

# Seminar - Izvještaj trenutnog stanja

Mihael Nikić

March 2017

## 1 Čitanje rada/ova

Radovi koje treba pročitati:

- ***Your Sentiment Precedes You: Using an author's historical tweets to predict sarcasm*** - DONE
- Harnessing context incongruity for sarcasm detection - **DJELOMIČNO**: pročitana poglavlja 3. i 4.
- Ostali radovi:
  - TODO

## 2 Skup podataka

Učitavanje i analiziranje skupova podataka iz rada:

- 8000 tweets marked with #sarcasm, - **Nepoznata lokacija???**
- **Manually annotated corpus consisting 2278 tweets** - DONE\* (80% tvitova nije moguće bilo dohvatiti???)
- Lexicon L1 - Pang and Lee (2004) - **Nepoznata lokacija???**,
- **Lexicon L2 - Mohammad and Turney (2013).** - DONE

## 3 Implementacija

### 3.1 Contrast-based predictor

#### 3.1.1 Implicit contrast-based predictor

Potrebno je učiniti:

1. Za prikupljenih 8000 tvitova pretpostaviti da su sarkastični,
2. Izdvojiti 3-grams u 1-grams,
3. Odabrati fraze koje se pojavljuju najmanje tri puta,
4. Pronaći kontrast riječ-fraza u tvitu (*Joshi et al.(2015)*)

#### 3.1.2 Explicit contrast

Potrebno je učiniti:

1. Pronaći kontrast riječ-riječ u tvitu (*Joshi et al.(2015)*) uz pomoć leksikona

### 3.2 HISTORICAL TWEET-BASED PREDICTOR - IN PROGRESS

Potrebno je učiniti:

1. POS tagirati rečenicu u promatranom tvitu, - DONE\* (nije još sklopljeno u razred/metodu)
2. Izvući NNP sekvencu kao ciljanu frazu u promatranom tvitu, - DONE\*(nije još sklopljeno u razred/metodu)
3. Skinuti kompletni timeline od autora koristeći Twitter API, - DONE
4. Odabrati tvitove koje sadrže ciljanu frazu,
5. Za odabrane stare tvitove učiniti sljedeće:
  - Ocijeniti sentiment tvita odnosno rečenice uz pomoć leksikona i pravila za negaciju, veznike i sl.

- Glasanjem utvrditi koji sentiment je dominantan u odabranim tvitovima
6. Proglasiti promatrani tvit sarkastičnim ako je sentiment promatranog tvita **različit** od onoga u starim tvitovima.
  7. Ponavljati postupak za sve ostale NNP sekvence dok se nužni uvjet za proglašenje tvita sarkastičnim ne ispuni

### 3.3 Integrator

Potrebno je napraviti četiri vrste integratora

1. "Only historical tweet-based" - Integrator koji uzima u obzir izlaze "Historical tweet-based" prediktora. U slučaju da se ciljna fraza nije spominjala u starim tvitovima, tvit će biti proglašen ne-sarkastičnim,
2. "OR" - Integrator koji proglašava tvit sarkastičnim ako to učini **bilo koji** od prediktora,
3. "AND" - Integrator koji proglašava tvit sarkastičnim ako to učine **oba** prediktora,
4. "Relaxed-AND" - Sličan prethodnom, s tim da za razliku od prethodnog ne uzima u obzir izlaze "Historical tweet-based" prediktora ako on nije pronašao niti jedan stari tvit u kojem se pojavljuje ciljani sentiment.

## 4 Eksperimenti

Za svaku kombinaciju različitog leksikona i integratora potrebno je učiniti sljedeće:

1. Izračunati preciznost, odziv i F1-score,
2. Usporediti vrijednosti sa onima iz Riloff et. al (2013)

## 5 Pisanje rada

Potrebno je napisati sljedeće:

- Detaljan opis članka koji se sastoji od (pot)poglavlja:

– TODO

- Opis (re)implementacije modela
- Analiza repliciranih eksperimenata
- *DODATNO: napraviti dodatne eksperimente koji nisu bili u izvornom članku*