
PdfàSom
--o0o--
Version du 2 janvier 2021

Yvon Massé
ymasse@gnomonique.fr

SOMMAIRE

1	Présentation-----	1
2	Licences-----	1
3	Installation de PDFtk Server-----	2
4	Installation de PdfàSom-----	3
5	Lancement de PdfàSom-----	3
6	Fichier .som-----	4
6.1	Création-----	4
6.2	Généralités - sections obligatoires-----	5
6.3	Table de renvois-----	6
6.3.1	Pagination classique-----	7
6.3.2	Pagination par folio-----	7
6.3.3	Pagination par double page-----	7
6.4	Propriétés-----	7
7	Historique des versions-----	8

1 **Présentation**

PdfàSom permet d'ajouter un sommaire dans un fichier pdf, ou modifier celui déjà existant, dans le but de faciliter la navigation et la lecture du pdf. PdfàSom utilise pour cela un programme annexe, PDFtk Server, qu'il faut aussi installer.

Le sommaire se déclare dans un fichier texte (son extension est .som) qui est facile à renseigner et à modifier. PdfàSom se charge ensuite d'intégrer le sommaire dans le fichier pdf à l'aide de PDFtk Server.

PdfàSom et PDFtk Server fonctionnent sur PC Windows et Mac OS X.

2 **Licences**

PdfàSom est un « échangiciel ». Vous pouvez le copier, le communiquer à vos amis et l'utiliser gratuitement. En échange, si vous mettez en place un sommaire pour une des numérisations proposées sur les bibliographies Biblio-Gno ou Bibli-Horlo, vous devrez transmettre le fichier .som à l'auteur du programme (ymasse@gnomonique.fr) qui le mettra à disposition de tous les internautes sur les sites Biblio-Gno ou Bibli-Horlo en regard de la numérisation correspondante.

PdfàSom est fourni en l'état, sans garantie d'aucune sorte. Toutefois, si vous constatez un quelconque dysfonctionnement n'hésitez pas à contacter l'auteur du programme qui s'efforcera d'apporter les modifications nécessaires.

PDFtk Server est un logiciel libre sous licence GPL.

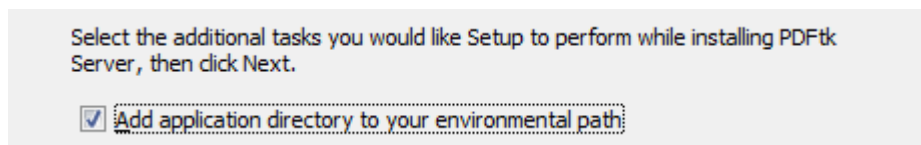
3 Installation de PDFtk Server

Aller sur le site officiel et suivre les instructions de la page :

<https://www.pdflabs.com/tools/pdftk-server/>

Les utilisateurs de Windows peuvent télécharger et installer indifféremment PDFtk Server ou [PDFtk Free](#). Dans le second cas ils disposeront en plus d'une application leur permettant de manipuler (assembler, dissocier, repaginer, etc.) les fichiers pdf.

Lors de l'installation sous Windows, il est important de ne pas décocher la case suivante :



Les utilisateurs de Mac qui ont une version 10.11 ou supérieure d'OS X doivent télécharger une nouvelle version en utilisant le lien direct suivant :

https://www.pdflabs.com/tools/pdftk-the-pdf-toolkit/pdftk_server-2.02-mac_osx-10.11-setup.pkg

Dans tous les cas, pour s'assurer que PDFtk Server est bien installé et disponible, il suffit de taper dans une console la commande `pdftk` qui doit renvoyer le message suivant :

SYNOPSIS

```
pdftk <input PDF files | - | PROMPT>
[ input_pw <input PDF owner passwords | PROMPT> ]
[ <operation> <operation arguments> ]
[ output <output filename | - | PROMPT> ]
[ encrypt_40bit | encrypt_128bit ]
[ allow <permissions> ]
[ owner_pw <owner password | PROMPT> ]
[ user_pw <user password | PROMPT> ]
[ flatten ] [ need_appearances ]
[ compress | uncompress ]
[ keep_first_id | keep_final_id ] [ drop_xfa ] [ drop_xmp ]
[ verbose ] [ dont_ask | do_ask ]
```

