

PREGLED TEHNIKA I POSTUPAKA PORAVNAVANJA ONTOLOGIJA³

SAŽETAK

Ontologije čine model za predstavljanje znanja na formalan, strojno čitljiv način. Posljednjih godina razvijene su brojne ontologije, a njihovo semantičko povezivanje je nužan preduvjet za ostvarivanje interoperabilnosti između agenata, servisa i korisnika koji koriste različite ontologije. Zbog velikog broja ontologija i njihove složenosti, bilo bi zahtjevno proces poravnavanja ontologija obavljati ručno, stoga se razvijaju sustavi za automatsko i poluautomatsko poravnavanje ontologija. Cilj ovog rada je dati pregled tehnika koje omogućavaju poravnavanje te prikazati kako se one kombiniraju u okviru postupka poravnavanja ontologija. U radu su prikazani različiti teorijski pristupi klasifikacije tehnika poravnavanja ontologija. Pored toga, opisane su osnovne faze u okviru kojih se odvija postupak poravnavanja ontologija. Na kraju je dan pregled sustava koji su razvijeni za povezivanje i poravnavanje ontologija u okviru kojega je prikazano koje skupine tehnika su implementirane u pojedinim sustavima. Pregled i analiza postojećih tehnika i postupaka poravnavanja te njihovih implementacija dovode do zaključka da još uvijek ima prostora za bolje iskorištavanje postojećih tehnika, definiranje novih pristupa i unapređivanje cjelokupnog postupka poravnavanja ontologija.

Ključne riječi: poravnavanje ontologija, uspoređivanje ontologija, semantička sličnost, tehnike poravnavanja ontologija, sustavi za poravnavanje ontologija

1. UVOD

Ontologija je oblik predstavljanja znanja o stvarnom svijetu. Prema definiciji T. Gubera, ontologija je eksplicitna specifikacija konceptualizacije (Gómez-Pérez et al., 2004:6). Iz perspektive računarskih znanosti, ontologija se odnosi na skup osnovnih reprezentacijskih funkcija pomoću kojih se može oblikovati model određene domene znanja (Lim et al., 2011). Osnovne reprezentacijske funkcije u ontologiji sadrže klase, attribute i relacije između klasa, a koriste se za modeliranje znanja pojedinih domena primjene. Ontologije su stoga postale vrlo važno područje istraživanja u razvoju informacijskih sustava temeljenih na znanju jer predstavljaju sredstvo za formalno oblikovanje strukture nekog sustava, odnosno prikazivanje relevantnih elemenata ontologije i odnosa među njima i to u strojno čitljivom formatu. Područje inženjeringa ontologija bavi se različitim postupcima razvoja ontologija, učenja ontologija, povezivanja ontologija i reinženjeringa ontologija (Gómez-Pérez et al., 2004).

¹ Mag. inf., E-mail: kjuricic@gmail.com

² Dr. sc., docent, Odjel za informatiku, Sveučilište u Rijeci, Radmile Matejčić 2, Rijeka, Hrvatska. E-mail: amestrovic@inf.uniri.hr

³ Datum primitka rada: 7. 2. 2013; datum prihvatanja rada: 3. 4. 2013

Poravnavanje ontologija (engl. *ontology alignment*) je postupak u kojem se za svaki element iz jedne ontologije pronalazi odgovarajući, istoznačni element u drugoj ontologiji, ukoliko takav postoji (Euzenat, Shvaiko, 2007). Poravnavanje ontologija podrazumijeva proces uspoređivanja ontologija (engl. *ontology matching*) koji se odnosi na pronalaženje semantičkih odnosa ili podudaranja između entiteta različitih ontologija. U nekim izvorima uspoređivanje (*matching*) se definira kao proces nalaženja podudarnosti (*alignment*) između entiteta iz različitih ontologija, dok je poravnanje izlazni rezultat tog procesa. U drugim izvorima ne pravi se razlika između pojmova *alignment* i *matching*. U ovom radu opisuju se generalno postupci poravnavanja, a samim time i postupci uspoređivanja ontologija.

Osnovna motivacija za razvoj postupaka poravnavanja ontologija je nadvladavanje različitosti u terminologiji te omogućavanje povezivanja ontologija. Konačni cilj je postizanje automatizirane komunikacije između heterogenih izvora znanja. Različiti sustavi uobičajeno koriste različitu terminologiju za opisivanje iste domene. Kako bi se automatizirala komunikacija između takvih sustava, potrebno je nadvladati semantičku heterogenost u terminologiji koja se koristi. Poravnavanje ontologija nužan je preduvjet za ostvarivanje interoperabilnosti⁴ između sustava, agenata i servisa koji koriste različite ontologije te za razvoj aplikacija i sustava kao što su sustavi za integraciju informacija, *peer-to-peer* sustavi, usluge weba koje se temelje na semantičkim tehnologijama i slično. Štoviše, korisnici žele pristupiti znanju prikazanom u brojnim ontologijama na transparentan način kako bi olakšali svoje aktivnosti pretraživanja ili pregledavanja informacija. Kombiniranjem informacija sadržanih u različitim ontologijama dolazi do stjecanja novih uvida u pohranjeno znanje. Stoga uspoređivanje, poravnavanje i pridruživanje ontologija predstavljaju ključne postupke u domeni inženjeringa ontologija.

Teorijska razmatranja i klasifikacija tehnika i postupaka poravnavanja u početku su se razvijala u domeni baza podataka (Rahm, Bernstein, 2001; Doan, Halevy, 2005;). U zadnjih dvadesetak godina postojeće tehnike poravnavanja i uspoređivanja počinju se primjenjivati u domeni inženjeringa ontologija. Značajan utjecaj na razvoj postupaka poravnavanja ima i područje procesiranja prirodnog jezika, strojno učenje i dubinska analiza podataka. Različite metode i tehnike teorijski su opisane u brojnoj literaturi (Rahm, Bernstein, 2001; Fensel, 2001; Davies et al., 2006; De Brujin, 2006; Ehrig, 2007; Euzenat, Shvaiko, 2007; Euzenat et al., 2008; Hepp, 2008; Bellahsene et al., 2011; Gargouri, Jaziri, 2011). Pored toga, prikazani su rezultati primjene tehnika poravnavanja ontologija (Granitzer, 2010; Jain et al., 2010; Euzenat et al., 2010; Euzenat et al., 2011; Shvaiko, Euzenat, 2012). Poravnavanje ontologija je složen, težak i vremenski zahtjevan zadatak, posebno u slučajevima velikih ontologija i ontologija koje imaju složene odnose među podacima (Romero et al., 2009). Istinski potencijal poravnavanja ontologija dolazi do izražaja kada su procesi razmjene informacija automatski integrirani. Stoga je cilj razviti tehnike koje će omogućiti visok stupanj automatizacije procesa poravnavanja ontologija.

Cilj ovoga rada je dati pregled tehnika koje omogućavaju poravnavanje te prikazati kako se one kombiniraju u okviru postupka poravnavanja ontologija. Opisano je jedanaest tehnika poravnavanja ontologija i prikazane su dvije moguće klasifikacije tih elementarnih tehnika. Analizirani su pristupi kombiniranja različitih tehnika u vidu sekvencijalne agregacije i paralelne agregacije tehnika, te

⁴ Semantička interoperabilnost podrazumijeva mogućnost dijeljenja podataka i znanja tako što je poznato i razumljivo značenje podataka (Bittner et al., 2005).

