

Robust Named Entity Recognition (NER) in Idiosyncratic Domains

Eine automatische Identifikation und Klassifikation von Eigennamen

Gliederung

1 Einführung

2 Ein neues Neuronales Netzwerk

3 Dataxis

4 Evaluation

1.1 Was sind Eigennamen?

Personennamen: *“Sebastian Arnold”*

Ortsnamen: *“Leipzig”*

Institutionen: *“Universität Leipzig”*

Zeitpunkt: *“2018”*

Beispiel

“Jim bought 300 shares of Acme Corp. in 2006.”

1.1 Was sind Eigennamen?

Personennamen: *“Sebastian Arnold”*

Ortsnamen: *“Leipzig”*

Institutionen: *“Universität Leipzig”*

Zeitpunkt: *“2018”*

Beispiel

“**[Jim]**_{Person} bought 300 shares of **[Acme Corp.]**_{Organization} in **[2006]**_{Time}.”

1.2 Wozu braucht man NER?

Entity Linking

Slot filling tasks

Googles Knowledge Graph

Beispiel: „***Paris** ist the capital of **France***“

1.2 Wozu braucht man NER?

Entity Linking

Slot filling tasks

Googles Knowledge Graph

Beispiel: „*Buche ein Hotel in **Berlin**, check in am **27.01.2018** und check out am **29.1.2018**“*“

1.2 Wozu braucht man NER?

Entity Linking

Slot filling tasks

Googles Knowledge Graph



Löwe

Tier

Der Löwe ist eine Art der Katzen. Er lebt im Unterschied zu anderen Katzen in Rudeln, ist durch die Mähne des Männchens gekennzeichnet und ist heute in Afrika sowie im indischen Bundesstaat Gujarat beheimatet. [Wikipedia](#)

Gattung: Eigentliche Großkatzen (Panthera)

Ordnung: Raubtiere (Carnivora)

Lebenserwartung: 10 – 14 Jahre (Erwachsener, In der Wildnis)

Geschwindigkeit: 80 km/h (Maximum, In kurzen Stößen)

Gewicht: Männlich: 190 kg (Erwachsener), Weiblich: 130 kg (Erwachsener)

Höhe: Männlich: 1,2 m (Erwachsener, Schulterhöhe), Weiblich: 1,1 m (Erwachsener, Schulterhöhe)

Länge: Männlich: 1,7 – 2,5 m (Körperlänge), Weiblich: 1,4 – 1,8 m (Körperlänge)

Beispiel Knowledge Graph für den Suchbegriff „Löwe“

<https://www.google.de/search?client=opera&q=löwe&sourceid=opera&ie=UTF-8&oe=UTF-8>

1.3 Related Works

Tagger: HMMs, MEMMs, CRFs

[McCallum und Li 2003] und [Collins 2002]

1.3 Related Works

Tagger: HMMs, MEMMs, CRFs

[McCallum und Li 2003] und [Collins 2002]

Neuronale Netzwerke

- Ont-Hot encoding
- Skip-Gram-Modelle

→ Kombinationen aus beidem

1) Zu verschiedene Textdomänen!

2) Zu viele, benötigte Trainingsdaten!

Gliederung

1 Einführung

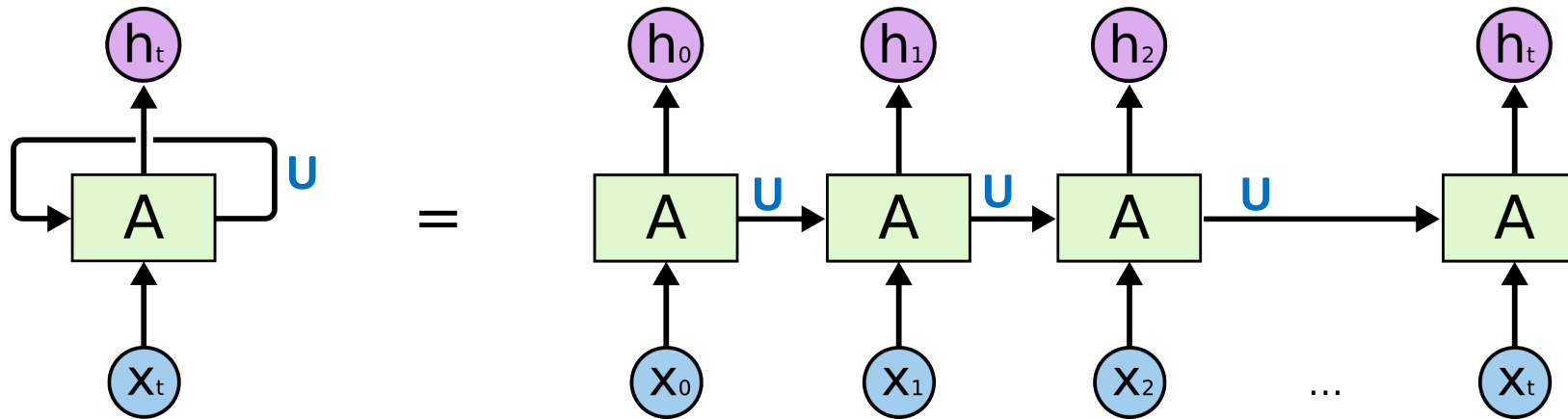
2 Ein neues Neuronales Netzwerk

3 Dataxis

4 Evaluation

2.1 Wiederholung RNN

$$h_t = \phi(Ax + Uh_{t-1})$$



Ein aufgefaltetes RNN

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

2.1 Wiederholung RNN

Long-Term Dependency Problem

“I grew up in France an speak ?”

vs “I grew up in France... I speak fluent ?”

2.1 Wiederholung RNN

Long-Term Dependency Problem

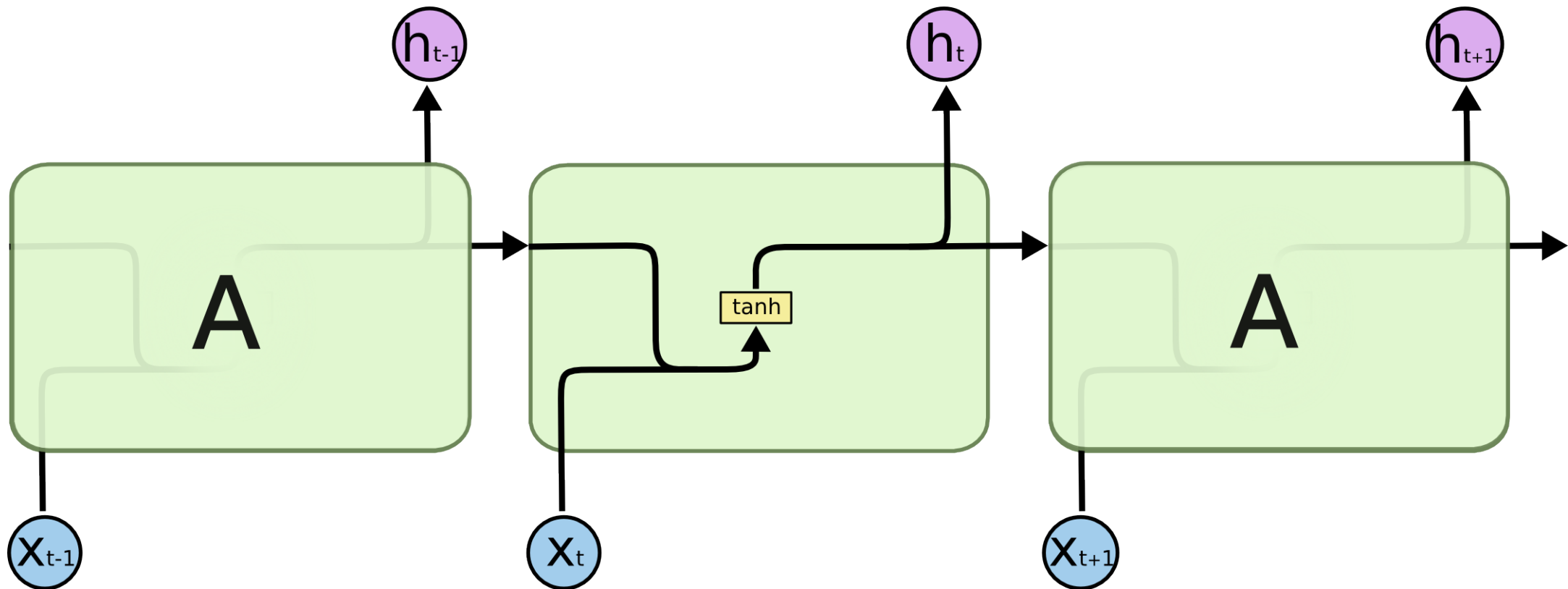
“I grew up in France an speak ?”

vs “I grew up in France... I speak fluent ?”

