

# An Open-source Web-based Application for Development of Resources and Technologies in Underresourced Languages

Siddharth Singh<sup>1,3</sup>, Shyam Ratan<sup>2,3</sup>, Neerav Mathur<sup>1,3</sup>, Ritesh Kumar<sup>3</sup>

<sup>1</sup>CTrans, Dr. Bhimrao Ambedkar University, Agra, India, <sup>2</sup>CALTS, University of Hyderabad, Hyderabad, India, <sup>3</sup>UnReal-TecE LLP, Agra, India



[unreal.tece@outlook.com](mailto:unreal.tece@outlook.com)

<https://github.com/unrealtecellp/life>

<http://life.unreal-tece.co.in>

## Dashboard

LIFE by UnReal-TecE

Welcome, life\_admin Logout

Linguistic Field Data Management and Analysis System [LIFE]

Create New Project

Questionnaire Data Lexicon Models

Change Active Project

Active Project: LDCIL Admin Lexicon

View/Edit Project Share Project Download Dictionary Manage Project

User Manual

## Create Questionnaire

LIFE by UnReal-TecE

Welcome, calls Logout

Q\_36

Q2023110125623373025

Domain

Generation

Elicitation Method

Narration

Prompt Brj-Devanagari

Prompt Audio Transcription Brj-Devanagari

Audio Instruction Brj-Devanagari

Prompt Image Brj-Devanagari

Image Instruction Brj-Devanagari

Prompt Multimedia Brj-Devanagari

Multimedia Instruction Brj-Devanagari

Target

Reflections and Suggestions

## ASR and VAD Interfaces

Project Name: D\_Transcription\_5

Generate Transcription for this file

It will overwrite any boundaries created earlier

Filename: A20231121003411025367\_eng2.wav

Audio Duration: 0.06

Model: Hindi-openal/whisper-small

Target Script: Devanagari

Transcribe

Project Name: D\_Speed-TB\_Phase\_1\_Braj\_Bhasha\_Narration

Make Boundaries for this file

It will overwrite any boundaries made earlier using this model

Source: karishakyadv1830\_16782227

Filename: A20230714112402011014\_281474976757233.wav

Audio Duration: 0.19

Boundary Pause Threshold (between 0.00001-10.0 seconds)

