Mašta ili sjećanje?
Otkrivanje razlika u kognitivnom doživljaju putem podatkovne znanosti

Fani Sentinella-Jerbić

Sadržaj

1. Uvod i motivacija

2. Povezano istraživanje

3. Moj rad

4. Rezultati i diskusija
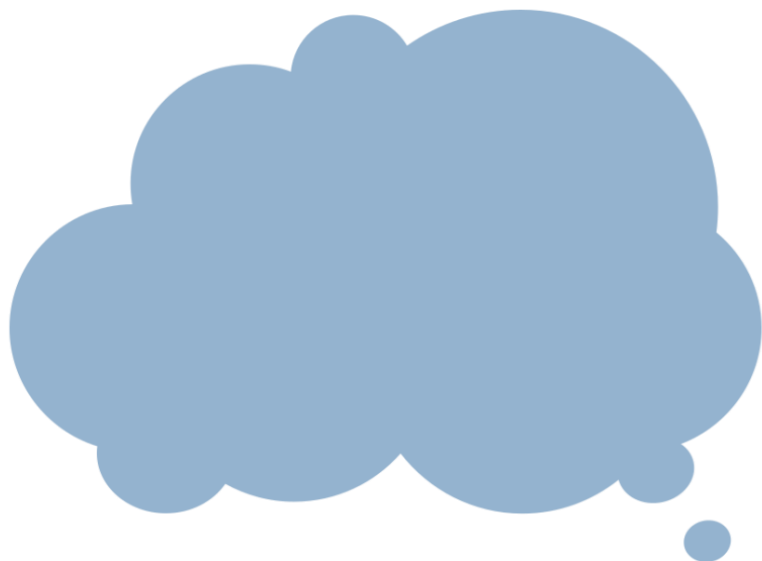
5. Literatura

Sadržaj

1. Uvod i motivacija

2. Povezano istraživanje

3. Moj rad
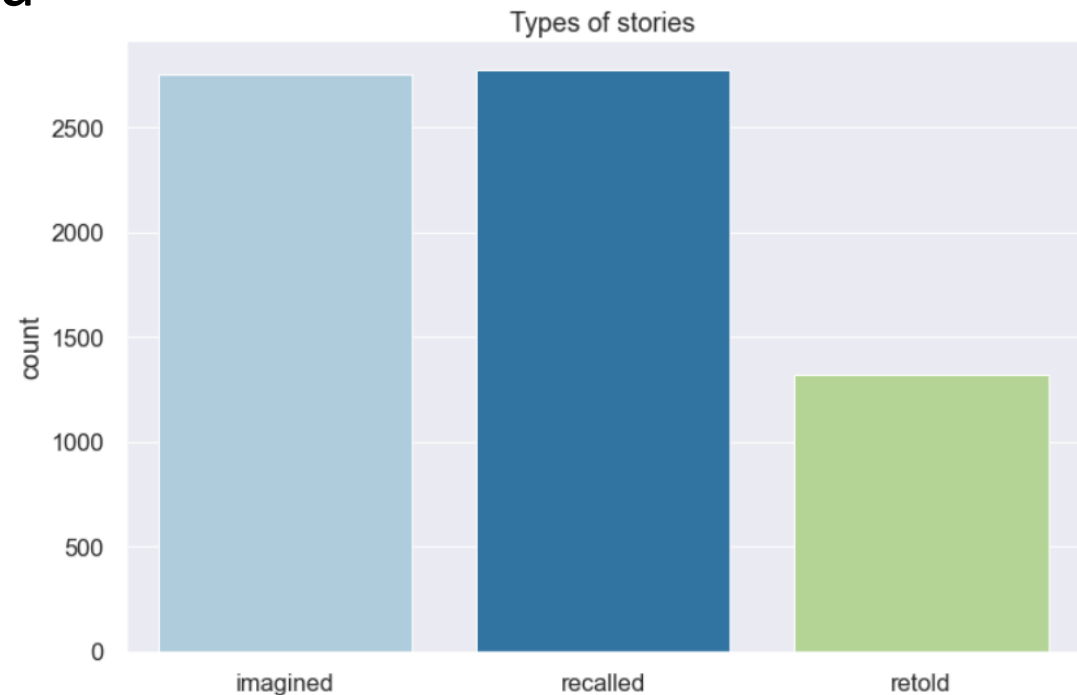
4. Rezultati i diskusija

5. Literatura

Uvod i motivacija

Sadržaj

Povezano istraživanje

- Skup podataka *Hippocorpus*
  - 6854 priča na engleskom jeziku
    - stvarne priče
    - izmišljene priče
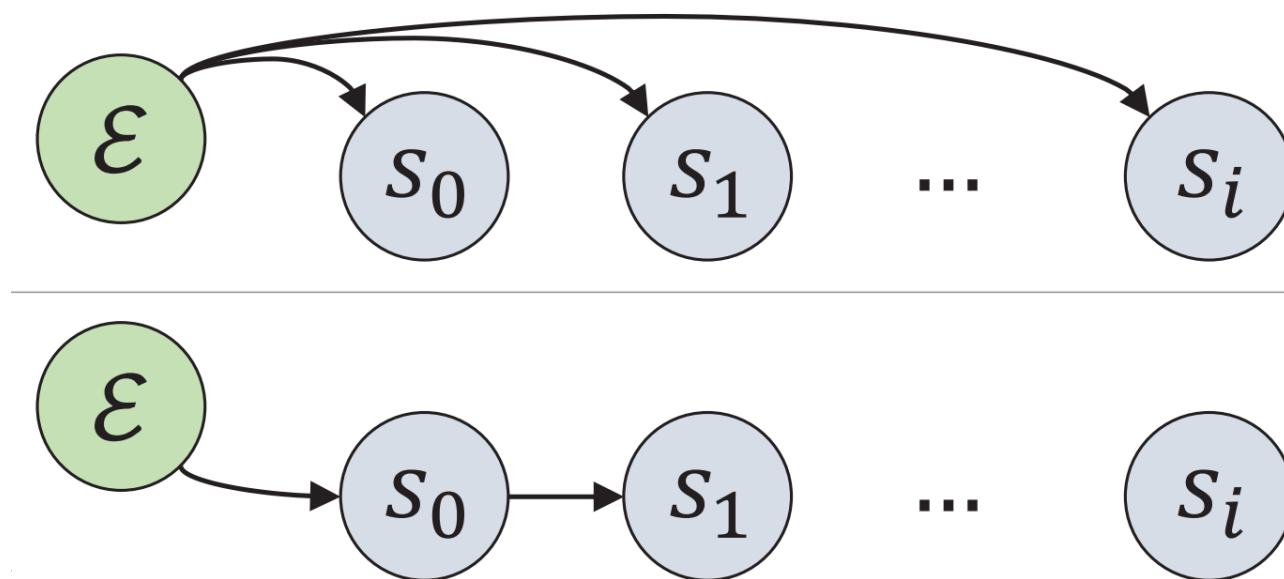   - prepričane priče

Types of stories

Primjer

Last week I won my first Brazilian Jiu Jitsu match! I was so nervous about competing and I can't believe I actually won! I can't describe how amazing it felt to win! I worked so hard training for this tournament and my hard work paid off.

I recently competed in a Brazilian Jiu Jitsu tournament. It took a lot of discipline to prepare for. I had to diet and lose about 15 pounds. I trained about 6-7 days out of the week. There were at least 4 jiu jitsu training sessions and 1 judo session.

Hippocorpus, MS Open Data, https://msropendata.com/datasets/0a83fb6f-a759-4a17-aaa2-fbac84577318

Tijek narativa

Konceptualni prikaz vjerojatnosnih modela vreće i modela lanca, preuzeto iz [1].

