



Co-financed by the European Union  
Connecting Europe Facility

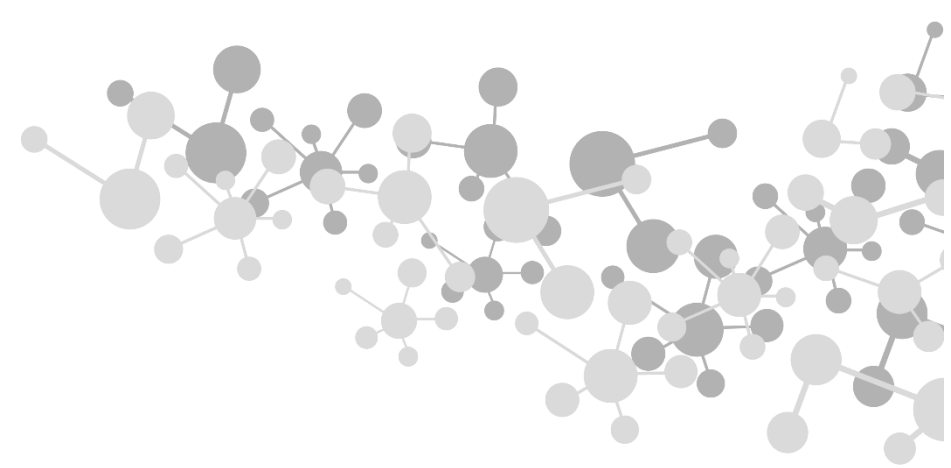
# Nevronski prevajalnik

## Ang $\longleftrightarrow$ Slo

Matjaž Rihtar  
AI Lab, Institut Jožef Stefan

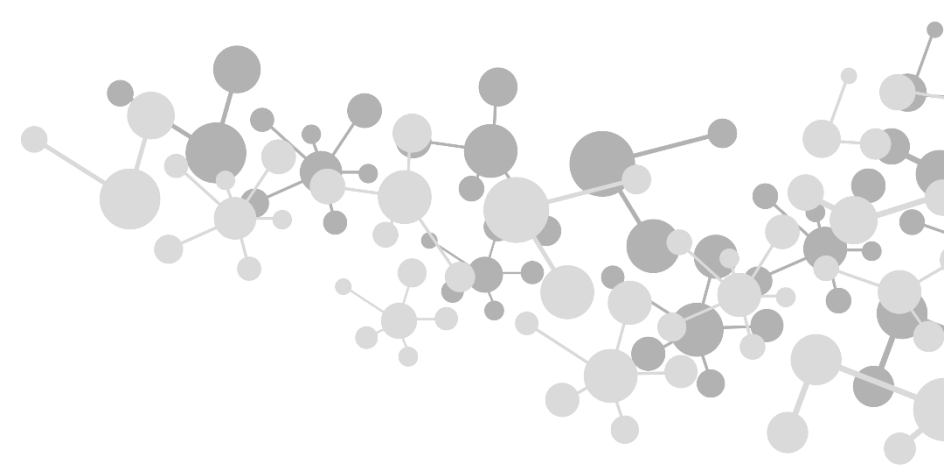
Ljubljana, 24. april 2018





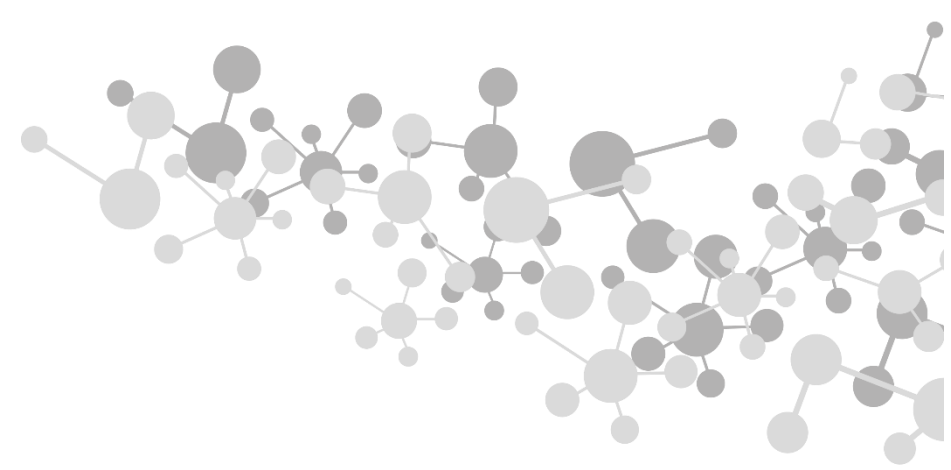
# Nevronski prevajalnik: Namen

- Do sedaj: na osnovi pravil in statističnih metod (Moses)
- Z nevronskimi prevajalniki dosežemo boljše rezultate
  - 'Google translate' preklopil novembra 2016
- Pogoji:
  - velik korpus
  - dobra strojna oprema (obvezno grafične kartice z veliko pomnilnika)



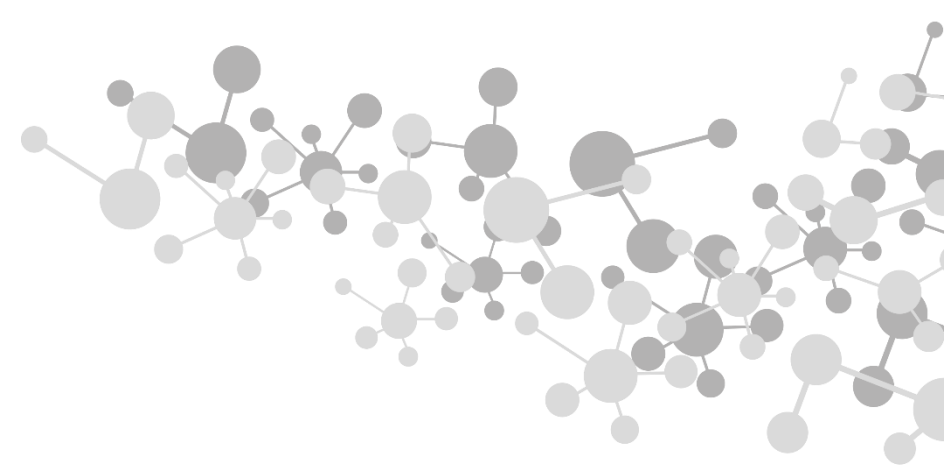
