

Multilevel Text Alignment with Cross-Document Attention

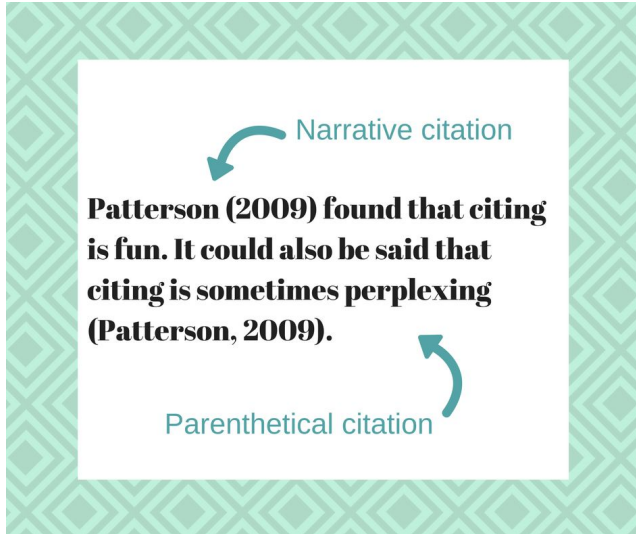


Xuhui Zhou, Nikolaos Pappas, Noah A. Smith

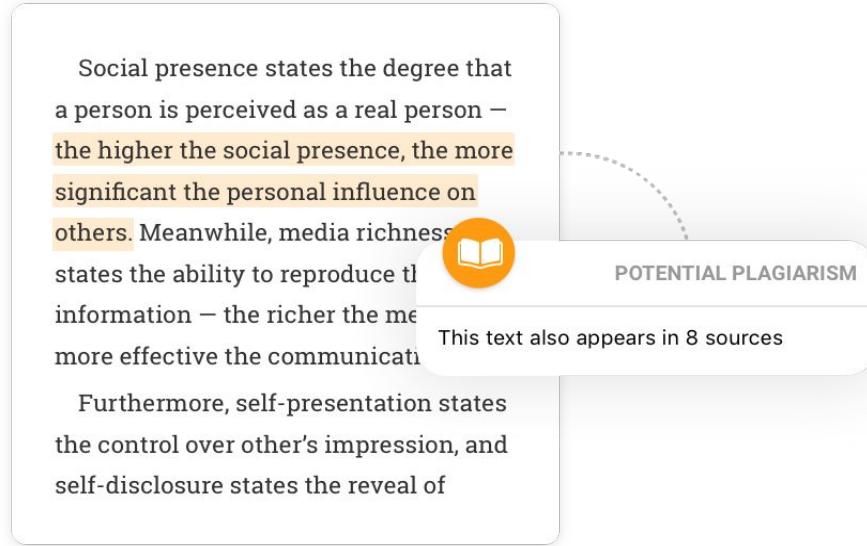
EMNLP 2020



Task: Comparing Documents



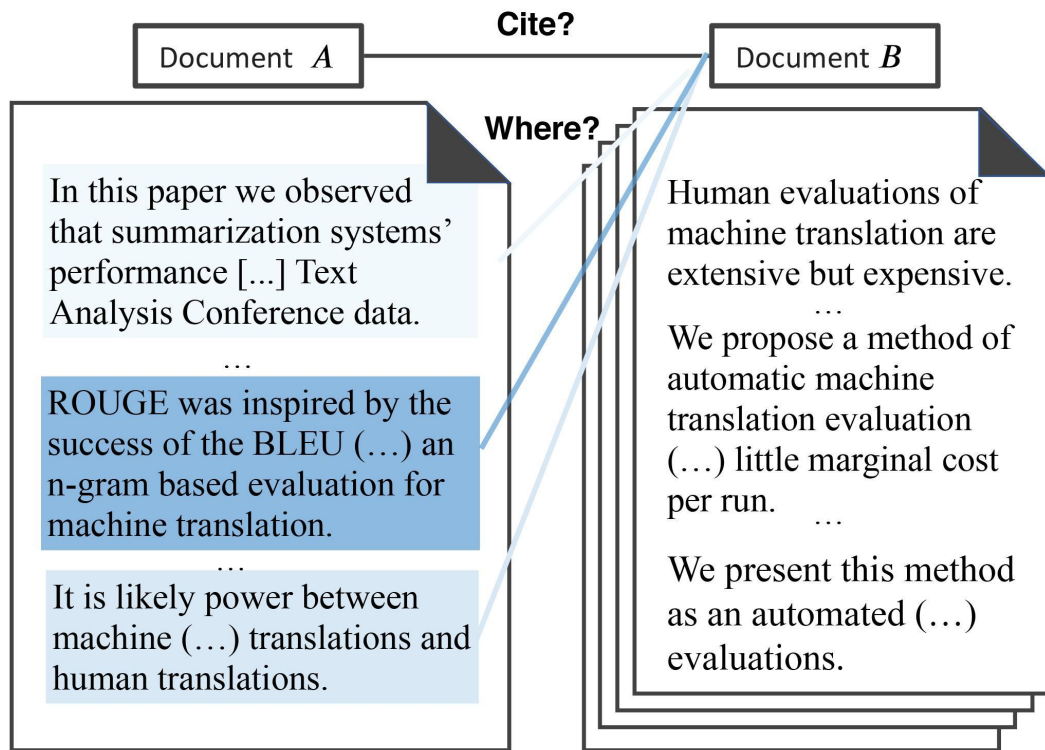
Citation Recommendation



