Socio-political events of Conflict and Unrest

Helene Bøsei Olsen

Language Technology Group

April 9 2024

Contents

1. Peace science infrastructure (PSI)
2. UCDP database
3. Event extraction
4. A Survey of Available Datasets

Peace science infrastructure (PSI)

PSI

- Interdisciplinary collaboration:
  - PRIO (Peace Research Institute Oslo)
  - LTG, UiO
  - The Department of Peace and Conflict Research (DPCR), Uppsala University
  - The Department of Political Science, UiO

Peace science infrastructure (PSI)

Motivation

- Social sciences rely on high-quality event data.
  Who did what to whom, where and when
- Currently: a growing number of databases
- Created to answer research questions
- Often annotated by human experts: time-consuming, inefficient use of resources and expensive

Goal

- Develop a robust infrastructure capable of automatically extracting and structuring event data and establishing shared standards and tools to serve as the foundation for a new industry standard in peace research.

Contents

1. Peace science infrastructure (PSI)
2. UCDP database
3. Event extraction
4. A Survey of Available Datasets

UCDP database

Uppsala conflict data program (UCDP)

- Recording armed conflicts
- One of the largest and most widely used event dataset
- Manually annotated by human experts
- News articles as sources

- Huge amount of data
- High quality data
- Closed domain: Armed conflict leading to fatalities at specific time and place.

UCDP database

SANAA (Reuters) - Islamic State's Yemen branch claimed responsibility for a car blast outside a mosque used by the Houthi group on Monday that police sources and medics said killed five people and wounded seven others. The blast hit the Moeed mosque in Sanaa, which is used by the Houthis, the police sources said, and is close to the house of one of their leaders. The Houthis are pushing for more rights for Yemen's Zaydi Shi'ite Muslims, regarded as heretics by the militant Sunni Islamic State.

Side_a	Government of North Yemen
Side_b	Islamic State
Date	2015-07-20
Unknown	5
Low	5
High	5

Contents

1. Peace science infrastructure (PSI)
2. UCDP database
3. Event extraction
4. A Survey of Available Datasets

Event extraction

Event extraction

- Event representations are complex structures describing a change of state, referring to a specific occurrence of something that happens at a particular time and place, involving several participants and other relevant entities [Doddington et al., 2004]

At daybreak on the 9th, the terrorists set off a truck bomb attack in Nazareth.

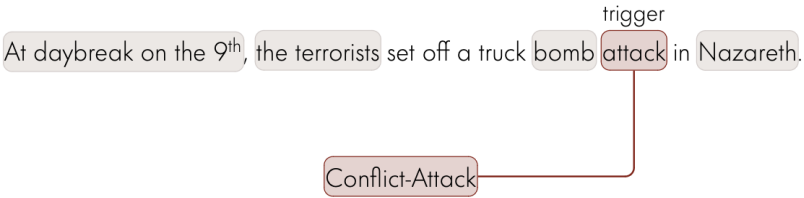
Event extraction

Event trigger

- A word or a phrase that most clearly expresses the occurrence of the event
  [Doddington et al., 2004]

Event arguments

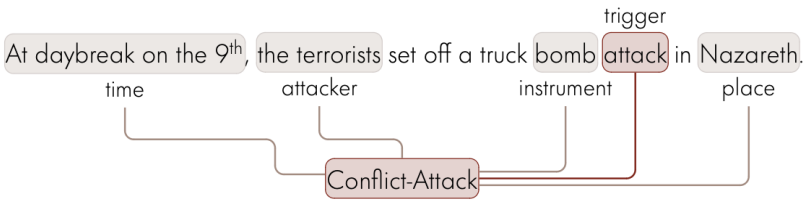
- Entities that are part of the event. Includes participants and attributes
  [Doddington et al., 2004]



Event extraction

Argument role

- The relationship between an event and its participants [Doddington et al., 2004]



