Cross-document Event Extraction and Tracking: Task, Evaluation, Techniques and Challenges

Heng Ji[*]   **Ralph Grishman**[†]   **Zheng Chen**[*]   **Prashant Gupta**[‡]

*Computer Science Department, Queens College and The Graduate Center,
The City University of New York
†Computer Science Department, New York University
‡Computer Science Department, Indian Institute of Information Technology Allahabad
hengji@cs.qc.cuny.edu

Abstract

This paper proposes a new task of cross-document event extraction and tracking and its evaluation metrics. We identify important person entities which are frequently involved in events as 'centroid entities'. Then we link the events involving the same centroid entity along a time line. We also present a system performing this task and our current approaches to address the main research challenges. We demonstrate that global inference from background knowledge and cross-document event aggregation are crucial to enhance the performance. This new task defines several extensions to the traditional single-document Information Extraction paradigm beyond 'slot filling'.

Keywords

Information Extraction, Cross-document Extraction, Event Extraction

1. Introduction

Consider a user monitoring or browsing a multi-source news feed, with assistance from an Information Extraction (IE) system. Various events are evolving, updated, repeated and corrected in different documents; later information may override earlier more tentative or incomplete facts. In this environment, traditional single-document IE would be of little value; a user would be confronted by thousands of *unconnected* events with tens of thousands of arguments. Add to this the fact that the extracted results contain *unranked, redundant* and *erroneous* facts and some crucial facts are *missing*, and it's not clear whether these IE results are really beneficial. How can we take proper advantage of the power of extraction to aid news analysis? In this paper we define a new cross-document IE task beyond 'slot filling' to generate more *coherent, salient, complete* and *concise* facts.

A high-coherence text has fewer conceptual gaps and thus requires fewer inferences and less prior knowledge, rendering the text easier to understand [1]. In our task, text coherence is the extent to which the relationships between events in a text can be made explicit. We aim to explore various new methods of identifying the logical connections among events to reveal this coherence. We aim to provide a more coherent presentation by linking events based on shared arguments. In the news from a certain period some entities are more central than others;

we propose to identify these *centroid entities,* and then link the events involving the same centroid entity on a time line. In this way we provide coherent event chains so that users can more efficiently review and analyze events, such as tracking a person's movement activities and trends. This will offer a richer set of views than is possible with document clustering for summarization or with topic tracking.

To sum up, the specific goals of this paper are to:
- Formulate a tractable but challenging task of cross-document IE, producing a product useful for browsing, analysis, and search;
- Propose a set of metrics for this task;
- Present a first cut at a system for performing this task;
- Lay out the potential research challenges and suggest some directions for improving this system's performance.

2. Traditional Single-document IE and Its Limitations

We shall start by illustrating, through the ACE [1] event extraction task, the limitations of traditional single-document IE.

2.1 Terminology and Task

ACE defines the following terminology:

entity: an object or a set of objects in one of the semantic categories of interest, e.g. persons, locations, organizations.

mention: a reference to an entity (typically, a noun phrase)

relation: one of a specified set of relationships between a pair of entities.

event: a specific occurrence involving participants, including 8 types of events, with 33 subtypes; for the purpose of this paper, we will treat these simply as 33 distinct event types.

event mention: a phrase or sentence within which an event is described.

event trigger: the main word which most clearly expresses an event occurrence.

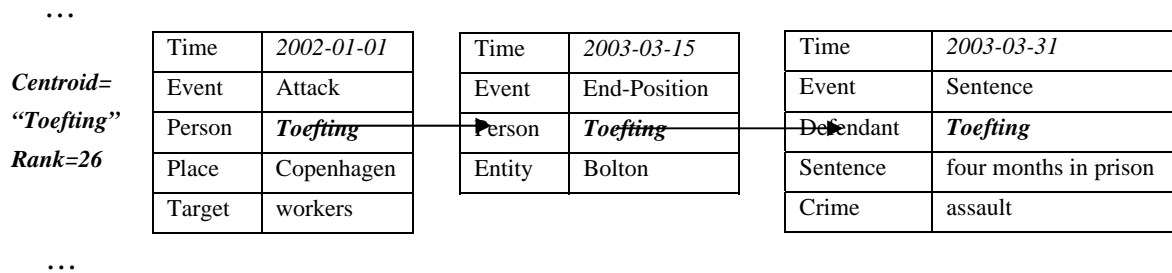**event argument**: an entity involved in an event with some specific role.

