Aspect-Driven News Summarization

Josef Steinberger, Hristo Tanev, Mijail Kabadjov, and Ralf Steinberger

Joint Research Centre, European Commission, Via E. Fermi 2749, Ispra (VA), Italy
firstname.lastname@jrc.ec.europa.eu

Abstract. A summary of any event type is only complete if certain information aspects are mentioned. For a court trial, readers will at least want to know who is involved and what the charges and the sentence are. For a natural disaster, they will ask for the disaster type, the victims and other damages. Will a co-occurrence or frequency-based sentence extraction summariser automatically provide the requested information, or are the results better if an information extraction (IE) system first detects the summary-crucial aspects? To answer this question, we compared the performance of a purely co-occurrence-based method with a system that additionally makes use of targeted IE. As each event type requires different information aspects and not all of them were covered by the existing IE software, we used a tool that learns semantically related terms to cover the remaining aspects. The comprehensive evaluation in the TAC'2010 competition showed that event extraction is indeed beneficial for summarisation performance, and that summary quality is directly related to IE quality. Our integrated system was ranked among the top systems participating at TAC.

1 Introduction

Our main goal is to produce succinct multilingual summaries within the Europe Media Monitor (EMM)[1] framework. The news collator gathers around 100,000 news articles every day from various news sources and continuously groups them, producing topic-homogeneous news clusters for each of a set of 40+ languages. There are thus many news clusters in various languages, varying in size from two to more than a hundred articles. Multi-document summarization systems can potentially reduce this big bulk of highly redundant news data and obtain one succinct text which summarizes the most important content.

Evaluation of multi-document summarisation is difficult and time-consuming. Teams participating in the summarisation task of the Text Analysis Conferences TAC, organised by the US National Institute of Standards and Technology, benefit from a thorough evaluation of the output of competing systems on a standard test set. While the task in TAC'2009 simply was to produce a concise summary of a cluster of related news articles, TAC'2010 requested that a given list of core information aspects for different event types be addressed in the automatic summaries. This ambitious and challenging requirement is congruous with current, IE-aware trends in the field of summarization [1, 2].

[1] http://emm.jrc.it/overview.html

In this paper, we present a novel approach to combining standard extractive summarization techniques with higher-level information extraction in a neat and unified manner. By submitting results produced by both this new approach and the standard technique to the TAC'2010 competition, we received a detailed comparative evaluation of both methods, giving us insight in the relative benefits of either approach.

One successful approach to standard summarization (e.g., yielding scores in the top 10% at previous TACs) builds on the Latent Semantic Analysis (LSA) paradigm. Proponents of this approach to summarization include [3, 4]. Being, by definition, a language-independent approach which is one of the core requirements in our setup, we decided to adopt it as a foundation for building an IE-aware summarizer. Additionally, from the news collator project for which we are building the summarization system, a mature multilingual event extraction (EE) system [5] was made available to us. Coincidentally, it was purpose-built for a very similar domain to that of the TAC corpora and as such, it by definition covered several of the aspects specified in the summarization track of TAC'10. In order to cover some of the remaining aspects of the TAC'10 track, we in addition used a system implementing statistical distributional semantics methods to learn new terms lexically similar to an initial seed of terms [6].

In the remainder of this paper we firstly discuss related work (section 2), then, in section 3, we describe the information extraction tools we used, followed by the description of our hybrid IE-aware summarization approach in section 4. Next, in section 5 we present a detailed analysis of the results obtained at TAC'10, and finally, we conclude and give pointers to future work.

2 Related Work

There is several related work carried out in the past which tried to exploit the potential of using information extraction in summarization. As a pioneering effort, the SUMMONS system [7], which summarized the results of MUC-4 IE systems in the terrorism domain, was the first to suggest using IE in a summarization system, though no evaluation was carried out. In [8] another system that combined information extraction and summarization was presented. Even though the potential improvement in content coverage when simulating the output of the IE system was demonstrated, using the actual output of the IE system was not good enough. Another attempt to use IE and summarization in a sequential pipeline was proposed in [9]. The system dynamically determined the focus of the article (mainly based on the analysis of entity mentions), which in turn determined the specific information that was extracted. However, the study arrived at inconclusive results. In [2] an approach that used templates conceived from the rhetorical structure of scientific papers was proposed. The templates guided the search for appropriate sentences in the source text. In [1] a new set of features based on low-level, atomic events that describe relationships between important actors in a document or set of documents was presented. Using the

event-based features resulted in an improvement in summary quality over using lexical features, but also in avoiding summary redundancy.

