

F GORAZD SUHADOLNIK  
finance@finance.si

**P**rihodnost umetne inteligence je v globljem razumevanju sveta, pravi Marko Grobelnik, raziskovalec računalniških znanosti, projektni vodja iz Laboratorija za umetno inteligenco na Inštitutu Jožefa Stefana, eden od naših najboljših strokovnjakov za analizo podatkov, natančneje besedilno rudarjenje oziroma analiziranje velikih količin tekstovnih podatkov.

Grobelnik pravi, da gre pri analizi podatkov za temeljno tehnologijo, ki lahko rešuje veliko zahtevnih problemov v raznih kontekstih. V okviru umetne inteligence je pomembna pri strojnem učenju (machine learning) in podatkovnem rudarjenju (data mining). »Danes ni relevantnega področja znanosti in inženirstva, ki ne bi uporabljalo dela analitičnih tehnik, gotovo to velja za spletne in jezikovne tehnologije, najbolj priljubljena področja pa so upravljanje odnosov s strankami, varnost in upravljanje proizvodnje.«

#### Računalnikne razume zapletenosti našega sveta

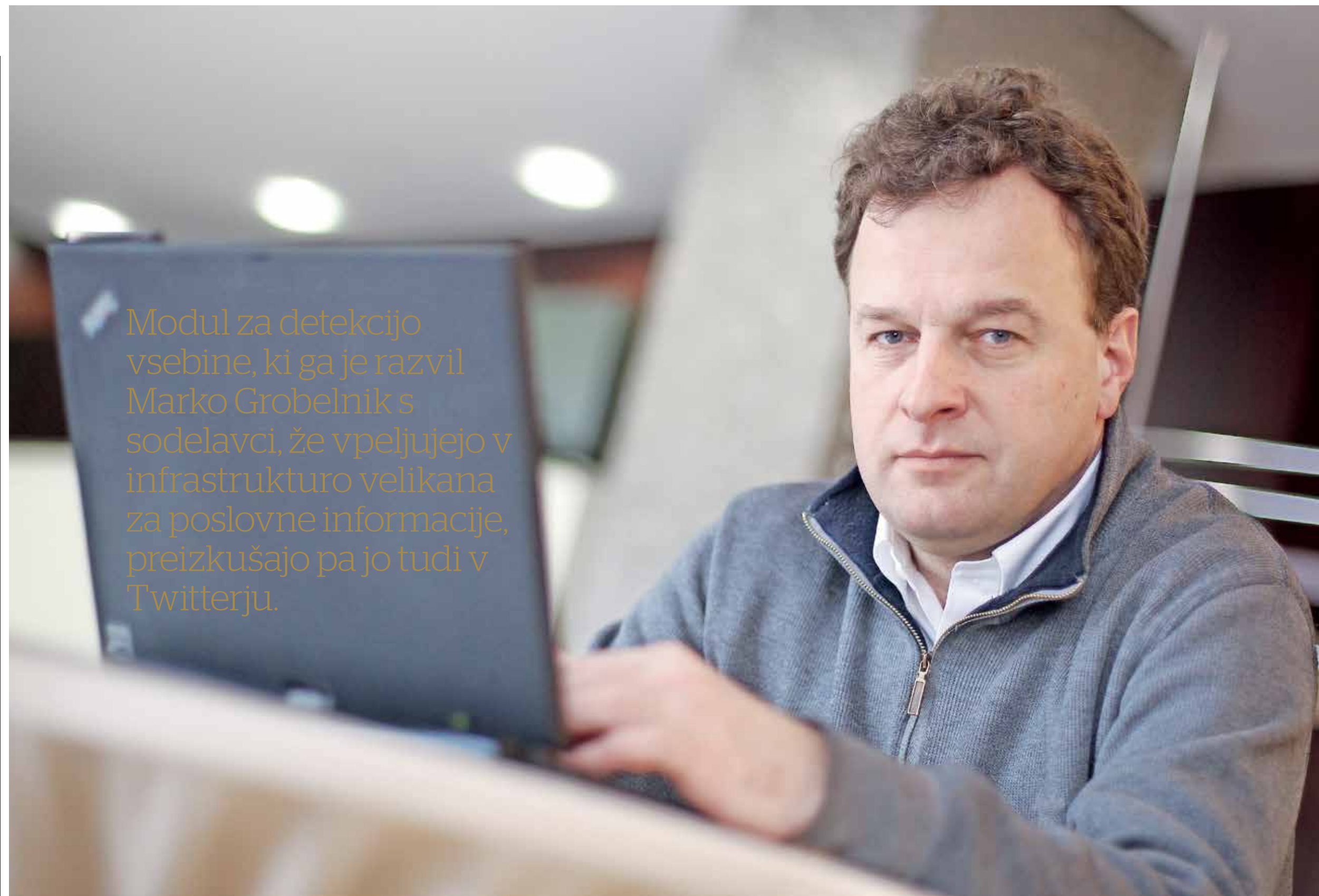
V razvoj umetne inteligence globalno največ vlagajo industrija, vojska in varnostne agencije, eno od bolj podarjenih področij pa je obvladovanje velikanskih količin podatkov – big data oziroma data science. Precej daljnosežne posledice pri delu s podatki ima uporaba »semantičnih tehnologij«, ki omogočajo povezovanje in razumevanje ter s tem kakovostne sinergije podatkov v sistemih. »Večna težava, s katero se ukvarjamo tudi na IJS, je razumevanje besedila. Težava je na videz preprosta, vendar celo petletni otrok veliko bolje razume besedilo od najzmogljivejšega računalnika, ki pač ne razume konteksta. Še vedno nam manjka jezik za opis kompleksnosti sveta – kako ga opisati, da bi ga računalnik razumel,« pravi Grobelnik.

