

Skripta za inteligentne sisteme

2015/2016

by Ladybrune

Inteligentni sistemi

- skripta za klasično polaganje ispita, školska 2013/14 (by Stepke) -

NAPOMENA: U nastavku se nalaze odgovori na pitanja za klasično polaganje ispita koji se primenjuju se počev od januarskog roka 2014. godine. Za njihovo rešavanje korišćena je odgovarajuća literatura sa predavanja i vežbi (zimski semestar školske 2013/2014) koju možete naći na zvaničnom sajtu predmeta

EDIT : Dopunjeno za školsku 2015/2016 po pitanjima by Ladybrune

Pitanja iz oblasti EKSPERTNI SISTEMI . (ES)

1. Kako definišemo ES?

Odgovor:

Ekspertni sistem (ES) - je računarski program kojim se emulira rešavanje problema na način na koji to čini *ekspert* (čovek).

2. Navesti i objasniti svrhu osnovna tri modula (dela) ES-a.

3. Koja je svrha baze znanja i šta ona sadrži?

4. Koja je svrha radne memorije i šta ona sadrži?

5. Koja je svrha mehanizma za zaključivanje?

Odgovor na 2,3,4 i 5. pitanje:

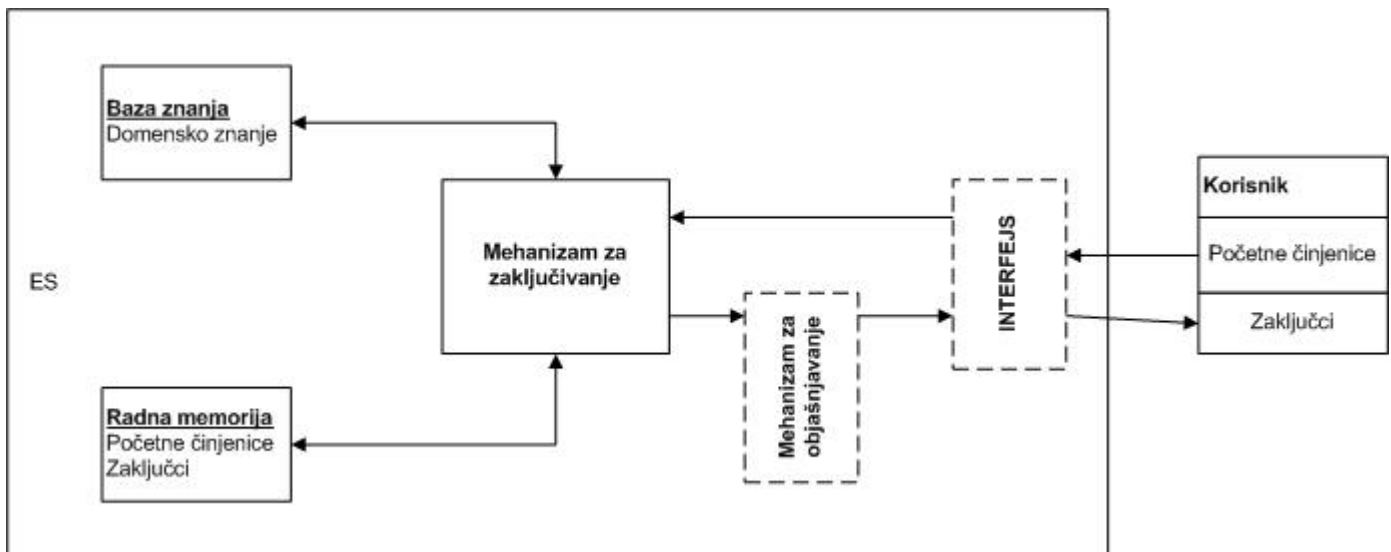
Osnovna tri modula (dela) ES-a su:

- 1) **Baza znanja** - sadrži *domensko znanje* koje MORA biti formalizovano (da bi računar mogao da ga koristi),
- 2) **Radna memorija** - sadrži *činjenice* i *zaključke* (zaključci predstavljaju činjenice nastale kao posledica rezonovanja).
- 3) **Mehanizam za zaključivanje** - kombinuje znanje iz baze znanja i činjenice iz radne memorije i stvara nove zaključke (tako omogućava *automatizovano rezonovanje*).

6. Nacrtati arhitekturu ES.

Odgovor:

Arhitektura ekspertnog sistema data je na sledećoj slici:



7. Koji su osnovni uslovi koje neki program mora da zadovolji da bi mogao da se nazove ES?

Odgovor:

Osnovni uslovi koje neki program mora da zadovolji da bi mogao da se nazove **ES** su da:

1. sadrži **ekspertska znanje** iz neke **oblasti**
2. omogućava **automatizovano rezonovanje**

8. Koja su osnovna dva dela svakog pravila? Navesti i ukratko objasniti.

Odgovor:

Osnovna dva dela svakog pravila su **IF** i **THEN** deo, koji imaju ulogu da povežu **uslov** (*premisu*) sa **zaključkom**, na primer:

IF

Auto neće da "upali" (*uslov - premisa*)

THEN

Kvar može da bude u električnom sistemu (*zaključak*)

Uslov (*premisu*) može da bude i složena, tj. da je čine više jednostavnih premisa povezanih logičkim operatorima **AND**, **OR** i **NOT**.

9. Objasniti šta je ulančavanje pravila i napisati konkretan primer koji sadrži makar tri pravila koja se ulančavaju.

Odgovor:

Ulančavanje pravila se postiže time što **zaključak** jednog pravila predstavlja **uslov** (*premisu*) drugog pravila, na primer:

IF

Auto neće da "upali" AND Napon na akumulatoru < 12V

THEN

Akumulator je prazan (*zaključak*)

IF

Akumulator je prazan (*uslov - premisa*)

THEN

Napuni akumulator

IF

Auto neće da "upali" AND Napon na akumulatoru = 12V

THEN

Anlaser je neispravan (*zaključak*)

IF

Anlaser je neispravan (*uslov - premisa*)

THEN

Zameni anlaser

*** **NAPOMENA:** treće ulančavanje smisliti po želji

10. Od čega zavisi izbor tehnike za zaključivanje?

Odgovor:

Izbor tehnike zaključivanja zavisi od korišćene **tehnike za predstavljanje znanja**.

Najpopularnije tehnike za zaključivanje su:

1. **Ulančavanje unapred** (Forward chaining)
2. **Ulančavanje unazad** (Backward chaining)

i mogu da se koriste isključivo u kombinaciji sa pravilima.

11. Navesti i ukratko objasniti osnovne korake ulančavanja unapred.

Odgovor:

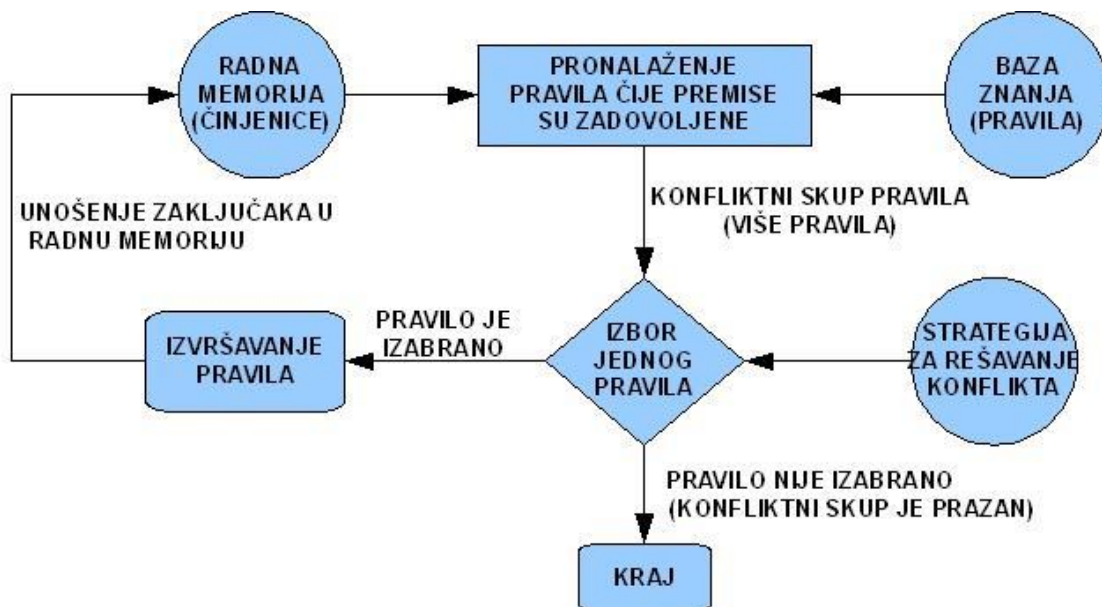
Osnovni koraci **ulančavanja unapred** su:

1. **Korak 1** - pronaći sva pravila čiji **uslovi** (premise) su zadovoljene (ova pravila čine *konfliktni skup*).
2. **Korak 2** - iz *konfliktnog skupa* izabrati samo jedno pravilo (korišćenjem *strategije za rešavanje konflikta*). Ako je *konfliktni skup prazan*, to je *kraj*.
3. **Korak 3** - izvršiti *izabrano pravilo* (uneti zaključke tog pravila kao činjenice u radnu memoriju) i ići na → **Korak 1**.

12. Nacrtati algoritam za ulančavanje unapred.

Odgovor:

Algoritam za ulančavanje unapred dat je na sledećoj slici:



13. Navesti bar tri različite strategije za rešavanje konflikta.

Odgovor:

Strategije za rešavanje konflikta kod ulančavanja unapred su:

1. Izbor prvog pravila
2. Izbor pravila sa najvišim prioritetom
3. Izbor najspecifičnijeg pravila (sa najsloženijim uslovom - premisom)
4. Izbor pravila koje se odnosi na najskorije dodate činjenice
5. Svako pravilo može samo jednom da se izvrši

*** Najčešće se koristi više strategija odjednom.

14. Koja je uloga strategije za rešavanje konflikta u okviru algoritma za ulančavanje unapred?

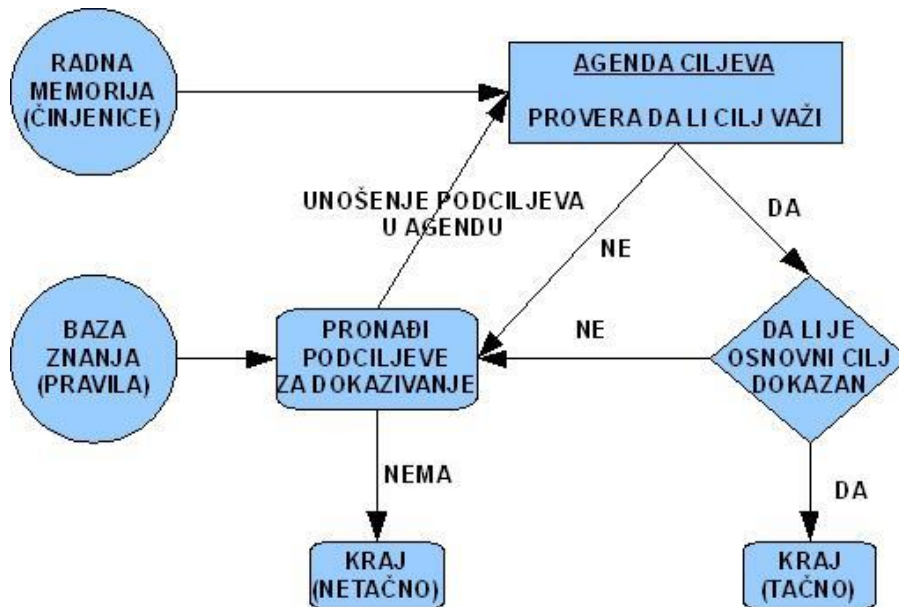
Odgovor:

Uloga **strategije za rešavanje konflikta** u okviru algoritma za **ulančavanje unapred** jeste da se iz *konfliktnog skupa* (sva pravila čiji *uslovi* - *premise* su zadovoljene) izabere samo jedno pravilo.

15. Nacrtati algoritam ulančavanja unazad.

Odgovor:

Algoritam za ulančavanje unazad dat je na sledećoj slici:



16. Koje vrste objašnjenja može da pruži mehanizam za objašnjavanje? Ukratko opisati svaku vrstu objašnjenja.

Odgovor:

Vrste objašnjenja koje mehanizam za objašnjavanje može da pruži su:

1. **ZAŠTO** - objašnjenje o tome zašto ES postavlja određeno pitanje.
2. **KAKO** - objašnjenje o tome kako je ES stigao do rešenja.

17. Navesti bar 4 oblasti primene ES.

Odgovor:

Ekspertni sistemi se primenjuju u sledećim **oblastima**¹:

- 1) Upravljanje industrijskim procesima
- 2) Praćenje rada medicinskih uređaja
- 3) Autonomno kretanje vozila (na zemlji i vodi)
- 4) Automatski piloti
- 5) Upravljanje satelitima
- 6) Nadgledanje instalacija
- 7) Operativno i taktičko upravljanje vojnim operacijama na bojnom polju
- 8) Analize složenih finansijskih transakcija itd.

18. Navesti osnovne uloge u razvoju ES.

19. Koja je uloga eksperta u razvoju ES?

20. Koja je uloga inženjera znanja u razvoju ES?

21. Koja je uloga korisnika u razvoju ES?

Odgovor na 18,19,20 i 21. pitanje:

Osnovne uloge u razvoju ES su:

1. **Ekspert** - koji daje ("pozajmljuje") svoje znanje i pomaže pri proveri (testiranju) znanja

2. **Inženjer znanja** - koji:

- a) Vodi intervju sa ekspertom i iz njega "izvlači" znanje
- b) Vrš izbor:
 - Odgovarajućih tehnika za predstavljanje znanja
 - Odgovarajućih tehnika za zaključivanje
 - Razvojnog alata
- c) Formalizuje, formuliše i "sređuje" ekspertovo znanje
- d) Testira ES

3. **Korisnik** - koji:

- a) Koristi gotov ES
- b) Učestvuje u formiranju zahteva
- c) Može da učestvuje u testiranju i pisanju dokumentacije za ES

22. Nacrtati proces razvoja ES.

Odgovor:



Pitanja iz oblasti MAŠINSKO UČENJE

1. Kako definišemo mašinsko učenje?

Odgovor:

Mašinsko učenje se definiše kao sposobnost softverskog sistema da:

- generalizuje na osnovu prethodnog iskustva (podataka), i
- koristi generalizacije kako bi pružio odgovore na pitanja koja se odnose na podatke koje pre nije sretao

2. U kojim slučajevima (tj. za koje vrste problema) je mašinsko učenje posebno korisno? Ukratko objasniti.

Odgovor:

Mašinsko učenje je posebno korisno za sledeće vrste problema:

- 1) Kod zadataka koje ljudi rešavaju vrlo lako, a pri tome nisu u mogućnosti da precizno (algoritamski) opišu kako to rade. **Primer:** prepoznavanje slika, zvuka, govora
- 2) Kod zadataka gde se mogu definisati algoritmi za rešavanje, ali su ti algoritmi vrlo složeni i/ili zahtevaju velike baze znanja. **Primer:** automatsko prevođenje
- 3) U mnogim oblastima gde se kontinuirano prikupljaju podaci sa ciljem da se iz njih “nešto sazna”.
Primer: u medicini (podaci o pacijentima i korišćenim terapijama), sportu (o odigranim utakmicama i igri pojedinih igrača), marketingu (o korisnicima/kupcima i tome šta su kupili, za šta su se interesovali, kako su proizvode ocenili itd.)

3. Koje su osnovne karakteristike nadgledanog mašinskog učenja?

Odgovor:

Osnovne karakteristike **nadgledanog mašinskog učenja** su da algoritam za učenje dobija:

1. Skup ulaznih podataka (x_1, x_2, \dots, x_n) i
2. Skup željenih/tačnih vrednosti, tako da za svaki ulazni podatak x_i , imamo željeni/tačan izlaz y_i

Zadatak mašine je da “nauči” kako da novom, neobeleženom ulaznom podatku dodeli tačnu izlaznu vrednost.

Izlazna vrednost može biti:

- a) *Labela* (tj. nominalna vrednost) - reč je o klasifikaciji
- b) *Realan broj* - reč je o regresiji

4. Koje su osnovne karakteristike nenadgledanog mašinskog učenja?

Odgovor:

Osnovne karakteristike **nenadgledanog mašinskog učenja** su da:

1. Nemamo informacija o željenoj izlaznoj vrednosti
2. Algoritam dobija samo skup ulaznih podataka (x_1, x_2, \dots, x_n)

Zadatak algoritma je da otkrije paterne tj. skrivene strukture/zakovitosti u podacima.

5. Koje su osnovne karakteristike učenja uz podsticaje?

Odgovor:

Osnovne karakteristike **učenja uz podsticaje** su da:

- 1) Program (agent) deluje na okruženje izvršavanjem niza akcija
- 2) Akcije utiču na stanje okruženja, koje povratno utiče na agenta pružajući mu povratne informacije koje mogu biti “nagrade” ili “kazne”
- 3) Cilj agenta je da nauči kako da deluje u datom okruženju tako da vremenom maximizira nagrade (ili minimizira kazne)

6. Koji su osnovni koraci procesa mašinskog učenja?

Odgovor:

Osnovni koraci procesa **mašinskog učenja** su:

- 1) Prikupljanje podataka potrebnih za formiranje dataset-ova za obuku, validaciju i testiranje algoritama mašinskog učenja
- 2) Priprema podataka, što tipično podrazumeva “čišćenje” i transformaciju podataka
- 3) Analiza rezultujućih dataset-ova, i njihovo, eventualno, dalje unapređenje kroz selekciju/transformaciju atributa
- 4) Izbor jednog ili više algoritama za obuku nad kreiranim trening dataset-om
- 5) Evaluacija izabranih algoritama na dataset-u za validaciju
- 6) Izbor algoritma koji će se koristiti (na osnovu rezultata *koraka* 5) i njegovo testiranje na test dataset-u

1. Kojim faktorima je, generalno posmatrano, uslovljen odabir algoritma mašinskom učenja?

Odgovor:

Odabir algoritma mašinskom učenja je, generalno posmatrano, uslovljen sledećim *faktorima*:

1. **Vrstom problema** koji rešavamo,
2. **Karakteristikama skupa atributa** (features):
 - a) Tipom atributa i stepenom homogenosti tipova atributa
 - b) Stepenom međuzavisnosti (korelisanosti) atributa
3. **Obimom podataka** koji su nam na raspolaganju

8. Koja je uloga validacije u procesu mašinskog učenja?

Odgovor:

Uloga **validacije** u *procesu mašinskog učenja* je da se:

- a) Izabere najbolji model/algoritam između više kandidata
- b) Odredi optimalna konfiguracija parametara modela
- c) Izbegnu problemi over/under-fitting-a

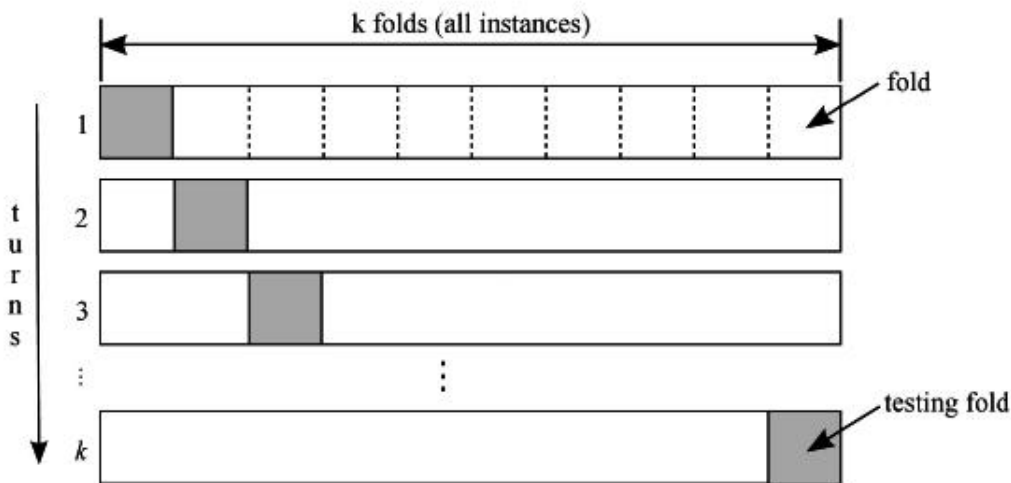
9.Šta je kros-validacija (cross-validation)? Ukratko objasniti kako funkcioniše.

Odgovor:

Kros-validacija (*cross-validation*) je pristup za efikasno korišćenje raspoloživih podataka.

Ona funkcioniše na sledeći način:

1. Raspoloživi skup podataka za trening se podeli na K delova ili podskupova (folds)
 - Najčešće se uzima 10 podskupova (10 fold cross validation)
2. Zatim se obavlja K iteracija (trening + validacija) algoritma, a u svakoj iteraciji:
 - Uzima se 1 deo podataka za potrebe validacije, a ostatak ($K-1$ deo) se koristi za učenje
 - Bira se uvek različiti podskup koji će se koristiti za validaciju



Pri svakoj iteraciji računaju se performanse algoritma.

Na kraju se računa prosečna uspešnost na nivou svih K iteracija - tako izračunate mere uspešnosti daju bolju sliku o performansama algoritma.

Ukoliko su rezultati u svih K iteracija vrlo slični, smatra se da je procena uspešnosti algoritma pouzdana.

