

LIETUVIŲ KALBOS MORFOLOGIŠKAI IR SINTAKSIŠKAI
ANOTUOTI TEKSTYNAI

Santrauka

Anotuoti tekstynai – pagrindiniai ištekliai, be kurių neapsieinama plėtojant kalbos technologijas, kompiuterizuojant kalbą. Nuo anotuotų tekstynų dydžio, kokybės, parengimo principų neretai priklauso ir kitų įrankių kūrimo galimybės. Straipsnyje apžvelgti Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centre parengti du anotuoti lietuvių kalbos tekstynai: morfologiškai anotuotas tekstynas MATAS ir sintaksiškai anotuotas tekstynas ALKSNIS. Pristatyta jų struktūra, naudotos anotavimo pažymos, anotavimo programos. Daugiau dėmesio skirta naujesniam anotuotam tekstynui ALKSNIS, paminėti sunkiausi jo anotavimo aspektai. Taip pat daug dėmesio straipsnyje skirta paieškai abiejuose tekstynuose per ANNIS sistemą. Pateiktos konkrečios taisyklės ir jų kombinacijos, leidžiančios atlikti tiek paprastąją (ieškoti konkrečios žodžio formos, antraštinės formos, tam tikros sintaksinės funkcijos ir pan.), tiek sudėtinę paiešką (ieškoti iš kelių žodžių sudarytų junginių, kelių gramatinių kategorijų kombinacijų ir pan.).

ESMINIAI ŽODŽIAI: tekstynas, automatinė morfologinė analizė, automatinė sintaksinė analizė, kalbos technologijos.

AGNĖ BIELINSKIENĖ, LOIČ BOIZOU, ERIKA RIMKUTĖ
Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centras

V. Putvinskio g. 23-216, 44243 Kaunas

agne.bielinskiene@vdu.lt

erika.rimkute@vdu.lt

lboizou@gmail.com