Suggested Values: For sentence: 0.3; For words: 0.01; For phones: 0.001

0.3

Minimum Boundary Size (between 0.5-10.0 seconds) 2.0

Overwrite my Boundaries

Make Boundaries

## Transcribe and Annotate

LIFE by UnReal-TecE

Welcome, The\_SpecD\_IL\_Project Logout

Start Time: 0.9914132840549736

End Time: 2.1058825823005876

Total Time: 1.114465296245614

Devanagari: ५ का जाइ रंजी ऐ ।

Transcription in Devanagari

५ का जाइ रंजी ऐ ।

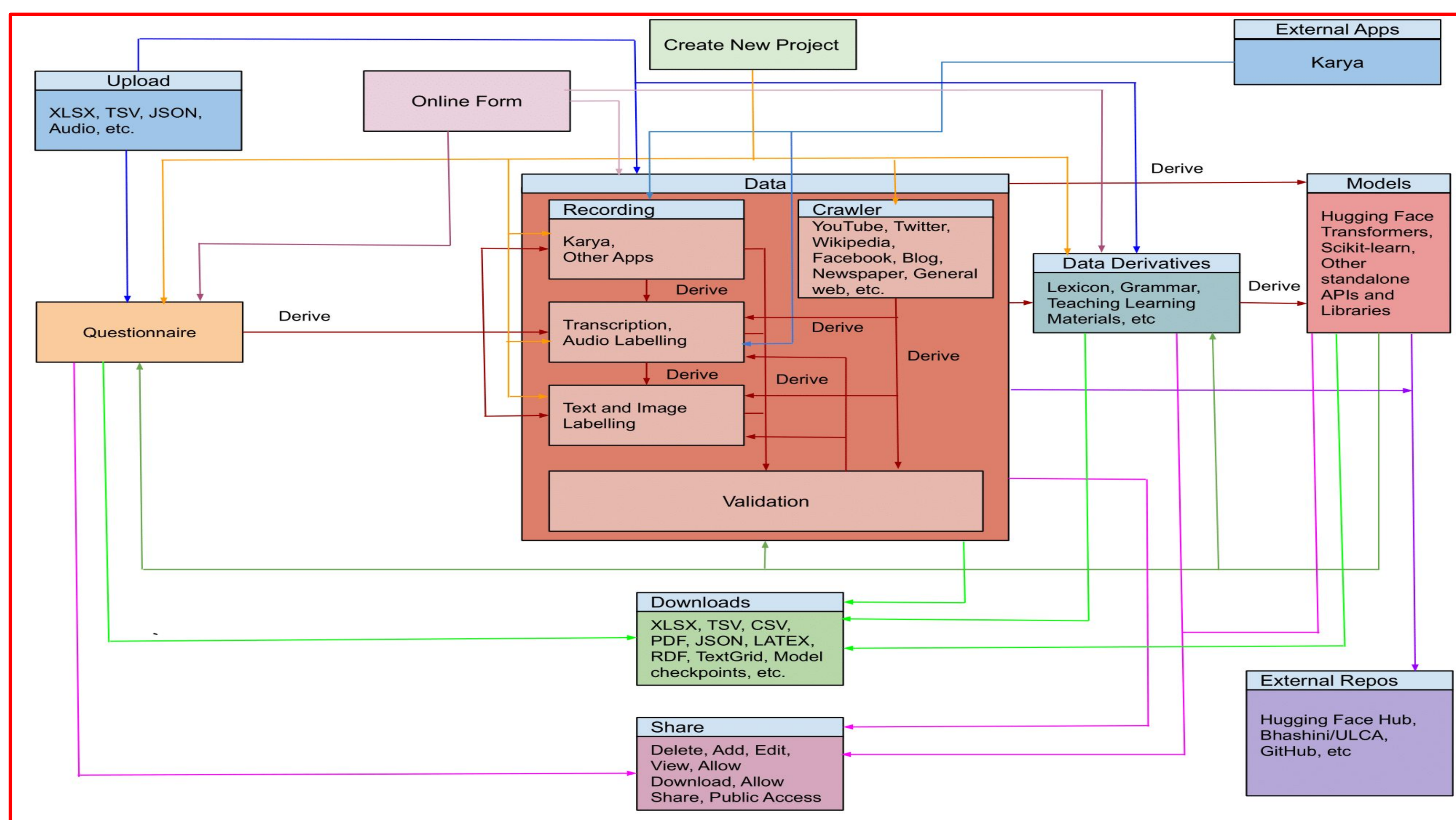
Add Translation

Translation in English

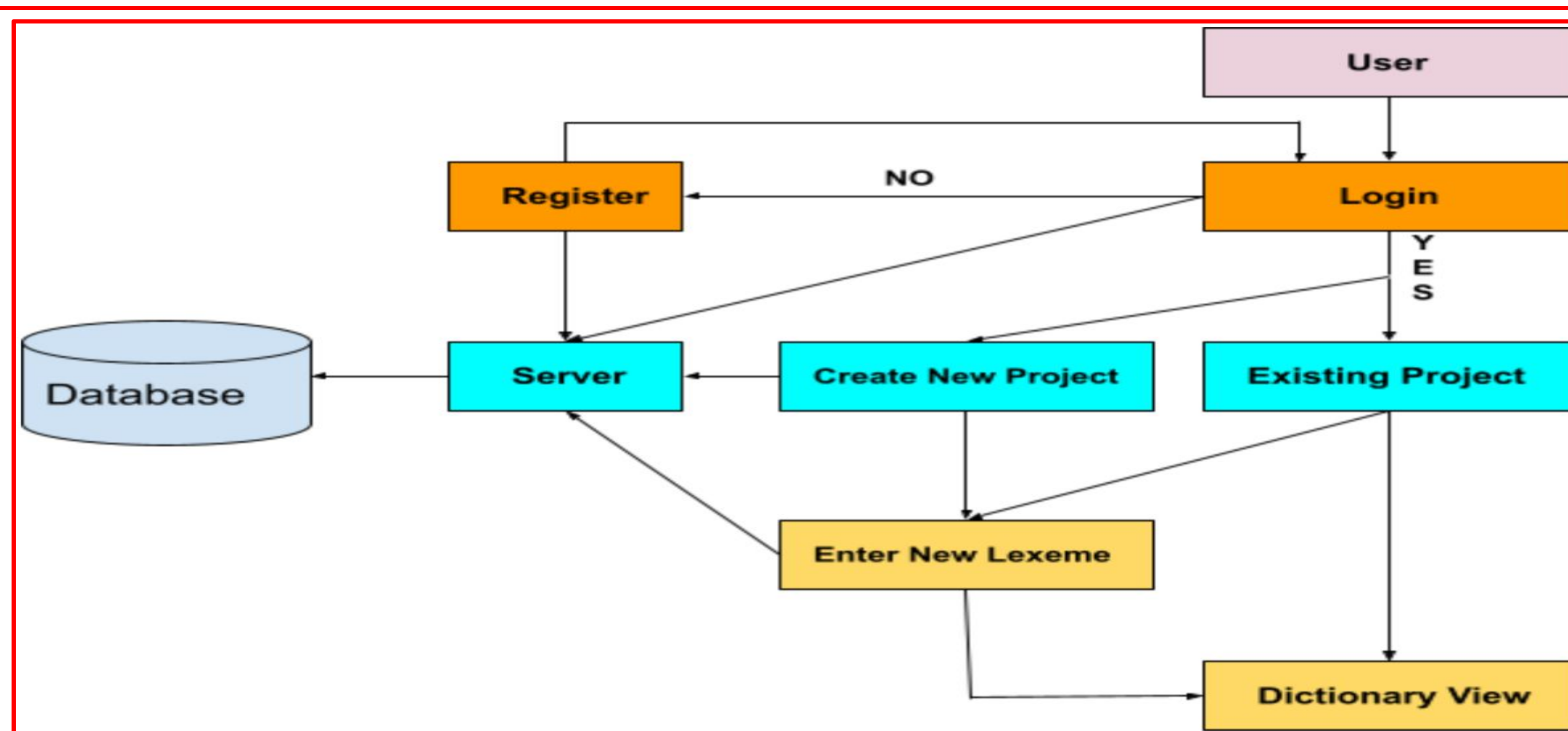
Where is he going ?

Comments

WORK FLOWS



MODEL DIAGRAM



BUILD LEXICON

LIFE by UnReal-TecE

Welcome, shyam Logout

Project Name : Sumi Dictionary

Search

Total Records: 13

lexemeld	headword	Grammatical Category	Gloss(eng)	Gloss(hin)	View	Edit	Delete
L2023032409362161769	irua	Verb	awake		View	Edit	Delete
L20230324094305061705	ij	Verb	become		View	Edit	Delete
L20230324094447470363	ovekju	Adverb	before		View	Edit	Delete
L20230324094614923969	nlu	Noun	belly		View	Edit	Delete