Cilj kros validacije je da:

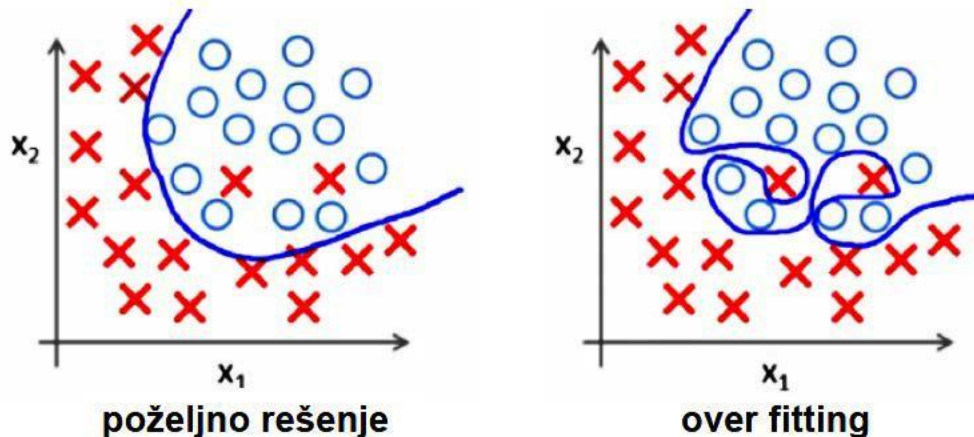
- spreči problem overfitting
- učini predikcije generalnijim

Unapređuje metodu deljenja dataseta procentualno time što smanjuje varijansu među podacima

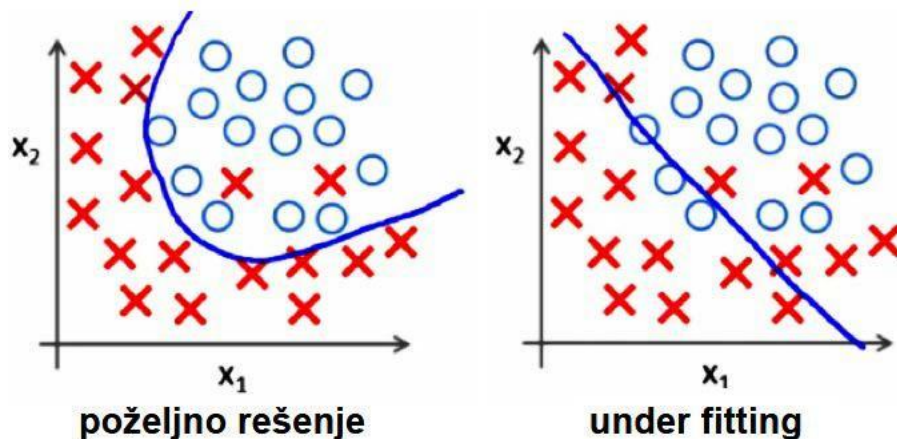
10. Šta označavaju termini *over-fitting* i *under-fitting*? Ukratko objasniti.

Odgovor:

Problem prevelikog podudaranja (*over-fitting*) se odnosi na situaciju u kojoj algoritam savršeno nauči da prepoznaje instance iz trening seta, ali nije u mogućnosti da prepozna instance koje se i malo razlikuju od naučenih:



Problem nedovoljnog podudaranja (*under-fitting*) se odnosi na slučaj kad algoritam ne uspeva da aproksimira podatke za trening, tako da ima slabe performance čak i na trening setu:



11. Šta označavaju termini *bias* i *variance* (u kontekstu modela mašinskog učenja)? Ukratko objasniti.

Odgovor:

Bias (*under-fitting*) i *variance* (*over-fitting*) su dve bitne pojave koje je potrebno razmotriti pri kreiranju modela mašinskog učenja.

Variance ukazuje na to u kojoj meri bi se kreirani model m. učenja promenio ukoliko bi došlo do promene podataka u korišćenom skupu za trening. Generalno, što je metoda m. učenja složenija/fleksibilnija, to će *variance* biti veći.

Bias se odnosi na grešku koja se javlja u slučaju korišćenja vrlo jednostavnog modela za potrebe rešavanja složenog realnog problema. Generalno, što je metoda m. učenja složenija/fleksibilnija, to će *bias* biti manji.

12. Šta je stratifikovana kros-validacija ? Navesti prednosti/nedostatke u odnosu na klasičan pristup kros validaciji.

Odgovor:

Vidi prethodno šta je kros-validacija (pitanje 9)! Stratifikovana kros-validacija obezbeđuje da se u svakom delu (fold) nalazi odgovarajuća proporcija podataka

13. Šta je normalizacija i koja je njena uloga u postupku pripreme podataka za primenu algoritama mašinskog učenja?

Odgovor:

Normalizacija je svođenje vrednosti na neki opseg (obično 0 - 1)

14. Šta je standardizacija i koja je njena uloga u postupku pripreme podataka za primenu algoritama mašinskog učenja?

Odgovor:

Standardizacija je svođenje srednje vrednosti na 0, a standardne devijacije na vrednost 1

15. Šta je diskretizacija i koja je njena uloga u postupku pripreme podataka za primenu algoritama mašinskog učenja?

Odgovor:

Diskretizacija je proces transformacije numeričkih podataka u nominalne tako što se numeričke vrednosti smeštaju u odgovarajuće grupe kojih ima konačan broj.

16. Koji su najčešći pristupi diskretizaciji u postupku pripreme podataka za priču algoritama mašinskog učenja?

Odgovor:

Najčešći pristupi diskretizacije su:

- Nenadgledani pristupi:
 - Jednake širine opsega (Equal-width binning)
 - Jednaka pojavljivanja u opsezima (Equal-frequency binning)
- Nadgledani pristup – uzima u obzir klase

17. Šta je diskretizacija sa jednakim širinama opsega (Equal-width binning) u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?

Odgovor:

Jednake širine opsega (eng. *Equal-width binning*) deli opseg mogućih vrednosti na N podopsega iste širine.

$\text{širina} = (\text{maks. vrednost} - \text{min. vrednost}) / N$

Obično se prvi i poslednji opsezi proširuju kako bi uključili vrednosti van opsega.

18. Šta je diskretizacija sa jednakim pojavljivanjima u opsezima (Equal-frequency binning) u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?

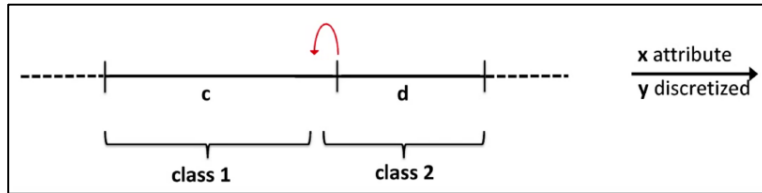
Odgovor:

Jednaka pojavljivanja u opsezima (eng. *Equal-frequency ili equal-height binning*) deli opseg mogućih vrednosti na N podopsega gde svaki podopseg sadrži isti broj instanci.

19. Šta je nadgledana diskretizacija u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?

Odgovor:

Šta ako sve instance u jednom binu pripadaju jednoj klasi, a sve instance drugog bina pripadaju drugoj klasi osim prvog koji pripada prvoj klasi?



Nadgledana diskretizacija uzima u obzir i klasu. Jedan od pristupa je koristiti entropiju. Uzima se razdelnik sa najmanjom entropijom (najvećom informacionom dobiti)

20. Šta je selekcija atributa i kada se primenjuje u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?

Odgovor:

Selekcija atributa (eng. Attribute Selection ili Feature Selection) je proces odabira podskupa relevantnih atributa koji će se koristiti. Primenuje se u slučajevima kada se u datasetu nalaze atributi koji su redundantni ili nerelevantni.

- Redundantni atributi su oni koji ne pružaju nikakve dodatne informacije u odnosu na već selektovane attribute.
- Nerelevantni atributi su oni koji ne pružaju nikakve informacije u datom kontekstu.

21. Koje su prednosti selekcije atributa u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?

Odgovor:

Suvišni atributi mogu degradirati performanse modela. Prednosti selekcije atributa:

- Pобољшава čitljivost modela time što se model sastoji samo iz relevantnih atributa
- Kraće vreme treniranja
- Povećana generalizacija time što smanjuje mogućnosti za overfitting

Najbolji način za selekciju atributa je ručno ukoliko se dobro poznaje problem koji se rešava. I automatizovani pristupi selekcije daju dobre rezultate.

22. Koji su pristupi selekciji atributa u kontekstu pripreme podataka za primenu algoritama mašinskog učenja? Ukratko objasniti sve pristupe

Odgovor:

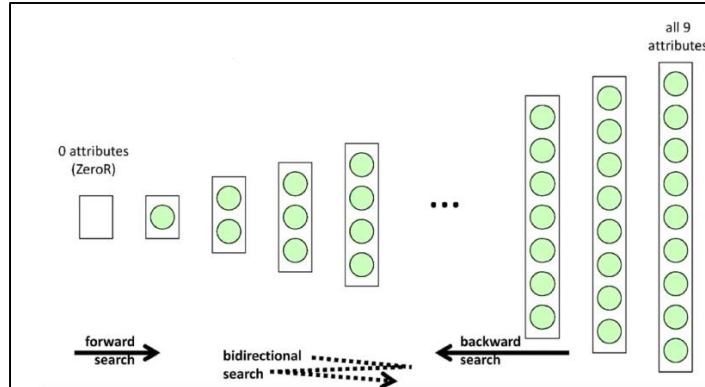
Postoje dva pristupa:

- *Filter* metoda – koriste se procene na osnovu generalnih svojstava podataka
- *Wrapper* metoda – podskupovi atributa se evaluiraju primenom algoritma mašinskog učenja koji će se koristiti nad skupom podataka. Naziv Wrapper se koristi iz razloga što je algoritam učenja “zapakovan” u samom procesu selekcije. Biće odabran onaj podskup atributa za koje dati algoritam učenja daje najbolje rezultate.

23. Navesti metode pretrage kod selekcije atributa u kontekstu pripreme podataka za primenu algoritama mašinskog učenja. Objasniti ukratko svaki metod.

Odgovor:

- Exhaustive search (512 podskupova atributa)
- Best First: *Forward, Backward, Bi-directional*
 - *searchTermination* atribut određuje koliko podskupova koji ne poboljšavaju performance testirati pre nego što prekine pretragu



24. Kako definišemo zadatak klasifikacije?

Odgovor:

Zadatak klasifikacije je određivanje klase kojoj neka instanca pripada (instanca je opisana vrednošću atributa, a skup mogućih klasa je poznat i dat).

25. Navesti mere koje se tipično koriste za procenu uspešnosti modela klasifikacije.

Odgovor:

Mere koje se tipično koriste za procenu uspešnosti modela klasifikacije su:

1. **Matrica zabune** (*Confusion Matrix*)
2. **Tačnost** (*Accuracy*)
3. **Preciznost** (*Precision*) i **Odziv** (*Recall*)
4. **F mera** (*F measure*)
5. **Površina ispod ROC krive** (*Area Under the Curve - AUC*)

26. Kako definišemo i izračunavamo Preciznost (Precision) i Odziv (Recall) kao mere uspešnosti modela klasifikacije?

Odgovor:

Precision = $TP / \text{no. predicted positive} = TP / (TP + FP)$

Npr. od svih poruka koje su *označene kao spam* poruke, koji procenat čine poruke koje su stvarno spam

Recall = $TP / \text{no. actual positive} = TP / (TP + FN)$

Npr. od svih poruka koje su *stvarno spam* poruke, koji procenat poruka je detektovan/klasifikovan kao spam

		Predicted Class		
		Yes	No	
Actual Class	Yes	TP	FN	TP = True Positive FP = False Positive
	No	FP	TN	TN = True Negative FN = False Negative

27. U kakvom su odnosu mere Preciznost (Precision) i Odziv (Recall)? Ukratko objasniti.

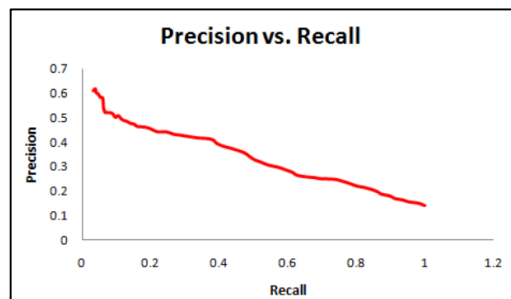
Odgovor:

ako želimo da povećamo Odziv, smanjićemo Preciznost, i obrnuto.

Mere **Preciznost** (*Precision*) i **Odziv** (*Recall*) su u "konfliktu":

- Možemo razviti sistem koji neće praviti mnogo grešaka (visoka preciznost), ali će propustiti da prepozna puno relevantnih informacija (nizak odziv)

- Alternativno, možemo staviti akcenat na odziv i propustiti manje relevantnih informacija, ali po ceni pravljenja više grešaka



28. U kom slučaju se Tačnost (Accuracy) ne može smatrati pouzdanom merom uspešnosti klasifikacionog modela? Ukratko objasniti.

Odgovor:

U slučaju vrlo neravnomerne raspodele podataka između klasa (tzv. skewed classes), ova mera je nepouzdana.

Na primer: U slučaju klasifikacije poruka na spam vs. not-spam, možemo imati skup za trening sa 0.5% spam poruka.

Ako primenimo "klasifikator" koji svaku poruku svrstava u not-spam klasu, dobijamo tačnost od 99.5%.

29. Navesti karakteristike F mere (F measure). U čemu se ogleda razlika između F mere i F1 mere?

Odgovor:

F mera kombinuje **Preciznost** (*Precision*) i **Odziv** (*Recall*) i omogućuje jednostavnije poređenje dva ili više algoritama.

Računa se kao:

$$F = (1 + \beta^2) * Precision * Recall / (\beta^2 * Precision + Recall)$$

* Parametar β kontroliše koliko više značaja će se pridavati *Odzivu* u odnosu na *Preciznost*.

U praksi se najčešće koristi tzv. **F1 mera** („balansirana“ **F mera**) koja daje podjednak značaj i *Preciznosti* i *Odzivu*:

$$F1 = 2 * Precision * Recall / (Precision + Recall)$$

30. Navesti karakteristike mere AUC (Area Under the Curve).

Odgovor:

Karakteristike mere **AUC** (*Area Under the Curve*) su sledeće:

- 1) Meri diskriminacionu moć klasifikatora tj. sposobnost da razlikuje instance koje pripadaju različitim klasama
- 2) Primenjuje se za merenje performansi binarnih klasifikatora
- 3) Vrednost za **AUC** se kreće u intervalu **0-1**
- 4) Za metodu slučajnog izbora važi da je **AUC = 0.5**, a što je **AUC** vrednost klasifikatora **> 0.5**, to je klasifikator bolji
- **0.7-0.8** se smatra prihvatljivim; **0.8-0.9** jako dobrim; sve **> 0.9** je odlično

31. Zadatak

Za dati opis jednog konkretnog zadatka binarne klasifikacije (npr. klasifikacija email poruka na spam i not-spam), interpretirati matricu zabune (confusion matrix), tj. objasniti značenje polja matrice (vrednosti TP, TN, FP i FN) u kontekstu datog konkretnog zadatka. Potrebno je takođe, izračunati vrednosti sledećih mera:

Tačnost (Accuracy), Preciznost (Precision) i Odziv (Recall).

$$Accuracy = (TP + TN) / N$$

gde je:

- TP – True Positive; TN – True Negative
- N – ukupan broj uzoraka (instanci) u skupu podataka

	Predicted Class		
	Yes	No	
Actual Class	Yes	TP FN	TP = True Positive FP = False Positive
	No	FP TN	TN = True Negative FN = False Negative

u TP bi pisali broj *spam* poruka koje su klasifikovane kao *spam*.

u TN bi pisali broj not-spam poruka koje su klasifikovane kao *not-spam*.

u FP bi pisali broj *not-spam* poruka koje su klasifikovane kao *spam*.

u FN bi pisali broj *spam* poruka koje su klasifikovane kao *not-spam*.

Formule za odziv i preciznost su date u pitanju 26. a formulu za tačnost sam ovde gore dopisala.

32. Napisati i objasniti Bayes-ovo pravilo (potrebno je objasniti značenje svih elemenata pravila)

Odgovor:

$$P(H|E) = P(E|H) * P(H) / P(E)$$

- H – hipoteza (*hypothesis*)
- E – opažaj (*evidence*) vezan za hipotezu H , tj. podaci na osnovu kojih bi trebalo da potvrdimo ili odbacimo hipotezu H
- $P(H)$ – verovatnoća hipoteze H (*prior probability*)
- $P(E)$ – verovatnoća opažaja tj. stanja na koje ukazuju prikupljeni podaci
- $P(E|H)$ – (uslovna) verovatnoća opažaja E ukoliko važi hipoteza H
- $P(H|E)$ – (uslovna) verovatnoća hipoteze H ukoliko imamo opažaj E
-

33. Koje pretpostavke uvodi Naive Bayes algoritam (koje ga i čine naivnim)?

Odgovor:

$P(o|c)$ - verovatnoća da u klasi c 'zateknemo' objekat o – nije tako jednostavno odrediti i tu uvodimo pretpostavku koja NB algoritam čine "naivnim"

Pretpostavke koje uvodi Naive Bayes algoritam su:

- objekat o predstavljamo kao skup atributa (features) koji ga opisuju (x_1, x_2, \dots, x_n)
- atributi koji opisuju objekat o su međusobno nezavisni tj. objekat o možemo posmatrati kao prost skup atributa

34. Navesti osobine Naive Bayes algoritma.

Odgovor:

Osobine Naive Bayes algoritma su sledeće:

- 1) Veoma je brz i efikasan
- 2) Najčešće daje dobre rezultate
 - često se pokazuje kao bolji ili bar podjednako dobar kao drugi, sofisticiraniji modeli
- 3) Nije memorijski zahtevan
- 4) Ima vrlo mali afinitet ka preteranom podudaranju sa podacima za trening (overfitting)
- 5) Pogodan kada imamo malu količinu podataka za trening
- 6) "Otporan" na nevažne attribute
 - atributi koji su podjednako distribuirani kroz skup podataka za trening, pa nemaju veći uticaj na izbor klase
- 7) Namenjen primarno za rad sa nominalnim atributima, a u slučaju numeričkih atributa:
 - koristiti raspodelu verovatnoća atributa (tipično Normalna raspodela) za procenu verovatnoće svake od

vrednosti atributa

- uraditi diskretizaciju vrednosti atributa
- N - ukupan broj uzoraka (instanci) u skupu podataka

35. Šta je osnovni (baseline) klasifikator?

Odgovor:

Osnovni klasifikator klasifikuje svaku instancu u klasu sa najvećom verovatnoćom.

36. Objasniti osnovnu ideju na kojoj su zasnovana klasifikaciona stabla

Odgovor:

- Podela prostora atributa kojima su objekti opisani u više različitih i međusobno nepreklopljenih regiona R_1, R_2, \dots, R_n
 - prostor atributa je p -dimenzionalni prostor koga čine moguće vrednosti p atributa (x_1, x_2, \dots, x_p) kojima su dati objekti opisani
- Za novi objekat X , određuje se pripadnost jednom od regiona $R_1 \dots R_n$ na osnovu vrednosti atributa (x_1, x_2, \dots, x_p) kojima je X opisan
- Klasa novog objekta će biti ona klasa koja dominira (*majority class*) u regionu R_j u koji je X svrstan

37. Opisati postupak rekurzivne binarne podele prostora atributa na regione (u kontekstu kreiranja klasifikacionih stabala)

Odgovor:

Osnovne karakteristike ovog pristupa:

- *top-down* pristup: kreće od vrha stabla, gde sve (trening) instance pripadaju jednoj (zajedničkoj) regiji, a zatim sukcesivno deli prostor atributa na regione
 - *greedy* pristup: pri svakom koraku, najbolja podela se određuje na osnovu stanja u tom koraku, odnosno, ne uzima se u obzir šta će biti u narednim koracima, tj koja bi to podela mogla dovesti do boljih rezultata u nekom narednom koraku
- Algoritam razmatra svaki atribut x_j ($j=1, p$) i svaku tačku podele s_j za taj atribut, i bira onu kombinaciju koja će podeliti prostor atributa u dva regiona $\{X/x_j > s_j\}$ i $\{X/x_j < s_j\}$ tako da se minimizuje greška klasifikacije

38. Koji se kriterijumi često koriste za podelu prostora atributa na regione (u kontekstu kreiranja klasifikacionih stabala) ?

Odgovor:

- greška pri klasifikaciji (Classification Error Rate).
- Gini index- često se opisuje kao mera čistoće čvora (node purity). Čisti čvorovi su oni u kojima visok procenat instance pripada istoj klasi. Mala vrednost za Gini indeks ukazuje na čiste čvorove.
- Cross-entropy- predstavlja meru čistoće čvora (što je vrednost manja, to je čvor čistiji)

39. Šta je orezivanje stable odlučivanja (tree pruning)? Zašto se radi i na osnovu čega određujemo u kojoj meri je potrebno orezati stablo?

Odgovor:

Velika klasifikaciona stabla, tj. stabla sa velikim brojem terminalnih čvorova (listova), imaju tendenciju over-fitting-a (tj. prevelikog uklapanja sa trening podacima). Ovaj problem se može rešiti 'orezivanjem' stabla, odnosno odsecanjem nekih terminalnih čvorova.

Kako ćemo znati na koji način i u kojoj meri treba da 'orežemo' stablo? Preporuka je primenom kros validacije (*cross validation*) utvrditi koje podstablo daje najmanju grešku pri klasifikaciji.

40. Koje su prednosti i nedostaci stabala odlučivanja?

Odgovor:

- Prednosti:
 - Mogu se grafički predstaviti i jednostavno interpretirati
 - Mogu se primeniti kako na klasifikacione, tako i regresivne probleme
 - Mogu se primeniti i u slučaju da atributi imaju nedostajuće vrednosti
- Nedostaci:
 - Daju slabije rezultate (manje tačne predikcije) nego drugi pristupi nadgledanog m. učenja

41. Napisati formulu za računanje informacione dobiti u kontekstu stabla odlučivanja (zajedno sa legendom pojmova) I objasniti kako se ona koristi prilikom procesa kreiranja stabla.

