

LitSOM. Kartierung russischer Gegenwartsliteratur

- 1 Einführung
- 2 Methode
- 3 Resultate
 - Klassifizierung
 - Kartierung
- 4 Resümee

1 Einführung

2 Methode

3 Resultate

- Klassifizierung
- Kartierung

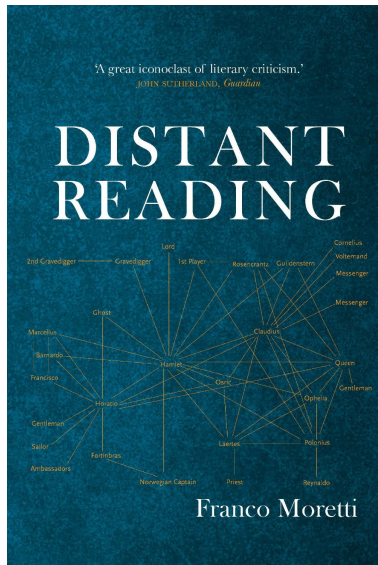
4 Resümee

Das Problem

120 Romane von acht zeitgenössischen russischen Autoren

- ▶ 8.794.483 Wörter
- ▶ 109 Megabytes
- ▶ ~ 114.295 Seiten
- ▶ entspricht einem Papierstapel, der 5.7 Meter hoch ist
- ▶ 586 Stunden Lesevergnügen (= 24 Tage)

Die Lösung



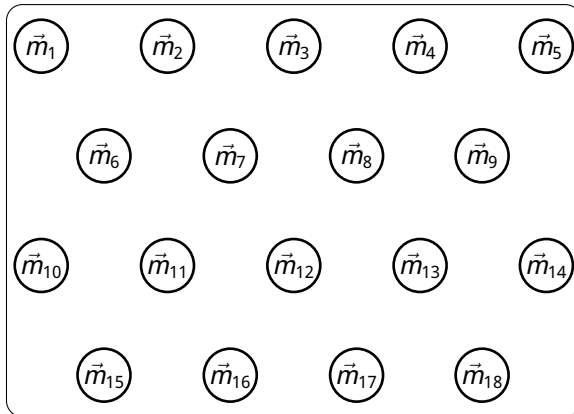


A. A. Markov (1856 – 1922)

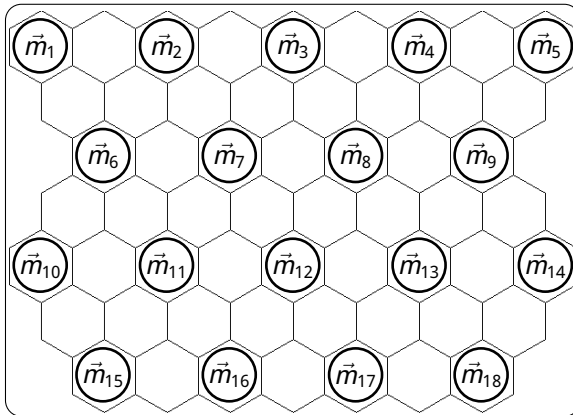
Пример статистического
исследования над текстом
Евгения Онегина,
иллюстрирующий связь
испытаний в цепи (1913)

- 1 Einführung
- 2 Methode**
- 3 Resultate
 - Klassifizierung
 - Kartierung
- 4 Resümee

Self-Organizing Maps (SOM)

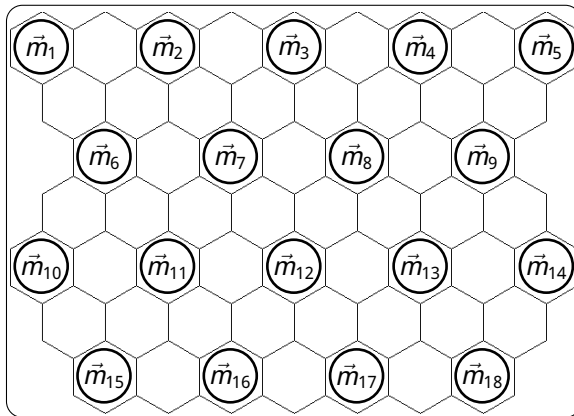


Self-Organizing Maps (SOM)



Self-Organizing Maps (SOM)

$\vec{X} \xrightarrow{?}$



Warum SOM?

