



Sujet sur Wikisource:Forum des nouveaux/Flow

[Ajouter des langues](#)

[Sujet](#) [Discussion](#)

[Voir l'historique](#) [Outils](#)

[< Wikisource:Forum des nouveaux/Flow](#)

Djvu obligatoire ou pdf possible ?

12 commentaires • il y a 5 ans

Valp

Bonjour, Je domine la production de djvu à partir de pdf, car j'ai ABBYY. Mais djvu est moins bon, et surtout perd les index. La question est : peut-on sur Wikisource se contenter de pdf, formater un texte et peaufiner l'OCR à partir d'une source en pdf ?

il y a 5 ans

ElioPrri

Bonjour Valp !

Le consensus actuel est de privilégier le plus possible les DjVu, et de demander si besoin la conversion du PDF en DjVu à un autre contributeur (par exemple [ici](#)). À cela plusieurs raisons :

- les fichiers DjVu sont, de manière générale, beaucoup plus légers que les fichiers PDF à résolution équivalente, donc plus facilement importables et téléchargeables, et demandant moins de temps pour afficher les pages ;
- de plus, une fois le fichier téléchargé, il existe beaucoup de logiciels libres (DjVuToy, ou, si on ne craint pas les lignes de commandes, DjVuLibre) pour modifier les DjVu (par exemple si un contributeur veut intervertir, ajouter ou supprimer des pages, les faire tourner, etc., après import du fichier), en tout cas beaucoup plus que pour les PDF, qui demandent souvent de passer par un site Internet, où le temps de traitement peut facilement exploser avec la taille du fichier ou la mauvaise qualité de la connexion ;
- ensuite (voilà le motif avancé plus spécifiquement sur Commons) les fichiers PDF ont tendance à conserver des données sur leur créateur ou leur dernier modificateur, surtout si on se sert de logiciels propriétaires, comme Adobe, données qui pourraient être récupérées à des fins commerciales ou malveillantes ;
- enfin, les quelques inconvénients pouvant avoir lieu dans la conversion (perte des index, notamment, que vous soulignez, mais aussi des renvois internes ou des signets) sont vraiment inoffensifs pour Wikisource, puisque ici on ne consulte jamais les fichiers eux-mêmes, mais seulement une page à la fois, et que, ainsi, on ne fait usage ni des signets, ni des hyperliens, qui pourraient être présents dans le fichier originel.

Certes, il peut être utile de conserver dans certains cas le fichier PDF, souvent de meilleure résolution, par exemple pour en extraire les illustrations. Mais même dans ce cas, il est souvent encore meilleur de

télécharger la page intéressante directement sur le site de la bibliothèque, qui proposent de plus en plus des téléchargements en HD pour les pages isolées.

Cependant, il me semble que le consensus peut évoluer rapidement, et que plusieurs contributeurs souhaiteraient comme vous voir les PDF acceptés comme format standard des fac-similés.

En espérant ne pas avoir dit trop de bêtises 😊

Modifié il y a 5 ans

Valp

Bonjour @ElioPrri et merci, Je n'ai pas la prétention de vouloir faire accepter pdf comme "standard" ! Je demande seulement s'il est permis et possible d'effectuer des Transclusions à partir d'un pdf. A ce stade, je crois comprendre que oui. Pouvez-vous le confirmer ?

il y a 5 ans

ElioPrri

Cela est possible, mais déconseillé, autant ici que sur Commons. Honnêtement, à part un gain de temps pour vous (puisque vous savez convertir en DjVu ; mais rien ne vous empêche quand même de demander la conversion à un autre contributeur), je ne vois aucun avantage à télécharger en PDF. Y a-t-il une raison spécifique qui fasse que vous soyez obligé de téléverser un PDF ? Pourriez-vous mettre en lien le livre que vous souhaitez importer ?

