

# Annotation Guidelines

Attached you shall find the sentences of code-mixed data:

Sample-codemixed.csv: <https://bit.ly/3SBsIxA>

Use appropriate files as mentioned in the tool overhead.

## Steps to follow common for all the tools:

1. Start a timer at the beginning of the annotation of every sentence and record the time taken for every annotated sentence(in seconds). There should be 10-time entries for the 10 corresponding sentences.
2. Store and update all timings in the timesheet.
3. End the timer on completion of annotation and submit the annotation time you took in the Google form attached (this form shall be circulated at a later stage, for now, note down the annotation timings in a self-created text file).
4. We shall be performing **word-level** annotation. For example, in a sentence “इसके अलावा संदेश ऐप (Sandesh App) और सार्वजनिक ऑफिसों में अटेंडेंस लगाने के लिए आधार ऑथेटिकेशन को ऑप्शनल कर दिया गया है।”, even though the sequence of three words is Hindi you will have to annotate the three words individually, like [इसके=Hindi], [अलावा=Hindi], [संदेश=Hindi]. Do not cluster continuous similar annotations.
5. Follow this rule strictly.
6. The sentence annotation time needs to be in **seconds (eg 87s)**, same for the installation and configuration time needs to be in a **second** format (eg 235s).

## Markup

<https://getmarkup.com/dashboard>

1. Create a new account on Markup. (**Note:** At the start of this step, start the timer and end the timer when you finish the configuration step to step 7 and note down the timing under header, “Configuration Time”)(You can enter the “Installation Time” = 0, for this tool).
2. Create a new session.
3. Upload a configuration file present in the drive>Markup directory for POS Tagging.

The screenshot shows the 'Config Creator' interface. On the left, under 'Entities', there is a list of parts of speech: NOUN, PROPN, PRON, VERB, ADJ, ADP, ADV, CONJ, DET, PART, PART-NEG, PART-WH, X, and NUM. Below this, under 'Attributes (optional)', there is a section for a 'Related entity' (set to 'NOUN'), an 'Attribute name' (empty), and an 'Attribute values' input field ('Start typing to add attribute values'). A checkbox for 'Allow custom attribute values' is checked. At the bottom right of the left panel is a blue 'Export & Use' button. On the right, under 'Live Preview', the entities are shown as colored circles: NOUN (orange), PROPN (green), PRON (pink), VERB (yellow), ADJ (cyan), ADP (orange), ADV (green), CONJ (grey), DET (yellow), PART (pink), PART-NEG (grey), PART-WH (cyan), X (cyan), and NUM (grey).

#### 4. Upload Sample-codemixed.txt available in the drive.

The screenshot shows the annotation interface. At the top, there are three buttons: 'Exit setup', 'Settings', and a green 'Annotate →' button. Below this, the 'Documents' section shows a file named 'Sample-codemixed.txt' with '500 annotations'. There are two buttons: 'Upload annotations' (grey) and 'Upload documents' (blue). The main area displays a sentence in Hindi: "पहले भी ला चुका है ऐसा फिचर 2017 में वॉट्सऐप ने अपने स्टेटस फंक्शन के लिए My contacts except... ऑप्शन ला चुका है। स्कॉटलैंड की प्रधानमंत्री (First Minister) निकोला स्टुर्जन ने पुलिस और इमीग्रेशन की कार्रवाई को एकत्रफा और गलत करार दिया। PFI पर पिछले साल सिटिजनशिप अर्डरमेंट एक्ट (CAA) के खिलाफ हुए प्रदर्शनों को फाइनेंस करने का आरोप लगा है। "पहले सभी गोट TMC को मिलते थे, लेकिन बीते पांच सालों में BJP ने अपनी पकड़ इस गांव में मजबूत की है।" इंडियन प्रीमियर लीग (IPL) के बचे हुए मैच और भारत की मेजबानी में होने वाला टी-20 वर्ल्ड कप UAE में होना लगभग तय है। "B Com की पढ़ाई भी कंप्लीट नहीं कर पाई, क्योंकि उस वक्त मेरी फैमिली की बजह से पापा ज्यादा पसंद नहीं करते थे।" "DRDO के मुताबिक, यह किट शरीर में SARS-CoV-2 के वायरस और इससे लड़ने वाले प्रोटीन न्यूक्लियो कैमिड (S&N) दोनों की मौजूदगी का पता लगा सकती है।" फिलहाल ICICI बैंक 5 से

5. Click on Annotate.
6. Select individual words, select an entity (on the left) and then click on add annotation. Continue this process for every word in the sentence. Here is a small example for the same.

--

<input type="radio"/>	NOUN	<input type="radio"/>	PROPN	<input type="radio"/>	PRON
<input type="radio"/>	VERB	<input type="radio"/>	ADJ	<input type="radio"/>	ADP
<input type="radio"/>	ADV	<input type="radio"/>	CONJ	<input type="radio"/>	DET
<input type="radio"/>	PART	<input type="radio"/>	PART-NEG		
<input type="radio"/>	PART-WH	<input type="radio"/>	X	<input type="radio"/>	NUM

**3 Add attributes**  
The specific properties of the entity.  
Suggested: NA  
Select entity to see attributes

**4 Ontology**  
The concept you want to map to the entity.

Ontology

Concept

Add annotation

**2 Select entity**  
The high-level concept you are annotating.  
Suggested: NA

NOUN PROPN PRON VERB ADJ ADP ADV CONJ DET PART PART-NEG PART-WH X NUM

**3 Add attributes**  
The specific properties of the entity.  
Suggested: NA  
Select entity to see attributes

**4 Ontology**  
The concept you want to map to the entity.

Ontology Concept

**ADV**

- जल्द (0 attributes)
- हाल (0 attributes)
- जहाँ-जहाँ (0 attributes)

**CONJ**

- और (0 attributes)
- और (0 attributes)
- और (0 attributes)

**DET**

Repeat the same process for the Token-level Language Identification Task.

**1 Select text**  
Highlight the document text you want to annotate.  
Highlight the text you want to annotate.

**2 Select entity**  
The high-level concept you are annotating.  
Suggested: NA

hindi english unidentified

**3 Add attributes**  
The specific properties of the entity.  
Suggested: NA  
Select entity to see attributes

**4 Ontology**  
The concept you want to map to the entity.

Ontology

**Sample-codemixed.txt**

**Search documents**

पहले मी ला चका है ऐसा फोबर 2017 में वॉट्सएप ने अपने स्टेट्स फंक्शन के लिए My contacts except ... आशन ला चुका है।

स्कॉटलैंड की प्रधानमंत्री ( First Minister ) निकोला स्टूर्जन ने पुलिस और इमोग्रेशन की कारवाई को एकत्रफा और गलत करार दिया।

PFI पर पिछले साल सिस्टिजनशिप अमेडमेंट एवट ( CAA ) के खिलाफ हुए प्रदर्शनों को फाइनेंस करने का आरोप लगा है।

"पहले सभी बोट TMC को मिलते थे, लेकिन बीते पांच सालों में BJP ने अपनी पकड़ इस गांव में मजबूत की है।"

इंडियन प्रीमियर लीग ( IPL ) के बचे हुए मैच और भारत की मैजबानी में होने वाला टी - 20 वर्ल्ड कप UAE में होना लगभग तय है।

" B Com की पढ़ाई भी कंफ्रीट नहीं करते थे।"

**पहले**

**सभी**

**को**

**मिलते**

**थे**

**लेकिन**

**बीते**

- Click on options and select data.txt (ann) and export the file. Submit the annotation file to the admin.
- Exit session.

