuál es la que describe un problema más parecido al que tiene, y encontrar (si existe) una solución que pueda satisfacerle.

Para facilitar esta tarea de los técnicos informáticos, nuestro proyecto tiene como objetivo la implementación de una herramienta que recupere la información disponible en hilos de discusión de foros técnicos de manera automática, y que a partir de un análisis basado en un modelo de calidad pertinente, permita clasificar dicha información y entregar a los usuarios un ranking de posibles soluciones para problemas técnicos recurrentes.

### Palabras Clave

Foros de discusión, Reuso de conocimiento, Modelos de Calidad.

### Contexto

La línea presentada está inserta en el contexto de los siguientes proyectos y acuerdos de cooperación:

- UNCo-FAI: “Reuso de Conocimiento en Foros Técnicos”, Subproyecto del Programa “Desarrollo Orientado a Reuso” que se encuentra en proceso de acreditación por la Universidad Nacional del Comahue.
- Acuerdo de cooperación con el Grupo Alarcos, Escuela Superior de Informática, Universidad de Castilla-La Mancha, Ciudad Real, España

## Introducción

En los últimos años se han producido importantes cambios en la disciplina de gestión del conocimiento, debido tanto a los avances en las tecnologías Web como al crecimiento desmedido (y muchas veces solapado en el tiempo) de las fuentes de información en el mundo entero. En particular, las organizaciones actuales hacen cada vez más esfuerzos para reutilizar el conocimiento, definiendo estrategias para tener catalogadas y reutilizar soluciones ya probadas en problemas recurrentes [8].

Respecto al conocimiento que está disponible en la Web hay algunos desafíos que necesitan ser considerados. Por un lado la necesidad de asegurar la *confiabilidad* de la información disponible, debido a que al ser la Web una plataforma de naturaleza libre no existen restricciones en la información publicada (y que en un gran número de veces no es debidamente verificada). Por otro lado, la distribución en un ambiente sin límites hace difícil la *usabilidad* (que sea fácil de usar), y por consiguiente la *utilidad* (que sea útil), del conocimiento subyacente, por lo que se presenta un desafío respecto a la captura y recuperación de información precisa y actualizada [9].

Desde este punto de vista, puede considerarse a la Web actual como un lugar de encuentro que, mediante el uso de herramientas colaborativas de distintos tipos (foros de discusión, blogs, wikis, etc.), posibilita debates sobre temas técnicos y, de alguna manera, se convierte en una nueva base para la revisión del conocimiento de la comunidad participante.

De las herramientas colaborativas encontradas en la Web, nuestro proyecto se enfoca en los *foros de discusión*, que

tienen la capacidad de permitir que un usuario de la comunidad pregunte sobre un problema particular o pida una opinión, y que por otro lado, al permanecer la pregunta y las respuestas visibles al resto de la comunidad, permite compartir el conocimiento creado por dicha comunidad de aprendizaje.

Dado que existen en la Web muchos foros de discusión sobre la misma temática, se pueden hallar preguntas y respuestas similares diseminadas en varios de ellos, por lo que generalmente es necesario navegar por varios hilos hasta dar con una solución correcta. Con esta idea en mente, nuestro proyecto tiene como objetivo principal favorecer el reuso de la información contenida en dichas conversaciones existentes en la Web, con el valor agregado de un análisis de calidad de dichas fuentes de información.

Se debe destacar que nuestro proyecto relaciona más de un área de investigación. Por ejemplo, más allá de la participación en de los miembros del grupo de investigación en Ingeniería del Software GIISCo, es interesante el intercambio que se realiza con otros investigadores del Departamento de Teoría de la Computación, especialmente en el estudio y aplicación de algoritmos de análisis de lenguaje natural, sentiment analysis, aprendizaje automático, etc. Esta apertura permite un trabajo colaborativo interdepartamental que enriquece a todas las áreas involucradas.

Respecto a trabajos relacionados con el análisis de lenguaje natural en foros de discusión, existe una propuesta de Tigelaar et al [14] que se enfoca en la simplificación de contenido de hilos de discusión, resumiéndolos de manera automática con un prototipo basado en lenguaje natural. Este trabajo, si bien es

un gran aporte para el análisis de los hilos de ejecución, no se enfoca en determinar si el conocimiento será de interés para la persona que lo consulta, como es el objetivo de nuestra propuesta.