Odgovor:

$$Gain(A, S) = H(S) - \sum_{j=1}^v \frac{|S_j|}{|S|} \cdot H(S_j) = H(S) - H(A, S)$$

gde je:

- $H(S)$ – entropija celog skupa instanci S
- $|S_j|$ – broj instanci sa j -tom vrednošću atributa A
- $|S|$ – ukupan broj instanci u skupu S
- v – skup vrednosti atributa A
- $H(S_j)$ – Entropija podskupa instanci sa atributom A
- $H(A, S)$ – Entropija atributa A

Koji atribut odabrati ? **Biramo atribut sa najvećom informacionom dobiti.** Ponavljamo postupak za svaku granu.

42. Šta je ARFF fajl i koje elemente sadrži?

Odgovor:

Attribute-Relation File Format – ARFF . ARFF je Tekstualni fajl. Atributi mogu biti : numerički i nominalni.

```
@relation TPONTPNom

@attribute Outlook {sunny, overcast, rainy}
@attribute Temp. {hot, mild, cool}
@attribute Humidity {high, normal}
@attribute Windy {'false', 'true'}
@attribute Play {no, yes}

@data
sunny, hot, high, 'false', no
sunny, hot, high, 'true', no
overcast, hot, high, 'false', yes
...
```

43. Kako definišemo zadatak klasterizacije?

Odgovor:

Zadatak klasterizacije jeste grupisanja instanci, tako da za svaku instancu važi da je sličnija instancama iz svoje grupe (klastera), nego instancama iz drugih grupa (klastera).

44. Kako se vrši procena sličnosti instanci u postupku klasterizacije?

Odgovor:

Sličnost instanci se procenjuje primenom neke od mera za računanje:

- sličnosti (npr. Kosinusna sličnost ili koeficijent korelacije) ili
- udaljenosti dve instance (npr. Euklidska ili Manhattan udaljenost)

45. Opisati osnovne korake K-Means algoritma.

Odgovor:

Osnovni koraci **K-Means** algoritma su:

1. Inicijalni izbor težišta klastera, slučajnim izborom
 - težišta se biraju iz skupa instanci za trening, tj. **K** instanci za trening se nasumično izabere i proglasi za težišta
2. Ponoviti dok algoritam ne konvergira ili broj iteracija $\leq \text{max}$:
 - 1) *Grupisanje po klasterima*: za svaku instancu iz skupa za trening, $i = 1, m$, identifikovati najbliže težište i dodeliti instance klasteru kome to težište pripada
 - 3) *Pomeranje težišta*: za svaki klaster izračunati novo težište uzimajući prosek tačaka (instanci) koje su dodeljene tom klasteru

46. Koji se kriterijumi tipično koriste za procenu kvaliteta klastera formiranih u procesu klasterizacije?

Odgovor:

Kriterijumi koji se tipično koriste za procenu kvaliteta klastera formiranih u procesu *klasterizacije* su:

1. **Međusobna udaljenost težišta** - što su težišta dalje jedno od drugog, to je stepen preklapanja klastera manji, i njihov kvalitet viši
2. **Standardna devijacija pojedinačnih instanci** u odnosu na *težište* - što je standardna devijacija manja, to su instance tešnje grupisane oko težišta i klasteri se smatraju boljim
3. **Suma kvadrata greške** unutar **klastera** (within cluster sum of squared errors) - daje kvantitativnu meru za procenu kvaliteta kreiranih klastera

47. Kako se prevazilazi problem K-Means algoritma uslovljen nasumičnim (random) inicijalnim izborom težišta? Ukratko objasniti.

Odgovor:

Problem **K-Means** algoritma uslovljen nasumičnim (random) inicijalnim izborom težišta se prevazilazi korišćenjem **višestruke nasumične inicijalizacije**, koja omogućuje da se izbegnu situacije koje **K-Means** dovode u *lokalni minimum*.

Sastoji se u sledećem:

```
for i = 1 to n { // n obično uzima vrednosti od 50 do 1000
    Nasumično odabрати inicijalni skup težišta;
    Izvršiti K-Means algoritam;
    Izračunati funkciju koštanja (cost function)
}
Izabrati instance algoritma koja daje najmanju vrednost za f. koštanja
```

Ovaj pristup daje dobre rezultate ukoliko je broj klastera relativno mali (2 - 10), a za veći broj klastera ne bi ga trebalo koristiti.

48. Koji su pristupi za procenu broja klastera (tj. parametra K) pri primeni K-Means algoritma?

Odgovor:

Pristupi za procenu broja klastera (tj. parametra K) pri primeni **K-Means algoritma** su:

1. U slučaju da **posedujemo znanje o fenomenu/pojavi** koju podaci opisuju:
 - Pretpostaviti broj klastera (K) na osnovu domenskog znanja
 - Testirati model sa $K-1$, K , $K+1$ klastera i uporediti grešku
2. Ukoliko **ne posedujemo znanje o fenomenu/pojavi** :
 - Krenuti od malog broja klastera i u više iteracija testirati model uvek sa jednim klasterom više
 - U svakoj od iteracija, uporediti grešku tekućeg i prethodnog modela i kad smanjenje greške postaje

zanemarljivo, prekinuti postupak

49. Koja je osnovna karakteristika probabilističke klasterizacije?

Odgovor:

instance *ne* pripadaju jednom i samo jednom klasteru, već svaka instanca sa određenom verovatnoćom pripada svakom od klastera

50. Opisati osnovne korake Expectation Maximization (EM) algoritma

Odgovor:

Sastoji se iz dva osnovna koraka:

- **E (expectation) korak** – u ovom koraku podrazumevamo da znamo vrednosti parametara modela i na osnovu njih, za svaku instancu, računamo verovatnoću pripadanja svakom od klastera
- **M (maximization) korak** – na osnovu datih instanci, računamo (ponovo) vrednosti parametara modela; maksimizacija se odnosi na usklađivanje (parametara) modela sa datim podacima Ovi koraci se ponavljaju sve dok algoritam ne počne da konvergira

Ovi koraci se ponavljaju sve dok algoritam ne počne da konvergira

Pitanja iz oblasti NEURONSKE MREŽE

1. Šta je neuronska mreža?

Odgovor:

Neuronska mreža je *paralelni distribuirani procesor* koji ima prirodnu sposobnost čuvanja i korišćenja iskustvenog znanja.

Sličnost sa mozgom se ogleda kroz dve osobine:

- Mreža stiče znanje kroz proces učenja
- Znanje se čuva u vezama između neurona (sinaptičkim težinama)

2. Koje su osnovne komponente veštačkog neurona?

Odgovor:

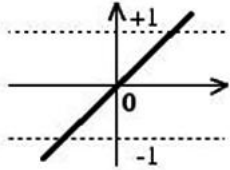
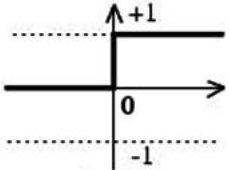
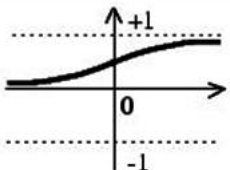
Osnovne komponente **veštačkog neurona** su:

1. **Ulazna funkcija sumiranja**
2. **Funkcija transfera**
3. **Ulazi sa težinskim koeficijentima**
4. **Izlaz**

3. Navesti i nacrtati grafike osnovnih funkcija prenosa koje se koriste u neuronskim mrežama.

Odgovor:

Osnovne funkcije prenosa koje se koriste u **neuronskim mrežama** su:

Linearna	
Odskočna	
Sigmoidna	

4. Navesti osnovne karakteristike neuronskih mreža.

Odgovor:

Osnovne karakteristike **neuronskih mreža** su:

1. Imaju sposobnost učenja
2. Imaju sposobnost generalizacije
3. Otporne na pogrešan ulaz i šum

5. Navesti tipične probleme za koje probleme se koriste neuronske mreže.

Odgovor:

Tipični problemi za koje se koriste **neuronske mreže** su:

- 1) **Klasifikacija**
- 2) **Prepoznavanje** (oblika, govora, vektora...)
- 3) **Aproksimacija**
- 4) **Optimizacija**
- 5) **Obrada signala**
- 6) **Modeliranje sistema**
- 7) **Predviđanje**
- 8) **Kontrola i upravljanje**

6. Navesti tipične slučajeve kada su neuronske mreže kandidat za primenu.

Odgovor:

Tipični slučajevi kada su **neuronske mreže** kandidat za primenu su:

- Kada nema jasno definisanog matematičkog modela ili drugog rešenja
- Kada je potrebna otpornost na nepotpun ili pogrešan ulaz
- Kada je potrebna sposobnost učenja
- Visokodimenzionalnost
- Kada se sa NM postižu bolji rezultati nego sa alternativnim rešenjima (npr. odziv u realnom vremenu, tolerancija

na greške)

7. Objasniti u čemu se sastoji učenje kod neuronskih mreža.

Odgovor:

Učenje kod **neuronskih mreža** se sastoji od podešavanja težina veza tako da mreža dobije željeno ponašanje/funkcionalnost.

8. Koji su osnovni problemi u primeni neuronskih mreža?

Odgovor:

Osnovni problemi u primeni **neuronskih mreža** su:

- Nedostatak semantike u strukturi
- Da li je neki problem uopšte rešiv sa NM
- Problemi sa određivanjem arhitekture i treningom za određenu primenu
- Plastičnost /stabilnost

9. Navesti nekoliko vrsta neuronskih mreža sa prostiranjem signala unapred.

Odgovor:

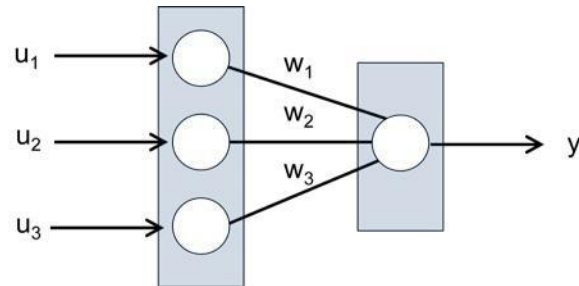
Vrste **neuronskih mreža** sa *prostiranjem signala unapred* mogu biti:

1. **ADALINE** - odlikuju je:

- Linearna funkcija transfera
- Linearna kombinacija ulaza:

$$y = w_1u_1 + w_2u_2 + \dots + w_nu_n$$

- Učenje metodom najmanjih kvadrata



2. **LMS učenje** - može izraziti kroz sledeće jednačine:

- Greška izlaznog neurona za **p**-ti uzorak iz skupa za trening

$$\epsilon_p = d_p - y_p$$

- Promena težine veze proporcionalno grešci

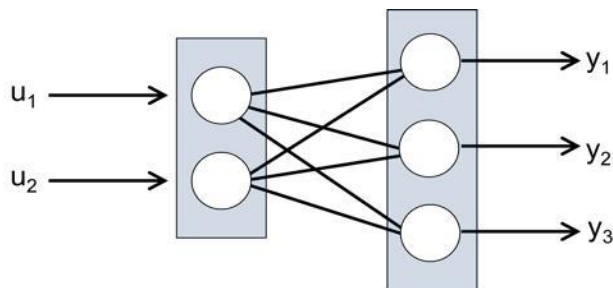
$$w_{ji}(k+1) = w_{ji}(k) + \mu \epsilon(k) u_{ji}(k)$$

- Ukupna greška mreže za sve uzorke iz skupa za trening (kriterijum za zaustavljanje treninga, tj. mreža je naučila kada je greška svedena na prihvatljivu meru)

$$E = \frac{1}{2n} \sum_{p=1}^n \epsilon_p^2$$

3. **PERCEPTRON** - odlikuju je:

- Step funkcija transfera
- *Perceptron learning* - prvi algoritam za učenje nelinearnih sistema
- Samo za linearno separabilne probleme



11. Šta je osnovno ograničenje kod neuronske mreže tipa Perceptron?

Odgovor:

Osnovno ograničenje kod **neuronske mreže** tipa **Perceptron** je to što je ona predviđena samo za linearno separabilne probleme.

10. Koji algoritam se koristi za učenje neuronskih mreža tipa višeslojni Perceptron (Multi Layer Perceptron) ?

Odgovor:

Algoritam koji se koristi za učenje neuronskih mreža tipa **višeslojni Perceptron** (Multi Layer Perceptron) je **Backpropagation algoritam** čije su osobine:

- Služi za *trening višeslojnog perceptrona* - može da podešava težine u skrivenim slojevima
- Predstavlja *supervizorni algoritam* koji se zasniva na LMS algoritmu
- Višeslojni perceptron sa Backpropagation algoritmom predstavlja *univerzalni aproksimator*

12. Opisati proceduru rešavanja problema pomoću neuronskih mreža.

Odgovor:

Procedura *rešavanja problema* pomoću **neuronskih mreža** se sastoji od:

- 1) Prikupljanja i pripreme podataka
- 2) Treninga mreže
- 3) Testiranja mreže
- 4) Određivanja optimalnih parametara mreže i treninga eksperimentalnim putem (broj neurona, broj slojeva neurona, parametri algoritma za učenje, podaci za trening)

Pitanja iz oblasti TEXT MINING i PREPOZNAVANJE ENTITETA u TEKSTU

1. Kako definišemo Text Mining?

Odgovor:

Text Mining je:

- Primena računarskih metoda i tehnika u cilju *ekstrakcije relevantnih informacija* iz teksta
- *Automatsko otkrivanje značenja / znanja* sadržanih u tekstu
- Primena računara za *otkrivanje novih, nepoznatih informacija i znanja*, kroz *automatizovanu ekstrakciju informacija* iz velikog broja različitih *nestrukturiranih* tekstualnih sadržaja

2. U čemu se ogleda složenost nestrukturiranog teksta?

Odgovor:

Nestrukturirani tekstualni sadržaji su opšte prisutni: knjige, finansijski i razni drugi poslovni izveštaji, različita poslovna dokumentacija i prepiska, novinski članci, blogovi, wiki, poruke na društvenim mrežama, ...

Razlozi zbog kojih razumevanje teksta predstavlja vrlo složen problem su ti što je prirodni jezik:

- Pun višesmislenih reči i izraza
- Zasnovan na korišćenju konteksta za definisanje i prenos značenja
- Pun fuzzy, probabilističkih izraza
- Baziran na zdravorazumskom znanju i rezonovanju
- Pod uticajem je i sam utiče na interakcije među ljudima

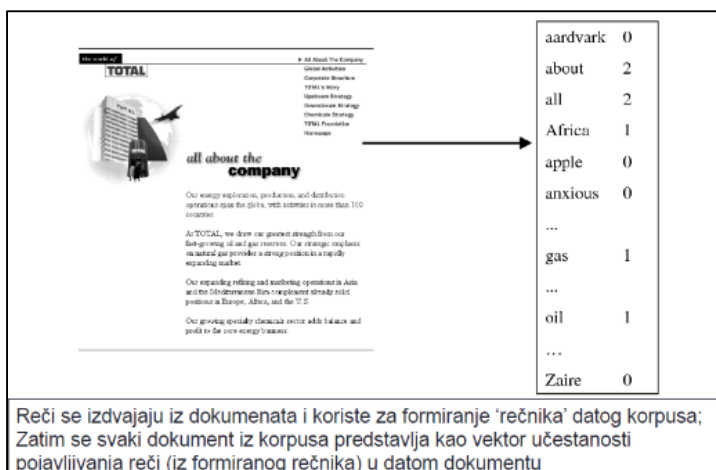
3. Navesti osnovne karakteristike Bag of Words modela .

Odgovor:

Tekst se tretira kao prost skup reči. Pristup zasnovan na sledećim (nerealnim) pretpostavkama:

- reči su međusobno nezavisne,
- redosled reči u tekstu je nebitan

Iako je zasnovan na nerealnim pretpostavkama i vrlo jednostavan, ovaj pristup se pokazao kao vrlo efektan i intenzivno se koristi u TM-u.



4. Navesti osnovne karakteristike Vector Space modela.

Odgovor:

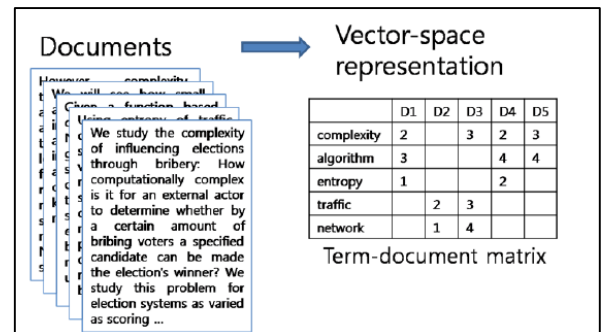
Generalizacija Bag of Words modela. Umesto fokusa isključivo na pojedinačne reči, fokus je na *termine*, pri čemu termin može biti jedna reč ili niz reči. Umesto da se kao mera relevantnosti termina za dati dokument koristi isključivo učestanost pojavljivanja termina u tekstu, koriste se i drugi oblici procene relevantnosti (težine) termina (više o tome kasnije). Ako korpus (*korpus* je kolekcija dokumenata koji su predmet analize) sadrži n termina ($t_i, i=1, n$), dokument d iz tog korpusa biće predstavljen vektorom: $d = \{w_1, w_2, \dots, w_n\}$, gde su w_i težine pridružene terminima t_i . Ovako kreirani vektori predstavljaju osnovu za formiranje matrice termina i dokumenata (*Term Document Matrix*).

5. Šta je to Term Document Matrix ? Ukratko objasniti.

Odgovor:

Term Document Matrix (TDM) je matrice dimenzija $m \times n$ u kojoj:

- Redovi ($i=1, m$) predstavljaju termine iz korpusa
- Kolone ($j=1, n$) predstavljaju dokumente iz korpusa
- Polje ij predstavlja težinu termina i u kontekstu dokumenta j



6. Zašto se i kako radi normalizacija teksta (u kontekstu Text Mining-a) ?

Odgovor:

Normalizacija se vrši da bi transformisali različite oblike jednog istog termina u osnovni, 'normalizovani' format

Npr.: Apple, apple, APPLE -> apple

Intelligent Systems, Intelligent systems, Intelligent-systems -> intelligent systems

Kako se radi:

- Primena jednostavnih pravila:
 - Obrisati sve znake interpunkcije (tačke, crtice, zareze,...)
 - Prebaciti sve reči da budu napisane malim slovima
- Primena rečnika, npr. WordNet, za zamenu sinonima zajedničkom klasom/konceptom. Npr. "automobile, car" -> vehicle

7. Objasniti smisao Zipf-ovog zakona i njegovu primenu u Text Mining-u.

Odgovor:

Zapaženo je da veliki broj reči ima veoma malu frekvencu pojavljivanja a jako mali broj reči se veoma često pojavljuje u tekstu. To je formalizovano Zipf-ovim zakonom: frekvencija bilo koje reči u datom korpusu je obrnuto proporcijalna njenom rang-u u tabeli frekvencija (tog korpusa).

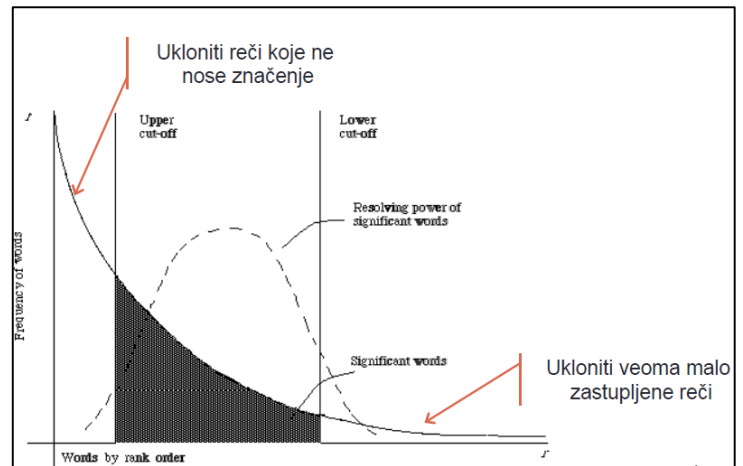
Reči pri vrhu tabele frekventnosti predstavljaju značajan procenat svih reči u korpusu, ali su semantički (gotovo) beznačajne

- Primeri: the, a, an, we, do, to

Reči pri dnu tabele frekventnosti čine najveći deo vokabulara datog korpusa, ali se vrlo retko pojavljuju u dokumentima

- Primer: dextrosinistral

Ostale reči su one koje najbolje reprezentuju korpus i treba ih uključiti u VSM model



8. U čemu se sastoji i koja je svrha lematizacije teksta (lemmatization) ?

Odgovor:

Lematizacija je jedan od pristupa za smanjenje varijabiliteta reči izvučenih iz nekog teksta, kroz svođenje reči na njihov osnovni/ koreni oblik. Koristi morfološki rečnik i primenjuje morfološku analizu reči, kako bi svela reč na njen osnovni oblik (koren reči definisan rečnikom) koji se naziva lema. Npr., argue, argued, argues, arguing -> argue

9. U čemu se sastoji i koja je svrha stemovanja teksta (stemming) ?

Odgovor:

Stemovanje je jedan od pristupa za smanjenje varijabiliteta reči izvučenih iz nekog teksta, kroz svođenje reči na njihov osnovni/ koreni oblik. Koristi heuristiku i statistička pravila za odsecanje krajeva reči (tj poslednjih nekoliko karaktera), gotovo bez razmatranja lingvističkih karakteristika reči. Npr., argue, argued, argues, arguing -> argu

10. Koja je osnovna ideja IDF (Inverse Document Frequency) metrike i kako se izračunava ?

Odgovor:

IDF je jedan od pristupa za dodelu težina terminima u VSM-u. Ideja je da treba dodeliti veće težine neuobičajenim terminima tj. onima koji nisu toliko prisutni u korpusu.

