

Un Método Sistemático para Diseñar Instrumentos de Evaluación Aplicado a Calidad de Contenidos Web

Guillermo Covella, Juan Oliveto, Luis Olsina

GIDIS_Web, Facultad de Ingeniería, UNLPam.
Calle 9 esq. 110, General Pico, La Pampa, Argentina.
guillermo.covella@gmail.com, [olivetoj, olsinal]@ing.unlpam.edu.ar

Abstract: La importancia que ha ido cobrando el estudio de la calidad de la información en entornos digitales, se debe principalmente al incremento del volumen de recursos a los que el usuario final tiene acceso. Sin embargo, la Web como medio de comunicación ha tenido un lento crecimiento en la evaluación de la calidad de los contenidos que distribuye y presenta. La facilidad y libertad con que pueden publicarse contenidos Web hacen necesaria la adopción de una serie de criterios que ayuden a filtrar los contenidos y a discernir la adecuación, credibilidad, autoridad, y en definitiva, la calidad de las contenidos que se publican. En este trabajo presentaremos un método que permite la construcción de cuestionarios como un instrumento de evaluación de calidad de información, aplicado a evaluar contenidos de aplicaciones Web 1.0, 2.0 y siguientes eras. Se presenta también aspectos de validación de un cuestionario.

Palabras clave: Calidad de Contenido, Cuestionario, ISO25010, Evaluación, Web 2.0.

1 Introducción

A través de la evolución de las aplicaciones Web hemos visto como se han transformado en aplicaciones cada vez más complejas, cubriendo el más amplio espectro de funcionalidades de las aplicaciones software junto a la edición, distribución y presentación de contenidos. Sin embargo, desde el punto de vista de la evaluación de la calidad, esas aplicaciones Web que desde el principio de su existencia (hace ya veintiún años) eran aplicaciones estáticas y prácticamente orientadas a presentar contenido, la calidad de los contenidos o de la información que ofrecían era una cuestión relevante, pero no tenida en cuenta en comparación con aspectos de calidad de interfaces o usabilidad.

Por entonces, la utilidad de la existencia de la Web se justificaba en la posibilidad de compartir información en un formato visual más rico, independientemente de la plataforma en que se generara y visualizara. Es decir, dos décadas atrás, el valor agregado de la Web estaba en la posibilidad de compartir más y mejor información, sin necesidad de homogeneizar la tecnología subyacente. Luego, en lo que podría llamarse la era Web 1.0, se puso énfasis en el desarrollo de aplicaciones

transaccionales que permitieran el reemplazo de aplicaciones tradicionales, aprovechando la plataforma distribuida de Internet. En esta etapa, la calidad de contenido se volvió relevante también al interior de las organizaciones, con el crecimiento de las Intranets y Extranets corporativas, motorizadas por estrategias y dominios de aplicación como e-commerce, e-banking, e-procurement, e-learning, etc. De manera que a la potencia de la tecnología necesaria para soportar las aplicaciones se añadía la necesidad de que el contenido publicado fuera confiable, preciso, actual y originado en fuentes seguras para un desenvolvimiento exitoso del negocio. También aspectos relevantes relacionados como accesibilidad y usabilidad fueron cuestionados y luego considerados sistemáticamente por Bevan [1, 2], Nielsen *et al.* [12, 13], entre otros.

En nuestros días, la dinámica de la Web 2.0, impulsada por el éxito de las redes sociales, actualiza la importancia que tiene la calidad de los contenidos [17], básicamente porque los usuarios, anteriormente sujetos pasivos, se han transformado en editores, moderadores y consumidores del contenido que se publica en la Web. Ahora, como la tecnología ya no es tanto una limitación no es descabellado pensar que, en gran medida, el éxito de un sitio Web moderno depende, inicialmente, de una buena idea y se sostiene luego en la calidad de su contenido.

Por lo tanto, vemos de un modo sintético como la calidad de contenidos atraviesa la historia de las aplicaciones Web y constituye uno de sus pilares. Su importancia se realiza actualmente en la era Web 2.0, y nuestra preocupación es cómo medirla, cómo evaluarla y, eventualmente, cómo mejorarla. En ese sentido discutiremos más adelante, en la sección 3, algunos enfoques y propuestas que nos permitan avanzar de manera sistemática sobre esos aspectos de medición, evaluación y mejora, teniendo en cuenta trabajos anteriores [16, 17].

