

Marco para el Intercambio de Documentos en Universidades Españolas mediante Estándares Abiertos

Conferencia de Rectores de las Universidades Españolas
Comisión de TIC

10 de julio de 2006

Índice general

I	Estándares abiertos en las universidades españolas	1
1.	Introducción	2
1.1.	Estándares abiertos	2
1.2.	El uso de estándares abiertos en la CRUE	3
1.3.	Iniciativas existentes de apoyo a los estándares en entidades públicas	4
2.	Formatos recomendados para universidades españolas	5
2.1.	Terminología	5
2.2.	Requisitos para la selección de formatos	5
2.3.	Formatos recomendados	6
II	Estudio de los formatos abiertos existentes	7
3.	Documentos de texto	8
3.1.	OpenDocument	8
3.2.	Formato XML de Microsoft	9
3.3.	RTF	10
3.4.	Texto simple	11
3.5.	HTML	11
3.6.	Lenguajes de marcado de texto: T _E X/L ^A T _E X, DocBook	12
3.7.	PDF	12
4.	Hojas de cálculo	14
4.1.	Valores separados por comas (CSV)	14
4.2.	OpenDocument	14
4.3.	Formatos XML de Microsoft	14
4.4.	PDF	15
5.	Presentaciones	16
5.1.	OpenDocument	16
5.2.	PDF	16
III	Material adicional	17
	Bibliografía	18

Resumen

En el presente documento se define un marco, basado en estándares abiertos, para el intercambio de información digital en ofimática (documentos de texto, hojas de cálculo, presentaciones) entre universidades.

En la primera parte se definen los estándares abiertos, justificando su uso para el intercambio de información en el entorno universitario. Se establecen los requisitos que deberían verificar los formatos utilizados y, en función de éstos se recomiendan, en la sección 2.3, una serie de estándares para el intercambio de información entre universidades: OpenDocument (ISO 26300) y PDF/A (ISO 19005). Esta sección es, por tanto, el apartado fundamental del documento.

La segunda parte contiene un estudio detallado sobre formatos y estándares, en el que se basan las recomendaciones contenidas en dicha sección. El estudio consiste en un análisis de los formatos abiertos (o, al menos, con especificaciones públicas) más extendidos en la actualidad, junto a una evaluación del grado de adaptación a los requisitos previamente establecidos.

Parte I

Estándares abiertos en las universidades españolas

Capítulo 1

Introducción

En la pasada reunión del Grupo de Trabajo permanente de la CRUE [1] sobre tecnologías de la información y las comunicaciones (CRUE-TIC [2]) de 21 de febrero de 2006, en Oviedo, se acordó realizar un documento que definiera una primera versión (habrá lógicamente muchas con el paso del tiempo) de formatos para intercambio de documentos abiertos entre universidades españolas. El objetivo: realizar un estudio riguroso para su aprobación en la siguiente reunión de CRUE-TIC del 27 de junio en Tarragona.

La redacción fue asumida, en su vertiente técnica, por el grupo Iris-Libre, integrado por miembros de la comunidad de RedIRIS y en el que participan representantes de diversas universidades españolas. El presente documento es fruto de ello.

Su redacción se apoya en distintos estudios realizados con anterioridad en materia de estándares abiertos, especialmente en el estudio sobre formatos realizado por la Universidad de Cádiz [6].

Se permite la distribución del presente documento con licencia *Creative Commons Reconocimiento - CompartirIgual* [37]

1.1. Estándares abiertos

El uso de estándares ha sido vital para el desarrollo tecnológico de nuestra sociedad. Sin la adopción de un marco de trabajo común y neutral, no habría resultado posible la interoperación entre equipos tecnológicos de distintos modelos y fabricantes: especificaciones tan sencillas como el diámetro de un tornillo y tan complejas como las de un equipo de radio o televisión, pasando por aquellas que han hecho posible la existencia de Internet, todas ellas comparten un aspecto distintivo: la estandarización.

Un *estándar abierto* consiste en una serie de especificaciones públicas, completamente mantenidas por una entidad neutral y sin ánimo de lucro, cuya propiedad intelectual se ofrece de forma libre de regalías y sin restricciones en cuanto a su uso y reutilización.

En concreto, adoptaremos la definición de estándares abiertos de la comisión IDABC de la Comunidad Europea [33]:

“Las siguientes son las mínimas características que una especificación y sus documentos de apoyo deben tener para ser denominados estándares abiertos:

1. El estándar ha sido adoptado y es mantenido por una entidad sin ánimo de lucro, y su sucesivo desarrollo tiene lugar sobre la base de un proceso de decisión abierto a todas las partes interesadas (consenso o decisión por mayoría, etc.).
2. El estándar se ha publicado y el documento con la especificación del mismo se encuentra disponible de forma gratuita o bien por un precio simbólico. Se debe permitir a cualquiera su copia, distribución y uso sin cargo o con un precio simbólico;
3. La propiedad intelectual -por ejemplo, posibles patentes presentes- del estándar (o de alguna de sus partes) se ofrece de forma irrevocable libre de regalías (royalty-free basis);

4. No hay restricciones en cuanto a la reutilización del estándar.”

Y, con objeto de no dejar fuera de esta definición formatos y protocolos tan estandarizados y abiertos a lo largo de la historia de la informática como puedan ser PVM, NFS y RPC, entre otros semejantes que son propuestos por empresas, ampliaremos la definición anterior con la siguiente salvaguarda:

El punto 1 no será de obligado cumplimiento en el caso de que exista una implementación de referencia libre¹ del estándar completo (y solamente en ese caso).

