

Tacita

Orodje za pomoč pri anonimizaciji sodb

Institut Jožef Stefan

Laboratorij za umetno inteligenco

Center za prenos znanja na področju
informacijskih tehnologij



Anonimizacija sodb

Pred javno objavo je treba sodbe zaradi varovanja zasebnosti anonimizirati – odstraniti iz teksta vse omembe podatkov, preko katerih bi lahko identificirali določene osebe vpletene v postopku.

Razvoj sistema, ki bo s tehnikami strojnega učenja pomagal sodnikom pri anonimizaciji z avtomatsko detekcijo delov besedila za anonimizacijo.



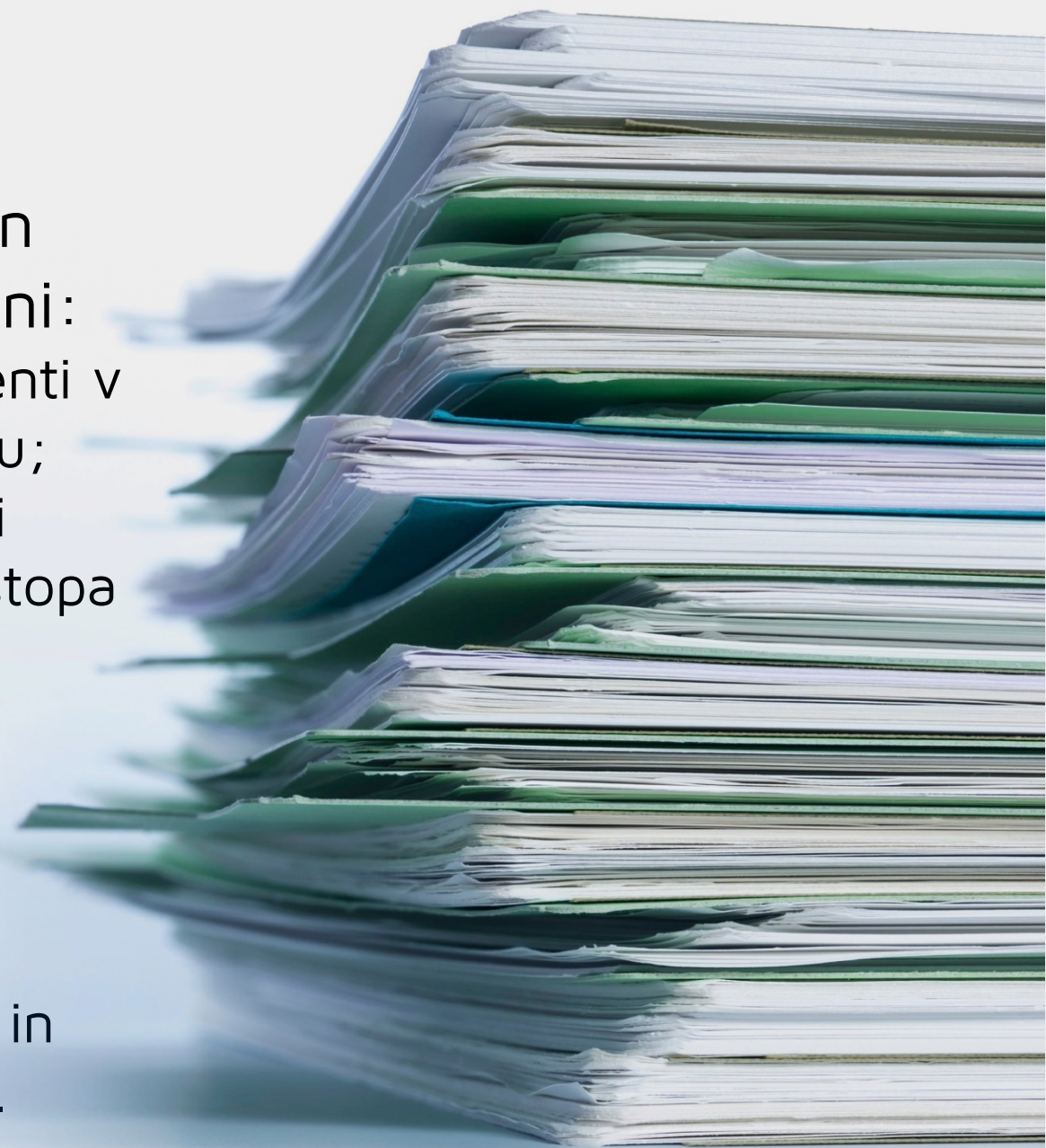
Podatki

Učna množica sestoji iz **2840 sodb** Vrhovnega in Višjega sodišča v Ljubljani:

- **neanonimizirani** dokumenti v open office (.odt) formatu;
- **anonimizirani** dokumenti dostopni preko HTTP dostopa v JSON formatu (<http://sodnapraksa.si>).

