# Auto Multiple Choice - Bug # 728: taille fichiers pdf trop élevée avec pdfform

| | | | |
|---|---|---|---|
| **Status:** | Closed | **Priority:** | High |
| **Author:** | Julien Varaldi | **Category:** | LaTeX |
| **Created:** | 05/07/2020 | **Assignee:** | |
| **Updated:** | 04/20/2021 | **Due date:** | |

**Description:**

Bonjour,
j'utiliser AMC avec l'option pdfform pour faire des pdfs cliquables, avec quelques champs ouverts (\Textfield) et une pre-association avec les etudiants. Je crée ainsi des copies uniques (car tirage dans nombres aleatoires avec FP) pour mes 190 etudiants. J'ai rencontré plusieurs probleme sur ce chemin.
1- j'ai du passer a xelatex car avec les autres compilateurs, j'obtenais systematiquement l'erreur "tex capacity exceeded, sorry [pdf buffer=5000000]". Ca marche pour 10 etudiants avec n'importe quel compilateur, mais pour les 190, je n'ai trouve que cette solution
2- quand je compile avec un mini fichier csv (pour tester la procédure, avec 10 etudiants), j'obtiens des pdfs individuels de taille tout a fait raisonnable (300kb). Par contre, quand je compile avec le csv contenant l'ensemble des 190 etudiants, les pdfs pèsent 3.4 Mb chacun, en dépit du fait qu'ils ont la meme allure que les premiers...!
Je ne me l'explique pas et ca pose probleme ensuite pour le telechargement des etudiants.
3- J'avais jusque la l'habitude de modifier le bareme apres la tenue de l'examen. Ici, j'ai beau essayer de modifier le barème (dans le tex source et en cochant mise a jour des baremes), rien ne change.

Pourriez-vous m'aider (surtout pour les points 2 et 3... ca me serait bien utile en vu des examens de la fin du mois...!)
Merci beaucoup de votre aide précieuse
Je joins le tex source
Julien
PS: j'utilise la version dev sur mac os 10.15 Catalina, que j'ai pu installer grace a vos conseils !

## History

**05/08/2020 02:57 pm - Frédéric Bréal**

Pour le point 2, je ne suis pas qualifié mais pour le 3, pouvez-vous préciser un ou deux exemples qui ne fonctionnent pas : numéro de la question dans le tex et les modifications ?

**05/08/2020 06:47 pm - José OUIN**

Bonjour,

Il manque le fichier "liste_Genetique2.csv" pour pouvoir tester votre fichier "source.tex".
J'ai également un problème de dépassement de capacité avec pdfform + listecsv + pythontex + pré-association.
https://project.auto-multiple-choice.net/boards/4/topics/9733

Cordialement.
J/O

**05/08/2020 07:14 pm - Julien Varaldi**

*- File liste_Genetique2.csv added*

Bonjour

merci pour votre reponse.

Je viens de realiser que je ne suis pas capable de reproduire ce que je pensais avoir observé…(pour le point 3). Donc ce point est caduque : la mise a jour du bareme fonctionne tres bien sur Linux et Mac. Désolé, je n'aurais peut etre pas du envoyer ce mail si tardivement…

Pour les deux autres points par contre, je maintiens mes dires (je viens de refaire un test pour le 2eme point) !

Entre temps, j'ai pu constater un autre point qui n'est pas optimal pour les etudiants : les textes que les etudiants ont rentré dans les champs libres (\TextField) disparaissent de leur copie corrigée. Ce n'est pas dramatique, puisque ce qui compte c'est la boite cochée par le correcteur, mais il me semble que ca serait plus satisfaisant pour l'etudiant de pouvoir retrouver sa composition. Savez-vous si c'est possible ?

J'ai mis le fichier liste_Genetique2.csv en attaché (en anonymisant les noms des etudiants...)

