A Consolidated Open Knowledge Representation for Multiple Texts

Rachel Wities∗1, Vered Shwartz1, Gabriel Stanovsky1, Meni Adler1, Ori Shapira1, Shyam Upadhyay2, Dan Roth2, Eugenio Martinez Camara3, Iryna Gurevych3,4 and Ido Dagan1

1Bar-Ilan University, Ramat-Gan, Israel
2University of Illinois at Urbana-Champaign, IL, USA
3Ubiquitous Knowledge Processing Lab (UKP), Technische Universität Darmstadt
4Ubiquitous Knowledge Processing Lab (UKP-DIPF), German Institute for Educational Research

{rachelvov,vered1986,gabriel.satanovsky,meni.adler,obspl8}@gmail.com
{upadhya3,danr}@illinois.edu, {camara,gurevych}@ukp.informatik.tu-darmstadt.de, dagan@cs.biu.ac.il

Abstract

We propose to move from Open Information Extraction (OIE) ahead to Open Knowledge Representation (OKR), aiming to represent information conveyed jointly in a set of texts in an open text-based manner. We do so by consolidating OIE extractions using entity and predicate coreference, while modeling information containment between coreferring elements via lexical entailment. We suggest that generating OKR structures can be a useful step in the NLP pipeline, to give semantic applications an easy handle on consolidated information across multiple texts.

1 Introduction

Natural language understanding involves identifying, classifying, and integrating information about events and other propositions mentioned in text. While much effort has been invested in generic methods for analyzing single sentences and detecting the propositions they contain, little thought and effort has been put into the integration step: how to systematically consolidate and represent information contributed by propositions originating from multiple texts. Consolidating such information, which is typically both complementary and partly overlapping, is needed to construct multi-document summaries, to combine evidence when answering questions that cannot be answered based on a single sentence, and to populate a knowledge base while relying on multiple pieces of evidence (see Figure 1 for a motivating example). Yet, the burden of integrating information across multiple texts is currently delegated to downstream applications, leading to various partial solutions in different application domains.

This paper suggests that a common consolidation step and a corresponding knowledge representation should be part of the “standard” semantic processing pipeline, to be shared by downstream applications. Specifically, we pursue an Open Knowledge Representation (OKR) framework that captures the information expressed jointly in multiple texts while relying solely on the terminology appearing in those texts, without requiring pre-defined external knowledge resources or schemata.

As we focus in this work on investigating an open representation paradigm, our proposal follows and extends the Open Information Extraction (OIE) approach. We do that by first extracting textual predicate-argument tuples, each corresponding to an individual proposition mention. We then merge these mentions by accounting for proposition coreference, an extended notion of event coreference. This process yields consolidated propositions, each corresponding to a single fact, or assertion, in the described scenario. Similarly, entity coreference links are used to establish reference to real-world entities. Taken together, our proposed representation encodes information about events and entities in the real world, similarly to what is expected from structured knowledge representations. Yet, being an open text-based representation, we record the various lexical terms used to describe the scenario. Further, we model information redundancy and containment among these terms through lexical entailment.

In this paper we specify our proposed representation, while specifying the involved annotation sub-tasks from which our structures are composed. We then describe our annotated dataset, of news headline tweets about 27 news stories, which is

∗Corresponding author
The first to be jointly annotated for all our required sub-tasks. We also provide initial predicted baseline results for each of the sub-tasks, pointing at the limitations of current state of the art.1

Overall, our main contribution is in proposing to create a consolidated representation for the information contained in multiple texts, and in specifying how such representation can be created based on entity and event coreference and lexical entailment. An accompanying contribution is our annotated dataset, which can be used to analyze the involved phenomena and their interactions, and as a development and test set for automated generation of OKR structures. We further note that while this paper focuses on creating an open representation, by consolidating Open IE propositions, future work may investigate the consolidation of other semantic sentence representations, for example AMR (Abstract Meaning Representation) (Banarescu et al., 2013), while exploiting similar principles to those proposed here.

2 Background: Relevant Component Tasks

In this section we describe the prior annotation tasks on which we rely in our representation, as described later in Section 3.