L20230324095017817086	ahju	FusedPreposition	behind		View	Edit	Delete
L20230324095140669190	axo	Adverb	down		View	Edit	Delete
L20230324095308851182	jeu	Verb	fly		View	Edit	Delete
L20230324095412537484	asa	Noun	hair		View	Edit	Delete
L20230419224709824800	mo da	Noun	Boy		View	Edit	Delete
L20230419224915211152	mo di	Noun	Girl		View	Edit	Delete
L20230419225228191553	croa	Noun	Boy		View	Edit	Delete
L20230419225304399643	croi	Noun	Girl		View	Edit	Delete
L20230419225445851188	lugal	Noun	wife		View	Edit	Delete

CRAWL DATA

LIFE by UnReal-TecE

Welcome, calls Logout

Project Name : D\_Youtube\_Hindi\_Data

Youtube API Key

Data Type

Crawl Mode

Add Videos/Channels Id for Data

Crawl

## Annotate text and images

LIFE by UnReal-TecE

Welcome, calls Logout

File Name: youtube\_speed\_35\_sai

00: TOKEN-LEVEL:

01: NOUN (N):

02: PRONOUN (P):

03: DEMONSTRATIVE (DM):

04: COMMENTS:

05: APPENDIX:

06: DEMONSTRATIVE (DM):

07: COMMENTS:

TECH STACK



Acknowledgements

LDC-12, CML, METTY, Govt of India

Karya Inc

CSDR