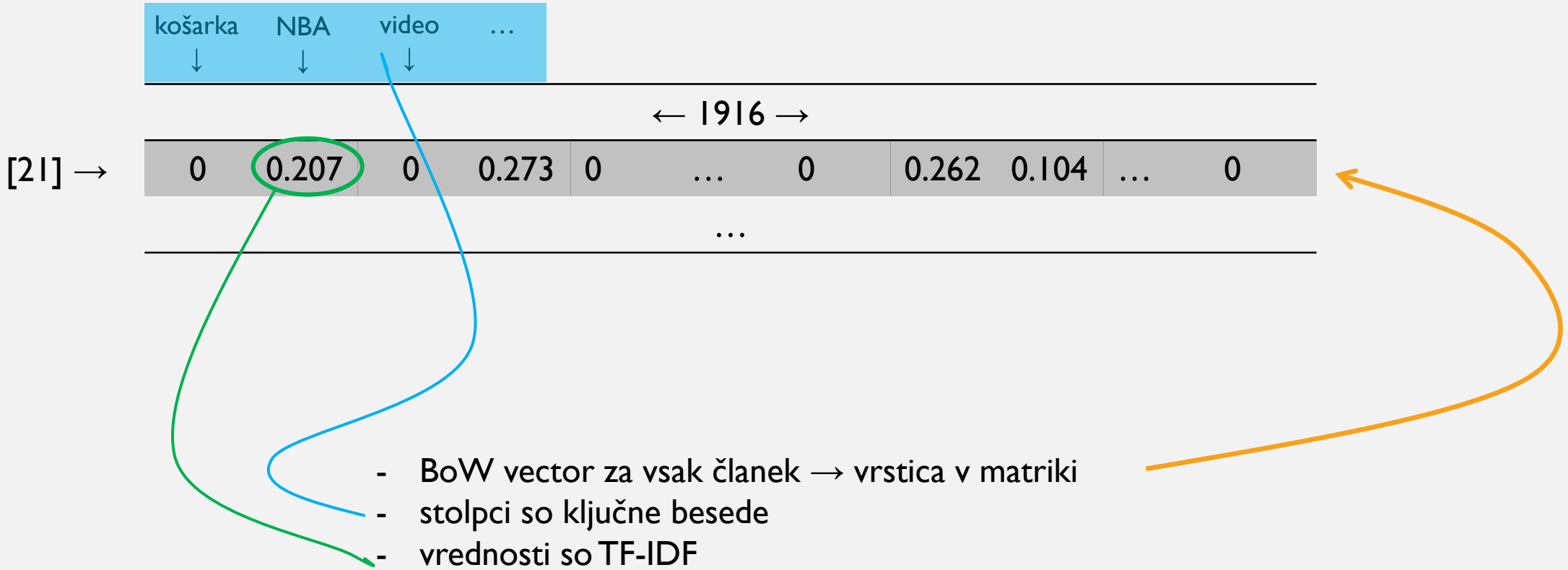
