

RAZUMEVANJE TEKSTA:

OSNOVNI POJMOVI; IZAZOVI;

MOGUĆNOSTI PRIMENE

Jelena Jovanović

Email: jeljov@gmail.com

Web: <http://jelenajovanovic.net>

Pregled izlaganja

- Analiza i razumevanje teksta:
 - Zašto je potrebno/bitno? Koje su oblasti primene?
 - U čemu se ogleda složenost razumevanja teksta?
 - Koji su osnovni pristupi analizi i razumevanju teksta?

Razumevanje teksta: primeri primene

- Kontekstno-zasnovana prevera pravopisa
 - Provera i korekcija sintaksne i gramatičke ispravnosti teksta
- Semantičko pretraživanje
 - Naprednije u odnosu na tradicionalno pretraživanje zasnovano na ključnim rečima
- Ekstrakcija informacija
 - Ekstrakcija entiteta i njihovih međusobnih relacija iz teksta najrazličitijeg tipa (izveštaji, Web stranice, status updates,...)
- Mašinsko (automatsko) prevođenje

Razumevanje teksta: primeri primene

- Novi interfejsi
 - Sistemi zasnovani na dijalogu
- Poslovne primene:
 - reputation management
 - kontekstualizovano oglašavanje
 - poslovne analitike
 - ...

Razumevanje teksta: Izazovi

Složenost prirodnog jezika

Primeri:

Mary and Sue are sisters.

Mary and Sue are mothers.

Joe saw his brother skiing on TV. *The fool...*

... didn't have a jacket on!

... didn't recognize him!

Razumevanje teksta: Izazovi

Primeri

I **deposited** \$100 in the **bank**.

The river **deposited** sediment along the **bank**.

Today hundreds of **planes land** daily on JFK **runway**.

Planes that once were parked on his **land** now are rolling down the **runway**.

Razumevanje teksta: Izazovi

Primeri ukazuju da je prirodni jezik:

- Pun višesmislenih reči i izraza
- Zasnovan na korišćenju konteksta za definisanje i prenos značenja
- Pun fuzzy, probabilističkih izraza
- Baziran na zdravorazumskom znanju i rezonovanju
- Pod uticajem je i sam utiče na interakcije među ljudima

Analiza i razumevanje teksta: Izazovi

Slojevita struktura jezika:

- Koje reči su korišćene (u datom tekstu)?
- Koje fraze su prisutne?
- Koje reči i na koji način modifikuju jedna drugu?
- Koje je (bukvalno) značenje korišćenih reči i fraza?
- Šta se može zaključiti iz činjenice da je neko rekao nešto u datoj situaciji?
- Kakva reakcija se očekuje?

Analiza i razumevanje teksta: Izazovi

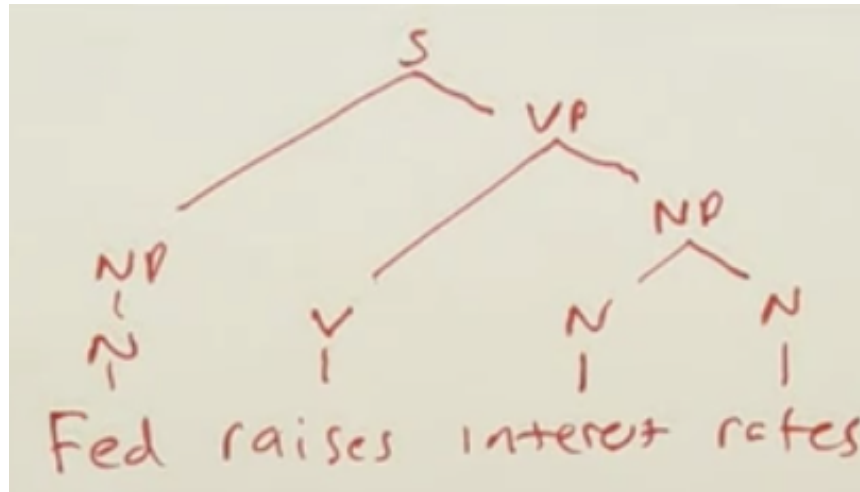
Vrsta analize	Opis	Primer
Morfologija	prepoznavanje reči i njihovih različitih formi	use, uses, user – različiti oblici jedne reči
Sintaksa i Gramatika	prepoznavanje tipa reči	There are 5 rows in the table. – rows je imenica; She rows 5 times per week. – rows je glagol
	razumevanje kako su reči povezane	Bob went out; <i>he</i> needed some fresh air. – Zamenica <i>he</i> se odnosi na Bob-a.
Semantika	razumevanje značenja reči (na osnovu konteksta)	The car <i>driver</i> was injured. vs. The <i>driver</i> was installed in the computer

Analiza i razumevanje teksta: Osnovni pristupi

■ Osnovni pristupi tj. načini modelovanja jezika:

– Logički modeli

- Zasnovani na detaljnoj lingvističkoj analizi teksta i formiranju apstraktnog modela strukture rečenica (tipično u formi stabla)
- Lingvistička analiza vodi ka ručnom kreiranju modela



Primer stabloidne strukture rečenice

Analiza i razumevanje teksta: Osnovni pristupi

■ Osnovni pristupi tj. načini modelovanja jezika:

– Stohastički modeli

- Zasnovani na verovatnoći pojavljivanja pojedinačnih reči ili niza od n reči (najčešće 2-4 reči)*
- Ovi modeli se “uče” tj. kreiraju se kroz primenu algoritama m. učenja nad velikim korpusima teksta

– Hibridni pristup

- Kombinuje logički i stohastički pristup
- Npr. pridruživanje verovatnoća pojedinim elementima stabloidnog modela strukture rečenice

* niz od n reči sa pridruženom verovatnoćom pojavljivanja naziva se ***n-gram***

Preporuka

Tema *Natural Language Processing* u okviru *Udacity.com* kursa *Introduction to Artificial Intelligence*

URL: <https://www.udacity.com/course/cs271>

Predavanje na temu *Natural Language Processing* održano u okviru *International Summer School on Semantic Computing, Berkeley 2011*

URL: http://videlectures.net/sssc2011_martell_naturallanguage/

(Anonimni) upitnik za vaše kritike,
komentare, predloge:

<http://goo.gl/cqdp3l>