Tijek narativa

Last week I won my first Brazilian Jiu Jitsu match! I was so nervous about competing and I can't believe I actually won! I can't describe how amazing it felt to win! I worked so hard training for this tournament and my hard work paid off.

I recently competed in a Brazilian Jiu Jitsu tournament. It took a lot of discipline to prepare for. I had to diet and lose about 15 pounds. I trained about 6-7 days out of the week. There were at least 4 jiu jitsu training sessions and 1 judo session.

Količina konkretnih događaja

Last week I won my first Brazilian Jiu Jitsu match! I was so nervous about competing and I can't believe I actually won! I can't describe how amazing it felt to win! I worked so hard training for this tournament and my hard work paid off.
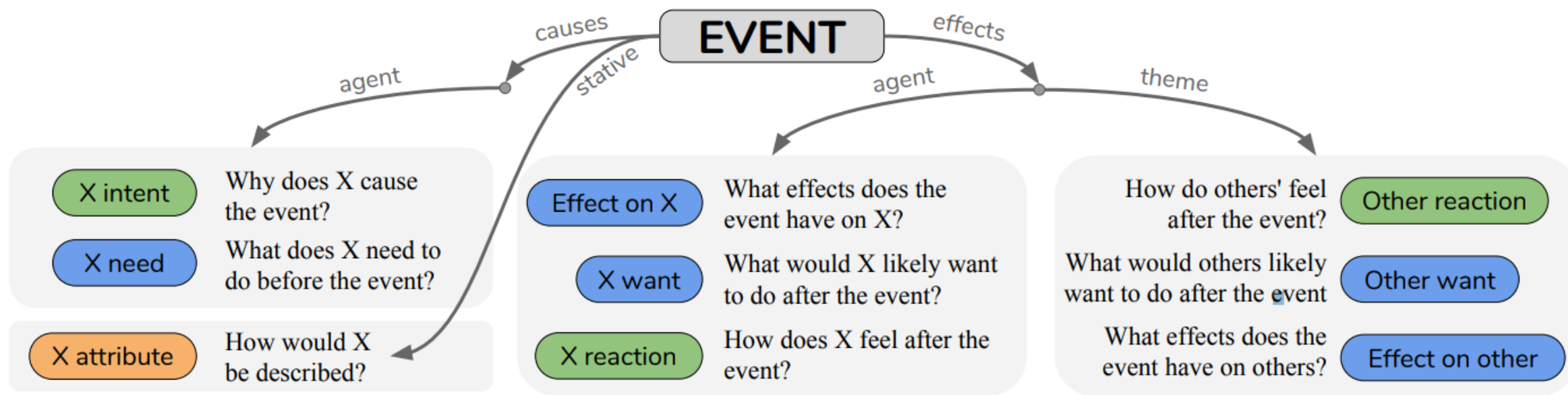
I recently competed in a Brazilian Jiu Jitsu tournament. It took a lot of discipline to prepare for. I had to diet and lose about 15 pounds. I trained about 6-7 days out of the week. There were at least 4 jiu jitsu training sessions and 1 judo session.

Količina konkretnih događaja

Last week **I won my first Brazilian Jiu Jitsu match**! I was so nervous about competing and I can't believe I actually won! I can't describe how amazing it felt to win! **I worked so hard training** for this tournament and my hard work paid off. I was so happy after winning and has motivated me to keep training in Jiu Jitsu.

I recently competed in a Brazilian Jiu Jitsu tournament. It took a lot of discipline to prepare for. I **had to diet** and **lose about 15 pounds**. I **trained about 6-7 days out of the week**. There were at least **4 jiu jitsu training sessions** and **1 judo session**. In addition to this, I also **traveled to other gyms** to spar against other jiu jitsu players.