Contents

1 Peace science infrastructure (PSI)

2 UCDP database

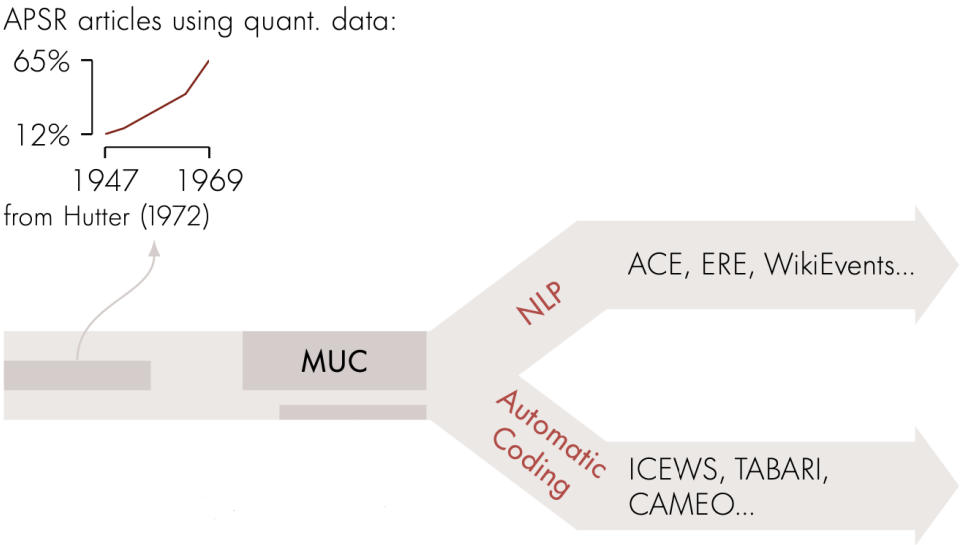
3 Event extraction

4 A Survey of Available Datasets

Socio-political Events of Conflict and Unrest: A Survey of Available Datasets

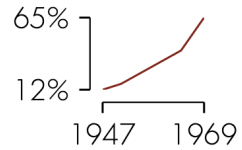
Helene Bøsei Olsen* Étienne Simon* Erik Velldal Lilja Øvrelid
University of Oslo, Language Technology Group

Two separate fields - similar goal



Two separate fields - similar goal

APSR articles using quant. data:

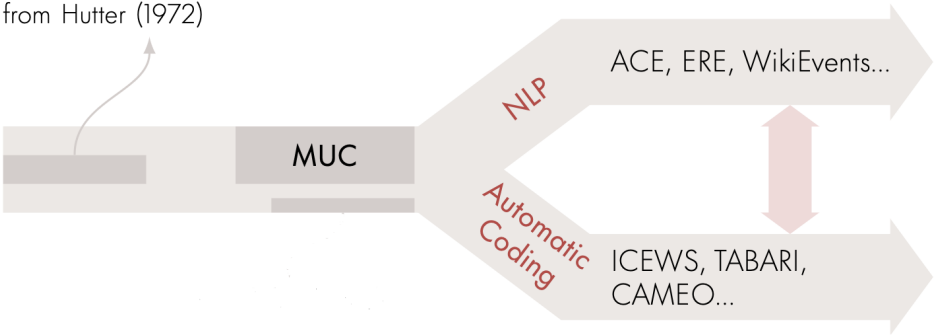


from Hutter (1972)

Our Paper

Jointly describing event datasets:

- in natural language processing
- in political science.



Overview of survey

Scope

Domain of social and political conflict and unrest.
Focus on the most central datasets, while keeping a diversity of approaches.

Overview of survey

Scope

Domain of social and political conflict and unrest.
Focus on the most central datasets, while keeping a diversity of approaches.

NLP Event Extraction Datasets

ACE2005, DocEE, Light&Rich ERE, MAVEN, MEE, MUC-4, WikiEvents. . .

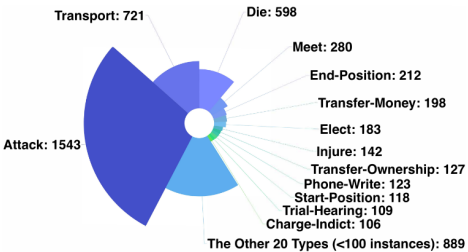
Overview of survey

Scope

Domain of social and political conflict and unrest.
Focus on the most central datasets, while keeping a diversity of approaches.

NLP Event Extraction Datasets

ACE2005, DocEE, Light&Rich ERE, MAVEN, MEE, MUC-4, WikiEvents. . .



Overview of survey

Scope

Domain of social and political conflict and unrest.
Focus on the most central datasets, while keeping a diversity of approaches.

NLP Event Extraction Datasets

ACE2005, DocEE, Light&Rich ERE, MAVEN, MEE, MUC-4, WikiEvents. . .