3 Information Extraction for Summarization

We describe here information extraction components we used to capture the required summary information. For capturing highly frequent topics in a cluster we use in addition to lexical features (words and bigrams) also person, organization and location entity mentions discovered by our entity recognition and disambiguation tools. For capturing the category-related aspects we used our event extraction system and the tool for automatic learning of semantic classes.

3.1 Entity Recognition and Disambiguation

Within the EMM's NewsExplorer project[2] multilingual tools for geo-tagging [10] and entity disambiguation [11] were developed. We used both systems to extract information about mentions of the entities in the TAC clusters. The extracted features were used as additional features in co-occurrence calculation but also to capture several aspects (places of events and persons involved in investigations).

3.2 Aspects Identified by NEXUS

NEXUS is an event extraction system which analyzes news articles reporting on violent events, natural or man-made disasters (see [5] for detailed system description). The system identifies the type of the event (e.g., flooding, explosion, assassination, kidnapping, air attack, etc.), number and description of the victims, as well as descriptions of the perpetrators and the means, used by them. For example for the text "Three people were shot dead and five were injured in a shootout", NEXUS will return an event structure with three slots filled: The *event type* slot will be set to *shooting*; the *dead victims* slot will be set to *three people*; and the *injured* slot will be set to *five*. Event extraction is deployed as a part of the EMM family of applications, described in [12].

NEXUS relies on a mixture of manually created linguistic rules, linear patterns, acquired through machine learning procedures, plus domain knowledge, represented as domain-specific heuristics and taxonomies. For example, one of the linear patterns for detection of *dead victims* is *[PERSON-GROUP] were shot dead* . The *[PERSON-GROUP]* phrases are recognized by a finite-state grammar. Event type detection is done through a combination of keywords, a Naive Bayes statistical classifier and several domain-specific rules.

NEXUS has been used to analyze online news in several languages and showed reasonable levels of accuracy [5].

We found out that some of the aspects, relevant to the summarization task, correspond to the information extracted by NEXUS. In particular, the aspects

[2] `http://emm.newsexplorer.eu/`

"What happened", "Perpetrators" and "Who affected" have corresponding slots in the event structures of NEXUS.

In our summarization experiments we ran the event extraction system on each news article from the corpus and we mapped extracted slots to summarization aspects. This was done in the following way: The event type (e.g., terrorist attack) was mapped to the aspect "What happened"; the slot "Perpetrator" was mapped to the aspect "Perpetrators"; and the values for the aspect "Victims" were obtained as a union of the event slots: "Dead victims", "Injured", "Arrested", "Displaced", "Kidnapped", "Released hostages" and "People, left without homes". At the end, from a fragment like: "three people died and many were injured", the system will extract two values for the aspect "Who affected", namely "three people" and "many".

3.3 Learning Lexica for Aspect Recognition

Ontopopulis is a system for automatic learning of semantic classes (see [6] for algorithm overview and evaluation). As an input, it accepts a list of words, which belong to a certain semantic class, e.g. "disasters", then it learns additional words, which belong to the same class. Ontopopulis is a multilingual adaptation of a syntactic approach described earlier in [13]. This approach accepts one or several seed sets of terms, each belonging to a semantic class; then, it finds other terms, which are likely to belong to the same semantic class.

Ontopopulis extracts for each semantic class a list of context features, n-grams which tend to occur with the seed set for this class. Each n-gram has a statistical score assigned to it. At the end, for each semantic class, the system finds other terms, which tend to co-occur with its context features. These terms are considered as candidate terms for the corresponding semantic class. For example, if we want to learn words from the class "natural disaster", we can give to Ontopopulis the following seed set *earthquake, flooding, tsunami*. Then, the system learns terms like *mudslides, landslide, tornado, cyclone, flash floods, fire, wildfires*, etc.

