



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

IEVA MAKACKAITĖ

**TEISINIŲ DOKUMENTŲ AUTOMATINIO SEMANTINIO
ANOTAVIMO GALIMYBIŲ TYRIMAS**

Baigiamasis magistro projektas

Darbo vadovas

doc. dr. R. Butkienė

KAUNAS, 2018

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

**TEISINIŲ DOKUMENTŲ AUTOMATINIO SEMANTINIO
ANOTAVIMO GALIMYBIŲ TYRIMAS**

Baigiamasis magistro projektas
Informacinių sistemų inžinerijos studijų programa (kodas 621E15001)

Vadovas

doc. dr. R. Butkienė
2018-05-18

Recenzentas

doc. dr. Lina Čeponienė
2018-05-18

Projektą atliko

Ieva Makackaitė
2018-05-18

KAUNAS, 2018



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

(Fakultetas)

Ieva Makackaitė

(Studento vardas, pavardė)

Informacinių sistemų inžinerijos studijų programa, 621E15001

(Studijų programos pavadinimas, kodas)

„Teisinių dokumentų automatinio semantinio anotavimo galimybių tyrimas“
AKADEMINIO SAŽININGUMO DEKLARACIJA

20 18 m. gegužės 18 d.
Kaunas

Patvirtinu, kad mano, **Ievos Makackaitė**, baigiamasis projektas tema „Teisinių dokumentų automatinio semantinio anotavimo galimybių tyrimas“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

Makackaitė, Ieva. Teisinių dokumentų automatinio semantinio anotavimo galimybių tyrimas. Magistro baigiamasis projektas / vadovas doc. dr. Rita Butkienė; Kauno technologijos universitetas, Informatikos fakultetas.

Mokslo kryptis ir sritis: Informatikos inžinerija, technologijos mokslai

Reikšminiai žodžiai: *semantika, anotavimas, teisė, kodeksai, ontologija, preciziškumas*

Kaunas, 2018. 98 p.

SANTRAUKA

Lietuvos Respublikos teisinių dokumentų sistema – struktūrizuota, tačiau labai plati. Analizuojant teisinius dokumentus ir norint atpažinti pagrindines teisės normos sąvokas straipsniuose, susiduriama su rankiniu anotavimu, kuris reikalauja didelės investicijos laiko atžvilgiu. Po atliktos semantinio anotavimo esamų sprendimų analizės ir teisinių dokumentų analizės, realizuotas naujas sprendimas – teisinių dokumentų automatinio semantinio anotavimo sistema. Sukurto sprendimo tikslas – optimizuoti ir palengvinti rankinį teisinių dokumentų anotavimą. Programa anotuoja kodeksus straipsnio lygyje, spalvomis išskirdama tekste pagrindines teisės normos sąvokas, taip palengvindama sąvokų atpažinimą naudotojui. Sprendime įgyvendinta sąsaja su *semantika.lt*. Testavimo ir eksperimento rezultatai bei apskaičiuoti jautrumo ir preciziškumo kriterijai parodo, kad sukurta sistema ganėtinai korektiškai atpažįsta teisės normos sąvokas straipsnyje. Ateityje tobulinant sistemą siūloma praplėsti anotacijų taisykles.

Makackaite, Ieva. Feasibility Study on Automatic Semantic Annotation of Legal Documents: Master's thesis in Information Systems Engineering / supervisor assoc. prof. Rita Butkiene. The Faculty of Informatics, Kaunas University of Technology.

Research area and field: Informatics Engineering, Technology Science

Key words: *semantic, annotation, legal, documents, precision*

Kaunas, 2018. 98 p.

SUMMARY

The system of legal documents of the Republic of Lithuania is structured though very broad. When analyzing legal documents and in order to identify the basic terms of the law in the articles, it encounters manual annotation, which requires a large investment in time. After a semantic automatic annotation analysis of existing solutions and analysis of legal documents, a new solution was implemented - an automatic semantic annotation system of legal documents. The goal of the new solution is to optimize and facilitate the work of people in the field of law. The program annotates the codes in the article, highlighting the basic concepts of law, thus facilitating the recognition of the terms for the user. The solution implements the interface with *semantika.lt*. The results of the test, the experiment and the criteria of sensitivity and precision are calculated, which shows that the system recognizes the terms of the norm in the article correctly. For the future system improvements, it is advised to extend the rules of annotations.

TURINYS

Turinys	5
Lentelių sąrašas	7
Paveikslų sąrašas	8
Terminų ir santrumpų žodynas	9
Įvadas	10
1. Semantinio anotavimo proceso ir naudojamų įrankių analizė	12
1.1 Analizės tikslas	12
1.2 Tyrimo objektas, sritis ir problema	12
1.3 Lietuvos Respublikos teisinių dokumentų ir automatinio semantinio anotavimo tyrimo objektų analizė	13
1.4 Semantinis anotavimas	15
1.5 Tyrimo objekto naudotojų analizė	20
1.6 Semantinio anotavimo įrankinių analizė (Lietuvos ir tarptautiniu mastu)	21
1.7 Darbo tikslas, uždaviniai, planas ir siekiami privalumai	26
1.8 Siekiamo sprendimo apibrėžimas	26
1.9 Analizės išvados	26
3. Teisinių dokumentų automatinio semantinio anotavimo sistemos reikalavimų specifikacija IR projektas	28
2.1 Reikalavimų specifikavimas	28
2.2 Teisinių dokumentų struktūros modelis	37
2.3 Teisinių dokumentų automatinio semantinio anotavimo sistemos naudotojo sąsajos modelis	40
2.4 Lietuvos Respublikos teisinių dokumentų automatinio semantinio anotavimo formalus sprendimo aprašas	41
2.5 Sprendimo reikalavimų ir projekto apibendrinimas	46
3. Lietuvos respublikos teisinių dokumentų automatinio semantinio anotavimo sprendimo realizacijos projektas	48
3.1 Teisinių dokumentų automatinio semantinio anotavimo sistemos loginė architektūra	48
3.2 Duomenų bazės modelis	49
3.3 Teisinių dokumentų automatinio semantinio anotavimo programos diegimas	50
3.4 Realizacijos projekto apibendrinimas	50
4. Teisinių dokumentų automatinio semantinio anotavimo sistemos realizacija ir testavimas	51
4.1 Teisinių dokumentų automatinio semantinio anotavimo sistemos realizacija ir veikimo aprašas	55
4.2 Teisinių dokumentų automatinio semantinio anotavimo sistemos testavimas	55
4.3 Sprendimo įgyvendinimo ir testavimo apibendrinimas	57

5. Eksperimentinis Lietuvos respublikos teisinių dokumentų automatinio semantinio anotavimo sistemos tyrimas	58
5.1 Eksperimento planas	58
5.2 Eksperimento rezultatai	59
5.3 Sprendimo veikimo ir savybių analizė, kokybės kriterijų įvertinimas	65
5.4 Sprendimo taikymo rekomendacijos	74
6. Rezultatų apibendrinimas ir išvados	76
7. Literatūra	78
8. Priedai	80
8.1. Priedas. 2.4 skyriuje pateiktų taisyklių pavyzdžiai straipsniuose	80
8.2. Anotacijos JSON failas	86
8.3. Eksperimento duomenys	89

LENTELIŲ SĄRAŠAS

1.1 lentelė. Vartotojo tipo ir savybių lentelė	20
1.2 lentelė. Problemos, su kuriomis susiduria vartotojai.	21
1.3 lentelė. Esamų sprendimų palyginimas	25
2.1 lentelė. „PA 1 Įkelti teisinį dokumentą“	32
2.2 lentelė. „PA 2 Nurodyti teisinio dokumento meta informaciją“	33
2.3 lentelė. „PA 3 Struktūrizuoti dokumentą“	34
2.4 lentelė. „PA 4 Anotuoti straipsnį automatiškai“	35
2.5 lentelė. „PA 5 Išsaugoti suanotuoto straipsnio anotaciją“	36
3.1.lentelė. Metodų paskirtis	49
4.1 lentelė. Sukurto sprendimo pagrindinės funkcijos	52
4.2 lentelė. Taisyklių testavimas Administracinių nusižengimų kodekse	56
5.1 lentelė. Eksperimento duomenys	60
5.2 lentelė. Administracinių nusižengimų kodekso eksperimentino rezultatai.....	60
5.3 lentelė. Darbo kodekso eksperimento rezultatai	61
5.4 lentelė. Civilinio kodekso eksperimento rezultatai	62
5.5 lentelė. Administracinių nusižengimų kodekso kriterijų skaičiavimas.....	63
5.6 lentelė. Darbo kodekso kokybės kriterijų skaičiavimas.....	64
5.7 lentelė. Civilinio kodekso kokybės kriterijų skaičiavimas.....	64
5.8 lentelė. Administracinių nusižengimų kodekso ir Darbo kodekso kokybės kriterijų skaičiavimas.....	64
5.9 lentelė. Darbo kodekso ir Civilinio kodekso kokybės kriterijų skaičiavimas.....	65
5.10 lentelė. Administracinių nusižengimų, Darbo kodekso ir Civilinio kodekso kokybės kriterijų skaičiavimas	65
8.1. lentelė. Sankcijos skaičiavimai	89
8.2. lentelė. Subjekto skaičiavimai.....	90
8.3. lentelė. Hipotezės skaičiavimai.....	91
8.4. lentelė. Objekto skaičiavimai	91
8.5. lentelė. Dispozicijos skaičiavimai.....	92
8.6. lentelė. Subjekto skaičiavimai.....	93
8.7. lentelė. Objekto skaičiavimai	93
8.8. lentelė. Teisių skaičiavimai.....	94
8.9. lentelė. Pareigų skaičiavimai.....	95
8.10. lentelė. Hipotezių skaičiavimai	95
8.11. lentelė. Dispozicijų skaičiavimai	96
8.12. lentelė. Hipotezės skaičiavimai.....	96
8.13. lentelė. Dispozicijos skaičiavimai.....	97
8.14. lentelė. Pareigos skaičiavimai	97
8.15. lentelė. Objekto skaičiavimai	97
8.16. lentelė. Subjektų skaičiavimai.....	97
8.17. lentelė. Teisių skaičiavimai.....	98

PAVEIKSLŲ SĄRAŠAS

1.1 pav. Semantinio tinklo technologijų stekas [6]	20
1.2 pav. Teksto anotacijos pavyzdys naudojant Brat įrankį [18]	21
1.3 pav. Teksto anotacijos pavyzdys naudojant Brat įrankį [18]	22
1.4 pav. GoNTogle architektūra [19]	23
1.5 pav. Semantinio anotavimo pavyzdys naudojant GoNTogle [19]	24
2.1 pav. Pagrindiniai Lietuvos Respublikos kodeksų automatinio semantinio anotavimo sistemos reikalavimai	29
2.2 pav. Reikalavimai LR teisinių dokumentų automatinio semantinio anotavimo	30
2.3 pav. Reikalavimai LR teisinių dokumentų įkėlimui į sistemą	29
2.4 pav. Grafinės sąsajos reikalavimai	31
2.5 pav. Panaudojimo atvejų diagrama	32
2.6 pav. „PA 1. Įkelti LR kodeksą“ veiklos diagrama	33
2.7 pav. „PA 2. Nurodyti teisinio dokumento meta informaciją“ veiklos diagrama	34
2.8 pav. „PA 3. Struktūrizuoti dokumentą“ veiklos diagrama	35
2.9 pav. „PA 4. Anotuoti dokumentą automatiškai“ veiklos diagrama	36
2.10 pav. „PA 5. Atsisiųsti suanotuotą dokumentą“ veiklos diagrama	37
2.11 pav. Protégė įrankio langas	37
2.12 pav. Dokumento esybių-ryšių modelis	38
2.13 pav. Teisinio elemento klasifikacija	39
2.14 pav. Sankcijos klasifikacija	39
2.15 pav. Administracinės sankcijos klasifikacija	39
2.16 pav. Teisinio ryšio elemento klasifikacija	40
2.17 pav. Sistemos navigacijos planas	41
2.18 pav. Taisyklių ieškojimo principas	46
3.1 pav. Sistemos loginė architektūra	48
3.2 pav. Detalizuotas sistemos loginė architektūra	49
3.3 pav. Duomenų bazės modelis	49
3.4 pav. Teisinių dokumentų automatinio semantinio anotavimo programos diegimas	50
4.1 pav. LR teisinių dokumentų automatinio semantinio anotavimo sistemos pagrindinis langas	52
4.2 pav. JSON failo, document dalies struktūra	53
4.3 pav. JSON failo, annotations dalies struktūra	53
4.4 pav. JSON failo, version dalies struktūra	53
4.5 pav. JSON failo, annotation dalies struktūra	54
4.6 pav. JSON failo struktūra	54
4.7 pav. Administracinių nusižengimų kodekso 100 straipsnio anotacija	55
4.8 pav. Darbo kodekso 99 straipsnio anotacija	55
5.1 pav. Administracinių nusižengimų kodekso, 675 straipsnio anotacija	61
5.3 pav. Administracinių nusižengimų kodekso, 54 straipsnio anotacija	61
5.4 pav. Aibės jautrumo ir preciziškumo skaičiavimui [22]	62
5.5 pav. Anotacijų atpažinimas Administracinių nusižengimų kodekse	66
5.6 pav. Anotacijų atpažinimas Darbo kodekse	67
5.7 pav. Anotacijų atpažinimas Civiliniame kodekse	68
5.8 pav. Kokybės kriterijų palyginimas Administracinių nusižengimų kodekse	69
5.9 pav. Kokybės kriterijų palyginimas Darbo kodekse	70
5.10 pav. Kokybės kriterijų palyginimas Civiliniame kodekse	71
5.11 pav. Bendras Administracinių nusižengimų ir Darbo kodesų kokybės kriterijų palyginimas	72
5.12 pav. Bendras Civilinio ir Darbo kodesų kokybės kriterijų palyginimas	72
5.13 pav. Bendras Administracinių nusižengimų, Civilinio ir Darbo kodesų kokybės kriterijų palyginimas	73

TERMINŲ IR SANTRUMPŲ ŽODYNAS

1. OWL (angl. *Web Ontology Language*) - semantinio tinklo ontologijų užrašymo kalba.
2. RDF (angl. *Resource Description Framework*) – išteklių aprašymo frameworkas.
3. SPARQL (angl. *Simple Protocol and RDF Query Language*) - RDF užklausų kalba.
4. XML (angl. *eXtensible Markup Language*) - Duomenų apsikeitimo formatas.
5. API (angl. *application programming interface*) – Aplikacijų programinė sąsaja.
6. JSON – atviro standarto formatas, perduodantis duomenų objektus.
7. Dispozicija – pagrindinė ir privaloma teisės normos dalis, kurioje įtvirtinta elgesio taisyklė, subjektų teisės ir pareigos.
8. Sankcija – teisės normos dalis, nurodanti poveikio priemones, taikomas, kai nesilaikoma teisės normų reikalavimų.
9. Hipotezė – teisės normos dalis, nurodanti faktines aplinkybes, kurioms atsiradus galioja teisės norma.
10. Teisės norma – kompetentingų asmenų arba jų įgaliotų institucijų priimta, sankcionuota, formaliai apibrėžta privaloma taisyklė, kurios vykdymas garantuojamas santykio dalyvių poveikio mechanizmais.

IVADAS

Darbas priklauso informacinių sistemų inžinerijos magistrantūros studijų programai.

Lietuvos valstybė yra nepriklausoma demokratinė respublika. Lietuvos piliečių elgesio normos yra griežtai apibrėžtos teisiniuose dokumentuose, kuriuos sudaro 11 kodeksų visumoje apimančių apie 4000 straipsnių. Kiekvienas straipsnis susideda iš bendrai nustatytų teisės normos dalių, tačiau bandant jas atpažinti, susiduriama su žodinėmis dviprasmybėmis, kas sąvokų atpažinimą ir straipsnio esmės suvokimą daro sudėtingesniu.

Semantinė anotacija yra naudojama siekiant „panaikinti“ žodines kalbos dviprasmybes. Šiame darbe siekiama automatiškai semantiškai anotuoti LR teisinius dokumentus.

Darbo problematika ir aktualumas

Lietuvos Respublikos teisinių dokumentų sistema yra labai plati ir sudėtinga. Kiekvienas kodekso straipsnis susideda iš teisės normos dalių, kurias sudėtinga atpažinti nestruktūrizuotame tekste. Dažnai gilinantis į straipsnį tenka jį perskaityti keletą kartų, kol teksto kontekstas tampa suprantamas. Dėl šios priežasties analizuojant teisinius dokumentus ir ieškant juose pagrindinių teisės normos dalių susiduriama su rankiniu anotavimu, kuris reikalauja didelės investicijos laiko atžvilgiu. Suanotuotas tekstas reikalingas semantinei paieškai, ko pasekoje dėl nevykstančio teisinių dokumentų automatinio semantinio anotavimo, teisiniai dokumentai anotuojami rankiniu būdu.

Problema aktuali asmenims, kurie gyvenimiškose situacijose susiduria su teisiniais dokumentais ir asmenims, kurių darbo sritis speciliizuota į teisinius dokumentus. Problema pasižymi tęstinumu, kadangi automatinis semantinis anotavimas Lietuvos Respublikos teisiniams dokumentams nevyksta.

Darbo tikslas ir uždaviniai

Darbo tikslas – sudaryti galimybę automatiškai semantiškai anotuoti Lietuvos Respublikos teisinius dokumentus.

Tiksliui įgyvendinti išsikelti šie uždaviniai:

1. išanalizuoti LR teisinius dokumentus (struktūrą, sąsajas tarp žodžių);
2. sukurti ontologijos srities schemą;
3. išanalizuoti esamas semantinio anotavimo priemones (metodai, įrankiai, standartai ir formatai);
4. sudaryti automatinio semantinio anotavimo algoritmą;
5. eksperimentiškai ištirti sudarytą algoritmą ir pateikti išvadas.

Darbo rezultatai ir jų svarba

Atlikta esamų sprendimų analizė ir sukurta teisinių dokumentų automatinio semantinio anotavimo sistema, kuri atpažįsta ir sužymi teisės normos sąvokas straipsniuose. Lietuvoje teisinių dokumentų automatinis semantinis anotavimas nebuvo įgyvendintas anksčiau, todėl šis sukurta

sprendimas sumažina laiko sąnaudas atliekant anotavimą rankiniu būdu ir suanotuoja tekstą pagal teisės normos sąvokas.

Darbo struktūra

Darbą sudaro šios pagrindinės dalys:

- lyginamoji semantinio anotavimo įrankių analizė;
- ontologijų ir atvaizdavimo kalbų analizė;
- teisinių dokumentų struktūros analizė;
- kuriamo sprendimo reikalavimų specifikavimas bei projektinė dalis;
- sprendimo realizacija;
- eksperimentas ir rekomendacijos;
- galutinės išvados bei pastebėjimai.

1. SEMANTINIO ANOTAVIMO PROCESO IR NAUDOJAMŲ ĮRANKIŲ ANALIZĖ

1.1 Analizės tikslas

Analizės tikslas – išanalizuoti teisinių dokumentų struktūrą, semantinio anotavimo procesą bei įrankius. Analizė reikalinga siekiant rasti geriausią problemos sprendimą, kurį įgyvendinus būtų atliekamas teisinių dokumentų automatinis semantinis anotavimas.

1.2 Tyrimo objektas, sritis ir problema

Tyrimo problema

Lietuvos Respublikos teisinių dokumentų sistema yra labai plati ir sudėtinga. Kiekvienas kodekso straipsnis susideda iš teisės normos sąvokų, kurias sunku atpažinti nestruktūrizuotame tekste. Dažnai gilinantis į straipsnį tenka jį perskaityti keletą kartų, kol teksto kontekstas tampa suprantamas. Dėl šios priežasties analizuojant teisinius dokumentus ir ieškant juose pagrindinių teisės normos dalių susiduriama su rankiniu anotavimu, kuris reikalauja didelės investicijos laiko atžvilgiu. Suanotuotas tekstas reikalingas semantinei paieškai, ko pasekoje dėl nevykstančio teisinių dokumentų automatinio semantinio anotavimo, teisiniai dokumentai anotuojami rankiniu būdu.

Problema aktuali asmenims, kurie gyvenimiškose situacijose susiduria su teisiniais dokumentais ir asmenims, kurių darbo sritis specilizuota į teisinius dokumentus. Problema pasižymi tęstinumu, kadangi automatinis semantinis anotavimas Lietuvos Respublikos teisiniams dokumentams nevyksta.

Tyrimo objektas

Lietuvos Respublikos teisinių dokumentų sistema, kurią sudaro kodeksai. Kiekvienam Lietuvos piliečiui aktualiausi kodeksai yra tie, su kuriais susiduria kasdieniškose situacijose, kaip pavyzdžiui: dirbančiam žmogui turėtų būti aktualu už kokio pobūdžio darbo laiko apskaitos pažeidimus galioja sankcijos. Administracinių nusižengimų 100 straipsnis atitinka prieš tai minėtą situaciją. Straipsnis nėra suanotuotas tai reiškia, kad jame nepažymėtos pagrindinės teisės normos sąvokos, todėl norint sužymėti pagrindinius „akcentus“, tai reikia atlikti rankiu būdu.

Pateikiamas nesuanotuotas Administracinių nusižengimų 100 – asis straipsnis.

100 straipsnis. Darbo laiko apskaitos pažeidimas

1. Darbuotojų, išskyrus dirbančius nekintančių darbo dienos (pamainos) trukmės ir darbo dienų per savaitę skaičiaus darbo laiko režimu, darbo laiko nežymėjimas darbo laiko apskaitos žiniaraštyje, taip pat į darbo laiko apskaitą privalomai įtrauktino darbuotojų darbo laiko (viršvalandžių, darbo laiko švenčių dieną, poilsio dieną (jeigu jis nenustatytas pagal grafiką), naktį, dėl papildomo darbo) nežymėjimas darbo laiko apskaitos žiniaraštyje arba žinomai neteisingų duomenų apie įmonėse, įstaigose, organizacijose dirbančių asmenų į darbo laiko apskaitą privalomai įtrauktiną darbo laiką įrašymas į darbo laiko apskaitos žiniaraštį užtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo vieno šimto penkiasdešimt iki vieno tūkstančio keturių šimtų penkiasdešimt eurų.

2. Šio straipsnio 1 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai, užtraukia baudą nuo vieno tūkstančio keturių šimtų iki trijų tūkstančių eurų [2].

Tyrimo metu kodeksai bus analizuojami straipsnio lygyje, juose ieškant pagrindinių teisės normos dalių.

Tyrimo sritis

Tyrimo sritis – metodai, įrankiai, standartai ir formatai, skirti lietuviško teksto semantinio anotavimo automatizavimui.

1.3 Lietuvos Respublikos teisinių dokumentų ir automatinio semantinio anotavimo tyrimo objektų analizė

Lietuvos Respublikos teisinių dokumento analizė

Teisiniai dokumentai turi griežtą struktūrą, teisės normas, teisės ryšius, objektus ir subjektus.

Teisės norma – kompetentingų asmenų (fizinių arba juridinių), organizacijų (pvz., valstybės, bažnyčios, tarptautinės organizacijos) arba jų įgaliotų institucijų (pvz., valstybėje – parlamento) priimta, sankcionuota, formaliai apibrėžta privaloma taisyklė, kurios vykdymas garantuojamas santykio dalyvių poveikio mechanizmais (išoriniais arba vidiniais). Teisės norma yra socialinė norma, todėl teisė veikia kartu su kitais visuomenės reguliatoriais kaip religija, moralė, dorovė, etika ir kitais.

Lietuvos Respublikos teisiniuose dokumentuose teisės normą sudarančios sąvokos:

- **hipotezė** – teisės normos dalis, nurodanti faktines aplinkybes, kurioms atsiradus galioja teisės norma. Ja nurodomos sąlygos (gyvenimiškos aplinkybės), kurioms esant galima įgyvendinti poelgio taisyklės – vykdyti, laikytis, taikyti šias taisykles. Pagal savo struktūrą hipotezės gali būti: santykinai apibrėžtos; visiškai apibrėžtos; absoliučiai apibrėžtos.
- **dispozicija** – pagrindinė ir privaloma teisės normos dalis, kurioje įtvirtinta elgesio taisyklė, subjektų teisės ir pareigos. Ja nurodoma pati elgesio taisyklė – veikimas ar neveikimas, kurią galioja atlikti teisės norma ir kuria privalo atlikti tie, kuriems norma yra adresuojama. Dispozicijų rūšys: visiškai ir santykinai apibrėžtosios, blanketinės.
- **sankcija** – teisės normos dalis, nurodanti poveikio priemonės, taikomas, kai nesilaikoma teisės normų reikalavimų. Atsižvelgiant į poveikio rūšį, sankcijos gali būti: baudžiamosios, administracinės, drausminės ir turtinės [3], [4].

Lietuvos Respublikos teisinių dokumentų sistema yra labai plati, kiekvieną teisinį kodeksą sudaro keli šimtai straipsnių. Lietuvos Respublikos kodeksai turi bendrą struktūrą, tačiau kiekvienas kodeksas turi savitą rašymo stilių. Dėl duomenų gausos šiame darbe plačiau nagrinėjami du kodeksai: Administracinių nusižengimų kodeksas ir Darbo kodeksas [23].

Kiekvieną teisinį dokumentą sudaro dalys, dalis susideda iš skyrių, skyriai iš straipsnių. Detalizuota teisinių dokumentų struktūra pateikiama antroje tiriamojo darbo dalyje, 2.3 skyriuje. Analizės metu pasirinkus atitinkamus kodeksus buvo išanalizuota jų straipsnių struktūra.

Analizuojant Administracinių nusižengimų kodekso straipsnius pastebėta, kad šio kodekso straipsniuose galima įžvelgti dėsningą teisės normų sąvokų išdėstymą. Žemiau esančiame pavyzdyje pateikiama vieno pasirinkto straipsnio struktūra ir anotacijos, kurios buvo atliktos rankiniu būdu.

Pasirinktas Administracinių nusižengimų kodekso Specialiosios dalies 12 skyriaus 99 straipsnis, kuris suanotuotas rankiniu būdu.

99 straipsnis. Darbo užmokesčio apskaičiavimo ir mokėjimo tvarkos pažeidimas

1. Lietuvos Respublikos darbo kodekse, kolektyvinėje arba darbo sutartyje nustatytos darbo užmokesčio apskaičiavimo ir mokėjimo tvarkos pažeidimas

užtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo vieno šimto penkiasdešimt iki vieno tūkstančio keturių šimtų penkiasdešimt eurų.

2. Šio straipsnio 1 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai,

užtraukia baudą nuo vieno tūkstančio keturių šimtų iki trijų tūkstančių eurų.

3. Tyčinis Lietuvos Respublikos darbo kodekse, kolektyvinėje arba darbo sutartyje nustatytos darbo užmokesčio apskaičiavimo ir mokėjimo tvarkos pažeidimas arba darbo užmokesčio ir kitų su darbo santykiais susijusių išmokų, neįtrauktų į buhalterinės apskaitos dokumentus, išmokėjimas

užtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo dviejų tūkstančių septynių šimtų iki šešių tūkstančių eurų [2].

Kiekvienas straipsnio punktas, turi tas pačias anotacijas. Žemiau detalizuojamas pirmas straipsnio punktas:

sankcijos hipotezė - Lietuvos Respublikos darbo kodekse, kolektyvinėje arba darbo sutartyje nustatytos darbo užmokesčio apskaičiavimo ir mokėjimo tvarkos pažeidimas;

teisės ryšys – užtraukia;

objektas – baudą;

dispozicija - užtraukia baudą;

subjektas - darbdaviams ar kitiems atsakingiems asmenims;

sankcija - nuo vieno šimto penkiasdešimt iki vieno tūkstančio keturių šimtų penkiasdešimt eurų.

Iš atliktos LR kodekso ir straipsnių analizės galima daryti išvadas, kad kai kurios anotacijos yra statinės, griežtai apibrėžtos ir atsikartoja didžiojoje dalyje straipsnių, pavyzdžiui: vieną iš dispozicijų Administracinių nusižengimų kodekse galima apibrėžti kaip dviejų žodžių junginį „užtraukia baudą“,

kuri tame pačiame straipsnyje atsikartoja tris kartus, o visame Administracinių nusižengimų kodekse 1504 kartus.

Atlikus Lietuvos Respublikos kodeksų analizę galima teigti, kad dokumentai turi griežtas normas ir struktūrą, teisės normos dalis bei ryšius, tačiau lyginant kelis kodeksus tarpusavyje pastebima, kad kodeksai išsiskiria savo rašymo stiliumi bei teisės normos dalių dėsningumu [2]. Dėl šių priežasčių automatiniam semantiniam anotavimui taisyklės bus kuriamos teisinėms normoms rasti, išnagrinėjus kiekvieną kodeksą individualiai.

1.4 Semantinis anotavimas

Semantinio anotavimo tikslas – panaikinti teksto dviprasmybes. Semantinis anotavimas tekstą suskirsto į frazes arba žodžius. Norint atlikti semantinį anotavimą tekstui yra pritaikomos lietuvių kalbos gramatikos taisyklės. Semantinis tekstų anotavimas priskiriamas natūralios kalbos apdorojimų uždavinių grupei. Semantinio anotavimo metodo esmė – gauti semantiškai struktūrizuotą tekstą iš nestruktūrizuoto teksto.

Semantinis anotavimas su ontologija susijęs tuo, kad ontologija semantiniame anotavime atskiria, kokią ontologijos klasę nagrinėjamas žodis atitinka, su kokiais kitais žodžiais galima jį susieti ir su kokiais ne. Nors tiriamajame darbe ontologijų klasės tiesiogiai nenaudojamos, tačiau sudarytos ontologijų klasės skirtos išsigryninti pagrindines teisės normos dalis.

Semantinės anotacijos gali būti sukurtos naudojant tradicinius dokumento anotacijos įrankius arba naujesnius metodus tokius kaip semantikos „wikis“, semantikos „blogs“. Šiuo metu neegzistuoja vienas modelis, kuris būtų skirtas šioms skirtingoms anotacijų rūšims [5], [6].

Semantinį anotavimą sudarantys komponentai:

- leksinė analizė – teksto skaidymas į vienetus (fragmentus, sakinius, frazes, žodžius, skyrybos ženklus ir kt.);
- morfologinė analizė – kalbos dalių nustatymas išskaidytam tekstui;
- lemavimas – pagrindinių žodžių formų nustatymas žodžiams;
- esybių atpažinimas – žodžių ir jų junginių klasifikavimas į apibrėžtas kategorijas;
- sintaksinė analizė – sakinių skaidymas į frazes bei kalbos dalis, nusakant jų tarpusavio sąryšius.

Tiriamajame darbe apibrėžtai problemai spręsti, semantinis anotavimas skirtas sąvokų dviprasmybių panaikinimui t. y. teisės normos dalių atpažinimui tekste. Anotacijų taisyklėms aprašyti naudojama leksinė ir gramatinių kategorijų analizė, lemavimas bei sintaksinė analizė. Teisinių dokumentų semantinio anotavimo tikslas – kuo didesnis atpažintų esybių kiekis, kuriose skirtingi žodžiai bei junginiai suklasifikuojami į teisės normos dalis.

Kompiuterinė ontologija

Teisinių dokumentų semantiniame anotavime ontologija yra naudojama kaip terminų žodynas, apibrėžiantis pagrindines teisės normos dalis bei jų prasmę.

Tiriamajame darbe ontologija sukurama Protégé įrankio pagalba, kaip priemonė indentifikuoti teisės normos dalis ir jų sąryšius. Ontologija nebus aprašoma OWL kalba, vizualizacija pateikiama esybių-ryšių modeliu (žiūrėti 2.3 skyrių).

Filosofijoje pasak Aristotelio terminas „ontologija“ turi prasmę „būties teorija“, „visos tikrovės svarbiausių ypatybių nagrinėjimas“. Čia „ontologija“ yra sinonimas terminui „metafizika“ (Aristotelis IV a.pr.m.e.).

Kompiuterinė ontologija – tai tam tikros srities sąvokų visumos specifیکavimas išreikštu pavidalu [7].

Ontologijos specifіkuoja nagrinėjimo srities:

- konceptus, esybių (daiktų bei reiškinių) tipus;
- konceptų hierarchijas, esybių tipų tarpusavio sąryšius, priklausomybes;
- aksiomas, taisykles, dėsningumus apie esybių tipus ir sąryšius.

Sudarant ontologijų klases pagal Lietuvos Respublikos teisinių dokumentų struktūrą, išanalizuota LKIF ontologijų biblioteka, kuri apima teisinių dokumentų sritį. LKIF analizės tikslas – įvertinti ar šis sprendimas tinkamas sudaryti ontologijos modelį tiriamojo darbo dalykinei sričiai.

LKIF Branduolys (angl. LKIF Core)

LKIF tai ontologijų biblioteka, kuri apima teisinių dokumentų sritį.

Abstrakčios sąvokos

Abstrakčios sąvokos yra apibrėžiamos penkiuose artimai susijusiuose modeliuose:

- *viršūnė (angl. top)* – LKIF viršūnės ontologija yra pagrįsta LRI – branduolio aukščiausiuoju lygiu, bet turi mažesnę ontologinę prasmę. Skiria mažiau apribojimų aukščiausių kategorijų subklasėms viršūnės kategorijose.
- *vieta (angl. place)* – vietos modulis, kuris iš dalies įgyvendina reliatyvių vietų teoriją (Donnelly, 2005) OWL DL.
- *mereologija (angl. mereology)* – šis modulis apibrėžia dalių ryšius, komponentus, tokias sąvokas, kaip narystė ir kt.

Pagrindinės sąvokos

Bazinio lygio sąvokos yra paskirstytos per keturis modulius:

- *procesas (angl. process)* – proceso modulis plečia LKIF viršūnės ontologijos modulį su pakeitimų apibrėžimais, procesais ir fiziniais objektais. Tai įveda apribotą komplektą ypatybių tam, kad apibūdintų dalyvaujančius procesų vaidmenis;
- *vaidmuo (angl. role)* – vaidmens modulis apibrėžia vaidmenų tipologiją (episteminiai vaidmenys, funkcijos, asmens vaidmenys, organizacijos vaidmenys) ir žaidimų turtą (angl. plays-property) tam, kad susietų vaidmens pildytoją su vaidmeniu;
- *veiksmas (angl. action)* – veiksmo modulis apibūdina žodyną. Veiksmai yra procesai, kurie yra įvykdyti tam tikro atstovo (veiksmo aktorius);
- *posakis (angl. expression)* – posakio modulis apibūdina žodyną apibūdinimu, tvirtinimu ir logišku požiūriu (tikėjimas, ketinimas), kvalifikacija ir žiniasklaida. Tai, be to, plečia vaidmens modulį su skaičiumi ar episteminiais vaidmenimis, ir yra pagrindas normų apibrėžimui.

Teisinės sąvokos

Šitos pagrindinės grupės yra išplėtos trijų modulių, kurie formuoja teisėtą ontologiją:

- *teismo ieškinys (angl. legal – action)* – teismo ieškinio modulis plečia veiksmo modulį su daug teisėtų sąvokų, susietų su veiksmu ir atstovu, tokių kaip vieši aktai, vieši kūnai, juridinis asmuo, fizinis asmuo ir kt.
- *teisėtas vaidmuo (angl. legal – role)* – teisėtas vaidmens modulis plečia vaidmens modulį su mažu skaičiumi teisėtų sąvokų, susietų su vaidmenimis, teisininko profesija ir kt.
- *norma (angl. norm)* – pirmiausia normos modulis yra ištęsiamas posakio modulio, kur normos yra apibrėžiamos kaip kvalifikacijos. Be to šis modulis apibrėžia daug teisėtų šaltinių, pavyzdžiui teisėti dokumentai, įprastas įstatymas ir kt.

