

LA TYPOGRAPHIE du WEB

6- Le format PDF et son usage sur le web

Jean-Claude Sohm / EFPG

(créé le : 20 juillet 2000 / révisé le : 19 janvier 2001)

URL : <http://cerig.efpg.inpg.fr/icg/dossiers/typo/chap6.htm>

Copyright © CERIG/EFPG 1996-2001

Le logiciel Acrobat et le format PDF

Le logiciel Acrobat permet de créer des documents au format PDF. Ce format a été imaginé par l'éditeur Adobe pour l'échange de documents numérisés. Quelles que soient la machine et la plate-forme, le format PDF conserve la présentation d'origine du document (typographie, mise en page, images), et permet de l'imprimer facilement. Le document est enregistré page par page ("page indépendance"), et il est éditable... assez peu commodément, il faut le reconnaître.

Le format PDF est de plus en plus utilisé dans les arts et industries graphiques. Adobe a présenté le format PDF comme le successeur probable de PostScript en PAO, et pour l'instant les deux formats co-existent. Pour certains, ces deux formats apparaissent comme complémentaires plutôt que comme concurrents dans le domaine de la PAO.

Lorsque le web s'est développé, l'éditeur Adobe a nourri l'espoir de voir PDF supplanter HTML. Il a enrichi PDF de la possibilité d'inclure des liens hypertextes, des formulaires, et des images gif animées. Le format PDF, cependant, n'a jamais pu concurrencer sérieusement le HTML, parce qu'il n'a pas été conçu à l'origine pour le web. D'où un certain nombre de défauts, que les internautes ne lui ont pas pardonné :

- la taille des fichiers est souvent plus élevée. Ce défaut s'est particulièrement manifesté avec les premières versions de PDF ;
- il est nécessaire de télécharger une version récente du lecteur (Acrobat Reader), qui fonctionne ensuite comme module externe (plug-in) ;
- il est impossible d'exploiter un fichier requérant un module externe, et de ce fait les nouveautés du web (Flash, le streaming, etc.) ne peuvent pas être utilisées dans une page PDF ;
- la page PDF s'adapte à la largeur de la fenêtre du navigateur par homothétie, et non par réarrangement du texte comme le fait la page HTML, etc.

Finalement, le format PDF a trouvé sa niche sur le web : il est couramment utilisé pour **enregistrer les pages destinées à être imprimées** plutôt que visualisées, et pour lesquelles une reproduction de bonne qualité est requise. En d'autres termes, sur le web, les formats HTML et PDF ne sont pas concurrents, mais complémentaires. Au premier les pages destinées au téléchargement, à l'affichage et à la consultation en direct ; au second les pages destinées à l'impression, et à la consultation en différé.

Les concepteurs de sites web qui créent des pages au format PDF peuvent en profiter pour soigner leur typographie ; celle-ci sera transmise à l'internaute, affichée, et imprimée comme telle. Le format PDF résout le problème de la transmission de la police avec le document... du moins pour les pages qui sont plutôt destinées à être imprimées.

PDF et le web

Il existe trois bonnes raisons pour créer deux versions (une HTML et une PDF) de la même information destinée à être diffusée sur le web :

- les pages HTML sont conçues pour être visualisées sur un écran, et non pour être imprimées. Elles contiennent des éléments qui sont inutiles, ou qu'il n'est pas d'usage d'utiliser, ou qui ne fonctionnent pas, sur support papier ;
- la fonction "Impression" des navigateurs fonctionne de manière médiocre, et ce d'autant plus que la page HTML considérée est plus complexe ;
- il est pénible de consulter un document de grande taille sur l'écran d'un moniteur. Beaucoup d'internautes préfèrent se voir offrir la possibilité de le télécharger, puis de l'imprimer, en tâche de fond.

Passer de l'affiché à l'imprimé. Les pages HTML contiennent des éléments qui sont spécifiques du web (barres de navigation, liens hypertextes), et qu'il est inutile de reproduire sur papier. De plus, le web est beaucoup plus décoré que l'imprimé : sur une page HTML, personne ne s'étonne de trouver des bannières, des fonds colorés ou décorés, des plages de couleurs vives, et du texte qui n'est pas noir. Sur support papier, le style est généralement plus sobre ; on soigne davantage le confort de lecture, et on limite le plus souvent les débordements de la communication graphique à la page de couverture. Enfin, une page web peut contenir de l'information multimédia (sons, images animées, vidéo), que l'on ne peut pas transcrire sur papier, et dont il n'est pas utile de garder la trace quand on imprime la page.