Predprocesiranje:

- ekstrakcija besedila;
- poravnava anonimizirane in neanonimizirane različice.



Pristop

Naučen statistični model, na podlagi katerega iz oblike in konteksta besede v besedilu napovemo ali jo je potrebno anonimizirati.

zaradi 12 kaznivih dejanj trgovine z ljudmi po drugem in prvem odstavku 175. člena v zvezi z 20. členom KZ-1 in kaznivega dejanja zlorabe prostitucije po tretjem in prvem odstavku 175. člena v zvezi z 20. členom KZ-1 ter zoper obdolženo [redacted] zaradi petih kaznivih dejanj trgovine z ljudmi po drugem in prvem odstavku 175. člena v zvezi z 20. členom KZ-1 in kaznivega dejanja zlorabe prostitucije po tretjem in prvem odstavku 175. člena v zvezi z 20. členom KZ-1 ter zoper obdolženo [redacted] zaradi dveh kaznivih dejanj trgovine z ljudmi po drugem in prvem odstavku 175. člena v zvezi z 20. členom KZ-1. Obdolženci so v priporu po 201. členu Zakona o kazenskem postopku (v

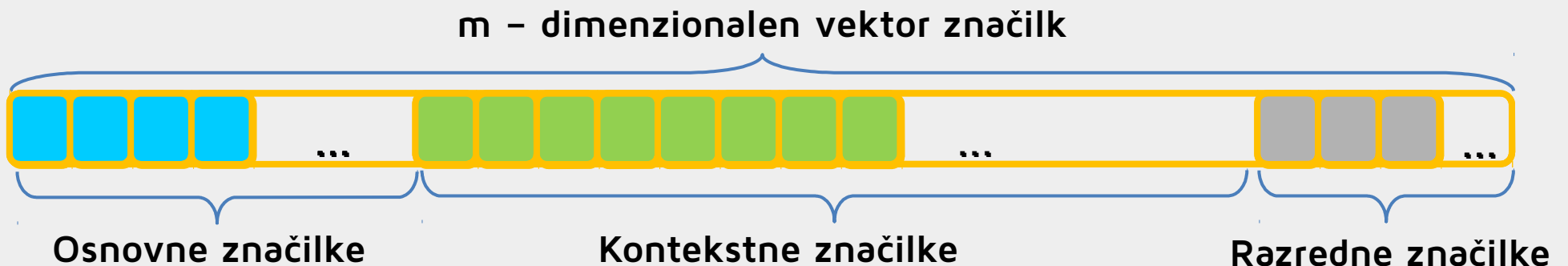
Pristop - kaj anonimiziramo?

»Obdolžena ~~Metka Irena NOVAK~~, hči ~~Janeza~~ in ~~Micke~~, rojene ~~Vidmar~~, rojeno ~~6.10.1941~~ v ~~Brežicah~~, EMŠO: ~~1234567891011~~, stanuje v ~~Murglah 123~~, ~~Ljubljana~~ ... Na podlagi tretjega odstavka 105. člena ZKP se oškodovanko ~~Micko Vidmar~~ ... vozilo ~~Opel Zafira 2.2 TDI~~, št . šasije ~~ABCDEFGH 1234567890~~ ... rušenje nelegalno postavljenega objekta na parcelno št. ~~12/3~~ k. o. ~~Ljubljana~~ ... njegovi sopotniki ~~Ehan Jasrov~~, ~~Maria Jasrova~~ in ~~Vasvije Ramad~~ utrpeli telesne poškodbe ... Zoper tožene stranke ~~Avto-moto zveza Slovenije~~, d. d. je bila ... Vrhovna državna tožilka Barbara Brezigar ...«