Struktūros moduliai

Šiose teisės grupėse yra du moduliai, kurie yra kaip pagrindinis dviejų struktūrų žodynas: pasikeitimas ir taisyklės.

- *Pasikeitimas (angl. modification)* – pasikeitimo modulis yra laiko modulio ir teisinio ieškinio modulių ištęsimas. Laiko modulis yra išplėstas su gausiais intervalais ir momentais, apibūdinamas efektyvumą. Veiksmo modulis yra išplėstas su pasikeitimų tipologija.
- *Taisyklės (angl. rules)* – taisyklių ir argumentacijos modulis apibrėžia centrinius vaidmenis argumentacijai ir apibūdina žodyną LKIF taisyklėms.

Pagrindinė ir išplėstinė ontologija

Apibendrinant, LKIF sudaro dvylika sujungtų abstraktaus, pagrindinių ir teisėto lygmens modulių ontologijos modulį.

Du struktūros moduliai yra laikomi kaip LKIF išplėstinis ontologijos modulis [8].

Remiantis LKIF buvo bandoma pritaikyti Lietuvos Respublikos teisiniams dokumentams. LKIF sprendimas pasirinktai dalykinei sričiai netiko, kadangi jis apibrėžia konkrečių teisinių bylų nagrinėjimą ir ontologija skirta teisės normos pažeidimų išvedimui. Žvelgiant į ateitį LKIF galėtų būti naudojamas sudarant teisinius dokumentus.

Naudojamų įrankių (*Protégé*, *TopBraid*) atvaizdavimo kalbų (OWL2) ir ontologijų saugyklų analizė

Naudojamų įrankių analizėje detalizuojami įrankiai, skirti ontologijoms kurti. Protégé įrankio pagalba sudaromos pagrindinės ontologijų klasės skirtos LR teisiniams dokumentams. Naujo sprendimo realizacijoje nesiekama kurti semantinio tinklo, apsiribojama teisinių dokumentų automatinio semantiniu anotavimu, todėl ontologija nebus kuriama taikomojoje programoje.

Protégé – tai nemokamas atviro kodo ontologijos kūrimo įrankis. Šis įrankis yra vienas iš populiariausių ir plačiausiai naudojamų ontologijos kūrimo įrankių. Įrankiu naudotis galima dvejomis galimybėmis: Protégé Desktop arba WebProtégé. Protégé leidžia kurti, valdyti, grafiškai atvaizduoti ontologijas bei jų hierarchinius ryšius. Grafinio įskiepio OWL Viz dėka, Protégé leidžia atvaizduoti ontologiją grafiškai, kas gerokai palengvina darbą ontologijų specialistui. Protégé leidžia eksportuoti ontologijas į įvairius formatus, tame tarpe ir populiariausius OWL, RDF, XML. Dar vienas Protégé privalumas – integruotas Jena karkasas. Tai Java kalba parašytas karkasas, turintis RDF/OWL paslaugų programavimo sąsają. Jena karkasas gali apdoroti SPARQL užklausas, kurios yra neatsiejamos nuo semantinio žiniatinklio.

WebProtégé pasižymi šiomis savybėmis:

- W3C standartų sukalbamumas;
- paprasta naudojimosi grafinė sąsaja;
- sekamos ir peržiūrėtos istorijos keitimas;
- tinkle suformuojamas specialus srities redagavimas;
- daugialypiai įkėlimo/parsiuntimo formatai;
- kryžminis susiejimas su Protégé Desktop [9].

***TopBraid* [10]**

TopBraid naudojamas siekiant išvystyti ontologijos modelius, formuoti duomenų šaltinio integraciją ir sukurti semantikos žiniatinklio paslaugas.

TopBraid pasižymi šiomis savybėmis:

1. ontologijų ir RDF duomenų redagavimu;
2. semantikos internetinių puslapių aplikacijų vystymu;
3. importuojami duomenys;

4. eksportuojami duomenys;
5. dirbama su „Data Back Ends“;
6. dirbama su XML ir lentelių failais;
7. RDF;
8. duomenų integracija.

OWL 2 Web ontologijos kalba

OWL (angl. Web Ontology Language) yra semantinio tinklo ontologijų užrašymo kalba, praplečianti RDF formatu pateikiamos struktūrizuotos informacijos išraiškingumą. OWL siekia detalai apibrėžti RDF žodynuose naudojamų sąvokų reikšmes ir jų tarpusavio sąryšį. Šis sąryšis bendrinio atveju vadinamas ontologija. OWL 2 suteikia klases, ypatybes, asmenis, duomenų reikšmes ir yra saugoma kaip semantikos tinklo dokumentai. Ši kalba gali būti panaudojama su informacija, kuri parašyta RDF pagal sudėtingumą, aksiomų kiekį ir ekspresyvumą, skiriami 3 pagrindiniai OWL 2 kalbos profiliai:

- OWL 2 EL – skirtas aprašyti ontologijoms, pasižyminčioms dideliu sąvybių ir klasių kiekiu;
- OWL 2 QL – skirtas aprašyti ontologijoms, pasižyminčioms dideliu klasių egzempliorių kiekiu;
- OWL 2 RL – skirtas aprašyti ontologijoms, kuomet reikalingas aukštesnis ekspresyvumo laipsnis.

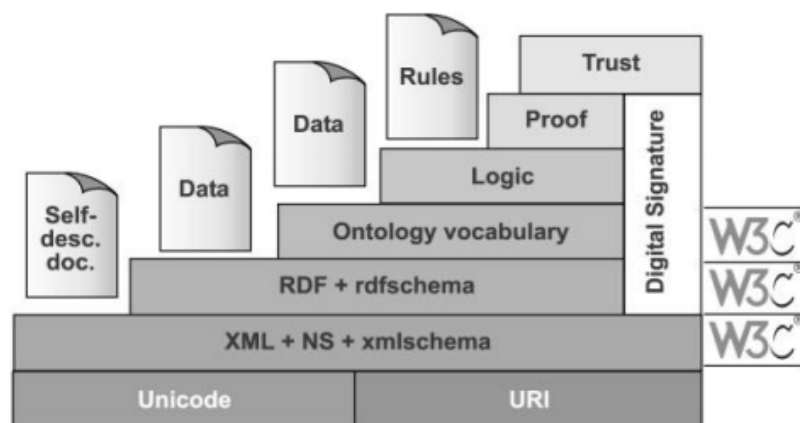
RDF – meta duomenų vaizdavimo būdas, kuris naudoja mažą modeliavimo konstrukcijų rinkinį ir neatitinka taikymo srities žodyno.

RDF schema išplečia modeliavimo konstrukcijų rinkinį (klasė – poklasė-tipas, savybė – subsavybė, domenai – sfera), įgalina apibrėžti taikymo srities žodyną, organizuoja šį žodyną kaip apibendrinimų hierarchiją. RDF-S išreiškiamoji galia yra maža (be tiksliai apibrėžiamos prasmės, be loginio išvedimo modelio) [11], [12].

Semantinio tinklo technologijų stekas

Semantinio tinklo technologijų steke atvaizduojamos naujausios kartos semantinio tinklo technologijos.

Semantinio tinklo technologijų stekas – tai semantinio tinklo įgyvendinimui būtinos technologijos ir standartai, kurie turi savo hierarchiją. Kiekvienas sluoksnis išnaudoja hierarchiškai žemesniame sluoksnyje esančių technologijų teikiamas galimybes. Žemiau pateikiamas semantinio tinklo technologijų stekas.



1.1 pav. Semantinio tinklo technologijų stekas [13]

Semantinio tinklo technologijų steką galima išskaidyti į tris pagrindines dalis:

- WWW tinklo technologijos (Unicode, URI, XML);
- semantinio tinklo technologijos (RDF, RDFS, OWL, SPARQL);
- koncepcinės semantinio tinklo technologijos, kurios būtinos semantiniam tinklui įgyvendinti (Logic, Proof, Trust) [13], [14].

Semantinio tinklo technologijos tiriamojo darbo sprendimo realizacijoje nebus naudojamos. Grynas semantinis anotavimas uždeda sąvokai anotaciją ir žymą į ontologijų individą. Teisinių dokumentų automatinio semantinio anotavimo sprendimo realizacijoje sprendimas bus panašus, nes sąvokoms uždedamos anotacijos.

Tiriamajame darbe semantinis anotavimas orientuojamas į anotacijų taisyklių aprašymą, kurios skirtos teisės normos dalių atpažinimui.

1.5 Tyrimo objekto naudotojų analizė

Teisinių dokumentų automatinis semantinis anotavimas gali būti naudingas: teisininkams, Seimo nariams, dėstytojams, teisės studijų studentams ir kitiems su šia dalykine sritimi susijusiems žmonėms.

1.1 lentelė. Vartotojo tipo ir savybių lentelė

Vartotojo tipas	Vartotojo savybės
Teisininkai, Seimo nariai, dėstytojai, teisės studijų studentai, buhalteriai	Dalykinės srities specialistai, kurie dirba su teisiniais dokumentais Seime, universitetuose ir kitose susijusiose įstaigose. Šios srities specialistai nėra programuotojai, tačiau jie gerai suvokia teisinių dokumentų struktūras, geba atrasti sąsajas ir esybes tarp žodžių.
Programuotojai	Šios srities specialistai geba automatizuoti lietuviško teksto semantinį anotavimą, aprašydami esybes, kintamuosius ir kt.

1.2 lentelė. Problemos, su kuriomis susiduria vartotojai.

Problema	Kaip viskas vyksta dabar
Sugaištama daug laiko anotuojant teisinius dokumentus rankiniu būdu.	Teisiniai dokumentai turi savitą terminologiją, tokio teksto anotavimas vyksta rankiniu būdu, kurį atlikdami vartotojai sugaišta daug laiko.
Nėra sukurtas įrankis LR teisinių dokumentų automatiniam semantiniam anotavimui.	Norint turėti suanotuotą teisinį dokumentą, to negalime padaryti automatinio būdu, rinkoje nėra alternatyvių sprendimų.

1.6 Semantinio anotavimo įrankinių analizė (Lietuvos ir tarptautiniu mastu)

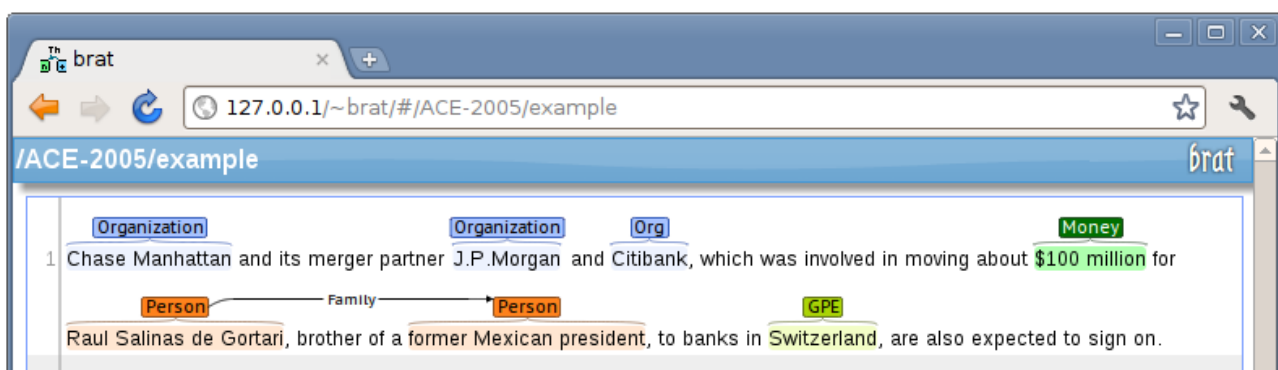
Esamų problemos sprendimo metodų analizėje plačiau nagrinėjami ir aprašomi automatinio semantinio anotavimo įrankiai, labiausiai detalizuojami įrankiai: Brat rapid ir GoNTogle. Pasirinktuose įrankiuose yra aiškiai pažymimos ir atvaizduojamos teksto anotacijos. Naujai kuriamas sprendimas, planuojamas realizuoti atsižvelgiant į šių įrankių anotacijų atvaizdavimo pavyzdį.

Brat rapid – teksto automatinio anotavimo įrankis

Brat – tai įrankis, esantis internetiniame puslapyje, skirtas teksto anotavimui.

Brat įrankis yra skirtas struktūrizuotam teksto anotavimui. Norint automatiškai apdoroti tekstą su šiuo įrankiu, tekstas privalo turėti aiškią struktūrą.

Žemiau esančiame paveikslėlyje pateiktas paprastas pavyzdys, kuriame atvaizduojamas suanotuotas sakinyss, identifikuojant kelių subjektų tipus ir jų tarpusavio ryšius.

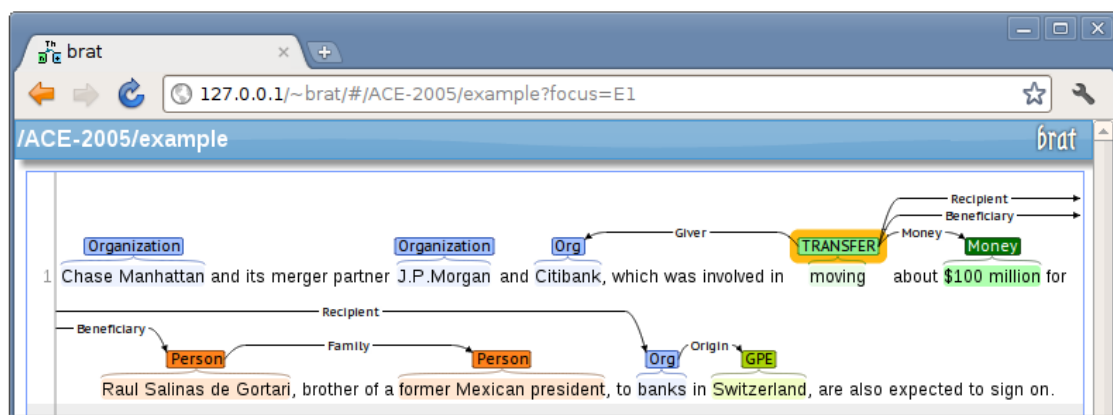


1.2 pav. Teksto anotacijos pavyzdys naudojant Brat įrankį [15]

Šioje iliustracijoje vaizduojamos dvi pagrindinės antoacijų kategorijos:

- teksto anotacijos. Jos iliustracijoje pažymėtos kaip „Organization“ and „Person“;
- ryšiai tarp anotacijų, iliustracijoje pažymėta kaip „Family“.

Brat rapid įrankis taip pat palaiko asociacijas, kurios gali susieti bet kokį skaičių kitų anotacijų, turinčių panašių savybių kitose specifinėse rolėse. Ši anotacijos kategorija gali būti naudojama plačiai, vienas iš pavyzdžių „įvykio anotacija“, iliustracijoje tokia anotacija pažymėta „TRANSFER“



1.3 pav. Teksto anotacijos pavyzdys naudojant Brat įrankį [15]

Kiti anotacijų tipai ir sąvybės gali būti išsamiau apibrėžti naudojant atributus, kuriuos galima nustatyti anotacijose, pavyzdžiui, pažymėti įvykį kaip faktinį ar spekuliacinį, arba pažymėti, kad subjektas paminėtas kaip nuoroda į grupę ar asmenį. Brat taip pat įgyvendina keletą funkcijų, kurios remiasi natūralios kalbos apdorojimo technologijomis, siekiant atsižvelgti į žmonių pastabas [15].

GoNTogle: Įrankis skirtas semantiniam anotavimui ir paieškai (angl. GoNTogle: A tool for semantini annotation and search)

GoNTogle leidžia vartotojams gauti skirtingų formato dokumentų anotacijas, pritaikant ontologijos sąvokas. Taip pat šis įrankis kuria automatinius anotacijų pasiūlymus, kurie yra pagrįsti tekstiniu panašumu į ankstesnių dokumentų anotacijas. GoNTogle jungia raktažodžius ir semantinę paiešką, siūlo išplėstas ontologijų užklausų galimybes.

Anotacijos apima duomenų žymėjimą pagal prasmę t. y. ontologijų klasėmis, dėl šios priežasties duomenys įgyja teisingas prasmes. Anotacijų duomenys gali pagerinti paieškos galimybes, kadangi atlikus anotavimą, vartotojas gali ieškoti informacijos naudojant ne tik paprastus ir tradicinius raktažodžius, bet naudojant gerai apibrėžtas bendrąsias sąvokas, kurios apibūdina reikalingas informacijos sritis.

GoNTogle palaiko rankinį ir automatinį anotavimą keliems dokumentų tipams, galimi dokumentų formatai: doc, pdf, rft, txt, odt, sxw. Dokumentai anotuojami naudojant ontologijų klases. Priešingai nei kiti anotacijų sprendimai, šio sprendimo tikslas – įdiegti lengvai naudojamą įrenginį skirtą dokumentų anotacijoms ir paieškomis, kuris pilnai:

- a) palaikytų peržiūrėjimą ir populiarių dokumentų tipų anotavimą, išlaikant jų pradinius formatus;
- b) dalintųsi anotacijomis;
- c) vykdytų paiešką kombinuojant raktinius žodžius ir semantinę paiešką.

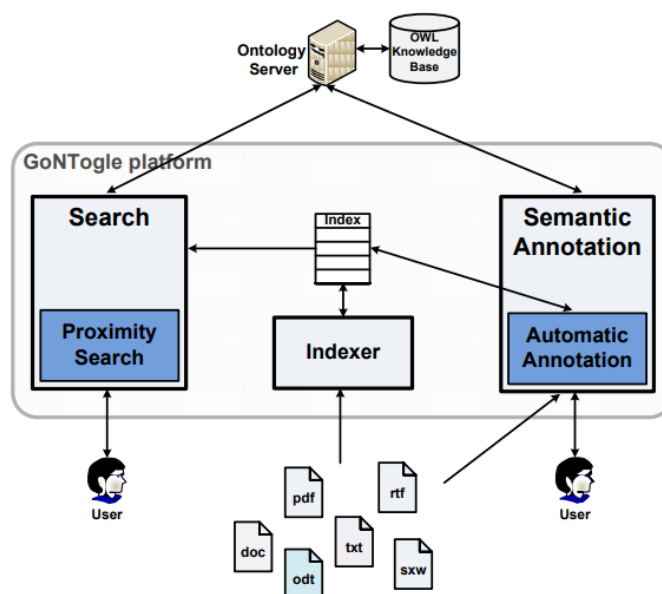
Pagindinės GoNTogle įrankinio sąvybės:

- leidžia vartotojams atidaryti ir peržiūrėti plačiai naudojamus dokumentų formatus, pavyzdžiui: .doc ir .pdf, išlaikant originalų formatą;
- suteikia paprastą ir intuityvų dokumentų ar jo dalių anotavimą, naudojant OWL ir RDF/S ontologijas.
- suteikia automatinį anotavimo mechanizmą, pagrįstą sudarytais modeliais, kurie paremti vartotojų anotavimo istorija, siekiant anotavimą pritaikyti atsižvelgiant į vartotoją;
- įrankis paremtas serverių architektūra, kur dokumentų anotacijos yra saugomos centrinėje saugykloje. Šis sprendimas suteikia bendradarbiavimo aplinką, kurioje vartotojas gali anotuoti ir ieškoti dokumentų;
- įrankis sujungia raktinius žodžius ir semantinę paiešką, teikia išplėstinę paiešką.

GoNTogle sistemos apžvalga

1.4 pav. pateikiama GoNTogle architektūra. Sistema yra suskirstyta į keturis pagrindinius komponentus:

- a) semantinis anotacijos komponentas, kuriame pateikiamos priemonės, susijusios su semantine dokumentų anotacija. Jį sudaro trys moduliai: dokumento peržiūra, ontologijos peržiūra ir anotacijos redaktorius;
- b) ontologijos serverio komponentas, kuris saugo semantines dokumentų anotacijas OWL ontologijų pavidalo formose;
- c) indeksavimo komponentas, kuris atsakingas už dokumentų indeksavimą;
- d) paieškos komponentas, kuris leidžia vartotojams ieškoti dokumentų naudojant tiek teksto (raktažodžių paieška), tiek semantinės (ontologijų paieška) informaciją.



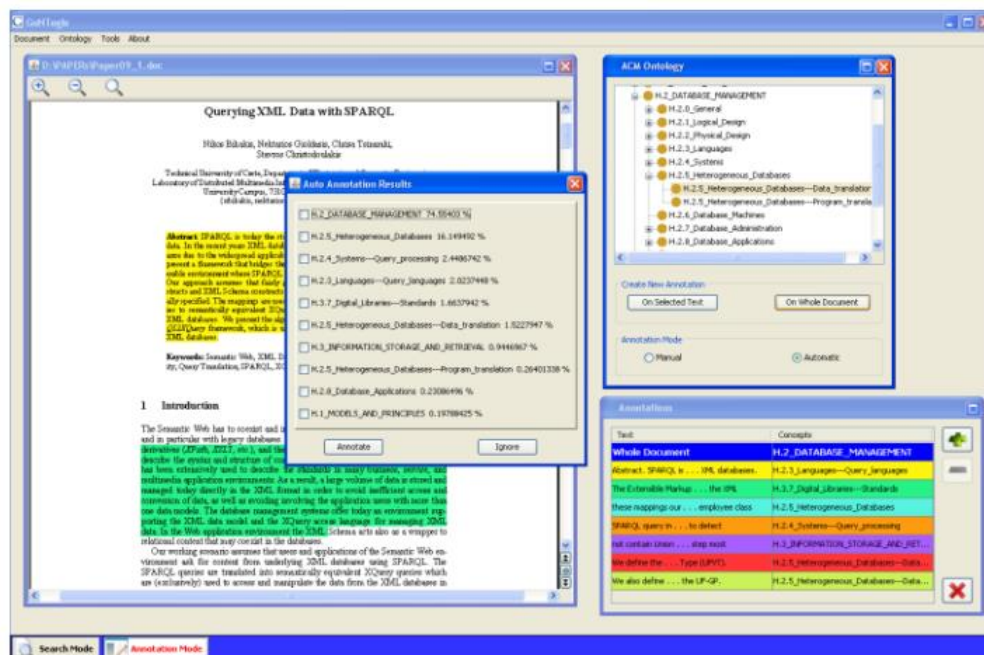
1.4 pav. GoNTogle architektūra [16]

GoNTogle semantinis anotavimas

Semantinio anotavimo komponentas turi dvi pagindines funkcijas:

- a) pilna viso dokumento anotacija;
- b) anotacija dokumento dalims.

Be to, vartotojas turi galimybę pasirinkti tarp rankinio ir automatinio anotavimo. Žemiau pateiktame 1.5 pav. atvaizduojamas semantinio anotavimo taikomosios programos langas [16].



1.5 pav. Semantinio anotavimo pavyzdys naudojant GoNTogle [16]

LKSSAIS (semantika.lt)

Lietuvių kalbos sintaksinės ir semantinės analizės informacinė sistema. LKSSAIS paskirtis – sumažinti lietuvių rašytinės kalbos išteklį ir analizės priemonių trūkumą, padidinti jų prieinamumą ir tokiu būdu sudaryti geresnes sąlygas asmenims ir organizacijoms vartoti lietuvių rašytinę kalbą elektroninėje erdvėje. Paslauga leidžia morfologiškai ir sintaksiškai suanotuoti įkeltą lietuvišką tekstą. vartotojas turi galimybę pasirinkti anotavimo rūšį: tik morfologinio lygio anotavimo grandinėle arba morfologinė-sintaksinė anotavimo grandinėle. Semantika.lt leidžia atlikti lietuviškų interneto svetainių straipsnių semantinę analizę ir paiešką politikos, ekonomikos ir verslo bei viešojo administravimo sričių tekstuose pateikiant klausimus SBVR struktūrizuota lietuvių kalba. [17].

GATE

Tai atviro kodo kalbos ir teksto inžinerijos platforma, skirta teksto semantinei analizei, apimanti onotologijų klases. GATE yra informacijos išgavimo sistema, turinti daug plėtinių, kurie leidžia sistemą adaptuoti įvairiems kitiems poreikiams. GATE leidžia anotuoti ir analizuoti grafiškai

konferencijų grandinėle. Šis įrankis turi F – vertės automatinį apskaičiavimą. Vartotojui norint naudotis šiuo įrankiu, reikalingos gilesnės žinios semantiniame anotavime [18].

ontotext.com

Šis įrankis paremtas semantinėmis technologijomis, leidžiantis geriau valdyti teksto turinį, sąvokų atradimą ir semantinę paiešką. Ontotext paslauga analizuoja tekstą, ištraukia sąvokas, atpažįsta temas, raktinius žodžius ir tarpusavio sąryšius, taip pat geba suvienodinti panašius subjektus. Suanotuoto teksto rezultatas apima metaduomenis, suderintus su žinių grafiku, kuris apima viso turinio valdymą [19].

Esamų sprendimų palyginamoji analizė

Žemiau esančioje lentelėje pateikiama semantinio anotavimo įrankių lyginamoji analizė. Pasirinkti įrankiai: GATE, UIMA, Ontotex.lt, Semantika.lt, Brad rapid, GoNtogle.

1.3 lentelė. Esamų sprendimų palyginimas

Palyginimo kriterijus	GATE [18]	Ontotex.com [19]	Semantika.lt [17]	Brat rapid [15]	GoNtogle [16]
Anotavimui skirtų tekstų failo įkėlimas arba teksto įkopijavimas	+	+	+	+	+
Prisijungimas prie teksto semantiniame anotavimui skirtų programų per serverį.	+	+	+	+	+
Paieška ontologijose SPARQL užklauso	+	+	+	-	+
Internetinių svetainių semantinė analizė	+	+	+	-	-
Anotacijų išskyrimas spalvomis	+	+	-	+	+
Ontologijų klasių išskyrimas ir atvaizdavimas	+	+	-	+	+
Ryšų tarp anotacijų atvaizdavimas	+	+	-	+	+
Teisinių dokumentų semantinis anotavimas	-	-	-	-	-

1.7 Darbo tikslas, uždaviniai, planas ir siekiami privalumai.

Tikslas

Sudaryti galimybę teisinių dokumentų automatiniam semantiniam anotavimui ir atpažinti didesnę kiekį esybių bei ryšių.

Uždaviniai:

1. išanalizuoti pasirinktus teisinius dokumentus (struktūra, sąsajos tarp žodžių);
2. sukurti ontologijos srities schemą;
3. išanalizuoti esamas semantinio anotavimo priemones (metodai, įrankiai, standartai ir formatai);
4. sudaryti automatinio semantinio anotavimo algoritmą;
5. eksperimentiškai ištirti sudarytą algoritmą ir pateikti išvadas.

1.8 Siekiamo sprendimo apibrėžimas

Šiame darbe siekiama sukurti Lietuvos Respublikos teisinių dokumentų automatinio semantinio anotavimo įrankį. Įrankis skirtas teisės normos dalių atpažinimui vieno straipsnio lygyje. Sprendimo realizacijai nebus naudojamos semantinio tinklo technologijos (RDF, RDFS, OWL, SPARQL), kadangi semantiškai anotuojant nesiekama kurti ontologijos subgrafo. Rastos anotacijos nevirs ontologijos individu. Teisinių dokumentų automatinio semantinio anotavimo siekiama sukurti didesnę kiekį anotacijų taisyklių, kurios gebėtų automatiškai atpažinti teisės normos sąvokas kodeksų straipsniuose ir semantinio anotavimo komponentais apibrėžti anotacijų taisykles. Rastos anotacijos turi būti išskiriamos tekste spalvomis arba kitu ženkliniu, kad būtų paprasčiau naudotojui atpažinti suanotuotą tekstą.

1.9 Analizės išvados

1. Atlikus Lietuvos Respublikos teisinių dokumentų analizę, galima daryti išvadą, kad kodeksai yra sudaryti iš dalių, dalys iš skyrių, skyriai iš straipsnių. Gilesnė analizė atlikta straipsnio lygyje, po kurios pastebėta, kad kiekvieno kodekso straipsniai turi griežtas struktūras bei sąsajas tarp žodžių. Straipsnius sudaro nustatytos teisės normos dalys ir teisės ryšio elementai. Lyginant kelis Lietuvos Respublikos kodeksus pastebėta, kad kodeksai tarpusavyje skiriasi rašymo stilistika ir turi tik dalinai panašią straipsnių struktūrą. Remiantis nustatytais teisės normos dalių apibrėžimais bus kuriamos anotacijų taisyklės pagrindinėms teisės normos dalims atpažinti.
2. Analizės metu buvo palyginti du įrankiai skirti ontologijoms kurti – Protégé ir TopBraid. Populiariausias ir daugiau naudojamas įrankis yra Protégé. Jis pasižymi paprasta naudojimosi grafine sąsaja, vienas iš šio įrankio privalumų – leidžia atvaizduoti ontologiją grafiškai.

Įrankas Protégé buvo išbandytas juo naudojantis ir sudarant ontologijas Lietuvos Respublikos teisinių dokumentų anotacijoms. Ontologija nėra būtina tiriamojo darbo tikslui pasiekti, tačiau tai yra plačiai naudojama esamuose sprendimuose ir gali būti kaip nauja alternatyva ateityje tobulinant bei plečiant automatinio semantinio anotavimo taikomąją programą.

3. Atlikus LKIF branduolio analizę, buvo siekiama sudaryti Lietuvos Respublikos teisinių dokumentų struktūros modelį, tačiau šis sprendimas nebuvo tinkamas dalykinei sričiai, nes LKIF apibrėžia konkrečių teisinių bylų nagrinėjimą ir ontologija skirta teisės normos pažeidimų išvedimui. Žvelgiant į ateitį LKIF galėtų būti naudojamas sudarant naujus teisinius dokumentus.
4. Atlikus semantinio anotavimo įrankių analizę, nuspręsta, kad tiriamojo darbo sprendimui įgyvendinti bus kuriamas naujas įrankis, skirtas teisinių dokumentų automatiniam semantiniam anotavimui. Sprendimo įgyvendinimui reikalinga sąsaja su semantika.lt, iš kurios bus gaunamos pagrindinės žodžių gramatinės kategorijos, kadangi tai vienintelis įrankis skirtas lietuviško teksto semantiniam anotavimui.
5. Kadangi kuriamo įrankio tikslas aprašyti kuo daugiau bendrinių taisyklių, skirtų teisės normos dalių atpažinimui, todėl sprendimo įgyvendinimui nebus naudojamos semantinio tinklo technologijos (RDF, RDFS, OWL, SPARQL). Išanalizavus esamus sprendimus tarptautinėje rinkoje priimta išvada, kad kuriamame sprendime reikėtų ne tik automatiškai atpažinti didesnę kiekį esybių, tačiau aiškiai ir vartotojui suprantamai atvaizduoti rastas anotacijas, labiausiai tinkami sprendimai yra įgyvendinti Brat rapid ir GoNTogle įrankiuose.

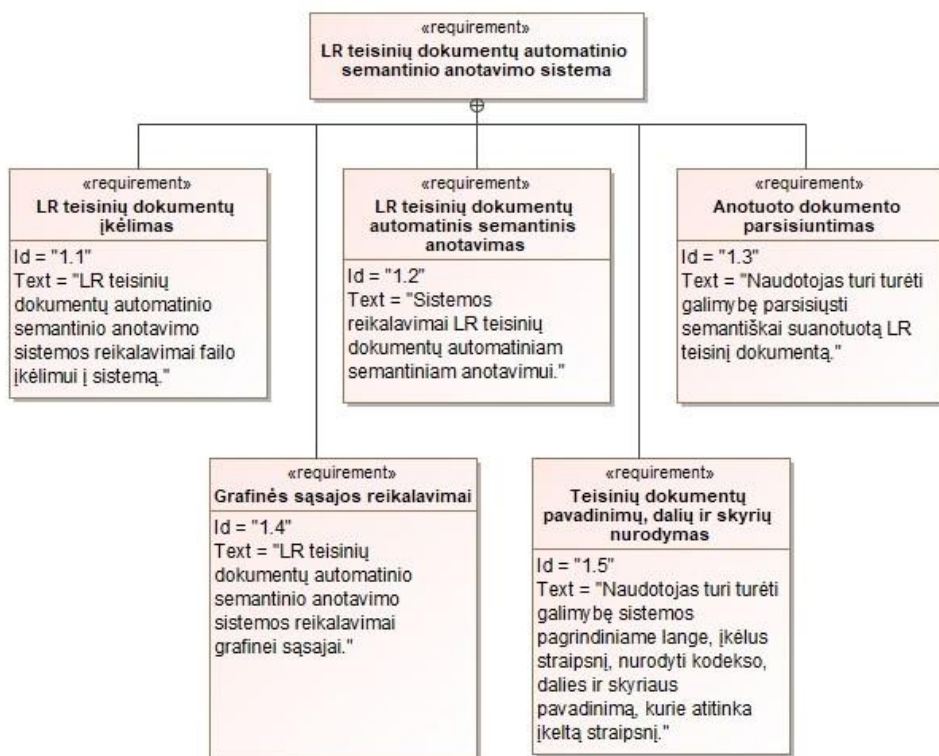
6. TEISINIŲ DOKUMENTŲ AUTOMATINIO SEMANTINIO ANOTAVIMO SISTEMOS REIKALAVIMŲ SPECIFIKACIJA IR PROJEKTAS

2.1 Reikalavimų specifikuojimas

Reikalavimų specifikuojimas, funkciniai reikalavimai

Lietuvos Respublikos teisinių dokumentų automatinio semantinio anotavimo sistemos reikalavimai skirstomi į šias dalis:

- 1.1 LR teisinių dokumentų įkėlimas – sistemoje privalo būti galimybė vartotojui įkelti teisinį dokumentą. Apsibrėžta, kad teisinių dokumentų įkėlimas yra apribojamas straipsnio lygyje t. y. į sistemą gali būti įkeliamas pasirinkto LR kodekso konkretus vienas straipsnis.
- 1.2 LR teisinių dokumentų automatinis semantinis anotavimas – tai yra vienas iš pagrindinių sistemos reikalavimų. Sistema pagal aprašytas taisykles turi automatiškai suanotuoti įkeltą straipsnį, atpažįstant pagrindines anotacijas pagal žodžių lemas, gramatines kategorijas ir kt. Reikalavimas detalizuotas 2.3 pav.
- 1.3 Anotuoto dokumento parsisiuntimas – sistemai suanotavus dokumentą, naudotojas turi turėti galimybę parsisiųsti anotaciją. Anotacija išsaugoma kompiuteryje, pasirinktame kataloge. Parsiunčiamo failo formatas - JSON.
- 1.4 Grafinės sąsajos reikalavimai – tai reikalavimai skirti taikomosios programos grafinei sąsajai, jie detalizuojami 2.4 pav.
- 1.5 Teisinių dokumentų pavadinimų, dalių ir skyrių nurodymas – naudotojas įkėlęs straipsnį į sistemą turi turėti galimybę sistemoje pasirinkti LR kodekso pavadinimą, nurodyti dalį ir skyrių. Kadangi sistema pagal įkelto straipsnio turinį negali automatiškai atpažinti straipsnio kilmės, todėl šis reikalavimas skirtas – užfiksuoti meta duomenis ir juos išsaugoti JSON formate, kai parsisiunčiama dokumento anotacija.