En particular, el aporte de este trabajo es un método sistemático para la construcción de un cuestionario que permita la evaluación de calidad de contenido de sitios Web por parte de usuarios finales. El procedimiento para alcanzar este objetivo considera, sin embargo, la participación de usuarios expertos en la fase inicial. A partir de las respuestas de los expertos se selecciona un número reducido de preguntas, para ser incluidas luego en el cuestionario para usuarios finales.

Este trabajo está organizado de la siguiente manera: a continuación se discuten los trabajos relacionados para la presente propuesta. En la sección 3.1, se presenta un modelo de calidad de contenidos que representa la base para la construcción del cuestionario. Luego, en la sección 3.2, se detalla la propuesta de construcción de un instrumento de evaluación subjetiva –cuestionario– teniendo en cuenta lo siguiente: presentación del conjunto de pasos sistemáticos para diseñar un instrumento de evaluación validado por expertos y destinado a usuarios finales, e instanciación del cuestionario, validado previamente a partir de un modelo de calidad de referencia de contenidos Web. Más adelante, en la sección 3.3, se presenta una experiencia realizada con 6 usuarios (evaluadores) expertos que respondieron el cuestionario, para luego, en la sub-sección 3.3.1 documentar y resaltar los resultados correspondientes. Finalmente en la sección 4 se remarcan las conclusiones y las líneas de avance.

2 Trabajos Relacionados/Discusión

El diseño, la medición y la evaluación de calidad de contenido para la Web, en base a un modelo conceptual [17] y un marco de referencia [16] no es frecuente en la literatura relacionada. Más bien se encuentran propuestas donde se trata de establecer o fijar un piso de “calidad” en base a heurísticas, reglas prácticas y sentido común. En este sentido, Tate *et. al.* [20] han elaborado una propuesta de evaluación basada en checklists sobre ciertas características claves y comunes, como por ejemplo autoridad, exactitud, actualidad, entre otras, con un destacable trabajo de campo y cobertura de los dominios más importantes, pero no está claro si hay un modelo de calidad subyacente que la sustente.

Si se deseara evaluar o establecer la calidad de contenido en base a esta propuesta, la pregunta que puede formularse sería: ¿es factible elaborar, de un modo sistemático, nuevos cuestionarios para diferentes dominios o sub-dominios específicos, además de los ejemplos incluidos en la cita? Sin ser necesariamente negativa, la respuesta será distinta para distintas audiencias y dependerá de la interpretación –subjetiva- de los interesados.

Con otra perspectiva, y considerando las particularidades de las aplicaciones Web [7], se valora especialmente en nuestro trabajo una propuesta que refleje la influencia e interdependencia que tienen para las aplicaciones Web actuales, las distintas perspectivas de la calidad de software, en relación a la de calidad de contenido.

Por lo general, un programa de software tradicional es una mezcla de funciones y datos, pero una aplicación Web posee además orientación a contenido, es decir, que está diseñado para proporcionar información de distintos medios (y por lo general, no estructurada semánticamente). También un componente de sitio Web, como por ejemplo un carrito de compra, permite editar una cantidad de artículos y volver a calcular los precios (aplicando una función a ciertos datos). Puede decirse así que la información es la interpretación significativa de los datos, para un uso y en un contexto determinado. En ese sentido, en [17, 18] los autores han desarrollado una propuesta construida en base al modelo ISO 9126-1 de calidad de producto software [9] (y ahora ampliado para la familia ISO 25010 [10]) proponiendo extender y articular el modelo ISO con una nueva característica, a saber: calidad de Contenido. Argumentan que el conjunto de seis características propuestos originalmente, es decir, funcionalidad, usabilidad, confiabilidad, eficiencia, mantenibilidad y portabilidad, no son adecuadas para (o no tenían por objeto) especificar o evaluar los requisitos de calidad de la información de un artefacto software.

