



Monitoring of AI regulations

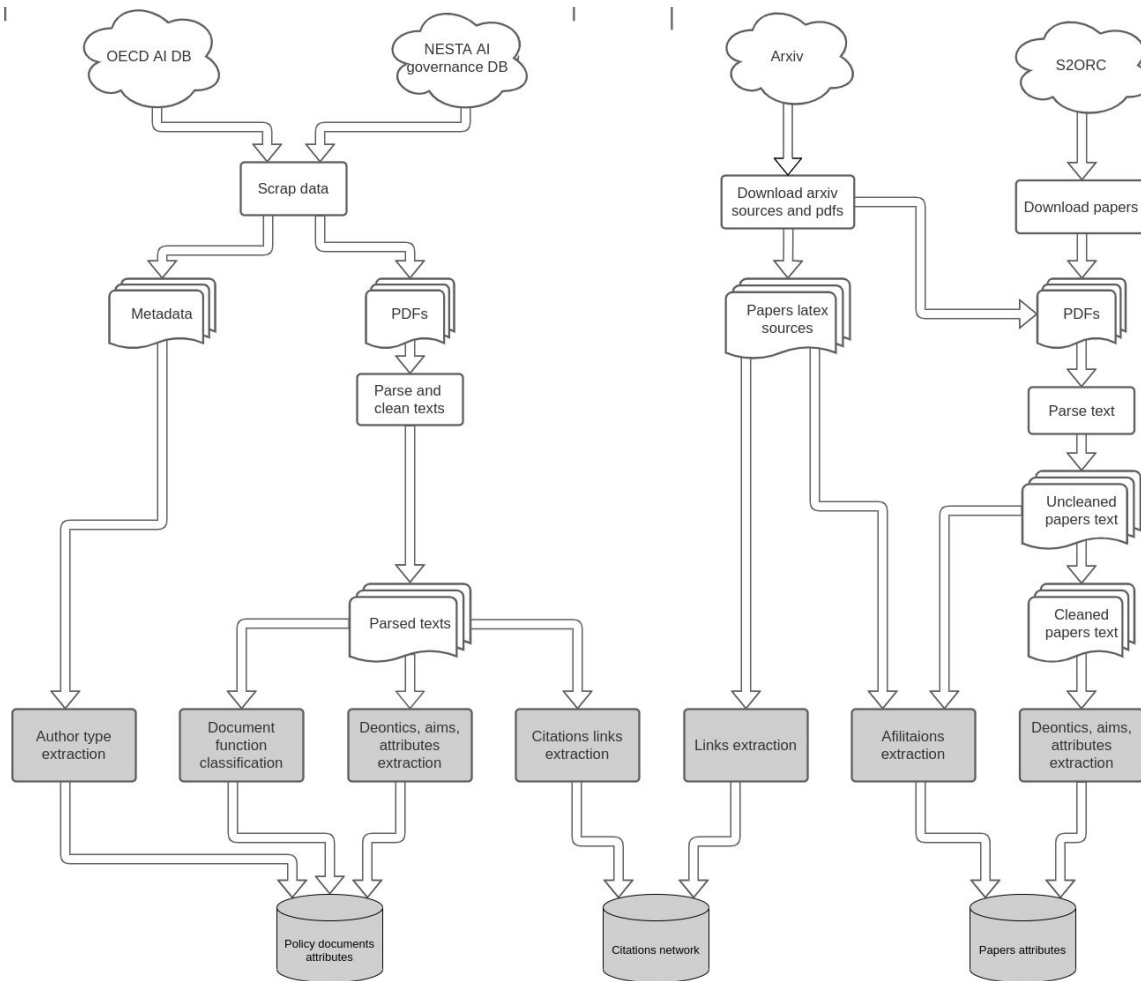
Stanisław Giziński, Hanna Zdulska



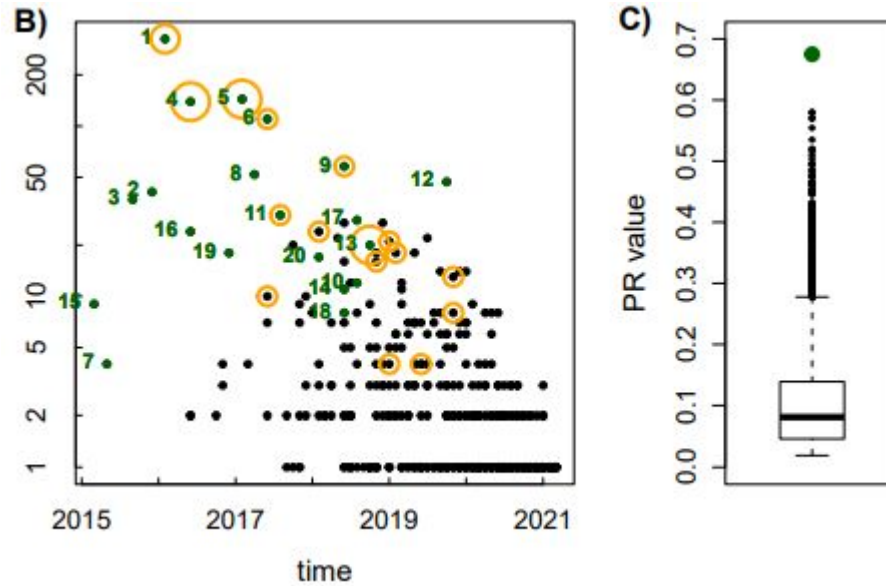
Questions

1. How scientific papers influence public policies?