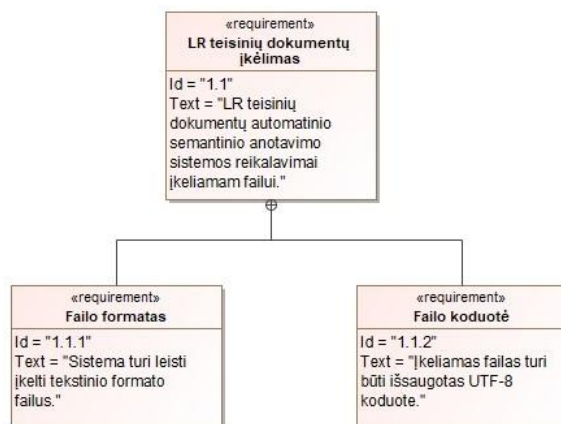


2.1 pav. Pagrindiniai Lietuvos Respublikos teisinių dokumentų automatinio semantinio anotavimo sistemos reikalavimai

Reikalavimai failo įkėlimui į sistemą

Reikalavimas „1.1 LR teisinių dokumentų įkėlimas“ išskaidomas į kitus smulkesnius reikalavimus:

- 1.1.1 Failo formatas – kuriant sistemą yra apsibrėžiama kokio formato failas gali būti įkeliamas į sistemą. Pasirinktas txt formatas.
- 1.1.2 Failo koduotė – kadangi atliekamas lietuviško teksto automatinis semantinis anotavimas, todėl įkeliamas failas privalo būti išsaugotas UTF-8 koduote.

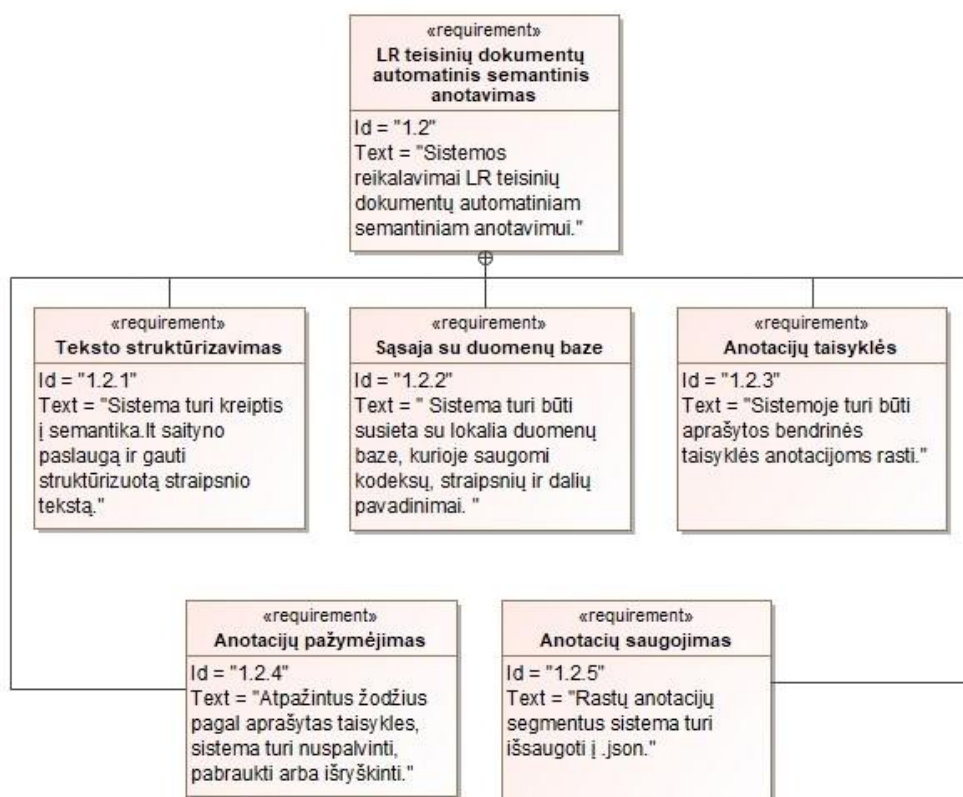


2.2 pav. Reikalavimai LR teisinių dokumentų įkėlimui į sistemą

Reikalavimai automatiniam semantiniam anotavimui

Reikalavimas 1.2. „LR teisinių dokumentų automatinis semantinis anotavimas“ skaidomas į kitus reikalavimus:

- 1.2.1 Teksto struktūrizavimas – įkėlus failą į sistemą, ji turi kreiptis į semantika.lt saityno paslaugą, iš kurios būtų grąžinamas struktūrizuotas tekstas. Tekstas išskaidomas į segmentus, pastraipas ir sakinius bei grąžinamos kiekvieno žodžio visos galimos gramatinės kategorijos.
- 1.2.2 Sąsaja su duomenų baze – sistema turi būti susieta su duomenų baze, kurioje būtų saugomi kodeksų, dalių ir skyrių pavadinimai, kadangi naudotojas pagrindiniame lange turi nurodyti pagrindinius straipsnio meta duomenis. Meta duomenys yra išsaugomi JSON faile.
- 1.2.3 Anotacijų taisyklės – sistemoje turi būti aprašytos bendrinės taisyklės skirtos teisės normos dalims rasti. Taisyklėms yra naudojamos žodžių gramatinės kategorijos ir lemos.
- 1.2.4 Anotacijų pažymėjimas – kai sistemoje pagal aprašytus metodus žodžiams pritaikomos anotacijų taisyklės, sistemoje turi būti aprašytas kitas metodas, kuriame kiekviena rasta anotacija būtų nuspalvinama ar kitaip pažymima pagal jos rūšį.
- 1.2.5 Anotacijų saugojimas – žodžiams pritaikius aprašytas taisykles, sistemoje turi būti aprašytas kitas metodas, kuris skirtas anotacijų segmentų išsaugojimui JSON formato faile.

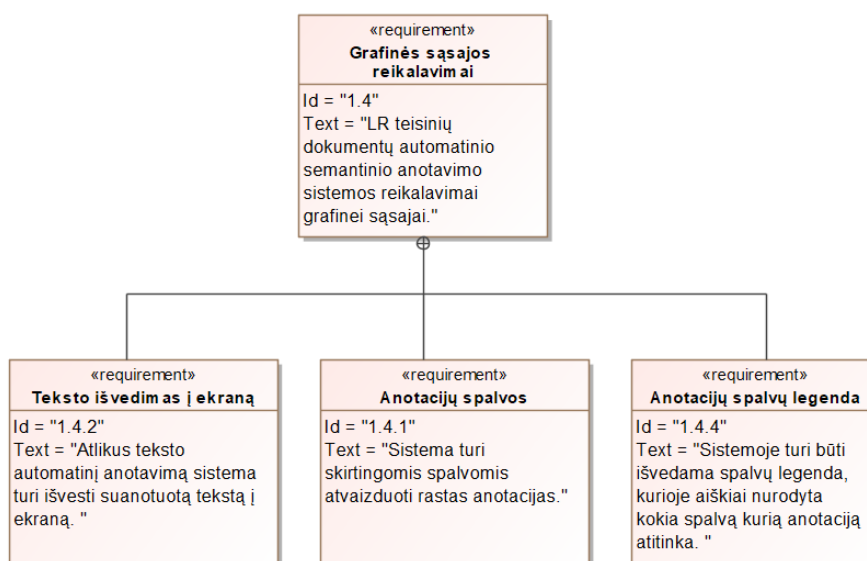


2.3 pav. Reikalavimai LR teisinių dokumentų automatinio semantinio anotavimo

Reikalavimai sistemos grafinei sąsajai

Žemiau pateiktame reikalavimų modelyje detalizuojami grafinės sistemos reikalavimai:

- 1.4.1 Anotacijų spalvos – suanotavus įkeltą straipsnį, sistema turi sužymėti rastas anotacijas skirtingomis spalvomis, kad naudotojas galėtų aiškiai matyti, kokios anotacijos buvo atpažintos straipsnyje. Neatpažintas tekstas paliekamas juoda spalva.
- 1.4.2 Teksto išvedimas į ekraną – sistemai atlikus automatinį semantinį teksto anotavimą, suanotuotą tekstą turi išvesti pagrindiniame sistemos lange.
- 1.4.3 Anotacijų spalvų legenda – sistemos pagrindiniame lange turi būti išvedama anotacijų legenda, kurioje pažymima kokią spalvą kokia anotaciją atitinka. Kad vartotojas lengviau suprastų, kurie žodžiais kokias anotacijas atitinka.

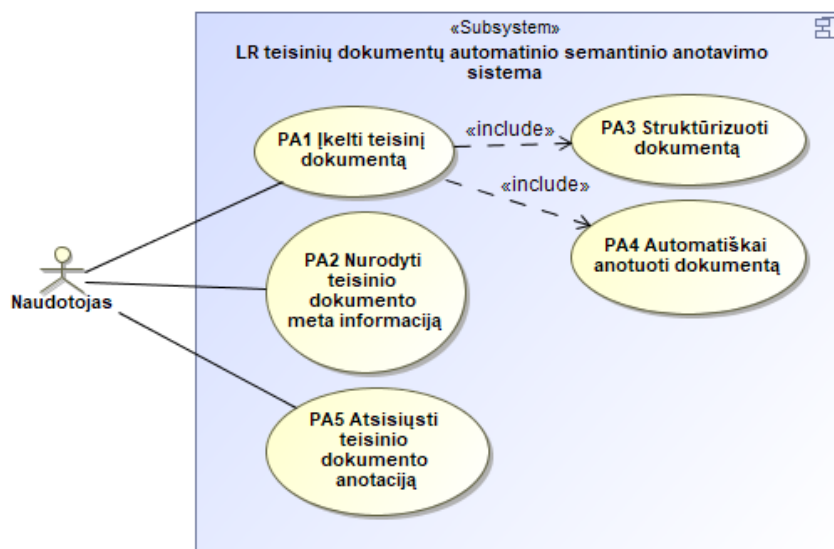


2.4 pav. Grafinės sąsajos reikalavimai

Panaudojimo atvejai

Panaudojimo atvejų modelyje pateikiami pagrindiniai panaudos atvejai skirti Lietuvos Respublikos teisinių dokumentų automatinio semantinio anotavimo sprendimo įgyvendinimui. Modelyje išskiriamas vienas aktorius – sistemos naudotojas ir sistemos bei aktoriaus atliekamos funkcijos.

Kiekvienas panaudojimo atvejis detalizuojamas atskiroje lentelėje, įvertinant pagrindinius aspektus.

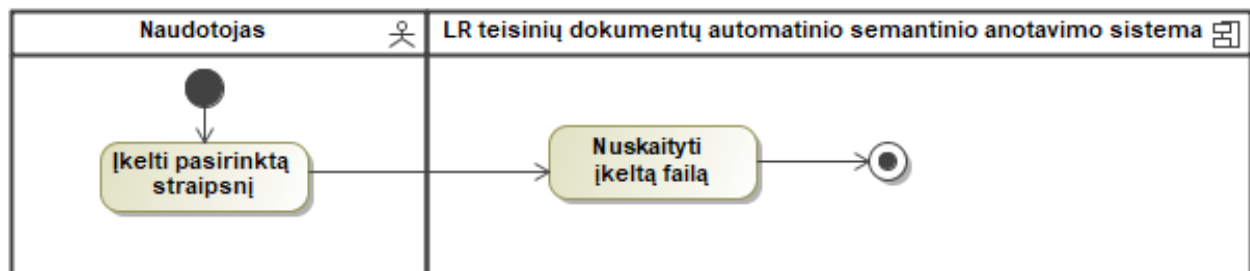


2.5 pav. Panaudojimo atvejų diagrama

Panaudojimo atvejų detalizacija

2.1 lentelė. „PA 1 Įkelti teisinį dokumentą“

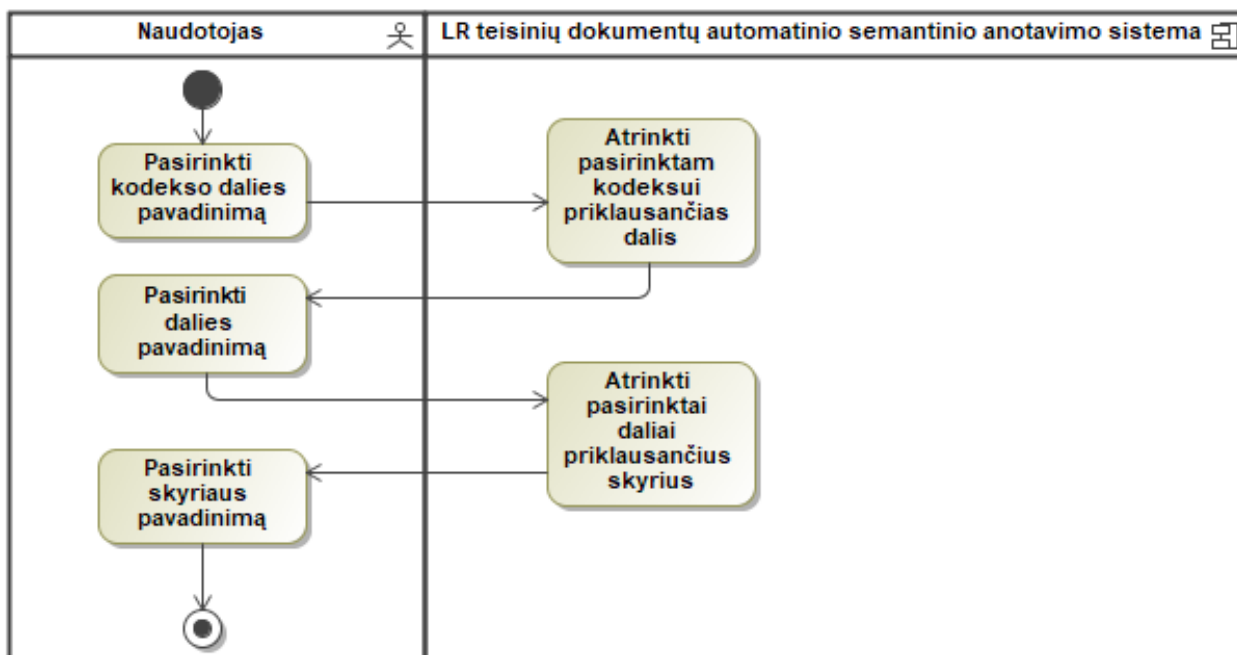
PA 1. Įkelti teisinį dokumentą	
Tikslas/uždavinys. Leisti įkelti į sistemą straipsnį.	
Aprašymas. Šis panaudojimo atvejis skirtas LR teisinių kodeksų, konkrečių straipsnių įkėlimui.	
Prieš-sąlyga	Naudotojas įsijungia Lietuvos Respublikos teisinių dokumentų automatinio anotavimo taikomąją programą.
Aktorius	Naudotojas
Sužadinimo sąlyga	Naudotojui reikalingas semantiškai anotuotas konkretus pasirinkto Lietuvos Respublikos kodekso straipsnis.
Pagrindinis scenarijus	
Pagrindinis įvykių srautas	Sistemos reakcija
<ul style="list-style-type: none"> Naudotojas pasirenka straipsnį, kurio anotacija yra reikalinga; sistemoje įkelia pasirinktą straipsnį. 	Įkėlus straipsnį sistema tikrina ar įkeltas failas teisingo formato t. y. išsaugotas kaip tekstinis failas „.txt“ ir jo koduotė UTF-8. Įkėlus teisingo formato failą, sistema leidžia atlikti sekančius veiksmus, įkėlus neteisingo formato failą sistema nebeleidžia atlikti jokių sekančių veiksmų.
Po-sąlyga Įkėlus straipsnį į sistemą, naudotojas sistemos lange, laukuose: „Kodekso pavadinimas“, „Dalies pavadinimas“, „Skyriaus pavadinimas“, pasirenka teisinio dokumento meta informaciją.	Naudotojas sėkmingai įkelia pasirinktą straipsnį iš Lietuvos Respublikos kodekso ir gali sėkmingai tęsti sekančius žingsnius.



2.6 pav. „PA 1. Įkelti LR kodeksą“ veiklos diagrama

2.2 lentelė. „PA 2 Nurodyti teisinio dokumento meta informaciją“

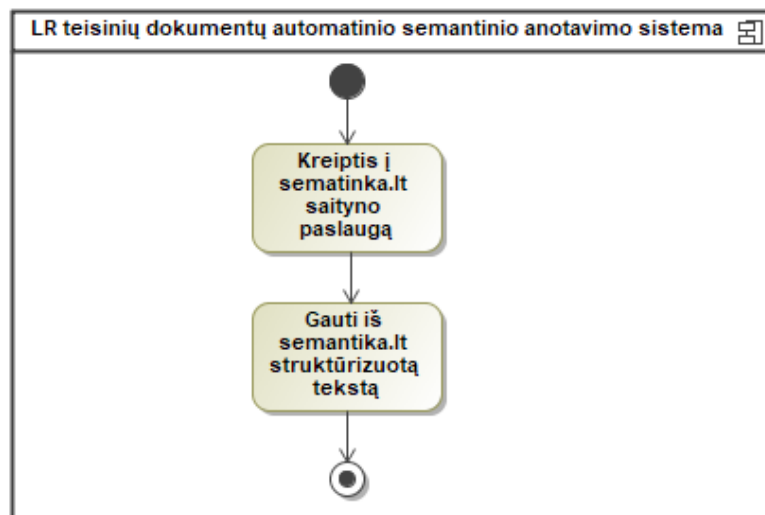
PA 2. Nurodyti Lietuvos Respublikos teisinio dokumento meta informaciją	
Tikslas/uždavinys. Įkėlus straipsnį į sistemą, laukuose „Kodekso pavadinimas“, „Dalies pavadinimas“ ir „Skyriaus pavadinimas“ reikia pasirinkti kodekso, kurio straipsnis buvo įkeltas, meta informaciją tam, kad išsaugant anotaciją vartotojas turėtų tikslią straipsnio kilmę.	
Aprašymas. Šis panaudojimo atvejis skirtas patikslinti, kokio kodekso straipsnis buvo įkeltas į sistemą, siekiant duomenis išsaugoti straipsnio anotacijoje.	
Prieš-sąlyga	PA1
Sužadinimo sąlyga	Kadangi teisinė sritis yra labai didelė ir kiekvienas teisinis kodeksas turi labai platų spektrą straipsnių, todėl išsaugojus straipsnio anotaciją yra svarbu turėti meta duomenis dėl informacijos atsekamumo.
Pagrindinis scenarijus	
Pagrindinis įvykių srautas	Sistemos reakcija
Įkeltas straipsnis į sistemą, nurodytas kodekso pavadinimas, nurodytas dalies pavadinimas, nurodytas skyriaus pavadinimas.	Pagrindiniame sistemos lange, sistema leidžia pasirinkti kodekso pavadinimą, pagal tai išfiltruojamos susijusios dalys. Pasirinkus konkrečią dalį išfiltruojamas skyrių sąrašas.
Po-sąlyga	Sistema išsaugojo naudotojo pasirinktą meta informaciją.



2.7 pav. „PA 2. Nurodyti teisinio dokumento meta informaciją“ veiklos diagrama

2.3 lentelė. „PA 3 Struktūrizuoti dokumentą“

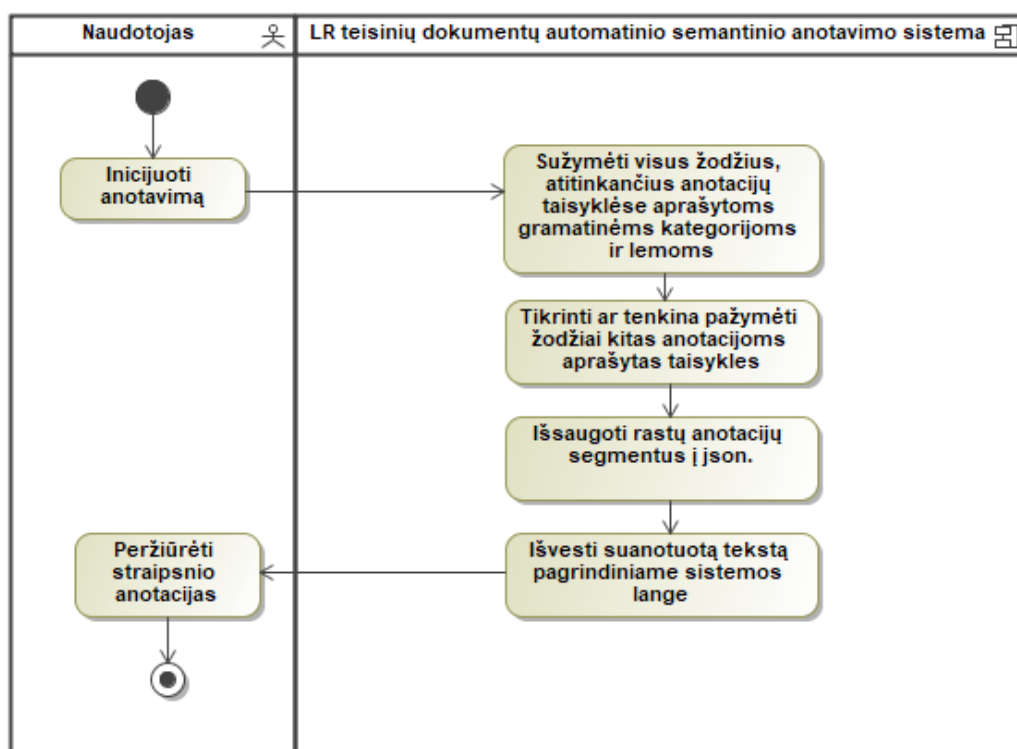
PA 3. Struktūrizuoti dokumentą	
Tikslas/uždavinys. Struktūrizuoti įkelto straipsnio tekstą, kad būtų galima pritaikyti aprašytas anotacijų taisykles.	
Aprašymas. Šis panaudojimo atvejis skirtas straipsnio teksto struktūrizavimui. Straipsnis suskaidomas į segmentus, pastraipas ir sakinius, išskiriamos kiekvieno žodžio lemos ir žodžių gramatinės kategorijos.	
Prieš-sąlyga	PA1
Sužadinimo sąlyga	Siekiant turėti struktūrizuotą tekstą, automatinio semantinio anotavimo sistema turi kreiptis į semantika.lt saityno paslaugą.
Pagrindinis scenarijus	
Pagrindinis įvykių srautas	Sistemos reakcija
Į sistemą įkeltas tekstinis failas, kurio koduotė UTF-8.	Įkėlus failą į sistemą kreipiamasi į sematinka.lt saityno paslaugą (angl. web service)
Po-sąlyga	Sistemai grąžinamas rezultatas iš semantika.lt t. y. struktūrizuotas failas, išskaidytas į segmentus, pastraipas, sakinius, lemas ir gramatines kategorijas. Struktūrizuotam tekstui taikomos anotacijų taisyklės.



2.8 pav. „PA 3. Struktūrizuoti dokumentą“ veiklos diagrama

2.4 lentelė. „PA 4 Anotuoti straipsnį automatiškai“

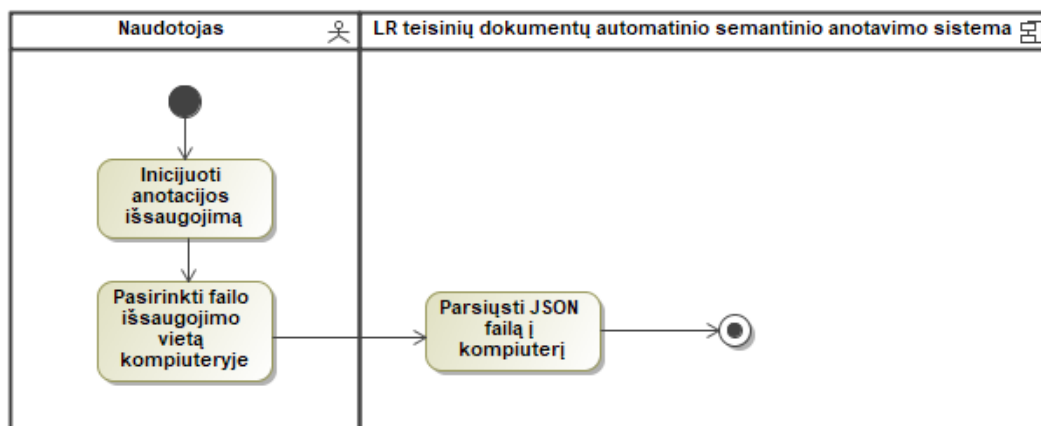
PA 4. Anotuoti straipsnį automatiškai	
Tikslas/uždavinys. Semantiškai suanotuoti įkeltą straipsnio tekstą pagal aprašytas anotavimo taisykles.	
Aprašymas. Šis panaudojimo atvejis skirtas įkelti ir struktūrizuoti straipsnio teksto automatiniam anotavimui t. y. teisės normų sąvokų atpažinimui straipsnyje.	
Prieš-sąlyga	PA 1, PA 3
Sužadinimo sąlyga	Naudotojui reikalingas suanotuotas kodekso straipsnis.
Pagrindinis scenarijus	
Pagrindinis įvykių srautas	Sistemos reakcija
Įkėlus straipsnį į sistemą, nurodžius kodekso pavadinimą, dalį ir dalies skyrių, spaudžiama „Anotuoti“.	Sistema struktūrizuotam straipsnio tekstui pritaiko aprašytas anotacijų taisykles. Anotacijos pritaikomos, jeigu teisės normos sąvokos atitinka taisyklėse aprašytas gramatines kategorijas ir lemas arba tenkina kitas nurodytas sąlygas, kurios yra skirtingos kiekvienai teisės normos sąvokai.
Po-sąlyga	Įkeltas straipsnio tekstas suanotuotas ir rezultatas išvedamas sistemos pagrindiniame lange. Kiekviena anotacija yra pažymima skirtinga spalva, vienodos anotacijos pažymimos ta pačia spalva.



2.9 pav. „PA 4. Anotuoti dokumentą automatiškai“ veiklos diagrama

2.5 lentelė. „PA 5 Išsaugoti suanotuoto straipsnio anotaciją“

PA 5. Išsaugoti suanotuoto straipsnio anotaciją	
Tikslas/uždavinys. Leisti išsisaugoti suanotuotą straipsnį JSON formatu pasirinkame kompiuterio kataloge	
Aprašymas. Šis panaudojimo atvejis skirtas automatinio semantinio anotavimo rezultato išsaugojimui.	
Prieš-sąlyga	PA 4
Aktorius	Naudotojas
Sužadinimo sąlyga	Naudotojui reikalingas semantiškai suanotuotas straipsnis
Pagrindinis scenarijus	
Pagrindinis įvykių srautas	Sistemos reakcija
<ul style="list-style-type: none"> • Paspaudžiama išsaugoti anotaciją; • pasirenkamas katalogas, kuriame norima išsaugoti anotaciją; • išsaugoma. 	Naudotojui pasirinkus funkciją išsaugoti anotaciją, sistema anotacijų rezultatą sudėlioja į struktūrizuotą formatą ir išsaugo JSON formatu.
Po-sąlyga	Suanotuotas, įkeltas kodekso straipsnis ir rezultatas išsaugojamas kompiuteryje JSON formatu.



2.10 pav. „PA 5. Atsisiųsti suanotuotą dokumentą“ veiklos diagrama

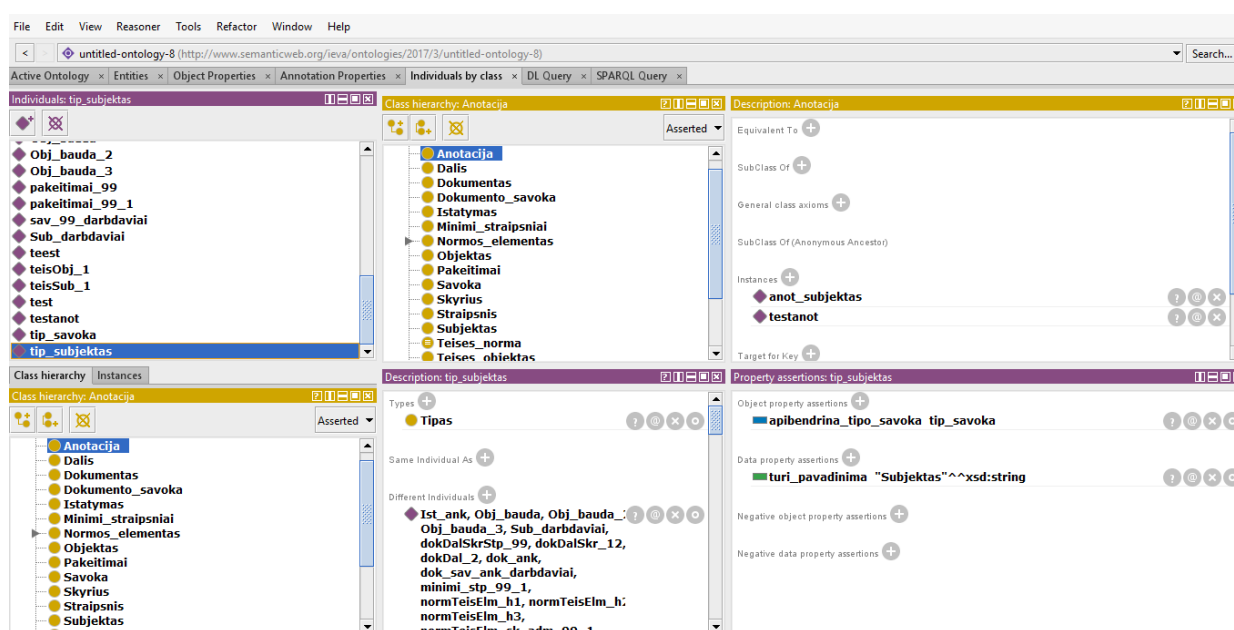
2.2 Teisinių dokumentų struktūros modelis

UML esybių-ryšių modelis

Šioje dalyje pateikiamas esybių-ryšių modelis, kuris buvo kartu kurtas su Mantu Bataičiu ir taip pat atvaizduojamas jo tiriamajame darbe „Teisinių dokumentų automatizuoto semantinio anotavimo galimybių tyrimas“. Modelis buvo sudarytas remiantis straipsniu „Ontologijos modelis teisinėms normoms, kuriant ir naudojant įstatymus“ [20] ir naudojant įrankį Protégé.

Žemiau esančiame paveikslėlyje pateikiamas Protégé įrankio langas „Individual by class“, kuriame matomos ontologijų klasės, kuriami individai bei jų savybės ir sąryšis. Pavyzdyje pateiktos ontologijų klasės LR teisiniams dokumentas.

Dalyje „Class hierarchy“ sukuriamos pagrindinės ontologijų klasės, kurios apibrėžia teisės normos dalis ir dokumento struktūrą. Dalyje „Individuals“ sukuriami klasėms priskirti konkretūs straipsnių individai. „Property assertions“ apibrėžia duomenų savybes, kurias turi klasė, ir kitus klasės apibrėžiančius objektus.



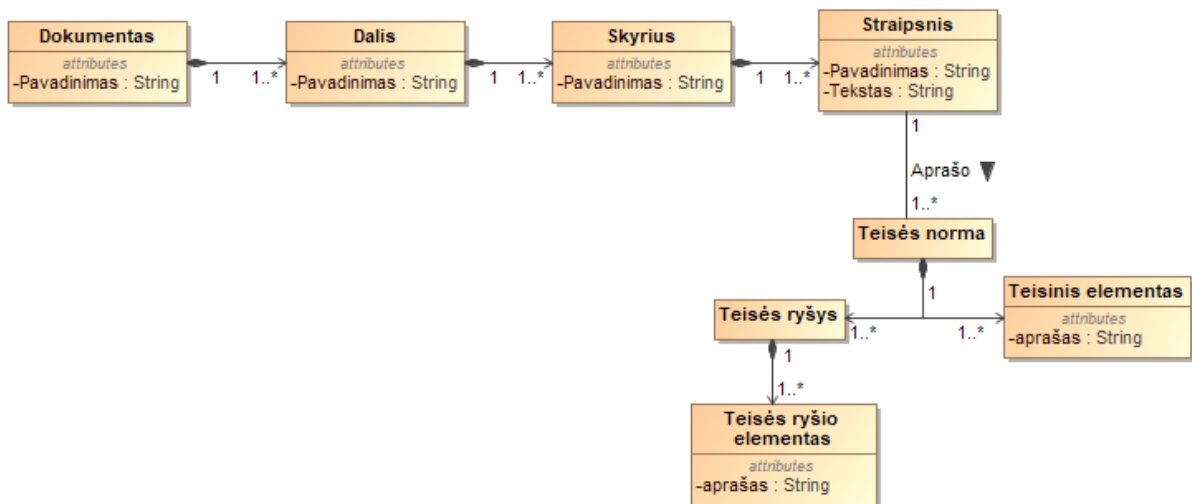
2.11 pav. Protégé įrankio langas

Lietuvos Respublikos kodeksai skaidomi į dalis, skyrius ir straipsnius.

Dokumentas gali turėti vieną ar daugiau dalių, kiekvieną dalį sudaro vienas ar daugiau skyrių, o skyrius turi vieną ar daugiau straipsnių. Kai kuriuose kodeksuose skyriai yra skaidomi į skirsnius, tačiau tai tiriamajame darbe akcentuojama nebus.

Kiekviename straipsnyje apibrėžiama teisinė norma, kurią sudaro teisės ryšys ir teisės elementas. Teisės ryšys ir teisės elementas yra dvi ontologijų klasės, pagal kurias bus kuriamos taisyklės anotacijoms rasti.

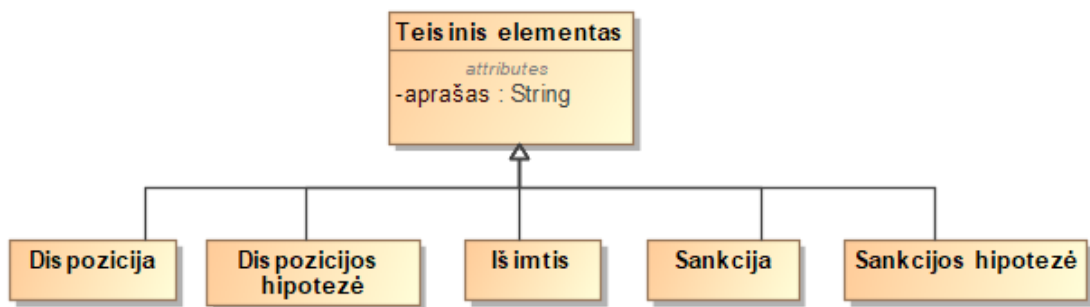
Teisės ryšio elementas detalizuojamas 2.16 pav., o teisės elementas 2.13 pav. Iš pateiktų modelių pastebima, kad „dispozicija“ priskiriama teisės elementui, bet „pareiga“ ir „teisė“ yra priskiriamos teisės ryšio elementui, tačiau remiantis pirmoje dalyje pateikta teisės normos sąvokų struktūra - dispoziciją sudaro pareigos ir teisės, ko pasekoje anotacijų taisyklės bus kuriamos abiem ontologijų klasėms.



2.12 pav. Dokumento esybių-ryšių modelis

Teisiniuose dokumentuose teisinį elementą sudaro: dispozicija, dispozicijos hipotezė, išimtis, sankcija, sankcijos hipotezė. Tai yra pagrindiniai teisiniai elementai. Realizuojamame sprendime hipotezė nebus skaidoma į atskiras dispozicijos ir sankcijos hipotezes.

