

# O formato MS-Word non é un bon formato para o intercambio de documentos.

Ramón Flores

6 de febreiro de 2002

## 1 Introducción

Desde ha algun tempo resulta excesivamente frecuente o envío de mensaxes de correo con anexos consistentes en arquivos MS-Word, i.e. con arquivos no formato propio dese procesador de textos. Esta práctica, se ben frecuente, non é moito recomendábel, xa que o dito formato non é, do meu punto de vista, unha boa elección para transmitir información textual, tal e como se vé en seguida.

Non se trata neste documento sobre a idoneidade do programa MS-Word como procesador de textos, aínda que eu persoalmente non goste del<sup>1</sup> e non o use, senón da (im)pertinencia do formato propio deste programa para o intercambio de documentos. O normal é que un procesador de textos poda gravar un mesmo documento en varios formatos<sup>2</sup>, e un/unha usuári@ intelixente debe escoller cal é o formato máis axeitado para cada caso.

## 2 O formato ideal

Un bon formato para a troca de informacións textuais debería ser:

- **Intelixíbel.** I.e. debería poder ser lido, e a ser posíbel mesmo manipulado, en calquer tipo de computador, rolando calquer sistema operacional. De maneira a ser entendido por todo o mundo sen problemas.
- **Confiábel.** Ou sexa, debería reproducir fielmente as informacións que carrega, nen máis nen menos da que @ autor@ desexa enviar, e mantendo a forma dada inicialmente pol@ autor@. Ademais non debería comprometer a seguridade nen do emisor nen do receptor.
- **Lixeiro.** I.e. a razón entre o tamaño ocupado polo arquivo e a información contida debería manter-se baixa.
- **Resistente a erros.** Ou sexa, a perda dalguns poucos bytes non debería dar lugar a unha perda masiva da información contida.

## 3 Carências do formato MS-Word

Infelizmente o formato do MS-Word non cumpre nengun dos requisitos anteriores. Vexamos porqué:

---

<sup>1</sup>Non son o único, vexa-se por exemplo o artigo de Allin Cottrell Word Processors: Stupid and Inefficient, <http://www.ecn.wfu.edu/~cottrell/wp.html>.

<sup>2</sup>Por exemplo, este artigo foi escrito con o procesador de documentos libre LyX <http://www.lyx.org>, o cal permitiu gravá-lo como Texto Simple, en L<sup>A</sup>T<sub>E</sub>X, no seu formato propio, en HTML, en DVI, en Postscript e en PDF.

### 3.1 O formato MS-Word non ten un alto grau de intelixibilidade.

- **Requer software específico e propietario.**

Este é un formato propietario, binario e fechado, polo cal só o procesador de textos MS-Word o entende cabalmente, e para iso nen sempre. Se ben o procesador de textos MS-Word está amplamente espallado, é un aplicativo propietario, e non especialmente barato, que ademais só ten versións para dous sistemas operacionais, o Windows e o Mac OS. Portanto non todo o mundo o ten instalado no seu computador. A Internet construíu-se procurando que os distintos tipos de computadores pudesen comunicarse entre si, para todo o mundo poder estar a vontade. Enviar arquivos en formatos propietarios presupondo, sen información confiábel, cal é o software que usa @ recipiendári@ agride ese espírito de procurar a intercomunicación, introducindo ruído na rede. máis

- **Ha problemas entre versións.**

Mesmo para os usuarios de MS-Word, ha problemas de compatibilidade entre as diversas versións do aplicativo. Así utilizar as versións máis modernas impede que os usuarios de versións máis antigas podan ler as súas mensaxes.<sup>3</sup>

Usar versións antigas non soluciona sempre o problema, se ben é práctica máis segura, pois aínda así poden aparecer alteracións molestas no formato. Os problemas entre Windows e Mac OS son loxicamente maiores.

- **Distintas impresoras = Distintas aparicións.**

Aínda que @ recipiendári@ dun ficheiro MS-Word teña a versión axeitada do procesador de textos, mesmo así pode haber problemas. A aparición dun documento MS-Word depende da impresora que teña o sistema, ou máis ben do driver correspondente. Portanto ao formatar un documento MS-Word, tal vez con unha aparición moito traballada para atinxir unha impresión dada n@ leitor@, non se pode estar 100% certo de cal vai ser a aparición do dito documento nun outro computador con unha impresora diferente.

### 3.2 O formato MS-Word non é confiábel.

É ben coñecido que a Microsoft non se ten caracterizado pola elaboración de produtos moito seguros. No caso que nos ocupa isto evidencia-se en dous puntos, as facilidades dadas para crear virus e a súa falta de discrição.

- **Virus**

O MS-Word posúe un sistema de macros bastante potente e non moito seguro. De maneira que hoxe en día é o vector máis común de propagación de virus electrónicos. Enviar arquivos MS-Word habitualmente e sen necesidade é unha maneira moito eficiente de favorecer a propagación de virus. Se só fosen enviados os ficheiros MS-Word precisos, moitísimos menos que os actualmente enviados, recibir un ficheiro deste tipo non esperado sería xa un aviso de perigo, e a difusión dos virus sería moito máis dificultosa. Realmente quer fomentar o espallamento de virus pola rede?

