WHAT CAN WE LEARN FROM ALMOST A DECADE OF FOOD TWEETS

LATVIJAS UNIVERSITĀTE ANNO 1919 UNIVERSITY OF LATVIA

THE UNIVERSITY OF TOKYO

Uga Sproģis
University of Latvia

Matīss Rikters
The University of Tokyo
matiss@logos.t.u-tokyo.ac.jp

ABSTRACT

The **Latvian Twitter Eater Corpus** is a set of tweets in the narrow domain related to food, drinks, eating and drinking. We also separate two sub-corpora of question and answer tweets and sentiment anno-tated tweets. We analyse contents of the corpus and demonstrate use-cases for the sub-corpora by training domain-specific question-answering and sentiment-analysis models using data from the corpus.

SENTIMENT CLASSIFICATION

5 Training Datasets

4 Data Processing Variations

2 Model Configurations

Training Data	TE	MP	MP.PE	TE.MP	All	TE.MP.RV.PE
Naive Bayes	53.21	43.32	45.72	56.55	59.63	58.02
Perceptron	53.07	52.67	53.47	57.87	57.33	58.27
Stemmed						
Naive Bayes	53.74	46.39	50.67	58.16	60.56	**61.23**
Perceptron	56.67	53.73	54.13	60.00	56.93	57.73
Lemmas						
Naive Bayes	53.88	45.45	49.60	56.42	58.42	59.63
Perceptron	54.41	51.07	53.07	57.35	56.95	56.95
Stemmed Lemmas						
Naive Bayes	54.41	45.99	49.33	57.62	59.63	59.63
Perceptron	53.34	51.47	52.67	58.29	56.68	57.09

MAIN KEYWORDS

taste	**lunch**	beet	potato	mandarin	sweet
eat	**feast**	bun	cabbage	sauce	mushroom
breakfast	**drink**	carrot	candy	pancake	onion
dine	**treat**	chips	sour cream	dumpling	chocolate
dinner	**nom**	vegetable	cream soup	gingerbread	tea
bite	**appetite**	meat	cake	rice	tomato
meal	orange	Hesburger	drink	salad	grape
food	apple	coffee	McDonald's	ice cream	strawberry

galas
zirga
skandālu
produkcija
vispār
Islanda
neatrod
tikai
pārjautāja
ārzemnieks
veikalā
Kas
Latvijas
Rīgā
tūrists
vēršas
pārdevējs
DimaDaugavpils
DelfiLV
Forevers
pelmeni
atbildēja
tiesām
kādā
šorīt
galu
pie
beer
nav
pirka
izmeklējot
nīzivijis
PVD
Es
ka
un
ir
var
sure
pārdevēja
galu
Neapolitāns
skandāls
bridis
bet
tur
pēc
desās
Apollo
Rimi
TikNeveikli
labākā
skandāla
praise
alzoomas
produktā
Peeteris
negribu
lielīties
darbības
maltas
pārtikas
konstatēja
Latvija
Latvijas
Ornoots
Kurtu
konstatē
atrasta
pārstrādes
mums
Ja
weird
ja
Lv
Labākais
Labi
Labs
Labu
Labā
Lauku
Lauku
Lauku