

Annotation de claims dans un corpus d'articles de recherche: premiers pas avec Doccano

Clémentine Bleuze

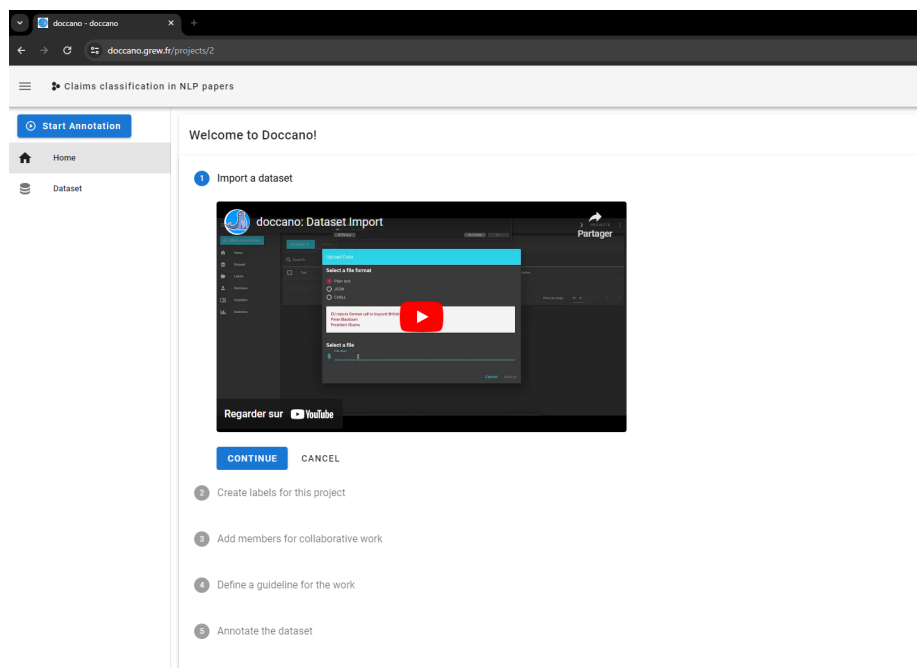
Avril 2024

1 Connexion à Doccano

Aucune installation n'est nécessaire puisque Bruno et Vincent (merci à eux !) ont installé Doccano à l'adresse suivante: doccano.grew.fr. Vous devez vous connecter grâce aux identifiants et mots de passe fournis.

2 Accès au projet

Vous arrivez ensuite directement sur une page contenant le projet intitulé *Claims classification in NLP papers*. Cliquez sur le projet: vous arrivez sur sa page d'accueil. En haut à gauche, cliquez sur le bouton bleu "start annotation".



3 Commencer les annotations

3.1 Prise en main des fonctionnalités

Vous arrivez sur la page d'annotation, contenant:

- au centre: la phrase à annoter (cadre **bleu**)
- juste au-dessus: les étiquettes disponibles avec les raccourcis clavier correspondants (cadre **noir**)
- en haut à droite: la progression de vos annotations dans l'ensemble du dataset (cadre **jaune**)
- en bas à droite: les métadonnées de la phrase à annoter (intitulé de la section contenant la phrase annotée + phrase précédente et phrase suivante) (cadre **rouge**)

The screenshot shows the annotation interface for 'Claims classification in NLP papers'. The central text area contains the sentence: "When not considering an LM, it is possible to efficiently find the best translation for an input sentence using the CKY+ algorithm (Chappelier et al., 1998)." The toolbar above the text area includes icons for validation (a cross), guidelines (a document), and navigation (arrows). The sidebar on the right shows a progress bar indicating 0% completion and a table of metadata.

Key	Value
doccano_art_id	55
sentence_id	1
current_sentence_section	Synchronous Context-Free Grammars
previous_sentence_section	Synchronous Context-Free Grammars
previous_sentence	At translation time, the decoder searches for the target sentence that maximizes the derivation probability, which is defined as the sum of the scores of the rules used in the derivation, and the log of the language model (LM) probability over the target string.
next_sentence_section	Synchronous Context-Free Grammars
next_sentence	When using an LM, the

Au-dessus des étiquettes, les icônes les plus importantes sont celles permettant de **valider une annotation** (la croix indique que l'annotation n'a pas encore été validée), consulter les **guidelines**, laisser un **commentaire** (en cas de doute sur la bonne étiquette, ou d'une quelconque remarque), et consulter les **raccourcis clavier** pour une navigation plus rapide entre les documents.



3.2 Commencer à annoter

Pour annoter un document, vous devez donc:

- 1) Lire la phrase et vous référer aux **guidelines** pour trouver la ou les catégories adéquates. Les métadonnées sont là pour vous **aider à contextualiser** la phrase: vous pouvez avoir accès à l'intitulé de la section de laquelle elle provient (**sec**), ainsi qu'aux phrases et sections précédentes (**prev_sent**, **prev_sec**) et suivantes (**next_sent**, **next_sec**) dans le texte de l'article.

À savoir: Les phrases issues d'un même article vous sont présentées séquentiellement, dans leur bon ordre d'apparition. Notez qu'il est normal que vous ne rencontriez pas l'ensemble des phrases d'un article, car certaines sections ont été exclues: il est donc normal de passer, par exemple, d'une section d'introduction à une section de résultats expérimentaux.

- 2) En cas d'hésitation, laisser un **commentaire** expliquant les éléments ambigus. S'il vous est impossible de choisir une catégorie, vous pouvez passer à l'étape 4.
- 3) Sélectionner (à la souris ou au clavier) la ou les **étiquettes** (un *checkmarck* apparaît alors, voir image de la page précédente).
- 4) **Valider** l'annotation (icône avec la croix ou touche "entrée"). Sans cette étape, votre progression ne sera pas mise à jour (cadre en haut à droite de l'écran).
- 5) Passer au document suivant (bouton ">" en haut à droite ou touches directionnelles du clavier).

Bonne chance !

(Bien sûr, je reste disponible pour vos questions)