

Sentiscope: A System for Sentiment Analysis in Daily Horoscopes

Danijela Merkle* and Željko Agić**

*Department of Linguistics, **Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
{dmerkler, zagic}@ffzg.hr

Overview

- develop a system for sentiment analysis in daily horoscopes (in Croatian)
- automatically collect texts from the web on a daily basis
- detect polarity phrases in horoscopes and estimate overall sentiment from them

Unatoč tome što vam preciznost i odgovornost danas nisu najjača strana, *<p>šarmom ćete izbjeći</p>* sve moguće posljedice. U odnosu s partnerom imajte više sluha za njegove potrebe i želje jer bi vas ignoriranje moglo dovesti u *<n>nezavidan položaj</n>*.

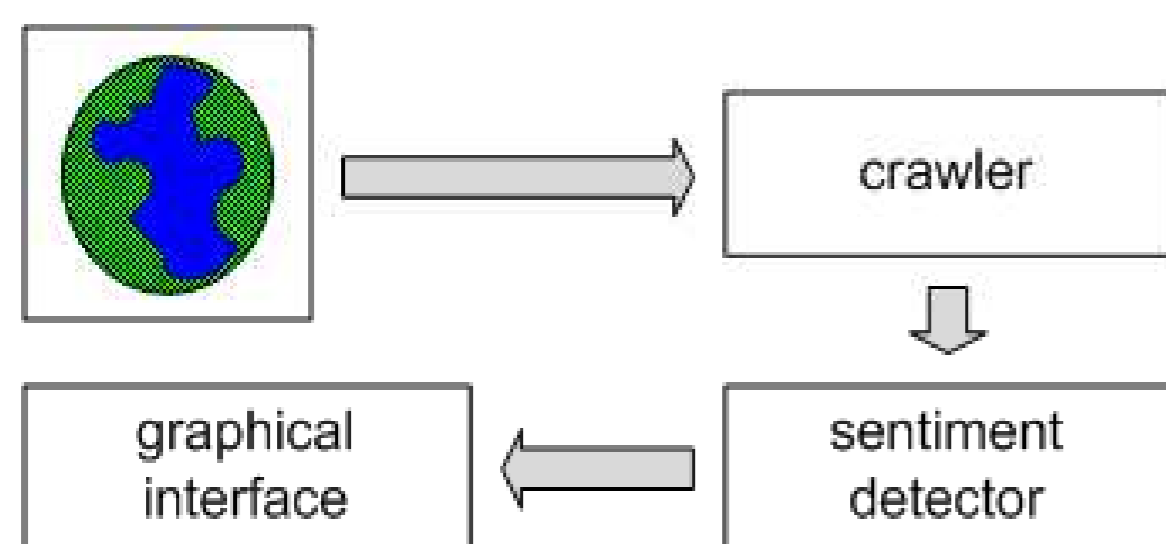


67% 30% 3%

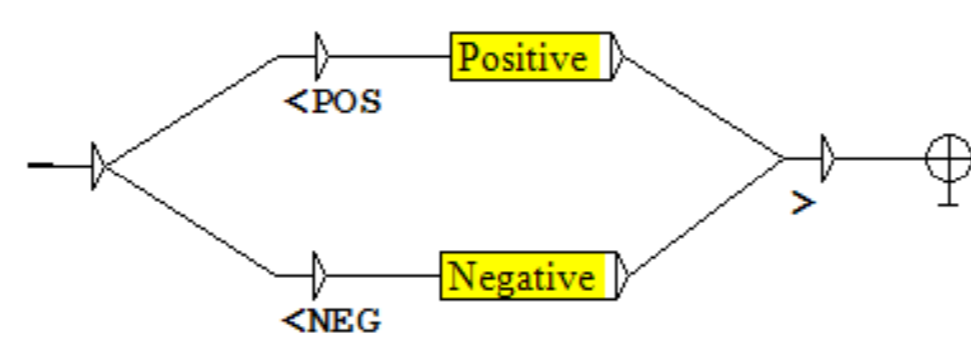


81% 14% 5%

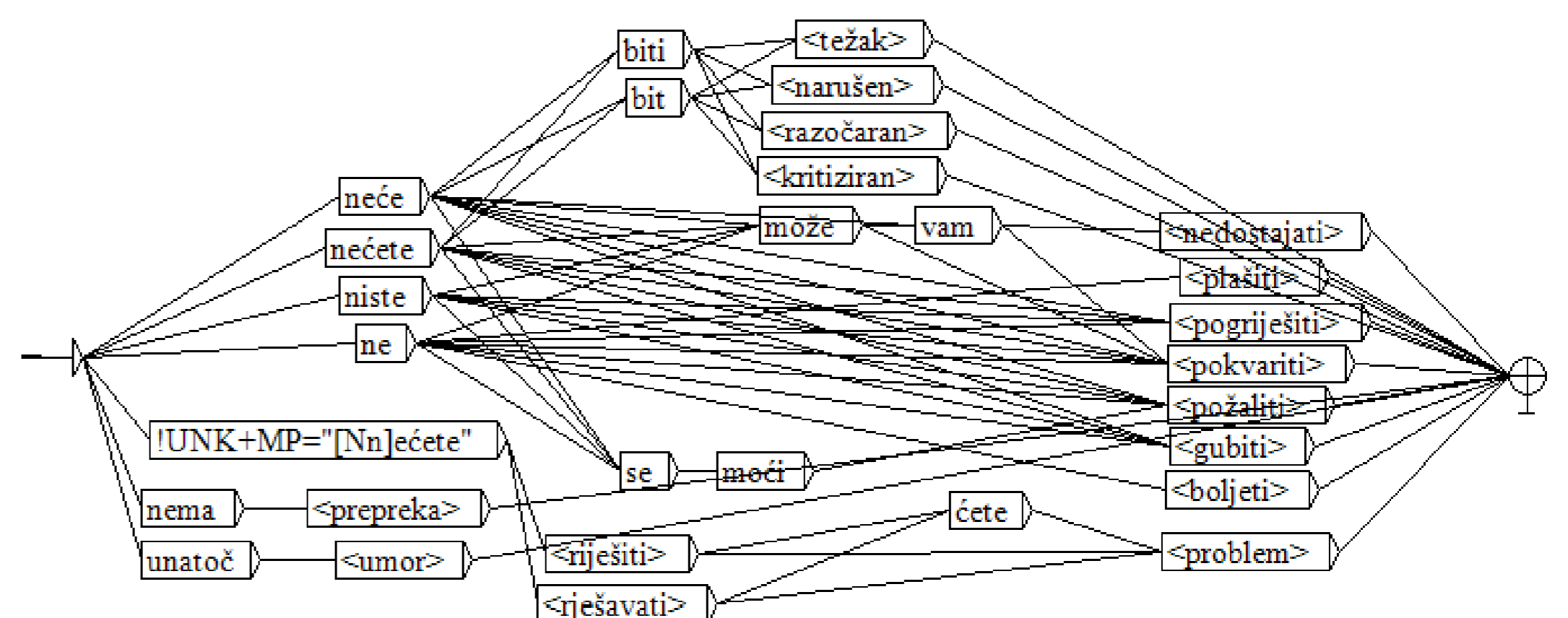
System design and implementation



System overview



NooJ top grammar



Positive phrase detection example

Experiment and results

Annotator agreement

	■	■	■	Σ
■	94	0	26	120
■	1	82	31	114
■	18	4	77	99
Σ	113	86	134	333