2. How influence is distributed across different affiliations?
3. How policymakers and researchers talk about AI?

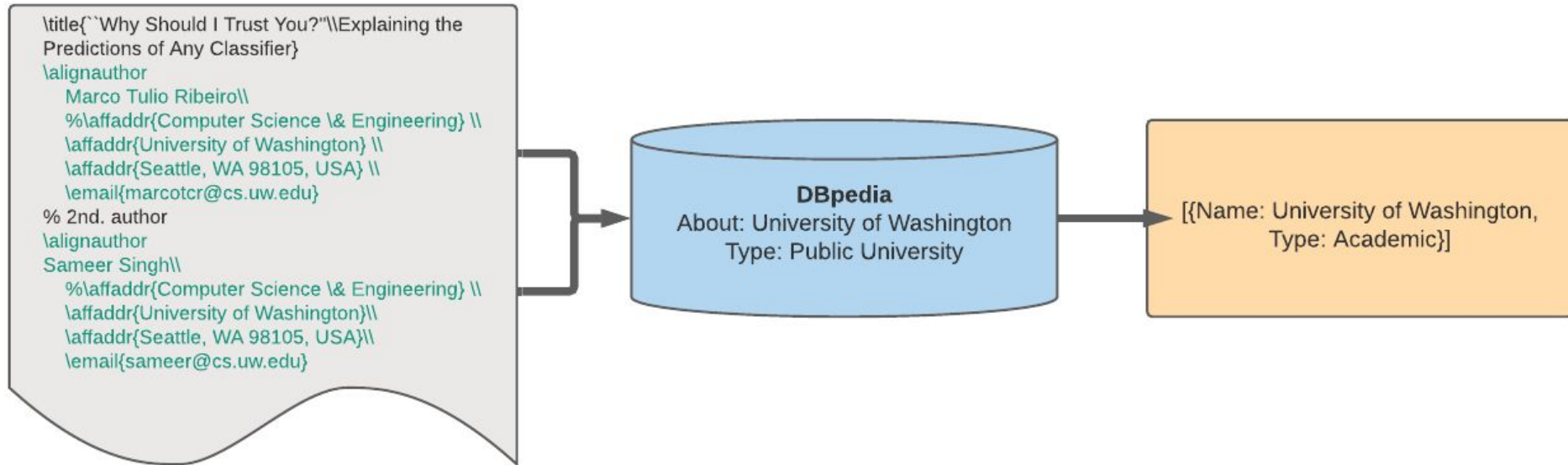




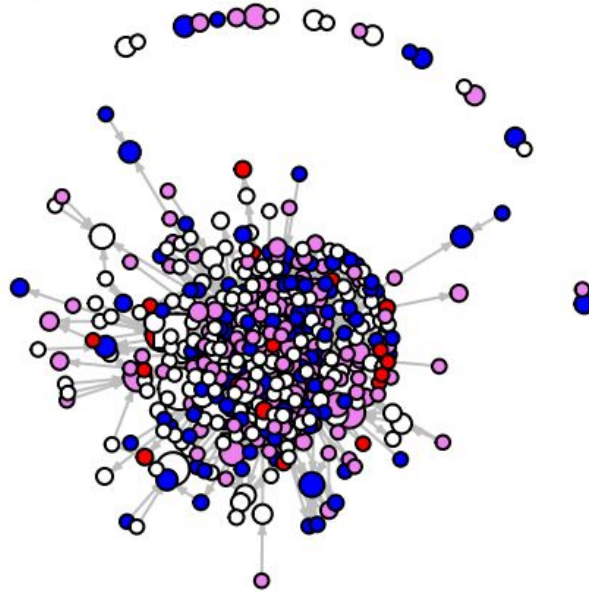
Using Pagerank as a proxy for document importance



