## *Entropics*

Compte-rendu de recherche des produits visant à convertir des documents en pdf

## *Exploration des différentes possibilités de conversion automatique de documents en format PDF sur un serveur Windows*

### **Hugo PONS, 7 Juin 2021**

# Introduction

Le système de gestion de documents d’Entropics est désormais conçu pour fonctionner sur site web, et il faut permettre aux utilisateurs de pré-visualiser leurs documents facilement sur cette nouvelle interface. La grande majorité des fichiers à pré-visualiser sont des documents texte, tableurs ou bien présentation de la suite Microsoft Office (Word, Excel et Powerpoint). Pour les afficher, le choix a été pris de convertir tous ces fichiers en PDF, qui est un format de fichier imprimable et lisible par beaucoup de lecteurs.

**Formats d’entrée :** .docx .xlsx .pptx (formats MS Office)

.odt .ods .odp (formats OpenDocument) et plus si possible

**Format de sortie :** .pdf

**Requis primordiaux :**

• doit fonctionner en mode service,

• doit être fiable et/ou réputé,

• ne doit pas planter

# Grille d’évaluation

Les critères qui ont été retenus pour comparer les différents outils sont les suivants :

* Robustesse (fiabilité et présence de bugs/plantages)
* Prix - Licence
* Formats d’entrée
* Performance (Vitesse et volume de conversion)
* Réputation
* Intégration
* Outils tiers requis
* Support
* Autres fonctionnalités - Description
* Précision

L’échelle d’évaluation pour les critères de stabilité, performance, réputation, précision et support est :

Médiocre, Mauvais(e), Moyen(ne), Bon(ne), Très bon(ne)

Le critère d’intégration indique le travail à effectuer en développement et implémentation afin d’utiliser le produit pour répondre aux besoins expliqués plus haut. Ses valeurs possibles sont Facile (pour des logiciels répondant directement au besoin par exemple), Moyenne (outils finis mais où il reste du travail d’intégration) et Difficile (pour les SDK entre autre).

# Produits

## 1. PDFCreator Server

## Résumé et lien

<https://www.pdfforge.org/pdfcreator/editions/pdfcreator-server>

Logiciel réputé et stable, le plus connu. Version PDFCreator Server qui permet d’utiliser le logiciel en mode service sur un serveur Windows. Cela ne demande pas d’outils tiers à part .NET, il faut simplement un serveur Windows 2008 et plus, et configurer le dossier destination de la queue d’impression

PDFCreator vend le logiciel de conversion, avec des mises à jour, un support mail…

## Politique et prix

960,00€ /par serveur par an

## Avantages

Bonne réputation, performances hautes pour un volume élevé de documents,

## Inconvénients

Prix relativement élevé, logiciel de haut niveau = pas de contrôle

## 2. pdffiller

## Résumé et liens

Pdffiller est un outil qui permet de convertir des fichiers Office en PDF, et qui propose une API pour les développeurs (Java ou PHP).

« Convert PDFs to and from DOCX, XLSX, PPTX, TXT and JPEG formats. »

A priori les formats disponibles en entrée contiennent bien Docx, Xlsx et Pptx mais il faudrait vérifier.

DocIO : <https://help.syncfusion.com/file-formats/docio/overview> pour Word

XlsIO : <https://help.syncfusion.com/file-formats/xlsio/overview> pour Excel

Presentation : <https://help.syncfusion.com/file-formats/presentation/overview> pour PowerPoint.

Pas d’interface utilisateur, support 24h/24

Fonctionne indépendamment des logiciels de la suite Office et utilise le Framework .NET

Fondé en 2001, pour l’instant compatible avec toutes les versions de Microsoft Word, Microsoft Excel et Microsoft Powerpoint.

## Politique et prix

1. ~~Pour utiliser ces bibliothèques il faut une licence Syncfusion « File Formats » qui se trouve dans la partie « Individual Platforms ».~~

~~Le prix est de 995$/an la première année puis 445$/an, par développeur.~~

~~Si beaucoup (plus de 3 / 4) de développeurs, voir option 2.~~

1. ~~Dans la partie Unlimited License, Syncfusion propose des licences illimitées et globales à une entreprise, prix inconnu.~~

La licence Community Edition est gratuite et éligible aux petites entreprises avec 5 développeurs ou moins et un chiffre d’affaires inférieur à $1,000,000 USD.