Where:

```
<operation> may be empty, or:
[ cat | shuffle | burst | rotate |
  generate_fdf | fill_form |
  background | multibackground |
  stamp | multistamp |
  dump_data | dump_data_utf8 |
  dump_data_fields | dump_data_fields_utf8 |
  dump_data_annots |
  update_info | update_info_utf8 |
  attach_files | unpack_files ]
```

For Complete Help: `pdftk --help`

4 Installation de PdfàSom

Sous Windows, lancer le programme [install_pdfasom_v2janv21.exe](#) et suivre les indications.

Sous OS X ouvrir l'image disque [pdfasom_v2janv21.dmg](#) et déplacer les éléments qui se trouvent à l'intérieur :

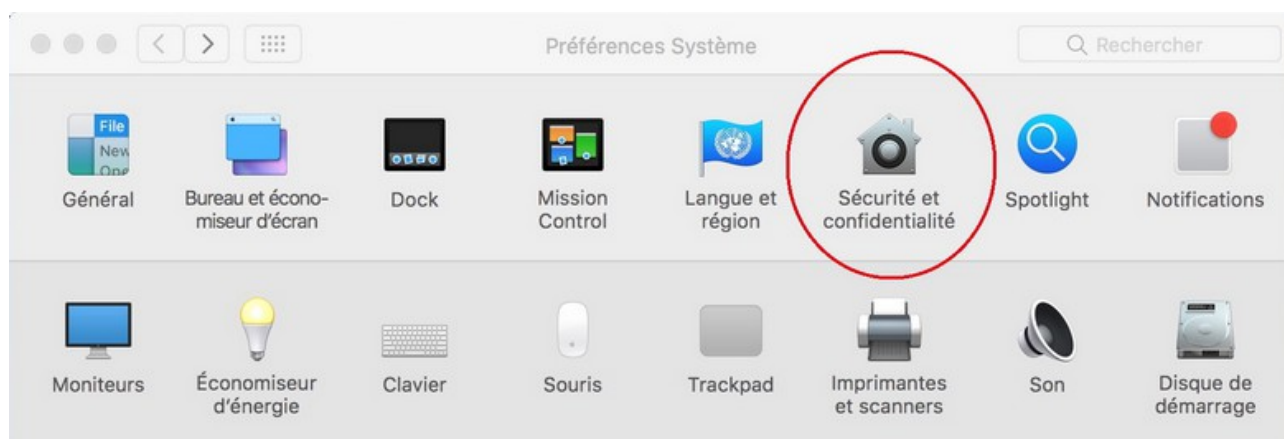
- L'application PdfàSom vers le dossier Applications
- Les fichiers Exemple.som, Exemple_sc.som et Exemple.pdf vers votre dossier de travail (par exemple le dossier PdfàSom à créer dans Documents)