Clearly, the system output needs to be manually cleaned, in order to build an accurate lexicon. Since the terms are ordered by reliability (more reliable terms are at the top), the user can review the list, starting at the top, deciding where to stop on the basis of his/her availability or the quality of the list around the point reached within the list. The unrevised items are discarded. Another possibility is to skip the manual reviewing process and take all the terms up to a certain threshold. This approach, however, cannot guarantee very high accuracy.

We learned 4 lexicons, using Ontopopulis, followed by manual cleaning. Each lexicon was relevant to a specific summary aspect. The four aspects covered by our lexicons are: "Damages", "Countermeasures", "Resource", and "Charges". Here we give a short sample from each of the learned lexicons:

1. Damages: damaged, destroyed, badly damaged, extensively damaged, gutted, torched, severely damaged, burnt, burned

2. Countermeasures: operation, rescue operation, rescue, evacuation, treatment, assistance, relief, military operation, police operation, security operation, aid
3. Resource: water, food, species, drinking water, electricity, gas, forests, fuel, natural gas
4. Charges: rape, kidnapping, aggravated, murder, attempted murder, robbery, aggravated assault, theft, armed robbery

The words and multi-word terms from these four lexicons were used to trigger the corresponding summary aspects.

4 Sentence Extraction Based on Co-occurrence and Aspect Information

In this section we describe how the extracted information is combined with lexical features to produce summaries that contain frequently mentioned information (derived from co-occurrence analysis) as well as the required aspects.

4.1 LSA-based Co-occurrence Information

Originally proposed by [3] and later improved by [14], this approach first builds a term-by-sentence matrix from the source, then applies Singular Value Decomposition (SVD) and finally uses the resulting matrices to identify and extract the most salient sentences.

The LSA approach to summarization first builds a term-by-sentence matrix from the source, then applies Singular Value Decomposition (SVD) and finally uses the resulting matrices to identify and extract the most salient sentences. SVD finds the latent (orthogonal) dimensions, which in simple terms correspond to the different topics discussed in the source.

More formally, we first build matrix $\mathbf{A}$ where each column represents the weighted term-frequency vector of sentence j in a given set of documents. The weighting scheme we found to work best is using a binary local weight and an entropy-based global weight (for details see [14]). If we generalize the notion of term to entail, in addition to words, also entities we can obtain a semantically enriched representation.

After that step Singular Value Decomposition (SVD) is applied to the above matrix as $\mathbf{A} = \mathbf{USV}^T$, and subsequently matrix $\mathbf{F} = \mathbf{S} \cdot \mathbf{V}^T$ reduced to r dimensions[3] is derived. This matrix that is passed to the sentence selection phase represents the topics of the cluster identified by co-occurring features.

[3] The degree of importance of each 'latent' topic is given by the singular values and the optimal number of latent topics (i.e., dimensions) r can be fine-tuned on training data.

4.2 Aspect Information

We use the aspects identified by the information extraction tools to boost the co-occurrence-based scores of the sentences that contain the aspects relevant to the corresponding cluster category. For each article cluster we build an aspect-by-sentence matrix **P** which contains boolean values to store the aspects' presence/absence in sentences. For each cluster category a different set of aspects is applied. This matrix is used in the sentence selection process then.

4.3 Sentence Selection

Input to the sentence scoring/selection is formed by matrices **F**, containing information about the most important topics within the cluster, and **P**, containing aspect information.

