

Fiche 14 - Recommenders

Sommaire

1. Description de la tâche

1.1. Définition

1.2. Pré-requis

1.3. Créer une couche d'annotation pour un recommander

2. Retour sur expérience

Mots-clés : importation, exportation, sauvegarde

1. Description de la tâche

1.1 Définition

Les recommenders fournissent une prise en charge des annotations en proposant des étiquettes potentielles. En d'autres termes, il s'agit d'une fonctionnalité d'apprentissage automatique proposée par INCEpTION. Les annotations faites par un recommander peuvent être acceptées ou rejetées par l'utilisateur. Grâce à l'interaction entre l'utilisateur et l'outil, ce dernier apprend à améliorer la qualité de ses prédictions.

1.2 Pré-requis

Afin de pouvoir utiliser la fonction Recommenders il faut au préalable disposer de :

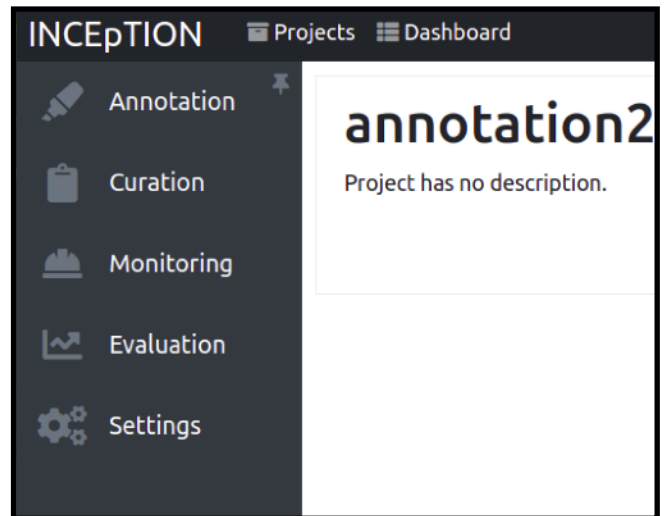
- Un projet déjà créé.
- Un texte (minimum) déjà annoté avec la couche d'annotation qui devra être géré automatiquement. L'annotation ne doit pas être aboutie, mais plus il y a d'annotations, mieux l'outil fonctionnera. L'annotation par le recommander se fera dans tous les documents du même projet.

Pour apprendre comment créer un projet, voir Fiche 3 - Création de projet

1.2 Créer une couche d'annotation pour un recommander

Étape 1 : Cliquer sur le projet

Étape 2 : Cliquer sur *Settings*



Étape 3 : Cliquer sur Recommenders
(en bas à gauche de l'image 2)

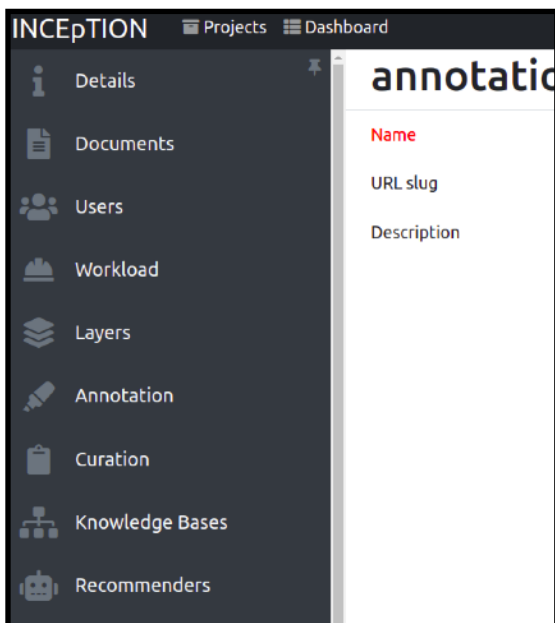


Image 1 : Menu d'un projet INCEpTION - Settings.

Image 2 : Menu des réglages d'un projet INCEpTION - Recommenders.