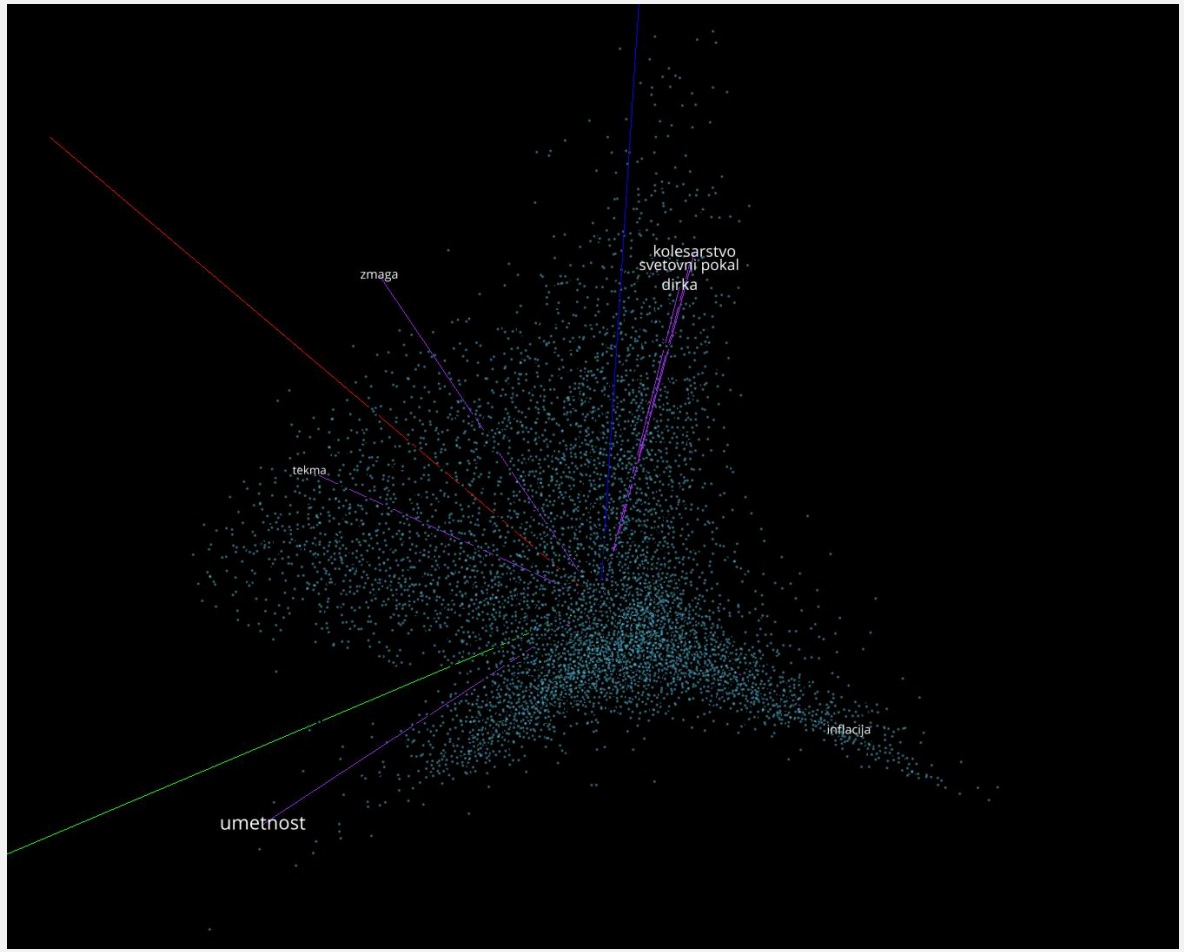
Bag of Words predstavitev