---

[1] http://www.nist.gov/speech/tests/ace/

Centroid=
"Toefting"
Rank=26

…

Time	2002-01-01
Event	Attack
Person	*Toefting*
Place	Copenhagen
Target	workers

Time	2003-03-15
Event	End-Position
Person	*Toefting*
Entity	Bolton

Time	2003-03-31
Event	Sentence
Defendant	*Toefting*
Sentence	four months in prison
Crime	assault

…

Figure 1. Example of Ranked Cross-document Temporal Event Chains

event time: an exact date normalized from time expressions and a role to indicate that an event occurs before/after/within the date.

For example, for the following text:

Barry Diller on Wednesday quit as chief of Vivendi Universal Entertainment, the entertainment unit of French giant Vivendi Universal.

A single-document ACE IE system should detect the following information:

entity: person: {Barry Diller, chief}; …

mention: person: "Barry Diller"; …

relation: part-whole: "the entertainment unit" is part of "French giant"

event mention: Personnel_End Position event: "Barry Diller on Wednesday quit as chief of Vivendi Universal Entertainment."

event trigger: quit ; **event time**: Wednesday (2003-03-05).

event argument: position: chief**; person: "Barry Diller"**

2.2 Evaluation Metrics

As for other ACE tasks, the ACE 2005 official evaluation scorer can produce an overall score called "ACE value" for event extraction. However, most of the ACE event extraction literature (e.g. [3]; [4]) used a simpler argument-based F-measure to evaluate ACE event extraction, and we will adapt this measure to our task.

2.3 Limitations

In the ACE single-document event extraction task, each event mention is extracted from a single sentence. The results are reasonably useful for hundreds of documents. However, when we apply the same system to process much larger corpora, the net result is a very large collection of events which are:

(1) Unconnected. Related events (for example, "Tony Blair's foreign trips) appear unconnected and unordered.

(2) Unranked. Event mentions are presented in the order in which they appear in the corpus and considered equally important. It will be beneficial to rank the myriad events by some salience criteria. Centroid-based multi-document text summarization (e.g. [5]; [6]) detects the 'centroid' words that are statistically important to a cluster of documents, and then ranks sentences by incorporating centroid word confidence values. We will adopt the same strategy to rank event arguments.

(3) Redundant. More critically, many events are frequently repeated in different documents. Cross-document event aggregation is essential, in order to enable the users to access novel information more rapidly.

(4) Erroneous and Incomplete. Like many other NLP applications, ACE event extraction systems faced a 'performance ceiling', -- they barely exceeded 50% F-score on argument labeling. Some extraction errors came from limitations on the use of facts already extracted from other documents because of the single-document extraction paradigm.

3. A New Cross-document IE Task

As one initial attempt to address the limitations as described above, we shall propose a cross-document IE task. Since this task is quite new to the IE community, there is no baseline system covering all the aspects for comprehensive comparison. Therefore it is valuable to develop new task standards (section 3.1) and scoring metrics (section 3.2). We shall elaborate the motivations for these changes over the traditional IE task.

3.1 Terminology and Task Definition

We extend the ACE terminology from single document to cross-document setting, and define the following new terminology:

centroid entities: N person entities most frequently appearing as arguments of events.

temporal event chain: a list of temporally-ordered events involving the same centroid entity.

