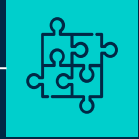


Analyse et implémentation d'un étiqueteur morpho-syntaxique du français

Par :
ALAOUI-YOUSFI Khaoula
CHAMROUK Laïla
ES-SEBBAR Imane
MAJDOUL Kaoutar

TABLE DES MATIÈRES



01

Introduction

Présentation du sujet, définition du cahier des charges



02

Conception

Présentation des données



03

Réalisation

Présentation de l'aspect technique du projet ; code



04

Gestion de projet

Plannings, choix techniques



05

Conclusion

Perspectives et démonstration

Introduction

01

ÉTUDE DU SUJET

- Analyse du langage naturel devient de plus en plus important
- L'évolution des technologies implique une demande plus forte pour l'étiquetage des mots



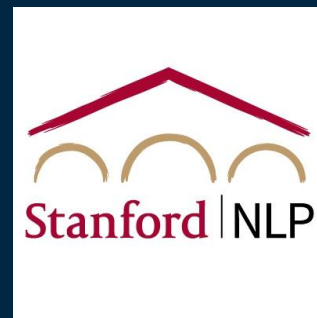
ÉTUDE DU SUJET

- Étiquetage morpho syntaxique = Part Of Speech tagging : attribuer une étiquette grammaticale à chaque mot d'un texte
- Travail complexe : des ambiguïtés existent au sein du français, on doit donc s'intéresser à la définition de chaque mot ET à leur contexte



ANALYSE DE L'EXISTANT

- NLTK
- Stanford Tagger
- Les tagger ne sont pas courant en français



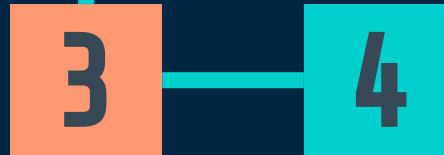
CAHIER DES CHARGES

Analyse et
compréhension
du sujet



Prise en main
des données

Conception



Implémentation

Conception

02



ANALYSE DES DONNÉES

- Données principales : JeuxDeMots
- Réseau lexical RezoJDM



Derniers fichiers disponibles par type de relation (51 fichiers)

- r_raff_sem (1) 11112020-LEXICALNET-JEUXDEMOTS-R1.txt (2,25 MB)
- r_raff_morpho (2) 11122020-LEXICALNET-JEUXDEMOTS-R2.txt (3,79 MB)
- r_domain (3) 11102020-LEXICALNET-JEUXDEMOTS-R3.txt (64,72 MB)
- r_pos (4) 06092020-LEXICALNET-JEUXDEMOTS-R4.txt (345,57 MB)
- r_syn (5) 11182020-LEXICALNET-JEUXDEMOTS-R5.txt (87,41 MB)

ANALYSE DES DONNÉES : exemple sur le mot chat

Tableau_noeud

```
// les noeuds/termes (Entries) : e;eid;'name';type;w;'formatted name'
e;150;'chat';1;5342
e;171869;'Adj: ';4;50
e;146888;'Adj:Mas+SG';4;50
e;146882;'Ver:Inf';4;50
e;212235;'Ver: ';4;50
e;2586627;'Gender:Mas';4;0
e;2586629;'Number:Sing';4;0
e;146885;'Nom:Mas+SG';4;50
e;171870;'Nom: ';4;50
```

Tableau_relation

```
// les relations sortantes : r;rid;node1;node2;type;w
r;12088360;150;171869;4;-60
r;12088359;150;146888;4;-55
r;390106829;150;146882;4;-25
r;390106830;150;212235;4;-25
r;53971059;150;2586627;4;45
r;53985457;150;2586629;4;45
r;213702;150;146885;4;104
r;484206;150;171870;4;345
```