Plagiarism Detection

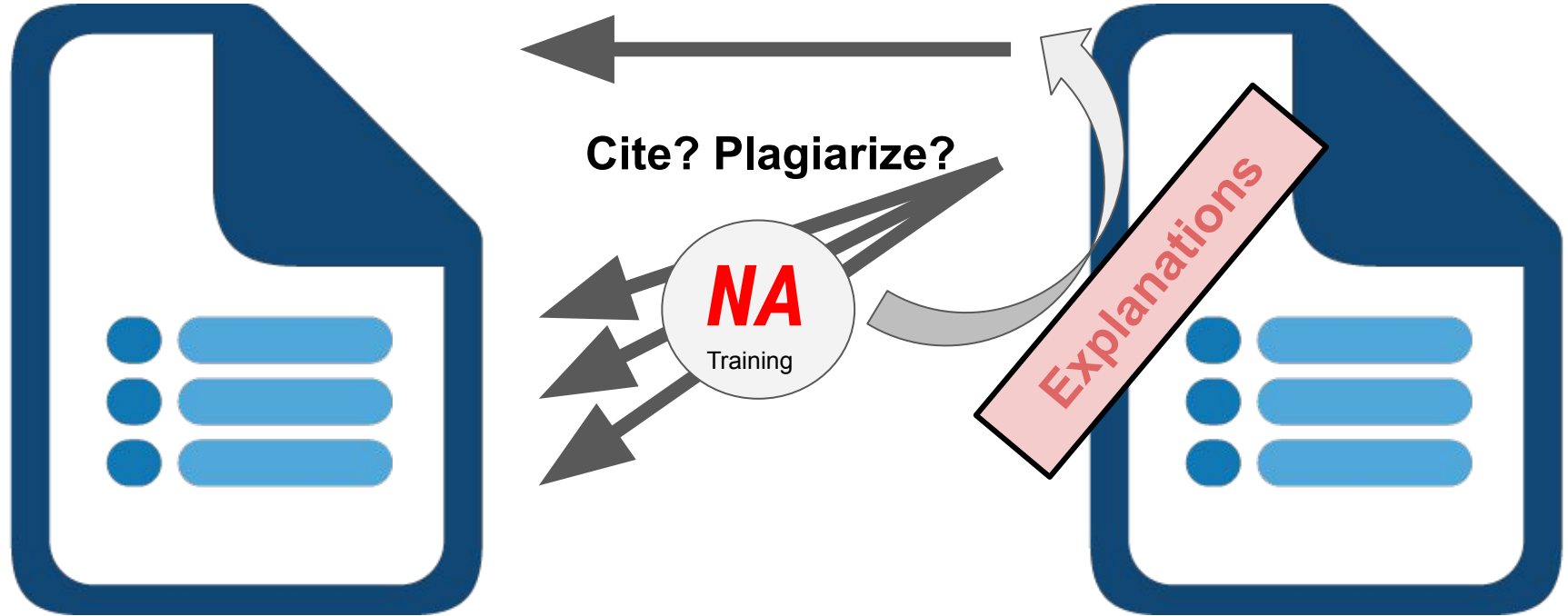
And More ...

Multilevel document alignment

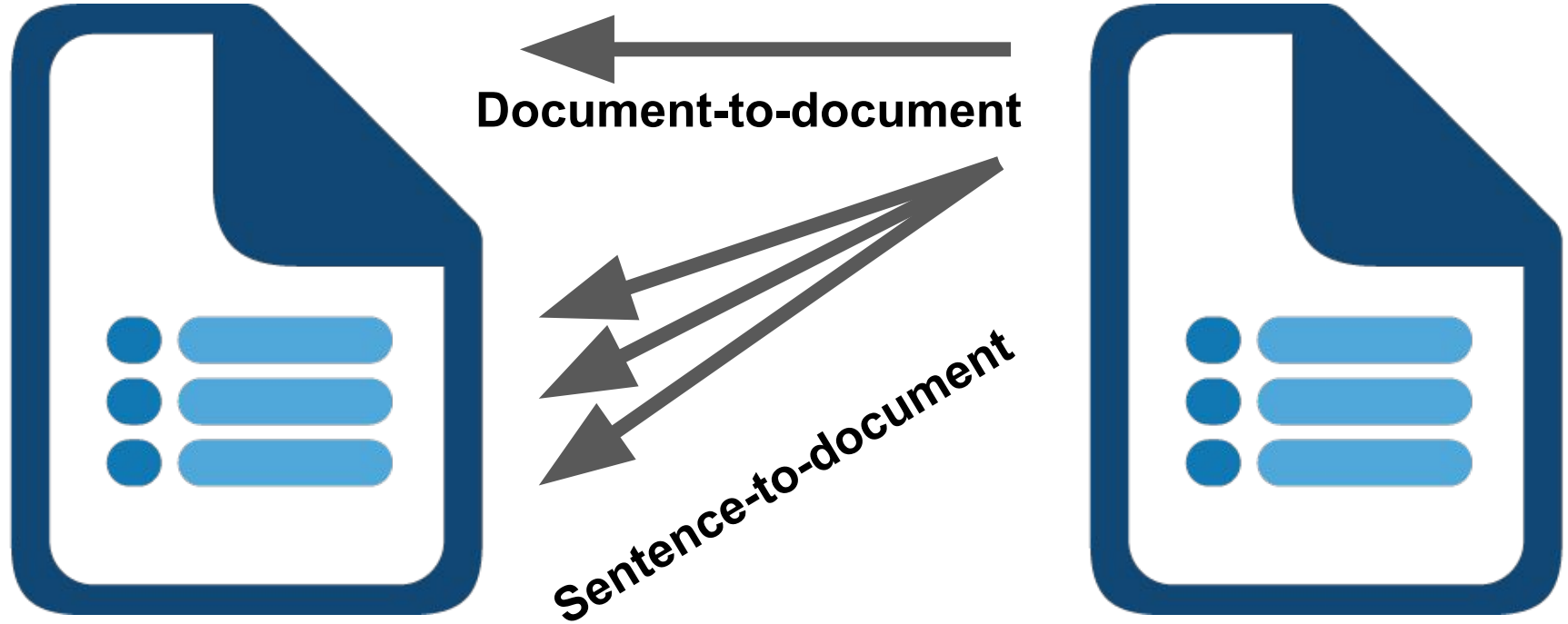


- Whether document A should cite document B
- Some sentences in A support that relationship
- Helps us understand model's decision