Our cross-document IE task is defined as follows:

Input: A test set of documents

Ouput: Identify a set of centroid entities, and then for each centroid entity, link and order the events centered around it on a time line. For example, Figure 1 presents a temporal event chain involving "Toefting".

3.2 Evaluation Metrics

We introduce the following new measures to gauge the effectiveness of a cross-document IE system.

(1) Centroid Entity Detection

To measure how well a system performs in *selecting the correct centroid entities* in a set of documents, we compute the Precision, Recall and F-measure of the top N centroid entities identified by the system as a function of N (the value of N can be considered as reflecting the 'compression ratio' in a summarization task):

- *A centroid entity is correctly detected* if its substring matches a reference centroid.

In the reference the centroids are the top N entities ranked by the number of events in which that entity appears as an argument.

For those correctly identified centroid entities, we will use a standard ranking metric, normalized Kendall tau distance [7], to evaluate how a system performs in ranking:

- Normalized Kendall tau distance (Centroid Entities) = the fraction of correct system centroid entity pairs out of salience order.
- Centroid Entity Ranking Accuracy = 1- Normalized Kendall tau distance (Centroid Entities)

(2) Browsing Cost: Incorporate Novelty/Diversity into F-Measure

It's important to measure how well a system performs at *presenting the events involving the centroid entities accurately*. The easiest solution is to borrow the argument based F-Measure in the traditional IE task. However, as we pointed out in section 2.3(3), many events are reported redundantly across multiple documents, we should incorporate novelty and diversity into the metric and assign penalties to the redundant event arguments. We define an evaluation metric *Browsing Cost which is* similar to the Search Length i metric [8] for this purpose:

- *An argument is correctly extracted* in an event chain if its event type, string (the full or partial name) and role match any of the reference argument mentions.
- *Two arguments in an event chain are redundant* if their event types, event time, string (the full or partial name) and roles overlap.
- *Browsing Cost (i)* = the number of incorrect or redundant event arguments that a user must examine before finding i correct event arguments.

If an event chain contains more redundant information, then the browsing cost is larger. We examine the centroid entities in rank order and, for each argument, the events in temporal order, inspecting the arguments of each event.

(3) Temporal Correlation: Measure Coherence

Since the traditional IE task doesn't evaluate event ordering, we shall use the correlation metric to evaluate how well a system performs at *presenting the events in proper temporal order for each event chain*. Assume the event chain *ec* includes a set of correct arguments *argset*, then the *temporal correlation* is measured by:

- *Temporal Correlation (ec)* = the correlation of the temporal order of *argset* in the system output and the answer key.

The overall system performance is measured by the average value of the temporal correlation scores of all the event chains. In assessing temporal correlation, we should also take into account the number of argument pairs over which temporal order is measured:

- *Argument recall* = number of unique and correct arguments in response / number of unique arguments in key

The general idea follows the event ordering metric in TempEval [9], but we evaluate over event arguments instead of triggers because in our task the representation

of a node in the chain is extended from an event trigger to a structured aggregated event including fine-grained information such as event types, arguments and their roles. Also similar to TempEval we focus more on the temporal order of events instead of the exact date associated with each individual event. This is different from other time identification and normalization tasks such as TERN[2]. We believe for a cross-document IE task, the exact date normalization results are less crucial. In some cases the system can insert the events into the correct positions in the chains even by only detecting rough date periods (e.g. "a few weeks ago"). Our temporal correlation metric is able to assign appropriate credit to these cases.

4. A Cross-document IE System

We have developed a system performing this new cross-document IE task.

