

Laboratorij za analizu teksta i inženjerstvo znanja

Text Analysis and Knowledge Engineering Lab

Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva
Unska 3, 10000 Zagreb, Hrvatska



Zaštićeno licencijom

Creative Commons Imenovanje-Nekomercijalno-Bez prerada 3.0 Hrvatska

<https://creativecommons.org/licenses/by-nc-nd/3.0/hr/>

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1151

**Plitko semantičko parsanje
tekstova na hrvatskome jeziku**

Krešimir Baksa

Zagreb, srpanj 2015.

Zagreb, 6. ožujka 2015.

Predmet: **Analiza i pretraživanje teksta**

DIPLOMSKI ZADATAK br. 1151

Pristupnik: **Krešimir Baksa (0036460125)**

Studij: Računarstvo

Profil: Računarska znanost

Zadatak: **Plitko semantičko parsanje tekstova na hrvatskome jeziku**

Opis zadatka:

Plitko semantičko parsanje (engl. shallow semantic parsing) ili označavanje semantičkih uloga jest postupak prepoznavanja i klasifikacije dijelova rečenica prema njezinoj predikatnoj strukturi. Semantičko parsanje središnji je zadatak u semantičkoj obradi prirodnog jezika i njegovo je rješavanje nezabilazno za razvoj sustava za razumijevanje jezika. Međutim, zbog višeznačnosti prirodnog jezika, ono je ujedno i vrlo težak zadatak. Razvijeni modeli uglavnom se oslanjaju na postupke nadziranog strojnog učenja koji iziskuju nemale količine ručno označenih podataka.

U okviru diplomskoga rada potrebno je proučiti formalizam semantičkih uloga PropBank prema Palmer i dr. (2005) te odgovarajuće pristupe za semantičko parsanje, s naglaskom na pristupe temeljene na nadziranom strojnom učenju i slijednom označavanju. Razraditi nadzirani model za semantičko parsanje u tekstovima na hrvatskome jeziku temeljen na formalizmu PropBank, uz može bitna pojednostavljenja. Razraditi postupak za poluatوماتsko semantičko parsanje temeljeno na sintaktički označenom hrvatskom web-korpusu uz primjenu odgovarajućih heuristika, te dodatno ručno označiti ispitni skup podataka. Razviti programsku implementaciju postupaka te provesti iscrpno vrednovanje na ispitnome skupu, uključivo analizu značajki i usporedbu s referentnim modelima. Radu priložiti izvorni i izvršni kod razvijenog sustava, označene skupove podataka i potrebnu dokumentaciju te citirati korištenu literaturu.

Zadatak uručen pristupniku: 13. ožujka 2015.

Rok za predaju rada: 30. lipnja 2015.

Mentor:

Doc. dr. sc. Jan Šnajder

Predsjednik odbora za
diplomski rad profila:

Djelovođa:

Doc. dr. sc. Tomislav Hrkać

Prof. dr. sc. Siniša Srblić

SADRŽAJ

1. Uvod	1
2. Plitko semantičko parsanje	2
2.1. Definicija problema	2
2.2. pristupi plitkog semantičkog parsanja	3
2.2.1. pristup temeljen na projektu FrameNet	3
2.2.2. pristup temeljen na projektu ProbBank	6
3. Priprema podataka	9
3.1. Korpus	9
3.2. Izrada semantičkih okvira	11
3.3. Označavanje podataka	12
4. Modeli za plitko semantičko parsanje na hrvatskom jeziku	15
4.1. Značajke	15
4.2. Priprema podataka	16
4.3. Osnovni model	17
4.4. Modeli temeljeni na nadziranom strojnom učenju	17
4.4.1. Logistička regresija	17
4.4.2. Stroj potpornih vektora	18
5. Vrednovanje	21
5.1. Mjere	21
5.2. Odabir značajki	21
5.3. Rezultati	23
6. Zaključak	29
Literatura	31

1. Uvod

U području obrade prirodnog jezika postoje pristupi koji vrlo efikasno mogu izvršiti sintaktičko parsanje prirodnog jezika¹. Sintaktičkim parsanjem možemo izvući veliku količinu informacija korisnih u obradi prirodnog jezika, ali za neke primjene te informacije nisu dovoljne jer ne možemo u potpunosti odrediti semantiku neke rečenice. Za razliku od sintaktičkog parsanja, plitko semantičko parsanje (engl. *shallow semantic parsing*) ili označavanje semantičkih uloga je postupak prepoznavanja i klasifikacije dijelova rečenice prema njezinoj predikatnoj strukturi, drugi riječima plitkim semantičkim parsanjem odgovaramo na pitanja: tko je napravio što nekome, gdje, kada, zašto i kako. Semantičko parsanje je središnji zadatak u semantičkoj obradi prirodnog jezika i njegovo rješavanje je nezaobilazno za razvoj sustava za razumijevanje jezika.

Postoji nekoliko zadataka u obradi prirodnog jezika gdje je semantička obrada ključna, a neki od tih zadataka su: izvlačenje informacija, odgovaranje na pitanja, sažimanje ili strojno prevođenje prirodnog jezika. Zbog višeznačnosti prirodnog jezika zadatak semantičkog parsanja je ujedno i vrlo težak zadatak, a razvijeni modeli se uglavnom oslanjaju na postupke nadziranog strojnog učenja koji iziskuju nemale količine ručno označenih podataka.

U ovom je radu obavljeno ručno označavanje manjeg dijela teksta na hrvatskom jeziku iz novinske domene iz korpusa *hrWac*² te je napravljeno vrednovanje modela koji se temelje na nadziranom strojnom učenju.

U poglavlju 2. opisani su pristupi i metode plitkog semantičkog parsanja, poglavlje 3. opisuje postupke pripreme podataka, poglavlje 4. opisuje modele korištene. Dobiiveni rezultati su opisani u poglavlju 5., a u poglavlju 6. je izveden zaključak.

¹Agić et al. (2013) za hrvatski jezik, Nivre i Scholz (2004) za engleski jezik.

²<http://nlp.ffzg.hr/resources/corpora/hrwac/>

2. Plitko semantičko parsanje

2.1. Definicija problema

Plitko semantičko parsanje zadatak je kojim se u osnovi pokušava odrediti uloga članova rečenice te odnos između članova. Ovaj problem se rješava dodjeljivanjem semantičkih uloga članovima rečenice, tj. označavanjem nekih dijelova rečenice određenim oznakama koje definiraju semantičku ulogu.

[U finalu_{LOC}] su [ih_{PAT}] čekali [košarkaši Zagreba_{AGT}].

U navedenom primjeru možemo uočiti da je u rečenici definiran član koji vrši neku radnju (u ovom slučaju ta radnja je čekanje) te je taj član označen oznakom *AGT*.¹ Osim člana koji vrši radnju također je u rečenici navedeno i nad kim se vrši radnja te je taj član označen oznakom *PAT*.² Također, oznakom *LOC*³ su označeni članovi rečenice kojima je definirano mjesto radnje. Iz ovog jednostavnog primjera vidimo da je semantičkim oznakama odgovoreno na nekoliko pitanja (tko čeka koga i gdje). Iako bi se iz ovog primjera moglo zaključiti da dodjeljivanje semantičkih uloga nema razlike u odnosu na sintaktičko parsanje rečenice jer vršitelj neke radnje je, prema definiciji, subjekt, onaj ili ono nad čime se vrši radnja je objekt, priložna oznaka mjesta je lokacija, itd.

[Poslodavac_{AGT}] je povećao [plaću_{PAT}] zaposlenicima.

Zaposlenicima se povećala [plaća_{PAT}].

Prikazani primjer objašnjava razliku između sintaktičkih i semantičkih uloga članova u rečenici. U prvoj rečenici *plaća* je sa sintaktičke strane objekt dok je u drugoj

¹AGT (engl. *Agent*) označava semantičku ulogu vršitelja radnje.

²PAT (engl. *Patient*) označava semantičku ulogu nad kojom se radnja vrši.

³LOC (engl. *Location*) označava semantičku ulogu kojom je definirano mjesto radnje.

rečenici ona subjekt. Bez obzira na sintaktičke uloge u obje rečenice *plaća* je semantički gledano ono nad čime se radnja vrši (u ovom slučaju radnja je *povećati*). Valja naglasiti i to da je u obje rečenice glagolsko stanje predikata radno ili aktivno.

Levin (1993) u svojem istraživanju navodi kako su semantička i sintaktička struktura rečenice povezane u smislu da skupina sintaktičkih struktura (koje se nazivaju sintaktički okviri) zapravo definira određenu semantičku strukturu (semantički okvir) te tako određuje i broj i vrstu semantičkih uloga koje se koriste u tom semantičkom okviru. Prema toj teoriji Levin klasificira glagole s obzirom na sposobnost glagola da se pojavi u jednom ili više parova sintaktičkih okvira koji imaju neko slično značenje. Ovaj rad je osnova daljnjem istraživanju semantičkih svojstava glagola i razvoju boljih razreda koje mogu definirati semantiku glagola i njegove semantičke uloge za ostale riječi u rečenici. Na temelju tog istraživanja postoje razni pristupi za semantičko parsanje i plitko semantičko parsanje, a dva najpoznatija pristupa opisana su u sljedećem poglavlju.

Priprema podataka i samo semantičko parsanje prirodnog jezika sastoji se od nekoliko koraka. Prvi korak je određivanje razreda kojem glagoli mogu pripadati. U tom koraku može biti raznih pristupa, a dva su najzastupljenija za engleski jezik pa i za ostale jezike. Prvi pristup temeljen je na radu Fillmore (1976) (kasnije proširena u Baker et al. (1998) i Johnson (2002)) dok je drugi pristup predstavljen u radu Palmer et al. (2005). Drugi korak pripreme podataka za semantičko parsanje jest označavanje podataka. U ovom radu priprema i označavanje podataka napravljeno je prema uzoru na Palmer et al. (2005) uz dodatna pojednostavljenja.

2.2. Pristupi plitkog semantičkog parsanja

2.2.1. Pristup temeljen na projektu FrameNet

FrameNet i PropBank su projekti čiji je cilj definirati metodologiju za označavanje semantičkih uloga te izraditi relativno velik korpus koji je označen određenom metodologijom. FrameNet je jedan od raširenijih i dugotrajnijih projekata u kojem je zadatak bio stvaranje semantičkih okvira koji će se kasnije moći koristiti za označavanje korpusa sa semantičkim ulogama. Razni glagoli (i druge vrste riječi) mogu u nekim slučajevima imati semantički slično ili isto značenje ili pripadati jednoj ili više semantičkih grupa, tj. semantičkim okvirima. Metodologija FrameNet-a je da se prvo definira semantički okvir na višoj razini apstrakcije kojim se označava grupa glagola koji mogu pripadati tom okviru i semantičke uloge koji se nazivaju elementi okvira. Primjerice

za semantički okvir *TRGOVINA* (engl. *COMMERCE*) postoje elementi ili semantičke uloge: *KUPAC* (engl. *BUYER*), *DOBRO* (engl. *GOOD*), *PRODAVAČ* (engl. *SELLER*), *NOVAC* (engl. *MONEY*). Glagoli kao što su kupiti (engl. *buy*), prodati (engl. *sell*), platiti (engl. *sell*), itd. pripadaju ovom semantičkom okviru. Za svaki takav okvir u bazi podataka projekta FrameNet postoji objašnjenje tog semantičkog okvira. Općenito za prirodni jezik vrijedi da neka riječ ima više značenja pa se zbog toga dodatno definira *leksička jedinica* (engl. *lexical unit*) koja pripada određenom semantičkom okviru. Iz ovog slijedi da svaki semantički okvir sadrži nekoliko leksičkih jedinica kojima se definira točno jedno značenje tog okvira. Također je moguć slučaj da jedan semantički okvir sadrži podokvir koji opet u sebi sadrži leksičku jedinicu te je tako pokriveno svako značenje okvira.

Elementi (tj. semantičke uloge) koji pripadaju određenom semantičkom okviru nazivaju se u projektu FrameNet *jezgreni elementi* (engl. *core frame elements*). Uz jezgrene elemente definirane su semantičke uloge koje nisu vezane uz specifičan semantički okvir već mogu pripadati većini semantičkih okvira. Takve semantičke uloge se prema FramNet-u nazivaju *nejezgreni elementi* (engl. *non-core frame elements*). Nejezgreni elementi (semantičke uloge) označavaju mjesto, vrijeme, uzrok i druge slične semantičke uloge. U bazi podataka FrameNet-a za svaki semantički okvir postoji opis svakog jezgrenog elementa i svakog mogućeg nejezgrenog elementa za taj semantički okvir.

U tablici 2.1 prikazan je sažeti primjer jednog semantičkog okvira iz baze podataka FrameNet. Iz sadržaja okvira može se izdvojiti nekoliko dijelova okvira. Na zaglavlju okvira se nalazi naziv okvira koji je ujedno i smislen, vezan uz semantičko značenje samog okvira. Nakon naziva okvira slijedi polje detaljnog semantičkog opisa okvira. Sljedeće polje okvira sadrži elemente okvira, tj. semantičke uloge. Radi jednostavnosti, u ovom primjeru su navedene samo neke semantičke uloge iako ih u stvarnoj bazi ima više te su podijeljene u dvije skupine: jezgrene i nejezgrene (engl. *core and non-core elements frame elements*). Uz semantičke uloge, u semantički okvir pridodan je i jedan primjer označene rečenice te popis leksičkih jedinica (engl. *lexical units*).