- ▶ Einfach zu erklären, einfach zu realisieren
- ▶ Seit Jahrzehnten in Verwendung
- ▶ Erlaubt überwacht und unüberwacht Lernen
- ▶ Bringt seine Visualisierung gleich mit

Kodierung der Romane

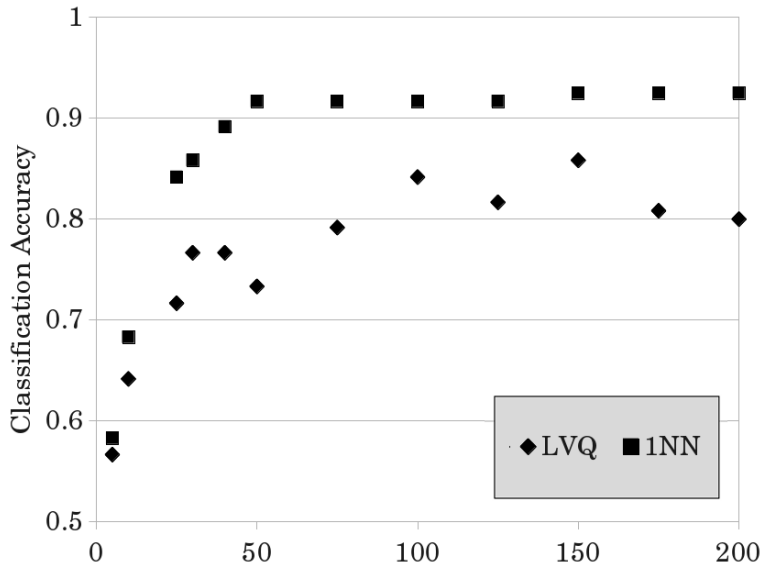
	Wort	Sharov	Očered'	Omon Ra	Kazus K.
1	i	3.8%	1.2%	3.8%	4.0%
2	v	2.9%	0.8%	2.7%	2.8%
3	ne	2.2%	2.1%	1.5%	2.1%
4	na	1.8%	0.7%	2.1%	1.5%
5	čto	1.3%	1.0%	1.4%	1.1%
6	s	1.2%	0.4%	1.2%	1.3%
7	to	0.8%	0.2%	0.2%	0.3%
8	ja	0.8%	3.2%	2.4%	0.5%
9	on	0.8%	0.2%	0.8%	0.8%
10	kak	0.7%	0.6%	0.1%	0.7%
:	:	:	:	:	:
150	čeloveka	0.05%	0.07%	0.02%	0.04%

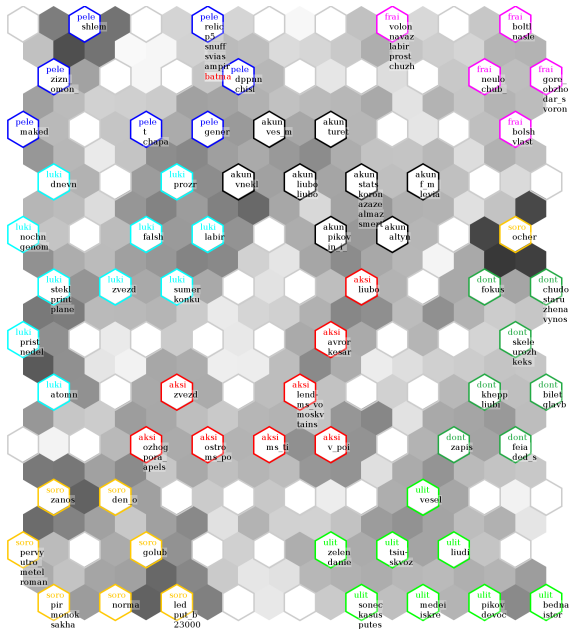
- 1 Einführung
- 2 Methode
- 3 Resultate**
 - Klassifizierung
 - Kartierung
- 4 Resümee