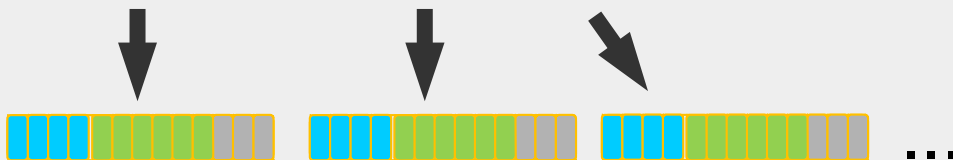
Pristop – predstavitev besed

Vsako **posamezno** besedo opišemo z množico značilk (značilka = *feature* (eng.)).

Množico značilk predstavimo z visoko-dimenzionalnim vektorjem.



»Obdolžena ~~Metka Irena NOVAK~~, hči Janeza in Micke ... «



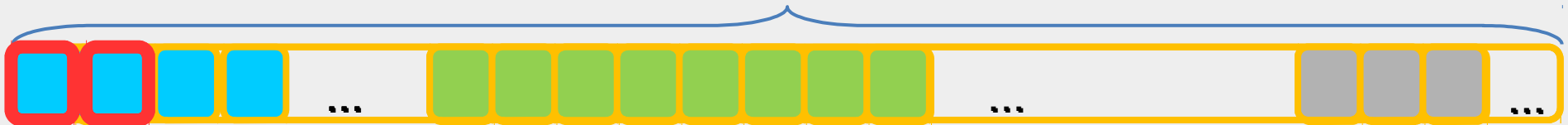
Pristop – izbor značilk

Značilka: **JE V SLOVARJU IMEN/PRIIMKOV?**

Značilka: **JE V SLOVARJU KRAJEV?**

Značilka: **OBLIKA BESEDE.**

m – dimenzionalen vektor značilk



Značilke – kontekst

Vreča besed:

Vzamemo okno besed (štiri pred in dve za).

Primer:

„... vrhovni tožilec in obdolženec Janez Goriški, rojen ...“ - anonimizirano

„... obdolženec, stanujoč na naslovu Poljanska ulica ...“ - anonimizirano

„... podjetje, na naslovu Poljanska ulica ...“ - neanonimizirano

N-grami:

Uni-grami za anonimizacijo:

„... obdolženi...“ „... EMŠO ...“

„... priča ...“ „... stranke ...“

„... priglašeni ...“ „... št. ...“

„... pokojni ...“ „... črpalke ...“

Bi-grami za anonimizacijo:

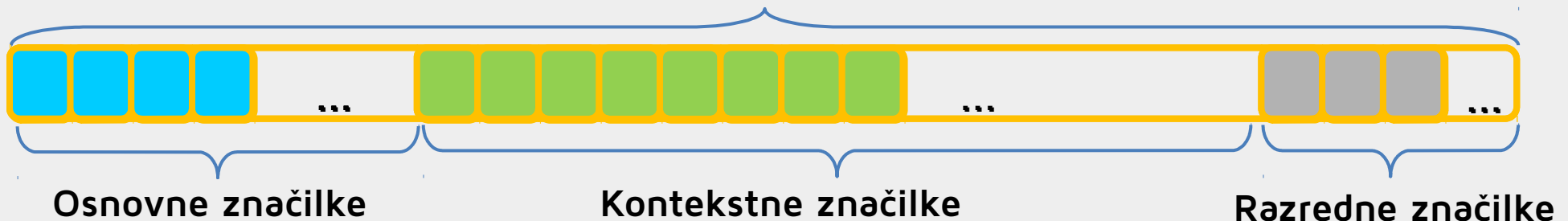
„ ... tožene stranke ...“

„ ... nova upraviteljica ...“

„ ... izvedenškega mnenja ...“

„ ... program univerze ...“

m – dimenzionalen vektor značilk



Značilke – razred prejšnjih

Standardna klasifikacija predpostavlja medsebojno **nepovezane** in **neodvisne** besede.

