On the Creation of a Security-Related Event Corpus

Martin Atkinson, Jakub Piskorski, Hristo Tanev, Vanni Zavarella
Text and Data Mining Unit, Directorate I: Competencies
Joint Research Centre of the European Commission
Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
`firstname.lastname@ec.europa.eu`

Abstract

This paper reports on an effort of creating a corpus of structured information on security-related events automatically extracted from on-line news, part of which has been manually curated. The main motivation behind this effort is to provide material to the NLP community working on event extraction that could be used both for training and evaluation purposes.

1 Introduction

Due to a rapid proliferation of textual information in digital form various security-related organisations have recently acknowledged the benefits of deploying techniques to automate the process of extraction of structured information on events from free texts (Appelt, 1999; Ashish et al., 2006; Ji et al., 2009; Piskorski and Yangarber, 2013). Examples of current capabilities of such techniques for the extraction of disease outbreaks, crisis situations, cross-border crimes and computer security events from on-line sources are given in (Grishman et al., 2002; King and Lowe, 2003; Tanev et al., 2008; Yangarber et al., 2008; Atkinson et al., 2011; Gao et al., 2013; Danilova and Popova, 2014; Ritter et al., 2015).

This paper reports on the creation of a corpus of structured information on security-related events automatically extracted from online news over a period of 8 years, part of which has been manually curated. The main drive behind this endeavour is to provide material to theNLP community working on event extraction, which could be used in various ways, e.g., for: (a) carrying out evaluations of detection and extraction of security-related events from online news (human-curated data), (b) training event type classifiers, (c) learning domain-specific terminology, (d) creating full-fledged inline or stand-off annotations with event-centric information based on the automatically extracted event templates.

Other efforts on the creation of corpora with event-related annotation of various nature include: GDELT (Leetaru and Schrodt, 2013), FactBank (Saurí and Pustejovsky, 2009), ICEWS (Ward et al., 2013), EventCorefBank (Cybulska and Vossen, 2014), ASTRE (Nguyen et al., 2016) and (Hong et al., 2016). Contrary to most other initiatives our corpus contains aggregated information on events extracted at news cluster level without provision of links to concrete phrases in news articles from which the information was inferred.

Section 2 briefly presents our news event extraction system. Section 3 reports on an evaluation thereof to provide insights on the quality of extraction. Section 4 provides some corpus statistics.

2 Event Extraction System Description

Our event extraction system has been primarily designed to help analysts from international institutions to automate the process of gathering intelligence on security-related events from online news. It is capable of extracting information on different types of crises, such as political violence, social turmoil, natural and man-made disasters. We briefly describe the core elements of the event extraction system, while a more detailed description can be found in (Tanev et al., 2008; Piskorski et al., 2008; Tanev et al., 2009).

The event extraction system runs on top of Europe Media Monitor (EMM), a large-scale news aggregation engine that gathers articles from ca. 7000 sources in 60 languages on a 24/7 basis (Atkinson and Van Der Goot, 2009). The system takes as input a set of news articles on the same topic, called a *news cluster* which are pro-

duced every 10 minutes by the news aggregation engine. The output of the event extraction is an event description template corresponding to the main event reported in the cluster and includes two main slots: *Event type* and *Geolocation*, and other event-type specific descriptive and numerical slots, e.g., *Perpetrators, Dead victims, Number injured, Displaced, Targeted infrastructure*.

In the first step, each article in the cluster is linguistically pre-processed to produce a more abstract representation of the text, including, i.a., tokenization, sentence splitting, NER, and labeling of key terms like action words (e.g. *kill*, *shoot*).

Our event extraction system is applied only on the title and lead sentences of each article assuming that articles are written using the *inverted pyramid style*, the dominant paradigm in modern journalism (Pöttker, 2003). Although one might potentially report on a relevant event in the final paragraphs of an article, our system has not been designed to capture them.

Next lexico-semantic patterns for the extraction of one or two slots in the event template are applied to parse more complex phrases, which express different actions and situations whose results are death, injury or other effects on people, e.g. *five people were injured*, *the boss of Cosa Nostra was found dead*. These patterns (several hundreds) were semi-automatically acquired using a bootstrapping approach (Riloff, 1996; Yangarber et al., 2000) described in more detail in (Tanev et al., 2008, 2009).

Since information about events is scattered over different articles, in the next step cross-document information validation and fusion heuristics are deployed, e.g., majority voting-like heuristics described in (Piskorski et al., 2008). To give a more precise example, in the context of extracting descriptive slots, among the phrases that apear as a filler of a given slot in the event templates extracted from all articles in the cluster, the most frequent one is selected.