Teisiniai elementai sudaro teisės normos visumą. Dispozicijos hipotezė ir sankcijos hipotezė nurodo faktines aplinkybes, kurioms atsiradus galioja teisės norma. Dispozicija yra pagrindinė ir privaloma teisės normos dalis, kurioje įtvirtinta elgesio taisyklė, subjektų teisės ir pareigos. Sankcija nurodo poveikio priemones, taikomas, kai nesilaikoma teisės normų reikalavimų.



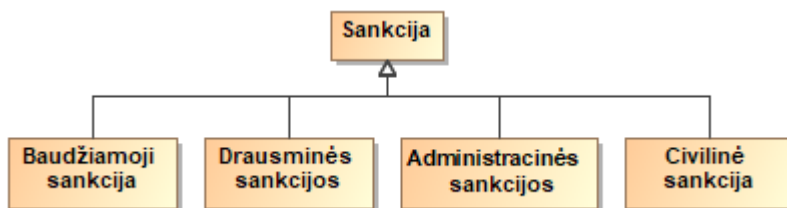
2.13 pav. Teisinio elemento klasifikacija

Sankcija nėra sudedamasis kiekvienos teisės normos teksto elementas, tai yra elementas, kuris būtinas kiekvienos normos loginės struktūros elementui.

Sankcija yra skirstoma į šias rūšis:

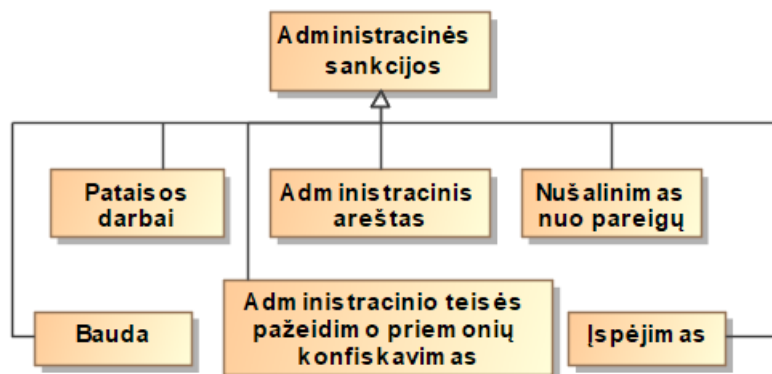
- baudžiamoji sankcija;
- civilinė sankcija;
- drausminė sankcija;
- administracinė sankcija.

Kadangi sankcijų rūšių spektras yra platus, todėl automatiniam sankcijų radimui bus aprašytos taisyklės tik administracinėms sankcijoms.



2.14 pav. Sankcijos klasifikacija

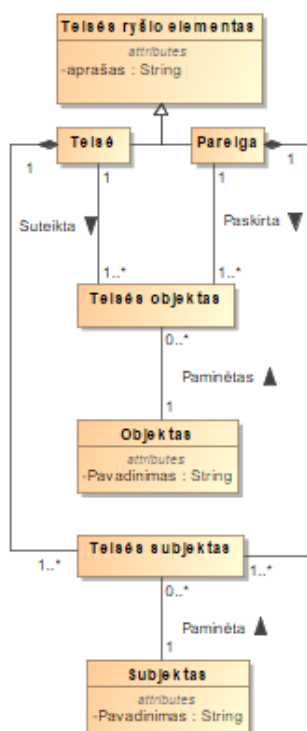
2.15 pav. pateikiama administracinės sankcijos detalizacija. Klasifikacijoje galima pastebėti, kad administracinė sankcija yra skaidoma į daug smulkesnių rūšių, tačiau po atliktos Administracinių nusižengimų kodekso straipsnių analizės, sprendimo įgyvendinimui pasirinkta labiausiai straipsniuose apibrėžta ir aiškiausiai pateikta administracinės sankcijos rūšis – bauda.



2.15 pav. Administracinės sankcijos klasifikacija

Teisės ryšio elementas yra ontologijos klasė. Šį elementą sudaro pareigos ir teisės, kurios yra įtraukiamos į dispoziciją, t. y. teisinį elementą, kitaip įvardinant – kitą ontologijos klasę. Pareigos ir teisės yra susiję su teisiniu objektu bei teisiniu subjektu.

Analizuojant kodeksų straipsnių struktūrą pastebėta, kad prie teisių arba pareitų turi būti norodyta kokiam subjektui už kokį objektą priskiriami teisės ryšio elementai. Kiekviena teisė ir pareiga gali turėti nuo 1 iki daug subjektų ir objektų.



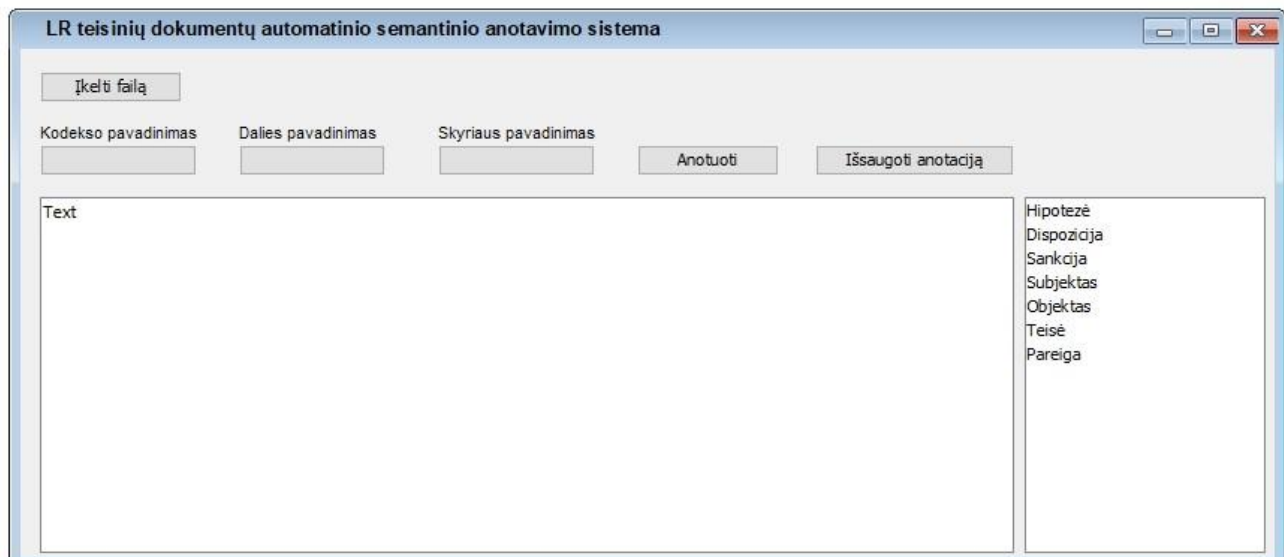
2.16 pav. Teisinio ryšio elemento klasifikacija

2.3 Teisinių dokumentų automatinio semantinio anotavimo sistemos naudotojo sąsajos modelis

Kuriamoje automatinio semantinio anotavimo sistemoje bus vienas pagrindinis langas, kuriame sistemos naudotojas galės atlikti pagrindines funkcijas:

- įkelti failą, pasirinktą iš kompiuterio;
- pasirinkti iš sąrašo LR kodekso pavadinimą;
- pasirinkti iš sąrašo LR kodekso dalies pavadinimą;
- pasirinkti iš sąrašo LR kodekso dalies skyriaus pavadinimą;
- anotuoti t. y. paspaudus ant šios funkcijos bus atliekamas automatinis straipsnio anotavimas;
- saugoti anotaciją – tai anotacijos parsisiuntimas JSON formatu.

Pagrindiniame lange išvedamas suanotuotas straipsnio tekstas, bei rastų anotacijų legenda, kurioje pažymima, kokia spalva ar kitu ženklinimu išskiriama kiekviena anotacija. Ši dalis reikalinga, kad sistemos naudotojas aiškiai ir suprantamai suvoktų straipsnyje atpažintas teisės normos sąvokas.



2.17 pav. Sistemos navigacijos planas

2.4 Lietuvos Respublikos teisinių dokumentų automatinio semantinio anotavimo formalus sprendimo aprašas

Formalus sprendimo aprašas parengtas pagal pirmos eilės predikatų logikos taisykles, aprašant anotacijų radimo taisykles, kurios buvo sudaromos po atliktos straipsnių analizės.

Taisyklės kuriamos išanalizavus Administracinių nusižengimų ir Darbo kodeksus.

DISPOZICIJA (1)

Kiekviena dispoziciją d_1 susideda iš dviejų šalia esančių žodžių junginio „užtraukia baudą“.

Z_1 – aibė, kurią sudaro žodžių junginys „užtraukti baudą“, žodžių gramatinės kategorijos: [„užtraukti“, „Vgmp3p--n--ni-“], [„bauda“, „Ncfsan-“].

D – dispozicijų aibė

$$\forall z_1 \in Z_1 \Rightarrow \exists d_1 \in D. \text{dispozicija}_1(d_1, z_1) \quad (2.1)$$

Tokia dispozicijos taisyklė pasirinkta, kadangi Administracinių nusižengimų kodekse į paiešką įvedus raktažodžius „užtraukia baudą“, buvo rasti 1504 paieškos rezultatai. Remiantis šiais skaičiais galima spręsti, kad šio kodekso straipsniuose tokio formato dispozicija atsikartoja dideliame kiekyje straipsnių. Darbo kodekse tokio tipo dispozicija nebuvo pastebėta.

Pavyzdys pateikiamas 8.1.1. priede.

DISPOZICIJA (2)

Kiekviena dispozicija d_2 susideda iš keturių šalia esančių žodžių jungio „užtraukia įspėjimą arba baudą“.

Z_2 – aibė, kurią sudaro žodžių junginys „užtraukti įspėjimą arba baudą“, žodžių gramatinės kategorijos: [„užtraukti“, „Vgmp3p--n--ni-“], [„bauda“, „Ncfsan-“], [„arba“, „Cg“], [„įspėjimas“, „Ncmsnn-“]

D – dispozicijų aibė

$$\forall \check{z}_2 \in Z_2 \Rightarrow \exists d_2 \in D. \text{dispozicija}_2(d_2, \check{z}_2) \quad (2.2)$$

Ši taisyklė Administracinių nusižengimų kodekso straipsniuose atsikartoja 146 kartus, aprašyta taisyklė netinkama Darbo kodekso dispozicijoms rasti.

Pavyzdys pateikiamas 8.1.3. priede.

DISPOZICIJA (3)

Kiekviena dispozicija d_3 yra *tekstas*, kuris prasideda žodžiu „turi“ ir tęsiasi iki taško, kurio gramatinės kategorijos [[".", "Tp"]].

D – dispozicijų aibė

T_1 – straipsnio tekstas nuo žodžio „turi“ iki taško, žodžio „**turi**“ gramatinės kategorijos [["turėti", "Vgmp3p--n--ni-"]]

$$\forall t_1 \in T_1 \Rightarrow \exists d_3 \in D. \text{dispozicija}_3(d_3, t_1) \quad (2.3)$$

Pavyzdys pateikiamas 8.1.4. priede.

DISPOZICIJA (4)

Kiekviena dispozicija d_4 yra *tekstas*, kuris prasideda žodžiu „privalo“ ir tęsiasi iki taško, kurio gramatinės kategorijos [[".", "Tp"]].

D – dispozicijų aibė

T_2 – straipsnio tekstas nuo žodžio „privalo“, kurio gramatinės kategorijos [["privalėti", "Vgmp3p--n--ni-"]]

$$\forall t_2 \in T_2 \Rightarrow \exists d_4 \in D. \text{dispozicija}_4(d_4, t_2) \quad (2.4)$$

Pavyzdys pateikiamas 8.1.5. priede.

DISPOZICIJA (5)

Kiekviena dispozicija d_5 yra *tekstas*, kuris prasideda žodžiu „gali“ ir tęsiasi iki taško, kurio gramatinės kategorijos [[".", "Tp"]].

D – dispozicijų aibė

T_3 – straipsnio tekstas nuo žodžio „gali“, kurio gramatinės kategorijos [["galėti", "Vgmp3p--n--ni-"]]

$$\forall t_3 \in T_3 \Rightarrow \exists d_5 \in D. \text{dispozicija}_5(d_5, t_3) \quad (2.5)$$

Pavyzdys pateikiamas 8.1.6. priede.

DISPOZICIJA (6)

Kiekviena dispozicija d_6 yra *tekstas*, kuris prasideda žodžiu „negali“ ir tęsiasi iki taško, kurio gramatinės kategorijos [[".", "Tp"]].

D – dispozicijų aibė

T_4 – straipsnio tekstas nuo žodžio „negali“, kurio gramatinės kategorijos [["negalėti", "Vgmp3p--y--ni-"]] iki taško [[".", "Tp"]].

$$\forall t_4 \in T_4 \Rightarrow \exists d_6 \in D. \text{dispozicija}_6(d_6, t_4) \quad (2.6)$$

Pavyzdys pateikiamas 8.1.7. priede.

TEISĖ

Kiekviena teisė r susideda iš žodžių jungio „turi teisę“ ir veiksmažodžio bendraties.

Z_3 - sąvokų aibė, kurią sudaro žodžių junginys „turi teisę“. Žodžių gramatinės kategorijos [["turėti", "Vgmp3p--n--ni-"]] [["teisė", "Ncfsan-"]]

Z_4 – sąvokų aibė, kurią sudaro veiksmažodžių bendratys, jų gramatinės kategorijos [["Vgi-----n--n--"]]

R – teisių aibė

$$\forall \check{z}_3 \in Z_3, \forall \check{z}_4 \in Z_4. \text{šalia}(\check{z}_3, \check{z}_4) \Rightarrow \exists r \in R. \text{teisė}(r, \check{z}_3, \check{z}_4) \quad (2.7)$$

Pavyzdys pateikiamas 8.1.8. priede.

PAREIGA

Kiekviena pareiga p susideda iš žodžio „privalėti“ ir veiksmažodžio bendraties.

Z_5 - sąvokų aibė, kurią sudaro žodis **privalėti**. Žodžio gramatinės kategorijos [["privalėti", "Vgmp3p--n--ni-"]]

Z_6 – sąvokų aibė, kurią sudaro veiksmažodžių bendratys, jų gramatinės kategorijos [["Vgi-----n--n--"]]

P – pareigų aibė

$$\forall \check{z}_5 \in Z_5, \forall \check{z}_6 \in Z_6. \text{šalia}(\check{z}_5, \check{z}_6) \Rightarrow \exists p \in P. \text{pareiga}(p, \check{z}_5, \check{z}_6) \quad (2.8)$$

Pavyzdys pateikiamas 8.1.9. priede.

SANKCIJA

Kiekvieną sankciją s yra **tekstas**, kuris prasideda žodžiu „nuo“, teksto viduryje turi žodį „iki“ ir pasibaigia žodžiu „eurų“.

Z_7 – sąvokų aibė, kurioje yra žodis **nuo**. Žodžio gramatinės kategorijos [["nuo", "Sgg"]]

Z_8 – sąvokų aibė, kurioje yra žodis **iki**. Žodžio gramatinės kategorijos [["iki", "Sgg"]]

Z_9 - sąvokų aibė, kurioje žodis **euras**. Žodžio gramatinės kategorijos [["euras", "Ncmpgn-"]]

S – sankcijų aibė

T_5 – sakinio dalis

$$\forall \check{z}_7 \in Z_7, \forall \check{z}_8 \in Z_8, \forall \check{z}_9 \in Z_9, \forall t_5 \in T_5. \text{tekstas su žodžiais}(\check{z}_7, \check{z}_8, \check{z}_9, t_5) \Rightarrow \exists s \in S. \text{sankcija}(s, \check{z}_7, \check{z}_8, \check{z}_9, t_5) \quad (2.9)$$

Pavyzdys pateikiamas 8.1.10. priede.

HIPOTEZĖ

Kiekviena hipotezė h yra *tekstas*, kuris prasideda nuo straipsnio punkto pradžios ir tęsiasi iki dispozicijos d .

D – dispozicijų aibė

H – hipotezių aibė

T_6 – sakinio dalis nuo straipsnio punkto pradžios iki dispozicijos.

$$\forall t_6 \in T_6, \forall d \in D. \text{tekstas_iki_dispozicijos}(d, t_6) \Rightarrow \exists h \in H. \text{hipotezė}(h, t_6) \quad (2.10)$$

Pavyzdys pateikiamas 8.1.2. priede.

SUBJEKTAS (1)

Kiekvienas subjektas c_1 yra tekstas, kuris yra tarp dispozicijos d ir sankcijos s .

D – dispozicijų aibė

S – sankcijų aibė

C – subjektų aibė

T_7 – sakinio dalis tarp dispozicijos ir sankcijos

$$\forall d \in D, \forall s \in S, \forall t_7 \in T_7. \text{tekstas_esantis_tarp}(d, s, t_7) \Rightarrow \exists c_1 \in C. \text{subjektas}_1(c_1, t_7) \quad (2.11)$$

SUBJEKTAS (2)

Kiekvienas subjektas c_2 yra žodžiai, kurie turi lemą – asmuo.

C – subjektų aibė

Z_{10} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „asmuo“

$$\forall \check{z}_{10} \in Z_{10} \Rightarrow \exists c_2 \in C. \text{Subjektas}_2(c_2, \check{z}_{10}) \quad (2.12)$$

Pavyzdys pateikiamas 8.1.11. priede.

SUBJEKTAS (3)

Kiekvienas subjektas c_3 yra žodžiai, kurie turi lemą – darbuotojas.

Z_{11} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „darbuotojas“

C – subjektų aibė

$$\forall \check{z}_{11} \in Z_{11} \Rightarrow \exists c_3 \in C. \text{subjektas}_3(c_3, \check{z}_{11}) \quad (2.13)$$

Pavyzdys pateikiamas 8.1.12. priede.

SUBJEKTAS (4)

Kiekvienas subjektas c_4 yra žodžiai, kurie turi lemą – darbdavys.

Z_{12} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „darbdavys“

C – subjektų aibė

$$\forall \check{z}_{12} \in Z_{12} \Rightarrow \exists c_4 \in C. \text{subjektas}_4(c_4, \check{z}_{12}) \quad (2.14)$$

Pavyzdys pateikiamas 8.1.13. priede.

OBJEKTAS (1)

Kiekvienas objektas o_1 susideda iš trijų šalia esančių žodžių, kurie turi lemas: darbas, laikas, apskaita.

Z_{13} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „darbas“

Z_{14} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „laikas“

Z_{15} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „apskaita“

O – objektų aibė

$$\forall \check{z}_{13} \in Z_{13}, \forall \check{z}_{14} \in Z_{14}, \forall \check{z}_{15} \in Z_{15}. \textit{šalia}(\check{z}_{13}, \check{z}_{14}, \check{z}_{15}) \Rightarrow \exists o_1 \in O. \textit{objektas}_1(o_1, \check{z}_{13}, \check{z}_{14}, \check{z}_{15}) \quad (2.15)$$

Pavyzdys pateikiamas 8.1.14. priede.

OBJEKTAS (2)

Kiekvienas objektas o_2 susideda iš trijų šalia esančių žodžių, kurie turi lemas: nelaimingas, atsitikimas, darbas.

Z_{16} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „nelaimingas“

Z_{17} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „atsitikimas“

Z_{13} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „darbas“

O – objektų aibė

$$\forall \check{z}_{16} \in Z_{16}, \forall \check{z}_{17} \in Z_{17}, \forall \check{z}_{13} \in Z_{13}. \textit{šalia}(\check{z}_{16}, \check{z}_{17}, \check{z}_{13}) \Rightarrow \exists o_2 \in O. \textit{objektas}_2(o_2, \check{z}_{16}, \check{z}_{17}, \check{z}_{13}) \quad (2.16)$$

Pavyzdys pateikiamas 8.1.15. priede.

OBJEKTAS (3)

Kiekvienas objektas o_3 susideda iš dviejų šalia esančių žodžių, kurie turi lemas: darbas, užmokestis.

Z_{13} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „darbas“

Z_{18} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „užmokestis“

O – objektų aibė

$$\forall \check{z}_{13} \in Z_{13}, \forall \check{z}_{18} \in Z_{18}. \textit{šalia}(\check{z}_{13}, \check{z}_{18}) \Rightarrow \exists o_3 \in O. \textit{objektas}_3(o_3, \check{z}_{13}, \check{z}_{18}) \quad (2.17)$$

Pavyzdys pateikiamas 8.1.16. priede.

OBJEKTAS (4)

Kiekvienas objektas o_4 susideda iš dviejų šalia esančių žodžių, kurie turi lemas: darbas, sąlyga.

Z_{13} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „darbas“

Z_{19} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „sąlyga“

O – objektų aibė

$$\forall \check{z}_{13} \in Z_{13}, \forall \check{z}_{19} \in Z_{19}. \textit{šalia}(\check{z}_{13}, \check{z}_{19}) \Rightarrow \exists o_4 \in O. \textit{objektas}_4(o_4, \check{z}_{13}, \check{z}_{19}) \quad (2.18)$$

Pavyzdys pateikiamas 8.1.17. priede.

OBJEKTAS (5)

Kiekvienas objektas o_5 yra žodis, kuris turi lemą „bauda“.

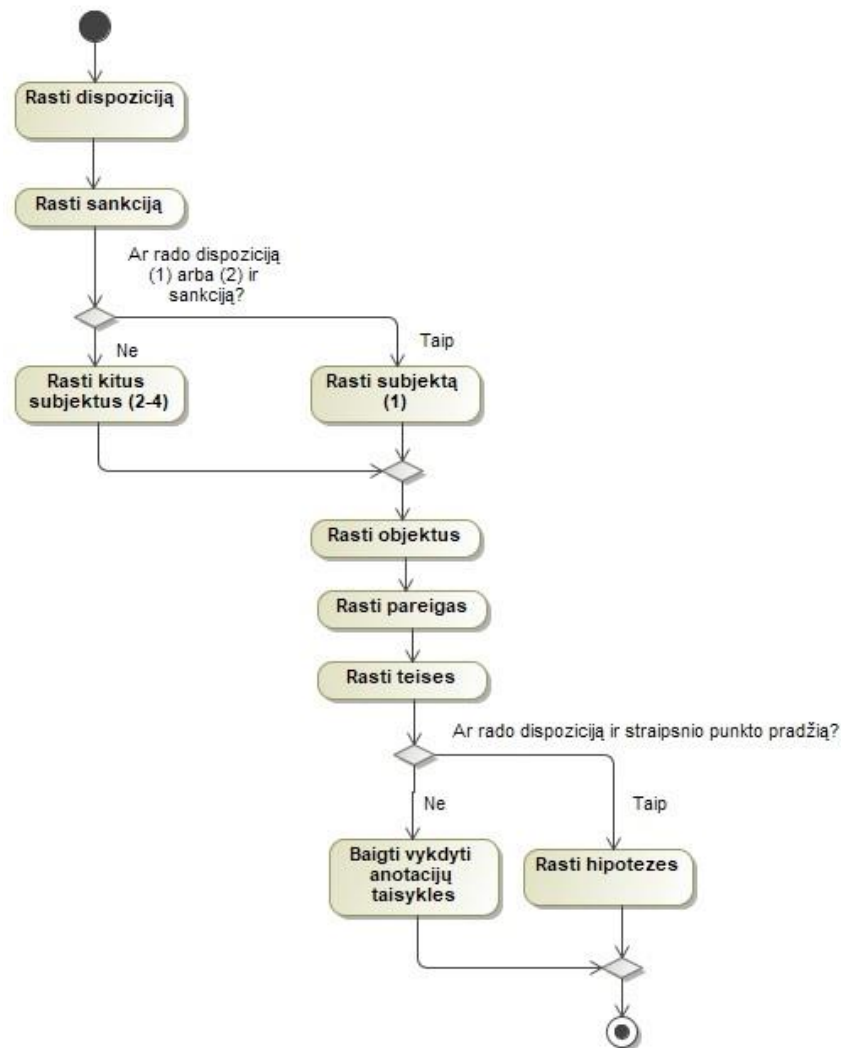
Z_{20} – sąvokų aibė, kurią sudaro žodžiai, turintys lemą „bauda“

O – objektų aibė

$$\forall \check{z}_{20} \in Z_{20} \Rightarrow \exists o_5 \in O. \textit{objektas}_5(o_5, \check{z}_{20}) \quad (2.19)$$

Pavyzdys pateikiamas 8.1.18. priede.

2.18 pav. atvaizduojamas taisyklių taikymo principas, kuriame pateikiama anotacijų radimo tvarka. Anotacijų taisyklės vykdomos viena po kitos iš eilės. Subjektas (1) ir hipotezė turi išskirtines taisykles, kadangi jų radimas priklauso nuo to ar tekste buvo rastos kitos anotacijos, kaip: dispozicija ir sankcija.



2.18 pav. Taisyklių ieškojimo principas

2.5 Sprendimo reikalavimų ir projekto apibendrinimas

1. Reikalavimų specifikavimo dalyje buvo sudaryti ir detalizuojami pagrindiniai LR teisinių dokumentų automatinio semantinio anotavimo sistemos reikalavimai:

- 1.1 LR teisinių dokumentų įkėlimas – reikalavimas papildomai detalizuotas jam priklausančiais kitais reikalavimais;
- 1.2 LR teisinių dokumentų automatinis semantinis anotavimas – reikalavimas papildomai detalizuotas jam priklausančiais kitais reikalavimais;
- 1.3 Anotuoto dokumento parsisiuntimas;
- 1.4 Grafinės sąsajos reikalavimai – reikalavimas papildomai detalizuotas jam priklausančiais kitais reikalavimais;

- 1.5 Teisinių dokumentų pavadinimų, dalių ir skyrių nurodymas.

Remiantis šiais sistemos reikalavimais bus atliekama sprendimo realizacija.

2. Sudarytas panaudojimo atvejų modelis, kuriame aiškiai atvaizduojamos naudotojui galimos funkcijos: įkelti failą; nurodyti kodekso, dalies ir skyriaus pavadinimus; parsisiūti anotaciją. Kiekvienas panaudojimo atvejis detalizuotas atskira lentele ir veiklos diagrama.
3. Sudarant esybių ryšių modelį išdetalizuota teisinių dokumentų struktūra, pagrindiniai teisinių normų elementai ir ryšiai. Pagal sudarytą modelį matome, kad teisinė norma susideda iš dviejų ontologijų klasių: teisinio elemento ir teisės ryšio elemento. Remiantis šiomis ontologijų klasėmis bus kuriamos anotacijų taisyklės, tačiau pilnai realizuoti esybių-ryšių modelio nepavyks, kadangi kuriamame sprendime teisiniai dokumentai anotuojami straipsnio lygyje ir ontologijų klasės nebus atskirai išskiriamos.
4. Navigavimo dalyje pateikiama planuojamo sprendimo vizualizacija t. y. pagrindinis sistemos langas ir jame esančios funkcijos.
5. Formalus sprendimo aprašas paremtas pirmos eilės predikatų logikos taisyklėmis, kurios skirtos apibrėžti kiekvienos anotacijos radimo sąlygas. Anotacijų taisyklės – tai bendrinės taisyklės, skirtos pagrindinėms teisės normos sąvokoms rasti. Taisyklės apašytos po atliktos Administracinių nusižnegimų kodekso ir Darbo kodekso straipsnių analizės.

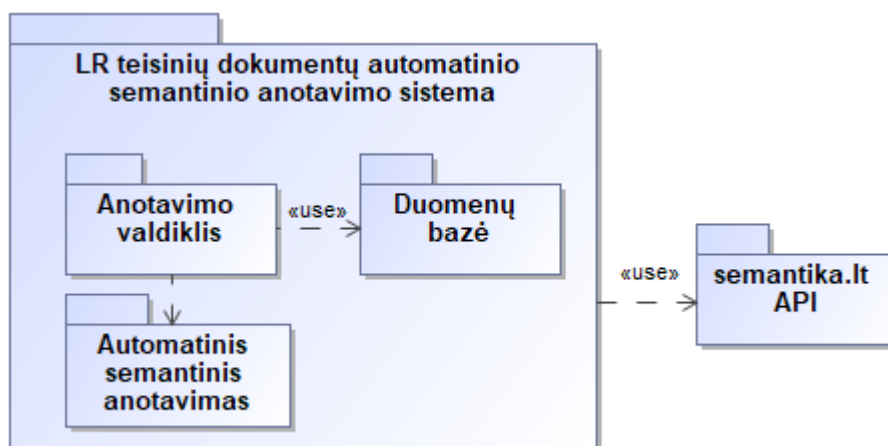
3. LIETUVOS RESPUBLIKOS TEISINIŲ DOKUMENTŲ AUTOMATINIO SEMANTINIO ANOTAVIMO SPRENDIMO REALIZACIJOS PROJEKTAS

Teisinių dokumentų automatinio semantinio anotavimo sprendimo realizacija susideda iš šių dalių:

- sistemos loginė architektūra;
- sistemos detalizuota loginė architektūra;
- duomenų bazės modelis;
- diegimo modelis.

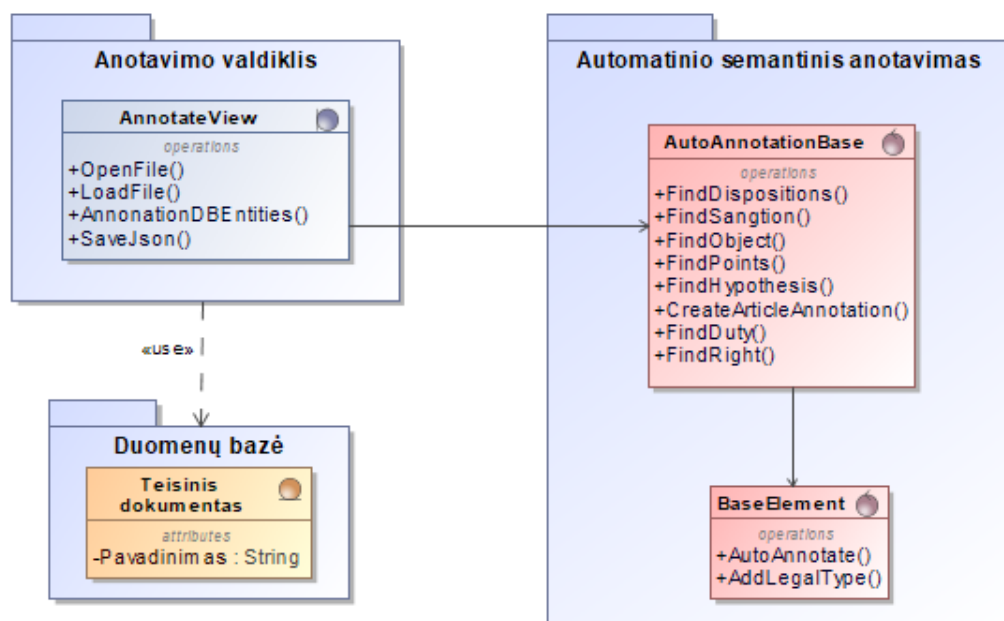
3.1 Teisinių dokumentų automatinio semantinio anotavimo sistemos loginė architektūra

Sistemos loginėje architektūroje atvaizduojamas pagrindinis, kuriamo sprendimo paketas – LR teisinių dokumentų automatinio semantinio anotavimo sistema, kuri turi sąsają su semantika.lt API. Duomenų bazėje yra saugomi kodeksų, dalių ir skyrių pavadinimai, o iš semantika.lt API yra grąžinamas struktūrizuotas tekstas JSON formatu.



3.1 pav. Sistemos loginė architektūra

Detalizuotoje sistemos loginėje architektūroje pateikiami sistemos pagrindinių klasių metodai. Anotavimo valdiklį sudaro viena AnnotateView klasė, kurioje metodas ReadFile() yra skirtas failo įkėlimui ir nuskaitymui. Metodas LoadFile() skirtas kreipimuisi į semantika.lt saityno paslaugą ir iškviečia automatinio semantinio anotavimo klasę AutoAnnotationBase. AutoAnnotationBase klasėje aprašyti metodai, skirti anotacijoms rasti. BaseElement klasėje metodas AutoAnnotate() nusplavina ir pažymi anotacijas tekse, o metodas AddLegalType() sudeda anotacijas į JSON. Metodas SaveJson() atlieka suanotuoto failo parsisiuntimą JSON formatu. Straipsnio meta duomenys, kurios naudotojas nuorodo pagrindiniame sistemos lange, paimami iš lokalios duomenų bazės.



3.2 pav. Detalizuota sistemos loginė architektūra

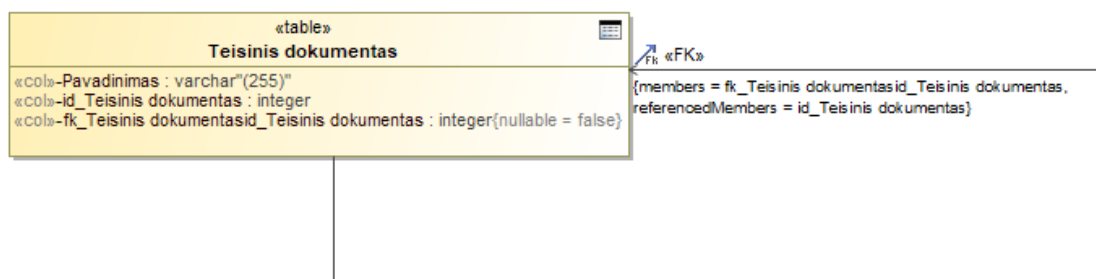
Žemiau pateiktoje lentelėje paaiškinta metodų, esančių AutoAnnotationBase() klasėje, paskirtis.

3.1.lentelė. Metodų paskirtis

Metodas	Paskirtis
FindDisposition()	Dispozicijų taisyklės
FindSangtion()	Sankcijos taisyklė
FindObject()	Subjektų ir objektų taisyklės
FindDuty()	Pareigos taisyklė
FindRight()	Teisės taisyklė
FindPoints()	Sakinio pabaigos ieškojimo taisyklė
FindHyphotesis()	Hipotezės ieškojimo taisyklė
CreateArticleAnnotation()	Straipsnio pavadinimo ieškojimo taisyklė

3.2 Duomenų bazės modelis

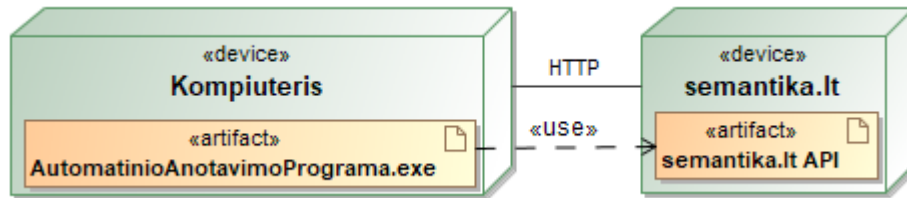
Sistema siejama su lokalia duomenų baze, kurioje yra rekursinė lentelė, susidedanti iš kodeksų, dalių ir skyrių pavadinimų. Šie duomenys reikalingi nustatyti straipsnio, įkeliamo į automatinio semantinio anotavimo sistemą, meta duomenis.



3.3 pav. Duomenų bazės modelis

3.3 Teisinių dokumentų automatinio semantinio anotavimo programos diegimas

„AutomatinioAnotavimoPrograma.exe“ – teisinių dokumentų automatinio semantinio anotavimo taikomoji programa, kuri įdiegiama į kompiuterį įkėlus exe failą. Naudotojui norint įsidiegti taikomąją programą kompiuteryje turi būti Windows operacinė sistema ir interneto ryšys, kad sistema galėtų kreiptis į semantika.lt API.