Así, recordamos que los principales estándares abiertos informáticos han sido creados gracias a implementaciones de referencia libres. Ejemplos notorios de ello han sido TCP/IP con BSD, las primeras versiones de HTTP y HTML, o el SMTP con Sendmail, todos ellos núcleo vital de la actual Internet.

En el ámbito de las TIC, los estándares abiertos garantizan la interoperabilidad entre agentes emisores y receptores de la información, con independencia del software utilizado por ellos, evitando la dependencia de un único fabricante, fomentando la competitividad en el mercado (y con ello el ahorro de costes) e incentivando la neutralidad tecnológica y la innovación.

Por otra parte, la existencia de unas especificaciones públicas, abiertas y completas certifica el acceso en el futuro a la información generada en la actualidad, independientemente de los avatares que el destino haya designado para los programas que se utilicen hoy para su almacenamiento digital. La perdurabilidad de los documentos es un requisito especialmente significativo en el caso de las administraciones y entidades públicas, para los que la legislación actual exige su adecuada conservación durante periodos que pueden llegar a ser de varias decenas de años, una eternidad en términos de vida y perdurabilidad de programas y arquitecturas hardware de ordenador.

Por último, dentro del ámbito de las administraciones y entidades públicas, los estándares abiertos facilitan su interacción con la sociedad, sin imponer marcas o fabricantes de software y hardware concretos.

1.2. El uso de estándares abiertos en la CRUE

En tanto que entidades públicas, todas las consideraciones anteriores son aplicables al marco de las universidades españolas. En concreto, las universidades tienen la obligación de facilitar el acceso, por parte de alumnos, profesores y personal de administración y servicios, a cualquier tipo de información institucional que publiquen de forma electrónica. Por otra parte, si no se define un marco adecuado para la interoperabilidad entre universidades, estaremos obstaculizando el intercambio de documentos entre ellas. Por ejemplo:

- Cuando una unidad administrativa de una universidad determinada debe enviar un informe a una segunda, situada en una universidad diferente, ¿qué formato debe utilizar para tener la confianza de que la unidad receptora pueda acceder correctamente a la información?
- Cuando un profesor confecciona una presentación para un encuentro universitario, debe tener garantías de que la universidad en la que este se realice cuente con los programas adecuados para visualizarla. ¿En qué formato debe generar la presentación?

Una posibilidad para evitar este tipo de problemas es asumir que todas las universidades deben utilizar los programas de ofimática que lideran actualmente el mercado. En este caso, se asume la dependencia tecnológica por parte del conjunto de las universidades españolas de una empresa determinada, a la vez que la codificación de su información relevante en formatos cerrados a los que no se podrá acceder, ni hoy ni en el futuro, si esta empresa y la evolución del mercado así lo dispusieran.

Una segunda posibilidad consiste en acordar el uso de una serie de formatos neutrales, abiertos y estandarizados por organismos internacionales independientes, de tal forma que cualquier

¹Licencia libre: <http://www.opensource.org/licenses/>

programa elaborado por cualquier empresa, pueda leerlos y escribirlos en igualdad de condiciones, hoy y en el futuro. Esta opción es, sin duda, la más adecuada, aunque pueda presentar el inconveniente de que los programas de ofimática más utilizados en este momento no utilizan formatos estándares, por lo que se hace necesario utilizar otro tipo de programas, con el coste que esto pueda representar (fundamentalmente, costes de formación, si asumimos que el estándar tiene implementaciones de referencia libres).

Aun así, tras hitos como la estandarización de los formatos PDF/A (como ISO 19005) y de OpenDocument (como ISO 26300) y la existencia de programas capaces de leer y escribir este tipo de formatos estándares con garantías, junto a su diversidad y madurez, ya es factible el plantearse la conveniencia de asumir el coste para la migración a estos por parte de las universidades.

1.3. Iniciativas existentes de apoyo a los estándares en entidades públicas

En los últimos años, son varias las entidades y administraciones públicas que han expresado de forma oficial su apoyo a los estándares abiertos y que han legislado en el sentido de promover su uso de forma efectiva. La madurez del entorno de ofimática *OpenOffice.org* y la publicación de *Open Document* como estándar ISO han sido un hito muy importante para lo anterior. A continuación, se enumeran algunos de los casos más representativos:

El gobierno de Bélgica aprobó, en 2004, una serie de directivas y recomendaciones para el uso de formatos abiertos, así como modalidades para la migración a estándares [40].

El estado de Massachusetts (EEUU) fue el primer estado norteamericano en publicar, en 2005, una normativa para el uso de OpenOffice y formatos abiertos [39].

El gobierno de Dinamarca ha aprobado en junio de 2006 una resolución que supone el compromiso de utilizar estándares abiertos en la administración pública antes de que termine el año 2008 [41].

