

Uvod u obradu prirodnog jezika

Uvod

Branko Žitko

prevedeno od: Dan Jurafsky, Chris Manning

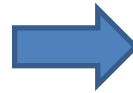
Što je obrada prirodnog jezika?

- **Odgovaranje na pitanja (Question answering)**

- IBM-ov Watson je pobijedio u kvizu Jeopardy

Voditelj

Fox show featuring characters named Itchy and Scratchy

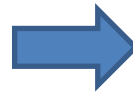


Watson

Who are the Simpsons?

Voditelj

William Wilkinson's "An Account of the Principalities of Wallachia and Moldavia" inspired this author's most famous novel



Watson

Who is Bram Stoker?



Što je obrada prirodnog jezika?

- **Ekstrakcija informacija (Information extraction)**

Subject: Sastanak zavoda

Date: 8. veljače 2012.

To: Branko Žitko

Sastanak zavoda će se održati sutra u sobi 30 od 12:00 do 13:30.

Pozdrav



Događaj: Sastanak zavoda

Datum: 09.02.2012.

Početak: 12:00

Kraj: 13:30

Mjesto: Soba 30

Što je obrada prirodnog jezika?

- **Ekstrakcija informacija i sentimentalna analiza**



Veličina i težina

- ✓ Lijepo i kompaktno za nositi
- ✓ Dok je kamera mala i lagana, ona teške, glomazne, profesionalne
- ✗ Kamera izgleda veoma krhka, stoga morate biti veoma oprezni

Već na prvi pogled, kvaliteta izrade ostavlja dobar dojam. Tijelo je čvrsto i odaje pozitivan **dojam** kojeg samo malo kvare pojedine tipke sa "gnjecavim" osjećajem pri pritisku. Kotačić za odabir načina rada se vrti relativno lako te ga je moguće nehotice zakrenuti pri vađenju ili vraćanju aparata u džep ili torbicu, te je potrebno ponekad obratiti pozornost na njega prije fotografiranja. Zakretni LCD zaslon je obješen na potpuno metalnu konstrukciju koja je vrlo čvrsta i djeluje kao da može izdržati godine svakodnevnog zakretanja i navlačenja u svim smjerovima. Ponekad nas je smetao **nedostatak većeg gripa** tako da će E-PL5 djelovati nespretno u većim rukama, no to je subjektivno. U redakciji se ipak svi slažemo da najbolje ležanje u ruci pružaju Sony NEX-5 i NEX-6 koji imaju izbočeni grip. Glavni **izbornik** je lagan za korištenje, a mnogobrojne opcije logično raspoređene. Najbitnije opcije poput ISO vrijednosti, balansa bijele i sličnih postavki se podešavaju putem brzog izbornika koji se poziva "OK" tipkom. Iako PEN E-PL5 daje odlične fotografije već u automatskom modu i prikladan je za korištenje i kompletnim amaterima, svoju punu vrijednost pokazuje u rukama naprednih korisnika.

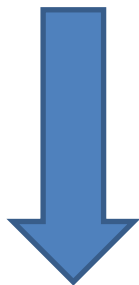
Olympus naime nikad nije škrtario sa opcijama pa tako E-PL5 ima niz detalja koje nedostaju mnogim profesionalnim DSLR aparatima, kao što je bežična kontrola do tri skupine bljeskalica, podešavanje brzine kontinuiranog okidanja od 1 do 8 fotografija u sekundi, pixel mapping, mogućost potpune preraspodjele kontrola na tipkama, direktni odabir fokusnih točaka i njihove veličine, mijenjanje smjera rotacije kontrolnog kotačića, čak 6 vrsta bracketinga (ISO, WB, Flash, Art filteri, HDR, AE), promjena EV i ISO koraka, antivibracijska odgoda okidanja sa podesivim vremenom, zasebna korekcija sva tri načina mjerenja svjetla... opcijama nema kraja!

Što je obrada prirodnog jezika?

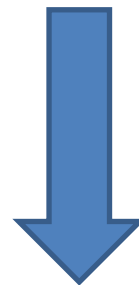
- **Strojno prevođenje (Machine translation)**
 - Google Translate

这不过是一个时间的问题

potpuno automatski



This however is a matter of time



kao pomoć
ljudskim prevoditeljima

To, **ipak**, je pitanje vremena

- međutim
- ma kako
- kako god
- ma koji

Jezične tehnologije

Uglavnom riješene

Detekcija SPAM-a (SPAM detection)

napravi dijagram klasa.



buy V1AGRA ...



Označavanje dijelova teksta (Part of speech tagging)

ADJ

ADJ

NOUN

VERB

ADV

Colorless green ideas sleep furiously.

Prepoznavanje imenovanih entiteta (Named entity recognition)

PERSON

Po mišljenju **Orsata Frankovića**, direktora studija
Laboratorium, kad **Hrvatska** postane...

ORG

LOC

Jezične tehnologije

Dobro napreduju

Sentimentalna analiza

Najbolji restoran u Splitu!



Konobar nas je ignorirao 20 minuta!



Određivanje smisla riječi (Word sense disambiguation)

Trebam nove baterije za mog **miša**.



Strojno prevođenje (Machine translation)

今天是美好的一天

Danas je lijep dan.

Rješavanje odnosa (Coreference resolution)

On je ulovio muhu u letu.



Parsanje (Parsing)

Mogu vidjeti Šoltu s prozora!



Ekstrakcija informacija (Information extraction)

Pozvani ste na zabavu
sljedeći Petak 27 Svibnja
u 8:30



Zabava
27.05.2012
20:30

Jezične tehnologije

Još uvijek jako teško

Odgovaranje na pitanja (Question answering)

Kolika je učinkovit ibuprofena u smanjenju povišene temperature kod bolesnika sa akutnom bolešću?

Parafraziranje

XYZ je stekao ABC jučer.

ABC je preuzeo XYZ

Sažimanje (Summarization)

Fondovi idu prema gore
Dionice su skočile
Cijene nekretnina rastu



Ekonomija je dobra

Dijalog

Gdje se prikazuje
Gospodar prstenova?



U kinu Central u
19:30. Želite li
rezervirati kartu?



Obrada prirodnog jezika

- Višeznačnost čini obradu prirodnog jezika teškom

Brat mi je slomio ruku

Mnogi naučnici **sumnjaju** da se u pećini kriju značajna arheološka blaga

Iskopao je crno zlato u dvorištu

Popila je kavu i vruću čokoladu sa **šlagom**