

Exploring the Usefulness of Cross-lingual Information Fusion for Refining Real-time News Event Extraction: A Preliminary Study

Jakub Piskorski

Institute of Computer Science
Polish Academy of Sciences
Warsaw, Poland
Jakub.Piskorski
@ipipan.waw.pl

Jenya Belayeva, Martin Atkinson

Joint Research Centre
of the European Commission
Ispra, Italy
{Jenya.Belayeva, Martin.Atkinson}
@jrc.ec.europa.eu

Abstract

Nowadays, many influential facts are reported multiple times by different sources and in different languages. This paper presents the results of an experiment on deploying cross-lingual information fusion techniques for refining the results of a large-scale multilingual news event extraction system. An evaluation on a test corpus consisting of 618 event descriptions which refer to 523 real-world events revealed that the description of circa 10% of the events extracted by the mono-lingual systems could be refined. In particular, an overall gain of 6,4% and 4,8% in recall and precision against the best mono-lingual system could be obtained respectively.

1 Introduction

The goal of event extraction is to automatically identify events in free texts and to derive structured and detailed information about them. In the past, a vast bulk of the research focused on the development of mono-lingual event extraction systems that operate on single documents without taking any advantage of global evidence, i.e., without reusing the knowledge acquired in the process of extracting information from other topically-related documents. The advantages of going beyond the classical single-document extraction and exploiting information redundancy to validate facts have recently been explored by various research groups (Downey et al., 2005; Finkel et al., 2005; Ji and Grishman, 2008; Lee et al., 2010; Liao and Grishman, 2010; Mann, 2007; Patwardhan et al., 2007; Poibeau et al., 2008; Yangarber

and Jokipii, 2005; Yangarber, 2006). Since nowadays many influential facts are not only reported multiple times by different sources, but also in different languages, the importance of the ability to aggregate and fuse information across documents in several languages is becoming paramount (Ji, 2010). Several experiments on cross-lingual information extraction have been reported (Chen et al., 2009; Sudo et al., 2004; Lee et al., 2010), however, they mainly focused on cross-lingual bootstrapping of ML-based event extraction systems.

This paper presents the results of an experiment aiming at exploring the usefulness of cross-lingual information fusion for refining the results of a real-time multilingual news event extraction engine that is deployed in a large-scale online news monitoring platform. To be more precise, we explored: (a) what fraction of event descriptions extracted could potentially be merged and refined through cross-lingual information fusion; and, (b) whether gain in precision/recall could be obtained. In principle, there are two ways of approaching cross-lingual information fusion in the context of multilingual news event extraction: (1) translate all news articles into one common language for which a high-performance event extraction system exists (e.g., English), and run that system on the translated news (including cross-article fusion), or (2) run mono-lingual event extraction on the native language news articles, then translate (normalize) automatically extracted event descriptions into one common language, and subsequently, perform information fusion. In this paper we explore the latter approach.

The remaining part of this paper is organized as follows. First, the real-time event extraction engine is presented in Section 2. Next, the creation and statistics of the test corpus are described in

Section 3. Subsequently, Sections 4 and 5 present the cross-lingual fusion technique and the results of the experiments. We end with some conclusions in Section 6.

2 Real-time Event Extraction Engine

First, news articles are gathered by Europe Media Monitor (EMM) (Atkinson et al., 2009), a large-scale media monitoring platform¹, which currently retrieves a vast bulk of news articles per day from over 2500 news sources in all major languages. The news articles harvested in a last 4-hour time window are grouped into clusters according to content similarity, using hierarchical agglomerative clustering in a manner as described in (Piskorski et al., 2011).² Then news article clusters are categorised using filters, which consist of boolean combinations of multilingual keywords and some metadata.

Next, each cluster is processed by NEXUS, the core event extraction engine, which initially performs shallow linguistic analysis, including, i.a., fine-grained tokenization, sentence splitting, domain-specific dictionary look-up (e.g., for the detection of numerical expressions, quantifiers, person titles, and for the labeling of key terms indicating unnamed person groups), and morphological analysis. In particular, for morphological analysis an extended version of the full-form MUL-TEXT³ lexica are used.