IDF se određuje na osnovu kompletnog korpusa i opisuje korpus kao celinu, a ne pojedinačne dokumente. Izračunava se primenom formule:

$$\text{IDF}(t) = 1 + \log(N/\text{df}(t))$$

N – broj dokumenata u korpusu

df(t) – broj dokumenata koji sadrži termin t

11. Koja je osnovna ideja TF-IDF (Term Frequency - Inverse Document Frequency) metrike i kako se izračunava ?

Odgovor:

TF-IDF je jedan od pristupa za dodelu težina terminima u VSM-u. Ideja je da treba vrednovati one termine koji nisu uobičajeni u korpusu (relativno visok IDF), a pri tome imaju nezanemarljiv broj pojavljivanja u datom dokumentu (relativno visok TF). Izračunava se primenom formule:

$$\text{TF-IDF}(t) = \text{TF}(t) \times \text{IDF}(t)$$

gde je : $\text{TF}(t) = c(t,d)$

$c(t,d)$ - broj pojavljivanja termina t u dokumentu d

$\text{IDF}(t) = 1 + \log(N/df(t))$

N – broj dokumenata u korpusu, $df(t)$ – broj dokumenata koji sadrži termin t

12. Navesti prednosti i nedostatke Vector space Modela.

Odgovor:

Prednosti : Intuitivan, Jednostavan za implementaciju, U studijama i praksi se pokazao kao vrlo efektan

Nedostaci : Nerealna pretpostavka nezavisnosti termina u tekstu, Zahteva puno angažovanja oko usklađivanja (*tuning*) parametara modela:

- izbor metrike za računanje težine termina
- izbor metrike za računanje sličnosti vektora (dokumenata)

13. Navesti osnovne karakteristike tzv. *instance-based* metoda učenja.

Odgovor:

Osnovne karakteristike ove metode su:

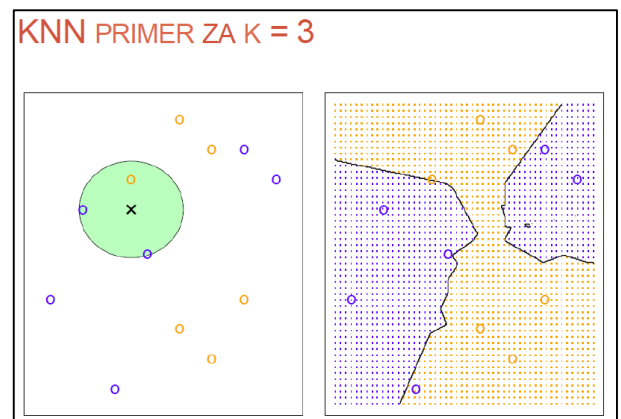
- Sve instance iz skupa za trening se čuvaju u memoriji
- Kada se pojavi nova instanca koju je potrebno klasifikovati
 - Pronalazi se k trening instanci koje su "najbliže" novoj instanci
 - Klasa nove instance se procenjuje na osnovu k najbližih trening instanci

14. Navesti osnovne karakteristike kNN (k Nearest Neighbors) metode mašinskog učenja.

Odgovor:

Osnovne karakteristike kNN metode su :

- Jedna od najbazičnijih *Instance-based* metoda m. učenja
- Posmatra sve instance kao tačke u n -dimenzionalnom prostoru
 - n je broj atributa kojima su instance opisane
- Koristi odgovarajuću metriku za računanje blizine/udaljenosti instanci
- Klasifikuje instancu tako što bira najpopularniju klasu među najbližim susedima te instance



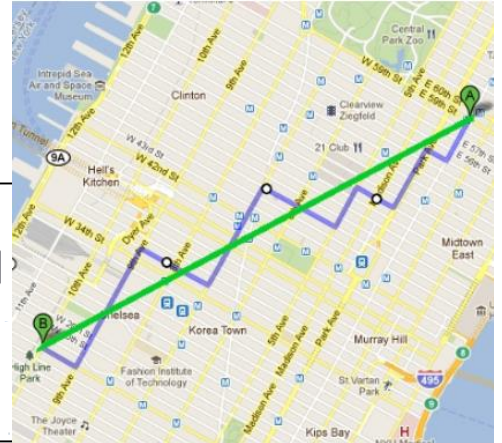
15. Kako se procenjuje blizina instance prilikom primene kNN (k Nearest Neighbors) metode?

Odgovor:

Blizina instance se procenjuje pomoću Euklidske i Manhattan (taxi-cab) distance. One su najpopularnije metrike, ali, zavisno od konkretnog problema klasifikacije, koristite se i druge metrike, izvedene iz ovih osnovnih.

Euklidska distanca	Manhattan (taxi-cab) distanca
$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$ <p>*x_i i x_j su instance čija se udaljenost računa *$a_1(x), a_2(x), \dots, a_n(x)$ je vektor atributa (<i>features</i>) kojim je svaka instanca predstavljena</p>	$d(x_i, x_j) = \sum_{k=1}^n a_{ik} - a_{jk} $ <p>x_i i x_j su instance; $a(x_i)$ i $a(x_j)$ su vektori atributa (<i>features</i>) koji opisuju x_i i x_j</p>

Euklidska vs Manhattan distanca



16. Kako se određuje parameter k prilikom primene kNN (k Nearest Neighbors) metode?

Odgovor:

- Postupkom kros-validacije pronaći K koje garantuje dobre performance klasifikatora, a pri tome ne dovodi do problema
- *over-fitting-a*
- Najčešće se za K bira neparan broj, kako bi se bez dileme mogla izabrati dominantna klasa među K najbližih suseda
- Generalno, što K ima manju vrednost, to je kNN metoda fleksibilnija, odnosno sklonija *over-fitting-u*

17. Zbog čega se koristi kNN algoritam sa težinskim faktorima i koji su najčešći pristupi određivanju težinskih koeficijenata?

Odgovor:

Koristiti se kako bi se uzela u obzir udaljenost suseda od nepoznate instance. Najčešće se svakom susedu dodeljuje težinski faktor $1/d$, gde je d udaljenost suseda od nepoznate instance.

18. Kada koristiti kNN metodu klasifikacije, koje su prednosti, a koji nedostaci?

Odgovor:

Koristi se kada imamo manje od 20 atributa i dosta podataka za trening.

Prednosti su: Treniranje je brzo, Može naučiti kompleksne funkcije, Nema gubitaka informacija

Nedostaci su: Spor odziv, Nerelevantni atributi unose veliku grešku

19. Šta je Cost Sensitive klasifikacija? Objasniti kako može uticati na preciznost i odziv.

Odgovor:

Primenjuje se kada imamo izrazito nebalansiran dataset (eng. skewed dataset) po klasi. Npr. u datasetu sa 10000 instanci sa dve moguće klase, imamo 100 instanci sa prvom klasom, a preostalih 9900 sa drugom klasom.

Cost Sensitive klasifikacija kažnjava FP (false positive) ili FN (false negative) i to utiče na preciznost i odziv. (vidi formule na strani 13 ;).

20. U čemu se sastoji razlika između prepoznavanja imenovanih entiteta u tekstu (*named entity recognition*) i semantičkog indeksiranja (*semantic indexing, entity linking*) ?

Odgovor:

Sistem koji radi prepoznavanje imenovanih entiteta identifikuje pominjanje entiteta i odredi njihov tip, i to je sve. Npr: "Belgrade is the capital of Serbia.". Sistem bi mogao da utvrdi da **Belgrade** i **Serbia** predstavljaju entitete i da odredi da su to City i Country respektivno, ali neće znati na koji konkretan grad se termin *Belgrade* odnosi niti će znati na koju konkretnu državu se termin *Serbia* odnosi.

Sistem koji radi sečantičko indeksiranje ne samo da prepozna pojavljivanje entiteta u tekstu već i jedinstveno identifikuje te entitete povezujući ih sa odgovarajućim konceptima iz baze znanja (npr. Wikipedia i DBpedia). S obzirom da indeksiranje teskta konceptima iz baze znanja obezbeđuje mogućnost interpretiranja semantike (značenja) prepoznatih entiteta, zato se zove semantičko indeksiranje.

21. U čemu se sastoji razlika između semantičkog indeksiranja (*semantic indexing, entity linking*) i identifikacije ključnih tema/konceptata u tekstu ?

Odgovor:

Identifikacija ključnih tema/konceptata je slično semantičkom indeksiranju ali razlike su sledeće:

- Teme/koncepti se ne vezuju za pojedinačne reči i/ili fraze u tekstu, već za ceo tekst
- Nakon identifikacije tema, vrši se njihovo rangiranje po značajnosti

22. Koji se atributi (features) obično koriste (kao sastavni deo modela nadgledanog m. učenja) za prepoznavanje entiteta u tekstu?

Odgovor:

Atributi (*features*) koje se obično koriste (kao sastavni deo modela nadgledanog m. učenja) za **prepoznavanje entiteta u tekstu** su:

1) **Atributi** koji se odnose na **pojedinačne reči**: dužina reči, prisutnost velikih slova, vrsta reči, učestanost pojavljivanja reči u dok. za trening, prisutnost znakova interpunkcije, pozicija reči u rečenici itd.

2) **Atributi** koji se odnose na **okruženje reči**: opseg okruženja, vrsta reči u okruženju i slično

23. Koji su to osnovni oblici procesiranja teksta koji se primenjuju pri kreiranju skupa atributa za prepoznavanje entiteta u tekstu? Ukratko opisati.

Odgovor:

Osnovni oblici procesiranja teksta radi kreiranja skupa atributa:

- Tokenizacija – podela teksta na tokene (= elementarne jedinice sadržaja)
- Eliminisanje stop-words- Stop-words su reči koje (same po sebi) ne nose informaciju
- Part-of-Speech (POS) tagging
 - oblik analize teksta u kome se svakoj reči pridružuje tag koji opisuje ulogu te reči (imenica, zamenica, glagol,...)
 - Primer:

“And now for something completely different”

[('And', 'CC'), ('now', 'RB'), ('for', 'IN'), ('something', 'NN'), ('completely', 'RB'), ('different', 'JJ')]

RB -> Adverb; NN -> Noun, singular or mass; IN -> Preposition, ...

24. Šta je glavna prepreka primeni metoda nadgledanog m. učenja za potrebe identifikacije entiteta u tekstu? Koji su alternativni pristupi i zašto?

Odgovor:

Glavna prepreka primeni **metoda nadgledanog m. učenja** za potrebe *identifikacije entiteta u tekstu* je to što je priprema dovoljno velikog skupa anotiranih dokumenata (korpusa) potrebnog za **training** prilično zahtevan zadatak.

Alternativni pristupi su:

1. **Polu-nadgledano** i
2. **Nenadgledano mašinsko učenje**

jer:

- ne zahtevaju anotirani skup dokumenata
- tradicionalno su imali slabije performanse u odnosu na pristupe nadgledanog m. učenja, ali su nova rešenja sve bolja

25. Šta je Bootstrapping? Ukratko objasniti kako funkcioniše u kontekstu prepoznavanja entiteta u tekstu.

Odgovor:

Bootstrapping je popularna tehnika polu-nadgledanog mašinskog učenja koja podrazumeva mali stepen “nadgledanja”, tipično u formi inicijalno zadatog skupa primera, potrebnog za pokretanje procesa učenja.

On u kontekstu *prepoznavanja entiteta u tekstu* funkcioniše na sledeći način:

- inicijalno, korisnik zadaje mali broj primera tj. naziva različitih entiteta
- sistem kreće sa analizom teksta i pokušava da identifikuje elemente koji karakterišu kontekst zadatih primera, zatim, pokušava da identifikuje druga pojavljivanja entiteta na osnovu identifikovanih karakteristika konteksta
- proces učenja se ponovo primenjuje polazeći od novo-otkrivenih instanci (entiteta), što vodi otkrivanju novih relevantnih konteksta
- ponavljajući ovaj proces, veliki broj naziva entiteta i konteksta u kojima se ona pojavljuju će biti otkriven

26. Kako se baza znanja može iskoristiti za kreiranje trening seta za semantičko indeksiranje? Ukratko objasniti.

Odgovor:

Uzmimo za primer, Wikipedia-u :

Svaki termin koji predstavlja interni link u Wikipedia-i –zvaćemo ga *anchor* – tretira se kao potencijalni entitet. Svaki *anchor* obezbeđuje nekoliko trening instanci:

- Jedan pozitivan primer: destinacija linka (Wikipedia stranica), odnosno “pravo” značenje datog anchor termina u datom kontekstu
- više negativnih primera: sve ostale moguće destinacije linka, odnosno ostala moguća značenja datog anchor termina

27. Koji su osnovni koraci procesa semantičkog indeksiranja u slučaju pristupa koji kombinuju mašinsko učenje i baze znanja?

Odgovor:

Osnovni koraci su:

1) Entity spotting & candidate selection – identifikacija termina koji bi mogli označavati entitete (*entity-mentions*) i selekcija mogućih entiteta iz baze znanja za svaki *entity-mention*

2) Disambiguation – izbor “najboljeg” entiteta za svaki *entity-mention*, tj. izbor entiteta koji najbolje odražava semantiku datog termina u datom kontekstu

3) Filtering – filtriranje rezultata u cilju eliminacije irelevantnih entiteta

28. Proces semantičkog indeksiranja sastoji se iz tri osnovne faze. Ukratko opisati prvu fazu ovog procesa : koji je njen zadatak, i kako se obično realizuje?

Odgovor:

Zadaci prve faze procesa prepoznavanja entiteta su:

- identifikovati tzv. *entity-mentions* u ulaznom tekstu, tj, delove teksta (pojedinačne reči i izraze) koji označavaju entitete;
- identifikovati u bazi znanja (npr., Wikipedia ili DBpedia) skup mogućih entiteta za svaki *entity-mention*

Ova faza se tipično realizuje kao dictionary look-up task

- Formira se rečnik putem ekstrakcije labela i opisa svih entiteta sadržanih u izabranoj bazi znanja
- Wikipedia i DBpedia se najčešće koriste kao baze znanja, odnosno kao izvori iz kojih se ekstrahuju labela i opisi entiteta
- Rečnik može sadržati, za svaki entitet, i različite statistike
 - o npr. relevantnost određene labela za određeni entitet

-----ova dva pitanja koja slede ne postoje u skripti, ali pošto su pitanja 28. i 29. identična na spisku pitanja-----
-----možda su pogrešili pa su mislili na neko od ova dva pitanja :-----

29. Proces semantičkog indeksiranja sastoji se iz tri osnovne faze. Ukratko opisati drugu fazu ovog procesa : koji je njen zadatak, i kako se obično realizuje?

Odgovor:

Cilj ove faze: za svaki *entity-mention*, selektovati jedan ili više entiteta koji mu po svom značenju (semantici) najviše odgovaraju

– selekcija se radi iz, obično povećeg, skupa kandidata identifikovanih u prethodnoj fazi procesa.

Postoji više različitih pristupa za realizaciju ove faze; neki od najčešće primenjivanih:

- *Popularity-based (mention-entity) prior*
- *Context-based approach*
- *Collective disambiguation*

30. Proces semantičkog indeksiranja sastoji se iz tri osnovne faze. Ukratko opisati treću fazu ovog procesa.

Odgovor:

Cilj ove faze je da se iz skupa rezultata uklone oni entiteti koji najverovatnije ne bi bili relevantni korisniku

– npr., entiteti koji se odnose na neke opšte koncepte ili oni koji su samo marginalno povezani sa glavnim temom teksta.

Pitanja iz oblasti PREDSTAVLJANJE ZNANJA U FORMI GRAFA/MREŽE i WEB-a PODATAKA

1. Navesti osnovne karakteristike *DBPedia* baze znanja.

Odgovor:

Vrsta otvorenog, javno dostupnog grafa podataka i znanja. Mašinski čitljiva verzija Wikipedia-e. Podaci preuzeti iz Wikipedia-e su:

- Strukturirani: predstavljeni u formi {subjekat-predikat-objekat} tripleta pogodnih za procesiranje
- Semantički opisani: semantika svakog elementa tripleta je eksplicitno definisana => može se direktno interpretirati od strane računara

2. Navesti osnovne karakteristike *WikiData* baze znanja.

Odgovor:

Wikidata je projekat fondacije Wikimedia : to je slobodna, kolaborativna, multijezička, sekundarna baza podataka, koja prikuplja strukturane podatke da bi pružila podršku za Wikipedia-u, Wikimedia-u Commons, drugim Wikimedia projektima, i još puno toga.

Sekundarna baza podataka: Wikidata može snimiti ne samo stanja već i njihove izvore reflektujući raznovrsnost dostupnog znanja i podržavajući ideju proverenosti.

Prikuplja strukturane podatke: dozvoljava lako ponovno upotrebljavanje podataka od strane Wikimedijinih projekata i drugih, i omogućava računarima lako procesiranje i "razumevanje".

3. Navesti osnovne karakteristike *Knowledge Vault* baze znanja.

Odgovor:

Probabilistička baza znanja koja bi trebalo da sadrži svo činjenično znanje koje Web poseduje, kao i da se uvećava i menja kako se Web razvija i menja. Predstavljanje znanja:

- koristi {subjekat-predikat-objekat} triplete (poput DBpedia-e),
- svaki triplet ima pridruženi *confidence score* koji predstavlja procenjenu verovatnoću da je triplet tačan

Kombinuje:

- automatizovanu ekstrakciju činjenica iz Web stranica (neizvesno, neprovereno znanje)
- znanje preuzeto iz postojećih baza znanja (provereno, validirano znanje)

4. Navesti osnovne karakteristike *NELL (Never Ending Language Learner)* sistema.

Odgovor:

NELL je sistem koji se razvija u okviru Read the Web projekta. Kontinuirano (24/7) izvršava 2 zadatka:

- *Reading task*: identifikuje u Web stranicama nove instance klasa i relacija iz svoje baze znanja, i tako proširuje bazu
- *Learning task*: uči da “bolje čita”, odnosno da vrši precizniju ekstrakciju informacija iz Web stranica
 - modeli mašinskog učenja na kojima je NELL zasnovan kontinuirano se iznova obučavaju koristeći stalno sve veću bazu znanja kao izvor podataka za trening

5. Navesti i kratko obrazložiti osnovne karakteristike tradicionalnog koncepta Web-a (Web dokumenata).

Odgovor:

- 1) Dizajniran za direktno korišćenje od strane ljudi Fokus je na prezentaciji sadržaja na način što pogodniji ljudima, kroz definisanje načina prikaza informacija i postavljanje linkova ka srodnim sadržajima. Semantika sadržaja je dostupna ljudima, ali ne (direktno) i programima...
- 2) Primarni objekti su dokumenti i multi-medija
- 3) Prilično nizak stepen strukturiranosti objekata
- 4) Linkovi između dokumenata (ili njihovih delova)
- 5) Implicitna semantika sadržaja i linkova
- 6) Analogija sa globalnim fajl sistemom

6. Navesti i kratko obrazložiti osnovne karakteristike Web-a podataka (*Web of Data*).

Odgovor:

- 1) Dizajniran za ljude koje ‘opslužuju’ programi
- 2) Primarni objekti su resursi (Resurs je bilo šta što se može jedinstveno identifikovati tj. ima svoj URI. Npr., Beograd je resurs, čiji je URI: <http://dbpedia.org/resource/Belgrade>) i opisi resursa
- 3) Visok stepen strukturiranosti objekata
- 4) Linkovi su između dokumenata i između resursa
- 5) Eksplicitna semantika sadržaja i linkova
- 6) Analogija sa globalnom bazom podataka

7. Šta je URI? Ukratko objasniti.

Odgovor:

URI (*Uniform Resource Identifier*) identifikuje stvari koje opisujemo i omogućava lako povezivanje podataka sa različitim izvorima. Ako se na dva različita mesta kreiraju podaci koristeći *isti URI*, to znači da se govori o istoj stvari.

8. Šta je RDF? Ukratko objasniti.

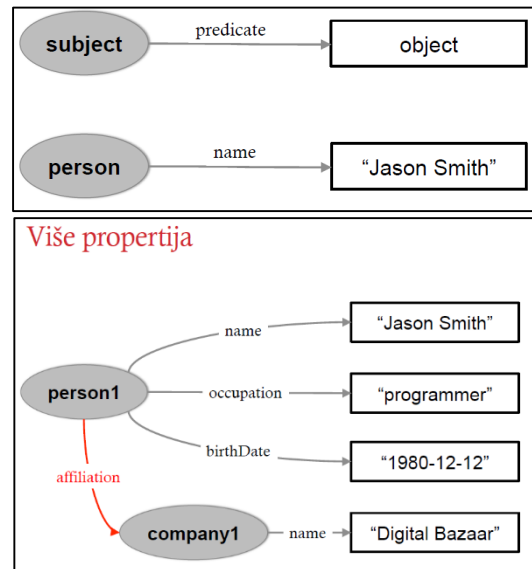
Odgovor:

RDF (Resource Description Framework) je *W3C standard* za opis podataka na Web-u, odnosno jednostavan *model* (zasnovan na *grafu*) koji opisuje relacije između "stvari" (resource). Jedna od tri osnovne tehnologije semantičkog veba (pored SPARQL i OWL). RDF je graf.

