

L'optimisation de fichiers PDF et Word pour le référencement

[Retour au sommaire de la lettre](#)

Google et Yahoo! indexent, depuis de nombreuses années, les fichiers PDF et Word (Doc). Mais comment faire en sorte que ce type de fichier soit le mieux classé possible dans les résultats de ces moteurs ? En d'autres termes, comment optimiser un fichier PDF ou DOC pour ces deux moteurs de recherche majeurs ? Voici les résultats de quelques tests que nous avons menés ces dernières semaines qui devraient vous aider à penser "moteurs" lorsque vous rédigez vos contenus mis en ligne sous un format autre que l'HTML...

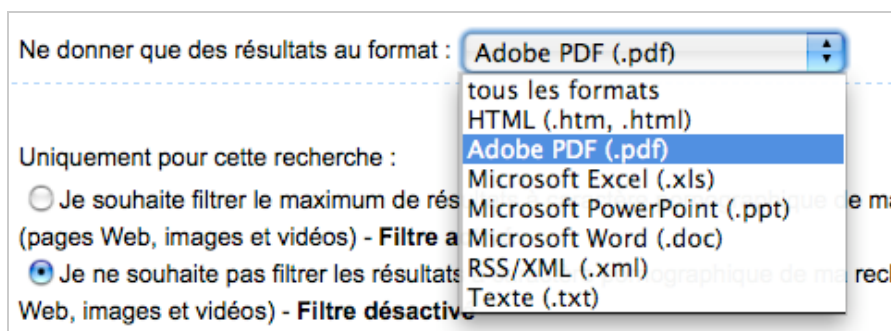
Nous avons, en octobre 2003, consacré un article à l'optimisation des fichiers PDF et Word pour les moteurs de recherche (<http://abonnes.abondance.com/archives/2003-10/>). Comme cela faisait cinq ans que cet article avait été écrit, et conscient qu'une forte demande existait à l'heure actuelle à ce sujet (de nombreuses questions sont posées sur les fichiers PDF, notamment, dans les formations que nous animons), nous avons entrepris d'écrire un nouvel article à ce sujet, basé sur un certain nombre de tests que nous avons effectués sur Google et Yahoo!. En voici les résultats.

Prise en compte par les moteurs

Tout d'abord, il est important de bien comprendre que les moteurs de recherche indexent sans problème les fichiers PDF (terminaison ".pdf") et Word (terminaison ".doc" pour la plupart). Pour les visualiser, vous pouvez :

- Sur **Google**, utiliser la syntaxe "filetype:". Exemple : **abondance filetype:pdf** (<http://www.google.fr/search?hl=fr&q=abondance+filetype%3Apdf>) ou **abondance filetype:doc** (<http://www.google.fr/search?hl=fr&q=abondance+filetype%3Adoc>) pour indiquer le filtre adéquat au moteur.

- Sur **Yahoo!**, utiliser la recherche avancée du moteur (<http://fr.search.yahoo.com/web/advanced?ei=UTF-8&p=>) et opter pour le choix "**Format de fichiers > Ne donner que des résultats au format :**" qui propose notamment ces deux possibilités, entre autres :



Chacun de ces deux moteurs indique, dans ses résultats, le format des fichiers trouvés devant leur titre par la mention [PDF] sur les deux outils, [DOC] sur Google et [MICROSOFT WORD] sur Yahoo! :

[PDF] [Uranium: l'abondance au rendez-vous](#)

Format de fichier: PDF/Adobe Acrobat - [Version HTML](#)

Une **abondance** planétaire. Cette **abondance** est d'autant plus profitable ... Uranium: l'**abondance**. au rendez-vous. i le nucléaire contribue pour 17% de ...

www.cea.fr/var/plain/storage/original/application/07eade098598d926eccc4627b584b173.pdf - [Pages similaires](#) - [À noter](#)

Fichier PDF sur Google

[PDF] [Laissez-vous conter la vallée d'Abondance](#)

107k - Adobe PDF - [Afficher en html](#)

place de l'église, **Abondance** ... plaine d'Offaz, **Abondance**. Du lundi au vendredi de 10h à 12h et de 14h à 18h ... du village, **Abondance**. tous les mercredis ...

www.abbaye-abondance.org/docs/divers/1215011255.pdf

Fichier PDF sur Yahoo!

