

Named Entity Recognition

Sequenzklassifikation mit Hidden Markov Models und dem
Perzeptron-Algorithmus

Dustin Beyer

Web Technologie und Informationssysteme

04.08.2010

Gliederung

- 1 Motivation
- 2 Problemstellung
- 3 Methode
- 4 Implementierung
- 5 Experimente

Gliederung

1 Motivation

2 Problemstellung

3 Methode

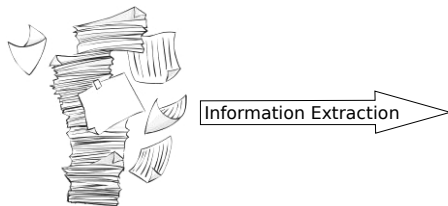
4 Implementierung

5 Experimente

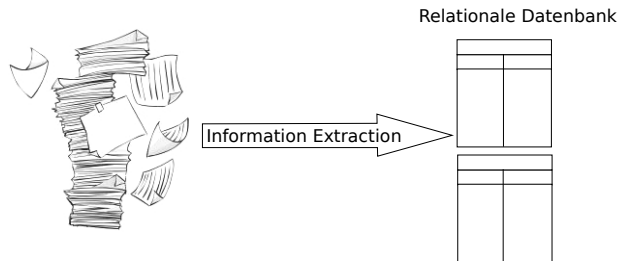
Information Extraction



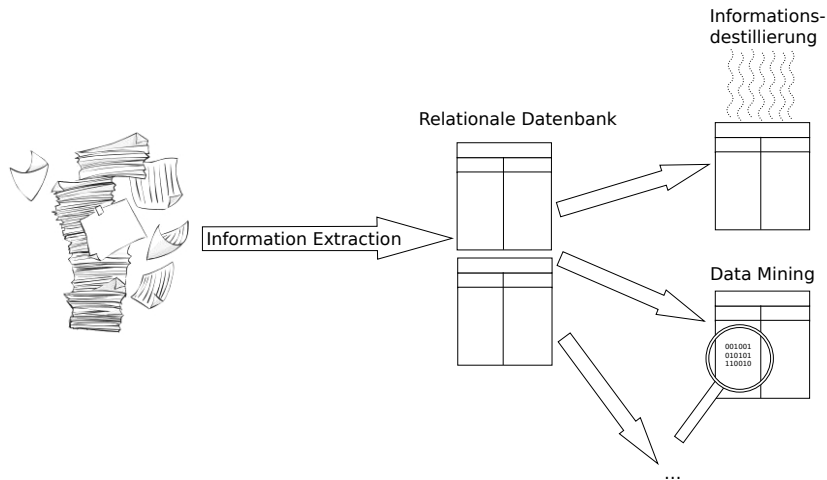
Information Extraction



Information Extraction



Information Extraction



Gliederung

- 1 Motivation
- 2 Problemstellung**
- 3 Methode
- 4 Implementierung
- 5 Experimente

Problemstellung

Beispiel

Offene Stadtinformationssysteme

VON Christiane Schulzki-Haddouti 28. MAI 2010 UM 11:53 UHR

Das im Herbst erst gegründete OpenData Network hat mit OpenBerlin.net ein kleines Projekt gestartet, das Keim eines offenen Stadtinformationssystems werden könnte. Es enthält im Moment Maßnahmen des Konjunkturpakets II sowie statische Karteninformationen zur Sozialstruktur Berlins. Die Daten bzw. Bilder stammen aus dem FIS-Broker der Stadt Berlin.

OpenBerlin.net basiert auf dem Projekt Mapnificent des Berliner Informatikers Stefan Wehrmeyer, der unter anderem auch die Daten des öffentlichen Nahverkehrs sowie Kriminalitätsdaten verwendet. [...] Wehrmeyer selbst ließ sich vom britischen Projekt Mapumental inspirieren, die noch in private beta ist.

