

Publicación científica de revistas electrónicas en formato EPUB

ALEXA RAMÍREZ VEGA

FELIPE ABARCA FEDULLO

Instituto Tecnológico de Costa Rica (TEC)



Resumen

Como respuesta a las nuevas demandas tecnológicas, los editores de revistas científicas se han dado a la tarea de incorporar las publicaciones científicas en distintos formatos digitales como PDF (formato más utilizado), HTML enriquecido, EPUB y XML. Esto con el objetivo de beneficiarse de las ventajas de cada uno de ellos, ya sea para aumentar la visibilidad en la web, generar indicadores de uso, mejorar la accesibilidad de los contenidos que faciliten la lectura a los usuarios. Por otra parte, en Costa Rica las universidades públicas se han dado a la tarea de fortalecer sus publicaciones periódicas mediante capacitaciones a editores en diversas herramientas y buenas prácticas editoriales. Esto ha hecho que algunos editores publiquen los artículos de las revistas en formatos variados, distintos al PDF, con el objetivo de aumentar la visibilidad, dar mayor valor agregado a la revista y adaptarse a las nuevas tendencias. De esta manera, surge la necesidad de identificar las revistas de Costa Rica que han incursionado en la publicación de otros formatos y especialmente en EPUB, el cual es un formato sencillo de generar y da la posibilidad de adaptarse a móviles, pero la generación del documento correcto, a simple vista, no implica que las máquinas (lectores de EPUB), la web y los motores de búsqueda los interpreten correctamente. Dado lo anterior, se realiza un estudio sobre los diversos formatos de publicación utilizados en las revistas de universidades públicas de Costa Rica, específicamente los contenidos disponibles en los portales de revistas de cada universidad. Además, se hace énfasis en el análisis de los contenidos en formato EPUB, con el objetivo de identificar problemas generales, posibles fallas de diseño o edición, así como establecer pautas y buenas prácticas para la publicación de contenidos en este formato.

Palabras clave

EPUB; publicación científica; revistas electrónicas.

Abstract

In response to new technological demands, editors of scientific journals have been given the task of incorporating scientific publications in different digital formats such as PDF (most used format), enriched HTML, EPUB and XML. This with the aim of benefiting from the advantages of each of them, either to increase the visibility on the web, generate usage indicators, improve the accessibility of contents that facilitate reading to users. On the other hand, in Costa Rica public universities have been given the task of strengthening their periodical publications through trainings to editors in diverse tools and good practices editors. This has led some publishers to publish journal articles in a variety of formats, other than PDF, with the aim of increasing visibility, giving added value to the journal and adapting to new trends. In this way, the need arises to identify the magazines of Costa Rica that have ventured in the publication of other formats and especially in EPUB, which is a simple format to generate and gives the possibility to adapt to mobile, but the correct generation of the document, at a glance, does not mean that the machines (EPUB readers), the web and the search engines interpret them correctly. Given the above, a study is carried out on the different publication formats used in the public university magazines of Costa Rica, specifically the contents available in the portals of magazines of each university. In addition, emphasis is placed on the analysis of content in EPUB format, with the aim of identifying general problems, possible design or editing failures, as well as establishing guidelines and good practices for publishing content in this format.

Keywords

EPUB; scientific publish; electronic journals.

Introducción

Hoy en día los ambientes digitales donde se desenvuelven los usuarios de Internet evolucionan cada vez rápido, esto impacta directamente a los creadores de contenidos digitales tanto comerciales como académicos. De esta manera, las publicaciones científicas en formato electrónico deben adaptarse a los diferentes ambientes, ajustando los documentos a nuevos formatos digitales, los cuales permitan satisfacer las necesidades de los

usuarios, las exigencias tecnológicas y nuevas tendencias de la web (Lux & Pérez, 2017).

De esta manera, ya no es suficiente publicar artículos científicos en formato PDF o HTML no enriquecido (carente de hipervínculos), las tecnologías móviles demandan la creación de formatos adaptables a todo tipo de dispositivo, sin importar el sistema operativo (software) o la terminal utilizada (hardware). Aunado a esto, los lectores digitales exigen contenido ágil, interactivo y que brinden mayor facilidades de consulta, almacenamiento y procesamiento de la información (Orosa & García, 2016).