Dentro de nuestro país, existen algunas iniciativas locales, por ejemplo las siguientes:

La Universidad de Cádiz aprobó en el año 2004 una normativa que obliga a todos sus integrantes a comenzar a utilizar formatos abiertos a la hora de intercambiar documentos oficiales[4].

La Junta de Andalucía se ha comprometido a crear, antes del fin de 2006, una nueva norma para el uso obligatorio de estándares abiertos por parte de la misma, para avanzar así hacia una progresiva sustitución y abandono de los formatos y protocolos privativos [42]

Capítulo 2

Formatos recomendados para universidades españolas

2.1. Terminología

En el presente documento se empleará la siguiente catalogación:

Formato cerrado o propietario. Su especificación no está publicada, o lo está parcialmente, perteneciendo sus derechos a una o varias empresas que la mantienen oculta o incluso bajo leyes de secreto industrial. Habitualmente se requieren técnicas de ingeniería inversa para que otras herramientas lo utilicen. Ejemplos: DOC (MS-Word), XLS (MS-Excel), PPT (MS-PowerPoint), DWG (Autocad).

Formato abierto o estándar abierto. Sus especificaciones están totalmente publicadas y pertenecen a alguna organización sin ánimo de lucro, con el propósito de regular el formato. Suele existir una implementación de referencia que ayuda a desarrollar software (libre o no) totalmente compatible. Ejemplos: HTML 4.01 de W3C (ISO 15445), OpenDocument (ISO 26300), PDF/A (ISO 19005, generado a partir de la versión 1.4 del formato PDF publicado por Adobe).

Formato con especificaciones públicas o formato publicado. Sus derechos pertenecen a una o varias empresas, pero su especificación está publicada totalmente, existiendo la posibilidad de su uso sin restricciones por parte de otras empresas distintas. Para implementar algunas especificaciones, sin embargo, sí se presentan restricciones, en algunos mercados, debidas a patentes de software. En todo caso, este tipo de formatos tiene el problema de que su futuro depende, al igual que en el caso de los formatos cerrados, de los avatares y decisiones de la empresa propietaria. Ejemplos: RTF de Microsoft, especificaciones Postscript y PDF de Adobe.

2.2. Requisitos para la selección de formatos

A la hora de determinar que un formato es adecuado para el intercambio de información en el marco de las universidades españolas, se han utilizado dos baremos:

1. **El formato es un estándar abierto.** Por los motivos expuestos en secciones anteriores, se ha dado una especial relevancia a que el formato en cuestión pueda considerarse como un estándar abierto.
2. **Disponibilidad de programas lectores y generadores adecuados** para la generación y lectura de este formato, valorando las siguientes características:
 - a) **De buena calidad.** Los usuarios no deben verse forzados a utilizar programas de baja calidad, pues esto les supondría perjuicios obvios, disminuiría su rendimiento y su motivación.

- b) **Muy difundidos.** Es deseable evitar la necesidad de capacitación y adaptación a nuevas herramientas, así como generar a los destinatarios más trabajo del mínimo necesario.
- c) **Disponibles en las plataformas más usadas** en puestos de trabajo. Resulta imprescindible no limitar la posibilidad de elección de la plataforma del puesto de trabajo por parte los usuarios, especialmente en el caso del hogar de los alumnos y de los ordenadores del personal docente e investigador. Deben existir implementaciones adecuadas de las herramientas necesarias para las más utilizadas: MS Windows, GNU/Linux y Macintosh, en lo referido a sistemas operativos, y x86, AMD64 y PPC en lo referente a arquitecturas hardware.
- d) **Variados y accesibles, a ser posible con licencia libre.** De este modo, se garantiza la independencia de un único proveedor de software y con ello la posibilidad de elegir por parte del usuario que lo desee. Así, no se fuerza a los destinatarios a adquirir productos comerciales concretos, y no se fomenta la copia ilegal de software sino la competencia y la calidad.

2.3. Formatos recomendados

En virtud de los estudios realizados por distintas universidades españolas, debatidos en el seno de IRIS-Libre y recogidos en las secciones 3, 4 y 5, se establecen las siguientes conclusiones sobre documentos ofimáticos de tipo texto, hojas de cálculo y presentaciones:

Se recomiendan los siguientes formatos para el intercambio de información entre universidades:

- **OpenDocument** para documentos de texto, hojas de cálculo y presentaciones, especialmente en el caso de que deban ser, eventualmente, modificados por el receptor. Véase 3.1.
- **PDF** para documentos de texto, hojas de cálculo y presentaciones que sean de lectura exclusiva para el receptor, especialmente en el caso de o que deban persistir durante largo tiempo. Más explícitamente, se recomienda utilizar la versión ISO PDF/A (véase el apartado 3.7) o, siempre que esto no sea posible, la versión 1.4. Se rechazan, sin embargo, las funcionalidades avanzadas de las últimas versiones de PDF de Adobe.

Se aceptan los documentos de texto sencillo (véase 3.4), recomendándolos solamente para situaciones donde el formato visual sea muy poco importante, como mensajes de correo electrónico o borradores. Además, se acepta HTML/XHTML (véase 3.5) en versiones fijadas por el W3C [7], siendo, por supuesto, el formato que debe ser usado en páginas web. Obsérvese que algunas aplicaciones actualmente muy extendidas, como MS-Word o MS-Frontpage, no generan HTML estándar.