Por lo tanto, a la cuestión central sobre cómo se puede especificar la calidad de contenido para aplicaciones Web desde el punto de vista de calidad interna y externa, indican que se debe tratar también con la característica de calidad de Contenido, que a su vez pueden subdividirse en exactitud, adecuación, accesibilidad y cumplimiento de cuestiones legales.

En la sección siguiente se discutirá cómo en base a este aporte puede diseñarse un modelo de evaluación de calidad de contenido como un esquema general, adaptable a todo tipo de aplicaciones Web. Al tratarse de un modelo integrado, significa que puede usarse tanto con aplicaciones Web 1.0 como Web 2.0 para diferentes tipos de dominio. Por esta razón, cuando se decide aplicarlo a una determinada aplicación Web, debe crearse una instancia sobre la base de parámetros más específicos, por

ejemplo atributos mensurables y las métricas correspondientes o ítems de cuestionario. Deben establecerse también criterios de decisión relacionados para los indicadores considerando el dominio específico, su necesidad de información y la perspectiva de los usuarios potencialmente involucrados.

Este enfoque proporciona, a nuestro juicio, una base sustentable para evaluar calidad de contenido para la Web, generalizable y robusta, en base a un estándar de referencia. En definitiva, en este trabajo tomamos como referencia este modelo, con el fin de diseñar sistemáticamente cuestionarios para evaluar calidad de contenido Web.

3 Evaluando Calidad de Contenido para Aplicaciones Web

Esta sección introduce el modelo de calidad empleado como referencia para el diseño del cuestionario y luego presenta los pasos del diseño en sí, para mostrar finalmente los resultados del procedimiento; es decir, el cuestionario destinado a usuarios finales.

3.1 El modelo de calidad

El modelo de calidad de contenido que empleado (desarrollado en [17, 19]) surge como respuesta a la necesidad de evaluar todas las dimensiones de las aplicaciones Web 2.0: un complejo compendio entre datos/contenido, funcionalidad y servicios que tienen algunas semejanzas con las aplicaciones de software tradicional, pero también sus propias particularidades.

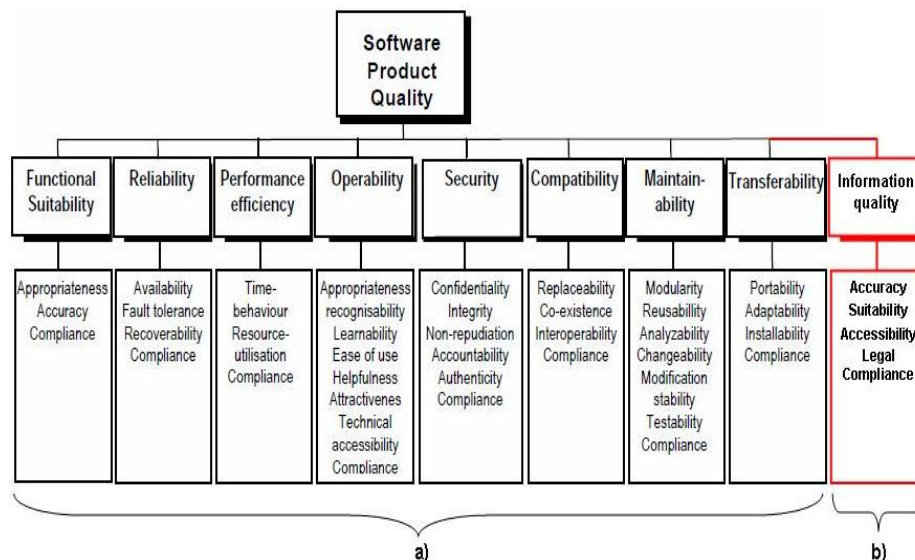


Fig. 1. a) Modelo de calidad Interna/Externa según ISO 25010; b) Agregando Calidad de Información (Information Quality)

Tabla 1. Definición de las sub-características para Calidad de Contenido o de Información

