

Autorska ljuska inteligentnog tutorskog sustava temeljena na prirodnom jeziku

Branko Žitko

Fakultet prirodoslovno-matematičkih znanosti i kineziologije

Teslina 12, 21000 Split, Hrvatska

branko.zitko@pmfst.hr

Sažetak - Inteligentni tutorski sustavi (ITS) i njihove autorske ljuske spadaju u kategoriju sustava e-učenja koji služe prilikom učenja i poučavanja. Potreba za učinkovitom komunikacijom između korisnika i ovakvih sustava definiraju novu potkategoriju ITS-ova koji se temelje na prirodnom jeziku. Računalni tutor koji je u stanju komunicirati s učenikom na prirodnom jeziku je sve sličniji ljudskom tutoru čime se olakšava učenje i poučavanje. U ovom radu je opisana autorska ljuska LASIT kod koje se generiranjem prirodnog jezika prikazuje nastavni sadržaj, a dijalogom testira znanje učenika.

Ključne riječi - inteligentni tutorski sustav, autorska ljuska, obrada prirodnog jezika, ontologija, tutorski dijalog

I. UVOD

Vrsta sustava e-učenja koji pokazuju prilagodljivost učeniku tijekom procesa učenja, poučavanja i testiranja znanja i vještina, su svakako inteligentni tutorski sustavi (ITS). Princip takvih sustava temelji se na ideji učenja "jedan naprama jedan", odnosno tutorskog učenja i poučavanja. ITS je zasnovan na znanju [1] i to na tri grupe znanja: (i) područno znanje je znanje kojim se ističe određeno područje koje je objekt učenja, (ii) pedagoško znanje je znanje o podučavanju, koje se preko nastavnog sadržaja vezuje za područno znanje, (iii) model studenta prati stanje usvojenog znanja i vještina učenika.

Navedene grupe znanja koriste tri modula sustava: (i) modul stručnjaka, (ii) modul učitelja (iii) modul učenika. Ova tri modula su povezana s modulom komunikacije [2] koji predstavlja sučelje prema učeniku, odnosno ostvaruje okolinu učenika. ITS-ovi su često usko specijalizirani za neko područno znanje ili nemaju autorska okruženja pomoću kojih će se oblikovati područno znanje ili pedagoško znanje. Autorske ljuske za ITS-ove [3] spadaju u kategoriju sustava za e-učenje i njihova značajka je pružanje autorskog okruženja stručnjaku za oblikovanje područnog znanja, autorskog okruženja učenika za oblikovanje nastavnog sadržaja i okruženja učeniku koje je zapravo sam ITS [4].

Učenik kao ciljni sudionik ITS-a uči se i poučava nad prethodno oblikovanom nastavnim sadržajem. Učitelj u

svom autorskom okruženju oblikuje nastavni sadržaj, što uključuje oblikovanje objekata nastavnog sadržaja koji se temelje na prethodno oblikovanom područnom znanju. Stručnjak je vrsta korisnika autorske ljuske ITS-a koji oblikuje područno znanje.

Ovakav postupak oblikovanja područnog znanja i nastavnog sadržaja koristi autorska ljuska zasnovana na modelu sustava Tutor-Expert System (TEx-Sys). [5] Područno znanje u modelu TEx-Sys je prikazano semantičkim mrežama s okvirima, dok je ideja prikaza nastavnog sadržaja temeljena na Sharable Content Object Reference Model (SCORM) [6]. Obilježje ovakvog prikaza nastavnog sadržaja ujedinjuje organizaciju nastavnog sadržaja temeljenu na SCORM-u, dok se objekti nastavnog sadržaja referenciraju na elemente područnog znanja, odnosno na čvorove semantičke mreže. Dvije su vrste objekata nastavnog sadržaja u modelu TEx-Sys koji se dijele po svojoj namjeni:

- 1) objekti nastavnog sadržaja namijenjeni učenju služe za prikazivanje znanja kao povezanih pojmova i
- 2) objekti nastavnog sadržaja namijenjeni testiranju dinamički generiraju pitanja koja se temelje na područnom znanju referenciranom u objektima nastavnog sadržaja namijenjenih učenju koji se nalaze u istoj hijerarhijskoj strukturi gdje i sam objekt nastavnog sadržaja namijenjen testiranju.

Proces učenja i poučavanja učenik započinje slobodnim izborom nekog objekta nastavnog sadržaja koji se hijerarhijski pronalazi kao list u stablu nastavnog sadržaja. Ako je učenik odabrao objekt nastavnog sadržaja namijenjenog učenju tada sustav prikazuje listu čvorova semantičke mreže koje taj objekt referencira u područnom znanju. Učeničkim izborom čvora sustav prikazuje veze s ostalim čvorovima te njegov okvir. U slučaju da je učenik odabrao objekt nastavnog sadržaja namijenjen testiranju, sustav dinamički generira pitanja i nudi odgovore na izbor. Proces dinamičkog generiranja pitanja koristi predložke pitanja unutar kojih se slučajnim izborom postavljaju elementi područnog znanja, odnosno čvorovi, veze i okviri semantičke mreže.

Navedeni način testiranja često generira pitanja koja su gramatički nepravilna, te zbunjuju učenika. Osim toga

učenik može jednostavno metodom eliminacije ponuđenih odgovora odabrati točan odgovor.