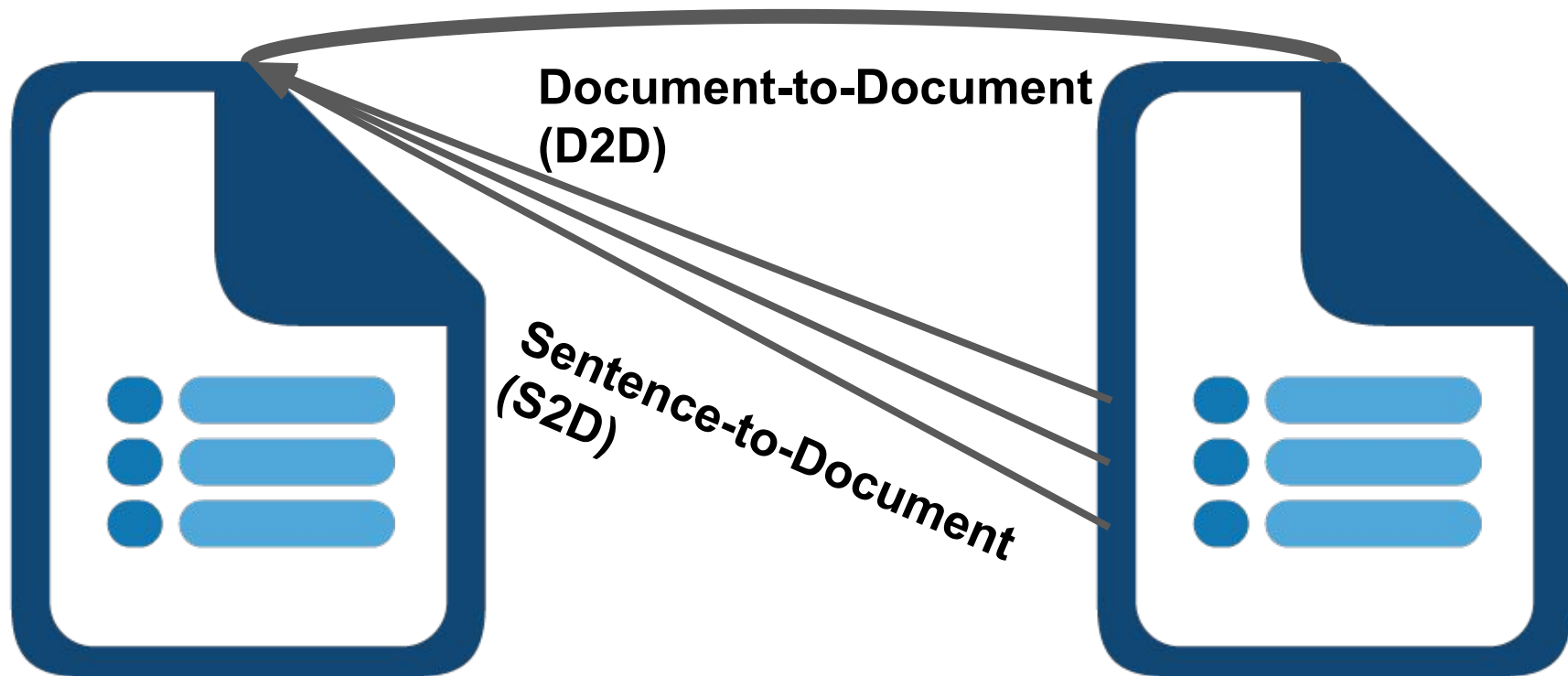
Multilevel document alignment



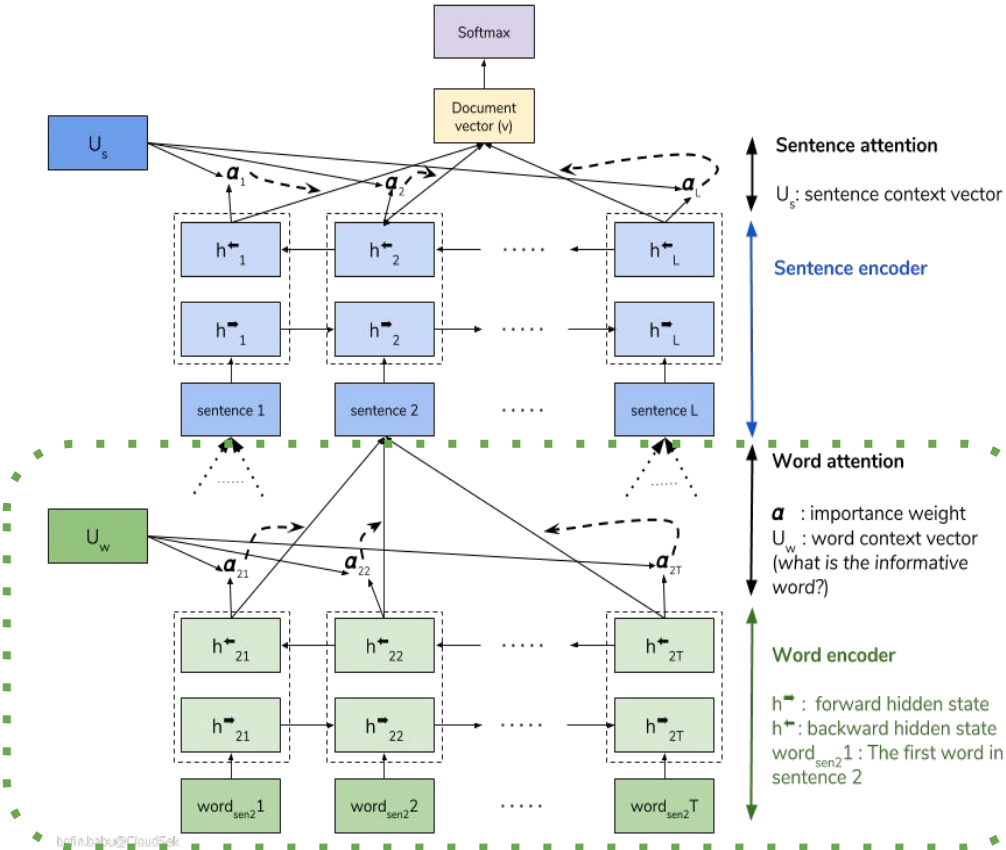
Multilevel document alignment



Want both document and fine-grained alignment



Hierarchical Document Encoders

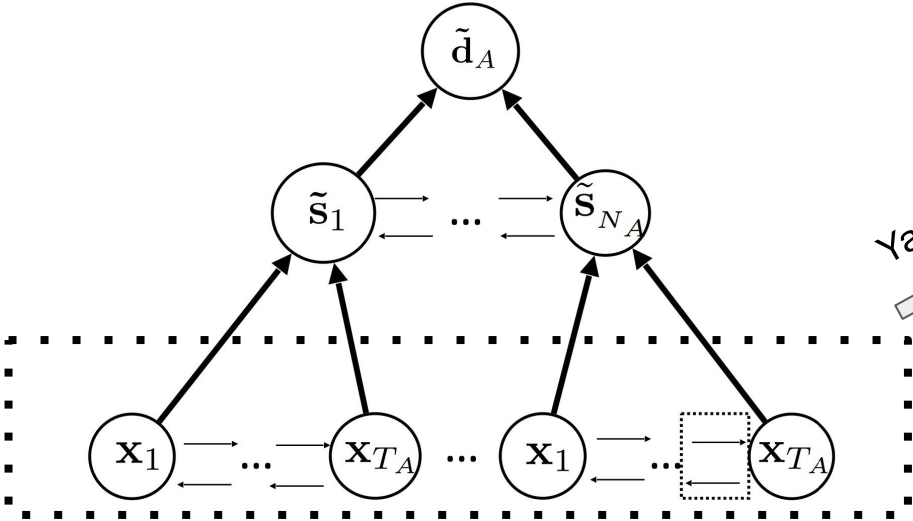


Pappagari et al. 2019



Hierarchical Document Encoders

Document A

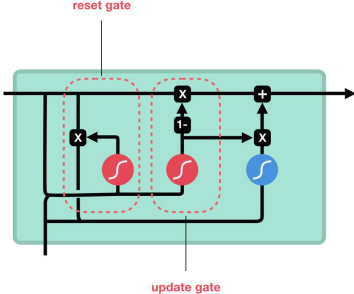


According to our view love is rewarded with pleasure.