Rezultat klasifikacije predhodnjih besed je lahko dobra značilka za trenutno besedo:

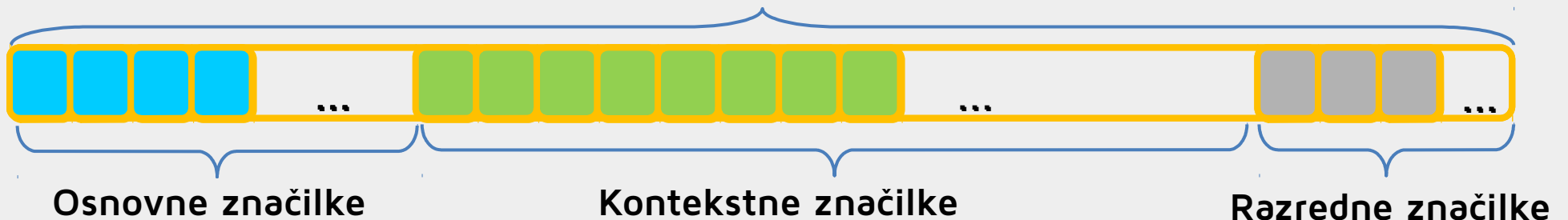
Primer:

„ ... hči **Janeza** in **Micke** ... ”

„ ... njegovi sopotniki **Ehan** **Jasrov**, **Haria** **Jasrova**... ”

Kot značilko vključimo še napovedani razred prejšnjih treh besed.

m – dimenzionalen vektor značilk



Pristop - besedni vektorji

```
In [1]: most_similar(u'obdolženec')
Out[1]:
[('obdolžen', 0.74),
 ('obdolži', 0.71),
 ('obdolženca', 0.71),
 ('obdolžila', 0.61),
 ('obdolžencu', 0.59),
 ('obtoženec', 0.59),
 ('mladoženec', 0.59),
 ('obsojenec', 0.56)]
```

```
In [1]: most_similar('janez')
Out[1]:
[('janezovak', 0.69),
 ('janezek', 0.63),
 ('janezov', 0.62),
 ('janezove', 0.62),
 ('janeza', 0.61),
 ('janezu', 0.61)]
```

```
In [1]: similarity('ljubljana', 'kranj')
Out[1]: 0.5901
```

```
In [2]: similarity('bmw', 'audi')
Out[2]: 0.4929
```

```
In [3]: similarity('dobro', 'slabo')
Out[3]: 0.6992
```

```
In [4]: similarity('obtoženec',
 'obtoženka')
Out[4]: 0.609
```

