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

Nicolay Rusnachenko, Huizhi Liang, Maxim Kolomeets, Lei Shi
{name.surname}@newcastle.ac.uk

Newcastle University

Outline

1. Information Retrieval and Large Texts
2. Existed Systems
3. Processing and Analysis of Large Texts

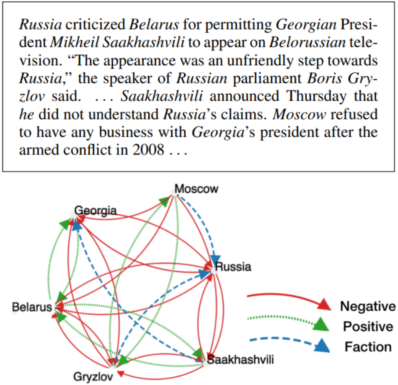
Information Retrieval (IR)

Information Extraction – one of the direction in Natural Language Processing (NLP) aimed on retrieving content from structuring textual information:

- **Objects** (entities, events)
- Establishing **relations** between objects (semantic, sentiment)[1]

[1] Iris Hendrickx et al. "SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals". In: *Proceedings of the 5th International Workshop on Semantic Evaluation*. Uppsala, Sweden: Association for Computational Linguistics, July 2010, pp. 33–38. URL: https://aclanthology.org/S10-1006.

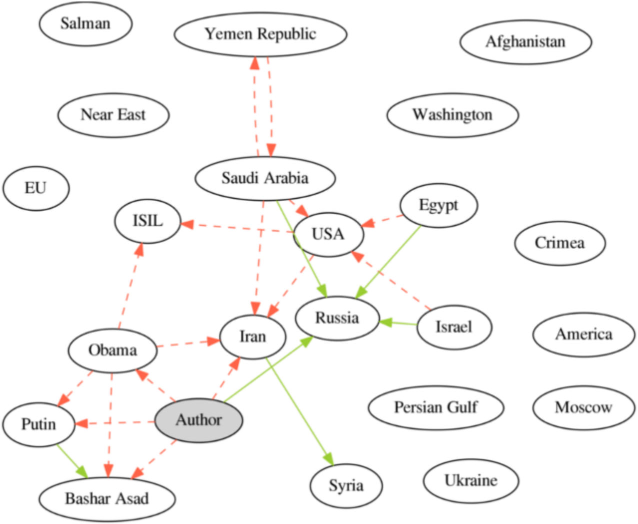
Large Document[2]



Russia criticized *Belarus* for permitting *Georgian* President *Mikheil Saakhashvili* to appear on *Belorussian* television. “The appearance was an unfriendly step towards *Russia*,” the speaker of *Russian* parliament *Boris Gryzlov* said. . . . *Saakhashvili* 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 . . .

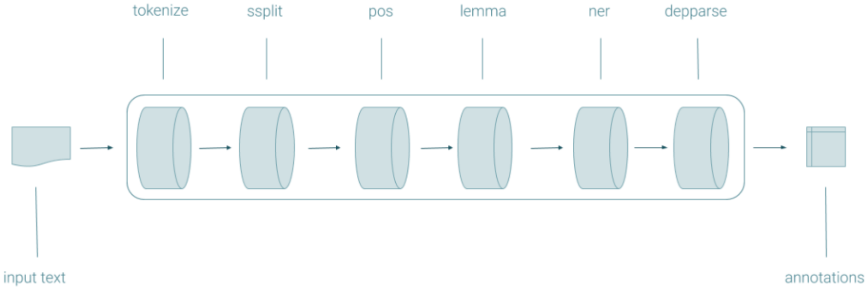
[2] Eunsol Choi et al. “Document-level Sentiment Inference with Social, Faction, and Discourse Context”. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 333–343. DOI: 10.18653/v1/P16-1032. URL: https://aclanthology.org/P16-1032.

Representation and ways to Analyse



Existed Systems

Pipeline-based Concept[3]



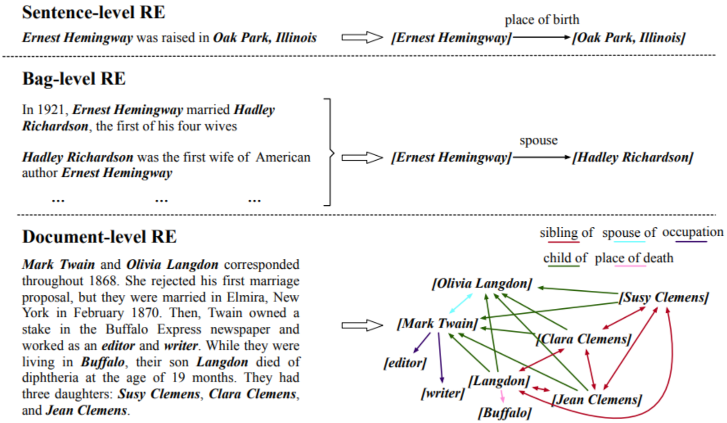
[3] Christopher Manning et al. "The Stanford CoreNLP Natural Language Processing Toolkit". In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 55–60. DOI: 10.3115/v1/P14-5010. URL: https://aclanthology.org/P14-5010.