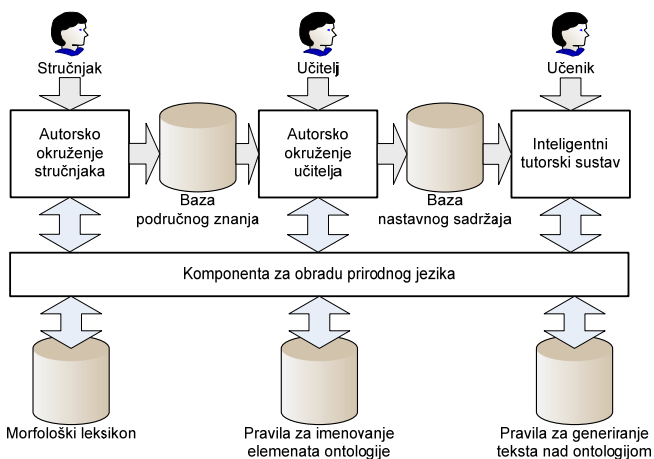
U brojnim provedenim anketama nad učenicima koji su koristili TEx-Sys ovo je bila česta zamjerka, koja ukazuje na potrebu primjene prirodnog jezika kod ovakvih sustava [7].

Zasnivanje ITS-a i pripadne autorske ljuske na prirodnom jeziku važan je korak u postizanju unapređenja ovakvih sustava. Karakteristika primjene prirodnog jezika u tutorskim sustavima zasniva se na komunikaciji između učenika i tutora. DIAG-NLP2 [8] i AutoTutor [9] su primjeri tutorskih sustava kod kojih se tijekom dijaloga učenika i računalnog tutora simulira temeljem analiziranih uzoraka stvarne komunikacije.

Generiranje i razumijevanje prirodnog jezika su dvije temeljne tehnologije primjenom kojih se može unaprijediti komunikacija između računalnog tutora i učenika. Povezivanjem nove komponente s komunikacijskim modulom ITS-a koja bi vršila obradu prirodnog jezika značajno bi se poboljšala komunikacija s učenicom [8].

Ideja izgradnje i povezivanje komponente za obradu prirodnog jezika s ITS-om može se proširiti povezivanjem iste s ostalim autorskim okruženjima u autorskoj ljuski ITS-a čime bi se nadasve olakšao proces oblikovanja područnog znanja i nastavnog sadržaja, a ne samo proces učenja, poučavanja i testiranja znanja.

II. AUTORSKA LJUSKA ZA INTELIGENTNO POUČAVANJE ZASNOVANA NA PRIRODNOM JEZIKU



Slika 1. Sustav LASIT

Autorska ljuska Language-based Authoring Shell for Intelligent Tutoring (LASIT) u svim svojim autorskim okruženjima primjenjuje određenu komunikaciju prirodnim jezikom (slika 1):

- 1) stručnjaku se omogućuje oblikovanje područnog znanja s povratnom vezom na prirodnom jeziku,
- 2) učitelj prilikom oblikovanja nastavnog sadržaja oblikuje objekte nastavnog sadržaja koristeći

rečenice na prirodnom jeziku generirane nad područnim znanjem i

- 3) učenik osim što mu se nastavni sadržaj prikazuje na prirodnom jeziku, također svoje znanje testira kroz dijalog s računalnim tutorom koji se obavlja na prirodnom jeziku.

Sustav LASIT je temeljen na modelu sustava TEx-Sys, odnosno osnovne postavke sustava, kao što su autorska okruženja, i temeljenost na grupama znanja su naslijeđene. Osnovne funkcije komponente za obradu prirodnog jezika su generiranje prirodnog jezika nad područnim znanjem i prepoznavanje fraza koje se koriste kod opisivanja elemenata područnog znanja. Generiranje rečenica na prirodnom jeziku se najčešće koristi prilikom iznošenja područnog znanja u nastavnom sadržaju učeniku. Dijalog između učenika i računalnog tutora, osim generiranja koristi i prepoznavanje prirodnog jezika prilikom učenikovog odgovaranja na pitanja računalnog tutora.

A. Autorsko okruženje stručnjaka i područno znanje

Komponenta za obradu prirodnog jezika u sustavu LASIT se u procesu generiranja i razumijevanja oslanja na tehnicu prikaza područnog znanja. LASIT koristi opisnu logiku zapisanu Web Ontology Language (OWL) [10] jezikom za opisivanje ontologija na Web-u.

Prednosti opisne logike nad semantičkom mrežom koja se koristi u sustavu TEx-Sys jest veća izražajnost [11]. Opisna logika opisuje činjenice preko aksioma. Činjenice kao elementi opisne logike pripadaju nekom opisu, a opisi se mogu slagati u taksonomiju, logički kombinirati, ograničavati po kvantifikaciji, vrijednosti i kardinalnosti svojstva, pri čemu svojstva čine binarnu relaciju između dva opisa. Osim toga, ako činjenice pripadaju nekom opisu koji je domena nekog svojstva, onda ta činjenica može imati i vrijednost tog svojstva. U OWL ontologiji aksiomi opisa definiraju klase, aksiomi relacija definiraju svojstva, a činjenice definiraju individue.

Stručnjak u svom autorskom okruženju oblikuje područno znanje imenovanjem klasa, svojstava i individua, te njihovim opisivanjem u skladu s opisnom logikom. Prilikom definiranja elemenata OWL ontologije, komponenta za obradu prirodnog jezika koristi pravila za imenovanje kako bi se spriječilo nepravilno nazivanje što bi imalo za posljedicu generiranje gramatički nepravilnih rečenica. Svakom promjenom područnog znanja, komponenta za obradu prirodnog jezika povratnom vezom generira rečenice, omogućujući stručnjaku praćenje oblikovanja područnog znanja.