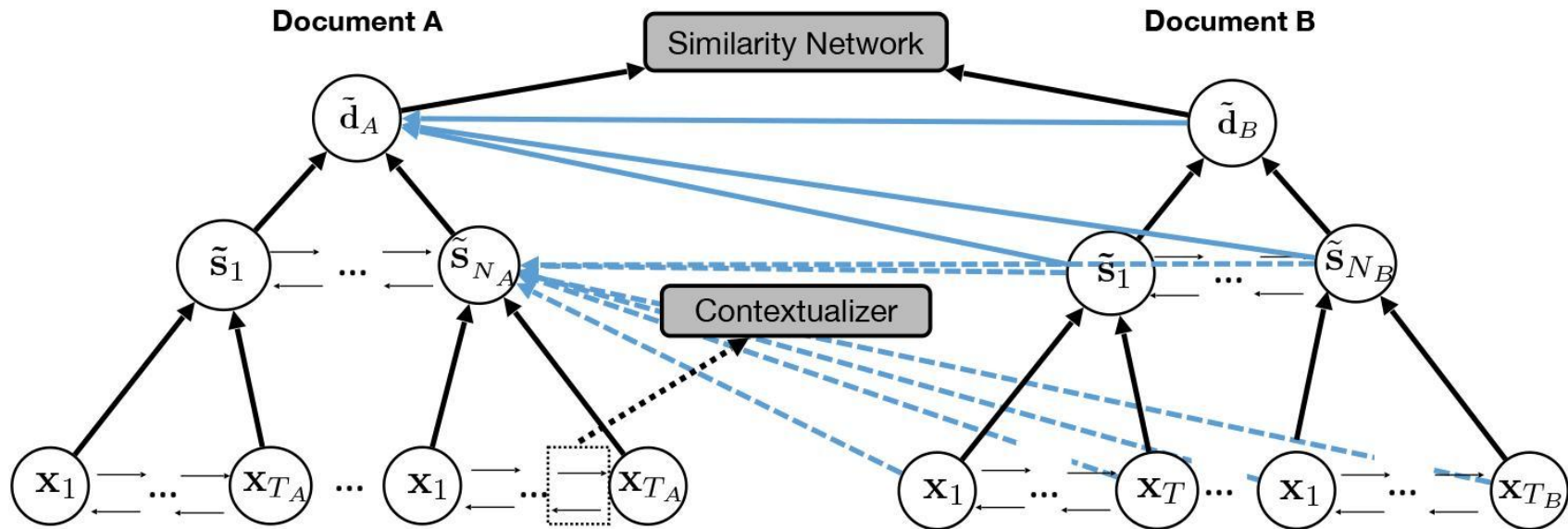
This object the bible attains by the commandments (...)

Yang et al., 2016

Pappagari et al. 2019



Cross-document attention (CDA)



According to our view love is rewarded with pleasure.

This object the bible attains by the commandments (...)

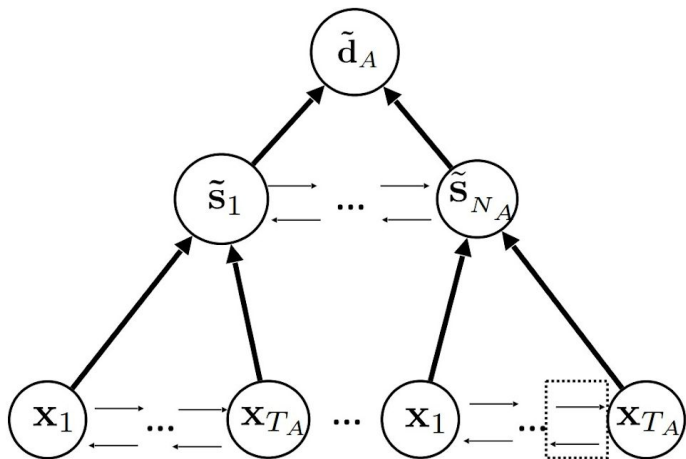
Proof of this that we find no delight in axioms (...)

The rabbis furthermore talk of decisive locations (...)

$$\tilde{d}_A = \text{affine} \left(\left[\mathbf{d}_A; \sum_{\mathbf{v} \in \mathcal{B}} \frac{\exp \mathbf{v}^\top \mathbf{d}_A}{\sum_{\mathbf{v}' \in \mathcal{B}} \exp \mathbf{v}'^\top \mathbf{d}_A} \mathbf{v} \right] \right)$$

Comparing documents

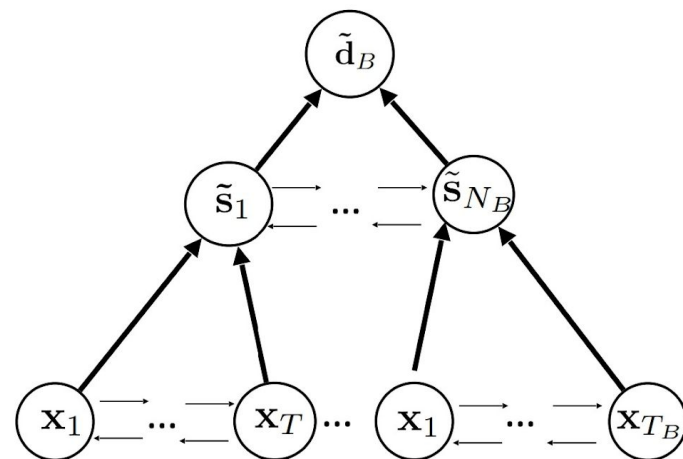
Document A



According to our view love is rewarded with pleasure.

This object the bible attains by the commandments (...)