Academia vs industry



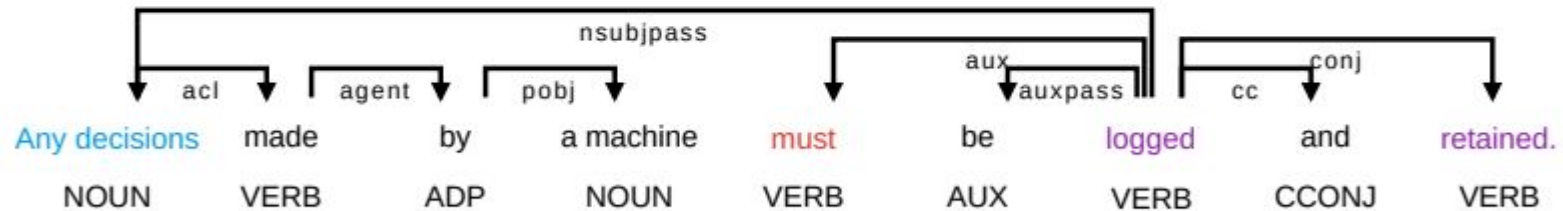
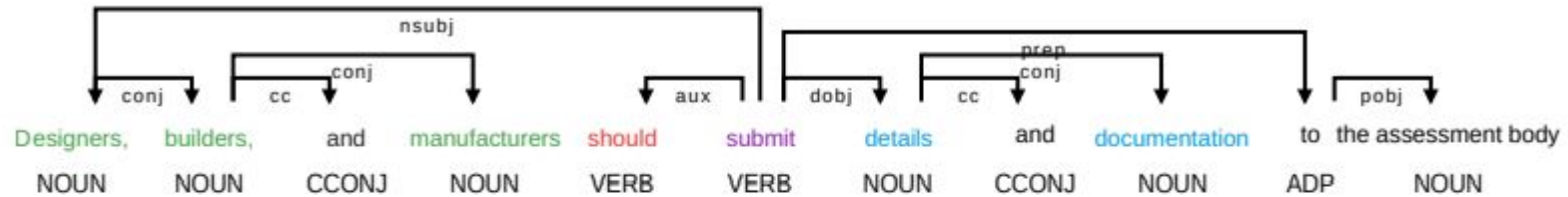
Academia vs industry