5 Lancement de PdfàSom

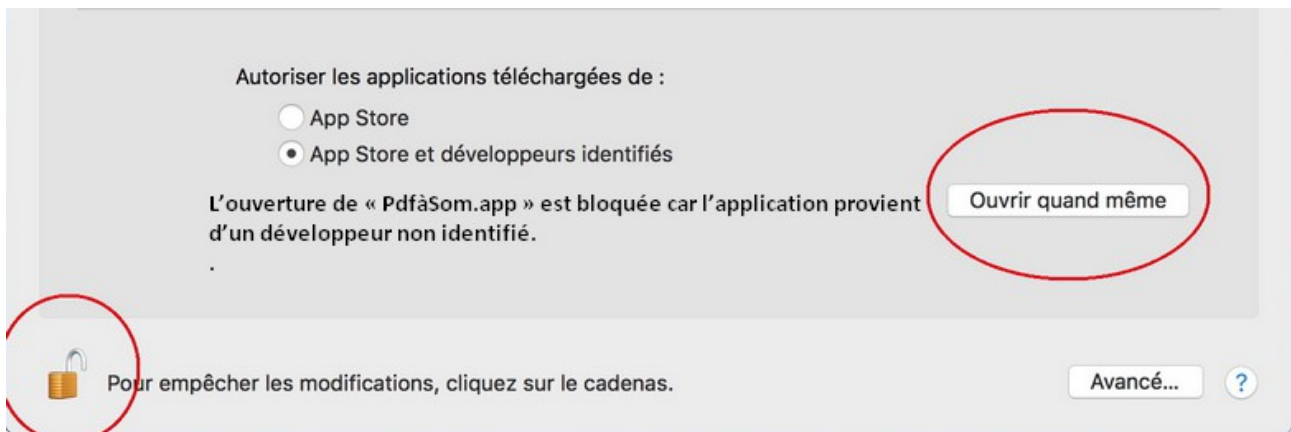
Sous OS X, la première fois qu'on lance PdfàSom on a ce message :



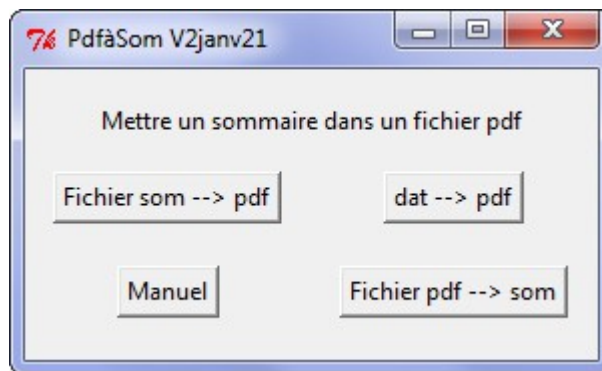
Il faut aller dans 'Préférences Système' et sélectionner 'Sécurité et confidentialité'.



Puis déverrouiller le cadenas avec le mot de passe de la session administrateur et 'Ouvrir quand même' l'application.



Au lancement, PdfàSom se présente de la façon suivante :



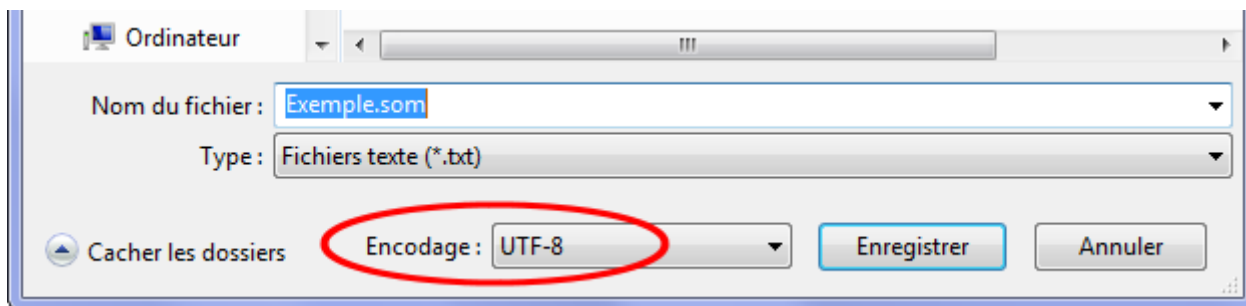
- Le bouton *Manuel* permet d'afficher ce manuel.
- Le bouton *Fichier som --> pdf* permet d'intégrer le sommaire dans le fichier pdf à partir du fichier .som en générant un fichier intermédiaire. Il est demandé de sélectionner le fichier .som puis le fichier pdf. Ouvrir le fichier Exemple.som (fourni avec PdfàSom) avec un éditeur de texte (Bloc-notes de Windows ou TextEdit d'OS X) pour savoir comment doit être renseigné un fichier .som.
- Le bouton *dat --> pdf* permet d'intégrer le sommaire à partir du fichier intermédiaire .dat. Il est demandé de sélectionner le fichier intermédiaire puis le fichier pdf.
- Le bouton *Fichier pdf --> som* permet d'extraire le sommaire d'un pdf dans un nouveau fichier .som dans le but de le modifier. Il est demandé de sélectionner le fichier pdf.

