



1 de julio de 2016 | Vol. 17 | Núm. 7 | ISSN 1607 - 6079

ARTÍCULO

ELEMENTOS BÁSICOS PARA LA INDEXACIÓN DE ARTÍCULOS EN LA RED: EL CASO DE ScIELO

(<http://www.revista.unam.mx/vol.17/num7/art54/>)

*Antonio de Jesús Granados Cervantes
(DGTIC, UNAM)*

ELEMENTOS BÁSICOS PARA LA INDEXACIÓN DE ARTÍCULOS EN LA RED: EL CASO DE SciELO

Resumen

En este texto se explica de manera concisa cómo se puede potenciar la visibilidad de los artículos científicos mediante su indexación en hemerotecas o bibliotecas electrónicas, como en el caso de SciELO, así como la importancia de la marcación de los metadatos gracias al lenguaje XML.

“ “ Indexar es una tarea que [...] se ha ido difundiendo para que los autores y el público en general puedan enriquecerse haciendo un buen uso de la red ” ”

Palabras clave: SciELO, indexación, visibilidad, marcación, XML.

BASIC ELEMENTS

Abstract

This text explains in a concise matter how to potentiate the visibility of scientific articles by its indexation in electronic libraries, such as the SciELO Project. It also illustrates the importance of the metadata's SciELO markup with the XML language.

Keywords: SciELO, indexation, visibility, markup, XML.

ELEMENTOS BÁSICOS PARA LA INDEXACIÓN DE ARTÍCULOS EN LA RED: EL CASO DE SciELO

Introducción

En la actualidad, la tecnología ha crecido de una manera sorprendente. Tareas que antes nos parecían imposibles hoy tienen soluciones sencillas que cualquier persona puede llevar a cabo y que, incluso, se han hecho parte de nuestra vida cotidiana.

También es cierto que el avance tecnológico va en aumento y que cualquier producto que hoy nos sorprende y consideramos innovador el día de mañana puede quedar obsoleto. El crecimiento que ha tenido la tecnología ha enriquecido a muchas áreas, por ejemplo, en la medicina, en la educación, en las construcciones, en la ciencia, entre otras.

Los textos digitales en la actualidad

Hoy día la publicación y consulta de artículos en la red se ha hecho una tarea cotidiana. A veces ya no es suficiente tener la versión impresa de algún libro o revista científica o cultural, pero es particularmente debido a la creación de dispositivos móviles, como los teléfonos inteligentes y las tabletas, que nos vemos obligados a adaptarnos a ellas.

La sociedad se está acostumbrando cada vez más a la innovación tecnológica, y la mayor parte de la población con acceso a internet que está percibiendo este cambio son los jóvenes. Cada vez ocurre más que, desde una edad temprana, se acostumbran a utilizar distintos dispositivos, sobre todo en la escuela y en el hogar.

A esta gran parte de la población es a quienes van dirigidos, en mayor medida, estos textos digitalizados, pues es más fácil fomentar el gusto de la lectura y de buscar información en textos científicos mediante el uso de estos dispositivos móviles.

Ante esto surge el interés no sólo de consultar textos, sino de que los artículos que publicamos sean también consultados y tengan buena visibilidad, pero ¿qué se necesita para esto? Al hablar sobre indexación, nos referimos a la visibilidad que tienen los objetos digitales en la red (entre ellos, los artículos), pues en ella existen millones de éstos y a nosotros nos interesa que nuestros artículos no se pierdan en la red.

El objetivo de la indexación es seleccionar y remarcar todos esas palabras importantes que conforman el artículo científico. Dentro de la red, estas palabras son conocidas como metadatos. Los desarrolladores de software o programadores utilizan estos metadatos para guardarlos dentro de una gran base de datos, con lo que permiten que los artículos no se pierdan o queden muy rezagados de otros que estén contenidos en la red. Además, también permiten que un artículo pueda ser enriquecido mediante algún otro, entre muchas otras tareas.

En otras palabras, podemos hacer que nuestros artículos sean de fácil acceso gracias a los metadatos, los cuales son palabras claves dentro de nuestros artículos o revistas científicas, que además de enriquecer nuestras publicaciones ayudan a los buscadores a localizar nuestros artículos dentro de todos los demás.

