Reference in povezave

V tem poglavju so zbrane relevantne reference in povezave na projekte, v katerih se je postopek segmentacije razvijal in uporabljal.

Projekti, na katerih se je označevalni sistem razvijal oz. uporabljal

JOS - Jezikoslovno označevanje slovenskega jezika: metode in viri

Sporazumevanje v slovenskem jeziku

Universal Dependencies

Janes - Viri, orodja in metode za raziskovanje nestandardne spletne slovenščine

Razvoj slovenščine v digitalnem okolju

Orodje Obeliks za tokenizacijo in stavčno segmentacijo

https://github.com/clarinsi/obeliks

Reference

Krek, Simon; Erjavec, Tomaž; Dobrovoljc, Kaja; Gantar, Polona; Arhar Holdt, Špela; Čibej, Jaka in Brank, Janez. The ssj500k training corpus for Slovene language processing. V: Fišer, D. in Erjavec, T. Jezikovne tehnologije in digitalna humanistika: zbornik konference: 24.-25. september 2020, Ljubljana, Slovenija. Ljubljana: Inštitut za novejšo zgodovino, 2020. Str. 24–33.

http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Krek-et-al_The-ssj500k-Training-Corpus-for-Slovene-Language-Processing.pdf [PDF]

Revision #7
Created 2023-11-12 17:37:52 UTC by Tina Munda
Updated 2023-11-15 11:33:45 UTC by Tina Munda