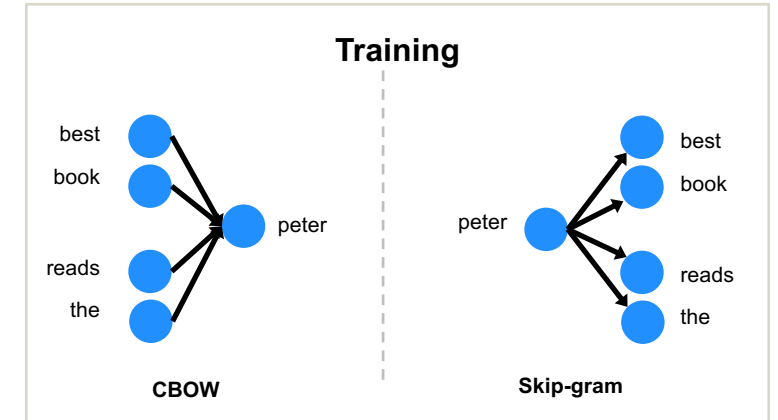
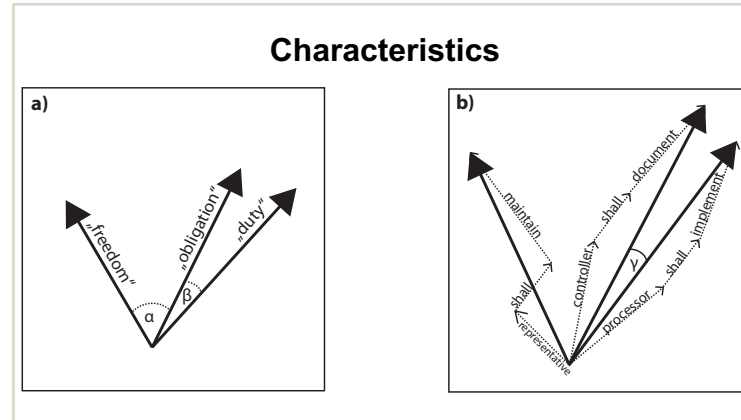
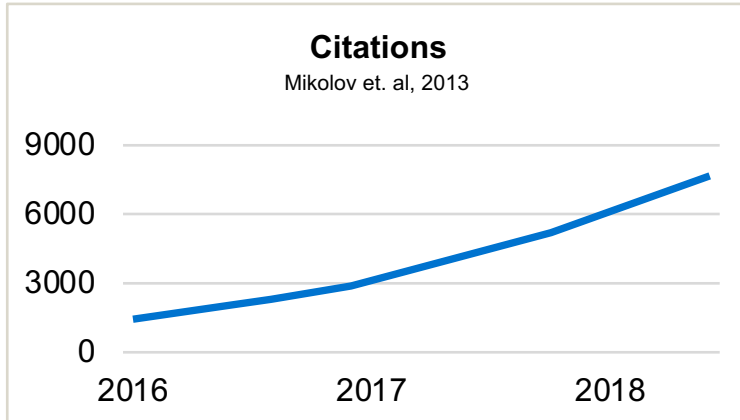


Semantic Text Matching (STM): Context-based recommendation of legal comment chapters to tenancy law contract clauses

Jörg Landthaler, 20. September 2018, München

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
www.matthes.in.tum.de

Word Embeddings: Brief Review



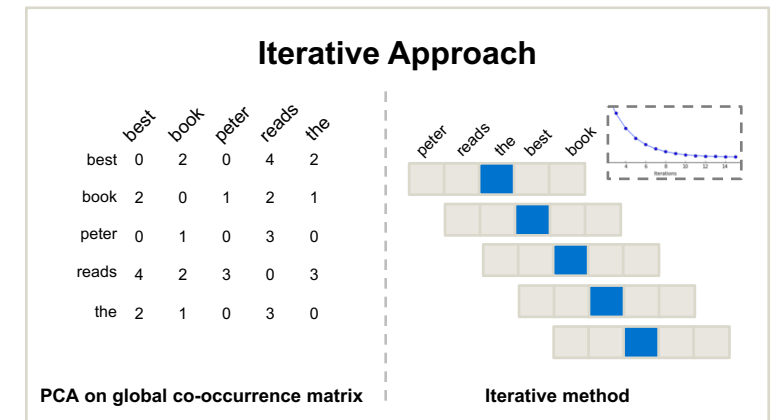
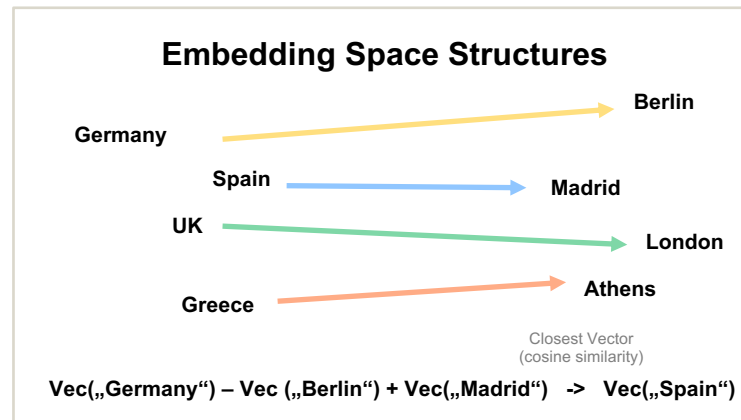
Word Representations

	Peter	reads	the	best	book
best	0	0	0	1	0
book	0	0	0	0	1
peter	1	0	0	0	0
reads	0	1	0	0	0
the	0	0	1	0	0