2.1 Open Information Extraction

Open IE (Open Information Extraction) (Etzioni et al., 2008) is the task of extracting coherent propositions from a sentence, each comprising a relation phrase and two or more argument phrases. For example, (plane, landed in, Ankara).

Open IE has gained substantial and consistent attention, and many automatic extractors were created (e.g., Fader et al. (2011); Del Corro and Gemulla (2013)). Open IE’s extractions were also shown to be effective as intermediate sentence-level representation in various downstream applications (Stanovsky et al., 2015; Angeli et al., 2015). Analogously, we conjecture a similar utility of our OKR structures at the multi-text level.

Open IE does not assign roles to the arguments associated with each predicate, as in other single-sentence representations like SRL (Semantic Role Labeling) (Carreras and Márquez, 2005; Palmer et al., 2010). While the former is not consistent in assigning argument slots to the same arguments across different propositions, the latter requires predefined thematic role ontologies. A middle ground was introduced by QA-SRL (He et al., 2015), where predicate-argument structures are represented using question-answer pairs, e.g. (what landed somewhere?, plane), (where did something land?, Ankara).

2.2 Coreference Resolution Tasks

In our representation, we use coreference resolution to consolidate mentions of the same entity or the same event across multiple texts.

Entity Coreference Entity coreference resolution identifies mentions in a text that refer to the same real-world entity (Soon et al., 2001; Ng and Cardie, 2002; Bengtson and Roth, 2008; Clark and Manning, 2015; Peng et al., 2015). In the cross-document variant, Cross Document Coreference Resolution (CDCR), mentions of the same entity can also appear in multiple documents in a corpus (Singh et al., 2011).

Event Coreference Event coreference determines whether two event descriptions (mentions) refer to the same event (Humphreys et al., 1997). Cross document event coreference (CDEC) is a variant of the task in which mentions may occur in different documents (Bagga and Baldwin, 1999).

Compared to within document event coreference (Chen et al., 2009; Araki et al., 2014; Liu et al., 2014; Peng et al., 2016), the problem of cross document event coreference has been relatively under-explored (Bagga and Baldwin, 1999; Bejan and Harabagiu, 2014). Standard benchmarks for
this task are the Event Coreference Bank (ECB) (Bejan and Harabagiu, 2008) and its extensions, that also annotate entity coreference: EECB (Lee et al., 2012) and ECB+ (Cytbulska and Vossen, 2014). See (Upadhyay et al., 2016) for more details on cross document event coreference.

Differently from our dataset described in Section 4, ECB and its extensions do not establish predicate-argument annotations. A secondary line of work deals with aligning predicates across document pairs, as done in Roth and Frank (2012). PARMA (Wolfe et al., 2013) treated the task as a token-alignment problem, aligning also arguments, while Wolfe et al. (2015) added joint constraints to align predicates and their arguments.

Using Coreference for Consolidation Recognizing that two elements are corefering can help in consolidating textual information. In discourse representation theory (DRT), a proposition applies to all co-referring entities (Kamp et al., 2011). In recognizing textual entailment (Dagan et al., 2013), lexical substitution of co-referring elements is useful (Stern and Dagan, 2012). For example, in Figure 1, sentence (1) together with the coreference relation between plane and jet entail that “Turkey forces down Syrian jet.”

2.3 Lexical Inference

Recognizing lexical inferences is an important component in semantic tasks, in order to bridge lexical variability in texts. For instance, in text summarization, lexical inference can help identifying redundancy, when two candidate sentences for the summary differ only in terms that hold a lexical inference relation (e.g. “the plane landed in Ankara” and “the plane landed in Turkey”). Recognizing the inference direction, e.g. that Ankara is more specific than Turkey, can help in selecting the desired granularity level of the description.

There has been consistent attention to recognizing lexical inference between terms. Some methods aim to recognize a general lexical inference relation (e.g. (Kotlerman et al., 2010; Turney and Mohammad, 2015)), others focus on a specific semantic relation, mostly hypernymy (Hearst, 1992; Snow et al., 2005; Santus et al., 2014; Shwartz et al., 2016), while recent methods classify a pair of terms to a specific semantic relation out of several (Baroni et al., 2012; Weeds et al., 2014; Pavlick et al., 2015; Shwartz and Dagan, 2016). It is worth noting that most existing methods are indifferent to the context in which the target terms occur, with the exception of few works, which were mostly focused on a narrow aspect of lexical inference, e.g. lexical substitution (Melamud et al., 2015).