Tableau noeud et relation liés

```
// les noeuds/termes (Entries) : e;eid;'name';type;w;'formatted name' // les relations sortantes : r;rid;node1;node2;type;w
e;150;'chat';1;5342
e;171869;'Adj: ';4;50
e;146888;'Adj:Mas+SG';4;50
e;146882;'Ver:Inf';4;50
e;212235;'Ver: ';4;50
e;2586627;'Gender:Mas';4;0
e;2586629;'Number:Sing';4;0
e;146885;'Nom:Mas+SG';4;50
e;171870;'Nom: ';4;50
r;12088360;150;171869;4;-60
r;12088359;150;146888;4;-55
r;390106829;150;146882;4;-25
r;390106830;150;212235;4;-25
r;53971059;150;2586627;4;45
r;53985457;150;2586629;4;45
r;213702;150;146885;4;104
r;484206;150;171870;4;345
```

FONCTIONNEMENT GÉNÉRAL

- En entrée : une phrase ; liste ordonnée de séquences POS valides (exemple : "*le chat mange*" a pour séquence valide "*det: nom: ver:*")
- Attribuer les POS aux mots qui n'ont qu'un POS unique
- Pour chaque mot de la phrase :
 - on recherche la séquence valide en partant du premier mot



Réalisation

03

RÉALISATION

1. Introduction

❖ Exemple:

Entrez la phrase:

le petit chat boit du lait

@@@@@ RESULTAT FINAL @@@@@

Sequence valide :

[['Det: ', ' Adj: ', ' Nom: ', ' Ver:Conjug ', ' Det: ', ' Nom: ']]

RÉALISATION

1. Fonction extraction

❖ JeuxDeMots (REZO-DUMP)

```
// les noeuds/termes (Entries) : e;eid;'name';type;w;'formatted name'
e;5714;'sortir';1;792
e;171870;'Nom: ';4;50
e;146885;'Nom:Mas+SG';4;50
e;2586627;'Gender:Mas';4;0
e;2586629;'Number:Sing';4;0
e;212235;'Ver: ';4;50
e;146882;'Ver:Inf';4;50

// les types de relations (Relation Types) : rt;rtid;'trname';'trgpname';'rthelp'
rt;4;'r_pos';'POS';Partie du discours (Nom, Verbe, Adjectif, Adverbe, etc.)

// les relations sortantes : r;rid;node1;node2;type;w
r;32444212;5714;171870;4;-51
r;32638300;5714;146885;4;-51
r;54164584;5714;2586627;4;-45
r;54204700;5714;2586629;4;-45
r;12110063;5714;212235;4;56
r;53439;5714;146882;4;109
```

RÉALISATION

1. Fonction extraction

❖ Filtrage

```
{  
['r', '12110063', '5714', '212235', '4', '56']  
['r', '53439', '5714', '146882', '4', '109']  
}
```

```
{  
['e', '5714', "'sortir'", '1', '792']  
['e', '212235', "'Ver:'", '4', '50']  
['e', '146882', "'Ver:Inf'", '4', '50']  
}
```

RÉALISATION

2. POS Unique

Les identifiants de noeuds positifs (noeuds figurant dans au moins une relation positifs):
['212235', '146882', '5714']

```
tabPOS = ["146889", "171869", "150504", "147628", "171870", "146911", "147826", "212235", "2354314"]
# adverb 146889
# adjective qualificative 171869
# conjunction of coordination, and of subordination 150504
# determinant, 147628
# common noun, proper noun, 171870
# preposition, 146911
# pronoun, 147826
# verb. 212235
# punctuation : 2354314
```


RÉALISATION

3. POS Multiple

le terme

Options

☐ Pas de relations sortantes ☐ Pas de relations entrantes ☐ Format com

Filtrage type relations : + -

Filtrage valeur : min max

Filtrage type noeuds : + -

Présentation de sortie : (ex: -rien-, cloud, nicecloud)

'bien'

