

**UČEŠĆE U COST AKCIJI
D-READING
STANJE RADOVA MAJ 2019.**

Cvetana Krstev & Ranka Stanković

Jerteh, 30. maj 2019.

Radna grupa WG1

- Srpski deo korpusa:
 - Odabrano je oko 100 dela – teško je reći, neka dela ispadaju sa spiska, neka se dodaju
 - Do sada je potpuno obrađeno u skladu sa zahtevima akcije **20** dela, postavljeno na sajt akcije
 - Dela koja su pročitana i čekaju konačnu obradu da budu uključena u korpus (dopuna rečnika, zaglavlje) – **8**
 - Dela koja su raspodeljena i čitaju se – **12** (**7** će sigurno završiti, ostale ću tokom juna blago podsetiti)
 - Dela koja su skanirana a još nisu raspodeljena: **22**
 - **U narednom periodu, regrutovanje novih akcijaša**

Kako izgleda radna verzija prezentacije korpusa

- <https://distantreading.github.io/ELTeC/index.html>
- Vidi se da naš korpus ne zadovoljava nijedan kriterijum reprezentativnosti:
 - Podjednaka zastupljenost 4 perioda
 - Podjednaka zastupljenost romana po dužini
 - Podjednaka (ili velika zastupljenost ženskih autora)
 - „kanoničnost“ – tu možda i ne stojimo loše
 - Ne više od 3 dela istoga autora – nemoguće dostići 100 bez Ignjatovića, Komarčića, Sremaca...
 - Videćemo šta će na kraju iz toga izaći

Dosadašnje i buduće aktivnosti

- Poslednji sastanak svih radnih grupa bio je u Lisabonu
- Joint Action Meeting, Lisbon, January 21-23, 2019
- Radionice:
 - Distant Reading Training School 2018: Workshop 1, Methods and Tools of Distant Reading Adapted to Multiple European Languages (**Mihailo**); Workshop 2: Theoretical Concepts and their Confrontation with Computational Methods, National University of Ireland, Galway, December 5-7, 2018.
- Training School Optical Character Recognition and Text Encoding for ELTeC, Würzburg University, April 16-17, 2018 (**Jelena A**)
- STSM, Lion grupa vezana za alat TXM (**Jelena J**)

Radna grupa WG2

- Na adresi <http://brat.jerteh.rs/> instalisan sistem za ručnu anotaciju imenovanih entiteta
- Urađena anotacija uzoraka kolekcija za 8 jezika po 20 romana
 - cze, eng, fra, hun, nor, por, slv, srp
- Još 2 jezika će priključiti
 - deu, ita
- Portal za transformaciju, obeležavanje i analizu imenovanih entiteta: <http://nerbeyond.jerteh.rs/>
 - BRAT <-> CoNLL02 <-> XML (inline)
 - Anotacija spaCy I StanfordNER
 - Statistička analiza i evaluacija
- Dalji plan za kolekciju na srpskom:
 - anotacija uzorka korpusa gramatičkim informacijama

Aktivnost grupe WG4

- „Diseminacija“ i promocija
- ideja je da svi jezici koji su zastupljeni, ili će biti zastupljeni, u korpusu napišu neki kratak tekst o nekom svom romanu iz korpusa kao vid upoznavanja ljudi sa delima raznih književnosti.
- Do sada su napisana 2 članka (Litvanija i Rumunija)
- <https://www.distant-reading.net/distant-reading-recommends-kuprelis-the-hunchback-by-ignas-seinius/>
- <https://www.distant-reading.net/distant-reading-recommends-iancu-jianu-by-n-d-popescu/>
- Za nas će takav članak napisati Vasilije Milnović iz UBSM koji aktivno učestvuje u akciji.
- Na moj predlog, to će biti roman „Nove“ Jelene Dimitrijević

Budući sastanci

Working group meetings Budapest, Hungary, on September 23 to 25, 2019.

- We plan to co-locate our next Training School with Working Group meetings during these days.
- Please note that this event is co-located with the "DH_Budapest 2019" conference organized by our Action member Gabor Palko from September 25 to 27, 2019.