Grobelnik poudarja, da je treba pri tem metode strojnega učenja razumeti predvsem kot pripomoček za obvladovanje kompleksnosti v podatkih, ki je ročno ne bi mogli obvladati v enakem času kot računalnik. »Včasih se zdi, da to ni povsem jasno, vendar računalnik ne počne ničesar, kar človek ne bi mogel početi ročno – razlika je le v hitrosti in obsegu reševanja problemov.«

Zelo aktualen problem je analiza družabnih omrežij in posledično povezovanje statistično-analitičnih tehnik s humanističnimi vedami. V bistvu gre za razumevanje delovanja družbe, njene dinamike in komunikacije na raznih ravneh, pravi Grobelnik. »S podrobnim opazovanjem sveta, ki ga omogočajo računalniki, se odpira možnost globljega razumevanja socialnih procesov in uporabe tega znanja pri napovedovanju ter povratnem vplivanju na družbo.«

#### Modul, ki »prepozna« več kot sto jezikov

Primer inovativne tehnologije, ki so jo pred nedavnim razvili v Laboratoriju za umetno inteligenco, je modul za detekcijo vsebine v dokumentih, ki »prepozna« več kot sto jezikov. Vendar dokumentov prej ni treba prevesti in je zato operacija izvedena veliko hitreje kot s strojnimi pre-



Modul za detekcijo vsebine, ki ga je razvil Marko Grobelnik s sodelavci, že vpeljujejo v infrastrukturo velikana za poslovne informacije, preizkušajo pa jo tudi v Twitterju.

# slovenski um, KI PRODAJA BLOOMBERGU, TWITTERJU ...

vajanjem. »Z vmesnim statističnim jezikom je mogoče zgraditi primerjavo med vsemi temi jeziki, kategorizacij, funkcije ... Gre za cross linguality, v slovensčini čezjezičnost oziroma povezovanje jezikov, prek katerega lahko primerjamo denimo slovenski in

kitajski dokument brez strojnega prevajanja,« pojasnjuje Marko Grobelnik. Modul za detekcijo vsebine že vpeljujejo v infrastrukturo vodilnega medijskega podjetja za poslovne in finančne informacije Bloomberg, preizkušajo pa jo tudi v Twitterju.

#### Register dogodkov, ki sestavi zgodbo

S čezjezičnim modulom kot osnovnim gradnikom razvijajo tudi sistem Event Registry za opazovanje globalne socialne dinamike (<http://eventregistry.org>). »Naš cilj je izde-

lava sistema, ki v realnem času zaznava večino svetovnih novic v raznih jezikih, iz tega oblikuje dogodnim gradnikom razvijajo tudi sistem Event Registry za opazovanje globalne socialne dinamike (http://eventregistry.org). »Naš cilj je izde-

V bazi sistema Event Registry je 40 tisoč glavnih medijev z vsega sveta, 250 tisoč najpomembnejših blogov in tvitov. Sistem zajema novice v bolj ali manj v realnem času (od 15 do 20 minut po objavi), jih zbira v sveženi oziroma grozd ter oblikuje v dogod-

ke z veliko članki. Tako povezuje članke o isti temi z vsega sveta in iz njih na koncu oblikuje zgodbe. »Takšna platforma bo podlaga za vrsto aplikacij, ki potrebujejo obvladovanje in razumevanje globalne socialne dinamike. Event Registry je zelo uporaben za ti-

skovne agencije, finančne institucije in okoljske organizacije. Marko Grobelnik ga je pred dvema mesecema predstavil na letnem srečanju organizacije IPCT (mednarodni svet za tisk in telekomunikacije), glavnega telesa za standardizacijo novičarske tehnologije, v katerem sodelujejo predstavniki vseh globalnih tiskovnih agencij. »Vsi vidijo v tem sistemu priložnost za korak naprej,« pravi Grobelnik.

