¿CÓMO DISEÑAR UN CORPUS DE CALIDAD? PARÁMETROS DE EVALUACIÓN[1]

Miriam Buendía Castro
José Manuel Ureña Gómez-Moreno
Universidad de Granada

Resumen

La compilación de un corpus ejerce una gran influencia en los resultados que se obtienen de cualquier tipo de investigación. A día de hoy, Internet se configura como la principal fuente abastecedora de textos para la elaboración de un corpus. No obstante, dado que cualquiera puede publicar en la Web sin ningún tipo de revisión, los investigadores deben asegurarse de que los textos provengan de sitios fiables, por lo que deben evaluar continuamente la calidad de los recursos textuales.

La presente investigación establece una serie de parámetros de evaluación que hemos usado para evaluar la validez de los sitios web. Nuestro protocolo de evaluación se compone de tres parámetros, a saber, autoridad, contenido y diseño, cada uno de los cuales se subdivide en una serie de subparámetros. Al aplicar este protocolo de evaluación a los textos extraídos del sitio web, la calidad del corpus está asegurada. El uso de este protocolo puede hacerse extensivo para asegurar la calidad de corpus de otros dominios.

Palabras clave: calidad de un corpus, Internet, evaluación

Abstract

The compilation of a corpus is an important factor which greatly influences the results obtained in any research study. While the Internet is currently the principal source of texts for corpus creation, the fact that anybody can publish on the web without any kind of revision means that researchers must always ensure that texts come from reliable sites, and must continually assess the quality of textual resources.

This paper describes a set of evaluation parameters used by the authors to assess website validity. Our evaluation protocol is composed of three parameters, namely authority, content and design, each of which is divided into a set of sub-parameters. By applying this evaluation protocol to website texts, corpus quality may be assured. In addition, the protocol may be extended to assure the quality of corpora in other domains.

Keywords: quality of a corpus, Internet, evaluation.

1. Esta investigación forma parte del proyecto *Ecosistema:* Espacio único de sistema de información ontológica y tesaurus sobre el Medio Ambiente (FFI2008-06080-C03-01/FILO) financiado por el Ministerio de Ciencia e Innovación.

1. Explosión web

Internet ha traído consigo una nueva forma de organizar y obtener información. La *World Wide Web* ofrece la posibilidad de acceder a cualquier tipo de información en cualquier momento. Esta es la razón por la que aspectos como la *visibilidad*, *accesibilidad* y *diseño universal* han cobrado un interés especial en los últimos años.

En la actualidad nos encontramos inmersos en lo que podemos denominar una "sobrecarga de información" (Jiménez Piano y Ortiz-Repiso Jiménez 2007: 18). La cantidad de información que circula por Internet en un día cualquiera es mayor que toda la información que había disponible a lo largo del siglo XIX (Austermühl 2001: 7). El inglés continúa monopolizando la red al representar el 45% del número total de páginas web. Otras lenguas europeas con un porcentaje destacado de páginas web son el alemán (5,90%), el francés (4,41%), el español (3,80%), el italiano (2,66%) y el portugués (1,39%)[2].

Como señalan Kilgarriff y Grefenstette (2003: 333), la WWW se ha convertido en un "fabulous linguists' playground". La Traducción no ha permanecido ajena a este desarrollo y la metodología utilizada en las fases de documentación y vaciado terminológico ha evolucionado sustancialmente. Si en el pasado los traductores se apoyaban casi exclusivamente en recursos lexicográficos y terminográficos —como los diccionarios generales y especializados—, para desempeñar su trabajo, ahora los córpora, cuando están correctamente diseñados, proporcionan al traductor información tanto lingüística como conceptual que no es posible encontrar en los diccionarios. Así pues, un corpus especializado resulta de gran utilidad en la traducción puesto que proporciona terminología especializada y actualizada, a la vez que aporta información sobre el uso combinatorio de los términos y su frecuencia.

Sin embargo, dada la naturaleza anárquica de los textos que se publican en Internet y puesto que no existe ninguna organización que coordine o supervise las publicaciones en línea, hay gran cantidad de páginas web que carecen de precisión, fiabilidad y validez (Austermühl 2001: 64). Resulta necesario, por tanto, desarrollar una estrategia de evaluación que asegure la calidad de los documentos que posteriormente pasarán a formar parte de un corpus, ya que esto condicionará la validez de la información que se pueda extraer. No obstante, los criterios no son prescriptivos y, como apunta Cooke (1999: 52), los lectores deberán seleccionar los criterios pertinentes en función de sus propias necesidades.

La presente investigación establece una serie de parámetros de evaluación que hemos usado para evaluar la validez de los sitios web. Nuestro protocolo de evaluación se compone de los parámetros *autoridad*, *contenido* y *diseño*, cada uno de los cuales se subdivide en una serie de subparámetros.

