

## Hacia un framework de evaluación de calidad de información en foros de discusión técnicos

Gabriela N. Aranda, Nadina Martínez, Pamela Faraci, Alejandra Cechich

Grupo GIISCo, Facultad de Informática, Universidad Nacional del Comahue  
Buenos Aires 1400 (8300) Neuquén, Argentina  
{gabriela.aranda|nadina.martinez}@fi.uncoma.edu.ar

**Resumen** Es una práctica común entre los técnicos informáticos, que utilicen los foros de discusión de Internet para abrir una pregunta sobre un problema particular y que luego los usuarios del foro respondan proponiendo una o más soluciones a dicho problema. De esta manera, los foros de discusión técnicos se transforman en plataformas colaborativas útiles para plantear dudas y compartir soluciones que puedan ser reutilizadas en situaciones similares, ya sea por los mismos técnicos que participaron de la discusión u otros. Sin embargo, dada la naturaleza de Internet, muchas preguntas y respuestas similares se encuentran diseminadas en distintos foros de discusión. Ante la necesidad de encontrar la solución que más se ajuste a un problema particular, es necesario navegar por distintos foros hasta descubrir la más adecuada para tal situación. Con el objetivo de asistir en la búsqueda de la mejor solución a un problema dado, en este artículo se presenta un conjunto de características relevantes para evaluar la calidad de la información contenida en los foros de discusión, proponiendo además un modelo conceptual para la clasificación de dicha información.

### 1. Introducción

Los foros de discusión son espacios públicos en Internet donde las personas discuten acerca de algún tema en particular. De esta manera, las personas pueden leer los mensajes que han dejado otros y contribuir a la discusión dejando su propio mensaje. De acuerdo a la clasificación propuesta por Ellis et al [1], un foro de discusión es una herramienta de soporte al trabajo grupal (*groupware*) que corresponde a la categoría de *sistemas de mensajes* (basados en mensajes textuales) y *asíncrona* (lo que significa que los participantes no necesitan estar conectados al mismo tiempo para interactuar). En ambas clasificaciones, los foros se comportan de manera similar al correo electrónico; la diferencia es que un mensaje dejado en un foro de discusión puede ser visto por cualquier persona que ingrese al foro, por lo que no es posible controlar quien lo lee y quien no (esto no sucede al enviar un correo electrónico a un grupo cerrado de destinatarios).

Durante la última década el paradigma de la Web ha ido evolucionando desde páginas que proveían información de manera estática hacia la actual Web 2.0, basada en la interacción y colaboración, donde las comunidades virtuales

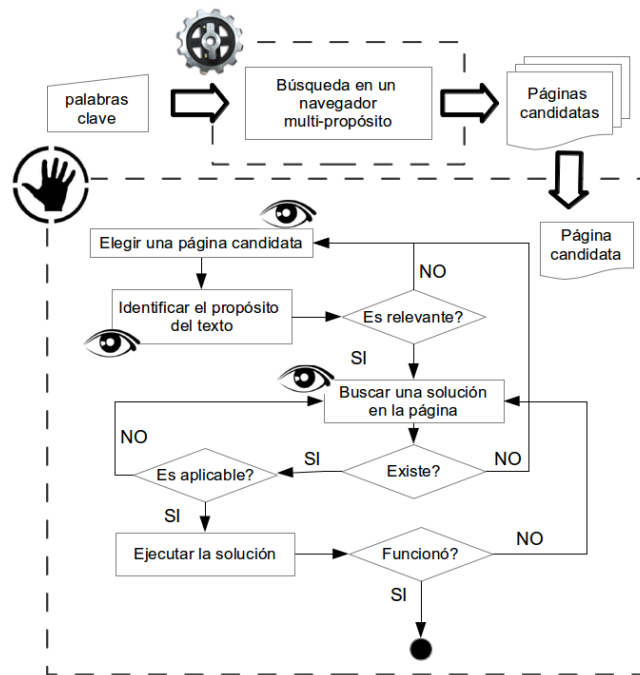
ganan protagonismo día a día. Este escenario ha dado lugar a la aparición de distintas plataformas para el intercambio de opiniones y recomendaciones entre usuarios de perfiles similares, como las wikis o los weblogs, y al crecimiento y consolidación de otras, como los foros de discusión, que ya existían como herramientas colaborativas en la Web [2]. Un ejemplo concreto es la creación de foros de discusión técnicos, donde una persona inicia un tema de debate con una pregunta y los miembros de la comunidad responden proponiéndole soluciones de acuerdo a sus experiencias previas. Si bien las soluciones propuestas atienden a la pregunta particular del usuario que abrió la discusión, el conjunto de mensajes queda disponible al público en general, y la o las soluciones propuestas pueden ser reutilizadas al surgir problemas similares.