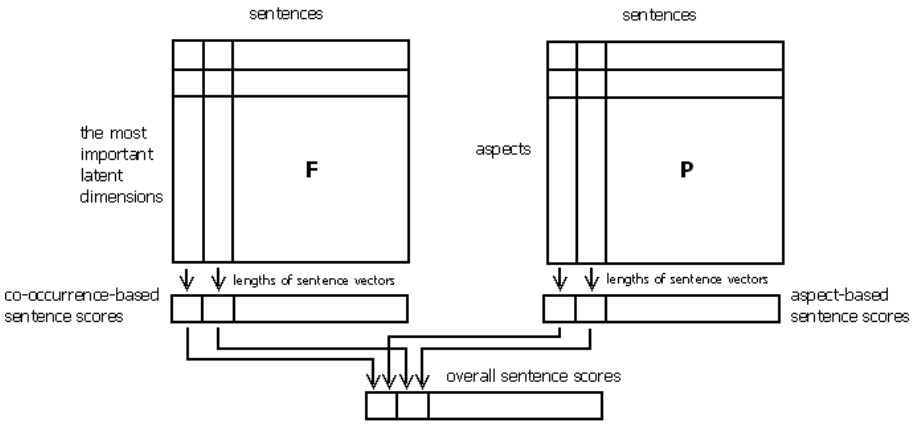

Fig. 1. Sentence selection process.

Sentence selection (see figure 1) starts with measuring the length of sentence vectors in matrix **F**. The length of the vector can be viewed as a measure for importance of that sentence within the top cluster topics. We call it 'co-occurrence sentence score'. For the aspect matrix (**P**) we do the same: measuring the length of sentence vectors. In this case the score corresponds to how many relevant aspects the sentences contain ('aspect-based sentence score'). The two scores are then combined in a way that the aspect-based score works as a booster for the co-occurrence score. The formula for the overall score computation is defined as follows:

$$o_j = |\mathbf{f}_j|(1 + |\mathbf{a}_j|^{bc}). \tag{1}$$

where o_j is the overall score of sentence j, $|\mathbf{f}_j|$ and $|\mathbf{a}_j|$ are its corresponding vectors lengths in matrices **F** and **P**. Coefficient bc can control the impact of aspects on the overall score.

The sentence with the largest overall score is selected as the first to go in the summary (its corresponding vector in $\mathbf{F}$ is denoted as $\mathbf{f}_{best}$, similarly $\mathbf{p}_{best}$ for $\mathbf{P}$). After placing it in the summary, the topic/sentence distribution in matrix $\mathbf{F}$ is changed by subtracting the information contained in that sentence:

$$\mathbf{F}^{(it+1)} = \mathbf{F}^{(it)} - \frac{\mathbf{f}_{best} \cdot \mathbf{f}_{best}^T}{|\mathbf{f}_{best}|^2} \cdot \mathbf{F}^{(it)}, \tag{2}$$

The vector lengths of similar sentences are decreased, thus preventing within-summary redundancy. For aspects, however, we wish to select diverse information as well. But we take a different approach for that. There are cases in which the same aspect should be repeated. For example, for a killing event we want to see the date of the killing and the date when the perpetrator was arrested. Another example are countermeasures. Both following snippets were found important in a model summary of TAC'09 data: *Russian rescue attempts to free and raise the submarine were unsuccessful. Russia requested international help.* Thus, we lower the influence of the aspects already contained in the summary but we do not zero it. Also, we do not use the same formula as in the case of matrix $\mathbf{F}$ because here we are in positive low-dimensional space in comparison with the positive/negative high-dimensional LSA latent space. We use the following formula to update each value in matrix $\mathbf{P}$:

$$\mathbf{p}_{i,j}^{(it+1)} = dc * \mathbf{p}_{i,j}^{(it)}, \quad \text{if} \quad \mathbf{p}_{i,best}^{(it)} > 0. \tag{3}$$

By dc we can control the fadeout of used aspects (a value from 0 to 1).

After the subtraction of information in the selected sentence the process continues with the sentence which has the largest overall score computed from updated matrices $\mathbf{F}$ and $\mathbf{P}$. The process is iteratively repeated until the required summary length is reached.

5 Results and Discussion

The task was to produce a 100-word summary for a set of 10 newswire articles for a given topic, where the topic falls into a predefined set of categories. This was similar to last year's task definition (TAC'09), but as oppposed to last year's event, this years participants (and human summarizers) were given a list of important aspects for each category, and a summary had to cover all those aspects, if possible. The summaries could also contain other information relevant to the topic.

There was an update part of the task this year as at TAC'08 and TAC'09: to write a 100-word update summary of a subsequent 10 newswire articles for the topic, under the assumption that the user has already read the earlier articles.