Se desaconsejan aquellos formatos que, aun siendo abiertos o publicados, estén poco extendidos entre los receptores finales, como Postscript, DjVu, DVI, DocBook, TEX/LaTeX (véase 3.6), así como aquellos formatos con especificaciones públicas para los que existan estándares abiertos que representen claras alternativas, como RTF (sección 3.3) y los esquemas de referencia XML de Microsoft [9]. Por último, se desaconsejan todos aquellos formatos cuyas especificaciones completas no estén disponibles públicamente, entre ellos los de MS-Office [8] (doc, xls, ppt...) y otros formatos propietarios binarios (wp o doc de WordPerfect, etc.).

Estas conclusiones se han obtenido como resultado del análisis realizado en la parte II, en función de los requisitos que fueron fijados en la sección 2.2

Parte II

Estudio de los formatos abiertos existentes

Capítulo 3

Documentos de texto

En esta sección se presentará un estudio de aquellos formatos para documentos de texto que pueden considerarse como estándares abiertos, o al menos aquellos cuyas especificaciones hayan sido publicadas por su propietario, estudiando tanto sus ventajas como sus puntos débiles.

En este estudio se excluyen algunos formatos de documentos muy relevantes, como el uso de forma nativa por las versiones actuales de Microsoft Office (.doc), el formato de ficheros de WordPerfect y otros. El motivo es que, al tratarse de formatos cerrados, no cumplen el primero de los dos requisitos enunciados en la sección 2.2 y ni siquiera pueden ser contemplados entre los formatos propietarios con especificaciones públicas.

El estudio no pretende ser exhaustivo pues, en función del segundo de dichos requisitos, el objetivo es atender a exclusivamente aquellos formatos que se consideren suficientemente extendidos. En primer lugar se estudian formatos de ficheros de texto expresamente orientados a la ofimática: OpenDocument y formatos Office Open XML y RTF de Microsoft. A continuación se estudian otros tipos de ficheros (texto sencillo, HTML, lenguajes de marcado...) que tienen un propósito más amplio. Por último se estudia PDF como exponente de formato para documentos perdurables no modificables y orientados a impresión.

3.1. OpenDocument

Se trata de un formato basado en XML cuyo origen está en el formato que usaba de forma nativa el programa OpenOffice.org. Durante varios años fue discutido públicamente, modificado, adaptado y, finalmente, estandarizado el 1 de mayo de 2005 por OASIS [22] (consorcio internacional sin ánimo de lucro orientado al desarrollo, la convergencia y la adopción de estándares en TICs). Su intención es que constituyera la base de un estándar para los formatos utilizados en las suites ofimáticas. Así, el 1 de mayo de 2006 fue aprobado como un estándar ISO (ISO/IEC 26300), siendo en la actualidad el único estándar internacional existente para documentos de texto orientados a la ofimática.

La extensión identificativa de los ficheros de texto OpenDocument es .odt.

Evaluación de requisitos

■ Requisito 1.

- Puesto que ha sido aprobado como estándar ISO, el formato OpenDocument puede definirse como un estándar oficial e internacional, cumpliendo por lo tanto el primero de los requisitos enunciados en la sección 2.2.
- Confirmando lo anterior, OpenDocument no es propiedad de una empresa de software concreta, sino de OASIS, un organismo de estandarización neutral y sin ánimo de lucro que aglutina a las mayores empresas y entidades del sector por lo que puede considerarse abierto al no existir limitaciones sobre su uso.¹

¹Entre sus integrantes están empresas como Microsoft, IBM, Sun, HP, Intel, Oracle, etc. Véase [23] y [24].

■ Requisito 2.

- Existen numerosas aplicaciones de calidad, tanto libres como comerciales privativas, que son capaces de leer y grabar ficheros utilizando este formato. Algunas de estas aplicaciones están disponibles para las plataformas más comunes. Por lo tanto, se verifican los requisitos 2a, 2c y 2d de la sección 2.2.
- En cuanto al requisito 2b (difusión de los programas que utilizan este formato), estos programas, aunque son los más extendidos que se basan en estándares abiertos, aún están poco difundidos, en términos relativos. A pesar de ello, es un formato emergente, del que se estiman actualmente unos 40 millones de usuarios en el mundo.

Comentarios

- Al ser un estándar internacional, está comenzando a ser recomendado como estándar en ofimática y utilizado por entidades y administraciones públicas de todo el mundo. De hecho, ya en mayo de 2004, el *Telematics between Administrations Committee (TAC)* de la Unión Europea publicó una serie de recomendaciones orientadas al fomento de un estándar como el que ahora existe. Igualmente, se puede comprobar que cumple con todos los requisitos del documento Marco de Interoperabilidad de la Comisión Europea (iDABC).
- Está basado en XML, respetando su filosofía (por ejemplo, separando presentación de contenido, en XML ambos, de binarios como imágenes) y diseñado para integrar todos los estándares XML previos, creando etiquetas nuevas solamente cuando no existe un estándar que ofrezca la funcionalidad necesaria [20]. Así, OpenDocument utiliza el núcleo de XHTML para enlaces, MathML para las fórmulas matemáticas, SVG para los gráficos vectoriales, SMIL para la multimedia, Dublin Core para los metadatos, ZIP para la compresión, etc.