Problemstellung

Beispiel

Offene Stadtinformationssysteme

VON [Christiane Schulzki-Hadduti](#) 28. MAI 2010 UM 11:53 UHR

Das im Herbst erst gegründete [OpenData Network](#) hat mit [OpenBerlin.net](#) ein kleines Projekt gestartet, das Keim eines offenen Stadtinformationssystems werden könnte. Es enthält im Moment Maßnahmen des [Konjunkturpakets II](#) sowie statische Karteninformationen zur Sozialstruktur [Berlins](#). Die Daten bzw. Bilder stammen aus dem [FIS-Broker der Stadt Berlin](#).

[OpenBerlin.net](#) basiert auf dem Projekt [Mapnificent](#) des Berliner Informatikers [Stefan Wehrmeyer](#), der unter anderem auch die Daten des öffentlichen Nahverkehrs sowie Kriminalitätsdaten verwendet. [...] [Wehrmeyer](#) selbst ließ sich vom britischen Projekt [Mapumental](#) inspirieren, die noch in private beta ist.

Problemstellung

Beispiel

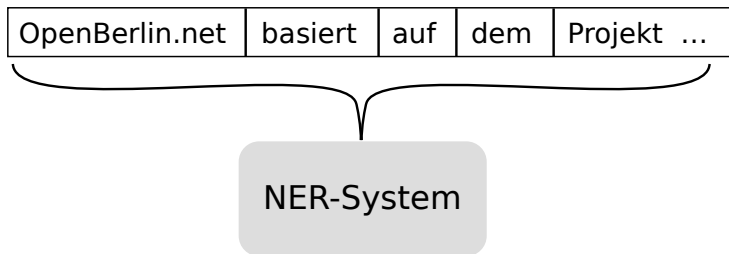
Offene Stadtinformationssysteme

VON [Christiane Schulzki-Haddouti](#) 28. MAI 2010 UM 11:53 UHR

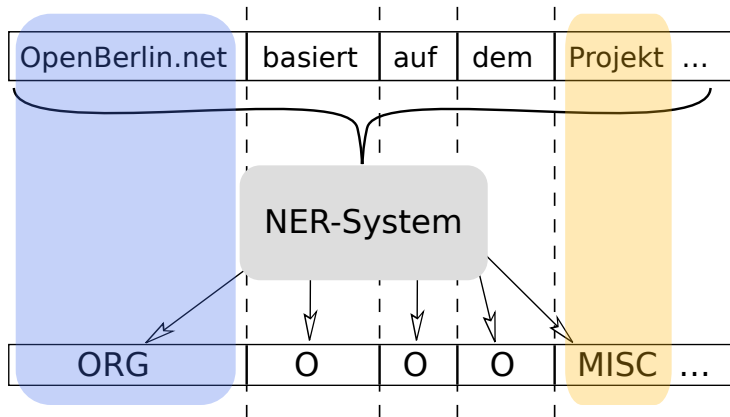
Das im Herbst erst gegründete [OpenData Network](#) hat mit [OpenBerlin.net](#) ein kleines Projekt gestartet, das Keim eines offenen Stadtinformationssystems werden könnte. Es enthält im Moment Maßnahmen des [Konjunkturpakets II](#) sowie statische Karteninformationen zur Sozialstruktur [Berlins](#). Die Daten bzw. Bilder stammen aus dem [FIS-Broker der Stadt Berlin](#).

[OpenBerlin.net](#) basiert auf dem Projekt [Mapnificent](#) des Berliner Informatikers [Stefan Wehrmeyer](#), der unter anderem auch die Daten des öffentlichen Nahverkehrs sowie Kriminalitätsdaten verwendet. [...] [Wehrmeyer](#) selbst ließ sich vom britischen Projekt [Mapumental](#) inspirieren, die noch in private beta ist.