Determining entailment between predicates is a different sub-task, which has also been broadly explored (Lin and Pantel, 2001; Duclaye et al., 2002; Szpektor et al., 2004; Schoenmackers et al., 2010; Roth and Frank, 2012). Berant et al. (2010) achieved state-of-the-art results on the task by constructing a predicate entailment graph optimizing a global objective function. However, performance should be further improved in order to be used accurately within semantic applications.

3 Proposed Representation

Our Open Knowledge Representation (OKR) aims to capture the consolidated information expressed jointly in a set of texts. In some analogy to structured knowledge bases, we would like the elements of our representation to correspond to entities in the described scenario and to statements (propositions) that relate them. Still, in the spirit of Open IE, we would like the representation to be open, while relying only on the natural language terminology in the given texts without referring to predefined external knowledge.

This section specifies our proposed structure, with a running example in Figure 2. The specification involves two aspects: the first is defining the component annotation sub-tasks involved in creating our representation, following those reviewed in Section 2; the second is specifying how we derive from these component annotations a consolidated representation. These two aspects are interleaved along the presentation, where for each step we first describe the relevant annotations and then how we use them to create the corresponding component of the representation.

3.1 Entities

To represent entities, we first annotate the text by entity mentions and coreference. Following the typical notion for these tasks, an entity mention corresponds to a word or multi-word expression that refers to an object or concept in the described scenario (in the broader sense of “entity”). Accordingly, we represent an entity in the described scenario by the coreference cluster of all its mentions. We represent the coreferring cluster of mentions by the multiset of its terms, keeping pointers
to each term’s mentions (see Entities in Figure 2; to avoid clutter, pointers are not presented in the figure). We note that we take an inclusive view which regards concepts as entities, for example the adjective Syrian is considered an entity mention that may corefer with Syria.

3.2 Proposition Mentions and Consolidated Propositions

To represent propositions, we first annotate Open IE style extractions, which we term proposition mentions. Each mention consists of a predicate expression, e.g. around verbs or nominalizations, and a set of arguments (see Proposition Mentions in Figure 2). We deviate slightly from standard Open IE formats by representing the predicate expression as a template, with place holders for the arguments (marked with brackets in the figure). This follows the common representation of predicates within predicate inference rules, as in DIRT (Lin and Pantel, 2001), and allows the span of entity arguments to correspond exactly to the entity term. Further, as typical in Open IE, modalities and negations become part of the lexical elements of the predicate. Notice that at this stage an argument mention is already associated with its corresponding entity. Further, we annotate implicit predicates when a predication between two entities is implied, without an explicit predicate expression, as common for noun modifications ($P_2$ in the figure). Nested propositions are represented by having one proposition mention as an argument of the other (e.g. “the [plane] was forced to [land in Ankara]”).

Figure 2: An illustration of our OKR formalism (a), with a corresponding graphical view of the consolidated structure (b). In (b), dashed lines connect entities to their instantiation within arguments, while allowing graph-traversal inferences such as: what is the relation between Turkey and Russia? Turkey intercepted a plane that carried ammunition from Russia (the path from $E_1$ to $E_6$ via the darker dashed lines).
To link different mentions of the same real world fact, we annotate proposition coreference, which generalizes the notion of event coreference to cover all types of predications (e.g., *John is Mary’s brother* would co-refer with *Mary is John’s sister*). This annotation specifies the coreference relation for a cluster of proposition mentions (denoted by the same proposition index $P_i$ in Figure 2), as well as an alignment of their arguments, (denoted by matching argument indexes within the same proposition cluster). We then consider a *proposition* to correspond to a coreference cluster of proposition mentions, which jointly describe the referred real-world fact.

Yet, a cluster of co-referring proposition mentions does not provide a succinct representation for the aggregated textual description of a proposition. To that end, we aggregate the information in the cluster into a *Consolidated Proposition*, composed of a *consolidated predicate* and *consolidated arguments*. Similar to entity representation, a consolidated predicate is represented by the set of all predicate expressions appearing in the cluster. A consolidated argument is specified by the set of all entities (or propositions, in case of having one proposition being an argument of another one) that occupy this argument’s slot in the different mentions. As with entities, each element in this representation is accompanied by a set of pointers to all its original mentions (omitted from the figure). A graphical illustration of this structure is given in Figure 2(b) (for now, ignore the arrows within some of the nodes).