4.1 System Overview

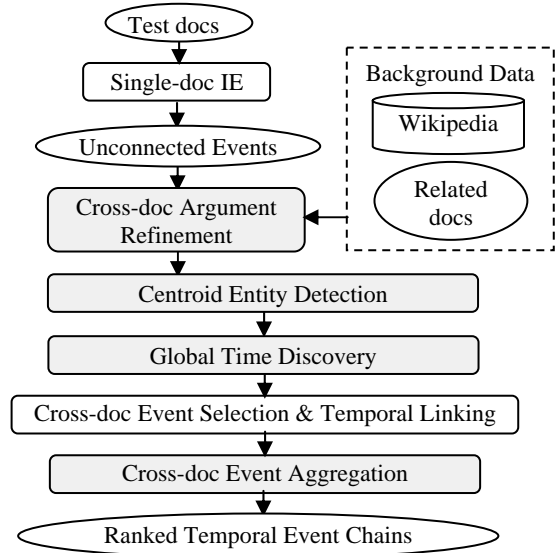Figure 2 depicts the overall architecture of our system.

Figure 2. Cross-document IE System Overview

4.2 Baseline Single-document IE System

We first apply a state-of-the-art English ACE single-document IE system [10] which can extract events from individual documents. This IE system includes entity extraction, time expression extraction and normalization, relation extraction and event extraction. The event extraction component combines pattern matching with a set of Maximum Entropy (MaxEnt) classifiers for trigger labeling, argument identification and argument classification. Each of these classifiers produces local confidence values [3].

4.3 What's New

The following sections will describe the various challenges in this new task and our current techniques to address them, including:

- More Salient: Detecting centroid entities using global confidence (section 5);
- More Accurate and Complete: Correcting and enriching arguments from the background data (section 6);
- More Concise: Conducting cross-document event aggregation to remove redundancy (section 7).

Except for the cross-document argument refinement techniques in section 6.1 which is based on prior work, all other components are newly developed in this paper.

5. Centroid Entity Detection

After we harvest a large inventory of events from single-document IE, we label those person entities involved frequently in events with high confidence as centroid entities. We first construct the candidates through a simple form of cross-document coreference (section 5.1) and then rank these candidates (section 5.2).

5.1 Cross-document Name Coreference

We merge two person name mentions $<mention_i, mention_j>$ into one candidate centroid if they satisfy either of the following two conditions:

- identified as coreferential by single-document coreference resolution; or
- in different documents, there is a $name_i$ referring to $mention_i$ and a $name_j$ referring to $mention_j$ (if several names, taking the maximal name in each document), and $name_i$ and $name_j$ are equal or one is a substring of the other.

Using this approach we can avoid linking "Rod Stewart" and "Martha Stewart" into the same entity. In the future we intend to exploit more advanced cross-document person name disambiguation techniques (e.g. [11], [12]) to resolve ambiguities.

5.2 Global Entity Ranking

Because the candidate entities are extracted automatically, and so may be erroneous, we want to promote those arguments which are both central to the collection (*high frequency*) and more likely to be accurate (*high confidence*). We exploit global confidence metrics to reach both of these goals. The intuition is that if an entity is involved in events frequently as well as with high extraction confidence, it is more salient.

Our basic underlying hypothesis is that the salience of an entity e_i should be calculated by taking into consideration both its confidence and the confidence of other entities $\{e_j\}$ connected to it, which is inspired by PageRank [13]. Therefore for each entity e_i, we construct a set of related entities as follows:

$\{n_j \mid n_j$ is a name, n_j and e_i are coreferential or linked by a relation; and n_j is involved in an event mention$\}$

Then we compute the salience of e_i based on local confidence lc by the baseline single-document event extraction, and select the top-ranked entities as centroid entities:

$$salience(e_i) = \sum_j \sum_k lc(n_j, event_k)$$

6. Cross-document Event Refinement

Any extraction errors from the baseline system, especially on name and time arguments, will be compounded in our new cross-document IE task because they are vital to centroid detection and temporal ordering. We shall exploit knowledge derived from the background data (related documents and Wikipedia) to improve performance.

6.1 Cross-document Argument Refinement

We apply the cross-document inference techniques as described in [3] to improve name argument labeling performance. We detect clusters of similar documents and aggregate similar events across documents, and then for each cluster (a "super-document") we propagate highly consistent and frequent arguments to override other, lower confidence, extraction results, by favoring interpretation consistency across sentences and related documents.