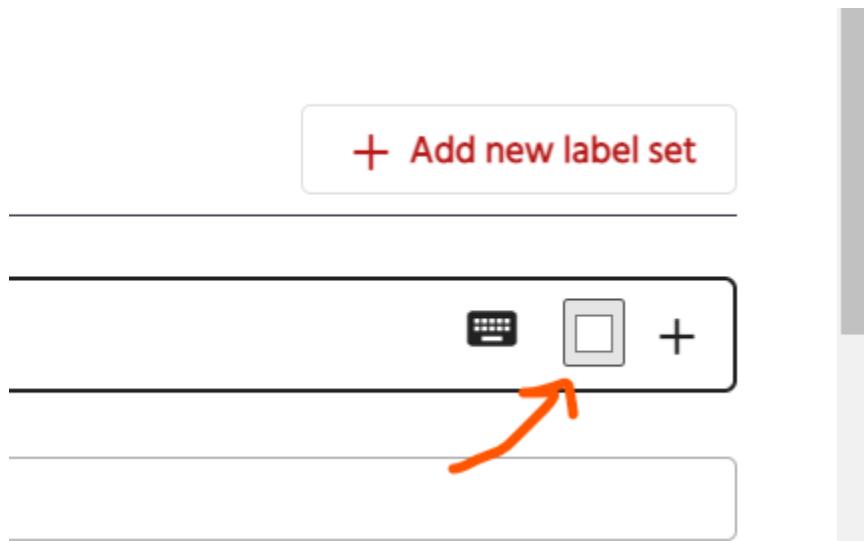
## Ubaii Tool

<https://ubaii.tools/>

1. Create a new account on the Ubaii Tool Website. (**Note:** At the start of this step, start the timer and end the timer when you finish the configuration step to step 9 and note down the timing under the header, “Configuration Time”)(You can enter the “Installation Time” = 0, for this tool).
2. Create a new project and add the project name.
3. Select span-based annotation.
4. Add the following entities for the Token-level Language Identification task.

The screenshot shows the Ubaii Tool's interface for managing labels. At the top, there are three tabs: HINDI (green), ENGLISH (red), and UNIDENTIFIED (purple). Each tab has icons for edit, delete, keyboard, and list. To the right is a search bar labeled "Search...". Below the tabs is a section for "Labels (3)". On the far right, there are two buttons: "Import labels" with a cloud icon and "Add new label set" with a plus icon.

Use this box to choose colours, different colours are acceptable, they are just for your visual difference.



5. Go next to relations and submit on step 5.
6. Now upload the Sample-codemixed.txt file (Select no-pre annotation).
7. Now annotate token-wise. First, select the entity and then select the word to annotate it.

The screenshot shows the ubiAI Text Annotation interface. At the top, there's a header with 'CodeMixing' (Text Annotation), 'Sample-codemixed.txt' (In Progress), and buttons for 'Entities', 'Relations', 'Classification', 'Predict' (with a dropdown), and 'Validate' (with a checked checkbox). Below the header is a toolbar with buttons for 'HINDI' (green), 'ENGLISH' (red), 'UNIDENTIFIED' (grey), 'Add label...', and a search bar. The main area displays a list of words with their language tags: 'इसके HINDI', 'अलावा HINDI', 'सदैर UNIDENTIFIED', 'ENGLISH HINDI', 'ऐप ENGLISH', '( UNIDENTIFIED )', 'Sandesh ENGLISH', 'App ENGLISH', 'UNIDENTIFIED', 'और HINDI', 'सार्वजनिक HINDI', 'ऑफिसों HINDI', 'ही HINDI'. Below this list is another row of words: 'अटेंडेंस ENGLISH', 'लगाने HINDI', 'के HINDI', 'लिए HINDI', 'आधार HINDI', 'ऑयोटिकेशन्स ENGLISH', 'को HINDI', 'ऑप्शनल ENGLISH', 'कर HINDI', 'दिया HINDI', 'गया HINDI', 'ही HINDI'. On the left side of the main area, there are several icons: a bookmark, a folder, a list, a gear, a magnifying glass, and a download icon.

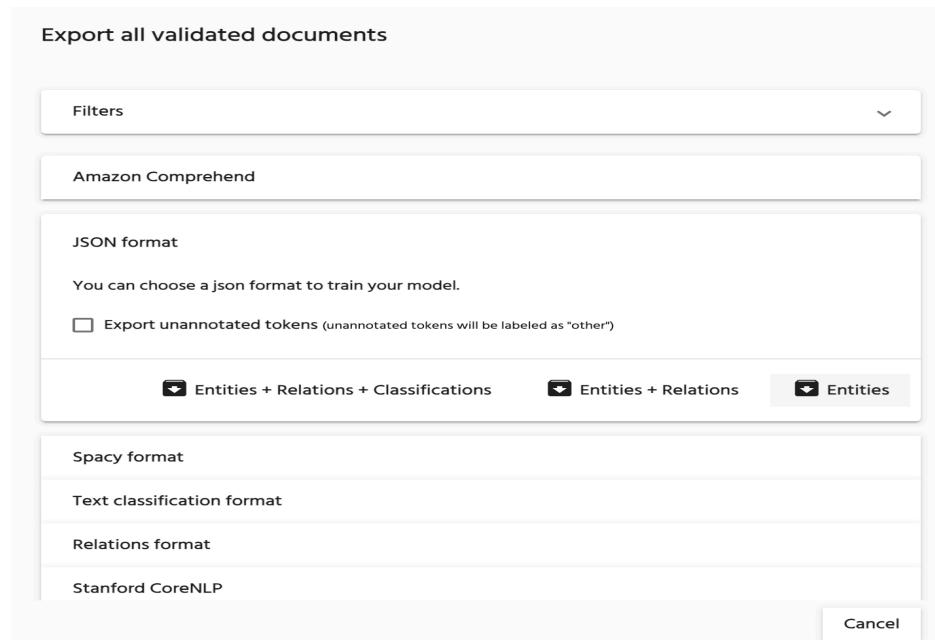
8. If you want help with annotations, then click on the predict button, which will help you with suggestions.

The screenshot shows the ubiAI interface with a sentence in Hindi: 'किसान HINDI वाची ने अपने इताके की 300 महिलाओं को 'Self Help Group' से जोड़ा और आर्थिक HINDI रूप से स्वावलंबी बनाने के लिए प्रेरित किया।' Below the sentence, a blue download icon is visible. To the right, there's a purple circular icon with a magnifying glass. The text below the sentence reads: 'इस अवसर पर हम देश भर में' FIT INDIA MOVEMENT ' launch करने वाले हैं।' Further down, it says: '" और इस कार्यक्रम में देशभर की IITs HINDI , IIIMs HINDI , Central Universities , NITs HINDI , जहाँ- जहाँ युवा HINDI पीढ़ी है , उन सबको live - connectivity HINDI के द्वारा'.

9. After annotation of all the sentences, validate all the sentences, then go to the left-hand side of the page, select the blue Export box and then click on the install button, Select JSON entities.

The screenshot shows the ubiAI navigation sidebar with the following options: 'Annotation' (selected), 'Documents', 'Pre-Annotation', 'Metrics', 'Models', and 'Project Settings'. At the bottom of the sidebar is a large blue 'Export' button with a download icon.