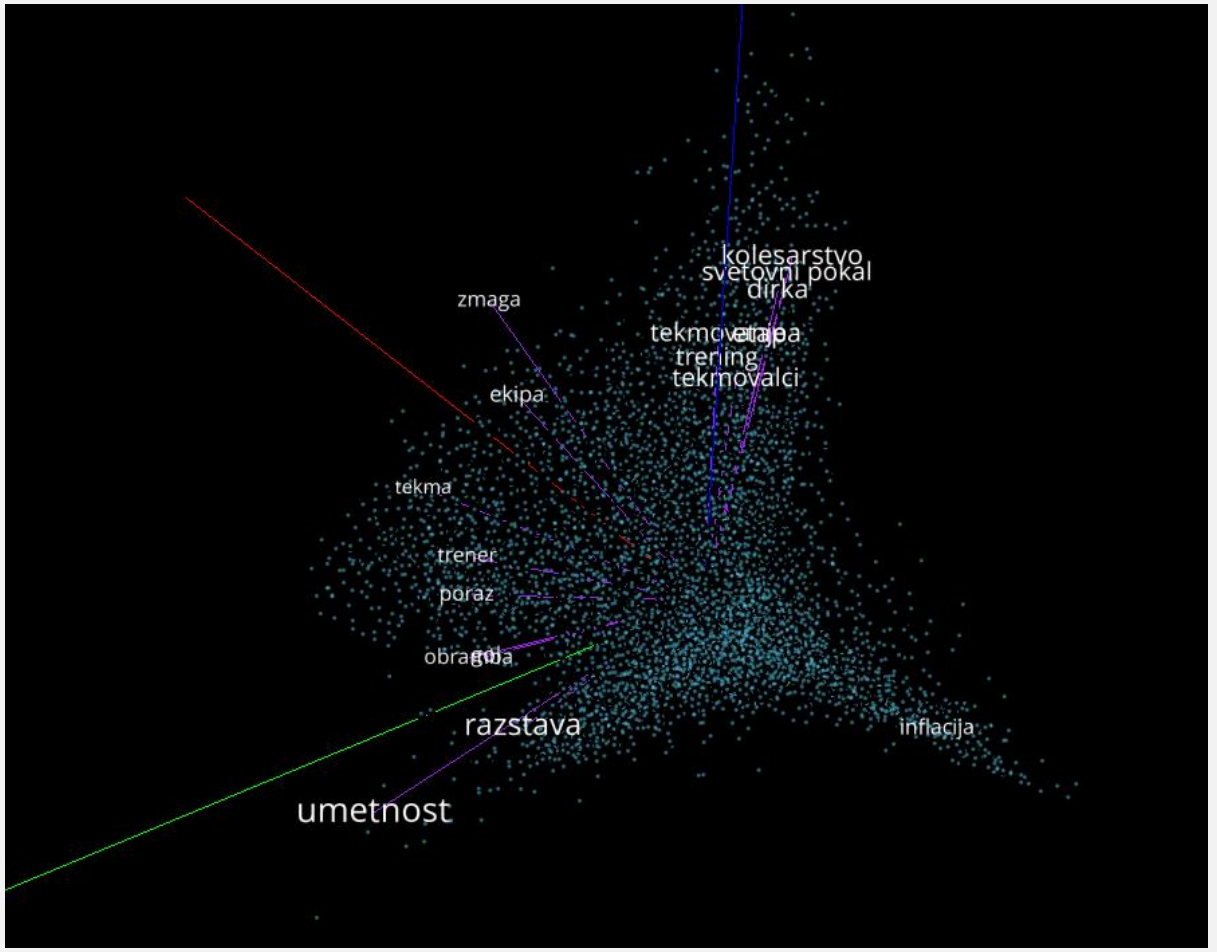
- BoW vector za vsak članek → vrstica v matriki
- stolpci so ključne besede
- vrednosti so TF-IDF

	komponenta 1	komponenta 2	komponenta 3	Σ
explained variance	0.00660031	0.00337756	0.00317879	0.01316
explained variance ratio	0.01123889	0.00575125	0.00541279	0.0224

z višjo mejo za izbiro nateznih koeficientov



z nižjo mejo za izbiro nateznih koeficientov



- izbira glede na vpliv na komponente PCA
- loadings:

```
array([[ 0.14822385,  0.03489821, -0.08163245],
       [ 0.05488558,  0.01639936, -0.05503689],
       [-0.00444451,  0.00926148,  0.00037815],
       ...,
       [ 0.00795837, -0.01690399,  0.03384755],
       [ 0.0040688 , -0.01096533,  0.01977564],
       [ 0.00416904, -0.00753452,  0.01626376]])
```

  vsaka vrstica je povezana z ključno besedo

- njihove norme:

```
000: 0.1727774605605261
001: 0.07943818328072684
002: 0.010279676910965888
003: 0.023825753554273018
004: 0.006684474849979278
005: 0.004661426100828256
006: 0.011317439394541327
007: 0.12773876681069238
008: 0.11460491381246926
009: 0.06140048318246006
```

- tudi po vplivih na posamezno komponento
- izbrano glede na threshold (najbolj vplivni)