<p>1. EXACTITUD (ACCURACY): grado en el que el contenido es fiable y libre de errores</p> <p>1.1 CORRECCIÓN: grado en que el contenido es confiable en el sentido de estar libre de errores</p> <p>1.2 CREDIBILIDAD: grado en que el contenido es de confianza, objetivo y verificable</p> <p>1.2.1 AUTORIDAD: grado en que la fuente de contenido es fidedigna</p> <p>1.2.2 OBJETIVIDAD: grado en que el contenido expresa hechos o información con independencia de la propia manera de pensar o de sentir (sin distorsión)</p> <p>1.2.3 VERIFICABILIDAD: grado en la cual el propietario o autor del contenido puede ser verificado</p> <p>1.2.4 ACTUALIDAD: grado en que el contenido puede ser identificado como actual, que atrae y ocupa la atención de los usuarios en un momento dado.</p> <p>2. ADECUACION (SUITABILITY)</p> <p>2.1. VALOR AGREGADO: grado en que el contenido puede ser novedoso, beneficioso y contribuye a que el usuario realice la tarea actual</p> <p>2.1.1 NOVEDOSO: grado en el cual el contenido es fresco y contribuye a tomar nuevas decisiones para el objetivo previsto por el usuario</p> <p>2.1.2 BENEFICIOSO: grado en el cual el contenido es ventajoso y contribuye para tomar nuevas decisiones para el objetivo previsto por el usuario</p> <p>2.1.3 REACTIVO: grado en el cual el contenido contribuye a interactuar y reaccionar para el objetivo previsto por el usuario</p> <p>2.2. COBERTURA: amplitud de temas cubiertos por el contenido y la profundidad con la que son tratados</p> <p>2.2.1. APROPIADO: grado en el cual el contenido aporta a una meta prevista del usuario</p> <p>2.2.2. COMPLETO: grado en el cual el contenido tiene suficiente cantidad de información para el objetivo previsto por el usuario</p> <p>2.2.3. CONCISO: grado en el cual el contenido está representado en forma compacta sin ser abrumadora</p> <p>2.3. CONSISTENTE: grado al cual el contenido es constante a un fragmento de información o a la página del sitio con respecto para el objetivo previsto por el usuario.</p> <p>3. ACCESIBILIDAD: capacidad de un producto Web a entregar información accesible para todos los usuarios tomando en cuenta aspectos técnicos y de representación</p> <p>4. CONFORMIDAD LEGAL: capacidad del producto Web para adherirse a los estándares, convenciones, y normas legales relacionadas con el contenido como también con los derechos de propiedad intelectual</p> <p>4.1 PRIVACIDAD</p> <p>4.2 COPYRIGHT</p> <p>4.3 CONFORMIDAD CON LAS LEYES</p>
--

Este modelo toma en cuenta estándares ISO, ya utilizados en otros contextos de evaluación [5, 16], integrándolos y extendiéndolos. Se trata de un modelo mixto que está basado en la familia ISO-2501X [8, 10], (antes familia ISO 9126-X [9]) orientadas a modelar, medir y evaluar calidad interna, externa y calidad en uso de producto software.

Según se observa en la Figura 1, el nuevo modelo para calidad interna y externa, está basado en la ISO 25010 en tanto emplea un esquema jerárquico del tipo categoría-subcategoría-atributo y lo extiende en cuanto añade una nueva categoría o característica, denominada Calidad de Contenido (Information Quality) a las ya prescriptas, a saber: Adecuación Funcional, Confiabilidad, Eficiencia, Operabilidad, Seguridad, Compatibilidad, Mantenibilidad y Transferibilidad.

A su vez esta nueva categoría tiene cuatro sub-categorías según se observa en tabla 1, a saber: Exactitud, Adecuación, Accesibilidad y Conformidad Legal. Luego, cada sub-categoría incluye una serie de sub-características o atributos que, eventualmente, pueden ser medidos, evaluados y comparados para realizar recomendaciones que orienten a la comprensión y mejora.

En este trabajo hemos optado por desarrollar un cuestionario, como instrumento de medición y evaluación, que constituye una instanciación de la característica calidad de contenido.

El cuestionario consta de un conjunto de ítems, en correspondencia con cada atributo identificado para cada sub-característica de la tabla 1, que serán valorados para establecer, finalmente, un indicador de calidad de contenido. En el caso en que se realizara un estudio más amplio en consideración de cantidad de características, con el fin de obtener un indicador global de calidad de un artefacto, se puede emplear los modelos de agregación apropiados, tal como realizado en [5, 16].

A seguir presentamos el diseño del cuestionario, considerando las particularidades del modelo discutido.