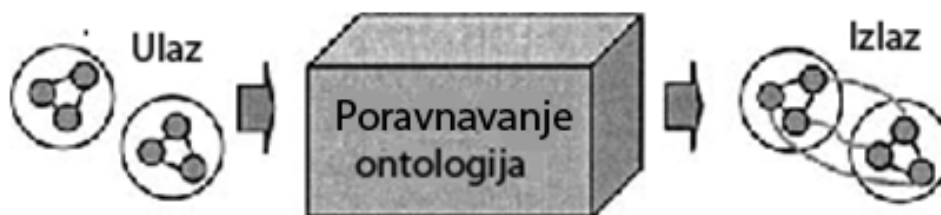
pristup koji uključuje različite mogućnosti intervencije korisnika. Kako bi se dobio uvid u složenost cjelokupnog postupka, opisane su faze općeg procesa poravnavanja. Na kraju je dan kratak pregled sustava koji imaju implementiran postupak poravnavanja ontologija. Sustavi za poravnavanje, uspoređivanje, povezivanje ontologija intenzivno se razvijaju zadnjih petnaestak godina. Danas je dostupan velik broj takvih sustava, a nude se različite mogućnosti i različita kvaliteta u postupku poravnavanja. Sa ciljem unapređivanja postupka poravnavanja razvija se OAEI (Ontology Alignment Evaluation Initiative) inicijativa (Euzenat et al., 2011) čiji je zadatak evaluirati i rangirati dostupne sustave za poravnavanje ontologija. U trenutno aktualnim istraživanjima (Euzenat 2007; Euzenat et al., 2008; Euzenat et al., 2009; Euzenat et al., 2011) koja se bave problemima evaluacije postojećih sustava za poravnavanje ontologija, postavlja se pitanje mogu li se postupci poravnavanja još dodatno unaprijediti i automatizirati. To je jedno od ključnih pitanja iz područja inženjeringa ontologija koje je trenutno aktualno. U ovom radu pokušat će se temeljem opisanih tehnika i postupaka analizirati postoje li mogućnosti unapređivanja postojećih tehnika i sustava.

U drugom poglavlju objašnjava se pojam sličnosti ontologija i njegova veza s pojmom poravnavanja. U trećem poglavlju opisane su tehnike i postupci definirani za poravnavanje ontologija, objašnjene su dvije različite klasifikacije tehnika za poravnavanje. U četvrtom poglavlju opisan je opći postupak poravnavanja ontologija i dan je pregled sustava u kojima su implementirane tehnike i postupci za poravnavanje ontologija. U petom poglavlju iznesena su zaključna razmatranja i moguće smjernice za nastavak istraživanja.

2. SLIČNOST ONTOLOGIJA I POSTUPAK PORAVNAVANJA

Postupak poravnavanja ontologija kao ulaz uzima dvije početne ontologije koje na različit način prikazuju istu domenu. Rezultat cjelokupnog postupka je skup poravnanja između dvije ulazne ontologije. Osnovna ideja postupka poravnavanja ontologija može se prikazati na sljedeći način (Ehrig, 2007):

Shema 1. Poravnavanje ontologija



Izvor: Ehrig (2007)

Osnovni cilj poravnavanja ontologija je nadvladati njihovu semantičku heterogenost. Međutim, iako se radi o različitim ontologijama, one moraju u određenoj mjeri pokazivati i sličnosti, odnosno moraju postojati i neka preklapanja među njima kako bi se uspješno mogli uspostaviti odnosi podudarnosti među pojedinim elementima tih dviju ontologija.

Ključni korak u postupku poravnavanja je uspoređivanje elemenata ontologije i izračunavanje sličnosti između tih elemenata. Stoga je za početak bitno definirati sličnost, odnosno mjeru sličnosti između ontologija. Jasno je da mjera sličnosti mora biti numerička vrijednost, no postoje različite tehnike i pristupi koji određuju kako se ona izračunava. Elementi koji se uspoređuju u procesu poravnavanja ontologija jesu najčešće klase, atributi, instance i relacije. Kao rezultat uspoređivanja pojedinačnih elemenata ontologija dobivaju se sličnosti na elementarnoj razini, a potom se može računati sličnost na razni cijelih ontologija.

Sličnost ontologija (engl. *ontology similarity*) odnosi se na usporedbu cjelovitih ontologija ili njihovih elemenata. Ta usporedba za rezultat daje numeričku vrijednost koja pokazuje viši ili niži stupanj sličnosti između dvaju elemenata. Funkcija sličnosti pridružuje parove između skupova elemenata dviju ontologija O_1 i O_2 i izražava njihovu sličnost iskazanu kao numeričku vrijednost koja pripada segmentu $[0,1]$.

Problem pronalaženja sličnosti pojavljuje se u različitim domenama istraživanja i nije nužno vezan uz ontologije i povezivanje ontologija, a mjere sličnosti koje se koriste u tehnikama za poravnavanje ontologija preuzete su iz područja matematike, statistike i umjetne inteligencije (strojno učenje, dubinska analiza podataka i dr.). Osim toga, različiti su postupci uspoređivanja, poravnavanja i semantičke integracije na razini shema podataka općenito (Rahm, Bernstein, 2001; Doan, Halevy, 2005; Giunchiglia 2007; Bellahsene et al., 2011), a potom su preuzeti kao tehnike za poravnavanje ontologija.

Budući da ontologija predstavlja konceptualizaciju određene domene, uspoređivanje elemenata dviju ontologija zapravo predstavlja više od same reprezentacije tih elemenata na razini sintakse. Pored troga, važno je uzeti u obzir i njihov odnos prema elementima iz stvarnoga svijeta na koje referiraju, odnosno njihovo značenje kao i njihovu svrhu u stvarnom svijetu, njihovu upotrebu. Kako bi se ostvarila takva sveobuhvatna usporedba dviju ontologija, koristi se semiotički pristup ontologijama prema kojem se sličnost promatra na trima razinama: na razini podataka ili simboličkoj razini, na razini ontologije ili semantičkoj razini i na razini konteksta ili pragmatičkoj razini.

2.1 Sličnost na razini podataka

Na prvoj razini elementi se uspoređuju uzimajući u obzir samo podatkovne vrijednosti određenih jednostavnih ili kompleksnih tipova podataka, kao što su cijeli brojevi i nizovi. Za uspoređivanje podatkovnih vrijednosti mogu se koristiti opće funkcije sličnosti kao što je, primjerice, određivanje relativne udaljenosti među brojevima ili daljine uređivanja (engl. *edit distance*) između nizova. Složeni tipovi podataka sastavljeni od jednostavnih tipova podataka zahtijevaju i složenija mjerila koja se sastoje opet od jednostavnih mjerila.

2.2 Sličnost na razini ontologije

Na drugoj razini, razini ontologije, promatraju se semantički odnosi među elementima ontologije. Ova se razina prema semantičkoj složenosti može podijeliti na četiri razine (Ehrig, 2007): semantičke mreže, deskriptivske logike, ograničenja ili restrikcije i pravila. Na najnižoj razini

ontologije se promatraju samo kao grafovi⁵ s konceptima i odnosima. Ta se razina nadograđuje deskriptivnim logikama, ograničenjima ili pravilima na najvišoj razini. Međutim, razina pravila u semantičkom webu općenito još nije dovoljno istražena pa je ta razina slabije podržana u sustavima za poravnavanje ontologija.

2.2.1 Sličnost na razini konteksta

Na najvišoj razini razmatra se način na koji se elementi ontologija koriste u vanjskom kontekstu. Taj proces podrazumijeva korištenje informacija koje se nalaze izvan ontologija. Kontekst se promatra kao lokalni model koji kodira korisnikov subjektivni pogled na domenu. Iako postoje mnogi konteksti u kojima se neka ontologija može promatrati (npr. kontekst u kojem je ontologija razvijena ili kontekst u kojem je izmijenjena), s gledišta određivanja sličnosti najvažniji je aplikacijski kontekst, odnosno kako se specifičan element neke ontologije koristi u kontekstu dane aplikacije. Dakle, sličnost između dvaju elemenata ontologije može se odrediti usporedbom njihova korištenja u sklopu aplikacije temeljene na ontologijama. Pojednostavnjeno, slični elementi imaju slične uzorke korištenja, tj. slični elementi se koriste u sličnom kontekstu. Kod otkrivanja sličnosti to se pravilo razmatra u oba smjera: ukoliko se dva entiteta koriste u istom (sličnom) kontekstu, onda su ti elementi slični i obrnuto, ukoliko se u dva konteksta koriste isti (slični) elementi, onda su ta dva konteksta slična.

3. TEHNIKE PORAVNAVANJA ONTOLOGIJA

Tehnike poravnavanja ontologija (engl. *ontology aligning techniques*) nazivaju se još i tehnike uspoređivanja ontologija (engl. *ontology matching techniques*), a podrazumijevaju izračunavanje sličnosti između dva elementa ontologija. Definirane su različite tehnike, metode, algoritmi i postupci pronalaženja sličnosti između dva entiteta, odnosno koncepta.