3.2. Formato XML de Microsoft

A partir de la versión 2003 de MS-Office, Microsoft anunció un formato basado en XML, cuyas especificaciones serían publicadas. Este ha evolucionado hasta el actual *MOOX* (Microsoft Office Open XML) [26], que teóricamente está destinado a ser el formato nativo de las próximas versiones de Microsoft Office.

Evaluación de requisitos

■ Requisito 1.

- No se trata de un formato abierto, sino de un formato propietario de Microsoft, cuyas especificaciones han sido publicadas, siendo esta empresa quien puede decidir de forma unilateral su futuro, en función de sus legítimos intereses.
- Por otro lado, Microsoft ha iniciado el proceso que, en el futuro, podría convertirlo en un segundo formato estándar para documentos en ofimática, en competencia con el ya existente ISO 26300 (OpenDocument). Una parte de la industria ha criticado a Microsoft por este motivo, considerando que la existencia de dos estándares paralelos podría llegar a ser contraproducente (ver [36]). Adicionalmente, la consultora Gartner Group ha pronosticado en su último informe sobre OpenDocument que Microsoft no conseguirá la normalización ISO para su formato [38].

■ Requisito 2.

- Hoy por hoy, estos formatos solo están disponibles en sus versiones previas (esquemas XML 2003), de forma no nativa y para las últimas versiones de MS-Office, aunque

existen actualizaciones a estas que les permiten «leer, editar y grabar ficheros en alguno de los formatos Open XML [de Microsoft]» (traducción libre de [26]). En particular, existe poca variedad de software capaz de utilizarlos y los existentes no están disponibles en todas las plataformas del mercado, por lo que no se cumplen los requisitos 2c ni 2d.

- De cualquier manera, el hecho de que sea posible que las últimas versiones de MS-Office puedan llegar a leer y grabar este tipo de formato es un punto positivo, pues significa que existe software de muy buena calidad que los utiliza (2a y 2b).

Comentarios

- La gran baza de este formato es la inmensa base de usuarios que, presumiblemente, tendrá si llega a convertirse en el formato utilizado de forma nativa por MS-Office, la *suite* ofimática que domina el mercado de forma absoluta.
- Está basado en XML, lenguaje de marcado ampliable o extensible desarrollado por el World Wide Web Consortium que constituye en la actualidad el lenguaje de bajo nivel idóneo para intercambiar información estructurada. Aunque se le achaca el no estar integrado con los estándares previos relacionados con esta tecnología (como XHTML) y respeta de forma dudosa su filosofía (por ejemplo, separación entre presentación y contenido, y XML de binarios, como imágenes). Producto de uno usar los estándares existentes es que sus especificaciones actuales abarquen 4.000 folios en vez de los poco más de 700 del estándar oficial OpenDocument. Especificaciones tan innecesariamente grandes dificultan severamente el que aparezcan implementaciones alternativas y competitivas a la del propietario del formato original, por lo que al final, acaban fomentando o realimentando los monopolios existentes.
- Los documentos guardados en este formato XML pueden aún contener datos binarios codificados que solamente se pueden manejar desde Windows y Office 2003, como objetos OLE.
- No todas las aplicaciones de MS-Office pueden hacer uso del formato XML: PowerPoint 2003 sólo guarda las presentaciones en formatos binarios propietarios, no XML.
- Algunas características avanzadas de MS-Office 2003 solo están disponibles cuando los documentos se guardan en su formato binario propietario.
- XML no es el formato «nativo» o predeterminado de MS-Office 2003: el usuario tiene que preocuparse de guardar el documento en este formato diciéndolo expresamente.
- Algunos aspectos legales no están claros; en concreto, es necesario realizar un estudio que asegure la no existencia de patentes que pudieran impedir la libre implementación de este formato por parte de la industria en aquellos mercados donde tengan carácter legal.

3.3. RTF

Siglas de Rich Text Format [25], o formato de texto enriquecido, es formato de texto, en el sentido de que el documento no se guarda en formato binario, sino que está compuesto de caracteres y órdenes de control, análogamente a \TeX . Ha evolucionado desde la versión 1.0 hasta la 1.7.

Evaluación de requisitos

- **Requisito 1.** No se trata de un formato abierto, sino de un formato propietario de Microsoft cuyas especificaciones han sido publicadas por esta empresa. Aunque esta proporciona documentación técnica sobre RTF, no existe ninguna normativa que especifique este formato. Por lo tanto, no se cumple el requisito 1

■ **Requisito 2.**

- Se trata de un formato muy difundido, en el sentido de que lo contemplan prácticamente todas las *suites* ofimáticas existentes, en todas las plataformas. Por lo tanto, se cumplen los requisitos 2a al 2d.