6.2 Global Time Discovery

About 50% of the event mentions produced by single-document IE don't include explicit time arguments. However, many documents come from a topically-related news stream, so we can recover some event time arguments by gleaning knowledge from other documents.

(1) Time Search from Related Documents and Wikipedia

We analyze the entire background data and store the extracted events into an offline knowledge base:

Event type, {argument entity$_i$, role$_i$ | i =1 to n}, Event time, global confidence

Then if any event mention in the test collection is missing its time argument, we can search for this event type and arguments in the knowledge base, seeking the time argument with the highest global confidence. In the following we give two examples for discovering time from related documents and Wikipedia respectively.

In the following example, the single-document IE system is not able to identify a time argument for the *"interview"* event in the test sentence. The related documents, however, do include the time "Wednesday" (which is resolved to 2003-04-09), so we can recover the event time in the test sentence.

[*Test Sentence*]
<entity>Al-Douri</entity> said in the <entity>AP
</entity> interview he would love to return to teaching but for now he plans to remain at the United Nations.
[*Sentence from Related Documents*]
In an interview with <entity>The Associated Press
</entity> <time>Wednesday<time> night, <entity>
Al-Douri</entity> said he will continue to work at the United Nations and had no intention of defecting.

For some biographical facts for famous persons, hardly any time arguments can be found from the news articles. However, we can infer them from the knowledge base extracted from Wikipedia. For example, we can find the time argument for the *start-position* event involving *"Diller"* in the following test sentence as "1966":

Relation	$Event_i$ Arguments	$Event_j$ Arguments	Centroid	Event Type	Event Time
Coreference	Entity[Ariel Sharon] Place [Jerusalem]	Entity[Sharon] Place[Jerusalem]	Powell	Contact-Meet	2003-06-20
Subset	Entity[Bush]	Entity[Bush] Place[Camp David]	Blair	Contact-Meet	2003-03-27
Subsumption	Destination[Mideast]	Destination[Egypt]	Bush	Movement-Transport	2003-06-02
Complement	Sentence[nine-year jail] Crime[corruption]	Adjudicator[court] Place[Malaysia] Sentence[nine-year prison]	Anwar Ibrahim	Justice-Sentence	2003-04-18

Table 1. Cross-document Event Aggregation Examples

[*Test Sentence*]

<person>Diller</person> **started** *his entertainment career at* *<entity>ABC</entity>*, *where he is credited with creating the* *``movie of the week'' concept.*

[*Sentence from Wikipedia*]

<person>Diller</person> *was hired by <entity> ABC</entity> in <time>1966</time> **and was soon placed in charge of negotiating broadcast rights to feature films.***

(2) Statistical Implicit Time Prediction

Furthermore, we exploited a time argument prediction approach as described in [14]. We manually labeled 40 ACE05 newswire texts and trained a MaxEnt classifier to determine whether a time argument from an event mention EM_i can be propagated to the other event mention EM_j. The features used include the event types of EM_i and EM_j, whether they are located in the same sentence, if so the number of time expressions in the sentence; whether they share coreferential arguments, if so the roles of the arguments. This predictor is able to propagate time arguments between two events which indicate some precursor/consequence, subevent or causal relation (e.g. from a "Conflict-Attack" event to a "Life-Die/Life-Injure" event).

7. Cross-document Event Aggregation

The degree of similarity among events contained in a group of topically-related documents is much higher than the degree of similarity within an article, as each article is apt to describe the main point as well as necessary shared background. Therefore we also take into account other events that have already been generated to maximize *diversity* among the event nodes in a chain and *completeness* for each event node. In order to reach these goals, a simple event coreference solution is not enough. We also aggregate other relation types between two events: Subset, Subsumption and Complement as shown in Table 1.

