

No	Text Corpus Specification	
1	Brief Description (approx. 500 characters)	Hindi – Telugu Parallel corpus developed under ILMT Hindi-Telugu Pilot by IIIT-Hyderabad (Final)
2	Linguality Type (Monolingual, Bilingual or Multilingual corpus)	Bilingual Data
3	Language	Hindi – Telugu
4	Domain of the Corpus (Corpora prepared from specific domains,e.g Tourism,General,Legal etc.)	Chemistry, Law, News & General, HealthCare, Education Others, open education books
7	Size of Text Corpus (tokens, sentences etc)	506178 Parallel Sentences for Hindi-Telugu
	Format of Text corpus (File Format text/xml)	text
8	Text Encoding	Utf-8
9	If Annotated, Annotation Type	Parallel data
10	Whether the corpus is validated/verified (Cleaned & Vetted)	Validated
12	Conformance to Standards/Best Practices	Human Verified
13	Data Source	NCERT BOOKS, NPTEL, NEWS , goverment healthcare sites, Open Education books
20	Funding Agency(if Any)	Meity , India
21	License Type	Attribution-NonCommercial CC BY-NC
22	Whether Copyrighted: If yes, please mention Copyright owner	Owner : LTRC , IIIT-Hyderabad, India
23	Sample Corpus (Mandatory for Upload)	Provided
24	Other Details (If Any):	