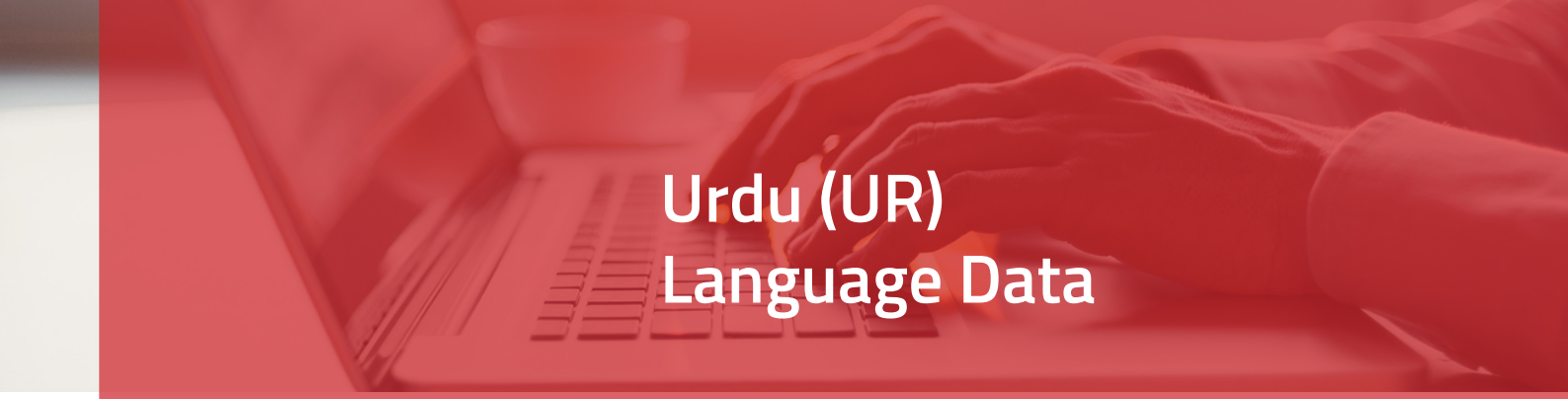


Urdu (UR) Language Data

by Bitext

bitext we help AI understand humans

INFLECTIONAL FORMS LIST

includes all the standard inflectional forms for nouns, verbs, adjectives, postpositions, conjunctions, etc. Each form will be annotated with the lemma (root form), POS, and morphological attributes (tense, mood, gender, number, person, case, possessive-gender, possessive-number, possessive-case).

DERIVATIONAL FORMS LIST

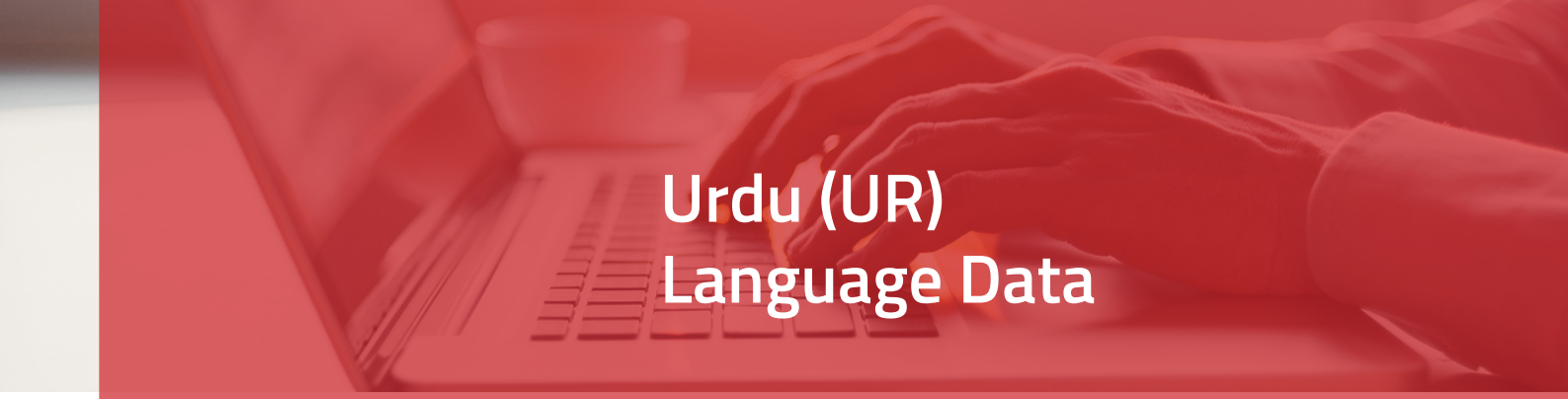
includes all the standard derivational forms including verbs derived from nouns, nouns derived from verbs and adjectives derived from nouns, and common compound words. Each form will be annotated with the lemma (root form), POS, and morphological attributes (tense, mood, gender, number, person, case, possessive- gender, possessive-number, possessive-case).

EXTENDED FORMS LIST

includes the result of extending the inflectional and derivational forms lists as a result of considering additional morphological phenomena such as common combinations of postposition suffixes. Each form will be annotated with the lemma (root form), POS, and morphological attributes (tense, mood, gender, number, person, case, possessive-gender, possessive-number, possessive-case).

About Bitext

Bitext has broken through the barriers that block multi-language text analysis. The company's Deep Linguistics Analysis Platform supports 77 languages at a lexical level and +20 at a syntactic leveland makes the company's technology available for a wide range of applications in Artificial Intelligence, text analytics and the new wave of products designed for voice interfaces such as chatbots and assistants.



Urdu (UR) Language Data

by Bitext

bitext we help AI understand humans

NAMED ENTITIES FORMS LIST

includes the data regarding named entities comprising person names, places, companies and organizations. Each form will be annotated with the lemma (root form), POS, and morphological attributes (tense, mood, gender, number, person, case, possessive-gender, possessive-number, possessive-case and entity-type).

info@bitext.com

www.bitext.com

FREQUENCY INDICATION

includes the data regarding the relative frequency of appearance for the words in the above lists in the given language. The relative frequency could be in the range of 0-255, or as requested.

OFFENSIVE LANGUAGE FLAG

includes information per word indicating if the word might be considered offensive in certain contexts.

VOLUME OF LANGUAGE DATA

- Total number of lemmas: 15,000 lemmas
- Total number of forms: 200,000 forms
  - Verbs: 20,000 forms (10%)
  - Nouns: 180,000 forms (89%)
  - Adjectives: 2,000 forms (1%)
  - Other: 1,000 forms (1%)