Respecto al reuso de conocimiento disponible en foros de discusión, existen dos propuestas: la primera [2] implementa un sistema recomendador en base a mensajes de un foro de discusión de un curso de Inteligencia Artificial. El recomendador busca y propone mensajes con contenido similar, escritos por estudiantes de años anteriores del mismo curso. Otra propuesta [4], presenta un método de clasificación de los mensajes de un foro de discusión de una universidad de acuerdo a una jerarquía de temas preestablecida. Respecto a las diferencias de ambas propuestas con nuestro enfoque, nuestro recomendador apunta a un dominio más amplio, involucrando usuarios con distinto conocimiento previo, en lugar de dominios de e-learning de un curso o universidad en particular. Además, ambas propuestas analizan los mensajes de un único foro, lo que permite asegurar que la información se encuentra en un formato estándar y que cualquier modificación puede ser prevista. Por el contrario, nuestra propuesta apunta a recolectar información de distintos foros, por lo que la heterogeneidad de la información a capturar y la posibilidad de cambios no programados es un desafío extra.

Respecto a la generación de algoritmos de ranking basados en la calidad de los atributos, el trabajo de Kuna et al [15] tiene cierta similitud con nuestro enfoque en cuanto a la necesidad de contar con métricas de calidad, mientras que dicha propuesta cuenta con una serie de índices conocidos y reconocidos por la comunidad investigadora (como el índice SJR para revistas, CORE para

conferencias, e índice H para los autores), en nuestro caso no contamos con índices apriori para calificar la calidad de los foros de discusión y de los usuarios que escriben en ellos, por lo tanto debemos proponer un conjunto de métricas propios con su correspondiente validación.

## Líneas de investigación y desarrollo

Este proyecto de investigación está enmarcado dentro del Programa de Investigación “*Desarrollo de Software Basado en Reuso*” conducido por el grupo GIISCo. El objetivo de dicho grupo de investigación es trabajar para ofrecer soporte en investigación y transferencia de tópicos relacionados con la Ingeniería de Software.

Además de la línea de este Subproyecto, que recibe el nombre de “*Reuso de Conocimiento en Foros de Discusión Técnicos*”, el programa actualmente desarrollado por el Grupo GIISCo aborda distintos aspectos del Reuso en otras dos grandes líneas (Reuso Orientado al Dominio y Reuso Orientado a Servicios).

Respecto al subproyecto “Reuso de Conocimiento en Foros de Discusión Técnicos”, se está trabajando en dos líneas de investigación relacionadas:

1. *Captura, análisis y procesamiento de la información disponible en foros de discusión técnicos*: en esta línea se está trabajando en la elaboración de un modelo de calidad que permita reutilizar el conocimiento disponible en Internet por parte de la comunidad de técnicos informáticos.
2. *Aplicación del conocimiento adquirido*: el objetivo de esta línea es utilizar el conocimiento

adquirido durante nuestra investigación, para analizar los foros de discusión de la Plataforma de Educación a Distancia de nuestra Universidad (PEDCO) y proponer mejoras en la enseñanza de los alumnos de nuestra Facultad.

## Resultados y objetivos

En ediciones anteriores del WICC [10] [11] se presentaron las expectativas y trabajos previos de nuestro grupo de investigación y de nuestra línea de investigación en particular.

Durante el año 2013, se trabajó en la definición de un modelo de calidad para foros de discusión técnicos (basado en estándares para la calidad de datos software [1] y otros modelos de calidad de datos en la Web [3][6][7]), aplicable a foros de discusión disponibles en Internet. Dicho modelo se publicó en ASSE 2013 [12]. Respecto a esta línea de investigación, se encuentra en proceso de escritura una tesina de Licenciatura que propone un conjunto preliminar de métricas basado en el modelo anterior. También se encuentra en proceso la implementación de una herramienta que almacenará la información capturada desde los foros en la Web y que implementará el cálculo del conjunto de dicho conjunto de métricas. En la validación de estas métricas y la definición de nuevas métricas se continuará trabajando a lo largo del proyecto.

Por otro lado, se realizó una encuesta entre usuarios de foros de discusión técnicos del ámbito de nuestra Universidad, a fin de determinar el tipo de foros que suelen visitar y en qué atributos o ítems de información se basan para determinar si un hilo de discusión está relacionado o no al problema que

origina la búsqueda en internet. Los resultados preliminares de esta encuesta se presentaron en CACIC 2013 [13]. Parte del trabajo futuro será mejorar esta encuesta y extenderla a usuarios de otras universidades nacionales y extranjeras.

