

Guide d'océrisation des articles en format PDF

Ce guide explique comment utiliser le module d'océrisation afin d'extraire le contenu plein-texte d'un article en format PDF.

Formulaire de description d'un article

Le processus d'océrisation est lancé par un clic sur le bouton [**Extraire le texte complet du PDF**] qui se trouve au bas du formulaire de description d'un article.

Notes : Il est recommandé d'enregistrer les changements à l'article avant de procéder à l'extraction du PDF.

Formulaire d'océrisation

Deux situations peuvent se présenter :

- 1) Le contenu a déjà été océrisé et est visible dans le champ **Contenu** de l'article (voir l'onglet **Contenu** dans le formulaire de description de l'article) :

Si vous souhaitez seulement apporter des corrections au texte, alors cliquez sur l'onglet **WORD** au haut de la moitié gauche du formulaire. Vous y retrouverez le texte déjà océrisé provenant du champ **Contenu** de cet article.

Vous pouvez alors continuer à appliquer vos corrections dans cet onglet (**WORD**).

Lorsque vous aurez terminé vos corrections, cliquez sur le bouton [**Récupérez le travail depuis WORD**]. Le texte dans l'onglet WORD sera alors transféré dans le champ **Résultat** dans la moitié droite du formulaire.

Si la présentation du texte doit encore être modifiée, retournez à l'onglet WORD pour faire vos corrections. Utilisez encore une fois le bouton [**Récupérez le travail depuis WORD**].

Si la présentation du texte vous convient, vous pourrez cliquer sur le bouton [**Fermer**] pour retourner au formulaire de description de l'article. Utilisez alors le bouton [**Valider**] pour enregistrer le résultat de l'océrisation dans le champ **Contenu** de l'article ou le bouton [**Annuler**] pour annuler les changements.

Avertissement : n'utilisez pas le bouton [**Extraire le texte**]. Ce bouton va lancer l'extraction du contenu du PDF et le copier dans l'onglet **WORD**, écrasant ainsi le travail de correction antérieur.

D'ailleurs un message vous en avertira et vous permettra de reculer.

Si vous avez choisi de continuer avec l'extraction mais souhaitez maintenant revenir au texte antérieur, il vous suffira de fermer le formulaire d'océrisation, puis d'ignorer les changements faits au champ **Contenu** via le bouton [**Annuler**] du formulaire de description de l'article.

2) Vous souhaitez extraire le contenu du PDF et le verser dans le champ **Contenu** de l'article :

Dès l'entrée dans le formulaire d'océrisation, cliquez sur le bouton [Extraire le texte]. Cela lancera le processus d'océrisation. Un message apparaîtra au milieu de l'écran pour indiquer que l'extraction est en cours. Cette opération peut être longue, il faudra donc être patient.

Lorsque l'extraction sera terminée, vous verrez le résultat de celle-ci apparaître dans l'onglet WORD dans la moitié gauche de l'écran. C'est à cet endroit que vous ferez toutes les corrections nécessaires.

Cliquez ensuite sur le bouton [**Récupérez le travail depuis WORD**]. Le texte dans l'onglet WORD sera alors transféré dans le champ **Résultat** dans la moitié droite du formulaire.

Si la présentation du texte doit encore être modifiée, retournez à l'onglet WORD pour faire vos corrections. Utilisez encore une fois le bouton [**Récupérez le travail depuis WORD**].

Si la présentation du texte vous convient, vous pourrez cliquer sur le bouton [**Fermer**] pour retourner au formulaire de description de l'article. Utilisez alors le bouton [**Valider**] pour enregistrer le résultat de l'océrisation dans le champ **Contenu** de l'article ou le bouton [**Annuler**] pour annuler les changements.

Les images :

Le portail de recherche permet de chercher un mot ou une expression dans la légende d'une photo ou d'une image. Le bouton [**Générer une balise image**] permet de générer une balise qui indique la présence d'une image dans le texte et qui mentionne la légende utilisée dans l'article.

[{{[image]JÉRÔME LE ROYER DE LA DAUVERSIÈRE}}

Cette balise ne montre pas la photo, uniquement la légende. Ainsi, en cherchant **Royer**, par exemple, et en spécifiant le lieu de recherche comme étant les images, le portail va montrer comme résultat tous les articles contenant une balise mentionnant **Royer** dans la légende. C'est en visitant chaque article du résultat que le chercheur pourra voir les images elles-mêmes dans le document PDF associé.

Pour générer la balise, il faut d'abord sélectionner la légende puis cliquer sur le bouton [**Générer une balise image**]. La balise sera construite comme ci-haut autour de la légende.

Les interlignes (entre les paragraphes) :

Il faut deux interlignes dans l'onglet WORD pour en obtenir un dans le champ **Contenu** de l'article. Le champ **Résultat** dans la moitié droite du formulaire permet justement de vérifier l'espacement correct des paragraphes par une ligne vide. Donc, avant de retourner au formulaire de l'article, il vaut la peine d'examiner le contenu du champ **Résultat**. Si quelque chose doit être modifié, on peut retourner à l'onglet WORD.

Aussi, il est préférable de supprimer les lignes vides au haut du texte, au dessus du titre.

La mention d'auteur :

Le nom de l'auteur, s'il y a lieu, doit paraître sous le titre ou le sous-titre et être précédé de la mention **Par** ou d'une barre oblique / :

Par Richard Ouellet

ou

/ Richard Ouellet.