Lorsqu'un internaute décide d'imprimer une page web, c'est pour conserver sous forme écrite l'essentiel de l'information qu'elle contient. Il n'est pas intéressé à garder trace des éléments de navigation... et encore moins des bannières publicitaires ! Il peut, certes, travailler par copier-coller vers un traitement de texte, et remettre en forme le nouveau document obtenu. Il peut également introduire la page dans un éditeur de pages web, et la remanier à son goût. Tout ceci, bien sûr, est long et fastidieux.

Pour rendre service à l'internaute qui veut conserver par écrit des informations provenant du web, certains concepteurs de sites offrent donc une copie "**printer-friendly**" des pages qu'ils mettent en ligne. Cette copie peut être simplement une version dépouillée de la page HTML. Mais si le document est quelque peu complexe, s'il comprend des tableaux et des figures, si sa mise en forme est soignée, et **si l'on veut soigner sa typographie**, le format PDF s'impose. Si l'internaute ne possède pas le lecteur autonome Acrobat, ce n'est pas un drame : il peut toujours se contenter d'imprimer la page web telle quelle... ou plutôt telle que son navigateur voudra bien la transmettre à l'imprimante.

Faciliter l'impression de l'information. Comme chacun sait, les navigateurs impriment mal, et même souvent très mal. Nous avons tous fait l'expérience de tableaux et de listes (parfois même de lignes de texte !) qui sont coupés lors du passage à la page suivante, de figures de fond de tableau ou de fond de cellule qui ne sont pas reproduites, de texte qui est tronqué à droite, et même de pages qui ne s'impriment pas du tout. Les navigateurs, il est vrai, ont une excuse : une page HTML n'a pas de hauteur définie -- et pas non plus de largeur définie lorsque la présentation proportionnelle est utilisée. Les navigateurs doivent donc découper ladite page en morceaux de taille A4 avant de l'imprimer, ce qui requiert un peu d'intelligence et d'initiative, qualités que les ordinateurs ne possèdent évidemment pas. Dès qu'une page web est un peu complexe, il est bon d'en proposer à l'internaute une copie, prédécoupée en pages A4, et enregistrée au format PDF -- copie dont on est sûr qu'elle s'imprimera correctement (la fonction "Impression" du lecteur autonome Acrobat fonctionne généralement de manière tout à fait satisfaisante).

Régler le problème des documents volumineux. Comme chacun peut le constater, il est plus agréable de consulter un document volumineux sous forme imprimée que sous forme affichée. De plus, si le document comporte de nombreuses figures, le temps de téléchargement peut devenir prohibitif pour une lecture directe en ligne. La solution souvent adoptée consiste à créer une page web contenant une forme abrégée du document (plan, résumé, résumé étendu), et de donner la possibilité à l'internaute de télécharger la version complète sous forme d'un fichier PDF. On peut aussi utiliser un format courant de traitement de texte mais, comme nous l'avons vu au chapitre 2, le fichier correspondant sera beaucoup plus lourd. Il faut absolument éviter d'utiliser le format propre à un logiciel de mise en page, et ce pour les deux raisons suivantes :

- seuls les spécialistes de la PAO disposent de ce type de logiciel sur leur machine ;
- il n'existe pas de lecteur autonome (plug-in) dont on puisse proposer le téléchargement à l'internaute.

Lorsqu'un document est volumineux, mais qu'il est dépourvu de figures, une solution pratique mais inélégante consiste à en faire un un fichier ASCII. Les seuls éléments de mise en page dont on dispose alors sont le retour-chariot et le saut de ligne. Inutile de parler de typographie dans ce cas, puisque le texte s'affiche avec une police à chasse fixe. À titre d'exemple, on pourra consulter le RFC de l'Internet Society qui définit la [version 1.1 du protocole HTTP](#). Ce document de 412 Ko est fort austère, mais à qui gère un site web, il rend service. Avant de l'imprimer, réfléchissez bien, car il s'étend sur plus de 150 pages !

En pratique, on peut se trouver dans deux situations différentes, selon l'origine du document que l'on désire enregistrer au format PDF pour le diffuser sur le web :

- le document a initialement été conçu pour être imprimé. Il a donc été créé soit dans un logiciel de traitement de texte, soit dans un logiciel de mise en page. La transformation en fichier PDF à l'aide de PDFWriter est rapide et ne présente pas de difficulté ;
- le document a initialement été conçu pour le web. Il a donc de très fortes chances d'avoir été créé directement dans un éditeur de pages HTML. Il faut alors lui faire subir quelques modifications, avant de l'enregistrer au format PDF à l'aide d'Acrobat. L'opération est plus longue et un peu plus compliquée que la précédente.