B. Autorsko okruženje učitelja i nastavni sadržaj

Model nastavnog sadržaja u sustavu LASIT je hijerarhijski organiziran kao stablo čiji se korijenski čvor zove korijenska agregacija, unutarnji čvorovi su agregacije, a listovi su objekti nastavnog sadržaja koji

mogu biti namijenjeni učenju ili testiranju znanja. Agregacija je element modela nastavnog sadržaja koja služi za njegovu organizaciju, dok objekti nastavnog sadržaja uključuju elemente područnog znanja. Učitelj definira elemente nastavnog sadržaja određivanjem njihovog položaja u modelu i zavisnosti s drugim elementima nastavnog sadržaja. Zavisnost nekog elementa nastavnog sadržaja ukazuje na elemente nastavnog sadržaja koje učenik mora poznavati kako bi pristupio tom elementu.

Jedino objekti nastavnog sadržaja osim položaja i zavisnosti imaju i sadržaj. Sadržaj objekta nastavnog sadržaja namijenjen učenju učitelj definira izborom rečenica iz nekog područnog znanja kojeg je prethodno stručnjak oblikovao. U tom procesu sudjeluje komponenta za obradu prirodnog jezika koja generira rečenice za odabrano područno znanje.

Prilikom definiranja sadržaja objekta nastavnog sadržaja namijenjen testiranju, učitelj odabire one objekte nastavnog sadržaja namijenjenih učenju koji će sadržavati elemente područnog znanja nad kojima će se vršiti testiranje znanja putem dijaloga.

C. Okruženje učenika i model studenta

Okruženje učenika, odnosno ITS prikazuje model nastavnog sadržaja u kojemu učenik odabire one elemente nastavnog sadržaja koji su mu ponuđeni. U modelu studenta se pri tome pamte učenikovi koraci, odnosno redosljed elemenata koje je odabrao. U slučaju da je učenik odabrao objekt nastavnog sadržaja namijenjen učenju, ITS mu prikazuje generirane rečenice na prirodnom jeziku. Ako je učenik odabrao objekt nastavnog sadržaja namijenjen testiranju, onda ITS započinje dijalog s učenikom. U ovom dijalogu računalni tutor prvo postavlja pitanje, na koje učenik odgovara. U slučaju da je odgovoran točan onda računalo postavlja drugo pitanje i tako dalje, a ako je odgovor netočan ili djelomično točan, računalo pokušava pojednostavniti početno pitanje ili postaviti potpitanje kako bi pomoglo učeniku u pronalaženju odgovora. Sva pitanja, potpitanja i odgovori ovog dijaloga se također zapisuju u model studenta što kasnije pomaže prilikom određivanja ocjene znanja učenika.

III. KOMPONENTA ZA OBRADU PRIRODNOG JEZIKA

Podrška za generiranje i razumijevanje prirodnog jezika u autorskoj ljsuci LASIT je implementirana u komponenti za obradu prirodnog jezika. Ova komponenta se može promatrati kao podsustav čija arhitektura sadrži morfološki leksikon sa svim oblicima riječi koje se koriste u LASIT-u, identifikatora imenovanih elemenata područnog znanja i generatora rečenica nad područnim znanjem.

A. Morfološki leksikon

Svi oblici promjenjivih vrsta riječi koje sustav LASIT koristi su zapisani u morfološkom leksikonu. Realizacija leksikona je temeljena na Hrvatskom Morfološkom Leksikonu (HML) [12].

Morfološki leksikon je zapravo baza podataka koja se sastoji od tri temeljna dijela: leksikona, tablica nastavaka promjena riječi i tablica preoblika. Prije izrade popisa preoblika i nastavaka potrebno je klasificirati riječi jezika, tako da riječi na čiji korijen pri istoj promjeni djeluju iste preobliske budu u istoj klasi, odnosno da imaju isti broj uzorka promjene. Riječi se klasificiraju s obzirom na to koje nastavke dodajemo na osnovu riječi i koje se preobliske pri tom događaju. Promjene koje djeluju na riječi su deklinacija kod imenica, pridjeva i zamjenica, konjugacija pridjeva i komparacija pridjeva i priloga.

Leksikon sadrži veliku količinu podataka, odnosno veliki broj riječi. Zapis u leksikonu se tvori od:

- leme odnosno osnovnog oblika riječi (kod imenica je osnova u nominativu jednine),
- osnove riječi odnosno korijena riječi,
- broja uzorka promjene koji ukazuje na promjenu koja će se primijeniti nad osnovom riječi,
- roda u imenica ako se ne može vidjeti iz uzorka ili oznaka neprijelaznosti glagola,
- broja u imenica
- značenja ukoliko je potrebno za razlikovanje dviju istih lema

Broj uzorka promjene podijeljen je na tri dijela, zbog toga što sve vrste riječi imaju neke od tri promijene: deklinaciju, konjugaciju, komparaciju, pa opći oblik izgleda ovako:

$$dek\{0..4\} / konj\{0..4\} / komp\{0..4\} \quad (1)$$

Preobliske su glasovne promjene, kod imenica se mogu definirati kao uređena četvorka (u, pr, b, pa) gdje je:

- u – uzorak promjene odnosno broj koji identificira promjenu,
- pr – paradigma,
- b – broj (jednina ili množina) i
- pa – padež.

Sve preobliske koje se koriste dijelimo u tri skupine, one koje izbacuju pojedina slova, zamjenjuju pojedina slova ili niz slova pojedinim slovom ili grupama slova i dodaju slova ili grupu slova na kraj ili prije kraja osnove.

Primjer preobliske je sibilarizacija koja je definirana

$$...p_n > ...P_n / p_n = [g, h, k], P_n = [z, s, c] \quad (2)$$

što znači ako je zadnje slovo p_n osnove neko od slova k, g ili h, onda ga zamijeni sa slovima z, s ili c, što je označeno sa slovom P_n preobličene osnove.