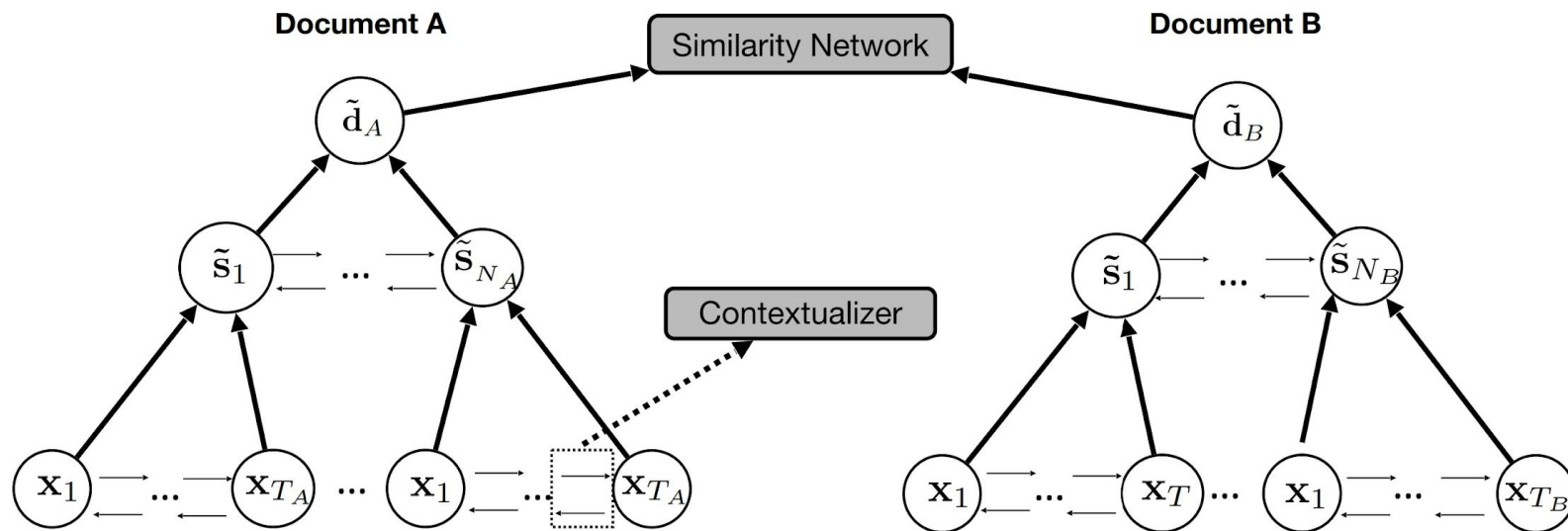
Document B



Proof of this that we find no delight in axioms (...)

The rabbis furthermore talk of decisive locations (...)

Cross-document attention (CDA)



According to our view love is rewarded with pleasure.

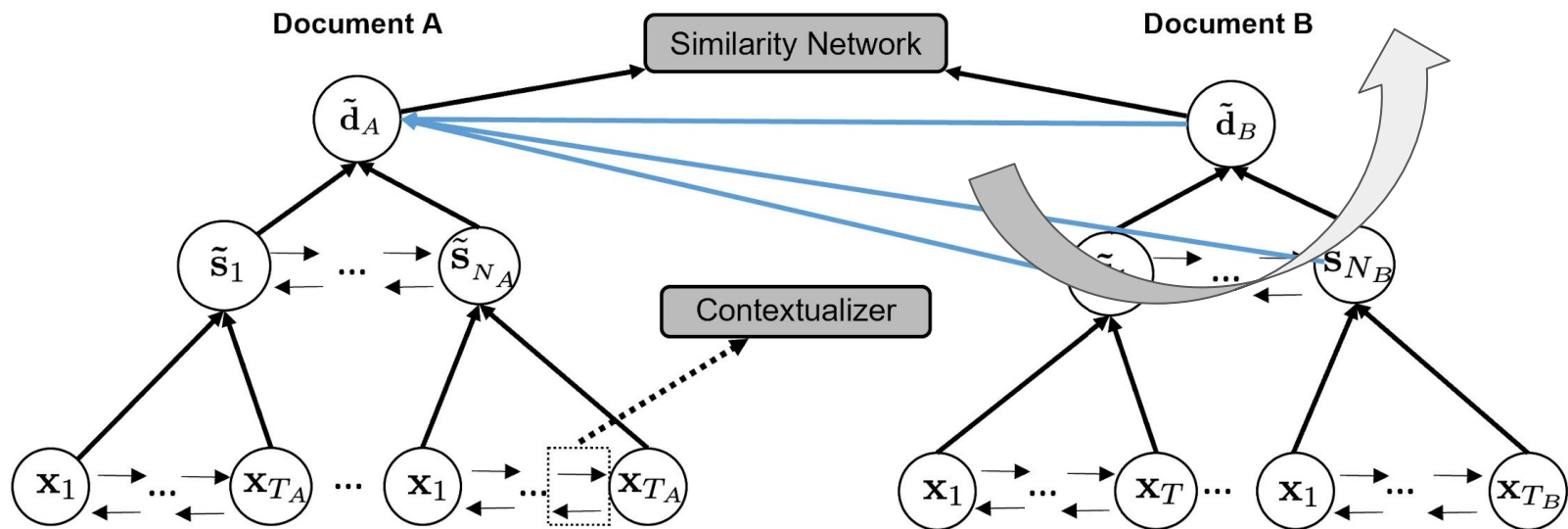
This object the bible attains by the commandments (...)

Proof of this that we find no delight in axioms (...)

The rabbis furthermore talk of decisive locations (...)

Cross-document attention (Shallow)

$$\tilde{\mathbf{d}}_A = \text{affine} \left(\left[\mathbf{d}_A; \sum_{\mathbf{v} \in \mathcal{B}} \frac{\exp \mathbf{v}^\top \mathbf{d}_A}{\sum_{\mathbf{v}' \in \mathcal{B}} \exp \mathbf{v}'^\top \mathbf{d}_A} \mathbf{v} \right] \right)$$



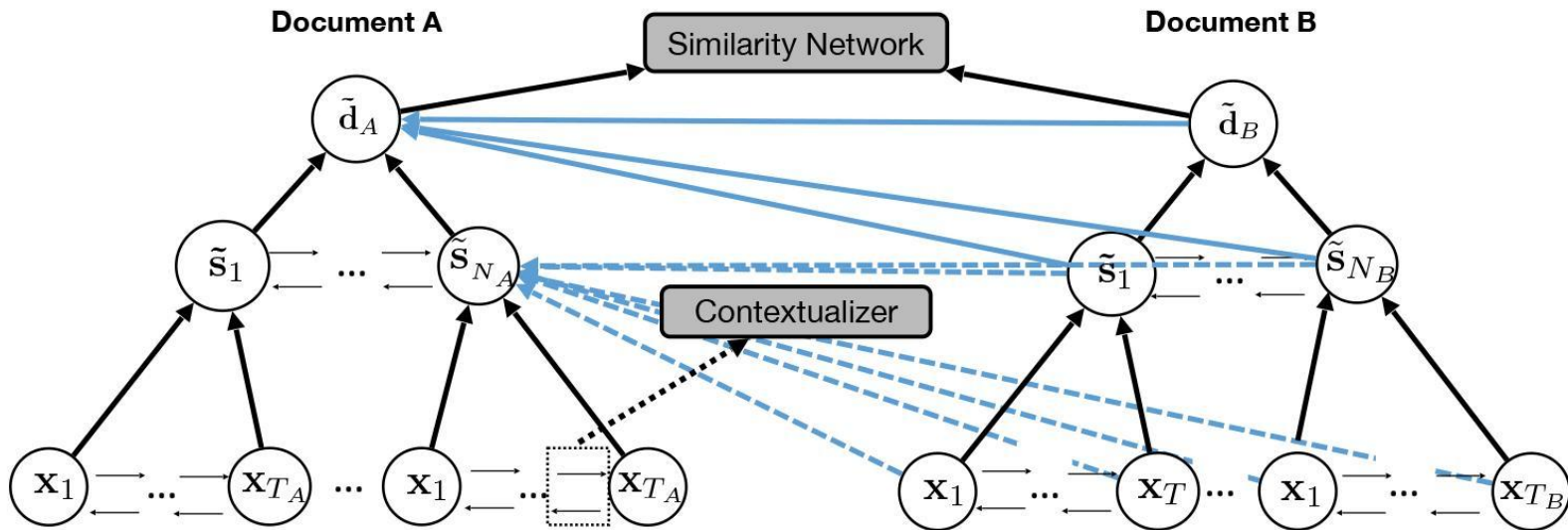
According to our view love is rewarded with pleasure.