Vanishing Gradient Problem

- Relationen über Multiplikationen erzeugen zu kleine Ableitungen

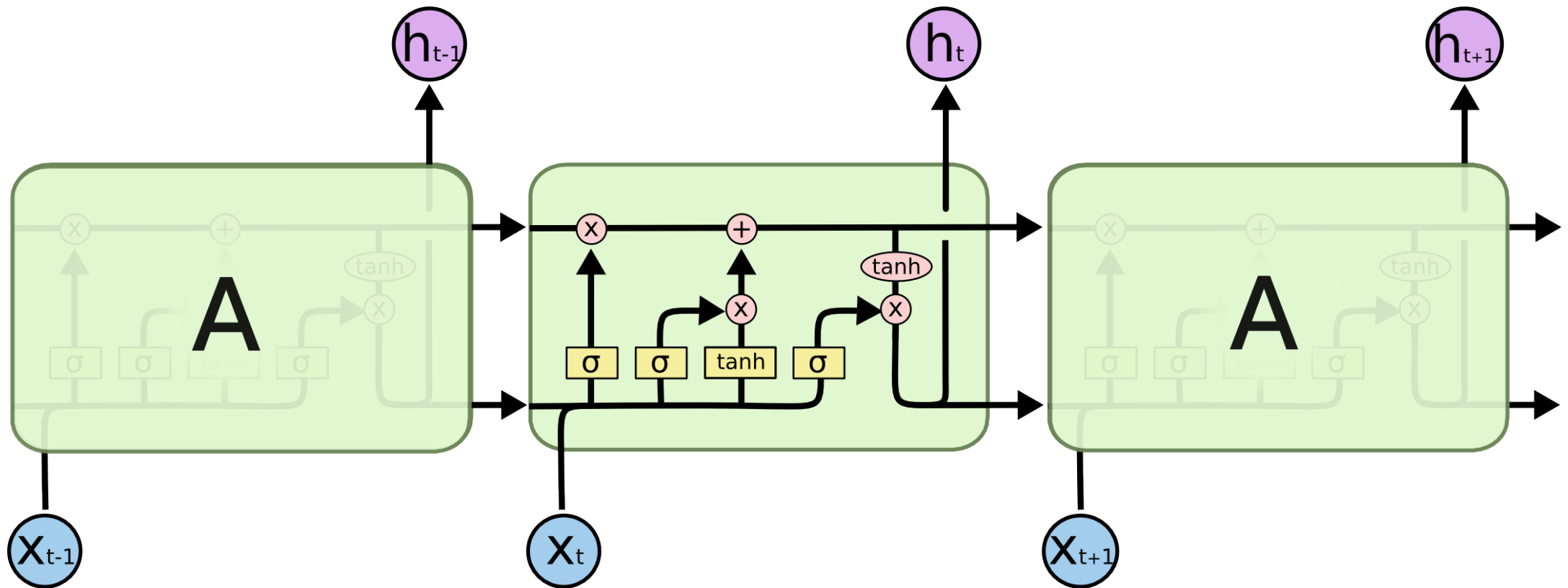
2.1 Wiederholung RNN



Sequenzielle Berechnung von h in einem RNN

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

2.2 LSTM (Long Short-Term Memory)



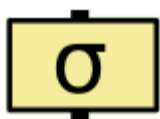
Aufbau der Zelle eines LSTMs

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

2.2 LSTM

\otimes = Punktweise Vektormultiplikation, z.B. $\begin{bmatrix} 0.5 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$

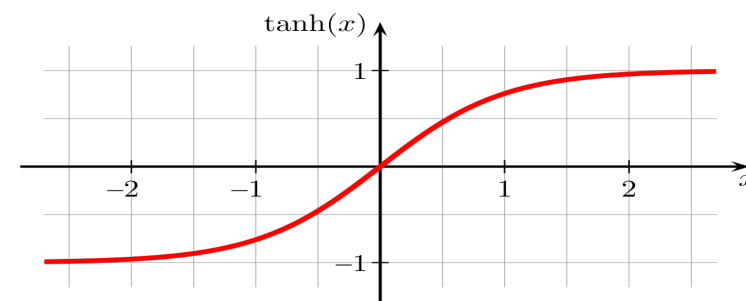
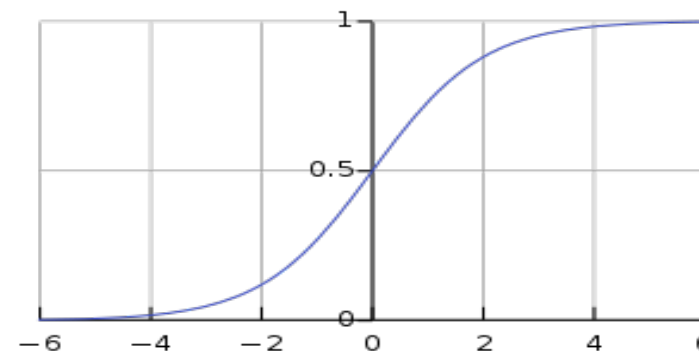
\oplus = Vektoraddition, z.B. $\begin{bmatrix} 0.5 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 3 \\ 3 \end{bmatrix}$