Uzorci u tablici preoblika imaju istu identifikacijsku oznaku kao i zapisi u tablici nastavaka. Nastavci za imenice su osim uzorkom promjene određeni i brojem paradigme, te brojem i padežom imenice.

Krajnje generiranje oblika riječi kod imenica je postupak gdje se za određenu riječ iz leksikona primjenjuje pravila preoblike nakon čega slijedi dodavanje određenog nastavka iz tablice nastavka.

Primjerice ako imamo imenicu oblak, njena lema i osnova su oblak. U tablici 1 je dan uzorak preoblike koje se primjenjuje na danu imenicu, dok tablica 2 određuje nastavke koji se dodaju na promijenjenu osnovu imenice.

Tablica 1. uzorak preoblike

Paradigma 1													
Jednina							Množina						
N	G	D	A	V	L	I	N	G	D	A	V	L	I
				3			2		2		2	2	2

Tablica 2. tablica osnova

Paradigma 1													
Jednina							Množina						
N	G	D	A	V	L	I	N	G	D	A	V	L	I
a	u		e	u	om	i	a	ima	e	i	ima	ima	

Preoblika identificirana brojem 2 je sibilizacija, dok je brojem 3 označena palatalizacija. Određeni oblik riječi je označen po Multext specifikaciji [13], koja za svaku vrstu riječi definira niz atributa koje je klasificiraju u određenu skupinu. Na primjer, riječ oblak je opća imenica srednjeg roda koja po Multext-u je specificirana sa Nmsn, a može biti i Nmsa. Atribut N (noun) označava da je riječ imenica, c (common) je oznaka opće imenica, m (maskulin) određuje rod, s (singular) određuje broj i n (nominative) odnosno a (akuzative) govori o padežu imenice.

Primjenom preoblika i dodavanjem nastavaka, svi oblici imenice oblak su dani u tablici 3.

Tablica 3. Oblici imenice oblak

Osnova	Preoblika	Nastavak	Multext
oblak	oblak	oblak	Nc sn
oblak	oblak	oblaka	Nc sg
oblak	oblak	oblaku	Nc sd
oblak	oblak	oblak	Nc sa
oblak	oblač	oblače	Nc sv
oblak	oblak	oblaku	Nc sl
oblak	oblak	oblakom	Nc si
oblak	oblac	oblaci	Nc pn
oblak	oblak	oblaka	Nc pg
oblak	oblac	oblacima	Nc pd
oblak	oblak	oblake	Nc pa
oblak	oblac	oblaci	Nc pv
oblak	oblac	oblacima	Nc pl
oblak	oblac	oblacima	Nc pi

Morfološki leksikon može za zadanu lemu dati sve oblike riječi, te može za oblik riječi dati lemu kojoj ta riječ pripada. Ovo su osnovne operacije morfološkog leksikona na kojoj se temelji generiranje i razumijevanje prirodnog jezika.

B. Pravila za imenovanje elemenata područnog znanja

Generiranje prirodnog jezika je proces automatskog stvaranja razumljivog teksta na Hrvatskom ili bilo kojem drugom ljudskom jeziku temeljem nekog ne lingvističkog prikaza informacija [14]. Takav skup informacija najčešće predstavlja ulazne podatke koje sustav za generiranje prirodnog jezika obrađuje koristeći znanje o jeziku i domeni primjene kako bi automatski generirao dokumente, izvještaje, objašnjenja, poruke i ostale vrste teksta. Sustav LASIT generiranje hrvatskog jezika vrši nad opisnom logikom, koja je ne lingvistički zapisana u obliku OWL ontologije. Formalizam OWL ontologije definira elemente ontologije i iznosi njihovu specifikaciju, a za samo imenovanje elemenata je zadužen stručnjak u određenom područnom znanju.

Klase, individue, objektna svojstva i podatkovna svojstva su osnovni građevni elementi OWL ontologije koji se moraju imenovati radi kasnije pravilne obrade prirodnim jezikom [15]. Klasa kao skup individua jest apstraktni pojam preko kojeg se grupiraju individue sa sličnim svojstvima.

Naziv klase u generiranoj rečenici može označavati subjekt ili nekakav subjektivni skup, dio predikata i objekt ili dio objektnog skupa. Sama opća imenica ili opća imenica vezana s ostalim riječima čini naziv klase, što se korištenjem Multext specifikacije zapisuje na sljedeći način:

$$\{As--b_1n\} \{Af-r_1b_1n\} \{Rgp\} \{Ncr_1b_1n\} \quad (3)$$

$$\{\{As--b_2g\} \{Af-r_2b_2g\} \{Rgp\} \{Ncr_2b_2g\}\}$$

$$\{Np\}$$

Riječ ili niz riječi u zagradama mogu se i ne moraju pojaviti u nazivu klase. Središnja riječ koja najčešće definira subjekt je podcrtana i ona eventualno definira ovisnost određenih Multext atributa s ostalim riječima. Ispred središnje riječi koja je opća imenica u nominativu jednine nalaze se posvojni pridjev koji može stajati ispred opisnog pridjeva. Prilog se također nalazi ispred središnje riječi. Iza središnje riječi se mogu naći opća imenica u genitivu ili vlastita imenica u nominativu. Opća imenica u genitivu također može imati ispred sebe posvojni i/ili opisni pridjev.

"Perin visoki rođak Ivan" je primjer niza riječi kojima se definira naziv klase čija specifikacija glasi:

$$Aspmnsn Afpmnsn Ncmsn Npmsn \quad (4)$$

Ovaj primjer uključuje jedan posvojni i opisni pridjev ispred središnje riječi i jednu vlastitu imenicu iza središnje riječi.

