

ARElight: Context Sampling of Large Texts for Deep Learning Relation Extraction



Nicolay Rusnachenko¹ Huizhi Liang¹ Maksim Kalameyets¹ Lei Shi¹

1. School of Computing, Newcastle University, Newcastle upon Tyne
 {nicolay.rusnachenko, huizhi.liang, maksim.kalameyets, lei.shi}@newcastle.ac.uk



Motivation

There are numerous tool sets (CoreNLP [1], OpenNRE [2]) aimed at a deep analysis of the single document. With this ARElight demo we bridge the gap with the lack of instruments aimed on deep analysis of large texts or massive collections of documents.

Focus of Application

Sentiment Analysis between mentioned named entities [3]:

Russia criticized Belarus for permitting Georgian President Mikheil Saakashvili to appear on Belorussian television. "The appearance was an unfriendly step towards Russia," the speaker of Russian parliament Boris Gryzlov said. ... Saakashvili announced Thursday that he did not understand Russia's claims. Moscow refused to have any business with Georgia's president after the armed conflict in 2008 ...



Implementation Details

Language: Python-3.9

Sampler: AREkit-0.24.0

NER: BERT_{OntoNotes5} (DeepPavlov)

Translator: googletrans PyPI

Classifier: BERT_{base} 3-class

(positive, negative, neutral), fine-tuned on Russian texts:

- RuSentRel 1.1

- RuAttitudes 2.0

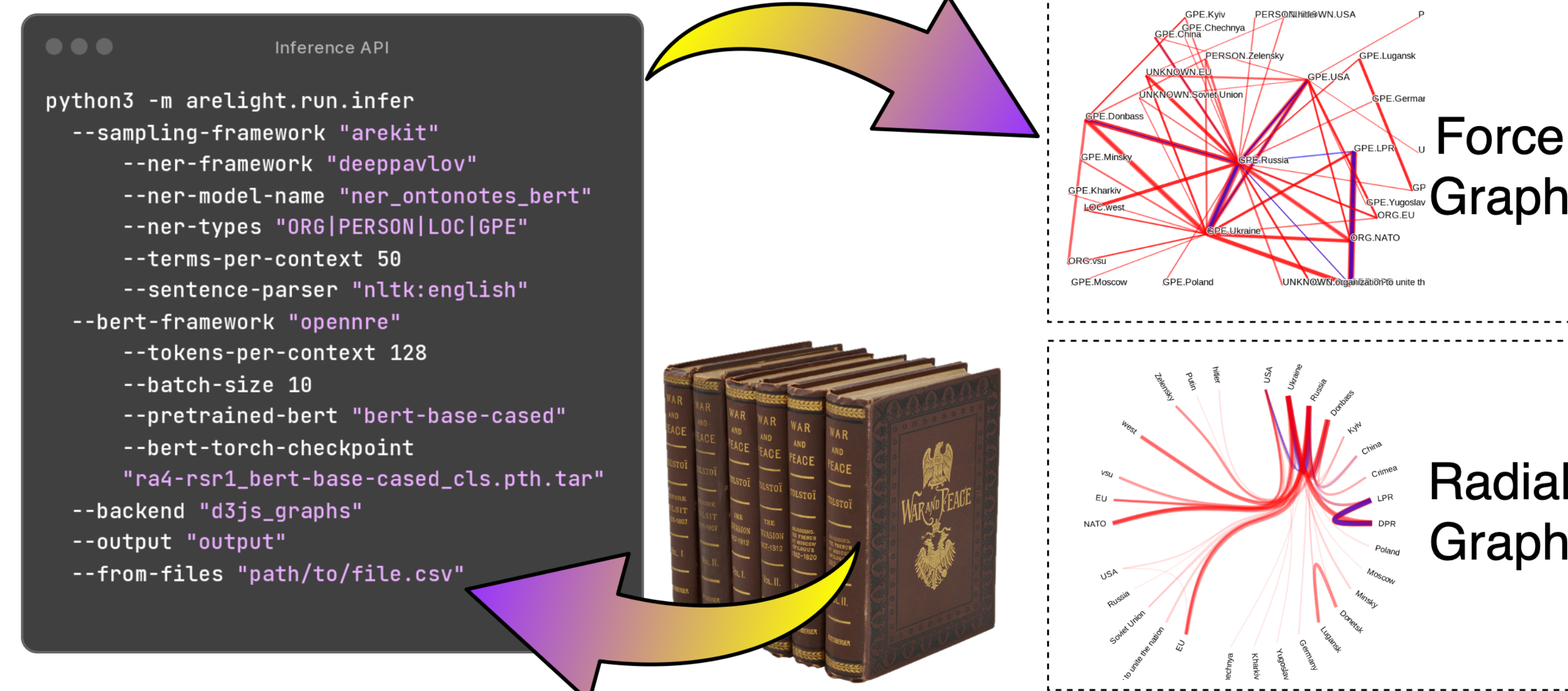
UI: Web + D3JS

`pip install git+https://github.com/nicolay-r/areligh@v0.24.0`

<https://github.com/nicolay-r/ARElight>



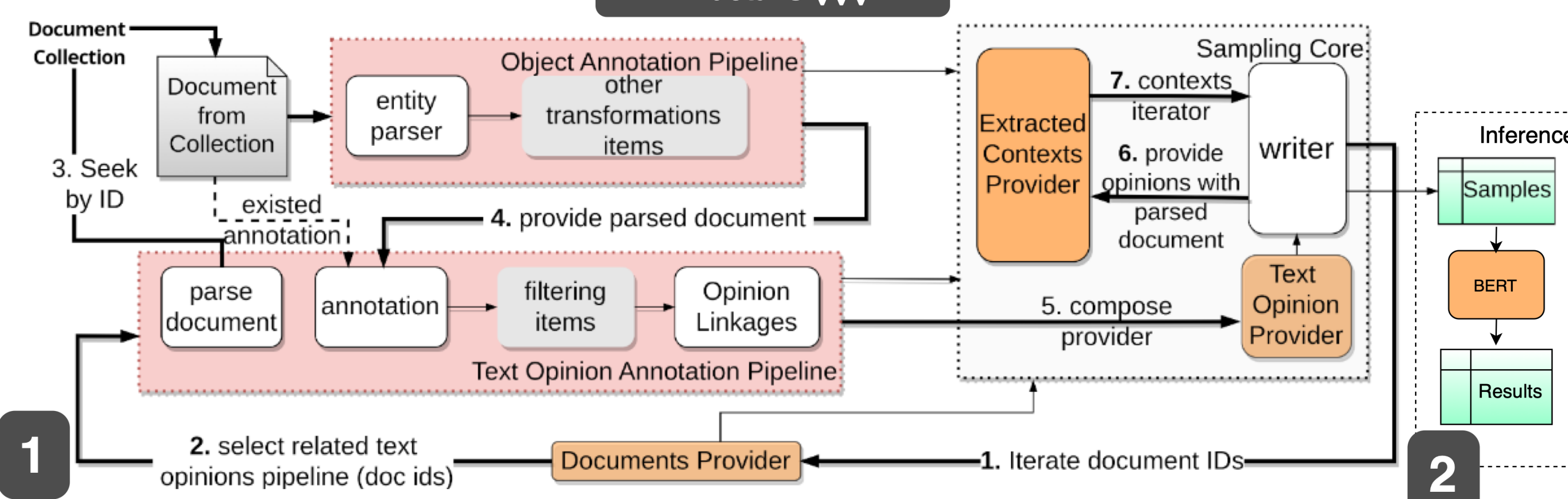
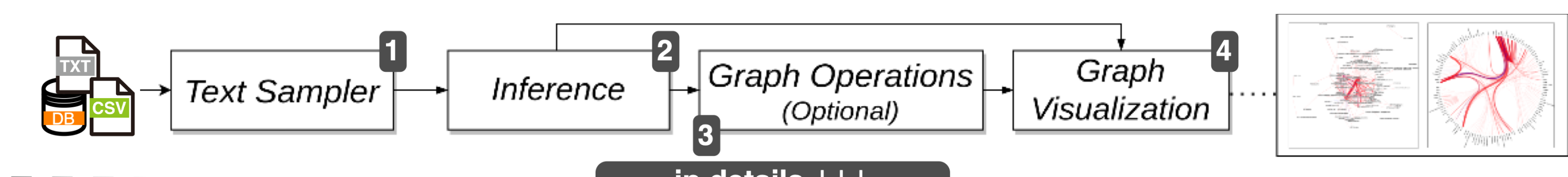
Concept



