

# Tutorial on Socio Political Event Extraction

**Hristo Tanev**

European Commission, Joint Research Centre

via Enrico Fermi 2749

Ispra, Italy

hristo.tanev@ec.europa.eu

## Abstract

The purpose of this tutorial is to introduce researchers and practitioners into the fast developing technology of event extraction. It is directed towards people who have some experience in natural language processing and little or no experience with the event detection and extraction technology. The tutorial will present state of the art approaches and resources in this area.

## 1 Introduction

The purpose of this tutorial is to introduce researchers and practitioners into the technology of event extraction. It is directed towards people who have some experience in other areas of NLP, for example information extraction, text classification, etc., but little or no experience with event detection and extraction. The tutorial will encompass presentation and a live Web demo of an event extraction system NEXUS.

The tutorial will present the architecture of a state-of-the-art event extraction system for detecting socio political events:

1. sources of information
2. the structure of the output event template
3. the stages of event extraction
4. the main problems of event extraction

We will make overview of the most important problems of event detection and extraction and the approaches for their solution.

The tutorial will present as an example a real life system, NEXUS, part of the Europe Media Monitoring project, which analyses socio political events, crimes, and disasters, and which works in several European languages.

We will give an overview with real life examples of linguistic rules like cascade finite state grammars

and regular expressions, as well as lexical resources for parsing of event specific information.

The tutorial will also give overview of the recent approaches in machine learning (ML) for classification of sentences and documents into event classes, as well as the use of ML for extraction of event arguments.

Finally, we will review existing public datasets of socio-political events, which can be used for training of event detection ML classifiers.

## 2 Audience

The tutorial will be directed towards computational linguists and NLP practitioners, who are involved or interested in the development of event extraction systems in the socio-political and crisis domain. Although there will be no practical exercises, the tutorial is aimed at researchers with interests in real world applications, rather than purely theoretical interests.

The tutorial audience is expected to have idea about the key concepts in Natural Language Processing (NLP), including evaluation measures and levels of automatic analysis. Additionally, participants are expected to have a rudimentary understanding of news media analysis and the content typically associated with conflict, unrest, and disasters.

## 3 About the presenter

Hristo Tanev is a researcher at the Text Mining Competence Centre at the Joint Research Centre. His main research is in the area of event extraction. He is a co-organizer of the Workshop on Challenges and Applications of Automated Extraction of Socio Political Events from Text (CASE), which has been collocated with EMNLP, ACL, RANLP, LREC and other prestigious conferences in NLP. He has also developed the NEXUS event extraction system, which is part of the Europe Media Monitor.

Hristo Tanev has worked in the Istituto Trentino di Cultura in the period 2001-2006, where he actively researched the topic of open domain Question Answering and worked on a PostDoc project "MoreWeb - Multilingual Question Answering on the Web".

More information on Hristo's current activities and projects is shown on his LinkedIn page <https://www.linkedin.com/in/hristotanev/>

## 4 Tutorial schedule

The tutorial will consist of two 45 minute long sessions.

Session 1: Approaches and overview of socio-political (SP) event extraction

- Overview of SP event extraction
- Live demo of Europe Media Monitoring and NEXUS
- Linguistic rules for event extraction and event detection
- Algorithms for learning rules and lexica

Session 2: Event classification and event extraction resources

- ML methods for event classification: text classification, using ML and deep learning
- Event taxonomies
- Event databases for conflict: ACLED, GDELT, etc.
- Annotated event corpora: ACE, EMM event corpus

## 5 Literature

Participants in the tutorial can get acquainted to survey works on event extraction, such as (Hogenboom et al., 2011), (Xiang and Wang, 2019).

We are going to use through the presentations examples from the NEXUS event extraction system, a description of this system is presented in (Tanev et al., 2008).

The tutorial will also look into event databases, such as ACLED (Raleigh et al., 2010), (Raleigh et al., 2012).

## References

- Frederik Hogenboom, Flavius Frasincar, Uzay Kaymak, and Franciska De Jong. 2011. An overview of event extraction from text. *DeRiVE@ ISWC*, pages 48–57.
- Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joachim Carlsen. 2012. Armed conflict location and event dataset (acled) codebook. *Center for the Study of Civil War, International Peace Research Institute, Oslo (PRIO)*.
- Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. Introducing acled: An armed conflict location and event dataset. *Journal of peace research*, 47(5):651–660.
- Hristo Tanev, Jakub Piskorski, and Martin Atkinson. 2008. Real-time news event extraction for global crisis monitoring. In *Natural Language and Information Systems: 13th International Conference on Applications of Natural Language to Information Systems, NLDB 2008 London, UK, June 24-27, 2008 Proceedings 13*, pages 207–218. Springer.
- Wei Xiang and Bang Wang. 2019. A survey of event extraction from text. *IEEE Access*, 7:173111–173137.