= Sigmoid Funktion: $\sigma(x) \in (0, 1)$



= Hyperbel Funktion: $\tanh(x) \in (-1, 1)$



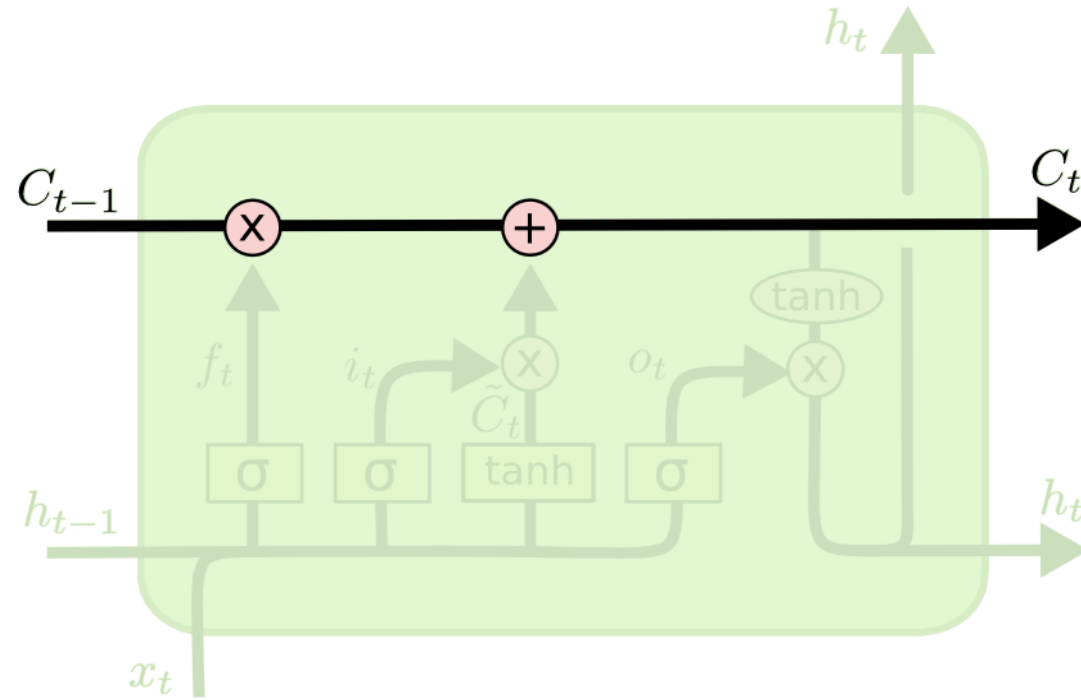
Oben: Sigmoid, unten: Tanh

<https://tinyurl.com/ycggxehs>,

<https://tinyurl.com/yb2hu99p>

2.2 LSTM

Speicher: Cell State

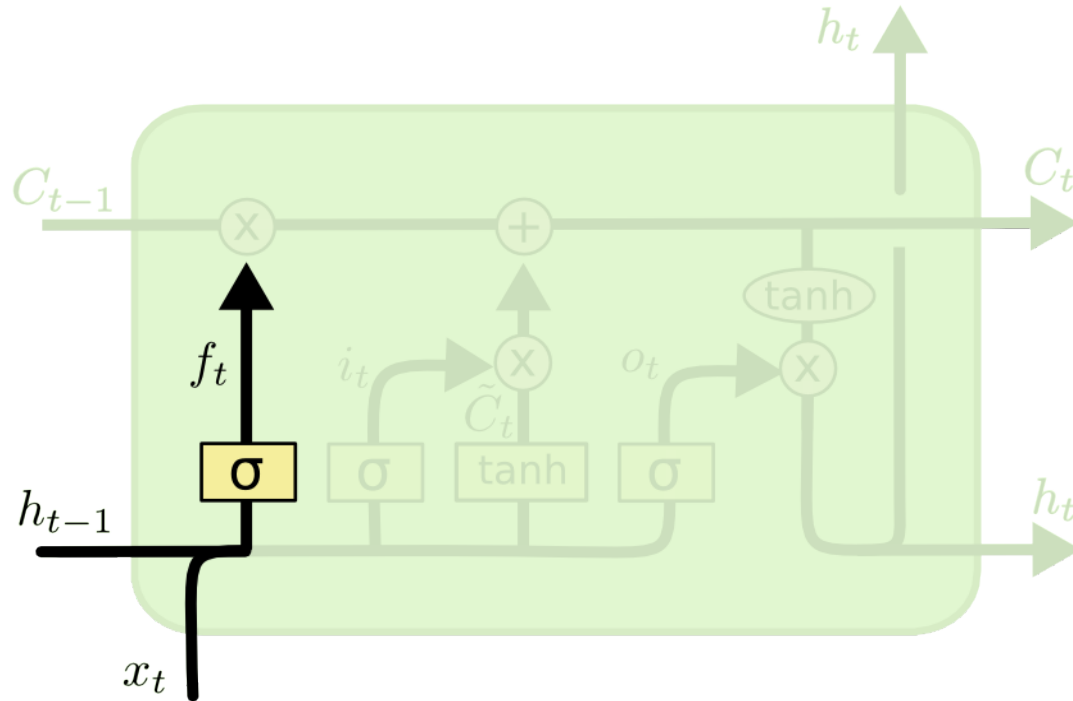


Sequenzielle Arbeit auf dem Speicher C_{t-1}

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

2.2 LSTM

Löschfunktion: Forget-Gate f_t



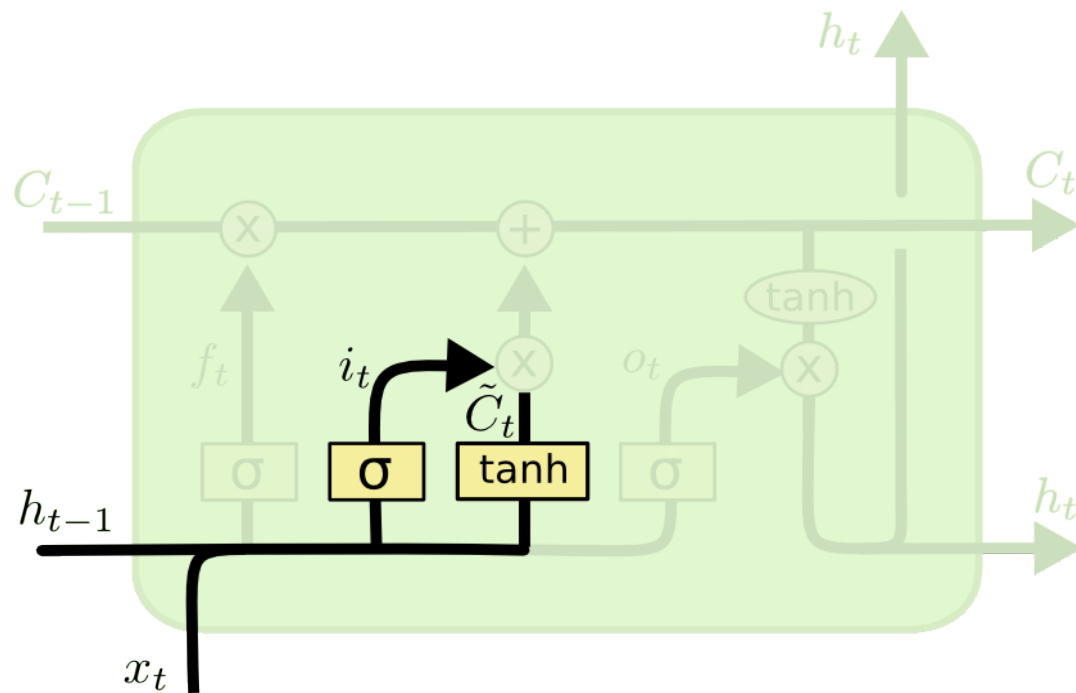
Berechnung zu löschender Informationen

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

2.2 LSTM

Speicherfunktion: Input-Gate (i_t) und Input (\tilde{C}_t)