Como respuesta a estas nuevas demandas tecnológicas, los editores de revistas científicas se han dado a la tarea de incorporar las publicaciones científicas en distintos formatos digitales como PDF (formato más utilizado), HTML enriquecido, EPUB y XML. Revistas como *Science* incluyen el texto completo de sus artículo en PDF y HTML, de esta manera, cada vez más publicaciones científicas publican sus contenidos en varios formatos con el objetivo de beneficiarse de las ventajas de cada de uno de ellos, ya sea para aumentar la visibilidad en la web, generar indicadores de uso, mejorar la accesibilidad de los contenidos que faciliten la lectura a los usuarios (Liberatore, Aparicio & Banzato, 2016).

Cada uno de estos formatos aporta valor agregado a la revista y los contenidos que se publican, facilitando la lectura, consulta, citación y hasta la generación de métricas convencionales y alternas. El formato PDF (Portable Document Format) es uno de los más utilizados en la publicación científica, ya que reproduce la apariencia exacta de la versión en papel de los documentos. Por su parte el HTML (HyperText Markup Language), es el lenguaje utilizado para marcar documentos en el web, es muy sencillo y permite describir hipertextos con enlaces que conducen a otros documentos o fuentes de información relacionadas, y con inserciones multimedia limitada (Travieso Aguiar, 2003).

El EPUB (Electronic PUBlication) es un formato de documento estándar orientado a su visualización en dispositivos tipo libro electrónico (e-book). Este formato presenta independencia del contenido y la forma. Además, se adapta a los diferentes tamaños de pantalla y fuentes, también se le llama "texto líquido". Esto permite su visualización en cualquier dispositivo redimensionando el contenido de manera dinámica, lo cual hace posible la interactividad mediante hipervínculos, y en su versión más reciente (EPUB 3)

se permite incrustar galería de imágenes y hasta elementos multimedia (De Giusti, Lira, Rodríguez Vuan & Villarreal, 2016).

El formato XML (Extensible Markup Language) es un lenguaje de marcado, o un meta-lenguaje (lenguaje que describe otros lenguajes), el cual hereda algunas características de HTML y describe el contenido de los documentos almacenados en formato electrónico. La característica más significativa de XML es el concepto de DTD (Document Type Definition), que especifica la jerarquía lógica de los documentos y puede facilitar la recuperación de información en la web (Altamura, Esposito, & Malerba, 2001).

En este sentido, en el 2000 se inició el funcionamiento de PubMed Central (PMC), repositorio digital desarrollado por el Centro Nacional para la Información (NCBI), el cual almacena artículos académicos en texto completo. Ahí nació la necesidad de crear estándares que permitieran la publicación de contenido académico en la web por medio de XML (Huh, Choi, & Kim, 2014). Esto surge con el objetivo de solventar las carencias de otros formatos existentes, separar el contenido del formato y reutilizar el contenido de manera ágil y eficiente. Para esto PMC creó en 2012 el estándar internacional JATS v1.0 (Journal Article Tag Suite), el cual se convirtió en el estándar internacional para marcaje de artículos científicos más reconocido (Huh, Choi, & Kim, 2014). Debido a lo anterior, otros sistemas de publicación adoptaron el estándar JATS (con algunas modificaciones) con el objetivo de sacar provecho de este tipo de publicación, tal es el caso de Scielo y Redalyc.

Todo esto da grandes posibilidades de publicación digital a los editores de revistas científicas, pero también supone mayor esfuerzo, conocimiento y dedicación para incursionar en la publicación en diversos formatos. Cada formato de publicación digital es diferente y tiene sus lineamientos o recomendaciones que permiten crear contenidos óptimos desde diversos programas como: Indesign, Scribus, procesadores de texto (Microsoft Word o LibreOffice), entre otros.