U projektu FrameNet se koristi korpus *the British National Corpus*, koji sadrži više od 100 milijuna riječi na engleskom jeziku. Za sada se u bazi podataka FrameNet nalazi oko 1,200 semantičkih okvira, 13,000 leksičkih jedinica i preko 190,000 označenih primjeraka rečenica. Detaljniji opis projekta FrameNet nalazi se u (Baker et al., 1998) i kasnije u (Johnson, 2002).

Primjena podataka iz baze podataka projekta Framenet za plitko semantičko parsiranje je za engleski jezik opisano u (Gildea i Jurafsky, 2002), gdje je postignuta točnost

Frame: COMMERCE_BUY	
Description:	These are words describing a basic commercial transaction involving a BUYER and a SELLER exchanging MONEY and GOODS, taking the perspective of the BUYER. The words vary individually in the patterns of frame element realization they allow. For example, the typical pattern for the verb BUY: BUYER buys GOODS from SELLER for MONEY.
Frame Elements:	
BUYER:	The BUYER wants the GOODS and offers MONEY to a SELLER in exchange for them.
SELLER:	The SELLER has possession of the GOODS and exchanges them for MONEY from a BUYER.
GOODS:	The FE GOODS is anything (including labor or time, for example) which is exchanged for MONEY in a transaction.
Example:	<i>Abby bought a car from Robin for \$5,000.</i> BUYER: <i>Abby</i> GOODS: <i>car</i> SELLER: <i>from Robin</i> MONEY: <i>for \$5,000</i>
Lexical Units:	buy.v, buyer.n, purchase [act].n, purchase.v, purchaser.n

Tablica 2.1: Primjer semantičkog okvira iz baze podataka FrameNet

od 82% u određivanju semantičkih uloga. U radu je korišteno oko 50,000 ručno označenih rečenica koje su korištene za treniranje automatskog parsera temeljenog na nadziranom strojnom učenju.

Osim za engleski jezik postoje projekti i za druge jezike koju su temeljeni na pristupu projekta Framenet. Za španjolski jezik također je razvijena FrameNet inačica baze podataka. Detaljniji opis španjolske inačice projekta Framenet se nalazi u radovima Subirats i Petruck (2003) i Subirats i Sato (2004). Od ostalih zastupljenijih jezika projekti Framenet su u procesu razvoja za Japanski (Ohara et al., 2003), Njemački (Boas, 2002) i Poljski (Zawislawska et al., 2008).

projekt Framenet za engleski jezik (pa i za druge jezike) relativno je velik projekt relativno je velik projekt i razvija se već dugi niz godina.⁴ Prema dosadašnjem saznanju trenutno nema projekata sličnih FrameNet-u za hrvatski jezik. S obzirom na opseg zadatka definiranog u FrameNet-u i količine znanja potrebnog za ostvarivanje ovog zadatka nije moguće da jedna osoba u relativno kratkom vremenu može ostvariti neke značajne rezultate, pa zbog toga ovaj pristup nije primijenjen u ovom radu.

2.2.2. Pristup temeljen na projektu ProbBank

PropBank je projekt predstavljen u Palmer et al. (2005) te ima isti cilj kao i FrameNet, definirati metodologiju za označavanje semantičkih uloga u prirodnom jeziku te razviti ručno označeni korpus na temelju te metodologije koji kasnije može poslužiti za razvoj modela za automatsko označavanje. Za razliku od projekta Framenet, PropBank se manje bavi semantikom glagola te se za dodatno objašnjenje semantike glagola koriste reference na VerbNet⁵ uz to, PropBank se primarno fokusira na glagole⁶ dok FrameNet pokriva i semantičke definicije za imenice i pridjeve. Osnovni zadatak PropBank-a je bio definirati dovoljan skup semantičkih okvira da se može označiti čitav korpus Penn Treebank.⁷

Postupak izrade semantičkih okvira u PropBank-u je jednostavniji od onog koji je definiran u FrameNet-u. Semantički okviri su izrađeni na temelju semantike pojedinog glagola te ne postoje pod-okviri. Za svako značenje glagola okvir sadrži kratak opis

⁴Projekt je započeo 1997. godine i još uvijek se obnavlja i nadograđuje FrameNet baza podataka kao i označeni korpus.

⁵VerbNet je jezični leksikon za glagole u kojem za svaki glagol postoji sintaktička i semantička informacija u obliku razreda definiranih u Levin (1993).

⁶U početnoj fazi projekta razvijao se samo skup semantičkih okvira za glagole, kasnije je razvijan skup za imenice i pridjeve.

⁷Penn Treebank (Marcus et al., 1993) je korpus na engleskom jeziku koji sadrži oko 4.5 milijuna riječi te je označen POS oznakama i sintaktičkim oznakama.

značenja tog glagola, a semantičke uloge manje ovise o značenju glagola i o semantičkom okviru kojem pripadaju. Semantičke uloge su generički određene pa se tako oznakom ARG0 određuje član koji vrši radnju na razini cijelog skupa semantičkih okvira (engl. *Agent*). Oznakom ARG1 se, generalno kroz cijeli korpus, određuje član nad kojim se radnja vrši (engl. *Patient*). Ostale semantičke uloge se dodaju ovisno o semantičkom okviru kojem pripadaju te kroz skup semantičkih okvira nema neke konzistencije. Time je povećana razina apstrakcije nad cijelim skupom semantičkih okvira, a ujedno smanjen broj semantičkih uloga. Prema Palmer et al. (2005) jedan od semantičkih okvira ima najviše šest semantičkih uloga (od ARG0 do ARG5), a najviše okvira sadrži od dvije do četiri semantičke uloge. S obzirom da je ovim pristupom povećana razina apstrakcije kod izgradnje semantičkih okvira prilikom označavanja korpusa, nije moguće u potpunosti izraditi preslikavanje na oblik oznaka definiran u FrameNet-u.

Predicate: <i>buy</i>		
Roleset id: buy.01	Description: <i>purchase</i>	vncls: 13.5.1
Roles: Arg0: <i>buyer</i> Arg1: <i>thing bought</i> Arg2: <i>seller</i> Arg3: <i>price paid</i>		
Example: <i>Consumers who buy at this level are more educated than they were.</i> Arg0: <i>Consumers</i> Rel: <i>buy</i> ArgM-MNR: <i>at this level</i>		

Tablica 2.2: Primjer semantičkog okvira iz PropBank baze podataka

Tablicom 2.2 prikazan je format semantičkog okvira za jedan glagol na engleskom jeziku. Okvir se može podijeliti u tri dijela: zaglavlje, semantičke uloge i označeni primjeri.

Zaglavlje okvira se sastoji od sljedećih polja:

- *Predicate*, koji označava za koji glagol je ovaj okvir namijenjen.
- *Roleset id*, čime se indentificira pojedinačni okvir u skupu okvira koji pripadaju određenom glagolu.

- *Description*, opis semantičkog okvira koji je često samo sinonim glagola koji taj okvir opisuje.
- *vncls*, skraćeno od (engl. *VerbNet class*) je poveznica na VerbNet razred.

Semantičke uloge u okviru su redom sažeto opisane u središnjem dijelu okvira nakon čega slijedi jedan ili više primjera označenih rečenica. U primjeru je radi jednostavnosti naveden samo jedan primjer iako je u stvarnoj bazi podataka projekta PropBank to rijedak slučaj. Osim semantičkih uloga, u označenim primjerima je također označen i predikat.

Prilikom označavanja korpusa u projektu PropBank objavljeno je nekoliko radova u kojima je napravljeno vrednovanje raznih modela na do onda označenom korpusu. Gildea i Palmer (2002) ukazuju na važnost sintaktičkih oznaka za razliku od korpusa koji ima dostupne samo plitke sintaktičke oznake. Gildea i Hockenmaier (2003) koriste značajke iz drugog korpusa (engl. *Combinatory Categorical Grammar, CCG*)⁸ te pokušavaju preslikati semantičke oznake na one koje su korištene u PropBank-u kako bi dobili bolje rezultate od modela temeljenih na TreeBank korpusu.

PropBank pristup označavanju semantičkih uloga je relativno jednostavniji od FramNet pristupa, a najveća prednost je što se fokusira na izgradnju semantičkih okvira za predikate i izgradnja dosta ovisi o primjerima iz korpusa. U projektu PropBank označeno je oko 1 milijun riječi iz financijske domene (*the Wall Street Journal*) od kojih su 3,633 jedinstveni glagoli. Prvi rezultati vrednovanja klasifikator na manjem skupu podataka imaju preciznost od 82.0% i odziv od 74.7% uz dodatne korištene oznake te ručno označenim sintaktičkim oznakama. Za sustav automatskog sintaktičkog i plitkog semantičkog parsera postignuta je ocjena preciznosti od 69.9% te odziva od 61.1%. Iako je označavanje podataka za ovaj projekt trajalo nekoliko godina⁹ te su u projekt uključeni mnogi stručnjaci na području lingvistike i obrade prirodnog jezika, moguće je u relativno kratkom vremenu izgraditi manji skup označenih rečenica u prirodnom jeziku te uz to formirati semantičke uloge za taj skup. Zbog jednostavnijeg pristupa ova metodologija je, uz neka pojednostavljenja, korištena u ovom radu za izradu semantičkih okvira na hrvatskom jeziku te označavanje primjera rečenica u tekstu iz novinske domene. Detaljniji pristup i pojednostavljenja koja su uvedena u postupak opisana su u sljedećem poglavlju.

⁸Steedman (2000)

⁹Projekt PropBank još uvijek traje te se rade dodatna istraživanja kako bi se omogućilo preslikavanje semantičkih uloga i okvira između FramNet projekta i PropBank projekta.

3. Priprema podataka

3.1. Korpus

U ovom radu korišten je korpus hrWaC (Ljubešić i Klubička, 2014). Korpus hrWaC je dobiven prikupljanjem tekstova dostupnih na Internetu na hr domeni te ti tekstovi prema sadržaju pripadaju novinskoj domeni. U prvoj verziji (v1.0) korpusa hrWaC prikupljeno je oko 1.2 milijardi riječi (Ljubešić i Erjavec, 2011), a kasnije u verziji 2.0 (Ljubešić i Klubička, 2014) korpus je povećan na 1.9 milijardi riječi na hrvatskom jeziku. Uz povećanje, korpus je lematiziran CST lematizatorom¹ (Jongejan i Dalianis, 2009) te su dodane sintaktičke oznake. Za plitke sintaktičke oznake korišten je Hun-Pos² (Halácsy et al., 2007), a za sintaktičke oznake alat mate-tools³ (Bohnet, 2010).

Korpus hrWaC korišten u ovom radu sadrži oko 1.2 milijarde riječi te oko 50 milijuna rečenica koje su označene prethodno navedenim oznakama. Podaci su spremljeni u formatu koji je zadan u zadatku *CoNLL-X Shared Task: Multi-lingual Dependency Parsing*.⁴ Za potrebe ovog rada prvo je bilo važno izdvojiti podskup rečenica koje će se u sljedećoj fazi označiti semantičkim ulogama. U konačnici je prema predviđanju kompleksnosti zadatka bilo potrebno izgraditi stotinjak semantičkih okvira pa je zbog višeznačnosti izdvojeno 76 glagola za koje su izrađeni semantički okviri. Broj rečenica u korpusu u kojima se nalaze rečenice kojima je korijenski glagol jedan od izdvojenih 76 iznosi 21 milijun što pokriva oko 43% korpusa što je očekivano (zbog Zipfovog zakona) jer su za izradu semantičkih okvira uzeti najučestaliji glagoli u korpusu.

Primjer 3.1 prikazuje format označenih podataka sa sintaktičkim oznakama. U primjeru 3.1 skicirano je sintaktično stablo dobiveno pomoću podataka dostupnih iz 3.1.

¹<https://github.com/kuhumcst/cstlemma>

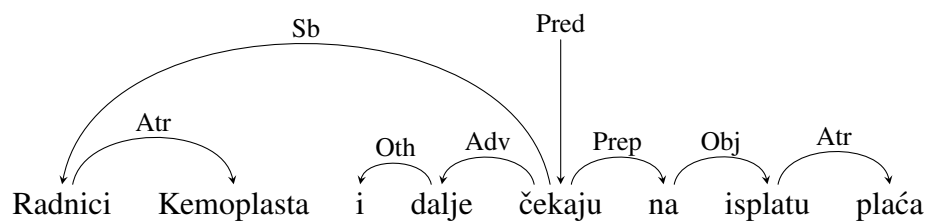
²<https://code.google.com/p/hunpos/>

³<https://code.google.com/p/mate-tools/>

⁴<http://ilk.uvt.nl/conll/>

ID	FORM	LEMMA	CPOSTAG	POSTAG	HEAD	DEPREL
1	Radnici	radnik	N	N-mpn	5	Sb
2	Kemoplasta	kemoplast	N	N-msg	1	Atr
3	i	i	C	Cc	4	Oth
4	dalje	daleko	R	Rgc	5	Adv
5	čekaju	čekati	V	Vmr3p	0	Pred
6	na	na	S	Sa	5	Prep
7	isplatu	isplata	N	N-fsa	6	Obj
8	plaća	plaća	N	N-fpg	7	Atr
9	.	.	Z	Z	0	Punc

Tablica 3.1: Primjer označene rečenice sintaktičkim oznakama prema zadatku *CoNLL-X Shared Task: Multi-lingual Dependency Parsing*



Slika 3.1: Sintaktičko stablo označene rečenice.