Term frequency

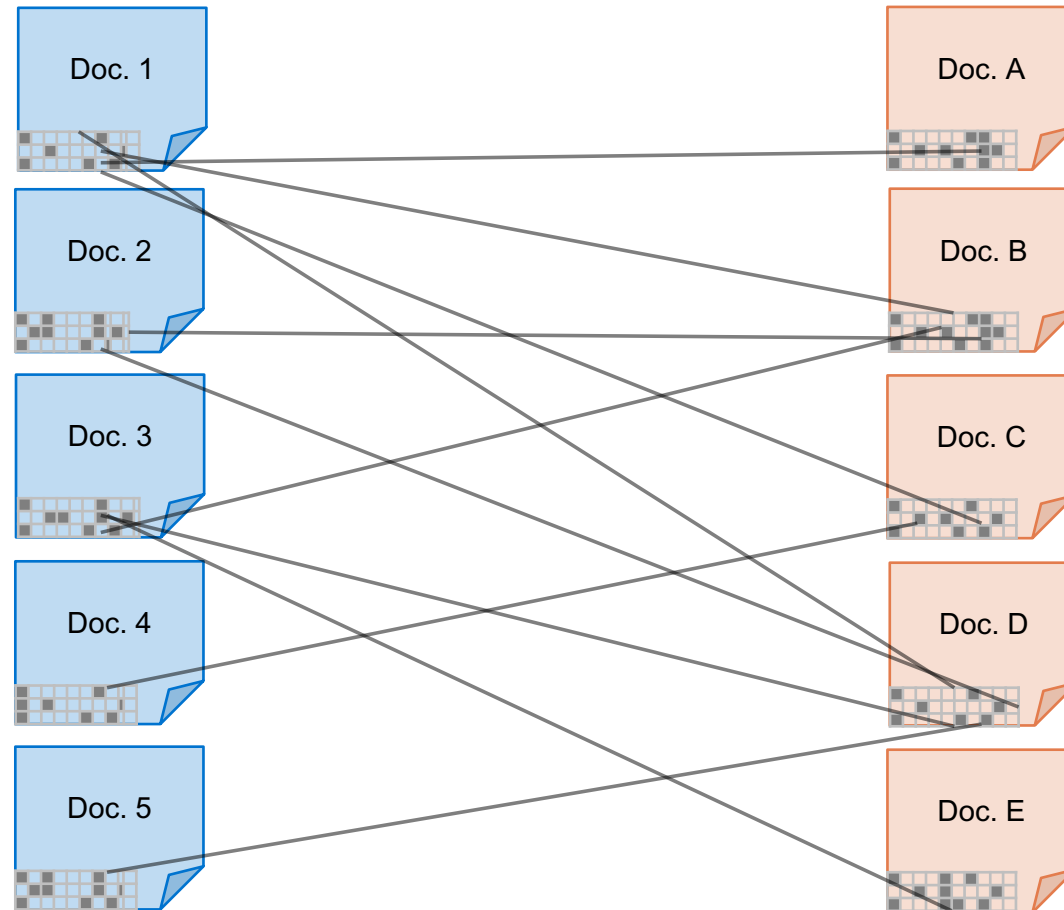
	Peter	reads	the	best	book
Peter	1.20	2.33	1.01	8.34	3.22
reads	0.51	-1.4	0.01	0.23	5.67
the	3.22	1.23	2.44	2.34	0.01
best	-2.0	6.23	3.84	1.04	-1.2
book	-1.2	1.01	-2.8	4.55	2.66

Word Embeddings



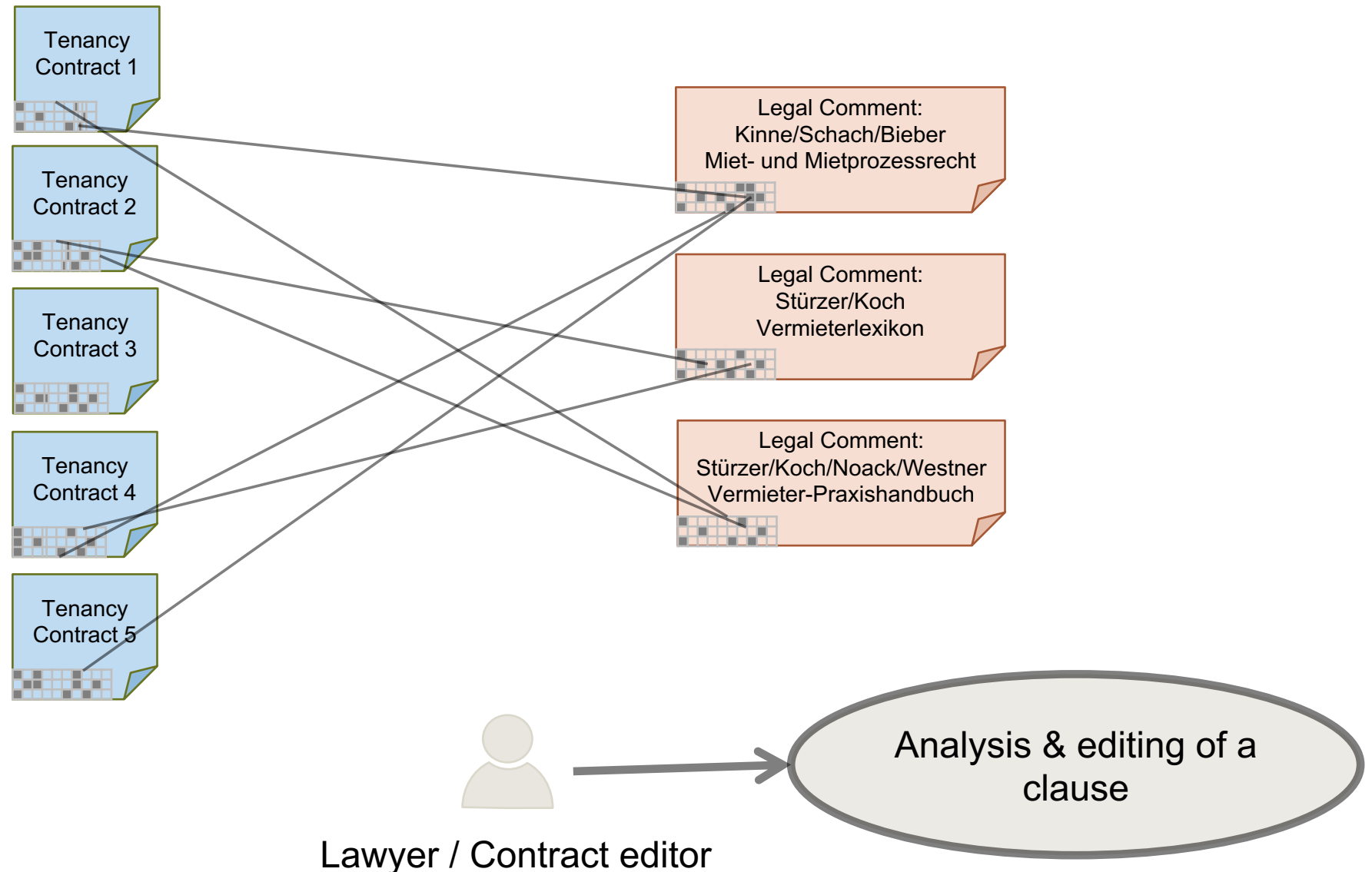
Mikolov et. al, *Efficient Estimation of Word Representations in Vector Space*, ICLR Workshop 2013,
 Mikolov et al., *Distributed representations of words and phrases and their compositionality*, Proceedings of the 26th International Conference on Neural Information Processing Systems, NIPS 2013
 Le & Mikolov, *Distributed Representations of Sentences and Documents*, Proceedings of the 31st International Conference on International Conference on Machine Learning, ICML'14
 Levy & Goldberg, *Neural word embedding as implicit matrix factorization*, Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS'12014

The Semantic Text Matching (STM) Problem



The identification of semantically or logically related text-fragments of different documents (from different document types). Example: Contract clauses have to satisfy provisions.