(id=114700 ; fe=bien ; type=1 ; niveau=48.5432 ; luminosité=3057 ; somme entrant

≈ 8 relations sortantes

1. bien -- r_pos #4: 160 -> Adv:
2. bien -- r_pos #4: 106 -> Nom:
3. bien -- r_pos #4: 62 -> Adj:
4. bien -- r_pos #4: 60 -> Nom:Mas+SG
5. bien -- r_pos #4: 57 -> Adj:InvGen+SG:InvGen+PL
6. bien -- r_pos #4: 45 -> Gender:Mas
7. bien -- r_pos #4: 45 -> Number:Sing
8. bien -- r_pos #4: 25 -> Int:

RÉALISATION

3. POS Multiple

Entrez la phrase:

le petit chat

le ::

POS_MULTIPLE

#####

petit ::

POS_MULTIPLE

#####

chat ::

POS_UNIQUE

#####

RÉALISATION

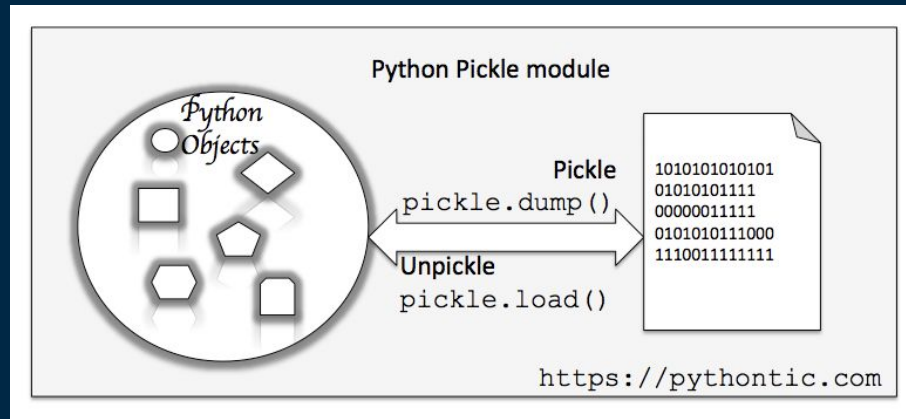
3. POS Multiple

```
le :: Number:Sing , Gender:Mas , Pro: , Pro:3Mas+SG , Det:Mas+SG , Pro:Pers , Pro:Pers:COD , I  
petit :: Gender:Mas , Number:Sing , Nom: , Adj:Mas+SG , Adv: , Nom:Mas+SG , Adj: , ;  
chat :: Gender:Mas , Number:Sing , Nom:Mas+SG , Nom: , ;
```

RÉALISATION

4. Mise en place de la mémoire cache

- Utilisation du module Pickle



RÉALISATION

4. Mise en place de la mémoire cache

```
if cache:
    chemin_absolu = os.path.dirname(os.path.abspath(__file__))
    if not os.path.isdir(chemin_absolu + '/cache'):
        try:
            os.mkdir(chemin_absolu + '/cache')
        except OSError:
            print('La création du dossier cache a échoué')

    fichier_cache = open(chemin_absolu + '/cache/' + word + '.pkl', 'wb')
    pickle.dump([categorie, noeudsPositifs, id], fichier_cache)
    fichier_cache.close()
    pos_unique(id, noeudsPositifs)
    #print("cat : ", categorie)

tableau_noeuds.clear()
tableau_relations.clear()
#return categorie
return "#####"
```



```
def extraction_cache(word: str, cache: bool):
    chemin_absolu = os.path.dirname(os.path.abspath(__file__))
    if not cache:
        return extraction(word, cache)
    elif cache and (not os.path.isdir(chemin_absolu + '/cache') or not os.path.isfile(
        chemin_absolu + '/cache/' + word + '.pkl')):
        return extraction(word, cache)
    elif cache:
        fichier = open(chemin_absolu + '/cache/' + word + '.pkl', 'rb')
        categorie = pickle.load(fichier)
        fichier.close()
        pos_unique(categorie[2], categorie[1])
    return "#####"
```