The defined categories and their aspects were the following:[4]

[4] For full definitions of the aspects see the official task guidelines at `http://www.nist.gov/tac/2010/Summarization/Guided-Summ.2010.guidelines.html`.

1. Accidents and Natural Disasters (1.1 WHAT, 1.2 WHEN, 1.3 WHERE, 1.4 WHY, 1.5 WHO AFFECTED, 1.6 DAMAGES, 1.7 COUNTERMEASURES),
2. Attacks (2.1 WHAT, 2.2 WHEN, 2.3 WHERE, 2.4 PERPETRATORS, 2.5 WHY, 2.6 WHO AFFECTED, 2.7 DAMAGES, 2.8 COUNTERMEASURES),
3. Health and Safety (3.1 WHAT, 3.2 WHO AFFECTED, 3.3 HOW, 3.4 WHY, 3.5 COUNTERMEASURES),
4. Endangered Resources (4.1 WHAT, 4.2 IMPORTANCE, 4.3 THREATS, 4.4 COUNTERMEASURES),
5. Investigations and Trials (5.1 WHO, 5.2 WHO INVOLVED, 5.3 WHY, 5.4 CHARGES, 5.5 PLEAD, 5.6 SENTENCE).

We used several types of information extraction for capturing the aspects. Several aspects were identified by our event extraction system:

- WHAT HAPPENED (used for aspects 1.1, 2.1, 3.1, 5.3): = type of event (e.g. 'bombing');
- WHO AFFECTED (1.5, 2.6, 3.2, 5.1) = number of victims/injured/ displaced etc. (we extracted a full string, not only a number, e.g. '200 soldiers killed');
- PERPETRATORS (2.4, 5.1).

We treated the aspect 5.1 in a special way. For several event types, like 'arrest' the affected person is the one who is investigated, however, for other types of events like 'killing' that person is the perpetrator. This is the reason why we used both WHO AFFECTED and PERPETRATORS slots for capturing the aspect.

The lexical lists of semantically similar terms were generated for capturing the following aspects:

- DAMAGES (1.6, 2.7);
- COUNTERMEASURES (1.7, 2.8, 3.5, 4.4);
- RESOURCE (4.1) = list of resources;
- CHARGES (5.4).

For the identification of temporal expressions (aspects 1.2, 2.2) we produced simple lists of month names etc. Now we work on including a proper temporal analysis.

In the case of aspect 5.2 we took advantage of the fact that we have information about person mentions in the text. This aspect was set in the case that there was a person mentioned in the particular sentence. We took the same approach for locations (1.3 and 2.3). All locations were considered as fillers of that aspect.

We did not deal with the most complex aspects (1.4, 2.5, 3.3, 3.4, 4.2, 4.3, 5.5, 5.6). We simply rely on the fact that they should be captured by the co-occurrence part of the sentence scorer if they seem to be important (frequently mentioned).

We submitted two runs. The first one (RUN-IE) is the complete proposed system: it combines co-occurrence and aspect information. The second run (RUN-CO) represents our baseline system: it uses only co-occurrence information (including lexical and entity co-occurrence). In the remainder of this discussion we refer to the former run as the IE run and the latter as non-IE run (but note that the non-IE run includes the named entity information).

The summaries were evaluated at NIST for content (based on Columbia University's Pyramid method [15]), readability/fluency and overall responsiveness. ROUGE [16] and BE [17] scores were also provided.

The total number of systems this year was 43 including two baselines. The 1st baseline (LEAD) was the first 100 words from the most recent document, the 2nd baseline was the output of the MEAD summarizer [18]. 23 groups participated.

We can analyze 3 types of results. The overall results compare the systems based on all 46 topics (clusters) - basic and update summaries. We have also results for each category. But also, we can see how well we identified each aspect (only pyramid scores are available).

5.1 Overall Results

Table 1 contains the overall TAC results for initial summaries. We report the results and corresponding ranks (in brackets) within all the 43 systems of the two best TAC systems, our two submissions, and the two baselines.