Semantic Text Matching (STM) in Tenancy Law: Contracts & Legal Comments



Three Perspectives on Semantic Text Matching (STM)

Implicit references view



Analogy to explicit references
/ citation networks

Graph view



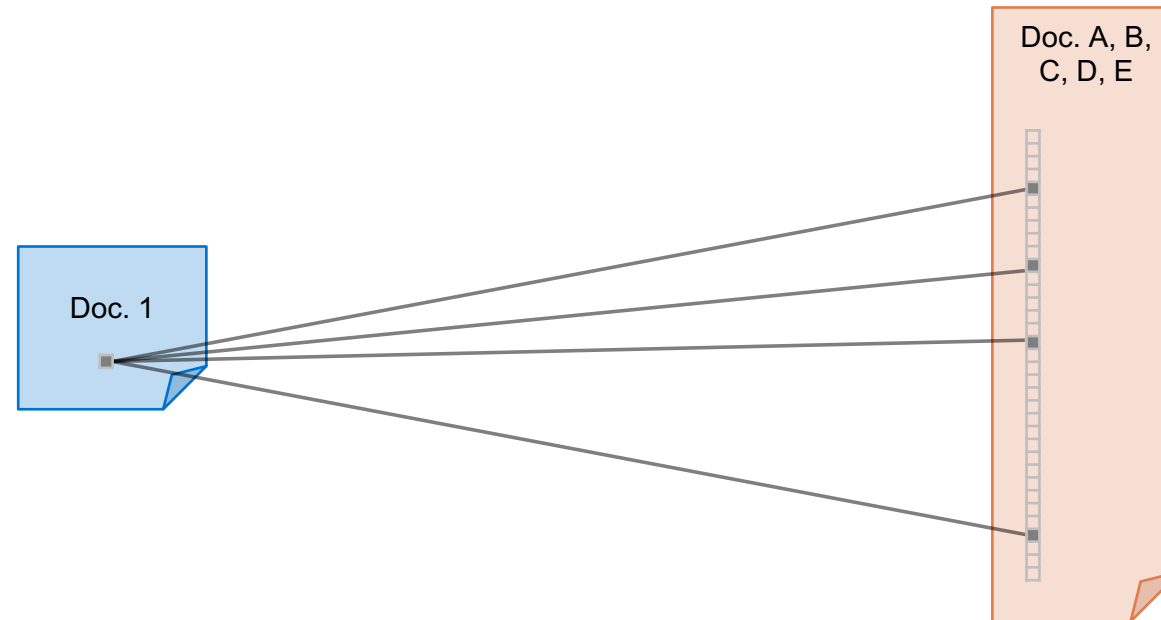
For two document types:
bipartite graph
(matching characteristics)

Search algorithm view



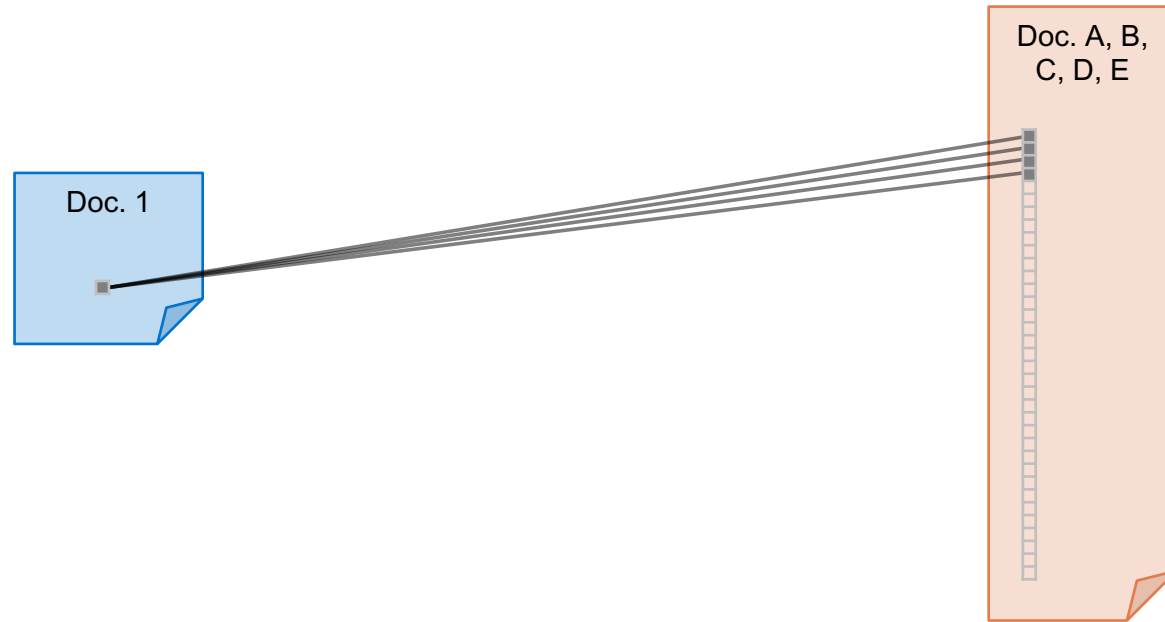
A reduction/simplification of
general search

General STM Solution Approach: Text Similarity Measures⁽¹⁾



- 1) Segmentation of the documents into text fragments, e.g. using a heuristic (e.g. newlines).
- 2) Pool all text-fragments of one document type.

General STM Solution Approach: Text Similarity Measures (2)



3) Application of a text similarity measure to rank all text-fragments of the pool with respect to one text fragment of documents 1, ..., N

If required repeat (sequentially or matrix-matrix operation) for all text-fragments of document 1, ..., N

Korpus: Stürzer/Koch/...: Vermieter-Praxishandbuch


Methode: Sentence Embeddings


Normalize


Stopsremoval

koelner-miete.de: Mietvertrag - Wohnraum


§1 Mietvertragsparteien

 Zwischen: - nachstehend "Vermieter" genannt - vertreten durch und: ausgewiesen durch Personalausweis Reisepass Nr. ausgestellt am durch ausgewiesen durch Personalausweis Reisepass Nr. ausgestellt am durch - nachstehend "Mieter" genannt vertreten durch wird folgender Mietvertrag geschlossen.








 Mehrere Mieter haften dem Vermieter gegenüber als Gesamtschuldner für alle Verpflichtungen aus dem Mietverhältnis in voller Höhe. Sie bevollmächtigen sich durch Unterzeichnung dieses Vertrages gegenseitig in dem unter § 25 dargelegten Umfang.