[doc] [TP n°8 Découverte et **abondance** des éléments chimiques](#)

Format de fichier: Microsoft Word - [Version HTML](#)

Info : l'**abondance** de chaque élément mesure la masse de l'élément en milligrammes par kg d'échantillon de la croûte terrestre. ...

www2.ac-lyon.fr/enseigne/physique/docs/fichiersftp/phys2000/stgenis/tp7corps-simples.doc - [Pages similaires](#) - [À noter](#)

Fichier Word sur Google

[MICROSOFT WORD] [www.abbaye-abondance.org/docs/divers/1226569937.doc](#)

53k - - [Afficher en html](#)

Cloître de l'Abbaye- BP 01- 74360 **Abondance**. Tel/fax : 04 50 81 60 54. Courriel : cloitre.abondance@orange.fr. www.abbaye-abondance.org ...

www.abbaye-abondance.org/docs/divers/1226569937.doc


Fichier Word sur Yahoo!

Côté indexation, pas de problème donc... Mais quelles informations les moteurs de recherche lisent-ils dans ces fichiers ?

Zones reconnues par les moteurs de recherche

Voici les différentes zones que l'on peut "remplir" dans un fichier Word ou PDF et la façon dont Google et Yahoo! les lisent ou non (notamment les "Propriétés" ou "métadonnées" que l'on peut fournir pour mieux décrire les documents) :

FICHIERS PDF (*.pdf)

		
Contenu textuel	OUI	OUI
Métadonnée Titre (TITLE)	OUI	NON
Métadonnée Sujet (SUBJECT)	NON	NON
Métadonnée Auteur (Author)	NON	NON
Métadonnée Mots Clés (Keywords)	NON	OUI
Url	OUI	OUI

FICHIERS WORD (*.doc)

		
Contenu textuel	OUI	OUI
Métadonnée Titre (TITLE)	OUI	NON*
Métadonnée Sujet (SUBJECT)	NON	NON
Métadonnée Manager	NON	NON
Métadonnée Auteur (Author)	NON	NON
Métadonnée Compagnie	NON	NON
Métadonnée Category	NON	NON
Métadonnée Mots Clés (Keywords)	NON	NON
Métadonnée Commentaires (Comment)	NON	NON
Url	OUI	OUI

* : nous avons trouvé quelques cas isolés où Yahoo! lisait la balise TITLE du document mais la plupart du temps, ce n'était pas le cas...

La situation est donc, ici, assez simple en termes de lecture des contenus et des métadonnées par ces deux moteurs majeurs :

- Les deux moteurs lisent les contenus textuels des deux formes de fichiers (ce n'est certes pas un scoop...).
- Ils détectent également les mots clés dans les urls. Il sera donc important de soigner les intitulés des noms de fichier en y incluant des mots clés pertinents par rapport au contenu de l'article ou du document.
- Pour les métadonnées, Google ne lit que le TITLE des fichiers PDF et Word. Il s'en sert d'ailleurs pour l'afficher comme titre dans ses résultats :

[\[PDF\] title azerty11](#)

Format de fichier: PDF/Adobe Acrobat - [Version HTML](#)

Yahoo! Search Assist ajoute des images à ses suggestions. La fonction Yahoo! Search Assist permet, comme "Suggest" sur Google, de proposer des mots clés ...

www.abondance.com/yahoo-search-assist.pdf - [Pages similaires](#) - [À noter](#)

"title azerty11" est le contenu de la métadonnée "TITLE" de ce fichier-test au format PDF.

Google ne lit pas d'autres métadonnées.

- Pour ces métadonnées, Yahoo! ne lit pas le TITLE mais lit les mots clés ("KEYWORDS") ajoutés au document PDF (mais pas Word). Il ne lit pas d'autre métadonnées.

Côté métadonnées, le travail sera finalement assez vite effectué sur vos fichiers : seule la balise TITLE devra être remplie, voire également la KEYWORDS pour Yahoo! (bien qu'il y ait de fortes chances pour que ce moteur octroie à cette zone un poids bien faible...).

Contenu des Snippets

Comment les deux moteurs affichent-ils leurs résultats lorsqu'il s'agit de fichiers PDF ou Word ?