2. Estos resultados corresponden a la investigación llevada a cabo por la Unión Latina: <http://dtil.unilat.org/LI/2007/es/resultados_es.htm> [Consulta: 25/06/09].

Este trabajo se enmarca dentro del proyecto *Ecosistema* (Espacio único de sistema de información ontológica y tesaurus sobre el Medio Ambiente), financiado por el Ministerio de Ciencia e Innovación. El principal objetivo de Ecosistema es la representación de la estructura conceptual del dominio de la gestión integrada de zonas costeras en forma de un tesauro visual de conceptos especializados, organizados en marcos dinámicos de conocimiento especializado. Ecosistema pretende, además, crear un diccionario electrónico multilingüe (inglés-español-alemán) en el dominio de la Ingeniería del Medio Ambiente. Del corpus compilado para el proyecto se han extraído los datos necesarios para establecer la estructura conceptual del dominio a través de la elaboración de jerarquías terminográficas. El corpus compilado también se ha utilizado en la elaboración de las definiciones que incluyen la información contextual que pertenece a cada término.

2. Calidad y evaluación de los recursos disponibles en línea

Antes de profundizar en los conceptos de calidad y evaluación y dado que este artículo proporciona una serie de parámetros de evaluación que hemos usado para evaluar la validez de los *sitios web*, consideramos necesario matizar el significado de términos como *página web* (webpage), *sitio web* (website) o *página de inicio* (homepage), que a menudo se utilizan indistintamente.

Una *página web* es cada documento individual que se visualiza en la pantalla. Así pues, el *sitio web* de una compañía u organización podrá estar compuesto de varias *páginas web.* La puerta de entrada a un *sitio web* se denomina *página de inicio* o *homepage*. La *página de inicio* sirve como punto inicial de navegación a través de un sitio web. Un *sitio web,* de esta forma, puede definirse como un conjunto de páginas web que comparten la misma página de inicio e hipervínculos internos y que constituyen un tipo de unidad documental (Jiménez Piano y Ortiz-Repiso Jiménez 2007: 33).

Tal y como señala Austermühl (2001: 52), "finding data on the world wide web is no problem at all. But finding reliable information is rather a difficult task. And finding the information you really need can be very time-consuming and often frustrating". La evaluación de un sitio web está íntimamente asociada a la calidad puesto que el objetivo de toda evaluación es establecer el máximo nivel de calidad. El término *calidad* se usa, a menudo, para denotar *buena calidad* o *alta calidad*. Referido a la información disponible en Internet, *calidad* hace referencia a fuentes que son precisas y fiables (Cooke 1999: 14). En este sentido, la norma ISO 8402-94 define *calidad* como "the set of characteristics of an entity that give that entity the ability to satisfy expressed and implicit needs".

Se ha escrito mucho acerca de los parámetros que determinan la calidad de los recursos digitales, pero a día de hoy todavía no existe un consenso al respecto. Según Sinclair (2005), cualquier selección debe hacerse atendiendo a unos criterios

y el primer paso en la elaboración de un corpus es determinar los criterios que se van a seguir para seleccionar los textos. Cooke (1999) establece un inventario de diez parámetros: (i) objetivo; (ii) cobertura; (iii) autoridad y reputación; (iv) precisión; (v) actualización y mantenimiento de la fuente; (vi) accesibilidad; (vii) presentación y disposición de la información; (viii) facilidad de uso de la fuente; (ix) comparación con otras fuentes; (x) la calidad global de la fuente. Tanto Auer (1999) como Alexander y Tate (1999) subrayan cinco parámetros: autoridad, cobertura, objetividad, precisión y actualización. Codina (2000), por su parte, habla de seis parámetros: autoridad, contenido, accesibilidad, ergonomía, luminosidad[3] y visibilidad, mientras que el inventario de Jiménez Piano y Ortiz-Repiso Jiménez (2007) se compone de búsqueda y recuperación, autoridad, contenido, administración de recursos y diseño.

Gordon-Murname (1999) llevó a cabo una investigación en la que examinó las políticas de evaluación de doce servicios de evaluación web. Destacó la falta de consenso entre los servicios de revisión y concluyó que el único parámetro en el que todos coincidían era en el de *contenido*[4].