RDF je baziran na tripletima (subjekat predikat objekat)

Elementi grafa

- Čvor (predstavljanje subjekta i objekta)
 - Resursi (predstavljaju se elipsom)
 - Literali (predstavljaju se pravougaonikom)
- Veza (predstavljanje predikata)

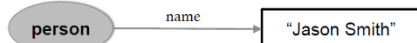


9. Napisati jedan (proizvoljan) RDF triplet i predstaviti ga: 1) grafički; 2) korišćenjem JSON-LD sintakse.

Odgovor:

triplet: person name "Jason Smith"

grafički:



JSON-LD:

```
{
  "http://schema.org/name": "Jason Smith"    !!!--- ovo proveriti nisam sigurna ---!!
}
```

10. Šta je RDFS? Ukratko objasniti.

Odgovor:

RDFS (*RDF Schema*) je šema koja ima ulogu u dodavanju semantike u RDF i kreiranju šeme podataka - vokabulara. Vokabular se definiše na isti način kao i podaci.

11. Navesti po čemu se koncept property-a u RDFS-u razlikuje od koncepta property-a u objektno-orijentisanim jezicima.

Odgovor:

Koncept **property**-a u RDFS-u se razlikuje od koncepta **property**-a u objektno-orijentisanim jezicima po tome što:

- **Propertiji** moгу imati svoju *hijerarhiju*
- Ne mogu se *overwrite*-ovati na nižem nivou hijerarhije

12. Korišćenjem RDFS-a može se definisati domen i opseg bilo kog property-a. Šta predstavlja domen, a šta opseg jednog property-a?

Odgovor:

Domen property-a - pokazuje na *klasu* (ili skup klasa) na koje se *relacija* može primeniti

Opseg property-a - predstavlja *klasu* (ili skup klasa) koje mogu predstavljati *vrednost relacije*

13. Ukoliko za neki property nisu definisani ni domen ni opseg, da li se i kako taj property može koristiti?

Odgovor:

I **domen** i **opseg** su opcioni.

Ukoliko **domen** nije definisan, *relacija* se može primeniti na bilo koju klasu. Ukoliko **opseg** nije definisan, *vrednost relacije* može biti bilo koja klasa.

14. U nastavku je data definicija nekoliko klasa i svojstava (properties) korišćenjem RDF Schema-e.

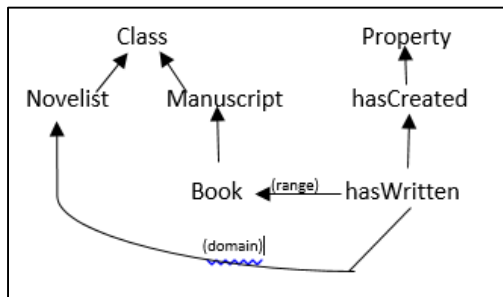
```
PREFIX ex: "http://example.org/vocabulary/"
ex:Novelist rdf:type rdfs:Class .
ex:Manuscript rdf:type rdfs:Class .
ex:Book rdf:type rdfs:Class ; rdfs:subClassOf ex:Manuscript .
ex:hasCreated rdf:type rdf:Property .
ex:hasWritten rdf:type rdf:Property ; rdfs:subPropertyOf ex:hasCreated ;
  rdfs:domain ex:Novelist ; rdfs:range ex:Book .
```

Takođe je dat i sledeći triplet :

```
ex:KenFollet ex:hasWritten ex:FallofGiants .
```

Napisati dodatne triplete koje bi program mogao da izvede iz datog tripleta, primenjujući dati mini RDFS model.

Napomena: na ispitu treba očekivati pitanje ovog tipa, ali ne potpuno identično pitanje. Konkretno, ono što će biti drugačije su klase i svojstva (properties) koje čine RDFS model, kao i dati triplet; međutim, oblici zaključivanja koje treba primeniti (definisani RDFS-om) da bi se izveli dodatni triplet ostaju identični.



Pravila zaključivanja:

Class: Y a Book => Y a Manuscript
 Property: X hasWritten Y => X hasCreated Y
 range: X hasWritten Y => Y a Book
 domain: X hasWritten Y => X a Novelist

iz poslednjeg reda u zadatku: ex:hasWritten rdf:type

Možemo izvesti: Novelist hasWritten Book

iz tripleta zaključujemo ex:KenFollet a ex:Novelist, ex:FallofGiants a ex:Book

Pa imamo i:

ex:KenFollet ex:hasCreated ex:FallofGiants , ex:FallofGiants a ex:Manuscript

15. Koje su osnovne karakteristike umreženih podataka na Web-u i u kom obliku su umreženi podaci prisutni na Web-u?

Odgovor:

Osnovne karakteristike umreženih podataka na Web-u su:

- da imaju jasno definisanu strukturu (= strukturirani podaci), i
- eksplicitno definisano značenje (= semantika podataka je data na način da se može direktno procesirati od strane programa)

Dva osnovna pojavna oblika ovakvih podataka:

- podaci umetnuti u Web stranice
- podaci sadržani u bazama podataka i dostupni na Web-u kroz različite oblike programskih interfejsa (RESTful APIs, query endpoints)

16. Za dodavanje strukturiranih podataka u Web stranice, potrebna su dva ključna elementa; navesti koji su to elementi.

Odgovor:

Za dodavanje podataka u Web stranice, potrebni su nam:

- RDFS vokabulari koji će omogućiti opisivanje sadržaja Web stranica u mašinski razumljivom formatu
- Način da proširimo HTML jezik tako da mašinski razumljivi opisi podataka budu sastavni deo Web stranice

17. Šta je to Schema.org ? Navesti osnovne karakteristike.

Odgovor:

Schema.org je:

- Vokabular za opisivanje podataka u mašinski razumljivom obliku; trenutno, najzastupljeniji vokabular na Web-u
- Inicijativa potekla od velikih Web kompanija: Google, Yahoo, Microsoft (Bing), Yandex
- Dalje se razvija kao community effort u okviru Web konzorcijuma
- Inicijalno omogućavao opis svega nekoliko osnovnih tipova sadržaja, vremenom se taj broj značajno uvećao

18. Šta je to Open Graph Protocol (OGP) ? Navesti osnovne karakteristike.

Odgovor:

Open Graph Protocol je:

- Vokabular koji je uveo Facebook kako bi omogućio prikupljanje dodatnih informacija o sadržajima koje korisnici Like-uju na Web-u
 - OGP vokabular u kombinaciji sa RDFa standardom za proširenje HTML-a, obezbeđuje eksplicitnu semantiku Like-ovanih sadržaja
 - Tako prikupljene informacije Facebook koristi kao input za dalji razvoj svog Entity Graph-a
- OGP omogućuje opis različitih tipova sadržaja popularnih među korisnicima Web-a, kao što su muzika, knjige, video zapisi, profili korisnika i sl

19. Predstaviti bilo grafički ili u JSON-LD sintaksi RDF triplete koji se mogu estrahovati iz sledećeg HTML segmenta (strukturirani podaci su umetnuti korišćenjem RDFa standarda).

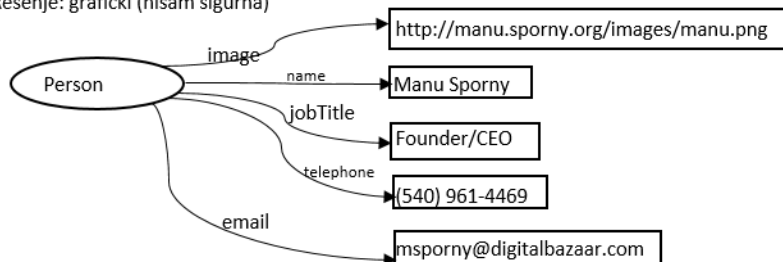
```
<div vocab="http://schema.org/" typeof="Person">
  <a property="image" href="http://manu.sporny.org/images/manu.png">
    <span property="name">Manu Sporny</span>
  </a>,
  <span property="jobTitle">Founder/CEO</span>
</div> Phone: <span property="telephone">(540) 961-4469</span> </div>
<div> E-mail:
  <a property="email"
    href="mailto:msporny@digitalbazaar.com">msporny@digitalbazaar.com
  </a>
</div>
</div>
```

Napomena: ovo je samo primer koda koji može biti dat u okviru ovog pitanja na ispit. Za vežbu možete koristiti primere na sajtu <http://rdfa.info/play/> gde je za svaki primer raspoloživo rešenje u formi grafa; za JSON-LD prikaz možete koristiti <http://rdf-translator.appspot.com/> gde ćete u Input Field tab uneti RDFa kod, zatim izabrati RDFa za Input i JSON-LD za Output, i onda klik na Submit dugme. Primer na ispit će biti vrlo sličan nekom od primera sa <http://rdfa.info/play/> stranice, samo jednostavniji.

Rešenje: JSON-LD

```
{
  "@context": {
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfa": "http://www.w3.org/ns/rdfa#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "schema": "http://schema.org/",
    "xsd": "http://www.w3.org/2001/XMLSchema#"
  },
  "@id": "",
  "rdfa:usesVocabulary": {
    "@id": "schema:"
  },
  "schema:email": {
    "@id": "mailto:msporny@digitalbazaar.com"
  },
  "schema:image": {
    "@id": "http://manu.sporny.org/images/manu.png"
  },
  "schema:jobTitle": "Founder/CEO",
  "schema:name": "Manu Sporny",
  "schema:telephone": "(540) 961-4469"
}
```

Rešenje: grafički (nisam sigurna)



20. Navesti osnovne principe na kojima se zasniva koncept linkovanih podataka (Linked Data).

Odgovor:

Osnovni *principi* na kojima se zasniva koncept **linkovanih podataka** (*Linked Data*) su:

1. Koristiti URI za jedinstvenu identifikaciju entiteta/objekata/pojava/...
2. Koristiti HTTP URI tako da se informacije o entitetima učine dostupnim posredstvom Web-a
3. Opisati entitete korisnim podacima primenom RDF modela. U te svrhe, preporučuje se korišćenje postojećih RDF vokabulara
4. Uspostaviti imenovane linkove ka drugim entitetima/objektima/pojavama...

21. Koje uslove moraju da ispunjavaju otvoreni podaci da bi bili ocenjeni sa 5 zvezdica u okviru Linked Open Data star scheme?

Odgovor:

Uslovi koje moraju da ispunjavaju *otvoreni podaci* da bi bili ocenjeni sa 5 zvezdica u okviru **Linked Open Data star scheme** su ¹ :

1. Omogućiti da podaci budu dostupni na Web-u (u bilo kom formatu)
2. Omogućiti da budu dostupni kao strukturirani podaci
3. Omogućiti da budu u nezaštićenom (*open*) formatu
4. Koristiti *URL*-ove za identifikaciju, kako bi mogli da usmeravaju ka tim podacima
5. Povezati podatke sa podacima ostalih korisnika, u odgovarajućem kontekstu

22. Čemu je namenjen SPARQL ASK upit i kakav tip rezultata vraća?

Odgovor:

SPARQL **ASK** upit je namenjen za proveru da li neki upit uopšte ima rešenje i ne vraća nikakvu informaciju o samom rešenju upita, već samo da li ono postoji. Tip rezultata koji vraća je **boolean** (*true/false*).

23. Čemu je namenjen SPARQL DESCRIBE upit i kakav tip rezultata vraća?

Odgovor:

SPARQL **DESCRIBE** upit vraća graf koji sadrži sve raspoložive triplete o resursu koji je mečiran u okviru *graf paterna* (tj. u *WHERE* delu upita).

24. Čemu je namenjen SPARQL CONSTRUCT upit i kog oblika je rezultat ovog upita?

Odgovor:

SPARQL **CONSTRUCT** upit je namenjen za kreiranje novih RDF grafova na osnovu postojećih tj. za *transformaciju RDF grafova*. Ovaj upit je za RDF graf isto što i XSLT za XML podatke.

⁵¹ "Linked data", http://en.wikipedia.org/wiki/Linked_data

25. Opisati dati SPARQL upit: šta bi trebalo da bude rezultat ovog upita i koji izvori podataka će se koristiti za izvršenje upita? URI dat u upitu (<<http://www.ldodds.com#me>>) je URI osobe tj resursa tipa foaf:Person.

