

Reuso de Conocimiento en Foros de Discusión Técnicos

Nadina Martinez Carod, Gabriela Aranda, 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: {nadina.martinez, gabriela.aranda, alejandra.cechich}@fai.uncoma.edu.ar

Resumen

Los técnicos informáticos suelen utilizar frecuentemente los foros de discusión en Internet para realizar una pregunta sobre un problema o una dificultad para llevar a cabo alguna tarea en particular. Dicha pregunta suele ser respondida por uno o más usuarios, proponiendo una serie de soluciones al problema en cuestión. Bajo este prisma, los foros de discusión técnicos se pueden considerar como plataformas colaborativas donde es posible plantear dudas y compartir soluciones probadas que pueden ser reutilizadas por otras personas en situaciones similares.

Sin embargo, la dinámica de Internet permite que muchas preguntas y respuestas similares, o al menos relacionadas al mismo ítem planteado, se encuentren diseminadas en diversos foros de discusión. Ante la necesidad de encontrar una solución para resolver un problema en particular, un técnico necesitar navegar por varios foros y conversaciones hasta descubrir el lugar donde el problema tratado es el más semejante a su dilema, y encontrar una solución que pueda satisfacerle.

A partir de este escenario, nuestro proyecto tiene como objetivo la definición de un modelo de calidad que pueda ser aplicado a foros de discusión técnicos así como la generación de conocimiento procedente de la información contenida en ellos, complementado con la implementación y puesta en marcha de un portal donde dicho modelo sea utilizado para facilitar la búsqueda de soluciones a problemas técnicos recurrentes.

Palabras Clave

Foros de discusión, Modelos de calidad, Gestión de Conocimiento.

Contexto

La línea presentada se 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.
- PAE-PICT-2312: Métodos y herramientas para sistemas masivamente distribuidos.

Investigaciones conjuntas con
ISISTAN-UNICEN

- Acuerdo de cooperación con el Grupo Alarcos, Escuela Superior de Informática, Universidad de Castilla-La Mancha, Ciudad Real, España

Introducción

El reuso del conocimiento ha ido evolucionando a lo largo de las últimas décadas. En particular, en las organizaciones actuales existe una tendencia cada vez más marcada a reutilizar el conocimiento, definiendo estrategias para el reuso de soluciones ya probadas en casos de problemas recurrentes[8].

Más aún, en los últimos años se están produciendo cambios importantes en la disciplina, debido a los avances en las tecnologías Web y al crecimiento desmedido (y muchas veces solapado en el tiempo) de las fuentes de información en el mundo entero.

En este contexto, si bien la Web abre las barreras físicas gracias a su naturaleza pervasiva, hay una serie de desafíos que necesitan ser considerados [9]:

- Las barreras cognitivas parecen cada vez mayores en lo que respecta a la “confiabilidad” del conocimiento que se encuentra en la web, esto es debido a que al ser de naturaleza libre no existen restricciones en la información, la cual en un gran número de veces no es debidamente verificada.
- La distribución desmedida en un ambiente sin límites, hace difícil la “usabilidad” (que sea fácil de usar), y por consiguiente la “utilidad” (que sea útil), de la información o del conocimiento

subyacente, enfocándose en el desafío referente a la captura y recuperación de información precisa y actualizada.

Teniendo en cuenta dichos desafíos, se puede considerar a la Web actual como un punto de encuentro de ideas, que posibilita la creación de debates técnicos y científicos mediante varias herramientas colaborativas (foros de discusión, blogs, wikis, etc.), y que son, en cierta forma, una nueva y exhaustiva forma de revisión de conocimiento para la comunidad participante.

