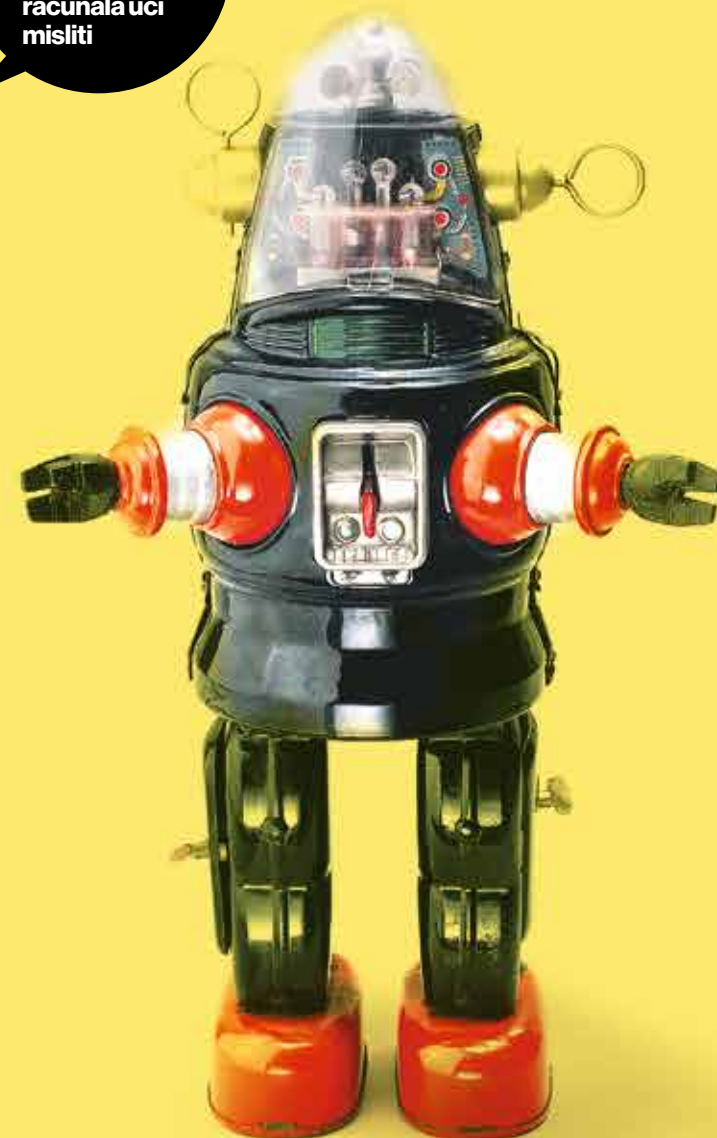


Šaptač robotima

Do 2020. sa svojim ću timom stvoriti umjetnu inteligenciju koja će analizirati što korisnik na društvenoj mreži doista misli, osjeća i namjerava. Kao u Zvezdanim stazama, gdje roboti razumiju ljude i s njima smisleno komuniciraju

PIŠE TOMISLAV NOVAK | SNIMIO MARKO MIŠČEVIĆ/EPH



Fantastični laboratorij dr. Jana Šnajdera, čovjeka koji računala uči misliti



BUDUĆNOST
PAMETNIH
STROJEVA

R

oboti. Inteligentni strojevi koji komuniciraju, žive i rade s ljudima.

Još otkako je kao dječak slušao očeve priče o robotima Jan Šnajder znao je čime se u životu želi baviti.

"Fascinirali su me roboti koji razmišljaju. Mentalno stanje strojeva", kaže mi dok razgovaramo u Zemrisovom TakeLabu na Fakultetu elektronike i računarstva.

"Ako stroj generira misli, te misli moraju se moderirati nečime što je slično jeziku. A jezik je sve, u njemu se kriju svi odgovori", dodaje.

TakeLab je kratica za Laboratorij za analizu teksta i inženjersko znanje, a Zemris za Zavod za elektroniku, mikroelektroniku, računalne i inteligentne sustave.

Doc. dr. sc. Jan Šnajder docent je na tom Zavodu.