10. Unzip the folder and upload the JSON file on drive.



Repeat the same process For POS Tagging:

The screenshot shows the ubiAI platform interface with various navigation tabs like 'PROJECTS', 'MODELS', 'COLLABORATION', 'ANALYSIS', and 'RESOURCES'. On the left, there are icons for project creation, settings, and download. The main area displays several examples of Indian text with their Part-of-Speech (POS) tags. The first example is: "Sustainable ADJ development NOUN के ADP लिए ADP के से DET उनके PRON traditions NOUN को ADP हमें PRON अपने PRON जीवन NOUN में ADP अपनाने". The second example is: "की ADP आवश्यकता NOUN है VERB , उनसे PRON कुछ ADJ रीखने NOUN की ADP जारी VERB है VERB |". The third example is: "Preventive ADJ health - care NOUN के ADP स्व- ADP में ADP योग NOUN ने VERB , नये ADJ सिरे NOUN से ADP दुनिया ADV - X HR ADV में ADP अपनी PRON पढ़चान NOUN बनाई VERB है VERB |". The fourth example is: "किसान ADJ चाची NOUN ने VERB अपने PRON इलाके NOUN की ADP 300 NUM महिलाओं NOUN के ADP Self ADJ Help ADJ Group NOUN से ADP जोड़ा VERB |". The fifth example is: "और CONJ आर्थिक ADJ स्व- NOUN से ADP स्वावलंबी NOUN बनाने VERB के ADP लिए ADP प्रेरित VERB किया VERB |".

Prediction result for POS Tagging:

The screenshot shows the ubiAI platform interface. At the top, there are navigation links: PROJECTS ▾, MODELS, COLLABORATION ▾, ANALYSIS, and RESOURCES ▾. On the right side, there are icons for profile, notifications, and search.

The main content area displays a sentence with part-of-speech (POS) tags and annotations:

"Made VERB In India की न केवल ग्लोबल इमार्ड NOUN हो बल्कि ग्लोबल acceptance NOUN भी हो, हमें ये सुनिश्चित NOUN करना है।"

Below this, another sentence is shown:

अब International Standard Time NOUN को हमारा Indian Standard Time NOUN 3 नेतोरेकड़ NOUN से भी कम की एकद्युर्मती NOUN लेवल से match NOUN कर रहा है।

Further down:

Forensic Science NOUN के हर छात्र NOUN के लिए human intelligence NOUN की बासीकियों NOUN को विकसित NOUN करना भी बहुत महत्वपूर्ण NOUN है।

And another:

विकासशील देशों NOUN में Sustainable Consumption NOUN, E - Commerce NOUN और Financial Services NOUN के संबंध में ये गाइडलाइंस NOUN बहुत महत्वपूर्ण हैं।

At the bottom:

" साथियों, Inflation NOUN पर लगाम NOUN लगाने की वजह से भी गरीब NOUN और मध्यम वर्ग NOUN के Consumers NOUN का आर्थिक NOUN फायदा NOUN हुआ है।"

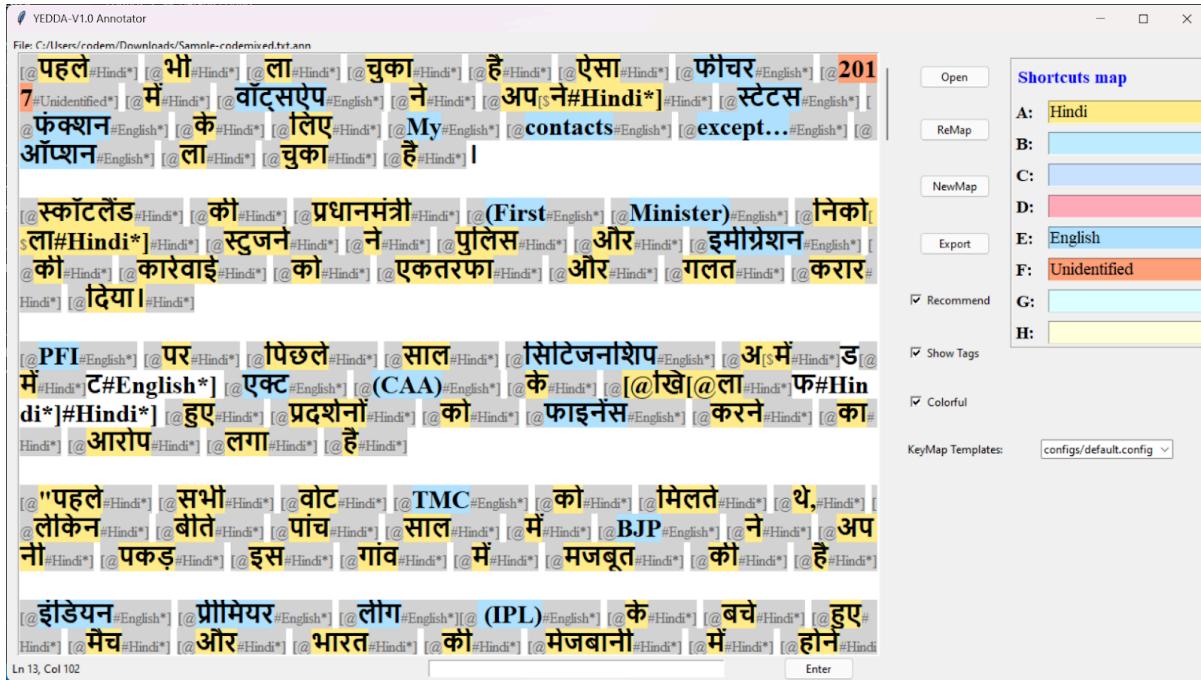
Finally:

इसके लिए National Digital Health Mission NOUN के तहत एक प्रभावी प्लेटफॉर्म NOUN पर भी काम चल रहा है।