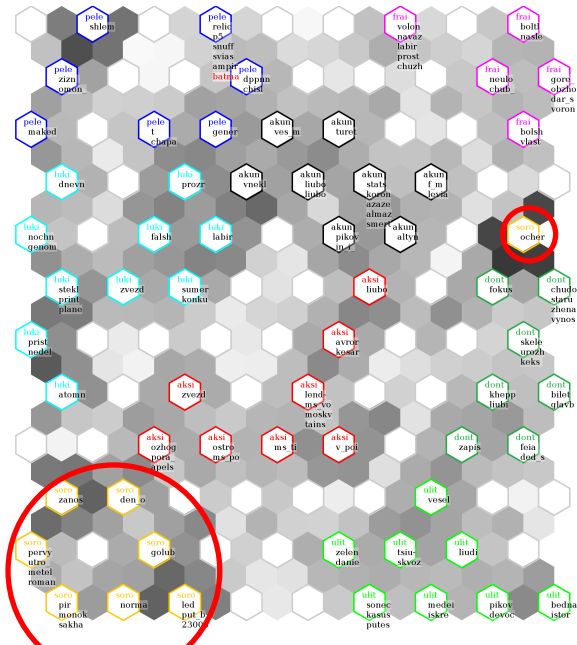
Beste Ergebnisse

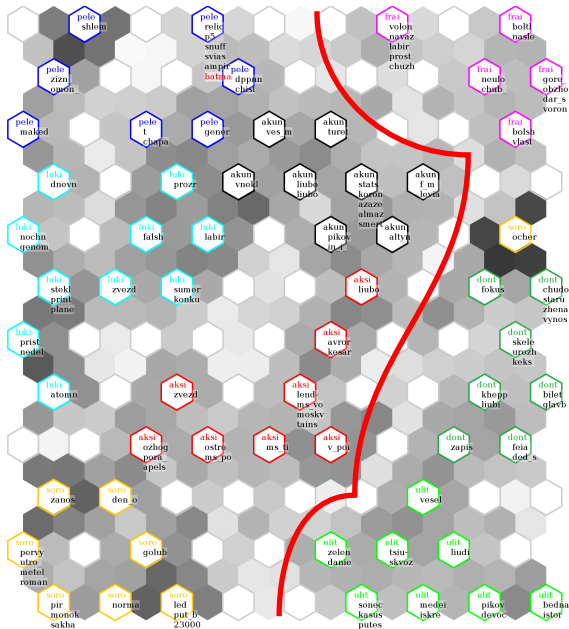
	Features	# Korrekt	SOM	1NN
1	150 Verben	103	85.83%	92.50%
2	100 Nomen	102	85.00%	95.83%
3	100 Verben	101	84.16%	91.67%
4	200 Sharov	100	83.33%	90.00%
5	175 Nomen	99	82.50%	95.00%
6	100 Sharov	98	81.67%	90.00%
„	150 Sharov	98	81.67%	90.83%
„	125 Nomen	98	81.67%	93.34%
„	125 Verben	98	81.67%	91.67%
10	75 Nomen	97	80.83%	94.16%
„	40 Nomen	97	80.83%	89.16%
„	175 Verben	97	80.83%	92.50%

Klassifizierung (mit Verben)