RÉALISATION

5. Séquences valides

Etape	Phrase	Mot traité et ses tags	Séquences valides	Réaction du programme
1	Le chat boit du lait	Le : Det: Det:Mas+SG Gender:Mas Number:Sing	Pro: Ver:Conjug Det: Nom: Det: Adj: Ver:Conjug Det: Nom: Det: Nom: Ver:Conjug Det: Nom:	Incorrect, je passe à la séquence suivante
2	Le chat boit du lait	Le : Det: Det:Mas+SG Gender:Mas Number:Sing	Pro: Ver:Conjug Det: Nom: Det: Adj: Ver:Conjug Det: Nom: Det: Nom: Ver:Conjug Det: Nom:	Correct, je continue avec cette séquence
3	Le chat boit du lait	chat : Nom: Nom:Mas+SG Gender:Mas Number:Sing	Pro: Ver:Conjug Det: Nom: Det: Adj: Ver:Conjug Det: Nom: Det: Nom: Ver:Conjug Det: Nom:	Incorrect, je reprend depuis le début avec la séquence suivante
4	Le chat boit du lait	Le : Det: Det:Mas+SG Gender:Mas Number:Sing	Pro: Ver:Conjug Det: Nom: Det: Adj: Ver:Conjug Det: Nom: Det: Nom: Ver:Conjug Det: Nom:	Correct, je continue avec cette séquence
5	Le chat boit du lait	Le : Det: Det:Mas+SG Gender:Mas Number:Sing	Pro: Ver:Conjug Det: Nom: Det: Adj: Ver:Conjug Det: Nom: Det: Nom: Ver:Conjug Det: Nom:	Correct, je continue avec cette séquence
6	Le chat boit du lait	chat : Nom: Nom:Mas+SG Gender:Mas Number:Sing	Pro: Ver:Conjug Det: Nom: Det: Adj: Ver:Conjug Det: Nom: Det: Nom: Ver:Conjug Det: Nom:	Correct, je continue avec cette séquence