6 Fichier .som

6.1 Création

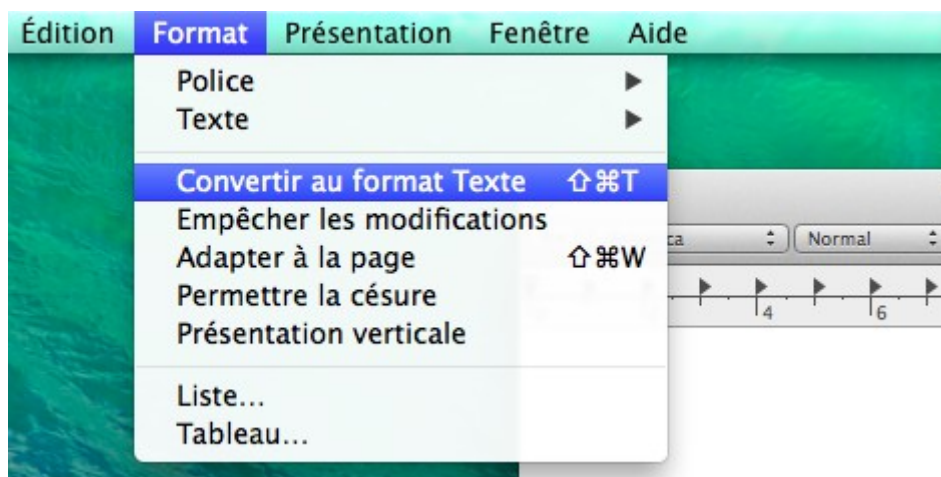
Le fichier .som est un fichier texte qui doit être encodé en utf-8.

Sous Windows, on peut utiliser le Bloc-notes pour le créer ou le modifier en prenant soin de sélectionner l'encodage UTF-8 au moment de l'enregistrement.

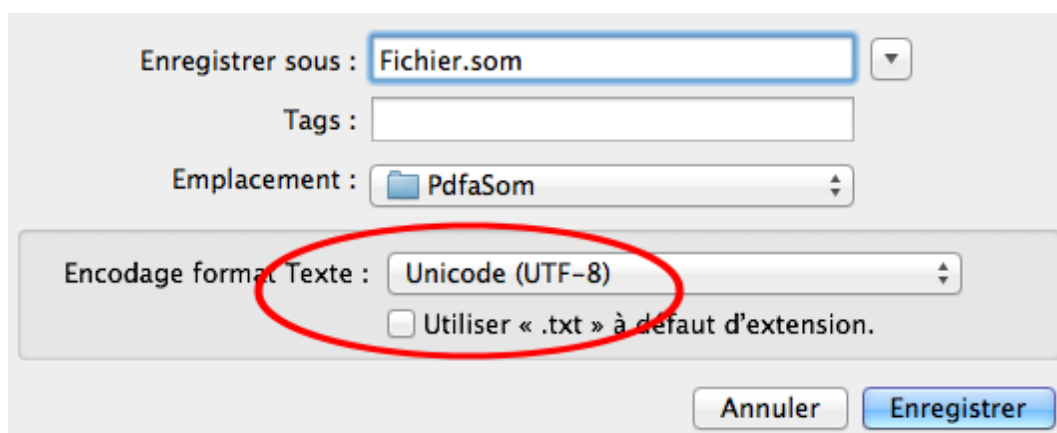


On utilisera aussi avec avantage l'éditeur de texte gratuit et très performant Notepad++

Sur Mac, on peut utiliser TextEdit. Pour créer le fichier .som, il faut utiliser le format texte :