3.2. Izrada semantičkih okvira

Postupak stvaranja semantičkih okvira odvijao se u dvije faze. Najprije je pomoću hrvatskog rječnika⁵ stvoren inicijalni skup semantičkih okvira koji je sadržavao opis semantičkog okvira, identifikacijsku oznaku te popis semantičkih uloga. S obzirom da je cijeli rad, pa tako i ovaj dio rada, radila jedna osoba koja nema izrazito široko znanje iz područja lingvistike ovaj korak u izradi semantičkih okvira je potrajao oko dva tjedna.

Sljedeća faza izrade semantičkih okvira sastojala se od toga da se pomoću dostupnih rečenica iz korpusa poprave nedostaci u semantičkim okvirima. Već prije je spomenuto da su semantički okviri izrađeni prema uzoru na projekt PropBank pa za razliku od PropBank okvira ovi okviri nisu sadržavali primjer označenog teksta jer za to nije bilo potrebe s obzirom da je označavanje korpusa također radila ista osoba koja je izradila semantičke okvire. S obzirom da se tokom ove faze vršilo ispravljanje semantičkih okvira i označavanje podataka ovaj korak je trajao oko tri tjedna.

Predikat: <i>završiti</i>
Id: <i>završiti.01</i>
Opis: <i>priveći kraju, dovesti do kraja; okončati; biti gotov;</i>
Argumenti: Arg0: <i>onaj tko je završio</i> Arg1: <i>što se dovodi do kraja</i>

Tablica 3.2: Primjer semantičkog okvira iz izrađenog skupa semantičkih okvira

Primjer 3.2 prikazuje format semantičkog okvira koji se koristio u ovom radu. U slučaju opisa semantičkih uloga, za razliku od većine semantičkih okvira u bazi podataka PropBank, korišten je nešto opširniji opis te je u većini slučajeva opis izrađen u prvom licu jednine muškog roda što ne znači da se ta semantička uloga ne odnosi na druga lica ili rodove.

Prema izdvojenim glagolima u konačnici je izrađeno 128 semantičkih okvira te većina glagola ima samo jedan ili dva semantička okvira, a najveći broj semantičkih okvira za pojedini glagol je četiri i to za glagol *odbiti*. Semantički okviri su mijenjani i ispravljani s obzirom na primjere rečenica u korpusu pa su neki semantički okviri uklonjeni iz skupa jer nije bilo primjera značenja tog glagola u rečenicama koje su bile označene semantičkim ulogama.

⁵<http://hjp.novi-liber.hr/index.php>

3.3. Označavanje podataka

Nakon određivanja semantičkih okvira sljedeća faza projekta je bila označavanje semantičkih uloga u rečenicama. Iz korpusa je izdvojeno 1,214 rečenica koje su označene sa semantičkim ulogama. Pri označavanju je sudjelovao samo jedan označivač. Prilikom označavanja cilj je bio označiti od 10 do 15 rečenica za svaki izdvojeni glagol za koje su izrađeni semantički okviri. Prilikom označavanja neki semantički okviri su bili loše formirani pa je tokom označavanja promijenjena struktura okvira. S obzirom da je korpus sastavljen pretežito od novinskih tekstova velik broj rečenica je imao sličnu formu što je za posljedicu imalo da neki semantički okviri nisu imali primjer rečenice gdje se koristi semantika koju opisuju ti okviri. Bez obzira na to što ti semantički okviri nisu imali primjer rečenice u izdvojenom skupu rečenica, oni nisu izostavljeni iz baze semantičkih okvira.

Uz semantičke uloge koje pripadaju semantičkim okvirima (ARG0-ARG4) u korpusu su članovi rečenice označeni i općenitim semantičkim ulogama (ARGM). Za razliku od PropBank projekta koji ima definiranih 12 oznaka kojima se određuju tzv. "modifikatori" (engl. *modifiers*) u ovom slučaju korpus je označeni sa samo 6 takvih definiranih modifikatora:

- Argm-neg određuje članove u rečenici koji su negacija nečega (npr. "ne", "nije", itd.);
- Argm-dir označava članove koji u semantičkom smislu određuju smjer;
- Argm-cau određuju one članove u rečenici koje imaju značenje uzroka neke radnje;
- Argm-loc određuju članove rečenice kojima je iskazano mjesto ili lokacija;
- Argm-tmp označava članove rečenice kojima je opisano vrijeme;
- Argm-mnr označava članove u rečenici koji pripadaju opisu načina na koji je neka radnja izvedena.

Tablica 3.3 prikazuje broj semantičkih uloga u cijelom označenom korpusu. S obzirom da je broj oznaka ARGM-DIR i ARG4 relativno malen u cijelom korpusu, te oznake nisu uključene u konačan korak vrednovanja automatskih klasifikatora jer se nakon podjele podataka na tri skupa dogodio slučaj da je broj riječi s tim oznakama premalen u jednom ili dva skupa pa ocjene automatskog klasifikatora za tu oznaku nemaju dovoljnu značajnost.

Tablice 3.4 i 3.5 prikazuju raspodjelu sintaktičkih uloga u odnosu na semantičke uloge i obratno (3.5) raspodjelu semantičkih uloga u odnosu sintaktičke. Iz tablice

Semantička uloga	Broj oznaka
Arg0	1399
Arg1	3280
Arg2	678
Arg3	145
Arg4	23
Argm-neg	64
Argm-dir	34
Argm-cau	609
Argm-loc	791
Argm-tmp	370
Argm-mnr	155

Tablica 3.3: Broj oznaka u korpusu

Uloga	Broj	Četiri najčešće sintaktičke uloge								Ostale
Arg1	3280	Atr	23.9	Obj	18.9	Sb	10.7	Pred	9.0	37.5
Arg0	1399	Sb	55.7	Atr	18.7	Ap	17.5	Obj	2.8	5.4
Argm-loc	791	Atr	39.4	Prep	32.4	Adv	14.1	Obj	6.1	8.1
Arg2	678	Atr	31.6	Obj	19.9	Prep	8.3	Pred	7.7	32.5

Tablica 3.4: Raspodjela sintaktičkih oznaka za najčešće semantičke uloge u postocima

Uloga	Broj	Četiri najčešće semantičke uloge								Ostale
Atr	2044	Arg1	38.3	loc	15.2	Arg0	12.8	Arg2	10.5	23.2
Sb	1296	Arg0	60.0	Arg1	27.1	Arg2	3.8	cau	2.9	6.2
Obj	1024	Arg1	60.6	Arg2	13.2	cau	6.2	loc	4.7	15.2
Prep	809	loc	31.6	Arg1	31.3	cau	11.0	tmp	9.3	16.8

Tablica 3.5: Raspodjela semantičkih oznaka za najčešće sintaktičke uloge u postocima

3.4 možemo uočiti da većina semantičkih uloga ARG1 ima sintaktičku oznaku OBJ ili ATR, što su oznake objekta i njegovih atributa, a semantička uloga ARG0 ima najčešće sintaktičke oznake SB i ATR što su oznake subjekta i njegovih atributa. Prema tim podacima iz tablice 3.4 izrađen je osnovni model.

4. Modeli za plitko semantičko parsanje na hrvatskom jeziku

4.1. Značajke

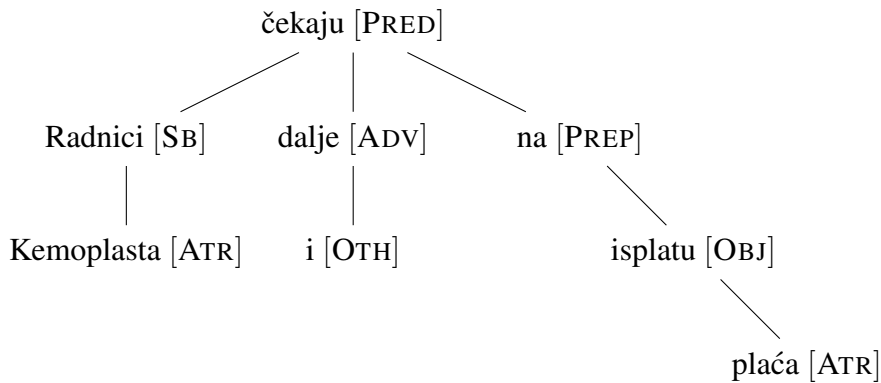
Prilikom treniranja modela temeljenih na strojnom učenju iz podataka je izvučeno nekoliko osnovnih značajki prema uzoru na značajke iz (Palmer et al., 2005). Choi i Palmer (2010) navode razliku između fraznih struktura (engl. *phrase structure*) i ovisnih struktura (engl. *dependency structure*) te opisuju vrlo efikasan heuristički algoritam pretvorbe oznaka iz izvornog fraznog stabla u ovisnosno stablo. U radu također navode da je prebacivanje podataka iz PropBank baze u ovisnosnu strukturu korisnije za modele temeljene na strojnom učenju jer takva struktura sintaktičkog stabla pruža mogućnost boljeg semantičkog razumijevanja prirodnog jezika. U ovom radu dostupno je upravo sintaktičko ovisnosno stablo za svaku rečenicu te su na temelju tog stabla prema uzoru na Choi i Palmer (2010) te Pradhan et al. (2004) korištene sljedeće značajke:

Osnovne sintaktičke značajke obuhvaćaju 6 sintaktičkih oznaka dobivenih parsanjem prema zadatku *CoNLL-X Shared Task: Multi-lingual Dependency Parsing*. Značajke koje ulaze u ovu skupinu su sljedeće: izvorna riječ (*FORM*), lematizirana riječ (*LEMMA*), jednostavan oblik POS oznake (*CPOSTAG*), prošireni oblik POS oznake (*POSTAG*), riječ roditeljskog člana u sintaktičkoj strukturi (*HEAD WORD*), sintaktička uloga u rečenici (*DEPREL*).

Značajka putanje po stablu (*PATH*) kojom se određuje niz oznaka sintaktičkih uloga koje se nalaze na putanji u stablu od korijena (predikata) do trenutne riječi, bez ograničenja na duljinu putanje. Prema primjeru 4.1, riječ *plaća* bi u tom slučaju imala sljedeću vrijednost značajke putanje: PRED→PREP→OBJ→ATR

Pozicija riječi (*BEFORE*) je jednostavna binarna značajka kojom se određuje prethodi li trenutna riječ u rečenici predikatu.

Sintaktičke oznake u prozoru (*WINDOW*) određuju skup značajki kojima su za



Slika 4.1: Primjer sintaktičkog ovisnosnog stabla za rečenicu iz korpusa

trenutnu riječ izdvojene sintaktičke oznake od tri riječi koje neposredno prethode trenutnoj riječi te isto tako od tri riječi koje neposredno slijede trenutnu riječ u rečenici.

Pripadnost semantičkom okviru¹ (*FRAME*) određuje semantičko značenje glagola koji se nalazi u korijenu sintaktičke ovisnosne strukture što ujedno određuje pripadnost glagola određenom semantičkom okviru ovisno o značenju tog glagola.

4.2. Priprema podataka

U ovom je radu semantičkim ulogama označeno 1,214 rečenica iz korpusa. Od cijelog skupa podataka oko 70% (850 rečenica, 10509 riječi) je korišteno kao skup podataka $\mathcal{D}$ za treniranje modela i u optimizaciji hiperparametara modela, 15% (181 rečenica, 2321 riječi) je korišteno za odabir značajki te je 15% (183 rečenica, 2394 riječi) korišteno u vrednovanju modela kao skup podataka za vrednovanje.

Za vrednovanje modela temeljenih na nadziranom strojnom učenju pa i za predo-bradu podataka i odabir hiperparametara korištena je biblioteka *scikit-learn*² za programski jezik Python u kojem je napisano cijelo programsko rješenje za sve automatske postupke u ovom radu. Gotovo sve korištene značajke imaju reprezentaciju u obliku niza znakova³ koje je potrebno preslikati u vektorski prostor. Za preslikavanje značajki u vektorski prostor korišten je *DictVectorizer*⁴ iz biblioteke *scikit-learn*.

¹Određivanje značenja glagola kao riječi nije dio zadatka u ovom radu te se smatra zasebnim zadatkom (engl. *word-sense disambiguation*) te je zbog toga oznaka pripadnosti semantičkom okviru kojom se posredno određuje značenje samog glagola uzeta u skup značajki.

²<http://scikit-learn.org/stable/>

³Dvije značajke imaju brojevnú reprezentaciju te ti brojevi poprimaju binarnu vrijednost 1 ili 0

⁴http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.DictVectorizer.html

4.3. Osnovni model

U poglavlju 2.1. razmatran je pristup da u općenitom slučaju semantička uloga ARG0 (kojim je označen vršioc radnje) pripada subjektu u rečenici, dok semantička uloga ARG1 (kojim je označeno ono nad čime se radnja vrši) pripada objektu. Također je pokazan slučaj u kojem ta pretpostavka ne vrijedi no s obzirom da su ti slučajevi u normalnom (novinskom) tekstu rijeđi može se generalizirati te oznakom ARG0 označavati subjekt, a oznakom ARG1 objekt. Na toj pretpostavci je izrađen osnovni model koji pomoću definiranih pravila automatski označava riječi semantičkim ulogama ARG0 i ARG1.