Entre más metadatos, mayor visibilidad tendrá el artículo. Si bien es cierto esta tarea no era nada fácil, es justamente gracias al crecimiento de la tecnología que esta tarea se ha vuelto cada vez más sencilla. Antes, los editores necesitaban de alguien que tuviera conocimientos en programación o en sistemas. Actualmente se ha estado trabajando para que puedan realizar estas tareas por sí mismos, tarea que no ha sido sencilla.

Indexando artículos con SciELO

Foto: phsymyst

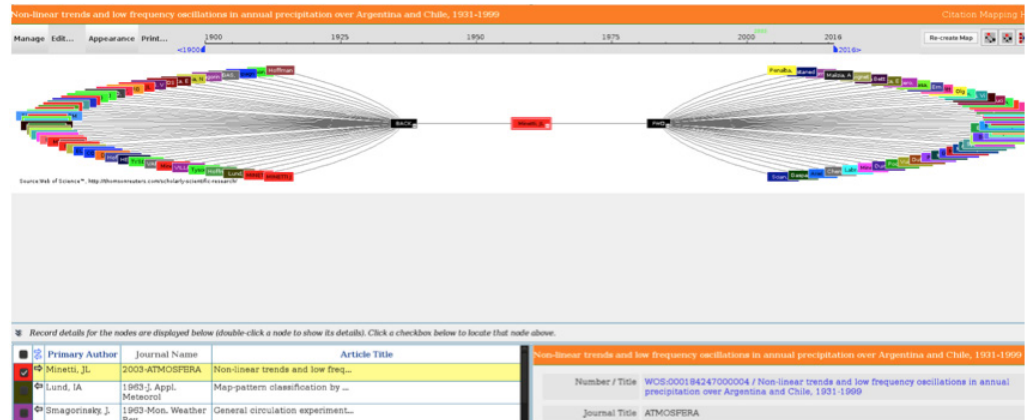
SciELO (Scientific Electronic Library Online) es un proyecto que nació en 1997 a partir de una iniciativa conjunta entre la FAPESP (Fundación de Apoyo a la Investigación del Estado de São Paulo) y la BIREME (Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud), a través de la cual se desarrolló una metodología común para la preparación, almacenamiento, diseminación y evaluación de la literatura científica en formato electrónico (SciELO México, s.f.).



SciELO es una biblioteca electrónica conformada por una red de colecciones de revistas científicas en texto completo, de acceso abierto y gratuito. Ayuda a la divulgación e incrementa la visibilidad de publicaciones científicas.

El SciELO Citation Index (SciELO CI), que forma parte de la plataforma del Web of Science (WoS), comenzó a operar a partir de enero de 2014. Este índice "comparte las mismas funciones, recursos y la navegabilidad de la interfaz de WoS, en conjunto con las otras bases de datos que integran la plataforma WoS. El desarrollo de SciELO CI es producto de la asociación del Programa SciELO/FAPESP con Thomson Reuters, propietaria de la plataforma WoS" (PACKER, 2014).

¿Qué obtiene una revista incluida en SciELO CI? De acuerdo con Álvarez Espinosa (s.f.), "La contabilización de citas (informe y mapa de citas) recibidas por la revista así como por los artículos publicados en ellas, provenientes de todas las revistas incluidas en las bases de datos del WoS, incluyendo la Colección Principal. La plataforma WoS ofrece además diferentes reportes bibliométricos".



Con esta finalidad, desde el 2012 ha estado promoviendo el uso del lenguaje XML (eXtensible Markup Language) para la estructuración de los textos completos y a partir del 2015 se adoptó para la operación de todas las revistas de SciELO.

Las ventajas de usar XML

¿Qué ventajas trae el uso del lenguaje XML? Es un lenguaje (o lenguajes) o metalenguaje que permite la definición de etiquetas para la especificación de partes significativas o de importancia para un texto, por ejemplo, el título, los autores, elementos gráficos como tablas e imágenes, entre otros. Asimismo, el uso de XML ayuda a los editores a la fácil detección de errores en sus mismos artículos.

A nivel computacional, aunque puede ser complejo el desarrollo de algoritmos que se encarguen leer estos archivos, se compensa con la rapidez que se obtiene. Esto es debido a la estructura de las etiquetas que estos archivos poseen.

El XML se utiliza para la estructuración de todos los elementos que tienen los textos en los procesos de edición contemporánea de los artículos y otros tipos de documentos. Cada elemento está definido por una etiqueta o tag. Por ejemplo:

```
<title>Cálculo diferencial e integral</title>
```

Es así como de textos marcados se pueden extraer todos estos metadatos que los componen y construir su referencia bibliográfica, es decir, por medio de esta extracción aseguramos que la referencia bibliográfica es fiel al texto del artículo para evitar errores de transcripción, además de que se evita el uso de otros procesos para la creación de una referencia bibliográfica.