Fichiers PDF :

[PDF] [title azerty11](#)

Format de fichier: PDF/Adobe Acrobat - [Version HTML](#)

Yahoo! Search Assist ajoute des images à ses suggestions. La fonction Yahoo! Search Assist permet, comme "Suggest" sur Google, de proposer des mots clés ...

www.abondance.com/yahoo-search-assist.pdf - [Pages similaires](#) - [À noter](#)

Le titre du fichier (en bleu souligné, à droite de la mention [PDF]) est constitué :

1. Par le contenu de la métadonnée TITLE s'il existe. Il semblerait que Google ne prenne pas en compte le TITLE si son contenu se termine par une terminaison de fichier (".doc", ".eps"...). Il faut donc que cette zone contienne de "vrais mots" et non pas un nom de fichier.
2. Un titre trouvé dans le contenu textuel du document si le TITLE n'est pas rempli ou s'il propose un nom de fichier.

La mention "Format de fichier: PDF/Adobe Acrobat" est indiquée ensuite, suivie d'un lien vers une version HTML du document, permettant de visualiser rapidement le contenu du fichier.

Le texte descriptif reprend le début du contenu textuel du document (ici : "Yahoo! Search Assist ajoute des images à ses suggestions. La fonction Yahoo! Search Assist permet, comme "Suggest" sur Google, de proposer des mots clés") ou une zone de texte identifiée comme pertinente par Google. A ce niveau, nous vous conseillons de créer des documents PDF dont la mise en pages est très simple pour la première page : titre, chapo, texte, avec un titre et un chapo très descriptif. La mise en page peut être plus sophistiquée par la suite, mais, si vous voulez maîtriser la façon dont Google "comprend" vos documents, un système de multi-colonnage n'est par exemple pas approprié, pour le début du document en tout cas... Car dans ce cas, c'est le contenu de la première colonne qui risque d'être pris en compte par Google, ce qui ne donnera pas obligatoirement le résultat escompté...

Fichiers Word :

[doc] [TP n°8 Découverte et abondance des éléments chimiques](#)

Format de fichier: Microsoft Word - [Version HTML](#)

Info : l'**abondance** de chaque élément mesure la masse de l'élément en milligrammes par kg d'échantillon de la croûte terrestre. ...

www2.ac-lyon.fr/enseigne/physique/docs/fichiersftp/phys2000/stgenis/tp7corps-simples.doc - [Pages similaires](#) - [À noter](#)

Titre du fichier : idem que PDF

Mentions : "Microsoft Word" + lien vers la version HTML

Texte descriptif : Idem que PDF



Fichiers PDF :

[PDF] [Laissez-vous conter la vallée d'Abondance](#)

107k - Adobe PDF - [Afficher en html](#)

place de l'église, **Abondance** ... plaine d'Offaz, **Abondance**. Du lundi au vendredi de 10h à 12h et de 14h à 18h ... du village, **Abondance**. tous les mercredis ...

www.abbaye-abondance.org/docs/divers/1215011255.pdf

Titre du fichier : Yahoo! tente de détecter un titre dans le texte du document et l'affiche dans ses résultats. Ainsi le fichier de test www.abondance.com/yahoo-search-assist.pdf est affiché ainsi par ce moteur :

[PDF] [Yahoo! Search Assist ajoute des images à ses suggestions](#)
92k - Adobe PDF - [Afficher en html](#)
La fonction Yahoo! **Search Assist** permet, comme "Suggest" sur Google, ...
<http://tools.search.yahoo.com/newsearch/searchassist>. <http://images.search.yahoo.com> ...
www.abondance.com/yahoo-search-assist.pdf

Le titre (ici : "Yahoo! Search Assist ajoute des images à ses suggestions") est extrait du contenu du document par Yahoo! Il faudra donc bien faire en sorte que le premier contenu trouvé par les moteurs corresponde au titre de l'article ou du document...

Mentions : Taille du fichier + "Adobe PDF" + lien vers la version HTML.

Texte descriptif : il est ici constitué par plusieurs extraits du texte, concaténés, chacun d'entre eux contenant les mots clés demandés. Le résultat n'est d'ailleurs pas toujours très heureux...

Fichiers Word :

[MICROSOFT WORD] www.abbaye-abondance.org/docs/divers/1226569937.doc
53k - - [Afficher en html](#)
Cloître de l'Abbaye- BP 01- 74360 **Abondance**. Tel/fax : 04 50 81 60 54. Courriel :
cloitre.abondance@orange.fr. www.abbaye-abondance.org ...
www.abbaye-abondance.org/docs/divers/1226569937.doc

Titre du document : la plupart du temps, ce sera l'url du document si Yahoo! ne trouve pas de titre pertinent. Là aussi, il vous faudra faire en sorte que le titre du document soit aisément identifiable (au début, en gras, en gros caractères, etc.) pour éviter le plus possible l'affichage de l'url qui n'est pas très "sexy"...