# Nevronski prevajalnik: Korpusi

- Angleško-slovenski korpusi na sistemu Opus (prost dostop):
  - OpenSubtitles2018 (21.1 M stavkov, 172.9 M pojavnic)
  - OpenSubtitles2016 (11.8 M stavkov, 144.2 M pojavnic)
  - DGT (3.2 M stavkov, 71.7 M pojavnic)
  - Europarl (0.6 M stavkov, 16.9 M pojavnic)
  - EMEA (1.1 M stavkov, 11.7 M pojavnic)
  - ~~EUbookshop (0.4 M stavkov, 17.1 M pojavnic)~~



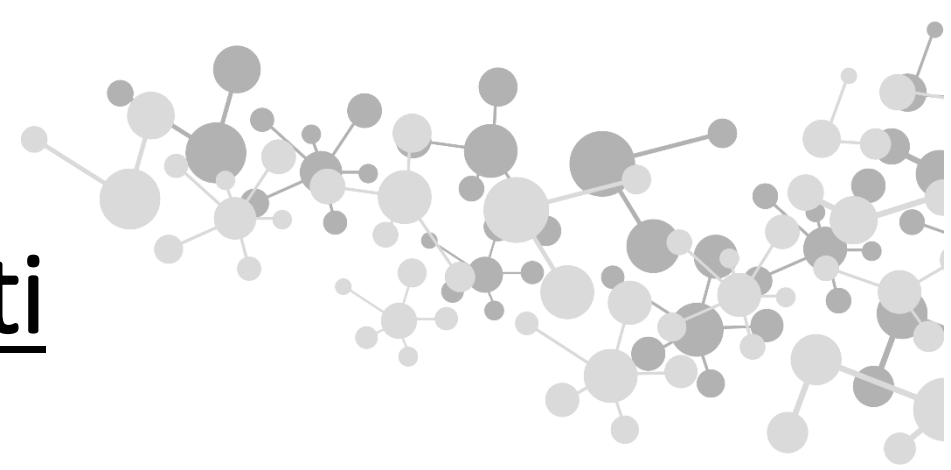
# Nevronski prevajalnik: Korpusi

- Skupaj:
  - 37.8 M stavkov, 417.4 M pojavnic
  - Še vedno premalo!
  - Primer – angleško-nemški korpus na sistemu Opus:
    - 95.4 M stavkov, 1354.4 M pojavnic