## Avantages

Coût plutôt faible si peu de développeurs (ou licence illimitée bien négociée ?), implémentation bas niveau donc plus de possibilités et de contrôle sur l’outil et ses, performances.

Pas d’expiration de licence à priori.

## Inconvénients

Pas sûr que ces bibliothèques soient mise à jour en même temps que la suite MS Office. Besoin d’implémenter nous même le service Windows qui utilisera cet outil pour convertir les fichiers, donc plus de travail.

## 3. Syncfusion

## Résumé et liens

Syncfusion est un éditeur de composants logiciels, qui propose une base de bibliothèques codées en C# et qui utilisent le framework .NET. Parmi toutes ces bibliothèques, certaines proposent des outils de manipulation de fichiers Word, Excel ou encore Powerpoint. En particulier, des fonctions permettent de convertir les fichiers en pdf.

DocIO : <https://help.syncfusion.com/file-formats/docio/overview> pour Word

XlsIO : <https://help.syncfusion.com/file-formats/xlsio/overview> pour Excel

Presentation : <https://help.syncfusion.com/file-formats/presentation/overview> pour PowerPoint.

Pas d’interface utilisateur, support 24h/24

Fonctionne indépendamment des logiciels de la suite Office et utilise le Framework .NET

Fondé en 2001, pour l’instant compatible avec toutes les versions de Microsoft Word, Microsoft Excel et Microsoft Powerpoint.

## Politique et prix

1. ~~Pour utiliser ces bibliothèques il faut une licence Syncfusion « File Formats » qui se trouve dans la partie « Individual Platforms ».~~

~~Le prix est de 995$/an la première année puis 445$/an, par développeur.~~

~~Si beaucoup (plus de 3 / 4) de développeurs, voir option 2.~~

1. ~~Dans la partie Unlimited License, Syncfusion propose des licences illimitées et globales à une entreprise, prix inconnu.~~

La licence Community Edition est gratuite et éligible aux petites entreprises avec 5 développeurs ou moins et un chiffre d’affaires inférieur à $1,000,000 USD.

## Avantages

Coût plutôt faible si peu de développeurs (ou licence illimitée bien négociée ?), implémentation bas niveau donc plus de possibilités et de contrôle sur l’outil et ses, performances.

Pas d’expiration de licence à priori.

## Inconvénients

Pas sûr que ces bibliothèques soient mise à jour en même temps que la suite MS Office. Besoin d’implémenter nous même le service Windows qui utilisera cet outil pour convertir les fichiers, donc plus de travail.

## 4. E-iceblue Spire

## Résumé et liens

Même principe que Syncfusion.

Pour Word : <https://www.e-iceblue.com/Introduce/word-for-net-introduce.html?gclid=EAIaIQobChMIrvSXv9uF8QIVCmAYCh0lOwLBEAAYASAAEgIi6PD_BwE#.YL4qyfkzZPY>

Pour Excel : <https://www.e-iceblue.com/Introduce/excel-for-net-introduce.html?gclid=EAIaIQobChMIqb_3jN2F8QIVifuyCh2pswFdEAAYASAAEgLx8fD_BwE#.YL4sc_kzZPY>

Pour Powerpoint : <https://www.e-iceblue.com/Introduce/presentation-for-net-introduce.html?gclid=EAIaIQobChMIpLnF9NyF8QIVrACiAx0fWwX_EAAYASAAEgJDRvD_BwE#.YL4sQfkzZPY>

## Politique et prix

1. Idem que précédemment, E-Icecube propose différents packs de licences et licences individuelles, et la mieux adaptée dans notre cas est la licence « Spire.Office for .NET ».

Pour un développeur et pour une « deployment location » (serveur ou application), la licence est à $1,899 et valable à vie, mais il faut la renouveler pour avoir des updates.

Pour un développeur et un nombre illimité de « deployment location », c’est $5,999.

Pour la première option, le prix du renouvellement est de $1139.4 avant expiration et $1424.25 après expiration.

