

Studienarbeiten zur internetbasierten Kommunikation mit KobRA-Verfahren an der Universität Mannheim

Institutionelle Einbindung

Projektmodul im Masterstudiengang Sprache und Kommunikation

- Studiengang der Philosophischen Fakultät unter Beteiligung der Lehrstühle Anglistik, Germanistik und Romanistik
- Eigenständige Durchführung einer zweisemestrigen Studienarbeit

Zwei Studienarbeiten im Themenfeld der internetbasierten Kommunikation (IBK)

- Laufzeit: HWS 2014 bis FSS 2015
- Betreuung: Harald Lüngen luengen@ids-mannheim.de
Christian Pöllitz christian.poelitz@cs.uni-dortmund.de
Angelika Storrer astorller@mail.uni-mannheim.de
Tassja Weber tasweber@mail.uni-mannheim.de

Datengrundlage beider Studienarbeiten

- Diskussionsseiten der deutschsprachigen Wikipedia aus dem Jahr 2013 (264 Mio. Tokens/550.000 Diskussionsseiten), verfügbar über das IDS Mannheim: <http://www1.ids-mannheim.de/kl/projekte/korpora/verfuegbarkeit.html>
- Werkzeug zur Datenerhebung: RapidMiner mit KobRA-Plugin
- Auswertung mit Excel

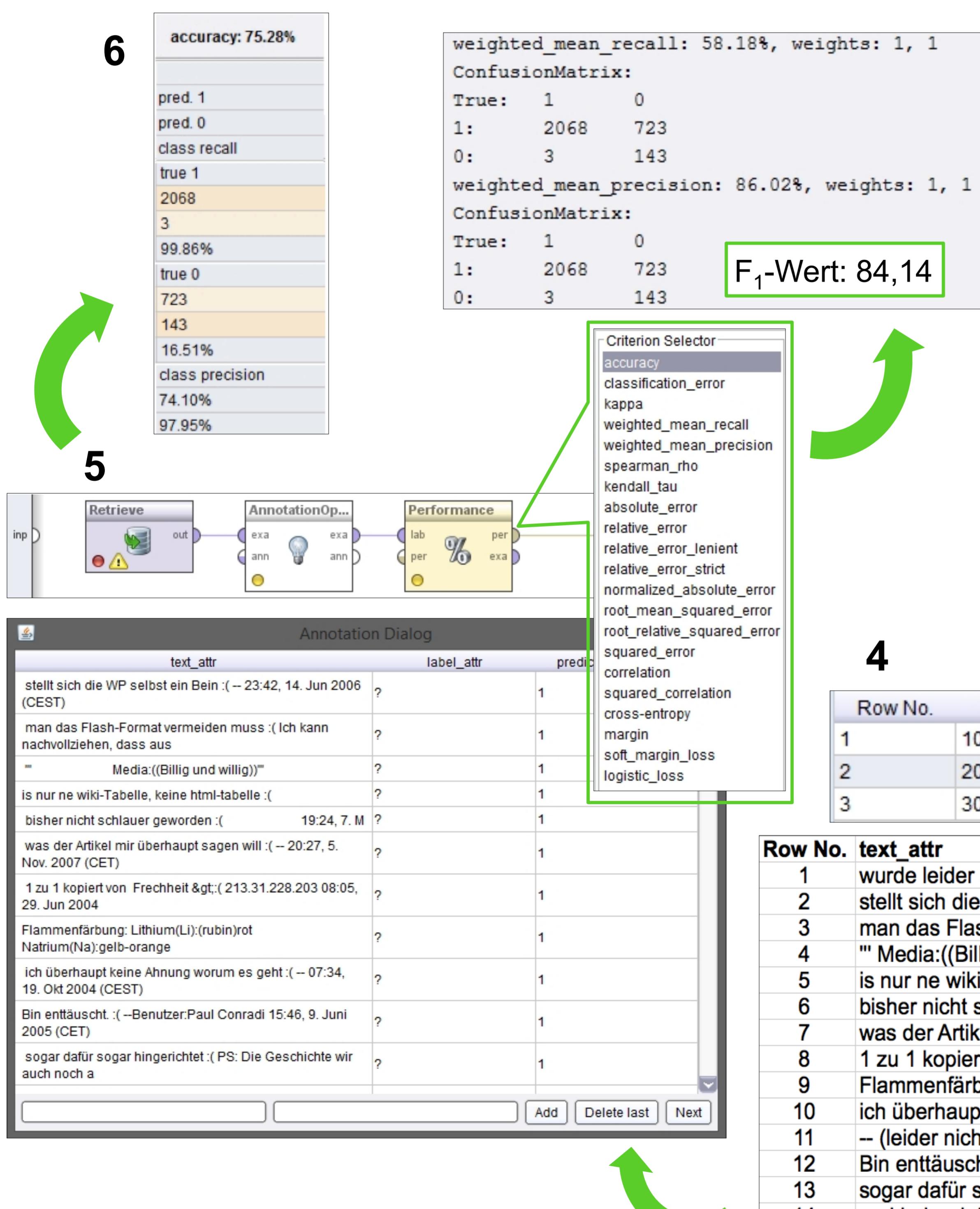
Disambiguierung mit RapidMiner und KobRA-Plugin: Fallstudie zum Active Learning