Otra característica que tienen los archivos XML es su capacidad de ser presentados en distintos formatos. Así, con el avance en la tecnología, SciELO aprovecha esta característica para presentar los artículos científicos en distintos dispositivos de lectura que existen actualmente, ya que las vistas son totalmente adaptativas. De esta forma no importa si la consulta se realizó desde una computadora o una tableta, o incluso desde nuestro teléfono inteligente, la información de los metadatos siempre se podrá ver y nada se perderá durante el proceso de visualización.



INTRODUCCIÓN
 La circulación en la capa superficial de los grandes giros es conducida principalmente por el viento (Huang y Renfrew 1994), que gira alrededor de los centros de alta presión llamados anticiclones. El movimiento vertical y horizontal en estos giros cumple un rol fundamental en el control de la interacción entre el océano y la atmósfera, y es de gran importancia para la comprensión de la circulación oceánica general y la variabilidad climática a diferentes escalas de tiempo. En la cuenca del Pacífico Sur, el Anticiclón Subtropical del Pacífico Sudeste (ASPS) es el forzamiento dominante del giro subtropical, el cual en su flanco derecho está constituido por el sistema de las corrientes del Pacífico sudeste conformado por la corriente de choque de Humboldt, corriente oceánica de Humboldt, corriente costera de Humboldt, contracorriente Perú-Chile, corriente subsuperficial Perú-Chile, corriente costera de Chile, corriente del Pacífico Sur y corriente ecuatorial del Sur (García et al. 2004, Penland et al. 2004).

El ASPS tiene ciclos estacionales, interanuales y decadales (interdecadales). A escala estacional, en el verano austral, el ASPS es más intenso frente a la costa centro-sur de Chile (aproximadamente 35° S) y en invierno es más intenso frente a la costa de La Serena (norte de Chile, aproximadamente 30° S) (García et al. 2004). Frente a las costas del Perú, el ASPS provoca vientos favorables a la surgencia que son más intensos en el invierno que en el verano. Estos vientos paralelos a la costa adectan agua hacia el norte y ocasionan un transporte de Ekman costa afuera, y se genera un consecuente afloramiento de estratos más profundos (García et al. 2004). El patrón estacional de los vientos tiene una variabilidad baja entre los 18° y 28° S (Huang et al. 1994), mientras que entre los 30° y 40° S, esta variabilidad se incrementa (Fig. 1) e incluso promueve hundimiento durante el invierno (García et al. 2004, Penland et al. 2004). Cuando el ASPS se fortalece, se incrementa la magnitud de los vientos provenientes del sur y, por tanto, hay un mayor transporte de aguas del polo y más eventos de surgencia (García et al. 2004). Por el contrario, al debilitarse el ASPS, los vientos provenientes del sur decrecen en magnitud y la incidencia del transporte de Ekman es menor.

Figura 1
 Magnitud (color) y dirección (vectores) del promedio anual del esfuerzo del viento en el Pacífico oriental con base en QuikSCAT (Pillay y Bentamy 2002). Las elipses muestran la varianza de la climatología mensual del esfuerzo del viento. Los contornos muestran el promedio anual de la presión del aire a nivel del mar (hPa) con base en NCEP/NCAR (Kalnay et al. 1996). La escala para la dirección del esfuerzo del viento y las elipses de varianza se muestran en la esquina superior derecha. La escala de la elipse representa un eje mayor de 0.03 Pa y un eje menor de 0.015 Pa, y la elipse rota 90° desde el norte. La A representa el centro anual del Anticiclón Subtropical del Pacífico Sudeste.