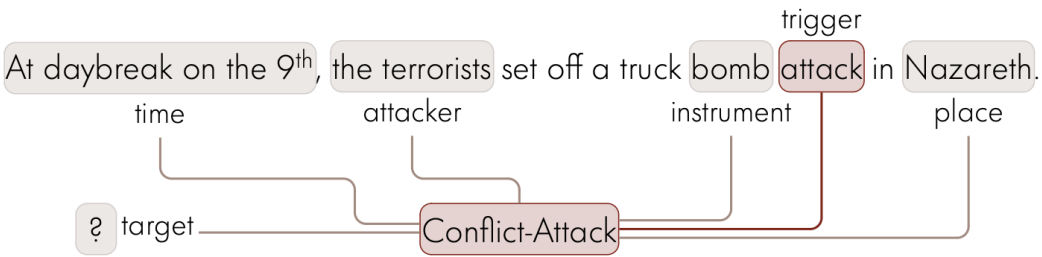
Socio-political Event Databases

Hand-annotated Databases: ACLED, COW, GTD, IBC, MMAD, NAVCO, SCAD, SPEED, UCDP GED, USPVD. . .
Automatically-extracted Databases: GDELT, ICEWS, PHOENIX, POLECAT, TERRIER. . .

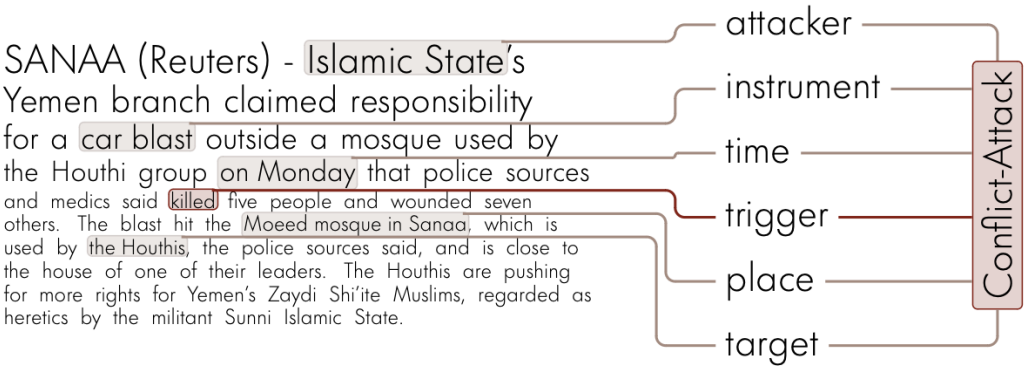
Analysis overview

1. Token-level Annotations
2. Source Availability
3. Abstraction Gap
4. Temporal Dynamic

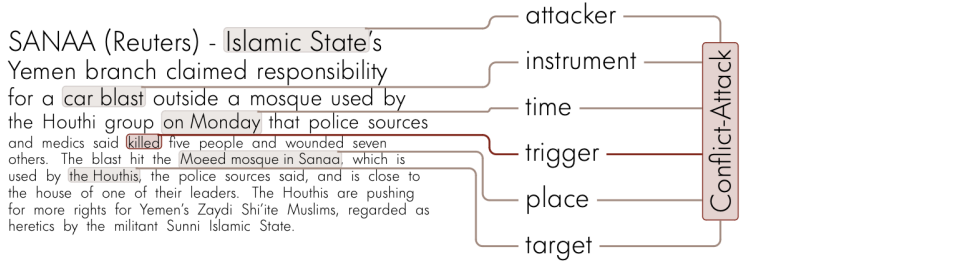
1. Token-level annotations

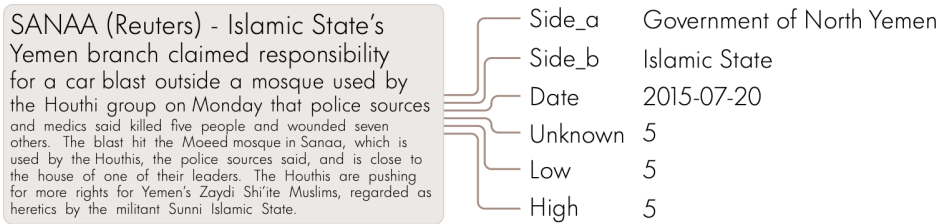


1. Token-level annotations



1. Token-level annotations





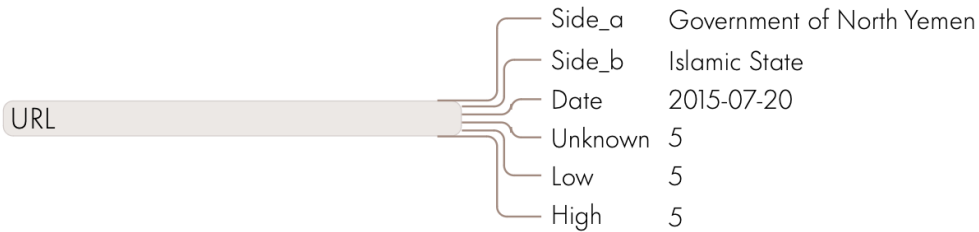
2. Source availability

SANAA (Reuters) - Islamic State's Yemen branch claimed responsibility for a car blast outside a mosque used by the Houthi group on Monday that police sources and medics said killed five people and wounded seven others. The blast hit the Moeed mosque in Sanaa, which is used by the Houthis, the police sources said, and is close to the house of one of their leaders. The Houthis are pushing for more rights for Yemen's Zaydi Shi'ite Muslims, regarded as heretics by the militant Sunni Islamic State.

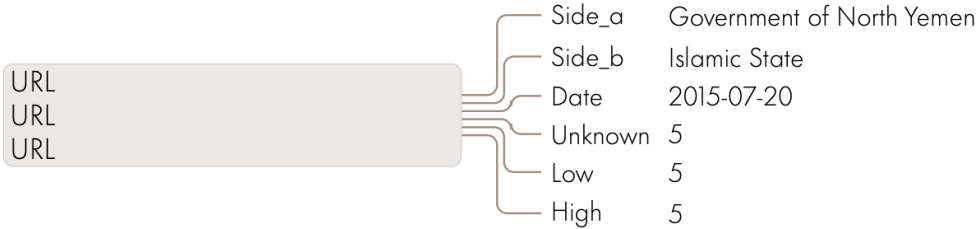
Side_a	Government of North Yemen
Side_b	Islamic State
Date	2015-07-20
Unknown	5
Low	5
High	5

Side_a	Government of North Yemen
Side_b	Islamic State
Date	2015-07-20
Unknown	5
Low	5
High	5

2. Source availability



2. Source availability



3. Abstraction gap

Definition:

Socio-political event databases record what happens in the world.

NLP event extraction datasets record what is in the text.

3. Abstraction gap: Disinformation

THE TIMES

DIEU ET MON DROIT

Wednesday October 11 2023 | thetimes.co.uk | No 74224

£2.80 £2.00 to subscribers