Iza kompliciranih imena krije se ustvari malena učionica D339-I u središtu tunnelskog kompleksa hodnika i liftova na trećem katu zgrade D na FER-u, i isprva ju je teško pronaći. Krcata je računalima, a Jan u njoj zajedno sa svojim studentima provodi desetak i više sati dnevno.

Na zidu između ostalih plaketa i priznanja visi i premijerina nagrada koju je potpisala tadašnja šefica Vlade Jadranke Kosor za sustav CADIAL, moćni i dobro organizirani pretraživač hrvatskih zakona i propisa. Voditeljica tog projekta ujedno je i šefica TakeLaba i profesorica na FER-u Bojana Dalbelo Bašić.

KNJIŽEVNICI Cadial je samo jedan u nizu hitova iz TakeLaba. Između ostalog za Konzum su takelabovci izgradili računalni sustav koji na temelju podataka o proizvodima s polica i njihovoj kupnji za Konzum predviđa koja će im akcija sniženja biti uspješna, a koja ne.

Ali sve to već je prošlost. A sa Šnajderom pričamo o budućnosti.

Jan dolazi iz književničke obitelji, otac mu je pisac Slobodan Šnajder, a majka Krana Tarle, kazališna i likovna umjetnica.

"Možda sam upravo zato što dolazim iz obitelji književnika toliko fasciniran jezikom", zaključuje.

Iako se danas bavi obradom prirodnog jezika, strojnim učenjem i umjetnom inteligencijom, a ima magisterij i doktorat iz računarstva, čini se da se u suštini bavi istim stvarima kao i njegovi roditelji samo u drugoj epohi.

Misterij jezika i komunikacije među ljudima fokus je cijele obitelji.

A način na koji se danas Jan Šnajder bavi jezikom doista je uzbudljiv.

Dobar primjer za to je projekt SenseHive.

Težak je 800 tisuća kuna, financiraju ga Hrvatska zaklada za znanost i FER, počeo je u listopadu, a trajat će pune tri godine. Na njemu uz Šnajdera surađuje još deset istraživača, s Filozofskog fakulteta u Zagrebu, Sveučilišta u Ljubljani i Stuttgartu, te s Instituta za hrvatski jezik i jezikoslovlje.

Kad završe, tamo negdje potkraj 2018. godine Hrvati bi trebali dobiti veliku online semantičku mrežu riječi, što je pojednostavljeno veliki internetski rječnik prirodnoga govora.

Istražuje se prvo hrvatski jezik, a onda i slovenski pa engleski.

Tražim ga da mi na razumljiv način objasni

interneta.

Svaki od tih ispitanika svojim trudom ili znanjem onda pridonosi nekoj većoj ideji ili projektu, često u ime općeg dobra.

Takav način prikupljanja znanja brz je i jeftin. Bolji od ispitivanja zaposlenika ili plaćenih 'pokusnih kunića'. Najbolji primjer za uspješni crowdsourcing je crowdfunding, kad se novac za početak poslovnog projekta preko stranica poput Kickstartera ili Indiegoga skuplja od velikog broja ljudi.

Time se sve češće koriste i startupovi u Hrvatskoj.

TRAGAČI "Kad završimo, imat ćemo alat koji će biti koristan ne samo za znanstvenike nego i za razvoj jezičnih tehnologija. A primjena mu je itekako komercijalna. Sve što u TakeLabu razvijamo ima praktičnu primjenu i komercijalnu svrhu", govori mi.

Pitam ga na koje bi se sve načine SenseHive mogao komercijalno upotrijebiti?

"Zamislite semantičku tražilicu koja će korisnicima donijeti bolje i preciznije rezultate. Ljudi će biti zadovoljniji, a i imat će pristup većem broju informacija."

Što je to semantička tražilica i zašto bi ma-

U sklopu projekta SenseHive Snajder i njegovi suradnici će tisućama ljudi postaviti milijun pitanja o višeznačnosti riječi. Pomocju strojnog učenja nakon toga će izgraditi veliki online rječnik prirodnog govora i semantičku tražilicu za hrvatski jezik. Zvuči kao da grade hrvatski Google

o čemu se zapravo radi. Riječ 'pojednostavljeno' često ćemo koristiti tijekom razgovora.