Gracias a que esta información se mantiene disponible en la Web, es común entre los informáticos que ante un problema o una duda sobre el uso de una sentencia en un lenguaje de programación, o de la configuración de alguna característica de un sistema operativo, etc., utilicen un navegador multipropósito (por ejemplo, *Google*) para encontrar una respuesta que les ayude a resolver su problema. El proceso puede resumirse aproximadamente de la siguiente manera:

1. El técnico realiza una búsqueda en el navegador, a partir de una serie de palabras clave, obteniendo por lo general un gran conjunto de enlaces a páginas Web de distinto formato (manuales, páginas de instituciones técnicas, blogs personales, foros de discusión, etc.), ordenados según los algoritmos de ponderación o posicionamiento propios del navegador.
2. Luego, el individuo realiza las siguientes tareas de manera manual:
  - a) Elige y abre las páginas cuyos títulos se parecen más a su pregunta (generalmente aquellos donde aparecen todas o la mayoría de las palabras clave y en un orden similar al esperado);
  - b) Va leyendo cada página, buscando entre el texto la pregunta exacta (para verificar si se parece a su duda inicial) y analiza la solución o soluciones propuestas;
  - c) Evalúa si las soluciones propuestas son relevantes para su problema y si la fuente es confiable. Luego elige alguna para probar.
3. Repite los pasos 2(b) y 2(c) hasta conseguir alguna solución que le satisfaga o repite la búsqueda (paso 1), ajustando el conjunto de palabras clave y volviendo a comenzar.

En la Figura 1 se grafica este proceso, identificando la etapa automática (búsqueda de páginas candidatas), de la etapa manual (búsqueda de información y prueba de posibles soluciones). También se identifica con un icono especial (que representa un ojo) aquellas tareas donde el usuario decide por cuál página candidata comenzar o seguir, y detecta visualmente los fragmentos correspondientes al objetivo de la página (pregunta principal) y las posibles soluciones.

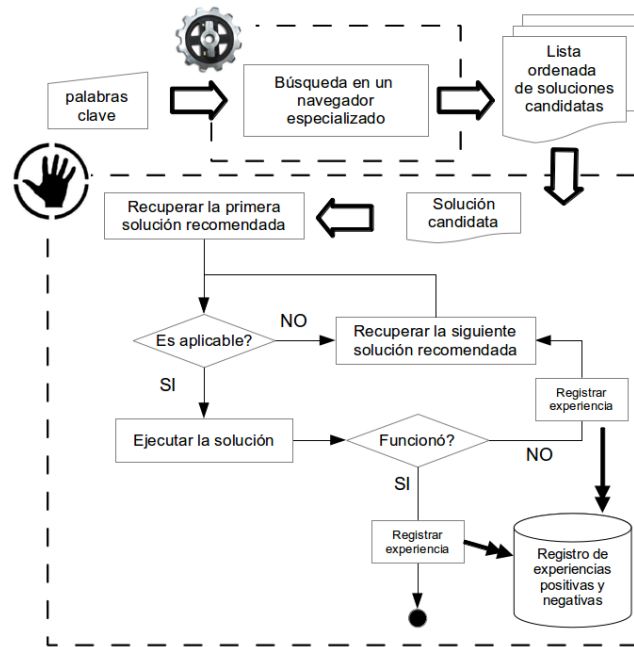
Considerando que en todos los casos se parte de información que está disponible de manera pública en la Web, el objetivo de nuestro trabajo es



**Figura 1.** Proceso de búsqueda de una solución técnica en base a un navegador multipropósito

automatizar la búsqueda de soluciones a preguntas técnicas y presentarlas de manera ponderada cercana a las expectativas y necesidades del técnico que hace la pregunta. Entre los distintos tipos de páginas Web donde es posible obtener soluciones a problemas técnicos, este trabajo se ha enfocado específicamente en reutilizar el conocimiento resguardado en foros de discusión. Esto es debido a que, al ser un entorno colaborativo, se puede contar con distintos puntos de vista del problema así como comentarios de otros usuarios del foro, que dan información adicional e incluso a veces informan si las soluciones propuestas han sido exitosas o no y en qué contexto. En base a esta información, el objetivo al que apunta nuestro trabajo es construir un navegador especializado en problemas técnicos que, a partir de un conjunto de palabras clave que representan la búsqueda inicial, retorne una lista ordenada de soluciones candidatas. El orden otorgado a las soluciones candidatas será determinado por medio de un proceso de evaluación de calidad de la información. Además, está previsto mantener una base de datos de las experiencias de los usuarios (después de seleccionar y aplicar las soluciones candidatas), como un mecanismo de mejora constante a partir de la retroalimentación realizada por los mismos usuarios. El proceso, basado en el uso del navegador especializado mencionado, se muestra en la Figura 2, donde

puede observarse que la tarea manual del técnico se reduce considerablemente, al contar con un análisis previo de los hilos de foros de discusión más relevantes y al análisis de la información contenida en ellos de manera automática.



**Figura 2.** Proceso de búsqueda de una solución técnica en base a un navegador especializado

Con dicho objetivo en mente, en la Sección 2 se describe una serie de modelos de calidad relacionados al producto software, y especialmente aquellos enfocados en datos. Posteriormente, en la Sección 3 se describe primero un modelo conceptual de la información contenida en un foro de discusión técnica, y a continuación se presenta un framework que considera un conjunto de características con las que se propone evaluar la calidad de la información en dicho contexto. Por último se presentan las conclusiones y líneas de trabajo futuro.