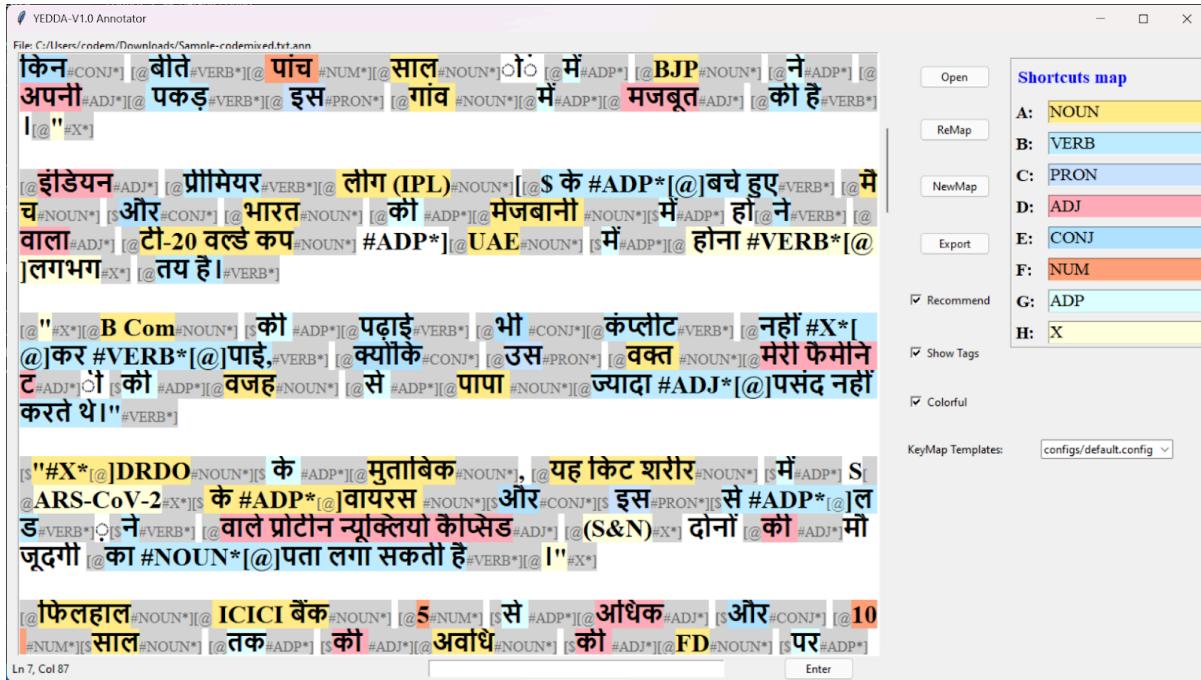
## YEDDA

<https://github.com/jiesutd/YEDDA>

1. Open the YEDDA folder and type command “python YEDDA.py” for Python version 3.9 or above and “python YEDDA.py” for Python version below 3.8 in the terminal to start the application. Hit Enter. (**Note:** At the start of this step, start the timer and end the timer when you finish the installation step and note down the timing under header, “Installation Time”)
2. Click on open and load data.txt to the YEDDA tool (Available in the Data Folder) (**Note:** At the start of this step, start the timer and end the timer when you finish the configuration step to step 6 and note down the timing under header, “Configuration Time”).
3. Update the shortcut maps using the new default.config file available in the attached drive link at the start of the document ( drive>YEDDA>default.config ). Upload this config file in the location YEDDA>config in the local file system.
4. Tick on Colourful to aid you in visual representations.
5. Perform annotations of the loaded sentences on word level, meaning annotate every word with respect to the language they pertain to. Identify them as unidentified (null) when you encounter anything similar to period (.), exclamation (!), question mark (?) etc. Select a word and then press a key to annotate it.
6. For example: select “Contacts” then hit key E to annotate it as English.



Repeat the same process For POS Tagging:

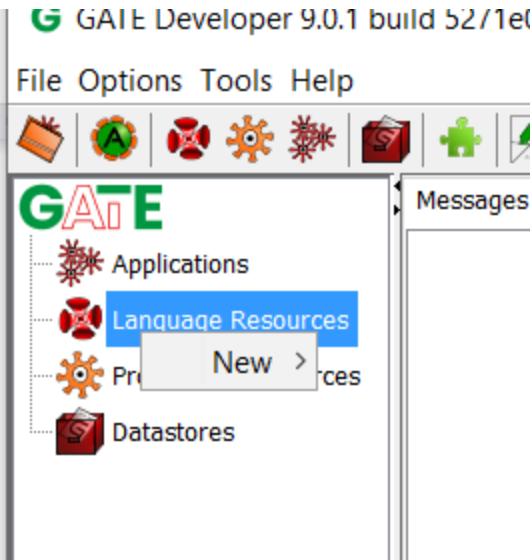


7. Click on export and then ok.
8. An 'ann' and a 'bmes' file shall be downloaded, and submit the files to the admin for further analysis.

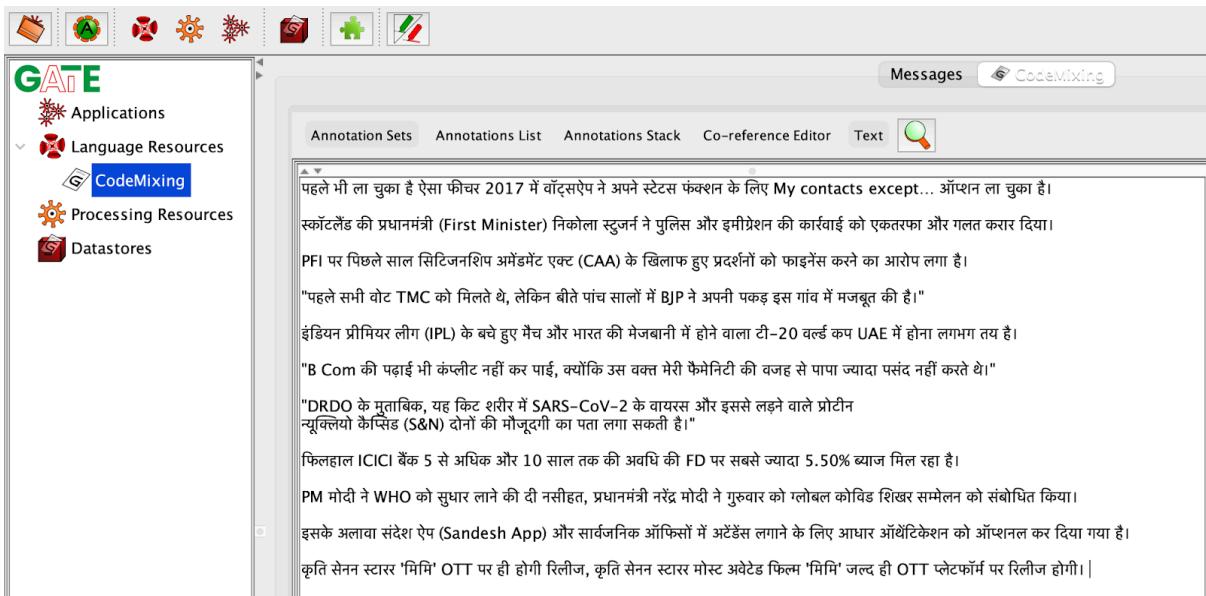
## GATE

<https://gate.ac.uk/download/#latest>

1. Follow all the installation procedures as the program driver asks. (**Note:** At the start of this step, start the timer and end the timer when you finish the installation step, note down the timing under header, "Installation Time").
2. Add the data to the GATE application (**Note:** At the start of the this step, start the timer and end the timer when you finish the configuration step to step 7 and note down the timing under header, "Configuration Time").

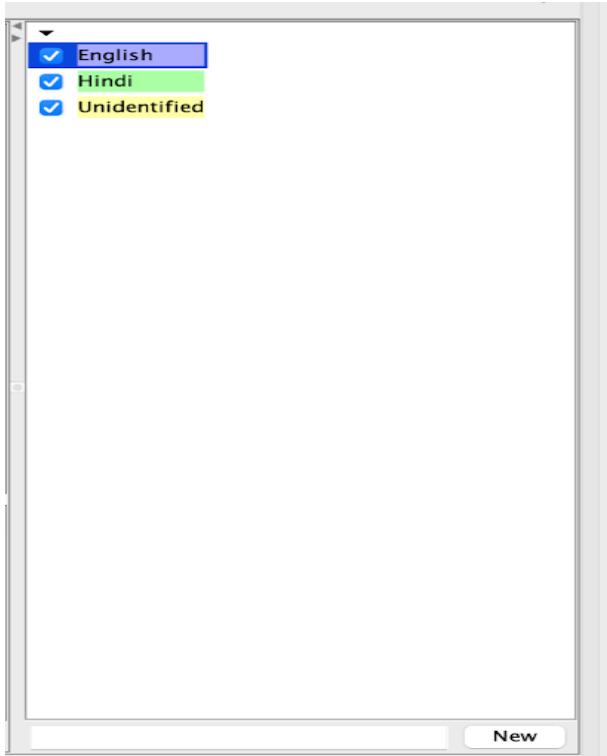


3. Click on the new and then gate document, name it and click on the newly created document.
4. Copy the contents from the data.txt file to the new gate document.
  - a. Click on annotation sets.