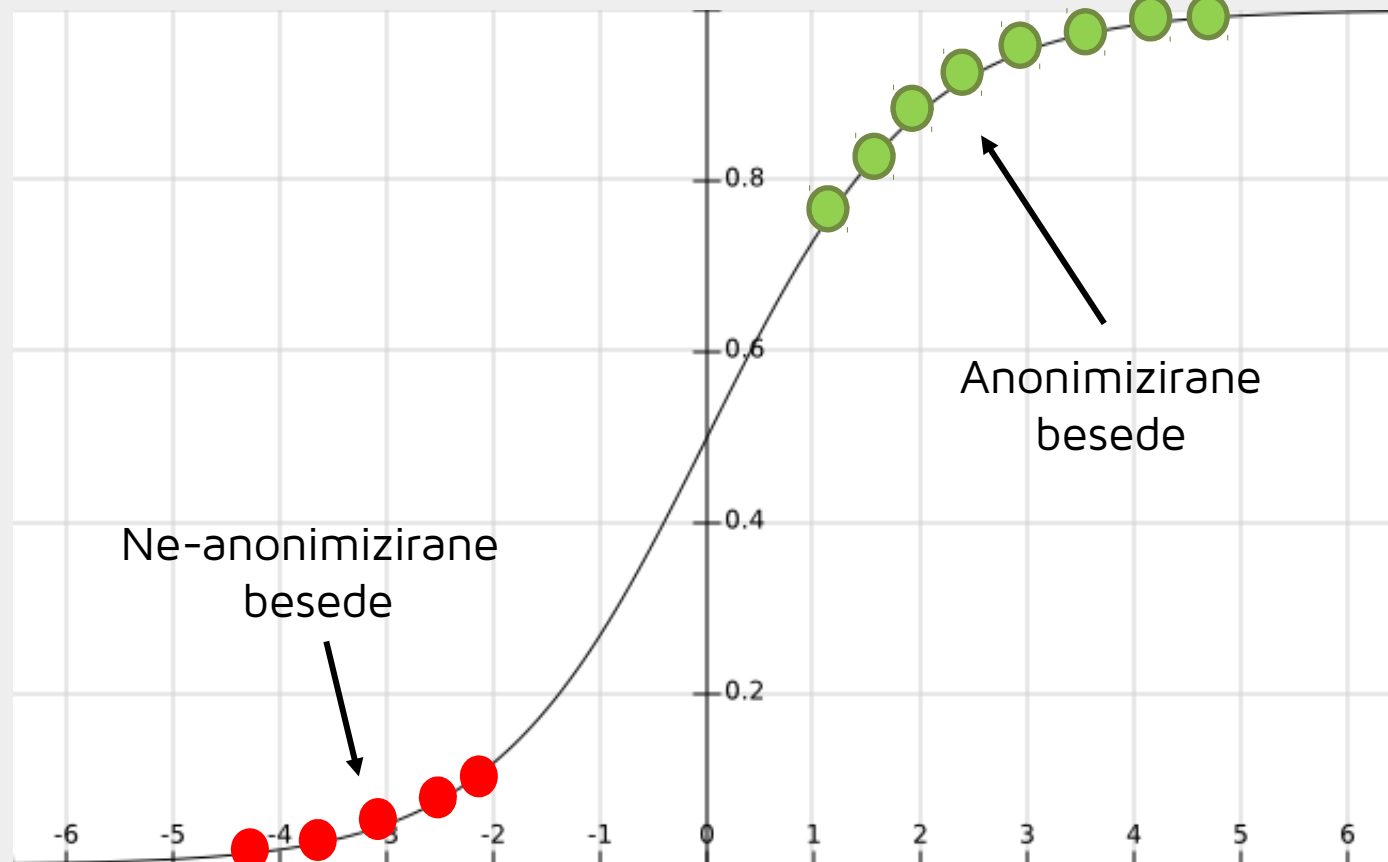
```
In [1]: 'evgeniju' in vocab
Out[1]: False
In [2]: most_similar('evgeniju')
[('evgenij', 0.86),
 ('evgeni', 0.81),
 ('evgenijo', 0.78),
 ('evgenija', 0.78),
 ('evgenijem', 0.76),
 ('evgenu', 0.73)]
```

Primer - anonimizacija

»Obdolžena ~~Metka Irena NOVAK~~, hči ~~Janeza~~ in ~~Micke~~, rojene ~~Vidmar~~, rojeno ~~6.10.1941~~ v ~~Brežicah~~, EMŠO: ~~1234567891011~~, stanuje v ~~Murglah 123~~, ~~Ljubljana~~ ... Na podlagi tretjega odstavka 105. člena ZKP se oškodovanko ~~Micke Vidmar~~ ... vozilo ~~Opel Zafira 2.2 TDI~~, št . šasije ~~ABCDEFGH 1234567890~~ ... rušenje nelegalno postavljenega objekta na parcelno št. ~~12/3~~ k. o. ~~Ljubljana~~ ... njegovi sopotniki ~~Ehan Jasrov~~, ~~Haria Jasrova~~ in ~~Vasvije Ramad~~ utrpeli telesne poškodbe ... Zoper tožene stranke ~~Avto-moto zveza Slovenije~~, d. d. je bila ... Vrhovna državna tožilka Barbara Brezigar ...«

Pristop - klasifikacija

Klasifikacija besed v dva razreda z uporabo metode **logistična regresija** (logistic regression).



Uspešnost sistema

5-kratno prečno preverjanje (*5-fold cross validation*).

Točnost: 72,8% od vseh primerov (besed), ki smo jih označili za anonimizacijo je pravih (pravilno pozitivni). Ostali so lažno pozitivni (jih ne bi smeli anonimizirati).

Priklic: 91,8% je delež resničnih pozitivnih primerov (besed) - se pravi smo (nepravilno) "pozabili" anonimizirati 8,2% primerov.