Target-Oriented Systems

$$\text{TEXT} \rightarrow \text{OBJECTS}$$

$$[\text{TEXT},\ \text{OBJECTS}] \rightarrow \text{RELATIONS}$$

OpenNRE[4]



Sentence-level RE

Ernest Hemingway was raised in ***Oak Park, Illinois***

Bag-level RE

In 1921, ***Ernest Hemingway*** married ***Hadley Richardson***, the first of his four wives

Hadley Richardson was the first wife of American author ***Ernest Hemingway***

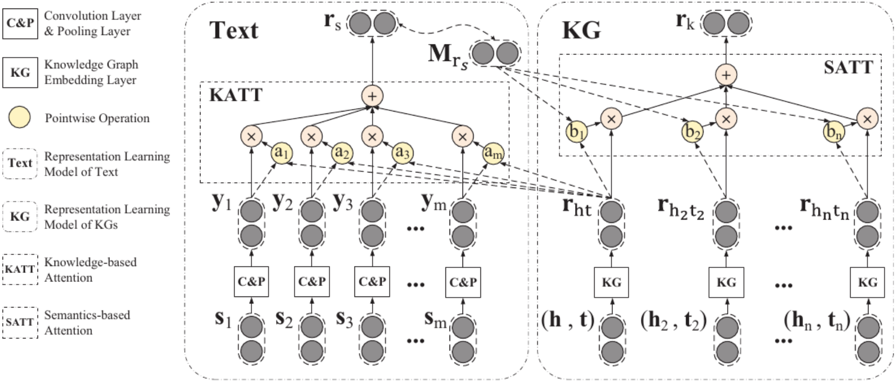
... ...

Document-level RE

Mark Twain and ***Olivia Langdon*** corresponded throughout 1868. She rejected his first marriage proposal, but they were married in Elmira, New York in February 1870. Then, Twain owned a stake in the Buffalo Express newspaper and worked as an ***editor*** and ***writer***. While they were living in ***Buffalo***, their son ***Langdon*** died of diphtheria at the age of 19 months. They had three daughters: ***Susy Clemens***, ***Clara Clemens***, and ***Jean Clemens***.

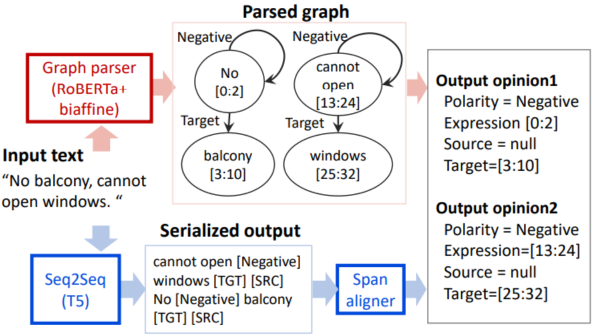
[4] Xu Han et al. "OpenNRE: An open and extensible toolkit for neural relation extraction". In: *arXiv preprint arXiv:1909.13078* (2019).

JointNRE[5]


C&P Convolution Layer & Pooling Layer
KG Knowledge Graph Embedding Layer
Pointwise Operation
Text Representation Learning Model of Text
KG Representation Learning Model of KGs
KATT Knowledge-based Attention
SATT Semantics-based Attention
Text
r_s
M_{r_s}
KATT
a_1
a_2
a_3
a_m
y_1
y_2
y_3
y_m
C&P
s_1
s_2
s_3
s_m
KG
r_k
SATT
b_1
b_2
b_n
r_{ht}
$r_{h_2t_2}$
$r_{h_nt_n}$
KG
(h , t)
(h_2 , t_2)
(h_n , t_n)

[5] Xu Han, Zhiyuan Liu, and Maosong Sun. "Neural Knowledge Acquisition via Mutual Attention between Knowledge Graph and Text". In: *Proceedings of AAAI*. 2018.