Por otra parte, en Costa Rica las universidades públicas se han dado a la tarea de fortalecer sus publicaciones periódicas mediante capacitaciones a editores en diversas herramientas y buenas prácticas editores. Esto ha hecho que algunos editores publiquen los artículos de las revistas en formatos variados, distintos al PDF, con el objetivo de aumentar la visibilidad, dar mayor valor agregado a la revista y adaptarse a las nuevas tendencias. De esta

manera, surge la necesidad de identificar las revistas de Costa Rica que han incursionado en la publicación de otros formatos y especialmente en EPUB, el cual es un formato sencillo de generar (ya que muchos programas permiten crearlos de manera fácil y sin conocimiento experto) y da la posibilidad de adaptarse a móviles, pero la generación del documento, a simple vista, correcto no implica que las máquinas (lectores de EPUB), la web y los motores de búsqueda los interpreten correctamente. Dado lo anterior, se realiza un estudio sobre los diversos formatos de publicación utilizados en las revistas de universidades públicas de Costa Rica, específicamente los contenidos disponibles en los portales de revistas de cada universidad. Además, se hace énfasis en el análisis de los contenidos en formato EPUB, con el objetivo de identificar problemas generales, posibles fallas de diseño o edición, así como establecer pautas y buenas prácticas para la publicación de contenidos en este formato.

Metodología

Para esta investigación se toma como base las revistas disponibles en los Portales de revistas de las universidades públicas de Costa Rica. Para cada revista se analiza el último número publicado de donde se toman los distintos formatos de publicación de ésta. De todas las revistas, se toman únicamente las que publican en formato EPUB, luego se descargan al menos dos documentos del último número en formato EPUB, que servirá como base para el análisis de su contenido y estructura.

Posteriormente, para cada documento EPUB descargado de cada revista se le realiza los siguientes análisis:

- Validación del documento EPUB por medio de herramienta EpubCheck¹. Esta es una herramienta de código abierto que valida los documentos EPUB. Este análisis valida si un documento cumple con las especificaciones del formato EPUB, tanto a nivel gramatical como ortográfico, lo cual garantiza que el documento funcione correctamente en los lectores de este formato. Si el documento no pasa la validación el sistema genera un informe de errores al final del proceso.

¹ Se utilizó la versión 4.0.2, disponible en <https://github.com/idpf/epubcheck>.

- Uso del documento EPUB en lectores de distintas plataformas y proveedores.
- Enlaces internos. Se verifica que los enlaces a elementos internos (notas al pie de página, citas, referencias a imágenes y cuadros, etc.) se encuentren activos y funcionando correctamente.
- Enlaces externos. Se verifica que los enlaces a fuentes externas se encuentren activos y funcionando correctamente (ausencia de enlaces rotos).
- Posición de imágenes. Los gráficos y figuras dentro del texto en formato líquido deben estar en una posición cercana al lugar donde se hace referencia a los mismos. Se debe evitar ubicar las imágenes al final de todo el texto.
- Espacios entre elementos (imágenes, tablas y textos). Se verifica que exista espacio entre los diferentes elementos (texto-imágenes, tablas-texto, entre párrafos). Se debe evitar la superposición entre elementos.
- Interactividad de tamaño de fuente. Esta es una de las características más importantes del formato EPUB, la cual permite aumentar o disminuir el tamaño de la fuente del texto, adecuando el documento a las necesidades del usuario. Se verifica que está opción funcione correctamente.
- Tabla de contenidos. Se verifica la existencia de la tabla de contenidos en el documento EPUB. Este elemento es fundamental para la navegabilidad del texto.
- Tipo de inserción de tablas. Se identifica la forma de inserción de las tablas en el documento, ya sea en forma de imágenes o texto tabulado.
- Tipografía incrustada y colores. Se verifica si la tipografía utilizada en el documento es visible en otras plataformas y si se encuentra incrustada.
- Contraste. Se verifica el contraste entre el color del fondo y la tipografía del documento, y si ésta es ajustable.

Finalmente, se recopilan los datos de las validaciones de manera que se agrupen los errores o fallas comunes y se logre identificar la fuente del problema y cómo solventarlo.

Resultados

De los datos extraídos de los Portales de revistas de las universidades públicas de Costa Rica se contabilizan 95 revistas académico-científicas, de

las cuales 28 publican en un formato diferente a PDF y de esas solo 10 publican en formato EPUB. Además, como se muestra en la figura 1, el formato más utilizado (diferente de PDF) es HTML.