Količina generalnog znanja

Taksonomija ako-onda rasuđivanja iz atlasa zdravorazumskog znanja ATOMIC, preuzeto iz [3].

Količina generalnog znanja

Last week I won my first Brazilian Jiu Jitsu match! I was so nervous about competing and I can't believe I actually won! I can't describe how amazing it felt to win! I worked so hard training for this tournament and my hard work paid off.

I recently competed in a Brazilian Jiu Jitsu tournament. It took a lot of discipline to prepare for. I had to diet and lose about 15 pounds. I trained about 6-7 days out of the week. There were at least 4 jiu jitsu training sessions and 1 judo session.

Količina generalnog znanja

Last week I won my first Brazilian Jiu Jitsu match! I **was so nervous** about competing and I **can't believe I actually won**! I can't describe how amazing it felt to win! I **worked so hard training** for this tournament and **my hard work paid off**.

I recently competed in a Brazilian Jiu Jitsu tournament. It took a lot of discipline to prepare for. I **had to diet** and **lose about 15 pounds**. I trained about 6-7 days out of the week. There were at least 4 jiu jitsu training sessions and 1 judo session.

Leksičke značajke

Last week I won my first Brazilian Jiu Jitsu match! I was so nervous about competing and I can't believe I actually won! I can't describe how amazing it felt to win! I worked so hard training for this tournament and my hard work paid off.

I recently competed in a Brazilian Jiu Jitsu tournament. It took a lot of discipline to prepare for. I had to diet and lose about 15 pounds. I trained about 6-7 days out of the week. There were at least 4 jiu jitsu training sessions and 1 judo session.

Leksičke značajke

Last week I won my first Brazilian Jiu Jitsu match! I was so **nervous** about competing and I **can't believe** I actually won! I can't describe how **amazing it felt** to win! I worked so hard training for this tournament and my hard work paid off.
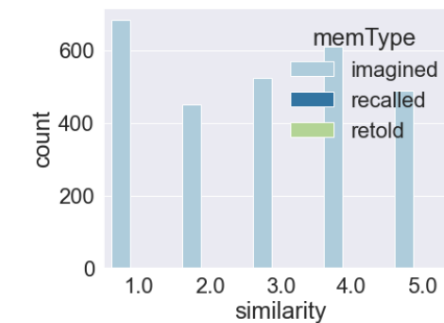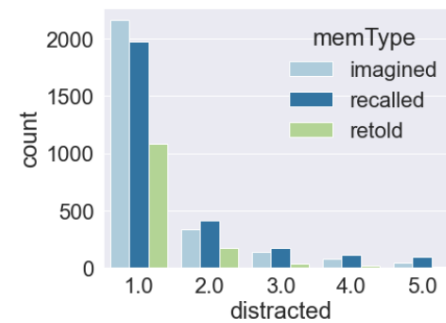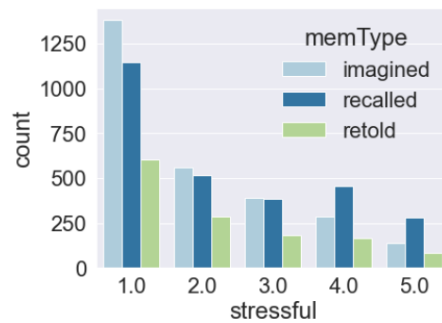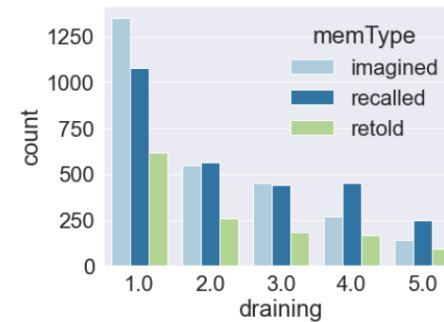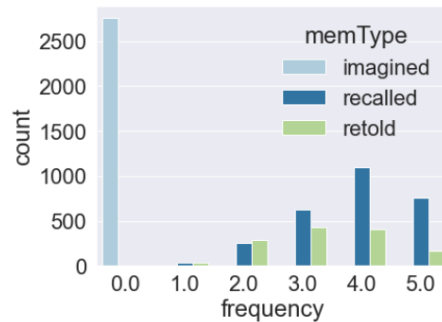
I recently competed in a Brazilian Jiu Jitsu tournament. It took a lot of discipline to prepare for. I had to diet and lose about 15 pounds. I trained about 6-7 days out of the week. There were at least 4 jiu jitsu training sessions and 1 judo session.