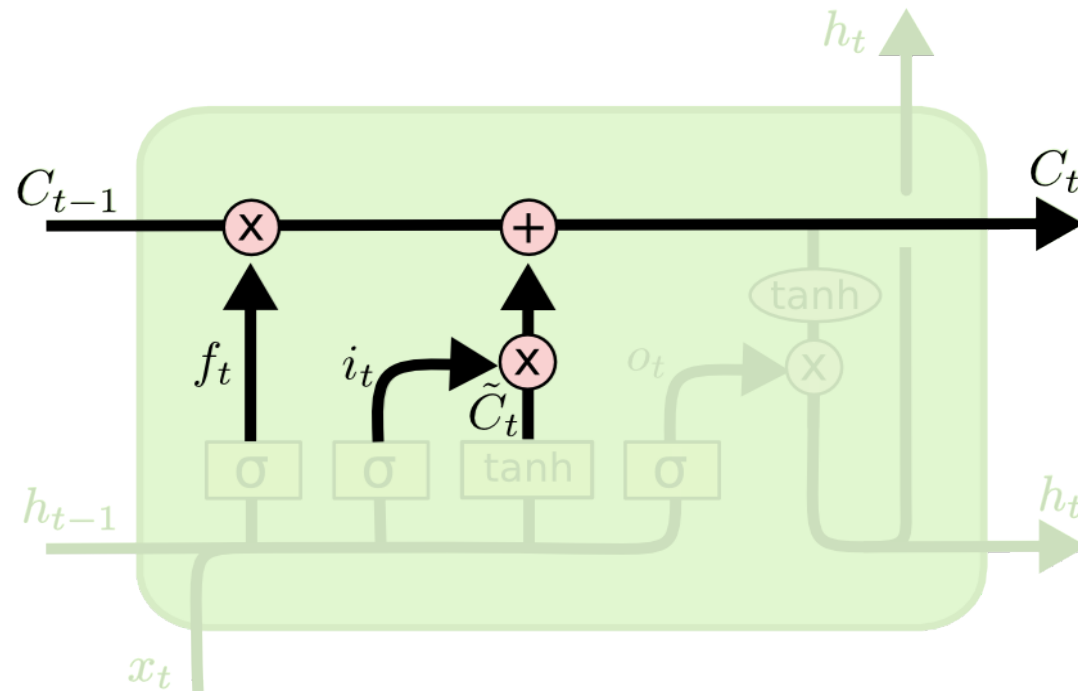
Berechnung neuer Informationen

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

2.2 LSTM

Funktionen auf Speicher anwenden



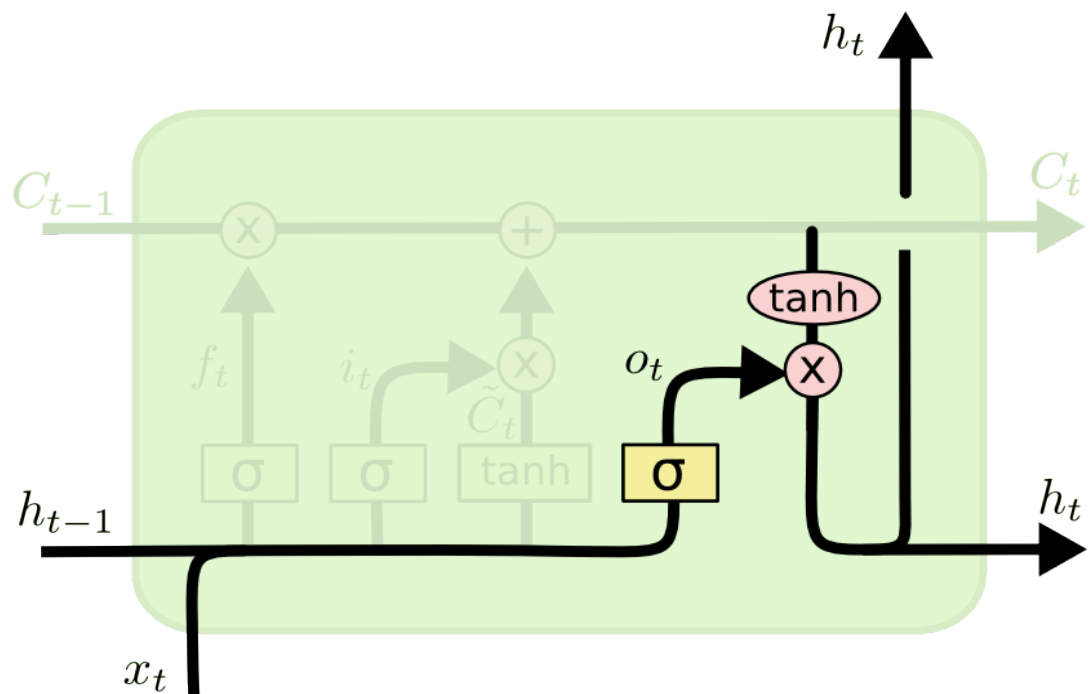
- 1) Lösche Speicher
- 2) Füge neue Informationen hinzu

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

2.2 LSTM

Output-Gate o_t und Hidden State h_t



Berechnung der Ausgabe des Layer: h_t

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

2.2 LSTM

RNN

- Speicher = Rückkopplungen
- Summe langer Multiplikation vieler Ableitungen → **geg. 0**

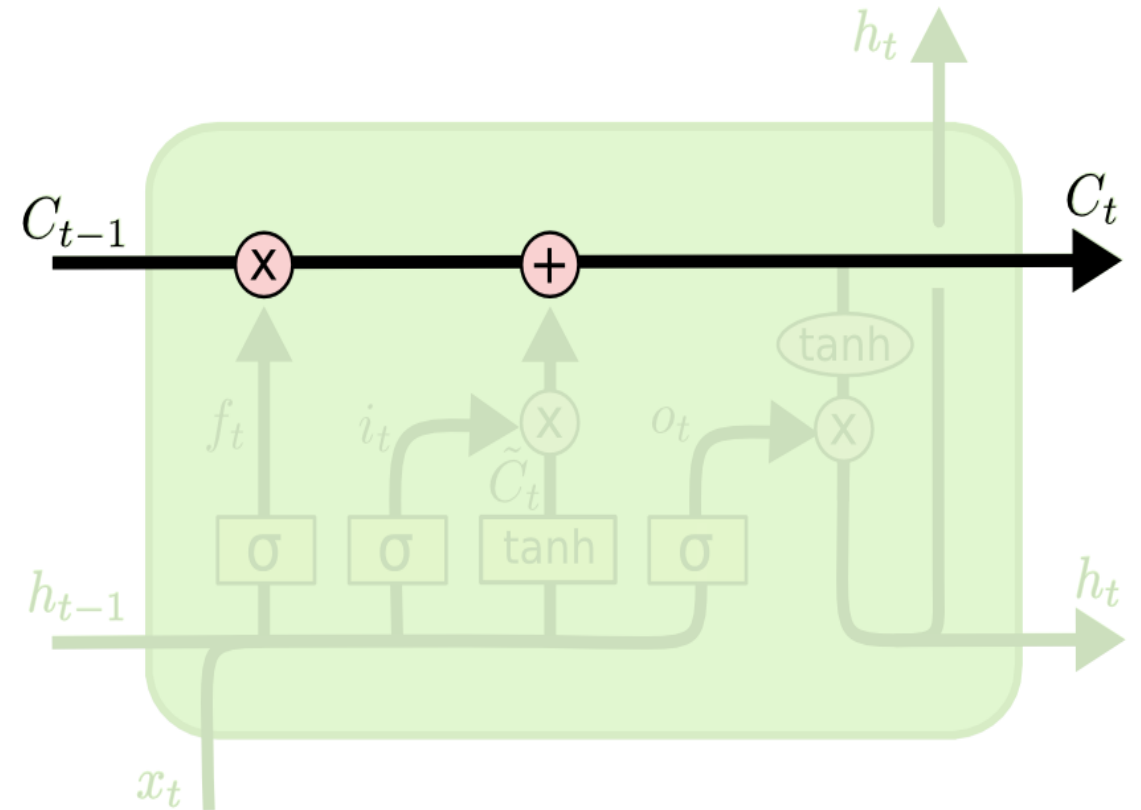
2.2 LSTM

RNN

- Speicher = Rückkopplungen
- Summe langer Multiplikation vieler Ableitungen → **geg. 0**

