

# Konvertovanje *plain-text* zakona u Akoma-Ntoso xml format

Tomislav Dobrički R1 18/2018

## **Opis problema:**

Rešava se problem konvertovanja pravnih dokumenata (zakona), pisanih na srpskom jeziku, iz „plain-text“ formata u mašinski čitljiv format. Glavni cilj projekta jeste da se napravi program koji konvertovanje vrši automatski. Rezultat konvertovanja treba da bude xml dokument koji je napisan po Akoma Ntoso standardu.

## **Opis relevantnosi problema:**

Pravni dokumenti koju su zapisani po Akoma Ntoso standardu imaju mnogo prednosti. Lakše se mogu pretražiti, referencirati, pronaći izmene itd. Konvertovanje dokumenata ručno bi zahtevalo mnogo vremena, zbog toga bi program koji to automatski radi učinio proces bržim a i lakšim za održavanje (odnosno kada se donose novi zakoni i menjaju stari). Dodato, verovatno postoji značajno manji broj istraživanja vezanih za pravne dokumente koji su pisani na srpskom jeziku, nego za neke ostale (u glavnom engleski).

## **Opis metoda rešavanja problema:**

Glavni problem prilikom konvertovanja dokumenata jeste pronalaženje delova koji su značajni (npr. reference na druge zakone, kazne, hijerarhija članova i tačaka itd.). U projektu će se pomenuti problem rešavati pomoću *text mining*-a. Za obučavajući skup će se koristiti podskup zakona koji se nalaze na Pravno Informacionom Sistemu Republike Srbije. Obučavajući skup će se ručno labelirati. Da bi podaci bili što raznovrsniji, potrebno je odabrati zakone iz različitih oblasti, ili dodatno, više različitih delova iz istih zakona. Za dobavljanje zakona će se izvršiti *scrape*-ovanje veb stranice Pravno Informacionog Sistema.