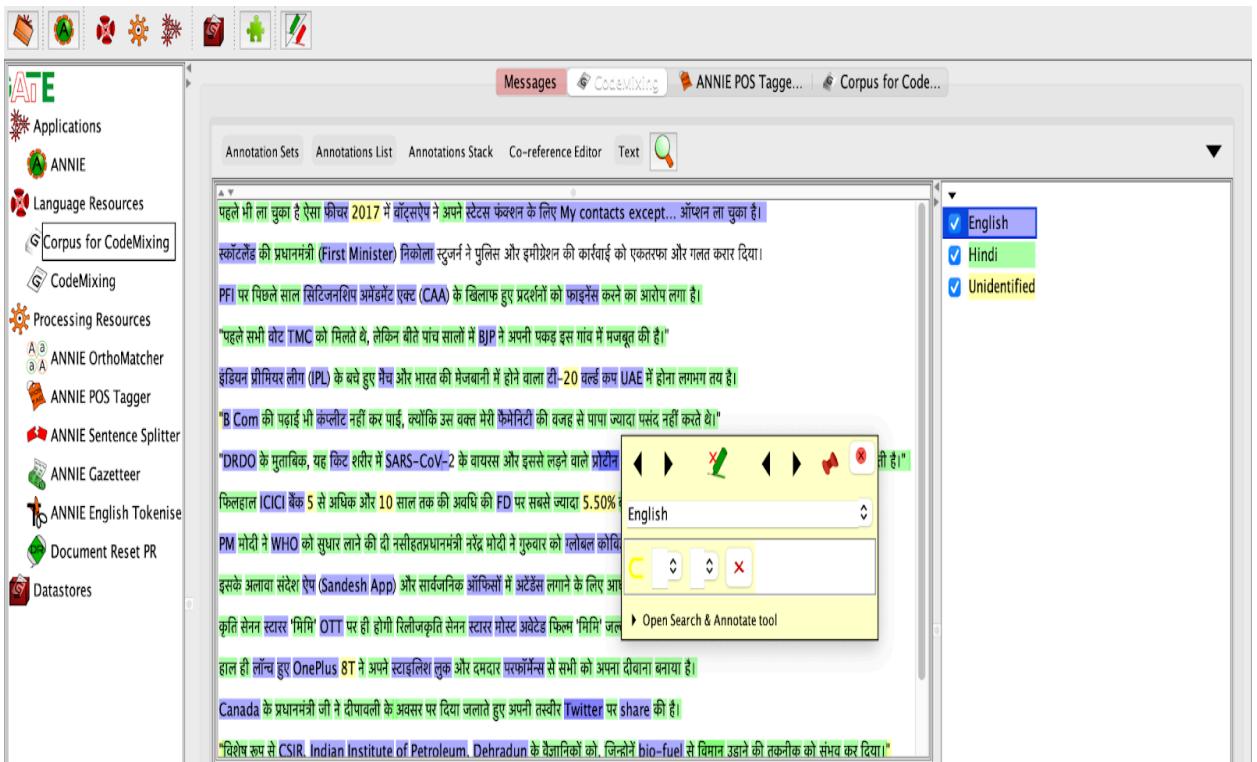


5. And add the following annotation entities:

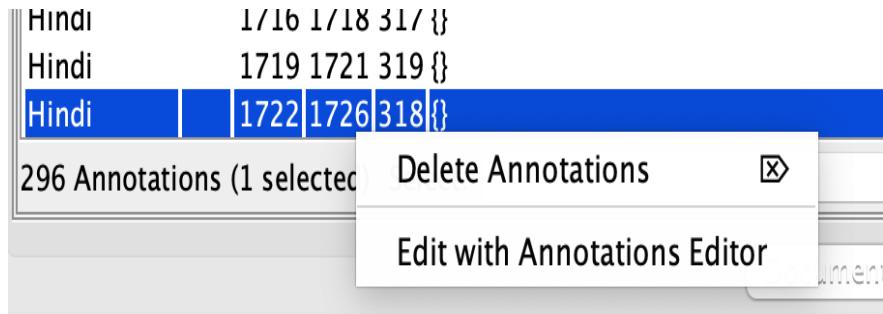
Hindi  
English  
Unidentified



6. Now perform **word-level** annotations, first select the entity on the right, then select the word and double click to annotate it.



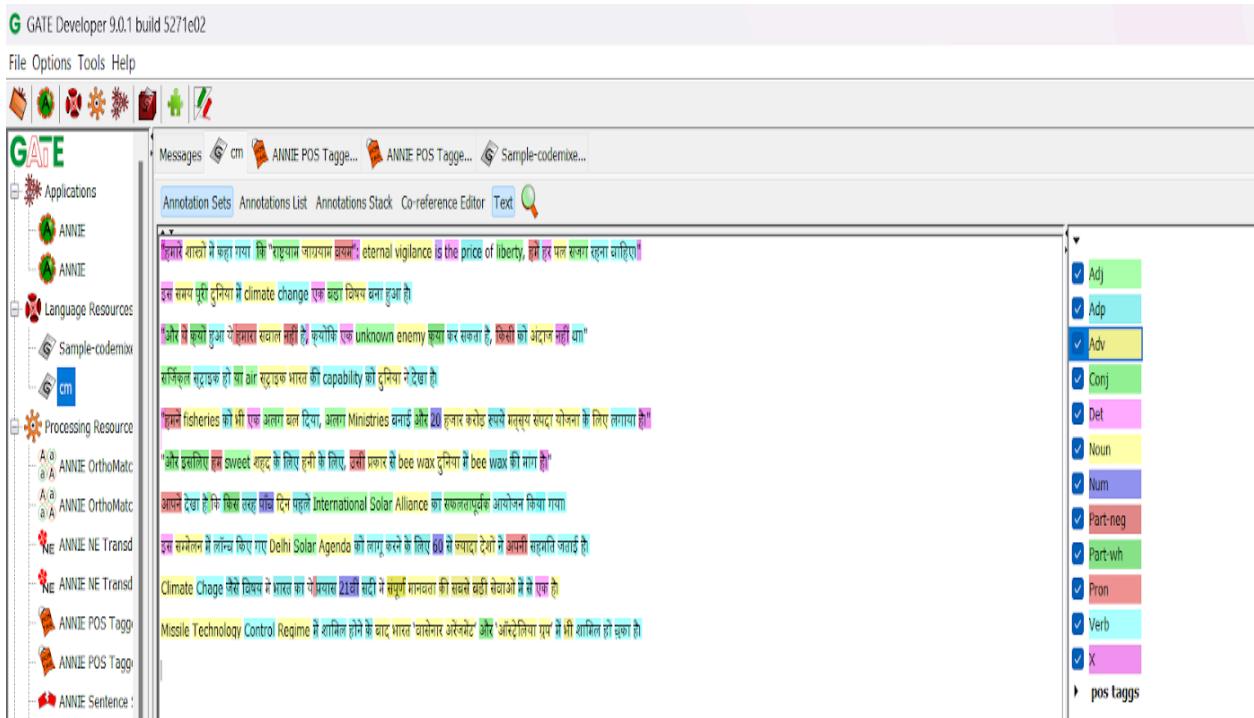
7. If you perform wrong annotation, double-click on that word to edit annotations with the Annotations editor.