Subsequently, a cascade of finite-state extraction grammars⁴ is applied on each article in the cluster. The low-level grammars are primarily used for the detection of small-scale structures (e.g., person groups, which might potentially constitute a slot filler). The higher-level grammars consist of simple linear 1/2-slot extraction patterns, similar to those in (Riloff, 1996), e.g., PER-GROUP <VICTIM> "was killed" assigns a group of persons followed by a phrase "was killed" the role of a victim. These patterns are applied only on the top sentences and the title of each article. The main rationale behind this

¹<http://press.jrc.it>

²The article feature vectors are simple word count vectors and no lemmatization is performed.

³<http://nl.ijs.si/ME/>

⁴A grammar consists of pattern-action rules, where the left-hand side of a rule is a regular expression over non-recursive typed feature structures (the recognition pattern), whereas the right-hand side constitutes a list of feature structures, which will be returned in case the recognition pattern is matched. See (Piskorski, 2007) for more details.

is that news articles are written in the inverted-pyramid style.⁵ Secondly, analysing the entire text might involve handling complex language phenomena (e.g., anaphora resolution), which is hard and requires knowledge intensive processing. In particular, in the context of developing an event extraction system capable of processing news in several languages tackling more complex language phenomena would involve a substantial effort to provide the necessary language-specific resources. Finally, if some crucial information can not be captured from one article in the cluster (due to the simplistic approach mentioned before), it might be extracted from other articles in the same cluster. Let us consider as an example the following sentence.

*'The United Nations says **Somali gunmen** who hijacked a U.N.-chartered vessel carrying food aid for tsunami victims **have released the ship** after holding it for more than two months.'*

The proper extraction of *Somali gunmen* as the actor of a RELEASE event would require some syntactical parsing to identify the relative clause that describes the *Somali gunmen*, otherwise the application of a linear extraction pattern might result in assigning the *tsunami victims* the actor role of the RELEASE event (incorrect). However, the title and the initial sentence of most of news articles on crisis-related events exhibit relatively simple syntactical structure, e.g., it would be more likely (based on empirical observations) that the same information as in our example is conveyed through a sentence like this:

'Somali gunmen have released the ship after holding it for more than two months.'

Consequently, the application of the pattern PER-GROUP <ACTOR> "have released" would yield a correct extraction of *Somali gunmen* as the actor of the RELEASE event.

Since the information about events is scattered over different articles, the last step consists of cross-article cluster-level information fusion in order to produce full-fledged event descriptions, i.e., information extracted locally from each single article in the same cluster is aggregated and validated. This involves: (a) disambiguation on entity roles (as a result of application of extraction patterns the same entity might be assigned different

⁵The most important parts of the story are placed in the beginning of the article and the least important facts are left toward the end.

roles), (b) computing an estimate of the total number of victims, and (c) event type classification, all accomplished through heuristics.⁶

It is important to note that NEXUS detects only the main event for each news article cluster ('one sense per discourse') (Gale et al., 1992), and 6 language-specific instances of the system have been developed to cover news in English, Italian, Spanish, French, Portuguese, and Russian. In particular, for each language extraction grammars and specialized lexica were acquired using weakly supervised ML techniques and validated by human experts. Noteworthy, certain part of the extraction grammars are shared among languages (Zavarella et al., 2008).

There are several differences in language-specific versions of NEXUS. Currently, Italian, French, Spanish and Portuguese versions fully rely on morphological analysis (MULTEXT), whereas Russian and English system instances do not, i.e., morphological features are not referred to in the extraction patterns. In addition, the Italian, Spanish and Portuguese systems deploy more (abstract) linguistic rules that constitute a partial parser of domain specific phrases. The overall number of extraction patterns used in the Italian, Spanish and Portuguese system varies from 100 to circa 400, whereas the English, French and Russian system deploy thousands of extraction patterns, mainly relying on surface-level text features. Another important difference is that the event type classification for English is done using a blend of category definitions and a statistical classifier, whereas the other 5 language-specific instances rely only on well-defined event category definitions. There are over 30 event category definitions, which can consist of a simple list of related keywords or a combination of lists of words. Most category definitions are defined using Boolean operators with optional proximity operator and wild cards. Alternatively, cumulative positive or negative weights and a threshold can be specified.