3.2 Diseño del cuestionario

La idea subyacente en el diseño del cuestionario fue capitalizar la experiencia en la construcción de un cuestionario exitoso en la industria para medir usabilidad, como realizado en [3]. El valor de este procedimiento fue luego corroborado por otros autores [22], quienes por otro método comprobaron la validez del instrumento.

Se decidió formular un lote de preguntas en relación al nivel más bajo de agregación que presente cada característica del modelo de calidad de Contenido. De este modo, si una característica presenta sub-características y atributos, las preguntas se formularon en referencia al último nivel. Así, por ejemplo, en el cuestionario hay una pregunta (id 5, en Tabla 3) respecto a Autoridad, que está en un segundo nivel del modelo (1.2.1 en Tabla 1) y otra (id: 2 en Tabla 3) respecto a Corrección que está en un nivel superior (1.1 en Tabla 1). En ambos casos se trata del nivel menor de agregación en esa rama del árbol.

El paso siguiente fue ordenar las preguntas, eliminar las que semánticamente representaran el mismo concepto y establecer un conjunto de criterios de aceptación, tal como pueden encontrarse en tabla 2 que pudieran evaluarse en forma binaria (cumple:1/no cumple:0); y así, empleando un sencillo procedimiento de cálculo,

Un Método Sistemático para Diseñar Instrumentos de Evaluación Aplicado a Calidad de Contenidos Web

establecer un indicador global de admisión que denominamos Peso de Inclusión (PdI).

El cálculo del PdI consistió en contar para cada ítem, cuántos unos obtuvo considerando todos los criterios, en una escala que iba de 1 a 5 como máximo, para el caso ideal. Los ítems que eran considerados innecesarios, por lo expuesto anteriormente, quedaban automáticamente fuera de la selección, sin puntuación. Los criterios de aceptación (CA) para cada ítem fueron los siguientes:

Tabla 2. Criterios de Aceptación (CA) para la selección inicial de Ítems del Cuestionario para Expertos

CA1	CA2	CA3	CA4	CA5	CA6	CA7
Es necesaria para la evaluación	No son necesarias más preguntas?	Los encuestados tienen la información necesaria para responder?	No se necesitan más detalles?	Es suficientemente general	Se puede responder en forma consistente con la realidad?	Pertenece a la subcategoría a la que se asigna?
Este criterio tiene en cuenta el nivel de detalle estrictamente necesario para la evaluación	Este criterio considera si no es necesario dividir el ítem en dos o más, agregando especificidad	Este criterio tiene por objetivo evitar ítem que requieren información fuera del alcance de los encuestados	Este criterio no permite la inclusión de ítem excesivamente generales	Este criterio evitaría puntualizar situaciones muy específicas en el ítem	Este criterio limitaría la presencia de ítem sensibles o incómodos para ser respondidos verazmente por los encuestados	Este criterio permite contrastar la pertenencia del ítem a la categoría asignada en la instancia del modelo de Calidad de Contenido

Una vez calculado el Peso de Inclusión para cada uno de los ítems se procedió a ordenarlos y a establecer un criterio de corte, por debajo del cual los ítems quedarían fuera del cuestionario para los expertos. Considerando los extremos de la escala adoptamos el punto medio, tres, como límite inferior para la admisión de ítem en el cuestionario. Este criterio de corte dejó vigentes un lote de 34 ítems a efectos de seleccionar aquellos que constituirían el cuestionario definitivo, a ser presentado a los evaluadores expertos.

El paso siguiente fue someter el cuestionario a una evaluación de 6 expertos, a quienes se solicitó la colaboración voluntaria, considerando sus antecedentes en el área de conocimiento de referencia. El número no es aleatorio, sino que responde a un estudio conocido y citado frecuentemente [14].

En el diseño del cuestionario se excluyó explícitamente la característica Usabilidad (Operabilidad en la Fig. 1), pues existen métodos [15] y guías para la selección de herramientas [23] bastante probados que permiten evaluaciones tanto objetivas como subjetivas y pueden integrarse al método de evaluación sin mayores inconvenientes.