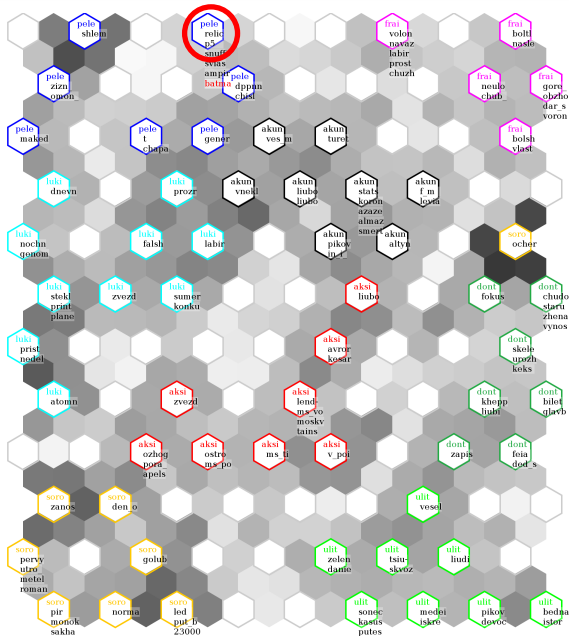






Batman Apollo (2013)

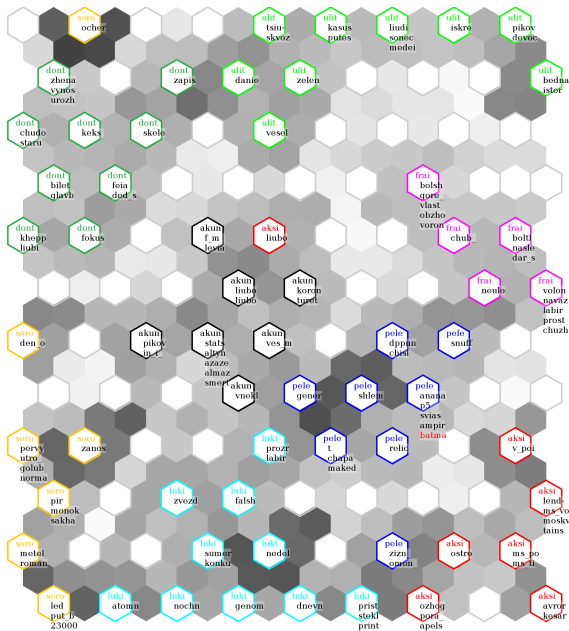




Batman Apollo (2013)

“

In diesem Buch agieren die Helden des Romans ‚Empire V‘, aber man muss diesen nicht unbedingt vorher gelesen haben. Sie können hier beginnen, und ‚Empire V‘ später lesen (Pelevin 2013:1, Übers. G.H.)





- 1 Einführung
- 2 Methode
- 3 Resultate
 - Klassifizierung
 - Kartierung
- 4 Resümee**

Resümee

- ▶ Klassifizierung funktioniert sehr gut
- ▶ Karten zeigen Beziehungen zwischen Texten auf
- ▶ Karten weisen auf Besonderheiten hin
- ▶ „Distant Reading“ plus „Close Reading“!

ToDoS

- ▶ Programm zur Verfügung stellen
(Open Source oder als Web App)
- ▶ Größere Anzahl an Romanen kartographieren...
- ▶ ... oder auf eine/n Autor/in konzentrieren

Quellen

- ▶ Howanitz, G. 2013. *LitSOM. Automatic Mapping of Russian Authors Using SOM*. Salzburg. [Masterarbeit]
- ▶ Kohonen, T. 2001. *Self-Organizing Maps*. Berlin.
- ▶ Moretti, F. 2000. "Conjectures on World Literature". *New Left Review* 1, 54-68.
- ▶ Pelevin, V. 2013. *Batman Apollo*. Moscow.
- ▶ Schreibman, S. et al. (Hg.) 2004. *A Companion to Digital Humanities*. Oxford.

web-texten.org/dhd2014



Bonus

Confusion Matrix für 150 Verben

		Klassifiziert							
		Aks	Aku	Don	Fra	Luk	Pel	Sor	Uli
Tatsächlich	Aks	75	8				8		8
	Aku	17	83						
	Don			100					
	Fra		8		92				
	Luk	17				83			
	Pel	8		8			67		17
	Sor	17				17		67	
	Uli							8	92