 Bestandteil dieses Vertrages ist die Hausordnung unter Anhang A Mietgegenstand

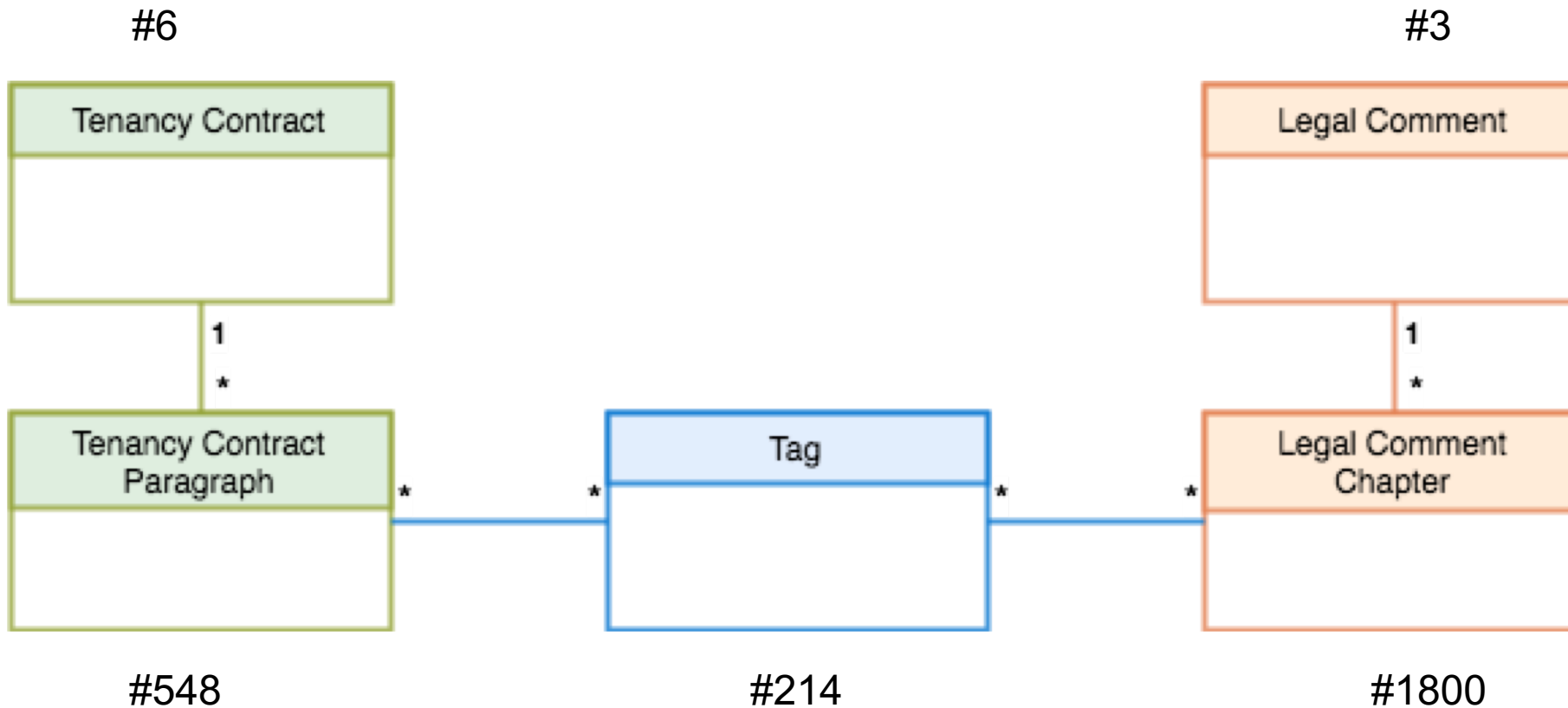
§2 Mietgegenstand

 Der Vermieter vermietet dem Mieter ausschließlich zu Wohnzwecken die im Hause im -geschoss (rechts/links/mitte) gelegene Wohnung, bestehend aus ___ Zimmern, ___ Küche, ___ Diele, ___ Flur, ___ Bad/Dusche, ___ WC, , ___ Balkon

Vorgeschlagene Literatur (11)

- (1) 2.1 Wohnraummietvertrag  Stürzer/Koch/...: Vermieter-Praxishandbuch
- (2) 2.2.1 Die Parteien des Mietvertrags  Stürzer/Koch/...: Vermieter-Praxishandbuch
- (3) 2.3 Garagenmietvertrag  Stürzer/Koch/...: Vermieter-Praxishandbuch
- (4) 11.7 Mietaufhebungsvereinbarung  Stürzer/Koch/...: Vermieter-Praxishandbuch
- (5) 2.1.1 Die Parteien des Mietvertrags  Stürzer/Koch/...: Vermieter-Praxishandbuch
- (6) 5.1 Mieterhöhung bei Wohnraum  Stürzer/Koch/...: Vermieter-Praxishandbuch
- (7) 1.1 Die Suche nach dem richtigen Mieter  Stürzer/Koch/...: Vermieter-Praxishandbuch

Creating a Dataset for a Quantitative Evaluation



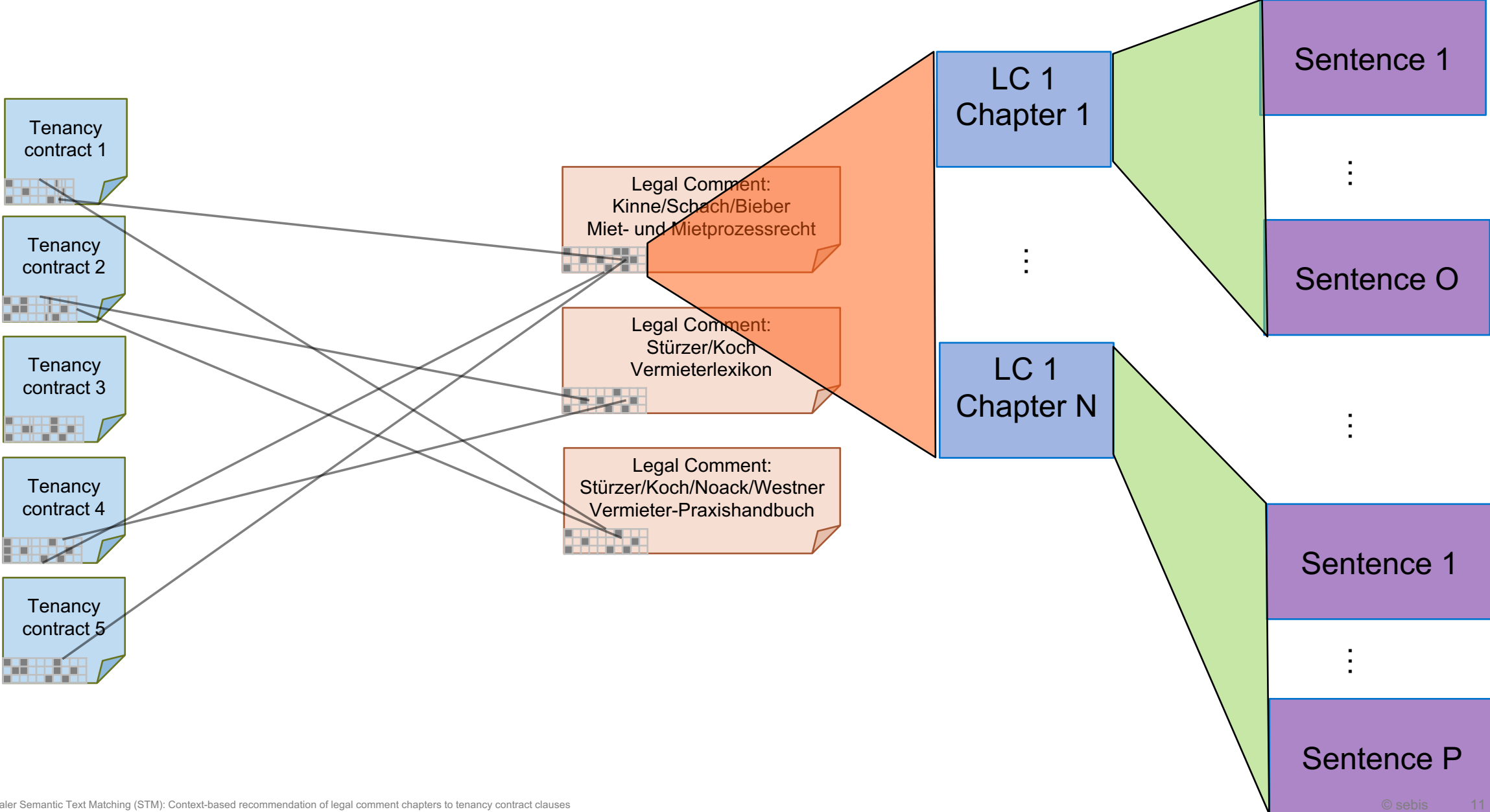
Results STM with TF-IDF and Word Embeddings (vector size=300, iter=100)