Lösungsansatz

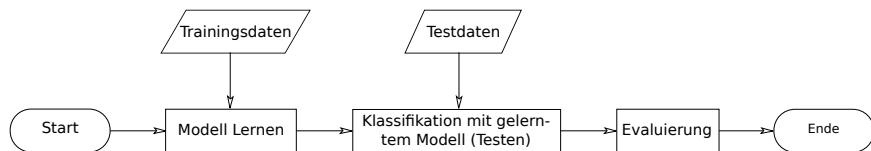


Lösungsansatz

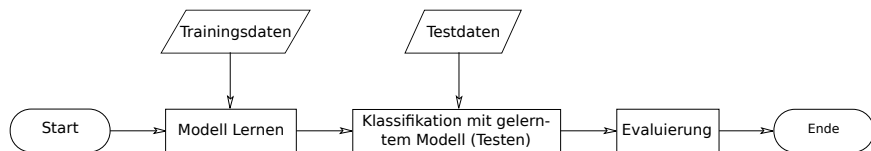


$$\mathcal{K} = \{LOC, ORG, PER, MISC, O\}$$

Maschinelle Lernverfahren

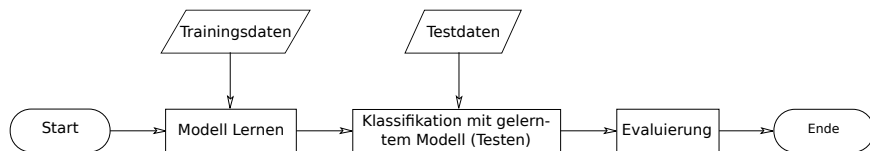


Maschinelle Lernverfahren



- Eingangssequenz: $X = x_1 x_2 \dots x_{|X|}$
- Klassensequenz: $C = c_1 c_2 \dots c_{|X|}$
- Modellierung: $P(C|X) = P(c_1 c_2 \dots c_{|X|} | x_1 x_2 \dots x_{|X|})$

Maschinelle Lernverfahren



- Eingangssequenz: $X = x_1 x_2 \dots x_{|X|}$
- Klassensequenz: $C = c_1 c_2 \dots c_{|X|}$
- Modellierung: $P(C|X) = P(c_1 c_2 \dots c_{|X|} | x_1 x_2 \dots x_{|X|})$
- Dekodierung: $\arg \max_C P(C|X)$

Gliederung

- 1 Motivation
- 2 Problemstellung
- 3 Methode**
- 4 Implementierung
- 5 Experimente

Hidden Markov Model

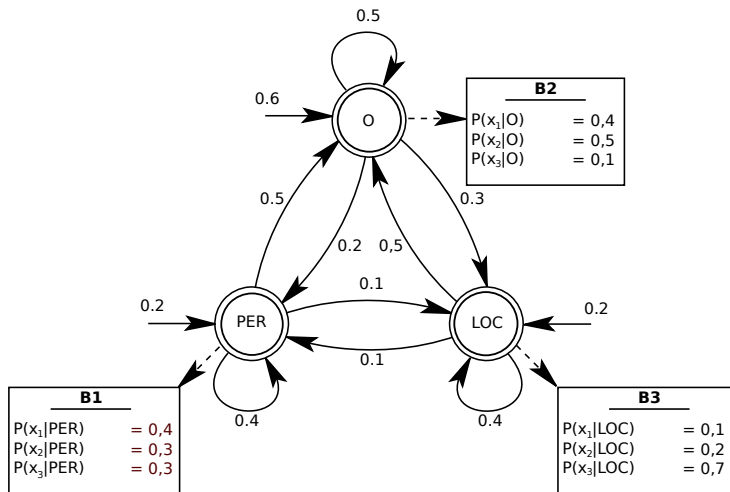
Modellierung:

$$P(C|X) \propto P(X, C) \text{ (Generativ)}$$

Perzeptron-Algorithmus

$$P(C|X) \approx h(X) \text{ (Diskriminativ)}$$

Methode



Hidden Markov Model

Modellierung:

$$P(C|X) \propto P(X, C) \text{ (Generativ)}$$

Perzeptron-Algorithmus

$$P(C|X) \approx h(X) \text{ (Diskriminativ)}$$

Hidden Markov Model

Perzeptron-Algorithmus

Modellierung:

$$P(C|X) \propto P(X, C) \text{ (Generativ)}$$

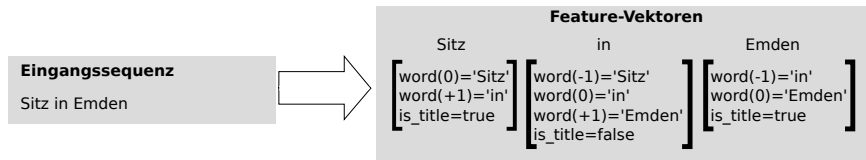
$$P(C|X) \approx h(X) \text{ (Diskriminativ)}$$