Tabla 1**.** Criterios elegidos por los doce servicios de evaluación

Criterio	**Evaluadores que lo incluyen en su lista**
Contenido	12
Diseño/Presentación/Formato	11
Frecuencia de actualización	8
Audiencia/Necesidades de la comunidad	7
Actualización	7
Sistema de medición	7
Autoridad	5
Disponibilidad/ Velocidad	5
Valor/Utilidad	5
Accesibilidad	4
Alcance	4
Coste	3

Son muchos los autores que destacan la importancia de la *autoridad*, *objetividad* y *actulización*, ya sea como parámetros (ej. Alexander y Tate 1999; Beckwith 2005; Duke University Libraries 2007; Kapoum 1998; Kirk 1996; Videon 1998) o como

3. Luminosidad es un término acuñado por Codina (2000) que se refiere al número de vínculos que una página web tiene a otras páginas web.

4. Los doce servicios que se analizaron fueron: *CyberStacks*, *Best Information on the Net (BIOTN)*, *Librarian's Index to the Internet*, *Scout Report*, *Argus Cleringhouse*, *Blue Web'N*, *Dow Jones Business Directory*, *Finding Business Research on the Internet*, *The PH Directory of Online Business Information 1998*, *Lycos Top and Magellan/MackKinley SelectSurf.*

indicadores del parámetro *contenido*. Este es el caso de Anderson, Allee, Grove y Hill (1999) y Kessinger (2008). Otros autores apuestan también por la *precisión* y el *alcance*, como es el caso de Beck (2009) y Schrock (2006). Las propuestas que se decantan por el *diseño* han aumentado notoriamente en los últimos años. Algunos de los autores que abogan por el diseño de un sitio web como una manera de garantizar la calidad son Gaffney (1998), Adreon, Catey y Strysick (2002) y Pearl K. Wise Library (2006).

3. Protocolo de evaluación de parámetros

Si el número de sitios web que consultáramos fuera pequeño, podríamos analizarlos basándonos en los 509 criterios propuestos por Wilkinson, Oliver y Bennett (1997). Sin embargo, en proyectos de investigación más grandes, como es nuestro caso, necesitamos resultados de calidad que puedan obtenerse de una forma rápida. Así pues, siguiendo las propuestas de Alexander y Tate (1999), Cooke (1999), Codina (2000) y Jiménez Piano y Ortiz-Repiso Jiménez (2007), nuestro protocolo de evaluación se basa en tres parámetros, a saber, *autoridad*, *contenido* y *diseño*.

3.1. Autoridad

Las principales dificultades inherentes a la evaluación de la calidad de la información en la web son: (i) su naturaleza descontextualizada; (ii) la falta de estandarización en la presentación de la información. Así pues, la identificación de los autores y de su estatus y reputación profesional resultan cruciales a la hora de garantizar la naturaleza científica de la información (Jiménez Piano y Ortiz-Repiso Jiménez 2007: 149).

La autoridad hace referencia al prestigio y al crédito que se reconoce a una persona o institución por su calidad y competencia en una área determinada (RAE 2001). La evaluación de la autoridad de un documento está basada en una variabilidad de factores, pero ante todo en el conocimiento y experiencia de aquellos responsables en producir la página. Así pues, una fuente será autoritaria si está escrita por un experto en la materia o si está producida por una institución con reconocido prestigio en el campo. Tal y como señala Austermühl (2001: 64-65), "the validity of an online source depends very much on the credibility of its author", por lo que una página anónima o sin autor gozará, en líneas generales, de poca credibilidad. A veces, incluso, cuando encontramos una página web a través de un buscador, resulta necesario retroceder en la URL para obtener información acerca del autor. Otra manera de comprobar la autoridad de una persona, en caso de que la información no se proporcione en la página web, es indagar en el *background* de dicha persona a través de un buscador como Google.

La dirección de un sitio web también puede ser indicativa de la credibilidad de su autoridad. La siguiente URL se ofrece a modo de ejemplo de cómo se organiza un sitio web: http://www.intute.ac.uk. La última parte "uk" recibe el nombre de *dominio de primer nivel* (top-level domain, TLD) y las otras partes se denominan *sub-dominios*. Las letras "uk" se refieren al "Reino Unido", las letras "ac" a "academia", la palabra "intute" es el nombre de la organización, y la primera parte de la URL –http://– designa el protocolo de comunicaciones que se usará para transferir la información requerida. HTTP es el protocolo estándar para la transferencia de documentos HTML en Internet (Austermühl 2001: 46).

