

# Computing Storylines: Automatic Approaches

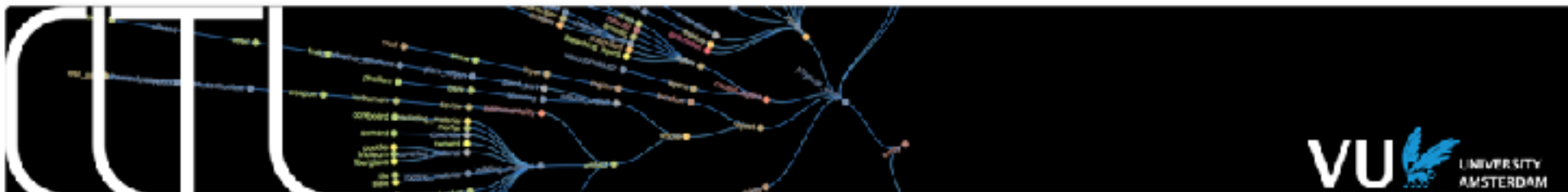
Tommaso Caselli & Piek Vossen  
t.caselli@{gmail.com; rug.nl}  
@tommaso\_caselli

*Storyline Coding and Annotation Workshop @Narrative Matters 2018*  
Enschede, July 02, 2018



university of  
 groningen

faculty of arts



# Overview

- The NWR solution
- PlotProb: Acquiring Plot Links

# The NWR Solution

- An event consists of an **action**, a **place**, a **time** and one or more **actors** (SEM - Van Hage et al., 2011)
- Events can be placed on **timelines**
- A **storyline** is a n-tuple of:
  - an ordered set of time points, **T {t1,t2,...tn}**
  - a set of events, **E {e1, e2,...,en}**
  - a set of explanatory relations, **R {r1, r2,...,rn}**, between the events in E