● academia
● industry

● academia & industry
○ undefined

Inspiration from Political Science - The Institutional Grammar



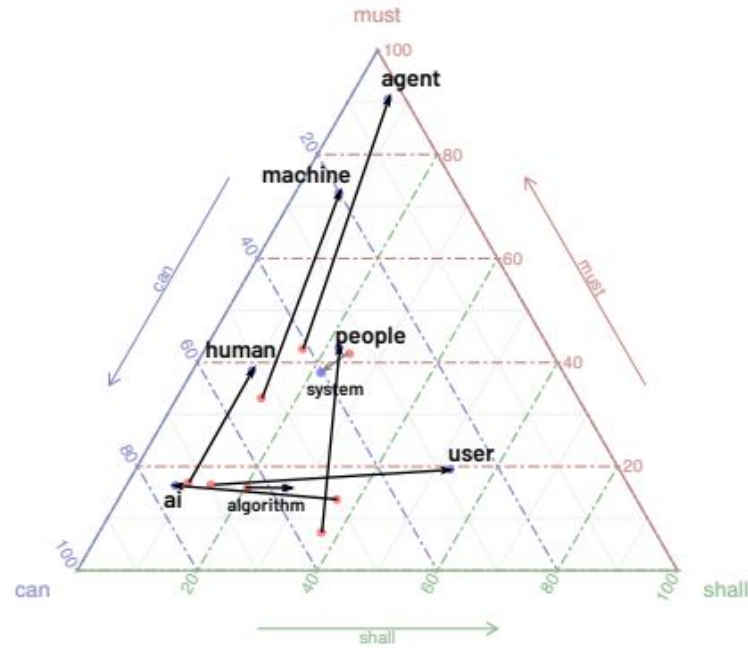
AIM

DEONTIC

ATTRIBUTE

OBJECT

How deontics change between papers and public policies



Challenge - different formats and languages of documents



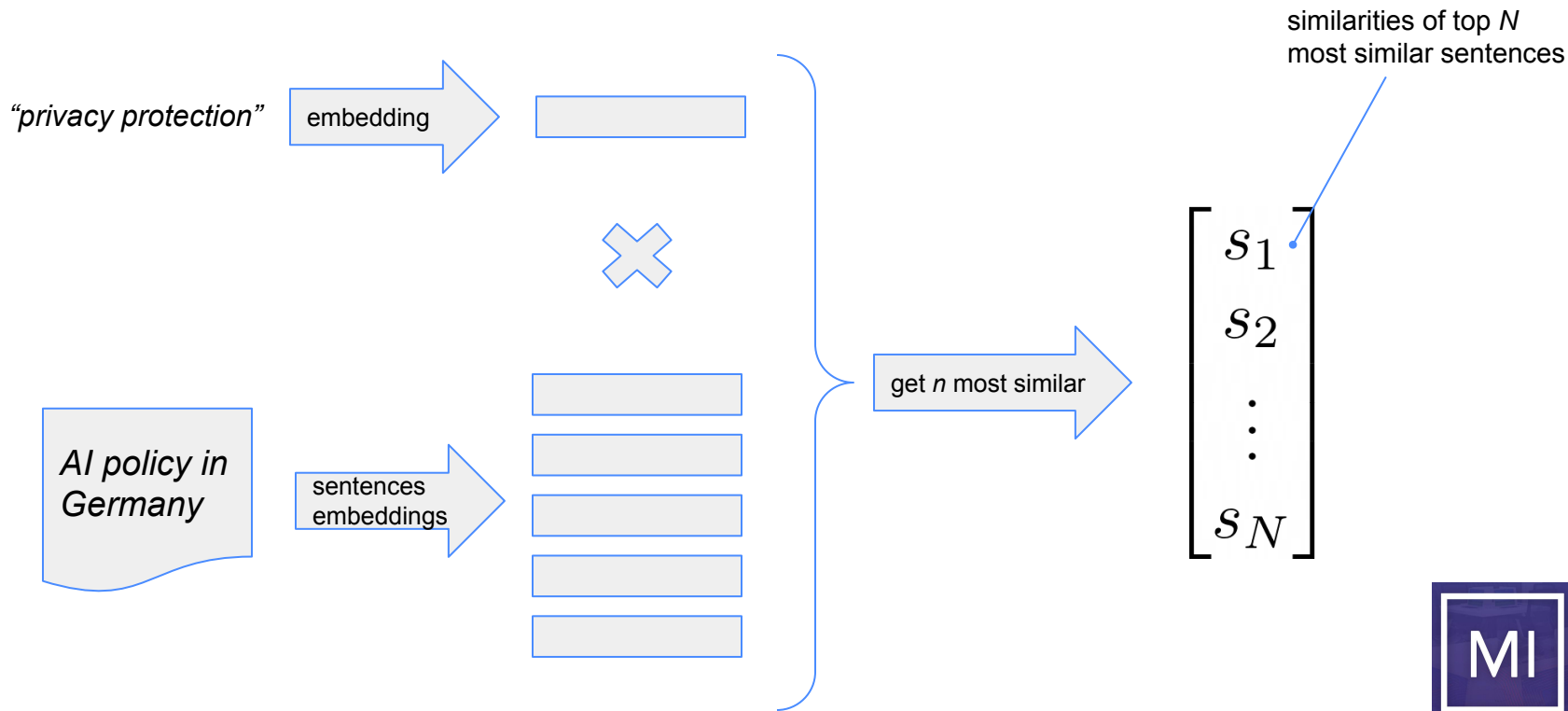
Solutions

1. Utilizing libraries for content extraction from web pages (dragnet, newspaper3k)
2. Language agnostic NLP



[illegible]