Wissensakkumulierung:

Bedingte Wahrscheinlichkeiten

Feature-Vektoren

Methode



Hidden Markov Model

Modellierung:

$$P(C|X) \propto P(X, C) \text{ (Generativ)}$$

Wissensakkumulierung:

Bedingte Wahrscheinlichkeiten

Dekodierung:

Viterbi-Algorithmus (Global)

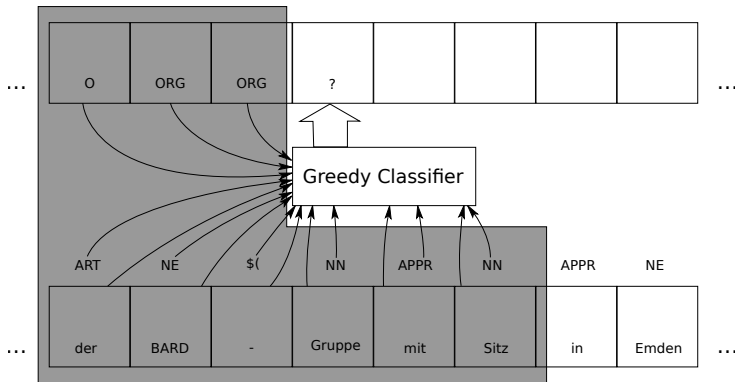
Perzeptron-Algorithmus

$$P(C|X) \approx h(X) \text{ (Diskriminativ)}$$

Feature-Vektoren

Greedy Search (Lokal)

Methode



Hidden Markov Model

Perzeptron-Algorithmus

Modellierung:

$P(C|X) \propto P(X, C)$ (Generativ)

$P(C|X) \approx h(X)$ (Diskriminativ)

Wissensakkumulierung:

Bedingte Wahrscheinlichkeiten

Feature-Vektoren

Dekodierung:

Viterbi-Algorithmus (Global)

Greedy Search (Lokal)

Stärken:

Geringe Laufzeit beim Lernen, Abwägen der Entscheidungen beim Dekodieren, Kann viel Information aus Klassensequenz beachten

Hohe Wissensakkumulierung möglich, sehr flexibel in der Featurewahl, geringe Laufzeit beim Dekodieren

Hidden Markov Model

Perzeptron-Algorithmus

Modellierung:

$P(C|X) \propto P(X, C)$ (Generativ)

$P(C|X) \approx h(X)$ (Diskriminativ)

Wissensakkumulierung:

Bedingte Wahrscheinlichkeiten

Feature-Vektoren

Dekodierung:

Viterbi-Algorithmus (Global)

Greedy Search (Lokal)

Stärken:

Geringe Laufzeit beim Lernen, Abwägen der Entscheidungen beim Dekodieren, Kann viel Information aus Klassensequenz beachten

Hohe Wissensakkumulierung möglich, sehr flexibel in der Featurewahl, geringe Laufzeit beim Dekodieren

Limitierungen:

Alles Wissen in bedingten Wahrscheinlichkeiten, sehr unflexibel in Featurewahl

Keine Konvergenz des Trainingsfehlers, wenn Daten nicht linear separierbar sind

Gliederung

- 1 Motivation
- 2 Problemstellung
- 3 Methode
- 4 Implementierung**
- 5 Experimente