3.3 El procedimiento de evaluación

El objetivo de este paso es detectar la pertinencia de los ítems propuestos para la evaluación prevista. El procedimiento de evaluación se realizó poniendo a disposición de los expertos el cuestionario con los 34 ítems (no mostrado aquí por cuestiones de espacio), basado en una escala tipo Likert de 7 puntos, acompañado de un instructivo sencillo para llenado en formato electrónico, durante un lapso de tiempo fijo, preestablecido entre dos fechas.

Los evaluadores consideraron el cuestionario en la evaluación de dos sitios Web pertenecientes al mismo dominio –e-government o gobierno electrónico- con las siguientes características: uno de ellos con clara orientación a la Web 2.0 y bien posicionado en un estudio comparativo independiente [6], y otro con evaluación de calidad inferior, ubicado en el último decil, del mismo índice.

La selección de los sitios a evaluar empleando un indicador externo, independiente, por un lado quita subjetividad a los artefactos de referencia para calibrar el instrumento de evaluación y, por otro lado, tendería a agrupar las respuestas en los extremos de la escala, generando un buen discriminador a la hora de seleccionar los ítem para el cuestionario final.

Una vez respondidos los dos cuestionarios por parte de cada evaluador se registraron las respuestas en una hoja de cálculo para luego computar el nivel de acuerdo en las respuestas, para cada ítem de cuestionario. Para ello se calculó la moda y la frecuencia de dicha moda. El primer indicador refleja la posición en la escala sobre la que los evaluadores tienen mayor nivel de acuerdo y el segundo, el grado de acuerdo alcanzado.

Por ejemplo, en el ítem (id. 3 de la Tabla 3): “Se indican las fuentes de los contenidos publicados para que puedan comprobarse con los datos originales”, las respuestas permitieron computar los siguientes valores: para el Sitio 1, moda=1 y frecuencia=5; para el sitio 2, moda=7 y frecuencia= 2. Esto se interpreta como que la mayor cantidad de respuestas para el sitio 1 fue 1 -totalmente de acuerdo- y que 5 evaluadores respondieron 1; para el sitio 2 la respuesta más seleccionada fue 7 -totalmente en desacuerdo-, pero sólo para dos evaluadores. Inicialmente puede inferirse que el nivel de acuerdo entre los evaluadores para el primer sitio es elevado, pero no así para el segundo, de modo que a priori no puede afirmarse que este ítem será incluido en el cuestionario, hasta tanto no se establezca un criterio de corte a partir desde donde refinar la selección de los ítems.

El paso siguiente fue determinar un valor de corte, por debajo del que se eliminarían los ítems que no cumplieran el criterio. Dicho valor se estableció en el valor superior más próximo al promedio de las frecuencias calculadas para cada ítem, entre ambos sitios. Esto permitiría dejar dentro de la selección inicial a aquellos ítems calificados con un nivel de acuerdo mayor por parte de los evaluadores. Como los promedios de las frecuencias oscilaron entre 2,5 y 4, se obtuvo 3,16 como promedio, ajustándose a 3,5 el valor de corte.

Por encima del valor de corte se encuentran 14 ítems, cuyos promedios de frecuencia de la moda va entre 3,5 y 4. Todos ellos formarán parte del cuestionario diseñado para usuarios finales. Pero si hubiera necesidad de recortar a menos cantidad de ítems el cuestionario final, se deberían seleccionar, para permanecer, aquellos cuyo nivel de acuerdo fuera mayor determinando posiciones más “extremas”, es decir,

sobre las puntas de la escala. Así, un ítem que tuviera un grado importante de acuerdo sobre el extremo inferior para el sitio 1 y sobre el extremo superior para el sitio 2 sería candidato a permanecer.

3.3.1 Resultados de la evaluación por usuarios expertos

Como resultado del procedimiento descrito anteriormente se obtuvieron los 14 ítems que forman parte del cuestionario para usuarios finales. Este cuestionario validado por el procedimiento, es de utilidad para evaluar de un modo individual o comparativo, un sitio Web del dominio para el que fue diseñado.

Además, la cantidad de ítems incluidos lo hace de una extensión apropiada para evitar respuestas por comodidad o semejanza. Es decir, el cuestionario resultante es conciso y, a la vez, tiene la suficiente amplitud como para dar cobertura a las características del modelo de calidad de contenido, en particular para el dominio de gobierno electrónico.