Primeri - uspešna anonimizacija

Vrhovno sodišče Republike Slovenije v senatu, ki so ga sestavljali vrhovni sodniki in sodnik predsednik ter [redacted] in [redacted] [redacted] sodnega [redacted] kot zapisnikarja, v kazenski zadevi [redacted] zaradi kaznivih dejanj po drugem in prvem odstavku 175. člena Kazenskega zakonika [redacted] Specializiranega državnega tožilstva RS št. [redacted] in [redacted]

Zoper obdolženega [redacted] se [redacted] obdolženo [redacted] in [redacted] pa iz pripornega razloga [redacted] kazenskem postopku podaljša pripor še za dva meseca, do vključno 27. 6. [redacted]

Pred Okrožnim sodiščem v Ljubljani teče zoper obdolžene [redacted] zaradi 12 kaznivih dejanj trgovine z ljudmi po drugem in prvem odstavku 113. člena v zvezi z 20. členom in kaznivega dejanja zlorabe prostitucije po tretjem in prvem odstavku 175. člena Kazenskega zakonika (v nadaljevanju KZ-1), zoper obdolženo [redacted] zaradi petih kaznivih dejanj trgovine z ljudmi po drugem in prvem odstavku 113. člena v zvezi z 20. členom KZ-1 in kaznivega dejanja zlorabe prostitucije po tretjem in prvem odstavku 175. člena KZ-1 ter zoper obdolženo [redacted] zaradi dveh kaznivih dejanj trgovine z ljudmi po drugem in prvem odstavku 113. člena v zvezi z 20. členom KZ-1. Obdolženci so v priporu [redacted] iz pripornih razlogov po 2. in 3. točki prvega odstavka 201. člena Zakona o kazenskem postopku (v nadaljevanju ZKP), obdolženi [redacted] in [redacted] pa iz pripornega razloga po 1. točki prvega odstavka 201. člena ZKP. Pripor zoper obdolžence teče od 27. 1. 2015 dalje. Obdolžencem je bil pripor p[red] Okrožnega sodišča I Ks 3724/2015 z dne 24. 2. 2015 do 27. 4. 2015, obdolženima [redacted] do 12.51 ure, obdolženi [redacted] pa do 13.05 ure. Okrožni državni tožilec j[redacted] odstavka 205. člena ZKP dne 9. 4. 2015 predlagal, da se obdolženim [redacted] podaljša pripor. V predlogu za podaljšanje pripora navaja, da obstaja zakonski utemeljen sum, da so obdolženci storili očitana jim kazniva dejanja. Navaja, da so še vedno podani priporni razlogi ter navaja okoliščine, na podlagi katerih ocenjuje, da sta za obdolženega [redacted] še vedno podana pripora razloga po 2. in 3. točki prvega odstavka 201. člena ZKP, za obdolženi [redacted] in [redacted] pa priporni razlog po 1. točki prvega odstavka 201. člena ZKP. Obrazlaga sorazmernost pripora in njegovo neogibnost za varnost ljudi ter potek kazenskega postopka. Kot razlog za podaljšanje pripora navaja, da preiskave iz objektivnih razlogov ni

Včasih zaradi napake ne upoštevamo dela besedila. Napaka se tipično pojavi, kadar je bilo besedilo med opombami pod črto, ali ko gre za tipkarsko napako.

Sivo besedilo v obstoječem sistemu ni objavljeno.

Zelena barva označuje pravilno anonimizirano besedilo.