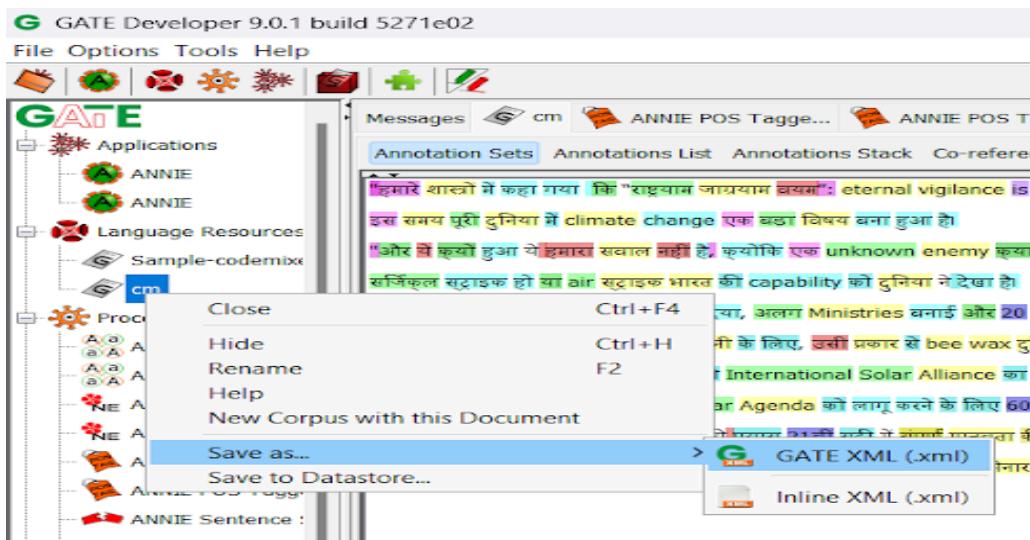
8. Click on the wrong annotation and then click on the green pen with the cancel icon to delete that annotation.



Repeat the same process for POS tagging:



## 9. Post annotation, export the annotation as the following gate XML document.



## 10.. Upload the XML file to the drive folder for GATE.

## INCEpTION

1. Navigate to INCeption Website: Go to [INCeption Project](#).
2. Create a new account on the INCeption Tool Website. (**Note:** At the start of this step, start the timer and end the timer when you finish the configuration step to step 9 and note down the timing under the header, "Configuration Time") (You can enter the "Installation Time" = 0, for this tool).
3. Create a new project and add the project name.

The screenshot shows the INCEpTION web application's project management screen. At the top, there is a dark navigation bar with the INCEpTION logo and a 'Projects' tab. Below this is a search bar with a magnifying glass icon and a placeholder 'Filter by name...'. Two buttons are visible: '+ New project ...' and 'Import project ▾'. The main area displays a list of projects with the following entries:

- demo - New project
- demo - New project

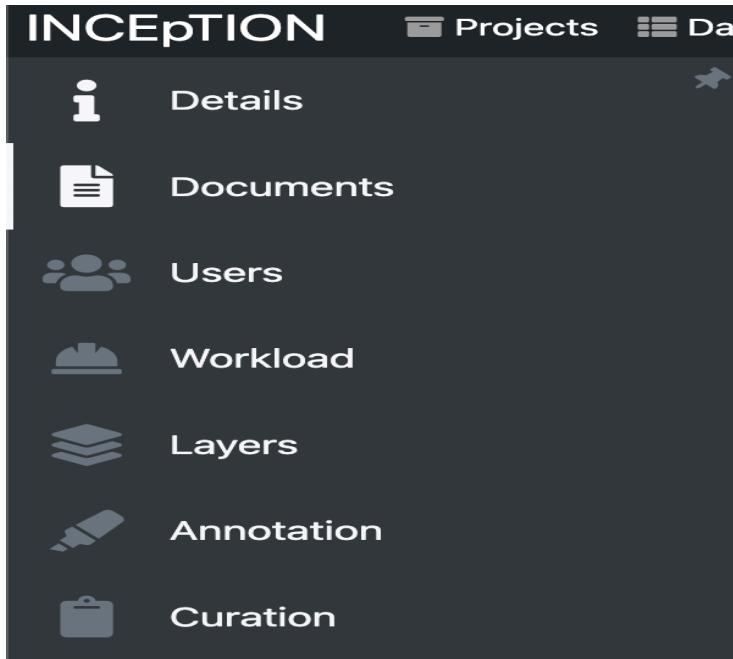
4. Add details and configurations of the Project.

The screenshot shows the INCEpTION web application's project configuration screen for the 'CodeMixing project'. The left sidebar contains icons for annotation editor, Brat annotation editor, annotation sidebar, and annotation search. The main configuration area includes the following sections:

- Annotation editor:** Editor dropdown set to 'Leave choice to annotator' (selected), checked box for 'Allow access to user-specific annotation editor preferences'.
- Brat annotation editor:** Default page size set to 10, Script direction set to 'left-to-right', unchecked box for 'Allow changing the script direction in the annotation editor'.
- Annotation sidebar:** Pinned groups section with a '+' button.
- Annotation sidebar:** Default tab set to 'None'.
- Annotation search:** Case sensitive checkbox checked, note: 'Only takes effect after index has been re-built.'

A 'Save' button is located at the bottom right of the configuration panel.

5. Go to "Documents" and click "Upload Document."



6. By selecting the add button, you can add multiple users to annotate the same document.

A screenshot of the 'CodeMixing project' annotation interface. The title 'INCEPtion' is at the top, followed by 'Projects' and 'Dashboard'. The main area shows the project name 'CodeMixing project'. On the left is a sidebar with icons for 'Details', 'Documents', 'Users', 'Workload', 'Layers', and 'Annotation'. In the center, there is a search bar with placeholder text 'Start typing here or select from the drop-down list...' and an 'Add' button with a plus sign. Below the search bar, the text 'demo [annotator, curator, manager]' is displayed. The 'Annotation' icon in the sidebar is highlighted with a blue border.

7. After logging in through user credentials, click on Annotation to start annotations. Now annotate token-wise for Language Identification. First, select the word and then select the layer(choose "Chunk" from the drop-down menu) to annotate it.

CodeMixing project Sample-codemixed.txt 41-50 / 50 sentences [doc 1 / 1]

Layer Chunk

Span Delete Clear ← →

Layer Chunk

Text देश

No links or relations connect to this annotation.