Event classification is done using: (i) detecting keyword combinations, e.g., if a word in: *soldiers, troops, tanks, marines, etc.* occurs in the vicinity of a word in: *attacked, destroyed, raided, etc.*, then *Armed conflict* type is inferred, (ii) type-preference heuristics, e.g., if the text talks about violence, but simultaneously arrested people were detected using some pattern, then *Arrest* is preferred to *Violence*, and (iii) SVM-based word ngram text classifier, which is applied, when the rule-based classification yields no result.

Our event types, e.g. *Armed Conflict, Terrorist Attack, Protest, Earthquake, etc.*, were chosen among those that have the strongest impact on the security of the society.

Finally, a keyword-based filter (semi-automatically created using bootstrapping lexical learning (Tanev and Zavarella, 2014)) is deployed to eliminate events that are vaguely related to some past security-related events, e.g., commemorations related to past natural disasters, political meetings with the purpose of resolving violence-related issues, fake threats of terrorist attacks.

Our event extraction system relies on lightweight linguistic processing vis-a-vis state-of-the-art systems that use more linguistic sophistication (Kilicoglu and Bergler, 2009; Chen et al., 2015) due to: (a) specifics of the environment our system used in, where the key feature is scalability, i.e., one has to be able to quickly extend the system to detect new event types and process news in many languages, and (b) the paramount importance of providing the analysts some sort of event-centric navigation structure to guide further reading and analysis, in whose context the high quality extraction of certain slots and extraction of very fine-grained information (e.g. guessing the most specific event location information versus guessing the administrative region in which an event happened) is of lower importance.

3 Evaluation

3.1 Test Data Set

For the purpose of evaluating the performance of the automatically extracted information we have first selected 15 event types, taken from the full list of 62 types that the system is designed to detect, reported in Annex (see Sec. 4.2). The chosen types are representative of 5 broader event categories: (a) Natural disasters: *Wildfire, Flood, Earthquake*, (b) Man-made disasters: *Maritime Accident, Explosion, Ordinary Man-Made Disaster*, (c) Violence: *Kidnapping/Hostage Taking, Shooting, Terrorist attack*, (d) Military-related: *Heavy Weapons Fire, Armed Conflict, Air/Missile Attack*, and (e) Socio-political: *Riot/Turmoil, Boycott/Strike, Public Demonstration*. Then, we have randomly collected 16 news clusters that the sys-

tem had tagged with each target event type, from system data between 1/05/2016 and 31/12/2016, for a total of 240 clusters.

For each event news cluster, the annotators were given the title and first two sentences of each of the 15 (max) latest articles, including duplicates. The rationale of this setting is that we intended to 'simulate' the limited amount of data an analyst is usually able to process in order to pick up the main facts of the event reported in an article set.

For each news cluster the annotators were then tasked to provide: (a) a ranked list of up to three event types, where higher rank is given to more specific event types applicable (e.g., riot vs. disorder) and to the main event reported in the cluster vis-a-vis background events mentioned in the cluster[1], (b) a non-ranked list of locations, each represented by an ID, where in case of two or more locations being in 'administrative' inclusion relation only the most specific one is retained, (c) for each event role descriptor slot a non-ranked list of all names and descriptions found in the text, and (d) for each event role amount slot a single integer or a span of integers reflecting the minimum and maximum values reported.

Gold standard was annotated by 4 annotators. We analyzed inter-annotator agreement (IAA) for the event type classification task on a sample of 120 clusters, by considering only the first type in the ranked lists, obtaining a Fleiss Kappa score of 0.7 (Fleiss, 1971).

3.2 Evaluation metrics

For the purpose of evaluating the performance of event extraction methods the research community has been predominantly using mention-based metrics and standards such as ACE (Doddington et al., 2004), where, e.g., the scores for extracted slot fillers are summed up over their mentions in text. However, motivated by the specific environment in which our event extraction system is used, we propose partly novel evaluation metrics that try to quantify from a user perspective the most relevant semantic dimensions of event information aggregated from multi-document sets. As an example, evaluating geo-coding as the task of locating events both on a geographical reference system

and an administrative unit hierarchy (rather than as a standard entity recognition task (Mandl et al., 2009)) allows to estimate its usefulness for spatial analysis of aggregated event data. For an analyst responsible for studying events that happened in a particular administrative region (e.g., country, state) an incorrect extraction of the place, although within the boundaries of the region assigned to him, still does provide some value, which should be awarded with a non-zero score.