Formula (4) definira pravilo za imenovanje klase kao subjekta ili dijela subjektivnog skupa, koje je jednako pravilo za imenovanje klase kao dijela predikata. Karakteristika ova dva pravila je što je središnja imenica u nominativu jednine. Kod pravila za imenovanje klase

kao objekta ili objektnog skupa središnja imenica može biti u svim ostalim padežima. Za ova pravila se uvodi pojednostavljeni način specificiranja u kojemu se ističe položaj naziva klase u odnosu na rečenicu i njen padež.

Pravilo iz formule (4) se skraćeno označava s K-sn gdje "K" ukazuje na klasu, "s" na subjekt ili subjektini skup, a "n" na nominativ. Primjer pravila za imenovanje klase kao objekta ili dijela objektnog skupa gdje je središnja imenica u akuzativu glasi:

$$K-oa ::= \{As--b_1a\} \{Af-r_1b_1a\} \{Rgp\} \{Ncr_1b_1a\} \{As--b_2g\} \{Af-r_2b_2g\} \{Rgp\} \{Ncr_2b_2g\} \{Np\} \quad (5)$$

Imenovanje individue je slično imenovanju klase, s razlikom što se pravila za imenovanje osim definiranja po ulozi individue u odnosu na dio rečenice, određuju i po vrsti središnje riječi, koja može biti opća imenica, vlastita imenica ili broj. Pravilo za imenovanje individue koja za središnju riječ ima opću imenicu je jednako pravilu za imenovanje klase. Ako je središnja riječ posebna imenica, onda pravilo glasi:

$$\{As--b_1n\} \{Af-r_1b_1n\} \{Rgp\} \{Npr_1b_1n\} \{As--b_2g\} \{Af-r_2b_2g\} \{Rgp\} \{Ncr_2b_2g\} \quad (6)$$

Broj kao središnja riječ kod individua ima jednostavno pravilo koje glasi:

$$M---n \quad (7)$$

Nazivi svojstava ontologije imaju ulogu predikatnog skupa u rečenici. Predikatni skup ovisi o vremenu i utječe na padež objektnog skupa. Zbog toga se pravila za imenovanje objektnog skupa definiraju za glagolsko vrijeme i za padež. Na primjer "je imao" je naziv objektnog svojstva koji je u perfektu i u akuzativu. Njega će prihvatiti pravilo koje glasi

$$Va-p3s-ap \ Vmp-3s-ap \quad (8)$$

Podatkovna svojstva se razlikuju od objektnih svojstava po tome što povezuju individuu s nekom vrijednošću primitivnog tipa podataka kao što je broj, datum ili istinitost. Njihova pravila se dijele po glagolskom vremenu, međutim druga podjela nije po padežu kao kod objektnih svojstava već se dijele u skupine ovisno o odgovoru na neko od sljedećih skupina pitanja: (i) koliko, (ii) kako, kada, (iii) koji, koju, koje. U četvrtu skupinu spadaju podatkovna svojstva kojima je vrijednost istinitosti, na primjer "je bijel", "nije bijel".

Kod podatkovnih svojstava karakteristično je da se nekada sama vrijednost svojstva prilikom generiranja rečenice ubacuje u niz riječi koje čine ime svojstva. Primjerice svojstvo "ima % godina" sadrži oznaku "%" koja označava mjesto gdje će se postaviti vrijednost svojstva. Pravilo za imenovanje ovog svojstva glasi

$$Vm-p3s-ap \ \% \ Nc-sg \quad (9)$$

odnosno ovo svojstvo spada u prvu skupinu koja odgovara na pitanje koliko i glagolsko vrijeme je prezent.

Pravila za imenovanje klasa, individua, objektnih svojstava i podatkovnih svojstava stvaraju osnovu za generiranje gramatički ispravnih rečenica. Ona se koriste prilikom određivanja pripadnosti niza riječi nazivu nekog elementa ontologije. Samo određivanje je proces koji za ulaz ima niz riječi, a za izlaz Multext specifikaciju niza riječi, koje se koristi prilikom generiranja rečenica na prirodnom jeziku.

C. Generiranje rečenica nad područnim znanjem

Generiranje prirodnog jezika nad ontologijom kao područnim znanjem je proces koji za ulaz ima ontologiju, a za izlaz rečenice na prirodnom jeziku. Ovaj proces se sastoji od nekoliko faza, pri čemu se početna ontologija transformira u apstraktno stablo jednostavnih rečenica. Postupak generiranja započinje ulazom izraza ontologije čija je struktura definirana OWL apstraktnom sintaksom [16]. Zapravo OWL ontologija je niz izraza koji se zovu direktive, a izraz može biti aksiom ili činjenica.

$$directive ::= axiom \ | \ fact \quad (10)$$

Činjenicama se između ostalog definira sintaksa individue na sljedeći način:

$$individual ::= 'Individual(' [individualID] \ { annotation } \ { 'type(' type ')'} \ { value } \)' \quad (11)$$

pri čemu se individua može i ne mora identificirati jednom URI referencom [individualID], a može i ne mora sadržavati vrijednosti svojstava {value} čija gramatika glasi:

$$value ::= 'value(' \ individualvaluedPropertyID \ individualID \)' \ | \ 'value(' \ individualvaluedPropertyID \ individual \)' \ | \ 'value(' \ datavaluedPropertyID \ dataLiteral \)' \quad (12)$$

odnosno vrijednost individue je uređeni par koji se sastoji od identifikatora svojstva i vrijednosti koja kod je objektnih svojstava opet individua, a kod podatkovnih svojstava je konstanta tipizirana primitivnim tipom podataka. Individua može biti nekog tipa { 'type(' type ')'} gdje se tip definira kao opis:

$$type ::= description \quad (13)$$

Opisima se između ostalog određuje klasa kojoj pripada neka individua.

$$\begin{aligned}
description ::= & classID \mid restriction \mid \\
& 'unionOf(\{ description \})' \mid \\
& 'intersectionOf(\{ description \})' \mid \\
& 'complementOf(\{ description \})' \mid 'oneOf(\{ \\
& individualID \})'
\end{aligned}
\tag{14}$$

Najjednostavniji opis se tvori identificiranjem predefimirane klase s *classID*. Opisi se mogu logički kombinirati operatorima unije *unionOf*, presjeka *intersectionOf* i komplementa *complementOf*, mogu se ograničiti na konačan skup individua s operatorom prebrojavanja *oneOf* ili ograničiti s vrijednošću nekog svojstva *restriction*.