## 2. Modelos de calidad

Un modelo de calidad define una serie de criterios para satisfacer las necesidades de los desarrolladores y usuarios finales. Dichos modelos son

utilizados para construir mejores productos y asegurar su calidad. A tal efecto se han propuesto varios modelos de calidad enfocados en el producto y proceso software. Dichos modelos descomponen la calidad jerárquicamente en una serie de características y subcaracterísticas que pueden usarse como lista de comprobación para los aspectos relacionados con la calidad. Un ejemplo ampliamente adoptado de modelo de calidad del producto software es el propuesto por la norma ISO 9126 [3] que presenta un modelo para *calidad interna y externa del producto software* basado en seis características divididas en un conjunto de subcaracterísticas, así como un modelo para *calidad en uso* basado en cuatro características. Sin embargo, en los últimos años, este estándar ha sido modificado por la norma ISO 25010 [4] que mantiene la misma estructura en cuanto a la definición del modelo y las características de su predecesor, aunque difiere en el agregado de algunas características y subcaracterísticas.

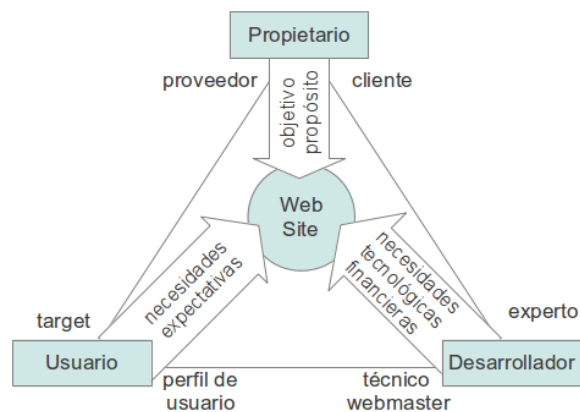
Como ejemplo de los modelos de calidad de datos, se destaca como pionero el modelo propuesto por Wang y Strong [5] que captura los aspectos de calidad de datos enfocado en los consumidores, es decir quienes utilizan la información. Dicho marco está formado por 20 dimensiones, las cuales están conformadas en uno o más factores de atributos de calidad.

Con un objetivo similar, el estándar ISO 25012 [6] define un modelo general para la calidad de datos mantenidos por un sistema de computadoras de manera estructurada. Este modelo categoriza los atributos de calidad en quince características, considerando dos puntos de vista (inherente y dependiente del sistema). El modelo propuesto en el estándar ISO 25012 ha sido recientemente analizado por Rafique et al [7] para considerar calidad de datos particularmente en aplicaciones Web, proponiendo una extensión que agrega dos características provenientes del modelo de Strong y Wang [5]. En las características de ambos modelos se profundizará en la Sección 3.3.

### 3. Definiendo calidad en foros de discusión técnicos

Antes de definir el modelo de calidad a utilizar para foros de discusión, es necesario tener en claro la funcionalidad y los distintos tipos de actores que participan en dichos foros. Para ello, se ha tomado como referente la clasificación de las necesidades de los actores relacionados con un sitio Web, según la propuesta de Mich et al [8]. En esta propuesta, se identifican tres tipos de actores de un sitio Web: *propietario*, *desarrollador* y *usuario*. El propietario es quien pide al desarrollador, en su rol de experto, que implemente el sitio Web, y el usuario es quien lo utilizará (identificado como blanco o *target* del sitio, que es la necesidad del propietario), como se muestra en la Figura 3.

Tomando como base dicho modelo, es importante destacar los distintos tipos de usuario que existen en un foro de discusión, cada uno con sus propias necesidades, y que no son los mismos que en otros tipos de sitio Web. Los tipos de usuario en un foro de discusión se pueden clasificar como [9]:

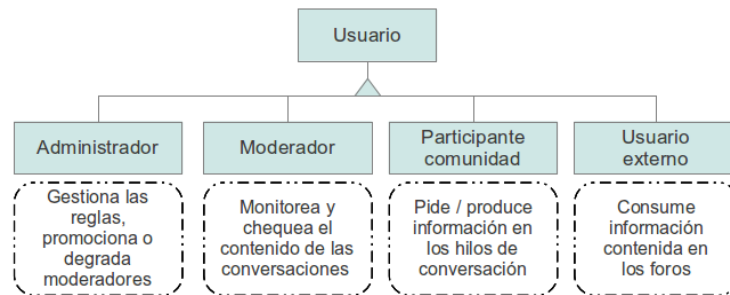


**Figura 3.** Necesidades de los actores relacionados con un sitio Web [8]

- *Usuario administrador*: Es el tipo de usuario con el nivel de permisos más alto. Puede manipular ciertos aspectos técnicos como controlar las reglas del foro, decidir qué participantes pueden (o no) ocupar el lugar de moderadores, manipular la creación de secciones nuevas, cambiar aspectos visuales de la página, etc.
- *Usuario moderador*: Es un usuario con privilegios que verifica que se cumplan las reglas de uso del foro, permite o no la publicación de ciertos mensajes y realiza monitoreos en busca de spams. Ocasionalmente responde a preguntas generales sobre el foro y atiende reclamos de otros usuarios.
- *Usuario participante*: Este tipo de usuario tiene permisos para escribir en el foro de discusión. Puede ser autor de una pregunta, o también puede responder preguntas realizadas por otros participantes.
- *Usuario externo*: Es el tipo de usuario con menor nivel de permisos. La actividad permitida es leer el contenido de los foros públicos y tener conocimiento de dicha información, es decir que actúa como un consumidor de la información contenida en el foro.

Esta clasificación y las actividades más significativas de cada rol, se resumen en la Figura 4.

Con el objetivo de reutilizar el conocimiento contenido en las conversaciones entre usuarios participantes de una comunidad virtual como es un foro de discusión sobre temas técnicos, la primera instancia es definir un modelo de calidad para la información contenida en dicho tipo de foros. En este sentido, es pertinente que el modelo se plantee considerando sólo el punto de vista del *usuario externo*, es decir, enfocándose en la calidad desde el punto de vista de la información y no de la funcionalidad que el sitio pueda o necesite proveer para el resto de los tipos de usuario.



**Figura 4.** Clasificación de los distintos tipos de usuarios en un Foro de Discusión Técnico

### 3.1. Revisión de foros de discusión técnicos

A efectos de definir un marco teórico para el estudio de la información contenida en los foros de discusión, se ha realizado una revisión formal de hilos de discusión reales. Dado que la cantidad de foros disponibles en la Web es muy amplia, en esta primera etapa se tomaron como base del estudio 6 foros (muestra que será extendida al continuar profundizando en este trabajo). Con el fin de determinar si existen diferencias significativas en foros para las comunidades anglosajona e hispanohablante, se seleccionaron 3 foros en español y 3 en inglés. Luego, dentro de cada foro elegido, se seleccionaron 6 hilos de conversación que cumplieran las siguientes características predefinidas:

1. *El hilo presentaba al menos dos mensajes:* un mensaje con pregunta principal y otro mensaje relacionado a éste. Este requisito excluye los hilos de conversación donde nadie intentó responder a la inquietud principal (es decir, no existen soluciones planteadas).
2. *Se priorizaron los hilos formados por la mayor cantidad de mensajes posibles:* Este requisito parte de la premisa de que cuanto más mensajes hay en un hilo es más fácil aprender sobre la dinámica de las conversaciones, ya que, además del mensaje principal, aparecen mensajes que proponen soluciones, requieren más información, etc.

En la Tabla 1 se presenta un resumen del estudio realizado sobre un total de 36 hilos de discusión, en la que se enumeran algunas características que suelen diferenciarse entre distintos foros.

### 3.2. Modelo conceptual de foros de discusión técnicos

En base al estudio explicado en la sección anterior, en la Figura 5 se describe un modelo conceptual del formato de los foros de discusión técnicos en la Web. Dicho diagrama considera la información disponible en un foro de discusión

**Tabla 1.** Características de los foros de discusión tomados como base del estudio

Foro	Idioma hilos	Objetivo del foro	Usuario		Muestra cant. visitas	Mensajes
			Reputación	Otra info		
1	Español	Lenguaje Java	Principiante, Miembro, Miembro experimentado, Moderador, Experto de Java, Gurú de Java	Likes, dislikes, thanks: recibidos de otros usuarios	Si	Texto, código, enlaces web
2	Español	Lenguaje Java	---	---	No	Texto, código, enlaces web
3	Español	Lenguaje Java	Guest, Member, Senior Member, Moderator, Hostage + Reputation Power (entero)	---	Si	Texto, código, enlaces web
4	Inglés	Lenguaje Java	Guru, Expert, Pro, Journeyer, Explorer, Newbie	Reciben puntos (reward points) de otros usuarios	Si	Texto, código, enlaces web
5	Inglés	Programación en general	Reputation score (valor numérico)	Gold, Silver and Bronze badges (recibidos de otros usuarios)	Si	Texto, código, enlaces web
6	Inglés	Programación en general	Reputation power (valor numérico)	---	Si	Texto, código, enlaces web

**Foros analizados:** (1) <http://forodejava.com>, (2) <http://www.javahispano.org/java-ee/>, (3) <http://www.java-forums.org>, (4) <https://forums.oracle.com>, (5) <http://stackoverflow.com>, (6) <http://forums.devshed.com>

desde el punto de vista del usuario externo, identificándose las entidades más importantes y sus atributos.