Framework

- Experimentierumgebung
- Datenstrukturen
- Klassifikationsalgorithmen:
 - ▶ Baseline
 - ▶ Unigram Hidden Markov Model
 - ▶ Bigram Hidden Markov Model
 - ▶ Perzeptron

Features im Perzeptron-Algorithmus

Features im Perzeptron-Algorithmus

Trainingsmenge

Sitz in Emden

Feature Template

Windows

word(-1,0) ,
part_of_speech(-1,0),
is_title(0,0)

Features im Perzeptron-Algorithmus

Trainingsmenge

Sitz in Emden

Feature Template

Windows

word(-1,0) ,
part_of_speech(-1,0),
is_title(0,0)



Vokabular

Feature

Index

word(0) = Sitz	→ 0
part_of_speech(0) = NN	→ 1
word(-1) = Sitz	→ 2
word(0) = in	→ 3
part_of_speech(-1) = NN	→ 4
part_of_speech(0) = APPR	→ 5
word(-1) = in	→ 6
word(0) = Emden	→ 7
part_of_speech(-1) = APPR	→ 8
part_of_speech(0) = NE	→ 9
is_title(0) = true	→ 10

Gliederung

- 1 Motivation
- 2 Problemstellung
- 3 Methode
- 4 Implementierung
- 5 Experimente**

Setup



Setup



- Datensatz aus dem CoNLL-2003 Shared Task.

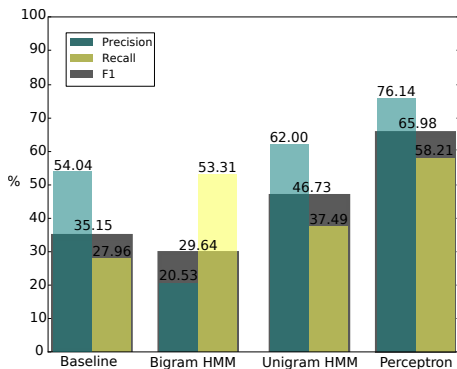
Deutsch	Sätze	Token
Training	12705	206931
Development	3068	51444
Test	3160	51943

Englisch	Sätze	Token
Training	14987	203621
Development	3466	51362
Test	3684	46435

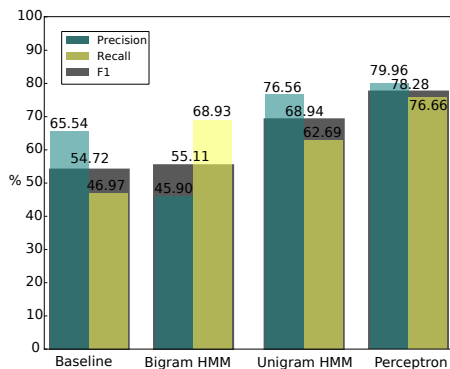
Deutsch	LOC	MISC	ORG	PER
Training	4363	2288	2427	2773
Development	1181	1010	1241	1401
Test	1035	670	773	1195

