Extractive Summarization using Inter- and Intra- Event Relevance

Wenjie Li, Mingli Wu and Qin Lu
Department of Computing
The Hong Kong Polytechnic University
{cswjli,csmlwu,csluqin}@comp
.polyu.edu.hk

Wei Xu and Chunfa Yuan
Department of Computer Science and
Technology, Tsinghua University
{vivian00,cfyuan}@mail.ts
inghua.edu.cn

Abstract

Event-based summarization attempts to select and organize the sentences in a summary with respect to the events or the sub-events that the sentences describe. Each event has its own internal structure, and meanwhile often relates to other events semantically, temporally, spatially, causally or conditionally. In this paper, we define an event as one or more event terms along with the named entities associated, and present a novel approach to derive intra- and inter- event relevance using the information of internal association, semantic relatedness, distributional similarity and named entity clustering. We then apply PageRank ranking algorithm to estimate the significance of an event for inclusion in a summary from the event relevance derived. Experiments on the DUC 2001 test data shows that the relevance of the named entities involved in events achieves better result when their relevance is derived from the event terms they associate. It also reveals that the topic-specific relevance from documents themselves outperforms the semantic relevance from a general purpose knowledge base like Word-Net.

1. Introduction

Extractive summarization selects sentences which contain the most salient concepts in documents. Two important issues with it are how the concepts are defined and what criteria should be used to judge the salience of the concepts. Existing work has typically been based on techniques that extract key textual elements, such as keywords (also known as significant terms) as weighed by their tf*idf score, or concepts (such as events or entities) with linguistic and/or statistical analysis. Then, sentences are selected according to either the important textual units they contain or certain types of inter-sentence relations they hold.

Event-based summarization which has emerged recently attempts to select and organize sentences in a summary with respect to events or sub-events that the sentences describe. With regard to the concept of events, people do not have the same definition when introducing it in different domains. While traditional linguistics work on semantic theory of events and the semantic structures of verbs, studies in information retrieval (IR) within topic detection and tracking framework look at events as narrowly defined topics which can be categorized or clustered as a set of related documents (TDT). IR events are broader (or to say complex) events in the sense that they may include happenings and their causes, consequences or even more extended effects. In the information extraction (IE) community, events are defined as the pre-specified and structured templates that relate an action to its participants, times, locations and other entities involved (MUC-7). IE defines what people call atomic events.

Regardless of their distinct perspectives, people all agree that events are collections of activities together with associated entities. To apply the concept of events in the context of text summarization, we believe it is more appropriate to consider events at the sentence level, rather than at the document level. To avoid the complexity of deep semantic and syntactic processing, we complement the advantages of statistical techniques from the IR community and structured information provided by the IE community.

We propose to extract semi-structured events with shallow natural language processing (NLP) techniques and estimate their importance for inclusion in a summary with IR techniques.

Though it is most likely that documents narrate more than one similar or related event, most event-based summarization techniques reported so far explore the importance of the events independently. Motivated by this observation, this paper addresses the task of event-relevance based summarization and explores what sorts of relevance make a contribution. To this end, we investigate intra-event relevance, that is action-entity relevance, and inter-event relevance, that is event-event relevance. While intra-event relevance is measured with frequencies of the associated events and entities directly, inter-event relevance is derived indirectly from a general WordNet similarity utility, distributional similarity in the documents to be summarized, named entity clustering and so on. Pagerank ranking algorithm is then applied to estimate the event importance for inclusion in a summary using the aforesaid relevance.

The remainder of this paper is organized as follows. Section 2 introduces related work. Sections 3 introduces our proposed event-based summarization approaches which make use of intra- and inter- event relevance. Section 4 presents experiments and evaluates different approaches. Finally, Section 5 concludes the paper.

2. Related Work

Event-based summarization has been investigated in recent research. It was first presented in (Daniel, Radev and Allison, 2003), who treated a news topic in multi-document summarization as a series of sub-events according to human understanding of the topic. They determined the degree of sentence relevance to each sub-event through human judgment and evaluated six extractive approaches. Their paper concluded that recognizing the sub-events that comprise a single news event is essential for producing better summaries. However, it is difficult to automatically break a news topic into sub-events.

