Ekstrakcija likova iz kratkih priča

Gorana Levačić

Tomislav Horina

***Sažetak*—U ovom radu smo radili klasifikator za likove u kratkim pričama. Koristili smo kratke priče s Gutenberg projecta.Priče smostalno označavali jer postojeća riješenja ne prepoznaju na primjer kralja kao lika.Promatrali smo ponašanje klasifikanirao na 4 načina zadavanja tesnog skupa.Vrste su :**

1) **bez ikakve promjene**
2) **priče bez interpukcijski znakova**
3) **sve riječi u lowercaseu**
4) **sve riječi u lowercaseu i bez interpukcijski znakova**

Isprobali smo algoritme klasifikacije koji su preporučeni za izradu tog modela – skriveni markovljev model(eng. Hidden Markov Model, HMM) i uvjetno slučajno polje(eng. Conditional Random Fields, CRF) te smo stanford Ner učili na naše ozanke. (sada tu treba ici nesto rezzultatima to treba nadopisati.)

I. Uvod

U ovom projektu bavit ćemo se ekstrahiranjem likova iz kratkih priča, konkretno priča za djecu. Taj problem pripada problemu ekstrakcije, odnosno identifikacije entiteta u tekstu, poznatiji pod engleskim nazivom named-entity recognition (NER). NER je podvrsta zadaće crpljenja obavijesti (information extraction), u kojoj se svakom elementu teksta pridjeljuje neki atribut. U općem sluaju imamo više atributa, na primjer osoba, lokacija, vrijeme, iznos novca i drugi, te više riječi može činiti entitet kojem se pridjeljuje jedan atribut. Jasnije je iz sljedećeg primjera:
Jim bought 300 shares of Acme Corp. in 2006.
$|Jim|_{person}$ bought 300 shares of $|AcmeCorp.|_{organization}$ in $|2006|_{time}$.
U našem slučaju imamo samo jedan atributa, lik, koji određuje tko su sve likovi u priči.

II. Podaci

Subsection text here.

III. Opis korištenih meotda

The conclusion goes here.

IV. Rezultati

hmm nevalja

Literatura

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.