Tag

(Chunk)

41 " " देश में कई स्थानों पर Dairy business से जो माहिलाएं जुड़ी हुई हैं, सीधे उनके bank account में पैसा transfer किया जाता है।

42 " अटल इनोवेशन मिशन ए आई एम – AIM के माध्यम से देश में Innovation एवं Entrepreneurship का कल्वर तैयार किया जा रहा है।

43 सरकार ने PMRF यानि Prime Minister Research Fellowship की घोषणा की है और मैं चाहूंगा कि आप सब इसका फायदा उठाएं।

44 "इसके तहत IIT, IISC, NIT जैसे संस्थानों के B Tech, M tech और M SC के Best Students को Prime Minister Fellowship दी जाएगी।

45 " इंटेलेक्युअल Property Rights पर, और Banks के बीच सहयोग से Business एनवाइरमेंट आसान हो रहा है।

46 " "Space Sector और Space Tech को लेकर आज भारत में जो बड़े Reforms हो रहे हैं, वो इसी की एक कड़ी है।

47 " आज ISRO की facilities को प्राइवेट सेक्टर के लिए Open किया जा रहा है।

48 "एक ऐसी strategy जो भारत के टेक्नोलॉजीकल एक्सपर्टोंज को आधार बनाकर, भारत को innovations का Global center बनाए।

49 " "देश में health और fitness को लेकर निरंतर awareness भी बढ़ रही है, और activeness भी बढ़ी है।

50 " Ease of Doing Business की रैंकिंग में भारत सिर्फ तीन वर्षों में 142 से 100वें नंबर पर पहुंच गया है।

CodeMixing project Sample-codemixed.txt 29-38 / 50 sentences [doc 1 / 2]

Layer Chunk

Span Delete Clear ← →

Layer Chunk

Text ट्रांसपोर्ट

No links or relations connect to this annotation.

Tag English

Unidentified Hindi Hindi Hindi English (Chunk) Unidentified English Hindi English Services के संबंध में ये गाइडलाइंस बहुत महत्वपूर्ण हैं।

29 " विकासशील देशों में Sustainable Consumption , E-Commerce और Financial Services के संबंध में ये गाइडलाइंस बहुत महत्वपूर्ण हैं।

30 " साथियों, Inflation पर लगाम लगाने की वजह से भी गरीब और मध्यम वर्ग के Consumers का आर्थिक फायदा हुआ है।

31 " इसके लिए National Digital Health Mission के तहत एक प्रभावी लेटरफॉर्म पर भी काम चल रहा है।

32 " Digital India अभियान की एक और खास बात रही कि इसमें Infrastructure के Scale और Speed , दोनों पर बहुत बल दिया गया।

33 " " One Nation , One Card , यानि देशभर में ट्रांसपोर्ट और दूसरी सुविधाओं के लिए पेमेंट का एक ही माध्यम , एक बहुत बड़ी सुविधा सिद्ध होने वाला है।

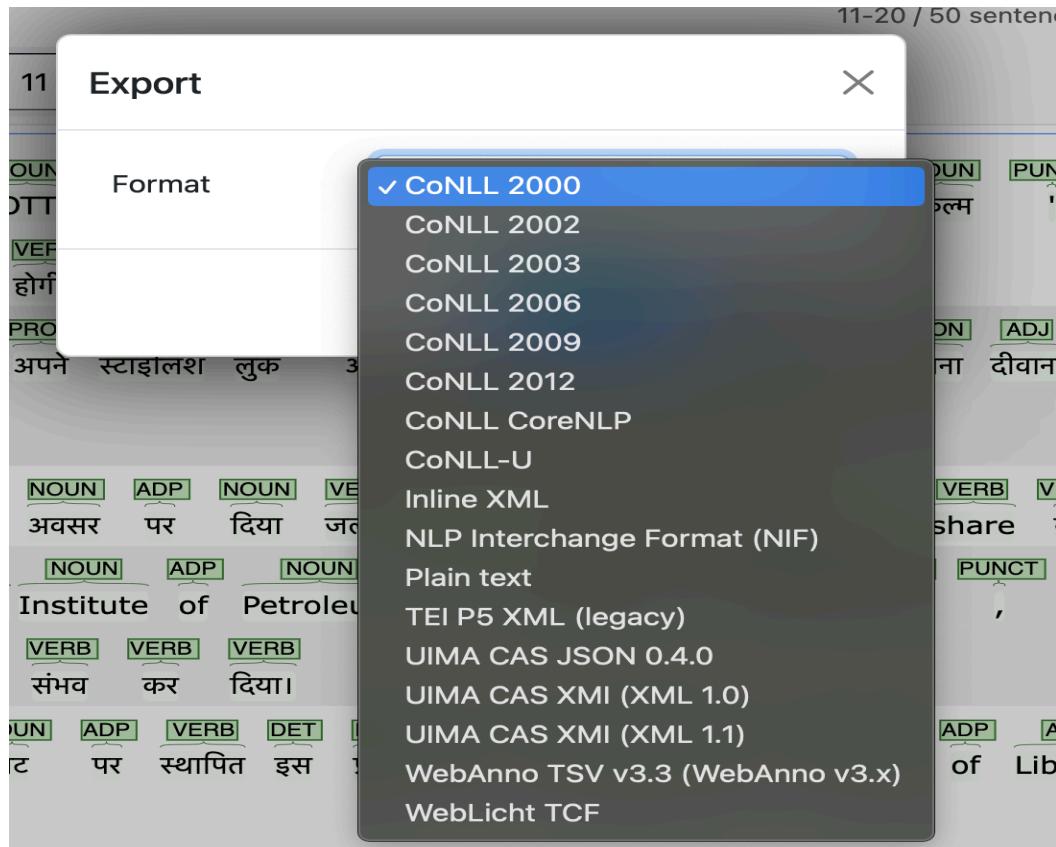
Repeat the same process for POS tagging.

The screenshot shows the INCEpTION annotation interface. On the left, there's a sidebar with 'Annotations' and 'Group by label' dropdowns, a 'Sort by score' toggle, and a 'Suggestions first' toggle. Below these are sections for 'ADJ', 'NOUN', 'VERB', 'PRON', 'ADP', 'NUM', 'PUNCT', and 'CCONJ'. The main area displays five sentences from a 'CodeMixing project' file named 'Sample-codemixed.txt'. The sentences are numbered 1 through 5. Each sentence is shown with tokens highlighted in green boxes, each labeled with its part-of-speech category. Sentence 1: पहले भी ला चुका है ऐसा फोचर 2017 में वॉट्सऐप ने अपने स्टेटस फंक्शन के लिए My contacts except... ऑप्शन ला चुका है। Sentence 2: PROPN ADP PROPN PUNCT ADJ NOUN PUNCT PROPN PROPN ADP NOUN CCONJ VERB स्कॉलॉड की प्रधानमंत्री ( First Minister ) निकोला स्टुर्जन ने पुलिस और इमीग्रेशन की कार्रवाई को एकत्रका और गलत करार दिया। Sentence 3: NOUN ADP ADJ NOUN NOUN VERB VERB PUNCT NOUN PUNCT ADP ADP VERB NOUN PFI पर पिछले साल सिटिजनशिप अमेडमेट एक्ट ( CAA ) के खिलाफ हुए प्रदर्शनों को फाइनेंस करने का आरोप लगा है। Sentence 4: X ADV DET NOUN NOUN ADP VERB VERB PUNCT CCONJ VERB NUM NOUN ADP NOUN " पहले सभी वोट TMC को मिलते थे , लेकिन बीते पांच सालों में BJP ने अपनी पकड़ इस गांव में मज़बूत की है। Sentence 5: ADJ VERB NOUN PUNCT NOUN PUNCT ADP VERB VERB NOUN CCONJ PROPN ADP " इंडियन प्रीमियर लीग ( IPL ) के बचे हुए मैच और भारत की मेजबानी में होने वाला टी - 20 वर्ल्ड कप UAE में होना लगभग तय है।

8. After completing the annotation, click on the download button to extract the annotations.

This screenshot shows the toolbar of the INCEpTION interface. It includes buttons for opening a folder, saving, navigating between files, and navigating between sentences. There are also buttons for zooming in and out, a search icon, and a settings gear icon. The top bar shows the project name 'CodeMixing project' and the file name 'Sample-codemixed.txt'.