LSTM

- Speicher = **Cell-State**
- Informationen lange erhalten



Cell-State eines LSTMs

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

2.3 Vom LSTM zum Bidirectional LSTM (BLSTM)

Geg.: *Trainingsdaten (Text)*

Ges.: *BLSTM*

2.3 Vom LSTM zum Bidirectional LSTM (BLSTM)

Geg.: *Trainingsdaten (Text)*

Ges.: *BLSTM*

BLSTM = 2 LSTMs

- 1. *LSTM*: Training auf Text
- 2. *LSTM*: Training auf reversiertem Text

Ergebnis: *Konkatenation beider Ausgaben*

Gliederung

1 Einführung

2 Ein neues Neuronales Netzwerk

3 Datexis

4 Evaluation

3.1 Trigramm Hashing

Trainingsatz

„John likes to watch football games.“



#jo	1	ike	7
joh	2	kes	8
ohn	3	es#	9
hn#	4	#to	10
#li	5	to#	11
lik	6

3.1 Trigramm Hashing

Trainingsatz

„John likes to watch football games.“



#jo	1	ike	7
joh	2	kes	8
ohn	3	es#	9
hn#	4	#to	10
#li	5	to#	11
lik	6



$Tri(John) =$

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

3.2 Softmax Classifier

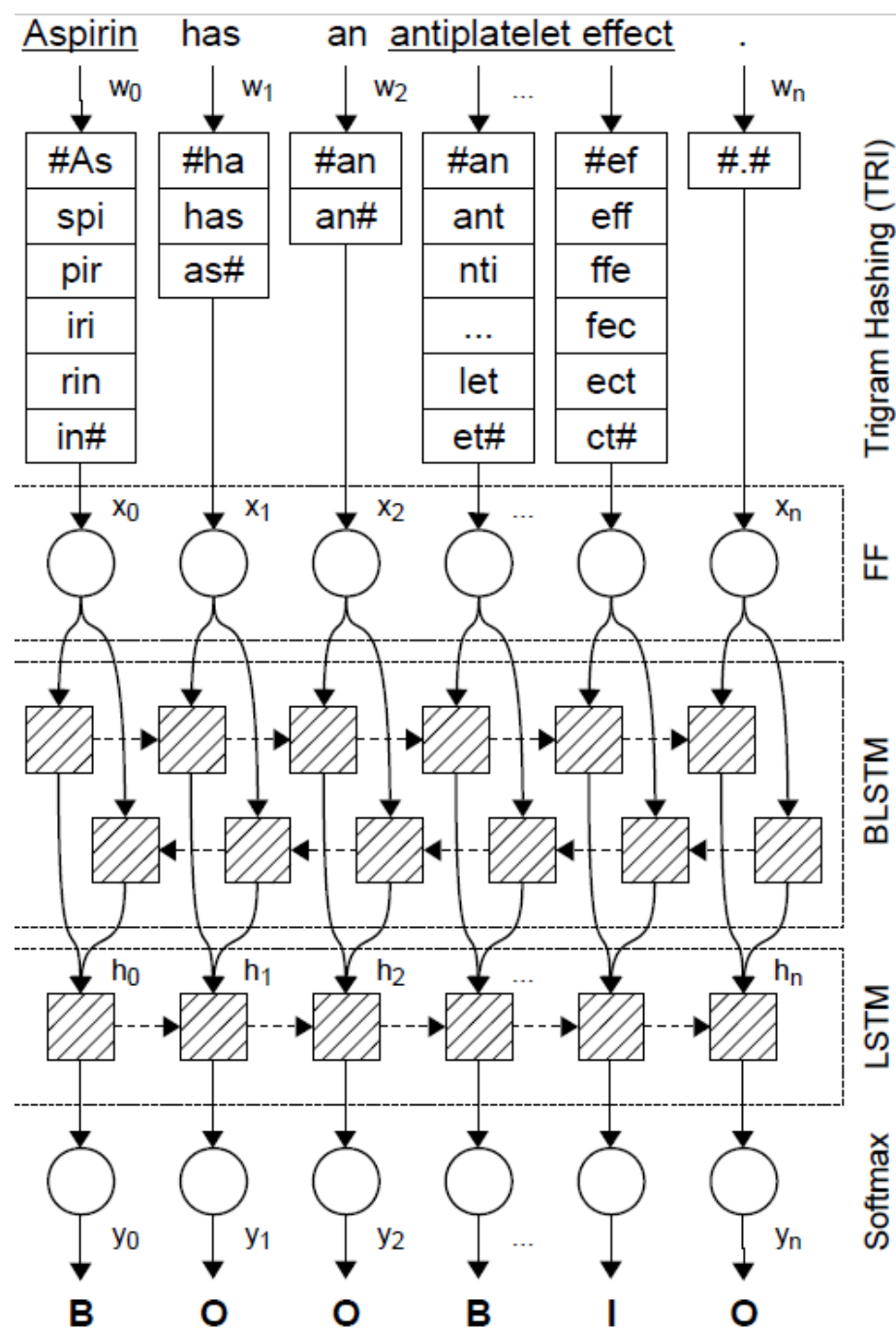
Ergebnis: $y_t \in \{B, I, O\} \rightarrow$ **Begin, Inside, Outside**

3.2 Softmax Classifier

Ergebnis: $y_t \in \{B, I, O\} \rightarrow$ **Begin, Inside, Outside**

Jim	bought	300	shares	of	Acme	Corp.	in	2006
↓	↓	↓	↓	↓	↓	↓	↓	↓
B	O	O	O	O	B	I	O	B

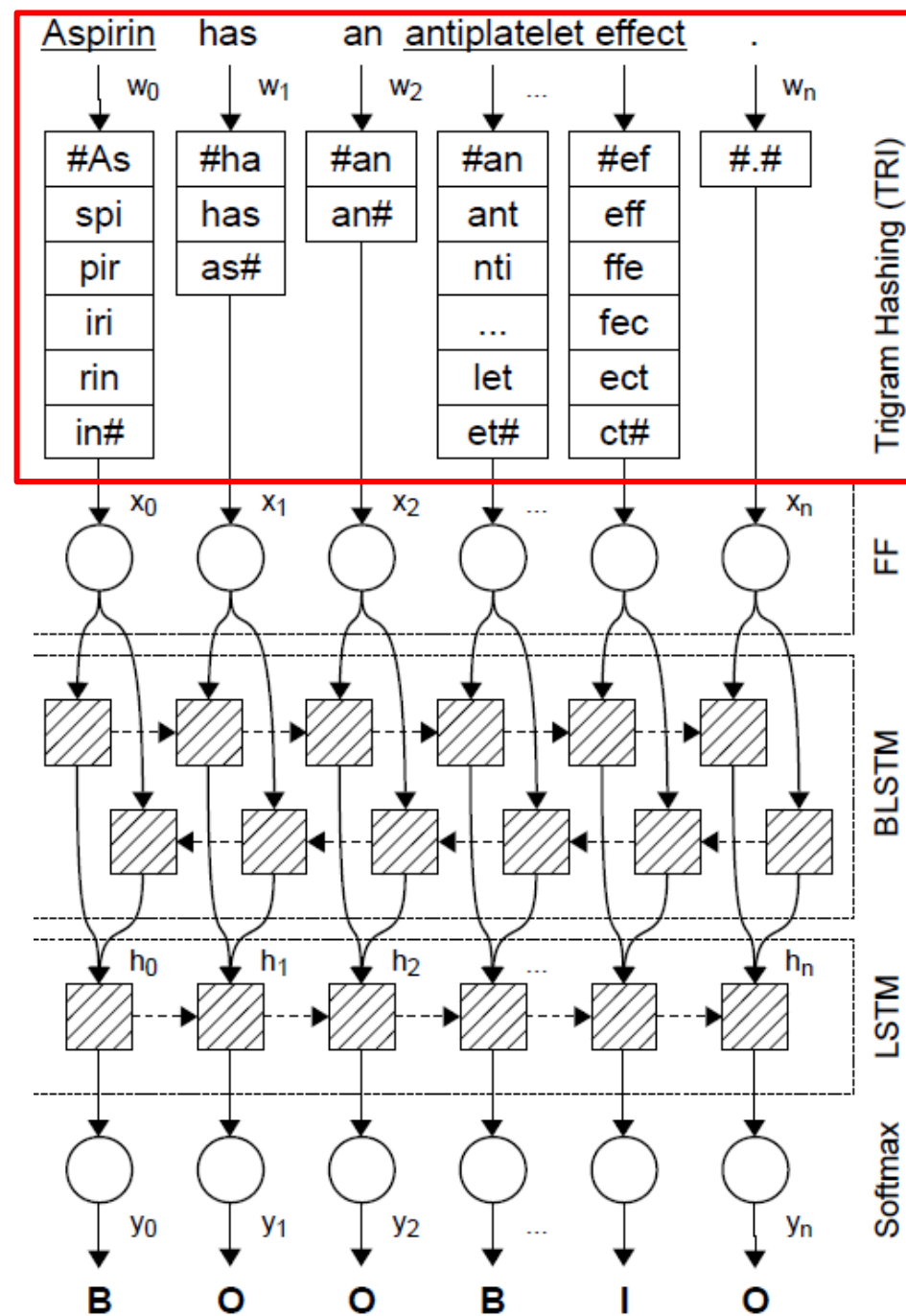
3.3 Architektur



Architektur des LSTM Netzwerks zur Erkennung der Named Entities

[Arnold et al. 2016]