Étape 4 : Cliquer sur Create

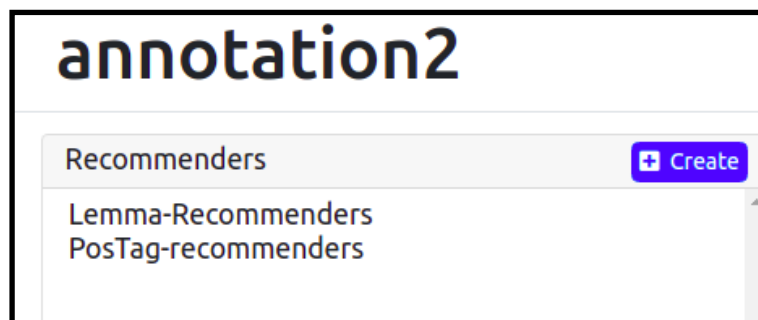


Image 3 : Réglages des recommenders - create

Fiche 14 - Recommenders

En cochant la case auto-generate, INCEPTION va générer un nom à la couche d'annotation automatiquement. La case peut être également décochée et le nom choisi par l'utilisateur.

Étape 5 : Dans le menu déroulant à côté de *Layer*, choisir la couche d'annotation sur laquelle l'annotation automatique (le *recommender*) va se baser :

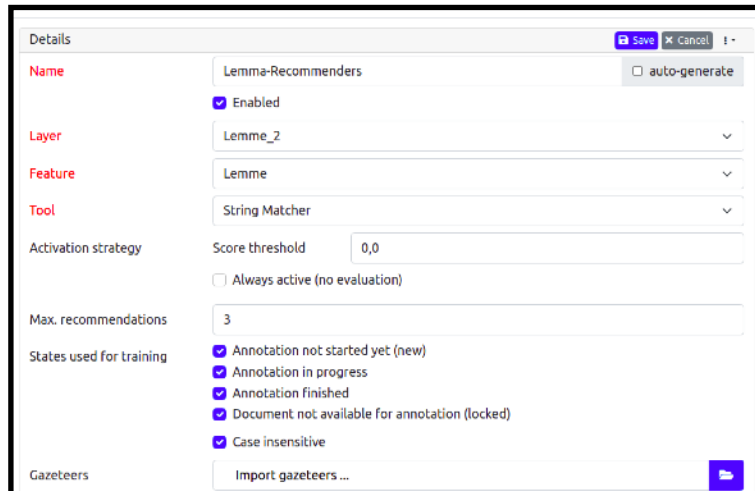


Image 4 : Détails des réglages des recommenders, expliqués dans étape 4-9

Étape 6 : Choisir un outil (*Tool*) pour le recommander, il en existe plusieurs :

- **String Matcher** : Cet outil donne uniquement des étiquettes aux mêmes chaînes de caractères qui ont déjà pu être annotées manuellement. Par exemple, si la chaîne “La ville rose” a été annotée comme une entité nommée (LOC), toutes les chaînes identiques vont être étiquetées avec LOC par le recommander. Il faut pourtant être conscient du fait que cet outil essaie d'identifier des mots qui correspondent selon les premières lettres d'un mot, pas le mot en entier. Un mot tel que “paramètres” peut donc se voir attribuer le lemme “par”.
- **Open NLP** :
 - **Sentence Classifier (OpenNLP Document Categorizer)** – Ce recommander est disponible pour les couches d'annotation au niveau de la phrase.
 - **Token Sequence Classifier (OpenNLP POS)** – Ce recommander apprend à marquer les tokens. Il est disponible pour les lemmes ainsi que pour les POSTags.
 - **Multi token sequence Classifier (OpenNLP NER)** – Ce recommander apprend à marquer plusieurs mots-formes. Il est disponible pour les entités nommées.
- **Named Entity Linker** : Cet outil tente de faire correspondre le contexte d'une entité nommée avec le contexte des entités candidates dans une base de données et ainsi suggérer une entité.
- **Recommenders externes** : Trois recommenders externes sont disponibles sur INCEPTION. Ils permettent tous d'utiliser des services web externes afin de générer des prédictions.

Fiche 14 - Recommenders

Étape 7 : Pour tester le recommander sur un texte, il faut retourner sur le *Dashboard* et choisir un texte (les documents disponibles se trouvent à droite dans la catégorie *Recent Activity*).

