

# Da li nas to AI razume?

Una Jaćimović

Matematička gimnazija

17. 05. 2023.

1. NLP i language modeli

2. Reprezentacija reči

3. Primena

Нова порука

Примаоци

Наслов

Dear Sir/Madam,

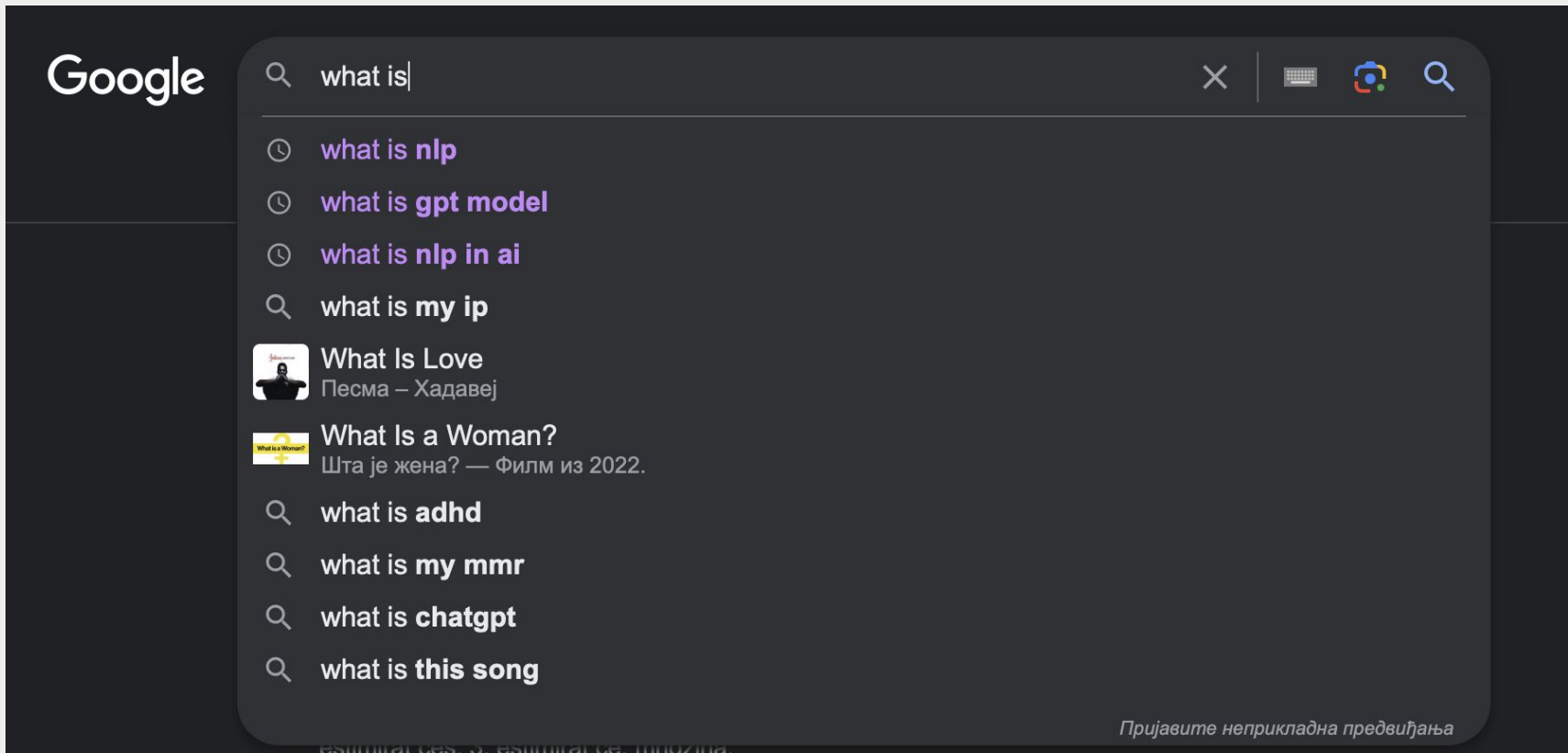
I hope this email finds you well



## Problemi:

- Prevođenje
- Sumarizacija teksta
- Klasifikacija teksta
- Prepoznavanje govora
- Izvlačenje informacija
- Generisanje teksta

- Procena verovatnoće za sekvencu reči



- Kako do modela?
- Šta su podaci?



1. NLP i Language modeli

2. Reprezentacija reči

3. Primena

- Mi imamo slova, a računari?