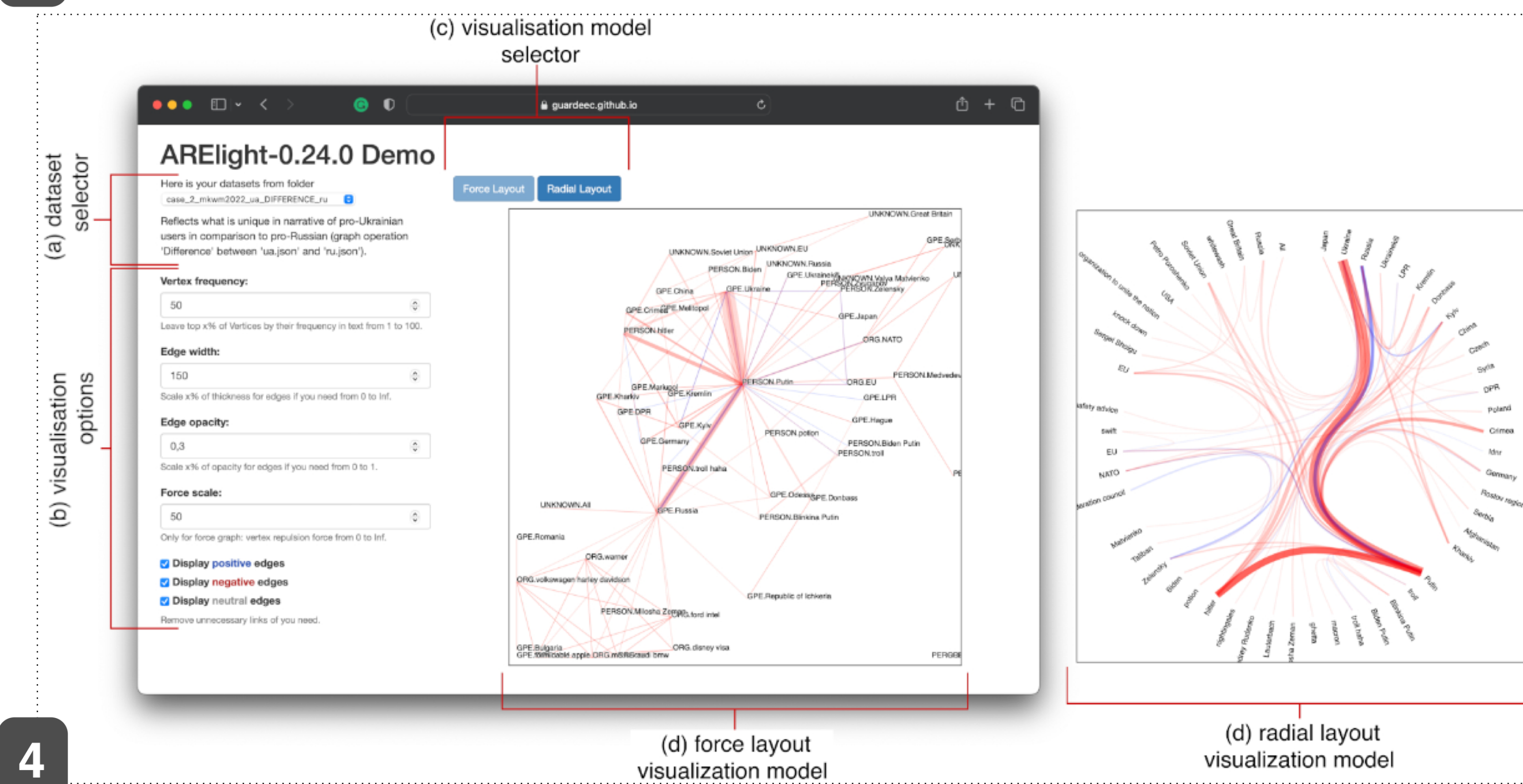
Memory-effective Multi-document Context Sampler

ARElight pipeline includes 4 modules:

- 1 sampling of text entity-pairs $\langle A, B \rangle$ context from large documents;
- 2 extract sentiment between entities $(A \xrightarrow{\text{sent}} B)$ via fine-tuned BERT_{base} model;
- 3 [optional] form graphs from $\{A \xrightarrow{\text{sent}} B\}$ and analyse with graph algorithms;
- 4 interactive web-visualisation with force- & radial-layout graphs.



	G_1		G_2		G
Union $W_e = W_{e1} + W_{e2}$ $W_v = \sum W_v(e)$ e.g. making graph of book from its sentences		\cup		$=$	
Similarity $W_e = \min(W_{e1}, W_{e2})$ $W_v = \sum W_v(e)$ e.g. what is common between 2 X accounts		\cap		$=$	
Difference $W_e = W_{e1} - W_{e2}$ $W_v = \sum W_v(e)$ e.g. unique in user ₁ compared to user ₂		$-$		$=$	



Web Demo

https://guardeec.github.io/areligh_demo/template.html



Demo includes 3 case-studies:

- 1 Analysis of "War & Peace" books by Leo Tolstoy Volumes 1-3.
- 2 VKontakte Social Network 1, 351, 478 Ru/Ua-war user comments.
- 3 Comparison of X/Twitter Rishi Sunak, Boris Johnson, and Elon Musk.

References

- [1] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the ACL*, pages 55–60. ACL, June 2014.
- [2] Xu Han, Tianyu Gao, Yuan Yao, Deming Ye, Zhiyuan Liu, and Maosong Sun. OpenNRE: An open and extensible toolkit for neural relation extraction. In *Proceedings of the 2019 EMNLP-IJCNLP*, pages 169–174. ACL.
- [3] Eunsol Choi, Hannah Rashkin, Luke Zettlemoyer, and Yejin Choi. Document-level sentiment inference with social, faction, and discourse context. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (2016)*.