(based on a 7 Day Print and Digital Subscription)

Hamas 'cut the throats of babies' in massacre

parliament, including a generation of new towns, by "bulldozing" through planning laws.

He said that Labour would also build homes on low-quality parts of the green belt, branding them the "grey belt" and a "dreary wasteland".

The Labour leader also defended his plans to borrow to invest, promising that the party would "steer the ship on industrial policy" as he argued that the state must "kick-start growth".

In a direct appeal to moderate Tories, he said: "If you are a Conservative voter who despairs of this, if you look in

3. Abstraction gap: Uncertainty in numbers

REUTERS®

LONDON, Sept 21 (Reuters) - Russia will draft 300,000 reservists to support its military campaign in Ukraine, Defence Minister Sergei Shoigu said on Wednesday in televised remarks.

In Moscow's first update on casualty numbers in almost six months, Shoigu said 5,937 Russian soldiers had been killed since the start of the conflict.

3. Abstraction gap: Uncertainty in language

Four fragile lives found ended in evacuated Gaza hospital

BY MIRIAM BERGER, EVAN HILL AND HAZEM BALOUSHA

JERUSALEM — The nurse in the

Sarah Davies, a spokeswoman for the International Committee of the Red Cross in Jerusalem, said the agency made no guarantees and could not safely reach the

wall" through broken buildings to reach the medical complex.

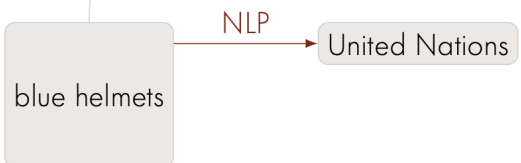
As he approached the neonatal intensive care unit, he said, he "started to smell a foul odor." He

3. Abstraction gap: Contextual world-knowledge

On October 17, 2022, 10 km from the Amachach camp (Tessalit) an IED exploded against blue helmets.