Ergebnisse

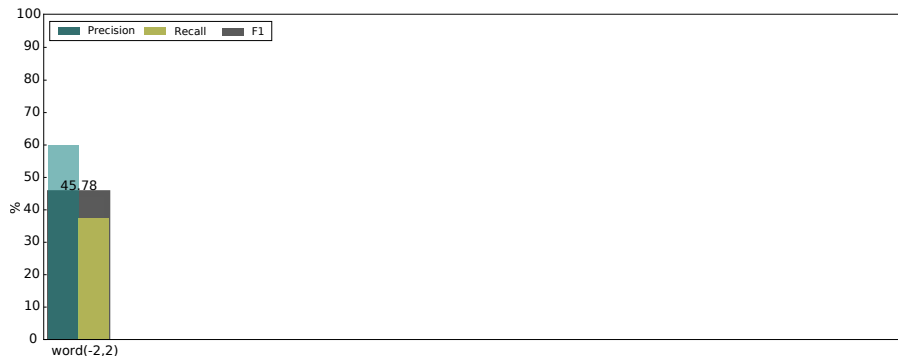
Deutscher Datensatz



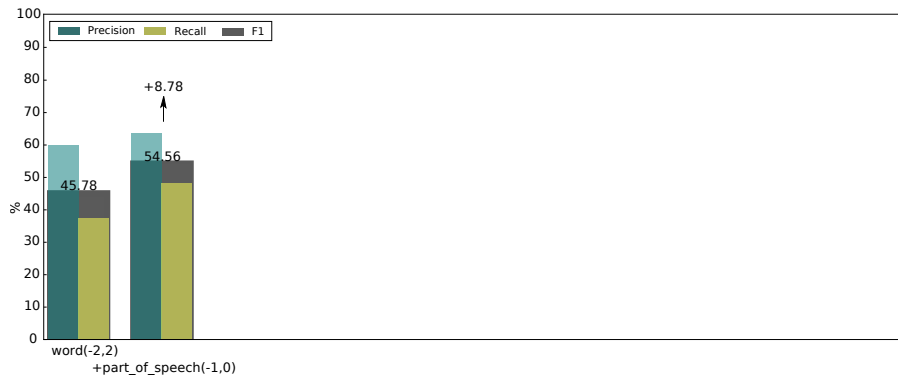
Englischer Datensatz



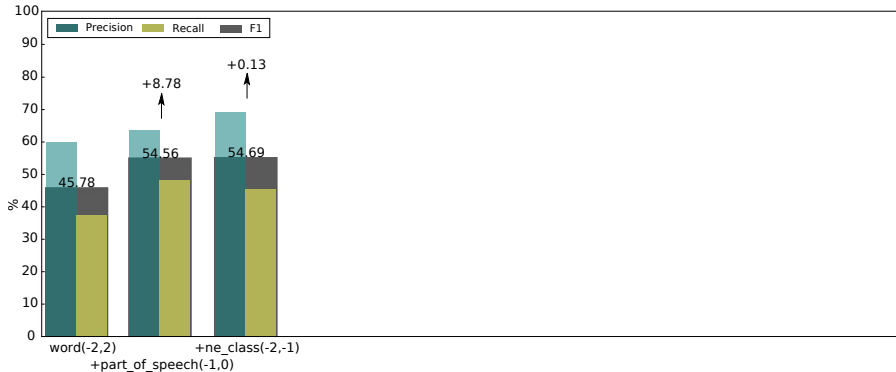
Einfluss der Features (Deutscher Datensatz)



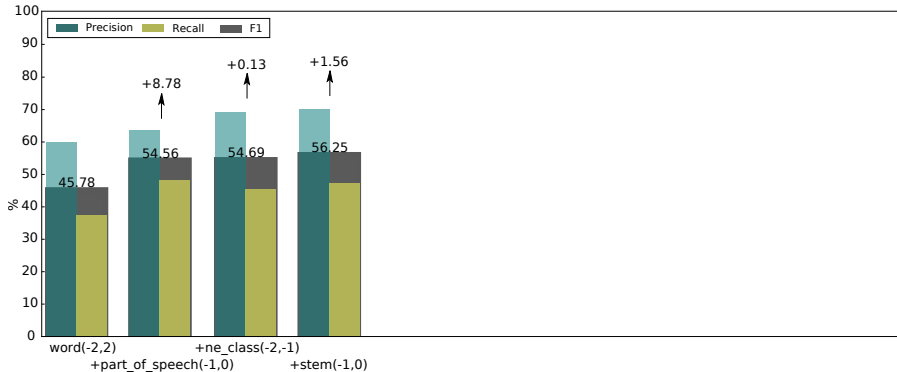
Einfluss der Features (Deutscher Datensatz)