Ostale semantičke uloge (ARG2, ARG3, itd.) u ovom modelu nisu korištene jer za njih nije strogo definirana uporaba i ovise o semantičkom okviru. Rezultati ovog osnovnog modela su korišteni kao donja granica za ostale modele temeljene na strojnom učenju. Primjer pseudokoda 4.1 opisuje jednostavnu heuristiku koja je korištena za osnovni klasifikator. Pomoću značajke putanje u sintaktičkom stablu moguće je izvući informacije o tome ako je trenutna riječ objekt, subjekt ili atribut jednom od njih. Za primjer iz slike 3.1 ovaj klasifikator bi označio rečenicu sljedećim oznakama:

[Radnici_{Arg0}] [Kemoplasta_{Arg0}] i dalje čekaju na [isplatu_{Arg1}] [plaća_{Arg1}].

Kod 4.1: Pseudokod označavanja riječi za osnovni model

```
oznaci(riječ):  
    ako je_subj(riječ) ili je_atr_od_subj(riječ):  
        vrati Arg0  
    ako je_obj(riječ) ili je_atr_od_obj(riječ):  
        vrati Arg1  
    inace:  
        vrati nista
```

4.4. Modeli temeljeni na nadziranom strojnom učenju

4.4.1. Logistička regresija

Logistička regresija (engl. *logistic regression*) je diskriminativni model koji za izlaz ima vjerojatnosno tumačenje pa je zbog toga ujedno i vjerojatnosni model. Uvođenjem nelinearne funkcije f može se poopćiti linearan regresijski model:

$$h(\mathbf{x}) = f(\mathbf{w}^T \mathbf{x} + w_0) \quad (4.1)$$

Funkcija f naziva se aktivacijska funkcija, a zadatak te funkcije je da linearnu funkciju preslika na jedinični interval. Za logističku regresiju je tako definirana funkcija logistička ili sigmoidalna funkcija:

$$\sigma(\alpha) = \frac{1}{1 + \exp(-\alpha)} \quad (4.2)$$

Vrijednost α je definirana kao:

$$\alpha = \mathbf{w}^T \mathbf{x} + w_0 = \tilde{\mathbf{w}}^T \tilde{\mathbf{x}} \quad (4.3)$$

Učenje modela logističke regresije se tako svodi na određivanje parametara $\tilde{\mathbf{w}}$, a optimizacija parametara se svodi na minimizaciju funkcije pogreške na skupu za učenje:

$$E(\tilde{\mathbf{w}}|\mathcal{D}) = - \sum_{i=1}^N \{y^{(i)} \ln(h(\mathbf{x}^{(i)})) + (1 - y^{(i)}) \ln(1 - h(\mathbf{x}^{(i)}))\} + \frac{\lambda}{2} \sum_{j=1}^n |w_j|^q \quad (4.4)$$

gdje je $\frac{\lambda}{2} \sum_{j=1}^n |w_j|^q$ regularizacijski izraz kojim se povećava funkcija pogreške za složenije modele čime se smanjuje slučaj prenaučivosti modela.

U ovom radu korištena je biblioteka `sklearn.linear_model.LogisticRegression`⁵ te su najbolji rezultati postignuti uz L2 regularizaciju ($q = 2$ u regularizacijskom izrazu)

4.4.2. Stroj potpornih vektora

Stroj potpornih vektora (engl. *support vector machine*, *SVM*) je diskriminativan model temeljen na nadziranom strojnom učenju te, za razliku od logističke regresije, nema vjerojatnosnu interpretaciju. U osnovnom obliku ovaj klasifikator je binaran, a osnovna ideja za izgradnju ovog modela je određivanje hiperravnine koja razdvaja primjere iz skupa za učenje u prostoru značajki (engl. *feature space*). Ovaj model je iznimno popularan u području obrade prirodnog jezika te je uvođenjem tzv. jezgrenog trika (engl. *kernel trick*) proširiv i na nelinearnu klasifikaciju.

Za zadani skup linearno razdvojivih primjera za učenje sljedećeg oblika:

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^p, y_i \in \{-1, 1\}\}_{i=1}^n \quad (4.5)$$

⁵http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Možemo definirati linearan model:

$$h(x) = \mathbf{w}^T \phi(\mathbf{x}) + w_0 \quad (4.6)$$

koji za skup $\mathcal{D}$ ima beskonačno mnogo rješenja no samo jedno rješenje prema kriteriju maksimalne margine (engl. *maximum margin*), što znači da, za linearno razdvojive primjere, postoji jedna hiperravnina čija je udaljenost do najbližeg primjera najveća moguća. U slučaju hiperravnine kojom su sve točke ispravno klasificirane vrijedi da je udaljenost između primjera i hiperravnine sljedeća:

$$\frac{y^{(i)} h(\mathbf{x}^{(i)})}{\|\mathbf{w}\|} = \frac{y^{(i)} (\mathbf{w}^T \phi(\mathbf{x}^{(i)}) + w_0)}{\|\mathbf{w}\|} \quad (4.7)$$

Prema tome maksimalna margina bit će jednaka udaljenosti najbližeg primjera te je time formuliran optimizacijski problem SVM klasifikatora:

$$\arg \max_{\mathbf{w}, w_0} \left\{ \frac{1}{\|\mathbf{w}\|} \min_i \{ y_{(i)} (\mathbf{w}^T \phi(\mathbf{x}^{(i)}) + w_0) \} \right\} \quad (4.8)$$

Ovaj optimizacijski problem moguće je riješiti metodom Lagrangeovih multiplikatora te je također moguće formulirati dualni problem kojim se smanjuje količina računanja.

U slučaju linearno nerazdvojivih primjera Cortes i Vapnik (1995) uvode formulaciju problema mekom marginom (engl. *soft margin*) te je time riješena klasifikacija linearno nerazdvojivih primjera. Također je uvođenjem jezgrenih funkcija od linearnog modela SVM-a moguće dobiti nelinearan model. Jezgrenom funkcijom može se zamijeniti umnožak dvaju primjera $\mathbf{x}$ i $\mathbf{x}'$ (koji se nalazi u definiciji dualnog problema) u prostoru značajki:

$$\kappa(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}') \quad (4.9)$$

Jezgrena funkcija mjeri sličnost dvaju vektora u nekom prostoru značajki. Umjesto preslikavanja vektora u prostor značajki i računanja sličnosti vektora kao skalarni produkt, može se izravno izračunati sličnost značajki pomoću jezgrene funkcije, što znači da se time može smanjiti računalna složenost i povećati broj dimenzija prostora značajki. Prilikom preslikavanja iz ulaznog prostora u prostor značajki možemo odabrati jednu od tri strategije:

- *Izravno oblikovanje*: odabrati preslikavanje ϕ , preslikati primjere u prostor značajki i trenirati model;

- *Izravno oblikovanje s jezgrenom funkcijom*: odabrati preslikavanje ϕ , izračunati jezgrenu funkciju i trenirati model s tom funkcijom;
- *Inverzno oblikovanje*: odabrati jezgrenu funkciju κ te izravno trenirati model tom funkcijom bez poznavanja funkcije preslikavanja ϕ .

U ovom radu korištena je treća strategija jer je jezgrenu funkciju lakše definirati te uz to biblioteka *sklearn.svm*⁶ nudi implementaciju nekih poznatih jezgrenih funkcija koje su opisane u nastavku:

- *Linearna jezgra*: $\kappa = (\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$
- *Plinomijalna jezgra*: $\kappa = (\mathbf{x}, \mathbf{x}') = (\gamma(\mathbf{x}^T \mathbf{x}') + r)^d$
- *Radijalne bazne funkcije (RBF)*: $\kappa = (\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|)$
- *Sigmoida*: $\kappa = (\mathbf{x}, \mathbf{x}') = \tanh(\gamma(\mathbf{x}^T \mathbf{x}') + r)$

Odabir najbolje jezgrene funkcije od četiri dostupne uključen je u algoritam odabira hiperparametara, tj. algoritam iscrpnog pretraživanja po rešetci. Za SVM najbolje rezultate postiže model s radijalnom baznom funkcijom čiji su rezultati prikazani u poglavlju 5.

Pradhan et al. (2004) koristi SVM za plitko semantičko parsanje engleskog jezika na PropBank korpusu te postiže vrijednost F1-mjere od 89.4%.

Oba modela temeljena na strojnom učenju opisana u ovom radu proširena su za višeklasnu klasifikaciju primjera klasifikacijom tipa jedan-naspram-jedan kod koje je potrebno naučiti po jedan binarni klasifikator za svaki par klasa.

⁶<http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC>

5. Vrednovanje

5.1. Mjere

U postupku vrednovanja korištene su tri mjere: preciznost (engl. *precision*), odziv (engl. *recall*), F1-mjera (engl. *F1-score*). Mjera preciznost (P) definirana je kao udio točno klasificiranih primjera u skupu pozitivno klasificiranih primjera:

$$P = \frac{TP}{TP + FP} \quad (5.1)$$

Mjera odziv (R) definirana je kao udio točno klasificiranih primjera u skupu svih pozitivnih primjera:

$$R = \frac{TP}{TP + FN} \quad (5.2)$$

Mjera F1¹ definirana je sljedećim izrazom:

$$F_1 = \frac{2TP}{2TP + FP + FN} \quad (5.3)$$

5.2. Odabir značajki

Iako biblioteka *scikit-learn* pruža implementaciju odabira značajki, ta implementacija u ovom slučaju nije korištena zbog jednog bitnog razloga, a taj je što ti implementirani algoritmi koriste već preslikane primjere u prostoru značajki, što znači da od nekoliko desetaka značajki ulaznih podataka te primjere preslikavamo u vektorski prostor dimezije oko 50,000 te u tom prostoru vršimo odabir značajki (tj. smanjujemo dimenziju tog prostora). Ovaj implementirani algoritam bi se mogao koristiti te bi se, nakon odbacivanja određenih značajki u prostoru značajki, moglo analizirati koje su to značajke na ulazu imale najviše utjecaja na preslikane dimenzije u prostoru značajki, no takvom metodom ne bi imali informaciju koliko određena značajka ima utjecaja na

¹Van Rijsbergen (1974)

rezultate modela. Implementiranim metodom odabira značajki dobijemo samo informaciju o tome koje su preslikane značajke odbačene (one koje imaju "najgori" utjecaj na rezultate modela) kako bi se smanjila dimenzija prostora. Ovime se ne može sa sigurnošću odrediti koja ulazna značajka ima negativan utjecaj na rezultate modela jer jedna ulazna značajka se općenito preslikava u nekoliko stotina ili više dimenzija u prostoru značajki, a tijekom izvođenja implementiranog algoritma odabira značajki samo neke od tih dimenzija će biti odbačene što ne znači da je ulazna značajka nužno loša za model.

Zbog problema preslikavanja ulaznih značajki u prostor značajki implementiran je jednostavan algoritam. Algoritmom se odbacuje jedna ulazna značajka te se dobiveni rezultati uspoređuju s obzirom na one sa svim značajkama.

Optimizacija hiperparametara prethodila je koraku odabira značajki. Za odabir značajki korišten je skup od 15% primjera iz ukupnog skupa primjera. Postupkom odabira značajki analiziran je utjecaj svake pojedinačne značajke na rezultate vrednovanja na prethodno spomenutom skupu.

SVM		Logistička regresija	
Značajka	Utjecaj (%)	Značajka	Utjecaj (%)
PATH	-0.57	PATH	-0.49
HEAD WORD	-0.43	POSTAG	-0.49
POSTAG	-0.31	BEFORE	-0.26
CPOSTAG	-0.30	FORM	-0.10
BEFORE	-0.29	CPOSTAG	-0.04
WINDOW	-0.04	DEPREL	0.10
FORM	0.08	LEMMA	0.36
DEPREL	0.16	HEAD WORD	0.66
LEMMA	0.19	WINDOW	0.66

Tablica 5.1: Utjecaj značajki F1-mjeri za modele temeljene na strojnom učenju

Tablica 5.1 prikazuje koliki utjecaj ima pojedina značajka na F1-mjeru za oba modela (SVM i logistička regresija). Prema ovim brojkama može se zaključiti da je utjecaj pojedine značajke vrlo malen na konačne rezultate te se može svrstati u područje statističke pogreške. Također se mora uzeti u obzir da sintaktičke oznake nisu ispravne za svaku riječ jer su dobivene automatskim sintaktičkim parsanjem.

Za oba modela je pokušano grupiranje značajki koje imaju zajednička obilježja (npr. POSTAG i CPOSTAG za SVM) i razmatran slučaj ponašanja modela s cijelom

grupom značajki i bez nje. U svim takvim slučajevima je F1-mjera modela pala za nešto veći iznos (5-10%), što upravo znači da su značajke povezane te isključivanje jedne donosi neznatne promjene u performansama modela no izbacivanjem cijele grupe značajki odbacujemo skup informacija koji može poslužiti modelu kod klasifikacije. Zbog toga je u postupku vrednovanja oba modela korišten cijeli skup značajki.

