

Méthodologie

Master : Sciences du langage



marine.delaborde@cyu.fr

Séance 2 : Traitement de texte

Organisation du cours

Liens utiles

- Fiche d'informations (à remplir une seule fois) : <https://forms.gle/tSBzQ6gvZeBVjsH37>
- Site du cours : <https://marine-delaborde.github.io/methodologie>

Les séances

2/4 séances

Séances	Points abordés
Séance 1 : 16/11/2022 13h30 - 15h30	Recherche documentaire
Séance 2 : 13/12/2022 14h30 - 16h30	Traitement de textes
Séance 3 : 18/01/2023 13h30 - 15h30	Interroger des corpus
Séance 4 : 25/01/2023 13h30 - 15h30	Examen

Application

Exercices pour aujourd'hui

1. Identifier 5 références à citer pour **définir** votre sujet de recherche
 - Formuler une ou plusieurs questions auxquelles ces références permettent de répondre
2. Identifier 5 références à citer pour **approfondir** votre sujet de recherche
 - Formuler une ou plusieurs questions auxquelles ces références permettent de répondre
3. Identifier 5 sources permettant de faire de la veille pertinente pour votre sujet de recherche

Le traitement de texte

Pour le mémoire de recherche

- Lequel choisir* ?
 - **Microsoft Word** (Microsoft Office) - Propriétaire
 - **Apple Pages** (iWork) - Propriétaire
 - **Google docs** (Google) - Propriétaire
 - **Writer** (OpenOffice / LibreOffice) - Licence libre
 - **Framapad** (Framasoft) - Licence libre
 - **Cryptpad** - Licence libre
 - ...
 - Autre solution :
 - **Latex** (+Sharelatex par exemple)
- **Templates CY** (pour les sciences de l'éducation mais peut servir d'inspiration) : <https://cyu.libguides.com/education/memoire-master>
 - Solution simple : suivre le modèle en appliquant une feuille de style imposée (Word, LibreOffice, PDF)
 - Penser à changer les logos...

* https://fr.wikipedia.org/wiki/Correspondance_entre_logiciels_libres_et_logiciels_propri%C3%A9taires

Politesse et interopérabilité

Quel format partager (et à qui) ?

- Si le document est finalisé : choisir le format .pdf
 - Seule solution pour conserver exactement la mise en page, les polices de caractères, les images, etc.
 - Quel que soit le système d'exploitation (OS)
- Si le document doit être modifié :
 - Poser la question du format (selon les OS, tous les traitements de texte ne sont pas compatibles)
 - Exemple : on évite d'envoyer un document .pages (Apple) sans avoir posé la question avant
 - Solutions collaboratives : Google doc, Framapad, Cryptpad, ShareLatex, etc.

Nommer un document

Conseils

- Un nom bref et explicite
 - Éviter les mots outils
 - Utiliser des abréviations connues
 - Mettre l'élément important en premier (soit la date, soit le nom)
- Suivi des versions
 - Indiquer un numéro de version dans le titre (VP : version provisoire, VF : version finale), mais pas seulement
 - Mettre la date dans le nom au format année, mois, jour (AAAAMMJJ)
 - Pour que l'affichage par ordre alphabétique soit aussi chronologique
 - Exemple : `memoire_NOM_20221201.pdf`
 - Utiliser un logiciel de gestion de versions comme GIT : <https://git-scm.com/doc>
- De manière générale, éviter caractères spéciaux : les accents, la ponctuation et les espaces
 - Éviter tout ce qui n'est pas alphanumérique pour un affichage correct pour tous les OS
 - Pour séparer les mot, utiliser underscore, tiret ou majuscule

Les expressions régulières = Regex = ER (ou RE)

Faire des recherches puissantes grâce aux patrons de chaînes de caractères

- Utilisation dans Word, LibreOffice, Google doc, etc.
- Outil de recherche dans des textes
- Permet de trouver des séquences correspondant à un **motif** (patron). Ex :
 - Même préfixe / suffixe / contexte
 - Dates
 - Adresses
 - Urls
- Exemples :
 - r?évolution
 - joli.*
 - (monsieur|madame)
 - \b\w+ta\b