Los dominios territoriales o geográficos normalmente usan los códigos asignados por la Organización Internacional para la Estandarización (ISO). Los códigos suelen consistir en las dos primeras letras del nombre del país original[5]. Los dominios de primer nivel más comunes son (Austermühl 2001: 47):

- **-.com**. Designa instituciones comerciales y es el dominio más extendido de la web.
- -.**mil.** Dominio que se usa para sitios web de carácter militar de los Estados Unidos.
- **-.net.** Designa empresas u organizaciones que actúan en calidad de proveedores de red o que tienen que ver con la administración de redes.
- **-.edu.** Originalmente designaba los sitios web de todas las entidades relacionadas con la educación de los Estados Unidos (universidades, institutos, escuelas, etc.). En la actualidad su uso ha quedado restringido a sitios web de la Universidad y de estudios superiores de cuatro años. Las escuelas y estudios superiores de dos años se registran como dominio "us".
- **-.gov.** Este dominio se reserva para agencias del gobierno federal de los Estados Unidos como el Departamento del Estado, el Senado, la Casa Blanca o la Biblioteca del Congreso.
- -.**org.** Dominio usado por diversas organizaciones, en especial instituciones internacionales como Naciones Unidas.
- **-.int:** Dominio que designa organizaciones establecidas por acuerdos internacionales.

Las páginas personales, cuyos dominios típicos son .name, .members, users, people, ~, %, merecen una atención especial. La mayoría de ellas no ofrecen más información que la de "este es mi gato" o "aquí es donde yo vivo" (Cooke 1999: 10). A veces, los autores crean y desarrollan un sitio personal en un trabajo concreto o universidad, pero después de unos años abandonan el sitio sin eliminar la infor-

5. La escuela Donald Bren de Ciencias de la Computación y de la Información de la Universidad de California ofrece una lista de todos los códigos de los dominios de los países: <http://ftp.ics.uci.edu/pub/websoft/wwwstat/country-codes.txt> [Consulta: 15/05/09].

mación. Cuando los usuarios intentan seguir los vínculos para buscar información, sólo encuentran *file not found* o una página obsoleta o desactualizada. Los dominios están pensados para ayudar a categorizar los recursos de Internet, ya que nos proporcionan información esencial que puede determinar la fiabilidad de una página. Esta es la razón por la que nos permiten aceptar o descartar una página, incluso antes de abrirla.

En cuanto a los dominios territoriales o geográficos, en lo que a nuestra investigación se refiere, aceptamos cualquier sitio web de calidad proveniente de cualquier país de habla hispana, inglesa o alemana, puesto que estas son las lenguas de trabajo de nuestro proyecto.

Las leyes de Copyright que hacen referencia al uso de la información electrónica y de la información disponible a través de Internet, varían dependiendo del país. Por ello, conviene que el autor proporcione información acerca del poseedor del copyright del material o los detalles acerca de cómo debe citarse la información en una publicación, o a quién debemos dirigirnos en caso de que necesitemos permisos de Copyright (Cooke 1999: 70).

Algunos investigadores sostienen, además, que una página será más fiable cuantas más visitas tenga. Nosotros no hemos tenido en cuenta la visibilidad de una página para la evaluación ya que la popularidad de un sitio no se corresponde necesariamente con su calidad. Un ejemplo claro lo encontramos en las páginas de contenido sexual. Aunque se encuentran entre las más visitadas de la web, esto poco tiene que ver con la calidad del contenido que ofrecen.

3.2. Contenido

El parámetro *contenido* lo hemos dividido en *cobertura*, *precisión*, *objetividad*, *actualización* y *audiencia*.

3.2.1. Cobertura

La cobertura de un sitio web se refiere a si el sitio web reúne la cantidad de información suficiente y de validez referida al tema.

3.2.2. Precisión

La precisión determina si la información es fidedigna y no contiene errores (Alexander y Tate 1999: 11). Así pues, es importante comprobar si los textos del sitio web no contienen errores gramaticales, tipográficos o de deletreo, y si hay referencias a otras fuentes de información.

Sinclair (2005) reconoce la existencia de errores en cualquier corpus. Advierte de que la precisión perfecta es, a menudo, sistemáticamente imprecisa. A modo de ejemplo señala que en un corpus de aproximadamente cien millones de palabras, con un 99% de precisión, contendrá, obviamente, más de un millón de errores.

3.2.3. Objetividad

La objetividad determina la expresión de hechos o de información sin distorsión por sentimientos personales u otros prejuicios (Alexander y Tate 1999: 13). Los principales indicadores que hay que tener en cuenta en la evaluación de la objetividad son los siguientes:

- ¿Resulta evidente el punto de vista del individuo u organización responsable de proporcionar información?
- ¿Cuenta la página con anuncios? En caso afirmativo, es importarte determinar hasta qué punto el anunciante puede influir en los contenidos de la información.
- ¿Están claramente citados los patrocinadores corporativos o sin ánimo de lucro del sitio web, en caso de tenerlos? ¿Hay vínculos a los sitios web de dichos patrocinadores con el objetivo de aprender más de ellos?

3.2.4. Actualización

La actualización de una fuente hace referencia a la renovación y a la puesta al día de la información (Cooke 1999: 63). Para determinar el grado de actualización de una fuente, se comprobarán las fechas de creación y de la última actualización del sitio web.

