ULL | Universidad
de La Laguna
**Departamento de Filología
Inglesa y Alemana**

131051120 / 161061122 INGLÉS TÉCNICO

Brian Jones: Open XML Formats
Whitepaper summarizing the Office Open XML standard

Ecma has now published a 14 page whitepaper that does an excellent job of describing the Office Open XML standard and the different goals and challenges TC45 had over the past year while working on the spec. I highly recommend everyone interested in office file formats take a look:

http://www.ecma-international.org/news/TC45_current_work/OpenXML%20White%20Paper.pdf

Rather than describe the whitepaper in my own words, I figured I'd just leverage the introduction as it does an excellent job of summarizing the overall purpose of the whitepaper:

Office Open XML (OpenXML) is a proposed open standard for word-processing documents, presentations, and spreadsheets that can be freely implemented by multiple applications on multiple platforms. Its publication benefits organizations that intend to implement applications capable of using the format, commercial and governmental entities that procure such software, and educators or authors who teach the format. Ultimately, all users enjoy the benefits of an XML standard for their documents, including stability, preservation, interoperability, and ongoing evolution.

The work to standardize OpenXML has been carried out by Ecma International via its Technical Committee 45 (TC45), which includes representatives from Apple, Barclays Capital, BP, The British Library, Essilor, Intel, Microsoft, NextPage, Novell, Statoil, Toshiba, and the United States Library of Congress (1).

This white paper summarizes OpenXML. Read it to:

- Understand the purposes of OpenXML and structure of its Specification

- Know its properties: how it addresses backward compatibility, preservation, extensibility, custom schemas, subsetting, multiple platforms, internationalization, and accessibility

- Learn how to follow the high-level structure of any OpenXML file, and navigate quickly to any portion of the Specification from which you require further detail

I also really like the second section in the whitepaper titled "Purposes for the Standard," as it helps in dealing with a lot of the questions I've received over the past several months. I think at times folks still aren't clear on the reasons we created this file format in the first place and then passed ownership of it to Ecma international.

131051120 / 161061122 INGLÉS TÉCNICO

OpenXML was designed from the start to be capable of faithfully representing the pre-existing corpus of word-processing documents, presentations, and spreadsheets that are encoded in binary formats defined by Microsoft Corporation. The standardization process consisted of mirroring in XML the capabilities required to represent the existing corpus, extending them, providing detailed documentation, and enabling interoperability. At the time of writing, more than 400 million users generate documents in the binary formats, with estimates exceeding 40 billion documents and billions more being created each year.

The original binary formats for these files were created in an era when space was precious and parsing time severely impacted user experience. They were based on direct serialization of in-memory data structures used by Microsoft® Office® applications. Modern hardware, network, and standards infrastructure (especially XML) permit a new design that favors implementation by multiple vendors on multiple platforms and allows for evolution.

Concurrently with those technological advances, markets have diversified to include a new range of applications not originally contemplated in the simple world of document editing programs. These new applications include ones that:

- generate documents automatically from business data;

- extract business data from documents and feed those data into business applications;

- perform restricted tasks that operate on a small subset of a document, yet preserve editability;

- provide accessibility for user populations with specialized needs, such as the blind; or

- run on a variety of hardware, including mobile devices.

Perhaps the most profound issue is one of long-term preservation. We have learned to create exponentially increasing amounts of information. Yet we have been encoding that information using digital representations that are so deeply coupled with the programs that created them that after a decade or two, they routinely become extremely difficult to read without significant loss. Preserving the financial and intellectual investment in those documents (both existing and new) has become a pressing priority.

The emergence of these four forces – extremely broad adoption of the binary formats, technological advances, market forces that demand diverse applications, and the increasing difficulty of long-term preservation – have created an imperative to define an open XML format and migrate the billions of documents to it with as little loss as possible. Further, standardizing that open XML format and maintaining it over time create an environment in which any organization can safely rely on the ongoing

stability of the specification, confident that further evolution will enjoy the checks and balances afforded by a standards process.

Various document standards and specifications exist; these include HTML, XHTML, PDF and its subsets, ODF, DocBook, DITA, and RTF. Like the numerous standards that represent bitmapped images, including TIFF/IT, TIFF/EP, JPEG 2000, and PNG, each was created for a different set of purposes. OpenXML addresses the need for a standard that covers the features represented in the existing document corpus. To the best of our knowledge, it is the only XML document format that supports every feature in the binary formats.

Tom Ngo, the editor of the whitepaper did an excellent job of summarizing what I've tried to convey numerous times in this blog (with varying degrees of success I admit). The industry absolutely needs OpenXML, and we've heard this repeatedly from our customers. It doesn't make it the "file format to replace all file formats," and we would never make such ridiculous claims. Instead it's simply an open standard that helps serve very real customer needs.

-Brian

Published Tuesday, October 24, 2006 9:41 AM by BrianJones
Filed under: Office Open XML Standard, Tech Interop
http://blogs.msdn.com/brian_jones/archive/2006/10/24/whitepaper-summarizing-the-office-open-xml-standard.aspx

TASKS

1. Watch the following You Tube video clip: World First I.T. Supporter (http://www.youtube.com/watch?v=vnAeO1N-G0s) and then answer the questions below:

2. What is the gist of video? Make a brief summary in 30-50 words.

3. As you may see, this video is a parody of modern Computer Customer Services. Although some terms are not used the usual way, please make a list with as many computer terms as you can spot, with their equivalent into Spanish.

4. Look up the definitions of XML and OOXML (Office Open XML).

5. You should have started your Computer & Information Technology glossary. The following fields should be included: English term, Spanish equivalent, definition, related hyperlinks, synonyms and antonyms. If you want, you may expand these fields. You are advised to work in groups, and advance at a rate of some 50 new words per week. By the end of the semester, your glossary should include some 400-500 words.

6. Write down the outline underlying this text.