The briefly sketched cluster-centric approach to news event extraction, the process of acquisition of language specific resources for NEXUS, and other particularities of NEXUS are given in (Tanev et al.,

⁶For instance, if the same entity has two roles assigned in the same news cluster, preference is given to the role assigned by the most reliable group of patterns, e.g., 2-slot extraction patterns are considered more reliable than 1-slot extraction patterns. In case of event type classification and victim counting heuristics similar in spirit to those described in (Piskorski et al., 2011) were used.

2008; Tanev et al., 2009; Piskorski et al., 2011). Some other effort aiming at constructing multilingual event extraction based on light-weight linguistic approach is presented in (Lejeune et al., 2010).

3 Corpus and Event Statistics

For exploring the potential of cross-lingual information fusion a corpus consisting of crisis-related event descriptions automatically extracted by NEXUS on 22 randomly selected (non-continuous) days in 2010 from news in 6 languages has been prepared. In particular, we focused on violent events and natural and man-made disasters. The set of slots we considered includes the following ones: TYPE, LOCATION, PERPETRATOR, DEAD, DEAD-COUNT, INJURED, INJURED-COUNT, KIDNAPPED, KIDNAPPED-COUNT, ARRESTED, WEAPONS.

The corpus consists of 618 event descriptions. Table 1 gives the statistics on the extracted event descriptions and news sources used. The 618 event descriptions extracted refer to 523 real-world events.

Language	#Event descriptions	#Slots filled in total	#Slots filled on average	#News sources
English	268	963	3.59	783
Spanish	129	454	3.52	174
French	77	273	3.55	224
Italian	50	172	3.44	68
Russian	52	158	3.04	178
Portuguese	42	137	3.26	55
All	618	2157	3.49	1482

Table 1: The statistics of the extracted events.

Out of the 523 events 51 were reported in more than one language. This accounts for circa 9,8% of all extracted events that could be potentially refined through cross-lingual information fusion. The 51 events reported in more than one language include: 33 violence events, 7 natural disasters, 9 man-made disasters and 2 other crisis-related events. Noteworthy, 350 events out of the 523 were detected in non-English news. In the latter group of 'non English' events only 7 were reported in more than two languages, which accounts for 2% of all events in this group. Hence, extraction from English news is crucial in the process of cross-lingual information fusion. The histogram in Figure 1 shows the number of languages in which news report on events in our corpus.

For the 51 events reported in more than one language we manually created the gold-standard

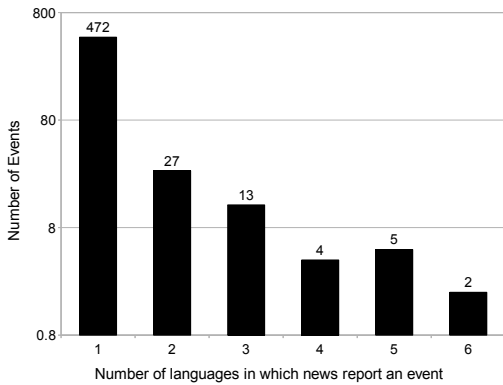


Figure 1: The histogram showing the number of languages in which news report on events.

stand-off annotations based on any information which could be found in news articles in all 6 languages. Furthermore, for the purpose of evaluating mono-lingual systems, we also created for each of the 51 events and a given language (i.e., for each mono-lingual news article cluster) a stand-off annotation based on information which could be found in that particular language only. In total, there were 4252 news articles that refer to the 51 events. In particular, the average number of news articles per cluster which correspond to an event in the set of the 51 events is: 29 (all languages), 55 (English), 16 (Spanish), 21 (Italian), 29 (French), 21 (Portuguese), and 8 (Russian). The annotation task (including the classification of the events) was jointly carried out by two annotators.