Modifié il y a 5 ans

Valp

@ElioPrri : La réduction en djvu altère la qualité des scans de vieux livres et leur OCR. L'indexation est utile quand il s'agit d'une compilation de lettres (correspondance). L'argument de la taille des fichiers est obsolète. Le format pdf a de facto cessé d'être propriétaire. Donc, sans besoin d'acquérir le statut de format "standard", pourquoi le format pdf continuerait-il à être "déconseillé" ?

il y a 5 ans

ElioPrri

Je ne peux que me répéter. Les PDF sur Commons ont toujours une icône « attention » signalant les problèmes de sécurité et de confidentialité de ce format ; parce que le format a beau être libre maintenant, les logiciels pour éditer ne le sont pas tous et peuvent encore encoder des données sur leur utilisateur. De même, si un contributeur futur doit modifier votre fichier (pages inversées, OCR), il vaut mieux que celui-ci soit un DjVu, format pour lequel on dispose de logiciels libres et de page de tutoriel sur Wikisource.

La limitation de la taille des fichiers reste importante (tout le monde ne dispose pas de connexions Internet rapides, les pages des PDF se chargent moins vite que les pages des DjVu, et il existe encore une limite de 100Mo sur Commons, aisément atteignable par les fac-similés en couleur de la BnF, par exemple), et la compression est inoffensive (sauf pour les images, qui de toute façon doivent être importées à part) : il suffit que le fichier soit lisible, sans se soucier de son plus ou moins grand niveau de détail. La compression, et le passage en noir et blanc, sont même souvent souhaitables, car, outre le gain de poids, cela améliore le contraste et, donc, la reconnaissance de caractères et la lisibilité.

Il est tout à fait possible de conserver dans la conversion en DjVu la couche texte si elle est de très bonne qualité (PDF to DjVu Gui, dont je fais usage régulièrement) ou la table des matières (PDF & DjVu Bookmaker, que je ne connais que de nom) du PDF — sachant que la table, les signets, les index vous seront inutiles, car vous ne pourrez pas consulter le fichier PDF entier sur Wikisource et y naviguer dedans, mais seulement le voir page à page, sous forme de JPG, sans liens internes ou hyperliens.

Il faut garder en tête que nous ne sommes pas une archive de numérisations (comme Internet Archive ou Hathitrust), visant la reproduction la plus exacte des *images*, mais que nous sommes une bibliothèque numérique, destinée à donner le *texte* des livres : les images ne sont que les outils nécessaires à la fiabilité de nos transcriptions, ce ne sont pas un but mais un moyen. On privilégie donc leur lisibilité et leur légèreté à leur fidélité au livre physique. Voilà pourquoi les fichiers entiers ne sont pas disponibles ni à la consultation ni au téléchargement sur Wikisource : il faut passer par Commons, ou, mieux, par la source indiquée dans la page Livre:. Ce sont les archives numérisées qui doivent se préoccuper de la qualité des fac-similés, pas nous.

Ces raisons font que le format DjVu est préférable. Mais vous pouvez tout à fait importer un PDF.

Modifié il y a 5 ans

Valp

@ElioPrri : La limitation à 100 Mo m'a convaincu et je vous remercie de vos explications.

Maintenant, pourriez-vous m'aider à créer une page des matières dans la page wikisource.../Livre:*nomfich*.djvu' ?

Je dispose de la liste de mes signets : nom du signet / page du djvu.

il y a 5 ans

ElioPrri

Je vais prendre un exemple pour être plus clair, avec [Livre:Hugo - Œuvres complètes, Impr. nat., En voyage, tome I.djvu](#) ; excusez-moi par avance si je vous explique quelque chose que vous connaissez déjà 😊. En gros, il faut procéder comme suit :

1. Commencez par repérer les numéros de page de la table des matières du livre, et par numéro de page, entendre non pas ceux imprimés au haut des pages, mais le numéro de page dans le fichier DjVu — dans cet exemple, 573 à 576. Pour cela, il est bon de commencer par modifier le champs « Pages » de la page Livre:, ce qui est maintenant assez intuitif avec le nouvel outil « Mode visuel pour modifier les pages ».
2. dans la page Livre:, il faut ensuite ajouter au champ « Sommaire », autant de fois que nécessaire, `{{Page:nomfich.djvu/n}}`, où *n* décrira les numéros de page du point précédent — à ce moment là, vous avez fini de modifier la page Livre: ; des liens rouges doivent apparaître dans la colonne de droite
3. ensuite, il faut créer chacune des pages de sommaire ; dans mon exemple, voici [la première](#) ; chaque entrée de la table des matières (avec titre et numéro de page) utilise le modèle `{{table}}`, dont je vous invite à lire la documentation. Il peut être utile de faire la mise en forme de manière presque automatique, en utilisant astucieusement la fonction Rechercher/Remplacer, ou ce [tableur partagé](#) .