Aksiomima kao direktivama se prvenstveno definiraju klase i svojstva ontologije.

$$\begin{aligned}
axiom ::= & 'Class(\{ classID \} [\'Deprecated\]) \\
& modality \{ description \})'
\end{aligned}
\tag{15}$$

Na primjer, sljedeći konkretni izraz u svojoj apstraktnoj sintaksi opisuje klasu #X preko identifikatora klase #A, unije klase B, komplementa klase #C i egzistencijalne restrikcije svojstva #P opisa nastalog presjekom klasa #D i #E.

$$\begin{aligned}
Class(\{ \#X \#A unionOf(\{ \#B \\
complementOf(\{ \#C \}) restriction(\{ \#P \\
someValuesFrom(IntersectionOf(\{ \#D \\
\#E \})) \}) \})
\end{aligned}
\tag{16}$$

Pretvorba izraza ontologije u niz rečenica zahtjeva pretvorbu izraza u apstraktno stablo rečenica. Za konstruiranje apstraktnog stabla rečenica koriste se čvorovi koje svrstavamo u četiri skupine:

- logičke operatore AND, OR i NOT,
- relacijske operatore SAME i NOTSAME,
- jednostavne rečenice KIND, NOTKIND, IS, NOTIS, INSTANCE, NOTINSTANCE, SIMPLE i NOTSIMPLE,
- veznici za složene rečenice koji se eventualno mogu transformirati u jednostavne rečenice WHICH, NOTWHICH,
- pomoćne riječi SOME, ALL, VALUE, MIN, MAX, i EQUAL,
- fraze (označeni s prefiksom "#") koje predstavljaju nazive elemenata OWL ontologije.

Transformacija apstraktno sintakse ontologije u stablo vrši se tako što se elementi izraza u apstraktnoj sintaksi pretvore u čvorove i grane stabla. Prvi red u tablici 4 prikazuje transformaciju izraza klase čiji je naziv označen s #X i koja ima N opisa u apstraktno stablo rečenica.

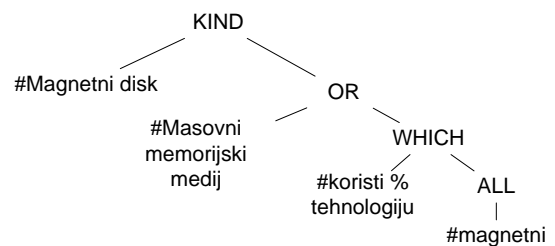
Tablica 4.

Class(#X description1 ... descriptionN)
<pre> KIND / \ #X OR / \ description1 ... descriptionN </pre>
unionOf(description1 ... descriptionN)
<pre> OR / \ description1 ... descriptionN </pre>
intersectionOf(description1 ... descriptionN)
<pre> AND / \ description1 ... descriptionN </pre>
restriction(#DP dataRestrictionComponent1 ... dataRestriscionComponentN)
<pre> WHICH / \ #DP AND / \ dataRestrictionComponent ... dataRestrictionComponent </pre>
minCardinality(#N)
<pre> MIN #N </pre>

Primjer izraza u apstraktnoj sintaksi

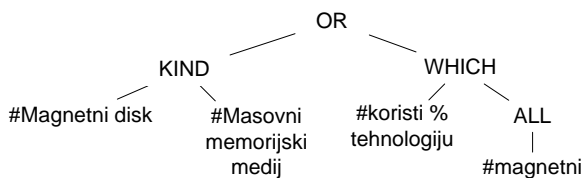
$$\begin{aligned}
Class(\{ \#Magnetni disk \#Masovni \\
memorijski medij restriction(\{ \#koristi \% \\
tehnologiju allValuesFrom(\{ \#Magnetski \}) \})
\end{aligned}
\tag{17}$$

nakon pretvorbe dobiva se apstraktno stablo sa slike 1.



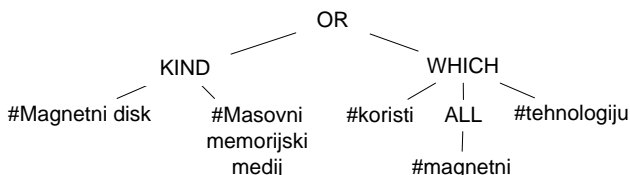
Slika 2.

Jednostavne rečenice i veznici za složene rečenice uvijek imaju dvije grane u apstraktnom stablu. Bitno je da se transformacijom apstraktnog stabla logički i relacijski operatori budu bliže korijenu stabla, fraze ili logička kombinacija fraza da budu listovi stabla, a jednostavne rečenice da imaju fraze u svoje dvije grane, kao što je prikazano na slici 3.