3.2.5. Audiencia

La audiencia es el lector meta, la persona para la que el autor escribe (Pearson 1998: 61).

Para nuestra investigación, seguimos la terminología adoptada por Pearson (1998: 35-39) que distingue entre *expert to expert communication* (comunicación entre especialistas), *expert to initiates* (comunicación entre especialista-principiante), *relative expert to the uninitiated* (comunicación entre (semi-)especialista y lego) y *teacher-pupil communication* (comunicación profesor-alumno). Hemos aunado estos tipos de comunicación en tres, a saber, *comunicación especializada*, *comunicación semiespecializada* y *comunicación divulgativa.*

La comunicación especializada es la que alberga la mayor densidad de términos especializados. El lenguaje usado diverge notablemente del lenguaje general. Contiene un vocabulario altamente especializado y los términos se usan de forma precisa, por lo que no se ofrece ninguna explicación de la terminología a menos que se esté redefiniendo un concepto existente o acuñando un término nuevo. Este entorno comunicativo experto es en el que encajan la mayoría de los textos incluidos en nuestro corpus. Está representado especialmente por las publicaciones de revistas especializadas, libros académicos y proyectos de investigación.

La comunicación semi-especializada se da cuando los expertos se comunican con otros que poseen conocimientos del campo, pero que no alcanzan el mismo nivel de profundidad. La densidad de términos es menor ya que se incluyen explicaciones de algunos términos que resultan problemáticos o desconocidos. En nuestro corpus, este tipo de comunicación lo representan principalmente los textos extraídos de las enciclopedias de carácter general o de los libros de texto de materias específicas.

La comunicación divulgativa contiene, en general, una cantidad muy inferior de términos especializados en comparación con los otros dos tipos de comunicación. En este tipo de comunicación no se presupone ningún conocimiento especializado, simplemente un buen dominio de la lengua de llegada. Este enfoque se da especialmente en revistas de ciencias divulgativas como *New Scientist* o en las columnas de algunos periódicos.

3.3. Diseño

"La presentación y la disposición de la información en la pantalla pueden influir en la facilidad con la que se asimila la información" (Cooke 1999: 72). Atendiendo al parámetro *diseño*, nos centramos en las *ayudas a la navegación*, *accesibilidad* y *presentación y gestión de la información*.

3.3.1. Ayudas a la navegación

Las ayudas a la navegación son elementos que ayudan al usuario a localizar la información en un sitio web y le permiten moverse fácilmente de página en página por el sitio (Alexander y Tate 1999: 50). Las ayudas a la navegación que hemos tenido en cuenta son los mapas del sitio o índices, los enlaces, el título, la disponibilidad de un motor de búsqueda interna en el sitio web y de una sección de ayuda.

Es importante comprobar si el sitio web dispone de un mapa del sitio o índice en la página de inicio o en una página directamente vinculada a la página de inicio. Alexander y Tate (1999: 52) aclaran que:

> *A site map is a display, often graphical, of the major components of a website. An index is a listing, often alphabetical, of the major components of a website. A site map or index provides a quick overview of the pages contained within the entire site, and each can be an important tool in determining the coverage of the site.*

Los hipervínculos también contribuyen a facilitar la navegación por el sitio. Así pues, es importante comprobar si en las páginas de un sitio web hay enlaces a la página de inicio, a las páginas situadas un nivel más arriba en la jerarquía —en caso de que un sitio web se organice por jerarquías—; y si se proporcionan "atajos" para navegar de forma más rápida.

En cuanto al título, es necesario comprobar si el título del navegador describe el contenido del sitio y si indica claramente el sitio web del que proviene la página, lo que puede conseguirse fácilmente mediante un logo. Asimismo es aconsejable que el título sea corto y único para el sitio.

3.3.2. Accesibilidad

En lo que respecta a la accesibilidad, es importante determinar si se puede acceder de forma rápida y fácil a la información. Además, en caso de que se precise un software adicional, el sitio debe proporcionar el vínculo correspondiente para poder descargar el software, así como las instrucciones pertinentes que nos guíen en la descarga y nos ayuden a comprender el funcionamiento del mismo. El coste por el acceso a la información es también un elemento que hay que tener en cuenta. A veces hay cargos por el acceso a la versión electrónica de documentos que se encuentran en papel, por lo que habrá que considerar si merece la pena pagar por acceder a la versión electrónica. Las restricciones de acceso, como los registros y contraseñas o la comprobación de ser miembro de una organización, pueden condicionar también la velocidad de acceso a un sitio web.