En tabla 3 detallamos los ítems del cuestionario resultante para usuarios finales en dicho dominio.

Tabla 3. Cuestionario para usuarios finales –de 14 ítems- luego del procedimiento de selección. Leyenda: **Id.:** representa el orden de los ítems en el cuestionario original –de 34 ítems- presentado a los evaluadores expertos.

Nº	Id.	Ítem
1	2	Hay indicación de que el contenido ha sido revisado por un editor para comprobar su exactitud?
2	3	Se indican las fuentes de los contenidos publicados para que puedan comprobarse con los datos originales?
3	5	Se indica la autoridad de la organización, persona o empresa identificada como responsable del sitio?
4	10	Si es posible agregar opiniones al contenido, existe algún mecanismo para informarlas como agresivas o inoportunas al moderador?
5	11	Se deja constancia sobre si el material publicado ha sido tomado de una fuente de información externa?
6	14	En caso de material sujeto a revisión, se incluye la fecha de la última revisión?
7	16	Se da de baja o se retira de publicación en forma regular el contenido antiguo o no actualizado?
8	17	Se indica en forma resaltada la incorporación de contenido nuevo en el sitio?
9	18	Si el contenido incluye información de actualidad, ¿se indica la frecuencia de actualización?
10	19	Si el contenido originalmente fue obtenido desde un programa emitido por radio o televisión, se indica claramente fecha y hora de la emisión?
11	20	Si el material original fue capturado de un informe impreso, ¿se indica claramente la edición de la cual se han tomado los datos?
12	22	No hay modo de opinar sobre los contenidos
13	24	Se ha unificado el formato internacional de fechas para el sitio?
14	31	El contenido de la información está compuesto por partes que no están correctamente articuladas?

4 Conclusiones y Trabajos Futuros

La complejidad de las aplicaciones Web actuales y sus características heredadas y propias como funcionalidad de programas, datos, información no estructurada; diversidad de audiencias, aspectos de internacionalización, de accesibilidad, de seguridad y legalidad, entre otras, hacen que la calidad de contenido continúe siendo un ciudadano de primera clase.

En este trabajo hemos propuesto un método sistemático para la construcción de cuestionarios que permitan evaluar la calidad de contenido de aplicaciones Web –ya sean de las eras Web 1.0 o 2.0-, considerando para ello un modelo de calidad subyacente, como discutido en la sección 3.1. Este enfoque hace que el método sea reusable y pueda emplearse en distintos contextos, audiencias, dominios y perfiles de usuario. En cada caso se requiere una instanciación del modelo (de calidad de contenido en nuestro caso), la participación de un reducido número de expertos, y la construcción de un instrumento destinado a evaluadores finales, en general usuarios del dominio al que pertenezca la aplicación Web a evaluar.

Para el diseño del cuestionario en sí hemos empleado un procedimiento diseñado por otros autores, adecuadamente validado luego por estudios comparativos independientes. La secuencia de pasos detallados en el procedimiento (en las secciones 3.2 y 3.3) conduce a la selección, debidamente justificada, de un conjunto de ítems que formarán parte del cuestionario para los usuarios finales. Como resultado del trabajo se obtienen dos entregables, a distinto nivel de abstracción, a saber: i) la secuencia de pasos para obtener un cuestionario destinado a evaluar calidad de contenido Web y, ii) un cuestionario de catorce ítems, “listo para usar”, a ser empleado en un estudio de sitios Web del dominio de gobierno electrónico.

En cuanto a trabajos futuros, a partir del cuestionario obtenido, se está diseñando un estudio comparativo de calidad de contenido Web para aplicaciones de gobierno electrónico en el ámbito municipal, con participación de usuarios finales como evaluadores. Por otra parte, se encuentra en desarrollo una herramienta para construcción, almacenamiento y análisis de cuestionarios de calidad de contenido, considerando como referencia aplicaciones de uso libre y gratuito.