Sadržaj

Moj rad

	mostSurprising	mainEvent
0	when I got concert tickets	attending a show
1	we saw the beautiful sky.	a concert.
2	she went into labor early	my sister having her twins a little early
3	finding out they were healthy.	meeting my twin niece and nephew.
		the consequences of going to burning man
		about my journey this year to Burning Man meta...
		a scary car ride where we swerved on our way t...
		a bad driving experience on the way to Acadia ...

Moj rad

|	story	memType
0	Concerts are my most favorite thing, and my bo...	imagined
1	The day started perfectly, with a great drive ...	recalled
2	It seems just like yesterday but today makes f...	imagined
3	Five months ago, my niece and nephew were born...	recalled
4	About a month ago I went to burning man. I was...	imagined
5	Burning Man metamorphoses was perfect. I am de...	recalled
6	Play stupid games, win stupid prizes road trip...	imagined
7	Dear Diary,I would say this happened about 3 m...	recalled
8	I wanted to write about one of the best days i...	imagined
9	Me and my girlfriend had gone to the Los Angel...	retold

Hippocorpus, MS Open Data, https://msropendata.com/datasets/0a83fb6f-a759-4a17-aaa2-fbac84577318

Moj rad

- Jednostavnije metode → slični zaključci?

model	broj slojeva	broj parametara
GPT-2	12	117 milijuna
BERT-Base	12	110 milijuna
RoBERTa	12	125 milijuna
CTRL	48	1.6 milijardi

Transformers: Pretrained models, HuggingFace, https://huggingface.co/transformers/v2.2.0/pretrained_models.html