For the preparation of the test corpus and annotation, in particular, for linking (manually) of event descriptions across languages, the Event Moderation Tool (EMT) described in (Atkinson et al., 2011) has been used. EMT provides GUI-based tools that can: retrieve automatically extracted event descriptions gathered over time according to a number of different criteria (e.g., event type, date of occurrence, language, source and location), edit, validate, group, translate, and export them into other knowledge repositories.

The test corpus comes from Internet news articles that EMM scrapes and analyses on the fly. The scraped information is governed by copyrights and therefore cannot be reproduced by any means or form without infringement. Hence, as of now, the only corpus that we can provide to the research community are the links to the original articles accompanied with some additional information, i.e., the corresponding event ID that we gen-

erated (for 51 events reported in at least two languages) as well as the language of the underlying news article. The resource file URL is available at: <http://emm-labs.jrc.it/CLEventResources.csv>.⁷

4 Cross-lingual Fusion

The information fusion process is divided into two steps. First, event descriptions extracted by mono-lingual systems are normalized, i.e., all non-numerical slot fillers are translated (converted) into English, whereas geographical names are mapped to their canonical forms using the multi-lingual *GeoNames*⁸ gazetteer. In the second step, for each event the corresponding normalized event descriptions are merged into one via the application of simple fusion methods. The computation of the value of each slot in the ‘fused’ event description is based on the following general assumption: *‘If a candidate slot value (returned by at least one of the mono-lingual systems) occurs frequently (more than once) as a filler of a given slot in a collection of event descriptions referring to a certain real-world event, and if this value was ‘on average’ extracted with high system confidence⁹, and if it refers to a more specific concept than the other values in the candidate slot filler set, that increases the likelihood that this slot value is correct’.*

Table 2 shows an example of system response (in as simplified form), i.e., event descriptions extracted by mono-lingual systems, and the result of cross-lingual fusion for an event related to U.S. drone strike that killed eight militants of German nationality in Islamabad.

We now present the fusion method more formally. First, let E denote an event. We denote the set of automatically extracted event descriptions that refer to E as $E_D = \{e_1, \dots, e_k\}$, where e_i is a set of slot-value pairs. The value of slot x in the event description e is denoted as $e(x)$. We extend this notion to a set of values for slot x in an event description collection $E_D(x) = \{v | \exists e \in E_D \wedge e(x) = v\}$. Next, let $E_D^{x=v} = \{e | e \in E_D \wedge e(x) = v\}$ be the set of event descriptions with certain value v for the slot

⁷It is important to note that some online media do not archive their news. As a consequence of this, a fraction of the links provided in the URL might become inactive relatively soon.

⁸<http://www.geonames.org/>

⁹‘on average’ meaning that the average system confidence was high

LANG	Event Type	Location	Dead (count)	Injured (count)
IT	<u>Air Attack</u>	<u>Islamabad</u>	German (3)	- (-)
EN	Armed Conflict	<u>Islamabad</u>	<u>German militants (5)</u>	- (-)
ES	Armed Conflict	<u>Islamabad</u>	<u>German militants (8)</u>	German militants (3)
RU	-	Pakistan	people (8)	- (-)
FR	-	Pakistan	insurgers (8)	- (-)
Fusion	Air Attack	Islamabad	German militants (8)	none (0)

Table 2: Cross-lingual fusion example. The underlined values were selected as slot fillers in the fusion process.