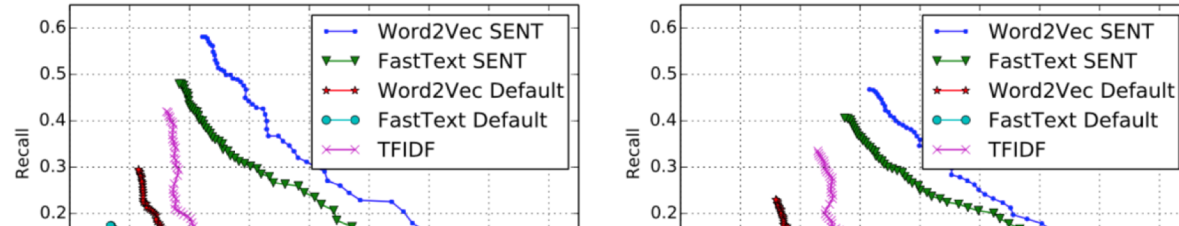
		Comment Nr. 3		SHR broad		SHR narrow	
		FRP	RPS	FRP	RPS	FRP	RPS
TF-IDF	CHAPTER Embeddings						84.49
	Lesson Learned 1) TFIDF performs better than Word2Vec CHAPTER Embeddings						59.51
	SP + Only POS-tagged nouns (=POSN)	7.21	28.74	2.76	369.62	5.43	254.74
Word Embeddings	CHAPTER Embeddings						311.94
	Lesson Learned 2) Training corpus extension for Word Embeddings does NOT always improve vector quality						-
	Trained on CC + BGB	17.70	42.23	6.09	475.69	6.42	336.97
	Trained on CC + BGB+ Wikipedia (1GB)	19.85	47.19	28.28	599.28	33.33	474.61
	SENTENCE Embeddings						26.09
	Lesson Learned 3) Our SENTENCE Embeddings significantly outperform TFIDF & CHAPTER Embeddings						02.84
	Trained on CC + SENT	25.85	58.17	3.25	255.33	3.33	233.29

FRP: Average First Ranking Position; RPS: Average Ranking Position (RP-Score)

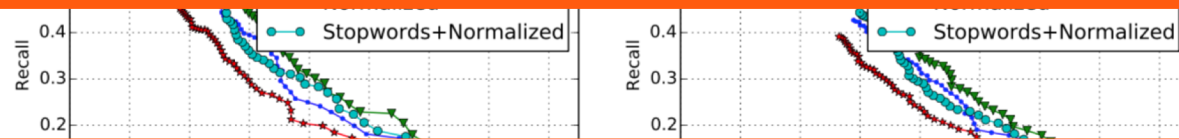
[La18a] Landthaler, J.; Scepankova, E.; Glaser, I.; Lecker, H.; Matthes, F.: Semantic Text matching of contract clauses and legal comments in tenancy law, IRIS: Internationales Rechtsinformatik Symposium, Salzburg, Austria, 2018

Additional Sentence Segmentation of the Legal Comment Chapters

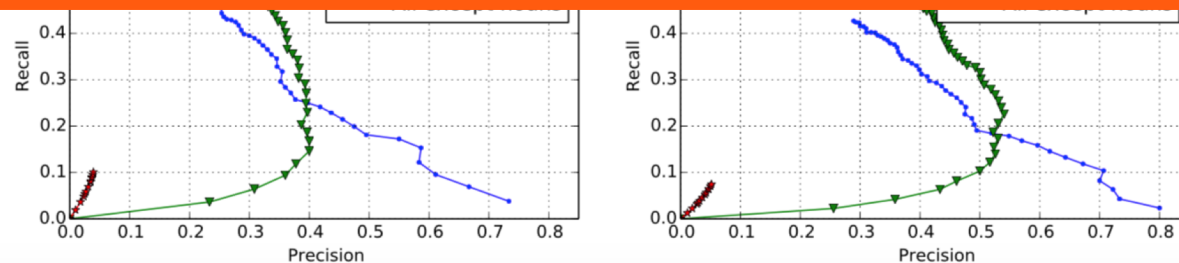




Lesson Learned 4) **Word2vec** outperforms FastText (subword-embeddings)



Lesson Learned 5) **Nouns** contribute most to the topical relatedness/similarity)