Mentions : Taille du fichier + "vide" + lien vers la version HTML.

Texte descriptif : idem que PDF

Quelques conseils d'optimisation

Pour bien optimiser vos fichiers PDF et Word, nous vous conseillons donc de suivre les conseils suivants :

- **Métadonnées :** remplissez la balise "TITLE" qui est prise en compte par Google et parfois par Yahoo!. Si vous avez un peu de temps, remplissez la balise "Keywords", qui est lue par Yahoo! (n'y passez pas non plus beaucoup de temps, cela n'en vaut que peu la peine)...

- **Contenu :** adoptez, pour la première page notamment, une mise en page simple et efficace : un titre en gras, en gros caractères, et un chapo de deux ou trois phrases, également en gras, le tout étant très descriptif et contenant les mots clés importants pour comprendre le document. Et oui, on est ici très proche de l'optimisation d'une page HTML "classique"... Notez bien que les "footer" et "header" (haut de page, bas de page) sont compris comme du texte par les moteurs. Évitez de les afficher sur la première page car les moteurs risquent de lire le contenu du header en premier, ce qui peut ne pas être pertinent (notez bien que cela peut également l'être...).

- **Indexation :** bien sûr, proposez, depuis vos pages web, des liens vers vos documents PDF et Word, avec des intitulés de lien explicites, pour leur donner "bonne réputation" (évitez donc les phrases du type "pour télécharger notre livre blanc sur référencement, [cliquez ici](#)" et préférez des formulations du type "Téléchargez notre [livre blanc sur le référencement](#)"). N'hésitez pas,

également, à proposer une page spéciale "téléchargement de documents" listant tous les fichiers disponibles sur votre site. Une bonne piste de départ pour les spiders des moteurs...

- **Nom du fichier** : utilisez des mots explicites pour nommer votre fichier ([assemblee-generale-2008.pdf](#), [charte-experts-comptables.doc](#), etc.) car ces mots sont lus et analysés par les moteurs.

- **Taille du fichier** : attention aux formats de fichiers, dépassant notamment le méga-octet. Des fichiers trop volumineux ne seront peut-être pas indexés en totalité. Préférez une suite de "petits fichiers" dans ce cas...

- **Accessibilité** : Word, notamment, propose de nombreuses fonctions visant à améliorer l'accessibilité des fichiers (ajout d'attribut "alt" aux images, etc.). N'hésitez pas à vous en servir, les critères d'accessibilité sont toujours très proches de ceux du référencement...

Voici donc pour ces quelques conseils qui, nous l'espérons, vous aideront à mieux optimiser vos fichiers pour les deux moteurs de recherche majeurs.

PS : notez que nous n'avons pas inclus Live Search, le moteur de recherche de Microsoft dans nos tests. En fait, nous attendons toujours, depuis plusieurs semaines, qu'il indexe nos fichiers de test, ce que Google et Yahoo! ont fait en quelques jours. Sacré Live...

Pour en savoir plus :

L'article que vous venez de lire est basé sur nos propres tests. Cependant, pour les conseils d'ordre général, nous avons parfois également puisé quelques informations dans les articles suivants que nous vous conseillons de lire avec la plus grande attention :

-> *Optimiser le référencement des fichiers PDF*

<http://s.billard.free.fr/referencement/?2007/09/14/422-optimiser-le-referencement-des-fichiers-pdf>

Traduction de l'article "*Eleven Tips For Optimizing PDFs For Search Engines*" disponible à l'adresse : <http://searchengineland.com/eleven-tips-for-optimizing-pdfs-for-search-engines-12156.php>

-> *Accessibilité et référencement des fichiers PDF*

<http://www.cmic.ch/2007/09/14/accessibilite-et-referencement-des-fichiers-pdf/>

-> *SEO Your PDF's*

<http://www.searchengineguide.com/senews/002430.html>

-> *Optimizing PDFs for SEO*

<http://www.smallbusinesssem.com/optimizing-pdfs-for-seo/288/>

Mais il en existe bien d'autres (n'hésitez pas à nous en signaler d'autres si vous en connaissez)...
Bonne optimisation !

Olivier Andrieu

Editeur du site Abondance.com

Réagissez à cet article sur le blog des abonnés d'Abondance :

<http://abonnes.abondance.com/blogpro/2008/12/loptimisation-de-fichiers-pdf-et-word.html>