Klasifikacija tehnika poravnavanja ontologija zasniva se na klasifikaciji tehnika uspoređivanja na shematskoj razini. Nekoliko je značajnih teorijskih pristupa koji razmatraju problem poravnavanja na shematskoj razini općenito i koji predstavljaju temelj danas aktualnih klasifikacija tehnika poravnavanja ontologija (Rahm, Bernstein, 2001; Wache, 2001; Kalfoglou and Schorlemmer, 2003). Aktualni radovi koji daju detaljan pregled različitih pristupa i klasifikacija tehnika poravnavanja ontologija jesu (Ehrig 2007; Euzenat, Shvaiko 2007; Shvaiko, Euzenat, 2012), a ostali radovi iz tog područja uglavnom citiraju navedene autore.

Ehrig analizira postojeće pristupe u teoriji poravnavanja te objašnjava algebarski pristup, pristup temeljen na toku informacija i okvire prevođenja. Nadalje, on opisuje dvije klasifikacije tehnika poravnavanja definirane prema dvije ortogonalne dimenzije. Horizontalna dimenzija odnosi se na razinu podataka, razinu ontologije i razinu konteksta. Vertikalna razina predstavlja znanje o domeni koje se može postaviti na bilo koju od horizontalnih dimenzija. Elementi horizontalne dimenzije podudaraju se s razinama sličnosti opisanim u drugom poglavlju, a tehnike koje pripadaju svakoj pojedinoj kategoriji podrazumijevaju izračunavanje sličnosti upravo za tu razinu podataka. Ehrig (2007) daje teorijski

⁵ Ontologija se može prikazati kao usmjereni aciklički graf (engl. *directed acyclic graph*, DAG)

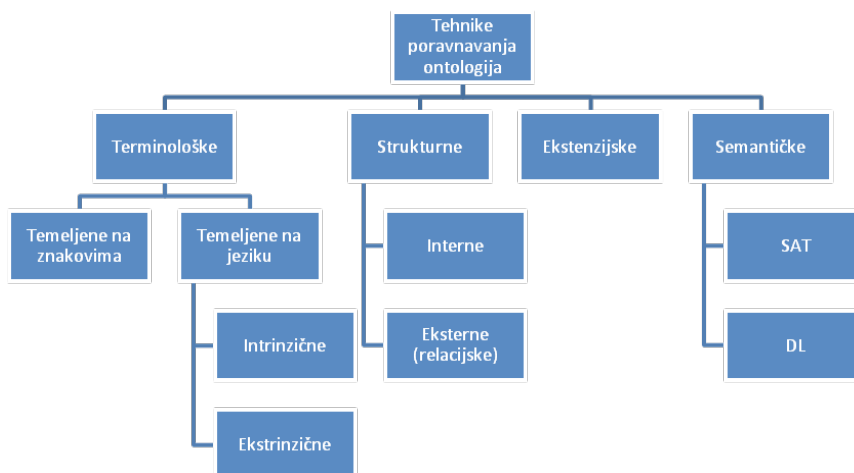
pregled tehnika, ali se ne opisuju konkretne tehnike, već se opisuju sustavi poravnavanja ontologija te se u okviru tih opisa spominju neke od tehnika poravnavanja.

Euzenat i Shvaiko detaljnije opisuju konkretne tehnike poravnavanja, a također navode i više mogućih klasifikacija tehnika poravnavanja, ovisno o različitim aspektima. Oni ističu tri bitne, međusobno nezavisne dimenzije klasifikacije: ulazni podaci, karakteristike procesa uspoređivanja, izlazni podaci. Zbog detaljnijeg prikaza tehnika poravnavanja, u iduća dva poglavlja opisane su dvije moguće klasifikacije tehnika poravnavanja preuzete većim dijelom iz radova Euzenata i Shvaika. Kroz taj opis dan je i detaljan pregled jedanaest konkretnih tehnika poravnavanja koje pokrivaju i tehnike implementirane u sustavima koje opisuju Euzenat (Euzenat, 2007), što je prikazano u četvrtom poglavlju.

3.1 Klasifikacija tehnika poravnavanja ontologija temeljena na vrsti ulaznih podataka

Po uzoru na klasifikaciju tehnika za povezivanje na shematskoj razini (Rahm, Bernstein, 2001), kao jednu od mogućih klasifikacija tehnika poravnavanja Euzenat i Shvaiko predlažu klasifikaciju temeljenu na vrsti ulaznih podataka (Euzenat, Shvaiko 2005; Euzenat, Shvaiko 2007). To je klasifikacija koja se (ponekad uz manje izmjene) navodi i citira u većini izvora koji se bave klasifikacijom tehnika i postupaka uspoređivanja (Castano et al., 2006; Giunchiglia et al., 2007; Ehrig 2007; Romero et al., 2009; Granitzer et al., 2010). Prema navedenoj klasifikaciji, osnovne skupine tehnika za poravnavanje ontologija na temelju sličnosti njihovih elemenata jesu terminološke tehnike ili tehnike temeljene na nazivima, strukturne tehnike, ekstenzijske tehnike i semantičke tehnike. Klasifikacija je prikazana na shemi 2, a potom je dan detaljan opis navedenih osnovnih skupina tehnika za poravnavanje ontologija.

Shema 2. Osnovna klasifikacija tehnika za poravnavanje ontologija



Izvor: Romero (2009)

3.1.1 Terminološke tehnike

Terminološke tehnike uspoređuju znakovne nizove, a mogu se primijeniti na imena, oznake ili komentare elemenata ontologije kako bi se pronašli slični elementi. Te se tehnike mogu koristiti za

uspoređivanje imena klasa kao i drugih elemenata ontologije. Razlikuju se dvije skupine tehnika: tehnike temeljene na znakovnim nizovima i tehnike temeljene na jeziku. Tehnike temeljene na jeziku mogu se dalje dijeliti na intrinzične i ekstrinzične.

Tehnike temeljene na znakovnim nizovima daju prednost strukturi nizova kao sekvenci znakova, odnosno slova, a temelje se na mjerama udaljenosti. Najpoznatija takva mjera je Levenshteinova udaljenost (Levenshtein, 1966) koja mjeri najmanji broj operacija (umetanje, brisanje, supstitucija) potrebnih da se jedan znakovni niz pretvori u drugi. Osim te mjere, koristi se još i Hammingova mjera (Hamming, 1950) te Jarova mjera (Jaro, 1989). Prije same usporedbe znakovnih nizova provode se postupci tzv. normalizacije, odnosno međusobne prilagodbe nizova prema nekom kriteriju (npr. zamjenjivanje svih velikih početnih slova malima, zamjenjivanje slova s dijakritičkim znakovima onima najslabijima i sl.).

Tehnike temeljene na znakovnim nizovima korisne su ako se koriste vrlo slični nizovi kako bi se označili isti koncepti i u tom će ih slučaju sustav prepoznati kao slične. Na primjer, vrlo lako se može uočiti sličnost među klasama *auto* i *automobil*. Međutim, neće se pokazati visok stupanj sličnosti između klasa *auto* i *vozilo*. Ukoliko se pojavljuju sinonimi, onda će među tim objektima biti detektiran vrlo nizak stupanj sličnosti. Osim toga, dva niza mogu biti vrlo slična, ali pritom označavati dva različita koncepta, npr. *hotel* i *hostel*, a isto vrijedi i za sve homonime. Također, problem se može javiti i kod korištenja riječi iz različitih jezika kako bi se označio isti element. Takve se varijacije mogu javiti unutar iste ontologije, ali su mnogo češće uspoređujemo li različite ontologije.

Intrinzične jezične tehnike ne uspoređuju samo izdvojene znakovne nizove, odnosno riječi, već dijelove teksta koji se mogu rastaviti na riječi (npr. *sredstvo prijevoza*, *prijevozno sredstvo*). Ovakve tehnike temelje se na procesiranju prirodnog jezika kako bi se iz teksta izdvojili termini sa značenjem. Primjer algoritma za takav oblik uspoređivanja dao je Porter (Porter, 1980). Uspoređivanje termina i njihovih odnosa pomaže u otkrivanju sličnosti između elemenata ontologija koji su definirani u obliku fraza. I kod ove tehnike potrebno je provesti postupke normalizacije koji se odnose na tri glavne vrste varijacija: morfološke (varijacije oblika riječi koja proizlazi iz istog korijena), sintaktičke (varijacije gramatičke strukture) i semantičke (varijacije u značenju, obično korištenjem nadređenog ili podređenog pojma). Kod primjene ovih tehnika elementi ontologija se reprezentiraju kao skupovi termina, a ne riječi.