En general un foro de discusión técnico (*Foro*) contiene varios hilos de discusión (*Hilo*). Cada *hilo* se genera cuando un usuario participante de la comunidad (*Usuario*) crea un nuevo tema de debate que surge generalmente a partir de una inquietud personal. Cada *hilo* se identifica por un *título*, que está generalmente relacionado con la pregunta principal, realizada por el usuario que inicia el hilo o tema de debate (esto suele ser un requisito mencionado en las reglas del foro y controlado por los moderadores de los foros). Si bien es cierto que no está presente en todos los foros, suele encontrarse disponible la información relacionada con la *cantidad de visitas* realizadas al *hilo*, es decir la cantidad de veces que la página fue accedida o visitada por un usuario participante o externo.

La estructura del *hilo* está formada por una serie de aportes. Cada aporte, llamado post o mensaje (*Mensaje*), es realizado por un *usuario* participante en una *fecha* en particular. A fin de poder analizar el contenido de cada *mensaje*, se ha considerado que un *mensaje* consta de uno o más *fragmentos*, donde cada fragmento puede tratarse de lenguaje natural (*texto*), código que puede ser ejecutado en un sistema operativo o compilado en un lenguaje de programación (*código*), o bien un enlace a una página Web donde una pregunta similar con posibles soluciones afines han sido propuestas (*enlace web*).

Adicionalmente, en base al análisis de los fragmentos que componen el *mensaje*, se definieron cuatro *tipos* de mensajes principales:

- Mensaje de *pregunta principal* (el primer mensaje que se creó, dando inicio al *hilo*),
- Mensaje que *propone una solución*,
- Mensajes que *rechazan* o *agradecen* una solución propuesta anteriormente.

Sin embargo, a partir de la observación de hilos reales que se explicó anteriormente, se incluyeron otros *tipos* de adicionales:



- Mensaje que *pide una aclaración* sobre un mensaje anterior,
- Mensaje que *añade datos* (en respuesta a un pedido del tipo de mensaje anterior).
- Mensajes *irrelevantes*, es decir aquellos que no presentan información importante para el debate.

Respecto a los usuarios, se considera que para cada *mensaje* se puede saber el *usuario* que lo escribió, del cual se conoce su nickname o nombre dentro de la comunidad (*nombre*). Si bien no es un dato presente en todos los foros de discusión, habitualmente se cuenta con más información sobre el usuario como su *rol*, *reputación* (generalmente asignada por los usuarios registrados) y la *cantidad de mensajes* que ha emitido en la historia de su participación en la comunidad.

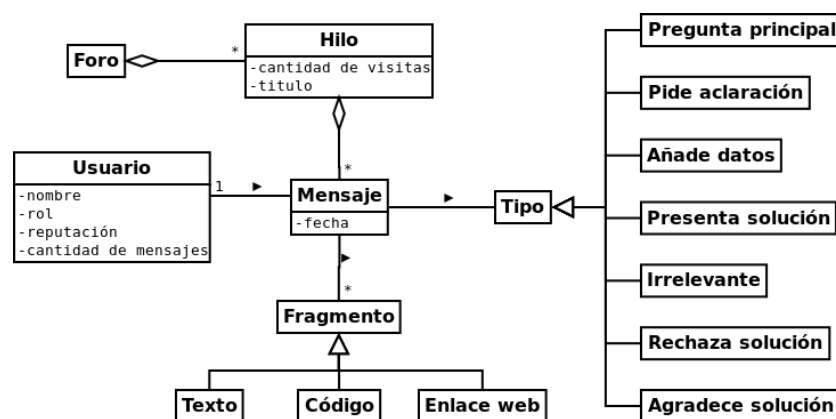
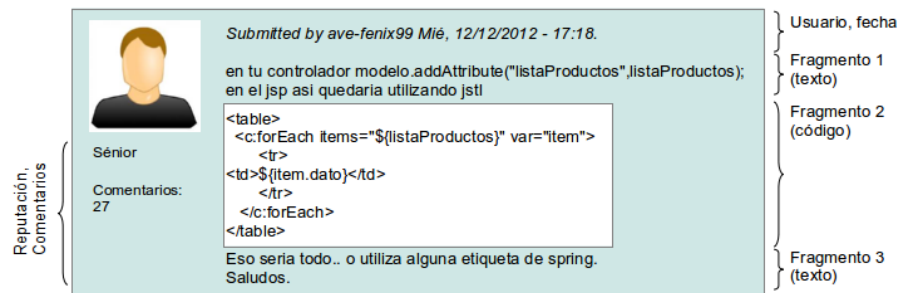


Figura 5. Diagrama de la información contenida en un foro de discusión técnico

En la Figura 6 se muestra un ejemplo de un mensaje que propone una solución. En la cabecera del mismo se observa que se cuenta con el nombre del usuario, la fecha y hora del envío. A la izquierda se encuentra información de su rol, reputación y la cantidad de comentarios realizados en el foro. Luego, el cuerpo del mensaje está formado por tres fragmentos, donde el primero y el último son de tipo texto y el segundo es de tipo código.