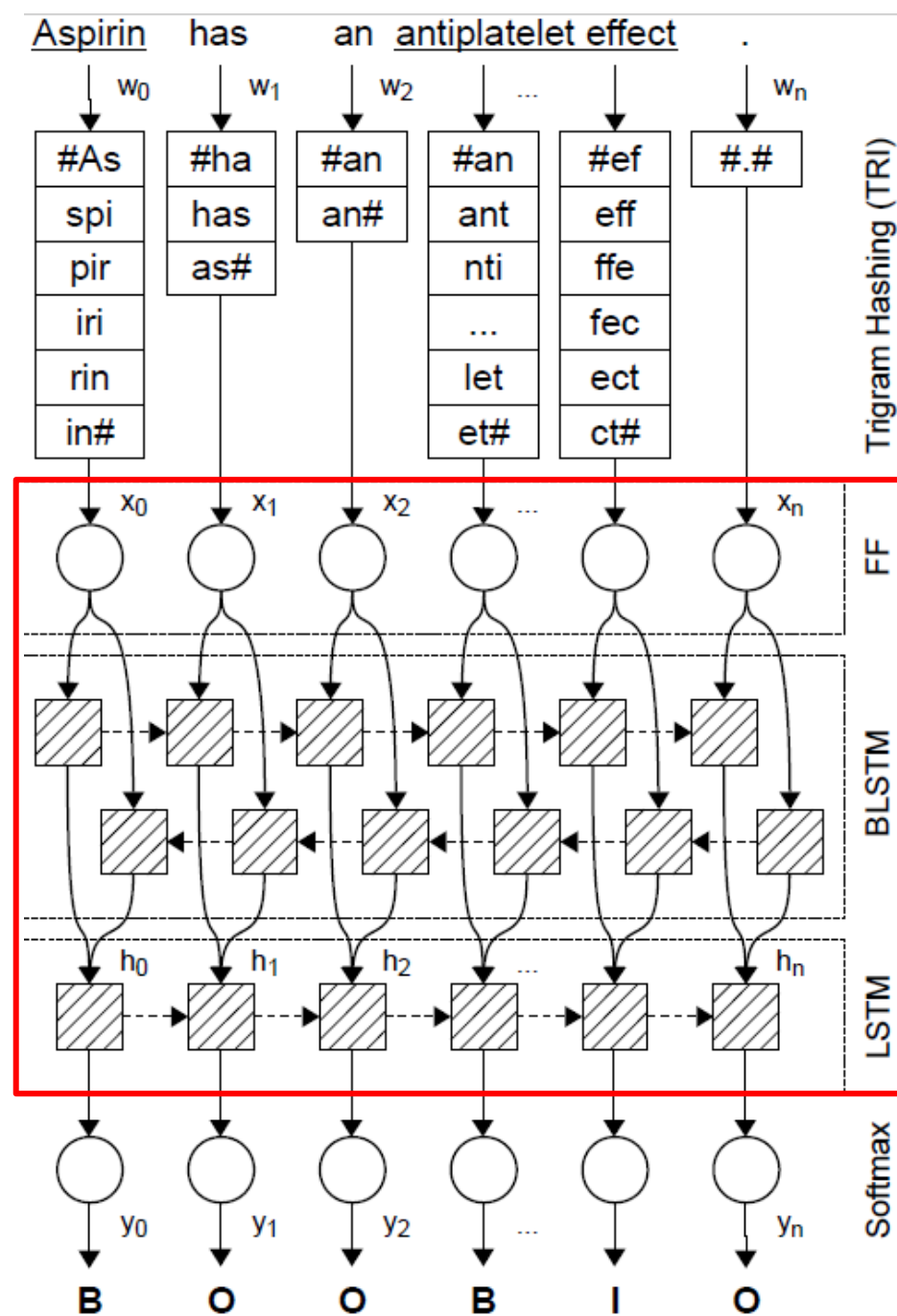
3.3 Architektur



Architektur des LSTM Netzwerks zur Erkennung der Named Entities

[Arnold et al. 2016]

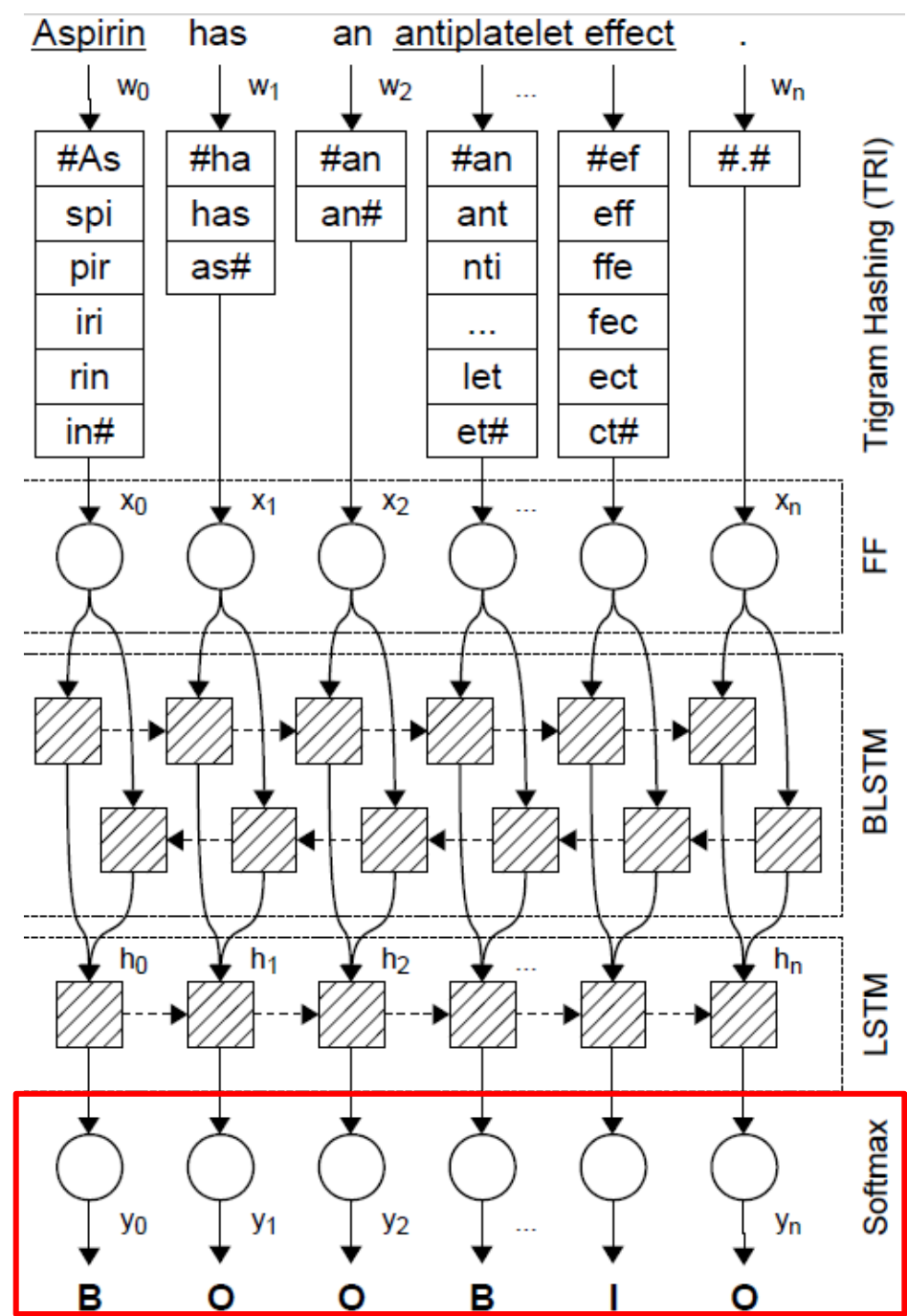
3.3 Architektur



Architektur des LSTM Netzwerks zur Erkennung der Named Entities

[Arnold et al. 2016]

3.3 Architektur



Architektur des LSTM Netzwerks zur Erkennung der Named Entities

[Arnold et al. 2016]