PDFWriter

PDFWriter est l'un des composants du logiciel Acrobat. Il fonctionne comme un module externe vis à vis des logiciels de traitement de texte ou de mise en page les plus courants. Il faut simplement veiller à l'installer sur la machine après ces derniers. La présence du module externe PDFWriter se manifeste de manière un peu singulière : lorsqu'on utilise la fonction "Fichier / Imprimer..." du logiciel de traitement de texte ou de mise en page, PDFWriter apparaît dans la liste des imprimantes disponibles.

On ouvre le document dans le logiciel qui a servi à le créer, puis on l'enregistre au format PDF à l'aide de PDFWriter en faisant comme si on l'imprimait. La présentation du document source n'est pas modifiée. On peut contrôler le résultat obtenu en ouvrant dans Acrobat (ou dans le lecteur autonome) le fichier PDF obtenu. Il est également bon de vérifier la taille de ce fichier avant de le proposer sur le web aux internautes.

La seule difficulté de cette procédure peut provenir des images. Il faut veiller à ce que le document d'origine contienne des images de résolution suffisante, sinon leur conversion par PDFWriter conduit à un résultat fort médiocre. Le logiciel, en effet, pixellise (si nécessaire), et ramène systématiquement la résolution à 72 dpi. Pour que le résultat soit correct, il faut que la résolution initiale (dans le cas d'une image matricielle) soit nettement supérieure à 72 dpi.

Le fichier "printer-friendly"

Dépouiller la page source. Il importe de créer d'abord une version expurgée de la page web, en éliminant :

- le fond de page. Un fond de page coloré ou décoré peut être fort plaisant sur le web -- du moins est-ce l'opinion de l'auteur de ces lignes. Mais à l'impression, le résultat est généralement désastreux. De plus, cela ralentit l'imprimante, et vide les cartouches d'encre à toute vitesse. L'élimination du fond de page peut être confiée au logiciel Acrobat si on le désire, mais c'est aussi simple de le faire dans l'éditeur de pages web ;
- les éléments de navigation : boutons, barres, etc. Ces éléments ne servent plus à rien lorsque la page est imprimée, puisqu'ils ne sont plus actifs ;
- la publicité (à moins que vous ne l'aimiez, bien entendu -- tous les goûts sont dans la nature) ;
- certains éléments décoratifs, qui rendent les pages web plus agréables, mais détonnent dans un document imprimé (bannière de haut de page, par exemple) ;
- les liens hypertextes, qui ne sont plus actifs lorsque le document est imprimé. Si vous êtes vraiment consciencieux, vous pouvez expliciter les URL correspondantes ;
- les animations, les sons, et toutes les informations non transposables au support imprimé.

Supprimer le halo. La suppression du fond coloré peut entraîner des problèmes avec les images. Pour éviter les effets de halo, les concepteurs consciencieux lissent le contour de certaines images sur un fond de même couleur que celui de la page web à laquelle elles sont destinées. Lorsque le fond devient blanc, le procédé se retourne contre son auteur : un halo coloré apparaît ! Il faut donc disposer d'un nouveau jeu d'images, lissées sur fond blanc. A titre d'exemple, les puces de couleur marron (660000) de la présente page web sont lissées sur fond jaune (FFFFE0). Pour créer la page "printer friendly", nous utilisons des puces de la même couleur marron, mais lissées sur fond blanc (FFFFFF).

Attention à l'image monopixel. Il est un type d'image que nous avons tendance à oublier, mais qui va se rappeler à notre attention si nous n'y prenons pas garde : il s'agit de l'image gif monopixel, qui sert à la mise en forme des pages web, en liaison avec les tableaux invisibles. Si l'image a été déclarée transparente, pas de problème. Si par contre elle possède la couleur du fond de page, il faut la blanchir pour la rendre de nouveau invisible.

Divers. Ces opérations terminées, il est bon d'écrire quelque part l'URL de la page web. Il faut en effet toujours garder trace de l'origine d'un document, qu'il provienne du web ou d'ailleurs. On peut confier cette tâche au logiciel Acrobat si on le désire.

Vous pouvez vous arrêter là si vous le voulez, et offrir la page HTML ainsi dépouillée comme page "printer-friendly" aux internautes qui visitent votre site. Mais vous pouvez leur offrir un meilleur service en transformant cette page web en fichier PDF, qui s'imprimera de manière plus satisfaisante. Le choix entre ces deux solutions dépend de la complexité de la page : la version HTML dépouillée suffit pour une page courte qui ne comporte ni figures, ni tableaux ; la version PDF est préférable lorsque la page source est plus complexe.

