Remarques sur l'HTML
de Proudon
Volume 1 de sa correspondances

Stage Lucie Slavik, Labex Obvil
mai-juillet 2020

Mardi 21 juillet 2020

Je compte 79 lettres. Elles ne sont pas numérotées.

Nettoyage de l'HTML
- balises meta à fermer si version oXygen trop ancienne
- br, hr supprimées
- Suppression du début (intro sur sa vie et son œuvre) et fin (table des matières)

Regex pour supprimer les titres de haut de pages
DE P\.-J\. PROUDHON\. [0-9]{1,3} = 88 matches
[0-9]{1,3} CORRESPONDANCE = 108 matches
[0-9]{1,3} CORRESPONDANCE = 4 matches
DE P\.-J\. PROUDHON\. [0-9]{1,3} = 36 matches
DE P\.-J\. PROUDHON\. [0-9]{1,3} = 21 matches
[0-9]{1,3} CORRESPONDANCE = 28 matches
[0-9]{1,3} CORRESPONDANCE = 19 matches
[0-9]{1,3} CORRESPONDANCE = 38 matches (pour recoller les mots séparés : tout rechercher, tout remplacer)
DE P\.-J, PROUDHON\. [0-9]{1,3} = 3 matches
CORRESP\. I\. [0-9]{1,3} = 8 matches
DE P\.-J\. PROUDHON\. [0-9]{1,3} = 2 matches

Le reste fait manuellement.

Enlever les faux paragraphes
( )([a-z]) tout rechercher $2 tout remplacer 201 matches

Corrections de fautes :
no au lieu de ne : 8 matches

Signature : P.-J. PROUDHON.
72 matches pour 79 lettres. On pourrait partir sur cela pour extraire les lettres.

Repérages du cahier des charges

MARQUEURS DES LETTRES

Absence de chiffres qui délimitent les lettres

Début de la lettre :

Lieu et date de l'écriture de la lettre en minuscule
Adresse au destinataire en petites capitales

Fin de la lettre

Signature en petites capitales (P.-J. Proudhon)

- Correspondance active
- Pas de distinction des lettres entre elles
- Alinéas marquent les paragraphes
- Systématiquement la date et le lieux à droite (« Lieu, jj mois AAAA. »)
- Le destinataire
  - indiqué en majuscules s'il est connu (« A M. TONNELIER »)
  - « A M. X*** » lorsque le destinataire n'est pas connu
- Structure de la lettre
  - la formule de politesse (opener/salute) est intégré au paragraphe de début
  - la formule de fin est distinguée par un paragraphe à part

La date est parfois incertaine : il est donc écrit 18..