- Untersuchungsgegenstand: Smiley :(in den Diskussionsseiten
- Motivation: IBK-spezifische Einheiten wie Smileys sind oftmals ambig, die Aussortierung von Falsch-Positiven (s.u.) ist notwendig:
 - 1) *parboiled* > [pa:(r)bɔld, engl.]
 - 2) *Aus: § 2 EBO:(Absatz 3) Die Vorschriften [...]*
 - 3) *einfache Intervalle [...] wie 1:2, 2:3, 3:4, n:(n+1)*

- Analyse und Ergebnisermittlung mit RapidMiner und KobRA-Plugin

Vorgehen:

1. Abfrage im Korpus (ggf. Festlegung der Zufallsstichprobe (sample size) und Annotationsrunden)
2. Durchführung des Active Learning (AL)-Verfahrens
3. Manuelle Annotation (Treffer vs. Pseudotreffer) im ausgegebenen *Annotation Dialog*; Annotationsrunden gemäß Festlegung
4. Anwendung des Verfahrens auf die restliche Stichprobe
5. Bewertung der automatischen Annotation (durch *Performance Classification*) mit anschließender manueller Nachannotation
6. Berechnung der Güte des Verfahrens (*Accuracy*)
7. Evaluation des Verfahrens: Berechnung des *F₁-Wertes*



Studienarbeiten im Bereich IBK

Inflektive in internetbasierter Kommunikation (Laura Herzberg)

- I) Literaturstudie zur historischen Entwicklung von Inflektiven
- II) Empirische Untersuchung zu den Inflektiven *seufz, grübel, räusper*

Fragestellung

Quantitative Analyse der kommunikativen Funktionen im Text: interpretativ, phatisch oder narrativ?

Ergebnis

Die interpretative Funktion tritt bei allen untersuchten Inflektiven am häufigsten auf.

Fragestellung

Treten die Inflektive wie die Interjektionen syntaktisch isoliert auf?

Ergebnis

Die Inflektive treten syntaktisch isoliert auf (Wortartenzugehörigkeit: Interjektionen).

Smileys in internetbasierter Kommunikation (Johannes Götz)

- I) Literaturstudie zur historischen Entwicklung von Smileys
- II) Empirische Untersuchung verschiedener Smileys

Fragestellung

Quantitative Analyse von vierzehn Smileys. Häufigkeitsverteilung?

Ergebnis

Der Smiley ;- tritt mit 80.555 Treffern am häufigsten auf, gefolgt von :-) mit 48.581 und ; mit 38.692 Treffern.

Fragestellung

Kann die Kodierung von Ironie als eine Primärfunktion der Smileys :-), ;-) und :-P angesehen werden?

Ergebnis

Nur jeder vierzehnte Smiley wird primär ironisch gebraucht. Die Ironie-Funktion ist somit nicht dominant.