Merci encore.
Julien

**05/09/2020 08:56 am - José OUIN**
*- File S__lection_059.png added*

Bonjour,

Il manque une (des) image(s) (voir capture).
Je pense que je vais arriver au même constat que vous.
J/O

**05/09/2020 03:26 pm - Julien Varaldi**
*- File EXEMPLE.png added*

*- File genealogie.pdf added*

Bonjour,
voici les fichiers manquants. Désolé.
Cordialement
Julien

**05/14/2020 06:40 pm - Hugues Bodiguel**
*- File THM2020_MF1.tex added*

Bonjour,

J'ai exactement le meme problème (point 2).
La taille des PDF individuels augmente significativement avec le nombre d'étudiants (100ko pour 10 copies, 300ko pour 50 copies, 1Mo pour 300).

Cdt,

Hugues

**12/18/2020 10:57 pm - Alexis Bienvenüe**
Pour le problème 2, cela vient de la méthode de découpage du PDF global utilisée par AMC : qpdf. En effet, lors du découpage, qpdf extrait quelques pages du PDF global, mais conserve toutes le données de formulaire (pdfform) du gros PDF, même si elles ne sont plus visibles. Et je n'arrive pas à faire mieux avec ghostscript ou pdftk…
pdfsam, mentionné par Joss OUIN (message#10578) semble faire le découpage correctement, mais je ne sais pas l'appeler (lui ou la bibliothèque java

"sejda" qu'il utilise) en ligne de commande. Si vous savez faire, je suis très intéressé pour une intégration dans AMC.

**12/18/2020 11:17 pm - Frédéric Bréal**

Est-ce que cette vidéo peut être une piste https://www.youtube.com/watch?v=eW8drPwWKV0 ?

**12/18/2020 11:34 pm - Alexis Bienvenüe**

@sejda-console@ est une piste, mais il y a me semble-t-il du travail : https://bugs.debian.org/cgi-bin/bugreport.cgi?bug=884377

**12/19/2020 04:14 pm - Alexis Bienvenüe**

Toujours pour le point 2:
@sejda-console@ n'est pas empaquetée dans debian, mais je propose une installation manuelle.
<pre>
cd
wget https://github.com/torakiki/sejda/releases/download/v3.2.85/sejda-console-3.2.85-bin.zip
unzip sejda-console-3.2.85-bin.zip
sudo ln -s ~/sejda-console-3.2.85/bin/sejda-console /usr/local/bin
</pre>
Ensuite, avec la version de développement d'AMC, il suffira de choisir @sejda-console@ pour _Préférences/Général/Impression/Méthode d'extraction_.

**12/30/2020 05:55 pm - Frédéric Bréal**

Pourquoi ne pas faire ceci ?
<pre>
sudo mv sejda-console-3.2.85 /usr/local/bin
</pre>

Avec le lien, il se peut que certain(e)s, moi le premier, efface le dossier sedja au bout d'un temps certain en oubliant son utilité.

**01/28/2021 08:50 pm - Alexis Bienvenüe**

*- Status changed from New to Closed*

*- % Done changed from 0 to 100*

**02/19/2021 12:47 pm - Alessandra Frabetti**

Bonjour,

Malgré l'installation de @sejda-console@, cette méthode d'extraction n'apparaît dans _Préférences/Général/Impression/Méthode d'extraction_.
Comment faire pour la sélectionner?

Cordialement,
Alessandra

**02/19/2021 02:32 pm - Frédéric Bréal**

Avez-vous installé la version test ?

**02/19/2021 03:41 pm - Alessandra Frabetti**

Quelle version test ? de quoi ?

Gérard Carpeaux wrote:
> Avez-vous installé la version test ?

**02/19/2021 03:55 pm - Alessandra Frabetti**
Ah, j'ai tout relu, c'est la version test de AMC que je dois installer ?
Bon, j'essaye.

**02/19/2021 03:57 pm - Alessandra Frabetti**
Non pardon, avant d'essayer, si j'installe la version de développement d'AMC, est-ce que je peux encore corriger les projets en cours ? ou bien il vaut mieux que j'attends de ne pas avoir de projets en cours ?

Alessandra Frabetti wrote:
> Ah, j'ai tout relu, c'est la version test de AMC que je dois installer ?
> Bon, j'essaye.

**02/19/2021 04:13 pm - Frédéric Bréal**
> ou bien il vaut mieux que j'attends de ne pas avoir de projets en cours ?
C'est préférable.