Run ID	Overall responsiveness	Linguistic quality	Pyramid score
16 (the best run in Overall resp.)	**3.17** (1)	3.46 (2)	0.40 (4)
22 (the best run in Pyramid score)	3.13 (2)	3.11 (13)	**0.43** (1)
RUN-IE (co-occurrence+aspects)	2.98 (10)	3.35 (4)	0.37 (18)
RUN-CO (co-occurrence only)	2.89 (19)	3.28 (6)	0.38 (13)
2 (baseline - MEAD)	2.50 (27)	2.72 (29)	0.30 (26)
1 (baseline - LEAD)	2.17 (32)	**3.65** (1)	0.23 (32)

Table 1. TAC'10 results of the Guided summarization task - initial summaries.

In the case of initial summaries the run that included aspects (run 25) performed better in the overall responsiveness and linguistic quality than the run based on co-occurrence only (run 31). It was slightly worse when evaluated by the Pyramid method. We do not report here the evaluation of the number of repetitions, but also in this qualitative measure the aspect-based run was better. The reason could be that we try to select diverse aspects here. Overall, both our systems were ranked high in linguistic quality. One reason could be that sentences that contain full entity mentions, which are used as features in the co-occurrence-based part of the sentence scorer, are getting higher scores. They are usually summary-worthy sentences and are less likely to contain anaphoric references to entities in the preceding context. Our systems performed better than both baselines, with the obvious exception of the LEAD baseline and linguistic quality (the summary is formed by a continuous text from one article).

The score differences between our systems and the best two listed systems (16 and 22) were not significant[5].

5.2 Category-focused Results

Now we continue with the discussion of the results for each category. We report the scores and ranks of both our systems in each cell of the table – the first score and rank correspond to Run 25 (with information extraction-based aspect capturing), the second to Run 31 (co-occurrence only).

Category	Overall responsiveness	Linguistic quality	Pyramid score
1. Disasters	3.00 (23) - **3.57 (2)**	3.43 (3) - 3.29 (5)	0.38 (23) - **0.43 (10)**
2. Attacks	**3.71 (3)** - 2.86 (22)	**3.29 (4)** - 3.00 (16)	**0.56 (6)** - 0.49 (18)
3. Health	**2.75 (6)** - 2.42 (21)	3.33 (6) - 3.25 (9)	0.30 (9) - 0.31 (7)
4. Resources	2.50 (25) - 2.60 (21)	3.60 (3) - 3.40 (6)	0.24 (29) - 0.27 (23)
5. Investigations	3.20 (6) - 3.30 (2)	3.10 (10) - **3.40 (2)**	0.45 (14) - 0.47 (5)

Table 2. Scores and ranks of our runs for each category (RUN-IE – RUN-CO).

In the case of the category "Accidents and natural disasters" the co-occurrence-only approach worked clearly better than the approach with IE. Our simpler run was ranked 2nd in overall responsiveness. The reason of the weaker performance of the IE-based run could be that several times the summarizer selected a sentence that mentioned a historical event, not the event that the cluster was focused on (like a previous earthquake in the same place).

On the contrary, in attacks we can see a really huge improvement with IE: 6th in Pyramids (compared to 18th), 3rd in overall resp. (compared to 22nd) and 4th in linguistic quality (compared to 16th). It could be explained by the fact that this category is the focus of the event extraction system.

In the 'health and safety' category we can notice an improvement when using IE, except for Pyramids. Overall, the runs were ranked high in that category. In the case of 'endangered Resources' the results were poor. We did not focus on this particular category. The linguistic quality, however, showed high levels also for this category.

In the last category, investigations and trials, the system without IE worked better but the differences in the scores were not significant. Our simpler system was ranked high: 2nd in both linguistic quality and overall responsiveness, and 5th in Pyramids.

[5] Here we omit the discussion of the results on update summarization, since our main interest is in the core summarization task.

5.3 Aspect-focused Results

In this section we focus on the most fine-grained results: how well each particular aspect was captured. We can use only Pyramid scores for this evaluation. We report the scores and ranks of our systems and the score of the best system. However, the best score refers to a different system for each aspect.