De las herramientas mencionadas, nuestro proyecto se enfoca en los foros de discusión, cuya característica principal es su capacidad para:

- permitir preguntar sobre un problema particular o pedir opiniones en ambientes informales,
- compartir conocimiento creado por una comunidad de aprendizaje

En los foros de discusión las opiniones y soluciones vertidas por los distintos usuarios pueden ser reutilizadas tantas veces como sea necesario. Cuando un usuario tiene un problema en especial y no encuentra una conversación directamente relacionada con él en el foro, escribe una nueva entrada (pregunta o planteo de un problema), que en general, es un pedido de aporte de soluciones. Luego, las respuestas (propuestas) de diferentes participantes van formando un hilo (thread) de discusión. Dichos hilos de discusión se convierten en una fuente de información que puede ser reutilizada en situaciones similares ya sea por los mismos técnicos que la han generado o por otros que acceden al foro en busca de respuestas.

Sin embargo, dada la naturaleza de Internet, muchas preguntas y respuestas similares se encuentran diseminadas en distintos foros de discusión, por lo que generalmente es necesario navegar por varios foros hasta dar con la solución correcta.

Nuestro proyecto tiene como objetivo principal favorecer el reuso de la información contenida en dichas conversaciones existentes en la Web.

De manera complementaria, nuestro proyecto planea aprovechar el trabajo de investigación realizado sobre foros de discusión públicos para proponer técnicas en pos de mejorar la enseñanza en las materias específicas del área de programación. Este último objetivo está proyectado cumplirse bajo la participación de los alumnos de nuestras materias en los foros de discusión de la Plataforma de Educación de la Universidad Nacional del Comahue (PEDCo). De esta manera los alumnos podrán mejorar su aprendizaje utilizando el conocimiento significativo producto de la información analizada

Una característica a destacar de nuestro proyecto es que esta relacionado con más de un área de investigación. Esta apertura nos permite trabajar en forma colaborativa e interdepartamental enriqueciendo las áreas involucradas. Por ejemplo, más allá de nuestra participación en el grupo de investigación de Ingeniería del Software, es muy interesante el intercambio realizado con los investigadores del Departamento de Teoría de la Computación, relacionados al estudio de algoritmos de análisis de lenguaje natural y sentiment analysis.

En cuanto a trabajos relacionados con propuestas de reuso de conocimiento en foros de discusión, debemos mencionar la de un sistema recomendador para

conocimiento colaborativo, desarrollado por Chen et al [2], que analiza automáticamente los mensajes de un foro de discusión de un curso de Inteligencia Artificial, en el objeto de proponer mensajes de contenido similar a aquellos de estudiantes de dictados anteriores del mismo curso. Respecto a los puntos en común con nuestro trabajo, esta propuesta evalúa la relevancia de los mensajes existentes en función de su similitud con el mensaje recientemente ingresado por un estudiante. El enfoque de Helic y Scerbakov [4] propone un método de clasificación de los mensajes de un foro de discusión de acuerdo a una jerarquía de temas. La principal diferencia entre nuestra perspectiva y las propuestas de Chen et al y Helic et al, es que mientras ambos fueron pensados para un dominio de aprendizaje colaborativo, nuestro recomendador apunta a un contexto más amplio, involucrando usuarios con distinto conocimiento previo del tema (background). Finalmente, y más importante, en los trabajos mencionados el foro utilizado es único, por lo tanto se puede asegurar que la información a analizar se encuentra en un formato estándar, contrariamente a nuestra propuesta la cual recolecta información de diferentes foros, por lo cual la heterogeneidad de formatos de la información a capturar es un desafío extra.

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 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).

Específicamente dentro del subproyecto de Reuso de Conocimiento en Foros de Discusión Técnicos, se destacan dos líneas de investigación relacionadas: La primera está enfocada en la captura, análisis y procesamiento de la información disponible en foros de discusión técnicos y un modelo de calidad relacionado que permita reutilizar el conocimiento disponible en Internet por parte de la comunidad de técnicos informáticos. La segunda línea está enfocada en la aplicación del conocimiento adquirido como parte de nuestra investigación, para hacer uso de la información de los foros de discusión de la Plataforma de Educación a Distancia de nuestra Universidad (PEDCO) y aprovecharla para proponer mejoras en la enseñanza de los alumnos de nuestra Facultad.