Later, atomic events were defined as the relationships between the important named entities (Filatova and Hatzivassiloglou, 2004), such as participants, locations and times (which are called relations) through the verbs or action nouns labeling the events themselves (which are called connectors). They evaluated sentences

based on co-occurrence statistics of the named entity relations and the event connectors involved. The proposed approach claimed to outperform conventional tf*idf approach. Apparently, named entities are key elements in their model. However, the constraints defining events seemed quite stringent.

The application of dependency parsing, anaphora and co-reference resolution in recognizing events were presented involving NLP and IE techniques more or less (Yoshioka and Haraguchi, 2004), (Vanderwende, Banko and Menezes, 2004) and (Leskovec, Grobelnik and Fraling, 2004). Rather than pre-specifying events, these efforts extracted (verb)-(dependent relation)-(noun) triples as events and took the triples to form a graph merged by relations.

As a matter of fact, events in documents are related in some ways. Judging whether the sentences are salient or not and organizing them in a coherent summary can take advantage from event relevance. Unfortunately, this was neglected in most previous work. Barzilay and Lapata (2005) exploited the use of the distributional and referential information of discourse entities to improve summary coherence. While they captured text relatedness with entity transition sequences, i.e. entity-based summarization, we are particularly interested in relevance between events in event-based summarization.

Extractive summarization requires ranking sentences with respect to their importance. Successfully used in Web-link analysis and more recently in text summarization, Google's PageRank (Brin and Page, 1998) is one of the most popular ranking algorithms. It is a kind of graph-based ranking algorithm deciding on the importance of a node within a graph by taking into account the global information recursively computed from the entire graph, rather than relying on only the local node-specific information. A graph can be constructed by adding a node for each sentence, phrase or word. Edges between nodes are established using inter-sentence similarity relations as a function of content overlap or grammatically relations between words or phrases.

The application of PageRank in sentence extraction was first reported in (Erkan and Radev, 2004). The similarity between two sentence nodes according to their term vectors was used to generate links and define link strength. The same idea was followed and investigated exten-

sively (Mihalcea, 2005). Yoshioka and Haraguchi (2004) went one step further toward event-based summarization. Two sentences were linked if they shared similar events. When tested on TSC-3, the approach favoured longer summaries. In contrast, the importance of the verbs and nouns constructing events was evaluated with PageRank as individual nodes aligned by their dependence relations (Vanderwende, 2004; Leskovec, 2004).

Although we agree that the fabric of event constitutions constructed by their syntactic relations can help dig out the important events, we have two comments. First, not all verbs denote event happenings. Second, semantic similarity or relatedness between action words should be taken into account.

3. Event-based Summarization

3.1. Event Definition and Event Map

Events can be broadly defined as "Who did What to Whom When and Where". Both linguistic and empirical studies acknowledge that event arguments help characterize the effects of a verb's event structure even though verbs or other words denoting event determine the semantics of an event. In this paper, we choose verbs (such as "elect") and action nouns (such as "supervision") as event terms that can characterize or partially characterize actions or incident occurrences. They roughly relate to "did What". One or more associated named entities are considered as what are denoted by linguists as event arguments. Four types of named entities are currently under the consideration. These are <Person>, <Organization>, <Location> and <Date>. They convey the information of "Who", "Whom", "When" and "Where". A verb or an action noun is deemed as an event term only when it presents itself at least once between two named entities.

Events are commonly related with one another semantically, temporally, spatially, causally or conditionally, especially when the documents to be summarized are about the same or very similar topics. Therefore, all event terms and named entities involved can be explicitly connected or implicitly related and weave a document or a set of documents into an event fabric, i.e. an event graphical representation (see Figure 1). The nodes in the graph are of two types. Event terms (*ET*) are indicated by rectangles and named entities (*NE*) are indicated by ellipses. They represent concepts rather than instances. Words in either their original form or morphological variations are represented with a single node in the graph regardless of how many times they appear in documents. We call this representation an event map, from which the most important concepts can be pick out in the summary.