### 3.3. Modelo de calidad para foros de discusión técnicos

Dado el objetivo principal de nuestro trabajo, que es obtener una lista de soluciones candidatas para resolver un problema técnico planteado por un usuario, y que dichas soluciones han de recopilarse a partir de uno o más hilos de conversación en varios foros de discusión técnicos, a continuación se analiza



**Figura 6.** Ejemplo de mensaje que propone una solución en más de un fragmento

como los atributos presentes en dichos hilos de discusión pueden utilizarse para aproximar el nivel de calidad de las soluciones propuestas en los mismos.

Dado que una solución es ante todo una pieza de información almacenada en un sitio Web en particular, se ha decidido tomar como base las características definidas en el estándar ISO/IEC 25012:2008 [6] para la calidad de datos, incluyendo las características *valor añadido* y *adecuación representacional*, de acuerdo a la propuesta de Rafique et al [7] para calidad de datos en sitios Web.

Como se explicó en la Sección 2, el estándar ISO/IEC 25012:2008 define dos puntos de vista para clasificar las dimensiones de calidad de los datos. Por un lado, el punto de vista *inherente* se refiere a evaluar la calidad en cuanto al grado de satisfacción de las características del dato y concerniente a la satisfacción de las necesidades cuando el dato es usado bajo condiciones que han sido especificadas, mientras que el punto de vista *dependiente del sistema* enfoca la evaluación de la calidad teniendo en cuenta el dominio tecnológico en el que se utilizan los datos (dispositivos de hardware, sistemas informáticos y otros programas). A continuación se presenta la definición de las características de calidad de acuerdo a si se corresponden con uno de los puntos de vista o con ambos:

#### 1. Características inherentes

- *Exactitud* (Accuracy): el grado en el cual el dato tiene atributos que representan correctamente el valor real del atributo en cuestión de un concepto o evento, en un contexto de uso específico.
- *Complejitud* (Completeness): el grado en el cual un dato (asociado con una entidad) tiene valores para todos los atributos e instancias de entidades relacionadas esperados, en un contexto de uso específico.
- *Consistencia* (Consistency): el grado en el cual el dato tiene atributos que son libres de contradicción y son coherentes con otros datos en un contexto específico de uso.
- *Credibilidad* (Credibility): el grado en el cual el dato tiene atributos que son considerados reales y creíbles por usuarios en un contexto de uso específico.
- *Actualidad* (Currentness): el grado en el cual el dato tiene atributos que son de una antigüedad correcta en un contexto de uso específico.

2. Características inherentes y dependientes del sistema

- *Accesibilidad* (Accessibility): el grado en el cual el dato puede ser accedido en un contexto de uso específico, en particular por personas que necesitan un soporte tecnológico o una configuración especial, debido a alguna discapacidad.
- *Conformidad* (Compliance): el grado en el cual el dato tiene atributos que se ajustan a normas, convenciones o regulaciones vigentes y reglas similares, relacionadas con la calidad de datos en un contexto de uso específico.
- *Confidencialidad* (Confidentiality): el grado en el cual el dato tiene atributos que aseguran que éste es sólo accesible e interpretado por usuarios autorizados, en un contexto de uso específico.
- *Eficiencia* (Efficiency): el grado en el cual el dato tiene atributos que pueden ser procesados y que proporcionan los niveles esperados de funcionamiento, usando cantidades apropiadas y tipos apropiados de recursos en un contexto de uso específico.
- *Precisión* (Precision): el grado en el cual el dato tiene atributos que son exactos o que proporcionan una correcta diferenciación en un contexto de uso específico.
- *Trazabilidad* (Traceability): el grado en el cual el dato tiene atributos que proporcionan un registro susceptible de auditoría, tanto para el acceso al dato como de cualquier cambio hecho al dato, en un contexto de uso específico.
- *Entendibilidad* (Understandability): el grado en el cual el dato tiene atributos que le permiten ser leído e interpretado por usuarios, y está expresado de manera apropiada respecto al lenguaje, símbolos y unidades utilizadas, en un contexto de uso específico.

3. Características dependientes del sistema

- *Disponibilidad* (Availability): el grado en el cual el dato tiene atributos que permiten que este sea recuperado por usuarios y/o aplicaciones autorizadas en un contexto de uso específico.
- *Portabilidad* (Portability): el grado en el cual el dato tiene atributos que le permiten ser instalado, substituido o movido de un sistema a otro, conservando la calidad existente en un contexto de uso específico.
- *Recuperabilidad* (Recoverability): el grado en el cual el dato tiene atributos que le permiten mantener y conservar un nivel especificado de operaciones y calidad, aún en caso de falla, en un contexto de uso específico.

4. Características agregadas de acuerdo a [7]

- *Valor Añadido* (Value Added): el grado en el cual el dato o la información son beneficiosos y proveen ventajas como producto de su utilización.
- *Adecuación Representacional* (Representational Adequacy): el grado en el cual el dato o la información se presenta de manera concisa, flexible y organizada, considerando los objetivos del usuario, para ayudarlo a alcanzar las metas que ha especificado.