Phrase vs. overall sentiment

	p	n	pn	p	n
■	410	27	23	85	27
■	19	321	15	19	53
■	142	145	67	117	115

Horoscope sentiment by source on 2012-05-18

aries	■	■	■	■	■	■	■
taurus	■	■	■	■	■	■	■
gemini	■	■	■	■	■	■	■
cancer	■	■	■	■	■	■	■
leo	■	■	■	■	■	■	■
virgo	■	■	■	■	■	■	■
libra	■	■	■	■	■	■	■
scorpio	■	■	■	■	■	■	■
sagittarius	■	■	■	■	■	■	■
capricorn	■	■	■	■	■	■	■
aquarius	■	■	■	■	■	■	■
pisces	■	■	■	■	■	■	■

Overall accuracy and confusion matrix

	■*	■*	■*	P	R	F1
■	40	3	17	0.677	0.666	0.671
■	2	25	17	0.555	0.568	0.561
■	17	17	30	0.468	0.468	0.468

Phrase detection accuracy

	P	R	F1
initial	0.371	0.283	0.321
development	0.435	0.469	0.451
test	0.413	0.393	0.402

Conclusions and future work

- overall system accuracy 0.566, phrase detection F1-score 0.402, annotator agreement 75.97%
- web-based interface for assesment of horoscope sentiment for daily horoscopes in Croatian
- comparison with statistical methods

Acknowledgement

The research within the project ACCURAT leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013), grant agreement no. 248347.

