

Problem Definition & Contributions

Fact Salience



Research Task

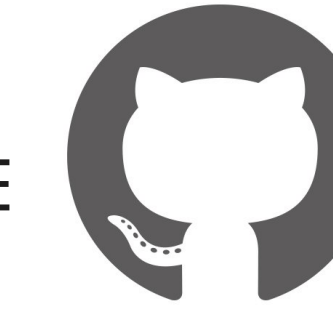
Generating a **machine-readable representation** of the most **prominent** information in a *text document* as a set of **salient open facts**



The 1st
Fact Salience System

Unsupervised and knowledge agnostic, based on **Open Information Extraction**, **PageRank** and **Clustering**

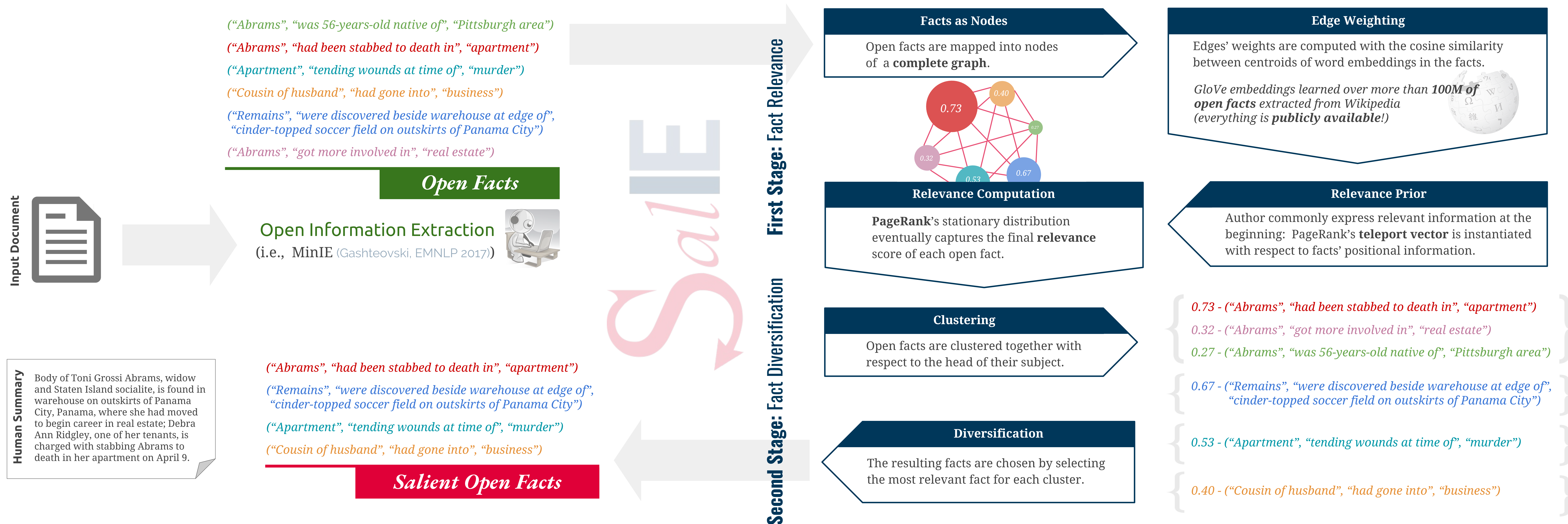
<https://github.com/mponza/SalIE>



Experiments

- **SalIE** outperforms baselines and text summarizers
- Open facts are an effective way to compress information

SalIE - Salient Open Information Extraction



Experiments

Method	ROUGE-1					ROUGE-L				
	1	2	3	4	5	1	2	3	4	5
Position	13.9	20.4	24.8	27.8	29.7	12.8	18.1	21.8	24.4	26.0
	10.1	15.3	19.3	22.3	24.5	9.6	13.8	17.2	19.8	21.7
TextRank	15.2	21.5	24.5	26.1	26.8	13.0	17.5	19.8	21.3	22.0
	11.3	17.2	20.3	22.2	23.3	10.1	14.3	16.7	18.2	19.2
Berkeley	8.5	18.0	25.4	30.4	34.1	8.00	16.3	22.5	26.7	29.7
	10.0	10.6	16.2	21.2	25.4	3.50	9.90	14.8	19.0	22.5
SalIE	17.1	24.2	28.0	30.0	30.9	15.3	21.2	24.3	26.0	26.8
	11.6	17.9	21.6	24.2	25.9	10.5	15.9	19.1	21.3	22.8

Experimental results on a real-world dataset (i.e., New York Times) where we evaluate the top- k extracted facts, with $k \in [1; 5]$. For each system we report the performance with MinIE's *safe* (top score of each row) and *aggressive* (bottom score of each row) modes. *More experiments in the paper!*

References

- Michele Banko, Michael J Cafarella, Stephen Soderland, Matthew Broadhead and Oren Etzioni. **Open Information Extraction from the Web**. In *Proceedings of the 2017 International Joint Conference on Artificial Intelligence, IJCAI 2017*, pages 2670-2676, 2017.
- Luciano Del Corro and Rainer Gemulla. **ClausIE: Clause-Based Open Information Extraction**. In *Proceedings of the 2013 International World Wide Web Conference, WWW 2013*, pages 355-366, 2013.
- Gabriel Stanovsky, Mausam, and Ido Dagan. **Open IE as an Intermediate Structure for Semantic Tasks**. In *Proceedings of the 2015 Annual Meeting of the Association for Computational Linguistics, ACL 2015*, pages 303-308, 2015.
- Kiril Gashteovski, Rainer Gemulla and Luciano Del Corro. **MinIE: Minimizing Facts in Open Information Extraction**. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017*, pages 2630-2640, 2017.