This object the bible attains by the commandments (...)

Proof of this that we find no delight in axioms (...)

The rabbis furthermore talk of decisive locations (...)

Cross-document attention (Deep)



According to our view love is rewarded with pleasure.

This object the bible attains by the commandments (...)

Proof of this that we find no delight in axioms (...)

The rabbis furthermore talk of decisive locations (...)

A Benchmark for Document Relation Prediction and Localization

	Pairs	Docs	Words		Sentences		
Dataset	count	count	avg	std	avg	std	
Radev et al., 2009	AAN	132K	13K	122.7	11.2	4.9	2.7
Bhagavatula et al., 2018	OC	300K	567K	190.4	16.3	7.0	3.5
Lo et al., 2020	S2ORC	190K	270K	263.7	19.2	9.3	5.9
Potthast et al., 2013	PAN	34K	23K	1569.7	90.4	47.4	66.1

Experiments

Models:

- **BERT-AVG**: document representation = the **average** of the sentence representations
- **BERT-HAN**: document representation = the **attention** of the sentence representations
- **GRU-HAN**: original HAN

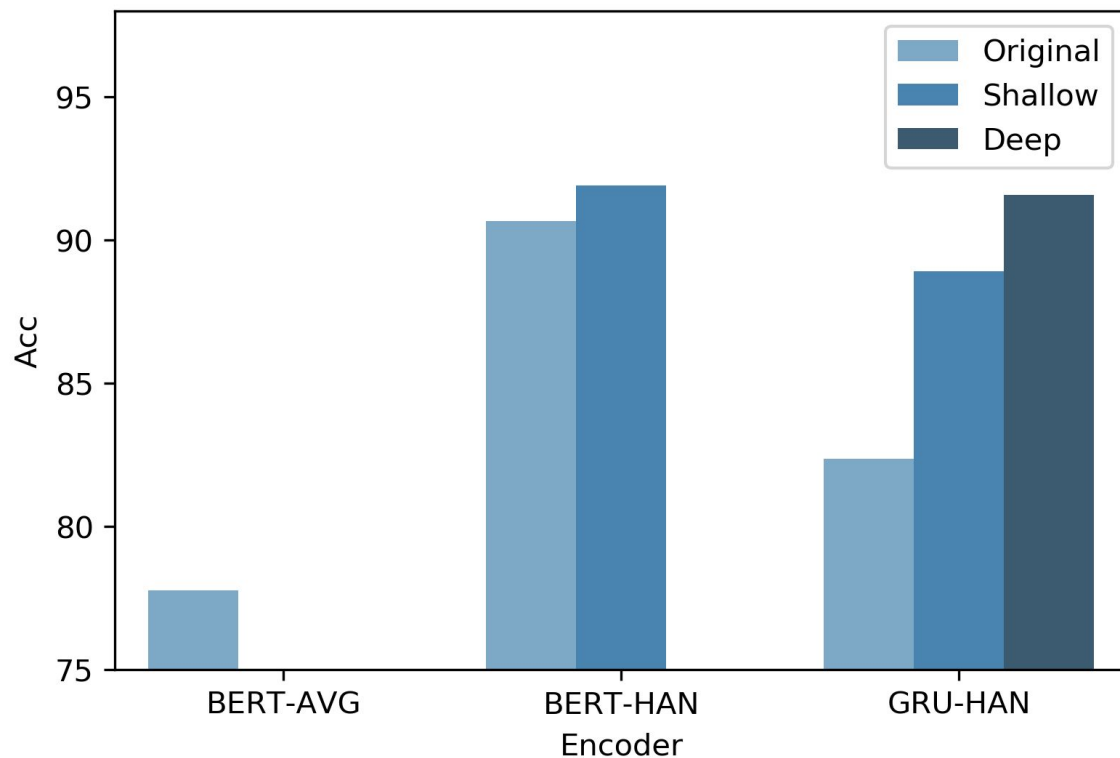
Evaluation:

- **Document-to-document**: F1
- **Sentence-to-document**: MRR

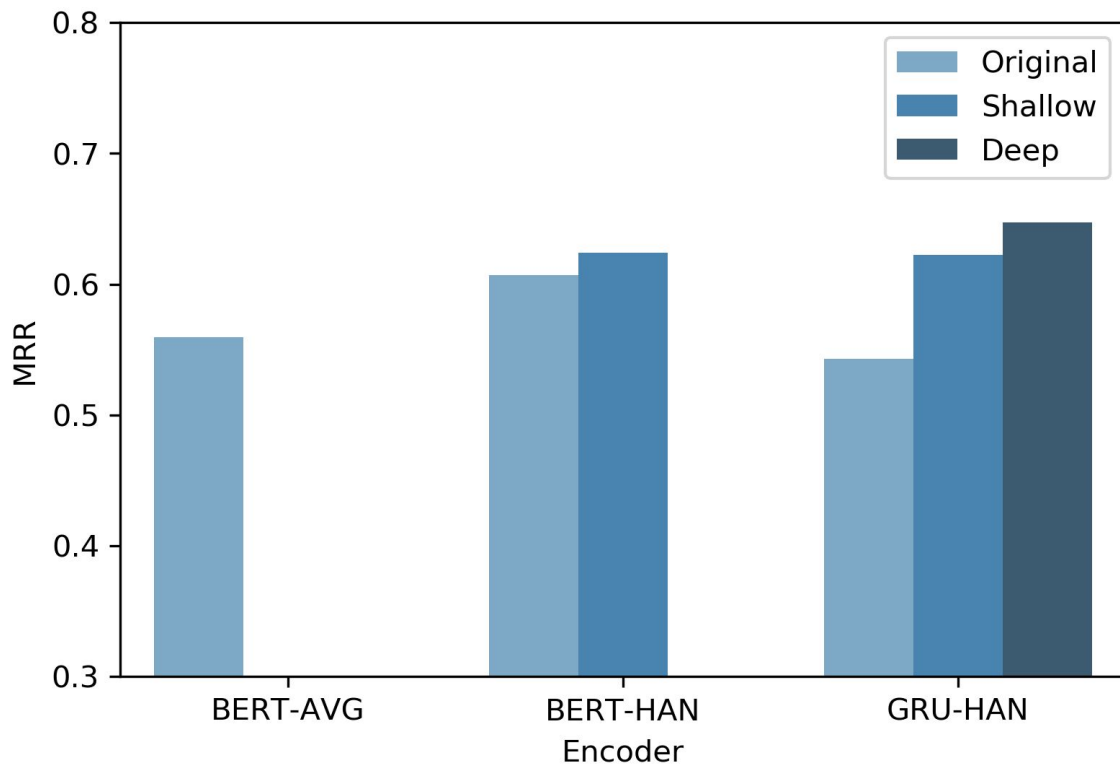
Results (Document-to-document)