Similarity search



Machine-generated analysis

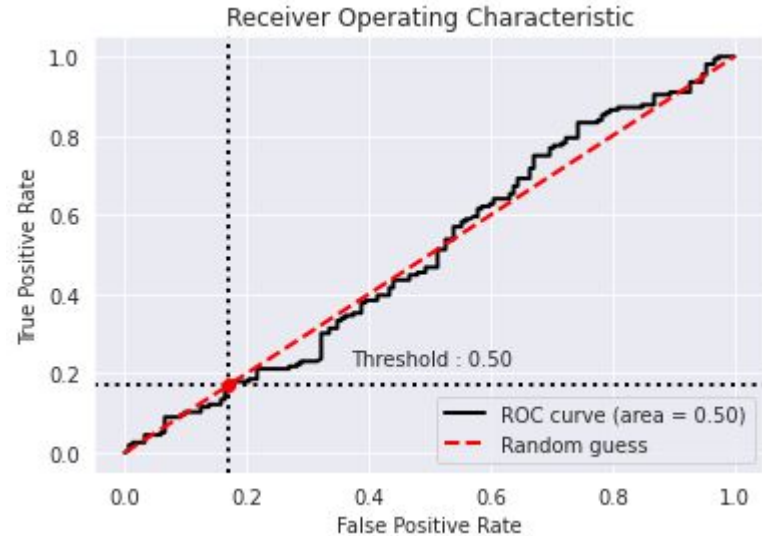
privacy protection	0.42	0.61	0.3	0.33	0.4	0.43	0.67	0.57	0.38	0.36	0.44	0.39	0.23	0.32
accountability	0.57	0.51	0.23	0.29	0.49	0.46	0.33	0.63	0.49	0.45	0.35	0.3	0.24	0.34
safety, cybersecurity	0.44	0.57	0.26	0.4	0.89	0.42	0.36	0.59	0.89	0.44	0.32	0.28	0.31	0.29
human oversight, control, auditing	0.53	0.73	0.3	0.33	0.58	0.37	0.4	0.44	0.4	0.35	0.35	0.33	0.32	0.26
explainability, interpretability	0.62	0.53	0.35	0.41	0.34	0.43	0.69	0.56	0.42	0.34	0.41	0.45	0.29	0.32
legislative framework, legal status of AI systems	0.53	0.41	0.31	0.29	0.28	0.5	0.34	0.41	0.44	0.29	0.34	0.35	0.29	0.3
responsible/intensified research funding	0.53	0.59	0.37	0.31	0.36	0.35	0.42	0.54	0.49	0.42	0.41	0.34	0.32	0.3
dual-use problem, military, AI arms race	0.46	0.46	0.37	0.4	0.35	0.45	0.42	0.51	0.37	0.4	0.39	0.31	0.37	0.3
human autonomy	0.42	0.54	0.28	0.63	0.45	0.42	0.32	0.47	0.45	0.3	0.33	0.23	0.26	0.31
certification for AI products	0.36	0.39	0.3	0.31	0.36	0.43	0.46	0.43	0.32	0.4	0.35	0.44	0.39	0.29
cultural differences in the ethically aligned design of AI systems	0.49	0.5	0.33	0.38	0.46	0.52	0.45	0.62	0.48	0.38	0.45	0.37	0.41	0.36
	0.41	0.38	0.26	0.42	0.34	0.35	0.32	0.47	0.37	0.38	0.41	0.32	0.28	0.32
	0.46	0.34	0.26	0.47	0.28	0.5	0.41	0.44	0.36	0.41	0.35	0.37	0.36	0.42
	0.51	0.52	0.36	0.47	0.43	0.49	0.52	0.62	0.53	0.51	0.45	0.39	0.38	0.41
	0.54	0.41	0.18	0.32	0.4	0.36	0.5	0.46	0.32	0.35	0.34	0.29	0.27	0.23
	0.39	0.45	0.37	0.34	0.37	0.38	0.5	0.48	0.43	0.39	0.39	0.31	0.35	0.37
	0.51	0.55	0.26	0.3	0.37	0.38	0.43	0.59	0.37	0.33	0.35	0.35	0.43	0.33
	0.55	0.42	0.23	0.46	0.42	0.46	0.57	0.7	0.45	0.49	0.49	0.3	0.3	0.29
	0.53	0.5	0.31	0.36	0.58	0.42	0.66	0.53	0.4	0.39	0.45	0.33	0.31	0.31
	0.27	0.31	0.17	0.28	0.3	0.24	0.41	0.37	0.24	0.3	0.39	0.14	0.2	0.3
	0.52	0.51	0.4	0.43	0.5	0.51	0.49	0.64	0.46	0.56	0.46	0.46	0.49	0.43
	0.43	0.4	0.37	0.3	0.3	0.33	0.44	0.42	0.28	0.44	0.41	0.32	0.29	0.33



Poor results, looking further

What to do:

- Better data cleaning
- Reformulation of the queries
- More than top 1 sentences from document