Gliederung

1 Einführung

2 Ein neues Neuronales Netzwerk

3 Dataxis

4 Evaluation

4.1 Trainingsdatensätze

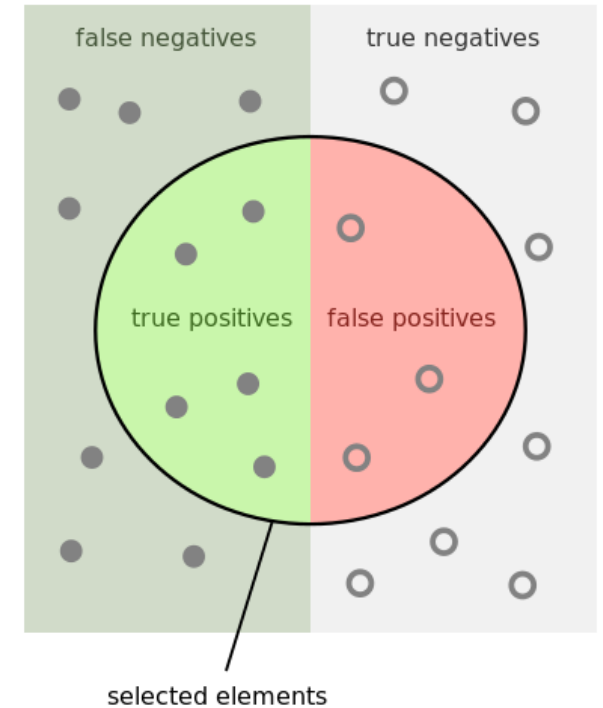
CoNNL2003	GENIA
<p>Germany I-LOC ... representative ... to ... the ... European I-ORG ... Union I-ORG ... ,s ... veterinary ... committee ... Werner I-PER ... Zwingmann I-PER ... said ... on ... Wednesday ...</p>	<p>Inhibition Inhibition NN B-NP O of of IN B-PP O NF-kappaB NF-kappaB NN B-NP B-protein activation activation NN I-NP O reversed reverse VBD B-VP O the the DT B-NP O anti-apoptotic anti-apoptotic JJ I-NP O effect effect NN I-NP O of of IN B-PP O isochamaejasmin isochamaejasmin NN B-NP O . . . O O</p>

4.2 NER-Style F1

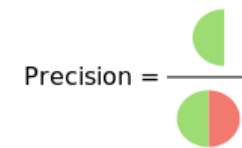
$$Prec = \frac{|Richtig\ selektierte\ Matches|}{|Alle\ selektieren\ Matches|} = \frac{\sum_{d \in D} tp_d}{\sum_{d \in D} (tp_d + fp_d)}$$

$$Rec = \frac{|Richtig\ selektierte\ Matches|}{|Richtige\ Matches|} = \frac{\sum_{d \in D} tp_d}{\sum_{d \in D} (tp_d + fn_d)}$$

$$F1 = \textit{Harmonisches Mittel} = \frac{2 * Prec * Rec}{Prec * Rec}$$

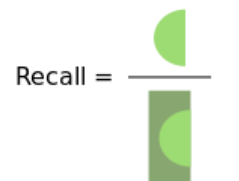


How many selected items are relevant?



Precision = $\frac{\text{green}}{\text{green} + \text{red}}$

How many relevant items are selected?



Recall = $\frac{\text{green}}{\text{green} + \text{light gray}}$

Grafische Darstellung von Precision und Recall

https://en.wikipedia.org/wiki/Precision_and_recall

4.2 NER-Style F1

= Schwaches Match reicht

[Arnold et al. 2016]

Richtig	BI00000BIIIO0B0B00000B
Vorhersage	OB00BI00B000000000B0000

4.2 NER-Style F1

= Schwaches Match reicht

[Arnold et al. 2016]

Richtig	BI00000BIII00B0B00000B
Vorhersage	OB00BI00B000000000B0000

4.2 NER-Style F1

= Schwaches Match reicht

[Arnold et al. 2016]

Richtig **BI00000BIIIOOB0B00000B**

Vorhersage **OB00BI00B0000000000B00000**

Bsp. Precision: $\frac{2}{2+2} = \frac{2}{4} = 0.5$

4.3 Vergleich aktueller Systeme

Common Test Sets		CoNLL2003			KORE50			ACE2004			MSNBC		
corpus		RCV-1			RCV-1			newswire			MSNBC news		
topic		news (en)			news (en)			news (en)			news (en)		
annotation guideline		named entities			named entities			all mentions			wikification		
Annotator	Method	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
Babelify	POS+DICT	53.8	70.4	61.0	72.1	73.6	72.9	12.1	42.2	18.8	43.3	77.8	55.6
DBpedia Spotlight	DICT	74.7	66.4	70.3	n/a	n/a	n/a	13.0	74.8	22.2	56.2	49.0	52.4
Entityclassifier.eu	Noun phrase	81.2	83.0	82.1	93.8	94.4	94.1	13.3	90.5	23.2	76.2	93.9	84.1
FOX	Ensemble	99.1	75.2	85.5	94.6	73.6	82.8	12.4	59.5	20.6	38.3	31.3	34.4
LingPipe MUC-7	LM+HMM	91.5	66.6	77.1	93.9	86.1	89.9	16.3	88.9	27.5	73.3	78.7	75.9
NERD-ML	Ensemble	59.9	72.0	65.4	70.4	82.6	76.0	19.6	47.4	27.7	69.7	57.2	62.8
Stanford NER	CRF+Dist	99.5	76.1	86.2	94.8	76.4	84.6	18.2	90.9	30.3	95.2	84.1	89.3
TagMe 2	DICT	68.3	47.7	56.2	66.5	88.2	75.8	23.2	71.6	35.0	55.6	38.0	45.2
DATEXIS-NER	BLSTM	87.8	94.6	91.0	92.6	95.1	93.8	13.3	98.0	23.4	73.3	98.2	83.9

Vergleich von 8 Systemen mit gleicher Konfiguration, die auf 2.000 annotierten Sätzen trainiert wurden. Als Datensätze dienen vier typische Nachrichten-Datensätze

[Arnold et al. 2016]

4.3 Vergleich aktueller Systeme

Common Test Sets		CoNLL2003			KORE50			ACE2004			MSNBC		
corpus		RCV-1			RCV-1			newswire			MSNBC news		
topic		news (en)			news (en)			news (en)			news (en)		
annotation guideline		named entities			named entities			all mentions			wikification		
Annotator	Method	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
Babelify	POS+DICT	53.8	70.4	61.0	72.1	73.6	72.9	12.1	42.2	18.8	43.3	77.8	55.6
DBpedia Spotlight	DICT	74.7	66.4	70.3	n/a	n/a	n/a	13.0	74.8	22.2	56.2	49.0	52.4
Entityclassifier.eu	Noun phrase	81.2	83.0	82.1	93.8	94.4	94.1	13.3	90.5	23.2	76.2	93.9	84.1
FOX	Ensemble	99.1	75.2	85.5	94.6	73.6	82.8	12.4	59.5	20.6	38.3	31.3	34.4
LingPipe MUC-7	LM+HMM	91.5	66.6	77.1	93.9	86.1	89.9	16.3	88.9	27.5	73.3	78.7	75.9
NERD-ML	Ensemble	59.9	72.0	65.4	70.4	82.6	76.0	19.6	47.4	27.7	69.7	57.2	62.8
Stanford NER	CRF+Dist	99.5	76.1	86.2	94.8	76.4	84.6	18.2	90.9	30.3	95.2	84.1	89.3
TagMe 2	DICT	68.3	47.7	56.2	66.5	88.2	75.8	23.2	71.6	35.0	55.6	38.0	45.2
DATEXIS-NER	BLSTM	87.8	94.6	91.0	92.6	95.1	93.8	13.3	98.0	23.4	73.3	98.2	83.9

Vergleich von 8 Systemen mit gleicher Konfiguration, die auf 2.000 annotierten Sätzen trainiert wurden. Als Datensätze dienen vier typische Nachrichten-Datsets

[Arnold et al. 2016]

4.3 Vergleich aktueller Systeme

Common Test Sets		CoNLL2003			KORE50			ACE2004			MSNBC		
corpus		RCV-1			RCV-1			newswire			MSNBC news		
topic		news (en)			news (en)			news (en)			news (en)		
annotation guideline		named entities			named entities			all mentions			wikification		
Annotator	Method	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
Babelify	POS+DICT	53.8	70.4	61.0	72.1	73.6	72.9	12.1	42.2	18.8	43.3	77.8	55.6
DBpedia Spotlight	DICT	74.7	66.4	70.3	n/a	n/a	n/a	13.0	74.8	22.2	56.2	49.0	52.4
Entityclassifier.eu	Noun phrase	81.2	83.0	82.1	93.8	94.4	94.1	13.3	90.5	23.2	76.2	93.9	84.1
FOX	Ensemble	99.1	75.2	85.5	94.6	73.6	82.8	12.4	59.5	20.6	38.3	31.3	34.4
LingPipe MUC-7	LM+HMM	91.5	66.6	77.1	93.9	86.1	89.9	16.3	88.9	27.5	73.3	78.7	75.9
NERD-ML	Ensemble	59.9	72.0	65.4	70.4	82.6	76.0	19.6	47.4	27.7	69.7	57.2	62.8
Stanford NER	CRF+Dist	99.5	76.1	86.2	94.8	76.4	84.6	18.2	90.9	30.3	95.2	84.1	89.3
TagMe 2	DICT	68.3	47.7	56.2	66.5	88.2	75.8	23.2	71.6	35.0	55.6	38.0	45.2
DATEXIS-NER	BLSTM	87.8	94.6	91.0	92.6	95.1	93.8	13.3	98.0	23.4	73.3	98.2	83.9