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX yago: <http://dbpedia.org/class/yago/>
SELECT ?person
FROM <http://www.ldodds.com/ldodds-knows.rdf>
WHERE {
  <http://www.ldodds.com#me> foaf:knows ?person .
  ?person foaf:familyName ?surname .
  SERVICE <http://dbpedia.org/sparql> {
    ?someone rdf:type yago:ComputerPioneers ;
    foaf:surname ?surname .
  }
}
```

Napomena: ovo je samo primer upita koji može biti dat u okviru ovog pitanja na ispitu. Upiti na ispitu će biti istog tipa kao dati upit, ali ne identični datom.

Rešenje :

Upit vraća sve osobe koje zna <<http://www.ldodds.com#me>>, čije je prezime isto kao i osobe koja je ComputerPioneer.

Izvori su i uri pored from-a i pored service-a i pošto se yago oslanja na wikipediu i wordnet.

Ključna reč SERVICE služi za definisanje dodatnih izvora podataka.

26. Kako se u domenu Analize društvenih mreža predstavljaju podaci o nekoj društvenoj mreži / grupi / zajednici ?

Odgovor:

Podaci o članovima društvene mreže i njihovim međusobnim relacijama se predstavljaju u formi (*društvenog*) grafa

- svaki član mreže predstavlja jedan čvor grafa
- relacije među članovima mreže su ivice grafa

27. Šta tipično predstavljaju ivice grafa koji se formira za potrebe Analize društvenih mreža? Navesti bar 3 moguća značenja koji se pridružuju ivicama grafa

Odgovor:

Ivice (društvene) mreže tipično odražavaju:

- Neki oblik društvene relacije (npr., prijateljstvo, rodbinske veze, poslovne konekcije, ...), ili
- Neki oblik društvene interakcije (npr., razmena poruka na chat-u, komunikacija putem email-a, ...), ili
- Posedovanje neke zajedničke osobine (npr., pohađanje iste škole, preferenca za isti tip filmova/muzike/knjiga/hrane,...)

28. Navesti nekoliko (minimum 3) načina dodele težina ivicama grafa koji se formira za potrebe Analize društvenih mreža.

Odgovor:

Ivicama se mogu dodati težine koje kvantifikuju intenzitet ili jačinu relacije između entiteta.

Težine mogu biti zasnovane na:

- Učestanosti interakcije (kvantifikovane npr., kroz broj poruka razmenjenih u nekom vremenskom periodu)
- Subjektivnoj oceni stepena povezanosti dve osobe
- Fizičkoj blizini ili udaljenosti dve osobe
- Nekoju kombinaciji prethodnih elemenata
- ...

29. Navesti nekoliko (minimum 3) pitanja na koje Analiza društvenih mreža daje odgovor.

Odgovor:

SNA (*Social Network Analysis*) omogućuje da se za datu društvenu mrežu utvrdi:

- Ko su najuticajniji/centralni članovi mreže?
- Koje se grupe mogu uočiti u mreži? U kojoj meri je mreža podeljena na manje, dobro povezane grupe?
- Kako se mreža razvija? Hoće li se mreža održati?
- Kako se ideje/informacije/virusi/... šire kroz mrežu?
- ...

30. Šta je to *degree centrality* (u kontekstu Analize društvenih mreža)? Navesti osnovne karakteristike ove metrike.

Odgovor:

Degree centrality je ključan indikator povoljne pozicije u mreži. Predstavlja ratio degree metrike datog čvora (tj. broja neposrednih suseda) i ukupnog broja čvorova u mreži. Korisna mera pri proceni/predikciji širenja informacija kroz mrežu. Potencijalni nedostatak: ne uzima u obzir kako su susedi povezani tj. kako su oni pozicionirani u mreži.

31. Šta je to *betweenness centrality* (u kontekstu Analize društvenih mreža)? Navesti osnovne karakteristike ove metrike.

Odgovor:

Betweenness centrality je ključan indikator povoljne pozicije u mreži. Omogućuje identifikaciju čvorova (tj. aktera) koji se često pojavljuju kao posrednici u komunikaciji između drugih čvorova mreže. Takođe, omogućuje identifikaciju tačaka gde može doći do "pucanja" mreže. Iako akteri sa visokom betweenness vrednošću imaju važnu ulogu *brokera* u mreži, često ostaju nezapaženi. Razlog: tipično nisu centralni ni u jednoj grupi u okviru mreže, već se nalaze na periferiji više grupa.

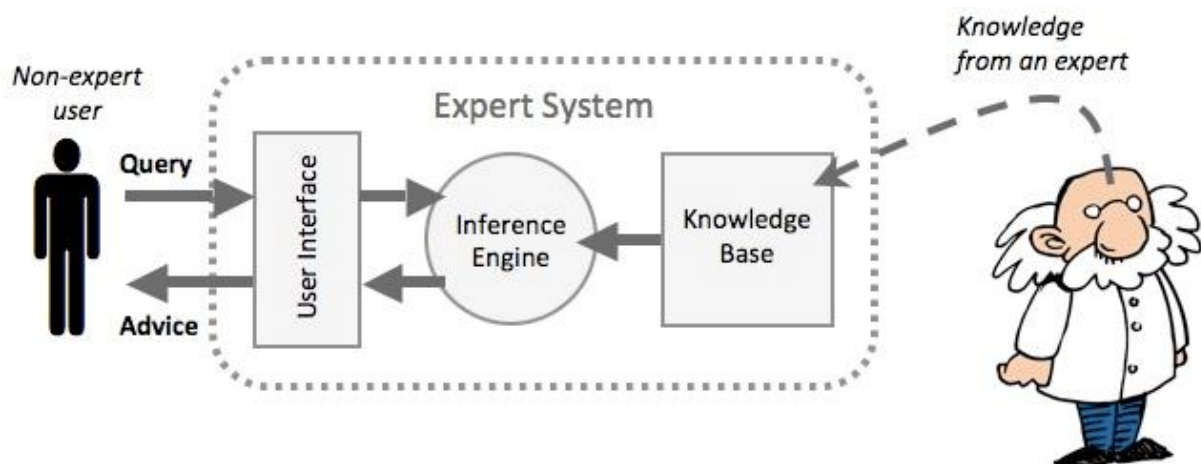
32. Šta je to *closeness centrality* (u kontekstu Analize društvenih mreža)? Navesti osnovne karakteristike ove metrike.

Odgovor:

Closeness centrality je ključan indikator povoljne pozicije u mreži. Ova metrika predstavlja prosečnu blizinu/udaljenost datog aktera od svih ostalih aktera u društvenoj mreži. Visoka vrednost ove metrike je tipična za čvorove koji se nalaze blizu centra lokalnih klastera (grupa) u većoj društvenoj mreži. Još neke karakteristike ove metrike:

- Odražava brzinu kojom informacije, polazeći od datog čvora mogu stići do ostalih čvorova mreže
- Akteri sa visokom vrednošću za closeness centrality
 - obično su uticajni članovi lokalne grupe kojoj pripadaju
 - najčešće su nepoznati ili slabo poznati na nivou mreže kao celine, ali su lokalno uticajni
 - omogućuju brzo širenje informacija kroz deo mreže kome pripadaju

Zadaci iz ekspertnih sistema



Zadaci ES

Zadatak 1

Radna Memorija	Baza znanja
1) Auto neće da “upali” 2) Napon na akumul. = 11V	PRAVILO 1: IF Auto neće da “upali” AND Napon na akumul. < 12V THEN Akumulator je prazan PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akumul. = 12V THEN Anlaser je неисправan PRAVILO 4: IF Anlaser je неисправan THEN Zameni anlaser

Ako je sadržaj radne memorije i baze znanja kao u tabeli i ako se zna da mehanizam za zaključivanje koristi ulančavanje unapred, odgovoriti na sledeća pitanja:

- Do kojih će sve zaključaka doći ES (dopisati u gornju tabelu sve nove činjenice tj. zaključke koji će biti dodati u radnu memoriju)?
- Kojim redosledom će da se izvršavaju pravila?
- Da li će se u nekom trenutku izvršiti PRAVILO 4?

Rešenje

(a) Posle dopune, tabela bi trebalo da izgleda ovako:

Radna Memorija	Baza znanja
1) Auto neće da “upali” 2) Napon na akumul. = 11V 3) <u>Akumulator je prazan</u> 4) <u>Napuni akumulator</u>	PRAVILO 1: IF Auto neće da “upali” AND Napon na akumul. < 12V THEN Akumulator je prazan PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akumul. = 12V THEN Anlaser je неисправan PRAVILO 4: IF Anlaser je неисправan THEN Zameni anlaser

- (b) 1. PRAVILO 1
2. PRAVILO 2

(c) NE

NAPOMENA: nije potrebno pisati objašnjenje na ispitu, već samo rešenje

Objašnjenje

(a) Ovaj primer je urađen korak po korak u okviru vežbi “ES - osnove”. Uz objašnjenje pratiti i

odgovarajuće slajdove (slajdovi 19-27).

U prvom ciklusu u konfliktni skup pravila može da uđe samo PRAVILO 1 tj. jedino je premisa ovog pravila zadovoljena - "Auto neće da "upali" i "Napon na akumul. < 12V" ("Napon na akumul. = 11V"). Ovo je jedino pravilo u konfliktnom skupu, pa će se ono izvršiti. Kada se izvrši, u radnu memoriju se dodaje THEN deo ovog pravila u vidu nove činjenice - "Akumulator je prazan".

U drugom ciklusu u konfliktni skup pravila može da uđe samo PRAVILO 2 - jedino je premisa ovog pravila zadovoljena "Akumulator je prazan". PRAVILO 1 se već izvršilo pa ne može ponovo da uđe u konfliktni skup i ne razmatra se. PRAVILO 2 je jedino pravilo u konfliktnom skupu, pa će se ono izvršiti. Kada se izvrši, u radnu memoriju se dodaje THEN deo ovog pravila u vidu nove činjenice - "Napuni akumulator".

U trećem ciklusu je konfliktni skup prazan - ni PRAVILO 3 ni PRAVILO 4 ne mogu da uđu u ovaj skup jer im premise nisu zadovoljene. Sa obzirom na to da je konfliktni skup prazan zaključivanje se završava.

(b) Na osnovu prethodnog, može se zaključiti da će se prvo izvršiti PRAVILO 1, pa onda PRAVILO 2 pa je odgovor:

1. PRAVILO 1
2. PRAVILO 2

(c) Na osnovu prethodnog, može se zaključiti da neće izvršiti ni PRAVILO 3 ni PRAVILO 4 pa je odgovor: NE.

Zadatak 2

Radna Memorija	Baza znanja
1) Auto neće da "upali" 2) Napon na akumul. = 12V	PRAVILO 1: IF Auto neće da "upali" AND Napon na akumul. < 12V THEN Akumulator je prazan PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da "upali" AND Napon na akumul. = 12V THEN Anlaser je neispravan PRAVILO 4: IF Anlaser je neispravan THEN Zameni anlaser

Ako je sadržaj radne memorije i baze znanja kao u tabeli i ako se zna da mehanizam za zaključivanje koristi ulančavanje unapred, odgovoriti na sledeća pitanja:

- (a) Do kojih će sve zaključaka doći ES (dopisati u gornju tabelu sve nove činjenice tj. zaključke koji će biti dodati u radnu memoriju)?
- (b) Kojim redosledom će da se izvršavaju pravila?
- (c) Da li će se u nekom trenutku izvršiti PRAVILO 4?

Rešenje

(a) Posle dopune, tabela bi trebalo da izgleda ovako:

Radna Memorija	Baza znanja
1) Auto neće da “upali” 2) Napon na akum. = 12V 3) <u>Anlaser je неисправan</u> 4) <u>Zameni anlaser</u>	PRAVILO 1: IF Auto neće da “upali” AND Napon na akum. < 12V THEN Akumulator je prazan PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akum. = 12V THEN Anlaser je неисправan PRAVILO 4: IF Anlaser je неисправan THEN Zameni anlaser

- (b) 1. PRAVILO 3
2. PRAVILO 4

(c) DA

Objašnjenje

(a) U prvom ciklusu u konfliktni skup pravila može da uđe samo PRAVILO 3 tj. jedino je premisa ovog pravila zadovoljena - “Auto neće da “upali” i “Napon na akum. = 12V”. Ovo je jedino pravilo u konfliktnom skupu, pa će se ono izvršiti. Kada se izvrši, u radnu memoriju se dodaje THEN deo ovog pravila u vidu nove činjenice - “Anlaser je неисправan”.

U drugom ciklusu u konfliktni skup pravila može da uđe samo PRAVILO 4 - jedino je premisa ovog pravila zadovoljena “Anlaser je неисправan”. PRAVILO 3 se već izvršilo pa ne može ponovo da uđe u konfliktni skup i ne razmatra se. PRAVILO 4 je jedino pravilo u konfliktnom skupu, pa će se ono izvršiti. Kada se izvrši, u radnu memoriju se dodaje THEN deo ovog pravila u vidu nove činjenice - “Zameni anlaser”.

U trećem ciklusu je konfliktni skup prazan - ni PRAVILO 1 ni PRAVILO 2 ne mogu da uđu u ovaj skup jer im premise nisu zadovoljene. Sa obzirom na to da je konfliktni skup prazan zaključivanje se završava.

(b) Na osnovu prethodnog, može se zaključiti da će se prvo izvršiti PRAVILO 3, pa onda PRAVILO 4 pa je odgovor:

1. PRAVILO 3
2. PRAVILO 4

(c) Na osnovu prethodnog, može se zaključiti da će se izvršiti PRAVILO 3 i PRAVILO 4 pa je odgovor: DA.

Zadatak 3

Radna Memorija	Baza znanja
1) Auto neće da “upali”	PRAVILO 1: IF Auto neće da “upali” AND Napon na akumul. < 12V THEN Akumulator je prazan PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akumul. = 12V THEN Anlaser je neispravan PRAVILO 4: IF Anlaser je neispravan THEN Zameni anlaser

Ako je sadržaj radne memorije i baze znanja kao u tabeli i ako se zna da mehanizam za zaključivanje koristi ulančavanje unapred, odgovoriti na sledeća pitanja:

- Do kojih će sve zaključaka doći ES (dopisati u gornju tabelu sve nove činjenice tj. zaključke koji će biti dodati u radnu memoriju)?
- Kojim redosledom će da se izvršavaju pravila?
- Da li će se u nekom trenutku izvršiti PRAVILO 2?

Rešenje

(a) Posle dopune, tabela bi trebalo da izgleda ovako (izgledaće potpuno isto):

Radna Memorija	Baza znanja
1) Auto neće da “upali”	PRAVILO 1: IF Auto neće da “upali” AND Napon na akumul. < 12V THEN Akumulator je prazan PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akumul. = 12V THEN Anlaser je neispravan PRAVILO 4: IF Anlaser je neispravan THEN Zameni anlaser

(b) Nijedno pravilo neće da se izvrši.

(c) NE

Objašnjenje

(a) U prvom ciklusu je konfliktni skup prazan - nijedno pravilo ne može da uđe u ovaj skup zato što nijedna premisa nije zadovoljena. Činjenica “Auto neće da upali” ne zadovoljava nijednu premisu.

Sa obzirom na to da je konfliktni skup prazan zaključivanje se završava i ES ne donosi nijedan zaključak. Sadržaj radne memorije se neće promeniti (tabela ostaje ista).

(b) Na osnovu prethodnog, može se zaključiti da se neće izvršiti nijedno pravilo pa ne postoji ni redosled izvršavanja pravila.

(c) Na osnovu prethodnog, može se zaključiti da se neće izvršiti PRAVILO 2 pa je odgovor: NE.

Zadatak 4

Radna Memorija	Baza znanja
1) Auto neće da “upali” 2) Napon na akum. = 11V	PRAVILO 1: IF Auto neće da “upali” AND Napon na akum. < 12V THEN Akumulator je prazan
Agenda ciljeva – Napuni akumulator	PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akum. = 12V THEN Anlaser je neispravan PRAVILO 4: IF Anlaser je neispravan THEN Zameni anlaser

Ako je sadržaj radne memorije i baze znanja kao u tabeli i ako se zna da mehanizam za zaključivanje koristi ulančavanje unazad, odgovoriti na sledeća pitanja:

- Kako će na kraju ulančavanja unazad da izgleda agendu ciljeva (dopisati u gornju tabelu sve nove ciljeve koji će biti dodati u agendu ciljeva)?
- Da li može da se dokaže osnovni cilj “Napuni akumulator”?
- Da li će se u nekom trenutku u razmatranje uzeti PRAVILO 2?

Rešenje

(a) Posle dopune, tabela će da izgleda ovako:

Radna Memorija	Baza znanja
1) Auto neće da “upali” 2) Napon na akum. = 11V	PRAVILO 1: IF Auto neće da “upali” AND Napon na akum. < 12V THEN Akumulator je prazan
Agenda ciljeva – Napuni akumulator – Akumulator je prazan – Auto neće da “upali” – Napon na akum. < 12V	PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akum. = 12V THEN Anlaser je neispravan PRAVILO 4: IF Anlaser je neispravan THEN Zameni anlaser

(b) DA

(c) DA

Objašnjenje

(a) Ovaj primer je urađen korak po korak u okviru vežbi “ES - osnove”. Uz objašnjenje pratiti i odgovarajuće slajdove (slajdovi 31-39).

U prvom ciklusu se proverava da li se možda osnovni cilj “Napuni akumulator” već nalazi u radnoj memoriji u vidu činjenice. Kada bi se nalazio, cilj bi bio dokazan i zaključivane bi se prekinulo. Pošto se ne nalazi (nedokazan je), traži se pravilo iz baze znanja koje ima ovaj cilj u svom THEN delu. To je PRAVILO 2. To pravilo implicira da je potrebno dokazati da je akumulator prazan da bi zaključak bio da je potrebno napuniti akumulator. Prema tome, “Akumulator je prazan” postaje novi cilj tj. dodaje se u agendu ciljeva.

U drugom ciklusu se proverava da li se možda cilj “Akumulator je prazan” već nalazi u radnoj memoriji u vidu činjenice. Pošto se ne nalazi, traži se pravilo iz baze znanja koje ima ovaj cilj u svom THEN delu. To je PRAVILO 1. To pravilo implicira da je potrebno dokazati da auto neće da upali i da je napon na akumulatoru manji od 12 volti da bi zaključak bio da je akumulator prazan. Prema tome, “Auto neće da upali” i “Napon na akumul. < 12V” postaju novi ciljevi (dodaju se u agendu ciljeva).

U trećem ciklusu se proverava da li važe ciljevi “Auto neće da upali” i “Napon na akumul. < 12V”. Činjenice iz radne memorije potvrđuju da važe ovi ciljevi – auto zaista neće da upali a napon je 11 volti. Samim tim, ova dva cilja su dokazana. Pošto ova dva cilja važe, važiće i sledeći, njima nadređeni cilj “Akumulator je prazan”. Pošto važi i ovaj cilj, sledi da važi i osnovni cilj - “Napuni akumulator”. Osnovni cilj je dokazan, pa se zaključivanje završava.

(b) Na osnovu prethodnog, može se zaključiti da se može dokazati osnovni cilj jer se mogu dokazati i svi podciljevi iz agende, pa je odgovor DA.

(c) Na osnovu prethodnog, može se zaključiti da će se PRAVILO 2 uzeti u razmatranje i to u prvom koraku tako da je odgovor DA.

Zadatak 5

Radna Memorija	Baza znanja
1) Auto neće da “upali” 2) Napon na akumul. = 12V	PRAVILO 1: IF Auto neće da “upali” AND Napon na akumul. < 12V THEN Akumulator je prazan
Agenda ciljeva – Napuni akumulator	PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akumul. = 12V THEN Anlaser je neispravan PRAVILO 4: IF Anlaser je neispravan THEN Zameni anlaser

Ako je sadržaj radne memorije i baze znanja kao u tabeli i ako se zna da mehanizam za zaključivanje koristi ulančavanje unazad, odgovoriti na sledeća pitanja:

- Kako će na kraju ulančavanja unazad da izgleda agenda ciljeva (dopisati u gornju tabelu sve nove ciljeve koji će biti dodati u agendu ciljeva)?
- Da li može da se dokaže osnovni cilj “Napuni akumulator”?
- Da li će se u nekom trenutku u razmatranje uzeti PRAVILO 3?

Rešenje

(a) Posle dopune, tabela će da izgleda ovako:

Radna Memorija	Baza znanja
1) Auto neće da “upali” 2) Napon na akumul. = 11V	PRAVILO 1: IF Auto neće da “upali” AND Napon na akumul. < 12V THEN Akumulator je prazan
Agenda ciljeva – Napuni akumulator – Akumulator je prazan – Auto neće da “upali” – Napon na akumul. < 12V	PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akumul. = 12V THEN Anlaser je neispravan PRAVILO 4: IF Anlaser je neispravan THEN Zameni anlaser

(b) NE

(c) NE

Objašnjenje

(a) U prvom ciklusu se proverava da li se možda osnovni cilj “Napuni akumulator” već nalazi u radnoj memoriji u vidu činjenice. Kada bi se nalazio, cilj bi bio dokazan i zaključivane bi se prekinulo. Pošto se ne nalazi (nedokazan je), traži se pravilo iz baze znanja koje ima ovaj cilj u svom THEN delu. To je PRAVILO 2. To pravilo implicira da je potrebno dokazati da je akumulator prazan da bi zaključak bio da je potrebno napuniti akumulator. Prema tome, “Akumulator je prazan” postaje novi cilj tj. dodaje se u agendu ciljeva.

U drugom ciklusu se proverava da li se možda cilj “Akumulator je prazan” već nalazi u radnoj memoriji u vidu činjenice. Pošto se ne nalazi, traži se pravilo iz baze znanja koje ima ovaj cilj u svom THEN delu. To je PRAVILO 1. To pravilo implicira da je potrebno dokazati da auto neće da upali i da je napon na akumulatoru manji od 12 volti da bi zaključak bio da je akumulator prazan. Prema tome, “Auto neće da upali” i “Napon na akumul. < 12V” postaju novi ciljevi (dodaju se u agendu ciljeva).

U trećem ciklusu se proverava da li važe ciljevi “Auto neće da upali” i “Napon na akumul. < 12V”. Činjenice iz radne memorije NE potvrđuju da važe ovi ciljevi. Prvi cilj “Auto neće da upali” je dokazan, ali drugi “Napon na akumul. < 12V” NIJE (napon je 12 volti). Da bi mogao da se dokaže nadređeni cilj, potrebno je da oba cilja važe. Pošto važi samo jedan, ne može se dokazati cilj “Akumulator je prazan”. Iz toga sledi da NE VAŽI ni osnovni cilj - “Napuni akumulator”. Osnovni cilj nije dokazan, pa se zaključivanje završava.

(b) Na osnovu prethodnog, može se zaključiti da se ne može dokazati osnovni cilj jer se ne mogu dokazati podciljevi iz agende, pa je odgovor NE.

(c) Na osnovu prethodnog, može se zaključiti da se PRAVILO 3 neće uzeti u razmatranje pa je odgovor NE.

Zadatak 6

Radna Memorija	Baza znanja
1) Auto neće da “upali” 2) Napon na akum. = 12V	PRAVILO 1: IF Auto neće da “upali” AND Napon na akum. < 12V THEN Akumulator je prazan
Agenda ciljeva – Zameni anlaser	PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akum. = 12V THEN Anlaser je неисправan PRAVILO 4: IF Anlaser je неисправan THEN Zameni anlaser

Ako je sadržaj radne memorije i baze znanja kao u tabeli i ako se zna da mehanizam za zaključivanje koristi ulančavanje unazad, odgovoriti na sledeća pitanja:

- Kako će na kraju ulančavanja unazad da izgleda agendu ciljeva (dopisati u gornju tabelu sve nove ciljeve koji će biti dodati u agendu ciljeva)?
- Da li može da se dokaže osnovni cilj “Zameni anlaser”?
- Da li će se u nekom trenutku u razmatranje uzeti PRAVILO 2?

Rešenje

(a) Posle dopune, tabela će da izgleda ovako:

Radna Memorija	Baza znanja
1) Auto neće da “upali” 2) Napon na akum. = 12V	PRAVILO 1: IF Auto neće da “upali” AND Napon na akum. < 12V THEN Akumulator je prazan
Agenda ciljeva – Zameni anlaser – Anlaser je неисправan – Auto neće da “upali” – Napon na akum. = 12V	PRAVILO 2: IF Akumulator je prazan THEN Napuni akumulator PRAVILO 3: IF Auto neće da “upali” AND Napon na akum. = 12V THEN Anlaser je неисправan PRAVILO 4: IF Anlaser je неисправan THEN Zameni anlaser

(b) DA

(c) NE

Objašnjenje

(a) U prvom ciklusu se proverava da li se možda osnovni cilj “Zameni anlaser” već nalazi u radnoj memoriji u vidu činjenice. Kada bi se nalazio, cilj bi bio dokazan i zaključivane bi se prekinulo. Pošto se ne nalazi (nedokazan je), traži se pravilo iz baze znanja koje ima ovaj cilj u svom THEN delu. To je PRAVILO 4. To pravilo implicira da je potrebno dokazati da je anlaser неисправan da bi zaključak bio da je potrebno zameniti anlaser. Prema tome, “Anlaser je неисправan” postaje novi cilj tj. dodaje se u agendu ciljeva.

U drugom ciklusu se proverava da li se možda cilj “Anlaser je neispravan” već nalazi u radnoj memoriji u vidu činjenice. Pošto se ne nalazi, traži se pravilo iz baze znanja koje ima ovaj cilj u svom THEN delu. To je PRAVILO 3. To pravilo implicira da je potrebno dokazati da auto neće da upali i da je napon na akumulatoru jednak 12 volti da bi zaključak bio da je anlaser neispravan. Prema tome, “Auto neće da upali” i “Napon na akumul. = 12V” postaju novi ciljevi (dodaju se u agendu ciljeva).

U trećem ciklusu se proverava da li važe ciljevi “Auto neće da upali” i “Napon na akumul. = 12V”. Činjenice iz radne memorije potvrđuju da važe ovi ciljevi – auto zaista neće da upali a napon jeste 12 volti. Samim tim, ova dva cilja su dokazana. Pošto ova dva cilja važe, važiće i sledeći, njima nadređeni cilj “Anlaser je neispravan”. Pošto važi i ovaj cilj, sledi da važi i osnovni cilj - “Zameni anlaser”. Osnovni cilj je dokazan, pa se zaključivanje završava.

(b) Na osnovu prethodnog, može se zaključiti da se može dokazati osnovni cilj jer se mogu dokazati i svi podciljevi iz agende, pa je odgovor DA.

(c) Na osnovu prethodnog, može se zaključiti da se PRAVILO 2 neće uzeti u razmatranje pa je odgovor NE.

Zadatak 7

Radna Memorija	Baza znanja
1) Korisnik želi kvalitetan aparat 2) Korisnik nije spreman da plati visoku cenu 3) Korisnik želi veliki ekran	<p>PRAVILO 1: IF Korisnik želi kvalitetan aparat AND Korisnik je spreman da plati visoku cenu THEN Robna marka je Sony</p> <p>PRAVILO 2: IF Robna marka je Sony AND Korisnik želi veliki ekran THEN Model televizora je Sony WS32</p> <p>PRAVILO 3: IF Korisnik želi kvalitetan aparat AND Korisnik nije spreman da plati visoku cenu THEN Robna marka je Samsung</p> <p>PRAVILO 4: IF Robna marka je Sony AND Korisnik želi mali ekran THEN Model televizora je Sony SS20</p> <p>PRAVILO 5: IF Robna marka je Samsung AND Korisnik želi veliki ekran THEN Model televizora je Samsung HD32</p>

Ako je sadržaj radne memorije i baze znanja kao u tabeli i ako se zna da mehanizam za zaključivanje koristi ulančavanje unapred, odgovoriti na sledeća pitanja:

- Do kojih će sve zaključaka doći ES (dopisati u gornju tabelu sve nove činjenice tj. zaključke koji će biti dodati u radnu memoriju)?
- Kojim redosledom će da se izvršavaju pravila?
- Da li će se u nekom trenutku izvršiti PRAVILO 4?

Rešenje

(a) Posle dopune, tabela bi trebalo da izgleda ovako:

Radna Memorija	Baza znanja
1) Korisnik želi kvalitetan aparat 2) Korisnik nije spreman da plati visoku cenu 3) Korisnik želi veliki ekran 4) <u>Robna marka je Samsung</u> 5) <u>Model televizora je Samsung HD32</u>	PRAVILO 1: IF Korisnik želi kvalitetan aparat AND Korisnik je spreman da plati visoku cenu THEN Robna marka je Sony PRAVILO 2: IF Robna marka je Sony AND Korisnik želi veliki ekran THEN Model televizora je Sony WS32 PRAVILO 3: IF Korisnik želi kvalitetan aparat AND Korisnik nije spreman da plati visoku cenu THEN Robna marka je Samsung PRAVILO 4: IF Robna marka je Sony AND Korisnik želi mali ekran THEN Model televizora je Sony SS20 PRAVILO 5: IF Robna marka je Samsung AND Korisnik želi veliki ekran THEN Model televizora je Samsung HD32

- (b) 1. PRAVILO 3
2. PRAVILO 5

(c) NE

Zadatak 8

Radna Memorija	Baza znanja
1) Korisnik želi kvalitetan aparat 2) Korisnik je spreman da plati visoku cenu 3) Korisnik želi mali ekran	PRAVILO 1: IF Korisnik želi kvalitetan aparat AND Korisnik je spreman da plati visoku cenu THEN Robna marka je Sony PRAVILO 2: IF Robna marka je Sony AND Korisnik želi veliki ekran THEN Model televizora je Sony WS32 PRAVILO 3: IF Korisnik želi kvalitetan aparat AND Korisnik nije spreman da plati visoku cenu THEN Robna marka je Samsung PRAVILO 4: IF Robna marka je Sony AND Korisnik želi mali ekran THEN Model televizora je Sony SS20 PRAVILO 5: IF Robna marka je Samsung AND Korisnik želi veliki ekran THEN Model televizora je Samsung HD32

Ako je sadržaj radne memorije i baze znanja kao u tabeli i ako se zna da mehanizam za

zaključivanje koristi ulančavanje unapred, odgovoriti na sledeća pitanja:

- (g) Do kojih će sve zaključaka doći ES (dopisati u gornju tabelu sve nove činjenice tj. zaključke koji će biti dodati u radnu memoriju)?
- (h) Kojim redosledom će da se izvršavaju pravila?
- (i) Da li će se u nekom trenutku izvršiti PRAVILO 4?

Rešenje

(a) Posle dopune, tabela bi trebalo da izgleda ovako:

Radna Memorija	Baza znanja
1) Korisnik želi kvalitetan aparat 2) Korisnik nije spreman da plati visoku cenu 3) Korisnik želi mali ekran 4) <u>Robna marka je Sony</u> 5) <u>Model televizora je Sony SS20</u>	PRAVILO 1: IF Korisnik želi kvalitetan aparat AND Korisnik je spreman da plati visoku cenu THEN Robna marka je Sony PRAVILO 2: IF Robna marka je Sony AND Korisnik želi veliki ekran THEN Model televizora je Sony WS32 PRAVILO 3: IF Korisnik želi kvalitetan aparat AND Korisnik nije spreman da plati visoku cenu THEN Robna marka je Samsung PRAVILO 4: IF Robna marka je Sony AND Korisnik želi mali ekran THEN Model televizora je Sony SS20 PRAVILO 5: IF Robna marka je Samsung AND Korisnik želi veliki ekran THEN Model televizora je Samsung HD32

- (b) 1. PRAVILO 1
- 2. PRAVILO 4

(c) DA

Zadatak 9

Radna Memorija	Baza znanja
1) Korisnik želi kvalitetan aparat 2) Korisnik nije spreman da plati visoku cenu 3) Korisnik želi veliki ekran	PRAVILO 1: IF Korisnik želi kvalitetan aparat AND Korisnik je spreman da plati visoku cenu THEN Robna marka je Sony PRAVILO 2: IF Robna marka je Sony AND Korisnik želi veliki ekran THEN Model televizora je Sony WS32
Agenda ciljeva – Model televizora je Sony WS32	PRAVILO 3: IF Korisnik želi kvalitetan aparat AND Korisnik nije spreman da plati visoku cenu THEN Robna marka je Samsung PRAVILO 4: IF Robna marka je Sony AND Korisnik želi mali ekran THEN Model televizora je Sony SS20 PRAVILO 5: IF Robna marka je Samsung AND Korisnik želi veliki ekran THEN Model televizora je Samsung HD32

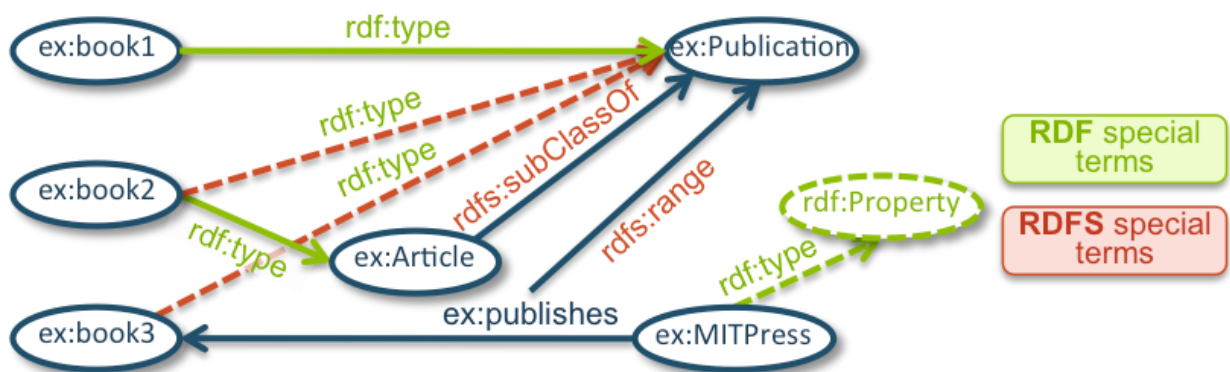
Ako je sadržaj radne memorije i baze znanja kao u tabeli i ako se zna da mehanizam za zaključivanje koristi ulančavanje unazad, odgovoriti na sledeća pitanja:

- Kako će na kraju ulančavanja unazad da izgleda agendu ciljeva (dopisati u gornju tabelu sve nove ciljeve koji će biti dodati u agendu ciljeva)?
- Da li može da se dokaže osnovni cilj “Model televizora je Sony WS32”?
- Da li će se u nekom trenutku u razmatranje uzeti PRAVILO 5?

Rešenje

(a) Posle dopune, tabela će da izgleda ovako:

Zadaci iz RDF & RDFS



1. Nacrtati grafički prikaz sledećeg RDF segmenta:

```
@prefix ex: <http://www.example.org/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<http://www.example.org/tutorials#RDFTutorial>
  dc:title "An Introduction to RDF" ;
  dc:date "23/01/2003" ;
  dc:creator _:staffID85740 .
_:staffID85740 ex:name "Brian McBride" ;
               ex:email "brian@yahoo.com" , "brian@hotmail.com" .
```

2. Nacrtati grafički prikaz sledećeg RDF segmenta:

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

_:peter rdf:type foaf:Person ;
        foaf:name "Peter Parker" ;
        foaf:gender "male" ;
        foaf:homepage <http://www.peterparker.com>;
        foaf:knows _harry .

_:harry foaf:firstName "Harry" ;
        foaf:surname "Osborn" ;
        rdfs:seeAlso <http://www.osborn.com/harry.rdf> .
```

3. Nacrtati grafički prikaz sledećeg RDF segmenta:

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<#www.cs.umd.edu/~hendler/2003/foaf.rdf#jhendler>
  rdf:type foaf:Person ;
  foaf:name "Jim Hendler" ;
  foaf:nick "jhendler" ;
  foaf:mbox_sha1sum "0b62d4242736e64be6138547c79a811b3e82fd52" ;
  foaf:holdsAccount [
    rdf:type foaf:OnlineAccount ;
    foaf:accountName "jhendler" ;
    foaf:accountServiceHomepage <http://trust.mindswap.org/FilmTrust/>
  ] .
```

4. Nacrtati grafički prikaz sledećeg RDF segmenta:

```
@prefix con: <http://www.w3.org/2000/10/swap/pim/contact#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<http://dbpedia.org/resource/Tim_Berners-Lee>
  con:organization <http://www.w3.org/data#W3C> ;
  con:office [
    con:phone "+1-617-253-5702" ;
    con:address [
      con:city "Cambridge" ;
      con:country "USA" ;
      con:postalCode "02139" ;
      con:street "32 Vassar Street" .
    ] .
  ] .
```

5. Nacrtati grafički prikaz sledećeg RDF segmenta:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix um: <http://example.org/usermodel/> .

<um:jeenbroekstra> a um:User ;
    foaf:firstName "Jeen" ;
    foaf:familyName "Broekstra" ;
    foaf:mbox <mailto:jeen.broekstra@aduna-software.com> ;
    rdfs:seeAlso <http://www.openrdf.org/people/foaf-jeen.rdf> .

<um:jeenbroekstra> um:rating [
    a um:ActorRating ;
    um:onActor <http://www.imdb.com/name/nm0000164/> ;
    rdf:value "6"^^xsd:integer
] .
```

6. Nacrtati grafički prikaz sledećeg RDF segmenta:

```
@prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#> .
@prefix talk: <http://www.w3.org/2004/08/Presentations.owl#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<#X_2005-09-08-349>
    rdf:type talk:Talk ;
    dc:date "2005-09-08" ;
    dc:title "Declaring Language in XHTML and HTML - Latest Thinking at the W3C" ;
    talk:event
        [ vcard:Country "USA" ;
          vcard:Locality "Orlando, Florida" ;
          dc:description "28th Internationalization & Unicode Conference" ;
          rdfs:seeAlso <http://www.global-conference.com/iuc28/>
        ] ;
    talk:presenter
        [ rdf:type foaf:Person ;
          foaf:name "Richard Ishida" ;
          foaf:mbox <mailto:ishida@w3.org>
        ] ;
    talk:type "talk".
```

7. U tekstu koji sledi dat je opis nekih klasa, njihovih svojstava i relacija. Takođe su opisane i neke instance ovih klasa. Potrebno je kreirati RDF model koji formalizuje znanje sadržano u datom tekstu. Model je potrebno predstaviti grafički i u jednoj od izabranih RDF sintaksi (RDF/XML ili Turtle).

Postoje publikacije.

Knjige, novinski clanci, strucni radovi su tipovi publikacija.

Publikacije imaju autore.

Autori su osobe.

Publikacije imaju naslov.

Postoji knjiga čiji je naslov "Weaving the Web", a autor osoba čije je ime "Tim Berners Lee".

Postoji stručni rad čiji je naslov "The Semantic Web" a autori:

- osoba čije je ime "Tim Berners Lee"
- osoba čije je ime "Jim Hendler"
- osoba čije je ime "Ora Lassila"

8. Korišćenjem SKOS klasa i properties kreirati RDF konceptualnu schemu koja formalizuje date podatke (tj. kreirati jednu mini taksonomiju vozila). Kreiranu konceptualnu schemu je potrebno predstaviti grafički i u jednoj od izabranih RDF sintaksi (RDF/XML ili Turtle).

Postoje vozila.
Putnička vozila i teretna vozila su vrste vozila.
Autobus, automobil i motocikl su tipovi putničkih vozila.
Kamion je vrsta teretnog vozila.
Minibus i double-decker su vrsta autobusa.
Pick-up je vrsta kamiona.
Za automobil se često koristi kraći termin 'auto';
za motocikl se koristi i termin 'motor';
za autobus se u žargonu koristi termin 'bus'.

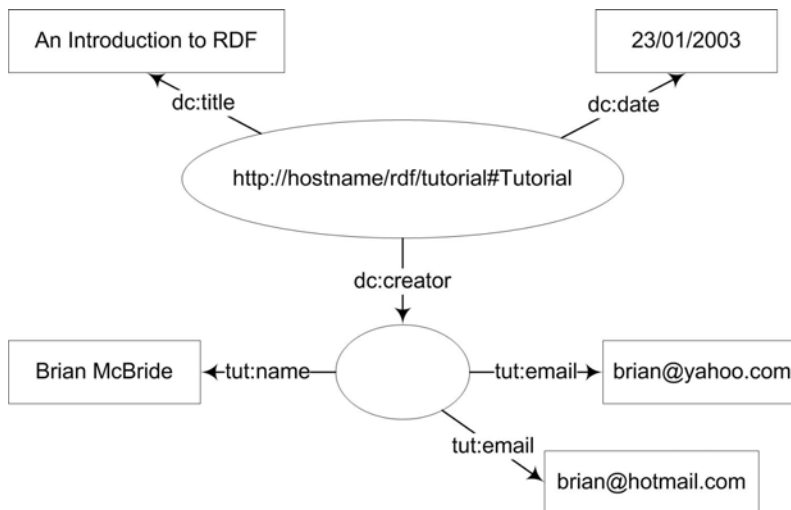
9. Korišćenjem FOAF klasa i properties kreirati RDF model koji formalizuje podatke o nekoj osobi date u tekstu koji sledi. Model je potrebno predstaviti grafički i u jednoj od izabranih RDF sintaksi (RDF/XML ili Turtle).

Osoba čiji URI je 'http://criptonita.com/~nacho/foaf.rdf#me '
zove se 'Nacho Arias';
nadimak joj je 'chipi';
homepage joj je 'http://criptonita.com/~nacho';
slika ove osobe je
'http://serv.inforg.uniovi.es/~webadmin/nacho/chipi.png';
ova osoba poznaje osobu koja se zove 'Sergio Fernandez',
čiji je nadimak 'wikier' i homepage 'http://www.wikier.org/';
takođe poznaje i osobu koja se zove 'Ivan Frade' i
o kojoj se više informacija može naći u fajlu
'http://frade.no-ip.info:2080/~ivan/foaf.rdf'.

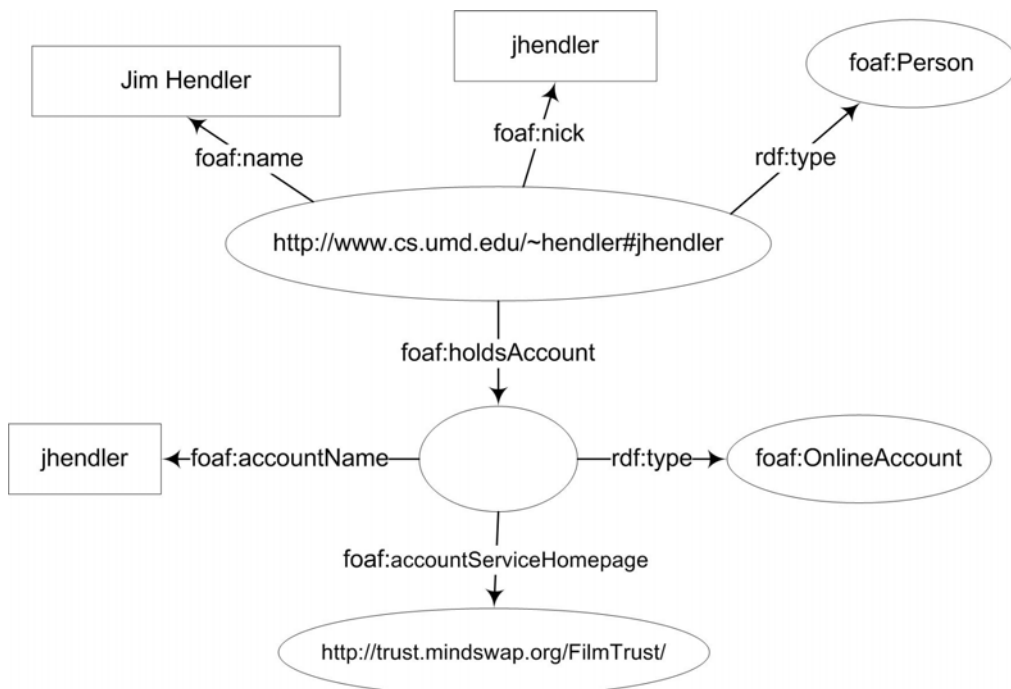
10. U tekstu koji sledi dat je opis nekih klasa, njihovih svojstava i relacija. Takođe su opisane i neke instance ovih klasa. Model je potrebno predstaviti grafički i u jednoj od izabranih RDF sintaksi (RDF/XML ili Turtle).

Postoje osobe i kućni ljubimci.
Kućni ljubici i osobe su živa bića.
Osoba može posedovati kućnog ljubimca.
Svako živo biće ima svoje ime.
Svako živo biće ima pol koji može biti muški ili ženski.
Kućni ljubimci imaju nadimke.
Mačka je kućni ljubimac.
Tom je mačka.
Tom je muškog pola.
John je osoba.
John je muškog pola.
John ima kućnog ljubimca Toma.
Tom ima nadimak 'Tommy'.

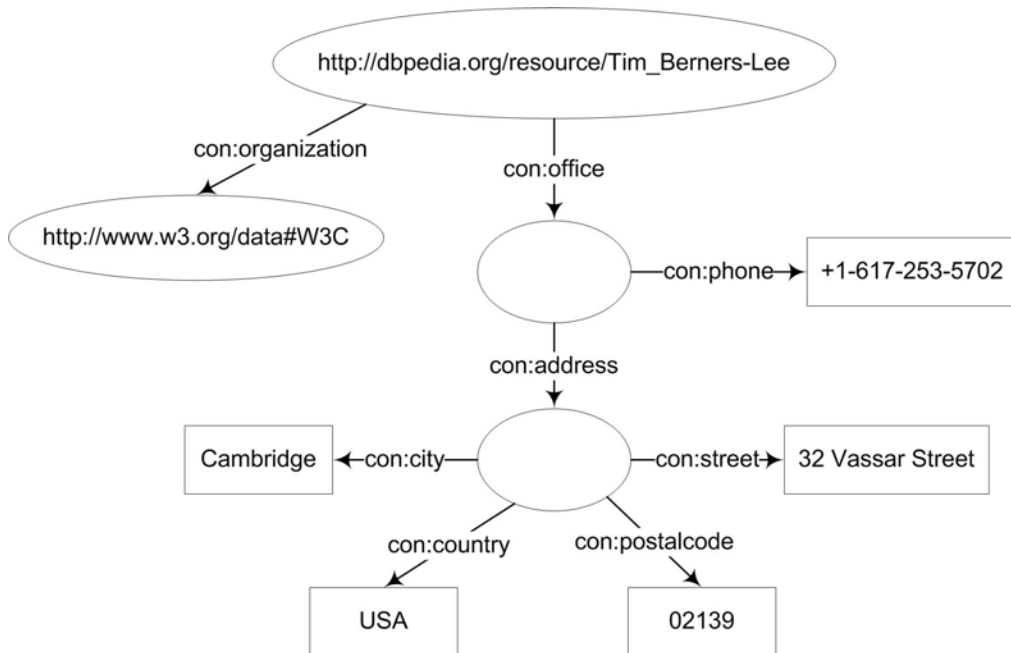
1.



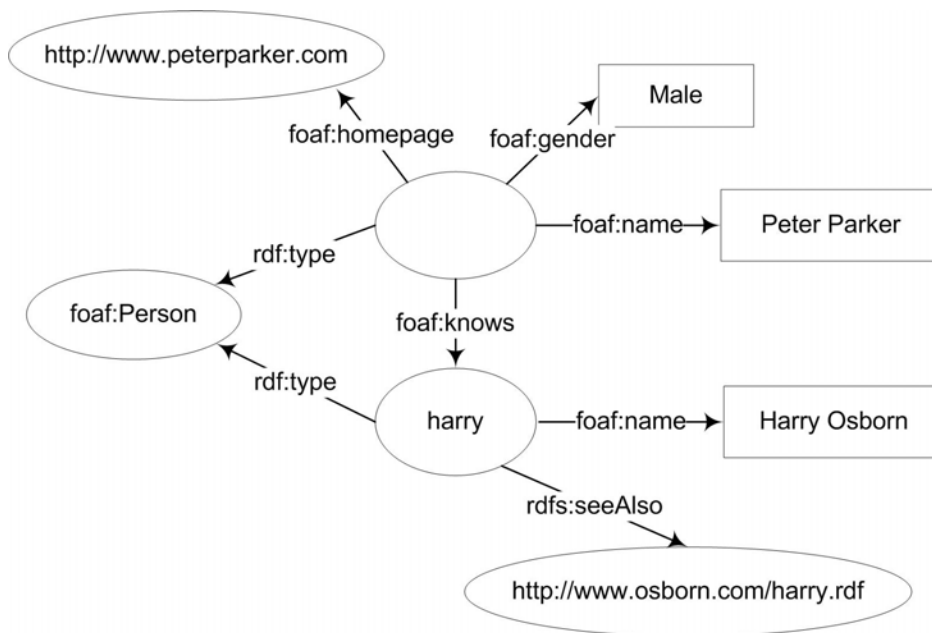
2.



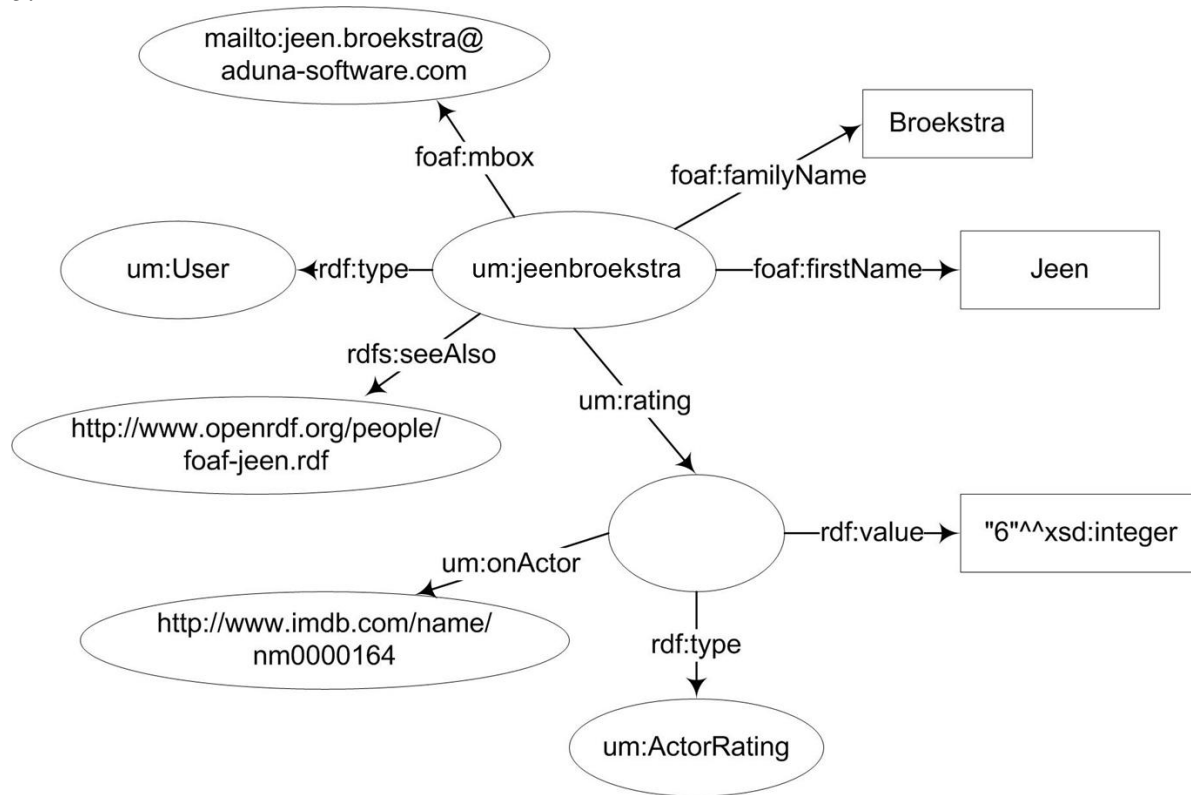
3.



4.