# Nevronski prevajalnik: Orodja

- Nematus (python, Theano)
  - Univerza v Edinburghu
- Marian (C++)
  - Univerza Adam Mickiewicz, Poznań
- Strojna oprema:
  - Unix, 2x Xeon, 128 GB RAM, 2x NVIDIA GeForce GTX 1080 Ti (11 GB RAM)

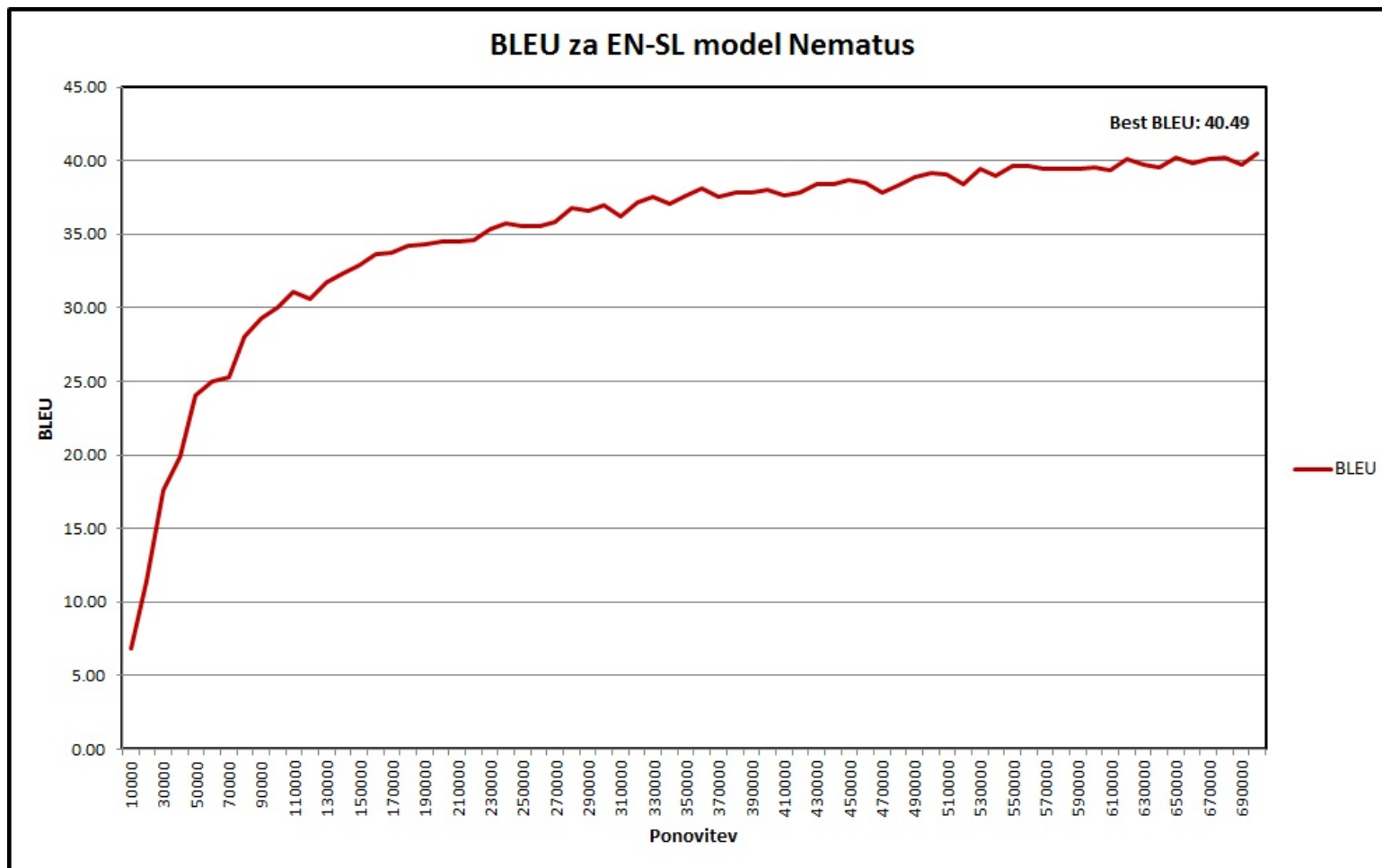
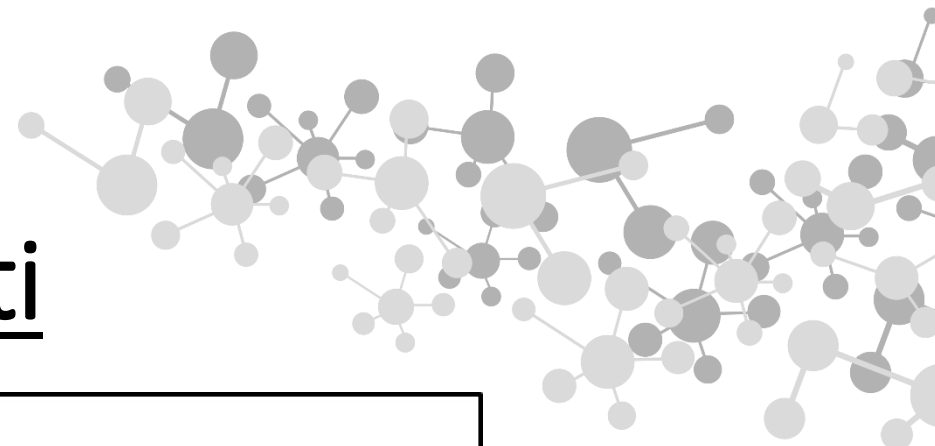