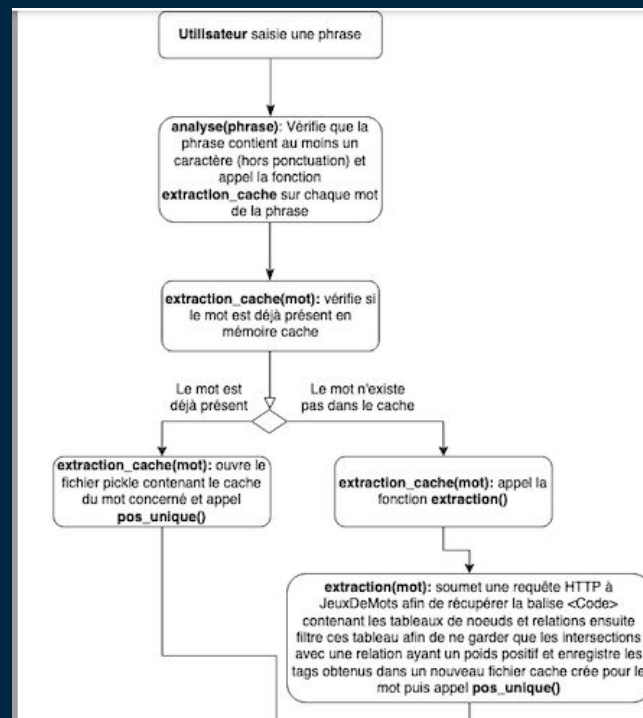
RÉALISATION

5. Séquences valides

7	Le chat boit du lait	boit : Ver:Conjug Ver: Ver:IPre+SG+P3 VerbalTime:Present VerbalMode:Indicatif VerbalNumber:SG VerbalPers:P3	Pro: Ver:Conjug Det: Nom: Det: Adj: Ver:Conjug Det: Nom: Det: Nom: Ver:Conjug Det: Nom:	Correct, je continue avec cette séquence
8	Le chat boit du lait	du : Pre:Det: Pre: Det: Det:Mas+SG Gender:Mas Number:Sing Adj:	Pro: Ver:Conjug Det: Nom: Det: Adj: Ver:Conjug Det: Nom: Det: Nom: Ver:Conjug Det: Nom:	Correct, je continue avec cette séquence
9	Le chat boit du lait	lait : Nom: Nom:Mas+SG Gender:Mas Number:Sing	Pro: Ver:Conjug Det: Nom: Det: Adj: Ver:Conjug Det: Nom: Det: Nom: Ver:Conjug Det: Nom:	Correct, je continue avec cette séquence
10	Le chat boit du lait	//FIN	Det: Nom: Ver:Conjug Det: Nom:	Je retourne cette séquence qui est valide