1. Il y a une version gratuite « Free Spire.Office for .NET » qui permet d’utiliser toutes les fonctionnalités de la licence gratuitement, mais il y a des limitations sur les documents :

Pour les fichiers Word, le pdf converti n’aura que les 3 premières pages de disponibles. Pour les Présentations, il y a une limite de 10 slides et tous les pdf sont bloqués à 10 pages de manière générale.

## Avantages

Mêmes avantages que Syncfusion. La licence peut coûter moins cher si renouvelée seulement lorsqu’il y a une mise à jour importante des fichiers Microsoft Office (tous les 4 ans par exemple ce qui ferait environ $474.75)

## Inconvénients

Quasiment équivalent à Syncfusion mais plus cher s’il y a besoin de renouveler.

## 5. CutePDF

## Résumé et liens

Cet outil utilise l’action « imprimer » de n’importe quelle application de l’ordinateur pour rediriger l’impression vers une imprimante virtuelle, qui crée le fichier pdf. C’est donc un peu compliqué à utiliser dans le cas d’un serveur même si la description de la version « PDF Converter SDK (Custom Redistribution) » laisse entendre que cela peut fonctionner en mode service (silencieusement). Requiert un interpréteur PS2PDF.

Le présence

Taille : 8 Mo.

« Allow developers to integrate PDF creation on the fly »

« Seamlessly support Terminal/Citrix Server and non-English Windows system »

<https://www.cutepdf.com/products/sdk/pdfwriter.asp>

## Politique et prix

Une licence par personne ou entreprise à acheter, au prix de $499 (US) pour avoir « Custom PDF Writer with programmatic access ». Cette licence est valable perpétuellement. Royalty free.

Cependant, utilisation de registry keys est à éviter.

## Avantages

Le prix est le plus bas.

## Inconvénients

Utilisation de registry keys à éviter, le support a l’air plutôt léger.

Avis préliminaire

L’outil et le site donnent l’impression d’être une copie de PDFCreator, l’obligation de passer par l’impression des documents et l’utilisation de registry keys font que cet outil ne sera pas retenu.

## 6. Win2PDF

## Résumé et liens

Dans le même style que CutePDF, Win2PDF utilise l’action « imprimer » de n’importe quelle application ou éditeur de texte de la machine pour rediriger l’impression vers une imprimante virtuelle, qui crée le fichier en PDF. Cette solution comporte des risques mais il faudrait tester pour vérifier que cela ne pose pas de problèmes dans le cas d’une utilisation à gros volume. La licence Terminal Server Edition laisse tout de même entendre qu’une utilisation pour un serveur est viable.

<https://www.win2pdf.com/server-features.html>

## Politique et prix

La licence la moins chère pour les serveurs est la licence « Win2PDF TSE » à 411,95€ qui permet de convertir les documents en pdf, mais ne permet pas de protéger avec un mot de passe ou encore d’ajouter un filigrane. Pour le faire il faudra choisir la licence « Win2PDF Pro TSE » à 782,70€.

Une licence est perpétuelle, mais l’accès à un support prioritaire est payant (20% du prix de la licence pour 1 an).

« The Win2PDF Server products do not support printer sharing »

## Avantages

Le coût de la licence n’est pas très élevé et il n’y a pas besoin de MS Office ou même .NET pour fonctionner.

## Inconvénients

Cet outil passe aussi par une imprimante virtuelle et la fonction imprimer des applications donc elle ne sera pas facilement utilisable (mais la version est conçue pour marcher sur serveur quand-même).

## 7. Pdf-tools 4-Heights Conversion Service

## Résumé et liens

Convertit de nombreux formats (Office et e-mails entre autre) vers PDF. Prend en compte les dossiers de surveillance, batch et API donc cet outil répond potentiellement au besoin de conversion automatique en tant que service Windows. Utilise un plugin .NET. Ce produit a besoin de MS Office pour convertir les fichiers de Microsoft Office.