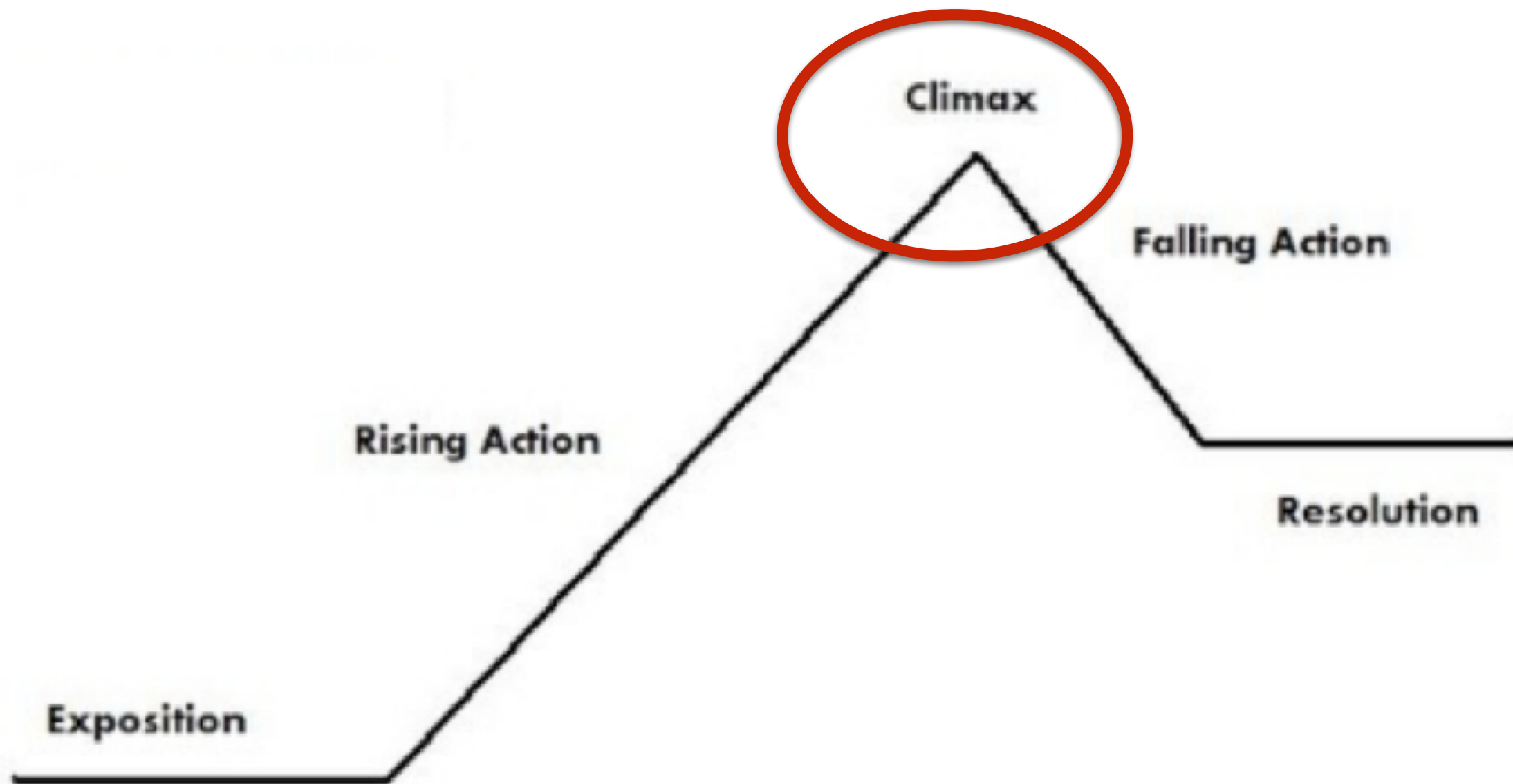
# The NWR Solution

- Storyline maximises the connectivity strength of the explanatory relations across sequences of events
- The connectivity strength can be calculated by means of rising and falling actions:
  - co-participation
  - enablement
  - causality
  - entailment
  - expectation

# The NWR Solution

- Something is missing....

# The NWR Solution



# The NWR Solution

- The climax is the central part of the plot structure
- The climax is the main event around which preceding and following events are connected

# The NWR Solution

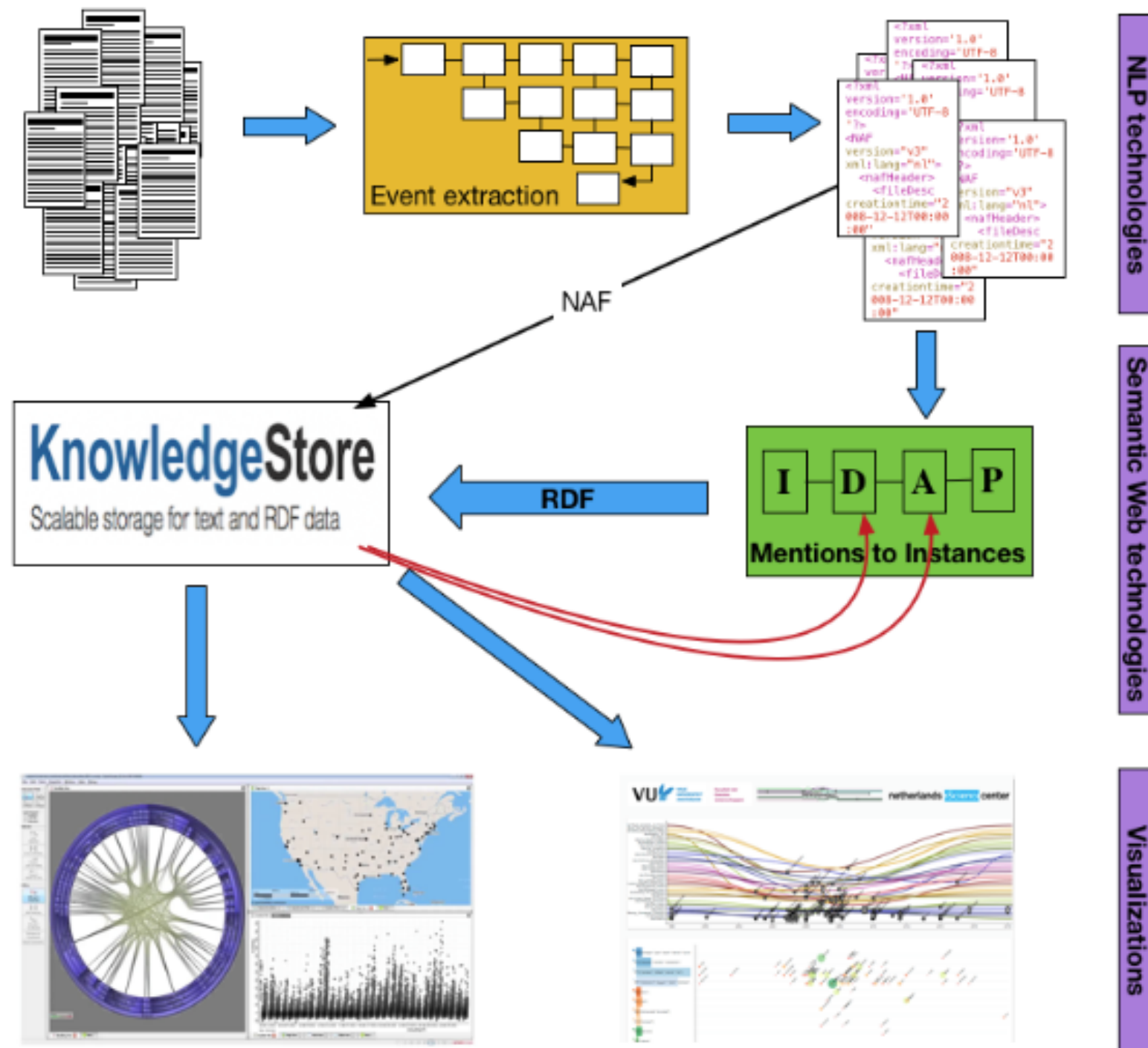
- Each event in a timeline can have a climax score
- The event with the maximum climax score is the starting point for the extraction of a storyline (climax)
- From the climax event, an iterative process can be applied to identify preceding/following events
  - maximise connectivity strength (bridging relations)