Resultados y objetivos

En [10] se ha presentado la visión de nuestra línea de investigación basada en el reuso como una actividad sistemática de ingeniería. Durante el año 2012, nuestro grupo GIISCo ha avanzado en el tratamiento del desarrollo de software basado en reuso desde los pilares de servicios, de productos, de datos y reuso de información, en colaboración con investigadores de ISISTAN, UNICEN; *reuso de conocimiento*, abordando el tema de información en foros técnicos.

Parte del trabajo realizado hasta el momento es 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 no restringidos disponibles en Internet. Dicho modelo está en proceso de ser publicado. Además, a partir de dicho modelo se ha comenzado a proponer un conjunto de métricas de calidad en cuya validación se trabajará a lo largo del proyecto.

De manera complementaria y paralela, se planea extender esta investigación al reuso de conocimiento en foros de discusión privados (de uso restringido a una comunidad en particular), específicamente en el contexto de la enseñanza de lenguajes de programación. En este sentido, uno de los objetivos de nuestro proyecto será proponer técnicas para mejorar la enseñanza en las materias del Departamento de Programación, a partir de la aplicación de técnicas de captura de información y generación de conocimiento desde los foros de discusión de la Plataforma de Educación a Distancia PEDCO, utilizada en las materias de dicho Departamento.

La información acumulada a partir de la interacción de los alumnos en los foros de discusión de la plataforma será analizada para generar conocimiento que pueda ser utilizado para mejorar el aprendizaje de los mismos alumnos en los siguientes cursos y de otros alumnos en el mismo curso. En este objetivo se combinan (en un nuevo dominio relacionado al campo de la gestión del conocimiento), experiencias obtenidas durante la ejecución del proyecto anterior, mediante la aplicación de técnicas cognitivas, conocidas como Modelos de Aprendizaje, que clasifican a las personas de acuerdo a

la manera que ellas perciben y procesan la información. Dichos modelos serán muy útiles para el objetivo de este proyecto, en la línea de investigación referida a la enseñanza de programación en la Facultad de Informática de nuestra Universidad.

Formación de recursos

humanos

El proyecto inicia una nueva línea de investigación respecto a la definición de un modelo de calidad y de gestión de conocimiento a partir de información contenida en foros de discusión técnicos. Para llevar a cabo estas metas, el proyecto está conformado por dos docentes del Departamento de Programación que han concluido en el año 2009 su Doctorado en Informática, y a dos docentes con dedicación simple, del Departamento de Ingeniería de Sistemas y de Programación que comienzan a formarse en investigación. También involucra a dos estudiantes que están desarrollando sus tesis de Licenciatura en Ciencias de la Computación. Además, se cuenta con 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. Finalmente se cuenta con la asesoría externa de una docente e investigadora de la Universidad de Castilla La Mancha, lo que permite asociar desarrollos y producciones entre ambas universidades. De esta manera el proyecto está conformado por un grupo interdisciplinario garantizando la pluralidad de puntos de vista y la cooperación entre las partes.

Referencias

- [1] ISO/IEC 25012:2008, Software product Quality Requirements and Evaluation (SQuaRE): Data quality model. 2008.
- [2] Weiqin Chen, Ricard Persen (2009), "A Recommender System for Collaborative Knowledge".
- [3] Coral Calero, Angélica Caro, Mario 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] Irfan Rafique, Philip Lew, Maisson Qanber 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] Richard Y. Wang, Diane 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 A. 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] Alejandra Cechich, Agustina Buccella, Andrés Flores, Gabriela Aranda, Nadina Martínez Carod, Juan Luzuriaga, Rodolfo Martínez, Marcelo Moyano, Rafaela Mazalu, Adriana Martin, Martín Garriga, Natalia Huenchuman. *Desarrollo Basado en Reuso, WICC 2012, XIV Workshop de Investigadores en Ciencias de la Computación*, Posadas, 2012