Firstly, we look at the aspects derived from NEXUS (table 3). Clearly, using type of event as capturing the 'what happened' aspect was not successful. An indicator like 'bombing' seems to be too general for capturing what happened. It could be left to LSA to cover this aspect by selecting the most frequent information. In the case of the aspect 'who affected' there was a large improvement for the attacks category. Roughly speaking, there was no effect in other categories. We noticed also an improvement in update summaries for this aspect. The IE run was successful in capturing also the 'perpetrators' aspect in comparison with the run without IE. Compared to other systems, however, the runs were ranked only slightly above the average.

Aspect	RUN-IE (rank)	RUN-CO (rank)	Best
1.1 WHAT (disasters)	0.60 (24)	**0.79 (3)**	0.89
2.1 WHAT (attacks)	0.74 (21)	0.79 (12)	0.88
3.1 WHAT (health)	0.33 (17)	0.36 (14)	0.58
5.3 REASONS (investigations)	0.46 (19)	**0.59 (6)**	0.67
1.5 WHO AFFECTED (disasters)	0.36 (25)	0.41 (23)	0.68
2.6 WHO AFFECTED (attacks)	**0.65 (2)**	0.54 (11)	0.66
3.2 WHO AFFECTED (health)	0.29 (6)	0.31 (4)	0.39
5.1 WHO (investigations)	0.67 (17)	0.65 (19)	0.96
2.4 PERPETRATORS (attacks)	**0.48 (18)**	0.34 (24)	0.69

Table 3. Pyramid scores and ranks of our runs for each aspect identified by the event extraction system.

Next, we look at the aspects derived from the lexical list generated by Ontopopulis (table 4). In the case of damages we can see worse results with IE in the category disasters . Treating all events in the cluster as equal probably led to selecting sentences, and subsequently also damages, concerned with non-central events. In attacks we can observe, that without IE we did not capture any damage (the score is 0), compared to the 4th best performance with IE. 'Countermeasures' was the category where the IE-based run was very successful in all four categories. It suggests the lexical lists were the right choice for treating this aspect. In resource descriptions there was a non-significant improvement with IE. In capturing charges the co-occurrence information itself performed better.

Among the aspects which were treated by other ways the only successful one was the 'who involved' aspect in investigations. Actually, giving a larger weight to all person mentions did a great job, ranking our IE-based submission

Aspect	RUN-IE (rank)	RUN-CO (rank)	Best
1.6 DAMAGES (disasters)	0.13 (26)	**0.38 (10)**	1.25
2.7 DAMAGES (attacks)	**0.50 (4)**	0 (30)	0.75
1.7 COUNTERMEASURES (disasters)	**0.34 (7)**	0.19 (29)	0.39
2.8 COUNTERMEASURES (attacks)	**0.34 (18)**	0.20 (32)	0.65
3.5 COUNTERMEASURES (health)	**0.31 (1)**	0.24 (7)	0.31
4.4 COUNTERMEASURES (resources)	**0.36 (5)**	0.29 (12)	0.50
4.1 WHAT (resources)	0.49 (19)	0.46 (25)	0.81
5.4 CHARGES (investigations)	0.33 (27)	**0.47 (11)**	0.72

Table 4. Pyramid scores and ranks of our runs for each aspect identified by generated lexical lists.

as the best one. Treating the place aspect the same way was not successful. For capturing time of the events the co-occurrence-driven approach worked well in the case of attacks (2nd).

There are several complex aspects on which we have not worked yet. However, we find that the co-occurrence analysis is able to capture some of those. For instance, we received top rank for identifying reasons for attacks, but in the 'importance of resource' aspect we did not capture anything.

6 Conclusion

We presented an approach to addressing multi-document summarization with an IE-aware perspective. The approach combines a co-occurrence-based summarization system with a mature multilingual event extraction system and a system for the automatic learning of semantically related terms tailored to recognise the required aspects. The results showed positive impact on the clusters that deal with the central focus of the event extraction system - criminal/terrorist attack. Regarding natural disasters, the IE system did not successfully distinguish the recent event from historic events mentioned in the same articles, with a negative impact on summary quality. This can be remedied by preferring information found at the beginning of the articles, or by performing a proper analysis of temporal information in the article. Regarding the coverage of new information aspects, not initially covered by the IE system used, we saw that the automatically generated word lists produced good information extraction and summary results. This shows that we can extend the IE system to more information aspects for which a reasonable base of seed terms can be identified. In the absence of IE patterns to recognise the crucial information aspects, it is more or less left to chance whether these important aspects are covered by the co-occurrence-based summary or not. Our next steps include running and evaluating our IE-aware summarization approach on languages other than English.