IDEA 1: Train model for assessing if issues is raised in document

similarities of top N most similar sentences

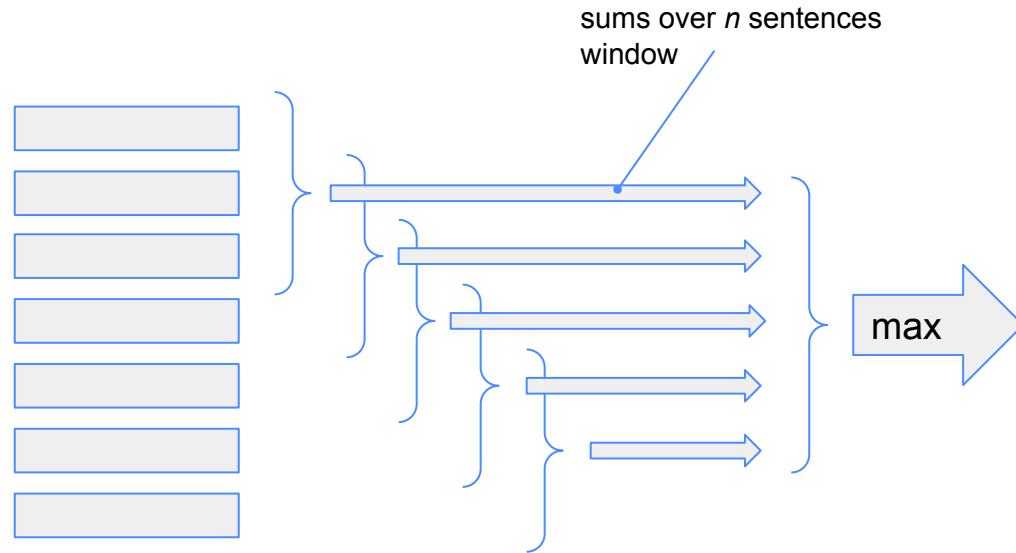
$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix}$$

regression, using each position similarity as feature

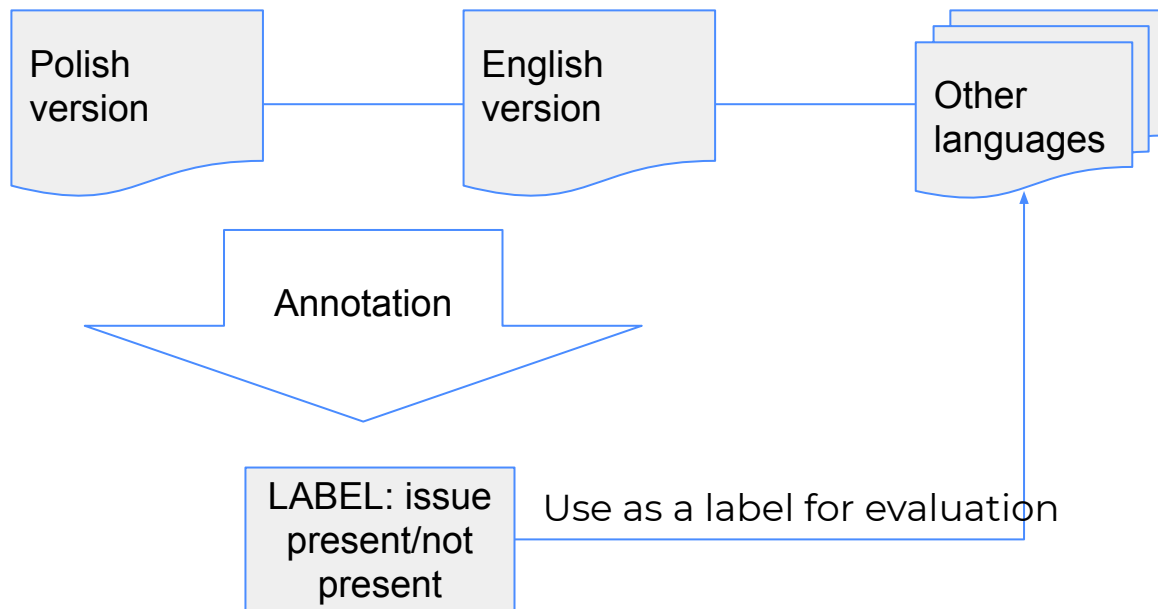
Is issue present in document?



IDEA 2: Aggregated similarity - rolling sum over window of N sentences similarities



How to evaluate it - Eurlex



<https://github.com/ModelOriented/AI-strategies-papers-regulations-monitoring>

<https://github.com/ModelOriented/MAIR>

