Creation of accessible EPUB documents by non-technical users: a long way to go

Jordi Roig
Student at Universitat de Barcelona
c/Melcior de Palau, 140
08010 - Barcelona
+34 934034017
jordi.roig@icloud.com

Mireia Ribera
Universitat de Barcelona
c/Melcior de Palau, 140
08010 - Barcelona
+34 934034017
ribera@ub.edu

1. ABSTRACT
This paper presents a quantitative evaluation of the conversion results from office documents to the EPUB format. As a general conclusion, the results of the conversions do not meet the standards nor preserve the accessibility information. Furthermore, tools are overdue in adopting the format EPUB3.

Categories and Subject Descriptors
I.7 Document and text processing, I.7.2 Document preparation, I.7.4 Electronic publishing

General Terms
Documentation

Keywords
Document conversions, EPUB3, Calibre, Sigil, digital accessibility, ATAG 2.0

2. INTRODUCTION
EPUB is a format to encode electronic documents, its version 3, EPUB3, has been designed considering the different ways to access information, with an inclusive vision suitable to everyone [2], from DAISY's format experience.

To create an EPUB3 document, in the publishing industry, usually a structured XML document and XSL transformations are used. In addition, this process is also suitable for the creation of other versions of the same content such as PDF, MOBI, etc. Nevertheless, academic authors and general public mostly use office tools and a posterior conversion to create EPUB documents.

3. PURPOSE
To evaluate the suitability of creating an EPUB document in accordance to the standards and accessibility good practices through office tools with a posterior conversion or directly through WYSWYG tools, but always without knowledge of code.

4. METHODOLOGY
To assess the validity of the resultant EPUB, firstly a selection of the most adopted mainstream author tools is carried out. Then, the elements to analyse are chosen and a document for each element is created with every selected tool.

4.1 Author tools selection
The main criterion for selecting the tools is its availability and high rate of adoption. Thus *MS Word* for Windows; *Pages* for OS X; and *LibreOffice Writer*, cross-platform, are selected. The EPUB cross-platform editor *Sigil* is also selected as it includes a text interface.

Solely *Pages* incorporates a converter to EPUB format. *MS Word* and *LibreOffice Writer* must use an add-in or external tools.

The add-in *Writer2epub* –W2E- [http://extensions.openoffice.org/en/project/writer2epub] is selected for *LibreOffice Writer* and the popular [3] *Calibre* program [http://calibre-ebook.com/] for *MS Word* (available on Windows and OS X).

At last, a two-step conversion is evaluated also on *MS Word* of Windows. In the first step, the add-in *Save As Daisy* [http://www.daisy.org/project/save-as-daisy-ms-word-add-in] converts the document to Digital Talking Books (DTBook) format and, in the second step, the *Tobi* program [http://www.daisy.org/project/tobi] converts the DTBook to EPUB format. Both tools are created within the accessibility community [1] and are selected for its special attention in regards to semantics.

The outcome of all the quoted combinations, except for *Save As Daisy* and *Tobi*, is an EPUB2 format document.

4.2 Testbed
This research does not include metadata nor structural elements such as preface, prologue or others. Quotes, mathematical formulae and graphics are also omitted.

The most common elements are listed, weighted and classified in basic (20%), secondary (10%) and complementary (5%) elements [4] [5] according to its accessibility relevance.

4.2.1 Basic elements
Headings. For someone with visual or motor disabilities the headings are fundamental elements for navigating the document. In the case of people with cognitive disabilities headings enable a global vision of the document. Six levels of HTML headings are considered, from the h1 to h6. Each level corresponds to a 3,3% of the 20% assigned.

Images: Image accessibility is essential for people with visual disabilities to access the content. Four possible scenarios are contemplated: decorative images without descriptions; images that require alternative texts; images that require alternative text and epigraph; and images that require a long description or a summary. Each type of image corresponds to a 5% of the 20% assigned.

Lists: In the same way as in headings, lists provide a better navigation for people with visual and motor disabilities. On the other hand, they are a key resource for people with cognitive disabilities to present the content in an understandable way. Ordered, unordered and mixed (a combination of both) lists are

DOI: http://dx.doi.org/10.1145/2829875.2829926

analysed. Each type of list corresponds to a 6,7% of the 20% assigned.

