

text2story: A Python Toolkit to Extract and Visualize Story Components of Narrative Text

Evelin Amorim¹ Ricardo Campos^{1,2} Alípio Jorge^{1,3} Pedro Mota³ Rúben Almeida¹

¹ INESC TEC

² University of Beira Interior

³ University of Porto

1 Introduction

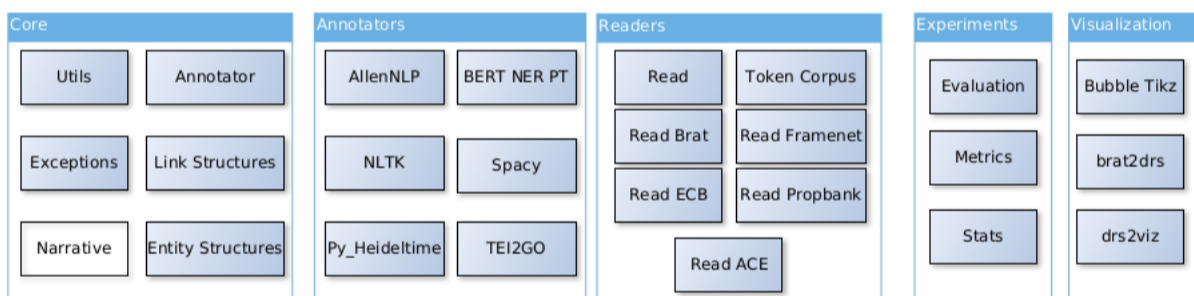
Narrative Definition

A sequence of events that are related to each other. The narrative structure encompasses **events, participants, and time expressions**.

Contributions 🚩

- it annotates text with off-the-shelf models, with an extensible annotator class;
- it reads some well-known annotation file format files;
- it automatizes batch experiments and their evaluations;
- it produces three types of visual representation of annotation, namely, Message Sequence Chart (MSC), Knowledge Graphs (KG), and Bubble Diagrams (BD).

2 The text2story package



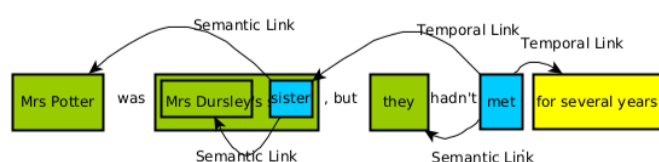
Example

Mrs Potter was Mrs Dursley's sister, but they hadn't met for several years; in fact, Mrs Dursley pretended she didn't have a sister, because her sister and her good-for-nothing husband were as unDursleyish as it was possible to be.

```

1 import text2story as t2s
2
3 narrative_doc = t2s.Narrative("en", doc, "
4     2023") # this is the narrative object
5 participants = narrative_doc.
6     extract_participants("spacy")
7 times = narrative_doc.extract_times("
8     py_heideltime")
9 events = narrative_doc.extract_events("
10    allennlp")
11 semanticrole_links = narrative_doc.
12    extract_semantic_role_links()

```



(a) Human labeling snippet text



(b) Automatic labeling snippet text

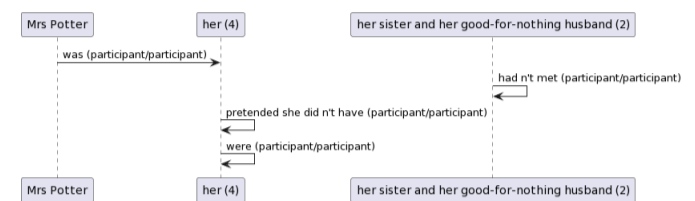


Figure: MSC representation built from the automatic labeling of a sentence of Harry Potter's book.

3 Experiments

Narrative Component	ACE		Lusa News	
	Train	Test	Train	Test
Participants	5,948	37,071	622	2,644
Events	585	3,692	524	2,332
Times	670	3,700	67	338
#token	34,208	213,273	3,707	16,805
#documents	80	455	20	90

		P_r	R_r	$F_{1,r}$
Time	TEI2GO	0.75	0.60	0.64
	Heideltime	0.68	0.53	0.57
	GPT-3	0.61	0.44	0.46
Participants	SRL	0.29	0.02	0.08
	SPACY	0.76	0.25	0.36
	GPT-3	0.68	0.52	0.56
Events	SRL	0.10	0.45	0.15
	GPT-3	0.16	0.079	0.08

		P_r	R_r	$F_{1,r}$
Time	TEI2GO	0.70	0.81	0.73
	Heideltime	0.70	0.80	0.73
	GPT-3	0.82	0.52	0.61
Participants	SRL	0.93	0.15	0.26
	SPACY	0.77	0.33	0.45
	GPT-3	0.70	0.77	0.72
Events	SRL	0.65	0.37	0.68
	GPT-3	0.51	0.71	0.57

Acknowledgements

This work is financed by FCT - Fundação para a Ciência e a Tecnologia, within project LA/P/0063/2020. The author also would like to acknowledge the project StorySense, with reference 2022.09312.PTDC (DOI 10.54499/2022.09312.PTDC).

Pypi