A consolidated proposition encodes compactly all possible textual descriptions for the referred proposition, which can be generated from its mentions taken jointly. Each description can be generated by picking one possible predicate expression and then picking one possible lexical choice for each argument. For example, $P_1$ may be described as *Turkey intercepted a plane*, *Turkey forces down a jet* etc. Some of these descriptions correspond to original mentions in the text, while others can be induced through coreference (as reviewed at the end of Section 2.2). The representation of a consolidated proposition thus does not depend on the particular way in which lexical choices were split across the different proposition mentions.

### 3.3 Lexical Entailment Graphs

The set of descriptions encoded in a consolidated proposition is highly redundant. To make it more useful, we would like to model the information overlap between different lexical choices. For example, we want to know that *Turkey intercepted a plane* is more general than, or equivalently, is entailed by, *Turkey intercepted a jet*. To that end, we annotate the lexical entailment relations between the elements in each component of our representation, that is, within each consolidated predicate, consolidated argument and entity. This yields a *lexical entailment graph* within each component (see figure 2), which models the information containment relationships between different descriptions.

Notice that in our setting the lexical entailment relation is considered within the given context (see Section 2.3). For example, *grounded* and *forced down* may not be generically synonymous, but they do convey equivalent information in a given context of forcing a flying plane to land. Contradictions are modeled to a limited extent, by annotating contradiction relations (in context) between elements of our entailment graphs, for example when different figures are reported for the number of casualties in a disaster. This is a natural representation, since contradiction is often modeled within a three-way entailment classification task. Modeling of broader cases of contradiction is left for future work.

The entailment graphs yield better modeling of the supporting text mentions (and their total count) for each possible description. For example, knowing that *Moscow* entails *Russia*, we can assume in $P_3$ two supporting mentions for knowing that the ammunition was carried from Russia, while having only one supporting mention for the more detailed information regarding Moscow being the origin. Such frequency support often correlates with confidence and prominence of information, which, together with generality modeling, may be very useful in applications such as multidocument summarization or question answering. Finally, the graphical view of our representation lends itself to graph-based inferences, such as looking for all connections between two entities, similar to aggregated inferences over structured knowledge graphs (see example in Figure 2(b)).

In summary, our open knowledge representation consists of the following: *entities*, generated
by detecting entity mentions and coreference; consolidated propositions, composed of consolidated predicates and arguments, which are generated by detecting proposition mentions and coreference relations between them; lexical entailment graphs for entities, consolidated predicates and consolidated arguments, which specify the inference relations between the elements within each of these components. This yields a compact representation of all possible descriptions of the statements jointly asserted by the set of texts, as induced via coreference-based inference, while tracking information containment between different descriptions as well as tracking their (induced) supporting mentions.

4 News-Related Tweets Dataset

Following the formal definition of our OKR structures, we compiled a corpus with gold annotations of our 5 subtasks (listed in Table 1). As outlined in the previous section, our structures follow deterministically from these annotations. Specifically, we make use of the news-related tweets collected in the Twitter Event Detection Dataset (McMinn et al., 2013), which clusters tweets from major news networks and other sources discussing the same event (for example, the grounding of a Syrian plane by the Turkish government). We chose to annotate news related tweets in this first dataset for several reasons: (1) they represent self contained assertions, (2) they tend to be relatively factual and succinct, and (3) by looking at several news sources we can obtain a corpus with high redundancy, which our representation aims to address.

We note that while this dataset exhibits a limited amount of linguistic complexity, making it suitable for a first investigation, it still represents a very practical use case of consolidating information in a large stream of tweets about a news story. This annotation serves two main purposes. First, it validates the feasibility of our annotation scheme in terms of annotator requirements, training and agreement. Second, to the best of our knowledge, this is the first time these core NLP annotations are annotated in parallel over the same texts. Following, this annotation has the potential of becoming a useful resource spurring future research into joint prediction of these annotations. For instance, predicate argument structures may benefit from co-reference signals, and entity extraction systems may exploit signals from lexical entailment.