"Želimo izgraditi mrežu prirodnog jezika tako da pitamo ljude: okej, ovo je riječ, možete li vi pronaći zamjenu za nju u ovoj rečenici tako da ne promijenite smisao te rečenice, da njeno značenje ostane isto? Govorimo o oko milijun takvih pitanja koja ćemo postaviti tisućama ljudi, izvornih govornika hrvatskog", objašnjava.

Tražit će ih, primjerice, da riječ avion zamijene riječju zrakoplov. Ili cesta s prometnica ili možda ulica.

Plan je da vide koliko sinonima i višeznačnica u realnom svijetu ima hrvatski jezik, ne onaj iz knjiga i rječnika, nego iz svakodnevnoga govora.

Ispitanike za SenseHive planiraju pronaći crowdsourcingom.

Termin skovan od izraza crowd (masa ljudi) i outsourcing (izdvajanje djelatnosti) označava popularni i jeftini model skupljanja ideja, ili usluga, u ovom slučaju odgovora na pitanja od velike grupe ljudi koji su volonteri ili honorarni suradnici, često putem

rio za nju, pita se David Amerland, poznati pisac, govornik i tech blogger u predgovoru svoje knjige Google Semantic Search.

Načini na koji koristimo internet i pretražujemo ga zbog informacija promijenio se unatrag nekoliko godina. Najtočniji i najbolji među njima, tvrdi Amerland, semantička je pretraga, odnosno mogućnost da ono što pretražujemo stavimo u kontekst.

Daje i konkretan primjer: ljudi više ne guglaju pojam 'restoran na Manhattanu', već sve češće konkretnije upite poput 'gdje jesti najbolju indijsku hranu na Manhattanu'.

Strojevi koji pogone internetske tražilice više se ne smiju zadovoljiti razumijevanjem samo ključnih riječi nego moraju razumjeti kako su te riječi povezane da bi dali bolji rezultat.

I to je ono što zanima Jana Šnajdera u projektu SenseHive: otići korak dalje, zaviriti iza zavjese i otkriti dublje značenje pojmova.

SenseHive još je važniji jer se iz njega može izroditi semantička tražilica hrvatskih riječi.

A to bi bila prava mala revolucija. ➔

Doc. dr. sc. Jan Šnajder sudjelovao je na brojnim domaćim i međunarodnim projektima, autor je i suautor 70 znanstvenih radova te mentor i sumentor na 60 diplomskih radova. Radi u TakeLabu na FER-ovu Zavodu Zemris i bavi se analizom jezika, strojnim učenjem i umjetnom inteligencijom



"Bit jezika je značenje. Da bismo mogli modelirati značenje jezika, moramo znati modelirati značenje riječi u tom jeziku. I zato istražujemo njihovu višeznačnost u ovom projektu. Hrvatski je mali jezik i za njegovo proučavanje i razvoj računalnih alata teško je dobiti financijska sredstva, barem u količini koju možemo dobiti za recimo engleski jezik, pa je svaki trud oko hrvatskog jako bitan."

Financiranje i općenito ulaganje u znanost nešto je što Šnajdera frustrira.

ZNANSTVENICI Iako je tek na početku znanstvene karijere, iza njega je poprilično dojmljiv CV.

Sudjelovao je na četiri znanstvena projekta Ministarstva znanosti, obrazovanja i športa, četiri međunarodna i dva domaća projekta.

Godinu dana bio je gostujući istraživač na Institutu za računalnu lingvistiku Sveučilišta u Heidelbergu.

Autor je ili suautor 70 znanstvenih radova u časopisima iz područja obrade prirodnog jezika i pretraživanja informacija.

Kao mentor ili sumentor pratio je studente na pripremi 60 preddiplomskih i diplomskih radova.

Prije pet godina primio je srebrnu plaketu "Josip Lončar" na FER-u za posebno istaknutu doktorsku disertaciju.

I to je samo dio njegova životopisa.

Pa opet ne prođe ni jedan tjedan, a da mu kroz glavu ne prođe ideja da otiđe iz Hrvatske.