3.4 pav. Teisinių dokumentų automatinio semantinio anotavimo programos diegimas

3.4 Realizacijos projekto apibendrinimas

1. Kuriant sistemos realizacijos projektą, pateikiami pagrindiniai modeliai: sistemos loginė architektūra, sistemos detalizuota loginė architektūra, duomenų bazės modelis ir diegimo modelis.
2. Loginės sistemos architektūros paketuose išskiriamos pagrindinės automatinio anotavimo sistemos dalys, sąsaja su lokalia duomenų baze bei semantika.lt API. Detalizuotoje sistemos loginėje architektūroje atvaizduojamos automatinio anotavimo sistemos klasės ir metodai.
3. Duomenų bazės modelis sudaromas iš vienos rekursinės duomenų lentelės, kurioje saugomi kodeksų, dalių ir skyrių pavadinimai.
4. Automatinio anotavimo programa yra exe failas, kuriam nėra reikalingos papildomos diegimo sąlygos, svarbiausia kompiuteryje būtų Windows operacinė sistema ir interneto ryšys. Kuriamas sprendimas turi sąsają su semantika.lt.

4. TEISINIŲ DOKUMENTŲ AUTOMATINIO SEMANTINIO ANOTAVIMO SISTEMOS REALIZACIJA IR TESTAVIMAS

4.1 Teisinių dokumentų automatinio semantinio anotavimo sistemos realizacija ir veikimo aprašas

Sprendimo realizacija

Sprendimo realizacija – Lietuvos Respublikos teisinių dokumentų automatinio semantinio anotavimo sistema.

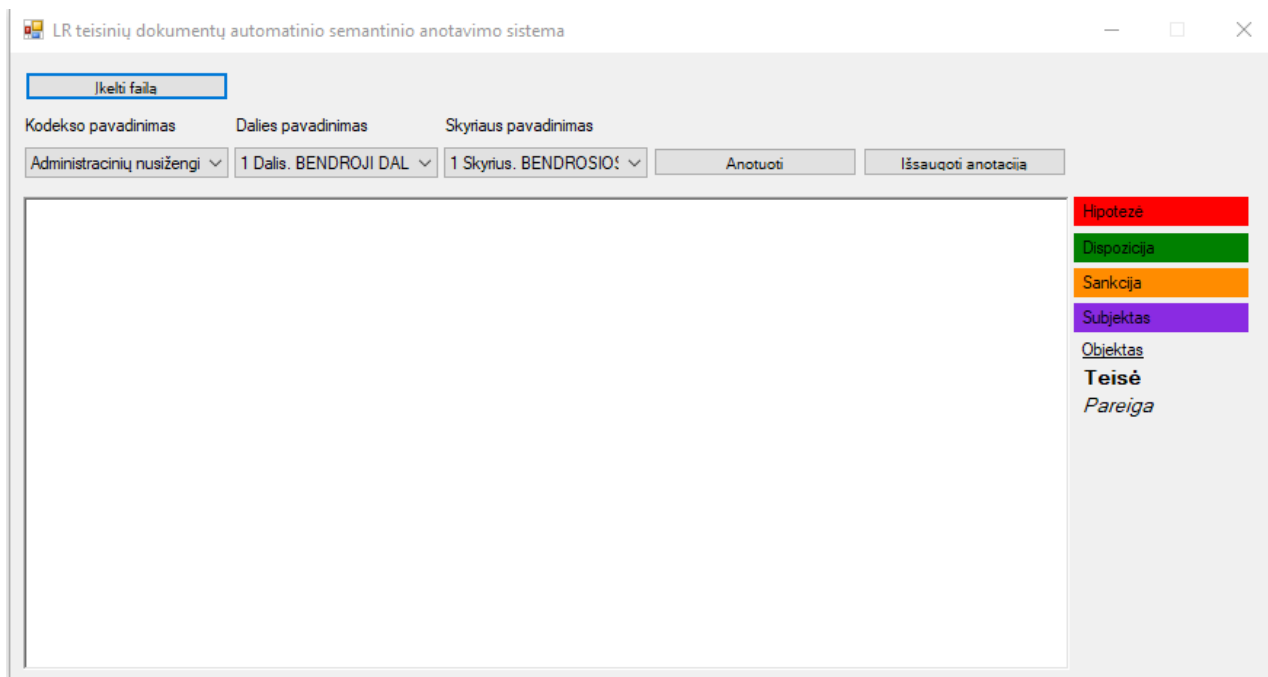
Teisinių dokumento automatinio semantinio anotavimo sistema – tai programa sukurta naudojant C# programavimo kalbą. Sprendimui įgyvendinti naudojamas virtualus duomenų apsikeitimo metodas t. y. žodžių morfemos, lemos, bei segmentai gaunami kreipiantis į semantika.lt saityno paslaugą (angl. web service).

Taip pat sistema turi sąsają su lokalia duomenų baze „SQL server express“, duomenų bazėje yra saugomi kodeksų pavadinimai, kodeksų dalių pavadinimai ir kodeksų skyrių pavadinimai. Šie įrašai yra susieti tarpusavyje, vadinasi sistemoje pasirinkus kodekso pavadinimą išfiltruojamos to kodekso dalys, o pagal pasirinktą dalį – skyriai.

Sistemos veikimo principas

Įkėlus failą į sistemą ir paspaudus funkciją „Anotuoti“, pirmiausiai kreipiamasi į semantika.lt saityno paslaugą, iš kurios grąžinamas struktūrizuotas ir su visomis žodžių lemomis bei žodžių gramatinėmis kategorijomis failas. Atlikus šią dalį programa pirmiausiai sužymi visus žodžius pagal metoduose aprašytas žodžių gramatinės kategorijas, pritaiko kitas anotacijų taisykles, nusplavina anotacijas tekste ir išsaugo JSON faile.

Sistema turi vieną pagrindinį langą, kuriame vartotojas gali matyti automatiškai suanotuoto teksto rezultatą. Siekiant, kad vartotojas lengviau atpažintų rastas anotacijas – anotacijų išskyrimo paaiškinimai pateikiami sistemos dešiniajame šone. Žemiau esančiame 4.1 pav. pateikiama vizualizacija, kaip atrodo sukurta sistema, kuria gali naudotis vartotojas.



4.1 pav. LR teisinių dokumentų automatinio semantinio anotavimo sistemos pagrindinis langas

4.1 lentelėje aprašytos realizuoto sprendimo funkcijos.

4.1 lentelė. Sukurto sprendimo pagrindinės funkcijos

Nr.	Funkcija	Veikimo principas
1	Įkelti failą	Paspaudus ant šio mygtuko atidaromas langas, kuriame iš kompiuterio pasirenkamas failas. Įkeliamo straipsnio failo formatas turi būti .txt formato.
2	Kodekso pavadinimas	Iš sąrašo pasirenkamas kodekso pavadinimas, kuriam priklauso įkeltas straipsnis.
3	Dalies pavadinimas	Pagal prieš tai pasirinktą kodekso pavadinimą, sistema atrenka dalis ir skyrius priklausančius kodeksui. Iš sąrašo pasirenkama kodekso dalis.
4	Skyriaus pavadinimas	Iš sąrašo pasirenkamas kodekso dalies skyriaus pavadinimas, kuriam priklauso įkeltas straipsnis. Skyriai išfiltruojami atitinkamai pagal pasirinktą kodekso dalį.
5	Anotuoti	Paspaudus „Anotuoti“ sistema pirmiausiai kreipiasi į sematinka.lt saityno paslaugą, iš kurios grąžinamas suskaidytas tekstas į segmentus, sakinius ir pastraipas, bei grąžinamos žodžių lemos ir morfologijos. Atlikus šį žingsnį programa pagal aprašytas anotacijų taisykles sužymi tekste visus žodžius pagal morfologijas ir tada pritaiko metoduose aprašytas anotacijų taisykles. Kiekviena anotacija yra išskiriama spalvomis, pabraukimu arba kitokiu šriftu bei grąžinama į .json formatą. Automatiškai suanotavus straipsnį, rezultatas išvedamas sistemos pagrindiniame lange, kuriame sistemos naudotojas gali atskirti rastas anotacijas pagal jų žymėjimą.
6	Išsaugoti anotaciją	Naudotojas turi galimybę suanotuotą straipsnį išsaugoti pasirinktoje kompiuterio lokacijoje .json formatu.

Json failo struktūra

Automatinio semantinio anotavimo sistemos sugeneruotas .json failas įkeliamas į kitą anotavimo sistemą, kuri aprašyta „Teisinių dokumentų automatizuoto semantinio anotavimo galimybių plėtimas“ tiriamjame darbe. Abi sistemos turi vienodą JSON struktūrą.

Žemiau pateikiama .json failo struktūra, pilna failo struktūra pateikiama 8.2. priede.

- „document“ – šioje dalyje išsaugoma įkelto straipsnio meta informacija: kodekso pavadinimas „name“, kodekso dalis „part“ ir kodekso dalies skyrius „section“. Šiuos duomenis vartotojas nurodo automatinio semantinio anotavimo sistemoje, pasirinkdamas reikšmes.

```
{
  "document": {
    "name": "Darbo kodeksas",
    "part": "2 Dalis. INDIVIDUALIEJI DARBO SANTYKIAI ",
    "section": "2 Skyrius. DARBO SUTARTIES SĄVOKA IR DARBO SUTARTIES SĄLYGOS "
  },
```

4.2 pav. JSON failo, document dalies struktūra

- „annotations“ – dalis išskiriama į kitas tris dalis:
 - „version“ : „lexv1.0“ – saugoma struktūrizuota teksto informacija, išskaidyta į segmentus „seg“, sakinius „s“ ir pastraipas „p“. Segmentai atitinka struktūrą [14,1], pirmas skaičius yra žodžio pirmas indeksas, antras skaičius nurodo žodžių skaičių segmente.

```
  "annotations": [
    {
      "version": "lexv1.0",
      "annotation": {
        "seg": [ ],
        "s": [ ],
        "p": [ ]
      }
    },
```

4.3 pav. JSON failo, annotations dalies struktūra

- „version“ : „morphologyv1.0“ – saugomos žodžių lemos ir morfologijos.

```
    {
      "version": "morphologyv1.0",
      "annotation": {
        "msd": [ ],
        "stem": [ ]
      }
    },
```

4.4 pav. JSON failo, version dalies struktūra

- „version“ : „teis1.0beta“ – saugomi rastų anotacijų segmentai, pagal kuriuos tekste yra pažymimi suanotuoti žodžiai. Dalis „type“ yra anotacijos pavadinimas, dalis „ref“ yra nuorodos į anotacijų segmentus.

```
{
  "version": "teis1.0beta",
  "annotation": {
    "legal": [
      {
        "type": "disposition",
        "ref": [ ]
      },
      {
        "type": "subject",
        "ref": [ ]
      },
      {
        "type": "object",
        "ref": [ ]
      },
      {
        "type": "hypothesis",
        "ref": [ ]
      }
    ],
    "doc": [ ]
  }
}
```

4.5 pav. JSON failo, annotation dalies struktūra

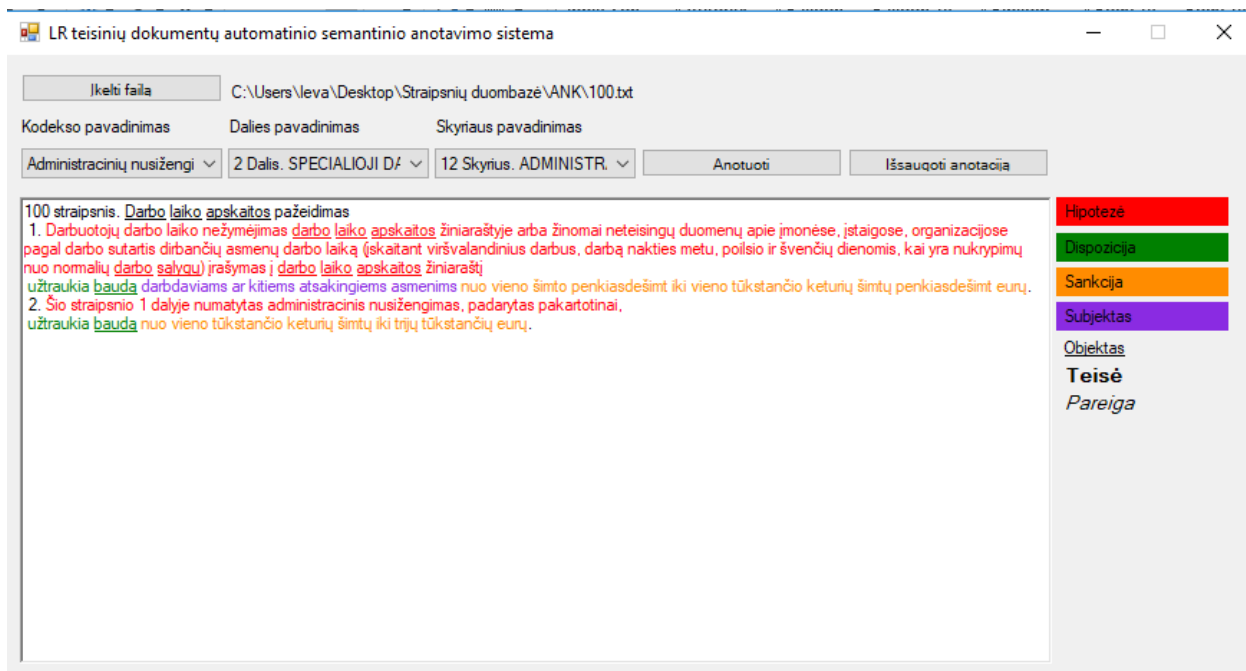
- „lang“ ir „enc“ dalyse yra grąžinama teksto kalba „lang“ bei koduotė „enc“ – UTF-8

```
],
"lang": { },
"enc": { }
}
```

4.6 pav. JSON failo struktūra

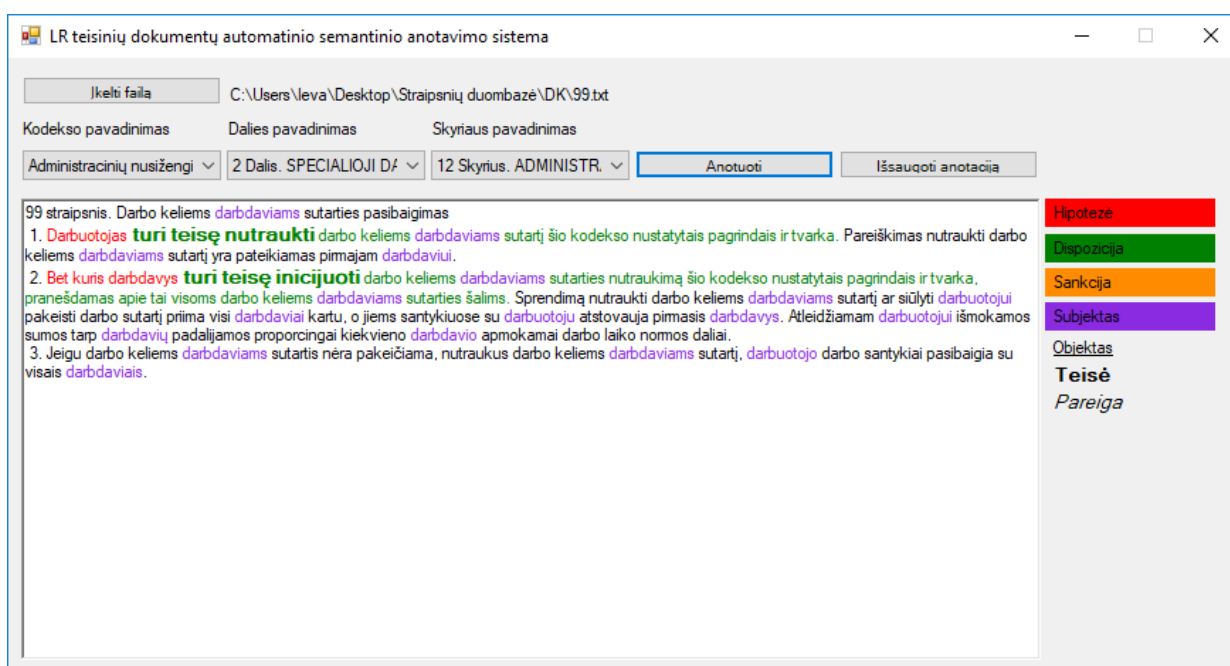
Žemiau pateikiami keli pavyzdžiai, kuriais norima parodyti, kaip atrodo suanotuoto straipsnio rezultatas.

Administracinių nusižengimų kodekso 2 dalies 12 skyriaus 100 straipsnis.



4.7 pav. Administracinių nusižengimų kodekso 100 straipsnio anotacija

Darbo kodekso 2 dalies 6 skyriaus 99 straipsnis.



4.8 pav. Darbo kodekso 99 straipsnio anotacija

4.2 Teisinių dokumentų automatinio semantinio anotavimo sistemos testavimas

Atliekant testavimą kiekvienai anotavimo taisyklei buvo pasirenkami konkretūs straipsniai, kuriuose anotacijos turėjo būti atpažintos, siekiant įvertinti ar programa taisyklės pritaiko korektiškai.

Testavimas vyko etapais, lygiagrečiai su sprendimo kūrimu. Testavimo proceso žingsniai:

- aprašoma nauja anotacijos taisyklė ir įkeliamas straipsnis į sistemą;

- suskaičiuojama kiek anotacijų sistema atpažino;
- suskaičiuojama kiek anotacijų turėjo atpažinti;
- 1-3 žingsniai kartojami su pasirinkta straipsnių imtimi;
- įvertinami gauti rezultatai, atitinkamai pagal tai taisyklės tobulinamos.

Testavimo duomenys – Administracinių nusižengimų kodekso ir Darbo kodekso straipsniai, kurie pasirinkti pagal anotacijoms aprašytas taisykles.

4.2 lentelė. Taisyklių testavimas Darbo kodekse

Anotacija	Taisyklės ir jų veikimo principai	Straipsniai	Rezultatai	
			Teisingai atpažinti	Turėjo atpažinti iš viso
Dispozicija	2.4 skyrius, 3 taisyklė	25, 28, 31, 35, 40, 50, 112, 140, 143, 158	24	24
	2.4 skyrius, 4 taisyklė	23, 31, 52, 90, 118, 155, 161, 189	14	18
	2.4 skyrius, 5 taisyklė	15, 25, 31, 35, 54, 58, 94, 156, 168	20	20
	2.4 skyrius, 6 taisyklė	31, 59, 68, 35, 153	7	7
Teisė	2.4 skyrius, 7 taisyklė	31, 40, 58, 80, 99	7	7
Pareiga	2.4 skyrius, 8 taisyklė	23, 24, 31, 115, 118, 153, 189, 197	15	15
Subjektas	2.4 skyrius, 13 taisyklė	58, 80, 86, 90, 96, 153	26	41
	2.4 skyrius, 14 taisyklė	28, 31, 90, 96	13	17
Objektas	2.4 skyrius, 15 taisyklė	115, 118, 120	14	14
	2.4 skyrius, 16 taisyklė	149, 157	2	2
	2.4 skyrius, 17 taisyklė	115, 118, 149, 150, 153, 156, 219	22	22
	2.4 skyrius, 18 taisyklė	23, 29, 33, 40	4	4
	2.4 skyrius, 19 taisyklė	217	3	3

4.3 lentelė. Taisyklių testavimas Administracinių nusižengimų kodekse

Anotacija	Taisyklės ir jų veikimo principai	Straipsniai	Rezultatai	
			Teisingai atpažinti	Turėjo atpažinti iš viso
Dispozicija	2.4 skyrius, 1 taisyklė	47, 84, 93, 95, 96, 99, 100, 104, 107, 133, 140	35	35
	2.4 skyrius, 2 taisyklė	49, 64, 72, 75, 90, 175, 187	12	12
Sankcija	2.4 skyrius, 9 taisyklė	47, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107	42	42
Subjektas	2.4 skyrius, 11 taisyklė	47, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107	26	26
	2.4 skyrius, 12 taisyklė	47, 97, 98, 100	12	14
Hipotezė	2.4 skyrius, 10 taisyklė	47, 95, 96, 99, 100, 104, 107, 133, 101, 102, 103, 105, 106, 107	39	39
Objektas	2.4 skyrius, 15 taisyklė	100	3	3
	2.4 skyrius, 16 taisyklė	97	4	4
	2.4 skyrius, 17 taisyklė	99	4	4
	2.4 skyrius, 18 taisyklė	100, 401	2	2
	2.4 skyrius, 19 taisyklė	96, 97, 98, 99, 100, 101, 102, 103, 104	28	28

Iš lentelėse pateiktų duomenų matome, kad rezultatai esantys stulpelyje „Teisingai atpažinti“ sutampa su rezultatais esančiais stulpelyje „Turėjo atpažinti iš viso“, išskyrus 12, 13, 14 taisykles. Tikrinant, kodėl paminėtų taisyklių anotacijos buvo rastos ne visos, pastebėta, kad sistema nepažymi subjektų hipotezėse, išvedant rezultatą ekrane, tačiau įrašo į JSON, to priežastis – kelių anotacijų persidengimas.

Testavimo rezultatai yra ganėtinai tikslūs, kadangi testuojant taisykles buvo pasirenkami straipsniai, kuriuose sistema tiksliai atpažintų anotacijas, priešingu atveju būtų sudėtinga įvertinti ar taisyklės pritaikomos korektiškai.

4.3 Sprendimo įgyvendinimo ir testavimo apibendrinimas

1. Sukurta LR teisinių dokumentų automatinio semantinio anotavimo sistema, kurioje aprašyta 19 taisyklių antacijoms rasti.
2. Lokalioje „SQL server express“ duomenų bazėje yra saugomi kodeksų, dalių ir skyrių pavadinimai.
3. Anotacijų taisyklės aprašytos naudojant žodžių gramatinės kategorijas bei lemas, kurios grąžinamos iš semantika.lt.
4. Sukurtas sprendimas pritaikytas 2 kodeksams: Administracinių nusižengimų kodeksui ir Darbo kodeksui, tačiau eksperimentas bus atliekamas ir Civiliniam kodeksui.
5. Sistema suanotuotą ir spalvomis išskaidytą tekstą išveda programos lange. Toks anotavimo rezultato pateikimas palengvina sistemos naudotojui atpažinti teisės normos sąvokas straipsnyje.
6. Sukurto sprendimo testavimui pasirinkti duomenys, su kuriais būtų ištestuotos visos aprašytos taisyklės. Pagal testavimo rezultatus galime priimti išvadą, kad taisyklės taikomos ganėtinai tiksliai, nes pastebimas mažas skirtumas tarp teisingai atpažintų ir pagal tai, kiek turėjo atpažinti pagal aprašytas taisykles reikšmių.

5. EKSPERIMENTINIS LIETUVOS RESPUBLIKOS TEISINIŲ DOKUMENTŲ AUTOMATINIO SEMANTINIO ANOTAVIMO SISTEMOS TYRIMAS

5.1 Eksperimento planas

Siekiant įvertinti pasirinkto Lietuvos Respublikos teisinių dokumentų automatinio semantinio anotavimo sistemos tikslumą, buvo atliekama eksperimentinė dalis. Šis dalis įgyvendinta remiantis eksperimento planu:

5.1.1. Pasirinkti Lietuvos Respublikos kodeksai:

- **Administracinių nusižengimų kodeksas.** Šis kodeksas pasirinktas, nes jame esantys straipsniai sudaryti iš griežtos struktūros. Galima aiškiai atskirti teisės normos sąvokas bei pastebėti jų dėsningumą straipsniuose. Anotacijų eiliškumo pavyzdys: hipotezė, dispozicija, objektas, subjektas, sankcija. Pagal šį kodeksą buvo sudarytos anotacijų taisyklės;
- **Darbo kodeksas.** Šis kodeksas pasirinktas dėl jo aktualumo ir svarbumo šiomis dienomis. Pagal šį kodeksą buvo papildytos jau sudarytos anotacijų taisyklės, kurios buvo sukurtos pagal Administracinių nusižengimų kodeksą. Administracinių nusižengimų kodekse, 2 dalyje, 12 skyriuje „Administraciniai nusižengimai, susiję su asmens darbo ir socialinėmis teisėmis“ yra straipsniai, kurie susiję su Darbo kodeksu. Pirmuose dviejuose punktuose įvardinti kodeksai yra pasirinkti neatsitiktinai, vienas iš tikslų – pagal rastas susijusių straipsnių objektų anotacijas, padaryti navigavimą tarp kodeksų. Navigavimas įgyvendinamas Manto Bataičio tiriamajame darbe „Teisinių dokumentų automatizuoto semantinio anotavimo galimybių tyrimas“;
- **Civilinis kodeksas [22].** Šis kodeksas pasirinktas, kadangi kodeksui nebuvo kuriamos specifinės taisyklės. Eksperimentinėje dalyje siekiama įvertinti, kiek sukurtas sprendimas atpažįsta anotacijų įkėlus straipsnį iš kodekso, pagal kurį nebuvo kuriamos taisyklės. Taip pat Civilinis kodeksas pasirinktas, kadangi jis yra vienas iš didžiausių kodeksų savo apimtimi, palyginti su kitais Lietuvos Respublikos kodeksais.

5.1.2. Kiekvienam kodeksui buvo pasirinkta skirtinga straipsnių imtis, atsižvelgiant į šiuos kriterijus:

- **straipsnio apimtis.** Kadangi straipsnių struktūrą sudaro konkretūs punktai, kurie turi aiškias numeracijas, todėl buvo siekiama pasirinkti straipsnius, kurie turėtų skirtingus punktų kiekius;
- **straipsnio kilmė.** Kiekvieną kodeksą sudaro skirtingos dalys, o kiekvieną kodekso dalį sudaro skirtingi skyriai. Analizuojant konkrečias kodeksų dalis ar skyrius tarpusavyje galima pastebėti tendenciją, kad gali skirtis straipsnių aprašymų stilistika skyriuose arba konkrečias anotacijas galima rasti tik atitinkamuose skyriuose. Įvertinant šią situaciją, eksperimentinėje dalyje siekiama pasirinkti straipsnius iš kodekso skirtingų dalių bei skyrių.

5.1.3. Pasirinkti straipsniai anotuojami dviem skirtingais būdais:

- **rankiniu būdu.** Rankinis anotavimas atliekamas žmogiškojo faktoriaus t. y. anotacijos randamos ir žymimos skaitant straipsnį;
- **automatinis anotavimas.** Kiekvienas straipsnis įkeliamas į sistemą ir automatiškai suanotuojamas.

5.1.4. Sudaromas anotacijų, kurios bus skaičiuojamos atliekant rankiniu ir automatinu būdu, sąrašas:

- hipotezė;
- sankcija;
- dispozicija;
- objektas;
- subjektas;
- pareiga;
- teisė.

5.1.5. Atliekamas rastų anotacijų skaičiavimas įvertinant ir lyginant rankinį anotavimą su automatinu. Randami šie parametrai:

- parametras, kiek anotacijų sistema atpažino pagal aprašytas taisykles;
- parametras, kiek anotacijų sistema turėjo atpažinti pagal aprašytas taisykles;
- parametras, kiek anotacijų sistema turėjo atpažinti, bet neatpažino pagal aprašytas taisykles;
- parametras, kiek anotacijų turėjo iš viso būti atpažintų straipsnyje.

5.1.6. Suskaičiuojami kiekvienos anotacijos absoliutiniai dydžiai, t. y. susumuojamos rastos anotacijos pagal kiekvieną parametą, pateiktą 5.1.5 punkte. Atliekamas kokybės kriterijų įvertinimas, kuris pateikiamas 5.3 dalyje.

5.1.7. Vizualiai atvaizduojami gauti eksperimentinės dalies rezultatai (pateikiamos diagramos).

5.1.8. Pateikiamos atlikto eksperimento išvados ir pritaikyto sprendimo rekomendacijos.

5.2 Eksperimento rezultatai

Eksperimentui atlikti buvo pasirinkti keli skirtingi kodeksai ir skirtingos straipsnių imtys. Žemiau esančioje lentelėje pateikiama bendra kodeksų ir straipsnių statistika.

5.1 lentelė. Eksperimento duomenys

Kodekso pavadinimas	Straipsnių skaičius
Administracinių nusižengimų kodeksas	31
Darbo kodeksas	19
Civilinis kodeksas	6

Eksperimento metu kiekvienam kodeksui ir anotacijoms apskaičiuojami parametrai, kurie nurodyti 5.1. skyriaus 5.1.5. punkte. Skaičiavimai atlikti straipsnio lygyje, tačiau tikslumo kriterijų skaičiavimai atlikti suminėms reikšmėms. Eksperimentas straipsnio lygyje pateikiama 8.3. prieduose.

Administracinių nusižengimų kodekso eksperimentinio rezultatai

5.2 lentelė. Administracinių nusižengimų kodekso eksperimentinio rezultatai

	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažino
Sankcija	157	166	173	16	0
Subjektas	179	183	254	75	66
Hipotezė	138	139	145	7	2
Objektas	145	147	599	454	2
Dispozicija	139	141	152	13	2

Atlikus Administracinių nusižengimų kodekso eksperimentą pastebėta, kad sukurta sistema daugiausiai neatpažino objektų ir subjektų, kadangi šios anotacijos kiekviename kodekso straipsnyje gali skirtis ir kuriant taisykles nepavyko atrasti bendro dėsningumo, pagal kurį būtų galima aprašyti taisykles. Taip pat galime pastebėti, kad geriausiai sistema, pagal aprašyta taisyklę, atpažįsta sankciją.

Eksperimento metu automatiškai anotuojant Administracinių nusižengimų kodekso straipsnius, pastebėta, kad šiame kodekse kai kurių straipsnių struktūra sutampa su Darbo kodekso straipsnių struktūra t. y. rastos anotacijos, kurios buvo aprašytos Darbo kodeksui. Iš žemiau pateikto 5.1 pav. matosi, kad iš 2.4 skyriaus atpažino šias taisykles: 3, 4, 5, 6, 9.

5. Teismas ar administracinio nusižengimo bylą ne teismo tvarka išnaginėjusi institucija (pareigūnas), atsižvelgdami į padaryto administracinio nusižengimo pobūdį, pažeidėjo kaltės formą ir rūšį, asmenybę, atsakomybę lengvinančias ir sunkinančias aplinkybes ir vadovaudamiesi teisingumo ir protingumo principais, gali paskirti mažesnę baudą negu šio kodekso specialiosios dalies straipsnio sankcijoje numatyta minimali bauda arba paskirti švelnesnę administracinę nuobaudą ar administracinio poveikio priemonę, negu numatyta šio kodekso specialiosios dalies straipsnio sankcijoje, arba administracinės nuobaudos ar administracinio poveikio priemonės neskirti. Teismas ar administracinio nusižengimo bylą ne teismo tvarka išnaginėjusi institucija (pareigūnas) kiekvieną savo sprendimą privalo motyvuoti. Administracinio nusižengimo bylą ne teismo tvarka išnaginėjusios institucijos (pareigūno) sprendimą sankcionuoja apylinkės teismo teisėjas.

5.1 pav. Administracinių nusižengimų kodekso 34 straipsnio anotacija

675 straipsnis. Nutarimų skirti **bauda** įvykdymo terminai ir tvarka

1. Nutarimus skirti **bauda** vykdo Valstybinė mokesčių inspekcija, antstoliai šiame kodekse, Civilinio proceso kodekse, Mokesčių administravimo įstatyme nustatyta tvarka.

2. Administracinės atsakomybės traukiamas asmuo **bauda** turi sumokėti ne vėliau kaip per keturiasdešimt dienų nuo nutarimo skirti **bauda** išsiuntimo ar išdavimo jam dienos, o apskundus nutarimą skirti **bauda**, – ne vėliau kaip per keturiasdešimt dienų nuo nutarties, kuria skundas nepatenkintas, išsiuntimo ar išdavimo dienos.

3. Atsižvelgdamas į administracinės atsakomybės traukiamo asmens materialinę padėtį, kitas reikšmingas aplinkybes, asmens prašymu, kai jis pateikia įrodymus (nekilnojamojo ir kilnojamojo turto pažymas, turto deklaracijas, pažymas apie **darbo užmokestį** ir gaunamas socialines išmokas ar pan.), teismas ar administracinio nusižengimo bylą ne teismo tvarka nagrinėjanti institucija (pareigūnas), priimdami nutarimą administracinio nusižengimo byloje, gali paskirtos **baudos** mokėjimą išdėstyti per laikotarpį iki dvejų metų, atsižvelgdamas į paskirtos **baudos** dydį. Asmens prašymas išdėstyti **baudos** mokėjimą, paduodamas administracinio nusižengimo bylą ne teismo tvarka nagrinėjančiai institucijai (pareigūnui) arba teismui ne teismo posėdžio metu, **privalo būti** rašytinis. Šis prašymas teismui ne teismo posėdžio metu paduodamas per instituciją, kurios pareigūnas atliko administracinio nusižengimo tyrimą. Po nutarimo priėmimo paskirtos **baudos** mokėjimo išdėstymo klausimai sprendžiami Mokesčių administravimo įstatyme nustatyta tvarka. Nesumokėtos **baudos** ar jos dalies mokėjimas dalimis negali būti išdėstomas asmeniui, kuris vengė atlikti to paties administracinio nusižengimo teisenoje jam paskirtus viešuosius darbus.

4. Teismas ar administracinio nusižengimo bylą ne teismo tvarka nagrinėjanti institucija (pareigūnas), skirdami **bauda**, atsižvelgdamas į administracinės atsakomybės traukiamo asmens sunkią materialinę padėtį, kitas reikšmingas aplinkybes, šio asmens prašymu, kai jis pateikia įrodymus (nekilnojamojo ir kilnojamojo turto pažymas, turto deklaracijas, pažymas apie **darbo užmokestį** ir gaunamas socialines išmokas ar pan.), ar sutikimu gali tuo pačiu nutarimu visą **bauda** ar jos dalį pakeisti viešaisiais darbais. Asmens prašymas pakeisti **bauda** (jos dalį) viešaisiais darbais, paduodamas administracinio nusižengimo bylą ne teismo tvarka nagrinėjančiai institucijai (pareigūnui) arba teismui ne teismo posėdžio metu, **privalo būti** rašytinis. Šis prašymas teismui ne teismo posėdžio metu paduodamas per instituciją, kurios pareigūnas atliko administracinio nusižengimo tyrimą.

Hipotezė
Dispozicija
Sankcija
Subjektas
Objektas
Teisė
Pareiga

5.2 pav. Administracinių nusižengimų kodekso 675 straipsnio anotacija

8. Už šio straipsnio 4, 5 dalyse numatytus administracinius nusižengimus gali būti skiriamas transporto priemonės konfiskavimas. Už šio straipsnio 1, 2 dalyse numatytus administracinius nusižengimus neturintiems teisės vairuoti transporto priemones asmenims gali būti skiriamas transporto priemonės konfiskavimas.