U oba modela najveći negativni utjecaj na F1-mjeru imala je značajka putanje po sintaktičkom ovisnosnom stablu jer s obzirom na većinu značajki može poprimiti vrlo veliki raspon vrijednosti (vrlo rijetko se putanje za sintaktičke članove u rečenici ponavljaju što ne pomaže klasifikatoru u generalizaciji).

5.3. Rezultati

Iz tablice 5.2 može se izračunati preciznost, odziv i F1 ocjena za osnovni model. Iako je u tablici zabune uključena prazna oznaka čime se označuju članovi rečenice koji ne pripadaju niti jednoj semantičkoj ulozi, te oznake ne ulaze u vrednovanje klasifikatora. Prema matrici zabune osnovni model postiže rezultate navedene u tablici 5.3. Iz dobivenih rezultata možemo zamijetiti da osnovni model ima veću preciznost od odziva u slučaju obje semantičke uloge, što znači da je model relativno jednostavan što dovodi do toga da lošije klasificira primjere u skupu svih pozitivnih primjera. Mogao se razmatrati slučaj u kojem osnovni model s oznakom ARG0 označava cijelu granu u sintaktičkom stablu koja u sebi sadrži sintaktičku oznaku subjekta SB, a oznakom ARG1 cijelu granu sintaktičkog stabla koja u sebi sadrži oznaku objekta OBJ, no time bi ujedno višestruko umanjili mjeru preciznosti modela² pa takvo pravilo nije razmatrano.

		Stvarno		
		∅	Arg0	Arg1
Klasificirano	∅	1420	54	340
	Arg0	86	142	52
	Arg1	143	19	138

Tablica 5.2: Matrica zabune za osnovni model

Tablica 5.4 prikazuje rezultate vrednovanja modela SVM. Za semantičke uloge Arg3 i Argm-neg vrijednosti mjera su ispale vrlo loše zbog toga što je tih primjera u skupu za učenje najmanje pa model nije uspio dobro generalizirati te semantičke uloge.

²Većina riječi u korpusu nema oznake što bi dovelo do velikog pada preciznosti modela

	Preciznost	Odziv	F1-mjera	Broj primjera
Arg0	50.71	66.05	57.37	215
Arg1	46.00	26.04	33.25	530
Ukupno	47.36	37.58	40.21	745

Tablica 5.3: Rezultati vrednovanja osnovnog modela u postocima

	Preciznost	Odziv	F1-mjera	Broj primjera
Arg0	72.81	73.49	73.15	215
Arg1	66.54	68.68	67.60	530
Arg2	58.89	63.10	60.92	84
Arg3	0.00	0.00	0.00	26
Argm-cau	17.14	7.06	10.00	85
Argm-loc	66.12	48.78	56.14	164
argm-mnr	100.00	25.00	40.00	24
Argm-neg	0.00	0.00	0.00	6
Argm-tmp	62.50	28.99	39.60	69
Ukupno	62.25	57.11	58.54	1203

Tablica 5.4: Rezultati vrednovanja modela SVM

U tablici se također može uočiti da model najbolje rezultate postiže za semantičke uloge Arg0, Arg1 i Arg2, što je očekivano jer tih primjera u skupu za učenje ima dovoljno da bi model mogao dobro generalizirati.

Tablica 5.5 prikazuje rezultate vrednovanja za model logistička regresija. I kod ovog modela kao i kod SVM-a rezultati su isti za semantičke uloge Arg3 i Argm-neg zbog istog razloga kao i kod modela SVM. I ovaj model kao i SVM najbolje rezultate postiže za uloge Arg0, Arg1 i Arg2.

	Preciznost	Odziv	F1-mjera	Broj primjera
Arg0	71.56	74.88	73.18	215
Arg1	65.20	70.00	67.52	530
Arg2	58.70	64.29	61.36	84
Arg3	0.00	0.00	0.00	26
Argm-cau	22.22	9.41	13.22	85
Argm-loc	67.54	46.95	55.40	164
Argm-mnr	100.00	16.67	28.57	24
Argm-neg	0.00	0.00	0.00	6
Argm-tmp	54.29	27.54	36.54	69
Ukupno	61.50	57.69	58.26	1203

Tablica 5.5: Rezultati vrednovanja modela logističke regresije

		Preciznost	Odziv	F1-mjera
Osnovni klasifikator	Arg0	50.71	66.05	57.37
	Arg1	46.00	26.04	33.25
	Ukupno	47.36	37.58	40.21
Logistička regresija	Arg0	71.56	74.88	73.18
	Arg1	65.20	70.00	67.52
	Ukupno	67,04	71,40	69,15
SVM	Arg0	72.81	73.49	73.15
	Arg1	66.54	68.68	67.60
	Ukupno	68,35	70.07	69,20

Tablica 5.6: Usporedba rezultata s osnovnim modelom

Tablica 5.6 prikazuje usporedbu rezultata modela temeljenih na nadziranom strojnom učenju s osnovnim modelom. Iz ove tablice očito je da modeli temeljeni na stroj-

nom učenju ostvaruju bolje rezultate u svim mjerama za vrednovanje. Model ostvaruje najbolje rezultate u preciznosti i F1-mjeri, dok model logistička regresija ostvaruje najbolje rezultate u mjeri odziv. Oba modela temeljena na strojnim učenjem za 20-30% nadmašuju rezultate osnovnog modela. Najveća razlika u između najboljeg i osnovnog modela je u odzivu i iznosi gotovo 34%. U usporedbi s rezultatima na engleskom jeziku ovi rezultati su lošiji, ali usporedivi. Palmer et al. (2005) navodi da je za engleski jezik koristeći automatski parsane sintaktičke oznake postignuta ocjena preciznosti od 69.9%, a ocjena odziva od 61.1% dok je s druge strane Pradhan et al. (2004) postiže mnogo bolje rezultate sa F1-mjerom od 89.4%.

	∅	Arg0	Arg1	Arg2	Arg3	Cau	Loc	Mnr	Neg	Tmp
∅	1013	32	128	15	12	44	53	11	4	32
Arg0	24	158	19	5	0	3	6	0	0	2
Arg1	96	19	364	7	11	18	21	4	2	5
Arg2	12	3	5	53	0	12	1	3	0	1
Arg3	0	0	0	0	0	0	0	0	0	0
Cau	16	2	5	0	1	6	3	0	0	2
Loc	20	1	7	3	1	2	80	0	0	7
Mnr	0	0	0	0	0	0	0	6	0	0
Neg	1	0	0	1	0	0	0	0	0	0
Tmp	9	0	2	0	1	0	0	0	0	20

Tablica 5.7: Matrica zabune za model SVM

Tablice 5.7 i 5.8 prikazuju matrice zabune za oba automatska klasifikatora te se pomoću njih može detaljnije analizirati gdje su najviše prisutne pogreške kod klasifikacije. Za oba modela vrijedi da najgore rezultate postižu za dvije semantičke uloge: ARG3 i ARG-M-NEG. Za semantičku ulogu ARG3 može se uočiti da su modeli klasificirali taj član bez semantičke uloge ili sa semantičkom ulogom ARG1, što znači da modeli nisu mogli, uz tako mali broj primjera, razlučiti razliku između tih semantičkih uloga.

Skup za vrednovanje sadrži vrlo malen broj primjera označenih oznakom ARG-M-NEG te dobiveni rezultati imaju premalo utjecaja da bi se moglo nešto generalno zaključiti o klasifikaciji takvih primjera. Jedine značajke koje su mogle indicirati da se radi o oznaci ARG-M-NEG su: *FORM* i *LEMMA*, što nije bilo dovoljno da model ispravno klasificira tih nekoliko primjera iz skupa za vrednovanje.

Oznaka ARG-M-CAU također je kod klasifikatora ostvarila lošije rezultate jer je, s

	∅	Arg0	Arg1	Arg2	Arg3	Cau	Loc	Mnr	Neg	Tmp
∅	1008	24	117	12	11	41	58	10	2	34
Arg0	27	161	22	4	0	4	4	1	0	2
Arg1	100	22	371	10	11	19	19	5	3	9
Arg2	11	4	5	54	0	11	3	3	0	1
Arg3	0	0	0	0	0	0	0	0	0	0
Cau	16	2	2	1	2	8	3	1	1	0
Loc	18	2	8	2	1	2	77	0	0	4
Mnr	0	0	0	0	0	0	0	4	0	0
Neg	0	0	1	1	0	0	0	0	0	0
Tmp	11	0	4	0	1	0	0	0	0	19

Tablica 5.8: Matrica zabune za model logistička regresija

obzirom na ostale generalne uloge, ulogu ARGM-CAU teže pronaći pa čak i kada ljudi označavaju tekst što je samo doprinjelo lošijim rezultatima automatskih klasifikatora.

Već prije je spomenuto kako je korišteni korpus iz novinske domene te sadrži rečenice koje su pisane određenim stilom i na taj način su oblikovane. Velik problem automatskim klasifikatorima zadaju složene rečenice čiji su korijenski predikati u sintaktičkoj strukturi glagoli: *pisati*, *govoriti*, *tvrditi*. Ovi glagoli u općenitom govoru se koriste na uobičajen način, kao što je navedeno sljedećim primjerima:³

Ana piše pismo Ivici.

Trener je govorio igračima da se povuku u obranu.

Dok se u normalnom govoru navedeni glagoli koriste u slučaju opisa radnje pisanje, govora ili izricanja tvrdnje, u novinskom tekstu ti glagoli najčešće služe za citiranje neke osobe ili institucije kako je to prikazano sljedećim primjerima iz korpusa:

Dikljani su ogorčeni, govori Milin.

Umjetniku su čak prijeli smrcu, piše Metro.

Ovim primjerima je prikazano kako jedna cijela rečenica koja je dio složene rečenice može biti u semantičkom smislu jedna semantička uloga, u ovim primjerima je

³Ovi primjeri se ne nalaze u korpusu te su izmišljeni samo kako bi se prikazala razlika u odnosu na primjere iz korpusa.

ta uloga ARG1. Iako to za osobu koja označava tekst nije nešto neuobičajeno, za modele temeljene na strojnom učenju ovakve rečenice uzrokuju šum u podacima jer bi u općenitom slučaju svaka jednostavna rečenica mogla imati više različitih semantičkih uloga, a dok su sastavljene u ovakav oblik tada cijela jedna jednostavna rečenica ima jednu semantičku ulogu.

Ovakve iznimne slučajeve gdje se pojavljuje šum u podacima zbog specifične upotrebe određenih glagola možda bi se moglo riješiti dodavanjem nekoliko značajki kojima bi se moglo odrediti radi li se o jednostavnoj ili složenoj rečenici te pripada li sintaktički svaki član rečenice korijenskom predikatu ili nekom drugom predikatu.

Također je potrebno naglasiti kako je jedan veliki nedostatak činjenica da je samo jedna osoba izrađivaja semantičke okvire te na temelju njih označavala korpus, čime su u označene primjere zasigurno unesene greške koje se mogu ispraviti tako da više osoba označi iste primjere te se napravi analiza grešaka. Pogreške u označavanju su se dodatno mogle smanjiti korištenjem korpusa koji je ručno označen sintaktičkim oznakama (Klubička i Ljubešić, 2014) jer automatski parser na testiranim podacima postiže mjeru točnosti od 80%.

6. Zaključak

Ovim radom prikazan je cjelokupni postupak pripreme semantičkih okvira, označavanja korpusa i vrednovanja modela temeljenih na nadziranom strojnom učenju. Za potrebe označavanja izrađen je, prema uzoru na projekt PropBank, skup semantičkih okvira, čime su određene semantičke uloge za pojedini glagol. Uz pomoć semantičkih uloga ručno je označen relativno malen skup rečenica koje su već prethodno sadržavale mnoge sintaktičke oznake dobivene automatskim parsanjem. Označene rečenice korištene su za uvježbavanje modela te kasnije njihovo vrednovanje.

Najbolji rezultati postignuti su modelom stroj potpornih vektora (s *RBF* jezgrenom funkcijom) čija F1-mjera iznosi 58.54. Uz automatske klasifikatore temeljene na strojnom učenju izvršeno je vrednovanje osnovnog klasifikatora temeljenog na pravilima. Vrednovanje osnovnog klasifikatora bilo je moguće izvršiti samo na podskupu semantičkih uloga i pokazano je da modeli temeljeni na strojnom učenju nadmašuju rezultate postignute osnovnim modelom.

Postignuti rezultati su zadovoljavajući, no s obzirom na rezultate nekih klasifikatora za druge jezike očito je da ima još mnogo mogućnosti za poboljšanje ovih modela. Postoji niz razloga koji su negativno utjecali na dobivene rezultate. Jedan od razloga je taj da je korišten korpus koji sadrži sintaktičke oznake dobivene automatskim sintaktičkim parsanjem te samim time uvodi greške već na sintaktičkoj razini.

Uz automatsko sintaktičko parsanje korpusa, dodatni šum u podacima je stvoren prilikom stvaranja semantičkih okvira i označavanja korpusa semantičkim ulogama jer je taj zadatak vršila samo jedna osoba čime je zasigurno uveden povećí broj grešaka u označenom korpusu. Za bolje označavanje korpusa semantičkim ulogama i bolju definiciju semantičkih okvira potrebno je uključiti više osoba (poželjno osobe iz područja lingvistike i drugih srodnih znanstvenih grana) u postupak čime bi se umanjio broj grešaka u ručnom označavanju.