»Navzven je Event Registry videti preprost, uporabniško prijazen, v ozadju pa je ogromno tehnologije: reševati mora vprašanja, kako sploh pridobiti vse te članke, kako besedila najprej obdelati lingvistično, potem semantično, potem čezjezično, iz tega oblikovati dogodke in potem vse skupaj vizualizirati. Vse to mora delovati nepretrgoma,« pojasnjuje sogovornik.

#### Razvija tehnologije, v katere tehnološki velikani mečejo milijarde

Grobelnik je z nekaterimi sodelavci tudi soustanovitelj CyCorp Europe, evropske veje ameriškega podjetja CyCorp, ki se že leta preživlja s tradicionalno umetno inteligenco, utemeljeno na globokem logičnem znanju o svetu. »Vse aplikacije tega sistema izhajajo iz razumevanja in razlage sveta. Četudi je reševanje problemov s takimi pristopi precej zahtevnejše in dražje, je možno po tej poti priti do rezultatov, za katere je navzven videti skoraj neverjetno, da bi jih izdelal računalnik. Tehnologija, ki jo uporablja CyCorp, poskuša razumeti svet v podrobnosti in v tem smislu prihaja do zelo zanimivih konceptualnih preskokov. Za nas pomeni sodelovanje s CyCorpom pomembno razširitev pristopov, ki jih uporabljamo pri reševanju vprašanj modeliranja sveta. Lahko bi rekli, da je pristop CyCorpa še vedno pristop prihodnosti, ta trenutek pa je še prezgodaj za široko komercialno uporabo tovrstnih tehnologij. Pri raziskovalnih projektih tovrstne pristope čedalje intenzivneje uporabljamo, Google, Microsoft, Facebook in IBM pa so pred nedavnim začeli intenzivno vlagati v tovrstne tehnologije (na primer Google Knowledge Graph in IBM Watson). V prihodnjih petih do desetih letih lahko zato pričakujemo pomembne preboje na področju globljega razumevanja sveta skozi orodja umetne inteligence,« pravi Grobelnik in poudarja, da so pri tem slovenski raziskovalci »na tem vlaklu, povezani z vsemi drugimi središči v svetu, v katerih se ukvarjajo s tem področjem.«

#### V deset letih 20 milijonov evrov

Medtem se v domovini pri prenosu znanja, pridobljenega z rudarjenjem podatkov, v tržne izdelke in storitve še vedno rado zaplete, pravi Grobelnik. »S komercializacijo imamo v Sloveniji precej težav zaradi pomanjkanja tovrstne tradicije. V zahodni Evropi je položaj nekoliko boljši, vendar v primerjavi z ZDA vsi skupaj zelo zaostajamo. Še vedno žal držijo opažanje, ki velja v Evropi, neformalno pa se pogosto sliši v Bruslju, da je cilj večine evropskih start-upov, da se prodajo na ameriški trg. Razlogov za to je več, vsekakor pa k temu pomembno pripomoreta evropska togotost in nenaklonjenost tveganim poslovnim potezom.«

V Sloveniji je sicer področje umetne inteligence dober razvit, saj je na IJS ter ljubljanski in mariborski



**V bazi sistema Event Registry je 40 tisoč glavnih medijev z vsega sveta, 250 tisoč najpomembnejših blogov in tvitov. Sistem zajema novice v bolj ali manj realnem času (od 15 do 20 minut po objavi), jih zbira v sveženi oziroma grozd ter oblikuje v dogodke z veliko članki.**

univerzi na tem področju okoli 150 polno zaposlenih. V Laboratoriju za umetno inteligenco jih je trenutno približno 50. »Če bi moral povzeti delovanje naše raziskovalne skupine (ailab.ijs.si), bi uporabil tri pojme: zavzetost, zaupanje in kompetentnost. Ključ do uspeha raziskovalne skupine pa so ponavadi aktual-

no znanje, ki se stalno obnavlja, denar, ki omogoča intenzivno in inovativno delo in navzočnost, ter aktivne povezave z glavnimi centri po vsem svetu.«