<Organization> America Online </Organization> was to buy <Organization> Netscape </Organization> and forge a partnership with <Organization> Sun </Organization>, benefiting all three and giving technological independence from <Organization> Microsoft </Organization>.
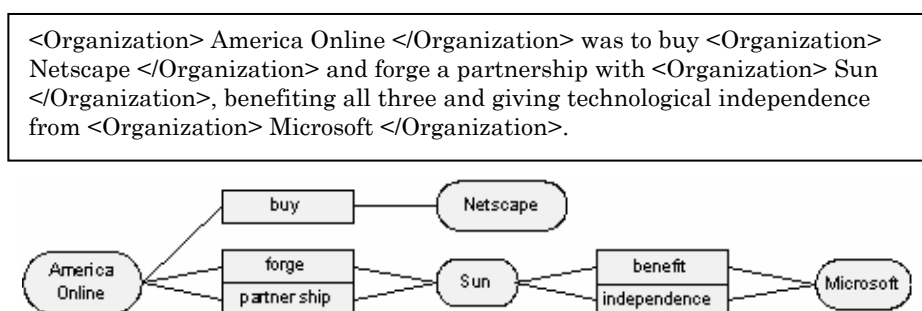

Figure 1 Sample sentences and their graphical representation

The advantage of representing with separated action and entity nodes over simply combining them into one event or sentence node is to provide a convenient way for analyzing the relevance among event terms and named entities either by their semantic or distributional similarity. More importantly, this favors extraction of concepts and brings the conceptual compression available.

We then integrate the strength of the connections between nodes into this graphical model in terms of the relevance defined from different perspectives. The relevance is indicated by $r(node_i, node_j)$, where $node_i$ and $node_j$ represent two nodes, and are either event terms (et_i) or named entities (ne_j). Then, the significance of each node, indicated by $w(node_i)$, is calcu-

lated with PageRank ranking algorithm. Sections 3.2 and 3.3 address the issues of deriving $r(node_i, node_j)$ according to intra- or/and inter-event relevance and calculating $w(node_i)$ in detail.

3.2 Intra- and Inter- Event Relevance

We consider both intra-event and inter-event relevance for summarization. Intra-event relevance measures how an action itself is associated with its associated arguments. It is indicated as $R(ET, NE)$ and $R(NE, ET)$ in Table 1 below. This is a kind of direct relevance as the connections between actions and arguments are established from the text surface directly. No inference or background knowledge is required. We consider that when the connection between an event term et_i and a named entity ne_j is symmetry, then $R(NE, ET) = R(ET, NE)^T$. Events are related as explained in Section 2. By means of inter-event relevance, we consider how an event term (or a named entity involved in an event) associate to another event term (or another named entity involved in the same or different events) syntactically, semantically and distributionally. It is indicated by $R(ET, ET)$ or $R(NE, NE)$ in Table 1 and measures an indirect connection which is not explicit in the event map needing to be derived from the external resource or overall event distribution.

	Event Term (*ET*)	Named Entity (*NE*)
Event Term (*ET*)	$R(ET, ET)$	$R(ET, NE)$
Named Entity (*NE*)	$R(NE, ET)$	$R(NE, NE)$

Table 1 Relevance Matrix

The complete relevance matrix is:

$$R = \begin{bmatrix} R(ET, ET) & R(ET, NE) \\ R(NE, ET) & R(NE, NE) \end{bmatrix}$$

The intra-event relevance $R(ET, NE)$ can be simply established by counting how many times et_i and ne_j are associated, i.e.

$$r_{Document}(et_i, ne_j) = freq(et_i, ne_j) \qquad \text{(E1)}$$

One way to measure the term relevance is to make use of a general language knowledge base, such as WordNet (Fellbaum 1998). WordNet::Similarity is a freely available software package that makes it possible to measure the semantic relatedness between a pair of concepts,

or in our case event terms, based on WordNet (Pedersen, Patwardhan and Michelizzi, 2004). It supports three measures. The one we choose is the function *lesk*.

$$r_{WordNet}(et_i, et_j) = similarity(et_i, et_j) = lesk(et_i, et_j)$$

$$\text{(E2)}$$

Alternatively, term relevance can be measured according to their distributions in the specified documents. We believe that if two events are concerned with the same participants, occur at same location, or at the same time, these two events are interrelated with each other in some ways. This observation motivates us to try deriving event term relevance from the number of name entities they share.