x . Furthermore, we denote systems’ confidence of extracting v as the value of $e(x)$ as $conf_e(x, v)$ ¹⁰. Let e^* denote the event description resulting from merging the event descriptions in E_D using fusion method M , which is defined as follows:

$$e^*(x) = \operatorname{argmax}_{v \in E_D(x)} Score_M(x, v)$$

where $Score_M(x, v)$ denotes a scoring function specific to method M . For filling non-numerical slots we used the following scoring function:

$$Score_M(x, v) = \sum_{e \in E_D^{x=v}} conf_e(x, v) \cdot \frac{1}{|E_D^{x=v}|} + \alpha \cdot |E_D^{x=v}| + \beta \cdot |\{v' \in E_D(x) \mid v' \supset v\}|$$

where $\alpha \geq 0$ is a factor determining the importance of the number of occurrences of v as a slot filler for x , and $\beta \geq 0$ is a factor which specifies the degree of boosting slot values, which happen to represent concepts that stand either in ‘is-subsumed-by’ or ‘is-part-of’ relation (denoted as ‘ \supset ’) with other concepts in the same slot value set.¹¹ The rationale of using the latter factor is that, intuitively, a ‘more-specific’ value co-occurring with a related ‘more-generic’ concept is more likely to be the correct slot filler among those two. For instance, in $E_D(LOCATION) = \{Spain, Andalucia, Algeciras\}$, *Algeciras* would be boosted by $\beta \cdot 2$ since *Algeciras* is a part of *Andalucia* and *Spain*. Hence, *Algeciras* gets

¹⁰The confidence is based on a combination of factors, e.g., the reliability of the pattern(s) used to extract a particular value (the likelihood that pattern extract the slot value correctly), the number of articles in which some patterns were triggered (frequency), the overall confidence of the language-specific instance of the event extraction system, etc.

¹¹A small in-house ontology was used for this purpose.

a higher chance of being selected as the location of the event. α and β were set differently for different slot types.

As for numerical slots, the fusion was done in a slightly different way. First of all, the definition of $Score_M(x, v)$ was simplified since the last part (β) does not apply to numbers, and secondly, in case of candidate values, which are significantly distant one from another we selected a maximum (provided that confidence of extracting it is higher than a pre-specified constant), based on a simple assumption that the event is most likely evolving and numbers change continuously, the highest being the more up-to-date one. It is not necessarily the case that the last news article on a certain event reports the most up-to-date figures since there is certain latency between reporting on a given event in different countries. Therefore, we chose the ‘maximum’ heuristic.

5 Experiments

We have applied the cross-lingual fusion technique presented in Section 4 on the corpus described in Section 3 and we measured extraction precision, recall and F-measure for each language-specific system instance and for the extraction based on cross-lingual information fusion. It is important to note that we assigned basically three scores (for non-numerical slots) for filling each slot: 0 (incorrect), 1 (correct), and 0.5 (partially correct), where ‘partially correct’ is assigned in cases where the slot fill represents a more generic concept than the one in the gold-standard, or in case of locations, if the slot fill refers to an administrative unit, which encompasses the specific place of an event, e.g., if the event happened in *Islamabad*, we assign the slot fill *Pakistan* the score ‘partially correct’.

	Event Type			Location		
	P	R	F	P	R	F
English	86.6	80.7	83.5	82.6	80.7	81.6
Spanish	80.4	61.7	69.8	85.5	85.5	85.5
French	83.3	70.0	76.1	68.8	66.0	67.4
Italian	67.8	63.3	65.5	86.7	86.7	86.7
Russian	81.3	28.3	42.0	61.4	58.7	60.0
Portuguese	73.7	59.2	65.7	87.5	55.6	66.7
FUSION	91.3	84.3	87.6	87.3	87.3	87.3

Table 3: Precision, recall and F-measure figures for the extraction of event type and location.

The overall precision and recall figures is shown in Figure 2. Compared to the performance of the best mono-lingual system a gain of 6,4% and

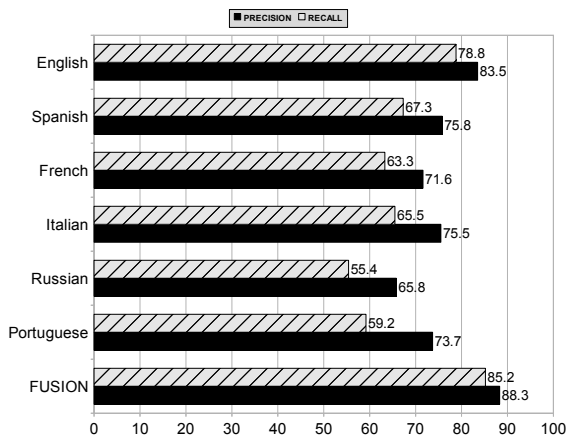


Figure 2: The overall precision (black solid bars) and recall figures for monolingual event extraction vs. event extraction refined by cross-lingual information fusion.