5.3 pav. Administracinių nusižengimų kodekso 54 straipsnio anotacija

Darbo kodekso eksperimento rezultatai

5.3 lentelė. Darbo kodekso eksperimento rezultatai

	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažino
Subjektas	185	253	320	135	73
Objektas	26	26	285	259	0
Teisė	7	8	28	21	3
Pareiga	13	13	23	10	1
Hipotezė	62	62	104	42	30
Dispozicija	106	106	163	57	17

Atlikus Darbo kodekso eksperimentą pastebėta, kad sukurtas sprendimas daugiausiai neatpažino subjektų ir objektų, kadangi jiems rasti buvo aprašytos tik kelios taisyklės (žiūrėti 2.4 skyrių), įvertinant situaciją, kad kiekvienam straipsniui tiek subjektai, tiek objektai yra skirtingi ir pagal kontekstą anotacijos gali persidengti. Taip pat atlikus analizę pastebėta, kad nėra aiškaus objektų dėsningumo straipsnyje, o žodžių lemos ir gramatinės kategorijos gali skirtis. Iš eksperimento rezultatų galima priimti išvadą, kad sistema Darbo kodekse tiksliausiai atpažino pareigas.

Civilinio kodekso eksperimento rezultatai

5.4 lentelė. Civilinio kodekso eksperimento rezultatai

	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
Hipotezė	16	16	23	7	10
Dispozicija	20	24	31	11	3
Pareiga	4	4	10	6	0
Objektas	0	0	58	58	0
Subjektas	5	9	49	44	2
Teisė	4	4	7	3	0

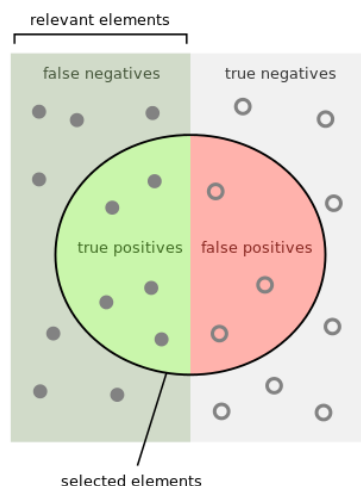
Atliekant analizę bei kuriant anotacijų taisykles Civiliniam kodeksui nebuvo aprašytos specifinės taisyklės. Eksperimento metu pastebėta, kad dauguma anotacijų taisyklių kaip: teisė, pareiga, dispozicija, atitinka tokias pačias taisykles, kurios buvo pritaikytos Darbo kodeksui. Atliekant Civilinio kodekso eksperimentą buvo pasirinkta maža įvairių straipsnių imtis, kadangi išsikeltas tikslas – patikrinti, kiek anotacijų sistema atpažįsta neanalizuotam kodeksui.

Vertinant atlikto eksperimento skaičiavimus ir sukurtą sprendimą automatiniai anotavimai yra svarbu paminėti, kad hipotezėse rasti subjektai nėra pažymimi, kadangi anotacijos persidengia ir dėl šios priežasties ne visas persidengiančias anotacijas pavyko sužymėti. Tačiau subjektų segmentai yra gražinami į JSON'ą, todėl įkėlus išsaugotą straipsnio anotaciją į automatizuoto anotavimo sistemą, subjektai, esantys hipotezėse, yra atpažįstami.

Jautrumas (angl. Recall) ir preciziškumas (angl. Precision)

Atlikus rankinį ir automatinį anotavimą bei apskaičiavus sumines reikšmes parametrų, kurie nurodyti 5.1. skyriuje, 5.1.5 punkte, su gautais rezultatais randami sistemos preciziškumo ir jautrumo rodikliai.

Preciziškumas (angl. *precision*) ir jautrumas (angl. *recall*)



5.4 pav. Aibės jautrumo ir preciziškumo skaičiavimui [21]

Precižiškumas ieškomas siekiant įvertinti santykį tarp teisingai atpažintų ir kiek automatiškai atpažino sistema. Šiuo kriterijumi vertinamas sistemos tikslumas aprašytoms taisyklėms.

Jautrumas ieškomas siekiant įvertinti santykį tarp teisingai atpažintų ir kiek iš viso anotacijų turėjo būti atpažinta. Šiuo kriterijumi vertinama, kokią dalį anotacijų sistema rado automatiškai lyginant su visomis rastomis anotacijomis, atlikus rankinį anotavimą.

$$\text{Precižiškumas (Precision)} = \frac{\text{Teisingai atpažinti}}{\text{Atpažinta iš viso}} \quad (5.1)$$

$$\text{Jautrumas (Recall)} = \frac{\text{Teisingai atpažinti}}{\text{Turėjo atpažinti iš viso}} \quad (5.2)$$

$$\text{Teisingai atpažinti} = \text{Atpažino pagal sukurtas taisykles} - \text{Klaidingai atpažinti} \quad (5.3)$$

Precižiškumo ir jautrumo reikšmės turi būti mažesnės arba lygios 1. Kuo reikšmė artimesnė 1, tuo geriau vertinamas sukurto sprendimo preciziškumas ir jautrumas.

Kokybės kriterijų skaičiavimo rezultatai

Administracinių nusižengimų kodekso kriterijų skaičiavimas

Pateikiami jautrumo ir preciziškumo koeficientų skaičiavimai anotacijoms, kurios buvo rastos Administracinių nusižengimų kodekse. Rezultatai apibendrinami ir įvertinami 5.3 skyriuje.

5.5 lentelė. Administracinių nusižengimų kodekso kriterijų skaičiavimas

Teisės normos dalis	Jautrumas	Precižiškumas
Sankcija	0.907514451	0.945783
Subjektas	0.44488189	0.617486
Hipotezė	0.937931034	0.978417
Objektas	0.238731219	0.972789
Dispozicija	0.901315789	0.971631

Darbo kodešo kokybės kriterijų skaičiavimas

Pateikiami jautrumo ir preciziškumo koeficientų skaičiavimai anotacijoms, kurios buvo rastos Darbo kodekse. Rezultatai apibendrinami ir įvertinami 5.3 skyriuje.

5.6 lentelė. Darbo kodešo kokybės kriterijų skaičiavimas

Teisės normos dalis	Jautrumas	Preciziškumas
Subjektas	0.35	0.442688
Objektas	0.09122807	1
Teisė	0.142857143	0.5
Pareiga	0.52173913	0.923077
Hipotezė	0.307692308	0.516129
Dispozicija	0.54601227	0.839623

Civilinio kodekso kokybės kriterijų skaičiavimas

Pateikiami jautrumo ir preciziškumo koeficientų skaičiavimai anotacijoms, kurios buvo rastos Civiliniame kodekse. Rezultatai apibendrinami ir įvertinami 5.3 skyriuje.

5.7 lentelė. Civilinio kodekso kokybės kriterijų skaičiavimas

Teisės normos dalis	Jautrumas	Preciziškumas
Hipotezė	0.26087	0.375
Dispozicija	0.548387	0.708333
Pareiga	0.4	1
Objektas	0	0
Subjektas	0.061224	0.333333
Teisė	0.571429	1

Administracinių nusižengimų kodekso ir Darbo kodekso kokybės kriterijų skaičiavimas

Pateikiami jautrumo ir preciziškumo koeficientų skaičiavimai bendroms Administracinių nusižengimų kodekso ir Darbo kodekso teisės normos dalims. Darbo kodekse nebuvo rasta sankcijų, bet Administracinių nusižengimų kodekse keliuose straipsniuose buvo rastos pareigos ir teisės. Kadangi tai buvo pavieniai atvejai, todėl į skaičiavimus šios teisės normos dalys nebuvo įtrauktos. Rezultatai apibendrinami ir įvertinami 5.3 skyriuje.

5.8 lentelė. Administracinių nusižengimų kodekso ir Darbo kodekso kokybės kriterijų skaičiavimas

Teisės normos dalis	Jautrumas	Preciziškumas
Subjektas	0.3919861	0.516055
Hipotezė	0.6746988	0.835821
Objektas	0.1911765	0.976879
Dispozicija	0.7174603	0.91498

Darbo kodekso ir Civilinio kodekso kokybės kriterijų skaičiavimas

Pateikiami jautrumo ir preciziškumo koeficientų skaičiavimai bendroms Darbo kodekso ir Civilinio kodekso teisės normos dalims. Rezultatai apibendrinami ir įvertinami 5.3 skyriuje.

5.9 lentelė. Darbo kodekso ir Civilinio kodekso kokybės kriterijų skaičiavimas

Teisės normos dalis	Jautrumas	Preciziškumas
Subjektas	0.31165312	0.438931298
Objektas	0.07580175	1
Teisė	0.22857143	0.666666667
Pareiga	0.48484848	0.941176471
Hipotezė	0.2992126	0.487179487
Dispozicija	0.54639175	0.815384615

Administracinių nusižengimų, Darbo kodekso ir Civilinio kodekso kokybės kriterijų skaičiavimas

Pateikiami jautrumo ir preciziškumo koeficientų skaičiavimai bendroms Administracinių nusižengimų, Civilinio ir Darbo kodeksų teisės normos dalims. Rezultatai apibendrinami ir į įvertinami 5.3 skyriuje.

5.10 lentelė. Administracinių nusižengimų, Darbo kodekso ir Civilinio kodekso kokybės kriterijų skaičiavimas

Teisės normos dalis	Jautrumas	Preciziškumas
Subjektas	0.365971	0.512359551
Hipotezė	0.639706	0.801843318
Objektas	0.179406	0.976878613
Dispozicija	0.702312	0.896678967

5.3 Sprendimo veikimo ir savybių analizė, kokybės kriterijų įvertinimas

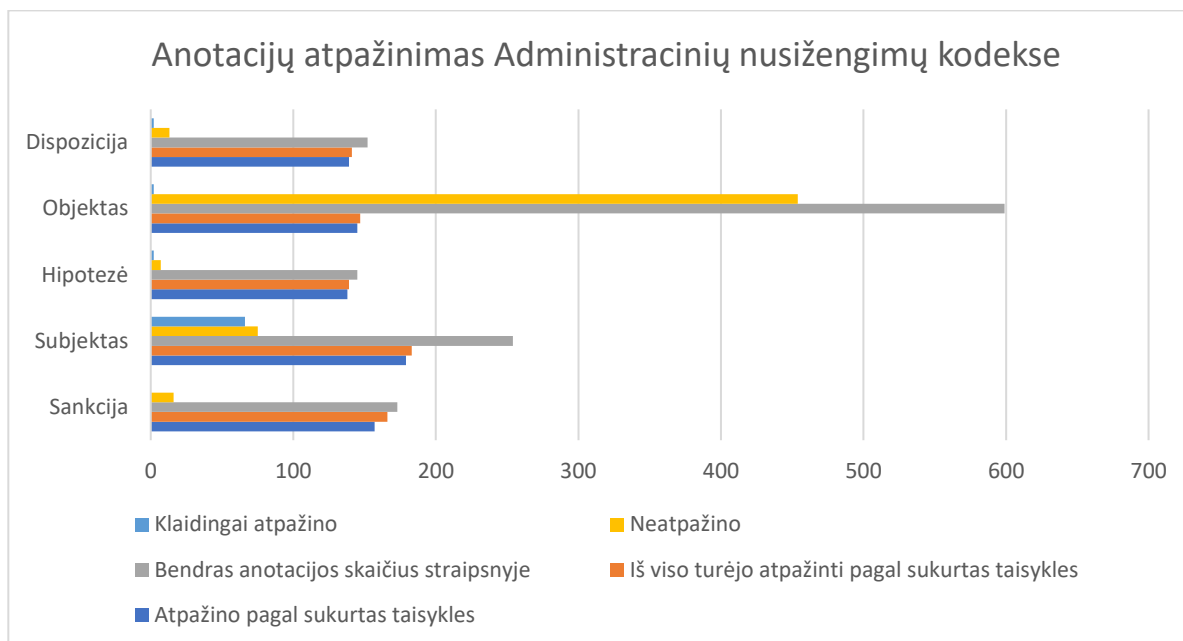
Sprendimo veikimo įvertinimas pasirinktiems kodeksams

Sprendimo veikimas kiekvienam kodeksui yra pateikiamas atskiromis stulpelinėmis diagramomis, siekiant parodyti kiek kiekvienam kodeksui anotacijų buvo atpažinta ir turėjo būti atpažinta pagal 5.1 skyriuje, 5.1.5 punkte įvardintus parametrus.

Rankinio anotavimo, kuris buvo atliktas paprastuoju būdu t. y. žmogiškajam faktoriui anotuojant straipsnis, rezultatus apibūdina parametras „Bendras anotacijų skaičius straipsnyje“.

Automatinio anotavimo, kurį atliko sukurtas sprendimas, rezultatus apibūdina parametras „Atpažino pagal sukurtas taisykles“.

Administracinių nusižengimų kodeksas



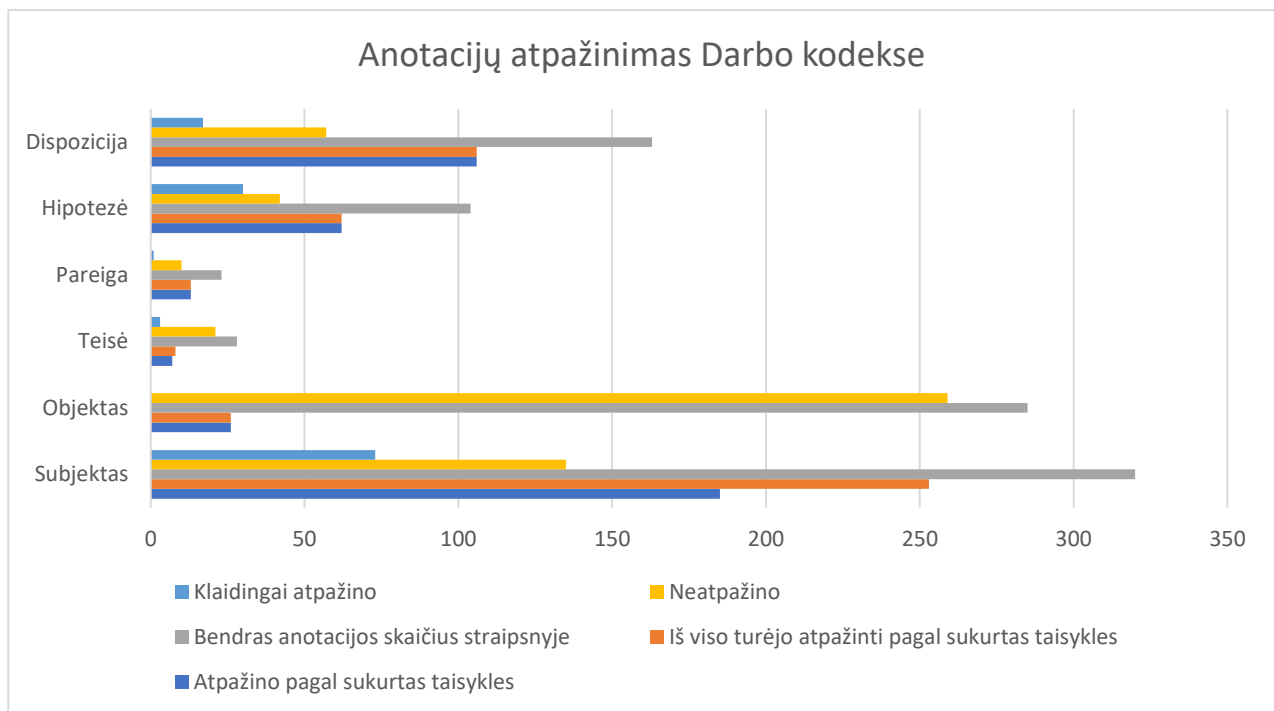
5.5 pav. Anotacijų atpažinimas Administracinių nusižengimų kodekse

Diagramoje pateikiami bendri Administracinių nusižengimų kodekso tiek rankinio, tiek automatinio anotavimo rezultatai. Iš gautų rezultatų apie sukurtą sprendimą galime daryti šias prielaidas:

- **dispozicija** – matome, kad lyginant šios anotacijos skaičiavimus: kiek sistema atpažino su kiek turėjo atpažinti ir koks bendras anotacijų skaičius straipsnyje, rezultatai skiriasi neženkliai. Taip pat, labai mažą dalį sudaro klaidingai atpažintos ir netapažintos anotacijos. Vadinasi sistema dispoziciją atpažįsta pakankamai tiksliai. Šių rezultatų prielaida – sistemoje buvo aprašytos bendrinės dispozicijų taisyklės, kurių atsikartojamumas straipsniuose pakankamai didelis;
- **objektas** – lyginant parametrus: kiek sistema atpažino su kiek turėjo atpažinti ir koks bendras anotacijų skaičius straipsnyje, pastebime, kad mažiausias skirtumas yra tarp reikšmių kiek atpažino ir kiek turėjo atpažinti pagal aprašytas taisykles, tačiau ženklus skirtumas lyginant su bendru objektų skaičiumi straipsniuose ir klaidingai atpažintais. Šių rezultatų prielaida – sistemoje buvo sukurtos taisyklės tik konkreitiems objektams rasti;
- **hipotezė** – šios teisės normos sąvokos parametrų: kiek sistema atpažino, kiek turėjo atpažinti ir bendras anotacijų skaičius straipsnyje, reikšmės yra panašios. Neatpažintų ir klaidingai atpažintų anotacijų skaičius yra nedidelis. Galime daryti prielaidą, kad sistema pakankamai tiksliai straipsniuose atpažįsta hipotezę;
- **subjektas** – rezultatų išvadas galime prilyginti objektų išvadoms (2 punktą), tačiau išsiskiria klaidingai atpažintų anotacijų skaičius. Taip yra todėl, kad subjektų nenuspalvina hipotezėse, tačiau juos suranda ir įrašo į .json;

- sankcija – šios anotacijos rezultatų išvadas galime prilyginti prie dispozicijos (1 punktą) parašytoms išvadoms.

Darbo kodeksas



5.6 pav. Anotacijų atpažinimas Darbo kodekse

Diagramoje pateikiami bendri Darbo kodekso tiek rankinio, tiek automatinio anotavimo rezultatai. Iš gautų rezultatų apie sukurta sprendimą galime daryti šias prielaidas:

- dispozicija – lyginant kiek sistema atpažino ir kiek turėjo atpažinti pagal sukurta taisykles pastebime, kad šių parametrų skaičiavimai sutampa, tačiau ženkliai skiriasi bendras dispozicijų skaičius straipsnyje. Iš šių rezultatų galime daryti prielaidą, kad sistema tiksliai atpažino dispozicijas, kurioms aprašytos taisyklės, tačiau pagal kodekso straipsnių kontekstą yra daugiau dispozicijų, kurioms reikėtų aprašyti taisykles;

- hipotezė – šiai anotacijai rasti buvo taikoma tokia pati taisyklė, kuri aprašyta Administracinių nusižengimų kodeksui. Parametrų skaičiavimo rezultatui prielaidas galime daryti panašias kaip dispozicijos (1 punktą), tačiau reikėtų išskirti klaidingai atpažintų anotacijų skaičių, kadangi straipsniuose ne visada punktas prasideda hipoteze. Pastebėta, jog kai kuriuose straipsniuose punktas pasibaigia hipoteze, tokios anotacijos nebuvo rastos;

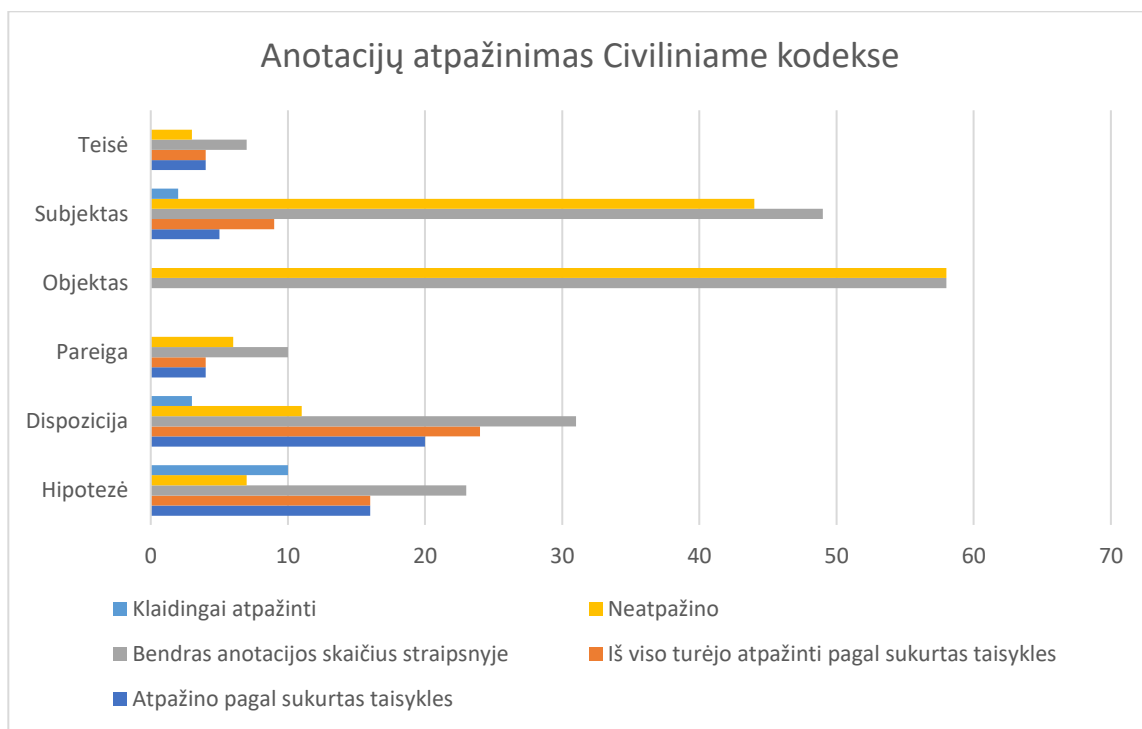
- pareiga – tai viena iš mažiausiai rastų anotacijų straipsniuose. Galime pastebėti, kad bendras šios anotacijos skaičius straipsniuose yra didesnis nei sistema atpažino, tačiau lyginant rodiklius kiek sistema turėjo atpažinti ir kiek atpažino, matome, kad rezultatai yra apylygiai. Vadinasi sistema randa pareigas, kurioms aprašytos taisyklės, bet gilinant į straipsnių kontekstą pareigoms rasti reikėtų sukurti daugiau bendrinių taisyklių;

- teisė – šios anotacijos automatinio radimo rezultatų išvadas būtų galima prilyginti prie pareigos aprašytoms įžvalgoms (3 punktas). Tačiau iš pateiktos diagramos pastebime didesnį skaičių prie parametro „Neatpažino“, nes atliekant eksperimentą pastebėti, kad teisės radimui yra kitų dėsningai atsikartojančių bendrinių taisyklių, kurios nebuvo įgyvendintos. Dėl šios priežasties anotacijos atpažinimo rezultatai yra prastesni;

- objektas – šios anotacijos eksperimento rezultatų išvadas, kiek objektų buvo automatiškai rasta Darbo kodekse, galime prilyginti objektų radimo Administracinių nusižengimų kodekso eksperimento rezultatams;

- subjektas – tai viena iš daugiausiai randamų anotacijų straipsniuose tiek automatinio, tiek rankiniu būdu. Tai galime pastebėti lyginant parametrus: kiek buvo atpažinta, kiek turėjo atpažinti pagal sukurtas taisykles ir koks bendras anotacijų skaičius. Taip pat patebime, kad subjektas kaip ir objektas yra anotacijos, kurias sistema atpažįsta netikslingiausiai, to priežastis – sudėtinga automatizuoti šių anotacijų tikslų radimą, nes subjektai turi daug įvairių skirtingų taisyklių, ko pasekoje kiekvienas žodis atitinkamai gali turėti skirtingas lemas ar gramatines kategorijas.

Civilinis kodeksas



5.7 pav. Anotacijų atpažinimas Civiliniame kodekse

Civilinio kodekso eksperimento rezultatus galime prilyginti Darbo kodekso eksperimento rezultatams. Kadangi atliekant automatinį kodekso straipsnių anotavimą pastebėta, kad šiame kodekse rastos anotacijos atitinka Darbo kodeksui aprašytų anotacijų taisykles, išskyrus objektus – automatinio anotavimo metu nebuvo rastas nė vienas objektas.

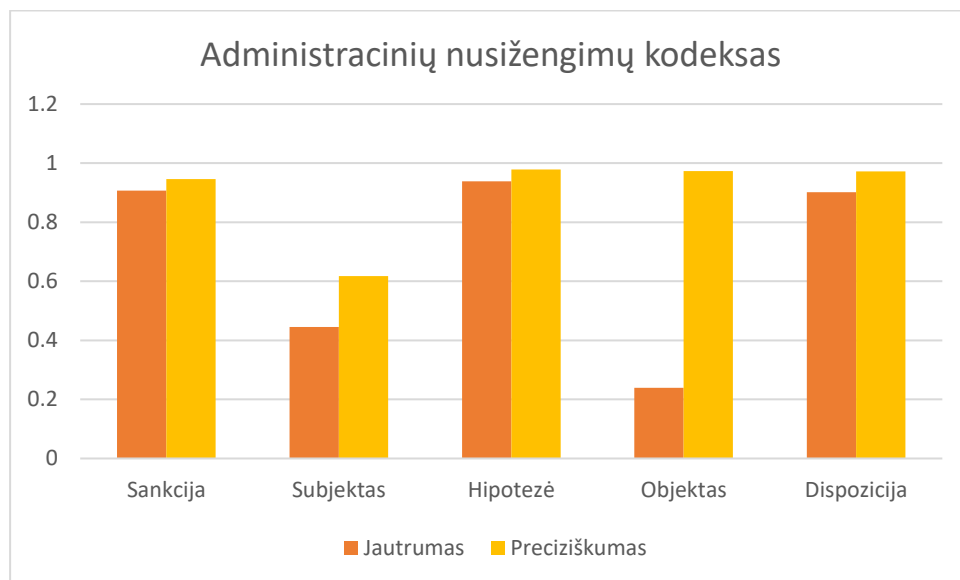
Civilinis kodeksas buvo pasirinktas siekiant išbandyti, kiek sukurtas sprendimas automatiškai suanotuoja kodeksų straipsnius, kurių analizė nebuvo atlikta, todėl iš gautų eksperimento rezultatų

galime daryti išvadą – sukurtame sprendime aprašytos anotacijų taisyklės yra bendrinės, tinkamos keliems kodeksas.

Kokybės kriterijų: preciziškumo ir jautrumo, įvertinimas

Kokybės kriterijų įvertinimas pateikiamas stulpelinėmis diagramomis pagal 5.2 skyriuje pateiktas lenteles.

Administracinių nusižengimų kokybės kriterijų įvertinimas



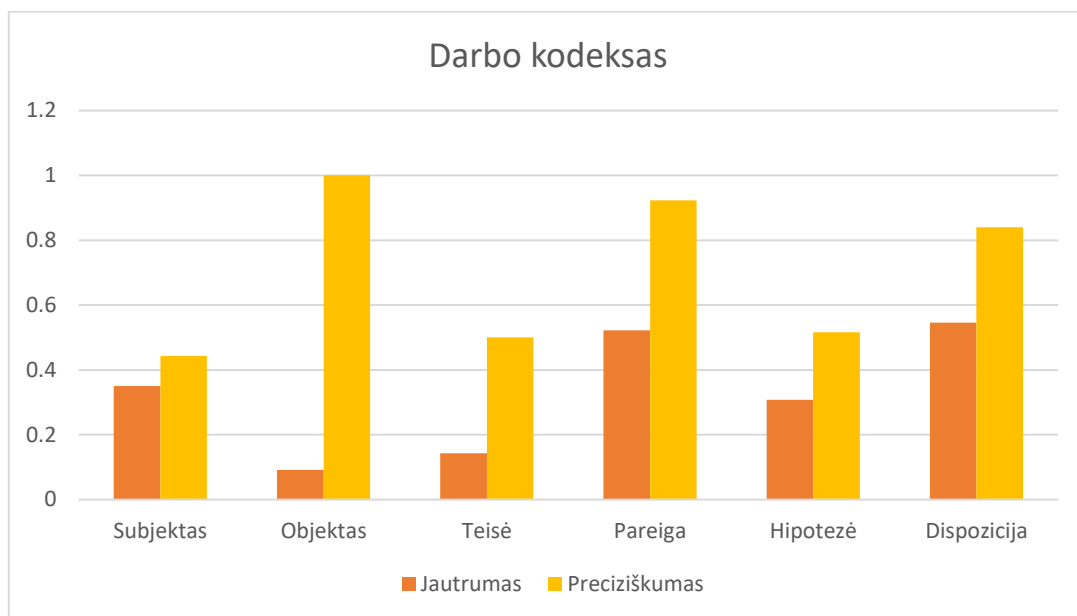
5.8 pav. Kokybės kriterijų palyginimas Administracinių nusižengimų kodekse

Iš pateiktų rezultatų matome, kad preciziškumo reikšmė kiekvienai anotacijai yra didesnė nei jautrumo. Vadinasi sistema automatiškai tiksliau atpažįsta teisės normos sąvokas, kurioms aprašytos taisyklės.

Tiksliausiai atpažįstamos anotacijos: sankcija (jautrumas 0.907514451, preciziškumas 0.945783), hipotezė (jautrumas 0.937931034, preciziškumas 0.978417), dispozicija (jautrumas 0.901315789, preciziškumas 0.971631). Šioms anotacijoms buvo aiškiausiai aprašytos taisyklės, dauguma straipsnių turi tikslią ir loginę terminų seką, pagal kurią galima aprašyti bendrines teisės normos sąvokoms skirtas anotacijų taisykles.

Netiksliausiai atpažįstamos anotacijos: subjektas (jautrumas 0.44488189, preciziškumas 0.617486), objektas (jautrumas 0.238731219, preciziškumas 0.972789). Preciziškumo koeficientas yra ženkliai didesnis už jautrumo, kadangi buvo aprašytos tikslių objektų ir subjektų taisyklės, kuriuos sistema atpažino, tačiau šioms anotacijoms nėjo aprašyti bendrinių taisyklių, nes tiek objektai, tiek subjektai straipsniuose yra „išsibarstę“ hipotezėse bei dispozicijose be logiškos sekos.

Darbo kodekso kokybės kriterijų įvertinimas



5.9 pav. Kokybės kriterijų palyginimas Darbo kodekse

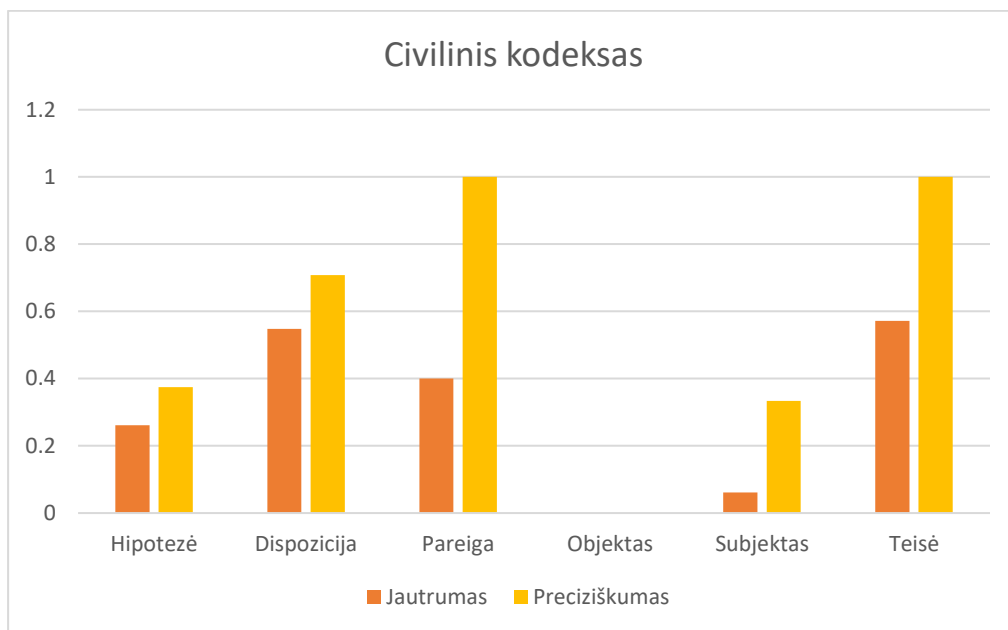
Atlikus Darbo kodekso eksperimentą ir apskaičiavus kokybės kriterijus, galime pastebėti, kad preciziškumo koeficiento reikšmės yra didesnės už jautrumo koeficiento reikšmes.

Didžiausias pokytis tarp preciziškumo ir jautrumo koeficientų matomas prie objekto. Tokia ryškaus kriterijų eikšmių skirtumo priežastis – aprašytos anotacijų taisyklės tik konkretiems objektams (žiūrėti 2.4 skyriaus 15, 16, 17, 18, 19 taisyklės). Sistema automatiškai atpažino šiuos objektus, todėl preciziškumo koeficientas yra 1, o jautrumo koeficientas 0,09.

Geriausiai atpažintos anotacijos: dispozicija, pareiga. Dispozicijai rasti buvo sukurta daugiausiai taisyklių (žiūrėti 2.4 skyriaus 3, 4, 5, 6 taisyklės), pareigos radimui buvo sukurta viena taisyklė (žiūrėti 2.4 skyrių).

Hipotezei ieškoti Darbo kodekse nebuvo aprašytos naujos taisyklės, taikoma tokia pati taisyklė kaip Administracinių nusižengimų kodeksui (žiūrėti 2.4 skyrių). Atliekant testavimą ir eksperimentą pastebėta, kad straipsniai gali turėti atvirštinę logiką t. y. pradžia sakinio prasideda subjektu ir pasibaigia hipoteze. Taisyklė buvo aprašyta, kai pradžia sakinio prisideda hipoteze, dėl šios priežasties dalis subjektų atpažįstama kaip hipotezė.

Civilinio kodekso kokybės kriterijų įvertinimas



5.10 pav. Kokybės kriterijų palyginimas Civiliniame kodekse

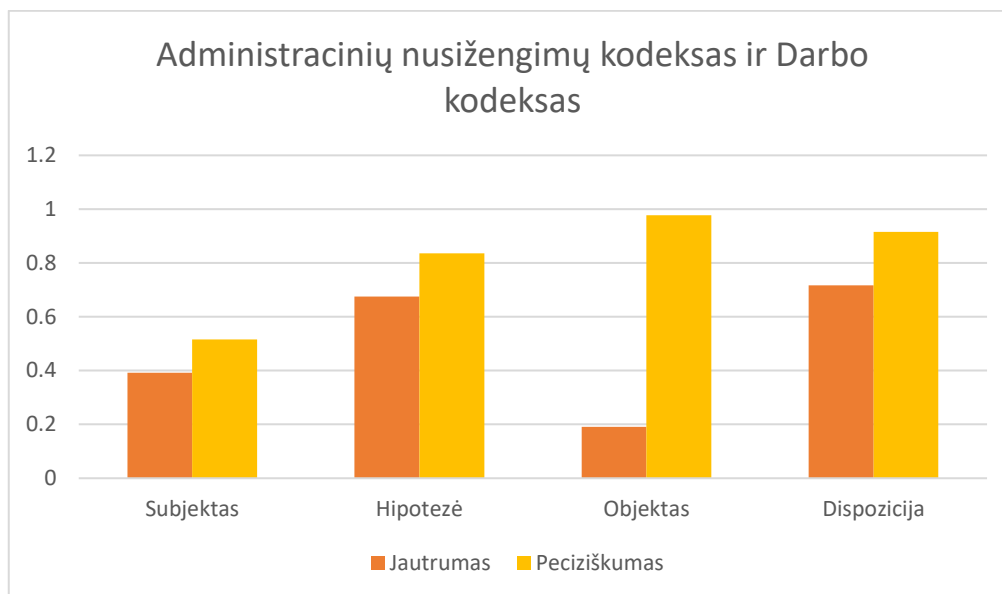
Apskaičiavus Civilinio kodekso preciziškumo ir jautrumo koeficientus iš pateiktų rezultatų galima pastebėti, kad Civiliniame kodekse objekto preciziškumo ir jautrumo koeficientai lygūs 0, kadangi automatiškai anotuojant nebuvo atpažintas nė vienas objektas, nors rankiniu būdu anotuojant rasti 58 objektai. Aprašyti objektai Administracinių nusižengimų ir Darbo kodeksams pasirinktuose šio kodekso straipsniuose nebuvo rasti.