Ekstrinzične jezične tehnike služe se vanjskim jezičnim izvorima. Različite vrste jezičnih izvora mogu se koristiti kako bi se ustanovila sličnost među pojmovima, npr. leksikoni, višejezični rječnici, semantičko-sintaktički leksikoni, glosari, terminologije. Ti izvori mogu biti definirani za jedan jezik ili biti specifično oblikovani za određenu domenu. Pomoću takvih izvora mogu se lakše riješiti problemi s pojavljivanjem sinonimima. Uzimanjem u obzir objašnjenja značenja riječi sadržanog u rječniku povećava se mogućnost nalaženja pravih kandidata za poravnanje. Slično je i s homonimima, hiponimima i hiperonimima.

3.1.2 Strukturne tehnike

Osim uspoređivanja naziva ili oznaka elemenata, u ontologijama se može uspoređivati i njihova struktura. Pritom se može uspoređivati unutarnja struktura elemenata, odnosno pored naziva i

anotacija pojedinih elemenata. Također, može se provesti i usporedba elemenata s drugim elementima s kojima je taj element povezan. U prvom slučaju radi se o internoj ili unutarnjoj strukturi, a u drugom o relacijskoj ili vanjskoj strukturi. Unutarnja struktura je definicija elementa bez odnosa prema drugim elementima, dok je relacijska struktura skup odnosa koje određeni element ostvaruje prema drugim elementima. Prvi je tip strukture istaknutiji kod uspoređivanja shema, dok je drugi tip važniji u uspoređivanju formalnih ontologija i semantičkih mreža. Prema dvama opisanim tipovima strukture mogu se razlikovati i dvije vrste tehnika: tehnike temeljene na unutrašnjoj strukturi i tehnike temeljene na vanjskoj strukturi.

Tehnike temeljene na unutarnjoj strukturi elemenata koriste kriterije kao što su skup svojstava tog elementa, doseg svojstava (atributi i relacije), njihova tranzitivnost ili simetrija i sl., kako bi se izračunao stupanj sličnosti među elementima. U dvjema ontologijama mogu se naći elementi s usporedivim unutarnjim strukturama ili svojstvima, tj. mnogi različiti tipovi objekata mogu imati svojstva s istim tipovima podatka. Stoga se ove tehnike češće koriste za kreiranje podudarnih grupa elemenata nego za otkrivanje podudarnosti među pojedinačnim elementima. Ove se tehnike često kombiniraju s terminološkim tehnikama.

Korištenjem tehnika temeljenih na strukturi odnosa među elementnima, ontologija se promatra kao graf, a traženje podudarnosti među elementima takvih grafova u vezi je s traženjem izomorfnih struktura grafa. Usporedba sličnosti između dva elementa iz dviju ontologija ovdje se temelji na odnosima koje ti elementi uspostavljaju prema drugim elementima u ontologijama, odnosno što su elementi sličniji, to sličniji trebaju biti i elementi koji su s njima povezani. Taksonomijska struktura, tj. graf oblikovan prema odnosima *subClassOf* je temeljna okosnica za prikazivanje ontologija kao grafova. Ovdje se može pretpostaviti da su određene klase slične ukoliko su i njihove nadređene ili podređene klase slične. Ako su nadređene klase iste, onda su i aktualne klase međusobno slične, a ako su podređene klase iste, onda su i uspoređivane klase također slične. Osim taksonomijske strukture, koristi se i mereološka struktura, tj. struktura koja odgovara odnosu cjeline i njezinog dijela.

3.1.3 Ekstenzijske tehnike

Ekstenzijske tehnike uspoređuju ekstenzije elemenata, koje mogu biti načinjene od drugih elemenata, npr. instanci, kao i povezanih izvora, npr. popisanih dokumenata. Te se tehnike razlikuju ovisno o tome dijele li dvije ontologije iste izvore ili ne dijele. Instance predstavljaju dobru priliku za uspoređivanje ontologija. Ukoliko dvije ontologije dijele isti skup instanci, njihovo poravnavanje je uvelike olakšano. Ako dvije klase dijele potpuno isti skup instanci, postoji velika vjerojatnost da te klase predstavljaju kandidate za poravnanje. Ekstenzijske tehnike koriste razne statističke metode, a dijelom su preuzete iz područja dubinske analize podataka i strojnog učenja. Analiza podataka i statistički pristupi koriste reprezentativni skup populacije kako bi se našle pravilnosti i neslaganja između ontologija. To su tehnike koje predlažu grupiranje elemenata i izračunavaju udaljenosti između njih. Tu spadaju tehnike kao što su izračunavanje distribucije frekvencija, naivni Bayesov algoritam, izračunavanje vektorskih udaljenosti, klasifikacije na temelju udaljenosti, formalna analiza koncepata i dr.

3.1.4 Semantičke tehnike

Semantičke tehnike temelje se na semantičkoj definiciji ontologija. Ključna je karakteristika semantičkih tehnika uključivanje deduktivnih tehnika. Takvim tehnikama treba prethoditi faza koja će pružiti

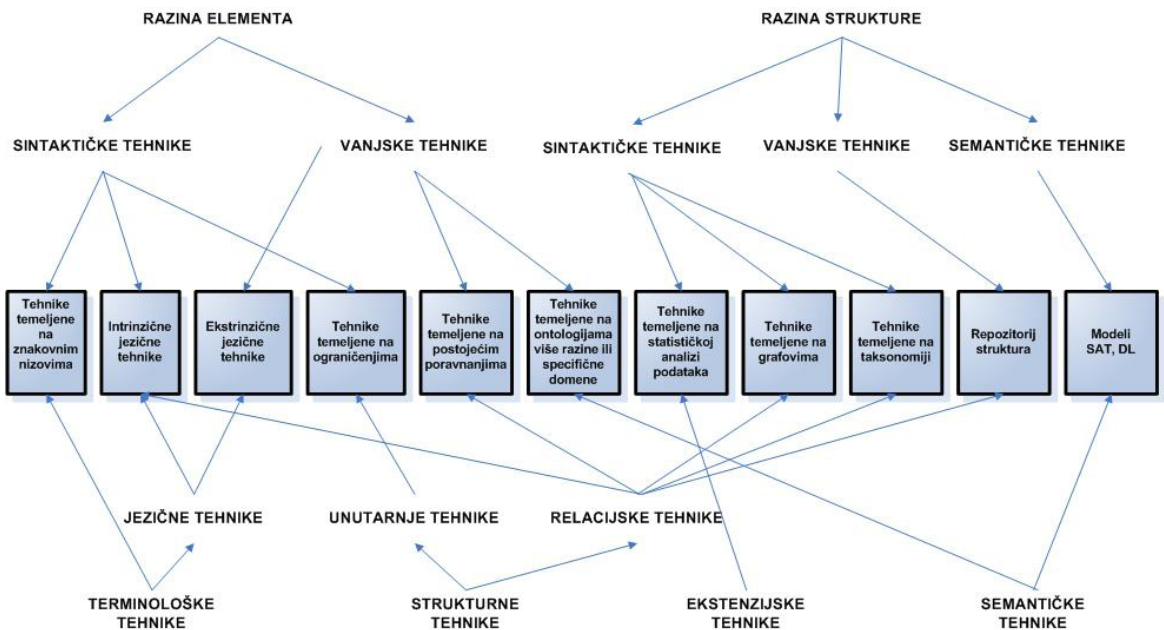
određene čvrste temelje, odnosno elemente za koje je utvrđeno da su, primjerice, ekvivalentni, temeljem usporedbe njihovih naziva ili neke druge tehnike. Semantičke tehnike funkcioniraju kao pojačivači tih korijenskih poravnanja. Zbog toga se u semantičke tehnike uključuju i pojedine tehnike za „učvršćivanje“ ontologija. One se temelje na korištenju postojećih resursa, odnosno eksternih formalnih ontologija za iniciranje poravnanja koja se zatim mogu dodatno razmatrati pomoću deduktivnih tehnika i metoda. Takve neposredne ontologije mogu definirati opći zajednički kontekst ili pozadinsko znanje za ontologije koje se uspoređuju, što omogućava razjašnjenje višestrukih mogućih značenja određenih pojmova. Razlikuju se dvije osnovne skupine tehnika: tehnike zadovoljivosti (*SAT techniques*) i tehnike bazirane na deskriptivnoj logici (*DL techniques*). Kako semantičke tehnike ne mogu same pronaći podudarnosti, jedan od važnih izazova pri korištenju tih tehnika je njihovo integriranje s induktivnim tehnikama.