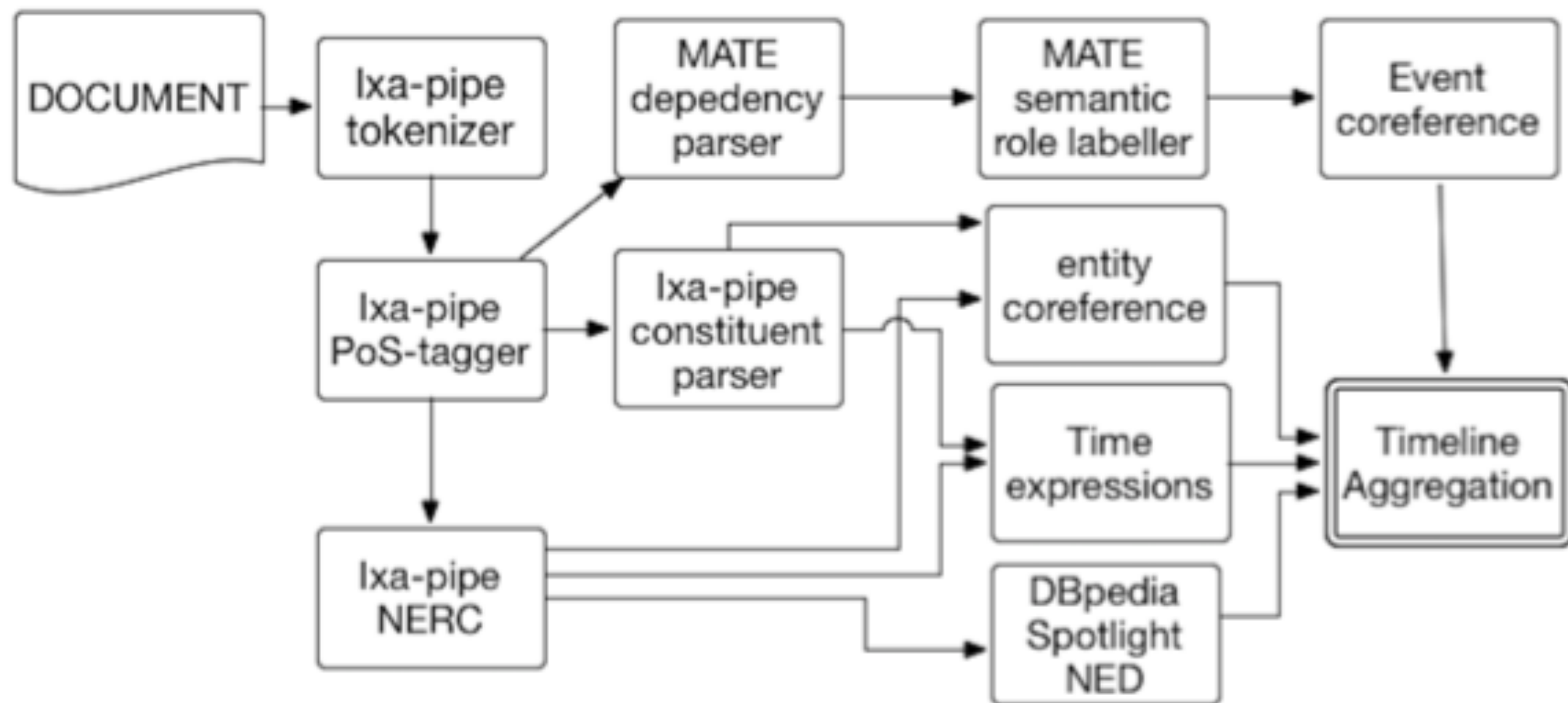
# The NWR Solution

- **Identify** events (*what*), participants (*who*), temporal (*when*) and location (*where*) information
- **Aggregate** data from different sources on the basis of semantic information
- **Reason** over events and participants
- **Visualize** events and participants to reproduce the plot structure