T5 graph-based transformer by Hitachi[6]



[6] Gaku Morio et al. "Hitachi at SemEval-2022 Task 10: Comparing Graph- and Seq2Seq-based Models Highlights Difficulty in Structured Sentiment Analysis". In: *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Seattle, United States: Association for Computational Linguistics, July 2022, pp. 1349–1359. DOI: 10.18653/v1/2022.semeval-1.188. URL: https://aclanthology.org/2022.semeval-1.188.

Attention

For input $X \in R^N$:

- $O(N^2)$ original self-attention[7] computation complexity;

How to address this problem:

1. Sparse version of Self-attention;
2. #1 with Global Attention;
3. **Structuring**[8] – limit attention on sentences, paragraphs, etc. via masking.

512 (BERT, T5) → 1K (ETC), 4K/8K/16K (LongT5), 32K (ChatGPT4)

[7] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[8] Joshua Ainslie et al. “ETC: Encoding long and structured inputs in transformers”. In: *arXiv preprint arXiv:2004.08483* (2020).

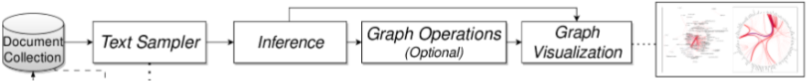
Summary of Limitations

- Pipeline-based:
  - is considered for the **whole document**.[3]
- Target-oriented:
  - Input Size Limitations.[7] (512-32K tokens at present)

Demo

Overall Demo Concept

Text Sampler – extract small portions of text (frames)[9] from (i) large document (samples) and/or (ii) collection of documents[10].

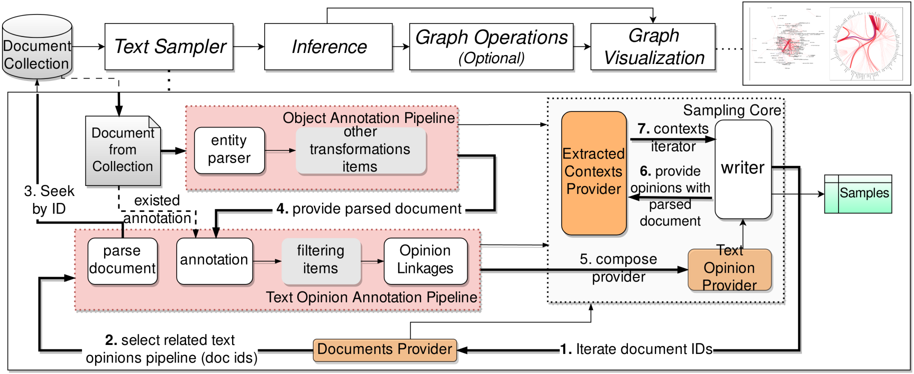


We consider sentiment analysis problem with classes: **positive**, **negative**.

[9] Heike Adel et al. "DERE: A task and domain-independent slot filling framework for declarative relation extraction". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 2018, pp. 42–47.

[10] Adam Roberts et al. "Scaling Up Models and Data with t5x and seqio". In: *arXiv preprint arXiv:2203.17189* (2022). URL: https://arxiv.org/abs/2203.17189.

Architecture of the Sampler and Overall Workflow



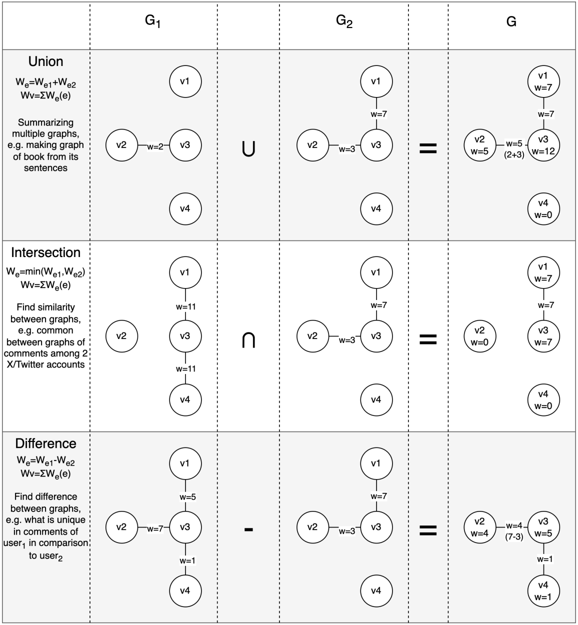
Two declarative pipelines[1] for separate annotation of **Objects** and **Relations**.