"Nisam nezadovoljan, ali znanstvenici u inozemstvu puno lakše dolaze do puno većih sredstava za svoja istraživanja. Mislim da se u Hrvatskoj u znanost ne ulaže dovoljno. Srećom, naši studenti su izuzetno talentirani i samo zbog toga i tvrdoglavo truda možemo parirati stranim zemljama", kaže mi.

S druge strane, upravo je način na koji se mladi znanstvenici u Hrvatskoj moraju boriti s preprekama i nedostatkom novca od njega učinio, vjeruje, kreativnijeg čovjeka, otpornijeg na probleme.

Kad ne radi, a to je rijetko, Šnajder se bavi jogom ili trči oko Jaruna. Pobornik je zdravog lifestylea: ne pije, ne puši, ne voli kavu, a ne čita, kaže, ni upola koliko bi htio. Karijane košulje i crvene starke neizostavni su mu modni izbori.

Jedva je našao vremena za razgovor za Globus. Unatrag nekoliko tjedana u velikoj je gužvi. Kad govori, brz je, elokventan i efikasan, baš kao računalni alati za obradu jezika koje proučava.

Uz SenseHive priprema još jedan veliki projekt. I ovog puta riječ je o pretraživanju goleme količine podataka na internetu. Big data ovdje je prikladan buzzword, kaže mi Šnajder. O čemu se radi?

Projekt će se baviti pretraživanjem i modeliranjem događaja.

"Događaji su ono što u medijima donosi najveću informacijsku vrijednost. Nas zanima kako ih na računalu možemo modelirati, i događaje i protagoniste i okolnosti ili lokacije. I opet idemo korak dalje, želimo događaje koje pretražujete povezivati s drugim sličnim događajima. A to cijeloj priči daje kontekst koji bi onda nudili korisnicima", objašnjava.

Evo primjera.

Kad bi netko pretraživao uhićenje Sanadera, svatko s dovoljno živaca i vremena može o tom događaju i sam pronaći golem broj informacija. Međutim, zamislite servis koji bi uz to korisniku kao informaciju ponudio i sva slična uhićenja, po lokacijama ili drugim odrednicama poput vremena i slično.

I onda mu to formirao u sadržaj.

U teoriji stroj bi tako čovjeku mogao brzinom munje dati informaciju u kojem mjesecu je uhićenje najviše političara, na kojim lokacijama, kakvo je tada bilo vrijeme i tko zna što još.

Sve to, naravno, mogao bi i korisnik sam, ali bi mu trebalo jako puno vremena i truda da to učini i pitanje je kakav bi bio rezultat. Ovako čitavu priču dobivaju serviranu nakon jednog klika.

"Pa vi ste zapravo ubojice novinarstva", kažem mu, a on se smije.

Jan je sin književnika Slobodana Šnajdera i kazališne i likovne umjetnice Krune Tarle. Kaže da je zato od uvijek opsjednut proučavanjem jezika, a robote je zavolio jer mu je otac kao dječaku pričao o njima. Star Trek, Kubrickova Odiseja u svemiru i sf-romani Stanislava Lema inspirirali su ga da postane znanstvenik



"Ne prođe ni tjedan a da ne maštam kako radim u inozemstvu. U znanost ulažemo premalo, ali srećom zemlja smo talentiranih i upornih studenata. S druge strane borba s birokracijom i manjkom novca ojačala nas je i učinila kreativnijim istraživačima"

"Baš suprotno. Novinari bi od takvog servisa imali puno koristi, kao i istraživači povijesnih i društvenih događaja ili fenomena. Podatke za projekt crpit ćemo sa stranih portala za engleski jezik i domaćih za hrvatski. A njih kreiraju novinari", objašnjava mi.

Koliko se to razlikuje od uobičajenog gulanja?

Znatno.

Jer takelabovci žele povezivati događaje i među njima pronaći vremenske i uzročno-posljedične veze. Šnajder to naziva recommendation engineom, odnosno sustavom koji će predviđati što želite pronaći.

"Otvoreno je pitanje što je točno događaj. I kad ga netko pretražuje kako mu omogućiti sve relevantne informacije, ne zaustaviti ga prerano, a opet ga ne zatrpiti s previše informacija."