[~~https://www.pdf-tools.com/pdf20/fr/produits/pdf-converter-validation/document-converter/~~](https://www.pdf-tools.com/pdf20/fr/produits/pdf-converter-validation/document-converter/)

<https://www.pdf-tools.com/pdf20/fr/produits/pdf-converter-validation/conversion-service/>

## Politique et prix

La politique et la licence sont encore à voir car la société propose de prendre contact et ne fait que des offres sur mesure.

## Avantages

Dans l’Entreprise Edition une API et un service Internet sont disponibles donc facilité d’intégration dans un environnement existant.

## Inconvénients

A besoin de MS Office pour fonctionner.

## 8. JODConverter

## 9. Aspose

## Résumé et liens

Aspose.Total propose des API de manipulation de formats pour plusieurs environnements. En particulier, elle propose des librairies .NET de manipulation des formats Word, Excel et Powerpoint et de les convertir en PDF.

<https://products.aspose.com/total/net/>

## Politique et prix

Pour un développeur et une location de déploiement, la licence permettant d’utiliser les bibliothèques Aspose pour Word Excel et PowerPoint avec .NET est à $2999 avec support gratuit, et la licence est perpétuelle (mises à jour pour un an seulement).

Les versions « Free Trials » sont limitées (150 appels à l’API par mois) donc pas très intéressantes sachant qu’il y a la « Temporary License » qui permet de tester tous les packages que l’on souhaite gratuitement pendant un certain temps.

## Avantages

Aspose est très utilisé et réputé pour convertir les fichiers avec beaucoup de précision. Le support semble réactif et l’outil semble fiable.

## Inconvénients

Le prix est tout de même très élevé en comparaison avec d’autres outils qui pourraient faire la même chose.

## 10. Conversion passant par HTML

## Résumé et liens

Cette solution est gratuite et ne passe pas par un logiciel /une bibliothèque payante mais elle demande plus de travail et ne sera peut-être pas assez efficace ou précise. Elle fonctionnerait pour la conversion d’un fichier docx et il faut encore regarder/tester pour la conversion de fichiers Excel et Présentation.

Dans le cas d’un fichier Docx, il faut d’abord convertir ce document en HTML. Pour ce faire nous utiliserons le package open source OpenXMLSDK-PowerTools pour .NET. Ensuite, il faut convertir ce nouveau fichier en un document final en PDF, en utilisant une bibliothèque libre également telle que DinkToPdf.

<https://stackoverflow.com/questions/46580718/convert-word-doc-and-docx-format-to-pdf-in-net-core-without-microsoft-office-in>

## Politique et prix

Ce procédé n’utilise que des packages et bibliothèques libres de droit donc il est totalement gratuit.

## Avantages

Il n’y a pas de licence à payer et il est possible de gérer directement quels formats sont pris en compte.

## Inconvénients

Il y a des risques que cela ne fonctionne pas facilement ou que la précision finale du pdf ne soit pas exceptionnelle (mais jusqu’où la précision est-elle importante dans le cadre d’une prévisualisation ?). Il faudra donc tester et développer pour avoir un résultat satisfaisant. Il n’y a pas de support ou de service de maintenance non plus en cas de bug. Pas sûr que cela fonctionne pour les formats autres que docx. Peu robuste.

11. Conversion avec LibreOffice

## Résumé et liens

LibreOffice permet de convertir des fichiers de Microsoft Office en PDF. Actuellement il n’y a pas d’API pour .NET donc il faut utiliser l’exécutable *soffice* en ligne de commande pour convertir. LibreOffice peut normalement bien fonctionner en service sur un serveur Windows donc il répond bien au problème.

## Politique et prix

Comme précédemment, LibreOffice et soffice en particulier sont des outils libres et donc gratuits d’utilisation. Il n’y a pas de licence à acheter.

## Avantages

Idem, solution gratuite et également plutôt stable (LibreOffice est très utilisé). Intégration assez facile et autres fonctionnalités sur les documents. Beaucoup de formats en entrée (MS Office, Open Document, HTML, …)

Inconvénients

Solution déjà testée à Entropics et qui ne peut pas être retenue car les conversions plantent de temps en temps. Le service devait être tué à la main, et pendant ce temps les conversions en attente n’étaient pas traitées. La précision n’était bonne non plus.