U četvrtom poglavlju prikazana je zastupljenost navedenih skupina tehnika u implementiranim sustavima za poravnavanje ontologija. Pokazana je puno veća zastupljenost terminoloških i strukturnih tehnika u odnosu na ekstenzijske i semantičke tehnike, pa je to razlog zbog čega su navedene tehnike nešto detaljnije opisane u ovom poglavlju.

3.2 Klasifikacija tehnika poravnavanja ontologija temeljena na zrnatosti ulaznih podataka

Druga dimenzija klasifikacije odnosi se na zrnatost⁶ ulaznih podataka, a ovisno o tome primjenjuje li se tehnika na razini elemenata ili na razini strukture razlikuju se dvije skupine tehnika: tehnike temeljene na razni elemenata i tehnike temeljene na razini strukture. Pored podjele na razini elemenata i na razini strukture, klasifikacija prikazuje grupiranje prema tome jesu li sintaktičke, semantičke ili vanjske. Detaljni prikaz klasifikacije prikazan je na shemi 3 koja ujedno prikazuje detaljnu klasifikaciju tehnika poravnavanja s različitih aspekata.

Shema 3. Detaljna klasifikacija tehnika za poravnavanje ontologija



Izvor: Euzenat i Shvaiko (2007)

⁶ Granularnost

3.2.1 Tehnike temeljenje na razini elemenata

Tehnike temeljene na razini elemenata podrazumijevaju izračunavanje semantičke sličnosti između elemenata. Razlikuju se sintaktičke i vanjske tehnike, ovisno o tome uspoređuju li se elementi ontologije na razini sintakse ili se gledaju vanjski odnosi s drugim elementima.

Tehnike temeljene na razini elemenata klasificirane kao sintaktičke tehnike uključuju sve već opisane terminološke tehnike. Vanjske tehnike jesu tehnike temeljene na ograničenjima koje analiziraju ograničenja postavljena kao tipovi, brojnost atributa i ključevi.

U vanjske tehnike temeljene na razini elemenata ubrajaju se i tehnike temeljene na postojećim poravnanjima koje podrazumijevaju ponovno korištenje postojećih, prethodno definiranih poravnanja.

Još jedna skupina tehnika iz te kategorije jesu tehnike temeljene na ontologijama više razine ili specifične domene. To su tehnike koje omogućavaju korištenje postojećih općih ontologija kao što su Cyc, SUMO, DOLCE, FMA i sl. za izračunavanje sličnosti.

3.2.2 Tehnike temeljenje na razini strukture

Tehnike temeljenih na strukturi izračunavaju semantičku sličnost analizirajući odnose koje entiteti ili njihove instance definiraju unutar strukture.

Tu se ubrajaju tehnike uspoređivanja temeljene na grafovima koje se odnose na sinaktičke tehnike. Tehnike temeljene na grafovima podrazumijevaju algoritme koji promatraju ulaznu ontologiju kao označeni graf. Određivanje sličnosti za dva čvora zasniva se na analizi njihovih pozicija unutar grafa. Rješavanje problema uspoređivanja na razini strukture primjenom grafičkih tehnika rješava se pronalaženjem homomorfizama grafova, odnosno pronalaženjem najvećeg zajedničkog usmjerenog podgraфа.

Tehnike temeljene na taksonomiji također spadaju u sintaktičke tehnike, a odnose se na traženje veza *a-kind-of* i *is-a*⁷ oblika kao što je prethodno navedeno u poglavlju 3.1.2.

Tehnike zasnivane na repozitoriju struktura jesu vanjske tehnike temeljene na strukturi, a funkcioniraju tako da se čuvaju uparene ontologije i dijelovi ontologija zajedno s njihovim sličnostima. Prilikom izračunavanja sličnosti dvaju elemenata prvo se provjerava postoji li već izračunata sličnost pohranjena u repozitoriju. Time je ukupno izračunavanje brže.

Tehnike temeljene na modelima obuhvaćaju skupinu semantičkih tehnika koja je opisana u poglavlju 3.1.4.

Tehnike temeljene na analizi podataka i statističkim metodama opisane su u prethodnom poglavlju kao ekstenzijske tehnike, a prema drugoj klasifikaciji pripadaju skupini sintaktičkih tehnika. Primjenjuju se za rezoniranje s velikim količinama podataka, u slučaju kada postoji reprezentativni uzorak populacije.

⁷ Standardni termini koji označavaju relacije biti podklasa (*subClassOf*) ili biti element klase preuzeti iz područja semantičkih mreža

3.3 Kombiniranje tehnika za poravnavanje ontologija

Tehnike opisane u prethodnom poglavlju mogu se na različite načine kombinirati sa ciljem dobivanja konačnog rezultata poravnavanja. Sukladno tome definirani su različiti pristupi i strategije za poravnavanje ontologija iz čega proizlaze različiti postupci poravnavanja ontologija.

Uz tehnike poravnavanja postoji još nekoliko elemenata bitnih za implementaciju postupka poravnavanja i dobivanje konačnog rezultata. Jedan od aspekata kombiniranja tehnika poravnavanja je njihovo međusobno kombiniranje na razini izračunavanja sličnosti dvaju elemenata, drugi aspekt je kombiniranje izračunatih sličnosti na elementarnoj razini kako bi se dobio rezultat konačnog poravnavanja, a treći bitan aspekt je kombiniranje automatiziranih tehnika s intervencijom korisnika. U ovom poglavlju ukratko su prikazane različite mogućnosti kombiniranja tehnika.

Sličnost dva elementa ontologija izračunava se na elementarnoj razini primjenom samo jedne od navedenih tehnika, a potom je potrebno izvesti složeniji rezultat koji se temelji na kombiniranju više različitih tehnika. Taj postupak naziva se kompozicija ili agregacija tehnika poravnavanja. Postoje različite formule (Euzenat i Shvaiko, 2007; Ehrig, 2007) definirane za izračunavanje agregacije sličnosti na temelju danih heterogenih sličnosti. Dva osnovna pristupa za kombiniranje različitih tehnika poravnavanja, odnosno uspoređivanja, jesu sekvencijalna kompozicija i paralelna kompozicija. Kod sekvencijalne kompozicije različite tehnike poravnavanja izvode se sekvencijalno, jedna za drugom. U tom slučaju ima smisla prvo primijeniti tehnike koje se zasnivaju na znakovnim nivoima, potom tehnike temeljene na strukturi i na kraju tehnike temeljene na semantici. Drugi pristup kombiniranja tehnika poravnavanja je paralelna kompozicija, odnosno izvođenje nekoliko tehnika paralelno, a potom se odabire jedan od rezultata prema nekom kriteriju (najmanja sličnost ili najveća sličnost) ili se objedinjuju sve dobivene vrijednosti sličnosti i izračunava konačna sličnost prema nekoj definiranoj formuli.

Budući da se sličnost može definirati na razini cijelih ontologija, potrebno je definirati postupak za izračunavanje globalne sličnosti. Postupci za izračunavanje globalne sličnosti temelje se na algoritmima za pronalaženje sličnosti između grafova.