À l'enregistrement, sélectionner le format UTF-8 :



6.2 Généralités - sections obligatoires

Il est possible mettre des commentaires dans le fichier .som, ils ne sont pas pris en compte par PdfàSom et ils permettent de laisser des indications pour faciliter la compréhension ou une future modification du fichier. Si vous ouvrez le fichier Exemple.som fournie avec PdfàSom vous aurez, grâce aux commentaires, toutes les indications nécessaires pour comprendre sa structure. Le fichier Exemple_sc.som est le même fichier débarrassé de tous les commentaires, il donnera exactement le même résultat quand il sera intégré dans le fichier Exemple.pdf par PdfàSom.

La section principale du fichier .som est le sommaire. Il est constitué de la succession des titres de chapitre. Chaque titre occupe une ligne du fichier. Pour chaque titre est associé :

- un niveau hiérarchique défini par la tabulation. Celle-ci doit être effectuée par des espaces ou des caractères de tabulation, il est toutefois fortement recommandé de ne pas mélanger les deux types de caractère. Si vous ne maîtrisez pas ces notions, privilégiez les espaces en prenant par exemple 2 espaces par tabulation, de cette façon :

```
Titre 1 de niveau 1
  Titre 1 de niveau 2
    Titre 2 de niveau 2
      Titre 1 de niveau 3
        Titre 2 de niveau 3
          Titre 3 de niveau 2
            Titre 2 de niveau 1
              Titre 3 de niveau 1
```

Il faut ainsi saisir les titres de niveau 1 directement en début de ligne, précéder tous les titres de niveau 2 de 2 espaces, tous les titres de niveau 3 de 4 espaces, etc.

- Une page qui est indiquée au bout de la ligne de titre et séparée de celui-ci par le caractère '_', de cette façon :
Titre t de niveau n_page p

Les pages peuvent être indiquées de deux manières. La plus simple est un unique nombre entier correspondant au numéro de la page du pdf. Cette possibilité a toutefois l'inconvénient de demander une modification au niveau de chaque titre dès qu'on souhaite réutiliser le fichier .som pour une numérisation différente du même ouvrage. Dans ce cas, en effet, les pages du livre numérisé ne sont pas forcément aux mêmes pages du nouveau fichier pdf.

La seconde façon de renseigner les pages est d'utiliser une table de renvois qui présente plusieurs avantages :

- Au niveau des titres, les pages peuvent être celles de la véritable pagination du livre. Le fichier .som peut ainsi être directement renseigné par un copier-coller de l'OCR de la propre table des matières du livre.
- Pour intégrer le sommaire dans le fichier pdf d'une autre numérisation, il suffit de modifier la table de renvois
- La table de renvois permet de gérer la pagination par folio ou par double page (2 pages du livre sont numérisés sur une seule page du pdf)

La table de renvois est définie dans une section facultative du fichier .som. Les différentes sections du fichier sont séparées par une ligne constituée de 3 tirets : ---. Les sections sont, dans l'ordre :

- La section d'en-tête qui est obligatoire. Elle identifie le type de fichier et indique la présence ou l'absence des sections facultatives.
- La table de renvois qui est facultative. Pour plus de détail, voir le chapitre concerné.
- Une liste facultative des propriétés qui seront consultables avec le lecteur pdf, voir le chapitre concerné
- Le sommaire qui est obligatoire et que nous avons décrit plus haut

6.3 Table de renvois

Chaque ligne de cette table doit comporter 2 parties séparées par le caractère '_'. La partie de droite définit le ou les numéros des pages du pdf qu'on souhaite adresser, cette partie doit donc toujours être un nombre entier. La partie de gauche définit la façon de désigner la ou les pages correspondantes dans la section du sommaire.

Dans le cas le plus simple une ligne de la table définit une seule page du pdf, par exemple :

`ttr_5`

ce qui signifie qu'il faudra écrire `ttr` dans la section du sommaire pour adresser la page 5 du pdf.