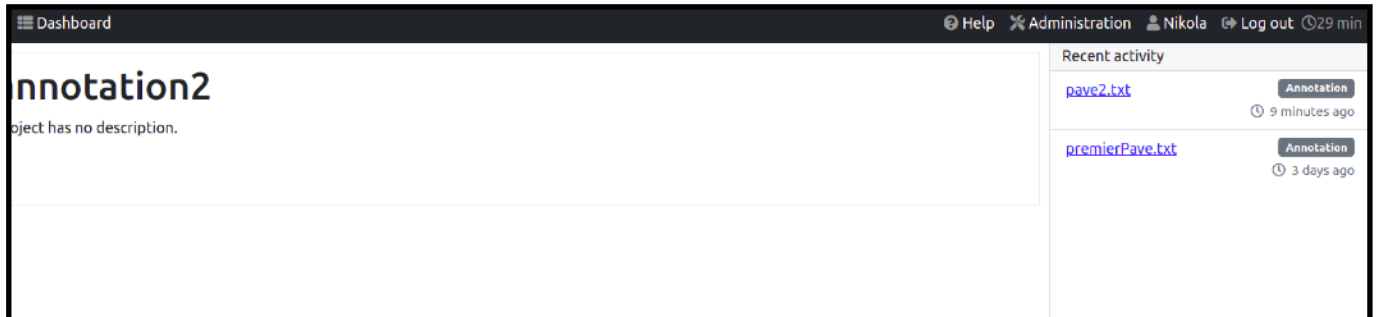


Image 5 : Dashboard d'un projet INECpTION

Étape 8 : Après avoir cliqué sur la flèche bleue, les suggestions d'annotations par le recommander apparaissent en gris dans les documents

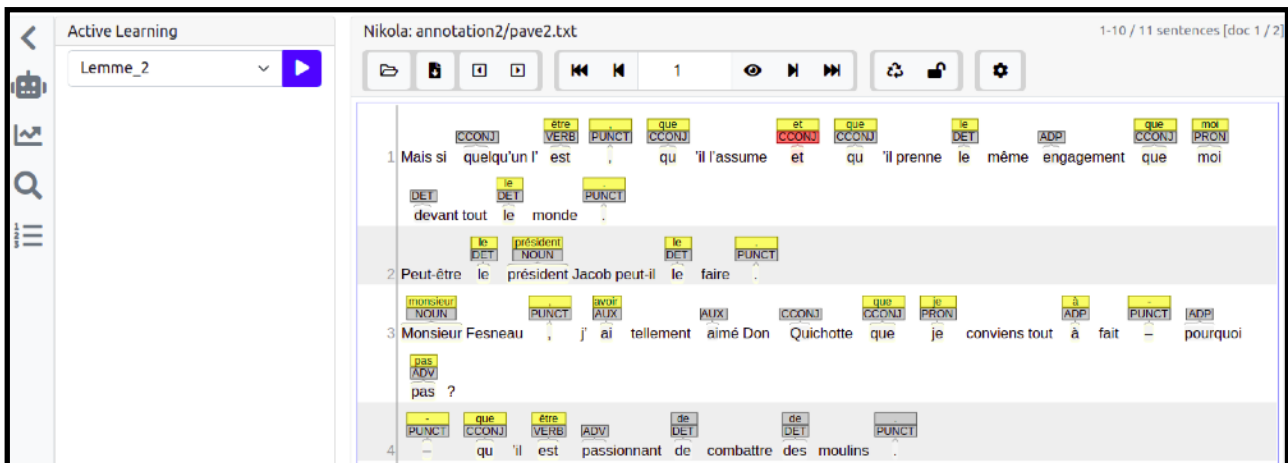


Image 6 : Exemple d'une annotation des lemmes faite par un recommander (en gris)

Fiche 14 - Recommenders

Les annotations faites par les recommenders doivent être acceptées ou rejetées manuellement. Cliquer une fois sur l'annotation pour la garder, et deux fois pour la rejeter. Les boutons colorés à gauche peuvent être utilisés dans ce même but :

