

Preprocessing

ClearTK / UIMA Based Pipeline
ClearNLP Wrappers

- Tokenizer
- POS Tagger
- Lemmatizer
- Dependency Parser

Normalization

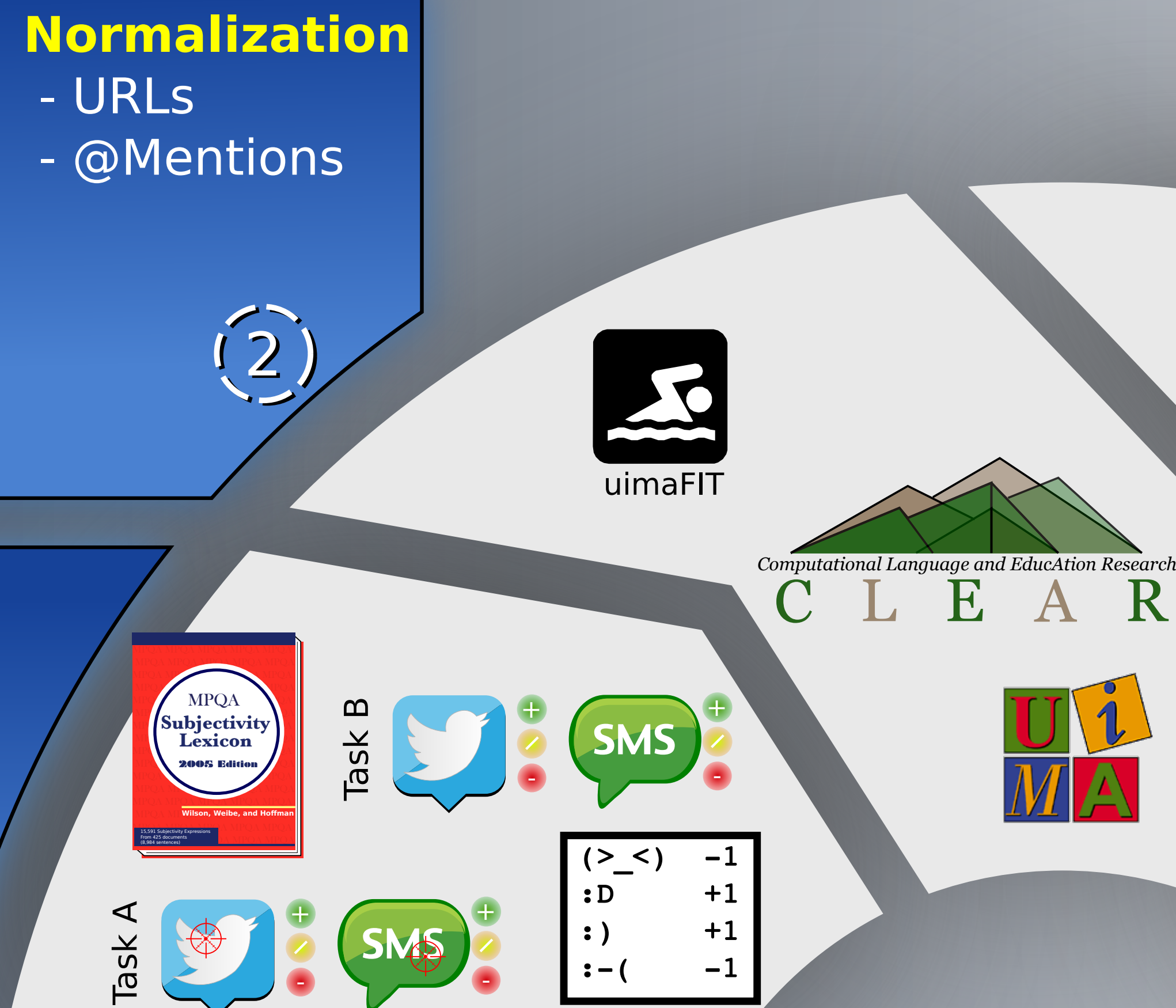
- URLs
- @Mentions

Data And Resources

Sentiment Labeled Tweets & SMS

- Message & Context Polarities

MPQA Subjectivity Lexicon
Negation Words Dictionary
Emoticons Dictionary



Task B Results		F ₊	Rank
Tweet	Unconstrained	0.641	5
	Constrained	0.608	12
SMS	Constrained	0.600	4
	Unconstrained	0.595	5

Task B: Message Polarity Classification

Model Features

- Polarized Bag-of-Words
- Word Polarities
- Emoticon Polarities
- Microblogging
- POS Tags Counts
- Syntactic Dependencies