Besides using cross-document name coreference to measure the similarity between a pair of arguments, we adopted some results from ACE relation extraction, e.g. using "PART-WHOLE" relations between arguments to determine whether one event subsumes the other. Earlier work on event coreference (e.g. [15]) in the MUC program was limited to several scenarios such as terrorist attacks and management succession. In our task we are targeting wider and more fine-grained event types.

8. Experimental Results

In this section we will describe our answer-key event chain annotation and then present experimental results.

8.1 Data and Answer-key Annotation

We used 106 newswire texts from ACE 2005 training corpora as our test set. Then we extracted the top 40 ranked person names as centroid entities, and manually created temporal event chains by two steps:
(1) Aggregated reference event mentions;
(2) Filled in the implicit event time arguments from the background data.

The annotations of (1) and (2) were done by two annotators independently and adjudicated for the final answer-key. In total it took one annotator about 8 hours and the other 10 hours. The inter-annotator agreement is around 90% for step (1) and 82% for step (2). We used 278,108 texts from English TDT-5 corpus and 148 million sentences from Wikipedia as the source as our background data. In these event chains there are 140 events with 368 arguments (257 are unique). The top ranked centroid entities are "Bush", "Ibrahim", "Putin", "Al-douri", "Blair", etc.

8.2 Centroid Entity Detection

First we use the arguments generated directly from single-document IE to detect 40 centroid entities, obtaining an F-measure of 55%. When we apply the cross-document name argument refinement techniques before argument ranking, the F-measure is enhanced to 67.5%, and we can cover all key centroid entities by using the top 76 system output arguments.

The ranking accuracy of the 40 correct system centroid entities is 72.95%. For comparison we computed two baselines: (1) random ranking: with accuracy about 42%; (2) ranked by the position where the first mentions of the centroid entities appear as event arguments in the test corpus, with accuracy 47.31%. We can see that our cross-document IE method achieved much higher accuracy than both baselines.

8.3 Browsing Cost

For all the system generated event chains which center around the 40 correct centroid entities, we present the browsing cost results in Figure 3.

Figure 3 indicates that for the event chains generated entirely from single-document IE, a user needs to browse 117 incorrect/redundant arguments before seeing 71

correct arguments. By adding cross-document event aggregation, the browsing effort is slightly reduced to seeing 103 incorrect/redundant arguments. The most notable result is that after applying cross-document name argument correction, the number of correct arguments is increased to 79 while the number of incorrect/redundant arguments is significantly reduced to 54. Global time discovery provided further gains: 85 correct arguments after seeing 51 incorrect/redundant ones. The final system resulted in a 60.7% user browsing effort reduction compared to the baseline before seeing 71 correct arguments; and extracted an additional 19.7% unique correct arguments.
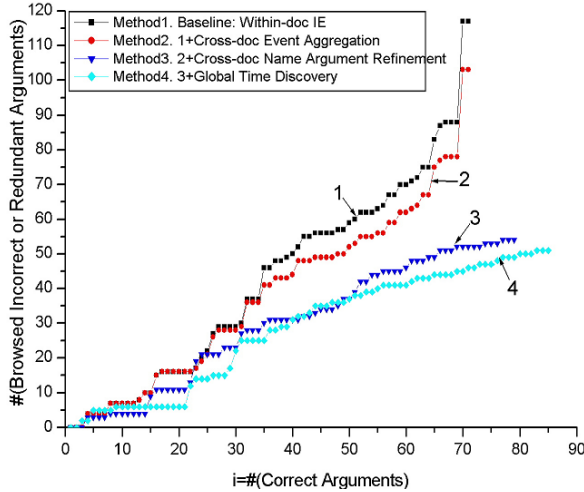

Figure 3. Argument Browsing Cost

8.4 Temporal Correlation

Table 2 shows the argument temporal ordering correlation score for each step. The 4 methods are listed in the legend for Figure 3. The difference among these methods is partly reflected by the number of scored argument pairs, as shown in the argument recall scores. As a first (crude) approximation, events can be ordered according to the time that they are reported. We treat this as our baseline.