Mažiausias santykis tarp preciziškumo ir jautrumo koeficiento apskaičiuotas dispozicijai (preciziškumas 0.548387, jautrumas 0.708333). Atlikus Civilinio kodekso eksperimentą pastebėta, kad šiame kodekse atpažįstama daugiausiai dispozicijų sutampančių su Darbo kodekso dispozicijomis.

Geriausi preciziškumo rezultatai apskaičiuoti pareigai (1) ir teisei (1), kadangi pasirinktuose straipsniuose buvo atpažintos visos teisės ir pareigos, kiek turėjo atpažinti pagal aprašytas anotacijų taisykles. Jautrumo koeficientai šių anotacijų yra mažesni (pareiga 0.4, teisė 0.571429), kadangi bendrai straipsniuose buvo rasta daugiau pareigų ir teisių, kurioms turėtų būti aprašytos kitos anotacijų atpažinimo taisyklės.

Civilinio kodekso eksperimento rezultatus vertinti galima tik kaip orientacinius skaičius, kadangi šio kodekso eksperimentui buvo pasirinkta maža straipsnių imtis.

Administracinių nusižengimų kodekso ir Darbo kodekso kokybės kriterijų įvertinimas



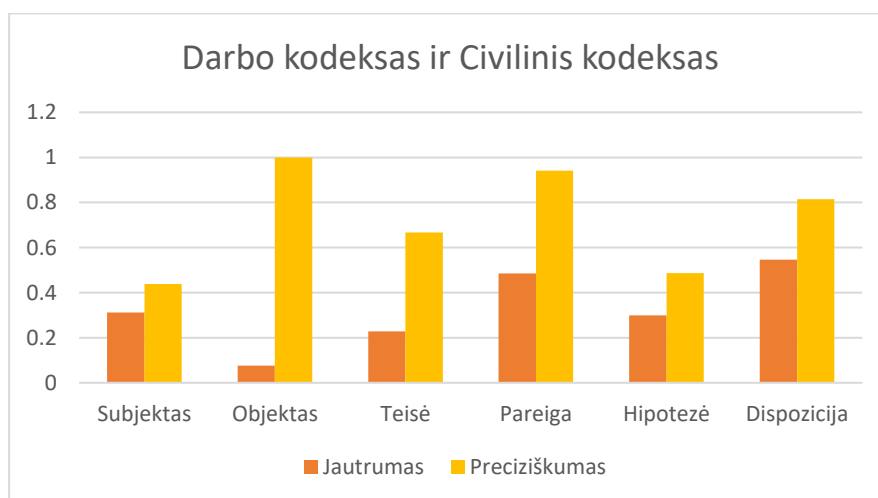
5.11 pav. Bendras Administracinių nusižengimų ir Darbo kodeksų kokybės kriterijų palyginimas

Susumavus Administracinių nusižengimų kodekso ir Darbo kodekso eksperimento rezultatus, buvo apskaičiuoti bendri jautrumo ir preciziškumo koeficientai. Eksperimento rezultatai sumuojami tik bendroms anotacijoms: subjektui, hipotezei, objektui ir dispozicijai.

Iš pateiktos diagramos galime pastebėti, kad lyginant sistemos preciziškumą ir jautrumą abiem kodeksams, geriausias santykis yra dispozicijos (jautrumas 0.7174603, preciziškumas 0.91498). Tokių kokybės kriterijų koeficientų reikšmių priežastis – dispozicijoms rasti buvo aprašyta daugiausiai anotacijų atpažinimo taisyklių.

Ženkiausią skirtumą tarp preciziškumo ir jautrumo galime pastebėti objektui (jautrumas 0.1911765 preciziškumas 0.976879). Šių rezultatų priežastis – tokia pati kaip buvo paminėta prie Administracinių nusižengimų kodekso ir Darbo kodekso objektų radimo, kokybės kriterijų vertinimo.

Darbo kodekso ir Civilinio kodekso kokybės kriterijų įvertinimas



5.12 pav. Bendras Civilinio ir Darbo kodeksų kokybės kriterijų palyginimas

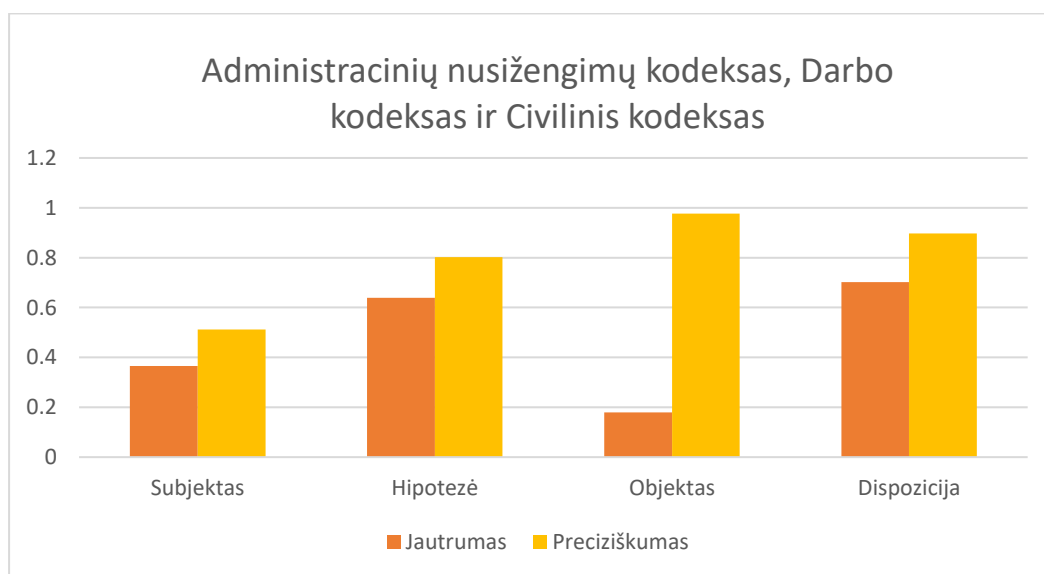
Susumavus Dabo kodekso ir Civilinio kodekso eksperimento rezultatus, buvo apskaičiuoti jautrumo ir preciziškumo koeficientai šioms anotacijoms: subjektas, objektas, teisė, pareiga, hipotezė, dispozicija.

Šių dviejų kodeksų eksperimento rezultatai yra sumuojami, kadangi abiems kodeksams buvo ieškomos tokios pačios teisės normos sąvokos.

Iš pateiktų rezultatų pastebime tą pačią tendenciją – didžiausias skirtumas tarp jautrumo ir preciziškumo koeficientų yra pastebimas objektui (jautrumas 0.07580175, preciziškumas 1). Priežastis panaši į paminėta ankstesniuose Administracinių nusižengimų kodekso ir Darbo kodeksų objektų radimo kokybės kriterijų vertinimą.

Geriausi rezultatai pastebimi ieškant dispozicijos anotacijų (jautrumas 0.54639175, preciziškumas 0.815384615), kadangi šiai anotacijai buvo sukurta daugiausiai bendrinių taisyklių, kurios tinkamos keliems kodeksams.

Administracinių nusižengimų kodekso, Darbo kodekso ir Civilinio kodekso kokybės kriterijų įvertinimas



5.13 pav. Bendras Administracinių nusižengimų, Civilinio ir Darbo kodeksų kokybės kriterijų palyginimas

Apibendrinant atlikto eksperimento rezultatus, buvo atliekami visų eksperimento metu automatinio ir rankinio būdu suanotuotų kodeksų jautrumo ir preciziškumo koeficientų skaičiavimai.

Kadangi lyginant Administracinių nusižengimų kodeksą, Darbo kodeksą ir Civilinį kodeksą, žinome, kad Administracinių nusižengimų kodeksas išsiskiria sankcijomis, Darbo kodeksas išsiskiria teisėmis ir pareigomis, todėl kokybės kriterijų skaičiavimai bendrai kodeksams atlikti tik šių anotacijų: subjektas, hipotezė, objektas, dispozicija.

Lyginant bendrus rezultatus su aukščiau pateiktais rezultatais galime pastebėti, kad situacija nepakito ties objektu – skirtumas tarp jautrumo ir preciziškumo reikšmių išliko didžiausias (jautrumas 0.179406, preciziškumas 0.97687861).

Geriausios kokybės kriterijų reikšmės rastos hipotezei (jautrumas 0.639706, preciziškumas 0.801843318) ir dispozicijai (jautrumas 0.702312, preciziškumas 0.896678967). Hipotezė ir dispozicija buvo vienos iš tiksliausiai rastos anotacijos Administracinių nusižengimų kodekse. Šis veiksnys nulėmė, kad bendrai kodeksams vertinant kriterijus, hipotezės ir dispozicijos kriterijų reikšmės padidėjo, lyginant su kiekvienam kodeksui rastų hipotezų ir dispozicijų kriterijų reikšmėmis.

Dispozicijos tikslumą nulėmė anksčiau paminėtas veiksnys – šiai anotacijai rasti buvo aprašyta daugiausiai bendrinių taisyklių.

Jautrumo ir preciziškumo kriterijų bendras vertinimas

Bendrai vertinant jautrumo ir preciziškumo koeficientus, atsižvelgus į kiekvieno kodekso ir kodeksų suminių eksperimento reikšmių apskaičiuotus kokybės kriterijus, galime daryti išvadą, kad preciziškumo reikšmė visoms anotacijoms yra didesnė už jautrumo reikšmę, ko pasekoje priimame teiginį, kad sistema ganėtinai tiksliai atpažįsta anotacijas, kurioms sukurtos taisyklės.

Anotacijoms, kurios turi žemiausią jautrumą ir aukščiausią preciziškumą, buvo aprašytos taisyklės konkrečioms teisės normos sąvokoms rasti, tačiau šių anotacijų (subjektų ir objektų) bendras skaičius straipsniuose yra didžiausias.

Anotacijos, kurių jautrumo ir preciziškumo koeficientų reikšmių skirtumas yra mažiausias, automatiškai atpažįstamos tiksliausiai ir automatiškai teisingai atpažintų anotacijų skaičius mažiausiai nukrypsta nuo bendro anotacijų kiekio straipsniuose.

5.4 Sprendimo taikymo rekomendacijos

1. Tolimesniam sistemos vystymui rekomenduojama įgyvendinti anotacijų žodyno failo importavimą, pavyzdžiui kodeksai turi aiškų subjektų žodyną atitinkantį kodekso kontekstą. Tokiems subjektas rasti galima sukurti taisyklę, kuriai būtų suimportuojamos žodžių lemos ar gramatinės kategorijos ir pagal tai automatiškai pažymimos anotacijos.
2. Reikėtų patobulinti hipotezės radimo taisyklę, kuri pagerintų automatinio anotavimo rezultatus. Kaip ir buvo minėta prieš tai, kai kuriuose straipsniuose, ypač Darbo kodekse, pastebėta, kad straipsnio punktas arba sakiny pasibaigia hipoteze ir prasideda subjektu. Šiam atvejui reikėtų sukurti atvirkštinę taisyklę. Taisyklės būtų galima papildyti daugiau atsižvelgiant į straipsnių struktūrą ieškant dėsningumo, bet toks pritaikymas turėtų būti skirtas kiekvienam kodeksui individualiai.

3. Rekomenduojama papildyti anotacijos taisyklę „teisės“ radimui. Įgyvendintam sprendime teisė traktuojama kaip trijų žodžių junginys: turi+teisę+veiksmažodžio bendratis. Kad teisė būtų atpažinta visi šie žodžiai turi būtų vienas šalia kito. Taisyklę galima patobulinti ieškant žodžių junginio „turi teisę“ ir papildomai pirmo žodio, kuris yra veiksmažodžio bendratis. Tai reiškia, kad tarp junginio „turi teisę“ ir veiksmažodžio bendraties gali būti kiti žodžiai ar anotacijos. Ši tendencija ypač pastebėta Darbo kodekse, atliekant sukurto sprendimo eksperimentą.
4. Žvelgiant iš vartotojo pusės būtų patogiau, jeigu vartotojas turėtų galimybę atsisiųsti suanotuotą dokumentą ne tik .json formatu.
5. LR teisinių dokumentų automatinio semantinio anotavimo sistemą būtų galima sujungti su kita anotavimo sistema, kuri aprašyta Manto Bataičio tiriamajame darbe „Teisinių dokumentų automatizuoto semantinio anotavimo galimybių plėtimas“.

6. REZULTATŲ APIBENDRINIMAS IR IŠVADOS

1. Atlikus teisinių dokumentų analizę nustatyta, kad kodeksai yra sudaryti iš dalių, dalys iš skyrių ir skyriai iš straipsnių. Bendrai kodeksai turi vieningą teisės normos dalių terminologiją, tačiau kiekvieno kodekso straipsniuose gali būti kitokia teisės normos dalių seka. Taip pat kiekviename kodekse teisinės normos atitinka skirtingą žodyną pagal kodekso tematiką. Dėl šių įvardintų priežasčių teisės normos sąvokoms atpažinti nepakanka tik bendrinių taisyklių, reikalingos ir konkrečiam teisiniui dokumentui (kodeksui) specializuotos taisyklės.
2. Atlikus analizę nustatytos esminės, teisinę normą apibrėžiančios sąvokos. Išbandytas Protégé įrankis, kurio pagalba aiškiau suvoktos teisės normos dalys ir sąryšiai, pagal tai sudarytas esybių-ryšių modelis dalykinei sričiai. Ontologijai sudaryti nenaudojama OWL ontologijų kalba, kadangi tiriamojo darbo tikslui pasiekti tai nebuvo reikalinga, tačiau yra plačiai naudojama esamuose sprendimuose ir gali būti kaip alternatyva ateityje tobulinant bei plečiant automatinio semantinio anotavimo taikomąją programą.
3. Išnagrinėtas LKIF – ontologijų branduolys, skirtas teisiniams dokumentams. Deja pateikta struktūra nėra tinkama LR kodeksams, nes LKIF apibrėžia konkrečių teisinių bylų nagrinėjimą ir ontologija skirta teisės normos pažeidimų išvedimui. Žvelgiant į ateitį LKIF galėtų būti naudojamas sudarant teisinius dokumentus.
4. Atlikta teksto semantinio anotavimo įrankių: Gate, ontotext.com, GoNTogle, Brat rapid, semantika.lt, lyginamoji analizė, kuri parodė, kad artimiausi kuriamam sprendimui įrankiai – GoNTogle ir Brat rapid, dėl anotacijų atvaizdavimo tekste. Po atlikto tyrimo priimta išvada, kad bus kuriamas naujas sprendimas, kurio įgyvendinimui planuojama sąsaja su semantika.lt, nes kiti egzistuojantys sprendimai nėra tinkami lietuviškam tekstui. Sprendimo įgyvendinimui nenaudojamos semantinio tinklo technologijos (RDF, RDFS, OWL, SPARQL). Semantiniu anotavimu siekiamas atpažinti kuo didesnis esybių kiekis, kuriose skirtingi žodžiai bei jų junginiai suklasifikuojami į teisės normos dalis.
5. Sistemoje sukurtos anotavimo taisyklės pagrindinėms teisės normos sąvokoms rasti. Taisyklės sukurtos naudojant pirmos eilės predikatų logiką. Aprašyta 19 anotacijų taisyklių, apimančių šias sąvokas: dispozicija, hipotezė, sankcija, teisė, pareiga, subjektas ir objektas. Sistemoje suanotuotas tekstas sužymimas skirtingomis spalvomis atitinkamai pagal rastas anotacijas ir išvedamas pagrindiniame taikomosios programos lange, siekiant naudotojui kuo aiškiau ir suprantamiau pateikti suanotuotą tekstą.
6. Atlikus eksperimentą ir apskaičiavus sistemos jautrumo ir preciziškumo koeficientus, galima teigti, kad sukurtas sprendimas ganėtinai tiksliai ir teisingai atpažįsta teisės normų anotacijas, kurioms aprašytos taisyklės. Preciziškumo koeficientas artėja prie 1.

7. Lyginant subjektų ir objektų bendrą anotacijų kiekį straipsniuose su automatiškai atpažintu anotacijų kiekiu, pastebėta, kad jautrumo koeficientas yra labai mažas, tai pasako, kad šioms anotacijoms rasti reikalingos specifinės taisyklės, kurios nebuvo aprašytos.
8. LR teisinių dokumentų automatinio semantinio anotavimo sistema gali padėti naudotojui, kuris susijęs su teisiniais dokumentais, sumažinti laiko sąnaudas atliekant kodeksų anotavimą rankiniu būdu. Automatinis semantinis anotavimas nestruktūrizuotą tekstą paverčia struktūrizuotu, pagal teisės normos dalis, spalvomis išskaidytu tekstu.
9. Atlikus teisinių dokumentų analizę, realizavus sprendimą ir jį eksperimentiškai įvertinus, galime priimti išvadą, kad dėl kodeksų bendro nevieningumo ir išsiskiriančios straipsnių struktūros bei gausios teisės normos sąvokų terminologijos, siekiant pilnai automatizuoti teksto anotavimą – reikalingas taisyklių papildymas, pagal kiekvieną kodeksą.
10. Plėtojant sprendimą ateityje rekomenduojama padidinti teisės normos dalių atpažinimo taisyklių kiekį, siekiant, kad sistema tiksliau automatiškai anotuotų daugiau kodeksų, ypač rekomenduojama tobulinti objektų ir subjektų taisykles. Žvelgiant iš vartotojo pusės – darbą su sistema palengvintų, jeigu straipsniai būtų saugomi duomenų bazėje ir pasirinkus konkretų kodeksą užkraunami sistemoje.

7. LITERATŪRA

- [1] „Lietuvos Respublikos seimas“, prieiga per:
https://www.smm.lt/uploads/documents/konkursai/kiti_konkursai/Verslumo%20rengini%C5%B3%20tvarka.pdf. [Žiūrėta 2016 10 20].
- [2] LIETUVOS RESPUBLIKOS SEIMAS. Lietuvos Respublikos administracinių nusižengimų įstatymas: 2016-03-25 įstatymu Nr. XII-2275 [Tinkle]. [Žiūrėta 2017-05-15]. Prieiga per:
http://www.infolex.lt/portal/start_ta.asp?act=doc&fr=pop&doc=336765&title=Lietuvos%20Respublikos%20administracini%C5%B3%20nusi%C5%BEengim%C5%B3%20kodeksas.
- [3] „Teisės normų rūšys“ [Tinkle] Prieiga per: <http://www.finaura.lt/?p=396> [Žiūrėta 2018 03 22]
- [4] „Teisės norma“ 15 sausio 2015 [Tinkle] Prieiga per:
https://lt.wikipedia.org/wiki/Teis%C4%97s_norma [Žiūrėta 2018 03 22]
- [5] CiteSeer. Prieiga per: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.97.7985> [Žiūrėta 2017 03 10]
- [6] „Semantic Web Services – Concepts, Technologies, and Applications“. Prieiga per:
https://books.google.lt/books?id=wglTWruS3UC&pg=PA135&lpg=PA135&dq=semantic+annotation+technology+stack&source=bl&ots=E4PsB14Zf1&sig=0TOgv2wIDIQ744KVU35Rzihw3r4&hl=lt&sa=X&ved=0ahUKEwi b2cmZhd_RAhWDfiwKHa6aBMgQ6AEISjAF#v=onepage&q=semantic%20annotation%20technology%20stack&f=false [Žiūrėta 2016 12 15]
- [7] Eyal Oren, Knud Hinnerk Möller, Simon Scerri, Siegfried Handschuh, and Michael Sintek „What are Semantic Annotations?“ 2006 [Tinkle] Prieiga per: <http://www.siegfried-handschuh.net/pub/2006/whatissemannot2006.pdf> [Žiūrėta 2018 04 15]
- [8] LKIF. Prieiga per: <http://www.estrellaproject.org/lkif-core/> [Žiūrėta 2017 03 10]
- [9] „Protege“ prieiga per: <http://protege.stanford.edu/> [Žiūrėta 2016 11 20].
- [10] „TopBraid Composer Maestro Edition“ Prieiga per: <http://www.topquadrant.com/tools/ide-topbraid-composer-maestro-edition/> [Žiūrėta 2016 11 20]
- [11] „OWL 2 Web Ontology Language Profiles“ Prieiga per: <https://www.w3.org/TR/owl2-profiles/> [Žiūrėta 2016 12 01]
- [12] „OWL 2 Web Ontology Language Document Overview“ Prieiga per:
<https://www.w3.org/TR/owl2-overview/>. [Žiūrėta 2016 12 01].
- [13] „Semantic Web Services – Concepts, Technologies, and Applications“. Prieiga per:
<https://books.google.lt/books?id=wglTWruS3UC&pg=PA135&lpg=PA135&dq=semantic+annotation+technology+stack&source=bl&ots=E4PsB14Zf1&sig=0TOgv2wIDIQ744KVU35Rzihw3r4&hl=lt&sa=X&ved=0ahUKEwi>

- [b2cmZhd_RAhWDfiwKHa6aBMgQ6AEISjAF#v=onepage&q=semantic%20annotation%20technology%20stack&f=false](#) [Žiūrėta 2016 12 15]
- [14] „Ontology“ [Tinkle] Prieiga per: <http://web.dfc.unibo.it/buzzetti/IUcorso2007-08/mdidattici/ontology-definition-2007.htm> [Žiūrėta 2018 16 05]
- [15] „Brat rapid teksto anotavimo įrankis“ [Tinkle] Prieiga per: <http://brat.nlplab.org/introduction.html> [2018 03 08]
- [16] Giorgos Giannopoulos, Nikos Bikakis, Theodore Dalamagas, Timos Sellis „GoNTogle: A Tool for Semantic Annotation and Search“ [Tinkle] Prieiga per: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.607.3724&rep=rep1&type=pdf> [Žiūrėta 2018 03 22]
- [17] „Lietuvių kalbos sitaksinės ir semantinės analizės informacinė sistema“. Prieiga per: <http://semantika.lt/> [Žiūrėta 2016 12 10]
- [18] „GATE - general architecture for text engineering“ Prieiga per: <https://gate.ac.uk/projects.html> [Žiūrėta 2016 12 10]
- [19] „Semantic technologies for smarter information retrieval and content management“. Prieiga per: <http://ontotext.com/> [Žiūrėta 2016 12 10]
- [20] “Ontological Model of Legal Norms for Creating and Using Legislation” Stevan Gostojić, Branko Milosavljević, and Zora Konjović DOI:10.2298/CSIS110804035G [Žiūrėta 2017 04 15]
- [21] „Precision and recall“ [Tinkle] Prieiga per: https://en.wikipedia.org/wiki/Precision_and_recall [Žiūrėta 2018 04 15]
- [22] LIETUVOS RESPUBLIKOS SEIMAS. Lietuvos Respublikos civilinis kodekso įstatymas: 2000-09-06 įstatymu Nr. VIII-1864 [Tinkle]. [žiūrėta 2018-03-18]. Prieiga per: <http://www.infolex.lt/ta/100228>
- [23] LIETUVOS RESPUBLIKOS SEIMAS. Lietuvos Respublikos darbo kodekso įstatymas: 2016-09-19 XII-2603 [Tinkle]. [žiūrėta 2017 06 11]. Prieiga per: http://www.infolex.lt/portal/start_ta.asp?act=doc&fr=pop&doc=368200&title=LR%20darbo%20kodeksas

8.PRIEDAI

8.1. Priedas. 2.4 skyriuje pateiktų taisyklių pavyzdžiai straipsniuose

8.1.1. 95. straipsnis. Nelegalus darbas (Administracinių nusižengimų kodeksas)

1. Nelegalus darbas

užtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo vieno tūkstančio iki penkių tūkstančių eurų.

2. Šio straipsnio 1 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai, užtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo penkių tūkstančių iki šešių tūkstančių eurų.

8.1.2. 45 straipsnis. Visuomenės sveikatos srities Europos Sąjungos reglamentų ar sprendimų, higienos norminių aktų ar kitų visuomenės sveikatos srities teisės aktų, Lietuvos Respublikos žmonių užkrečiamųjų ligų profilaktikos ir kontrolės įstatymo pažeidimas. (Administracinių nusižengimų kodeksas)

1. Visuomenės sveikatos srities Europos Sąjungos reglamentų ar sprendimų, higienos norminių aktų ar kitų visuomenės sveikatos srities teisės aktų, Lietuvos Respublikos žmonių užkrečiamųjų ligų profilaktikos ir kontrolės įstatymo pažeidimas užtraukia įspėjimą arba baudą asmenims nuo šešiasdešimt iki vieno šimto keturiasdešimt eurų ir baudą juridinių asmenų vadovams ar kitiems atsakingiems asmenims nuo vieno šimto keturiasdešimt iki šešių šimtų eurų.

2. Šio straipsnio 1 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai, užtraukia baudą asmenims nuo vieno šimto keturiasdešimt iki šešių šimtų eurų ir juridinių asmenų vadovams arba kitiems atsakingiems asmenims – nuo penkių šimtų penkiasdešimt iki vieno tūkstančio dviejų šimtų eurų.

3. Šio straipsnio 1 dalyje nurodytos veikos, sukėlusios pavojų išplisti pavojingoms ar ypač pavojingoms užkrečiamosioms ligoms, užtraukia baudą asmenims nuo trijų šimtų iki penkių šimtų šešiasdešimt eurų ir juridinių asmenų vadovams ar kitiems atsakingiems asmenims – nuo vieno tūkstančio keturių šimtų iki trijų tūkstančių eurų.

8.1.3. 49 straipsnis. Biocidinių produktų tiekimo rinkai, jų naudojimo, informacijos kaupimo ir (ar) saugojimo reikalavimų pažeidimas (Administracinių nusižengimų kodeksas)

1. Reglamente (ES) Nr. 528/2012 nustatyta tvarka neįteisintų biocidinių produktų naudojimas ar įteisintų biocidinių produktų naudojimas pažeidžiant Reglamente (ES) Nr. 528/2012 nustatytus reikalavimus

užtraukia įspėjimą arba baudą asmenims nuo trisdešimt iki vieno šimto keturiasdešimt eurų ir baudą juridinių asmenų vadovams ar kitiems atsakingiems asmenims – nuo dviejų šimtų keturiasdešimt iki vieno tūkstančio penkių šimtų eurų.

2. Šio straipsnio 1 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai,

užtraukia baudą asmenims nuo penkiasdešimt iki trijų šimtų eurų ir baudą juridinių asmenų vadovams ar kitiems atsakingiems asmenims – nuo keturių šimtų aštuoniasdešimt iki trijų tūkstančių eurų.

3. Biocidinių produktų klasifikavimo, pakavimo, ženklinimo reikalavimų, nustatytų Reglamente (ES) Nr. 528/2012, pažeidimas, išskyrus šio kodekso 308 straipsnio 16 dalyje numatytus pažeidimus, užtraukia baudą nuo trijų šimtų keturiasdešimt iki keturių tūkstančių trijų šimtų eurų.

4. Reglamente (ES) Nr. 528/2012 nustatytų reikalavimų biocidinių produktų gamintojui, autorizacijos liudijimų turėtojui kaupiti ir (ar) saugoti informaciją pažeidimas užtraukia baudą nuo penkių šimtų šešiasdešimt iki vieno tūkstančio dviejų šimtų eurų.

5. Biocidinių produktų tiekimo rinkai reikalavimų, nustatytų Reglamente (ES) Nr. 528/2012, pažeidimas, išskyrus šio kodekso 308 straipsnio 10 dalyje numatytus pažeidimus, užtraukia baudą nuo vieno tūkstančio penkių šimtų iki keturių tūkstančių trijų šimtų eurų.

8.1.4. 25 straipsnis. Teisingas informavimas ir konfidencialios informacijos apsauga (Darbo kodeksas)

1. Darbo sutarties šalys privalo viena kitai laiku pranešti apie bet kokias aplinkybes, galinčias reikšmingai paveikti sutarties sudarymą, vykdymą ir nutraukimą. Ši informacija turi būti pateikta teisinga, nemokamai ir darbo sutarties šalių nustatytais protingais terminais.

2. Darbo sutarties šalies kitai darbo sutarties šaliai šio kodekso, kitų darbo teisės normų ar sutarčių nustatytais atvejais perduodami dokumentai (pranešimai, prašymai, sutikimai, prieštaravimai ir kita) ir kita informacija turi būti pateikiami raštu. Dokumentų ir informacijos tinkamu pateikimu raštu laikomi tie atvejai, kada duomenys perduodami įprastai naudojamomis informacinių technologijų priemonėmis (elektroniniu paštu, mobiliaisiais įrenginiais ir kita) su sąlyga, kad įmanoma nustatyti informacijos turinį, jos pateikėją, pateikimo faktą ir laiką, taip pat sudarytos protingos galimybės ją išsaugoti. Jeigu darbo sutarties šalis nurodo pagrįstas abejones dėl šių abiejų sąlygų buvimo, įrodyti, kad jos buvo sudarytos, privalo darbdavys.

3. Darbo sutartis ir darbo teisės normos turi būti išdėstytos lietuvių kalba arba lietuvių kalba ir kita sutarties šalims priimtina kalba.

4. Šio straipsnio 1 ir 2 dalyse nurodytais atvejais informacija turi būti perduota lietuvių kalba. Prie jos gali būti pridedami vertimai į vieną ar kelias kalbas.

5. Pareigą saugoti konfidencialią informaciją (informaciją, kuri laikoma komercine (gamybine), profesine, valstybės ar tarnybos paslaptimi) ir atsakomybę už jos pažeidimą reglamentuoja įstatymai. Šio kodekso nustatyta tvarka darbo sutarties šalys gali sudaryti papildomus susitarimus dėl konfidencialios informacijos apsaugos.

8.1.5. 29 straipsnis. Darbuotojo profesinio tobulėjimo siekio gerbimas (Darbo kodeksas)

1. Darbdavys privalo apmokyti darbuotoją dirbti tiek, kiek tai būtina jo darbo funkcijai atlikti.

2. Darbdavys privalo imtis priemonių darbuotojų kvalifikacijai ir jų profesionalumui, gebėjimui prisitaikyti prie besikeičiančių verslo, profesinių ar darbo sąlygų didinti. Šiais tikslais šiame kodekse, darbo teisės normų ar šalių susitarimų nustatytais atvejais ir tvarka darbdavys sudaro sąlygas darbuotojui mokytis, tobulinti kvalifikaciją ir profesiskai tobulėti.

8.1.6. 37 straipsnis. Susitarimas dėl mokymo išlaidų atlyginimo (Darbo kodeksas)

1. Darbo sutarties šalys gali susitarti dėl darbdavio turėtų darbuotojo mokymo ar kvalifikacijos tobulinimo išlaidų atlyginimo sąlygų, kai darbo sutartis nutraukiama darbdavio iniciatyva dėl darbuotojo kaltės arba darbuotojo iniciatyva be svarbių priežasčių.
2. Atlygintos gali būti tik išlaidos, susijusios su darbuotojo žinių ar gebėjimų, viršijančių darbo veiklai keliamus reikalavimus, suteikimu. Susitarime gali būti nustatyta, ar į mokymo ar kvalifikacijos tobulinimo išlaidas įskaičiuojamos kitos komandiruotės išlaidos (kelionės, nakvynės ir kita).
3. Atlygintos gali būti tik darbdavio turėtos išlaidos per paskutinius dvejus metus iki darbo sutarties pasibaigimo, nebent kolektyvinėje sutartyje nustatytas kitoks terminas, kuris negali viršyti trejų metų.
4. Jeigu darbuotojas savo iniciatyva studijuoja, siekdamas studijų krypties bakalauro, magistro kvalifikacinio laipsnio ir (arba) profesinės kvalifikacijos pagal formaliojo profesinio mokymo programas, ir darbdavys apmoka visas ar ne mažiau kaip pusę šių išlaidų, darbo sutarties šalys papildomai gali susitarti dėl to, kad darbuotojas laikotarpiu, iki tęsia darbdavio lėšomis apmokėtas studijas, ir per trejus metus po šio laikotarpio pabaigos darbo sutartį savo iniciatyva be svarbių priežasčių gali nutraukti tik atlyginęs darbdaviui jo patirtas išlaidas.

8.1.7. 67 straipsnis. Terminuotos darbo sutarties sąvoka ir jos terminas (Darbo kodeksas)

1. Terminuota darbo sutartis – darbo sutartis, sudaroma tam tikram laikui arba tam tikrų darbų atlikimo laikui.
2. Terminuotos darbo sutarties terminas gali būti nustatomas iki tam tikros kalendorinės datos, tam tikram dienomis, savaitėmis, mėnesiais ar metais skaičiuojamam terminui, iki tam tikros užduoties įvykdymo ar tam tikrų aplinkybių atsiradimo, pasikeitimo ar pasibaigimo.
3. Terminuota darbo sutartis tampa neterminuota, kai darbo santykių buvimo laikotarpiu išnyksta aplinkybės, dėl kurių buvo apibrėžtas sutarties terminas.
4. Terminuotų darbo sutarčių nuolatinio pobūdžio darbams negali būti daugiau kaip dvidešimt procentų visų darbdavio sudarytų darbo sutarčių skaičiaus.

8.1.8. 99 straipsnis. Darbo keliams darbdaviams sutarties pasibaigimas (Darbo kodeksas)

1. Darbuotojas turi teisę nutraukti darbo keliams darbdaviams sutartį šio kodekso nustatytais pagrindais ir tvarka. Pareiškimas nutraukti darbo keliams darbdaviams sutartį yra pateikiamas pirmajam darbdaviui.
2. Bet kuris darbdavys turi teisę inicijuoti darbo keliams darbdaviams sutarties nutraukimą šio kodekso nustatytais pagrindais ir tvarka, pranešdamas apie tai visoms darbo keliams darbdaviams sutarties šalims. Sprendimą nutraukti darbo keliams darbdaviams sutartį ar siūlyti darbuotojui pakeisti darbo sutartį priima visi darbdaviai kartu, o jiems santykiuose su darbuotoju atstovauja pirmasis darbdavys. Atleidžiamam darbuotojui išmokamos sumos tarp darbdavių padalijamos proporcingai kiekvieno darbdavio apmokamai darbo laiko normos daliai.
3. Jeigu darbo keliams darbdaviams sutartis nėra pakeičiama, nutraukus darbo keliams darbdaviams sutartį, darbuotojo darbo santykiai pasibaigia su visais darbdaviais.

8.1.9. 23 straipsnis. Informacijos apie darbo santykių būklę teikimas (Darbo kodeksas)

1. Šio kodekso, kitų įstatymų ir kitų darbo teisės normų nustatyta tvarka darbdavys privalo teikti informaciją apie darbuotojus ir jų darbo sąlygas ar kitus darbo santykių aspektus juose nurodytoms kompetentingoms institucijoms.
2. Darbdavys, kurio vidutinis darbuotojų skaičius yra daugiau kaip dvidešimt, darbo tarybai ir profesinei sąjungai privalo teikti bent kartą per metus atnaujinamą informaciją apie:
 - 1) darbuotojų, išskyrus vadovaujančias pareigas einančius darbuotojus, nuasmenintus duomenis apie vidutinį darbo užmokestį pagal profesijų grupes ir lytį;
 - 2) kitą informaciją, kurią skelbti įpareigoja įstatymai, kolektyvinės sutartys, darbdavio ir darbo tarybos susitarimai.