5 Referencias

1. Bevan N.; Measuring usability as quality of use, *Software Quality Journal*, (4), pp 115-150, 1995.
2. Bevan N.; Usability is Quality in Use, *Proceedings 6th Int'l Conference on Human Computer Interaction*, Yokohama, 1995.
3. Brooke, J., SUS –A ‘quick and dirty’ usability scale; Digital Equipment Co Ltd., Reading, UK. <http://www.usability.serco.com/trump/documents/Suschapt.doc> .
4. Caro A., Calero C., Caballero I., Piattini M.; A proposal for a set of attributes relevant for Web portal data quality. V.16: 4 pp. 513-542, Kluwer A., MA, USA, 2008
5. Covella G.J, Olsina L.; Assessing quality in use in a consistent way, *Proc. 6th Int'l Conference on Web engineering*. ACM, New York, NY. ISBN: 1-59593-352-2
6. Fernández Arroyo N., Boix M.V., Pando D., Calamari M.; Índice Nacional de Páginas Web Municipales 2008, CIPPEC y Universidad de San Andrés Argentina, <http://www.cippec.org/Main.php?do=documentsDoDownload&id=352> 28/agosto/2009

Un Método Sistemático para Diseñar Instrumentos de Evaluación Aplicado a Calidad de Contenidos Web

7. Ginige A., Murugesan S.; The Essence of Web Engineering— Managing the Diversity and Complexity of Web Application Development, *IEEE MultiMedia* 8, 2, 2001.
8. ISO/IEC 25012; Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) -- Data quality model, 2008
9. ISO/IEC 9126-1; Information technology - Software product evaluation - Quality characteristics and guidelines for their use, 2001
10. ISO/IEC CD 25010 ; Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Quality model and guide, 2009
11. Katerattanakul P., Siau, K.; Measuring information quality of web sites: development of an instrument. In *Proc 20th Int'l Conference on Information Systems*. Association for Information Systems, Atlanta, GA, pp. 279-285, 1999.
12. Nielsen J., Molich R., Snyder C., Farrell S; E-Commerce User Experience, NNGroup, 2001.
13. Nielsen J.; *Designing Web Usability*; New Riders, 2000.
14. Nielsen, J; The Alertbox column; <http://www.useit.com/alertbox/> (1995-2010).
15. Nietzio, A., Strobbe, C., Velleman, E.: The Unified Web Evaluation Methodology (UWEM) 1.2 for WCAG 1.0. In *Proc. 11th Int'l Conference on Computers Helping People with Special Needs*. LNCS 5105, pp 394-401, 2008.
16. Olsina, L. Papa, F., Molina, H. Organization-Oriented Measurement and Evaluation Framework for Software and Web Engineering Projects. L.; *Proceedings Series: Lecture Notes in Computer Science*, Vol. 3579, D. Lowe, M. Gaedke (Eds.) pp. 42-52, 2005.
17. Olsina L., Sassano R., Mich L. Towards the Quality of Web 2.0 Applications, In *proc. of 8th Int'l Workshop on Web-oriented Software Technology (IWWOST 2009)* held at Int'l Congress on Web Engineering (ICWE09), San Sebastian, Spain, V. 493, pp. 3-15, CEUR (ceur-ws.org), ISSN 1613-0073, 2009.
18. Sassano R., Olsina L., Mich L.; Modelling Content Quality for the Web 2.0 and Follow-on Applications, In: *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*; Murugesan S. (Ed.) Ch. XXI, pp. 371-386, IGI Group, 2010.
19. Sassano R.; *Un Modello per la Qualità del Contenuto dei Siti Web 2.0*, Tesi di Laurea Specialistica in Net-Economy, UdSdT, Facoltà di Economia, 2007.
20. Tate M.A , Alexander J.E. ; *Web Wisdom: How To Evaluate and Create Information Quality on the Web*- ISBN-10: 0805831231. 2nd Ed. 2010.
21. Trochim W.M. K.; *The Research Methods Knowledge Base*, Cornell University, 2nd Ed. ISBN:0-9701385-9-8, 2001
22. Tullis T.S., Stetson J.N.; A Comparison of Questionnaires for Assessing Website Usability. http://www.usabilityprofessionals.org/conferences_and_events/upa_conference/2004/index.html
23. Web Accessibility Initiative (WAI), *Selecting Web Accessibility Evaluation Tools*, <http://www.w3.org/WAI/eval/selectingtools.html>