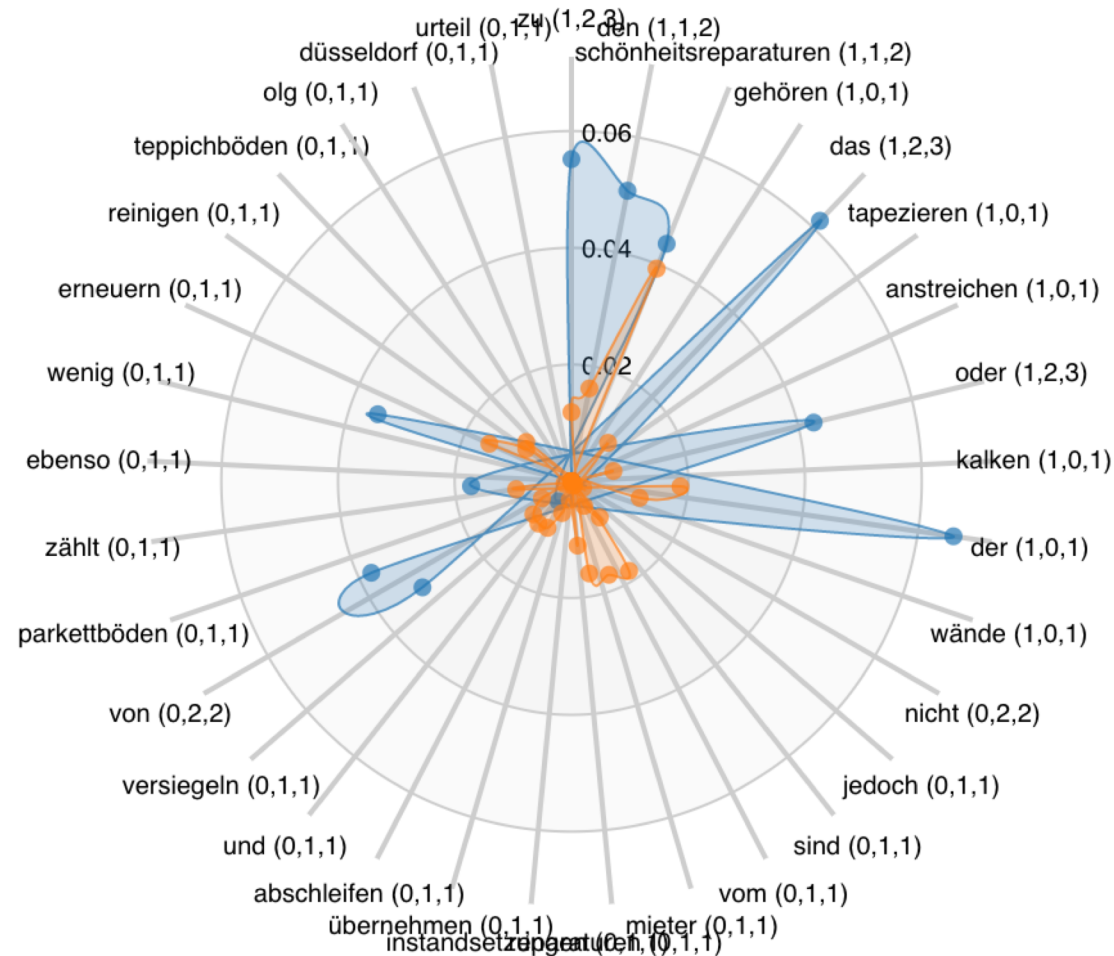
Visualization of XSTM for a Single Match

Input Query:

“Zu den Schönheitsreparaturen gehören das Tapezieren, Anstreichen oder Kalken der Wände”

Match Candidate:

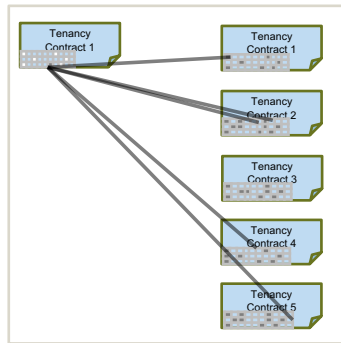
“4 Satz 4 II. BV wie folgt:
„Schönheitsreparaturen umfassen nur das Tapezieren, Anstreichen oder Kalken der Wände und Decken, das Streichen der Fußböden, Heizkörper einschließlich der Heizrohre, der Innentüren sowie der Fenster und Außentüren von innen.”



Sentence Query

Contribution is the difference of the similarities of query and qualifying text fragment when a single word/token is removed compared to similarity with all words. Occurrences of words/tokens in (Query, Sentence, Both)

Landthaler et al. 2018, submitted



A different STM problem

Das Halten kleiner Tiere

... aus Verträgen:

Der Mieter darf in der Wohnung Haustiere halten, soweit dies nach Anzahl und Größe der Tiere allgemein üblichen Vorstellungen entspricht.

Die Haltung anderer Tiere, insbesondere von Hunden und Katzen, ist nur nach vorheriger Zustimmung des Vermieters gestattet. Dies gilt auch für die zeitweilige Verwahrung von Tieren.

Die Tierhaltung in der Wohnung ist ohne Erlaubnis des Vermieters nur gestattet, sofern es sich um Kleintiere (z.B. Wellensittiche, Meerschweinchen, Hamster, Aquarienfische) handelt. Die Haltung größerer Tiere, dazu gehören auch Katzen und Hunde, bedarf der vorherigen Zustimmung des Vermieters.

Die Haltung von Hunden und Katzen sowie anderer als Tiere bedarf der Zustimmung des Vermieters.

Soweit keine Beeinträchtigungen gegenüber anderen Mietern entstehen, ist das Halten von Kleintieren (z.B. Meerschweinchen, Hamster, Wellensittich, Zierfische) zulässig.

Die Haltung von Kleintieren (wie z. B. Wellensittichen, Zierfischen, Hamstern, Kanarienvögeln u.ä.) ist zulässig, soweit es nicht zu Unzuträglichkeiten kommt.

Die Haltung anderer Tiere ist nur mit vorheriger schriftlicher Zustimmung des Vermieters gestattet. Eine etwaige Zustimmung gilt nur für den Einzelfall und kann widerrufen werden, wenn Unzuträglichkeiten eintreten.

Einer Zustimmung bedarf es nicht für solche Kleintiere, durch deren Haltung keinerlei Schäden, Belästigungen oder Gefährdungen entstehen können und keine Ekelgefühle bei Dritten hervorgerufen werden. Eine erteilte Zustimmung erlischt mit dem Tod oder der Abschaffung des Tieres. Dem Mieter ist untersagt, Tauben und andere Wildtiere vom Grundstück aus zu füttern.

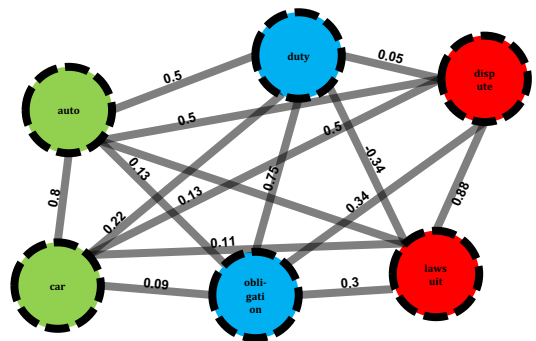
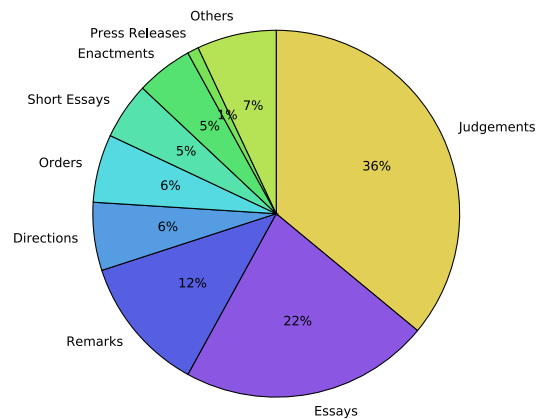
Die Haltung von Kleintieren ist dem Mieter ohne Zustimmung des Vermieters gestattet, soweit durch die Unterbringung in den Mieträumen eine Beeinträchtigung der Mietsache oder eine Belästigung von Hausbewohnern oder Nachbarn nicht gegeben ist.