Primeri - uspešno lovljenje napak

ponarejanja denarja po drugem in prvem odstavku 243. člena KZ-1, zoper obdolženega [redacted] za kaznivo dejanje ponarejanja denarja po drugem odstavku 243. člena KZ-1 in obdolženega [redacted] za kaznivo dejanje ponarejanja denarja po prvem odstavku 243. člena v zvezi z drugim odstavkom 20. člena KZ-1. Preiskovalna sodnica je v obširnih razlogih sklepa o uvedbi preiskave (strani 6-39) utemeljila utemeljenost suma obdolžencem očitanih kaznivih dejanj. Zoper obdolženega [redacted] pa je sklep o preiskavi z dne 15. 4. 2015 zaradi kaznivih dejanja ponarejanja denarja po drugem odstavku 243. člena KZ-1 in neupravičene proizvodnje in prometa s prepovedanimi denarnimi sredstvi v zvezi z 20. členom KZ-1 in neupravičene proizvodnje in prometa s prepovedanimi denarnimi sredstvi v zvezi z 186. člena KZ-1 pravno močan. Tudi po presoji Vrhovnega sodišča se utemeljijo sklepi o preiskavi v sklepih o preiskavi očitanih kaznivih dejanj med preiskavo z do sedaj izvedeno preiskavo. [redacted] ni pritrđiti navedbam zagovornika obdolženega [redacted] v odgovor na predlog državne tožilke, da se vloga obdolženca pri kaznivem dejanju zazna šele decembra 2014 in ne osem mesecev prej kot se mu to očita. Vrhovno sodišče je sledilo predlogu državne tožilke in pripor zoper obdolžene **Roda** [redacted] podaljšalo iz pripornega razloga po 3. točki prvega odstavka 201. člena ZKP. Po presoji Vrhovnega sodišča se okoliščine od zadnjega sklepa Okrožnega sodišča v Ljubljani z dne 30. 3. 2015) niso v ničemer spremenile. [redacted] (tiskanje [redacted] (dejanje ponarejanja denarja) ter okoliščine izvršitve teh kaznivih dejanj (tiskanje [redacted] bankovcev, razpečevanje in spravlanje v obtok tako ponarejenega denarja v [redacted] na dobro organiziranost, ozek in zaprt način delovanja udeležencev pri teh kaznivih dejanjih, pri čemer je imel vsak od obdolžencev točno določeno vlogo, vsi pa so bili motivirani s pridobitvijo protipravne premoženjske koristi. Obdolženi [redacted] je že bil obravnavan za premoženjska kazniva dejanja (sodba Okrožnega sodišča v Krškem z dne 19. 2. 2013 in z dne 9. 7. 2013, ko mu je zaradi kaznivih dejanj tatvine oziroma velike tatvine bil izrečen vzgojni ukrep). [redacted] je brez zaposlitve in se po lastni izjavi

Sistem je pravilno zaznal in anonimiziral tipkarsko napako.

Namesto Roka »Novaka«
je bilo napisano
Roda »Novaka«.

Delovanje sistema

Procesiranje ene sodbe je **hitro** (1s).

Ko so na voljo novi podatki, je sistem mogoče **ponovno učiti** in izboljšati njegovo delovanje. Proces ponovnega učenja traja nekaj ur.

Konsistentna raba sistema lahko pripomore k **poenoteni praksi anonimiziranja**, kar še dodatno izboljšuje delovanje sistema.

Uporabniški vmesnik

ANONIMIZACIJA

Shrani

Shrani in končaj

Prekliči

Predogled

Občutljivost (70%):

Osveži

Uvod

Vrhovno sodišče Republike Slovenije je v senatu, ki so ga sestavljali vrhovni sodnici in sodniki Janez Vlaj kot predsednik ter Karmen Iglič Stroligo, Vladimir Horvat, dr. Mateja Končina Peternel in mag. Rudi Štravs kot člani,

v pravdni zadevi tožeče stranke [redacted], ki jo zastopa [redacted] odvetnik v Ljubljani, zoper toženo stranko: 1. [redacted] in 2. [redacted] oba [redacted] ki ju zastopa [redacted] odvetnik v Domžalah,

zaradi plačila,

o reviziji tožene stranke zoper sodbo Višjega sodišča v Ljubljani I Cp 995/2014 z dne 15. 10. 2014, v zvezi s sodbo Okrožnega sodišča v Ljubljani IX Pg 4188/2012 z dne 3. 7. 2013,

na seji 1. julija 2015

Sklep

I. Revizija se zavrne.

II. Tožena stranka mora v 15 dneh, od vročitve te sodbe, povrniti tožeči stranki njene revizijske stroške v znesku 12.403,00 EUR z zakonskimi zamudnimi obrestmi, ki tečejo po izteku roka za izpolnitev obveznosti, določenega v tej točki izreka, do plačila.

Zaporedno

Po korenih

Besede po korenih

Republike

Republike

Republike

d

d

d

Ljubljana

Ljubljana

Ljubljani

Ljubljani

Ljubljani

odvetnik

odvetnik

odvetnik

[redacted]

[redacted]

[redacted]



Vprašanja . . .

Aljaž Košmerlj

aljaz.kosmerlj@ijs.si

Matej Kovačič

matej.kovacic@ijs.si

Patrik Zajec

patrik.zajec@ijs.si