

Projet final : Binome / Rapport LaTeX + Présentation + Démo

NLP ⇒ Natural language processing | TAL/N ⇒ Traitement automatique du langage naturel

Librairie python `nltk`

Moodle : ING-LANG

Scrap du Moodle

- Validation sur CC (3 petits projets par thématique) + 1 projet final.
Les projets doivent être rendus au plus tard le 20 mai 2022 et feront l'objet d'une présentation de 8-10' le lundi 23 mai 2022. Aucun délai supplémentaire ne sera donné. Pour ceux qui souhaitent rendre et présenter leur projet avant cette date je vous propose le jeudi 21 avril (vers 16h au bocal : à confirmer via mattermost).
Le cours abordera 3 thématiques :
- 1. *Analyse de données textuelles** (morphologique, grammaticale, syntaxique et sémantique)
 - annotation des textes (balisage)
 - parsing (pour l'annotation automatique)
- 2. *Création corpus**
 - crawling / [scraping](#)
 - création dataset
- 3. *Agents conversationnels**
 - chatbot avec AIML
 - introduction sur un modèle d'apprentissage automatique

Validation : contrôle continu (3 exos sur les trois thématiques) + projet final (les sujets font partie d'une de 3 thématiques : par exemple pour les projets:

a) dans la thématique d'analyse textuelle on peut envisager : reconnaissance automatique des Entités Nommées, reconnaissance automatique des personnages d'une œuvre, reconnaissance des événements chronologiques, résumé automatique, etc.

b) dans la thématique de création de corpus on peut envisager : le recueil de données textuelles dans les sites ciblés de par leur intérêt traitant un sujet particulier, par exemple des données sur projets collaboratifs (crowdfunding), la symptomatique d'une maladie (par exemple maladies cardiaques), etc.

c) dans la thématique de création d'un agent conversationnel on peut envisager des dialogues spécifiques à un domaine pré-défini (par exemple un bot qui explique tout sur la licence informatique de P8, ou un jeux en ligne, ou l'aide pour apprendre à coder un jeu en python ou autre langage de votre choix).

Les sujets sont donnés à titre d'exemple et ne sont pas exhaustifs. Vous devez me présenter votre proposition de projet pour validation.

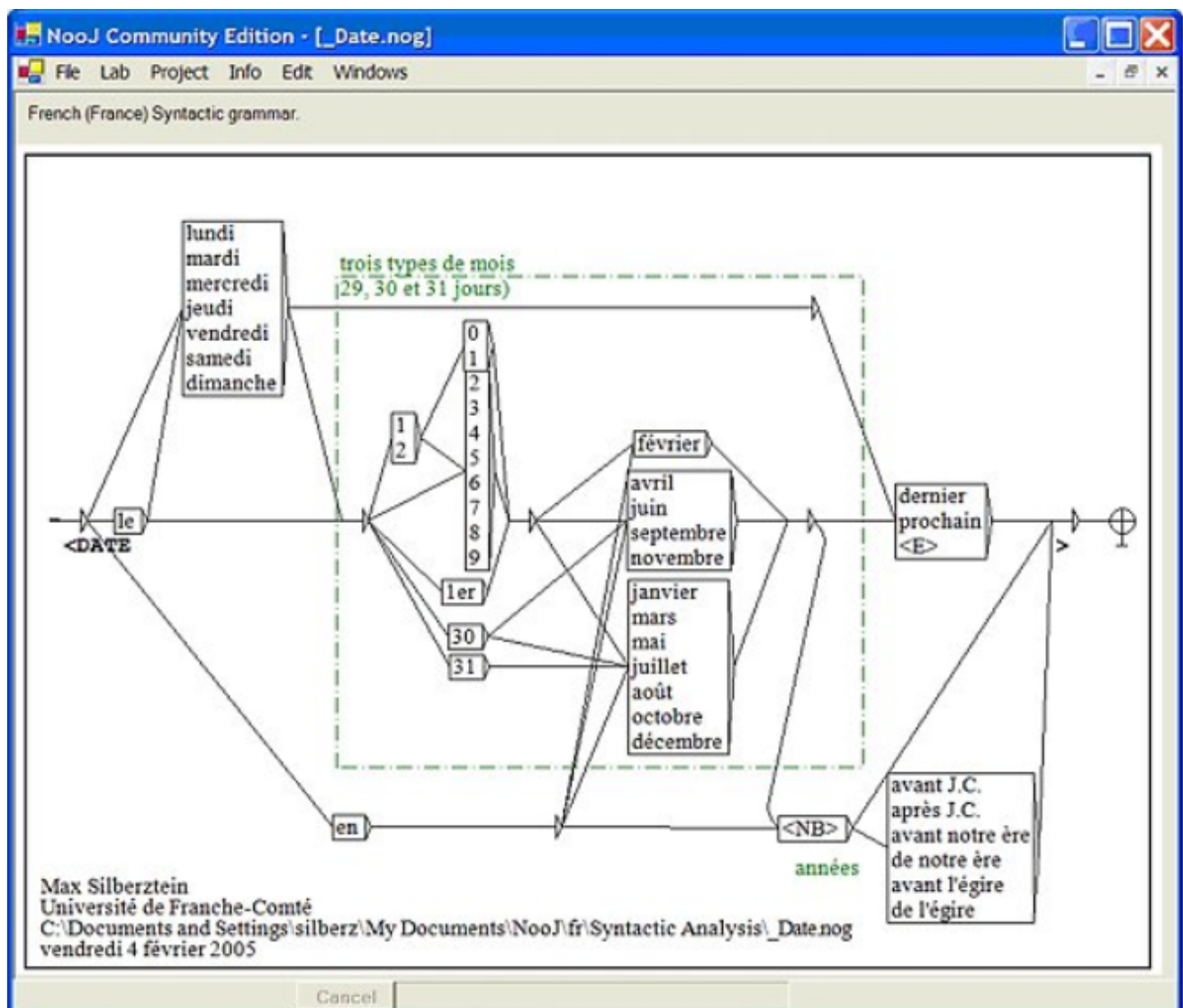
Les projets doivent être rendus au plus tard vendredi 20 mai 2022 et feront l'objet d'une présentation de 8-10' le lundi 23 mai 2022. Aucun délai supplémentaire ne sera donné. Pour ceux qui souhaitent rendre et présenter leur projet avant cette date je vous propose le jeudi 21 avril (vers 16h au bocal : voir messages sur mattermost).