References

1. Filatova, E., Hatzivassiloglou, V.: Event-based extractive summarization. In: Text Summarization Branches Out: Proceedings of the ACL-04 Workshop
2. Ellouze, M., Hamadou, A.: Relevant information extraction driven with rhetorical schemas to summarize scientific papers. In: Advances in Natural Language Processing. Lecture Notes in Computer Science, Springer Verlag (2002) 629–642
3. Gong, Y., Liu, X.: Generic text summarization using relevance measure and latent semantic analysis. In: Proceedings of ACM SIGIR, New Orleans, US (2002)
4. Steinberger, J., Poesio, M., Kabadjov, M.A., Ježek, K.: Two uses of anaphora resolution in summarization. Information Processing and Management **43**(6) (2007) 1663–1680 Special Issue on Text Summarisation (Donna Harman, ed.).
5. Tanev, H., Piskorski, J., Atkinson, M.: Real-time news event extraction for global crisis monitoring. In: Proceedings of 13th International Conference on Applications of Natural Language to Information Systems (NLDB 2008). (2008)
6. Tanev, H., Zavarella, V., Linge, J., Kabadjov, M., Piskorski, J., Atkinson, M., R.Steinberger: Exploiting machine learning techniques to build an event extraction system for portuguese and spanish. Journal Linguamatica: Revista para o Processamento Automatico das Linguas Ibericas (2010)
7. Radev, D., McKeown, K.: Generating natural language summaries from multiple on-line sources. Computational Linguistics **24**(3) (1998)
8. White, M., Korelsky, T., Cardie, C., Ng, V., Pierce, D., Wagstaff, K.: Multidocument summarization via information extraction. In: Proceedings of HLT. (2001)
9. Kan, M., McKeown, K.: Information extraction and summarization: Domain independence through focus types. Technical Report CUCS-030-99, Columbia University (1999)
10. Pouliquen, B., Kimler, M., Steinberger, R., Ignat, C., Oellinger, T., Blackler, K., Fuart, F., Zaghouani, W., Widiger, A., Forslund, A., Best, C.: Geocoding multilingual texts: Recognition, disambiguation and visualisation. In: Proceedings of LREC 2006
11. Pouliquen, B., Steinberger, R.: Automatic construction of multilingual name dictionaries. In Goutte, C., Cancedda, N., Dymetman, M., Foster, G., eds.: Learning Machine Translation. MIT Press, NIPS series (2009)
12. Steinberger, R., Pouliquen, B., der Goot, E.V.: An introduction to the europe media monitor family of applications. In: Information Access in a Multilingual World Proceedings of the SIGIR. (2009)
13. Tanev, H., Magnini, B.: Weakly supervised approaches for ontology population. In: Proceedings of the 11th conference of the European Chapter of the Association for Computational Linguistics (EACL). (2006)
14. Steinberger, J., Ježek, K.: Update summarization based on novel topic distribution. In: Proceedings of the 9th DocEng ACM Symposium, Munich, Germany. (2009)
15. Nenkova, A., Passonneau, R.: Evaluating content selection in summarization: The pyramid method. In: Proceedings of the Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL). (2004)
16. Lin, C.Y.: ROUGE: a package for automatic evaluation of summaries. In: Proceedings of the Workshop on Text Summarization Branches Out, Spain (2004)
17. Hovy, E., Lin, C., Zhou, L.: Evaluating duc 2005 using basic elements. In: Proceedings of the DUC. (2005)
18. Radev, D., Otterbacher, J., Qi, H., Tam, D.: Mead reducs: Michigan at duc 2003. In: Proceedings of DUC 2003. (2003)