Overall, we annotated 1257 tweets from 27 clusters. We release the dataset both in full OKR format, as well as ECB-like “light” format, containing only the annotated co-reference chains. Overall corpus statistics are depicted in Table 2.

4.1 Dataset Characteristics

An analysis of the annotations reveals interesting and unique characteristics of our annotated corpus.

First, the part of speech distribution of entities and predicates (Table 3) shows that our corpus contains a vast number of non-verbal predications (67%), and a relatively large number of adjectival entities, owing to the fact that our structure annotates concepts such as “northern” or “Syrian” as entities in an implicit relation.

Second, the average number of unique lemmas per entity and proposition chains (2.00 and 2.24, respectively) shows that our corpus exhibits a fair amount of non-trivial lexical variability.

Finally, roughly 96% of our entailment graphs (entity and proposition) form a connected component. This data provides an interesting potential for investigating and modeling lexical inference relations within coreference chains.

4.2 Annotation Procedure and Agreement

The annotation was performed by two native English speakers with linguistic academic background, which had 10 hours of in house training. The entire annotation process took 200 person-hours using a graphical tool purposely-designed to facilitate the incremental annotation for all subtasks. We employ the QA-SRL annotation methodology to help determining Open IE predicate and argument spans in the gold standard, for its intuitiveness for non-expert annotators (He et al., 2015). Five clusters were annotated independently by both annotators and were used to measure their agreement on the task. The other clusters were annotated by one annotator and reviewed by an expert.

We measure agreement separately on each annotation subtask. After each task in our pipeline we keep only the consensual annotations. For example, we measure entity coreference agreement only for entity mentions that were annotated by both annotators. For entity, predicate and argument mention agreement, we average the accuracy
of the two annotators, each computed while taking the other as a gold reference.

For entity, predicate, and argument co-reference we calculated coreference resolution metrics: the link-based MUC (Vilain et al., 1995), the mention-based $B^3$ (Bagga and Baldwin, 1998), the entity-based CEAf, and the widely adopted CoNLL $F_1$ measure which is an average of the three. For entity and proposition entailment we compute the $F_1$ score over the annotated directed edges in each entailment graph, as is common for entailment agreement metrics (Berant et al., 2010).

We macro-averaged these scores to obtain an overall agreement on the 5 events annotated by both annotators. The agreement scores for the two annotators are shown in Table 1, and overall show high levels of agreement. A qualitative analysis of the more common disagreements between annotators is shown in Table 4.

Overall, this shows that our parallel annotation is indeed feasible; agreement on each of the sub-tasks is relatively high and on par with reported inter-annotator agreement on similar tasks.

### Table 1: Inter-Annotator Agreement (top) and off-the-shelf state-of-the-art predicted performance (bottom, see Section 5) for the OKR subtasks: (1) Entity mention extraction (for prediction we use F1 score) (2) Entity co-reference (3) Proposition Extraction (predicate identification and argument detection) (4) Proposition Co-reference (predicate coreference and argument alignment), and (5) Entailment graphs (entity and proposition entailment; argument entailment figures are not presented due to very low statistics). † Numbers in parenthesis denote verbal vs. non-verbal predicates, respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>avg</td>
<td>acc</td>
<td>wMUC</td>
<td>$\delta^3$</td>
<td>CEAF</td>
</tr>
<tr>
<td>IAA Pred</td>
<td>.85</td>
<td>.87</td>
<td>.92</td>
<td>.90</td>
<td>.74 (93, 72$^\dagger$)</td>
</tr>
<tr>
<td>Pred</td>
<td>.58</td>
<td>.84</td>
<td>.89</td>
<td>.81</td>
<td>.41 (73, 25$^\dagger$)</td>
</tr>
</tbody>
</table>

### Table 2: Twitter dataset statistics. Distinct lemma terms per proposition chain were calculated only on explicit propositions. Average number of elements per argument chain measures how many distinct entities or propositions were part of the same argument.