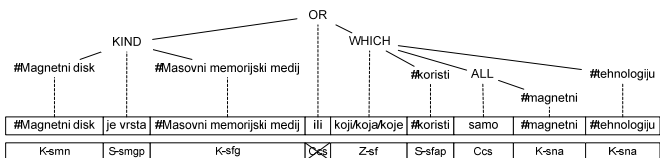
Slika 3.

Sljedeća faza transformacija razbija fraze koje sadrže zamjensku riječ i koje se nalaze u lijevoj grani jednostavne rečenice ili veznika za složene rečenice na dvije fraze kao što je prikazano na slici 4.



Slika 4.

Nakon izvršenih transformacija, čitanjem stabla s lijeva na desno dobiva se niz fraza kod kojih je morfologija fraze određena prema pripadnosti određenom čvoru i prema kojoj grani čvora pripadaju. Niz fraza na slici 5 je rezultat obilaska čvorova stabla.



Slika 5.

Fraza "Magnetni disk" ontološki određuje klasu čija morfološka specifikacija glasi K-smn, odnosno radi se o nazivu klase kojoj je središnja riječ u jednini, muškog roda i u nominativu. Nakon specifikiranja svih naziva, fraze i nepromjenjive riječi se nižu, te je ponekad potrebno da se određena riječ izbaci, kao što je izbačen veznik "ili".

U završnom koraku, za svaku riječ i za traženu specifikaciju fraza dobivenog niza, pomoću morfološkog leksikona se određuje potrebni morfološki oblik riječi i time je generiranje rečenice završeno, odnosno za naš primjer dobivamo rečenicu "Magnetni disk je vrsta masovnog memorijskog medija koji koristi magnetnu tehnologiju".

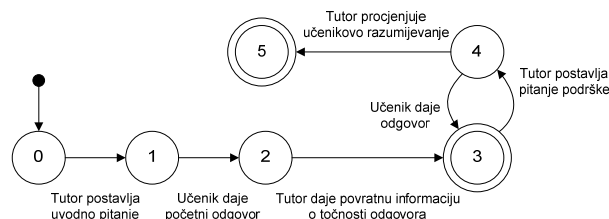
IV. DIJALOG U INTELIGENTNOM TUTORSKOM SUSTAVU

Komponenta za obradu prirodnog jezika je u uskoj suradnji s komunikacijskim modulom ITS-a. Osim što elemente područnog znanja prikazuje kao rečenice na prirodnom jeziku ima vrlo bitnu ulogu kod dijaloga

između računalnog tutora i učenika. Dijalog je osmišljen nad nekim skupom elemenata područnog znanja. Raspon dijaloga, odnosno skup elemenata područnog znanja odabire učitelj prilikom oblikovanja nastavnog sadržaja, odnosno oblikovanja objekta nastavnog sadržaja.

Testiranje znanja učenika vrši se metodom prekrivanja, gdje se znanje učenika procjenjuje kroz dijalog s računalnim tutorom. Prekrivanjem se uspoređuju tri baze područnih znanja: baza stručnjaka, baza problema i baza rješenja. Baza stručnjaka je ono područno znanje koje je učitelj namijenio učeniku za testiranje. Baza problema je manja od baze stručnjaka jer nastaje iz nje oduzimanjem elemenata područnog znanja. Baza rješenja nastaje nadopunom baze problema od strane učenika pri čemu se ta nadopuna ostvaruje dijalogom. Procjena i ocjena znanja učenika se, u osnovi, ostvaruje usporedbom ove tri baze znanja [5][17].

Tijek dijaloga između računalnog tutora i učenika se sastoji od niza tutorskih okvira [18] gdje je svaki tutorski okvir opisan kroz maksimalno 5 koraka (slika 6).



Slika 6. Okvir tutorskog dijaloga

Tutorski okvir započinje postavljanjem pitanja od strane računalnog tutora. Uvodno pitanje tvori se slučajnim odabirom predloška pitanja.

```

Class(#Optički disk #Masovni memorijski medij
restriction(#koristi % tehnologiju
allValuesFrom(#Optički)))
Class(#Dio magnetnog diska)
Class(#Magnetni disk #Masovni memorijski medij
restriction(#koristi % tehnologiju
allValuesFrom(#Magnetski)))
Class(#Dio magnetnog diska)
Individual(#Ploča type(#Dio magnetnog diska)
value(#ima % cilindar, 2))
Individual(#Glava type(#Dio magnetnog diska)
value(#čita, #Podatak) value(#piše, #Podatak))

```

Slika 7. Baza zadatka i baza problema

U primjeru na slici 7 je prikazana baza zadatka iz koje, ako se oduzmu fraze i vrijednosti kao što su #Masovni memorijski medij, #2 i #Glava, dobije se baza problema. Dijalog kreće tako što računalni tutor one izraze apstraktne sintakse u kojima nisu obrisane fraze ili vrijednosti prikazuje kao rečenice na prirodnom jeziku.

Tutor:	Optički disk je vrsta masovnog memorijskog medija koji koristi optičku tehnologiju.
--------	---

Slika 8. Računalni tutor generira rečenicu

Kod izraza kojima nedostaje fraza ili neka vrijednost računalni tutor će odmah postaviti pitanje ili će generirati uvodnu rečenicu i zatim postaviti pitanje.

Tutor:	Magnetni disk koristi magnetnu tehnologiju.
Učenik:	Što je još magnetni disk?
Tutor:	Masovni memorijski medij
Tutor:	Točno, magnetni disk je masovni memorijski medij koji koristi magnetnu tehnologiju.
Tutor:	Idemo dalje.

Slika 9. Generiranje uvodnog pitanja

Ako učenik odgovori točno, tutorski okvir se prekida i postupak se nastavlja za bazu problema sa slike 7.