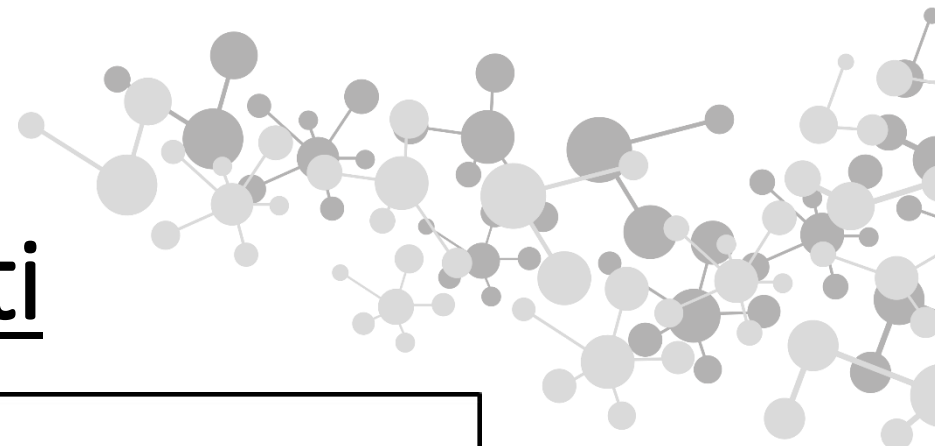
# Nevronski prevajalnik: Rezultati

- Ustvarjanje modelov
  - 10 dni na eni grafični kartici za model
- Izboljšave modelov
  - Dodatni visoko-kvalitetni korpusi (lahko manjši)
  - Domenska specializacija