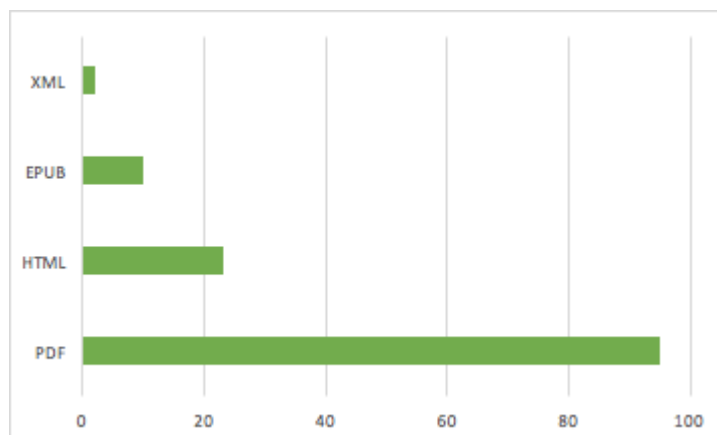


Figura 1. Formatos de publicación digital utilizados por las revistas

De las 10 revistas analizadas que publican en formato EPUB se tiene que solo tres de ellas pasó con éxito la validación de EpubCheck, donde se encontraron los siguientes errores o problemas considerados graves, ya que dificultan la lectura o acceso a los contenidos del documento:

- Documentos con URL rotos.
- Problemas mayores de código y etiquetado.
- Problemas en nomenclatura de imágenes y archivos.
- Nombres de imágenes con espacios.
- Nombres de imágenes con caracteres especiales.
- Nombres de hojas de estilo y otros archivos con espacios.
- Enlaces internos (notas al pie) perdidos.
- Metadatos incompletos, ausencia de metadato LANG y TITLE.
- Errores en enlaces a tablas de contenidos.
- Atributo ALT en blanco en etiqueta IMG.
- Errores de construcción de EPUB.

Por otra parte, se utilizaron dos lectores de EPUB, el iBooks y Adobe Digital Editions. Con iBooks sólo dos documentos fueron visualizados con errores mayores de diagramación o edición de los contenidos. Mientras que todos los documentos mostraron problemas en Adobe Digital Editions, ya sean errores menores y problemas graves como falta de interactividad y navegación invertida de los documentos. Además, con la visualización de

cada EPUB en los diferentes lectores de este formato, se encontró lo siguiente:

- Seis revistas tienen ausencia de tabla de contenidos en el documento EPUB, y una mostraba la tabla de contenidos, pero mal enlazada.
- En la biblioteca de los lectores, solo cuatro revistas mostraron la cubierta del artículo en la portada. Y dos de ellas no tienen cubierta del todo.
- Pies de página sin enlaces, URL externo no activo y enlaces externos rotos.
- Vista en dos páginas, no permite visualizar página por página el documento.
- Falta de interactividad de tipografía. No fue posible aumentar o disminuir el tamaño de la letra del texto.
- Falta de espacios después de tablas o figuras, esto genera la superposición de elementos.
- Imágenes al final del documento y desvinculadas totalmente del texto.
- Problemas de navegación, como enlaces que dirigen a otras partes del documento y no tienen retorno.
- Texto claro con fondo blanco, esto provoca bajo contraste.
- Las tablas se muestran en forma de imágenes, esto hace que el texto de la tabla no pueda ser navegable o rastreado.
- Problemas de estilos de textos. Se identificaron diversos tipos de letra en un mismo documento.
- EPUB generados con herramienta externas (convertidores).

De manera general, todos los documentos evaluados tienen problemas menores de diagramación-edición. En dos de ellos problemas mayores (muy graves) como navegación invertida e imágenes desplazadas al final del documento.

Conclusiones y recomendaciones

De manera general, los formatos de publicación de las revistas evaluadas siguen la tendencia internacional, donde mayormente se publica en formato PDF y en menor medida en EPUB y HTML. De esta manera, aunque la publicación electrónica implicaba, en un inicio, la reducción de costos de impresión; la publicación en diversos formatos digitales supone mayor

formación y recursos humano capacitado para generar los documentos en diversos formatos de manera adecuada y accesible para los lectores.