Grobelnik pravi, da jim denarja ne primanjkuje, lokalna kriza v Sloveniji pa ne vpliva na njihovo delo, saj pridobijo več kot 95 odstotkov sredstev na mednarodnem trgu. »Denarja za raziskave v Sloveniji nikoli ni bilo prav veliko, zato smo se usmerili v pridobivanje sredstev na evropski ravni in v tem smo v Sloveniji z naskokom najuspešnejši: v desetih letih smo pridobili več kot 40 evropskih projektov oziroma več kot 20 milijonov evrov. To nas uvršča tudi med najuspešnejše tovrstne skupine v Evropi. Tak pogon seveda generira veliko rezultatov, zato smo ustanovili nekaj odcpepljenih (spin-off) podjetij, ki del teh rezultatov razvijajo do tržnih izdelkov.« Med njimi sta sistem za hitro analizo velikih baz podatkov (QMiner) in omerjen sistem za analitsko spremljanje globalnega medijskega prostora. Produkti bodo predvidoma na trgu prihodnje leto.

## PTUJČAN BO V GOOGLU RAZVIJAL UMETNO INTELIGENCO



■ Ptujčan **Ruben Sipoš** je že v osnovni šoli uspešno tekmoval v računalništvu in matematiki, po končanem interdisciplinarnem študiju računalniške matematike v Ljubljani in prvih delovnih izkušnjah v Laboratoriju za umetno inteligenco na IJS pa se je odločil za »malo tveganje« in se prijavil na doktorski študij na sloviti ameriški univerzi Cornell. Doktorat je končan, jeseni ga čaka služba v kalifornijski centrali podjetja Google.

■ Večji del Sipoševih doktorskih nalog se osredotoča na strojno učenje in strnjevanje teksta (podatkov) oziroma raziskovanje, kako besedilo z avtomatskimi pristopi strniti v obliko, primernejšo za branje. »Poskus avtomatiziranja strnjevanja podatkov z uporabo računalnika, ki se pri tem na podlagi primerov tudi uči, pripomore k temu, da je v prihodnje ta postopek boljši, hitrejši in sam proizvajajo boljše strnjevanje povzetke,« pravi Ruben.

■ V nalogi se Ruben posebej posveča ocenam izdelkov oziroma vprašanju, kako ugotoviti, katere recenzije so za uporabnike koristne. »Tudi pri tem gre za strnjevanje. Če imamo tisoč recenzij, lahko uporabniku rečemo: teh pet se splača prebrati, v njih je strnjena vsebina vseh tisoč recenzij.«

■ Sipoš tudi pravi, da je pri analizi recenzij produktov med drugim ugotovil, da je odgovor na vprašanje, ali je recenzija uporabniku spletna strani pomagala pri odločitvi (Was This Review Helpful to You), odvisen tudi od tega, katere druge recenzije so bile v tistem trenutku prikazane v okolici obravnavane recenzije. »Če bi na primer brali recenzijo, okoli katere so vse druge zelo slabe, bi se vaše dojemanje trenutne recenzije spremenilo tako, da bi jo videli kot boljšo.« Dogajanje, pri katerem je glas za »da« ali »ne« odvisen od kakovosti recenzije in okoliščin prikaza, je mogoče po Rubenovih besedah opisati s precej preprostimi modeli, spletna strani, ki objavljajo recenzije, pa lahko tako izboljšajo izbiro in prikaz najboljših recenzij.

■ Sipošev model ima le eno »lepoto napako«: na spletnih straneh s porabniškimi ocenami kar mgolgi ponarejenih recenzij, napisanih po naročilu in za plačilo. »Na to temo se izvaja ogromno raziskav, v katerih ugotavljajo, katere recenzije so ponarejene, koliko jih je in kako to vpliva na porabnike,« odgovarja Ruben. »Udi ugotovitve iz mojega doktorata je mogoče uporabiti v obe smeri: če poznaš vedenje porabnikov, lahko bolje napišeš ponarejeno recenzijo, da jo je potem težje odkriti, po drugi strani pa dajejo dodatno znanje in nekaj več smernic, kako se lotiti odkrivanja ponarejenih recenzij. Večinoma gre za igro mačke in miši, napredek na eni strani sproži povečana prizadevanja za napredek na drugi strani.«

■ Ruben Sipoš namerava v Googlu nadaljevati delo, s katerim se ukvarja zadnja leta – umetno inteligenco, delo s teksti in povzetki.