Tables: Despite of being an unresolved element from the accessibility point of view, tables are considered a basic element since they are quite common elements in documents. Only simple tables with a header row, title and summary table are considered. Each element (well formed table structure, header, title and summary) corresponds to a 5% of the 20% assigned.

4.2.2 Secondary elements

Notes: They provide explanatory content to the document. Footnotes and endnotes are included. Each type of notes corresponds to a 5% of the 10% of points assigned.

4.2.3 Complementary elements

Text styles: Highlighting words and main sentences facilitate text understanding for people with cognitive disabilities and reading impairments. Bold, italic and underlined styles are considered. Each type of style corresponds to a 1,6% of the 5% assigned.

Links: Not as usual as in web sites since they will bring the user to an external application, outside the EPUB reader. Both links redirecting to web sites and links to emails are included. Each type of link corresponds to a 2.5% of the 5% of points assigned.

4.2.4 Test documents

For each one of the elements an independent document is created according to accessibility good practices [http://bd.ub.edu/grups/adaptabit/Interaccion2015/documents.zip]. The wording used is inspired in the fairy tale *The three little pigs*.

4.3 Evaluation

For every test document a theoretical EPUB is created in accordance to EPUB2 and EPUB3 specifications. The documents resulted of the conversion and the theoretical EPUB documents are compared both by their code and visually. The assessment is manual by the authors.

5. RESULTS

Table 1. Points assigned to conversions

	Headings 20%	Images 20%	Lists 20%	Tables 20%	Notes 10%	Text styles 5%	Links 5%	Total 100%
Pages	10 %	5 %	0 %	5 %	5 %	0.7 %	5 %	30.7 %
W2E	20 %	0 %	0 %	5 %	5 %	5 %	5 %	40 %
Sigil	20 %	0 %	20 %	0 %	0 %	1.5 %	5 %	46.5 %
Calibre	20 %	0 %	13.3 %	5 %	10 %	2.5 %	5 %	55.8 %
Save As Daisy+Tobi	20 %	15 %	20 %	15 %	5 %	5 %	5 %	85 %

5. CONCLUSIONS

Although the EPUB format has been widely adopted as a digital publishing format [6], it has not yet been adopted by the general public.

One cause may be that among the most common used office tools only *Pages* is able to export documents to EPUB. Furthermore, most of the add-in and external tools still convert to EPUB2 format, despite the fact that the EPUB3 version specifications included significant changes and was already approved in October 2011.

Results indicate that these conversions are not acceptable - especially when it comes to the accessibility criteria- and that they make impossible the creation of quality EPUB documents by the non-technical user.

The author of a document is not informed about the quality of the transformation beforehand and he does not receive any type of feedback during the conversion either. As a general rule the accessibility characteristics introduced in the office documents disappear during the conversion.

6. ACKNOWLEDGEMENTS

This research has been supported by the Spanish Ministry of Science and Innovation through the User Experience research project *InDAGuS-UX* (TIN2012-37826-C02-02).

7. REFERENCES

[1] DAISY Consortium "DAISY Pipeline 2 | DAISY Consortium". http://www.daisy.org/pipeline2

[2] Garrish, M. 2012. *Accessible EPUB 3: best practices for creating universally usable content*. O'Reilly.

[3] Goyal, K., 2015. "Calibre. Usage statistics" http://calibre-ebook.com/dynamic/calibre-usage

[4] Ribera, M., Pascual Almenara, A., Salse, M., Masip Ardévol, L., Granollers I Saltiveri, T., López González, J.M., Comas I Aleix, A.C. (2010). *Guia de contingut digital accessible.* Universitat de Lleida. http://diposit.ub.edu/dspace/handle/2445/29018

[5] Ribera, M., Salse, M., Alcalà Ponce de León, M., Splendiani, B., Centelles Velilla, M., Llerena, I., … Térmens i Graells, M. (2012, November 22). Plantillas para la creación de material docente accesible con herramientas ofimáticas y LaTex. CIUD (Congreso International Universidad y Discapacidad) Fundación ONCE. http://diposit.ub.edu/dspace/handle/2445/33169

[6] Zichuhr, K., 2014. E-reading rises as device ownership jumps. Pew Internet. http://www.pewinternet.org/files/2014/01/PIP_E-reading_011614.pdf