<table>
<thead>
<tr>
<th>POS</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Adj's</th>
<th>Impl.</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ent. Dist.</td>
<td>.85</td>
<td>.01</td>
<td>.09</td>
<td></td>
<td>.05</td>
</tr>
<tr>
<td>Pred. Dist.</td>
<td>.40</td>
<td>.33</td>
<td>.04</td>
<td></td>
<td>.18</td>
</tr>
</tbody>
</table>

### Table 4: Typical cases of annotator disagreements.

Annotated spans are denoted by square brackets, subscript denotes label for the mention (predicate, argument or entity).

### Table 3: Entity and Predicate distribution across part of speech tags: nouns, verbs, adjectives, non-lexicalized (implicit) and all others.

<table>
<thead>
<tr>
<th>Disagreement Type</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrasal verbs</td>
<td>placed to test</td>
</tr>
<tr>
<td>Nominalizations</td>
<td>[suspect]$_{2,7}$, plane</td>
</tr>
<tr>
<td>Entailment</td>
<td>fuel $\rightarrow$ gas vs. gas $\rightarrow$ fuel</td>
</tr>
</tbody>
</table>

### 5 Baselines

As we have shown in previous sections, our structure is derived from known “core” NLP tasks, extended where needed to fit our consolidated representation. Subsequently, a readily available means of automatically recovering OKR is through a pipeline which uses off-the-shelf models for each of the subtasks.

To that end, we employ publicly available tools and simple baselines which approximate the current state-of-the-art in each of these subtasks. For brevity sake, in the rest of the section we briefly describe each of these baselines. For a more detailed technical description see the OKR repository (https://github.com/vered1986/OKR).

For Entity Mention extraction we use the spaCy NER model\(^2\) in addition to annotating all of the nouns and adjectives as entities. For Proposition Mention detection we use Open IE propositions extracted from PropS (Stanovsky et al., 2016), where non-restrictive arguments were reduced following Stanovsky and Dagan (2016). For Proposi-

---

\(^2\)https://spacy.io/
tion and Entity coreference, we clustered the entity mentions based on simple lexical similarity metrics (e.g., lemma matching and Levenshtein distance), shown to be effective on our news tweets.

3 For Argument Mention detection we attach the components (entities and propositions) as arguments of predicates when the components are syntactically dependent on them. Argument Co-reference is simply predicted by marking co-reference if and only if the arguments are both mentions of the same entity co-reference chain. For Entity Entailment purposes we used knowledge resources (Shwartz et al., 2015) and a pre-trained model for HypeNET (Shwartz et al., 2016) to obtain a score for all pairs of Wikipedia common words (unigrams, bigrams, and trigrams). A threshold for the binary entailment decision was then calibrated on a held out development set. Finally, for Predicate Entailment we used the entailment rules extracted by Berant et al. (2012).

5.1 Results and Error Analysis

Using the same metrics used for measuring inter-annotator agreement, we evaluated how well the presented models were able to recover the different facets of the OKR gold annotations. The performance on the different subtasks is presented in Table 1 (bottom).

We measure the performance of each component separately, while taking the annotations for all previous steps from the gold human annotations. This allows us to examine the performance of the current component, alleviating any incurred errors from previous steps. Thus, we can identify technological “bottle-necks” – the steps which most significantly lower predicted OKR accuracy using current off-the-shelf tools.

First, we noticed that non-verbal predicates pose a challenge for current verb-centric systems. This primarily manifests in low scores for identifying entities, predicates and arguments. Many entity mention errors are due to nominalizations mistakenly annotated as entities. When excluding gold nominalizations, the entity mention baseline F1 score rises from 0.58 to 0.63. As mentioned earlier (Section 4.2) nominalizations were also one of the main challenges for the annotators. Furthermore, recognizing nominalizations and other non-verbal predicates, which are very common in our dataset (see Table 3), proves to be a difficult task. Indeed, we see a significant improvement in performance when comparing verbal predicate mention performance to non-verbal performance (accuracy of 0.73 vs. 0.25). Finally, argument identification was hard mainly because of inconsistences in verbal versus nominal predicate-argument structure in dependency trees.4

The low performance in predicate coreference compared to entity coreference can be explained by the higher variability of predicate terms. The argument co-reference task becomes easy given gold predicate-argument structures, as most arguments are singletons (i.e. composed of a single element).