Otro factor que hay que tener en cuenta en la accesibilidad es la estabilidad del sitio, es decir, si es estable o, si por el contrario, cambia con frecuencia, en cuyo caso, debe proporcionarse automáticamente el enlace a la nueva dirección.

3.3.3. Presentación y gestión de la información

El criterio de *presentación y gestión de la información* incluye la presencia de gráficos e imágenes en movimiento en el texto y el valor que añaden al texto. Además, en este criterio consideramos si una fuente se presenta de forma lógica y clara. Si contiene anuncios, resultará necesario comprobar si se usan de forma apropiada o si, por el contrario, distraen al usuario e interfieren en el principal propósito de la página.

4. Plantilla para la evaluación de los recursos web

Basándonos en el protocolo descrito en la sección anterior, proponemos la siguiente plantilla para la evaluación de los recursos web de los que posteriormente se extraerán los textos que pasarán a formar parte de nuestro corpus de textos especializados en el dominio del Medio Ambiente.

Tabla 2. Plantilla para la evaluación de recursos web

PARÁMETROS DE EVALUACIÓN PARA SITIOS WEB	
AUTORIDAD	
Autor	
Reputación y experiencia	
Forma de contacto – email, dirección, teléfono – con la organización, empresa o persona responsable del sitio	□ Sí □ No
Feedback – sugerencias, quejas, peticiones – para el autor	□ Sí □ No
URL (Dominios)	□ Personal: *~, %, users, .members, people, .name.* □ General: *.com, .mil, .net, .edu, .gov, .org, .int.* □ Dominios territoriales o geográficos: - □ Países de habla inglesa - □ Países de habla hispana - □ Países de habla alemana
Copyright	□ Sí □ No
CONTENIDO	
Cobertura	□ Información considerable y de validez atendiendo al tema
Precisión	□ Libre de errores gramaticales, tipográficos y de deletreo □ Referencia a otras fuentes de información
Objetividad	□ Evidencia del punto de vista del autor u organización responsable del sitio □ Anuncios - □ Influencia del anunciante sobre el contenido de la información □ Patrocinadores corporativos o sin ánimo de lucro claramente citados - □ Existencia de vínculos a los sitios de los patrocinadores corporativos o sin ánimo de lucro con el objetivo de poder aprender más de ellos
Actualización	□ Fecha de creación del sitio web _____ □ Fecha de la última actualización _____

(Cont.)

PARÁMETROS DE EVALUACIÓN PARA SITIOS WEB	
Audiencia	□ Comunicación especializada □ Comunicación semiespecializada □ Comunicación divulgativa
DISEÑO	
Ayudas a la navegación	□ Mapa del sitio o índice en la página de inicio □ Hipervínculos: - □ Vínculo a la página de inicio - □ Para sitios dispuestos jerárquicamente, vínculo a la página situada un nivel más arriba en la jerarquía - □ Atajos disponibles □ Motor de búsqueda interna □ Sección de ayuda □ Título del navegador: - □ Indica claramente la fuente del sitio de la que proviene - □ Describe claramente los contenidos de la página - □ Es corto y único para el sitio
Accesibilidad	□ Facilidad de acceso □ En caso de que se necesite un software adicional, facilidad de acceso para descargarlo □ Los vínculos funcionan bien □ Restricciones de acceso (registro, contraseñas, etc.) □ Coste de acceso □ Estabilidad del sitio □ Si cambia, se proporciona la transferencia automática al nuevo sitio
Presentación y gestión de la información	□ Existencia de gráficos e imágenes en movimiento que añaden valor al texto □ Presentación y disposición lógica y clara □ Anuncios: - □ Usados de forma apropiada - □ Distraen al usuario

5. Ejemplo de una página web de calidad

La presente sección describe un ejemplo de un sitio web de calidad que reúne la mayoría de los parámetros de evaluación considerados en la plantilla. El sitio pertenece a INTUTE, un servicio web gratuito que proporciona acceso a los mejores recursos para la educación y la investigación. El servicio ha sido creado por una red de universidades del Reino Unido y sus socios. A día de hoy ya contiene 123020 entradas[6].

6. <http://www.intute.ac.uk> [Consulta: 20/06/09].