➤ **ER** = définition de critères de recherche pour des chaînes de caractères

Expressions régulières			
Caractères spéciaux (espacement)		Ne pas confondre	
\r	Retour en début de ligne (retour chariot)	[]	<u>Ensemble</u> de caractères [^a-z] <u>Exclusion</u> d'un <u>ensemble</u> de caractères
\n	Saut de ligne	()	<u>Suite</u> de caractères
\f	Saut de page	{}	Intervalle (quantifieur)
\t	Tabulation horizontale		ou
\v	Tabulation verticale		
Classes de caractères		Quantifieurs (portent sur le caractère qui précède)	
.	N'importe quel caractère	?	0 ou 1
[a-z]	N'importe quelle lettre minuscule non accentuée	*	0 ou +
\d	Chiffre = [0-9]	+	1 ou +
\D	Tout sauf un chiffre = [^0-9]	Frontières	
\w	Caractère alphanumérique ou _ = [A-Za-z0-9_àéèëïïoùÀÉÊËËÏÔÙ]	^	Début de ligne
\W	Tout sauf un caractère alphanumérique ou _ = [^A-Za-z0-9_àéèëïïoùÀÉÊËËÏÔÙ]	\$	Fin de ligne
\s	Caractère d'espacement = espace + \t\r\n\f	\b	Frontière de mot
\S	Tout sauf un caractère d'espacement	\B	Ni début, ni fin de mot

Application

Exercice (vérifier avec votre traitement de texte)

1. Trouver une expression régulière qui reconnaît à la fois *verts* et *vertes*.
2. Trouver une expression régulière qui reconnaît à la fois *jour* et *journée* mais pas *bonjour*.
3. Que reconnaissent les expressions suivantes ?
 - 3.1. $\text{lo}\{2,6\}\text{in}$
 - 3.2. $(\text{anti}|\text{pré})?\text{révolutionnaires?}$

La bibliographie

Gestionnaire de références bibliographiques

- **Directement dans le traitement de texte**
- (Bibtex avec Latex)
- **Gestionnaires externes :**
 - **Zotero :** <https://bibliotheque.cyu.fr/version-francaise/se-former/tutoriels/outil-zotero/zotero-le-logiciel-de-gestion-bibliographique>
 - Logiciel gratuit de gestion de références bibliographiques
 - Permet de collecter, organiser et exporter des références bibliographiques et de créer une bibliographie de manière automatique
 - Nécessite une installation du logiciel mais aussi le plugin zotero connector
 - Permet d'ajouter automatiquement un article à la bibliothèque lors de la lecture d'un article sur internet par exemple
 - Possible de les ajouter à la main dans le logiciel aussi
 - Connection avec word aussi

La plagiat

Déclaration

- Qu'est-ce que le plagiat ? <http://www.univ-paris3.fr/lutter-contre-la-fraude-les-bonnes-pratiques-numeriques-683470.kjsp>
- Intégrer une déclaration anti-plagiat dans les devoirs / mémoires :

« Je déclare avoir rédigé ce travail sans aides extérieures ni sources autres que celles qui sont citées. Toutes les utilisations de textes préexistants, publiés ou non, y compris en version électronique, sont signalées comme telles. Ce travail n'a été soumis à aucun autre jury d'examen sous une forme identique ou similaire, que ce soit en France ou à l'étranger, à l'université ou dans une autre institution, par moi-même ou par autrui. »

Exemple issu du site internet de l'Université de la Sorbonne Nouvelle

Application

Exercice avec les regex

1. Récupérer le template du mémoire (sciences de l'éducation) sur le site de l'université
2. Ajouter les informations connues liées au mémoire (informations personnelles, sujet, etc.)
3. Supprimer les informations non nécessaires
4. Ajouter la bibliographie de la séance 1 (à l'aide du gestionnaire dans le traitement de texte ou de zotero)

Résultat à m'envoyer par mail avant la prochaine séance (18/01/2023).