Además, según la revisión de los documentos y sus respectivos códigos se evidencia que cuatro de ellas utilizan programas especializados de diagramación que permite la exportación nativa a EPUB (como Adobe Indesign). Mientras que el resto muestra evidencias de uso de convertidores ajenos al programa de diagramación. Estos sistemas suelen incluir elementos adicionales (considerados basura) en el código que pueden crear problemas de visualización, navegabilidad y estructura del EPUB, lo cual ocasiona diversos fallos en los lectores.

Dado lo anterior, se ve la necesidad de capacitar a los encargados de generación de contenidos en EPUB sobre las buenas prácticas de edición y diagramación para este formato. Algunas recomendaciones generales son:

- Utilizar un programa especializado para edición y diagramación de contenidos digitales, que permitan exportar de manera nativa a EPUB.
- Probar y revisar cada elemento del EPUB resultante en diversos lectores de este formato. Es fundamental revisar detalladamente el documento generado, ya que algunos de los archivos evaluados muestran problemas que pueden ser detectados con una revisión visual simple.
- Hacer énfasis en la revisión y validación de enlaces internos y externos de todo el documento. Se debe probar cada enlace que el documento incluya y validar que dirija al lugar correcto.
- Validar la estructura del EPUB con una herramienta como el EpubCheck. Esto nos permite obtener una revisión más profunda del código fuente del EPUB y minimiza los posibles errores en algunos lectores de este formato.
- Asignar estilos apropiados a textos, tablas e imágenes. Esto permite que la generación de las hojas de estilos se realicen de manera coherente con la diagramación deseada.
- Uso apropiado de colores de fondo y texto, esto permite adecuarse al brillo y contraste de los diversos dispositivos de lectura (tabletas, computadoras, teléfonos inteligentes).
- Optimizar la calidad de resolución de imágenes. No sacrificar calidad de imágenes por disminuir el peso del archivo.
- Verificar la inclusión de una imagen de portada para cada documento. Este se muestra en la bibliotecas de los lectores.

- Todos los nombres de archivos externos (especialmente imágenes) no deben incluir espacios o caracteres especiales, los cuales generan errores en el código.

Finalmente, como trabajo futuro sería importante realizar un estudio sobre el uso de EPUB a nivel Latinoamericano y la evaluación de estos documentos, así también incluir validación de la accesibilidad de EPUB de las revistas científicas como la realizada en De Giusti, Lira, Rodríguez Vuan & Villarreal, (2016).

Referencias

- ALTAMURA, O., ESPOSITO, F., & MALERBA, D. (2001). Transforming paper documents into XML format with WISDOM++. *International Journal on Document Analysis and Recognition*, 4(1), 2-17.
- DE GIUSTI, M. R., LIRA, A. J., RODRÍGUEZ VUAN, J. P., & VILLARREAL, G. L. (2016). Accesibilidad de los contenidos en un repositorio institucional: análisis, herramientas y usos del formato EPUB. *e-Ciencias de la Información*, 6(2). Disponible en <<http://sedici.unlp.edu.ar/handle/10915/53638>>
- HUH, S., CHOI, T. J., & KIM, S. H. (2014). Using Journal Article Tag Suite extensible markup language for scholarly journal articles written in Korean. *Science Editing*, 1(1), 19-23.
- LIBERATORE, G., APARICIO, A., & BANZATO, G. (2016). *Manual de gestión editorial de revistas científicas de ciencias sociales y humanas. Buenas prácticas y criterios de calidad*. REUN: Buenos Aires. Disponible en <<http://www.memoria.fahce.unlp.edu.ar/libros/pm.482/pm.482.pdf>>
- LUX, M., & PÉREZ, A. (2017). Reflexiones sobre la producción, circulación y uso de las publicaciones académicas en las ciencias sociales. *Anuario Colombiano de Historia Social y de la Cultura*, 44(1), 125-143.
- OROSA, B. G., & GARCÍA, X. L. (2016). El ebook busca en la lectura social la propuesta que impulse nuevos formatos de éxito. *Anales de Documentación*, 19(2).
- TRAVIESO AGUIAR, M. (2003). Las publicaciones electrónicas: una revolución en el siglo XXI. *acimed*, 11(2), 1-2.