Method	Temporal Correlation	Argument Recall
Baseline	3.71%	27.63%
Method1	44.02%	27.63%
Method2	46.15%	27.63%
Method3	55.73%	30.74%
Method4	70.09%	33.07%

Table 2. Temporal Correlation

As we can see, for news stories text order by itself is a poor predictor of chronological order (only 3.71% correlation with the true order). We can generally conclude that our cross-document IE-driven methods can produce significantly better temporal order than the baseline, and thus more coherent extraction results.

9. Related Work

To the best of our knowledge, no research group has combined ranking and linking for cross-document IE. Hence in this section, we present related work in other areas for ranking and linking separately.

Text summarization progressed from single-document to multi-document processing by centroid based sentence linking and ranking (e.g. [5], [6]). Accurate ranking techniques such as PageRank [13] have greatly enhanced information retrieval.

Recently there has been heightened interest in discovering temporal event chains, especially, the shared task evaluation TempEval [9] involved identifying temporal relations in TimeBank [17]. For example, [18] applied supervised learning to classify temporal and causal relations simultaneously for predicates in TimeBank. [19] extracted narrative event chains based on common protagonists. Our work is also similar to the task of topic detection and tracking [20] under the condition that each 'node' for linking is an event extracted by IE instead of a story.

Several recent studies have stressed the benefits of going beyond traditional single-document extraction and taking advantage of information redundancy. In particular, [3, 16, 21, 22, 23, 24, 25] have emphasized this potential in their work. As we present in section 6, the central idea of cross-document argument refinement can be applied to discover knowledge from background data, and thus significantly improve local decisions.

In this paper we import these ideas into IE while taking into account some major differences. Following the original idea of centering [2] and the approach of centering events involving protagonists [19], we present a similar idea of detecting 'centroid' arguments. We operate cross-document instead of single-document, which requires us to resolve more conflicts and ambiguities. In addition, we study the temporal event linking task on top of IE results. In this way we extend the representation of each node in the chains from an event trigger to a structured aggregated event including fine-grained information such as event types, arguments and their roles. Compared to [5, 6], we also extend the definition of "centroid" from a word to an entity; and target at linking extracted facts instead of sentences. On the other hand, we cannot simply transfer the traditional relevance or salience based ranking approaches in IR and multi-document summarization because of the incorrect facts extracted from IE. Therefore we incorporate quality into the ranking metric. Furthermore by incorporating global evidence we correct the original extracted facts and discover implicit time arguments.

10. Conclusion and Future Work

We have defined a new task of cross-document event extraction, ranking and tracking. These new modes can lay the groundwork for an improved browsing, analysis, and search process, and can potentially speed up text comprehension and knowledge distillation.

Then we presented and evaluated a system for performing this task. We investigated various challenging aspects of this new task and showed how to address them by exploiting techniques such as cross-document argument refinement, global time discovery and cross-document event aggregation. Experiments have shown that the performance of cross-document event chain extraction is significantly enhanced over the traditional single-document IE framework.

In this paper we presented event chains involving person entities, but this approach can be naturally extended to other entity types, such as tracking company start/end/acquire/merge activities. In addition we plan to automatically adjust cross-document event aggregation operations according to different compression ratios provided by the users. We are also interested in identifying more event types and their lexical realizations using paraphrase discovery techniques.

Acknowledgements

References

[1] Danielle S McNamara. 2001. Reading both High-coherence and Low-coherence Texts: Effects of Text Sequence and Prior Knowledge. *Canadian Journal of Experimental Psychology*.