We first introduce the metrics for the evaluation of event type and location extraction. Let $C = \{c_1, \ldots, c_n\}$ denote the set of input clusters of articles. Let also t_c (l_c) denote the event type (location) for cluster c returned by the system. Further, let T_c^G denote an ordered list of event types in the gold truth for cluster c, and let L_c^G denote an unordered list of event locations l_c^G for c in the gold truth[2].

For the evaluation of event type classification we use an adapted version of the Mean Reciprocal Rank (MRR) (Craswell, 2009) defined as follows:

$$MRR = \frac{1}{|C|} \sum_{c \in C}^{|C|} score(t_c)$$

where $score(t_c) = 1/rank(t_c)$ with $rank(t_c)$ denoting the rank of t_c in T_c^G, or $score(t_c) = 0$ if $t_c \notin T_c^G$. In our context the reciprocal rank of a system response for cluster c is the multiplicative inverse of the rank thereof in the gold truth.

For the evaluation of the event location extraction we define two basic metrics: Geographical Closeness (GC) and Administrative Closeness (AC) which are maximized over the gold truth locations. The GC metric is defined as follows:

$$GC(c) = \max_{l_c^G \in L_c^G} \frac{1}{\ln(dist(l_c, l_c^G) + e)}$$

where $dist(a, b)$ denotes the physical distance (in km.), between a and b, which is computed using the GEONAMES gazetteer[3];

The AC metric is a modification of WUP, the semantic metric presented in (Wu and Palmer, 1994), whose main aim is to reflect how close the system location response is to the corresponding gold truth location in the administrative hier-

---

[1] For 56 clusters the annotator estimated that no type from the list in Annex (see Sec. 4.2) could be inferred as main event of the cluster, which we marked with a fictitious NA event type tag; however in 27 cases secondary gold truth event(s) matched system response, producing a non-zero score.

[2] While the system returns the most relevant location of the event in c, semantically the event could sensibly be distributed over multiple locations.

[3] http://www.geonames.org

archy of geographical references. Let T_{GEO} denote the administrative hierarchy in the GeoNames gazetteer[4] and let $LCS(x,y)$ denote the lowest common subsumer for nodes x and y in T_{GEO}. AC is then defined as follows:

$$AC(c) = \max_{l_c^G \in L_c^G} \frac{2 \cdot \omega(LCS(l_c, l_c^G))^2}{\omega(l_c)^2 + \omega(l_c^G) \cdot \omega(l_c)}$$

where $\omega(v) = \sum_{i=0}^{depth(v)} \delta/2^i$ is a weighted depth of a node v in T_{GEO}, with δ empirically set to 10. The main intuition behind AC is to apply a higher penalty to system errors: (a) closer to the root of T_{GEO} (e.g., guessing wrong country is worse than guessing wrong city within a province), and (b) resulting from providing over-specific, false information vis-a-vis system responses being not as specific, but still encompassing, gold truth location (e.g. guessing only the region of a gold truth town).

We also compute Location Accuracy (LC) as a weighted harmonic mean of GC and AC, maximized over the gold truth locations:

$$LC_\beta = \max_{l_c^G \in L_c^G} \frac{2 \cdot \beta \cdot GC(c) \cdot AC(c)}{GC(c) + AC(c)}$$

where β was set to 1 in the evaluation.

For event slot descriptors we first distinguish two cases: definite description phrases are normalized and possibly merged to the morphological base forms of their noun/adjective components (e.g. '*three Iraqi militants*' and '*Iraqi militants*' are merged into '*Iraqi militant*', while all upper case phrases (supposedly person names) are kept as such. In the former case, if $descr_c$ is a system output descriptor for a certain role of event in cluster c and $descr_c^G$ is a gold standard descriptor for the same role, the match between $descr_c$ and $descr_c^G$ is computed as:

$$\max_{m \in descr_c^N \wedge n \in descr_c^{GN}} WUP(m,n)$$

where $descr_c^N$ and $descr_c^{GN}$ are the sets of all N-grams of $descr_c$ and $descr_c^G$, resp., and $WUP(m,n)$ is a WordNet-based semantic relatedness measure (Wu and Palmer, 1994). In the latter case, matches are computed as:

$$\max_{m \in descr_c^N \wedge n \in descr_c^{GN}} StringSim(m,n)$$

where $StringSim(m,n)$ is modification of the Jaro metric boosting agreeing initial characters (Winkler, 1999).