12. Foxit Reader

## Résumé et liens

Foxit Reader est un outil permettant à la base de lire des PDF (Foxit Reader), mais il existe un PDF SDK permettant de convertir les fichiers Office en PDF. Quick PDF Library 18 est une version allégée et beaucoup moins onéreuse qui pourrait répondre à nos besoins.

<https://www.debenu.com/products/development/debenu-pdf-library/>

Dans les termes de la licence, le support technique n’a pas l’air assuré du tout et il n’y a pas de garantie.

Pas sûr non plus que cet outil puisse convertir des formats Office vers PDF. La feature principale est « Create, render and print PDFs ». Donc je suppose que l’outil passe également par la fonction *imprimer* déjà présente dans les logiciels d’édition de documents (besoin d’en avoir un sur le serveur).

## Politique et prix

La licence Quick PDF est à $499 pour un développeur et une plateforme, et elle est royalty free et perpétuelle.

## Avantages

Ce produit ne coûte pas très cher et pourrait potentiellement répondre au problème simplement.

## Inconvénients

Il faut tester pour vérifier que ce SDK permette bien de convertir les fichiers comme il faut. Il n’y a pas de garantie et le support semble payant et peu présent. Cela implique également de développer le service et l’outil final soi-même.

## Décision

Le produit de Foxit n’a pas été retenue car elle ne réalise en fait pas vraiment de conversion mais manipule seulement des fichiers PDF.

?. Microsoft Office

## Résumé

Office n’est pas retenu dans cette recherche d’outil car c’est un logiciel vendu pour client et non pour serveur, ce qui implique des problèmes de licence et de stabilité pour l’intégrer en tant que service (ce n’est pas fait pour).

13. OSDC (Rainbow Office)

Ce logiciel répond parfaitement aux besoins, il est adapté aux serveurs et à la conversion de gros volumes de documents MS Office, sans recourir à ces logiciels ni même à Adobe Software.

<https://www.antennahouse.com/pricing#osdc>

## Politique et prix

La licence « Server License » perpétuelle est à $5000 (pour un processeur), avec $1000 en option pour une maintenance annuelle. Sinon la même licence coûte $2675 à l’année et comprend déjà le service de maintenance

## Avantages

Ce produit semble parfaitement adapté aux besoins et se concentre uniquement sur la conversion de documents Office en PDF, avec un souci de performance sur serveur. On peut donc s’attendre à ce que la performance et l’intégration soient très simples, et aussi à ce que la robustesse soit bonne.

## Inconvénients

Le prix de la licence est très élevé, le produit est le plus cher de tous. Il n’y a pas beaucoup d’avis également donc difficile de noter le support et la précision par exemple.

# Calcul des notes et coefficients

Dans le but de comparer les différents outils entre eux selon les critères retenus plus haut, j’ai créé un tableur Excel listant pour chaque produit sa valeur par rapport aux critères, selon la grille d’évaluation. Ces descriptions ont ensuite été transformées en notes sur 5 ou sur 3 pour noter numériquement chaque critère. Puis dans une autre feuille, des coefficients sont donnés (avec un coefficient par critère), qui correspondent à leur pondération dans la note finale.

Les notes globales font ainsi ressortir le niveau final de chaque produit et nous ne garderons donc que les meilleurs pour cette première sélection.

# Sélection des produits et conclusion

10/06 Cette première étape de recherche et d’explorations des différentes solutions visant à répondre au problème de conversion nous a amené à évaluer et noter une liste de produits. Le classement final est le suivant :

1. **PDFCreator Server**
2. **Aspose**
3. **Syncfusion**
4. LibreOffice
5. Pdf-tools
6. **OSDC (Rainbow PDF)**
7. E-iceblue Spire
8. CutePDF
9. Win2PDF
10. **Conversion en HTML**
11. Pdffiller

Les 4 outils sélectionnés pour la phase suivante (phase de test) sont donc PDFCreator Server, Aspose, Syncfusion, OSDC (Rainbow PDF). Et potentiellement la conversion par HTML que l’on pourrait tout de même essayer.

<https://blog.muhimbi.com/2009/12/converting-office-files-to-pdf-format.html>

Autre produit : Pdftron

https://www.pdftron.com/documentation/core/guides/features/office/