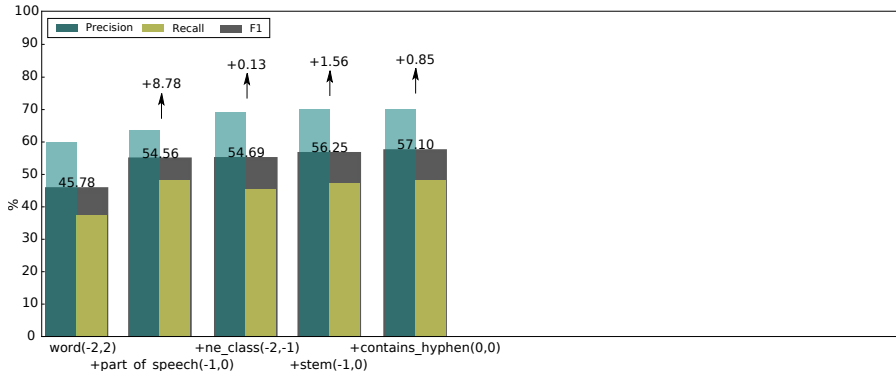
Einfluss der Features (Deutscher Datensatz)



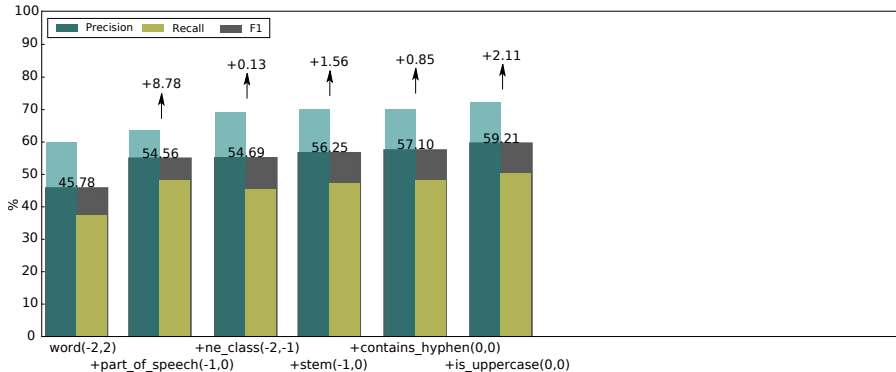
Einfluss der Features (Deutscher Datensatz)



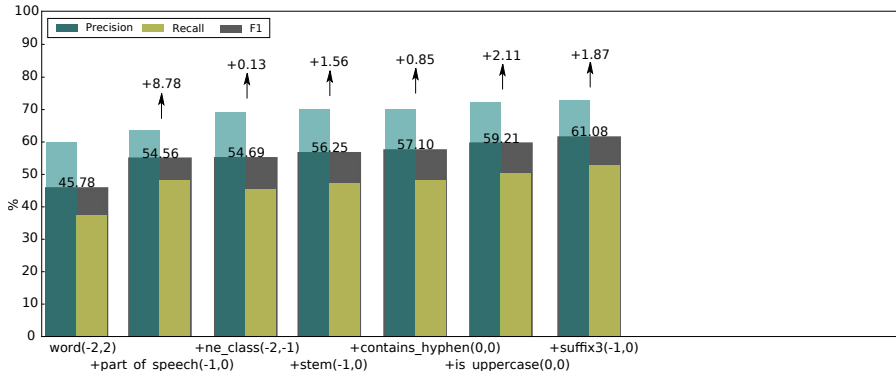
Einfluss der Features (Deutscher Datensatz)



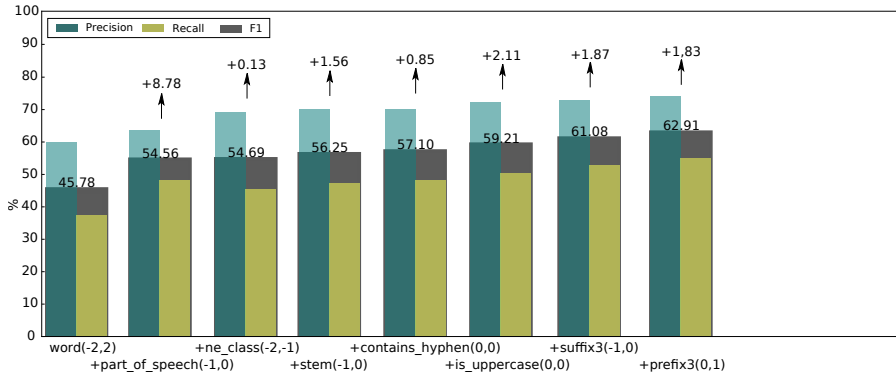
Einfluss der Features (Deutscher Datensatz)