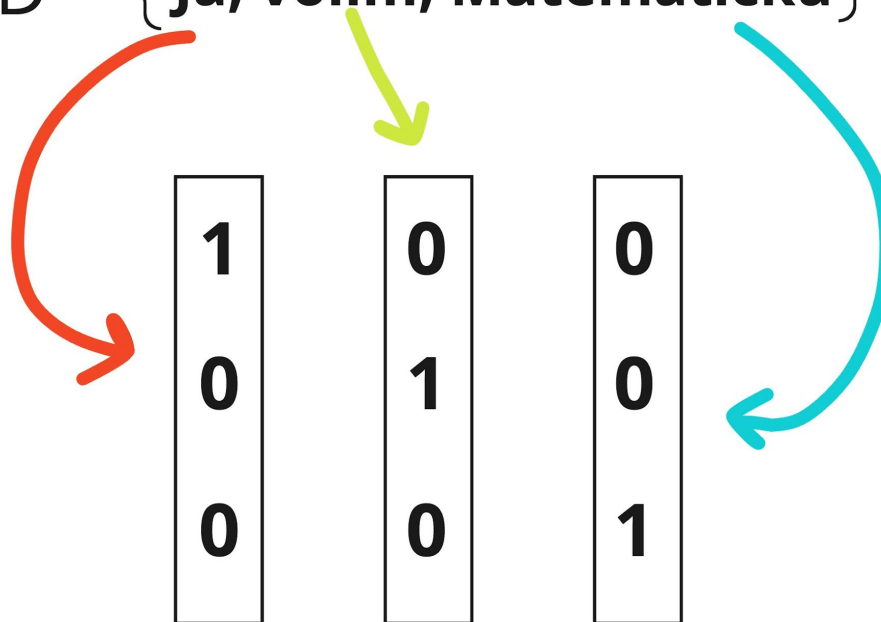




## One-hot vektori:

- Rečnik (vocabulary)
- Jedinica na mesto reči
- “unknown“ token

$D = \{ \text{Ja, volim, Matematičku} \}$

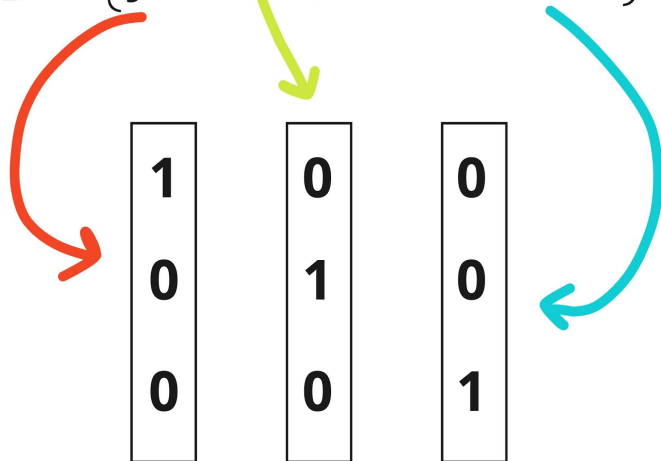


miro

Ali...

- dimenzije mogu biti ogromne
- nema prostora za nove podatke
- ne postoji veza između reči

$D = \{ \text{Ja, volim, Matematičku} \}$



miro



miro

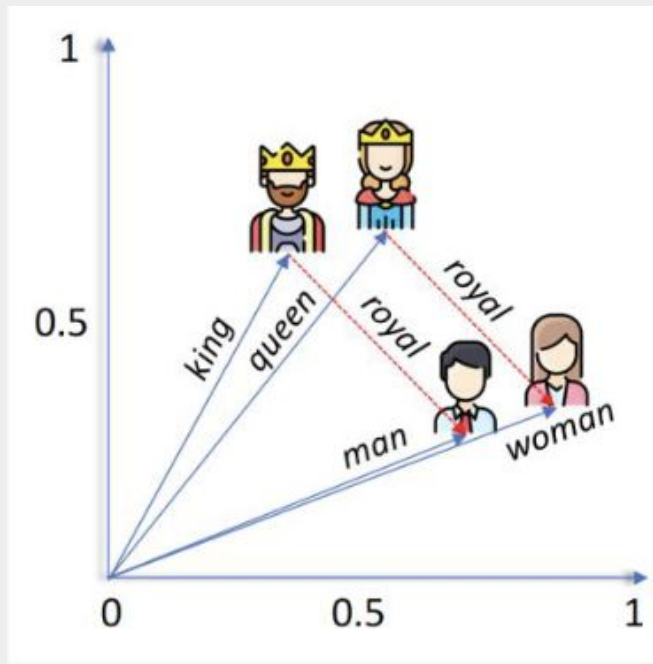
## Word-embedding

- učenje svojstava

	<b>Devojka</b>	<b>Dečko</b>	<b>Pas</b>	<b>Jabuka</b>	
Svojstva {	<b>Pol</b>	<b>-1</b>	<b>1</b>	<b>1</b>	<b>-0.5</b>
	<b>Godine</b>	<b>-0.01</b>	<b>0.01</b>	<b>0.03</b>	<b>0.04</b>
	<b>Hrana</b>	<b>0.02</b>	<b>0.02</b>	<b>-0.02</b>	<b>1</b>

miro





$$distanca_1 = Embedding[žena] - Embedding[muškarac]$$

$$distanca_2 = Embedding[žena] - Embedding[devojka]$$

$$distanca_1 > distanca_2$$

# Reprezentacija reči – opcija #2



Word embedding matrix

Word index

dimenzija vektora

<b>0.24</b>	<b>1.34</b>	<b>-0.27</b>	<b>0.18</b>
	...		
<b>0.73</b>	<b>0.3</b>	<b>0.45</b>	<b>0.23</b>