Que un particular envíe arquivos MS-Word con virus non acrescenta a súa estima entre os seus coñecidos. Mais no caso de ser unha institución a facé-lo, mesmo a ofrecé-los nas súas páxinas web, dá unha moito má imaxe da mesma.

- **Indiscricións**

Os documentos MS-Word inclúen máis informacións das que aparecen a primeira vista, e que en moitas ocasións @s autor@s dos documentos non desexarian foran fornecidas. Así ademais do que se le ao abrir o documento no procesador de textos MS-Word, o documento leva informacións sobre o autor e o computador no que foi redixido, e tamén informacións sobre versións antigas do documento, que nalguns casos pudesen ser confidenciais. Para ler esa información chega ás veces

---

<sup>3</sup>Ha quen diga que isto non é por acaso, senón procurado pola Microsoft para obrigar a@s usuári@s a actualizar os seus aplicativos. Que non o farian por necesidade de novas características técnicas, mais si para poder entender os arquivos da xente que os actualizaran.

con abrir o arquivo con un procesador de textos distinto do MS-Word. Outras veces son necesarios procedimentos máis elaborados<sup>4</sup>, mais en todo caso a información está no arquivo.

### 3.3 O formato MS-Word é pesado.

Con desmedida frecuencia envíanse mensaxes nun ficheiro MS-Word, que só contén texto simple, e que puidera ir moi ben nunha mensaxe de texto normal. Só que no primeiro caso ocupa centos de KB cando como texto simple ocuparía uns poucos KB. Se para acesar a internet ten que pagar, pense nisto antes de enviar ficheiros Word desnecesariamente. Se é afortunado e o acceso resulta-lle gratuito, sexa boa persoa e pense nos demais.

### 3.4 Resistencia aos erros do formato MS-Word.

A combinación procesador MS-Word formato MS-Word ten moi pouca resistencia aos erros. Unha pequena alteración do arquivo da lugar a que o procesador considere-o corrupto e non permita recuperar ningunha información. Mas neste caso o problema non é debido unicamente ao formato mais tamén ao procesador. Así utilizar outro procesador de textos<sup>5</sup>, ou o ollar o ficheiro con outras ferramentas<sup>6</sup>, pode permitir recuperar a maior parte da información.

## 4 Alternativas

Existen diferentes alternativas ao formato MS-Word, ningunha delas válidas para todas as ocasións. Da mesma maneira que non nos vestimos igual para unha boda e para xogar ao fútbol, non debemos utilizar o mesmo formato para diferentes finalidades.

### 4.1 Texto simple

Ha moitos casos onde o mellor é utilizarmos texto simple; unha carta comercial ou a un amigo, un artigo nun grupo de noticias da USENET, ou nunha lista de correos, etc. Sempre que posible debe preferirse o texto simple, pois é o formato máis intelixible, lixeiro e resistente a erros (sen redundancias), e 100% fiable. Ademais permite incluír facilmente nas respostas treitos da mensaxe contestada. Podemos dormir con smoking, mais é moito máis cómodo facelo con un pijama.

Caso de ser necesario enviar un documento con unha aparencia elaborada existen formatos máis adecuados que o MS-Word:

### 4.2 RTF

O RTF, acrónimo de Rich Text Format, é un formato que permite unha formatación moi elaborada, sendo propiedade da Microsoft. Mais a diferenza do formato MS-Word é aberto, pois as especificacións son públicas e facilmente verificábeis ao ser lexible polas persoas. É por iso que é lido e escrito correctamente por unha grande variedade de procesadores de texto. Ademais ao ser lexible polas persoas, sempre se pode recuperar moita información mesmo que for levemente danado.

Está pois indicado para cando se quer enviar un documento, que poda ser manipulado polo destinatario, e cuxa forma principal de lectura sexa impreso.

#### 4.2.1 Como gravar un arquivo con formato RTF no MW-Word?

Ir ao menu

Arquivo -> Gravar como -> Gravar como tipo:

---

<sup>4</sup>En Unix por exemplo pode-se usar o comando `strings` para obter texto "oculto" nun arquivo MS-Word.

<sup>5</sup>por exemplo o procesador libre Abiword, <http://www.abisource.com>.

<sup>6</sup>por exemplo o xa citado comando `strings`.

e escoller

Formato RTF (\*.rtf)

### 4.3 HTML

O HTML, acrónimo de Hypertext Markup Language, é un formato aberto, público e lexíbel pelas persoas, polo cal é entendido por moitos aplicativos informáticos; evidentemente todos os navegadores de internet, moitos procesadores de textos, clientes de correo, etc. É un formato pensado para ver documentos na tela, mais non é tan bon para documentos que se van imprimir. Tampouco é conveniente abusar deste formato. Algun coñecido cliente de correo envía por defeito mensaxes de correo duplas: texto e html, o cal é totalmente desnecesário se o que se quer enviar é unha mensaxe de texto.