Ako su informacije novac 21. stoljeća, onda Šnajder i njegovi suradnici žele da se nabrinu obogatite.

A to će učiniti pomoću strojnog učenja i umjetne inteligencije.

Roboti o kojima mu je u djetinstvu pričao otac bili su, čini se, pozitivci jer Šnajder, za razliku od mnogih uglednih znanstvenika i vodećih biznismena na planetu, ne vjeruje kako smo u opasnosti od umjetne inteligencije.

Ne misli da smo blizu stvaranja Skyneta koji će nas zbrisati sa zemlje.

"Otac mi zapravo nikad nije čitao. Koliko se sjećam, sve su te priče bile njegove, izmišljene, autorske", smije se.



ok Stephen Hawking, Elon Musk i Bill Gates potpisuju peticije protiv razvoja UI i boje se da je čovječanstvo u opasnosti, Šnajder ima drugačiji pogled na problem.

"Većina znanstvenika koji se bave umjetnom inteligencijom ne zamaraju se takvim crnim scenarijem, budući da bi on zapravo podrazumijevao razvoj tzv. jake umjetne inteligencije, odnosno inteligencije koja ima sposobnosti barem jednake ljudskima, uključivo mentalna stanja i svijest. Nije izvjesno da je takvo što uopće moguće i razmatranja te vrste trenutno su u domeni filozofije uma. U znanstvenoj zajednici koja se bavi umjetnom inteligencijom po tom pitanju ne postoji konsenzus. Opći je dojam, međutim, čak i ako to jest moguće, da smo od prave UI vrlo daleko", smatra.

Za njega je opasnost puno praktičnije naravi - a to je gubitak radnih mjesta na poslovima gdje umjetna inteligencija učinkovito može zamijeniti ljude. I oko toga se većina znanstvenika slaže: no ni to nije nužno negativno.

Šnajder smatra da se krađi poslova od strane kompjutera i robota možemo othrvati tako da mlade ljude obrazujemo za nova zanimanja, koja će aktualna postati za deset ili dvadeset godina.

UČENICI "Projekcije kažu da će mnoga od tih zanimanja biti vezana uz računarstvo, ponajviše umjetnu inteligenciju, analizu podataka i robotiku. U to obrazovanje treba ulagati sada, a ne za 20 godina, kad za to bude prekasno. Dakle, odgovor na prijetnju umjetne inteligencije je, paradoksalno, upravo inteligencija: sposobnost društva da se prilagodi novoj situaciji."

I tu se vraćam na početak priče.

U Hrvatskoj se ne ulaže dovoljno u obrazovanje i znanost. I zato smo u opasnosti.

Kad ga pitam što će on raditi za pet godina, odmah navodi projekt koji bi želio razviti, prijaviti na natječaj i onda ga uspješno završiti.

Projekt za sada nema ni ime.

Ali njime bi još dublje ušao u područje

umjetne inteligencije.

Želi analizirati razmišljanja, stavove i emocije ljudi na društvenim mrežama. Ljudima, ali i velikim kompanijama danas je vrlo bitno znati što drugi ljudi misle o političarima, proizvodima ili kontroverznim temama poput pobačaja. Velika količina tih razmišljanja u obliku podataka danas je dostupna na internetu.

"Živimo u vremenu kad se ljudi ne libe javno napisati sve što misle. Velika količina informacija tamo je vani, ali s njom ne možete puno ako ne automatizirate ekstrakciju tih podataka", kaže mi.

Njegova ideja je da objave korisnika na društvenim mrežama analizira dublje od onog što danas tehnologije rade.

Da opet zaviri iza zavjese ljudskog uma.

"Trenutačno se uglavnom analizira samo sentiment, što je subjektivan stav autora o nekome ili nečemu i onda se on određuje kao pozitivan ili negativan. Mi ponovno želimo korak dalje."

A to znači analizirati namjere onoga tko piše. Otkriti koje su njegove emocije u tom trenutku.

"Želiš čitati između redaka", prekidam ga.

"Točno to. I tim čitanjem spajali bismo njegove emocije, namjere ali i njegov stav, kako bi dobili mentalni model autora teksta ili objave na Facebooku ili Twitteru", odgovara.