Finally, while the performance of the predicate entailment component reflects the current state-of-the-art (Berant et al., 2012; Han and Sun, 2016), the performance on entity entailment is much worse than the current state-of-the-art in this task as measured on common lexical inference test sets. We conjecture that this stems from the nature of the entities in our dataset, consisting of both named entities and common nouns, many of which are multi-word expressions, whereas most work in entity entailment is focused on single word common nouns. Furthermore, it is worth noting that our annotations are of naturally occurring texts, and represent lexical entailment in real world coreference chains, as opposed to synthetically compiled test sets which are often used for this task.

While several tasks achieve reasonable performance on our datasets, most tasks leave room for improvement. These bottle-necks are bound to hinder the performance of a pipeline end-to-end system. Future research into OKR should first target these areas; either as a pipeline or in a joint learning framework.

6 Applications and Related Work

The need to consolidate information originating from multiple texts is common in applications that summarize multiple text into some structure, such as multi-document summarization and knowledge-base population. Currently, there is no
systematic solution, and the burden of integrating information across multiple texts is delegated to downstream applications, leading to partial solutions which are geared to specific applications.

**Multi-Document Summarization (MDS)** (Barzilay et al., 1999) is a task whose goal is to produce a concise summary from a set of related text documents, such that it includes the most important information in a non-redundant manner. While extractive summarization selects salient sentences from the document collection, abstractive summarization generates new sentences, and is considered a more promising yet more difficult task.

A recent approach for abstractive summarization generates a graphical representation of the input documents by: (1) parsing each sentence/document into a meaning representation structure; and (2) merging the structures into a single structure that represents the entire summary, e.g. by identifying coreferring items.

In that sense, this approach is similar to OKR. However, current methods applying this approach are still limited. Gerani et al. (2014) parse each document to discourse tree representation (Joty et al., 2013), aggregating them based on entity coreference. Yet, they work with a limited set of (discourse) relations, and rely on coreference only between entities, which was detected manually.

Similarly, Liu et al. (2015) parse each input sentence into an individual AMR graph (Banarescu et al., 2013), and merge those into a single graph through identical concepts. This work extends the AMR formalism of canonicalized representation per entity or event to multiple sentences. However, they only focus on certain types of named entities, and collapse two entities based on their names rather than on coreference.

**Event-Centric Knowledge Graphs (ECKG)** (Vossen et al., 2016; Rospocher et al., 2016) is another related work which represent news articles as graphs. Event nodes are linked to DBPedia (Auer et al., 2007), with the goal of enriching entities and events with dynamic knowledge. For example, an event describing the interception of the Syrian plane by Turkey will be linked to Syria and Turkey.

We propose that OKR can help the described applications by providing a general underlying representation for multiple texts, obviating the need to develop specialized consolidation methods for each application. We can expect the use of OKR structures in MDS to shift the research efforts in this task to other components, e.g. generation, and eventually contribute to improving state of the art on this task. Similarly, an algorithm creating the ECKG structure can benefit from building upon a consolidated structure such as OKR, rather than working directly on free text.

7 Conclusions

In this paper we advocate the development of representation frameworks for the consolidated information expressed in a set of texts. The key ingredients of our approach are the extraction of proposition structures which capture individual statements and their merging based on entity and event coreference. Coreference clusters are proposed as a handle on real world entities and facts, while still being self-contained within the textual realm. Lexical entailment is proposed to model information containment between different textual descriptions of the same real world components.

While we developed an “open” KR framework, future work may investigate the creation of similar models based on structures that do refer to external resources (such as PropBank, as in Abstract Meaning Representation – AMR). Gradually, fine grained semantic phenomena may be addressed, such as factuality, attribution and modeling sub-events and cross-event relationships. Finally, we plan to investigate performing the core annotation sub-tasks via crowdsourcing, for scalability.

Acknowledgments

This work was supported in part by grants from the MAGNET program of the Israeli Office of the Chief Scientist (OCS) and the German Research Foundation through the German-Israeli Project Cooperation (DIP, grant DA 1600/1-1), and by Contract HR0011-15-2-0025 with the US Defense Advanced Research Projects Agency (DARPA).

References

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging linguistic structure for open domain information extraction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages


Vincent Ng and Claire Cardie. 2002. Improving machine learning approaches to coreference resolution.