Referencias:
 Huang y Renfrew 1994
 Ventilation of the Subtropical North Pacific. J. Phys. Oceanogr., 1994
 Gubb et al. 1998
 Coastal ocean circulation off western South America. Coastal Ingress. The Sea, 1998
 Penland et al. 2004
 Satellite altimetry data reveal jet-like dynamics of the South Pacific. J. Climate
 Talbot 1972
 Synoptic meteorology of the Southern Hemisphere. Meteorology of the Southern Hemisphere, 1972
 Chiquet et al. 2007
 On differences in satellite wind products and their effects in estimating coastal upwelling processes in the south-east Pacific. Geophys. Res. Lett., 2007
 García et al. 1994
 Variabilidad del viento, nivel del mar y temperatura en la costa norte de Chile. Invest. Mar., 1994

En sí, el uso de XML no tiene límite y se puede definir cualquier lenguaje o estructura de marcación, pero se recomienda utilizar alguna estructura estandarizada. SciELO adoptó para este fin una norma internacional ya establecida llamada JATS (Journal Article Tag Suite), la cual tiene su origen en la Journal Archiving and Interchange Tag Suite creada por la National Library of Medicine en Estados Unidos, para marcar los textos de los artículos almacenados y publicados por PubMed Central (PMC). Las revisiones del conjunto de etiquetas PMC dieron origen a la JATS en la forma de una norma del National Information Standards Organization (NISO) de Estados Unidos, identificada como JATS: Journal Article Tag Suite, version 1.0 (ANSI/NISO Z39.96-2012) (PACKER et al., 2014).

Cabe mencionar que SciELO no utiliza JATS tal como fue desarrollado, pues, como su mismo nombre lo indica, XML es un lenguaje extensible, por lo que SciELO tomó JATS como su base y le agregó sus propias etiquetas. Éstas especifican los diferentes niveles de afiliación de los autores, incluyendo la universidad o facultad, por ejemplo. Agregó también una etiqueta que identifica la especificación de las agencias financiadoras de las investigaciones. Otra modificación importante es la de especificar detalladamente las referencias bibliográficas necesarias para el montaje de la base de datos bibliométricos.

La aplicación de la metodología de SciELO consta de 2 fases: la preparación de archivos para la marcación y la marcación XML. La primera fase consiste en darle formatos específicos de posición, fuente, estilo y tamaño a cada elemento del texto. La precisión que se le da a ésta fase es muy importante para darle un mayor nivel al reconocimiento automático a la marcación del XML.


La segunda fase es la marcación XML; ésta se realiza con un software especializado para dicha tarea, el cual es ejecutado en Office Word. Mediante su interfaz de usuario (GUI) permite la marcación de los elementos del texto con el que se está trabajando. SciELO provee la información necesaria sobre los formatos que cada elemento del texto debe llevar; también ofrece sus propias capacitaciones.

Finalmente, cabe mencionar que el uso de XML ayuda a la preservación digital, ya que conforme la tecnología siga creciendo el lenguaje XML seguirá siendo procesado, ya sea para almacenar, transferir o presentar información.

Conclusiones

Todos los artículos que hoy han sido indexados con SciELO han ganado gran visibilidad en la red desde cualquier parte del mundo, esto gracias al uso de JATS y la extensión que desarrolló el equipo de SciELO.

Es importante que los autores consideren indexar sus artículos en SciELO e incluso tomar las capacitaciones que ellos ofrecen, con el fin de que les quede más claro cómo generar estos archivos XML a partir de la información contenida en sus artículos. Además, es importante considerar que indexando sus artículos, preservan la información y enriquecen no sólo sus propios textos, sino los de muchos otros autores que pueden crear referencias del artículo.

Indexar es una tarea que poco a poco ha ido creciendo y que además se ha ido difundiendo para que los autores y el público en general puedan enriquecerse haciendo un buen uso de la red. 

Bibliografía

- [1] ÁLVAREZ ESPINOSA, José Luis, “Revista Nova Scientia en el Web of Science”, *Red Internacional de Investigación la Salle*, s.f. [en línea]: <http://www.riilsa.net/web/articulo.php?a=19>.
- [2] Equipo SciELO México, *Preparación de archivos para marcación XML*, 2016 [en línea]: http://www.scielo.org.mx/avaliacao/SciELO_Manual_XML_Preparacion_de_archivos.pdf.
- [3] PACKER, Abel L. “SciELO Citation Index en el Web of Science”, *SciELO en perspectiva*, febrero 28, 2014 [en línea]: <http://blog.scielo.org/es/2014/02/28/scielo-citation-index-en-el-web-of-science/>.
- [4] PACKER, Abel L. et al., “¿Por qué XML?”, *SciELO en perspectiva*, abril 4, 2014 [en línea]: <http://blog.scielo.org/es/2014/04/04/porque-xml/#.VDwJjvmSyPs>.
- [5] SciELO México: <http://www.scielo.org.mx/scielo.php>.