Comentarios

- RTF tiene un serio inconveniente: existen extensiones propietarias, y numerosas versiones que generan problemas de compatibilidad, incluso entre distintas aplicaciones de Microsoft: un documento RTF generado con una versión de MS-Word puede no poder abrirse desde una versión anterior.
- Las imágenes se guardan sin comprimir, lo que hace que los documentos con gráficos embotrados ocupen mucha memoria de almacenamiento.
- RTF no puede guardar macros, lo que también puede verse como una ventaja en cuanto a seguridad.
- Por último, parece que Microsoft va a abandonar este formato en favor de los esquemas de referencia de XML en las futuras versiones de MS-Office.

3.4. Texto simple

También conocido (incorrectamente) como «texto plano» o «ASCII» [15], este formato se compone de una secuencia de caracteres en alguna codificación normalizada por la organización ISO[14], como ISO-Latin1[16], ISO-Latin9 [17] o Unicode (UTF-8 [18]).

Evaluación de requisitos

- Cumple con todos los requisitos enumerados en la sección 2.2

Comentarios

- Es el formato adecuado para mensajes de correo electrónico.
- Se trata de un formato muy limitado, que no permite el resaltado del texto (salvo que se utilice algún lenguaje de marcado sencillo, como Texto Reestructurado [35])
- Su correcta visualización depende de que emisor y receptor utilicen los mismos códigos de caracteres. En Europa occidental, los formatos tradicionales ISO-8859-1 e ISO-8859-15 están siendo sustituidos por el «universal», UTF-8.

3.5. HTML

Es un lenguaje de marcas diseñado para hipertexto [19]. Evidentemente, se excluyen de este formato extensiones no especificadas por los estándares publicados por W3C, el consorcio encargado de ello.

Evaluación de requisitos

- Este formato cumple con todos los requisitos enumerados en la sección 2.2

Comentarios

- Recomendado para la publicación en la web.
- A la hora de su uso como formato para el intercambio de información estructurada, presenta algunos inconvenientes:
 1. El mismo documento puede verse distinto en distintos equipos o visores (navegadores).
 2. Algunas herramientas de generación de HTML muy extendidas generan código HTML propietario no compatible con la especificación estándar.
 3. Es común que un documento se componga de varios ficheros (gráficos, hojas de estilo, etc), por lo que su intercambio o publicación puede resultar confuso para el usuario.

3.6. Lenguajes de marcado de texto: T_EX/L^AT_EX, DocBook

Evaluación de requisitos

- Como ficheros de texto sencillo, cumplen con todos los requisitos enumerados en la sección 2.2. La única objeción que se puede encontrar es su propia difusión (requisito 2b), probablemente debida a que su curva de aprendizaje es, en principio, muy inclinada.

Comentarios

- Adecuados para ser para la generación de documentos en otros formatos de impresión o lectura, como Postscript o PDF (o incluso HTML, OpenDocument, etc). De hecho, T_EX/L^AT_EX puede considerarse como un «estándar de hecho» para la publicación técnica universitaria o científica de libros, revistas y artículos.
- Como en HTML, es habitual que un documento se componga de varios ficheros (gráficos, hojas de estilo, etc), por lo que su intercambio o publicación definitiva puede resultar algo confuso para el usuario.
- Como en el caso del texto sencillo, dependen de la codificación de caracteres utilizadas por el emisor y el receptor

3.7. PDF

Todos los formatos enumerados hasta ahora son y están orientados a ser editables, de forma que la persona que recibe la información puede editarla y modificarla con facilidad. Sin embargo, en la mayoría de las ocasiones la información difundida está orientada a ser visualizada o impresa solamente, y no necesita ser modificada por el receptor.

PDF (Portable Document Format, o formato transportable de documentos) [27] fue creado en 1996 por la compañía Adobe Systems, Inc.[28], derivado de Postscript (PS) [29], también de su propiedad. Se caracteriza porque el aspecto del documento es el mismo en pantalla que cuando se imprime en papel, independientemente de la escala y de la plataforma utilizadas. Es, por tanto, un formato orientado a impresión.

Diversos subconjuntos de la especificación PDF han sido estandarizados o están en proceso. Para lo que nos ocupa, conviene destacar el estándar ISO 19005-1:2005, publicado el 1 de octubre de 2005 y referido a documentos que deban preservarse en el tiempo; es decir, que se garantice que puedan leerse sin problemas en años venideros. Este estándar se conoce como PDF/A [30] y se basa en la referencia PDF 1.4 de Adobe Systems Inc., implantada en su producto Adobe Acrobat Reader 5. La principal característica es que toda la información debe estar incluida en el mismo archivo, por lo que no se admiten audio, vídeo, lanzamiento de programas, guiones Javascript, cifrado, tipos de letra externos, etc.

Otros subconjuntos de PDF estandarizados o en proceso son PDF/X (ISO 15930) para artes gráficas e impresión, PDF/E para intercambio de dibujos ingenieriles y PDF/UA para documentos accesibles universalmente.