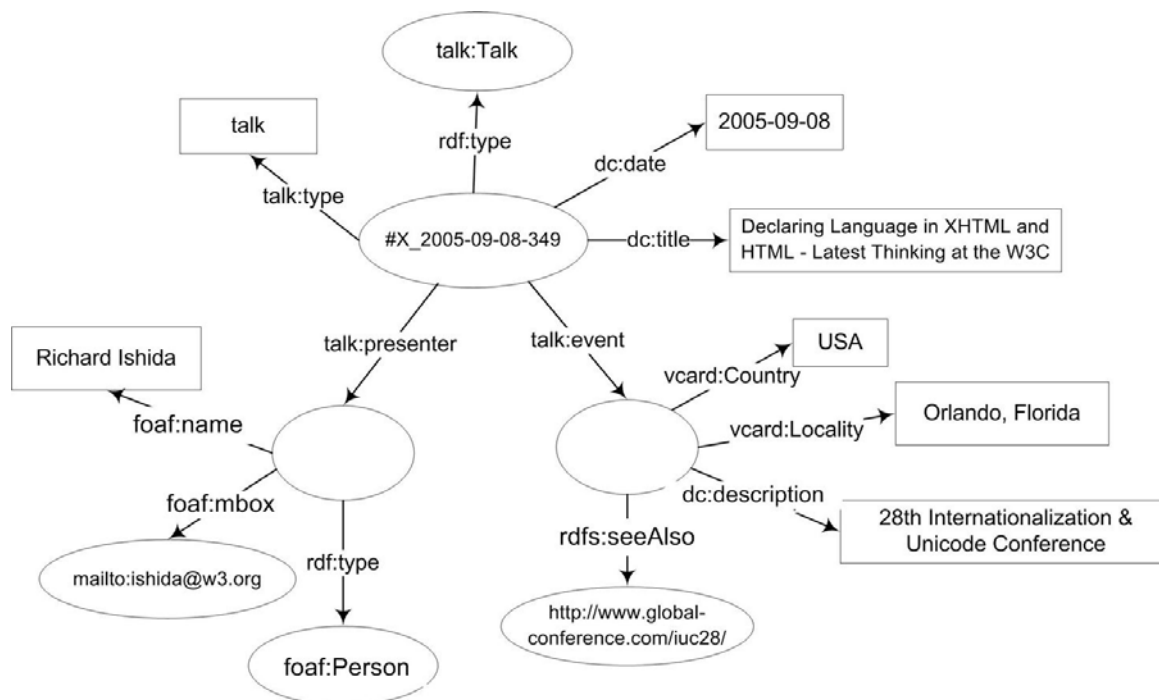


5.



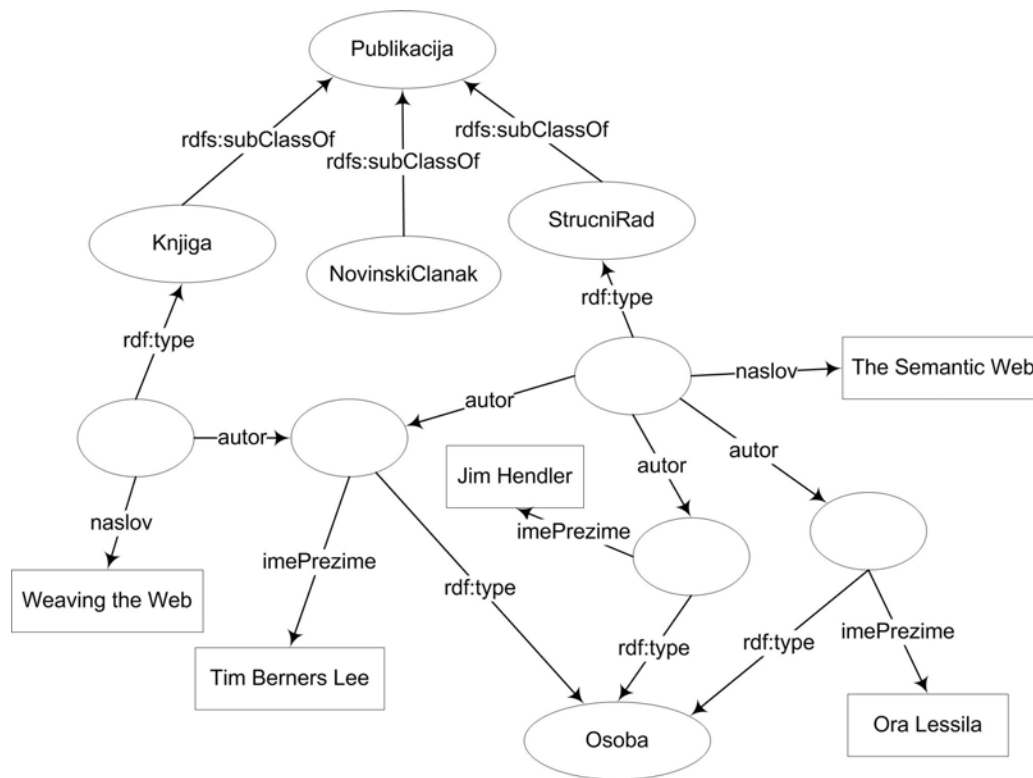
08.06.2011. Korekcija rešenja: centralni resurs – `um:jeenbroekstra` – je imao pogresan prefiks, tj. namespace alijas (umesto `um`, stajalo je `foaf`)

6.



08.06.2011. Korekcija rešenja: Neki properties prikazani na slici (tj. grafu) nisu bili uskladjeni sa postavkom zadatka; konkretno, unete su sledeće izmene: contact:fullname -> foaf:name; contact:mailbox -> foaf:mbox; ical:description -> dc:description; ical:url -> rdfs:seeAlso

7.



RDF model sa gornje slike predstavljen korišenjem TURTLE sintakse:

@prefix ex: <http://xmltehnologije/rdf/zadaci-za-vezbu/publikacije/> .
 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

ex:NovinskiClanak
 rdfs:subClassOf ex:Publikacija .

ex:autor
 rdfs:domain ex:Publikacija ;
 rdfs:range ex:Osoba .

<paper>
 rdf:type ex:StrucniRad ;
 ex:autor <author2> , <author3> , <author1> ;
 ex:naslov "The Semantic Web" .

ex:StrucniRad
 rdfs:subClassOf ex:Publikacija .

ex:Knjiga
 rdfs:subClassOf ex:Publikacija .

ex:imePrezime
 rdfs:domain ex:Osoba .

<author2>
 rdf:type ex:Osoba ;
 ex:imePrezime "Jim Hendler" .

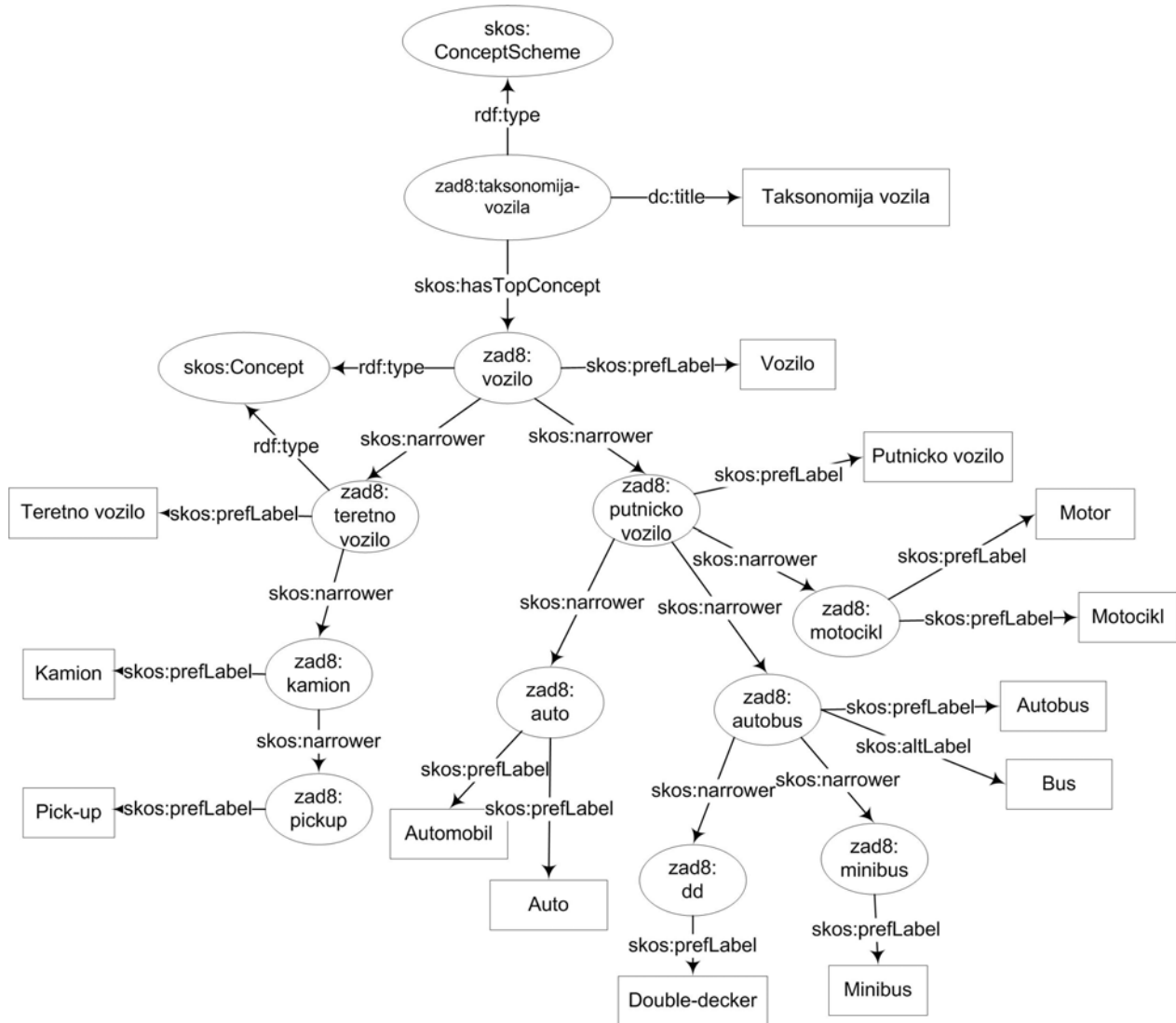
<author3>
 rdf:type ex:Osoba ;
 ex:imePrezime "Ora Lessila" .

<author1>
 rdf:type ex:Osoba ;
 ex:imePrezime "Tim Berners Lee" .

ex:naslov
 rdfs:domain ex:Publikacija .

<book1>
 rdf:type ex:Knjiga ;
 ex:autor <author1> ;
 ex:naslov "Weaving the Web"^^<http://www.w3.org/2001/XMLSchema#string> .

8.



Napomene:

- prefiks zad8 označava namespace 'http://inteligentni-sistemi/rdf/zadaci-za-vezbu/zadatak8'.
- prefiks skos označava namespace SKOS Core specifikacije "http://www.w3.org/2004/02/skos/core#"
- samo za dva resursa tipa skos:Concept je eksplicitno definisan tip (korišćenjem rdf:type property-a) iz razloga što bi predstavljanje ove relacije za ostale resurse koji su tipa skos:Concept učinilo sliku teško razumljivom.

RDF model sa gornje slike predstavljen korišćenjem TURTLE sintakse:

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix zad8 <http://xmltehnologije/rdf/zadaci-za-vezbu/zadatak8#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

zad8:kamion
rdf:type skos:Concept ;
skos:broader zad8:teretno-vozilo ;
skos:narrower zad8:pickup ;
skos:prefLabel "Kamion"^^xsd:string .

zad8:motocikl
rdf:type skos:Concept ;
skos:broader zad8:putnicko-vozilo ;
skos:prefLabel "Motor"^^xsd:string , "Motocikl"^^xsd:string .

zad8:putnicko-vozilo
rdf:type skos:Concept ;
skos:broader zad8:vozilo ;
skos:narrower zad8:motocikl , zad8:auto , zad8:autobus ;
skos:prefLabel "Putnicko vozilo"^^xsd:string .

zad8:minibus
rdf:type skos:Concept ;
skos:broader zad8:autobus ;
skos:prefLabel "Minibus"^^xsd:string .

zad8:taksonomija-vozila
rdf:type skos:ConceptScheme ;
dc:title "Taksonomija vozila"^^xsd:string ;
skos:hasTopConcept zad8:vozilo .

zad8:autobus
rdf:type skos:Concept ;
skos:altLabel "Bus"^^xsd:string ;
skos:broader zad8:putnicko-vozilo ;
skos:narrower zad8:minibus , zad8:dd ;
skos:prefLabel "Autobus"^^xsd:string .

zad8:teretno-vozilo
rdf:type skos:Concept ;
skos:broader zad8:vozilo ;
skos:narrower zad8:kamion ;
skos:prefLabel "Teretno vozilo"^^xsd:string .

zad8:auto
rdf:type skos:Concept ;
skos:broader zad8:putnicko-vozilo ;
skos:prefLabel "Auto"^^xsd:string , "Automobil"^^xsd:string .

zad8:vozilo
rdf:type skos:Concept ;
skos:narrower zad8:putnicko-vozilo , zad8:teretno-vozilo ;
skos:prefLabel "Vozilo"^^xsd:string .

zad8:pickup
rdf:type skos:Concept ;
skos:broader zad8:kamion ;
skos:prefLabel "Pick-up"^^xsd:string .

zad8:dd rdf:type skos:Concept ;
skos:broader zad8:autobus ;
skos:prefLabel "Double-decker"^^xsd:string .

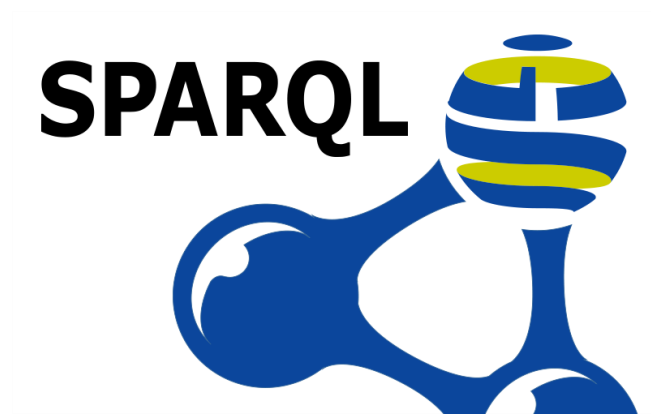
```

graph TD
    Central([http://criptonita.com/~nacho/foaf.rdf#me])
    Chipi([http://serv.inforg.uniovi.es/~webadmin/nacho/chipi.png])
    NachoAri([http://criptonita.com/~nacho])
    Sergio([Sergio Fernandez])
    Wikier([http://www.wikier.org/])
    Ivan([Ivan Frade])
    Frade([http://frade.no-ip.info:2080/~ivan/foaf.rdf])
    Person1([foaf:Person])
    Person2([foaf:Person])

    Central -- foaf:homepage --> Chipi
    Central -- foaf:nick --> ChipiBox[chipi]
    Central -- foaf:name --> NachoAriBox[Nacho Arias]
    NachoAri -- foaf:homepage --> Central
    Sergio -- foaf:knows --> Central
    Sergio -- foaf:name --> SergioBox[Sergio Fernandez]
    Sergio -- foaf:nick --> WikierBox[wikier]
    Sergio -- foaf:homepage --> Wikier
    Sergio -- rdf:type --> Person1
    Central -- rdf:type --> Person1
    Person1 -- foaf:knows --> Sergio
    Person1 -- foaf:knows --> IvanNode
    IvanNode -- foaf:name --> IvanBox[Ivan Frade]
    IvanNode -- rdfs:seeAlso --> Frade
    IvanNode -- rdf:type --> Person2
    Person2 -- rdf:type --> Person1
  
```

```
<?xml version="1.0" encoding="iso-8859-15"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
>
  <foaf:Person rdf:ID="http://criptonita.com/~nacho/foaf.rdf#me">
    <foaf:name>Nacho Arias</foaf:name>
    <foaf:nick>chipi</foaf:nick>
    <foaf:homepage rdf:resource="http://criptonita.com/~nacho"/>
    <foaf:depiction rdf:resource="http://serv.inforg.uniovi.es/~webadmin/nacho/chipi.png"/>
    <foaf:knows>
      <foaf:Person>
        <foaf:name>Sergio Fdez</foaf:name>
        <foaf:nick>wikier</foaf:nick>
        <foaf:homepage rdf:resource="http://www.wikier.org/" />
      </foaf:Person>
    </foaf:knows>
    <foaf:knows>
      <foaf:Person rdf:nodeID="frade">
        <foaf:name>Ivan Frade</foaf:name>
        <rdfs:seeAlso rdf:resource="http://frade.no-ip.info:2080/~ivan/foaf.rdf" />
      </foaf:Person>
    </foaf:knows>
  </foaf:Person>
</rdf:RDF>
```

Zadaci iz SPARQL-a



Napomena : da biste otvorili rdf dokumente idite na sajt <http://ai.fon.bg.ac.rs/osnovne/inteligentni-sistemi/pitanja-i-zadaci/> i skinite rar SPARQL i onda prevucite .rdf fajl u eclipse npr (tako sam ja uspela da ih otvorim)

1. Dat je RDF fajl 'W3CTalks.rdf' čiji je URL 'http://www.w3.org/2004/08/TalkFiles/2005/Talks.rdf'. Primenom SPARQL upita odrediti:

- a) imena i prezimena (contact:fullName) i email adrese (contact:mailbox) svih osoba koje su održale izlaganje (talk:Talk) koje je tipa 'tutorial';
- b) naslove (dc:title) izlaganja održanih (vcard:Locality) u Los Angeles-u sortirane u opadajućem redosledu datuma održavanja (dc:date);
- c) naslove izlaganja i imena i prezimena osoba koje su održale izlaganja vezana za w3c aktivnost (talk:w3cActivity) označenu kao 'Semantic Web Activity'.

2. Dat je RDF fajl 'dan-brickley-foaf.rdf' čiji je URL 'http://danbri.org/foaf.rdf'. Primenom odgovarajućih SPARQL upita:

- d) Pronaći slike (foaf:depiction ili foaf:img) osoba koje poznaje (foaf:knows) osoba čiji je email (foaf:mbox) danbri@danbri.org.
- e) Za osobu čiji je openId (foaf:openid) <http://danbri.org/>, odrediti email adrese (foaf:mbox) koje su još uvek aktuelne, tj. nisu zamenjene (dct:isReplacedBy) nekim drugim email adresama.
- f) Proveriti da li osoba čiji je homepage 'http://danbri.org/' ima nalog (foaf:OnlineAccount) na del.icio.us-u (<http://del.icio.us>).

3. Dat je RDF fajl 'Agenda_of_a_meeting.rdf' čiji je URL 'http://www.w3.org/2005/03dc-msp/agenda_62.rdf'. Primenom CONSTRUCT upita kreirati novi RDF graf (tj. model) koji će sadržati podatke samo o onim događajima (cal:Vevent) koji su održani (cal:location) u nekom od salona: 'Salon A', 'Salon B', 'Salon C',... U novom RDF grafu potrebno je za svaki takav događaj predstaviti samo njegov opis korišćenjem *dc:description* property-a. Opis događaja u ulaznom rdf fajlu dat je kao vrednost property-a *cal:description* i/ili *cal:summary*.

Rešenja

Zadatak 1.

a)

```
PREFIX contact: <http://www.w3.org/2000/10/swap/pim/contact#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX talk: <http://www.w3.org/2004/08/Presentations.owl#>
FROM <http://www.w3.org/2004/08/TalkFiles/2005/Talks.rdf>
SELECT DISTINCT ?name ?mbox
WHERE {
    ?talk rdf:type talk:Talk ;
          talk:type ?talkType ;
          talk:presenter ?person .
    ?person contact:fullName ?name ;
            contact:mailbox ?mbox .
    FILTER regex(?talkType, "tutorial", "i") .
}
```

b)

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX talk: <http://www.w3.org/2004/08/Presentations.owl#>
PREFIX vcard: <http://www.w3.org/2001/vcard-rdf/3.0#>
FROM <http://www.w3.org/2004/08/TalkFiles/2005/Talks.rdf>
SELECT ?title ?date
WHERE {
    ?talk rdf:type talk:Talk ;
          dc:title ?title ;
          dc:date ?date ;
          talk:event ?event .
    ?event vcard:Locality ?loc.
    FILTER regex (?loc, "Los Angeles", "i") .
} ORDER BY DESC (?date)
```

c)

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX talk: <http://www.w3.org/2004/08/Presentations.owl#>
PREFIX contact: <http://www.w3.org/2000/10/swap/pim/contact#>
PREFIX org: <http://www.w3.org/2001/04/roadmap/org#>
FROM <http://www.w3.org/2004/08/TalkFiles/2005/Talks.rdf>
SELECT ?title ?presenter
WHERE {
    ?talk rdf:type talk:Talk ;
          dc:title ?title ;
          talk:presenter ?p ;
          talk:w3cActivity ?a .
    ?p contact:fullName ?presenter .
    ?a org:name ?orgName .
    FILTER regex (?orgName, "Semantic Web Activity", "i") .
}
```

Zadatak 2.

a)

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
FROM <http://danbri.org/foaf.rdf>
SELECT ?picture
WHERE {
    ?dan rdf:type foaf:Person;
        foaf:mbox <mailto:danbri@danbri.org>;
        foaf:knows ?someone.
    {?someone foaf:img ?picture .}
    UNION
    {?someone foaf:depiction ?picture. }
}
```

b)

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dct: <http://purl.org/dc/terms/>
FROM <http://danbri.org/foaf.rdf>
SELECT ?email
WHERE {
    ?dan rdf:type foaf:Person;
        foaf:openid <http://danbri.org/>;
        foaf:mbox ?email.
    OPTIONAL {?email dct:isReplacedBy ?newEmail .}
    FILTER ( !bound(?newEmail) ) .
}
```

c)

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
FROM <http://danbri.org/foaf.rdf>
ASK
WHERE {
    ?person rdf:type foaf:Person ;
        foaf:homepage <http://danbri.org/> ;
        foaf:holdsAccount ?account.
    ?account foaf:accountServiceHomepage <http://del.icio.us>
}
```

Zadatak 3.

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX cal: <http://www.w3.org/2002/12/cal/ical#>
FROM <http://www.w3.org/2005/03dc-msp/agenda_62.rdf>
CONSTRUCT {
    ?event rdf:type cal:Vevent;
           dc:description ?desc .
} WHERE {
    ?event cal:location ?location .
           FILTER regex(?location, "Salon") .
    { ?event cal:description ?desc .}
    UNION
    { ?event cal:summary ?desc .}
}
```