The screenshot shows a mobile application interface for 'Lemme_2'. At the top, there is a dropdown menu with 'Lemme_2' and a close button. Below this is a 'Recommandation' section with the text: 'il est passionnant de combattre des mouli'. The word 'de' is highlighted with a black box. Underneath is a 'Suggestion' section with a text input field containing 'de'. Below that is a 'Score' section with a text input field containing '1' and a '(Δ 1)' label. The 'Lemme' section has a text input field containing 'de'. At the bottom of the main form are three colored buttons: a green button with a checkmark, a red button with a cross, and a yellow button with a plus sign. Below these buttons is a 'History' section with a scrollable list. The list contains two entries: the first entry shows the word 'pas' with a blue 'x' icon and a 'rejected' button; the second entry shows the word 'est' with a blue 'x' icon and an 'accepted' button.

Image 7 : Validation et historique des annotations des recommenders.

Les annotations acceptées ou rejetées se trouvent en dessous des boutons “valider”, “rejeter” et “ignorer”. Les choix peuvent être annulés grâce à la croix qui se trouve à côté des mots présents dans l'historique.

2. Retour sur expérience

Les recommenders ont été testés de deux façons :

- À partir de nos annotations manuelles faites avec les couches d'annotation créées par nous-même (cf. *Fiche 8 - Création d'une couche d'annotation*).
- À partir d'un document préalablement annoté par Stanza (cf. *Fiche 5 - Formats*).

Dans les deux cas, un recommender a été crée pour les lemmes et un autre pour les traits morpho-syntaxiques.

Comme on peut le voir dans l'image 6, certains mots sont ignorés, d'autres sont parsés partiellement. Par exemple, le nom *Quichotte* a été annoté en tant que CCONJ alors qu'il s'agit d'un nom propre. L'adverbe *passionnant* a été annoté comme ADV, mais cette étiquette ne concerne que la partie "pas". Étant donné que les lemmes sont annotés au niveau des caractères, l'outil *string matcher* ne prend en considération que le début du mot (voir étape 6).

Le deuxième type de recommender s'est entraîné sur un document annoté avec Stanza. Les couches d'annotation ne sont pas créées par nous-même, mais existent préalablement dans INCEpTION. Cependant, sa performance reste très faible :

1	CCONJ	DET	PRON	PUNCT	PRON	VERB	PRON	PUNCT	CCONJ	PUNCT	PRON	ADV	VERB	DET	NOUN	ADP	NOUN	PRON	PUNCT	PRON
	Mais	si	quelqu	,	un	lire	,	est	et	,	qu	il	lire	,	assume	et	qu	,	il	
	VERB	DET	NOUN	AUX	SCONJ	PRON	AUX	VERB	DET	NOUN	PUNCT									
	importe	le	même	être	que	moi	être	tout	le	féminisme	.									
2	ADJ	VERB	DET	VERB	PRON	VERB	PUNCT	DET	NOUN	PUNCT										
	bon	le	président	pouvoir	Jacob	peut-il	,	le	faire	.										
3	ADP	INTJ	PROPN	ADJ	PRON	PRON	ADV	VERB	ADV	VERB	PRON	VERB	SCONJ	PRON	NOUN	PRON	PRON	PRON	VERB	DET
	bonjour		Fesneau	il	ca	il	y	avoir	tout	avoir	on	bon	que	je	courant	tout	à	vrai	ce	fait
	NOUN	AUX	ADV																	
	pouvoir	pas	?																	
4	ADJ	PUNCT	PRON	PUNCT	PRON	CCONJ	ADV	ADP	VERB	NOUN	ADV	NOUN	PUNCT							
		,	qu	,	il	et	page	de	le	un	modification	.								
							haut		être											

Image 8 : Résultats d'un recommender entraîné sur un document annoté par Stanza.

Dans les deux scénarios présentés, nous estimons que les recommenders ne sont pas performants pour les couches lemmes et POS tags. En revanche, nous supposons que la performance du recommender est liée à la quantité de données annotées à la disposition de l'outil.

Suite aux annotations proposées par l'outil, l'utilisateur a la possibilité de rejeter, accepter ou alors ignorer l'annotation proposée. Ainsi, l'annotation peut être corrigée, ce qui améliorera la performance du recommender en question.