Tutor:	Ploča je dio magnetnog diska. Koliko ima cilindara?
Učenik:	3
Tutor:	Pokušaj ponovno.
Učenik:	2
Tutor:	Točno, ploča je dio magnetnog diska i ima 2 cilindra.

Slika 10. Dijalog s pitanjima podrške

Za svako postavljeno uvodno pitanje računalni tutor procjenjuje količinu pitanja podrške ovisno o poziciji fraze koja se izbrisala iz baze stručnjaka. Ako je učenik točno odgovorio na uvodno pitanje, dobio je maksimalni broj bodova za dani tutorski okvir. U slučaju da nije točno odgovorio na uvodno pitanje, i nakon i-tog pokušaja odgovorio na jedno od "n" pitanje podrške, tada se ukupnom broju bodova pribraja omjer "n" i "i" bodova.

Omjer ukupnog broja bodova i maksimalnog broja bodova predstavlja ocjenu za odabrani objekt nastavnog sadržaja namijenjen testiranju.

V. ZAKLJUČAK

Sustav LASIT komunikaciju s korisnicima temelji na prirodnom jeziku. Generiranje i prepoznavanje prirodnog jezika obavlja u podsustavu koji sadrži morfološki leksikon, pravila za imenovanje elemenata znanja i generator rečenica na hrvatskom jeziku. Autorska okruženja LASIT-a koriste komponentu za obradu prirodnog jezika prilikom oblikovanja područnog znanja, oblikovanja nastavnog sadržaja i kod učenja i poučavanja. Učeniku kao ciljnom korisniku sustava

učenje se sastoji od usvajanja tekstualnog i grafičkog prikaza znanja. Testiranje znanja učenika ostvaruje se kroz dijalog s računalnim tutorom.

REFERENCE

- [1] Self, J. (1990). *Theoretical foundations for intelligent tutoring systems*. Journal of Artificial Intelligence in Education 1 (4), pp. 3–14.
- [2] Burns, H. L., Capps, C. G. (1988). *Foundations of intelligent tutoring systems: an introduction*. Poison, M.C., Richardson, J. J. (Ed.) Foundations of intelligent tutoring systems, Lawrence Erlbaum, London, pp. 1-19.
- [3] Murray, T. (1999). *Authoring intelligent tutoring systems: an analysis of the state of the art*. International Journal of Artificial Intelligence in Education Vol. 10 (1999), pp. 98–129.
- [4] Stankov, S. (2003). Tehnološki projekt TP-02/0177-01. Web orijentirana inteligentna hipermedijalna autorska ljsuka. Ministarstvo znanosti i tehnologije republike hrvatske, 2003-2005.
- [5] Stankov, S. (1997). *Isomorphic Model of the System as the Basis of Teaching Control Principles in an Intelligent Tutoring System*. Doktorska disertacija. Fakultet elektronike strojarstva i brodogradnje, Sveučilište u Splitu, Split, Hrvatska.
- [6] ADL, (2006). *SCORM 2004*. 3rd Edition, Advanced Distributed Learning.
- [7] Stankov, S., Žitko, B., Grubišić, A., (2005). *Ontology as a Foundation for Knowledge Evaluation in Intelligent E-learning Systems*. International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL'05) in conjunction with 12th International Conference on Artificial Intelligence in Education (AI-ED 2005), July 18th, 2005, Amsterdam, The Netherlands, pp. 81-84
- [8] Di Eugenio, B., Fossati, D., Yu, D., Haller, S., Glass, M., (2005). *Natural Language Generation for Intelligent Tutoring Systems: a case study*. 12th International Conference on Artificial Intelligence in Education (AI-ED 2005), July 18th, 2005, Amsterdam, The Netherlands.
- [9] Graesser, A. C., Person, N. K., Harter, D. (2001) *Teaching Tactics and Dialog in AutoTutor*. International Journal of Artificial Intelligence in Education.
- [10] Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., Stein, L. A., Olin, F. W. (2004). *OWL Web Ontology Language Reference*. W3C Recommendation. 10 February 2004.
- [11] Stankov, S., Žitko, B., Grubišić, A. (2005). *Ontology as a Foundation for Knowledge Evaluation in Intelligent E-learning Systems*. International Workshop on Applications of Semantic Web

- Technologies for E-Learning (SW-EL'05) in conjunction with 12th International Conference on Artificial Intelligence in Education (AI-ED 2005), July 18th, 2005, Amsterdam, The Netherlands, pp. 81-84.
- [12] Tadić, M. (1994). *Računalna obrada morfologije hrvatskog jezika*. Doktorska disertacija. Filozofski fakultet Sveučilišta u Zagrebu, Zagreb 1994.
- [13] Erjavec, T. (editor) (2004) MULTEXT-East Morphosyntactic Specifications, Version 3.0.
- [14] Reiter, E., Dale, R. (2000). *Building Natural Language Generation Systems*. Cambridge University Press. ISBN-10: 0521620368
- [15] Fliedl, G., Kop, C., Vöhringer, J. (2007) *From OWL Class and Property Labels to Human Understandable Natural Language*. Natural Language Processing and Information Systems, Springer Berlin / Heidelberg, pp. 156-167.
- [16] Patel-Schneider, P. F., Hayes, P., Horrocks, I. (2004). *OWL Web Ontology Language Semantics and Abstract Syntax*. W3C Recommendation.
- [17] Stankov, S., Rosić, M., Žitko, B., Grubišić, A. (2007). (u tisku) *TEx-Sys model for building intelligent tutoring systems*. Computers & Education 2007.
- [18] Chi, M. T. H., Siler, S. A. , Jeong, H., Yamauchi, T., Hausmann, R. G. (2001) *Learning from human tutoring*. Cognitive Science , 25 , pp. 471-533.