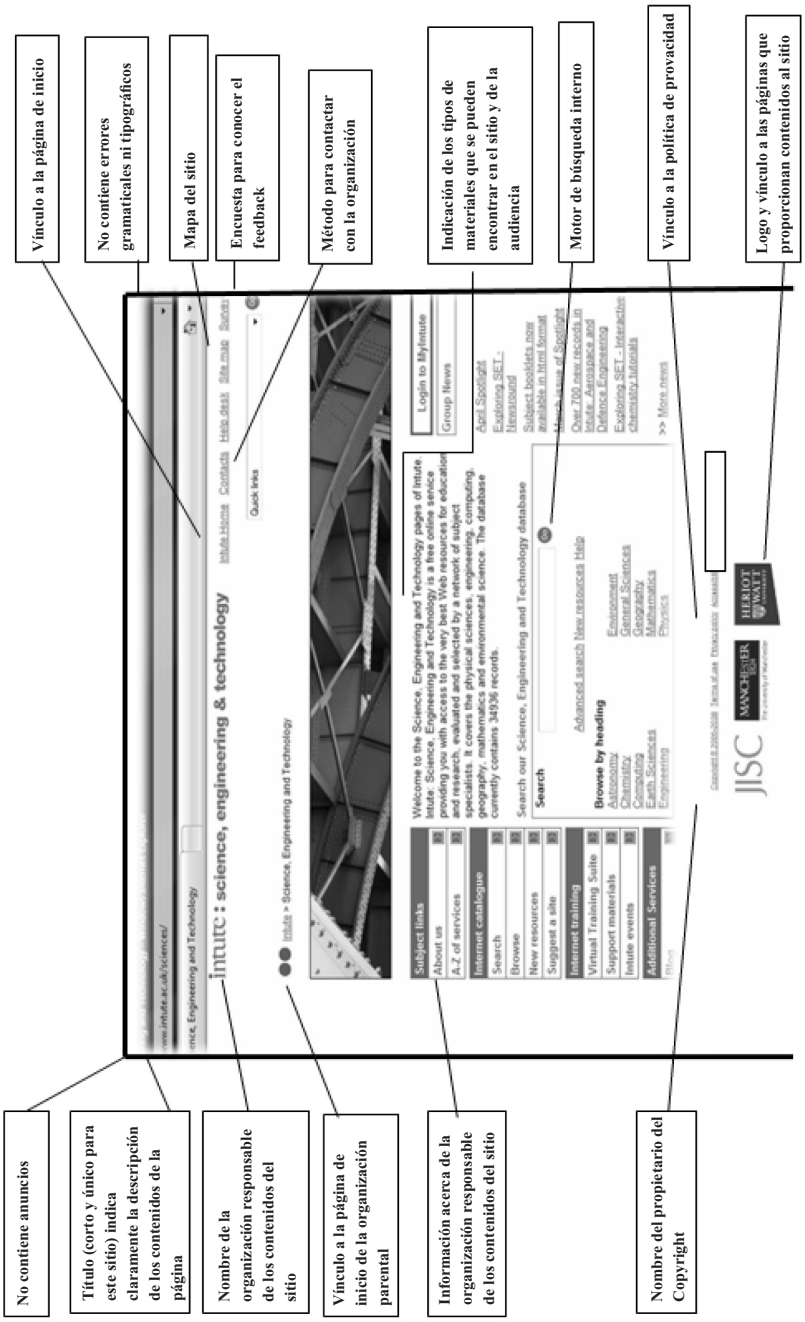

Ilustración 1: página web de calidad.

6. Conclusiones

El hecho de que Internet se haya convertido en la principal fuente de textos para la elaboración de un corpus significa que el concepto de calidad está directamente relacionado con la calidad de los sitios web. La presente investigación propone una serie de parámetros que hemos usado para evaluar la calidad de los sitios web a partir de los cuales obtendremos los textos para nuestro corpus. Nos hemos centrado principalmente en sitios web relacionados con la ingeniería medioambiental puesto que es el campo de estudio del proyecto de investigación Ecosistema.

Nuestro protocolo de evaluación se compone de los parámetros *autoridad*, *contenido* y *diseño*, cada uno de los cuales se subdivide en una serie de subparámetros. Atendiendo a la *autoridad*, evaluamos la reputación y experiencia de los autores o el dominio de la página web, entre otros. El parámetro *contenido* presta especial atención a la cobertura, precisión, objetividad, actualización y audiencia, mientras que el parámetro *diseño* se refiere especialmente a las ayudas a la navegación, accesibilidad y presentación y gestión de la información.

Este estudio confirma la importancia de seleccionar textos de calidad para la elaboración de un corpus representativo. Al aplicar este protocolo de evaluación a los textos, la calidad del corpus está asegurada, y por tanto, se asegura la calidad de la información conceptual, terminológica y colocacional disponible para el traductor. El presente protocolo puede hacerse extensivo para asegurar la calidad de corpus de otros dominios.

Bibliografía

Adreon, Heidi, Anne Catey y Kery Strysick (2002). An Educator's Guide to Credibility and Web Evaluation. Curriculum Technology Education Reform, University of Illinois [on line]. En <http://www.ed.uiuc.edu/wp/credibility-2002/index.html> [Consulta: 12/05/09].