	Non-numerical slots			Numerical slots		
	P	R	F	P	R	F
English	92.2	80.8	86.1	73.9	70.8	72.3
Spanish	84.6	69.1	76.1	62.9	51.5	56.6
French	91.6	71.7	80.4	54.1	46.4	50.0
Italian	85.0	51.5	64.1	53.8	41.1	46.6
Russian	91.6	55.0	68.7	75.0	53.5	62.5
Portuguese	83.3	63.1	71.8	50.0	40.0	44.4
FUSION	91.5	83.5	87.3	82.6	79.6	81.1

Table 4: Precision, recall and F-measure figures for the extraction of numerical and non-numerical slots.

4,8% respectively in the overall recall and precision could be obtained through cross-lingual fusion. Table 3 gives the precision, recall and F-measure for the extraction of the event type and location, whereas Table 4 gives the corresponding figures for the extraction of other non-numerical and numerical slots. As can be observed, a gain of 4-5% and 8% in precision and recall could be obtained for the extraction of event type and numerical slots respectively. The precision for extracting locations and non-numerical slots for the best-scoring mono-lingual system is better than the result of cross-lingual fusion. However, the recall for the same slot types is 0.6% and 2.7% respectively higher in case of cross-lingual fusion.

A small error analysis of cross-lingual fusion was carried out. In case of fusing event type information, it turned out that for 5 out of 51 events in our corpus none of the mono-lingual systems was able to assign any type information. Consequently, the cross-lingual fusion did not result in any improvement in case of those events, i.e., no

type information was assigned. In case of 2 other events, all of the mono-lingual systems returned incorrect event type information, which resulted in incorrect cross-lingual fusion. Furthermore, in case of 2 events, the cross-lingual fusion resulted in selection of an event type (extracted by at least one of the mono-lingual systems), which is related to the event type in the gold standard (partially correct extraction), but the latter was not detected by any mono-lingual system. Finally, for 1 event, the cross-lingual fusion resulted in selection of an incorrect event type, although the correct event type was detected by at least one of the mono-lingual systems. The analysis of fusing location information revealed that: (a) in case of 3 events a wrong location was selected, although at least one of the mono-lingual system returned the correct answer, (b) for 4 events the returned location was partially correct, and (c) for 2 events none of the mono-lingual systems provided a correct answer, and, consequently, the error was propagated in the fusion process.

6 Conclusions and Future Work

We presented the results of preliminary explorations on using cross-lingual information fusion to improve the recall/precision of a large-scale multilingual event extraction system. Circa 10% of event descriptions extracted by the mono-lingual systems could be refined, and a gain of 6,4% and 4,8% in the overall recall and precision could be obtained respectively. Since we limited the time window for grouping event descriptions referring to a given event to 1 day only the aforementioned figure of 10% constitutes an approximation of a lower bound for the fraction of crisis-related event descriptions, which can be potentially refined through cross-lingual information fusion. An effort is envisaged to create (multilingual) temporal event chains (Ji et al., 2009), which go beyond 1-day time window, for further explorations on the potential of cross-lingual information fusion for refining event extraction results.

Although the reported improvement in precision and recall appears to be promising, to better assess the actual impact of exploiting multilinguality for refining event descriptions an evaluation of the improvement achieved by merging information from different sources in the same language is planned too. In order to get a better insight into the real contribution of exploiting

news in each language a direct one-to-one comparison between the English system (the one with the highest impact) and each of the mono-lingual systems will be carried out too.

