Pour la réunion du mercredi 16/01 à 14H00.

Ce lot en HD est en double page. Je serai obligé d’océriser à partir des images en diffusion web donc en avec une qualité inférieure

<http://cnumdev.cnam.fr:81/ORIGS/ELECTRICITE/>

<http://cnumdev.cnam.fr:81/ORIGS/2003.02.28.ANNALES/>

On doit vérifier la compatibilité des langues, le nombre de fichier

Translitération

Rendre le volume de Science et vie de Pierre

Récupération de l’information du taux d’erreur et de la langue d’océrisation. Peut-être que cela risque de ralentir l’affichage (à tester)

Verrouiller les textes océrisés soumis au droit d’auteur

Supprimer les textes (et les fichiers xml des taux) océrisés vides comme <http://cnumdev.cnam.fr:81/DEV/CGI/sresrech.cgi?SCVIE.031/0>

On calcule le taux de réduction des images et l’intégrer dans la base de données

Relancer l’OCR sur français ancien / Impossible, l’océrisation bloque lorsque l’on choisit comme langue « français ancien » (il faut que j’enregistre cette phrase dans mes

« La langue de reconnaissance de l'OCR est le français ». Le mot français n’est pas dynamique mais statique

Les dictionnaires ABBYY sont destinés à une utilisation interne uniquement et ne peuvent pas être révélés aux utilisateurs.

D’autre part, les dictionnaires utilisés par Recognition Serveur, pour les langues avec prise en charge d’un dictionnaire, utilisent les mots les plus récurrents de la langue en question (le nombre de mots dépend de la langue), ainsi que les noms propres et noms communs récurrents.

--

Ce texte provient de Gallica :

Le texte affiché peut comporter un certain nombre d'erreurs. En effet, le mode texte de ce document a été généré de façon automatique par un programme de [reconnaissance optique de caractères (OCR)](https://gallica.bnf.fr/html/und/consulter-les-documents). Le taux de reconnaissance estimé pour cette page est de ...%.   
  
La langue de reconnaissance de l'OCR est le français.

---