In both cases, in order to penalize cases of role filler inversion, we score as 0 the matches of a system output role descriptor if it is lower than the max similarity with any of the other event role fillers in gold standard. Given the scores above, standard Precision, Recall and F1 measure are computed.

Finally, we record the root Mean Squared Error (MSE) of system output victim count values against gold standard, over all applicable roles[5].

3.3 Results

The evaluation results for the extraction of the event type and location are provided in Table 1. The overall results are good vis-a-vis the state-of-the-art results reported elswhere. A rudimentary error analysis of event type extraction revealed that somewhat worse results for *Violence*, *Socio-political* and *Military* categories were caused by the semantic 'proximity' of the event types contained in each of these categories. In particular, based on the low performance of extraction of *Explosion* events they were not included in the event corpus. The overall 0.4 score for GC corresponds to an average geographical error of around 9.2km from the gold standard location point, while the 0.49 for AC translates to a level of granularity between country and region levels.

The evaluation results for the extraction of the 'descriptor' and numerical slots are provided in Table 2, mF and MF columns for each role description task represent resp. the micro/macro average F1-measure. Extraction of numerical slots is quite accurate, except than for the *Dead* role, as dead counts are more likely to occur as cumulative figures in highly deadly events such as military conflicts; the systems often fails to separate them from real-time victim count updates.

---

[4] We distinguish between four levels: country, region, province and populated place.

[5] Gold truth count could sensibly be represented as an interval of max and min values, when the annotator can not pick up a unique figure among the ones mentioned in text; in those case error is computed wrt the interval boundary closest to system response

	All	Natural disaster	Man-made disaster	Violence	Military	Socio-political
MRR	0.71	0.84	0.8	0.64	0.67	0.62
GC	0.4	0.32	0.44	0.4	0.41	0.42
AC	0.49	0.45	0.56	0.11	0.53	0.41
LC	0.4	0.34	0.46	0.4	0.41	0.39

Table 1: Evaluation results for event type and location extraction for the different event type subsets.

SLOT	MF1	mF1	MSE
dead	0.46	0.62	14.6
injured	0.26	0.48	4.54
arrested	0.02	0.25	1.07
kidnapped	0.15	0.66	0.44
displaced	0.03	0.34	5.39
perpetrator	0.19	0.22	NA
weapon	0.23	0.53	NA

Table 2: Evaluation of descriptive/numerical slots.

4 Corpus release

4.1 Data sets

The current version of the corpus contains two sets: (a) moderated events (MOD) resulting from manual curation of the automated extractions in 6 languages by one trained human expert responsible for providing 'cleaned' data to the end-users, and (b) automatically extracted events (AUTO) from English news. The quantitative data on the MOD set containing 17536 event templates is given in Table 3. The (MOD) set was created during the period of 1/02/2009 to 18/08/2015. The breakdown of the events w.r.t. to languages covered is as follows: English (45.3%), Spanish (16.3%), Italian (12.0%), French (11,2%), Portuguese (7,7%) and Russian (7,5%).

The AUTO set contains ca. 600K events extracted from online news in English for the period 1/1/2009 to 1/4//2017. We have selected ca. 330K of the most 'reliable' event templates therefrom, i.e, whose extraction appears to be more accurate vis-a-vis other event types. The preliminary quantitative data on the corpus[6] is given in Table 4.

---

[6]The figures are subject to change since we are currently

TYPE	NUM	TYPE	NUM	TYPE	NUM
Arrest	2753	Terrorist attack	497	Maritime accid.	178
Disorder	2109	Earth quake	411	Stabbing	174
Man-made dis.	1971	Kidnapping	305	Physical attack	171
Trial	1903	Explosion	283	Hostage release	169
NONE	1510	Bombing	265	Hum. crisis	161
Armed conflict	1214	Air attack	253	Assassination	137
Medical	1117	Flood	223	Tropical storm	103
Shooting	906	Storm	185	OTHER	538

Table 3: Quantitative data on the MOD event set. The *OTHER* category includes all less frequent event types; *NONE* stands for events which include information on dead/injured, but whose type does not match any predefined event types

TYPE	NUM	TYPE	NUM	TYPE	NUM
Arrest	91	Armed confl.	18	Earthquake	6
Disorder/Protest	59	Flood	11	Air attack	5
Man-made dis.	42	Storm	11	Marritime accid.	3
Shooting	42	Kidnapping	8	Heavy weapon	2
Terrorist attack	36	Wildfire	7	fire	

Table 4: Quantitative data on the AUTO set (numbers of events are provided in thousands).