Alexander, Janet E. y Marsha A. Tate (1999). *Web Wisdom. How to Evaluate and Create Information Quality on the Web*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Anderson, P. F., Nancy Allee, Steve Grove y Sara Hill (1999). Development of a Web Evaluation Tool in a Clinical Environment. University of Michigan [on line]. En <http://www-personal.umich.edu/~pfa/pro/courses/WebEvalNew.pdf> [Consulta: 25/06/09].

Auer, Nicole J. (1999). Evaluating Internet Information. University Libraries at Virginia Tech [on line].

En <http://www.lib.vt.edu/help/instruct/evaluate/evaluating.html> [Consulta: 25/06/09].

Austermühl, F. (2001). *Electronic Tools for Translators*. Manchester: St. Jerome.

Beck, Susan E. (2009). The Good, The Bad y The Ugly: or, Why It's a Good Idea to Evaluate Web Sources. New Mexico State University Library [on line]. En <http://lib.nmsu.edu/instruction/evalcrit.html> [Consulta: 25/06/09].

Beckwith, Robert (2005). Web Page Evaluation Form. Otsego High School's Department of English [on line].

En <http://www.bgsu.edu/downloads/enrollment/file76381.pdf> [Consulta: 15/05/09].

Codina, Lluís (2000). Parámetros e indicadores de calidad para la evaluación de recursos digitales. En *Actas de las VII Jornadas Españolas de Documentación. La gestión del conocimiento: retos y soluciones de los profesionales de la información*, 135-144. Bilbao, España.

Cooke, Alison (1999). *A guide to finding quality information on the Internet: selection and evaluation strategies*. London: Library Association Publishing.

Duke University Libraries. (2007). Evaluating Web Pages. Duke University [on line]. En <http://library.duke.edu/services/instruction/libraryguide/evalwebpages.html> [Consulta: 30/06/09].

Gaffney, Gerry (1998). Website evaluation Checklist v1.1. [on line].

En <http://www.infodesign.com.au/ftp/WebCheck.pdf> [Consulta: 02/07/09].

Gordon-Murname, Laura (1999). Evaluating Net Evaluators. *Searcher* 7(2), 57-66.

ISO 8402:1994 Quality management and quality assurance. Vocabulary.

Jiménez Piano, Marina y Virginia Ortiz-Repiso Jiménez (2007). *Evaluación y calidad de sedes web*. Gijón: Ediciones Trea, S.L.

Kapoum, Jim (1998). Five criteria for evaluating web pages. Cornell University Library [on line]. En <http://www.library.cornell.edu/olinuris/ref/research/webcrit.html> [Consulta: 05/06/09].

Kessinger, Pamela (2008). Web Page Evaluation Checklist. Portland Community College [on line]. En <http://spot.pcc.edu/lrc/pam/webcheck.htm> [Consulta: 15/05/09].

Kilgarriff, Adam y Gregory Grefenstette (2003). Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29(3), 333-347.

Kirk, Elizabeth E. (1996). Evaluating Information Found on the Internet. The Sheridan Libraries, Johns Hopkins University, Baltimore [on line]. En <http://www.library.jhu.edu/researchhelp/general/evaluating> [Consulta: 25/05/09].

Pearl K. Wise Library (2006). Web Evaluation Form. Cambridge, Massachusetts [on line].

En <http://www.cpsd.us/CRLS/Library/PDFs/WebEvaluationForm.pdf> [Consulta: 10/08/09].

Pearson, Jennifer (1998). *Terms in Context*. Amsterdam/ Philadelphia: John Benjamins.

Schrock, Kathleen (2006). Critical Evaluation of a Web Site. Discovery Education Classroom Resources [on line].

En <http://school.discoveryeducation.com/schrockguide/pdf/evalhigh.pdf> [Consulta: 10/05/09].

Sinclair, John (2005). Corpus and Text-Basic Principles. En *Developing Linguistic Corpora: a Guide to Good Practice*, Martin Wynne (ed.), 1-16. Oxford: Oxford Books [on line]. En <http://ahds.ac.uk/linguistic-corpora> [Consulta: 25/06/09].

Videon, Carol (1998). WWW Checklist, evaluation of sources. Delaware County Community College [on line].

En <http://faculty.dccc.edu/~cvideon/wwweval.htm> [Consulta: 26/05/09].

Wilkinson, Gene L., Kevin M. Oliver y Lisa T. Bennett (1997). Evaluating the quality of Internet Information sources. Department of Instructional Technology, University of Georgia [on line]. En <http://www.iicm.tugraz.at/thesis/cguetl_diss/literatur/Kapitel06/References/Oliver_et_al._1997/Evaluating%20_the_Quality.html> [Consulta: 14/07/09].