<b>0</b>
<b>1</b>
<b>0</b>
<b>0</b>

dimenzija rečnika

miro

# Reprezentacija reči – opcija #2



Word embedding matrix

<b>0.24</b>	<b>1.34</b>	<b>-0.27</b>	<b>0.18</b>
	...		
<b>0.73</b>	<b>0.3</b>	<b>0.45</b>	<b>0.23</b>

Word index

<b>0</b>
<b>1</b>
<b>0</b>
<b>0</b>

dimenzija vektora

dimenzija rečnika

miro

# Reprezentacija reči – opcija #2



Word embedding matrix

Word index

dimenzija vektora

<b>0.24</b>	<b>1.34</b>	<b>-0.27</b>	<b>0.18</b>
	...		
<b>0.73</b>	<b>0.3</b>	<b>0.45</b>	<b>0.23</b>

<b>0</b>
<b>1</b>
<b>0</b>
<b>0</b>

dimenzija rečnika

miro



# Reprezentacija reči – opcija #2



Word embedding matrix

<b>0.24</b>	<b>1.34</b>	<b>-0.27</b>	<b>0.18</b>
	...		
<b>0.73</b>	<b>0.3</b>	<b>0.45</b>	<b>0.23</b>

Word index

<b>0</b>
<b>1</b>
<b>0</b>
<b>0</b>

dimenzija vektora

dimenzija rečnika

miro

## Kako do vektora?

- Treniranje modela
- Postojeći dataset
- Finetuning



Možeš li da mi sipaš čašu soka

Možeš

indeks

E

li

indeks

E

da

indeks

E

mi

indeks

E

sipaš

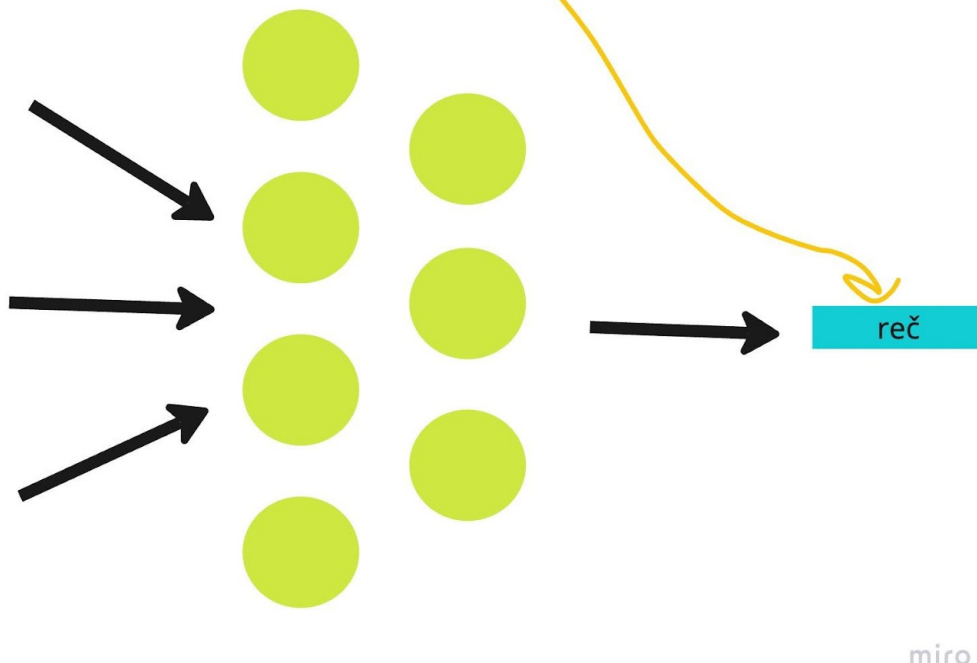
indeks

E

čašu

indeks

E



Možeš li da mi sipaš čašu soka