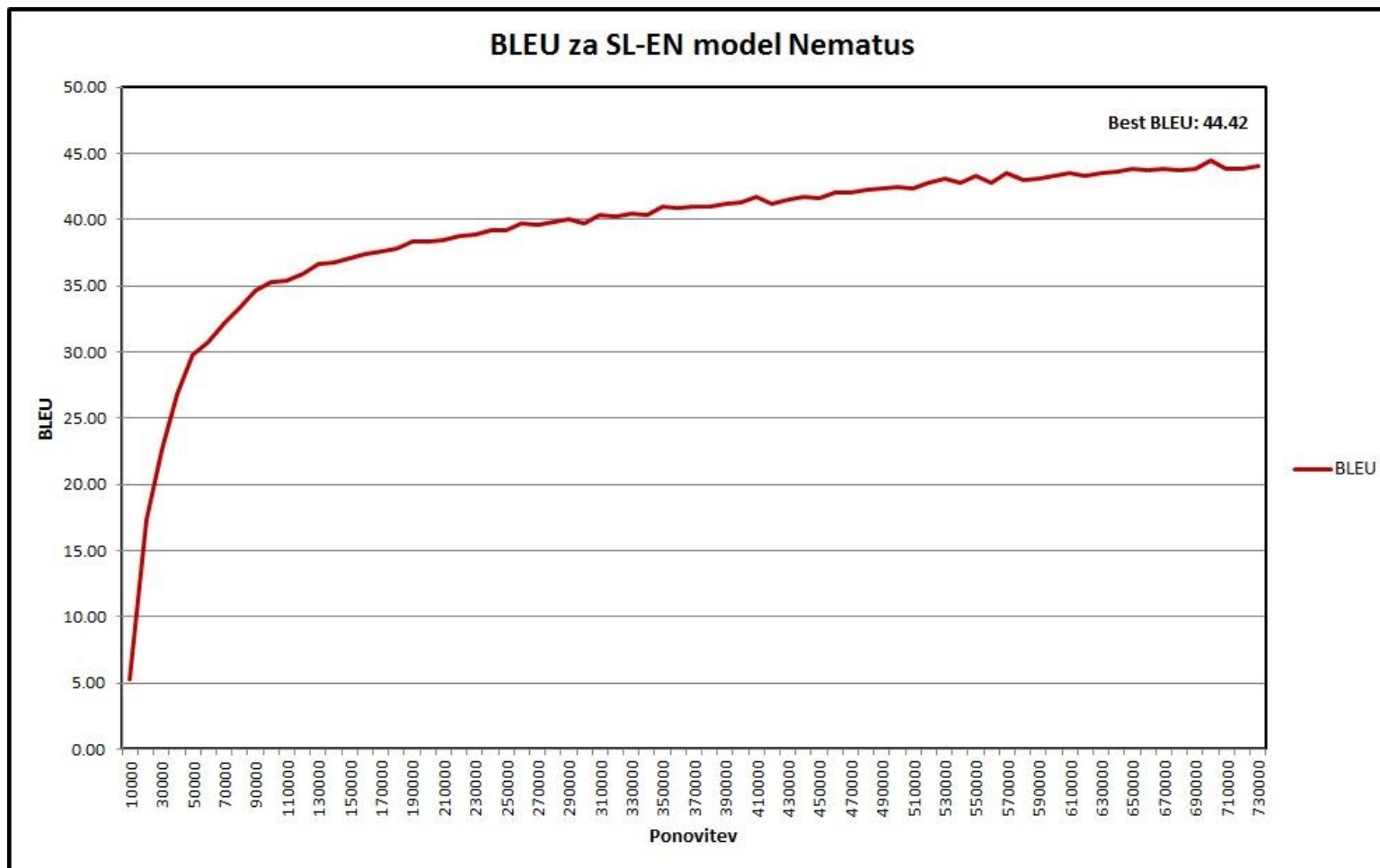


# Nevronski prevajalnik: Rezultati

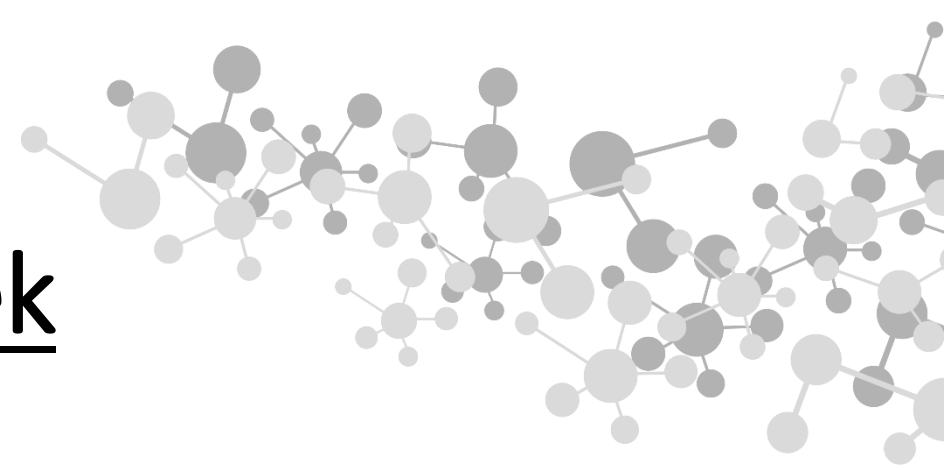




# Nevronski prevajalnik: Rezultati







# Nevronski prevajalnik: Povzetek

- Dober začetek!
- Potrebujemo več visoko-kvalitetnih korpusov