**02/19/2021 04:55 pm - Alessandra Frabetti**
Vu l'impasse où je me trouve, j'ai préféré courir le risque de devoir refaire les projets en cours pour pouvoir préparer les suivants urgents (ça roule à pleine vitesse dans nos UE) et voilà, ça marche !!!

La taille de mes 430 sujets pdf-éditables est passée de 1.5 MB à 150 kB chacun, c'est magique ! :-)

Merci beaucoup ! et espérons qu'il n'y a pas de gros bugs dans la version test ;-)

Alessandra

Gérard Carpeaux wrote:
> > ou bien il vaut mieux que j'attends de ne pas avoir de projets en cours ?
> C'est préférable.

**02/24/2021 10:20 am - Joël Brogniart**
Quand AMC est configuré avec qpdf, la commande générée pour le découpage des sujets est la suivante (aux chemins d'accès et aux numéros de page près).

<pre>
qpdf DOC-sujet.pdf --pages DOC-sujet.pdf 1-5 -- copie-0001.pdf
</pre>

J'ai fait quelques essais en modifiant les options qpdf sur des fichiers AMC-TXT avec 10, 100 et 300 copies et un fichier généré à partir des fichiers ci-dessus avec 10 et 188 copies. J'ai joué avec les options qpdf suivantes (extrait de la doc qpdf associée).

@--linearize@
Causes generation of a linearized (web-optimized) output file.

@--empty@
In place of infilename, the parameter --empty may be specified. This causes qpdf to use a dummy input file that contains zero pages.

@--remove-unreferenced-resources=yes@
Starting with qpdf 8.1, when splitting pages, qpdf is able to attempt to remove images and fonts that are not used by a page even if they are referenced

in the page's resources dictionary. When shared resources are in use, this behavior can greatly reduce the file sizes of split pages, but the analysis is very slow. In versions from 8.1 through 9.1.1, qpdf did this analysis by default. Starting in qpdf 10.0.0, if auto is used, qpdf does a quick analysis of the file to determine whether the file is likely to have unreferenced objects on pages, a pattern that frequently occurs when resource dictionaries are shared across multiple pages and rarely occurs otherwise. If it discovers this pattern, then it will attempt to remove unreferenced resources. Usually this means you get the slower splitting speed only when it's actually going to create smaller files. You can suppress removal of unreferenced resources altogether by specifying no or force it to do the full algorithm by specifying yes.

h3. Résultat

L'option @--empty@ est l'option qui donne des fichiers sensiblement de la même taille quel que soit le nombre de copies du fichiers source. En plus dans le cas des fichiers générés à partir des fichiers ci-dessus, le pdf généré s'ouvre et s'affiche beaucoup plus rapidement sur ma machine. Le fichier généré avec la commande originale d'AMC est très long à apparaitre et la navigation n'est pas rapide.

Par contre je n'ai pas fait de tests extensifs pour le remplissage du formulaire mais j'ai pu correctement cocher quelques cases.

**02/24/2021 10:25 am - Joël Brogniart**

Et j'ai oublié, avec @--empty@ ça donne la commande suivante.

<pre>
qpdf --empty --pages DOC-sujet.pdf 1-5 -- copie-0001.pdf
</pre>

**04/20/2021 07:14 pm - Franck Pommereau**

Bonjour,

Quand j'utilise qpdf avec l'option --empty, les cases à cocher n’apparaissent plus.
Est-ce que j'ai raté quelque chose ?

(En revanche, avec sejda-console le résultat est impeccable, ce qui me donne déjà une solution.)

Franck

## Files

| | | | |
|---|---|---|---|
| source.tex | 26.1 kB | 05/07/2020 | Julien Varaldi |
| liste_Genetique2.csv | 6.6 kB | 05/08/2020 | Julien Varaldi |
| S__lection_059.png | 168.5 kB | 05/09/2020 | José OUIN |
| EXEMPLE.png | 44.7 kB | 05/09/2020 | Julien Varaldi |
| genealogie.pdf | 14.4 kB | 05/09/2020 | Julien Varaldi |
| THM2020_MF1.tex | 3.4 kB | 05/14/2020 | Hugues Bodiguel |