

Online-Monitoring of Security-Related Events

Analysing multilingual media reports for crisis assessment

What does the application do?

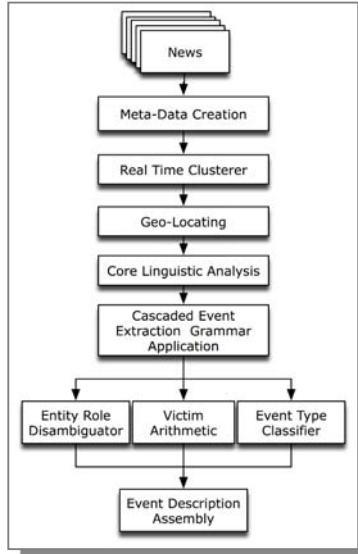
- Continuously **collect** multilingual media reports from web portals.
- Every 10 minutes, **cluster** the latest articles about the same event or subject ('live clusters').
- Identify and extract information on events** of the types: **violence**, **natural disaster** and **humanitarian crisis**.
- Display** the latest events on a map.
- Give access to extracted information and to full articles.



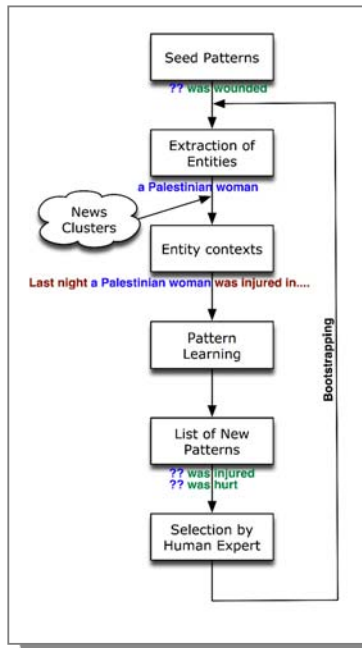
Features

- Languages:** English, Italian and French. More planned for 2008 and 2009.
- Shallow analysis**, simple (1 or 2-slot) extraction rules.
- Information aggregation** at cluster level, resolving contradictory information (e.g. on number of victims).
- Geo-tagging** of events: **place name disambiguation** using occurrence frequency in cluster, size of place and place name hierarchy (country, region, province, city).
- Event classification** is based on keywords, an event taxonomy, Naive Bayes classifiers and domain-specific rules.
- Pattern learning:** seed patterns extract information (e.g. victims); algorithm searches for paraphrases within cluster, exploiting redundancy in news reporting.
- Efficiency:** capable of processing large amounts of data in real time.

The processing chain



Pattern learning



Tanev, H. and P. Ozden-Wennerberg (in print). *Learning to Populate an Ontology of Violent Events*. In: Mining Massive Data Sets for Security. IOS Press.

Tanev, H., J. Piskorski, and M. Atkinson (2008). *Real-Time News Event Extraction for Global Crisis Monitoring*. Proc. of NLDB 2008, Springer LNCS 5039, 207-218.

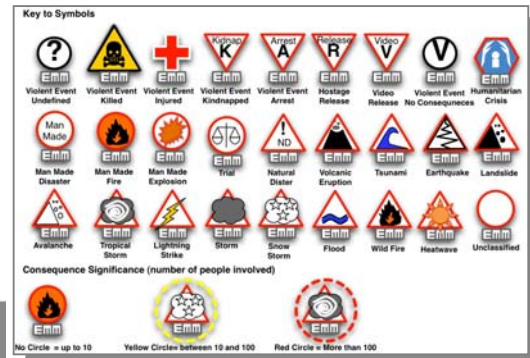
Piskorski, J., H. Tanev, M. Atkinson, and E. Van der Goot (2008). *Cluster-centric Approach to News Event Extraction*. Proc. of NISSI 2008.

Piskorski, J. (2007). *ExPRESS Extraction Pattern Recognition Engine and Specification Suite*. Proc. of FSMNLP 2007.

Atkinson M., Piskorski J., Pouliquen B., Steinberger R., Tanev H., & Zavarella V. (2008). *Online-monitoring of security-related events*. In Proc. of CoLing'2008.

Extracted information

- Victims: Number, Status** (injured, killed, kidnapped, ...), **Description** (e.g. *3 Palestinian women*)
- Perpetrators** (e.g. militant groups)
- Place** of the event
- Event type and weapons** (e.g. *bomb, missile, knife*)
- Event types:



Pattern engine / Formalism

- EXPRESS is a blend of JAPE (GATE) and XTDL (SProUT).
- LHS of the rule is a regular expression over flat feature structures.
- RHS specifies the output structure.
- Allows variables, labels, functional operators, grammar cascading, ...
- Multiple and nested labels (multiple actions)
- Rule sample:

```

injury-event -> ((person-group & [NAME: #name1, NUMBER: #num1]):injured1
token & [SURFACE: "and"]
(person-group & [NAME: #name2, NUMBER: #num2]):injured2
injured-phrase & [FORM: "passive"]
):event

-> injured1: victim & [NAME: #name1, NUMBER: #num1],
injured2: victim & [NAME: #name2, NUMBER: #num2],
event: injury & [VICTIM: #name, NUMBER: #count]
& #name = Concatenate(#name1, "and", #name2)
& #count = EstimateNumber(#num1, #num2).
  
```