## Formación de recursos humanos

El proyecto continúa la línea de investigación comenzada en 2012 con el objetivo de definir un modelo de calidad y de gestión de conocimiento a partir de información contenida en foros de discusión técnicos. Para ello, el proyecto está conformado por un grupo interdisciplinario que garantiza la pluralidad de puntos de vista y cooperación entre las partes.

Forman parte del proyecto:

- Dos docentes del Departamento de Programación, con dedicación exclusiva, que han concluido en el año 2009 su Doctorado en Informática,
- Dos docentes del Departamento de Ingeniería de Sistemas y de Programación con dedicación simple, que están comenzando a formarse en investigación.
- Dos estudiantes de Licenciatura en Ciencias de la Computación, que están desarrollando sus tesis de grado.
- La colaboración de una docente del Departamento de Teoría de la Computación de la misma Facultad, que está desarrollando su tesis de Doctorado sobre técnicas de análisis de lenguaje natural y provee asesoramiento sobre algoritmos de aprendizaje automático.

- La asesoría externa de una docente e investigadora de la Universidad de Castilla La Mancha, con experiencia en Gestión de Conocimiento, lo que permite asociar desarrollos y producciones entre ambas universidades.

## Referencias

- [1] ISO/IEC 25012:2008, Software product Quality Requirements and Evaluation (SQuaRE): Data quality model. 2008.
- [2] W. Chen, R. Persen (2009), “A Recommender System for Collaborative Knowledge”.
- [3] C. Calero, A. Caro, M. Piattini (2008), “An Applicable Data Quality Model for Web Portal Data Consumers”, World Wide Web, vol. 11, no. 4, pp. 465-484.
- [4] D. Helic, N. Scerbakov (2003), “Reusing Discussion Forums as Learning Resources in WBT Systems”.
- [5] J. Dorn (2010), “Social Software (and Web 2.0)”.
- [6] I. Rafique, P. Lew, M. Q. Abbasi, Zhang Li (2012), “Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model”, International Journal of Computer and Information Sciences, vol. 6, no. 1.
- [7] R. Wang, D. M. Strong (1996), “Beyond accuracy: What data quality means to data consumers”, Journal of Management Information Systems, vol. 12, no. 4, pp. 5-33.
- [8] Smith y Duffy (2001), Re-using knowledge: why, what and where. En Proceedings de 2001 International Conference on Engineering Design, Glasgow.
- [9] P. Di Maio (2009), Toward Pragmatic Dimensions of Knowledge Reuse and Learning on the Web. Proceedings of I-KNOW'09 and I-SEMANTICS'09, Graz, Austria.
- [10] A. Cechich, A. Buccella, A. Flores, G. Aranda, N. Martínez Carod, J. Luzuriaga, R. Martínez, M. Moyano, R. Mazalu, A. Martin, M. Garriga, N. Huenchuman. *Desarrollo Basado en Reuso*, WICC 2012, XIV Workshop de Investigadores en Ciencias de la Computación, Posadas, 2012
- [11] N. Martínez Carod, G. Aranda, A. Cechich, P. Faraci, C. Noda, M. Sagripanti. *Reuso de conocimiento en foros de discusión técnicos*. WICC 2013, XV Workshop de Investigadores en Ciencias de la Computación, Paraná, 2013
- [12] G. Aranda, N. Martínez Carod, P. Faraci, A. Cechich. *Hacia un framework de evaluación de calidad de información en foros de discusión técnicos*. ASSE 2013, 14th Argentine Symposium on Software Engineering, (JAIIO 2013, 42° Jornadas Argentinas de Informática), Córdoba, 2013.
- [13] N. Martínez Carod, G. Aranda, M. Sagripanti, P. Faraci, A. Cechich. *Análisis de la información presente en foros de discusión técnicos*. WIS 2013, X Workshop en Ingeniería del Software. (CACIC 2013, XIX Congreso Argentino de Computación), Mar del Plata, pp. 847-856, 2013.
- [14] A. Tigelaar, R. Op Den Akker and D. Hiemstra, *Automatic summarisation of discussion fora*, Natural Language Engineering, ISSN 1469-8110, Vol 16, Issue 02, pp. 161-192, 2010.
- [15] H. Kuna, M. Rey, J. Cortes, E. Martini, L. Solonezen, R. Sueldo, *Generación de un Algoritmo de Ranking para Documentos Científicos del Área de las Ciencias de la Computación*, WIS 2013, X Workshop en Ingeniería del Software. (CACIC 2013, XIX Congreso Argentino de Computación), Mar del Plata, pp. 787-796, 2013.