Pored automatiziranih kombinacija tehnika poravnavanja, važnu ulogu u postupcima poravnavanja ima i korisnik, odnosno inženjer znanja (ekspert). Korisnik može na više načina sudjelovati u postupku poravnavanja. Jedan od načina je njegovo sudjelovanje u prethodno opisanoj agregaciji sličnosti u smislu da se postavi dinamičko agregiranje sličnosti tako da korisnik određuje koje od tehnika će na koji način kombinirati. U tom slučaju korisnik mora jako dobro poznavati sustav i cjelokupni tijek poravnavanja. Drugi način intervencije korisnika je postavljanje početnih vrijednosti poravnavanja čime je definirano početno stanje na kojem će se temeljiti daljnji izračuni. I treće, korisnikova uloga je bitna na kraju postupka kada on daje povratnu informaciju o rezultatima postupka poravnavanja. Nakon korisnikove povratne informacije mijenjaju se parametri tehnika uspoređivanja i u iterativnom postupku generira se novi rezultat.

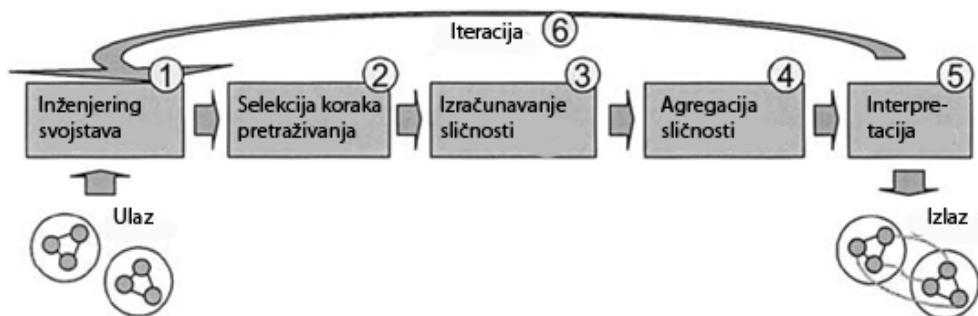
4. IMPLEMENTACIJA POSTUPKA PORAVNAVANJA

U ovom poglavlju opisan je opći postupak poravnavanja kao proces koji se odvija u šest faza (Ehrig, 2007). Nakon toga dan je pregled značajnijih sustava koji u nekom svom segmentu rješavaju problem poravnavanja ontologija.

4.1 Proces poravnavanja ontologija

Opći proces poravnavanja ontologija prikazan je na shemi 3, a obuhvaća ulaznu i izlaznu točku te šest glavnih koraka tog procesa (Ehrig, 2007; de Bruijn et al., 2006) opisanih u nastavku poglavlja. Kada je riječ o implementaciji procesa, ona varira od sustava do sustava, ali ovdje je prikazan opći postupak koji se nadovezuje na prethodno definirane teorijske pristupe.

Shema 4. Proces poravnavanja ontologija



Izvor: Ehrig (2007)

4.1.1 Ulaz

Ulaz čine dvije ili više ontologija koje trebaju biti međusobno poravnane. Kao što je navedeno u poglavlju 3.3, moguće je unijeti i prethodno poznata, ručno formirana poravnanja koja mogu pomoći u procesu traženja novih poravnanja.

4.1.2 Inženjering svojstava

U ovom, početnom, koraku selektiraju se mali isječci cjelokupne ontologije kako bi se opisao specifičan element ontologije. Ti isječci nisu proizvoljni konstrukti nego imaju specifično značenje unutar ontologije. U sljedećem koraku ta će se svojstva koristiti za usporedbu. Primjerice, proces poravnavanja može se temeljiti samo na podskupu osnovnih funkcija OWL-a, po mogućnosti samo na taksonomiji, ili čak samo na lingvističkim opisima elemenata (npr. oznaka „car“ - automobil za opisivanje koncepta o1:car).

4.1.3 Selekcija koraka pretraživanja

Izvođenje podudaranosti, odnosno poravnavanja među ontologinjama, odvija se u prostoru pretraživanja među odabranim kandidatima za poravnanje. U ovom koraku biraju se određeni kandidati, tj. parovi koncepata među kojima će se provesti izračunavanje stupnja sličnosti, dok se odbacuju ostali koncepti (npr. koncept *o1:car* uspoređuje se samo s konceptom *o2:automobile*, ali ne i sa *o2:hasMotor*).

4.1.4 Izračunavanje sličnosti

U ovom koraku izračunava se sličnost za pojedina dva elementa među kandidatima za poravnanje, na primjer:

$$\text{simlabel}(o1:car, o2:automobile) = \text{sim syntactic}(\text{„car“}, \text{„automobile“}) = 0$$

Kao rezultat uspoređivanja i izračunavanja sličnosti u ovom koraku daje se numerička vrijednost 0 ili 1, koja izražava stupanj sličnosti među uspoređivanim kandidatima. Ovdje stupanj sintaktičke sličnosti među izrazima *car* i *automobile* iznosi 0, odnosno ti su izrazi sintaktički potpuno različiti.

4.1.5 Agregacija sličnosti

Kao što je opisano u poglavlju 3.3, za neki par elemenata koji predstavljaju kandidate za poravnanje, općenito može postojati više vrijednosti koje iskazuju njihovu sličnost. Na primjer, jedna vrijednost iskazuje stupanj sličnosti između njihovih oznaka, druga vrijednost iskazuje sličnost njihovih odnosa prema drugim elementima, a treća vrijednost može iskazivati neki stupanj sličnosti dobiven na razini instanci. U ovom koraku primjenjuje se neki od pristupa komponiranja ili agregacije sličnosti (paralelna kompozicija, sekvencijalna kompozicija).

Primjer primjene paralelne kompozicije za izračunavanje kumulativne vrijednosti sličnosti kao računanje prosječne vrijednosti za tri različite vrijednosti sličnosti dobivene različitim tehnikama:

$$\begin{aligned} &(\text{simlabel}(o1:car, o2:automobile) + \\ &\text{simsubconcepts}(o1:car, o2:automobile) + \\ &\text{siminstances}(o1:car, o2:automobile)) / 3 = 0.5. \end{aligned}$$

4.1.6 Interpretacija

U fazi interpretacije naposljetku se koriste individualne ili kumulativne vrijednosti sličnosti kako bi se izvela poravnanja među elementima ontologije. Sličnosti se trebaju interpretirati, a mehanizmi koji se obično koriste u tu svrhu su korištenje granica (engl. *thresholds*) ili kombiniranje strukturalnih kriterija i kriterija sličnosti. Na kraju se dobije rješenje za poravnanje za odabrani par elemenata ukoliko se prethodno pokazala visoka vrijednost sličnosti među njima. Međutim, može se dogoditi da nisu pronađena valjana poravnanja među određenim elementima ontologije. U tom koraku potrebna je intervencija korisnika kako bi se odredile granice i odgovarajuća poravnanja.

4.1.7 Iteracija

Sličnost jednog para elemenata utječe na sličnost susjednog para elemenata, npr. ukoliko su neke instance jednake, to utječe na sličnost koncepata i obrnuto. Dakle, sličnost se prenosi kroz ontologije slijedom veza u ontologiji. Iteracija se provodi pomoću određenih algoritama, a u tom procesu preračunavaju se sličnosti kandidata za poravnanje na temelju sličnosti susjednih parova elemenata. To može dovesti do nove sličnosti, na primjer:

$$\text{sim}(o1:\text{car}, o2:\text{automobile}) = 0.85,$$

U konačnici će to rezultirati novim poravnanjem. Proces iteracije završava kada se više ne nalaze nova poravnanja.

4.1.8 Izlaz

Izlaz cjelokupnog prethodno opisanog procesa je prikaz poravnanja među elementima dviju ili više ontologija, npr. tablica s poravnanjima koja prikazuje odnos $\text{align}(O1,O2)$ te po mogućnosti i vrijednosti koje prikazuju stupanj podudarnosti među njima. Poravnanja među elementima mogu biti prikazana vizualno ukoliko alat u kojem je cjelokupni proces implementiran to omogućuje.

4.2 Pregled sustava za poravnanje ontologija

Rezultati različitih postupaka i tehnika poravnavanja ontologija jesu različite implementacije cjelokupnog procesa poravnavanja. U ovom poglavlju dan je kratki pregled značajnijih sustava koji uključuju postupak poravnavanja ontologija. Cilj prikaza sustava je prvenstveno prikazati zastupljenost određenih skupina tehnologija. U poglavlju se ne navode drugi tehnički detalji sustava, oni se mogu naći u literaturi (Ehrig 2007; Shvaiko, Euzenat, 2007; Shvaiko, Euzenat, 2012).