4.2 Format and Access

The current version of the corpus accompanied with additional information (including, i.a., list of event types and corresponding slots, instructions on how to access the underlying news stories from which the events were extracted, etc.) can be accessed at: http://labs.emm4u.eu/events.html

The corpus is available in two formats: (a) comma separated values (csv) and (b) JSON. The former contains only the following (reduced) data: unique event id; type of the event; event type category[7]; the date when it was detected; the title of the centroid article in the cluster; and the identified place name (including latitude/longitude and computed administrative path, where the first element therein provides most fine-grained information). The unique event id can be used to publicly access the articles in the cluster from which the event was extracted. The JSON format contains the full template structure including the descriptive slots: who was killed/injured; the perpetrators; the weapons used; any other descriptors that were identified for that particular event type.

It is envisaged to further extend the corpus through the provision of: (a) annotated data for new languages, (b) a new attribute reflecting extraction reliability (c) cross-language event links (Ji, 2010; Piskorski et al., 2011), and (d) additional access methods (e.g., KML).

References

Douglas Appelt. 1999. Introduction to Information Extraction. *AI Communications* 12:161–172.

Naveen Ashish, Doug Appelt, Dayne Freitag, and Dmitry Zelenko. 2006. *Proceedings of the Workshop on Event Extraction and Synthesis*. Held in conjunction with the AAAI 2006.

---

considering adding events of other types to the corpus. Updated information will be provided on the web page of the project (see Section 4.2).

[7]Currently there are six event type categories: *Natural Disaster*, *Man-mad Disaster*, *Civic-Political Action*, *Crime or Violence*, *Military Action* and *Other*

Martin Atkinson, Jakub Piskorski, Roman Yangarber, and Erik van der Goot. 2011. Multilingual Real-Time Event Extraction for Border Security Intelligence Gathering. In Uffe Kock Wiil, editor, *Open Source Intelligence and Counter-terrorism*. Springer, LNCS, Vol. 2.

Martin Atkinson and Erik Van Der Goot. 2009. Near Real Time Information Mining in Multilingual News. In *Proceedings of the 18th International World Wide Web Conference (WWW'2009)*. pages 1153–1154.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*. pages 167–176.

Nick Craswell. 2009. Mean Reciprocal Rank. In Ling Liu and M. Tamer zsu, editors, *Encyclopedia of Database Systems*, Springer US, page 1703.

Agata Cybulska and Piek Vossen. 2014. Using a Sledgehammer to Crack a Nut? Lexical Diversity and Event Coreference Resolution. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. European Language Resources Association (ELRA), Reykjavik, Iceland.

Vera Danilova and Svetlana Popova. 2014. Sociopolitical event extraction using a rule-based approach. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. Springer, pages 537–546.

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie Strassel, and Ralph M Weischedel. 2004. The Automatic Content Extraction (ACE) Program-Tasks, Data, and Evaluation. In *LREC 2004: : 4th International Conference on Language Resources and Evaluation*. volume 2, page 1.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5):378–382.

Jianbo Gao, Kalev H Leetaru, Jing Hu, Claudio Cioffi-Revilla, and Philip Schrodt. 2013. Massive media event data analysis to assess world-wide political conflict and instability. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, pages 284–292.

Ralph Grishman, Silja Huttunen, and Roman Yangarber. 2002. Real-time Event Extraction for Infectious Disease Outbreaks. In *Proceedings of HLT 2002*.

Yu Hong, Tongtao Zhang, Tim O'Gorman, Sharone Horowit-Hendler, Heng Ji, and Martha Palmer. 2016. Building a Cross-document Event-Event Relation Corpus. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016, LAW@ACL 2016, August 11, 2016, Berlin, Germany*.

Heng Ji. 2010. Challenges from Information Extraction to Information Fusion. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, Stroudsburg, PA, USA, COLING '10, pages 507–515.

Heng Ji, Ralph Grishman, Zheng Chen, and Prashant Gupta. 2009. Cross-document Event Extraction and Tracking: Task, Evaluation, Techniques and Challenges. In Galia Angelova, Kalina Bontcheva, Ruslan Mitkov, Nicolas Nicolov, and Nikolai Nikolov, editors, *RANLP*. RANLP 2009 Organising Committee / ACL, pages 166–172.