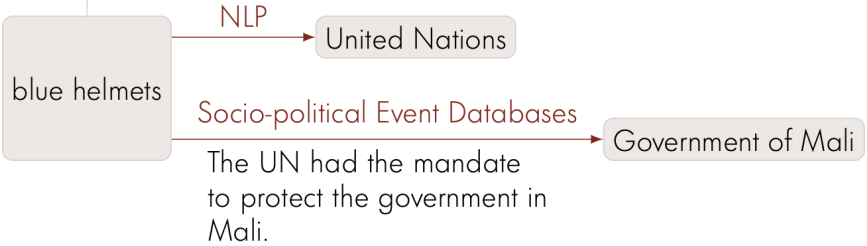
3. Abstraction gap: Contextual world-knowledge

On October 17, 2022, 10 km from the Amachach camp (Tessalit) an IED exploded against blue helmets.



3. Abstraction gap: Contextual world-knowledge

On October 17, 2022, 10 km from the Amachach camp (Tessalit) an IED exploded against blue helmets.



4. Temporal dynamic

In 2013

The security chief of Al-Oshah District in Amran was killed by fighters of Al-Huthi fighters on Wednesday.

- Government of Yemen
- Al-Houthis

4. Temporal dynamic

In 2013

The security chief of Al-Oshah District in Amran was killed by fighters of Al-Huthi fighters on Wednesday.

- Government of Yemen
- Al-Houthis

In 2016

Yemen's Houthi movement accused a Saudi-led coalition of launching air strikes that killed seven people on Sunday.

- Government of Yemen
- Forces of Hadi

Bridging the gap

Machine Coding efforts

Not a bridge:

- rely on rule-based models not using token-level annotations
- ignore the abstraction gap, overcounting events
- source text is not available

Bridging the gap

Machine Coding efforts

Not a bridge:

- rely on rule-based models not using token-level annotations
- ignore the abstraction gap, overcounting events
- source text is not available

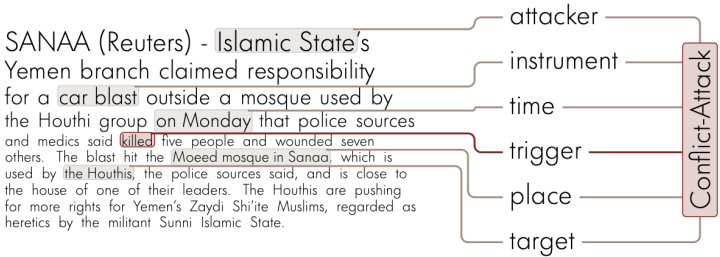
Notable efforts

- IBC-C, GLOCON & CASE shared tasks

Challenges

Document-level event extraction

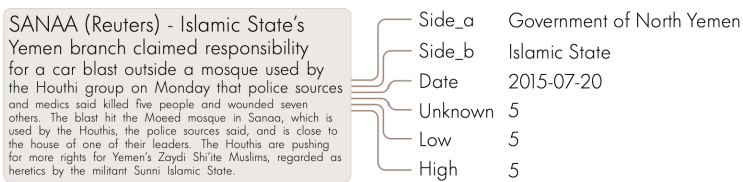
- Majority of current techniques for event extraction centre on extracting events at the sentence-level
- Available datasets on Document-level: DocEE, WikiEvents..



Challenges

Lack of text-span annotations

- How to leverage UCDP data for training event extraction models?
- Some recent efforts use only record-level annotations formulated as a constrained generative event extraction task [Lu et al., 2021, Wu et al., 2023]



Challenges

Evaluation

- Some errors more severe than others

More severe:

- The government of Iceland attacked China

Less severe:

- Output: Time = April 2023
- Gold: Time = 15. April 2023

References I

- Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S., and Weischedel, R. (2004).
  The automatic content extraction (ACE) program – tasks, data, and evaluation.
  In Lino, M. T., Xavier, M. F., Ferreira, F., Costa, R., and Silva, R., editors, *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).
- Lu, Y., Lin, H., Xu, J., Han, X., Tang, J., Li, A., Sun, L., Liao, M., and Chen, S. (2021).
  Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction.
  In Zong, C., Xia, F., Li, W., and Navigli, R., editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2795–2806, Online. Association for Computational Linguistics.

References II

Wu, T., Shiri, F., Kang, J., Qi, G., Haffari, G., and Li, Y.-F. (2023).
Kc-gee: knowledge-based conditioning for generative event extraction.
World Wide Web, pages 1–17.