

Fichiers au format .doc : où est le problème ?

Les logiciels de traitement de texte ont pour fonction de mémoriser un contenu comportant principalement du texte et, éventuellement, des dessins, et des images. On en attend finalement qu'ils nous permettent de façonner une mise en forme et de la restituer fidèlement à l'écran ou sur papier.

Mais cette fidélité de restitution s'appuie nécessairement sur l'élaboration de méta-données que le logiciel enregistre en annexe du fichier et qu'il exploite pour afficher ou imprimer comme attendu notre contenu (polices de caractères, configurations de pagination,...). Il peut aussi y incorporer les fruits de fonctionnalités supplémentaires telles que des statistiques sur le document, nombre de mots, nombres de signes typographiques, mémorisation des modifications successives, des identifications de la version du logiciel utilisé, du système d'exploitation, des données sur l'organisation de mon disque dur,... liste non-limitative, à la discrétion des créateurs du logiciel.

Avoir tout cela en conscience permet de comprendre pourquoi il est précieux, si l'on veut rester maître de ses productions bureautiques, que le logiciel de traitement de texte que l'on utilise soit tout à fait transparent quant aux méta-données qu'il ajoute à notre texte. Et voilà que l'usage du format « propriétaire » devient un problème grave. Par définition, un tel format est fait à partir de spécifications secrètes protégées par des brevets qui interdisent qu'on y ait accès. J'ai bien « saisi » mon texte, mais le fichier créé, en fait, ne m'appartient pas et je n'ai le droit d'en connaître que la face émergée.

Lorsque je l'envoie par e-mail ou que je le mets en ligne sur le web, je n'ai aucune conscience des informations me concernant qui sont de ce fait mises en circulation et librement accessibles avant tout pour l'éditeur du logiciel (Micro\$oft par exemple), mais aussi pour ceux à qui il a cédé des droits d'investigations (NSA, officines diverses...) ou pour les hackers débrouillards qui se sont donné les moyens de venir y fouiller.

Diffuser sans souci un fichier au format .doc, propriété privée de Micro\$oft, est aussi naïf que se croire pudiquement habillé en passant dans le sas détecteur du quai d'embarquement de l'aéroport. Bien pire, même, puisqu'avec le .doc, je ne sais même pas en quoi consiste ce que cachent mes habits !

Hum...! affaire à suivre.

Fichiers au format .doc : où est le problème ? (suite)

Heu... on disait quoi, déjà ? Ah oui ! Que le format d'un fichier définit la façon dont sont mentionnées les informations associées à un contenu numérisé. Elles sont destinées à ce que le logiciel soit capable de restituer celui-ci dans une forme accessible aux humains. Le format définit aussi la nature des informations associées, et c'est là que peuvent se jouer des manœuvres occultes à l'insu de l'utilisateur. Et de ceci ne se privent pas certains éditeurs de logiciels dont l'objectif ultime n'est pas de vous rendre service, mais de vous capter en tant que client en vous ligotant de la façon la plus serrée possible. La magouille est très habile : vous payez (cher !) pour être capté et manipulé, voire mené par le bout du nez quant au choix de votre système d'exploitation, de sa mise à niveau, du renouvellement de votre matériel, ... et pour n'avoir pas une totale maîtrise de vos productions numérisées : l'outil (disons, son propriétaire, qui n'est pas vous !) pourra à son gré imposer qu'on respecte ses droits.

Et de plus, la mise en circulation de fichiers dans un format propriétaire fait de vous un agent de propagande au service de ce propriétaire : « *Comment ! Vous n'utilisez pas la dernière version de Word ? N'êtes-vous pas un peu ringard ?* » (regard en coin, soupçonneux :-/)

Si vos destinataires, par légèreté, par défaut de servilité ou par souci d'indépendance ne se sont pas soumis au racket si intelligemment mis en place, ils peuvent bien être en difficulté pour accéder à vos contenus : soit on se soumet et on paye, soit on pirate (on est alors dans l'illégalité, passible de sanctions) soit... on prend les choses en main, et on utilise des logiciels libres.

En effet, l'intelligence collective des développeurs nombreux qui élaborent et améliorent sans cesse ces logiciels (on peut même les aider!), a bien-sûr rendu possible, outre de travailler joyeusement avec des formats de fichiers « ouverts », de retraiter les fichiers issus du monde « propriétaire » afin de les rendre librement accessibles à tous (format odf pour « open document format »).

Le contraire n'est évidemment pas vrai : les logiciels propriétaires-privatifs font tout pour essayer de faire croire que les formats libres seraient des formats exotiques marginaux abracadabrans : le diable ne s'est-il pas toujours fait fort d'affirmer que la vérité est un mensonge, et vice versa ?

Mais... à suivre !