Furthermore, we intend to explore the usefulness of deploying cross-lingual information fusion in the context of extracting other types of events. For instance, in (Atkinson et al., 2011) we elaborate on the specifics of reporting on border security-related events (e.g., illegal migration attempts, cross-border crimes, etc.) in online news, which revealed that suchlike events are intuitively less likely to benefit from cross-lingual information fusion.

Future work will also focus on exploring more elaborated fusion techniques (Ji and Grishman, 2008) and comparison with the approach based on translating news articles into one common language and running event extraction and information fusion on the translated articles. Although several authors reported that such an approach is error-prone due to inaccuracy of the state-of-the-art machine translation techniques, it has not been evaluated in the context of a cluster-centric and linguistically-lightweight approach to event extraction as described in this paper.

Our event extraction engine processes only the title and top sentences of each news article. However, processing additional ‘relevant’ sentences, which could be selected through deployment of some time-efficient sentence ranking measures (Litvak et al., 2010), might lead to a better coverage and is considered to be explored in the future. The inclusion of additional sentences in the event extraction process might also help to estimate the fraction of information which is being missed by the current event extraction engine.

With the emergence of social media, one can observe an ever growing trend of reporting on the same event in many different languages. For instance, the GLOBAL VOICES¹² is a community of bloggers and translators around the globe, who link and translate articles/posts on certain events and issues that are not usually present in international mainstream media. Therefore, we plan to carry out experiments on deploying cross-lingual information fusion techniques to refine event extraction from suchlike information sources.

Although the experiments reported in this paper are preliminary, we strongly believe that the

¹²<http://globalvoicesonline.org>

presented work and discussion constitutes useful source of information for researchers and practitioners working on advancing event extraction technology.

Acknowledgments

We are greatly indebted to all our colleagues in the OPTIMA action at the Joint Research Centre, who are working on *EMM*. In particular, we would like to thank Hristo Tanev and Vanni Zavarella, who are working on the core of NEXUS and language-specific versions of the system, and without whom the presented work would not have been possible.

References

- Martin Atkinson and Erik van der Goot. 2009. Near Real-Time Information Mining in Multilingual News. In *Proceedings of the 18th World Wide Web Conference 2009*.
- Martin Atkinson, Jakub Piskorski, Erik Van der Goot and Roman Yangarber. 2011. Multilingual Real-Time Event Extraction for Border Security Intelligence Gathering. Book chapter in: *Counterterrorism and Open Source Intelligence Series*. Lecture Notes in Social Networks, Vol. 2.
- Zheng Chen and Heng Ji. 2009. Can one Language Bootstrap the Other: A Case Study on Event Extraction. In *Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing*.
- Doug Downey, Oren Etzioni and Stephen Soderland. 2005. A Probabilistic Model of Redundancy in Information Extraction. In *Proceedings of IJCAI 2005*.
- Jenny Finakel, Trond Grenager and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of ACL 2005*.
- Gael Lejeune, Antoine Doucet, Roman Yangarber and Nadine Lucas 2010. Filtering News for Epidemic Surveillance: Towards Processing More Languages with Fewer Resources. In *Proceedings of 4th Workshop on Cross Lingual Information Access at COLING 2010*.
- William Gale, Kenneth Church and David Yarowsky 1992. One Sense per Discourse. In *Proceedings of the workshop on Speech and Natural Language*., pages 233–237.
- Silja Huttunen, Roman Yangarber and Ralph Grishman. 2002. Complexity of Event Structure in IE Scenarios. In *Proceedings of COLING 2002*.