[1] https://github.com/nicolay-r/AREkit/wiki/Task-Schemata

Inference

- Sentiment Relation Extraction[11]
- Using `OpenNRE`[4] and BERT-based models as inference.

[11] Nicolay Rusnachenko. "Language Models Application in Sentiment Attitude Extraction Task". Russian. In: *Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS), vol.33*. 3. 2021, pp. 199–222.



Union

$W_e=W_{e1}+W_{e2}$
$Wv=\Sigma W_e(e)$

Summarizing multiple graphs, e.g. making graph of book from its sentences

Intersection

$W_e=\min(W_{e1},W_{e2})$
$Wv=\Sigma W_e(e)$

Find similarity between graphs, e.g. common between graphs of comments among 2 X/Twitter accounts

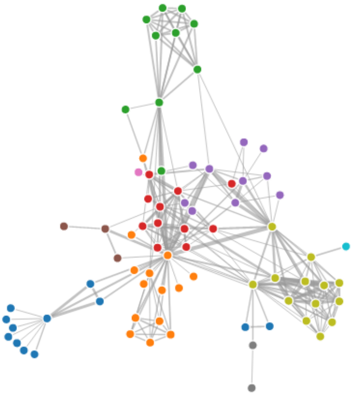
Difference

$W_e=W_{e1}\text{-}W_{e2}$
$Wv=\Sigma W_e(e)$

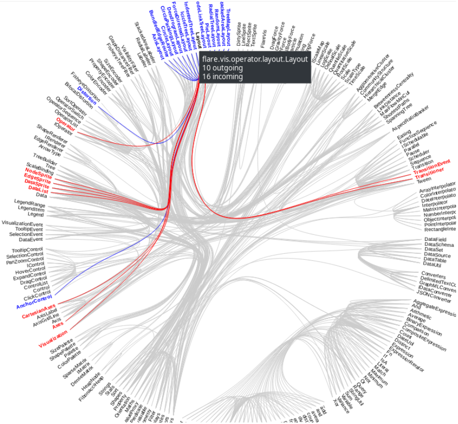
Find difference between graphs, e.g. what is unique in comments of user$_1$ in comparison to user$_2$

Serializing Graphs

Forced[a]



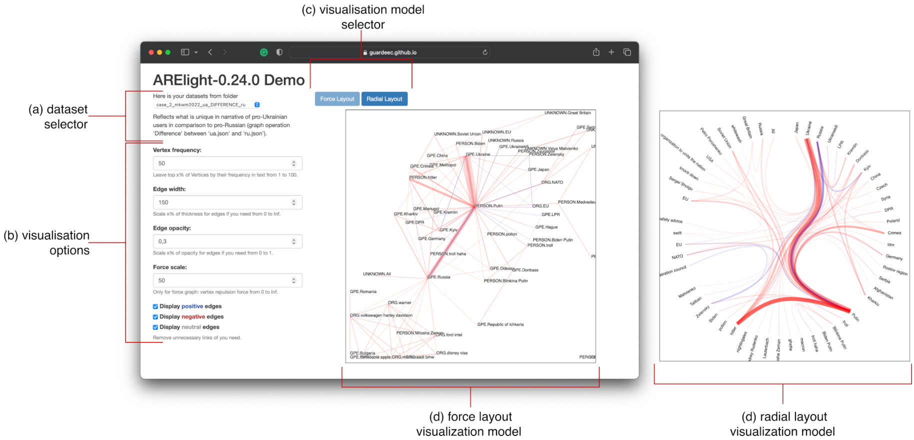
Radial[a]



[a] https://observablehq.com/@d3/force-directed-graph/2

[a] https://observablehq.com/@d3/hierarchical-edge-bundling

System Demo[2]



ARElight-0.24.0 Demo

Here is your datasets from folder
case_2_mkwm2022_ua_DIFFERENCE_ru

Reflects what is unique in narrative of pro-Ukrainian users in comparison to pro-Russian (graph operation 'Difference' between 'ua.json' and 'ru.json').

Vertex frequency: 50

Leave top x% of Vertices by their frequency in text from 1 to 100.

Edge width: 150

Scale x% of thickness for edges if you need from 0 to Inf.

Edge opacity: 0,3

Scale x% of opacity for edges if you need from 0 to 1.

Force scale: 50

Only for force graph: vertex repulsion force from 0 to Inf.

☑ **Display positive edges**

☑ **Display negative edges**

☑ **Display neutral edges**

Remove unnecessary links of you need.

[2] https://guardeec.github.io/arelight_demo/template.html

Thank you for attention!

`https://nicolay-r.github.io`