Nove značajke

- Broj znakova

- Broj riječi

- Broj rečenica

109

51

72

16

5

94

67

20

35

98

11

73
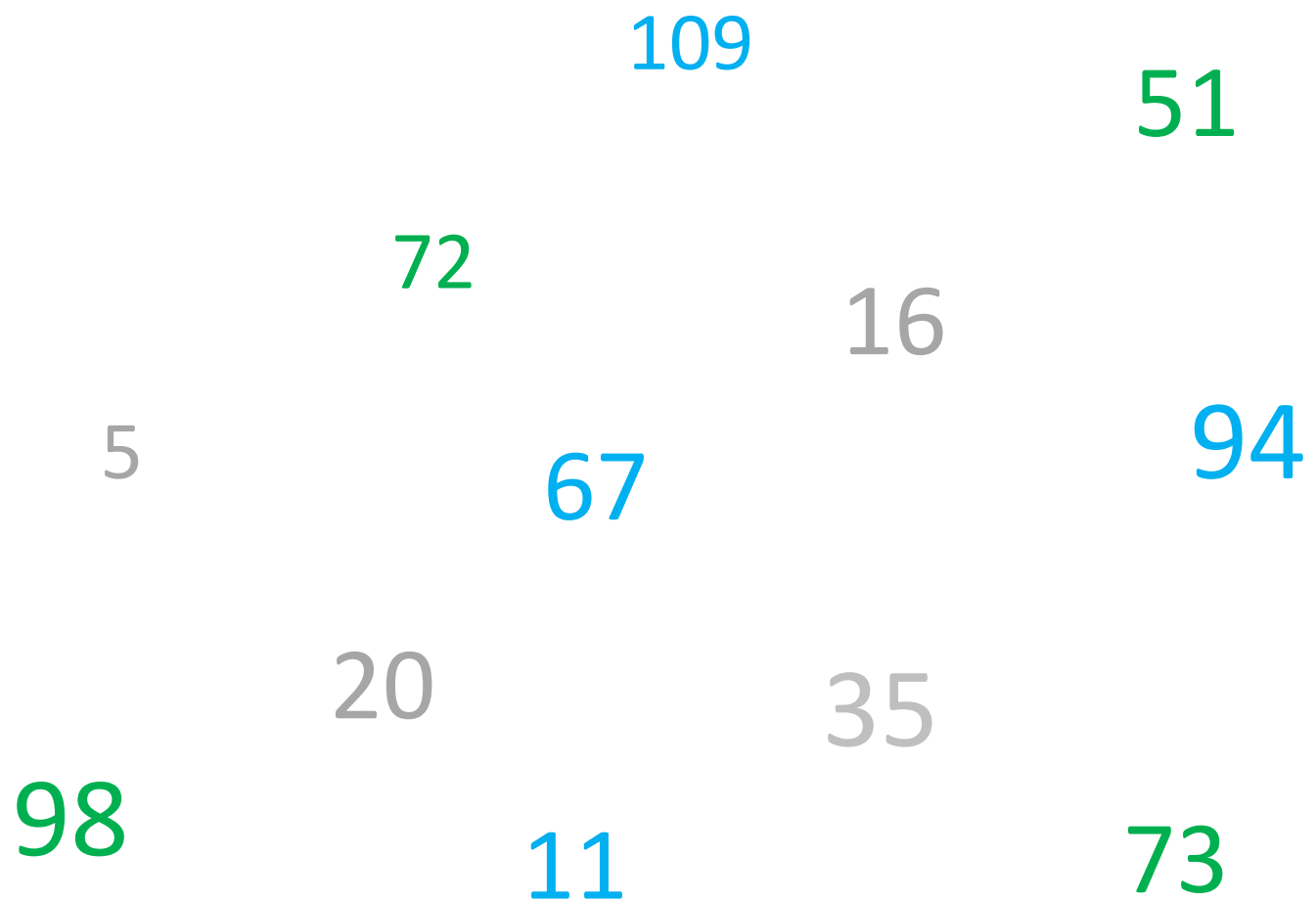
Nove značajke

- Broj jedinstvenih riječi

- Udio jedinstvenih riječi

- Leksička raznolikost

$$H(story) = -\sum_{w \in story} \frac{freq(w)}{len(story)} \log_2 \frac{freq(w)}{len(story)}$$