Možeš

indeks

E

li

indeks

E

da

indeks

E

mi

indeks

E

sipaš

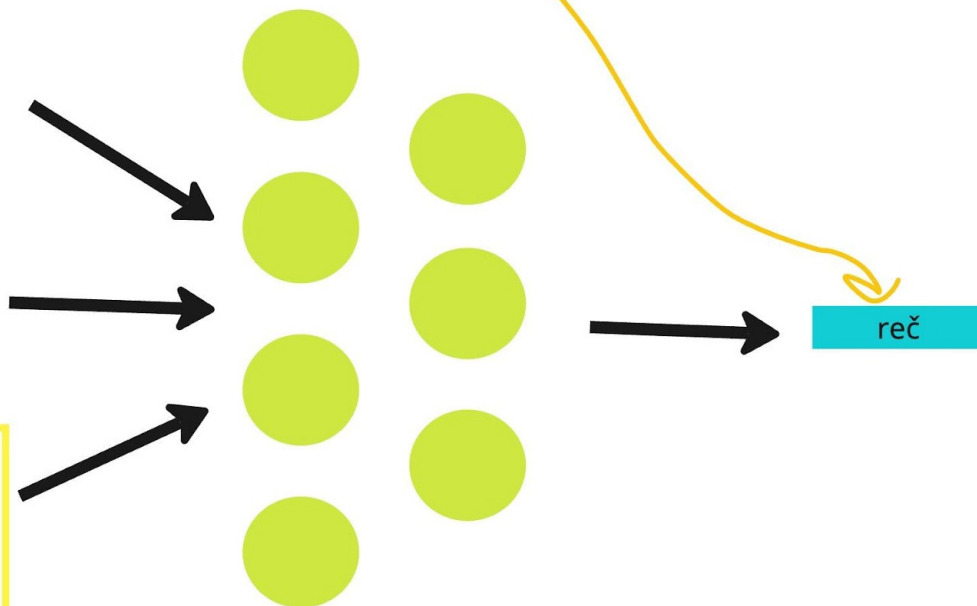
indeks

E

čašu

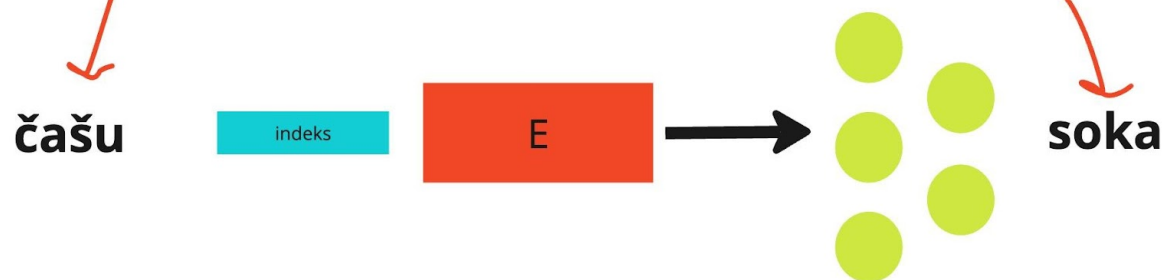
indeks

E

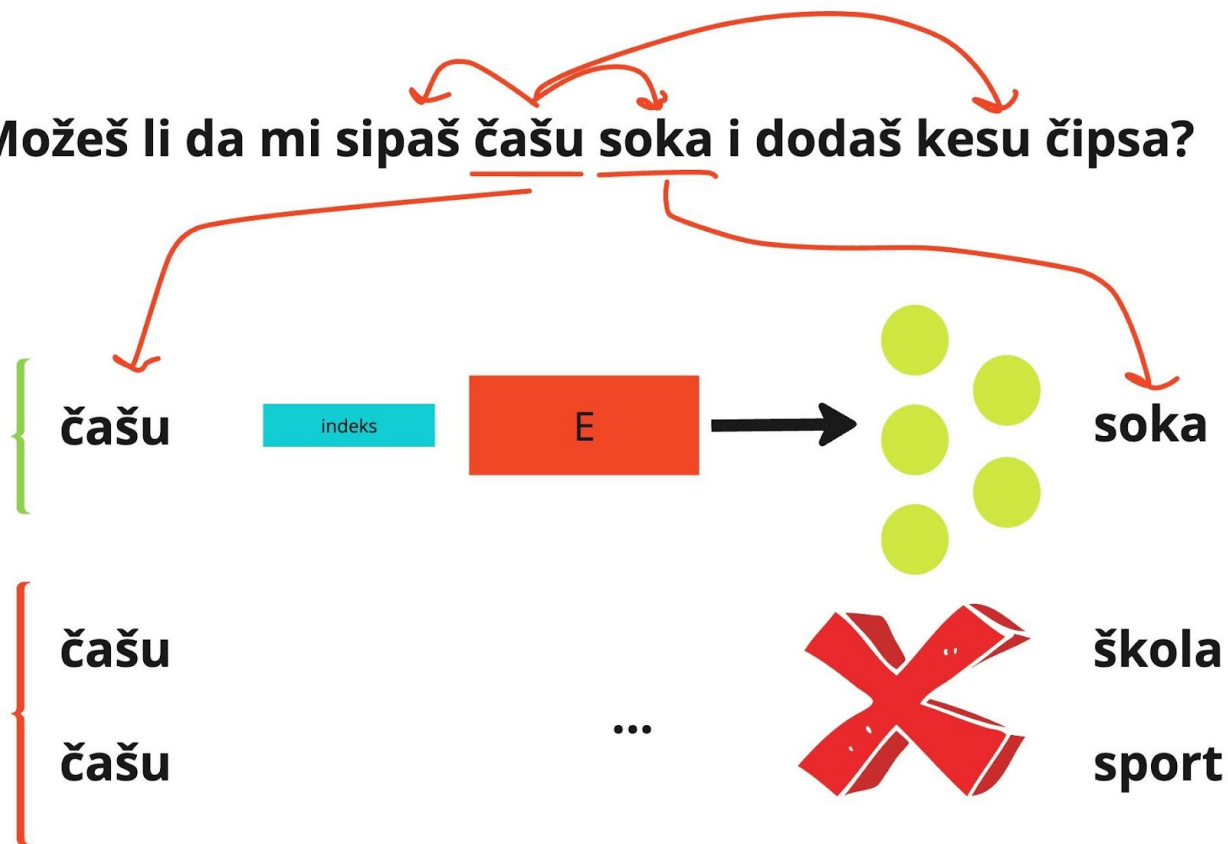


miro

Možeš li da mi sipaš čašu soka i dodaš kesu čipsa?



Možeš li da mi sipaš čašu soka i dodaš kesu čipsa?



1. NLP I language modeli

2. Reprezentacija reči

3. Primena

- Želimo da procenimo značenje teksta
- Imamo dostupne podatke

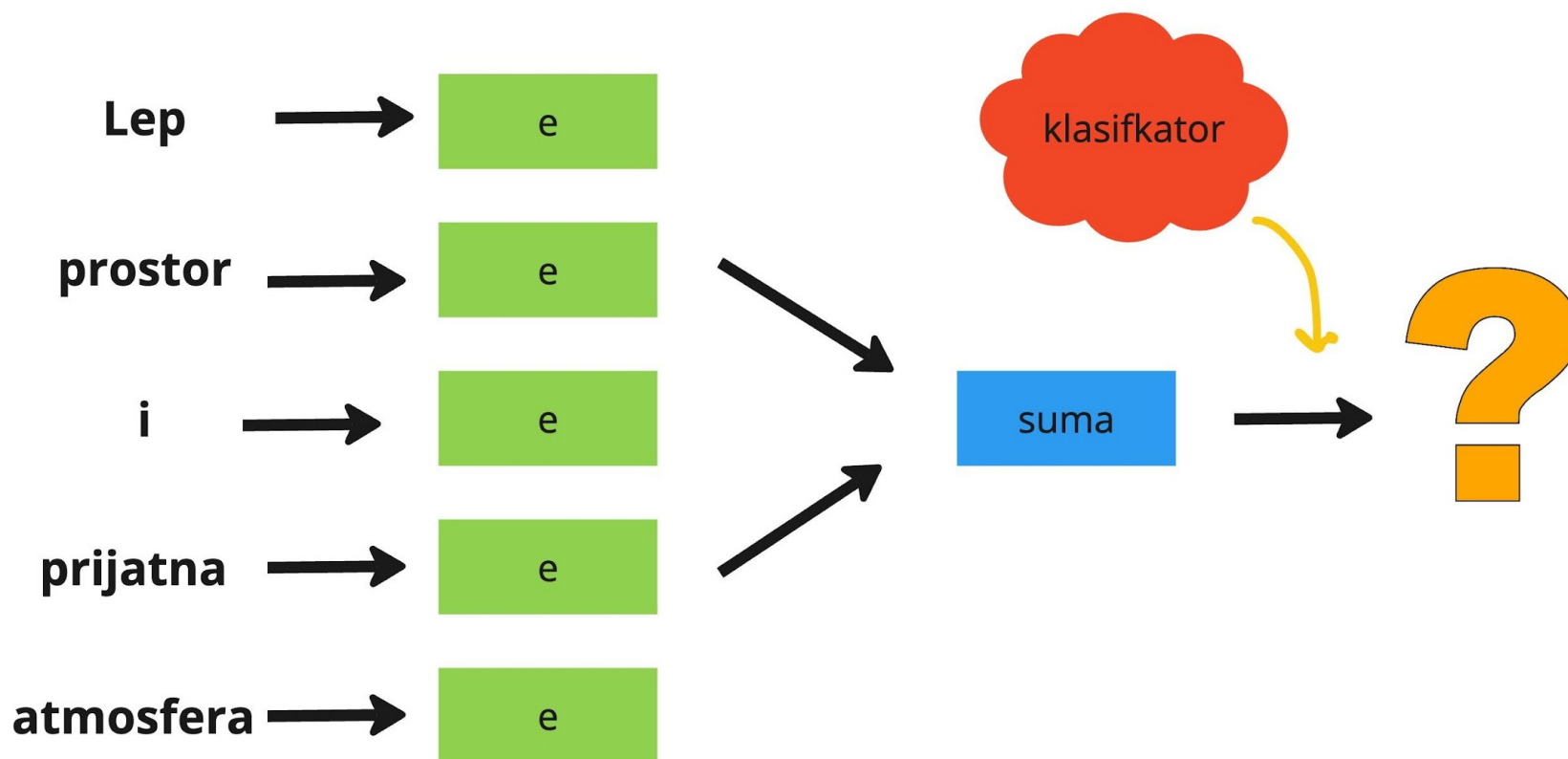






- Kako predstaviti tekst ?
- Šta sa dimenzijama?