Ce que je peux vous proposer, est de téléverser votre livre sur Commons, et je pourrais vous donner un coup de main pour mettre en route tout ça 😊 ; rien de bien difficile, mais il me sera plus commode de

donner des explications sur du concret. Par ailleurs, il ne faut pas hésiter à consulter des livres dont la table des matières ressemble au vôtre, et à en inspecter le code, et à le copier et ajuster à votre livre.

il y a 5 ans

Valp

@ElioPrri : Je vois. Je vous signale ce bon logiciel pour extraire la liste des signets d'un pdf : jpdfbookmarks-2.5.2 (portable ; menu Tools / Dump). Ensuite c'est un jeu d'enfant de construire les `{{Table | ... }}` qui feront le sommaire.

Je vous dirai quand j'aurai téléversé. Cordialement.

Modifié il y a 5 ans

Valp

@ElioPrri, c'est fait : [Livre:Leibniz-en.francais-Gerhardt.Math.1a7.djvu](#)

Un point qui m'échappe : peut-on modifier le sous-titre dans <pages HEADER ?

Modifié il y a 5 ans

ElioPrri

Ah je vois... Normalement on ne fait pas des montages comme ça ; plutôt, on importe les tomes entières, dont on ne corrige que les pages en français, en marquant les autres pages « Sans Texte » et en renvoyant au projet Wikisource de la langue correspondante grâce à `{{iwpage}}` (par exemple [cette revue](#), avec un article [en français](#) et le reste [en italien](#)). Cela permet de ne pas créer de doutes : le fac-similé est-il bien intègre, a-t-il été manipulé, etc. ?

Autre remarque : le [titre des fac-similés](#) est relativement stéréotypé : `Livre:Auteur - Titre, éd. Éditeur, t. Tome, Année.djvu` (les mentions éditeur et tome étant facultatives, et leur ordre peut fluctuer) ; cela permet de faire des recherches facilement pour trouver un fac-similé.

Je vois que nul besoin est de vous apprendre quoi que ce soit, car tout est bien rempli. 😊 Pour les transclusions, vous trouverez toutes les informations [ici](#). Il ne faut pas hésiter à consulter les tutoriels et les aides, qui mériteraient certes pour beaucoup un coup de jeune, mais qui dépannent vraiment au début, notamment [Aide:Publier un livre](#) et [Aide:Transclusion](#).

il y a 5 ans

Valp

@ElioPrri, oui je sais, normalement on ne fait pas des compilations comme cela, mais vu l'espèce d'occultation dont est victime cette partie de son oeuvre ("Newton...") alors qu'elle est, je crois, la plus actuelle et importante (en l'état de la science physique actuelle), je me permet ce petit coup de pouce.

L'important est de ne pas laisser de doute, donner le moyen de vérifier que le fac-similé est intègre, non manipulé, et c'est pourquoi je demandais s'il est possible de modifier un peu le sous-titre dans <pages HEADER. Car par exemple dans le 1er item, au lieu de "(p. 2-6)" je voudrais pouvoir figurer "(GM3a, p. 52-58)" ce qui serait plus correct.

il y a 5 ans

Répondre à « Djvu obligatoire ou pdf possible ? »

Les textes sont disponibles sous [licence Creative Commons Attribution-partage dans les mêmes conditions](#) ; d'autres conditions peuvent s'appliquer. Voyez les [Conditions d'utilisation](#) pour plus de détails.

[Politique de confidentialité](#) [À propos de Wikisource](#) [Avertissements](#) [Code de conduite](#) [Développeurs](#) [Statistiques](#)

[Déclaration sur les témoins \(cookies\)](#) [Version mobile](#)