Evaluación de requisitos

- **Requisito 1.** La versión PDF/A, es un estándar abierto y cumple, por lo tanto, el requisito 1.
- **Requisito 2.**
 - A pesar de que existen muchos programas capaces de leer y grabar este tipo de ficheros, algunos de ellos de gran calidad, pocos de ellos utilizan de forma explícita la versión estándar ISO PDF/A. Aun así, la mayor parte de estos programas utilizan la versión 1.4, que en la práctica presenta muy pocas diferencias con el estándar.
 - En resumen, si nos restringimos a PDF/A, no existen aún programas lectores y generadores adecuados y por tanto no se cumple el segundo de los requisitos establecidos. La versión 1.4 de PDF sí que cuenta con programas adecuados, pero no está estandarizada (aunque sí está publicada y es muy cercana a PDF/A).

Comentarios

- Las versiones posteriores a 1.4 admiten características que se alejan mucho del estándar PDF/A y del propósito de un formato de para documentos de texto persistentes: audio, vídeo, lanzamiento de programas, guiones Javascript, cifrado, tipos de letra externos, etc. Esto puede dar pie a muchas confusiones por parte del usuario.
- Como inconveniente, los ficheros PDF ocupan más espacio en disco que otros formatos analizados, entre otras razones por que contienen mapas de bits de los tipos de caracteres que se han usado en el documento.

Capítulo 4

Hojas de cálculo

En la presente sección se estudiarán los *formatos para hojas de cálculo que pueden considerarse como formatos o estándares abiertos*. Como en la sección 3, dedicada a documentos de texto, se excluyen los formatos propietarios y cerrados, aun estando tan extendidos como el uso de forma nativa por las versiones actuales de la hoja de cálculo Microsoft Excel (.xls), dominante en el mercado. Del mismo modo que en dicha sección, nos centraremos solamente en aquellos formatos que estén suficientemente extendidos.

4.1. Valores separados por comas (CSV)

Este tipo de ficheros, conocidos por las siglas CSV (del inglés “comma-separated values”), son ficheros de texto sencillo formados por datos dispuestos en forma de una tabla, cuyos campos están habitualmente separados por comas y encerrados entre comillas, esto último salvo que no haya ambigüedad. Las filas están separadas por caracteres de fin de línea.

Evaluación de requisitos

- Cumple con todos los requisitos enumerados en la sección 2.2

Comentarios

- Es un formato adecuado para producir tablas de datos de forma sencilla, por ejemplo para crear guiones (*scripts*) que exporten valores contenidos en bases de datos.
- Se trata de un formato muy limitado en el que todos sus datos son estáticos.
- Como documento de texto sencillo, su correcta visualización depende de que emisor y receptor utilicen los mismos códigos de caracteres (ver sección 3.4).

4.2. OpenDocument

El estándar ISO OpenDocument (ver apartado 3.1) incluye soporte para hojas de cálculo, siendo .odc la extensión identificativa a este tipo de ficheros. Las ventajas e inconvenientes del uso de OpenDocument en hojas de datos coinciden con las expresadas en el apartado antes citado, así como el análisis de requisitos. Es un formato de prestaciones avanzadas y de generación dinámica de contenidos.

4.3. Formatos XML de Microsoft

Como en el caso de OpenDocument, las distintas versiones de los formatos XML de Microsoft incluyen la creación de hojas de cálculo, teniendo un formato de ficheros específico para ello. Las

ventajas e inconvenientes se heredan de las que se comentaron en la sección 3.2. También es un formato avanzado y de generación de datos dinámica.

4.4. PDF

Por supuesto, para intercambiar documentos que contengan tablas de datos en forma de ficheros persistentes, destinados a ser impresos o que no deban ser modificados con facilidad, se puede utilizar el formato PDF, normalmente utilizando algún programa de hojas de cálculo que exporten a este tipo de ficheros. Se puede aplicar aquí el análisis realizado en el apartado 3.7. La información almacenada en este formato será únicamente de carácter estático.

Capítulo 5

Presentaciones

Por último, realizamos un estudio de los *formatos abiertos para presentaciones* más extendidos, excluyendo por imperativo de los requisitos fijados en la sección 2.2 todos los formatos propietarios y cerrados, como el usado por las versiones actuales del programa de presentaciones dominante, PowerPoint o KeyNote, dominante a su vez en el mundo Apple.

5.1. OpenDocument

El estándar ISO OpenDocument (ver apartado 3.1) incluye soporte para presentaciones, siendo `.odp` la extensión identificativa a este tipo de ficheros. Las ventajas e inconvenientes del uso de OpenDocument para presentaciones coinciden con las expresadas en el apartado antes citado. Las presentaciones creadas usando este formato podrán presentar dinamicidad e incluir funcionalidades avanzadas (vídeo, sonido, integración con su entorno, etc.)

5.2. PDF

Puesto que, normalmente, las presentaciones no deben ser modificadas con facilidad después de haber sido creadas, la opción de utilizar el formato PDF con este fin es realmente interesante. Para ello es suficiente utilizar un programa adecuado para generar la presentación y un visor de ficheros PDF a pantalla completa para reproducirla. Se puede aplicar el análisis realizado en 3.7, pero además existen algunas características distintivas:

Comentarios

- La posibilidad de utilizar presentaciones PDF cuyo «código fuente» haya sido creado mediante el lenguaje de marcado \LaTeX tiene una gran importancia en aquellos ámbitos científicos en los que pesa favorablemente la capacidad de este lenguaje para la creación de fórmulas matemáticas complejas
- Las presentaciones generadas en formato PDF serán estáticas y no dispondrán de funcionalidades avanzadas como el vídeo, el sonido, tablas de cálculo dinámicas integradas, gráficas dinámicas, integración con el entorno, etc.