8.1.10. 66 straipsnis. Farmacijos praktikos, su vaistais (vaistiniais preparatais) ar veikliosiomis medžiagomis susijusios veiklos sąlygų pažeidimas (Administracinių nusižengimų kodeksas)

1. Vaistų (vaistinių preparatų) prekybos tarpininkavimas nesilaikant nustatytų šios veiklos sąlygų užtraukia baudą nuo trisdešimt iki vieno šimto keturiasdešimt eurų.
2. Farmacijos praktikos sąlygų pažeidimas užtraukia baudą nuo trisdešimt iki dviejų šimtų devyniasdešimt eurų.
3. Šio straipsnio 2 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai, užtraukia baudą nuo dviejų šimtų aštuoniasdešimt iki šešių šimtų eurų.
4. Veikliųjų medžiagų, įskaitant eksportui skirtas veikliąsias medžiagas, gamyba, importas iš trečiųjų šalių ir platinimas nesilaikant nustatytų veiklos sąlygų užtraukia baudą nuo dviejų šimtų šešiasdešimt iki vieno tūkstančio penkių šimtų eurų.
5. Vaistų (vaistinių preparatų) pardavimo (išdavimo) gyventojams sąlygų nesilaikymas užtraukia baudą farmacinės veiklos vadovams nuo penkių šimtų iki dviejų tūkstančių keturių šimtų eurų.
6. Vaistinės veikla nesilaikant nustatytų licencijuojamos veiklos sąlygų, išskyrus pažeidimus, dėl kurių sustabdomas ar panaikinamas licencijos galiojimas, užtraukia baudą vaistinės veiklos ir gamybinės vaistinės veiklos licenciją turinčių juridinių asmenų vadovams nuo šešių šimtų iki dviejų tūkstančių trijų šimtų eurų.
7. Teisės aktuose nustatytų kvalifikuoto asmens arba farmacinės veiklos vadovo pareigų nevykdymas ar netinkamas vykdymas užtraukia baudą kvalifikuotiems asmenims arba farmacinės veiklos vadovams nuo aštuonių šimtų iki trijų tūkstančių eurų.
8. Vaistų (vaistinių preparatų) gamyba, žmogaus kraujo plazmos ruošimas, didmeninis platinimas nesilaikant nustatytų licencijuojamos veiklos sąlygų, išskyrus pažeidimus, dėl kurių sustabdomas ar panaikinamas licencijos galiojimas, užtraukia baudą farmacinės veiklos licenciją turinčių juridinių asmenų vadovams nuo aštuonių šimtų iki trijų tūkstančių eurų.
9. Už šio straipsnio 4, 5 dalyse numatytus administracinius nusižengimus gali būti skiriamas vaistų (vaistinių preparatų) ar veikliųjų medžiagų konfiskavimas.

8.1.11. 100 straipsnis. Darbo laiko apskaitos pažeidimas (Administracinių nusižengimų kodeksas)

1. Darbuotojų darbo laiko nežymėjimas darbo laiko apskaitos žiniaraštyje arba žinomai neteisingų duomenų apie įmonėse, įstaigose, organizacijose pagal darbo sutartis dirbančių asmenų darbo laiką (įskaitant viršvalandinius darbus, darbą nakties metu, poilsio ir švenčių dienomis, kai yra nukrypimų nuo normalių darbo sąlygų) įrašymas į darbo laiko apskaitos žiniaraštį užtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo vieno šimto penkiasdešimt iki vieno tūkstančio keturių šimtų penkiasdešimt eurų.
2. Šio straipsnio 1 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai, užtraukia baudą nuo vieno tūkstančio keturių šimtų iki trijų tūkstančių eurų.

8.1.12. 180 straipsnis. Darbdavių atstovavimas socialinėje partnerystėje darbdavio lygmeniu (Darbo kodeksas)

1. Darbdavys – fizinis asmuo – socialinėje partnerystėje dalyvauja, teises ir pareigas prisiima pats.
2. Darbdaviui – juridiniam asmeniui – socialinėje partnerystėje darbdavio lygmeniu atstovauja juridinio asmens vienasmenis valdymo organas arba jo įgalioti asmenys.

8.1.13. 9 straipsnis. Komandiruojamų darbuotojų darbo santykiams taikytina teisė (Darbo kodeksas)

1. Lietuvos Respublikos jurisdikcijai priklausančio darbdavio pavedimu laikinai dirbti užsienyje išsiųsto darbuotojo darbo santykiams Lietuvos Respublikos darbo teisės normos taikomos tiek, kiek jų nereglamentuoja tos užsienio valstybės, į kurią siunčiama laikinai dirbti, imperatyvios nuostatos. Į užsienio valstybę komandiruojamų darbuotojų darbo sąlygų ypatumus taip pat nustato kitos šio kodekso normos.
2. Laikinai į Lietuvos Respublikos teritoriją dirbti pagal užsienio valstybės jurisdikcijai priklausančio darbdavio pavedimą atsiųstam darbuotojui jo darbo sutarčiai taikytina teisė taikoma tiek, kiek jo darbo imperatyviai nereglamentuoja Lietuvos Respublikos darbo teisės normos.
3. Imperatyviomis nuostatomis laikomi įstatymai ir kiti teisės aktai ir (arba) kolektyvinės sutartys ar arbitražo sprendimai, kurie buvo paskelbti visuotinai taikytiniais ir nuo kurių pagal taikytinos teisės valstybės teisės aktus negalima nukrypti šalių susitarimu.

8.1.14. 23 straipsnis. Informacijos apie darbo santykių būklę teikimas (Darbo kodeksas)

1. Šio kodekso, kitų įstatymų ir kitų darbo teisės normų nustatyta tvarka darbdavys privalo teikti informaciją apie darbuotojus ir jų darbo sąlygas ar kitus darbo santykių aspektus juose nurodytoms kompetentingoms institucijoms.
2. Darbdavys, kurio vidutinis darbuotojų skaičius yra daugiau kaip dvidešimt, darbo tarybai ir profesinei sąjungai privalo teikti bent kartą per metus atnaujinamą informaciją apie:
 - 1) darbuotojų, išskyrus vadovaujančias pareigas einančius darbuotojus, nuasmenintus duomenis apie vidutinį darbo užmokestį pagal profesijų grupes ir lytį;

2) kitą informaciją, kurią skelbti įpareigoja įstatymai, kolektyvinės sutartys, darbdavio ir darbo tarybos susitarimai.

8.1.15. 115 straipsnis. Darbo laiko režimas taikant suminę darbo laiko apskaitą

1. Jeigu nustatyta suminė darbo laiko apskaita, dirbama darbo (pamainų) grafikuose nurodytu laiku, laikantis nustatytų maksimaliojo darbo laiko reikalavimų.
2. Darbo (pamainų) grafikai pranešami darbuotojams ne vėliau kaip prieš penkias darbo dienas iki jų įsigaliojimo. Jie gali būti keičiami tik išimtiniais atvejais, įspėjus darbuotoją prieš dvi darbuotojo darbo dienas. Darbo (pamainų) grafikus tvirtina administracija, darbo (pamainų) grafikų derinimo tvarką suderinusi su įmonėje veikiančia darbo taryba (darbuotojų atstovais) arba kolektyvinėje sutartyje nustatyta tvarka.
3. Darbdavys privalo užtikrinti tolygų darbuotojų keitimąsi pamainomis. Teisę pasirinkti pamainą per dvi darbo dienas nuo jų pranešimo turi asmenys, auginantys vaiką iki trejų metų, o asmenys, auginantys vaiką iki septynerių metų, jeigu yra tokia galimybė.
4. Darbdavys privalo sudaryti darbo (pamainų) grafikus taip, kad kiek įmanoma tolygiau paskirstytų darbuotojo darbo laiką per apskaitinį laikotarpį. Draudžiama skirti darbuotoją dirbti dvi pamainas iš eilės.
5. Jeigu apskaitinio laikotarpio pabaigoje darbuotojas dėl jam sudaryto darbo laiko režimo nėra išdirbęs bendros viso apskaitinio laikotarpio darbo laiko normos, už neįvykdytą darbo laiko normą jam sumokama pusė priklausančio išmokėti darbo užmokesčio.
6. Jeigu apskaitinio laikotarpio pabaigoje darbuotojas yra išdirbęs daugiau valandų, negu bendra viso apskaitinio laikotarpio darbo laiko norma, už viršytą darbo laiko normą jam apmokama kaip už viršvalandinį darbą arba darbuotojo prašymu viršytas darbo laikas, padaugintas iš skaičiaus 1,5, gali būti pridedamas prie kasmetinių atostogų laiko.
7. Dirbant pagal suminę darbo laiko apskaitą, darbo užmokestis mokamas už faktiškai išdirbtą laiką, išskyrus šio straipsnio 5 ir 6 dalyse nustatytus atvejus. Darbdavys turi teisę per kiekvieną apskaitinio laikotarpio mėnesį mokėti pastovų darbo užmokestį, nepaisydamas faktiškai išdirbtos darbo laiko normos, o galutinį atsiskaitymą už darbą per apskaitinį laikotarpį pagal faktinius duomenis atlikti apmokėdamas už darbą paskutinį apskaitinio laikotarpio mėnesį.

8.1.16. 97 straipsnis. Nelaimingo atsitikimo darbe nuslėpimas, nustatytos pranešimo ir ištyrimo tvarkos pažeidimas (Administracinių nusižengimų kodeksas)

1. Pranešimo apie nelaimingus atsitikimus darbe ar profesines ligas arba nelaimingų atsitikimų darbe ir profesinių ligų ištyrimo tvarkos pažeidimas užtraukia baudą darbdaviams ar juridinių asmenų vadovams nuo devyniasdešimt iki penkių šimtų devyniasdešimt eurų ir kitiems atsakingiems asmenims – nuo dvidešimt iki trisdešimt eurų.
2. Nelaimingo atsitikimo darbe nuslėpimas užtraukia baudą darbdaviams ar juridinių asmenų vadovams nuo trijų šimtų iki vieno tūkstančio keturių šimtų penkiasdešimt eurų ir kitiems atsakingiems asmenims – nuo vieno šimto keturiasdešimt iki septynių šimtų keturiasdešimt eurų.

8.1.17. 99 straipsnis. Darbo užmokesčio apskaičiavimo ir mokėjimo tvarkos pažeidimas (Administracinių nusižengimų kodeksas)

1. Lietuvos Respublikos darbo kodekse, kolektyvinėje arba darbo sutartyje nustatytos darbo užmokesčio apskaičiavimo ir mokėjimo tvarkos pažeidimas užtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo vieno šimto penkiasdešimt iki vieno tūkstančio keturių šimtų penkiasdešimt eurų.
2. Šio straipsnio 1 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai, užtraukia baudą nuo vieno tūkstančio keturių šimtų iki trijų tūkstančių eurų.
3. Tyčinis Lietuvos Respublikos darbo kodekse, kolektyvinėje arba darbo sutartyje nustatytos darbo užmokesčio apskaičiavimo ir mokėjimo tvarkos pažeidimas arba darbo užmokesčio ir kitų su darbo santykiais susijusių išmokų, neįtrauktų į buhalterinės apskaitos dokumentus, išmokėjimas užtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo dviejų tūkstančių septynių šimtų iki šešių tūkstančių eurų.

8.1.18. 46 straipsnis. Darbo sąlygų keitimas darbuotojo iniciatyva (Darbo kodeksas)

1. Kai šis kodeksas ar kitos darbo teisės normos nesuteikia darbuotojui teisės reikalauti pakeisti darbo sąlygas, darbuotojas turi teisę prašyti darbdavio pakeisti jo darbo sąlygas.
2. Atsisakymas patenkinti darbuotojo raštu pateiktą prašymą pakeisti būtinąsias ar darbo sutarties šalių suligtas papildomas darbo sutarties sąlygas turi būti motyvuotas ir pateiktas raštu ne vėliau kaip per penkis darbo dienas nuo darbuotojo prašymo pateikimo.
3. Darbdaviui atsisakius tenkinti darbuotojo prašymą pakeisti darbo sąlygas, dėl šių sąlygų pakeitimo darbuotojas pakartotinai gali kreiptis ne anksčiau kaip po vieno mėnesio nuo darbuotojo prašymo pakeisti darbo sąlygas pateikimo.
4. Darbdaviui sutikus su darbuotojo prašymu ar darbdaviui pateikus kitą pasiūlymą ir darbuotojui sutikus, laikoma, kad darbo sąlygos yra pakeistos, padarius atitinkamą darbo sutarties pakeitimą.

8.1.19. 102 straipsnis. Lietuvos Respublikos Europos darbo tarybų įstatymo pažeidimas (Administracinių nusižengimų kodeksas)

1. Lietuvos Respublikos Europos darbo tarybų įstatymo pažeidimas užtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo trisdešimt iki dviejų šimtų devyniasdešimt eurų.
2. Šio straipsnio 1 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai, užtraukia baudą nuo trijų šimtų iki penkių šimtų aštuoniasdešimt eurų.

8.2. Anotacijos JSON failas

```
{"document": {"name": "Administracinių nusižengimų kodeksas", "part": "1 Dalis. BENDROJI DALIS", "section": "1", "Skyrius. BENDROSIOS NUOSTATOS"}, "annotations": [{"version": "lexv1.0", "annotation": {"seg": "[0,3],[4,10],[14,1],[16,5],[22,5],[28,9],[3
```

8,10],[50,1],[51,1],[53,10],[64,5],[70,5],[76,11],[88,5],[94,5],[100,9],[110,12],[123,4],[128,7],[136,10],[147,7],[155,4],[160,7],[167,1],[169,9],[178,1],[180,14],[195,5],[201,5],[207,8],[216,9],[226,6],[233,5],[239,5],[245,1],[246,9],[256,15],[272,6],[278,1],[280,5],[286,7],[294,4],[298,1],[300,7],[308,2],[311,7],[319,8],[327,1],[329,3],[333,3],[337,9],[347,3],[351,8],[360,5],[366,6],[372,1],[374,8],[383,1],[385,5],[391,5],[397,9],[407,10],[419,9],[429,5],[435,11],[447,2],[450,7],[458,12],[471,8],[480,3],[484,5],[490,5],[496,13],[510,3],[514,5],[520,10],[531,7],[539,5],[545,13],[559,4],[563,1],[566,1],[567,1],[569,3],[573,10],[584,1],[586,6],[593,9],[603,15],[619,12],[631,1],[633,9],[643,12],[655,1],[658,9],[668,5],[674,3],[678,5],[684,10],[695,7],[703,5],[709,3],[713,5],[719,10],[730,4],[734,1]],"s":[[0,15],[16,32],[50,367],[419,145],[566,90],[658,77]],"p":[[0,48],[50,367],[419,145],[566,90],[658,77]]}},{"version":"morphologyv1.0","annotation":{"msd":[[["100","M----d-"],["straipsnis","Ncmsnn-"],[".","Tp"],["darbas","Ncmsgn-"],["laikas","Ncmsgn-"],["laikyti","Vgmp3s--n--ni-"],["laikyti","Vgmp3p--n--ni-"],["laikyti","Vgmp3---n--ni-"],["apskaita","Ncfsgn-"],["apskaita","Ncfpnn-"],["apskaita","Ncfpvn-"],["pažeidimas","Ncmsnn-"],["1","M----d-"],[".","Tp"],["darbuotoja","Ncfpgn-"],["darbuotojas","Ncmpgn-"],["darbas","Ncmsgn-"],["laikas","Ncmsgn-"],["laikyti","Vgmp3s--n--ni-"],["laikyti","Vgmp3p--n--ni-"],["laikyti","Vgmp3---n--ni-"],["nežymėjimas","Ncmsnn-"],["darbas","Ncmsgn-"],["laikas","Ncmsgn-"],["laikyti","Vgmp3s--n--ni-"],["laikyti","Vgmp3p--n--ni-"],["laikyti","Vgmp3---n--ni-"],["apskaita","Ncfsgn-"],["apskaita","Ncfpnn-"],["apskaita","Ncfpvn-"],["žiniaraštis","Ncmsln-"],["arba","Cg"],["arba","Qg"],["žinomai","Rgp"],["žinoti","Vgpp-sfpnndn-p"],["žinomas","Agpfdsn"],["neteisingas","Agpmpgn"],["neteisingas","Agpfpgn"],["duomuo","Ncmpgn-"],["apie","Sga"],["įmonė","Ncfpln-"],["","Tc"],["įstaiga","Ncfpln-"],["","Tc"],["organizacija","Ncfpln-"],["pagal","Sga"],["darbas","Ncmsgn-"],["sutartis","Ncfpan-"],["sutartis","Ncfsnn-"],["dirbti","Vgpp-pmanngn-p"],["dirbti","Vgpp-pfanngn-p"],["asmuo","Ncmpgn-"],["darbas","Ncmsgn-"],["laikas","Ncmsan-"],["laika","Ncfsan-"],["laikyti","Vgpp-pmannnn-p"],["laikyti","Vgpp-pfannnn-p"],["laikyti","Vgpp-nann-n-p"],["laikyti","Vgpp-pfannvn-p"],["laikyti","Vgpp-pmannvn-p"],["(","Tl"],["įskaityti","Vgap----n--n-"],["viršvalandinis","Agpmpan"],["darbas","Ncmpan-"],["","Tc"],["darbas","Ncmsan-"],["naktis","Ncfsgn-"],["metas","Ncmsin-"],["mesti","Vgmp1s--n--ni-"],["","Tc"],["poilsis","Ncmsgn-"],["ir","Cg"],["ir","Qg"],["šventė","Ncfpgn-"],["diena","Ncfpin-"],["","Tc"],["kai","Cg"],["kai","Qg"],["būti","Vgmp3s--n--ni-"],["būti","Vgmp3p--n--ni-"],["irti","Vgmp3s--n--ni-"],["būti","Vgmp3---n--ni-"],["irti","Vgmp3p--n--ni-"],["irti","Vgmp3---n--ni-"],["nukrypimas","Ncmpgn-"],["nuo","Sgg"],["normalus","Agpfpgn"],["normalus","Agpmpgn"],["normalė","Ncfpgn-

]],["darbas","Ncmmsgn-"],["sąlyga","Ncfpgn-"],["","Tr"],["įrašymas","Ncmsnn-
]],["į","Sga"],["darbas","Ncmmsgn-"],["laikas","Ncmmsgn-"],["laikyti","Vgmp3s--n--ni-
 "],["laikyti","Vgmp3p--n--ni-"],["laikyti","Vgmp3---n--ni-"],["apskaita","Ncfsgn-
 "],["apskaita","Ncfpnn-"],["apskaita","Ncfpvn-"],["žiniaraštis","Ncmsan-
 "],["užtraukti","Vgmp3s--n--ni-"],["užtraukti","Vgmp3p--n--ni-"],["užtraukti","Vgmp3---n--ni-
 "],["bauda","Ncfstan-"],["darbdavys","Ncmprdn-
 "],["ar","Cg"],["ar","Qg"],["ar","Og"],["kitas","Pgmpdn"],["atsakingas","Agpmpdn"],["asmuo
 ","Ncmprdn-"],["nuo","Sgg"],["vienas","Mcmmsgl-
 "],["vienas","Pgmsgn"],["vienas","Agpmsgn"],["šimtas","Mcmmsgl-
 "],["šimtas","Momsgln"],["penkiasdešimt","Mc---l-
 "],["iki","Sgg"],["iki","Cg"],["iki","Rgp"],["vienas","Mcmmsgl-
 "],["vienas","Pgmsgn"],["vienas","Agpmsgn"],["tūkstantis","Mcmmsgl-"],["keturi","Mcm-gl-
 "],["keturi","Mcf-gl-"],["šimtas","Mcmprgl-
 "],["šimtas","Mofpgln"],["šimtas","Mompgrln"],["penkiasdešimt","Mc---l-"],["euras","Ncmprgn-
 "],["","Tp"],["2","M---d-"],["","Tp"],["šis","Pgmsgn"],["straipsnis","Ncmmsgn-"],["1","M-
 ---d-"],["dalis","Ncfsln-"],["numatyti","Vgps-pfpnnan-p"],["numatyti","Vgps-smpnnvn-
 p"],["numatyti","Vgps-smpnnnn-p"],["administracinis","Agpmsnn"],["nusižengimas","Ncmsnn-
 "],["","Tc"],["padaryti","Vgps-pfpnnan-p"],["padaryti","Vgps-smpnnvn-p"],["padaryti","Vgps-
 smpnnnn-p"],["pakartotinai","Rgp"],["pakartoti","Vgp--sfnnndn-p"],["pakartoti","Vgb-----n--n-
 "],["","Tc"],["užtraukti","Vgmp3s--n--ni-"],["užtraukti","Vgmp3p--n--ni-"],["užtraukti","Vgmp3-
 --n--ni-"],["bauda","Ncfstan-"],["nuo","Sgg"],["vienas","Pgmsgn"],["vienas","Mcmmsgl-
 "],["vienas","Agpmsgn"],["tūkstantis","Mcmmsgl-"],["keturi","Mcm-gl-"],["keturi","Mcf-gl-
 "],["šimtas","Mcmprgl-
 "],["šimtas","Mofpgln"],["šimtas","Mompgrln"],["iki","Sgg"],["iki","Cg"],["iki","Rgp"],["trys","M
 cm-gl-"],["trys","Mcf-gl-"],["tūkstantis","Mcmprgl-"],["euras","Ncmprgn-
 "],["","Tp"]],{"stem":["100","straipsn","","Darb","laik","apskait","pažeidim","1","","Darbuot","d
 arb","laik","nežymėjim","darb","laik","apskait","žiniarašt","arb","žinom","neteising","duomen","api
 ","įmon","","įstaig","","organizacij","pagal","darb","sutart","dirb","asmen","darb","laik","(","įskait
 ant","viršvalandin","darb","","darb","nakt","met","","poils","ir","švent","dien","","kai","yr","nukry
 pim","nes","normal","darb","sąlyg","")","įraš","į","darb","laik","apskait","žiniarašt","užtrauk","baud",
 "darbdav","ar","kit","atsaking","asmen","nes","vien","šimt","penkiasdešimt","ik","vien","tūkst","ket
 ur","šimt","penkiasdešimt","eur","","2","","Šio","straipsn","1","dal","numatyt","administracin","nu
 sižengim","","padaryt","pakartotin","","užtrauk","baud","nes","vien","tūkst","ketur","šimt","ik","tri
 j","tūkst","eur","."]}},{"version":"teis1.0beta","annotation":{"legal":[{"type":"disposition","ref":[[62
 ,2],[94,2]]},{type":"sanction","ref":[[69,11],[96,9]]},{type":"subject","ref":[[64,5],[31,1],[68,1]]},


```
{ "type": "object", "ref": [[3,3],[13,3],[58,3],[53,2],[63,1],[95,1]] }, { "type": "hypothesis", "ref": [[9,53],[83,11]] }, "doc": [ { "type": "article", "ref": [[0,106]] }, { "type": "article_name", "ref": [[0,7]] } ] }, "lang": { "lithuanian": true }, "enc": { "detected_encoding": "utf-8", "UTF-8": "100 straipsnis. Darbo laiko apskaitos pažeidimas\r\n1. Darbuotojų darbo laiko nežymėjimas darbo laiko apskaitos žiniaraštyje arba žinomai neteisingų duomenų apie įmonėse, įstaigose, organizacijose pagal darbo sutartis dirbančių asmenų darbo laiką (įskaitant viršvalandinius darbus, darbą nakties metu, poilsio ir švenčių dienomis, kai yra nukrypimų nuo normalių darbo sąlygų) įrašymas į darbo laiko apskaitos žiniaraštį\r\nnužtraukia baudą darbdaviams ar kitiems atsakingiems asmenims nuo vieno šimto penkiasdešimt iki vieno tūkstančio keturių šimtų penkiasdešimt eurų.\r\n2. Šio straipsnio 1 dalyje numatytas administracinis nusižengimas, padarytas pakartotinai,\r\nnužtraukia baudą nuo vieno tūkstančio keturių šimtų iki trijų tūkstančių eurų.", "error": null } }
```

8.3. Eksperimento duomenys

8.3.1. Administracinių nusižengimų kodeksas

8.1. lentelė. Sankcijos skaičiavimai

Sankcija					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
47	6	6	7	1	0
50	2	2	2	0	0
57	3	3	3	0	0
63	10	10	11	1	0
85	9	11	11	2	0
95	2	2	2	0	0
96	4	4	4	0	0
97	3	4	4	1	0
98	3	3	3	0	0
99	3	3	3	0	0
100	2	2	2	0	0
101	4	4	4	0	0
102	2	2	2	0	0
103	4	4	4	0	0
104	4	4	4	0	0
105	4	4	4	0	0
106	1	2	2	1	0
107	2	2	2	0	0
125	2	2	4	2	0
130	9	9	9	0	0
139	15	17	17	2	0
208	3	3	4	1	0
226	1	1	2	1	0
235	12	15	15	3	0
299	4	4	4	0	0

308	26	26	26	0	0
353	12	12	12	0	0
370	2	2	2	0	0
551	2	2	2	0	0
554	1	1	1	0	0
600	0	0	1	1	0

8.2. lentelė. Subjekto skaičiavimai

Subjektas					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
47	9	9	11	2	3
50	0	0	0	0	0
57	8	9	9	1	1
63	0	0	0	0	0
85	11	11	26	15	1
95	4	4	4	0	0
96	7	7	11	4	5
97	6	6	6	0	0
98	8	8	12	4	6
99	4	4	4	0	0
100	1	3	3	2	2
101	5	5	9	4	4
102	2	2	4	2	0
103	5	5	9	4	4
104	5	5	8	3	3
105	5	5	8	3	3
106	2	2	2	0	0
107	2	2	2	0	0
125	0	0	0	0	0
130	15	15	25	10	10
139	26	26	26	0	0
208	0	0	6	6	0
226	2	2	4	2	2
235	15	15	15	0	0
299	0	0	4	4	0
308	29	29	32	3	20
353	0	0	0	0	0
370	2	2	4	2	0
551	0	0	2	2	0
554	1	1	1	0	1
600	5	6	7	2	1

8.3. lentelė. Hipotezės skaičiavimai

Hipotezė					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
47	4	4	4	0	0
50	2	2	2	0	0
57	3	3	4	1	0
63	10	10	11	1	0
85	5	5	5	0	0
95	2	2	2	0	0
96	4	4	4	0	0
97	2	2	2	0	0
98	3	3	3	0	0
99	3	3	3	0	0
100	2	2	2	0	0
101	4	4	4	0	0
102	2	2	2	0	0
103	4	4	4	0	0
104	4	4	4	0	0
105	4	4	4	0	0
106	2	2	2	0	0
107	2	2	2	0	0
125	2	2	4	2	0
130	9	9	9	0	0
139	7	7	7	0	0
208	4	4	5	1	0
226	2	2	2	0	0
235	5	5	5	0	0
299	5	5	6	1	0
308	23	23	23	0	0
353	12	12	12	0	0
370	2	2	2	0	0
551	2	2	2	0	0
554	0	1	1	1	0
600	3	3	3	0	2

8.4. lentelė. Objekto skaičiavimai

Objektas					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
47	3	3	10	7	0
50	2	2	5	3	0
57	3	3	16	13	0
63	10	10	37	27	0
85	5	5	13	8	0
95	2	2	4	2	0
96	5	5	23	18	0

97	6	6	11	5	0
98	3	3	8	5	0
99	6	6	9	3	0
100	6	6	9	3	0
101	4	4	9	5	0
102	2	2	5	3	0
103	4	4	10	6	0
104	4	4	13	9	0
105	4	4	9	5	0
106	2	2	5	3	0
107	2	2	6	4	0
125	2	2	14	12	0
130	9	9	21	12	0
139	9	9	30	21	0
208	3	3	81	78	0
226	0	1	13	13	1
235	5	5	20	15	0
299	4	4	19	15	0
308	24	24	120	96	0
353	12	12	26	14	0
370	2	2	5	3	0
551	2	2	23	21	0
554	0	1	4	4	1
600	0	0	21	21	0

8.5. lentelė. Dispozicijos skaičiavimai

Dispozicija					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
47	4	4	4	0	0
50	2	2	2	0	0
57	3	3	4	1	0
63	10	10	11	1	0
85	5	5	5	0	0
95	2	2	2	0	0
96	4	4	4	0	0
97	2	2	2	0	0
98	3	3	3	0	0
99	3	3	3	0	0
100	2	2	2	0	0
101	4	4	4	0	0
102	2	2	2	0	0
103	4	4	4	0	0
104	4	4	4	0	0
105	4	4	4	0	0
106	2	2	2	0	0
107	2	2	2	0	0

125	2	2	4	2	0
130	9	9	9	0	0
139	7	7	7	0	0
208	4	4	5	1	0
226	1	1	2	1	0
235	5	5	5	0	0
299	5	5	6	1	0
308	23	23	23	0	0
353	12	12	12	0	0
370	2	2	2	0	0
551	2	2	2	0	0
554	1	2	2	1	1
600	4	5	9	5	1

8.3.2. Darbo kodeksas

8.6. lentelė. Subjekto skaičiavimai

Subjektas					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
6	14	15	18	4	1
13	9	9	11	2	0
23	4	7	10	6	3
24	4	9	9	5	5
31	15	18	19	4	3
35	2	3	5	3	1
40	15	26	26	11	11
42	16	18	20	4	3
59	2	2	3	1	2
86	6	8	8	2	3
90	4	9	9	5	5
99	18	20	20	2	2
119	4	5	5	1	1
125	3	3	3	0	0
143	3	9	9	6	6
156	6	9	9	3	3
168	33	42	45	12	10
171	23	37	68	45	14
240	4	4	23	19	0

8.7. lentelė. Objekto skaičiavimai

Objektas					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
6	2	2	27	25	0
13	0	0	10	10	0
23	2	2	5	3	0

24	0	0	13	13	0
31	0	0	14	14	0
35	1	1	8	7	0
40	2	2	34	32	0
42	3	3	14	11	0
59	1	1	18	17	0
86	0	0	8	8	0
90	1	1	12	11	0
99	0	0	8	8	0
119	0	0	16	16	0
125	3	3	10	7	0
143	4	4	11	7	0
156	3	3	8	5	0
168	2	2	19	17	0
171	2	2	37	35	0
240	0	0	13	13	0

8.8. lentelė. Teisių skaičiavimai

Teisė					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
6	0	0	2	2	1
13	0	0	3	3	0
23	0	0	0	0	0
24	0	0	1	1	1
31	1	1	2	1	0
35	0	1	1	1	1
40	2	2	2	0	0
42	0	0	0	0	0
59	1	1	1	0	0
86	0	0	2	2	0
90	0	0	0	0	0
99	2	2	2	0	0
119	0	0	0	0	0
125	0	0	2	2	0
143	0	0	1	1	0
156	0	0	3	3	0
168	1	1	2	1	0
171	0	0	2	2	0
240	0	0	2	2	0

8.9. lentelė. Pareigų skaičiavimai

Pareiga					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
6	0	0	0	0	0
13	0	0	0	0	0
23	2	2	2	0	0
24	3	3	3	0	0
31	2	2	2	0	0
35	0	0	0	0	0
40	0	0	1	1	0
42	2	2	2	0	0
59	0	0	0	0	0
86	0	0	0	0	0
90	1	1	2	1	1
99	0	0	0	0	0
119	0	0	0	0	0
125	0	0	2	2	0
143	0	0	1	1	0
156	0	0	1	1	0
168	0	0	1	1	0
171	2	2	4	2	0
240	1	1	2	1	0

8.10. lentelė. Hipotezių skaičiavimai

Hipotezė					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
6	1	1	13	12	1
13	0	0	4	4	0
23	2	2	2	0	0
24	5	5	6	1	3
31	4	4	2	-2	2
35	6	6	3	-3	3
40	6	6	9	3	0
42	2	2	7	5	1
59	2	2	0	-2	2
86	2	2	2	0	1
90	2	2	2	0	2
99	2	2	3	1	2
119	3	3	6	3	5
125	0	0	1	1	0
143	3	3	4	1	2
156	2	2	2	0	0
168	5	5	8	3	3
171	12	12	24	12	2
240	3	3	6	3	1

8.11. lentelė. Dispozicijų skaičiavimai

Dispozicija					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
6	1	1	12	11	0
13	0	0	6	6	0
23	2	2	3	1	0
24	5	5	6	1	0
31	8	8	10	2	0
35	7	7	8	1	0
40	9	9	12	3	0
42	2	2	3	1	0
59	2	2	2	0	1
86	2	2	4	2	0
90	3	3	5	2	0
99	2	2	6	4	0
119	7	7	4	-3	3
125	1	1	3	2	3
143	3	3	4	1	0
156	4	4	5	1	0
168	10	10	17	7	0
171	35	35	42	7	9
240	3	3	11	8	1

8.3.3. Civilinis kodeksas

8.12. lentelė. Hipotezės skaičiavimai

Hipotezė					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
1.5	2	2	4	2	0
4.144	4	4	5	1	3
4.225	3	3	2	-1	2
4.238	2	2	4	2	1
6.729	2	2	6	4	1
6.1006	3	3	2	-1	3

8.13. lentelė. Dispozicijos skaičiavimai

Dispozicija					
Straipsnis	Atpažino pagal sukurta taisykles	Iš viso turėjo atpažinti pagal sukurta taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
1.5	2	5	5	3	0
4.144	5	6	8	3	0
4.225	3	3	3	0	0
4.238	2	2	5	3	0
6.729	5	5	7	2	0
6.1006	3	3	3	0	3

8.14. lentelė. Pareigos skaičiavimai

Pareiga					
Straipsnis	Atpažino pagal sukurta taisykles	Iš viso turėjo atpažinti pagal sukurta taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
1.5	1	1	3	2	0
4.144	3	3	6	3	0
4.225	0	0	0	0	0
4.238	0	0	0	0	0
6.729	0	0	1	1	0
6.1006	0	0	0	0	0

8.15. lentelė. Objekto skaičiavimai

Objektas					
Straipsnis	Atpažino pagal sukurta taisykles	Iš viso turėjo atpažinti pagal sukurta taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
1.5	0	0	10	10	0
4.144	0	0	13	13	0
4.225	0	0	6	6	0
4.238	0	0	8	8	0
6.729	0	0	15	15	0
6.1006	0	0	6	6	0

8.16. lentelė. Subjektų skaičiavimai

Subjektas					
Straipsnis	Atpažino pagal sukurta taisykles	Iš viso turėjo atpažinti pagal sukurta taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
1.5	0	0	3	3	0
4.144	0	1	13	13	0
4.225	0	0	3	3	0
4.238	0	1	5	5	0
6.729	2	3	14	12	1
6.1006	3	4	11	8	1

8.17. lentelė. Teisių skaičiavimai

Teisė					
Straipsnis	Atpažino pagal sukurtas taisykles	Iš viso turėjo atpažinti pagal sukurtas taisykles	Bendras anotacijos skaičius straipsnyje	Neatpažino	Klaidingai atpažinti
1.5	0	0	0	0	0
4.144	0	0	0	0	0
4.225	2	2	3	1	0
4.238	1	1	3	2	0
6.729	0	0	0	0	0
6.1006	1	1	1	0	0