#### 4.3.1 Como gravar un arquivo con formato HTML no MS-Word?

Ir ao menu

Arquivo -> Gravar como -> Gravar como tipo:

e escoller

Páxina Web (\*.htm; \*.html)

Contodo debe subliñar-se que o MS-Word produz ficheiros html de bastante pouca cualidade, moi pesados e non sempre respeitantes dos estándares. Neste aspecto é bastante mellor a cualidade dun ficheiro rtf producido por este procesador.

### 4.4 PDF

Acrónimo de Portable Document Format, o PDF é un formato aberto, binário e propiedade de Adobe. Mais as especificacións son públicas e ben publicadas, existindo aplicativos que len e escriben PDF en calquer sistema operacional. Se o que se desexa é enviar documentos con moi boa aparencia, que non dependa da impresora a usar, e que non poda ser alterado facilmente pol@ recipiendári@, este é o formato de eleición. De facto é usado cada vez máis pelas empresas para difundir información corporativa: catálogos, especificacións técnicas, etc. Por outra banda non é moi resistente a erros, ao ser binário, e tampouco é especialmente lixeiro.

O MS-Word non permite gravar en PDF, mais ha formas de facé-lo con software libre. Unha completa descrición de como o facer pode achar-se na páxina web Using Ghostscript to Make PDF Files, <http://www.sp.uconn.edu/~fdb94001/using-gs/make-pdf.html>.

## 5 Cando é apropiado empregar o formato MS-Word

Dicia anteriormente que non ha nengun formato apropiado en todas as circunstancias, senón que segundo a situación haberá que usar un ou outro. Existe tamén unha circunstancia propicia para empregar o formato MS-Word: cando se sabe positivamente que @ recipiendári@ ten o procesador de textos MS-Word, unha versión que poda manexar sen problemas o arquivo enviado, e ademais vai ora traballar no dito documento, ora empregar partes dele noutros.

## 6 Porqué deste documento

A verdade nunca gostei do procesador de textos MS-Word, polo cal nunca o tiveron no meu computador. Porén ha bastante xente que presupon que eu teño o dito procesador, e envía-me arquivos nese formato. E xa cansei de ver, despois de pasar bastante traballo para os abrir, que na maioría dos casos o formato era totalmente supérfluo.

Ha uns poucos meses batin nas miñas navegacións pola rede con un artigo de Jeff Goldberg, que partindo dunha situación semellante a miña tomara-se o traballo de redixir un alegación contra o abuso do envío indiscriminado de documentos en formato MS-Word. Pensei en traduci-lo, mais o fun deixando e ha un par de semanas Richard Stallman, o pai do software libre, deu a lume un artigo incitando a rebelar-se contra este abuso MS-Wordiano. Asi que comecei a traducir, mais xa postos abandonei a idea inicial e redixi un artigo próprio, bastante inspirado iso si no de Jeff Goldberg. Non pretendo portanto ser moi orixinal, tan só contribuir dentro das miñas posibilidades nesta campaña educativa. Pois diso se trata, dunha campaña para espallar boas maneiras na rede.

## 6.1 Páxinas relacionadas

- Jeff Goldberg, *MS-Word is Not a document exchange format* <http://www.goldmark.org/netrants/no-word/attach.html>.
- Richard Stallman, *We can put an end to Word attachments* [http://linxtoday.com/news\\_story.php3?ltsn=2002-01-11-002-20-OP](http://linxtoday.com/news_story.php3?ltsn=2002-01-11-002-20-OP).
- Martin Vermeer's, *Plaintext: In praise of practical e-mail hygiene*, <http://www.netby.dk/Oest/Europa-Alle/vermeer/plain.html>.
- Michael A. Gurski, *Why No Word?*, [http://userpages.umbc.edu/~mgursk1/rants/why\\_no\\_word.html](http://userpages.umbc.edu/~mgursk1/rants/why_no_word.html).
- Chris Mason, *Why no Word?* [http://www.unixzone.com/why\\_no\\_word.html](http://www.unixzone.com/why_no_word.html).
- Stéphane Bortzmeyer, *After Word: l'avenir du traitement de texte*, <http://www.internatif.org/bortzmeyer/afterword/afterword.html>.
- Juhapekka Tolvanen, *Miksi on typerää postittaa sähköpostin liitetiedostona MS Word -dokumenteja*, <http://www.cc.jyu.fi/~juhtolv/mswordmail.html>.
- Sam Steingold, *No Proprietary Binary Data Formats*, <http://www.podval.org/~sds/data.html>.
- Adam Barr, *Proposal: The Open Data Format Initiative*, <http://www.osopinion.com/perl/story/16034.html>.
- Scott Mcneal, *The Case for Open File Formats*, [http://www.computerworld.com/cwi/community/story/0,3201,NAV65-1797\\_STO64689,00.html](http://www.computerworld.com/cwi/community/story/0,3201,NAV65-1797_STO64689,00.html).
- Kevin Reichard, *Editor's Note: The Importance of Open File Formats*, <http://www.linuxplanet.com/linuxplanet/reviews/1934/1/>.