Fichiers au format .doc : où est le problème ? (suite 3)

Ainsi, les formats de fichiers 'propriétaires' mis en œuvre par Micro\$oft (doc, docx, xls...) sont coupables de piéger les utilisateurs : du fait que Micro\$oft modifie le format de fichier lors des changements de version, ses utilisateurs sont prisonniers d'un cercle vicieux qui les contraint à acheter chaque mise à jour, qu'ils désirent changer ou non. Ils pourront même s'apercevoir, dans quelques années, que les documents Word qu'ils rédigent aujourd'hui ne pourront plus être lus avec la version de Word qu'ils utiliseront alors.

- Ces formats sont inadaptés aux transferts et aux échanges. En effet, la mise en page d'un document dépend de la machine sur laquelle on le consulte. Ainsi un texte qui tient sur une page sur une machine avec une version de MS Word peut en occuper deux sur une machine avec une autre version. Les marges dépendent des imprimantes installées sur la machine. Il faut que le correspondant ait la même version pour être sûr qu'il verra la même chose que ce que vous avez préparé.

- Ces formats posent des problèmes de sécurité : ils peuvent véhiculer sournoisement des virus/macros

- Ces formats posent des problèmes de confidentialité du fait des informations incluses à l'insu de l'auteur.

Mais alors que faire ? Des envois au format TXT (texte brut), au format PDF (portable document format), au format ODT (Open document texte), HTML, etc. Tous ces formats sont documentés et connus, et des solutions libres existent pour les produire et les lire, quelle que soit la plate-forme matérielle et logicielle choisie. On en reparle, la prochaine fois ?

Fichiers au format .doc : où est le problème ? (épisode 4)

La preuve de l'inadéquation des formats de fichiers 'propriétaires' pour la publication et l'échange de documents numériques ayant été apportées dans les épisodes précédents, jetons un regard sur les diverses bonnes solutions accessibles à l'usager de la communication numérique, en ce qui concerne les documents textes.

Quel que soit le logiciel de traitement de texte utilisé, l'enregistrement du document que l'on a créé peut être fait rapidement par une commande "enregistrer" qui imposera le format de fichier 'par défaut' du logiciel en question. Avec Micro\$oft, aucune chance alors d'échapper au calamiteux doc, voire au docx. Pourtant l'enregistrement peut aussi toujours être fait par la commande "enregistrer sous" ('save as', in english) qui va proposer de choisir soi-même parmi une série de propositions de formats de fichiers dont certains sont tout à fait recommandables. Le plus rustique, TXT a l'avantage de passer partout et de ne s'alourdir par aucune méta-donnée, mais il ne peut, du coup, pas contenir de dessins ou d'images ni permettre de fixer une mise en page ou des effets de typographie en dehors de l'alternative majuscule/minuscule. Il s'avère donc insuffisant pour publier des textes. Le RTF (rich text format) format public quoi que créé par Micro\$oft, semble très avantageux tant que l'on s'en tient à des documents de textes avec mise en page. On rencontre cependant parfois quelques problèmes de restitution en passant d'un système à l'autre, liés au fait que Micro\$oft ne respecte pas toujours fidèlement ses propres spécifications. Le plus performant et potentiellement universel est sans doute l'ODT (open document texte), issu du monde du logiciel libre. C'est la raison pour laquelle Micro\$oft s'est longtemps refusé à le reconnaître. On le produira par défaut avec des logiciels comme Libre Office, mais il restera parfois illisible pour ceux qui n'auront pas fait le pas de commencer à se libérer de l'emprise de Micro\$oft.

C'est la raison pour laquelle la garantie de lisibilité universelle et irréprochable d'un document ne sera présente que s'il est « exporté » au format PDF (portable document format). L'exportation consiste à en créer une image morte (plus de souci de pagination ou de mise en forme à la réception). L'inconvénient ? Le document est momifié, on ne peut pas le modifier, il n'est pas destiné à un travail collaboratif d'élaboration. Ce format PDF, qui peut donc inclure à volonté dessins, images, voire sons, offre aussi, lors de la création du document, la possibilité de gérer son taux de compression. C'est une option dont on aurait tort de se priver puisqu'elle permet d'ajuster finement le compromis poids de fichier/qualité de rendu. Le nombre de kilo octets d'un fichier est en effet lourd... de conséquences sur le volume de stockage et sur le coût écologique de la transmission. Ce dernier, s'il est souvent occulté ou méprisé, n'en est pas moins réel : les communications électroniques (Internet) sont déjà devenues le plus gros poste de consommation électrique mondial. Une affaire à méditer si l'on se soucie de l'avenir de la planète.

Heu... bonnes vacances ?

Pierre Paccoud – <http://www.celibre.ovh>