..... -
..... und:
iesen durch Personalausweis
.....
... - nachstehend "Mieter" genannt

us dem Mietverhältnis in voller Höhe.
argelegten Umfang.

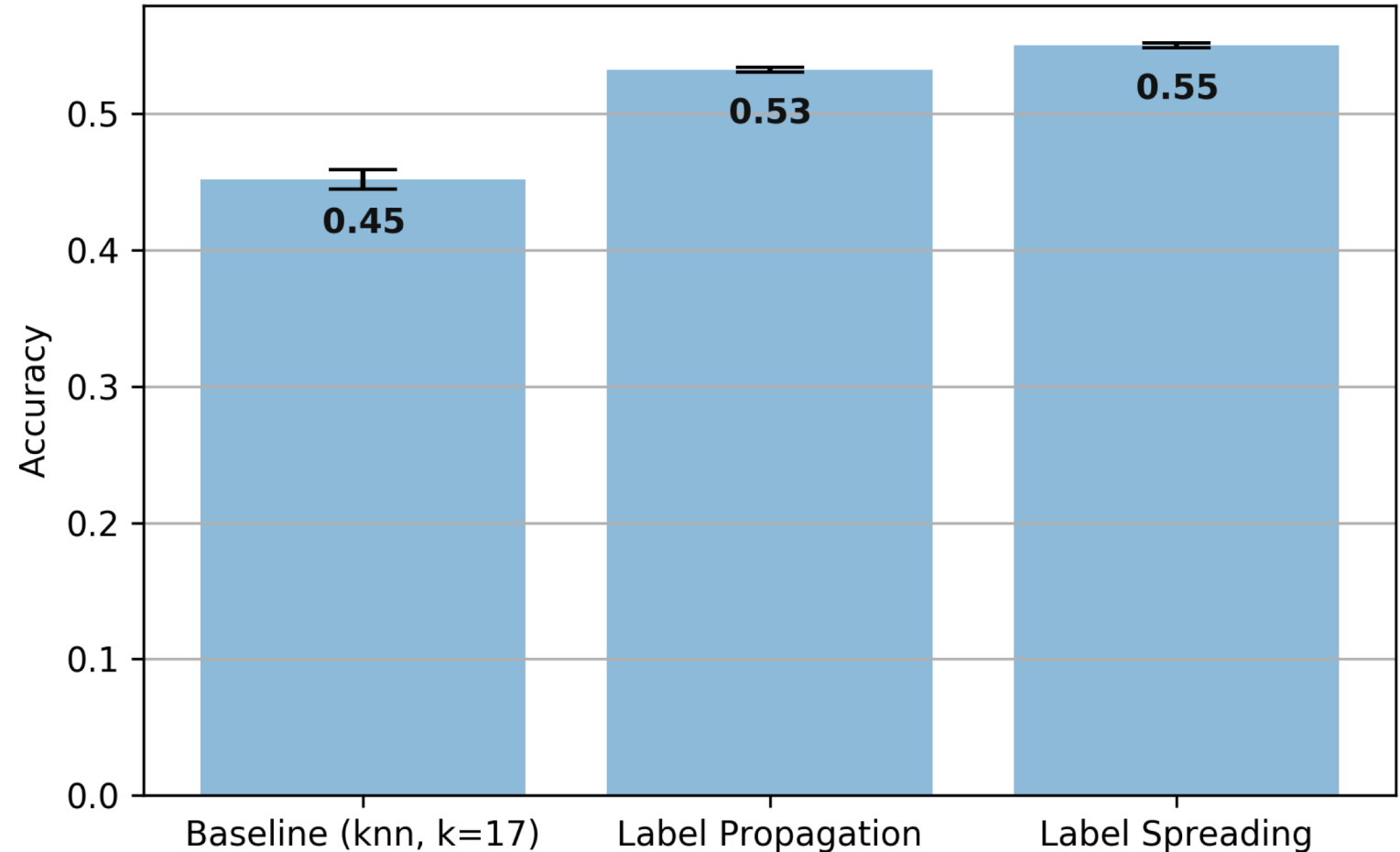
Outlook: Thesaurus Extension with Label Propagation (MT Markus Müller)

Datev Corpus, 1GB Tax Law
Thesaurus with 36.000 terms



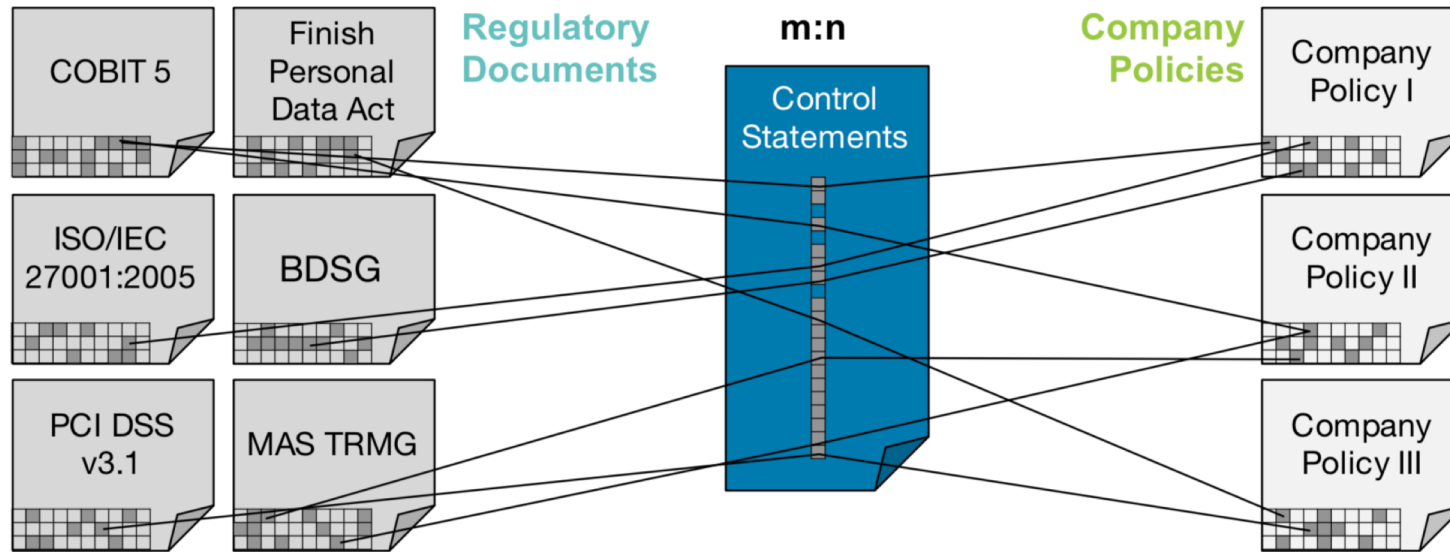
Label A: Synonym Group 1
Label B: Synonym Group 2
Label C: Synonym Group 3