[2] Barbara Grosz, Aravind Joshi, and Scott Weinstein. 1995. Centering: A Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics, 2(21).*

[3] Heng Ji and Ralph Grishman. 2008. Refining Event Extraction Through Unsupervised Cross-document Inference. *Proc. ACL-HLT 2008.*

[4] Zheng Chen and Heng Ji. 2009. Language Specific Issue and Feature Exploration in Chinese Event Extraction. *Proc. HLT-NAACL 2009.*

[5] Regina Barzilay, Noemie Elhadad and Kathleen R. Mckeown. 2002. Inferring strategies for sentence ordering in multi-document news summarization. *Journal of Artificial Intelligence Research*, v.17, pp. 35-55, 2002.

[6] Dragomir R. Radev, Hongyan Jing, Malgorzata Stys and Daniel Tam. 2004. Centroid-based Summarization of Multiple Documents. *Information Processing and Management*. 40 (2004) pp. 919-938.

[7] Maurice Kendall. 1938. A New Measure of Rank Correlation. Biometrica, 30, 81-89.

[8] W. S. Cooper. 1968. Expected search length: A Single Measure of Retrieval Effectiveness based on the Weak Ordering Action of retrieval systems. *Journal of American Society of Information Science*, 19(1), 30-41.

[9] Marc. Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz and James Pustejovsky. 2007. SemEval-2007 Task 15: TempEval Temporal Relation Identification. *Proc. ACL 2007 workshop on SemEval*, 2007.

[10] Ralph Grishman, David Westbrook and Adam Meyers. 2005. NYU's English ACE 2005 System Description. *Proc. ACE 2005 Evaluation Workshop*.

[11] K. Balog, L. Azzopardi, M. de Rijke. 2008. Personal Name Resolution of Web People Search. *Proc. WWW2008 Workshop: NLP Challenges in the Information Explosion Era (NLPIX 2008).*

[12] Alex Baron and Marjorie Freedman. 2008. Who is Who and What is What: Experiments in Cross-Document Co-Reference. *Proc. EMNLP 2008.*

[13] Lawrence Page, Sergey Brin, Rajeev Motwani and Terry Winograd. 1998. The PageRank Citation Ranking: Bringing Order to the Web. *Proc. the 7th International World Wide Web Conference.*

[14] Prashant Gupta and Heng Ji. 2009. Predicting Unknown Time Arguments based on Cross-event propagation. *Proc. ACL-IJCNLP 2009.*

[15] Amit Bagga and Breck Baldwin. 1999. Cross-document Event Coreference: Annotations, Experiments, and Observations. Proc. ACL1999 Workshop on Coreference and Its Applications.

[16] Gideon Mann. 2007. Multi-document Relationship Fusion via Constraints on Probabilistic Databases. *Proc. HLT/NAACL 2007.*

[17] J. Pustejovsky, P.Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, D. Radev, B.Sundheim, D. Day, L. Ferro and M. Lazo. 2003. The Timebank Corpus. *Corpus Linguistics*. pp. 647-656.

[18] Steven Bethard and James H. Martin. 2008. Learning semantic links from a corpus of parallel temporal and causal relations. *Proc. ACL-HLT 2008.*

[19] Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised Learning of Narrative Event Chains. *Proc. ACL-HLT 2008.*

[20] James Allan. 2002. Topic Detection and Tracking: Event-based Information Organization. Springer.

[21] Doug Downey, Oren Etzioni, and Stephen Soderland. 2005. A Probabilistic Model of Redundancy in Information Extraction. *Proc. IJCAI 2005.*

[22] Jenny Rose Finkel, Trond Grenager and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. *Proc. ACL 2005.*

[23] Roman Yangarber. 2006. Verification of Facts across Document Boundaries. Proc. International Workshop on Intelligent Information Access.

[24] Siddharth Patwardhan and Ellen Riloff. 2007. Effective Information Extraction with Semantic Affinity Patterns and Relevant Regions. *Proc. EMNLP-CONLL 2007.*

[25] Siddharth Patwardhan and Ellen Riloff. 2009. A Unified Model of Phrasal and Sentential Evidence for Information Extraction. 2009. *Proc. EMNLP 2009.*