Model Training & Parameters

- LIBLinear w/ L2-Logistic Regression
- Constrained: Supervised
- Unconstrained: Self-Training

Polarity Lexicon Expansion

1. Applied Task B classifier to approximately 475k unlabeled tweets.
2. Computed co-occurrence statistics between words and sentiment labels
3. Computed word polarity via PMI (Pointwise Mutual Information)
4. Merged new entries with MPQA Lexicon to create an expanded polarity lexicons

$$PMI(word, sentiment) = \log_2 \frac{p(word, sentiment)}{p(word)p(sentiment)}$$

$$polarity(word) = \text{sgn}(PMI(word, positive) - PMI(word, negative))$$

4. Merged new entries with MPQA Lexicon to create an expanded polarity lexicons

Sentiment Analysis on Twitter with Self-Training and Polarity Lexicon Expansion



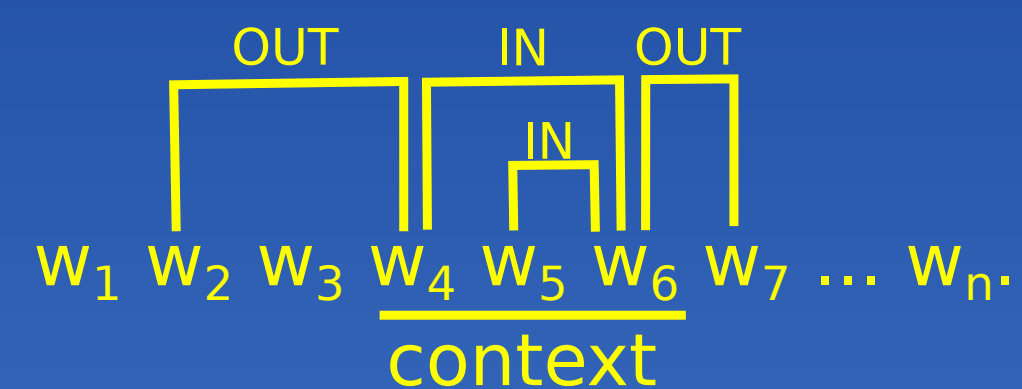
Lee Becker
David Skiba

George Erhart
Valentine Matula

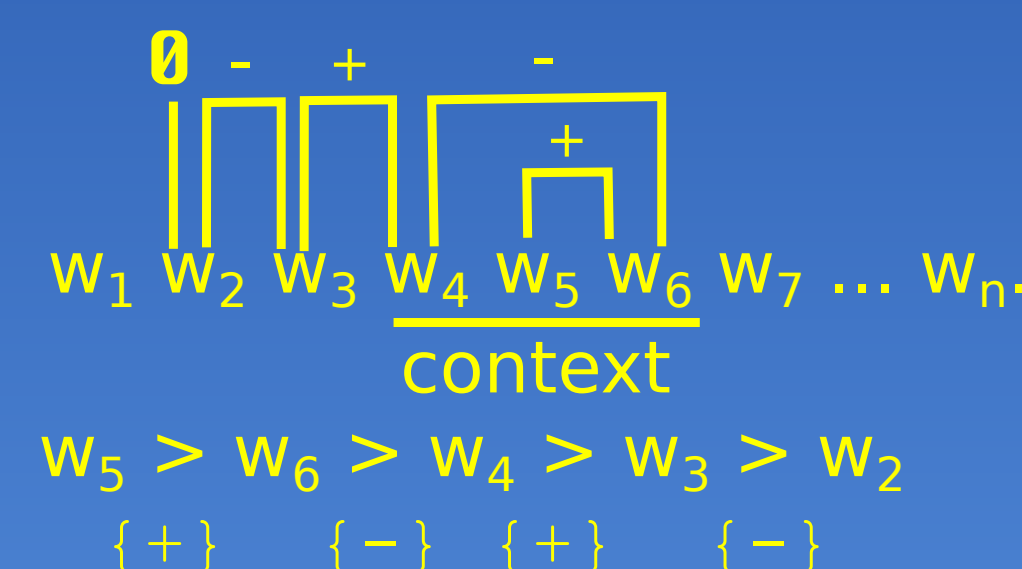
Task A: Contextual Polarity Disambiguation

Additional Model Features

- Scoped Dependencies



- Dependency and Sentiment Paths



Model

- LIBLinear Log. Reg.

Polarity Lexicons

- Constrained: MPQA
- Unconst.: Expanded

Task A Results		F ₊	Rank
Tweet	Unconstrained	0.874	2
	Constrained	0.870	3
SMS	Unconstrained	0.858	3
	Constrained	0.839	4

For more details
and to download resources
scan this QR code.