Kao što je ranije spomenuto, označen je relativno malen broj rečenica što, opet, ima utjecaj na lošije rezultate modela temeljenih na nadziranom strojnom učenju. Eventualnim proširenjem skupa označenih primjera omogućilo bi se modelima da imaju više

primjera za učenje čime bi postigli bolje rezultate.

Za buduće istraživanje na temelju ovog rada planira se koristiti korpus koji sadrži ručno označene primjere sa sintaktičkim oznaka koji se inicijalno koristio za uvježbavanje automatskih sintaktičkih parsera. Takve primjere je moguće pomoću nekoliko osoba ručno označiti kako bi se dobio neki skup podataka značajnije veličine. Također je u planu stvaranje ili analiza trenutnih semantičkih okvira. Kao druga opcija za semantičke okvire može biti izbor već gotovih semantičkih okvira dostupnih za hrvatski jezik, čime bi se umanjio broj grešaka prilikom označavanja. Za poboljšanje automatskih klasifikatora moguće je uvesti veći skup s proširenim značajkama, posebice onim značajkama kojima bi se mogla iskoristiti informacija o sintaktičkoj strukturi rečenice s obzirom da su trenutne značajke bile prilagođene drugačijoj sintaktičkoj strukturi.

Planirana poboljšanja u ovom istraživanju mogla bi dovesti do rezultata koji se mogu usporediti s rezultatima istraživanja provedena za druge jezike koja su vezana uz ovaj zadatak.

LITERATURA

- ‡ Agić, Danijela Merkler, i Daša Berović. Parsing croatian and serbian by using croatian dependency treebanks. U *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages*, 2013.
- Collin F Baker, Charles J Fillmore, i John B Lowe. The berkeley framenet project. U *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, stranice 86–90. Association for Computational Linguistics, 1998.
- Hans Christian Boas. Bilingual framenet dictionaries for machine translation. U *LREC*, 2002.
- Bernd Bohnet. Very high accuracy and fast dependency parsing is not a contradiction. U *Proceedings of the 23rd International Conference on Computational Linguistics*, stranice 89–97. Association for Computational Linguistics, 2010.
- Jinho D Choi i Martha Palmer. Retrieving correct semantic boundaries in dependency structure. U *Proceedings of the Fourth Linguistic Annotation Workshop*, stranice 91–99. Association for Computational Linguistics, 2010.
- Corinna Cortes i Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3): 273–297, 1995.
- Charles J Fillmore. Frame semantics and the nature of language*. *Annals of the New York Academy of Sciences*, 280(1):20–32, 1976.
- Daniel Gildea i Julia Hockenmaier. Identifying semantic roles using combinatory categorial grammar. U *Proceedings of the 2003 conference on Empirical methods in natural language processing*, stranice 57–64. Association for Computational Linguistics, 2003.
- Daniel Gildea i Daniel Jurafsky. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288, 2002.

- Daniel Gildea i Martha Palmer. The necessity of parsing for predicate argument recognition. U *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, stranice 239–246. Association for Computational Linguistics, 2002.
- Péter Halácsy, András Kornai, i Csaba Oravecz. Hunpos: an open source trigram tagger. U *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*, stranice 209–212. Association for Computational Linguistics, 2007.
- Mark Johnson. A simple pattern-matching algorithm for recovering empty nodes and their antecedents. U *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, stranice 136–143. Association for Computational Linguistics, 2002.
- Bart Jongejan i Hercules Dalianis. Automatic training of lemmatization rules that handle morphological changes in pre-, in-and suffixes alike. U *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, stranice 145–153. Association for Computational Linguistics, 2009.
- Filip Klubička i Nikola Ljubešić. Using crowdsourcing in building a morphosyntactically annotated and lemmatized silver standard corpus of croatian. U Tomaž Erjavec i Jerneja Žganec Gros, urednici, *Language technologies: Proceedings of the 17th International Multiconference Information Society IS2014*, Ljubljana, Slovenia, 2014.
- Beth Levin. *English verb classes and alternations: A preliminary investigation*. University of Chicago press, 1993.
- Nikola Ljubešić i Tomaž Erjavec. hrwac and slwac: Compiling web corpora for croatian and slovene. U Ivan Habernal i Václav Matousek, urednici, *Text, Speech and Dialogue - 14th International Conference, TSD 2011, Pilsen, Czech Republic, September 1-5, 2011. Proceedings*, Lecture Notes in Computer Science, stranice 395–402. Springer, 2011.
- Nikola Ljubešić i Filip Klubička. {bs,hr,sr}WaC – web corpora of Bosnian, Croatian and Serbian. U *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, stranice 29–35, Gothenburg, Sweden, 2014. Association for Computational Linguistics.
- Mitchell P Marcus, Mary Ann Marcinkiewicz, i Beatrice Santorini. Building a large

- annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2): 313–330, 1993.
- Joakim Nivre i Mario Scholz. Deterministic dependency parsing of english text. U *Proceedings of the 20th international conference on Computational Linguistics*, stranica 64. Association for Computational Linguistics, 2004.
- Kyoko Hirose Ohara, Seiko Fujii, Hiroaki Saito, Shun Ishizaki, Toshio Ohori, i Ryoko Suzuki. The japanese framenet project: A preliminary report. *Proceedings of PA-CLING'03*, stranice 249–254, 2003.
- Martha Palmer, Daniel Gildea, i Paul Kingsbury. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106, 2005.
- Sameer S Pradhan, Wayne Ward, Kadri Hacioglu, James H Martin, i Daniel Jurafsky. Shallow semantic parsing using support vector machines. U *HLT-NAACL*, stranice 233–240, 2004.
- Mark Steedman. *The syntactic process*, svezak 24. MIT Press, 2000.
- Carlos Subirats i Miriam Petruck. Surprise: Spanish framenet. U *Proceedings of CIL*, svezak 17, 2003.
- Carlos Subirats i Hiroaki Sato. Spanish framenet and framesql. U *4th International Conference on Language Resources and Evaluation. Workshop on Building Lexical Resources from Semantically Annotated Corpora. Lisbon (Portugal)*. Citeseer, 2004.
- Cornelis Joost Van Rijsbergen. Foundation of evaluation. *Journal of Documentation*, 30(4):365–373, 1974.
- Magdalena Zawisławska, Magdalena Derwojedowa, i Jadwiga Linde-Usiekniewicz. A framenet for polish. U *Converging Evidence: Proceedings to the Third International Conference of the German Cognitive Linguistics Association (GCLA'08)*, stranice 116–117, 2008.

Plitko semantičko parsanje tekstova na hrvatskome jeziku

Sažetak

U ovom radu definiran je skup semantičkih okvira za manji broj glagola koji su bili najzastupljeniji u dostupnom hrvatskom korpusu. Na temelju tih okvira označen je relativno malen skup rečenica iz korpusa te je izvršeno vrednovanje modela temeljenih na nadziranom strojnom učenju na zadatku plitkog semantičkog parsanja. Pokazano da na tako malom skupu primjera automatski modeli temeljeni na strojnom učenju postižu dobre rezultate $F1 = 58.54$. Na reduciranom skupu semantičkih uloga, modeli temeljeni na strojnom učenju nadmašuju rezultate osnovnog modela te u tom slučaju najbolju F1-mjeru postiže stroj potpornih vektora ($F1 = 69.20$)

Ključne riječi: obrada prirodnog jezika, hrvatski jezik, plitko semantičko parsanje, nadzirano strojno učenje, logistička regresija, stroj potpornih vektora.

Shallow Semantic Parsing of Croatian Text

Abstract

In this thesis a set of semantic frames for smaller number of most common verbs in available Croatian corpus was defined. Based on those semantic frames relatively small set of sentences was labeled with semantic roles and, based on those sentences, the evaluation of supervised machine learning models was performed on a task of shallow semantic parsing. It is shown that even on relatively small training set models based on supervised machine learning can perform well and achieve good results with $F1 = 58.54$. On a reduced set of semantic roles, machine learning models outperform baseline model and in that case the best F1-score is achieved with Support Vector Machine ($F1 = 69.20$).

Keywords: natural language processing, Croatian language, shallow semantic parsing, supervised machine learning, logistic regression, support vector machine.

Dodatak A

Semantički okviri

Predikat: *boriti*

Id: boriti.01

Opis: *nastojati savladati izravnog neprijatelja, protivnika ili suparnika*

Argumenti:

Arg0 - onaj koji se bori

Arg1 - protiv koga/čega se **Arg0** bori

Arg2 - sredstvo čime se **Arg0** bori

Arg3 - za koga/što se **Arg0** bori

Predikat: *čekati*

Id: čekati.01

Opis: *provoditi vrijeme koje je potrebno da tko ili što dođe ili da se što pojavi ili dogodi*

Argumenti:

Arg0 - onaj koji čeka

Arg1 - koga/što se čeka

Predikat: *čuti*

Id: čuti.01

Opis: *primiti/primati sluhom glas ili glasove*

Argumenti:

Arg0 - onaj koji čuje

Arg1 - što **Arg0** čuje

Predikat: *čuti se*

Id: čuti.02

Opis: *razg. osjetiti sluhom neki zvuk. Registrirati zvuk ili čuti glasine o negome.*

Argumenti:

Arg1 - tko/što se čuje

Predikat: *dati*

Id: dati.01

Opis: *dodijeliti nekome nešto.*

Argumenti:

Arg0 - onaj koji daje

Arg1 - što se daje

Arg2 - kome se daje

Predikat: *dobiti*

Id: dobiti.01

Opis: *primiti nešto.*

Argumenti:

Arg0 - osoba koja prima nešto

Arg1 - ono što **Arg0** prima

Arg2 - onaj koji daje nešto

Predikat: *dodati*

Id: dodati.01

Opis: *uvećati što u odnosu na već postojeću masu, volumen ili broj; dometnuti*

Argumenti:

Arg0 - onaj koji dodaje

Arg1 - što se dodaje

Arg2 - čemu se dodaje

Arg3 - rezultat dodavanja

Predikat: *dodati*

Id: dodati.02

Opis: *pružiti kome što; dati; u sportu prenijeti posjed nečeg nekom drugom; reći još nešto; dometnuti*

Argumenti:

Arg0 - onaj koji dodaje

Arg1 - što se dodaje

Arg2 - kome se dodaje

Predikat: *dokazati*

Id: dokazati.01

Opis: *činjenicama utvrditi istinitost, izvesti dokaz iz činjenica i njihovih veza*

Argumenti:

Arg0 - onaj koji dokazuje

Arg1 - što **Arg0** dokazuje

Arg2 - kome **Arg0** dokazuje

Arg3 - dokaz, čime je nešto dokazano

Predikat: *dokazati se*

Id: dokazati.02

Opis: *postupcima, radom, uspjesima postići dobro mišljenje o sebi među stručnjacima, u javnosti itd.*

Argumenti:

Arg0 - onaj koji se dokazuje

Arg1 - kome se **Arg0** dokazuje

Arg2 - u čemu se **Arg0** dokazuje

Predikat: *gledati*

Id: gledati.01

Opis: *imati pogled upravljen na koga ili što; promatrati*

Argumenti:

Arg0 - onaj koji gleda

Arg1 - tko/što se gleda

Predikat: *gledati*

Id: gledati.02

Opis: *odnositi se prema komu ili čemu, imati stav ili mišljenje o komu ili čemu*

Argumenti:

Arg0 - onaj koji ima gledište/stav

Arg1 - za što (o kome) ima **Arg0** stav

Arg2 - kakav stav **Arg0** ima

Predikat: *gledati*

Id: gledati.03

Opis: *nastojati*

Argumenti:

Arg0 - onaj koji nastoji

Predikat: *govoriti*

Id: govoriti.01

Opis: *riječima iznositi neki smisao (tekst, obično bez predložka, za razliku od čitati), razgovarati*

Argumenti:

Arg0 - onaj koji govori

Arg1 - slušatelj, kome se govori

Arg2 - subjekt o kome se govori

Predikat: *graditi*

Id: graditi.01

Opis: *podizati, zidati; izrađivati, praviti (o radnjama koje su nalik zidanju); stvarati nešto novo.*

Argumenti:

Arg0 - graditelj

Arg1 - što se gradi

Arg2 - sredstvo pomoću čega se gradi

Predikat: *graditi se*

Id: graditi.02

Opis: *praviti se kao da se nešto ne zna, izigravati neupućenoga u činjenice*

Argumenti:

Arg0 - onaj koji se gradi

Arg1 - od koga/čega se **Arg0** gradi

Predikat: *graditi se*

Id: graditi.03

Opis: *biti u procesu gradnje/podizanja.*

Argumenti:

Arg1 - što je u procesu gradnje

Predikat: *igrati*

Id: igrati.01

Opis: *igrati neku igru.*

Argumenti:

Arg0 - igrač(i)

Arg1 - igra, vrsta igre

Arg2 - suparnik, protiv koga se igra

Arg3 - osoba, momčad, ekipa s kojom **Arg0** igra

Predikat: *igrati se*

Id: igrati.02

Opis: *odvijati se*

Argumenti:

Arg1 - što se igra (odvija)

Predikat: *izjaviti*

Id: izjaviti.01

Opis: *dati izjavu, reći, riječima javno oblikovati neki trajniji ili službeni stav*