Parte III

Material adicional

Bibliografía

- [1] Conferencia de Rectores de las Universidades Españolas (CRUE). <http://www.crue.org/>
- [2] Grupos de trabajo de la CRUE. <http://www.crue.org/grupostrabajo-index.htm>
- [3] BOUCA n.º 15: http://www.uca.es/web/organizacion/normativa/documentos/boletines_2004/bouca_n0015.pdf
- [4] Normativa de Intercambio Institucional de Documentos en la UCA: http://softwarelibre.uca.es/normativa_iii
- [5] Área de Informática (CITI): <http://www2.uca.es/serv/ai>
- [6] Área de Informática de la Universidad de Cádiz. *Formatos para el intercambio institucional de documentos en la Universidad de Cádiz* http://softwarelibre.uca.es/docs/formatos_intercambio_documentos.pdf
- [7] Consorcio W3C: <http://www.w3c.org>
- [8] Microsoft Office: <http://www.microsoft.com/spain/office>
- [9] Microsoft Corp.: <http://www.microsoft.com/spain>
- [10] PDFCreator: <http://www.pdfcreator.de.vu>
- [11] Microsoft Windows: <http://www.microsoft.com/spain/windows>
- [12] OpenOffice.org: <http://es.openoffice.org>
- [13] GNU/Linux: <http://es.wikipedia.org/wiki/Linux>
- [14] ISO (*International Organization for Standardization*): <http://www.iso.org>
- [15] Código ASCII: <http://es.wikipedia.org/wiki/ASCII>
- [16] Código ISO-8859-1 o ISO-Latin1: <http://es.wikipedia.org/wiki/ISO-8859-1>
- [17] Código ISO-8859-15 o ISO-Latin9: http://en.wikipedia.org/wiki/ISO_8859-15
- [18] Código UTF-8: <http://es.wikipedia.org/wiki/UTF-8>
- [19] Lenguaje HTML: <http://es.wikipedia.org/wiki/Html>
- [20] OpenDocument: <http://es.wikipedia.org/wiki/OpenDocument>
- [21] Meta-lenguaje XML: <http://es.wikipedia.org/wiki/XML>
- [22] OASIS (*Organization for the Advancement of Structured Information Standards*): <http://www.oasis-open.org>

- [23] OASIS Foundational Sponsors http://www.oasis-open.org/about/foundational_sponsors.php
- [24] OASIS About <http://www.oasis-open.org/about/index.php>
- [25] Formato RTF: <http://es.wikipedia.org/wiki/RTF>
- [26] Esquemas XML de Microsoft Office 2003: <http://www.microsoft.com/office/preview/itpro/fileoverview.msp>
- [27] Formato PDF: <http://www.adobe.com/es/products/acrobat/adobepdf.html>
- [28] Adobe Systems, Inc.: <http://www.adobe.com/es>
- [29] Lenguaje PostScript: <http://en.wikipedia.org/wiki/PostScript>
- [30] Formato PDF/A: <http://en.wikipedia.org/wiki/PDF/A>
- [31] El Informe *Valoris (Comparative Assessment of Open Documents Formats Market Overview)*: <http://europa.eu.int/idabc/servlets/Doc?id=17982>
- [32] *TAC approval on conclusions and recommendations on open document formats*: <http://europa.eu.int/idabc/en/document/2592/5588>
- [33] Definición de Estándar Abierto por la comisión IDABC de la Comunidad Europea <http://europa.eu.int/idabc/en/document/3761>
- [34] Política de Formatos de Documentos, Universidad de la República (Uruguay): http://www.rau.edu.uy/universidad/csdi/proyectos/0311PoliticaFormatoDocumentos_1_0_0.pdf
- [35] Texto Restructurado: <http://docutils.sourceforge.net/rst.html>
- [36] Microsoft Office Open XML (Wikipedia) <http://en.wikipedia.org/wiki/MOOX>
- [37] Licencia Creative Commons Reconocimiento-CompartirIgual 2.5 Spain License <http://creativecommons.org/licenses/by-sa/2.5/es/>
- [38] Informe sobre OpenDocument de la consultora *Garner Group*. http://www.gartner.com/resources/140100/140101/iso_approvali_ofi_oasisi_opendoi_140101.pdf
- [39] Política de apoyo a los formatos abiertos en el estado de Massachusetts (EEUU) http://en.wikipedia.org/wiki/OpenDocument_adoption#Massachusetts
- [40] Apoyo a los formatos abiertos por parte del Gobierno de Bélgica <http://europa.eu.int/idabc/en/document/3146/360> (fuente: proyecto IDABC de la Unión Europea)
- [41] Apoyo a los formatos abiertos por parte del Gobierno de Dinamarca <http://ec.europa.eu/idabc/en/document/5650/469> (fuente: proyecto IDABC de la Unión Europea)
- [42] Compromiso de la Junta de Andalucía sobre el uso de estándares abiertos <http://www.juantomas.net/2006/06/06/andalucia-se-adelanto-a-los-daneses/>