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

--------------------------------

Présentation de BY :

<http://cnumdev.cnam.fr:81/test/test_echantillon_images/documents_a_oceriser/>

J’ai créé différents workflow pour les langues ; voir les différentes langues sur

----------

J’ai procédé au nettoyage des fichiers corrompus qui ne passait à l’océrisation avec adobe Photoshop et à l’aide du Poste de vérification (logiciel de correction de texte d’ABBYY OCR).

Fichiers corrompus, tâches

Récupération des données sur CD, disque dur

Tâches et fichiers corrompus :

<http://cnumdev.cnam.fr:81/DEV/CGI/redirv.cgi?8KY1.14>

<http://cnumdev.cnam.fr:81/test/test_echantillon_images/documents_a_oceriser/2018.11.07_ocr_a_ranger_plus_tard/Exceptions%20Folder/Ok,%20c'est%20r%C3%A9gl%C3%A9/2013.07.15.azentis.transport.periodique.musee/>

<http://cnumdev.cnam.fr:81/test/test_echantillon_images/documents_a_oceriser/2018.11.07_ocr_a_ranger_plus_tard/Exceptions%20Folder/Ok,%20c'est%20r%C3%A9gl%C3%A9/2007.LUM-ELECTRIQUE/P84.38/0402.png>

<http://cnumdev.cnam.fr:81/test/test_echantillon_images/documents_a_oceriser/2018.11.07_ocr_a_ranger_plus_tard/Exceptions%20Folder/2017.08.09_lot_industriel/P905.3/0265.T.260.png>

quid de la renumérisation de <http://cnumdev.cnam.fr:81/test/test_echantillon_images/documents_a_oceriser/2018.11.07_ocr_a_ranger_plus_tard/Exceptions%20Folder/ORIGS/2014.07.08.safig/lot5/HD_GRIS/POUPEER_dossiers/> ?

--------------

<http://cnumdev.cnam.fr:81/DEV/ocr/integral/xml/4DY4-07/>

Récupération dans les balises des fichiers xml, du taux d’erreur et de la langue sélectionnée pour la reconnaissance optique de caractère. La récupération dynamique des données risque de ralentir l’affichage (à tester)

<http://cnumdev.cnam.fr:81/DEV/CGI/fpage.cgi?SCVIE.001/142/100/180/20/160>

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

<http://cnumdev.cnam.fr:81/DEV/solr_not_entry3/boubacar_test/rech1.php>

J’ai développé un code calculant le taux de réduction des images (HD vers diffusion web) et j’intègrerai les résultats dans la base de données

Développement du code de génération du pdf contenant les mots océrisés géolocalisables sur la page

--------------------------------

Présentation de JPC :

Rendre le volume de Science et vie de Pierre

Ce lot en HD est en double page. J’ai dû océriser à partir des images en diffusion web, donc de qualité inférieure

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

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

L’océrisation avec la configuration langue « français ancien » bloque le logiciel ABBYY

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

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>

Translitération :

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 :

<https://gallica.bnf.fr/ark:/12148/bpt6k12005345/f17.item.texteImage>

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.

---

J’ai enrichi le défilement automatique des pages illustration en récupérant la balise illustration de l’OCR

Si vous souhaitez naviguez dans les fichiers océrisés : <http://cnumdev.cnam.fr:81/DEV/ocr>

Fichiers corrompus à renumériser :

<http://cnumdev.cnam.fr:81/DEV/CGI/fpage.cgi?8KY1.14/21/100/509/0/0>

Le fichier dans le cd 38 est également corrompu, on est obligé de renumériser la page.

/home/cnumdev/httpd/htdocs/ORIGS/2014.07.08.safig/lot5/HD\_GRIS2/POUPEER\_dossiers/

POUPEER54.3\0103.T.97

POUPEER54.11\0365.T.356 et 0366.T.357

POUPEER54.3\0491.T.477 et 0538.T.524

<http://cnumdev.cnam.fr:81/ORIGS/2014.07.08.safig/lot5/HD_GRIS2/POUPEER_dossiers/POUPEER54.3/0103.T.97.png>