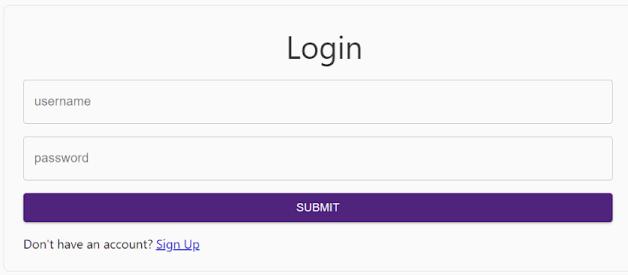
From the drop-down list, choose the appropriate option to download the annotation.



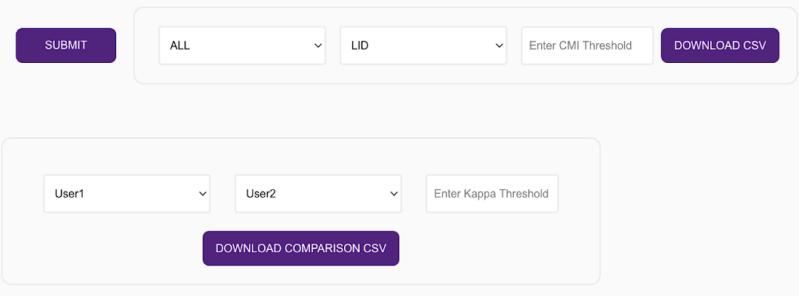
9. Upload the downloaded file to the drive folder for INCEpTION.

## COMMENTATOR

1. Download the COMMENTATOR folder from his link: <https://bit.ly/4dy3CTb>
2. Follow the instructions provided there to start the frontend server.
3. Navigate to the frontend folder. Inside the frontend folder, open the terminal and hit the command “npm i” or “npm install”. Make sure you have node installed on your machine. (<https://nodejs.org/en/download/>)
4. Now type command “npm start”. A new tab should open in your default browser, if not go to url, <http://localhost:3000/login>
5. Navigate to the backend folder and start the backend server according to the provided instructions.



6. Click on the signup button and create an admin account which will monitor the process.
7. Open MongoDB database and set `admin=True` to create a superuser/admin account.
8. After logging, upload the file which you would like to annotate by choosing the file, and after submitting, you can create another user account for annotating these sentences.



- Sign up for a user account and click on continue.
- Select the task from the task panel for which you want to annotate.

**Choose any Task to proceed**

Token-level language Identification

Token-level POS tagging

Matrix language Identification

Token-level entity labelling

Spelling correction & normalization

Parallel Translations

- For Token-level Language Identification: You shall see the sentences preloaded for you.

**COMMENTATOR** Edit Annotations USER ⚙️ #13

**Steps to Follow!**

- Select the **individual word tags** for the sentence.
- Individual word tags get colored according to the color convention below:
- 'English' - green 🟢
- 'Hindi' - yellow 🟡
- 'Unidentified' - blue 🟤
- If necessary, adjust the tags according to the context of the sentence.
- Click "Submit" to save your changes and proceed to the next sentence.

Canada के प्रधानमंत्री जी ने दीपावली के अवसर पर दिया जलाते हुए अपनी तस्वीर Twitter पर share की है।

Individual Word Tags							
Canada	के	प्रधानमंत्री	जी	ने	दीपावली	के	
अवसर	पर	दिया	जलाते	हुए	अपनी	तस्वीर	
Twitter	पर	share	की				

Feedback:

**SUBMIT**

12. Next, update individual word tags wherever necessary, the color strategy has been defined on the screen on the right along with the rules. If you have any feedback related to the dataset or the framework, you can simply add it there. Note the annotation time per sentence.

13. If you want to make any modifications then click on the Edit Annotations button, enter the number for which you want to annotate the sentence and make modifications, if any.

Sentence ID	Date	Sentence
14	2024-08-01	विशेष रूप से CSIR, Indian Institute of Petroleum, Dehradun के वैज्ञानिकों को, जिन्होंने bio-fuel से विमान उड़ाने की तकनीक को
15	2024-08-01	पुरुरात में नवदा नदी के तट पर स्थापित इस प्रतिभा की ऊँचाई अमेरिका के Statue of Liberty से दो गुने है।
16	2024-08-01	Sustainable development के लिए कैसे उनके traditions को हमें अपने जीवन में अपनाने की आवश्यकता है, उनसे कुछ सीखने की जरूरत है।
17	2024-08-01	Preventive health-care के रूप में योग ने, नये सिरे से दुनिया-भर में अपनी पहचान बनाई है।
18	2024-08-01	किसान चाची ने अपने डिलेक्ट की 300 महिलाओं को 'Self Help Group' से जोड़ा और आर्थिक रूप से स्वावलंबी बनाने के लिए प्रेरित किय
19	2024-08-01	इस अवसर पर हम देश भर में 'FIT INDIA MOVEMENT' launch करने वाले हैं।

14. Hit submit and inform post annotation of the 10 sentences.

Repeat the same for POS tagging:

**Steps to Follow!**

1. Automatic POS Tags are assigned to every lexicon.
2. Update individual POS tags.
3. Click on the dropdown menu, a list of tags will appear.
4. Choose the updated POS Tag.
5. Submit to load the next sentence.

**Individual POS Tag**

स्कॉटलैंड की प्रधानमंत्री (First Minister) निकोला स्टुर्जन ने पुलिस और इमीग्रेशन की कार्रवाई को एकतरफा और गलत करार दिया।

Feedback:  
Enter your feedback here

Sentence ID	Date	Sentence
1	2024-08-01	स्कोटलैंड की प्रधानमंत्री ( First Minister ) निकोला स्टुजर्न ने पुलिस और इमोरेशन की कार्रवाई को ...
2	2024-08-01	PFI पर मिले गाल सिटि�बंशिप अण्डपोट एक ( CAA ) के खिलाफ हुए प्रदर्शनों को काइरेस ...
3	2024-08-01	पहले सभी थोड़ी TMC को मिलते हैं , लेकिन यीते थांच शालों में BJP ने अपनी ...
4	2024-08-01	इंडियन प्रीमियर लीग ( IPL ) के बचे हुए मैच और भारत की योजनाएँ में होते ...
5	2024-08-01	B Com की चढ़ाई भी कॉलोन नहीं कर पाई , कर्याक्रियाएँ उस वक्त मरी केमेटी की ...
6	2024-08-01	DRDO के मुताबिक , यह किट शरीर में SARS - CoV - 2 के वायरस और ...

15. After completing the annotation, login using Admin credentials and download the results file by choosing the specific task for single or multi-user, click on the download button to extract the annotations. You can also calculate the Inter Annotator Agreement by selecting two usernames and entering the kappa threshold value.

Choose file	No file chosen	SUBMIT	ALL	LID	Enter CMI Threshold	DOWNLOAD CSV
User1	User2	Enter Kappa Threshold	DOWNLOAD COMPARISON CSV			