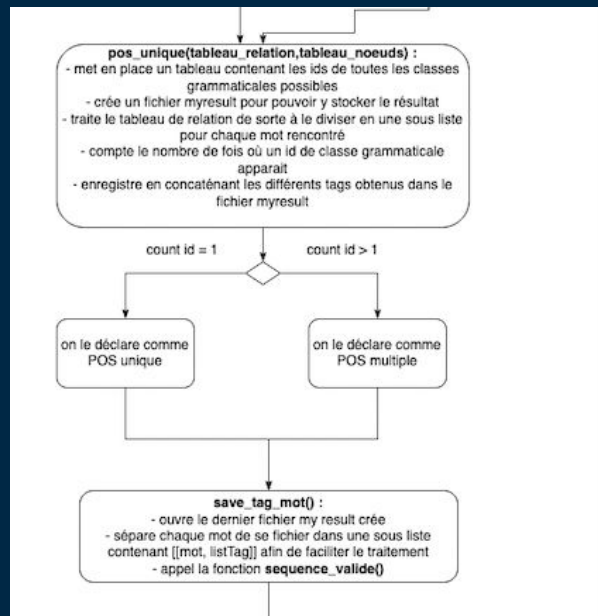
RÉALISATION

5. Déroulement



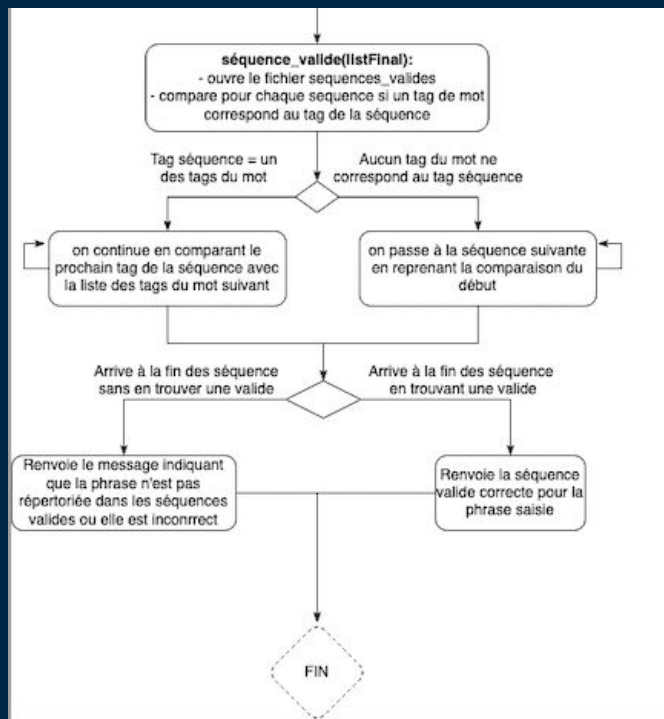
RÉALISATION

5. Déroulement

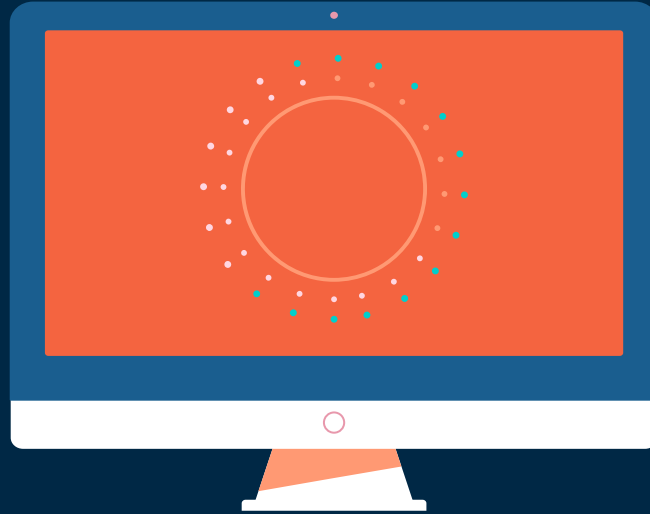


RÉALISATION

5. Déroulement



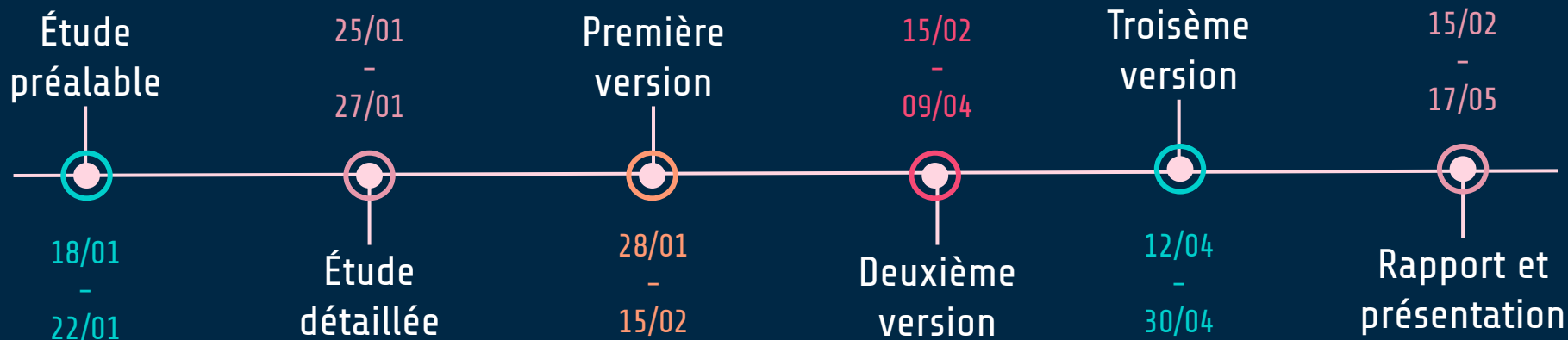
PLACE À LA DÉMONSTRATION



Gestion de projet

04

PLANNING INITIALE



PLANNING FINAL



CHOIX D'OUTILS



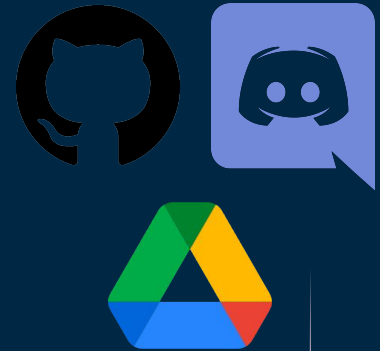
**Langage de
programmation**



**Environnement de
développement**



Outil de rédaction



Collaboration

Conclusion

05

Présenté par :
ALAOUI-YOUSFI Khaoula
CHAMROUK Laïla
ES-SEBBAR Imane
MAJDOUL Kaoutar

MERCI

Place aux questions