"To je jako ambiciozno", kažem.

"Pa pitao si gdje se vidimo za pet godina. E pa tu se vidimo. Da smo završili sustav koji bi automatski mogao prepoznati što je autor zapravo htio reći. A to dosad nitko nije uspio, barem koliko mi znamo. Iako sigurno jako puno ljudi i sada na tome radi."

Primjena je mnogo. Znalo bi se što ljudima smeta, što žele, kakvi su im stavovi. Komercijalna primjena je u sustavima poput e-gradani, e-Vlada, u centrima korisničke podrške na društvenim mrežama, pa čak i u prepoznavanju terorizma.

"Učinili bismo veliku stvar. Uključili bismo ljude u donošenje odluka i strategiju državnog upravljanja."

Međutim kad Šnajder kaže da na tome rade ili da su to učinili, priznajem mu, uopće mi nije jasno kako to 'rade'.

Što to znači?

"Vrlo pojednostavljeno?" pita me i smije se.

Kimam glavom.

Objašnjava mi da zajedno sa svojim studentima i suradnicima prvo jako puno vremena potroši proučavajući podatke koji su im dostupni na webu. Jednom kad nađu ono što im treba, a u ovim svim projektima o kojima smo razgovarali, to je tekst ili komad teksta, s njim se obrate ispitanicima.

"Pitam ih: što vi mislite da je autor htio reći, kako se osjećao."

Jednom kad tekst oboga novim informacijama, onda grade model na tehnikama strojnog učenja, odnosno izgradnji algoritama koji na temelju podataka koje obrađuju mogu stvarati predviđanja, odnosno 'razmišljati'.

"A to je jedna vrsta umjetne inteligencije", kaže Šnajder. U pozadini toga je mukotrpan rad na obradi jezika i statističkim modelima.

ROBOTI Dječake snove o robotima u srednjoj školi Šnajder je zamijenio Star Trekom. I tamo ga je najviše fascinirala mogućnost da ljudi mogu komunicirati sa strojevima i postavljati im kompleksna pitanja i na njih dobivati smislene odgovore.

"To i prizori iz Kubrickove Odiseje u svemiru u kojima astronauti komuniciraju sa superračunalom HAL-om, pa dijelovi Spielbergova filma AI koji su me se dojmili..." nabraja filmove koji su mu i danas inspiracija.

U literaturi voli djela Stanislava Lema, poljskog pisca sf-a, filozofa i satiričara. Lem je autor kultnog romana "Solaris", koji je dosad triput ekraniziran.

Šnajder posebno ističe knjigu koju preporučuje svim svojim studentima. Riječ je o "Gödel, Escher, Bach" Douglasa Hofstadtera.

U njoj se kroz biografije matematičara i logičara Kurta Gödela, umjetnika M.C. Eschera i kompozitora Johanna Sebastian Bacha izvlače teze o matematici, simetriji i inteligenciji.

"Riječ je o originalnom pogledu na jezik i ljudsko zaključivanje, ograničenja uma i umjetnu inteligenciju. Ova knjiga savršeno objašnjava umjetnu inteligenciju i granice ljudskog uma", tumači mi.

Šnajder vjeruje kako su rezultati koje hrvatski studenti postižu zapravo fantastični, s obzirom na uvjete u kojima funkcioniraju.

"Nemojmo biti naivni. Znanost košta. Ako želimo vrhunske rezultate i društvo znanja, onda u njega moramo ulagati. Mi smo po ulaganju na začelju Europe. Ako se to ne promijeni, šteta će uskoro biti nepopravljiva", zaključuje.

Prvi put ga ne pitam da mi pojasni što je želio reći.

Nemojmo biti naivni, znanost košta. Ako želimo društvo znanja, onda u njega moramo ulagati. Umjetna inteligencija najviše prijeti zapošljavanju. Odgovor na to, paradoksalno, upravo je inteligencija: sposobnost društva da se prilagodi novoj situaciji. Odmah moramo obrazovati mlade za poslove koji će postojati tek za 20 godina, kaže Jan Šnajder