Argumenti:

Arg0 - onaj koji izjavljuje
Arg1 - izjava, stav koji se izjavljuje
Arg2 - kome (za koga) se daje izjava
Arg3 - o kome/čemu se daje izjava

Predikat: *koristiti (se)*

Id: koristiti.01

Opis: *upotrebljavati koga ili što da bi se postigao neki rezultat*

Argumenti:

Arg0 - onaj koji koristi
Arg1 - tko/što se koristi

Predikat: *kupiti*

Id: kupiti.01

Opis: *dobiti u vlasništvo plaćanjem, određene cijene*

Argumenti:

Arg0 - onaj koji kupuje
Arg1 - što se kupuje
Arg2 - ono što se daje u zamjenu (novac, cijena)
Arg3 - od koga se kupuje

Predikat: *kupiti*

Id: kupiti.01

Opis: *pribirati ili uklanjati dizanjem s podloge*

Argumenti:

Arg0 - onaj koji skuplja
Arg1 - što se skuplja

Predikat: *misлити*

Id: misлити.01

Opis: *radom mozga oblikovati misao, oblikovati pojmove i sudove; rasuđivati, zaključivati*

Argumenti:

Arg0 - onaj koji misli
Arg2 - (misao) što se misli

Predikat: *nadati se*

Id: nadati.01

Opis: *imati nadu ili nade, gajiti nadu u što, očekivati što povoljno*

Argumenti:

Arg0 - onaj koji se nada
Arg1 - što/čemu se **Arg0** nada

Predikat: *nalaziti*

Id: nalaziti.01

Opis: *naći, pronaći*

Argumenti:

Arg0 - onaj koji nalazi

Arg1 - što je nađeno

Predikat: *nalaziti se*

Id: nalaziti.02

Opis: *biti smješten, zauzimati neki prostor*

Argumenti:

Arg1 - što je smješteno, što se nalazi

Predikat: *nalaziti se*

Id: nalaziti.03

Opis: *sastati se s nekim*

Argumenti:

Arg0 - onaj koji sastaje

Arg1 - s kime se **Arg0** sastaje

Predikat: *napraviti*

Id: napraviti.01

Opis: *izraditi, stvoriti, sastaviti; počiniti, uzrokovati*

Argumenti:

Arg0 - onaj koji izrađuje/stvara/sastavlja/uzrokuje

Arg1 - što se radi/napravilo

Arg2 - od čega, što se promjenilo

Predikat: *naučiti*

Id: naučiti.01

Opis: *steći znanje, vještinu, iskustvo*

Argumenti:

Arg0 - onaj koji je stekao znanje

Arg2 - što **Arg0** zna nakon sticanja znanja (što je **Arg0** naučio)

Predikat: *naučiti*

Id: naučiti.02

Opis: *prenijeti kome svoje znanje, vještinu ili iskustvo; podučiti*

Argumenti:

Arg0 - onaj koji podučava

Arg1 - koga **Arg0** podučava

Arg2 - što **Arg0** podučava **Arg1**

Predikat: *naučiti se*

Id: naučiti.03

Opis: *naviknuti se*

Argumenti:

Arg1 - onaj koji se naučio

Arg2 - (na) što se **Arg0** naučio

Predikat: *nositi*

Id: nositi.01

Opis: *tijelom (u pokretu) ili čim drugim držati što odozdo u svoj težini; mijenjati mjesto, izdržavati kao teret u kretanju; prenositi*

Argumenti:

Arg0 - onaj koji nosi/prenosi

Arg1 - što se nosi

Arg2 - pomagalo čime se nosi (u čemu se **Arg1** nosi)

Predikat: *nositi*

Id: nositi.02

Opis: *pren. voditi nešto*

Argumenti:

Arg0 - onaj koji vodi

Arg1 - što se vodi

Arg2 - kroz što se **Arg1** vodi

Predikat: *nositi*

Id: nositi.03

Opis: *biti odjeven u neki odjevni predmet*

Argumenti:

Arg0 - onaj koji nosi

Arg1 - što se nosi

Predikat: *nositi*

Id: nositi.04

Opis: *imati*

Argumenti:

Arg0 - onaj koji ima nešto

Arg1 - što **Arg0** ima

Predikat: *nuditi*

Id: nuditi.01

Opis: *predstavljati što da se prihvati, odbije ili razmisli o prijedlogu; izražavati spremnost ili iskazivati nakanu*

Argumenti:

Arg0 - entitet koji se predstavlja (tko nudi)

Arg1 - što je ponuđeno

Arg2 - kome je ponuđeno

Arg3 - u zamjenu za što je **Arg1** ponuđeno

Predikat: *objasniti*

Id: objasniti.01

Opis: *učiniti da što postane jasno, razumljivo; protumačiti, razjasniti, poučiti do u detalje, naučiti*

Argumenti:

Arg0 - onaj koji objašnjava

Arg1 - kome se objašnjava

Arg2 - što se objašnjava

Predikat: *odbiti*

Id: odbiti.01

Opis: *odbaciti kome što, ne usvojiti, ne prihvatiti, ne pristati*

Argumenti:

Arg0 - onaj koji odbija

Arg1 - koga/što **Arg0** odbacuje

Predikat: *odbiti*

Id: odbiti.02

Opis: *udarom odvojiti komad od cjeline, odvaliti*

Argumenti:

Arg0 - onaj koji odbija

Arg1 - što **Arg0** odbija

Arg2 - od čega **Arg0** odbija **Arg1**

Predikat: *odbiti se*

Id: odbiti.03

Opis: *odskočiti nakon udara o neku površinu (o predmetu); odbaciti se*

Argumenti:

Arg0 - tko/što se odbilo

Arg1 - od koga/čega se odbilo

Predikat: *odbiti*

Id: odbiti.04

Opis: *oduzeti od iznosa, ne uračunati*

Argumenti:

Arg0 - onaj koji odbija

Arg1 - koga/što **Arg0** odbija

Arg2 - od čega **Arg0** odbija

Arg3 - količina (cijena) koju **Arg0** odbija

Predikat: *omogućiti*

Id: omogućiti.01

Opis: *dati kome sredstva ili ovlasti da što učini*

Argumenti:

Arg0 - onaj koji omogućava

Arg1 - što se omogućava

Arg2 - kome se što omogućava

Predikat: *organizirati*

Id: organizirati.01

Opis: *spojiti/spajati, urediti ili sastaviti (što) u smislenu i funkcionalnu cjelinu, dati/davati čemu koherentan oblik*

Argumenti:

Arg0 - onaj koji organizira

Arg1 - koga/što se organizira

Arg2 - za koga/što se organizira

Predikat: *osigurati*

Id: osigurati.01

Opis: *učiniti sigurnim, zaštititi od opasnosti*

Argumenti:

Arg0 - onaj koji osigurava

Arg1 - tko/što se osigurava

Arg2 - od čega se osigurava

Arg3 - čime se osigurava što

Arg4 - za koga/što se **Arg1** osigurava

Predikat: *osigurati se*

Id: osigurati.02

Opis: *zaštititi se od opasnosti*

Argumenti:

Arg0 - ona koji se osigurava

Arg1 - od koga/čega se osigurava

Arg2 - čime se osigurava

Predikat: *osjećati*

Id: osjećati.01

Opis: *čuvstvovati čutjeti, imati, gajiti osjećaj prema kome*

Argumenti:

Arg0 - onaj koji osjeća

Arg1 - što **Arg0** osjeća

Predikat: *osjećati se*

Id: osjećati.02

Opis: *biti svjestan*

Argumenti:

Arg0 - tko se osjeća

Predikat: *ostati*

Id: ostati.01

Opis: *zadržati neko stanje, naći se u nekom drugom stanju.*

Argumenti:

Arg0 - tko/što ostaje

Arg2 - stanje u kojem ostaje

Predikat: *ostati*

Id: ostati.02

Opis: *zadržati se na jednom mjestu*

Argumenti:

Arg0 - tko/što ostaje

Predikat: *ostaviti*

Id: ostaviti.01

Opis: *odmaknuti od koga, napustiti nekoga*

Argumenti:

Arg0 - onaj koji se odmaknuo

Arg1 - od koga se **Arg0** odmaknuo

Predikat: *ostaviti*

Id: ostaviti.02

Opis: *prepustiti, dati kome što u naslijeđe, dati dojam nekome*

Argumenti:

Arg0 - onaj koji ostavlja

Arg1 - što se ostavlja

Arg2 - kome se ostavlja

Predikat: *ostaviti (se)*

Id: ostaviti.03

Opis: *odložiti što, odustati od nečega*

Argumenti:

Arg0 - onaj koji odložio

Arg1 - ono što je odloženo

Predikat: *otkriti (se)*

Id: otkriti.01

Opis: *pronaći što novo, pokazati nešto novo*

Argumenti:

Arg0 - onaj koji otkriva

Arg1 - tko/što je otkriveno

Arg2 - kome se otkriva

Predikat: *otkriti*

Id: otkriti.02

Opis: *skinuti nešto s koga ili s čega*

Argumenti:

Arg0 - onaj koji otkriva

Arg1 - što se otkriva s **Arg2**

Arg2 - onaj s koga se otkriva

Predikat: *otvoriti*

Id: otvoriti.01

Opis: *dovesti u fokus nešto; započeti nešto*

Argumenti:

Arg0 - onaj koji otvara

Arg1 - što se otvara

Arg2 - čime se otvara (instrument)

Predikat: *očekivati*

Id: očekivati.01

Opis: *vjerovati da će se što dogoditi, biti u iščekivanju čega, smatrati prikladnim da tko što čini*

Argumenti:

Arg0 - onaj koji očekuje

Arg1 - što se očekuje

Arg2 - od koga se očekuje

Predikat: *pisati*

Id: pisati.01

Opis: *bilježiti na papiru ili čemu drugom slova, brojke ili druge znakove; sastavljati i pismom bilježiti književno, znanstveno i sl. djelo; raditi, pridonositi pisanjem teksta; obraćati se pismeno, pismom se javljati, slati pismo*

Argumenti:

Arg0 - onaj koji piše

Arg1 - što se piše

Arg2 - za koga se piše

Arg3 - o kome/čemu se piše

Predikat: *pitati*

Id: pitati.01

Opis: *obraćati se riječima tražeći odgovor, postavljati pitanje ili pitanja*

Argumenti:

Arg0 - onaj koji pita

Arg1 - onaj koga se pita

Arg2 - što se pita (pitanje)

Predikat: *pitati (se)*

Id: pitati.02

Opis: *obavještavati se o komu; zanimati se, interesirati se za koga*

Argumenti:

Arg0 - onaj koji pita

Arg1 - koga se pita

Arg2 - (za) koga/što se pita

Predikat: *pitati*

Id: pitati.03

Opis: *tražiti što, potraživati*

Argumenti:

Arg0 - onaj koji traži

Arg1 - što se traži

Arg2 - od koga se traži

Predikat: *planirati*

Id: planirati.01

Opis: *praviti nacrt, plan, projekt, stvarati, razrađivati plan mjera i utvrđivati vrijeme za izvršenje, ostvarenje kakva zadatka*

Argumenti:

Arg0 - onaj koji planira

Arg1 - što **Arg0** planira

Predikat: *planirati se*

Id: planirati.02

Opis: *namjeravati, smjerati, kaniti*

Argumenti:

Arg1 - što se planira

Arg2 - izvor plana (od čega se planira)

Predikat: *pobijediti*

Id: pobijediti.01

Opis: *nadvladati protivnika, biti pobjednik, odnijeti pobjedu (u ratu, u nadmetanju itd.)*

Argumenti:

Arg0 - pobjednik

Arg1 - gubitnik

Arg2 - u čemu je **Arg0** pobedio **Arg1** (natjecanje, događaj, igra)

Predikat: *pojaviti se*

Id: pojaviti.01

Opis: *postati vidljiv, doći u društvo, doći na neko mjesto*

Argumenti:

Arg1 - onaj koji se pojavio

Predikat: *pokazati*

Id: pokazati.01

Opis: *učiniti da se što vidi, pogleda; obratiti čiju pozornost na koga, na što; učiniti jasnim, očitim, razumljivim; objasniti, dokazati*

Argumenti:

Arg0 - onaj koji pokazuje

Arg1 - kome se pokazuje

Arg2 - što se pokazuje

Predikat: *pokušati*

Id: pokušati.01

Opis: *jednim činom ili jednokratnom radnjom nastojati što učiniti; učiniti napor da se dosegne cilj*

Argumenti:

Arg0 - onaj koji pokušava

Arg1 - što se pokušava

Predikat: *pomoći*

Id: pomoći.01

Opis: *koristiti se, imati pomoć od koga*

Argumenti:

Arg0 - onaj koji pomaže

Arg1 - kome je pomoć pružana

Arg2 - za što (u čemu) je pomoć pružana

Predikat: *ponuditi*

Id: ponuditi.01

Opis: *predstaviti, predložiti komu da što prihvati, odbije ili razmisli; iskazati spremnost ili pokazati namjeru*

Argumenti:

Arg0 - onaj koji predlaže, iskazuje spremnost

Arg1 - što **Arg0** predlaže, iskazuje spremnost

Arg2 - kome **Arg0** predlaže, iskazuje spremnost

Predikat: *postići*

Id: postići.01

Opis: *doći do čega uz trud, dobiti ono za čim se težilo; steći kakav stupanj razvoja, veličine, brzine i sl.*