Vergleich von 8 Systemen mit gleicher Konfiguration, die auf 2.000 annotierten Sätzen trainiert wurden. Als Datensätze dienen vier typische Nachrichten-Datsets

[Arnold et al. 2016]

4.3 Vergleich aktueller Systeme

Common Test Sets		CoNLL2003			KORE50			ACE2004			MSNBC		
corpus		RCV-1			RCV-1			newswire			MSNBC news		
topic		news (en)			news (en)			news (en)			news (en)		
annotation guideline		named entities			named entities			all mentions			wikification		
Annotator	Method	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
Babelify	POS+DICT	53.8	70.4	61.0	72.1	73.6	72.9	12.1	42.2	18.8	43.3	77.8	55.6
DBpedia Spotlight	DICT	74.7	66.4	70.3	n/a	n/a	n/a	13.0	74.8	22.2	56.2	49.0	52.4
Entityclassifier.eu	Noun phrase	81.2	83.0	82.1	93.8	94.4	94.1	13.3	90.5	23.2	76.2	93.9	84.1
FOX	Ensemble	99.1	75.2	85.5	94.6	73.6	82.8	12.4	59.5	20.6	38.3	31.3	34.4
LingPipe MUC-7	LM+HMM	91.5	66.6	77.1	93.9	86.1	89.9	16.3	88.9	27.5	73.3	78.7	75.9
NERD-ML	Ensemble	59.9	72.0	65.4	70.4	82.6	76.0	19.6	47.4	27.7	69.7	57.2	62.8
Stanford NER	CRF+Dist	99.5	76.1	86.2	94.8	76.4	84.6	18.2	90.9	30.3	95.2	84.1	89.3
TagMe 2	DICT	68.3	47.7	56.2	66.5	88.2	75.8	23.2	71.6	35.0	55.6	38.0	45.2
DATEXIS-NER	BLSTM	87.8	94.6	91.0	92.6	95.1	93.8	13.3	98.0	23.4	73.3	98.2	83.9

Vergleich von 8 Systemen mit gleicher Konfiguration, die auf 2.000 annotierten Sätzen trainiert wurden. Als Datensätze dienen vier typische Nachrichten-Datsets

[Arnold et al. 2016]

4.3 Vergleich aktueller Systeme

Common Test Sets		CoNLL2003			KORE50			ACE2004			MSNBC		
corpus		RCV-1			RCV-1			newswire			MSNBC news		
topic		news (en)			news (en)			news (en)			news (en)		
annotation guideline		named entities			named entities			all mentions			wikification		
Annotator	Method	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
Babelify	POS+DICT	53.8	70.4	61.0	72.1	73.6	72.9	12.1	42.2	18.8	43.3	77.8	55.6
DBpedia Spotlight	DICT	74.7	66.4	70.3	n/a	n/a	n/a	13.0	74.8	22.2	56.2	49.0	52.4
Entityclassifier.eu	Noun phrase	81.2	83.0	82.1	93.8	94.4	94.1	13.3	90.5	23.2	76.2	93.9	84.1
FOX	Ensemble	99.1	75.2	85.5	94.6	73.6	82.8	12.4	59.5	20.6	38.3	31.3	34.4
LingPipe MUC-7	LM+HMM	91.5	66.6	77.1	93.9	86.1	89.9	16.3	88.9	27.5	73.3	78.7	75.9
NERD-ML	Ensemble	59.9	72.0	65.4	70.4	82.6	76.0	19.6	47.4	27.7	69.7	57.2	62.8
Stanford NER	CRF+Dist	99.5	76.1	86.2	94.8	76.4	84.6	18.2	90.9	30.3	95.2	84.1	89.3
TagMe 2	DICT	68.3	47.7	56.2	66.5	88.2	75.8	23.2	71.6	35.0	55.6	38.0	45.2
DATEXIS-NER	BLSTM	87.8	94.6	91.0	92.6	95.1	93.8	13.3	98.0	23.4	73.3	98.2	83.9

Vergleich von 8 Systemen mit gleicher Konfiguration, die auf 2.000 annotierten Sätzen trainiert wurden. Als Datensätze dienen vier typische Nachrichten-Datsets

[Arnold et al. 2016]

4.3 Vergleich aktueller Systeme

Common Test Sets		CoNLL2003			KORE50			ACE2004			MSNBC		
corpus		RCV-1			RCV-1			newswire			MSNBC news		
topic		news (en)			news (en)			news (en)			news (en)		
annotation guideline		named entities			named entities			all mentions			wikification		
Annotator	Method	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
Babelify	POS+DICT	53.8	70.4	61.0	72.1	73.6	72.9	12.1	42.2	18.8	43.3	77.8	55.6
DBpedia Spotlight	DICT	74.7	66.4	70.3	n/a	n/a	n/a	13.0	74.8	22.2	56.2	49.0	52.4
Entityclassifier.eu	Noun phrase	81.2	83.0	82.1	93.8	94.4	94.1	13.3	90.5	23.2	76.2	93.9	84.1
FOX	Ensemble	99.1	75.2	85.5	94.6	73.6	82.8	12.4	59.5	20.6	38.3	31.3	34.4
LingPipe MUC-7	LM+HMM	91.5	66.6	77.1	93.9	86.1	89.9	16.3	88.9	27.5	73.3	78.7	75.9
NERD-ML	Ensemble	59.9	72.0	65.4	70.4	82.6	76.0	19.6	47.4	27.7	69.7	57.2	62.8
Stanford NER	CRF+Dist	99.5	76.1	86.2	94.8	76.4	84.6	18.2	90.9	30.3	95.2	84.1	89.3
TagMe 2	DICT	68.3	47.7	56.2	66.5	88.2	75.8	23.2	71.6	35.0	55.6	38.0	45.2
DATEXIS-NER	BLSTM	87.8	94.6	91.0	92.6	95.1	93.8	13.3	98.0	23.4	73.3	98.2	83.9

Vergleich von 8 Systemen mit gleicher Konfiguration, die auf 2.000 annotierten Sätzen trainiert wurden. Als Datensätze dienen vier typische Nachrichten-Datensets

[Arnold et al. 2016]

4.4 Pro und Contra: Datexis

Pro






- **Hohe F1-Score** in verschiedenen Domänen
- **Resistent** gegenüber **unbekannten Wörtern** und Rechtschreibfehlern
- **Wenig Trainingsdaten** benötigt










Contra

- BIO-Labeling: **keine exakten Klassifizierungen** / Entity Linking

[Arnold et al. 2016]

Quellen

-  [Arnorld et al. 2016] Robust Names Entity Recognition in Idiosyncratic Domains
-  [Bengio et al. 2002] A Neural Probabilistic Language Model
-  [Collins2002] Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms
-  [Huang et al. 2013] Learning Deep Structural Semantic Models for Web Search using Clickthrough Data.
-  [McCallum und Li 2003] Early Results for Named Entity Recognition with Conditional Random Fields, Feature Induction and Web-Enhanced Lexicons

-  <http://2011.bionlp-st.org/home/genia-event-extraction-genia>
-  <https://deeplearning4j.org/lstm.html#vanishing>
-  <https://dialogflow.com/docs/dialogs>
-  https://en.wikipedia.org/w/index.php?title=Entity_linking&oldid=818722645
-  https://en.wikipedia.org/w/index.php?title=Named-entity_recognition&oldid=813025696
-  <https://en.wikipedia.org/w/index.php?title=One-hot&oldid=808377328>
-  <https://github.com/synalp/NER/tree/master/corpus/CoNLL-2003>
-  <https://machinelearningmastery.com/develop-bidirectional-lstm-sequence-classification-python-keras/>
-  <http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/> 03