Krenimo od nečeg jednostavnog...



Alii...

- red reči može biti problematičan
- pozitivne reči mogu da se pojave sa negacijom

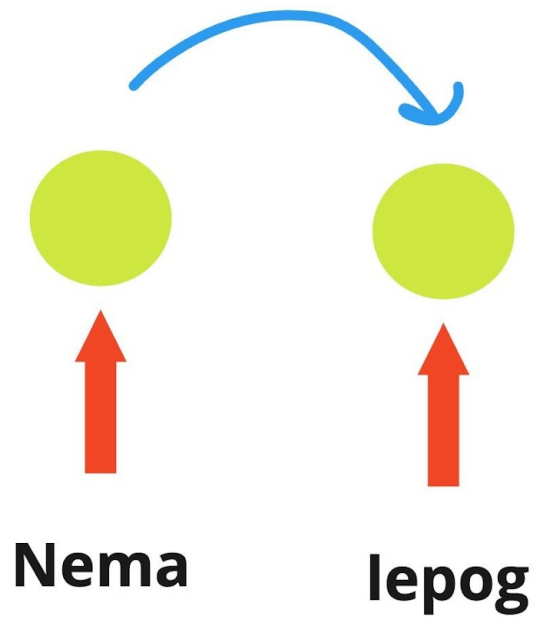


## RNNs

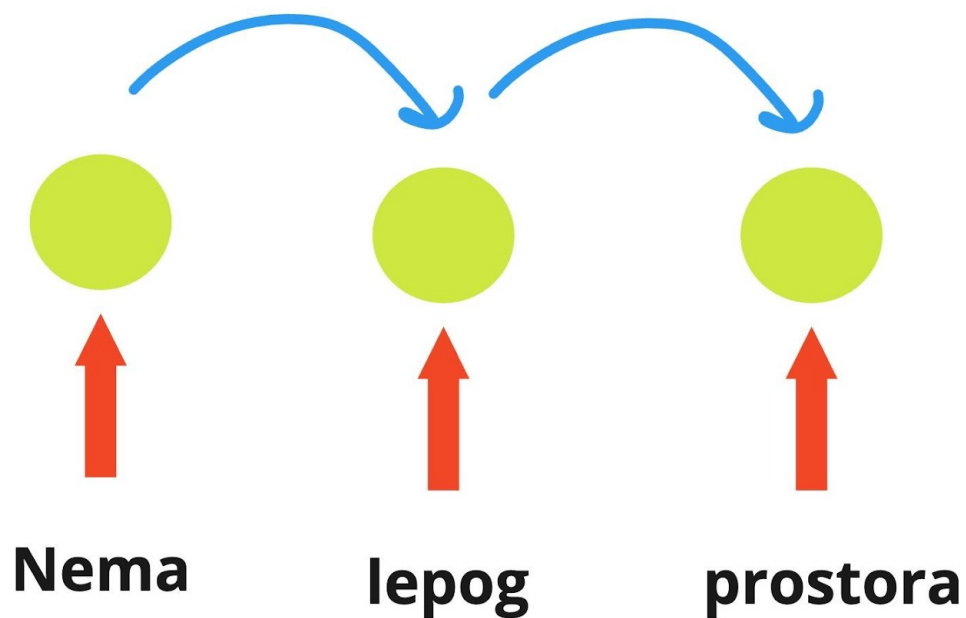