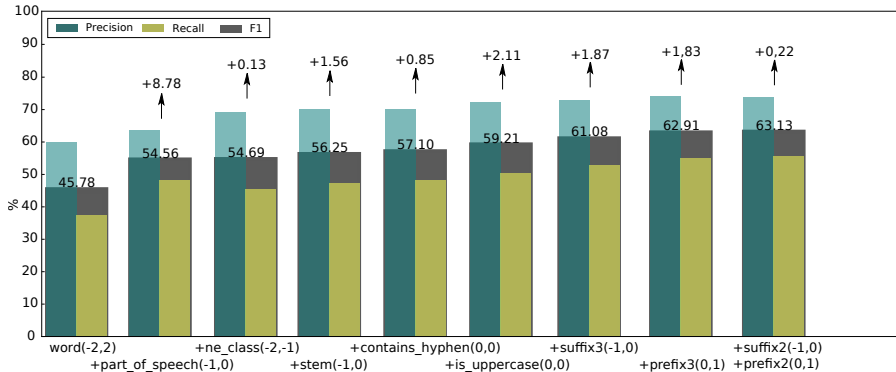
Einfluss der Features (Deutscher Datensatz)



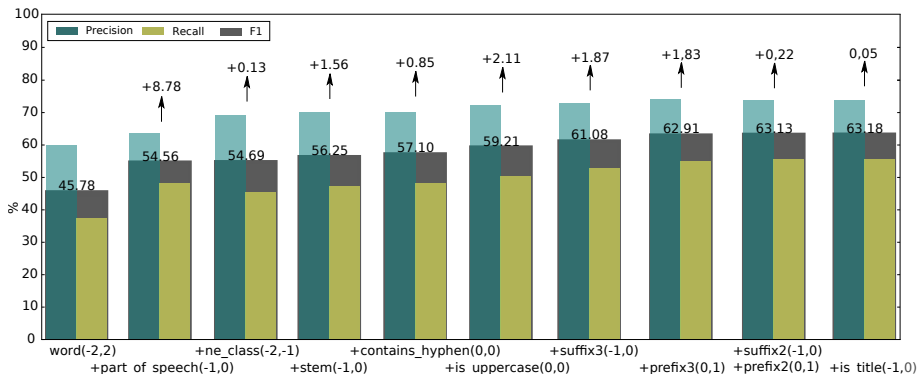
Einfluss der Features (Deutscher Datensatz)



Einfluss der Features (Deutscher Datensatz)



Einfluss der Features (Deutscher Datensatz)



Gesamtgewinn: **17,4 % F₁-Measure**

Größe des Featurevokabulars

- Deutsch: 223617
- Englisch: 74964

- Der Perzeptron-Algorithmus scheint sich besser für Named Entity Recognition zu eignen als Hidden Markov Models.
- Die Einbeziehung einer großen Menge an Features scheint eine Schlüsselqualifikation für Named Entity Recognition zu sein.
- Alle Algorithmen erzielen auf dem englischen Datensatz bessere Ergebnisse als auf dem deutschen.

Demonstration des Frameworks

- Speech and Language Processing

[Jurafsky, D. und Martin, J. H., Prentice Hall, 2008]

- Information Extraction: Distilling Structured Data from Unstructured Text

[McCallum, A., ACM, 2005]

- Markov Models for language-independent named entity recognition

[Malouf, R., CoNLL, 2002]

- Design challenges and misconceptions in named entity recognition

[Ratinov, L. und Roth, D., ACL, 2009]

- Discriminative Training Methods for Hidden Markov Models: Theory and Experiments
with Perceptron Algorithms

[Collins, M., NLP, 2002]

Danke für Ihre Aufmerksamkeit!