Thesaurus Extension



MT, Markus Müller, 2018, <https://wwwmatthes.in.tum.de/pages/1k57iqI63u5m5/Master-s-Thesis-von-Markus-Mueller>

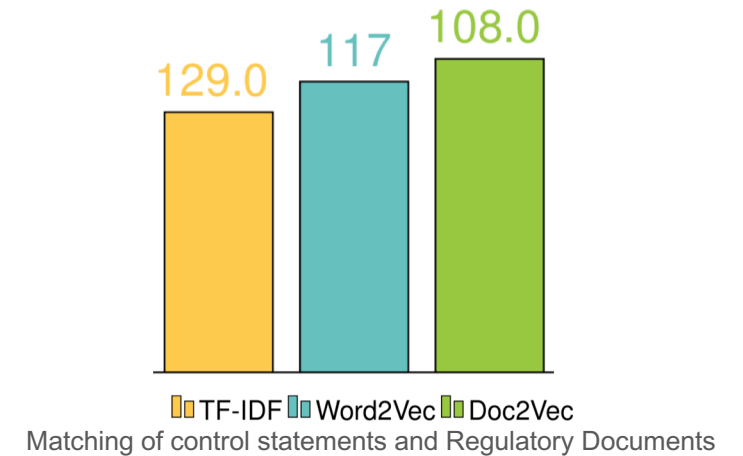
Semantic Text Matching in Compliance: MT Christoph Erl with Alyne



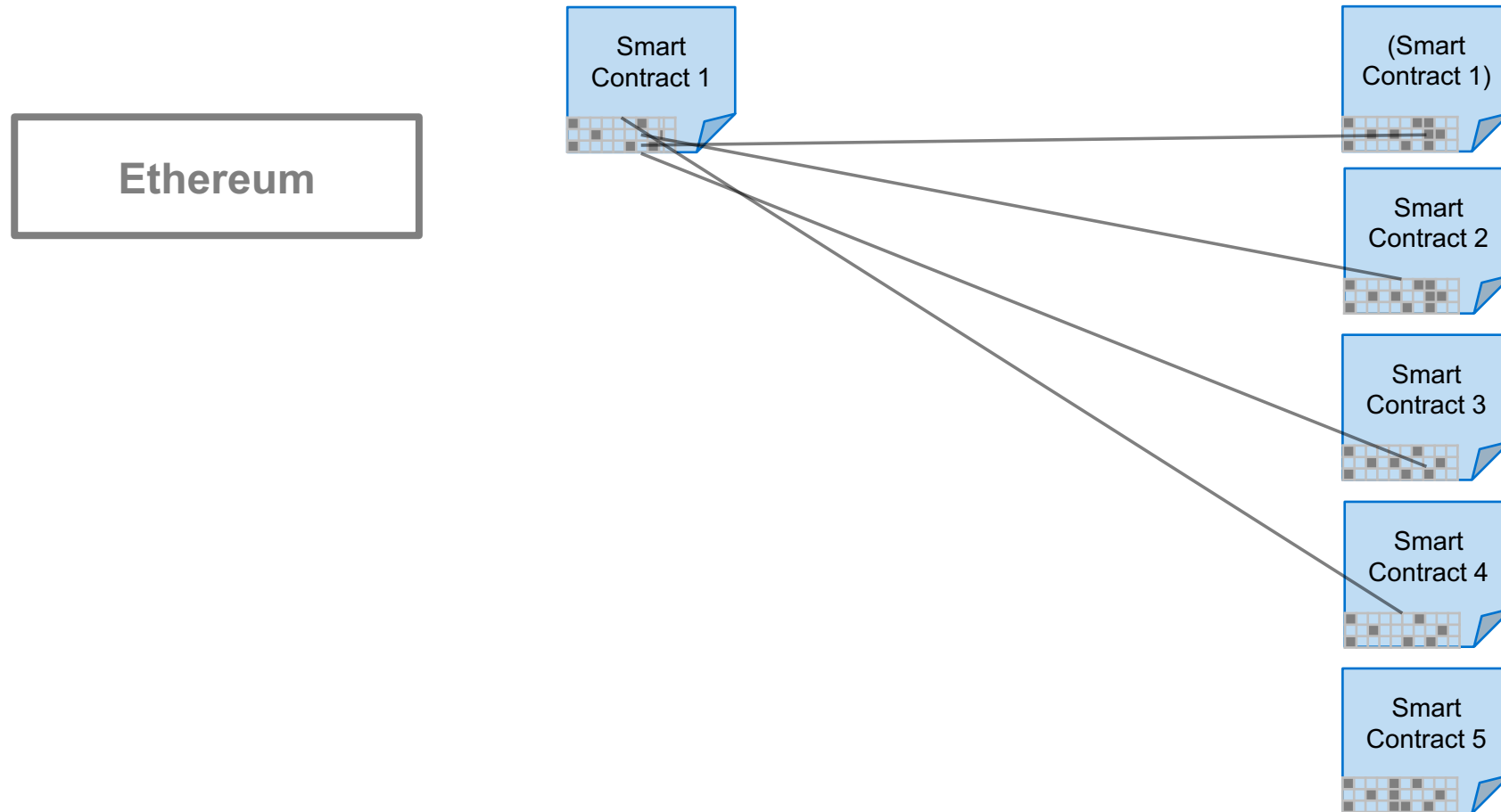
PCI DSS § 8.2.5
Do not allow an individual to submit a new password that is the same as any of the last four passwords [...].

Control Statement
User passwords shall be prevented from being changed to any of the previous x passwords.

Company Policy § 2.3.2
The user password [...] must not be same with any of his/her previous 4 passwords.



Outlook: Semantic Text Matching on Smart Contracts⁽²⁾



Semantic Text Matching of smart contracts or functions in smart contracts.

**The demo prototype will be
available after the presentations!**



Jörg Landthaler, M.Sc.

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business
Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289.17139
Fax +49.89.289.17136

joerg.landthaler@tum.de
www.matthes.in.tum.de