Encoder	CDA	AAN		OC		S2ORC		PAN	
		Acc	F_1	Acc	F_1	Acc	F_1	Acc	F_1
BERT-AVG	–	53.54	53.89	84.72	84.99	77.78	76.92	79.62	76.60
BERT-HAN	–	67.32	64.97	85.96	86.33	90.67	90.76	87.57	87.36
	SHALLOW	71.57	69.08	87.81	87.89	91.92	92.07	86.23	86.19
GRU-HAN	–	68.01	67.23	84.46	82.26	82.36	83.28	75.70	75.88
	SHALLOW	74.51	74.81	88.71	88.96	88.91	89.92	77.04	78.23
	DEEP	75.08	75.18	89.79	89.92	91.59	91.61	75.77	76.71

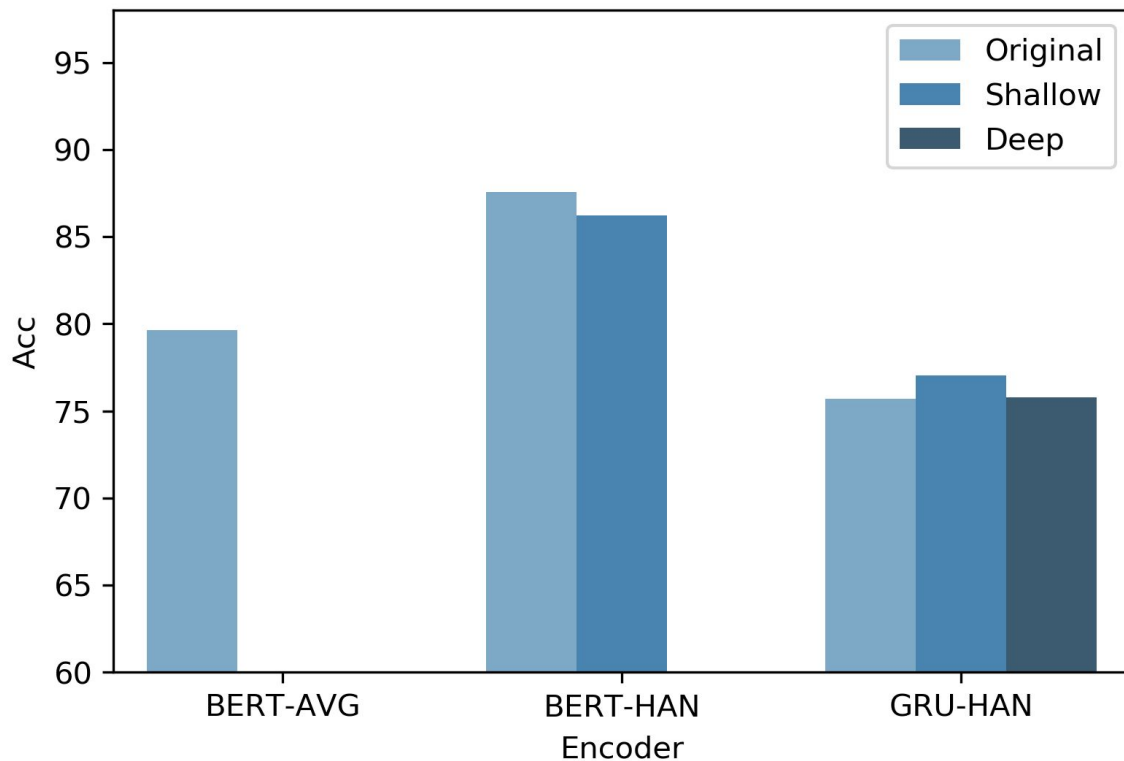
Results (Document-to-document; Citation Rec.)



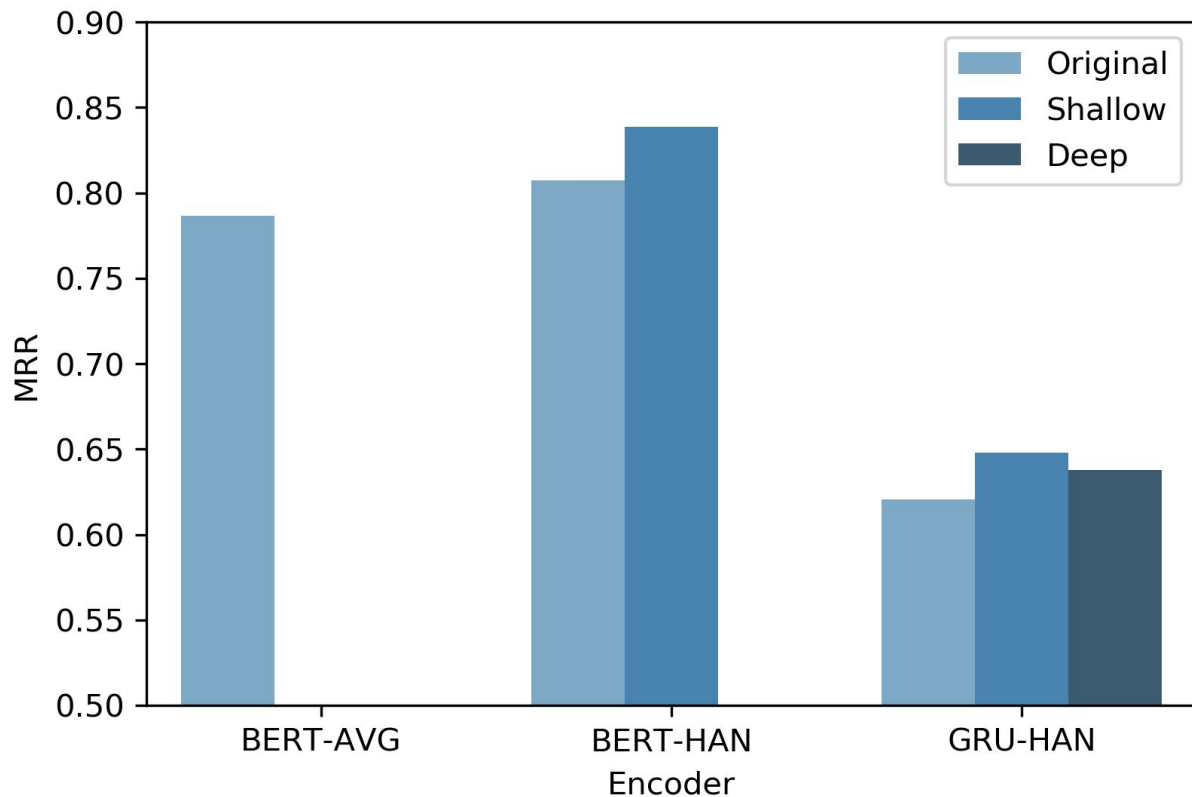
Results (Sentence-to-document; Citation Rec.)