Ambos puntos de vista (inherentes y dependiente del sistema) deben ser analizados para evaluar la calidad de las soluciones propuestas en un foro de discusión. Basada en estos conceptos, la Tabla 2 muestra la matriz de interrelación entre los atributos que se presentaron en la Sección 3.2 como parte del modelo conceptual para foros de discusión técnicos (filas) y las características de calidad deseadas (columnas) en las soluciones candidatas. La confección de dicha matriz se ha realizado a partir de la observación de distintos foros de discusión técnicos (como fue explicado en la Sección 3.1) y de la definición de las características de calidad deseadas, analizando la posibilidad de interrelación en el cruce de ambas dimensiones.

**Tabla 2.** Relación entre las entidades del modelo conceptual y las características de calidad deseadas en la solución

		Dimensiones de Calidad	Exactitud	Complejidad	Consistencia	Credibilidad	Actualidad	Accesibilidad	Conformidad	Confidencialidad	Eficiencia	Precisión	Trazabilidad	Entendibilidad	Disponibilidad	Portabilidad	Recuperabilidad	Valor Agregado	Adec. Represent.	
			Atributos																	
HILO	título		●	●																
	cantidad visitas		●	●							●									
USUARIO	nombre																			
	cantidad mensajes		●	●																
	reputación		●	●		●					●	●		●						
MENSAJE	fragmento	texto	●	●										●					●	
		código	●	●					●			●		●		●		●	●	
		enlace web	●	●										●		●			●	
	fecha						●													
	pregunta principal		●	●																
	añade datos		●	●							●	●	●	●	●				●	●
	irrelevante																			
	tipo	agradece solución		●	●		●					●	●	●	●		●		●	●
		presenta solución		●	●		●					●	●	●	●	●	●		●	●
		pide aclaración																		
rechaza solución																			●	

A continuación se detalla el análisis realizado para los atributos de la Tabla 2, y específicamente para los cruces que se encuentran resaltados.

Por ejemplo, en el caso del atributo *título* (representado en la primera fila de la tabla), se ha analizado su relación con la calidad de la o las soluciones presentes en el hilo. Para ello se ha considerado que aunque el título contenga todas las palabras clave esto no significa que el hilo esté directamente relacionado con la pregunta original del usuario. Por ese motivo, se ha marcado como válida la relación con la categoría *exactitud*, es decir que se presume su relación,

pero que ésta debe ser definida teniendo en cuenta la combinación con otros atributos. La misma observación se ajusta a la característica de *completitud*. Como resultado de este análisis, no puede asegurarse otro tipo de relación con las demás características a partir del *título* del hilo en cuestión.

Respecto a los atributos *mensaje con pregunta principal* y *mensaje que agrega información*, en estos casos la información que aportan ayudaría a definir mejor el grado de *exactitud* de la solución respecto al problema original del usuario, por lo cual la relación se ha marcado como posible. Por el contrario, el atributo *mensaje que agradece solución*, estaría relacionado con las características *exactitud*, *completitud* y *credibilidad*, aunque esto dependería de la información que puedan aportar otros atributos, como por ejemplo *mensaje de la pregunta principal*.

Respecto a las características *valor añadido* y *portabilidad*, estarían relacionadas con los *mensajes que agradecen o rechazan una solución*. La relación surgiría a partir de que dichos mensajes suelen agregar información referente al contexto donde la solución propuesta fue o no exitosa, por lo cual podría ser útil para decidir recomendar dicha solución en contextos similares.

En cuanto al atributo *fecha*, que corresponde al momento en que se publicó el mensaje, se presume que estaría relacionado con la característica de *actualidad*, si bien según la pregunta o el contexto en el cual ésta se realiza, podría no ser definitiva. Por ejemplo, una pregunta sobre el uso de una sentencia en un lenguaje de programación puede tener un grado de actualidad aceptable, aún cuando la solución haya sido propuesta en el pasado, si el cambio entre versiones no ha ocasionado modificación en la sintaxis de dicha sentencia.

Si bien la naturaleza de los foros de discusión hace que puedan coexistir conversaciones no confiables, el estudio está enfocado con la premisa que se utilizan foros técnicos de discusión cuya reputación está asegurada y donde existe un moderador que elimina mensajes fuera del ámbito técnico, tanto como mensajes erróneos o que no cumplen con las reglas del foro, por este motivo se considera que la información es confiable. Bajo esta perspectiva, la característica *credibilidad* estaría presente en todo momento. Sin embargo, una buena reputación de un usuario que propone una solución o la existencia de un *mensaje que agradece* la propuesta, podría afectar positivamente el grado de *credibilidad* de dicha solución.

#### 4. Trabajos relacionados

En cuanto a propuestas de reuso de conocimiento en foros de discusión, se encuentra la de un sistema recomendador para conocimiento colaborativo, desarrollado por Chen et al [10], que analiza automáticamente los mensajes de un foro de discusión de un curso de Inteligencia Artificial, para proponer mensajes con contenido similar 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. Otra propuesta similar es la de Helic

y Scerbakov [11], que propone un método de clasificación de los mensajes de un foro de discusión de acuerdo a una jerarquía de temas.