Pregled sustava dan je u Tablici 1. Posebno se navodi oblik ulaznih podataka koji može biti općenit oblik ontologije ili ontologija u nekom konkretnom jeziku ili shema ontologija. Ovisno o obliku ulaznih podataka, može se detektirati jesu li implementirane tehnike prema znatosti temeljene na razini podataka ili na razini strukture. Osim toga, navodi se kratki opis sustava s posebnim naglaskom na skupinu tehnika koja je implementirana (tehnike su kategorizirane prema klasifikaciji temeljenoj na ulaznim podacima). Uz naziv sustava dana je i godina u kojoj je uvedena prva verzija sustava ili kada se pojavljuju znanstveni radovi koji opisuju sustav. Neki od sustava unaprijeđeni su tijekom godina i dobili su nove verzije kao nasljednike polaznog sustava.

Prva skupina prikazanih sustava (prvih trinaest sustava) spadaju u skupinu starijih sustava o kojima ima puno podataka u dostupnoj literaturi, ali ne razvijaju se više tako intenzivno. Posljednjih sedam sustava (SAMBO, Falcon-AO, Dssim, RiMOM, ASMOV, Anchor-Flood, Agreement/Maker) aktivno se razvijaju u posljednjih pet godina i postižu najbolje rezultate na mjerenjima i evaluaciji tehnika poravnavanja koja se odvijaju u okviru OAEI (Ontology Alignment Evaluation Initiative) inicijative (Euzenat 2007; Euzenat et al., 2008; Euzenat et al., 2009; Euzenat et al., 2011; Shvaiko, Euzenat, 2012). Noviji sustavi su napredniji u smislu korištenja više različitih tehnologija poravnavanja i puno bolje rade s velikim količinama podataka (engl. *large-scale data*)

Tablica 1. Sustavi za poravnavanje ontologija

Naziv sustava	Oblik ulaznih podataka	Kratki opis sustava i implementirane tehnike
ONION (2000)	ontologije	terminološke tehnike; strukturne tehnike uključena korisnička validacija prijedloga poravnavanja koriste artikulacijska pravila temeljena na heuristici
PROMPT (2000)	RDF(S) ontologije	poluautomatski pristup temeljen na jednakostima oznaka; terminološke tehnike
Anchor – PROMPT (2001)	RDF(S) ontologije	unaprijedena verzija PROMPT algoritma; integriran u Protégé alat za razvoj ontologija terminološke tehnike; strukturne tehnike
Chimaera (2000)	OKBC	interaktivni alat; terminološke tehnike
FCA-Merge (2001)	Ontologije ontologije	poluautomatizirani pristup; temelji se na matematičkoj metodi formalne analize koncepata; terminološke tehnike; strukturne tehnike
GLUE (2003)	sheme/ontologije	višestrategijski pristup; temelji se na naučenim parovima; koristi statističke metode strukturne tehnike;
OLA (2004)	OWL Lite	terminološke tehnike; strukturne tehnike
SemInt (1994)	Sheme ontologija	koristi katalog baza i instanci; koristi neuronske mreže
DIKE (2000)	Sheme oblika entiteti-veze	automatski pristup; izračunavanje sličnosti na razini objekata i na razini atributa; strukturne tehnike
ARTEMIS (2001)	Sheme ontologija	terminološke tehnike; ekstrinznine jezične tehnike (WordNet); strukturalne tehnike
Cupid (2001)	Relacijske sheme, XML	terminološke tehnike, strukturne tehnike
COMA++ (2005)	Sheme	koristi različite kombinacije strategija terminološke tehnike (jezične tehnike, rječnici) strukturne tehnike
S-Match (2004)	Sheme	semantičke tehnike (SAT solver); terminološke tehnike (izračunavanje sličnosti, korištenje rječnika)
SAMBO (2006)	OWL	ograničen na ontologije za domenu biotehnologije terminološke tehnike (n-gram; edit-distance; UMLS; WordNet) strukturne tehnike (iterativno izračunavanje strukturne sličnosti temeljene na is-a i part-of relacijama) ekstenzijske tehnike (Bayesov naivni algoritam)
Falcon – AO (2005)	RDF(S), OWL	podijeli pa savladaj pristup terminološke tehnike; strukturne tehnike; ekstenzijske tehnike
DSim (2005)	OWL, SKOS	okvir za povezivanje ontologija temeljen na agentima (engl. agent-based) terminološke tehnike; strukturne tehnike; semantičke tehnike
RiMOM (2006)	OWL	terminološke tehnike; strukturne tehnike; ekstenzijske tehnike
ASMOV (2009)	OWL	ograničen na ontologije za domenu biotehnologije terminološke tehnike; strukturne tehnike; ekstenzijske tehnike; semantičke tehnike
Anchor-Flood (2009)	RDF(S), OWL	učinkovito radi s velikim ontologijama (large ontologies) terminološke tehnike; strukturne tehnike
AgreementMaker (2009)	XML, RDF(S), OWL,N3	terminološke tehnike; strukturne tehnike

Izvori: Ehrig (2007); Shvaiko, Euzenat (2007); Shvaiko, Euzenat (2012)

5. ZAKLJUČAK

U radu je dan pregled tehnika i postupaka poravnavanja ontologija. Na teorijskoj razini objašnjeni su različiti mogući aspekti klasifikacije tehnika poravnavanja. Prikazane su dvije osnovne klasifikacije prema kojima se tehnike poravnavanja grupiraju u nekoliko različitih skupina ovisno o tipu ulaznih podataka (terminološke tehnike, strukturne tehnike, ekstenzijske tehnike, semantičke tehnike) i ovisno o zrnatosti ulaznih podataka (tehnike temeljene na podacima, tehnike temeljene na strukturi). Klasifikacija tehnika poravnavanja i teorijska razmatranja postupaka poravnavanja čine temelje za razvoj područja istraživanja koje nije vezano samo uz područje ontologija već i baza podataka i inženjeringa znanja općenito. Za realizaciju konkretnog postupka poravnavanja bitno je definirati tehnike poravnavanja te formalizirati cjelokupni proces poravnavanja. U radu je opisano jedanaest konkretnih tehnika koje se koriste za poravnavanje ontologija i navedeni su mogući pristupi za njihovo kombiniranje. Također, prikazan je opći proces poravnavanja koji se izvodi u šest faza. Osim toga, dan je pregled značajnih sustava koji imaju implementirane različite postupke poravnavanja ontologija i koriste različite skupine tehnika.

Iz danog pregleda tehnika i cjelokupnog postupka može se zaključiti da je postupak poravnavanja ontologija složen i zahtjevan za realizaciju jer podrazumijeva kombiniranje različitih tehnika, pristupa i strategija. Same tehnike poravnavanja na najnižoj razini određuju sličnost između dva elementa dviju različitih ontologija. Mjera sličnosti može se na različite načine definirati, a uglavnom je potrebno kombinirati više različitih mjera sličnosti kako bi se dobio što precizniji rezultat. U radu su opisani mogući pristupi koji definiraju različite načine agregiranja postojećih tehnika. Sekvencijalno i paralelno komponiranje, odnosno agregacija predstavljaju dva osnovna pristupa u okviru kojih su dalje moguće različite strategije koje definiraju na koji način se izračunava konačna mjera sličnosti iz više mjera sličnosti dobivenih primjenom različitih tehnika poravnavanja.

Također, još jedan od zaključaka koji se nameće jest kako cjelokupni postupak nije moguće u potpunosti automatizirati. Nužna je intervencija korisnika, odnosno inženjera znanja koji provodi i nadgleda postupak. Korisnik može sudjelovati postavljanjem početnih poravnavanja, dinamičkim definiranjem kompozicije tehnika ili za evaluaciju automatski generiranih sličnosti u zadnjem koraku izračunavanja. Uloga korisnika je vrlo značajna, a trenutno aktualna istraživanja kreću se u smjeru uvođenja kolaborativnog pristupa za evaluaciju poravnanja (Sarasua et al, 2012; Shi et al, 2009) što predstavlja jedan od mogućih načina unapređivanja postojećih postupaka.

