ONJ - seminar 2

IMapBook

Žiga Simončič

Klemen Randl

Predprocesiranje

- Odstranitev pik, vejic ipd.
- Odstranitev besed, ki se pojavijo v vprašanju?
- Lematizacija
- Korenjenje?

Model A

- Kosinusna podobnost (vektorizacija s TF-IDF)
- OpenIE (StanfordNLP)
  - Ekstrakcija „trojčkov“
  - Koreferenčnost?
- Povprečje obeh metod

#	remove	use_length	Cosine	openie	F1 micro	F1 macro
1	False	False	True	0	73%	42%
2	False	False	True	1	18%	17%
3	False	False	True	2	18%	17%
4	True	False	True	0	72%	49%
5	True	False	True	1	17%	17%
6	True	False	True	2	17%	17%
7	False	True	True	0	71%	49%
8	False	True	True	1	14%	16%
9	False	True	True	2	14%	16%
10	True	True	True	0	65%	43%
11	True	True	True	1	13%	16%
12	True	True	True	2	13%	16%
13	False	False	False	1	11%	6%
14	False	False	False	2	11%	6%

Model B

- Enako kot model A (le več podatkov)
  - Kosinusna podobnost
  - OpenIE
  - Povprečje obeh

#	remove	Cosine	openie	F1 micro	F1 macro
1	False	True	0	67%	38%
2	True	True	0	62%	40%
3	False	True	1	65%	48%
4	False	True	2	60%	46%
5	True	True	1	50%	42%
6	True	True	2	39%	37%
7	False	False	1	63%	36%
8	False	False	2	58%	35%
9	True	False	1	45%	29%
10	True	False	2	34%	24%

Model C

- Tri rešitve
- 1. Podobnost glede na razdaljo besed v WordNetu
- 2. Kosinusna podobnost + OpenIE z dodanimi sinonimi (WordNet)
- 3. Pomembnost besed z dodanimi sinonimi

Model C1

#	remove	F1 micro	F1 macro
1	True	62%	43%
2	False	65%	46%

Model C2

#	remove	openie	no_of_synonyms	F1 micro	F1 macro
1	False	1	1	65%	48%
2	True	1	1	50%	42%
3	False	2	1	61%	47%
4	True	2	1	40%	37%
5	False	1	2	66%	48%
6	True	1	2	50%	42%
7	False	2	2	61%	46%
8	True	2	2	40%	37%
9	False	1	3	66%	48%
10	True	1	3	50%	42%
11	False	2	3	61%	46%
12	True	2	3	40%	37%

Model C3

#	importantWord	score05	score10	F1 micro	F1 macro
1	0.01	0.5	0.6	58%	43%
2	0.01	0.4	0.5	71%	52%
3	0.01	0.3	0.4	72%	41%
4	0.02	0.5	0.6	64%	42%
5	0.02	0.4	0.5	71%	49%
6	0.02	0.3	0.4	72%	40%
7	0.03	0.5	0.6	64%	39%
8	0.03	0.4	0.5	70%	45%
9	0.03	0.3	0.4	70%	38%

Iskanje najpomembnejših besed (primer)

- Q: How does Shiranna feel as the shuttle is taking off?

- A1: Shiranna feels both excited and nervous as the shuttle is taking off.
- A2: Nervous, but also excited to be with her mother.
- A3: she is excited and scared

- Predprocesiranje + odstranimo besede, ki se pojavijo v vprašanju

Iskanje najpomembnejših besed (primer)

- Q: How does Shiranna feel as the shuttle is taking off?

- A1: Shiranna feels both excited and nervous as the shuttle is taking off.
- A2: Nervous, but also excited to be with her mother.
- A3: she is excited and scared

Iskanje najpomembnejših besed (primer)

- Q: How does Shiranna feel as the shuttle is taking off?

- A1: both excite and nervous
- A2: nervous but also excite to with her mother
- A3: she excite and scare

- Dodatno odstranimo še besede kot so „and“, „her“, „she“, …

Iskanje najpomembnejših besed (primer)

- A1: both excite nervous
- A2: nervous but also excite to with mother
- A3: excite scare

Iskanje najpomembnejših besed (primer)

- A1: both excite nervous
- A2: nervous but also excite to with mother
- A3: excite scare

- Če se beseda pojavi v več odgovorih, je pomembna

- Za vsak odgovor:
  - Izračunaj povprečno podobnost z ostalimi odgovori – baseline
  - Izmenično odstranjuj besede in primerjaj podobnost
  - Če je podobnost manjša, je beseda pomembna

Iskanje najpomembnejših besed (primer)

- A1: both excite nervous
- A2: nervous but also excite to with mother
- A3: excite scare

- Če se beseda pojavi v več odgovorih, je pomembna

- Za vsak odgovor:
  - Izračunaj povprečno podobnost z ostalimi odgovori – baseline
  - Izmenično odstranjuj besede in primerjaj podobnost
  - Če je podobnost manjša, je beseda pomembna

Iskanje najpomembnejših besed (primer)

- Pomembne besede: [nervous, excite]
- (Prvotno vprašanje: How does Shiranna feel as the shuttle is taking off?)

- Preveri podobnost podanega odgovora s pomembnimi besedami

Povzetek

- Model A
  - Mikro: 72%
  - Makro: 49%
- Model B
  - Mikro: 65%
  - Makro: 48%
- Model C
  - Mikro: 71%
  - Makro: 52%