# The NWR Solution



# The NWR Solution



# StoryTeller: demo



<https://www.forbes.com/sites/brianrashid/2015/09/09/how-to-become-the-best-storyteller-on-the-planet/>

<http://nlesc.github.io/UncertaintyVisualization/>

# PlotProb:

## Acquiring Plot Links

- PlotProb: develop a knowledge base of events standing in plot link relations
- Acquiring this information from large collection of news data
  - aggregation per topic
  - collaborator: Wietse de Vries (BA - IK)

# PlotProb:

## Acquiring Plot Links

- Dataset:
  - news and documents about Natural Disasters
  - sources: Wikipedia articles, WikiNews articles, Reuters RCV1 corpus

Data Source	Docs	Lines	Tokens	avg lines/doc	avg tokens/doc
Wikipedia	6,418	131,079	7,587,076	20.4	1,182.2
WikiNews	790	12,180	328,160	15.4	415.4
Reuters RCV1	2,833	30,560	926,381	10.8	327.0
total	10,041	173,819	8,841,617	17.3	880.6

# PlotProb:

## Acquiring Plot Links

- Unsupervised approach
  - use of association measures: Mutual Information (PMI) & Causal Potential (CP)
- Events: restricted to verbs
- Actors: subject and objects
- Pairs of events in adjacent sentences

# PlotProb: Acquiring Plot Links

A hurricane **hit** the coast. Therefore, people had to **evacuate**. Most houses were **destroyed** by the hurricane.

**hit** (hurricane, coast)  
**evacuate** (people, coast)  
**destroy** (hurricane, houses)

Event representation

## Candidate Plot Links

**hit** (hurricane, coast) → **evacuate** (people, coast)  
**evacuate** (people, coast) → **destroy** (hurricane, houses)



# PlotProb:

## Acquiring Plot Links

- Lack of reference corpus for evaluation
- Not possible to use the Event StoryLine Corpus
  - plot links are acquired across documents
  - measures are generalised across documents
- Solution: manual evaluation - validate proposed pairs

# PlotProb: Acquiring Plot Links

## Natural Disaster events

<i>earthquake</i>	<b>STRIKE</b>	<i>country</i>
	▼	
<i>home</i>	<b>SUSTAIN</b>	<i>damage</i>

**Contingent**

**Not Contingent**

**Contingent, but  
wrong order**

<http://thesis.wietsedv.nl/annotator=NAME>

# PlotProb:

## Acquiring Plot Links

Metric	Pairing	Quarter	Arguments	Accuracy	Agreement
PMI	all events	1	yes	0.44	0.69
			no	0.48	0.57
		2	yes	0.28	0.28
			no	0.20	0.31
		3	yes	0.24	0.43
			no	0.36	0.30
	single event	1	yes	0.56	0.76
			no	<b>0.68</b>	0.58
		2	yes	0.32	0.35
			no	0.44	0.38
		3	yes	0.24	0.27
			no	0.64	0.25
CP	all events	1	yes	0.28	0.51
			no	0.36	0.34
		2	yes	0.28	0.49
			no	0.24	0.26
		3	yes	0.16	0.26
			no	0.16	0.57
	single event	1	yes	0.40	0.37
			no	0.28	0.34
		2	yes	0.16	0.27
			no	0.16	0.39
		3	yes	0.28	0.30
			no	0.20	0.33

# PlotProb:

## Acquiring Plot Links

- PMI seems to work better than CP
  - in contrast with previous works!
- Arguments negatively affects the evaluation

Subj	Event 1	Obj	Subj	Event 2	Obj	Contingent
nhc	ARCHIVE	version	person	RETRIEVE	body	Yes
	ARCHIVE			RETRIEVE		No
fire	BURN	quantity	firefighter	BATTLE	blaze	No
	BURN			BATTLE		Yes
loss	MULTIPLY	dike	damage	EXCEED	quantity	Yes
	MULTIPLY			EXCEED		No

# PlotProb:

## Acquiring Plot Links

- Improvements are needed:
  - apply a different methods to aggregate data (data2text)
  - remove Wikipedia pages (??)
- Explore new domains (Man-Made Disasters)
- Investigate document level Plot Links (benchmark corpus for evaluation)