nltk

nooj

TreeTagger

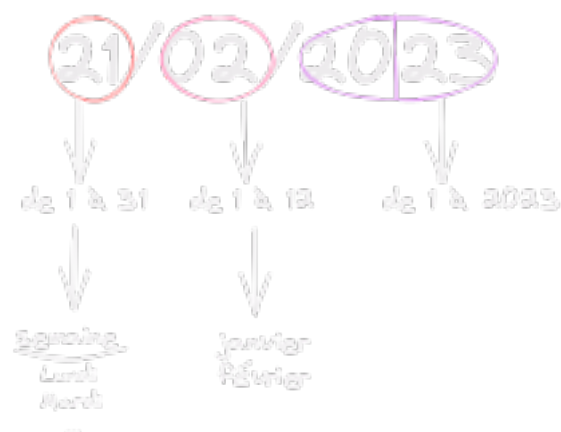
Prof :



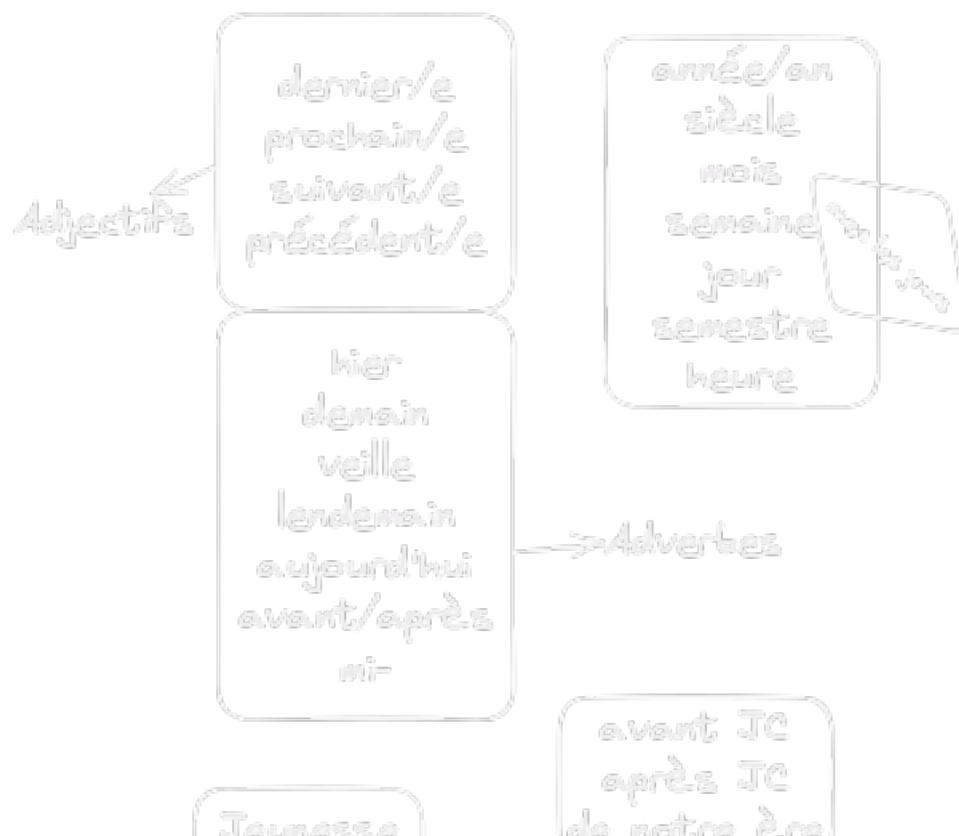
Liste suffixe/préfixe (non exhaustive) :

a -	-eur
in(n) -	-ité
di/e -	-tion
enti -	-esse
contre -	
i -	
mal -	
mé(s) -	
mi -	

Parse de date :



Parse pour la temporalité :



Vieillesse

je suis un

Tokenisation :

Tokens

j'ai vu l'homme en voiture bleue

j'ai

homme