**Nema**

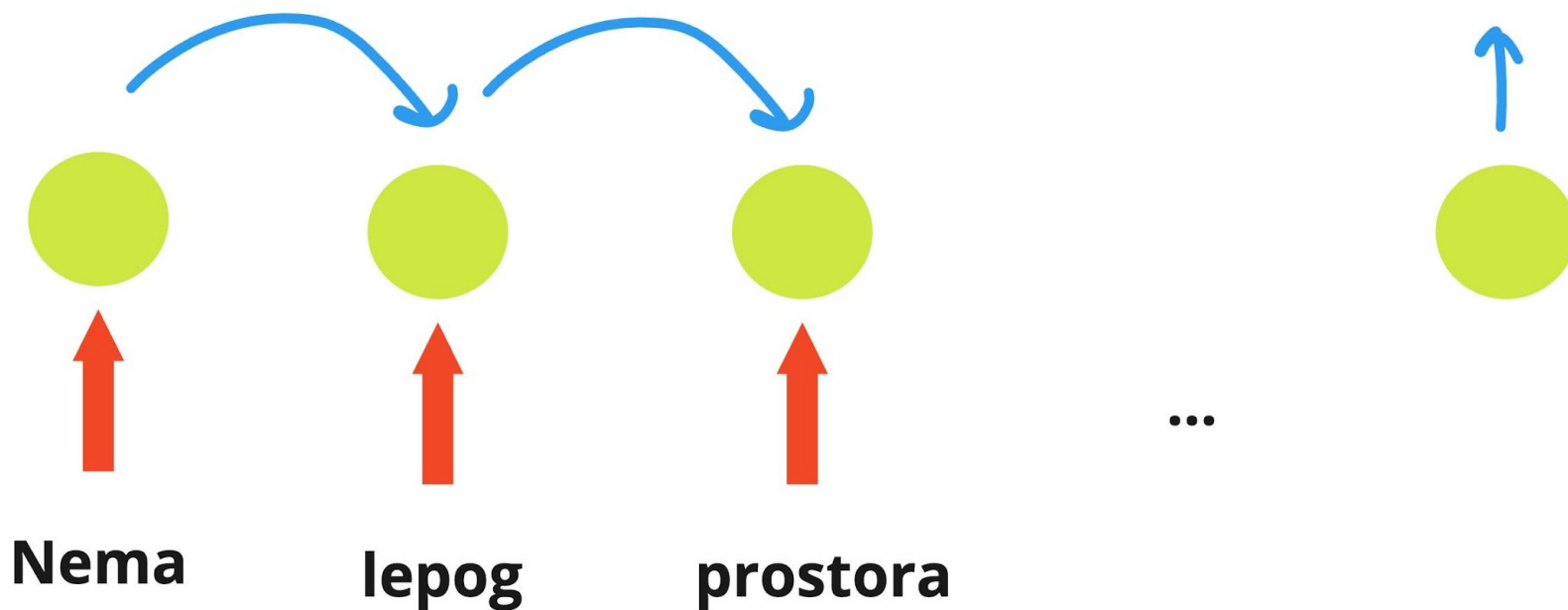
## RNNs



## RNNs



## RNNs

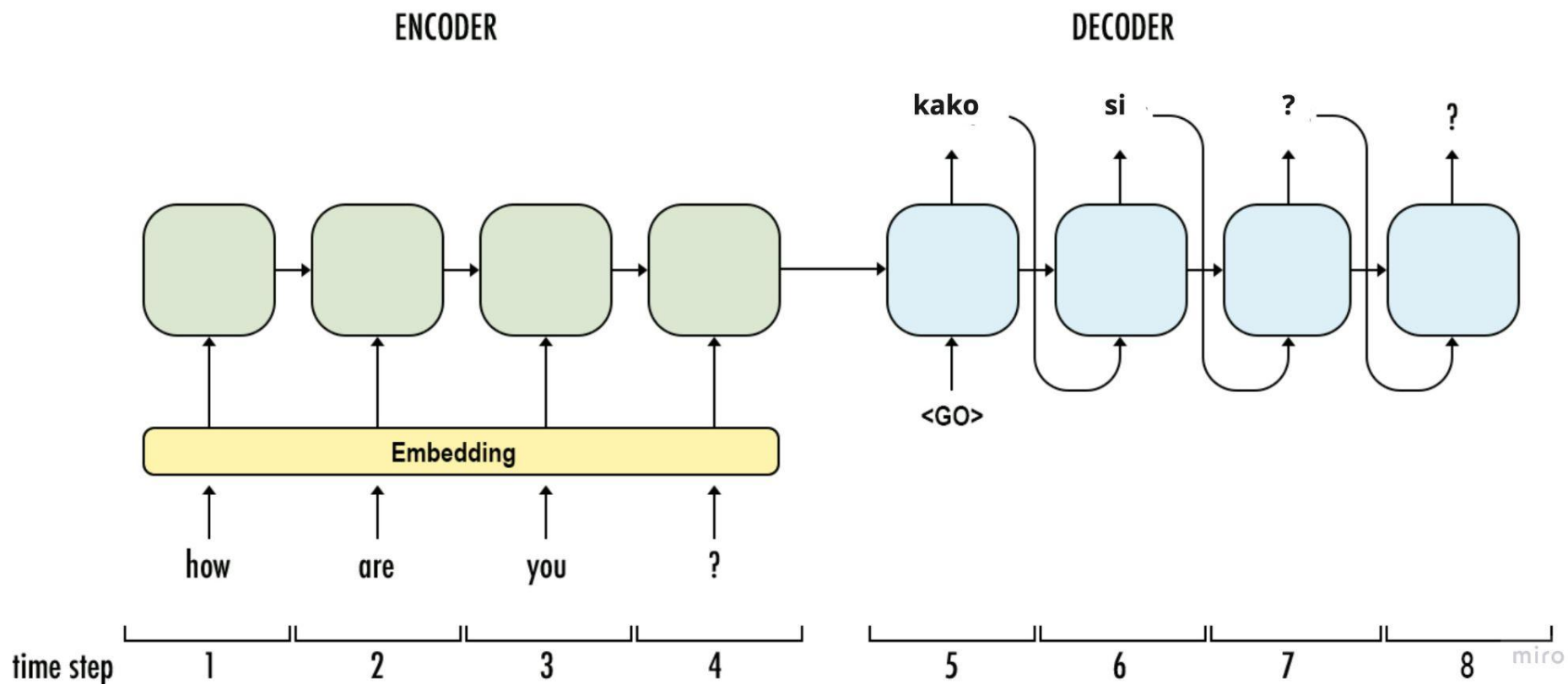


miro



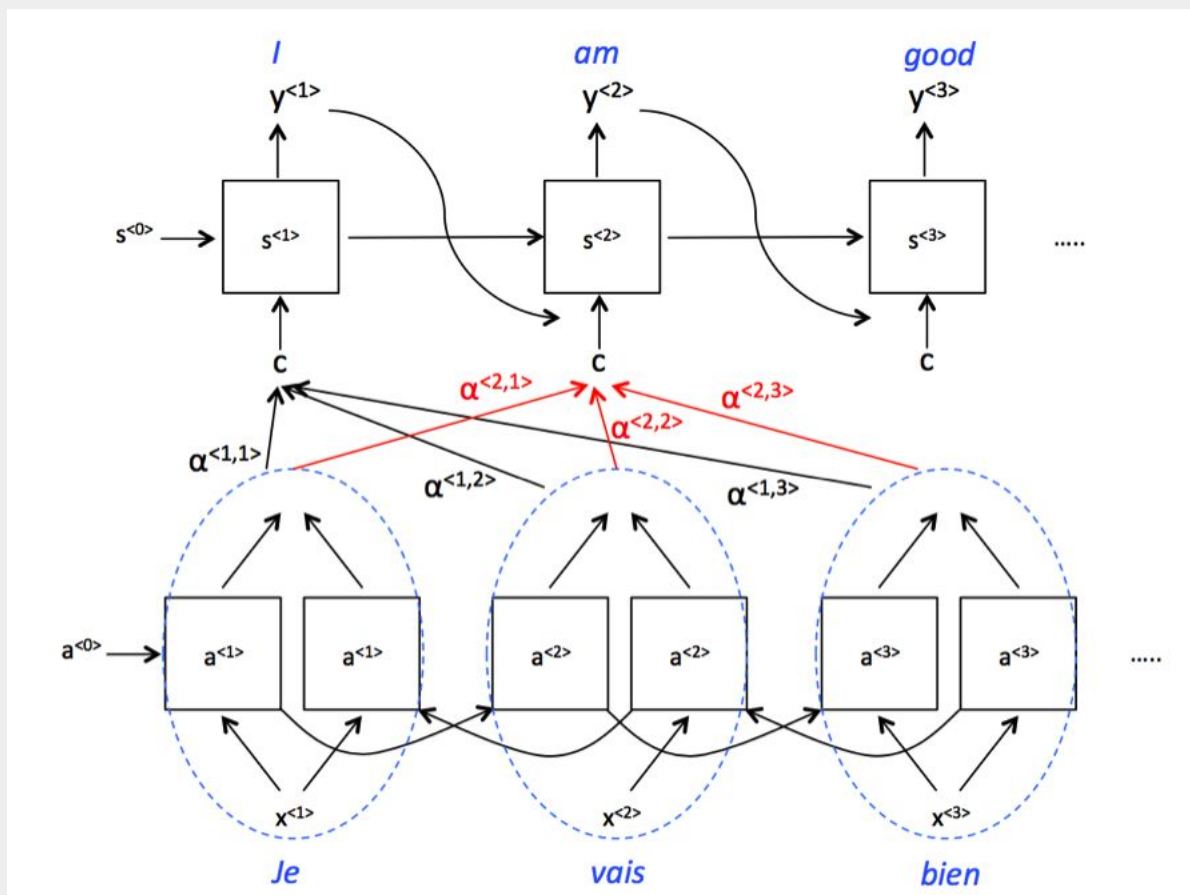
- Želimo da prevedemo rečenicu sa engleskog na srpski
- „enkodiramo“ da bismo razumeli „dekodiramo“
- „dekodiramo“ na srpski







- Ne pridajemo svim rečima isto značenje



- I dalje procesiramo reč po reč
- Inspiracija od CNN mreža



Hvala na pažnji!

Pitanja?