En cuanto a las diferencias entre nuestro enfoque y las propuestas anteriores, es que ambas están pensadas para un dominio de aprendizaje colaborativo, mientras que nuestro recomendador apunta a un contexto más amplio, involucrando usuarios con distinto conocimiento previo del tema (background). Finalmente, y más importante, en dichos trabajos el foro utilizado es único, por lo tanto se puede asegurar que la información a analizar se encuentra en un formato estandar, mientras que nuestra propuesta apunta a recolectar información de distintos foros, por lo tanto la heterogeneidad de formatos de la información a capturar es un desafío extra.

## 5. Conclusiones y trabajo futuro

En este trabajo se ha presentado un modelo conceptual para clasificar la información contenida en foros de discusión técnicos, identificando las entidades que forman parte del mismo, bajo el perfil del usuario que utiliza o consume la información, en su rol de usuario externo. A partir de dicha clasificación (y de la observación de varios hilos en foros de discusión técnicos), se ha propuesto una matriz que refleja las relaciones existentes entre las entidades del modelo conceptual y las dimensiones del estándar ISO/IEC 25012:2008 (extendido de acuerdo a la propuesta de Rafique et al [7]), como un primer paso para la definición de evaluación de la calidad de las soluciones presentes en foros de discusión técnicos. En general, al confeccionar esta matriz se analizaron las interrelaciones entre los atributos de información disponibles y las características de calidad de la solución deseada de manera aislada, por lo que probablemente estas asunciones se verían afectadas al profundizar el análisis de la combinación e interrelación de los atributos y las características entre sí. Por ejemplo, si se está analizando un hilo de discusión en particular, donde todas las palabras clave se encuentran en el título y en algunos mensajes de respuesta, los usuarios que proponen soluciones tienen una buena reputación, si alguien agradece la solución, y si además la cantidad de visitas al sitio es alta; en esta conjunción de casos hay una probabilidad más alta de que la solución propuesta sea exitosa. El análisis de la combinación de estos factores introduciría mejoras substanciales en la estimación de la calidad de las soluciones, por lo que se plantea como un próximo paso para la extensión del modelo de calidad propuesto. Además, en la actualidad, se está trabajando en el diseño de encuestas a usuarios de foros de discusión técnicos, que a partir de su aplicación permita aproximar el grado de interrelación entre los atributos de la información contenida en foros de discusión técnicos y la calidad de las soluciones propuestas en ellos. En base a dichos estudios, se planea establecer una serie de métricas e indicadores de calidad que sirvan para la detección automática de soluciones a problemas técnicos.

## Agradecimientos

Este trabajo está parcialmente soportado por el subproyecto “Reuso de conocimiento en foros de discusión técnicos”, correspondiente al Programa de Investigación F001 “Desarrollo orientado a reuso”, de la Universidad Nacional del Comahue (Neuquén, Argentina), y por el Proyecto PICT-2012-0045 “Mecanismos de soporte para grids híbridos orientados a servicios y técnicas de desarrollo de aplicaciones”.

## Referencias

1. C. A. Ellis, S. J. Gibbs, and G. L. Rein, “Groupware: Some issues and experiences,” *Communications of ACM*, vol. 34, no. 1, pp. 38–58, 1991.
2. J. Dorn, *Social Software (and Web 2.0)*, pp. 305–311. PA: Information Science Reference, 2010.
3. “ISO 9126-1:2001 Information technology - Software product quality, part 1: Quality model,” 2001.
4. “ISO/IEC 25010:2010, Systems and software engineering - Systems and software quality requirements and evaluation (SQuaRE): System and software quality models,” 2010.
5. R. Y. Wang and D. M. Strong, “Beyond accuracy: What data quality means to data consumers,” *Journal of Management Information Systems*, vol. 12, pp. 5–33, Mar. 1996.
6. “ISO/IEC 25012:2008, Software product quality requirements and evaluation (SQuaRE): Data quality model,” 2008.
7. I. Rafique, P. Lew, M. Q. Abbasi, and Z. Li, “Information quality evaluation framework: Extending ISO 25012 data quality model,” *International Journal of Computer and Information Sciences*, vol. 6, no. 1, 2012.
8. L. Mich, M. Franch, and G. Cilione, “The 2qcv3q quality model for the analysis of web site requirements,” *Journal of Web Engineering*, vol. 2, pp. 105–127, Sept. 2003.
9. G. Roquet Garcia, “Los foros de discusion en educacion,” *Siglo XXI: Perspectiva de la Educacion desde Am rica Latina*, no. 4, pp. 69–78, 1998.
10. W. Chen and R. Persen, “A recommender system for collaborative knowledge,” in *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, (Amsterdam, The Netherlands, The Netherlands), pp. 309–316, IOS Press, 2009.
11. D. Helic and N. Scerbakov, “Reusing discussion forums as learning resources in wbt systems,” in *Proceedings of the IASTED International Conference Computers and Advanced Technology in Education*, (Rhodes, Greece), pp. 223 – 228, 2003.