



# Maschinelle Sprachverarbeitung

## Übung

### Aufgabe 6: Evaluationsergebnisse

Mario Sanger

[mario.saenger@informatik.hu-berlin.de](mailto:mario.saenger@informatik.hu-berlin.de)

# Übung 6

---

- Implementation eines Verfahrens zur Erkennung von Genen in Texten
  - Modellierung als Sequenz-Labeling-Problem
  - Einsatz von Maschinellen Lernen
- Implementierung eines Conditional Random Fields (CRF)
  - Vielgestaltige Features: Wort- und Worteigenschaften, Dictionary-Matching, Kontextwörter, ....
- Wettbewerb: Die Lösung mit dem höchsten F1-Score gewinnt!

# Evaluationsverfahren

---

- Verwendung des BioCreative II Gene Mention Korpus
  - Veranstaltung von Shared Tasks im biomedizinischen Text Mining
  - 20 500 Sätze mit 20 703 Gen-Annotationen
- Evaluation mittels F1-Score
  - True positive nur bei exakter Übereinstimmung

	...	a	dosis	of	rabies	immunoglobulin	(	...	
Gold		0	0	0	B-NP	I-NP		0	
Pred		0	0	0	B-NP	I-NP		0	✓

# Evaluationsverfahren

---

- Verwendung des BioCreative II Gene Mention Korpus
  - Veranstaltung von Shared Tasks im biomedizinischen Text Mining
  - 20 500 Sätze mit 20 703 Gen-Annotationen
- Evaluation mittels F1-Score
  - True positive nur bei exakter Übereinstimmung

	...	a	dosis	of	rabies	immunoglobulin	(	...
Gold		O	O	O	B-NP	I-NP	O	
Pred		O	O	O	B-NP	<b>O</b>	O	



**Doppelte Bestrafung: 1 False Positive und 1 False Negative**

# Testergebnisse (Exakt)

---

<b>Gruppe</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>Rang</b>
Batorurowaiaru	0.736	0.670	0.701	2
DieBesteGruppe	0.721	0.563	0.633	5
LidlEffort	0.729	0.602	0.659	3
NotWorkingAsIntended	0.727	0.602	0.659	4
Pi	0.749	0.680	0.713	1

# Testergebnisse (Partiell)

---

<b>Gruppe</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>Rang</b>
Batorurowaiaru	0.883	0.821	0.851	2
DieBesteGruppe	0.876	0.702	0.779	5
LidlEffort	0.886	0.749	0.812	3
NotWorkingAsIntended	0.884	0.749	0.811	4
Pi	0.896	0.832	0.862	1

# Testergebnisse - Wettbewerbspunkte

---

<b>Gruppe</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>Rang</b>	<b>Punkte</b>
Batorurowaiaru	0.736	0.670	0.701	2	3
DieBesteGruppe	0.721	0.563	0.633	5	
LidlEffort	0.729	0.602	0.659	3	1
NotWorkingAsIntended	0.727	0.602	0.659	4	
Pi	0.749	0.680	0.713	1	5

# Testergebnisse – Weitere Korpora

Gruppe	BC2GM		GPRO		BioInfer		JNLPBA	
	F1	Rang	F1	Rang	F1	Rang	F1	Rang
Batorurowaiaru	0.701	2	0.634	2	0.820	2	0.759	1
DieBesteGruppe	0.633	5	0.560	5	0.777	5	0.733	5
LidlEffort	0.659	3	0.592	3	0.797	3	0.740	3
NotWorkingAI	0.659	4	0.591	4	0.797	3	0.739	4
Pi	0.713	1	0.650	1	0.842	1	0.755	2



# Testergebnisse – Cross-Corpus

---

<b>Gruppe</b>	<b>GPRO</b>		<b>BioInfer</b>		<b>JNLPBA</b>	
	<b>F1</b>	<b>Diff.</b>	<b>F1</b>	<b>Diff.</b>	<b>F1</b>	<b>Diff.</b>
Batorurowaiaru	0.397	-0.237	0.633	-0.186	0.621	-0.139
DieBesteGruppe	0.369	-0.191	0.621	-0.156	0.589	-0.143
LidlEffort	0.378	-0.214	0.639	-0.158	0.598	-0.142
NotWorkingAsIntended	0.382	-0.209	0.641	-0.156	0.599	-0.141
Pi	0.449	-0.201	0.682	-0.160	0.621	-0.135

# Testergebnisse – Cross-Corpus

---

<b>Gruppe</b>	<b>GPRO</b>		<b>BioInfer</b>		<b>JNLPBA</b>	
	<b>F1</b>	<b>Diff.</b>	<b>F1</b>	<b>Diff.</b>	<b>F1</b>	<b>Diff.</b>
Batorurowaiaru	0.397	62.6%	0.633	77.3%	0.621	81.8%
DieBesteGruppe	0.369	65.9%	0.621	79.9%	0.589	80.5%
LidlEffort	0.378	63.9%	0.639	80.1%	0.598	80.8%
NotWorkingAsIntended	0.382	64.6%	0.641	80.5%	0.599	81.0%
Pi	0.449	69.1%	0.682	81.0%	0.621	82.2%