Du HTML au PDF

Bien que conçu pour coopérer avec un logiciel de traitement de texte ou de mise en page, Acrobat offre la possibilité de transformer directement une page web en fichier PDF. Ce faisant, on se heurte aux trois difficultés suivantes :

- le découpage de la page HTML en pages A4 ;
- l'ignorance des styles par Acrobat ;
- l'ignorance de la justification.

Le découpage en pages A4. Une page web n'a pas de dimensions. Quand on l'affiche, c'est le navigateur qui détermine sa largeur et sa hauteur, en fonction des contraintes (largeur de la fenêtre, nature et taille des polices par défaut, etc.) qu'il rencontre sur la machine de l'internaute. Présentement, c'est Acrobat qui se charge de découper la page web en pages de taille définie, en fonction des consignes données par l'opérateur. Sauf exception rarissime, ce dernier choisit toujours le format A4 (ou quelque chose de très voisin), parce que c'est le format standard de la majorité des imprimantes.

Lors de cette opération, Acrobat fait de son mieux pour éviter de couper les tableaux et les images, et de créer des veuves et des orphelines. Il en résulte que certaines pages risquent d'être mal remplies et/ou mal présentées. Il n'y a pas d'autre solution que de retravailler le fichier HTML, par touches successives, de telle sorte que les pages A4 fournies par Acrobat aient une allure satisfaisante.

Les feuilles de style. Le logiciel Acrobat ne reconnaît pas les feuilles de style : les informations données par ces dernières sont donc perdues. Ce n'est pas bien grave pour l'instant, parce que les feuilles de style sont encore peu utilisées (voir le chapitre 5 du présent dossier), mais cela deviendra très gênant si l'éditeur Adobe ne corrige pas cette situation dans les deux ans à venir.

La justification. Acrobat reconnaît la justification lorsqu'elle se trouve dans un logiciel de traitement de texte ou de mise en page, mais il ne la reconnaît pas lorsqu'elle provient de l'attribut "align" de la balise <p> d'une page web. C'est pourquoi la version HTML de cette page vous apparaît justifiée (pas dans IE4 sur Mac, mais la version 5 a corrigé cela), alors que la version PDF apparaît alignée à gauche. N'en veuillez pas au CERIG, ce n'est pas sa faute !

Divers. Signalons pour terminer qu'un document imprimé doit être paginé. Acrobat se chargera de cette opération si on le lui demande.

Un peu de pratique

Convertir la page HTML dépouillée en fichier PDF. Lancer Acrobat, puis utiliser la fonction "Fichier / Ouvrir une page web...". La boîte de dialogue "Ouvrir une page web s'affiche.

Saisir l'adresse de la page web dans le champ URL. Si le serveur web correspondant se trouve sur la même machine, l'adresse commencera par : "http://localhost/.....". Dans la zone "Options", activer le bouton "Niveaux" et régler sa valeur à 1. Ne pas cocher les deux autres cases.

Presser le bouton "Options de conversion...", et la boîte de dialogue "Options de conversion", dotée de deux onglets, apparaît.

Dans l'onglet "Général", on trouve deux rubriques : "Types de contenu" et "Fichiers PDF résultants". Dans la première rubrique, choisir "HTML" bien entendu. Décocher toutes les cases de la seconde rubrique, à l'exception de celle intitulée "Insérer en-têtes et pieds de page" si l'on désire créer une pagination. En en-tête, le logiciel inscrit le titre de la page web -- celui qui figure entre les balises <TITLE> et </TITLE>. En pieds de page, le logiciel inscrit l'URL de la page web, la date et l'heure, et page. En utilisant l'outil "Retouche", on peut modifier le pied de page une fois la conversion effectuée.

Onglet "Mise en page". Choisir le format A4 et l'orientation "Portrait", bien entendu. Sauf cas rarissime, les pages web ne contiennent pas d'objet large (figure, tableau, etc.) nécessitant l'utilisation de la barre de défilement horizontale, et donc l'orientation "Paysage" lors du passage au PDF. Les marges indiquées s'ajoutent à celles qu'impose l'imprimante utilisée ; leur réglage n'est pas critique. Dans la rubrique "Échelle", il faut cocher "Ajuster le contenu à la largeur de la page", sinon le découpage en pages A4 ne s'effectue pas. Par précaution, décocher "Passer au mode paysage au-dessous de :", mais ce n'est pas important.