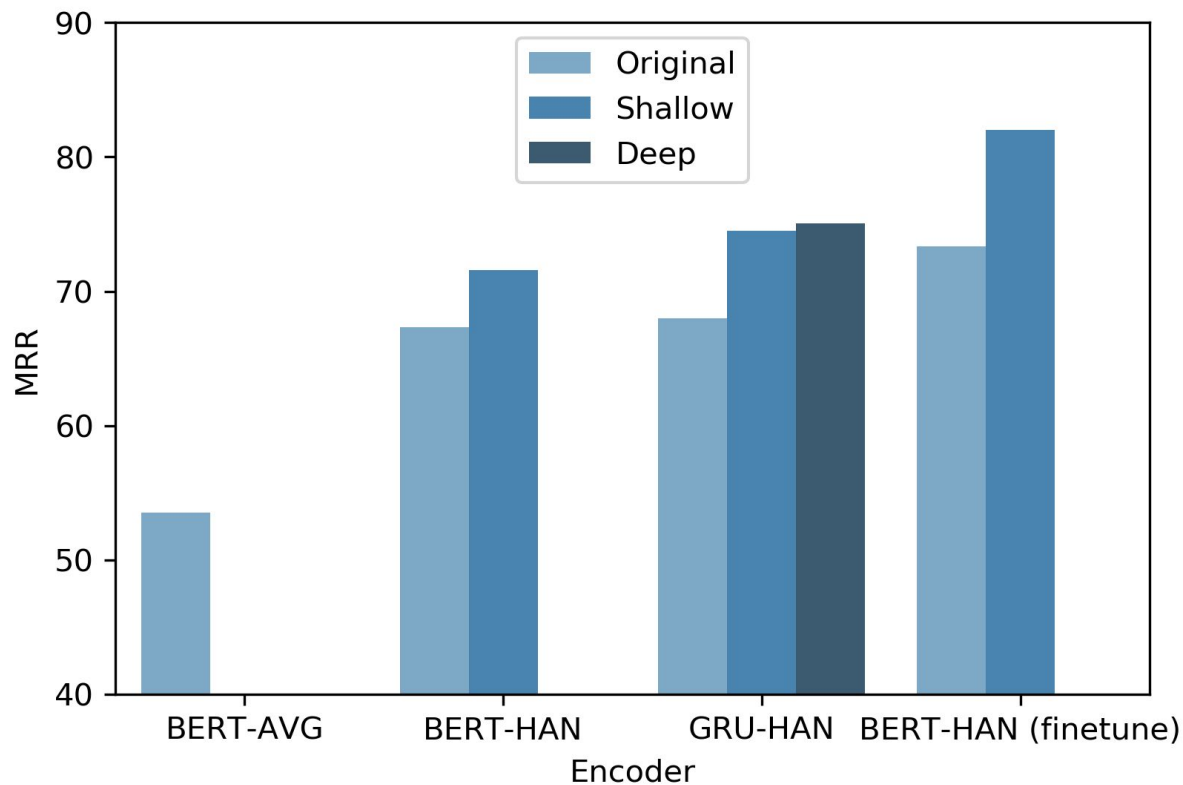
Results (Document-to-document; Plagiarism Det.)



Results (Sentence-to-document; Plagiarism Det.)



Fine tuning BERT



Results (Sentence-to-document)

Encoder	CDA	Attention Alignment				Cosine Alignment			
		MRR	P@10	P@5	P@1	MRR	P@10	P@5	P@1
Random	–	0.3611	46.03	39.30	28.83	0.3611	46.03	39.30	28.83
BERT-AVG	–	0.5596	72.67	64.24	47.22	0.5470	74.18	64.25	45.07
BERT-HAN	–	0.6068	82.75	70.67	51.20	0.5481	63.71	56.73	48.46
	SHALLOW	0.6240	84.95	73.14	52.51	0.6152	81.21	71.16	52.27
GRU-HAN	–	0.5430	74.94	63.38	44.75	0.4742	52.59	47.91	41.68
	SHALLOW	0.6225	84.17	73.11	52.42	0.5013	54.19	48.93	45.31
	DEEP	0.6474	86.37	76.11	54.90	0.6252	81.93	72.04	53.35

Results (Sentence-to-document)

Enc	CDA	MRR	P@10	P@5
Random	–	0.4215	44.23	43.28
BERT-AVG	–	0.7864	58.24	64.69
BERT-HAN	–	0.8072	60.36	68.94
	SHALLOW	0.8386	60.47	69.07
GRU-HAN	–	0.6205	50.72	51.90
	SHALLOW	0.6479	51.71	53.05
	DEEP	0.6378	52.07	53.82

Citing Document: The impressive achievements in image classification using deep neural networks at the turn of the decade precipitated a reemergence of interest in deep learning. Deep neural networks have achieved significant accuracy improvements in a broad spectrum of areas, including computer vision, natural language processing, and network analysis [...].(*) In order to improve the predictive accuracy of IOT applications, researchers employed deep learning to model complicated sensing tasks. [...] Yet their design uses traditional CNNs and RNNs, combining the real and imagery parts of complex-value inputs as additional features. To the best of our knowledge, STFNet is the first work that integrates neural networks with traditional time-frequency analysis, and designs fundamental spectral-compatible operations for Fourier-transformed representations. [...]

Source Document: We present DEEPWALK, a novel approach for learning latent representations of vertices in a network. These latent representations encode social relations in a continuous vector space, which is easily exploited by statistical models. [...] DEEPWORK is also scalable. It is an online learning algorithm which builds useful incremental results, and is trivially parallelizable. These qualities make it suitable for a broad class of real world applications such as network classification, and anomaly detection.

Suspicious Document: According to our view love is rewarded with pleasure. The pleasure we feel here below in intellectual work proves nothing, for it is due to the effort and the passing from potential knowledge to actual knowledge, i. e., to the process of learning. Proof of this is that we find no pleasure in axioms and first principles, which we know without effort, but the acquired intellect after the death of the body does not learn any new truths, hence can have no pleasure. (*) [...] Then it is prepared for immortality as a natural thing without regard to reward. The purpose of the soul as we showed is to love god. This object the bible attains by the commandments, which may be classified with reference to their significance in seven groups.

Source Document: Proof of this is that we find no delight in axioms and first values, which we understand without effort. But the came by intellect after the death of the body does not discover any new realities, therefore can have no pleasure. The rabbis furthermore talk of decisive locations of pay and penalty, which will not request to the came by intellect, since it is a separate matter and can have no place. [...]

Takeaways

- We introduce a publicly available English benchmark for both document-to-document and sentence-to-document tasks.
- We introduce cross-document attention to augment a family of hierarchical models, which demonstrates promising results.
- The sentence-to-document tasks are challenging and worthwhile for future research.

<https://xuhuizhou.github.io/Multilevel-Text-Alignment/>