La partie de gauche peut utiliser tous les caractères sauf '_' et ':'.
La partie de droite définit le numéro de la première page du pdf correspondant à la première page de la plage, les numéros sont ensuite incrémentés pour chaque nouvelle page de la plage. Quant à la partie de gauche, elle est constituée de 2 sous-parties séparées par le caractère '!'.

Dans le cas le plus complexe on peut définir une plage de pages dont la numérotation est régulière. La partie de droite définit le numéro de la première page du pdf correspondant à la première page de la plage, les numéros sont ensuite incrémentés pour chaque nouvelle page de la plage. Quant à la partie de gauche, elle est constituée de 2 sous-parties séparées par le caractère '!'.

La sous-partie de gauche définit 3 éléments qui sont écrits successivement sans séparateur :

- Un identificateur de la plage qui est facultatif. Cet identificateur devra être rappelé systématiquement dans la section du sommaire, il permet notamment de distinguer différentes portions de l'ouvrage numérisé qui ont une pagination similaire. L'identificateur peut utiliser tous les caractères sauf les chiffres, '_' et '!'.
- Le premier numéro de la plage, c'est obligatoirement un nombre entier
- Le type de pagination qui sera détaillé dans les chapitres suivants :
 - classique
 - par folio
 - par double page

La sous-partie de droite définit la dernière page de la plage, c'est obligatoirement un nombre entier.

6.3.1 Pagination classique

C'est la pagination que l'on trouve dans tous les ouvrages modernes : chaque page est numérotée dans l'ordre de lecture. C'est aussi la pagination par défaut de PdfàSom, il n'y a donc rien à préciser pour utiliser cette pagination. Ainsi, par exemple, la ligne suivante de la table de renvois :

`Page 3:5_9`

signifie qu'il faudra écrire `Page 3`, `Page 4` et `Page 5` dans la section du sommaire pour adresser respectivement les pages 9, 10 et 11 du pdf.

6.3.2 Pagination par folio

C'est la pagination que l'on trouve fréquemment dans les ouvrages du XVI^e siècle. C'est la feuille qui est numérotée et il faut distinguer le recto du verso. Pour préciser que l'on utilise ce type de pagination il faut ajouter un 'r' à la fin de la sous-partie de gauche. Ainsi la ligne suivante :

`fol.2r:3_7`

signifie qu'il faudra écrire `fol.2r`, `fol.2v`, `fol.3r` et `fol.3v` dans la section du sommaire pour adresser respectivement les pages 7, 8, 9 et 10 du pdf.

6.3.3 Pagination par double page

C'est la pagination qui devra être utilisée quand la numérisation comporte 2 pages successives de l'ouvrage pour 1 page de pdf. Pour préciser que l'on utilise ce type de pagination il faut ajouter un 'd' à la fin de la sous-partie de gauche. Ainsi la ligne suivante :

`p.3d:5_9`

signifie qu'il faudra écrire `p.3`, `p.4`, `p.5` et `p.6` dans la section du sommaire pour adresser respectivement les pages 9, 9, 10 et 10 du pdf.

6.4 Propriétés

À chaque ligne de cette section est associée une propriété. La ligne commence par le nom de la propriété et se termine par sa valeur, celle-ci est séparée du nom par le caractère '_'.

Les noms de propriété et leurs attributions sont définis par le format pdf et sont résumés dans le tableau suivant :

Nom	Fonction
Author	Auteur du document
Subject	Sujet du document
Keywords	Mots clefs associés au document
Creator	Application qui a créé le document original
Producer	Application qui a converti le document en PDF
CreationDate	Date de création du document
ModDate	Date de la dernière modification du document

Si on veut attribuer la date et l'heure courante aux propriétés CreationDate ou ModDate, il faut leur donner pour valeur : '\d'.

7 Historique des versions

V30nov20 :

- Version initiale

V2janv21 :

- Correction de bogues
- Meilleure gestion des erreurs