Nove značajke

- Broj *neznačajnih* riječi

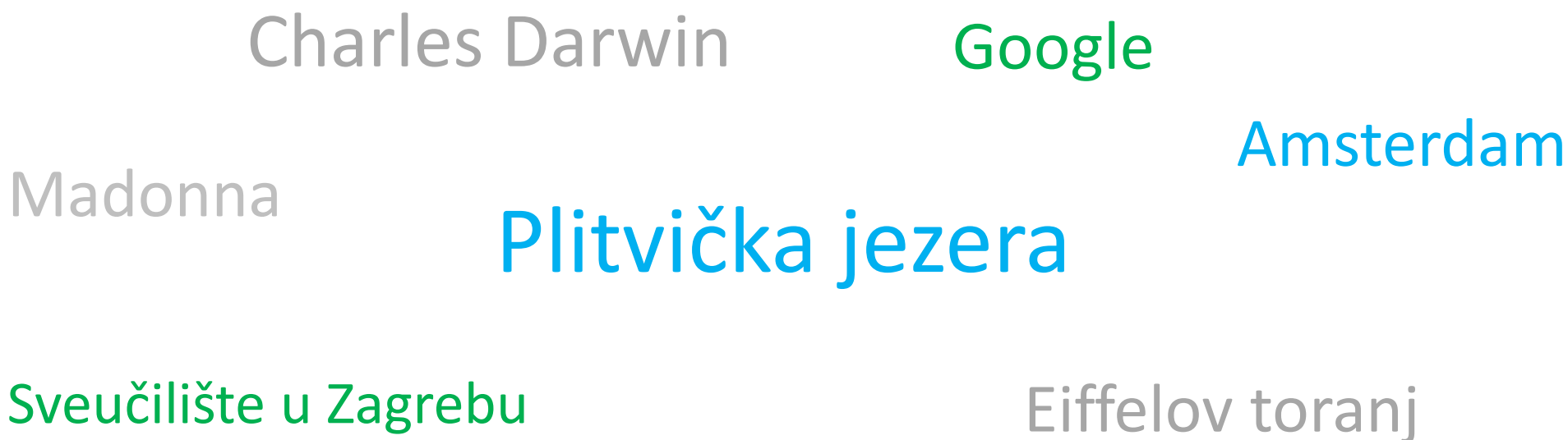- Udio *neznačajnih* riječi

but

the

on

for

and

until

also

than

about

Nove značajke

- Broj osobnih riječi – *I, me, my*

- Polarnost **+ –**

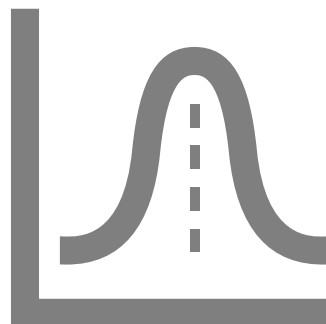- Subjektivnost

Nove značajke

- Broj imenovanih entiteta

Charles Darwin   Google

                                    Amsterdam

Madonna

Plitvička jezera

Sveučilište u Zagrebu   Eiffelov toranj

Sadržaj

Analiza novih značajki

- Testiranje hipoteza na razini jedne značajke
  - Studentov t-test
  - nema značajne razlike među vrstama priča

Analiza novih značajki

- Rangiranje značajki
  - **Broj znakova**
  - **Udio jedinstvenih riječi**

Narativizacija stvarnih događaja

- Binarna klasifikacija stvarnih i izmišljenih priča
  - Logistička regresija
  - Slučajna šume
  - Stablo odluke
  - Stroj potpornih vektora
  - K-susjeda

Narativizacija stvarnih događaja

- Binarna klasifikacija stvarnih i izmišljenih priča

  - **Logistička regresija – točnost 62.59%**

  - Slučajna šume

  - Stablo odluke

  - Stroj potpornih vektora

  - K-susjeda

Narativizacija stvarnih događaja

- Testiranje modela na prepričanim pričama
    - Točnost **58.38%**
        - → prepričane priče *liče* na izmišljene

Zaključak

- Potencijal jednostavnih metoda podatkovne znanosti

Literatura

1. Maarten Sap, Eric Horvitz, Yejin Choi, Noah A Smith, i James W Pennebaker. Recollection versus Imagination: *Exploring Human Memory and Cognition via Neural Language Models*. Association for Computational Linguistics, 2020.

2. Maarten Sap, Anna Jafarpour, Yejin Choi, Noah A Smith, James W Pennebaker, i Eric Horvitz. *Computational Lens on Cognition: Study of Autobiographical Versus Imagined Stories With Large-Scale Language Models*. arXiv preprint arXiv:2201.02662, 2022.

3. Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, and Yejin Choi. *Atomic: An atlas of machine commonsense for if-then reasoning.* In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, pp. 3027-3035. 2019.

4. Jekaterina Novikova, Aparna Balagopalan, Ksenia Shkaruta, i Frank Rudzicz. *Lexical features are more vulnerable, syntactic features have more predictive power.* arXiv preprint arXiv:1910.00065, 2019.