L'onglet "Général" possède un bouton "Options..." qui permet d'ouvrir la boîte de dialogue "Options de conversion des fichiers HTML". Cette boîte de dialogue possède deux onglets : "Disposition" et "Polices". Le premier onglet permet de régler de manière automatique quelques propriétés de la future page PDF (couleur du texte, de l'arrière-plan, des tableaux...). L'opinion de l'auteur est que ces propriétés sont beaucoup mieux réglées lorsqu'on crée la version dépouillée de la page HTML. Veiller cependant à cocher la case "Convertir les images". Le second onglet concerne les polices (titres, texte) : même remarque que ci-dessus. Ne pas cocher la case "Incorporer les polices de la plate-forme".

Prêt ? Déclancher l'opération en appuyant sur le bouton "Télécharger". Le logiciel Acrobat indique les opérations en cours, puis la première page du fichier PDF résultant s'affiche. Par prudence, enregistrer ! Puis procéder aux corrections si nécessaire, ou répéter l'opération en modifiant le fichier source si l'on n'est pas satisfait. Attention ! Après les corrections, enregistrer en utilisant "Fichier / Enregistrer sous...", sinon le fichier PDF prendra des proportions considérables (phénomène analogue à celui d'un traitement de texte).

Les polices. Le fichier PDF une fois créé dans Acrobat, il est instructif de jeter un coup d'œil dans la rubrique "Fichier / Informations sur le document / Polices...". Une boîte de dialogue intitulée "Informations sur les polices", contenant la liste des polices utilisées, s'ouvre. On apprend ainsi que le fichier PDF qui correspond à la présente page web utilise trois polices : Arial MT (pour le texte), Comic Sans MS Bold (pour les titres) et Haettenschweiler (pour l'en-tête). La taille du fichier PDF est telle (48,8 Ko) qu'il est évidemment impossible que ces trois polices y aient été incorporées intégralement. Seules les définitions des caractères effectivement utilisés ont été intégrées au fichier PDF. Quel dommage qu'on ne puisse pas faire la même chose en HTML ! Mais le sort du HTML est entre les mains du W3C, et il n'y a pas grand'chose de bon à attendre de cet organisme -- du moins est-ce l'opinion de l'auteur de ces lignes.

Incorporer le fichier PDF dans une page web. Rien de plus simple : on crée un lien hypertexte vers un fichier PDF comme vers un fichier HTML ou TXT, pour ne citer que ces deux exemples. Lorsque l'internaute clique sur ce lien, de deux choses l'une :

- ou son navigateur est doté du module externe "Acrobat Reader", et le fichier PDF s'affiche dans la fenêtre correspondante ;
- ou son navigateur n'est pas doté du module externe précité, et une boîte de dialogue s'ouvre, lui proposant de télécharger le fichier PDF et de l'enregistrer sur son disque dur.

Conclusion

Ainsi, le format PDF a trouvé sa place sur le web. Il est utilisé chaque fois que l'on veut donner à l'internaute la possibilité d'imprimer un document de bonne qualité : notice technique, manuel d'utilisation, rapport, document très long ou comportant beaucoup de figures, etc. C'est dans ce rôle que PDF connaît le succès sur le web, et plus de 110 millions de copies d'Acrobat Reader ont été téléchargées ou préinstallées à ce jour, toutes versions (1 à 4) confondues.

Fournir à l'internaute deux versions du même document, l'une optimisée pour l'affichage (page HTML), l'autre pour l'impression (page HTML "printer-friendly" ou fichier PDF), demande plus de travail qu'on ne pourrait le penser a priori. C'est le fameux problème de la publication "cross-platform", vu ici sous un jour particulier, et pour lequel il n'existe pas de solution miracle -- sauf chez les éditeurs qui ont du logiciel ad hoc à vendre, bien entendu.

Pour les documents relativement simples, ne comportant ni tableau ni figure, une version dépouillée de la page web est amplement suffisante pour créer un fichier "printer-friendly", que les navigateurs imprimeront sans gros problème. Pour les documents plus complexes, comportant figures et/ou tableaux, seul le fichier PDF peut garantir une bonne reproductibilité lors l'impression.

Le concepteur de site qui double sa page HTML d'une page PDF peut en profiter pour incorporer à cette dernière la typographie de son choix, et améliorer ainsi sa présentation. Tous ceux qui sont issus du milieu de la PAO apprécieront probablement de se retrouver ainsi en terrain familier.