***Pronalaženje sličnih presuda upotrebom metoda procesiranja prirodnog jezika***

Jelena Vlajkov

Fakultet tehničkih nauka

Univerzitet u Novom Sadu

[jelena.vlajkov@uns.ac.rs](mailto:jelena.vlajkov@uns.ac.rs)

Ilija Brdar

Fakultet tehničkih nauka

Univerzitet u Novom Sadu

[brdar.ilija@uns.ac.rs](mailto:brdar.ilija@uns.ac.rs)

Abstrakt –

....

1. UVOD

Anglosaksonsko pravo (eng. *common law*) ili precedentno pravo jeste pravni sistem koji preovlađuje u brojnim državama na engleskom govornom području, gde spadaju i SAD, Australija, UK. Pored zakona, ovaj pravni sistem zasnovan je i na presudama izrečenim u prošlosti (precedentima). Ukoliko je sličan spor rešen u prošlosti, sud je dužan da sledi obrazloženje iz prethodne presude. Ukoliko se proceni da je slučaj različit od prethodnih, sudije imaju ovlašćenje i dužnost da reše pitanje spora. [1]

Pretraga korpusa presuda u cilju nalaženja sličnih može predstavljati kompleksan i vremenski zahtevan postupak. Jedan model automatizacije postupka pretrage presuda je pretraživač (eng. *search engine*). Pretraživač kao ulaz prihvata korisnički upit (eng. *query*), a kao rezultat pretrage vraća presude koje odgovaraju postavljenom upitu. U slučaju primene pretraživača, korisnici moraju biti upoznati sa upitnim jezikom (eng. *query language*), što može predstavljati dodatno opterećenje. U cilju dostizanja veće efikasnoti pretrage, sam upit se može apstrahovati, tako da predstavlja ceo dokument. U ovom slučaju, pretraživač na ulazu dobija celu presudu, a kao rezultat vraća njoj slične presude iz celog korpusa presuda. Ovakav model pretraživača biće implementiran u ovom radu. Za implementaciju biće korišćene tehnike računarske inteligencije, konkretno one iz oblasti procesiranja prirodnog jezika (eng. *Natural Language Processing*, NLP).

Metode procesiranja prirodnog jezika imaju za cilj izvlačenje svojstava (eng. *features*) iz teksta koji je napisan u slobodnoj formi. Na ovaj način, svaki tekst se može predstaviti kao jedinstven *feature* vektor (eng. t*ext embedding*). Nakon što postoje *fature* vektori, možemo ih uporediti u cilju kvantifikacije njihove sličnosti. U ovom radu, sličnost se posmatra u kontekstu samog sadržaja presude. Kao tehnika za kvantifikaciju sličnosti, odabrano je kosinusno rastojanje.

Skup podataka na kojim se izvode eksperimenti preuzet je sa *Court Listener* veb sajta[[1]](#footnote-1) i čine ga prvostepene presude. Eksperimenti podrazumevaju primenu tehnika TF-IDF, word2vec, GloVe, TextRank i njihovo kombinovanje.

Evaluacija performansi u kontekstu problema koji ovaj rad rešava dosta je otežana s obzirom na nedostatak anotiranih podataka i variabilnosti značenja pojma *sličnost*. Zbog toga ne postoji klasičan vid evaluacije, već će ona izvodi poređenjem izlaza različitih NLP tehnika.

Drugo poglavlje ovog rada bavi se pregledom dosadašnje literature na temu sličnih problema. Treće poglavlje opisuje metodologiju izvođenja eksperimenata. Četvrto poglavlje izlaže dobijene rezultate, dok peto iznosi opšte zaključke ovog rada.

1. PRETHODNI RADOVI

...

1. METODOLOGIJA
2. REZULTATI
3. ZAKLJUČAK
4. LITERATURA

|  |  |  |
| --- | --- | --- |
| [1] |  | "Common law," 2 March 2022. [Online]. Available: https://en.wikipedia.org/wiki/Common\_law. |

1. https://www.courtlistener.com/api/bulk-info/ [↑](#footnote-ref-1)