Argumenti:

Arg0 - onaj koji postiže

Arg1 - što **Arg0** postiže

Arg2 - čime se **Arg1** postiže

Arg3 - za koga je **Arg0** postigao **Arg1**

Predikat: *potpisati*

Id: potpisati.01

Opis: *primiti kakve obveze, uvjete i sl. potvrđujući to svojim potpisom*

Argumenti:

Arg0 - onaj koji prima obveze

Arg1 - što **Arg0** prima

Predikat: *potpisati se*

Id: potpisati.02

Opis: *staviti svoj potpis*

Argumenti:

Arg0 - onaj koji se potpisuje

Arg2 - na što se **Arg0** potpisuje

Predikat: *potvrditi*

Id: potvrditi.01

Opis: *posvjedočiti pravilnost, točnost onoga što je rečeno, učinjeno; ovjeriti; na višoj razini vlasti ili mjerodavnosti prihvatiti odluku niže instance; sankcionirati*

Argumenti:

Arg0 - onaj koji potvrđuje

Arg1 - tko/što se potvrđuje

Arg2 - kome se potvrđuje

Predikat: *potvrditi se*

Id: potvrditi.02

Opis: *pokazati se točnim (o pretpostavci, hipotezi, mišljenju, predviđanju itd.). postati, dokazati se, pokazati se kao što*

Argumenti:

Arg0 - onaj koji se dokazuje

Arg1 - što se potvrđuje

Arg2 - kome se potvrđuje

Predikat: *povećati (se)*

Id: povećati.01

Opis: *učiniti da što postane veće, obilnije, brojnije*

Argumenti:

Arg0 - onaj koji povećava

Arg1 - tko/što se povećava

Arg2 - koliko se **Arg1** povećava

Arg3 - kome je **Arg1** povećani

Predikat: *pozvati*

Id: pozvati.01

Opis: *glasno zazvati; zavikati na koga da dođe; uputiti kome poziv; uglaviti za koga da dođe ili bude nazočan*

Argumenti:

Arg0 - onaj koji poziva

Arg1 - koga se poziva

Arg2 - kamo (na što) se **Arg1** poziva

Predikat: *pratiti*

Id: pratiti.01

Opis: *ići za kim, slijediti koga; uhoditi; ići neko vrijeme uz koga koji odlazi; ići do nekog mjesta s kime koji odlazi, praštajući se*

Argumenti:

Arg0 - pratitelj

Arg1 - onaj koji se prati

Predikat: *pratiti*

Id: pratiti.02

Opis: *pažljivo slušati, razumijevati, shvaćati, čitati o čemu, skupljati obavijesti o čemu*

Argumenti:

Arg0 - onaj koji prati

Arg1 - tko/što se prati

Predikat: *pratiti*

Id: pratiti.03

Opis: *glazb. pjevanjem ili sviranjem pridružiti se vokalnom ili instrumentalnom izvođenju pjesme itd.*

Argumenti:

Arg0 - onaj koji prati

Arg1 - tko/što se prati

Arg2 - u čemu **Arg0** prati **Arg1** (skladba, izvedba, pjesma...)

Predikat: *prihvatiti*

Id: prihvatiti.01

Opis: *uzeti što od onoga koji daje ili predaje, primiti na sebe, preuzeti, ponijeti; uzeti k sebi dajući sklonište, pružajući pomoć; primiti kao svoje, usvojiti, složiti se s čim, suglasiti se; primiti, uzeti što ponuđeno*

Argumenti:

Arg0 - onaj koji prihvaća

Arg1 - koga/što se prihvaća

Predikat: *pripadati*

Id: pripadati.01

Opis: *biti u nečijem vlasništvu; biti dio neke grupe/cjeline*

Argumenti:

Arg1 - onaj koji pripada

Arg2 - komu čemu **Arg1** pripada

Predikat: *priznati*

Id: priznati.01

Opis: *(koga, što) prihvatiti istinu, stvarnost ili postojanje koga ili čega; pol. izraziti prihvaćanje novonastala stanja ili odnosa*

Argumenti:

Arg0 - onaj koji prizna

Arg1 - koga/što se priznaje

Arg2 - kome se priznaje

Predikat: *priznati*

Id: priznati.02

Opis: *(što) preuzeti krivnju, odgovornost za što*

Argumenti:

Arg0 - onaj koji priznaje (krivac)

Arg1 - što **Arg0** priznaje

Arg2 - kome **Arg0** priznaje

Predikat: *priznati*

Id: priznati.03

Opis: *(komu što) uzeti u obzir [priznati studij u inozemstvu]; uvažiti*

Argumenti:

Arg0 - onaj koji uvažava

Arg1 - što **Arg0** priznaje

Arg2 - kome **Arg0** priznaje

Predikat: *promijeniti*

Id: promijeniti.01

Opis: *dati čemu drugi sadržaj, zamijeniti jedno drugim*

Argumenti:

Arg0 - onaj koji mijenja

Arg1 - koga/što **Arg0** mijenja

Arg2 - rezultat mijenjanja, u što se **Arg1** promjenio

Predikat: *promijeniti se*

Id: promijeniti.02

Opis: *postati drugačiji; izmijeniti se*

Argumenti:

Arg1 - onaj koji se mijenja

Arg2 - rezultat mijenjanja, u što se **Arg0** promjenio

Predikat: *pronaći*

Id: pronaći.01

Opis: *stvoriti, izmisliti što novo; iznaći, izumiti; naći (kao rezultat traženja).*

Argumenti:

Arg0 - onaj koji je pronašao

Arg1 - što je pronađeno

Predikat: *pronaći se*

Id: pronaći.02

Opis: *naći smisao u nekom poslu, životnoj situaciji i sl.*

Argumenti:

Arg1 - onaj koji se pronašao

Predikat: *raditi*

Id: raditi.01

Opis: *raditi na nečemu, biti zaposlen*

Argumenti:

Arg0 - radnik

Arg1 - posao koji se obavlja, projekt na kojem se radi

Predikat: *razgovarati*

Id: razgovarati.01

Opis: *riječima razmjenjivati misli, stavove itd., voditi s kim razgovor; govoriti*

Argumenti:

Arg0 - onaj koji razgovara

Arg1 - o kome/čemu je vođen razgovor

Arg2 - onaj s kim **Arg0** razgovara

Predikat: *reagirati*

Id: reagirati.01

Opis: *odgovoriti/odgovarati na poticaj primljen izvana, pokazati/pokazivati reakciju; pružiti/-pružati otpor; suprotstaviti/suprotstavljati se, djelovati u suprotnom smjeru*

Argumenti:

Arg0 - onaj koji reagira

Arg1 - na koga/što **Arg0** reagira

Predikat: *sudjelovati*

Id: sudjelovati.01

Opis: *djelovati zajedno s kime, imati udjela (u poslu itd.), biti sudionik*

Argumenti:

Arg0 - onaj koji sudjeluje

Arg1 - u čemu se sudjeluje

Arg2 - s kime/čime se sudjeluje

Predikat: *tražiti*

Id: tražiti.01

Opis: *truditi se da se pronađe ono što je izgubljeno, zametnuto ili još nepoznato ili ono što nije prisutno*

Argumenti:

Arg0 - onaj koji traži

Arg1 - tko/što se traži

Predikat: *tražiti*

Id: tražiti.02

Opis: *uporno moliti, zahtijevati, iskati*

Argumenti:

Arg0 - onaj koji traži

Arg1 - tko/što se traži

Arg2 - od koga se traži

Predikat: *tvrditi*

Id: tvrditi.01

Opis: *iznositi što kao istinu, kazivati nešto kao potpuno provjereno, kao sigurno*

Argumenti:

Arg0 - tko tvrdi

Arg1 - što se tvrdi

Arg2 - kome se tvrdi

Predikat: *upozoriti*

Id: upozoriti.01

Opis: *(koga na što) skrenuti čiju pozornost, reći mu da je važno*

Argumenti:

Arg0 - onaj koji upozorava

Arg1 - onaj koga se upozorava

Arg2 - na koga/što se **Arg1** upozorava

Predikat: *upozoriti*

Id: upozoriti.02

Opis: *(koga) reći kome da pazi (s obzirom na posljedice); opomenuti*

Argumenti:

Arg0 - onaj koji upozorava

Arg1 - onaj koji je upozoren

Arg2 - na što **Arg0** opominje

Predikat: *uživati*

Id: uživati.01

Opis: *osjećati zadovoljstvo, ugodu, udobnost, užitak*

Argumenti:

Arg0 - onaj koji uživa

Predikat: *uživati*

Id: uživati.02

Opis: *koristiti se čime, imati pravo na što*

Argumenti:

Arg0 - onaj koji uživa

Arg1 - što **Arg0** uživa

Predikat: *vidjeti*

Id: vidjeti.01

Opis: *uvidjeti, uviđati, razab (i)rati, shvatiti, shvaćati*

Argumenti:

Arg0 - onaj koji vidi (shvaća)

Arg1 - ono što **Arg0** shvaća

Predikat: *vidjeti (se)*

Id: vidjeti.02

Opis: *reagirati (na što) osjetilom vida, primjećivati, opažati očima, zapaziti, zapažati očima*

Argumenti:

Arg0 - onaj koji vidi

Arg1 - što se vidi

Predikat: *vjerovati*

Id: vjerovati.01

Opis: *držati, smatrati da je nešto onako kako se prikazuje, kako tko govori; misliti, držati, biti uvjeren*

Argumenti:

Arg0 - onaj koji vjeruje

Arg1 - što **Arg0** vjeruje

Predikat: *vjerovati*

Id: vjerovati.02

Opis: *imati povjerenja (u koga, u što), pouzdati se (u koga); vjerovati na riječ smatrati što istinitim ne tražeći dokaz*

Argumenti:

Arg0 - onaj koji vjeruje

Arg1 - kome **Arg0** vjeruje

Arg2 - što **Arg0** vjeruje **Arg1**

Predikat: *vjerovati*

Id: vjerovati.03

Opis: *biti vjernik, vjerovati u Boga; imati povjerenja (u koga, u što), pouzdati se (u koga)*

Argumenti:

Arg0 - onaj koji vjeruje

Arg1 - u koga/što **Arg0** vjeruje

Predikat: *voljeti*

Id: voljeti.01

Opis: *osjećati ili iskazivati naklonost, privrženost, odanost, prijateljstvo koje se osniva na zajednici ideala, krvnog srodstva i sl. osjećati prema kome ljubav, osjećati strastvenu privlačnost prema kome; biti sklon čemu, rado činiti, slušati, čitati, gledati, služiti se čim, imati sklonost prema zanimanju, poslu, igri*

Argumenti:

Arg0 - onaj koji voli

Arg1 - koga ili što **Arg0** voli

Predikat: *voziti*

Id: voziti.01

Opis: *upravljati prijevoznim sredstvom, kretati se (o vozilu), (koga) prevoziti*

Argumenti:

Arg0 - onaj koji vozi

Arg1 - što **Arg0** vozi prijevozno sredstvo

Arg2 - koga **Arg0** prevozi

Predikat: *voziti se*

Id: voziti.02

Opis: *kretati se u vozilu*

Argumenti:

Arg0 - onaj koji se vozi

Arg1 - u čemu se **Arg0** vozi

Arg2 - po čemu se **Arg0** vozi (cesta, autoput...)

Arg3 - s kime se **Arg0** vozi

Predikat: *zahtijevati*

Id: zahtijevati.01

Opis: *postavljati zahtjeve*

Argumenti:

Arg0 - onaj koji postavlja zahtjeve

Arg1 - zahtjev (što se zahtijeva)

Arg2 - kome **Arg0** postavlja zahtjeve

Predikat: *zaključiti*

Id: zaključiti.01

Opis: *izvesti zaključak, steći uvjerenje*

Argumenti:

Arg0 - onaj koji zaključuje

Arg1 - što se zaključuje

Predikat: *zaraditi*

Id: zaraditi.01

Opis: *ostvariti zaradu; zaslužiti svojim postupcima; razg. uhvatiti, dobiti, steći*

Argumenti:

Arg0 - onaj koji zarađuje

Arg1 - što **Arg0** zarađuje (novac, količina)

Arg2 - od koga **Arg0** zarađuje, na čemu se zarađuje

Arg3 - čime se **Arg1** zarađuje (prodajom, radom, itd.)

Predikat: *završiti*

Id: završiti.01

Opis: *priveći kraju, dovesti do kraja; okončati; biti gotov;*

Argumenti:

Arg0 - onaj tko je završio

Arg1 - što se dovodi do kraja

Arg2 - stanje u kojem je **Arg1** završio

Predikat: *završiti (se)*

Id: završiti.02

Opis: *dospjeti do nekog stupnja ili kraja u napredovanju, ob. kao osrednji ili loš ishod čega*

Argumenti:

Arg1 - onaj tko/što je završio

Arg2 - stanje u kojem je **Arg1** završio

Predikat: *znati*

Id: znati.01

Opis: *biti svjestan neke činjenice, shvatiti nešto; poznavati neki predmet, biti upućen u što; biti vješt u čemu; imati pojam o čemu; razumjeti, shvaćati*

Argumenti:

Arg0 - onaj koji zna

Arg1 - činjenica, koga/što **Arg0** zna