

TEPROLIN

Denumire: Tehnologii pentru procesarea limbajului natural - text (TEPROLIN)

Obiectivul general al acestui proiect este dezvoltarea unui set de tehnologii avansate pentru procesarea limbajului natural (text) în limba română: analiza morfologică, sintactică și semantică a textelor, cu adnotarea textului colectat în Proiectul 1 (COBILIRO) pe diferite nivele lingvistice (fonem, silabă, leamă, parte de vorbire, etc).

Aplicabilitate: Aceste tehnologii vor fi aplicate în sisteme de procesare și interpretare automată a textelor în limba română, în crearea modelelor de limbaj pentru interfețele de recunoaștere a vorbirii, respectiv în procesarea textului pentru realizarea interfețelor de sinteză din text a vorbirii.

Activitățile vor avea în vedere: Inventarierea atentă a producerea unei tehnologii integrate pentru procesarea limbajului natural în limba română, procesarea și adnotarea pe diferite niveluri lingvistice a corpusului bimodal generat în cadrul Proiectului 1 COBILIRO, evaluarea sistemelor de recunoaștere a vorbirii (Proiect 3, TADARAV) și de sinteză din text a vorbirii (Proiect 4, SINTERO) antrenate cu corpusul bimodal din Proiectul 1 COBILIRO. Valorizare și utilizare cecuri de tip A1, A2, B.

1.5. Definirea specificațiilor funcționale și arhitecturale ale platformei integrate și configurabile de prelucrare a textelor

1.6. Definirea modulelor software și a serviciilor oferite de proiect; identificarea adaptărilor pentru modulele NLP existente și a modulelor noi necesare

1.7. Realizarea adaptărilor necesare pentru modulele NLP existente, identificate în activitățile 1.5 și 1.6

1.8. Crearea și validarea (eventual cu corectările manuale necesare) a unui [lexicon specific](#) corpusului bimodal și încorporarea sa în lexiconul existent