- Heng Ji and Ralph Grishman. 2008. Refining Event Extraction through Cross-Document Inference. In *Proceedings of ACL 2008*, pages 254–262.
- Heng Ji, Ralph Grishman, Zheng Chen and Prashant Gupta. 2009. Cross-document Event Extraction and Tracking: Task, Evaluation, Techniques and Challenges. In *Proceedings of RANLP 2009*, pages 166–172.
- Heng Ji. 2010. Challenges from Information Extraction to Information Fusion. In *Proceedings of ACL 2008*, pages 507–515.
- Adam Lee, Marissa Passantino, Heng Ji, Guojun Qi and Thomas Huang. 2010. Enhancing Multi-lingual Information Extraction via Cross-Media Inference and Fusion. In *Proceedings of COLING 2010: Posters*, pages 630–638.
- Shasha Liao and Ralph Grishman. 2010. Using Document Level Cross-Event Inference to Improve Event Extraction. In *Proceedings of ACL 2010*, pages 789–797.
- Marina Litvak, Mark last and Menahem Friedman. 2010. A new Approach to Improving Multilingual Summarization using a Genetic Algorithm. In *Proceedings of ACL 2010*, pages 927–936.
- Gideon Mann. 2007. Multi-document Relationship Fusion via Constraints on Probabilistic Databases. In *Proceedings of HLT/NAACL 2007*.
- Martina Naughton, Nicholas Kushmerick and Joseph Carthy. 2006. Event Extraction from Heterogeneous News Sources. In *Proceedings of the AAAI 2006 workshop on Event Extraction and Synthesis*.
- Siddharth Patwardhan and Ellen Riloff. 2007. Effective Information Extraction with Semantic Affinity Patterns and Relevant Regions. In *Proceedings of EMNLP-CONLL 2007*.
- Jakub Piskorski. 2007. ExPRESS Extraction Pattern Recognition Engine and Specification Suite. In *Proceedings of the 6th International Workshop Finite-State Methods and Natural language Processing (FSMNL 2007)*.
- Jakub Piskorski, Hristo Tanev, Martin Atkinson, Erik van der Goot and Vanni Zavarella. 2011. Online News Event Extraction for Global Crisis Surveillance. In *Transactions on Computational Collective Intelligence*, Volume 5, Lecture Notes in Computer Science 6910, pages 182–212, Springer-Verlag Berlin Heidelberg.
- Thierry Poibeau, Horacio Sagion and Roman Yangarber. 2008. *Proceedings of the Workshop on Multi-source Multilingual Information Extraction and Summarization (MMIES 2008)*.
- Ellen Riloff. 1996. Automatically Generating Extraction Patterns from Untagged Text. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI-96)*, pages 1044–1049.
- Kiyoshi Sudo, Satoshi Sekine and Ralph Grishman. 2004. Cross-lingual Information Extraction System Evaluation. In *Proceedings of COLING 2004*.
- Hristo Tanev, Jakub Piskorski and Martin Atkinson. 2008. Real-Time News Event Extraction for Global Crisis Monitoring. In *Proceedings of NLDB 2008*. LNCS, Vol. 5039, pages 207–218, Springer.
- Hristo Tanev, Vanni Zavarella, Jens Linge, Mijail Kabadjov, Jakub Piskorski, Martin Atkinson and Ralf Steinberger. 2009. Exploiting Machine Learning Techniques to Build an Event Extraction System for Portuguese and Spanish. *Linguamatica Journal*, Vol.2, pages 55–66.
- Earl Wagner, Jiahui Liu, Larry Birnbaum, Kenneth Forbus and James Baker. 2006. Using Explicit Semantic Models to Track Situations Across News Articles. In *Proceedings of the AAAI 2006 workshop on Event Extraction and Synthesis*.
- Roman Yangarber and Lauri Jokipii. 2005. Redundancy-based Correction of Automatically Extracted Facts. In *Proceedings of HLT/EMNLP 2005*.
- Roman Yangarber. 2006. Verification of Facts across Document Boundaries. In *Proceedings of International Workshop on Intelligent Information Access, Helsinki*.
- Roman Yangarber, Clive Best, Peter von Etter, Flavio Fuart, David Horby and Ralf Steinberger. 2007. Combining Information about Epidemic Threats from Multiple Sources. In *Proceedings of International Workshop on Multi-source, Multilingual Information Extraction and Summarization at RANLP-2007*.
- Zavarella Vanni, Hristo Tanev and Jakub Piskorski. 2008. Event Extraction for Italian Using a Cascade of Finite-State Grammars. In *Proceedings of the 8th International Workshop on Finite-State Methods and Natural language Processing (FSMNL 2008)*.