Drugi mogući pristup unapređivanja postupaka poravnavanja je bolje iskorištavanje vanjskih izvora kao što su rječnici, leksikoni, tezauri te druge postojeće ontologije. Potrebno je istražiti mogućnosti implementacije potpunijih rječnika koji će imati mogućnost automatiziranog ažuriranja s novim podacima kao što su, na primjer, skupovi sinonima (eng. *synsets*). Cilj je automatizirati postupak poravnavanja ontologija u što većoj mjeri. Također, analizom primjena različitih skupina tehnika u pojedinim sustavima, pokazala se još uvijek slabija zastupljenost ekstenzijskih i semantičkih tehnika, pa tu također ima prostora za unapređivanje postojećih tehnika.

Temeljni zaključak koji se može izvesti iz danih razmatranja u ovome radu je taj da je unapređenja postupaka poravnavanja ontologija moguće. Štoviše, nova istraživanja mogu se razvijati u različitim smjerovima. S jedne strane, poboljšanje postupaka poravnavanja moguće je realizirati uvođenjem novih pristupa u kombiniranju automatskih tehnika s intervencijom korisnika. S druge strane, moguće je unaprijediti i bolje iskoristiti postojeće tehnike i razviti nove automatske postupke poravnavanja ontologija.

LITERATURA

- Bittner, T., Donnelly, M., Winter, S. (2004) „Ontology and Semantic Interoperability“, U: D. Prospero & S. Zlatanova (eds.), *Large-Scale 3D Data Integration*
- Bellahsene, Z., Bonifati, A., Rahm, E. (2011) *Schema Matching and Mapping*, Springer
- Castano, S., Ferrara, A., Montanelli, S. (2006) „Matching ontologies in open networked systems: Techniques and applications“, *Journal on Data Semantics V*, Springer Berlin Heidelberg, p. 25-63
- Davies, J., Studer, R., Warren, P. (ur.) (2006) *Semantic Web Technologies – Trends and Research in Ontology-based Systems*, Wiley
- De Bruijn, J. et al. (2006) „Ontology Mediation, Merging, and Aligning“. U: Davies, J.; Studer, R.; Warren, P. (ur.): *Semantic Web Technologies – Trends and Research in Ontology-based Systems*, Wiley, p. 95. – 113
- Doan, A., Halevy, A. (2005) „Semantic integration research in the database community: A brief survey“, *AI Magazine, Special issue on Semantic integration*, 26(1):83–94, p. 71, 153
- Ehrig, M. (2007) *Ontology Alignment – Bridging the Semantic Gap*, Berlin: Springer
- Euzenat, J. (2007) „Semantic precision and recall for ontology alignment evaluation“, In Proceedings of IJCAI, p. 348-353
- Euzenat, J., Mocan, A., Scharffe, F. (2008) „Ontology Alignments: An Ontology Management Perspective“, U: Hepp, M. et al. (ur.), *Ontology Management – Semantic Web, Semantic Web Services, and Business Applications*, Berlin: Springer, p. 177. – 206
- Euzenat, J., Shvaiko, P. (2007) *Ontology Matching*, Berlin: Springer
- Euzenat, J. et al. (2010) Results of the ontology alignment evaluation initiative 2009.
- Euzenat, J. et al. (2011) „Ontology Alignment Evaluation Initiative: Six Years of Experience“. U: Spaccapietra, S. (ur.), *Journal on Data Semantics XV*, Berlin: Springer, p. 158. – 192.
- Fensel, D. (2001) *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, Berlin: Springer
- Gargouri, F., Jaziri, W. (2010) *Ontology Theory, Management and Design: Advanced Tools and Models*, New York: Information Science Reference
- Giunchiglia, F., Yatskevich, M., Shvaiko, P. (2007) „Semantic matching: Algorithms and implementation“, In *Journal on Data Semantics IX*, Springer Berlin Heidelberg, p. 1-38
- Gómez-Pérez, A., Fernández-López, M., Corcho, O. (2004) *Ontological Engineering*, Berlin: Springer
- Granitzer, M. et al. (2010) „Ontology Alignment—A Survey with Focus on Visually Supported Semi-Automatic Techniques“, *Future Internet 2010*, Vol. 2, p. 238–258
- Hamming, R. W. (1950) „Error Detecting and Error Correcting Codes“. *Bell System Technical Journal*, 26(2): p. 147-160
- Hepp, M. (2008) „Ontologies: State of the Art, Business Potential, and Grand Challenges“. U: Hepp, M. et al. (ur.), *Ontology Management – Semantic Web, Semantic Web Services, and Business Applications*, Berlin: Springer, p. 3–22
- Jain, P. et al. (2010) „Ontology Alignment for Linked Open Data“. U: Patel-Schneider, P. F. et al. (ur.), *ISWC 2010, Part 1*, Berlin: Springer, p. 402-417

- Jaro, M. A. (1989) „Advances in record linking methodology as applied to the 1985 census of Tampa Florida“. *Journal of the American Statistical Society*, 64, p. 1183-1210
- Kalfoglou, Y., & Schorlemmer, M. (2003) „Ontology mapping: the state of the art“, *The knowledge engineering review*, 18(1), p. 1-31
- Levenshtein, V. I. (1966) „Binary codes capable of correcting deletions, insertions, and reversals“, *Soviet Physics Doklady*, 10, p. 707–710
- Lim, E., Liu, J., Lee, R. (2011) *Knowledge Seeker – Ontology Modelling for Information Search and Management*, Springer
- Porter, M. F. (1980) „An Algorithm for Suffix Stripping“. *Program*, 14(3), p. 130-137
- Rahm, E., Bernstein, P. (2001) „A Survey of Approaches to Automatic Schema Matching“. *VLDB Journal*, 10(4), p. 334-350
- Romero, M. M., Vázquez-Naya, J. M. (2009) „Ontology Alignment Techniques“. *Encyclopedia of Artificial Intelligence 2009*, p. 1290-1295
- Sarasua, C., Simperl, E., Noy, N. F. (2012) „CrowdMap: crowdsourcing ontology alignment with microtasks“ *In The Semantic Web–ISWC 2012*, p. 525-541
- Shi, F., Li, J., Tang, J., Xie, G.T., Li, H. (2009) „Actively learning ontology matching via user interaction“, In: *Proceedings of the 8th International Semantic Web Conference ISWC 2009*, p. 585–600
- Shvaiko, P., Euzenat, J. (2005). „A survey of schema-based matching approaches“, *Journal on Data Semantics*, IV, p.146–171
- Shvaiko, P., Euzenat, J. (2012) „Ontology matching: state of the art and future challenges“
- Wache, H. et al. (2001), „Ontology-based integration of information-a survey of existing approaches“, In *IJCAI-01 workshop: ontologies and information sharing*, Vol. 2001, p. 108-117

OVERVIEW OF ONTOLOGY ALIGNMENT TECHNIQUES AND METHODS³

ABSTRACT

Ontologies enable knowledge representation in a formal and in a machine-readable way. Numerous ontologies have been developed in recent years. Linking these ontologies semantically is an essential requirement in order to establish interoperability between agents, services or users working with them. Consequently, ontology alignment becomes a central issue, when building a world-wide Semantic Web. It would be too demanding to do it manually because of certain complexity, size and number of ontologies. Automatic or at least semi-automatic techniques have to be developed to reduce the burden of manual creation and maintenance of alignments. The main goal of this paper is to give an overview of ontology alignment techniques and to show how these techniques can be combined in the ontology alignment process. Different approaches in the ontology alignment techniques classification are represented. Furthermore, basic phases of ontology alignment process are described. An overview of ontology alignment and matching systems together with the alignment techniques implementation are represented. The analysis of existing techniques and their implementations lead to a conclusion that there is still room for improvement of the alignment process.

Key words: ontology alignment, ontology matching, semantic similarity, ontology alignment techniques, ontology alignment systems

¹ Mag. inf., E-mail: kjuricic@gmail.com

² PhD, Assistant Professor, Department of Informatics, University of Rijeka, Radmile Matejčić 2, Rijeka, Croatia. E-mail: amestrovic@inf.uniri.hr

³ Received: 7. 2. 2013; accepted: 3. 4. 2013