Halil Kilicoglu and Sabine Bergler. 2009. Syntactic dependency based heuristics for biological event extraction. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*. Association for Computational Linguistics, pages 119–127.

Gary King and Will Lowe. 2003. An Automated Information Extraction Tool For International Conflict Data with Performance as Good as Human Coders. *International Organization* 57:617–642.

Kalev Leetaru and Philip A Schrodt. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA Annual Convention*. Citeseer, volume 2.

Thomas Mandl, Paula Carvalho, Giorgio Di Nunzio, Fredric Gey, Ray Larson, Diana Santos, and Christa Womser-Hacker. 2009. GeoCLEF 2008: the CLEF 2008 cross-language geographic information retrieval track overview. *Evaluating Systems for Multilingual and Multimodal Information Access* pages 808–821.

Kiem-Hieu Nguyen, Xavier Tannier, Olivier Ferret, and Romaric Besanon. 2016. A Dataset for Open Event Extraction in English. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association (ELRA), Paris, France.

Jakub Piskorski, Jenya Belayeva, and Martin Atkinson. 2011. Exploring the Usefulness of Cross-lingual Information Fusion for Refining Real-time News Event Extraction: A Preliminary Study.

Jakub Piskorski, Hristo Tanev, Martin Atkinson, and Erik Van Der Goot. 2008. Cluster-Centric Approach to News Event Extraction. In *Proceedings of the 2008 Conference on New Trends in Multimedia and Network Information Systems*. IOS Press, Amsterdam, The Netherlands, The Netherlands, pages 276–290.

Jakub Piskorski and Roman Yangarber. 2013. Information extraction: Past, present and future. In Thierry Poibeau, Horacio Saggion, Jakub Piskorski, and Roman Yangarber, editors, *Multi-source, Multilingual Information Extraction and Summarization*, Springer Berlin Heidelberg, Theory and Applications of Natural Language Processing, pages 23–49.

Horst Pöttker. 2003. News and its communicative quality: The inverted pyramidWhen and why did it appear? *Journalism Studies* 4(4):501–511.

Ellen Riloff. 1996. Automatically Generating Extraction Patterns from Untagged Text. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 2*. AAAI Press, AAAI'96, pages 1044–1049.

Alan Ritter, Evan Wright, William Casey, and Tom Mitchell. 2015. Weakly supervised extraction of computer security events from twitter. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, pages 896–905.

Roser Saurí and James Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language resources and evaluation* 43(3):227.

Hristo Tanev, Jakub Piskorski, and Martin Atkinson. 2008. Real-Time News Event Extraction for Global Crisis Monitoring. In *Proceedings of NLDB 2008.*. pages 207–218.

Hristo Tanev and Vanni Zavarella. 2014. Multilingual Lexicalisation and Population of Event Ontologies: A Case Study for Social Media. In *Towards the Multilingual Semantic Web*, Springer, pages 259–274.

Hristo Tanev, Vanni Zavarella, Jens P. Linge, Mijail A. Kabadjov, Jakub Piskorski, Martin Atkinson, and Ralf Steinberger. 2009. Exploiting Machine Learning Techniques to Build an Event Extraction System for Portuguese and Spanish. *Linguamática* 1(2):55–66.

Michael D Ward, Andreas Beger, Josh Cutler, Matt Dickenson, Cassy Dorff, and Ben Radford. 2013. Comparing GDELT and ICEWS event data. *Analysis* 21:267–297.

William E. Winkler. 1999. The State of Record Linkage and Current Research Problems. Technical Report Statistical Research Report Series RR99/04.

Zhibiao Wu and Martha Palmer. 1994. Verbs Semantics and Lexical Selection. In *Proceedings of the 32Nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '94, pages 133–138.

Roman Yangarber, Peter Von Etter, and Ralf Steinberger. 2008. Content Collection and Analysis in the Domain of Epidemiology. In *Proceedings of DrMED 2008: International Workshop on Describing Medical Web Resources at MIE 2008: the 21st International Congress of the European Federation for Medical Informatics 2008*. Goeteborg, Sweden.

Roman Yangarber, Ralph Grishman, Pasi Tapanainen, and Silja Huttunen. 2000. Unsupervised Discovery of Scenario-level Patterns for Information Extraction. In *Proceedings of the Sixth Conference on Applied Natural Language Processing*. Association for Computational Linguistics, Stroudsburg, PA, USA, ANLC '00, pages 282–289.