Armenian (HY) Language Data

by Bitext

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.

INFLECTIONAL FORMS LIST

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

DERIVATIONAL FORMS LIST

includes all the standard derivational forms including superlatives. Each form will be annotated with the lemma (root form), POS, and morphological attributes (tense-mood, person, number, case, definiteness, degree, possessive person and possessive number).

EXTENDED FORMS LIST

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

Armenian (HY) Language Data

by Bitext

bitext — we help AI understand humans

✉ info@bitext.com

🌐 www.bitext.com

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, person, number, case, definiteness, degree, possessive person, possessive number and entity-type).

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: 6,000 lemmas
- Total number of forms: 150,000 forms
  - Verbs: 30,000 forms (20%)
  - Nouns: 120,000 forms (77%)
  - Adjectives: 2,000 forms (2%)
  - Other: 1,000 forms (1%)