$$r_{Document}(et_i, et_j) = |NE(et_i) \cap NE(et_j)| \quad \text{(E3)}$$

Where $NE(et_i)$ is the set of named entities et_i associate. $|\ |$ indicates the number of the elements in the set. The relevance of named entities can be derived in a similar way.

$$r_{Document}(ne_i, ne_j) = |ET(ne_i) \cap ET(ne_j)| \quad \text{(E4)}$$

The relevance derived with (E3) and (E4) are indirect relevance. In previous work, a clustering algorithm, shown in Figure 2, has been proposed (Xu et al, 2006) to merge the named entity that refer to the same person (such as Ranariddh, Prince Norodom Ranariddh and President Prince Norodom Ranariddh). It is used for co-reference resolution and aims at joining the same concept into a single node in the event map. The experimental result suggests that merging named entity improves performance in some extend but not evidently. When applying the same algorithm for clustering all four types of name entities in DUC data, we observe that the name entities in the same cluster do not always refer to the same objects, even when they are indeed related in some way. For example, "Mississippi" is a state in the southeast United States, while "Mississippi River" is the second-longest rever in the United States and flows through "Mississippi".

Step1: Each name entity is represented by
$ne_i = w_{i1}w_{i2}...w_{ik}$, where w_i is the ith word in it. The cluster it belongs to, indicated by $C(ne_i)$, is initialled by $w_{i1}w_{i2}...w_{ik}$ itself.
Step2: For each name entity
$ne_i = w_{i1}w_{i2}...w_{ik}$
For each name entity

$ne_j = w_{j1}w_{j2}...w_{jl}$, if $C(ne_i)$ is a sub-string of $C(ne_j)$, then

$C(ne_i) = C(ne_j)$.

Continue Step 2 until no change occurs.

Figure 2 The algorithm proposed to merge the named entities

Location	Person	Date	Organization
Mississippi	Professor Sir Richard Southwood	first six months of last year	Long Beach City Council
Mississippi River	Sir Richard Southwood	last year	San Jose City Council
	Richard Southwood		City Council

Table 2 Some results of the named entity merged

It therefore provides a second way to measure named entity relevance based on the clusters found. It is actually a kind of measure of lexical similarity.

$$r_{Cluster}(ne_i, ne_j) = \begin{cases} 1, & ne_i, ne_j \text{ are in the same cluster} \\ 0, & \text{otherwise} \end{cases}$$

(E5)

In addition, the relevance of the named entities can be sometimes revealed by sentence context. Take the following most frequently used sentence patterns as examples:

<Person>, a-position-name of <Organization>, does something.
<Person> and another <Person> do something.

Figure 3 The example patterns

Considering that two neighbouring name entities in a sentence are usually relevant, the following window-based relevance is also experimented with.

$r_{Pattern}(ne_i, ne_j)$

$$= \begin{cases} 1, & ne_i, ne_j \text{ are within a pre-specified window size} \\ 0, & \text{otherwise} \end{cases}$$

(E6)

3.3 Significance of Concepts

The significance score, i.e. the weight $w(node_i)$ of each $node_i$, is then estimated recursively with PageRank ranking algorithm which assigns the significance score to each node according to the number of nodes connecting to it as well as the strength of their connections. The equation calculating $w(node_i)$ using PageRank of a certain $node_i$ is shown as follows.

$$w(node_i) = (1-d) + d(\frac{w(node_1)}{r(node_i, node_1)} + ... + \frac{w(node_j)}{r(node_i, node_j)} + ... + \frac{w(node_t)}{r(node_i, node_t)})$$

(E7)

In (E7), $node_j$ ($j = 1,2,...t$, $j \neq i$) are the nodes linking to $node_i$. d is the factor used to avoid the limitation of loop in the map structure. It is set to 0.85 experimentally. The significance of each sentence to be included in the summary is then obtained from the significance of the events it contains. The sentences with higher significance are picked up into the summary as long as they are not exactly the same sentences. We are aware of the important roles of information fusion and sentence compression in summary generation. However, the focus of this paper is to evaluate event-based approaches in extracting the most important sentences. Conceptual extraction based on event relevance is our future direction.

4. Experiments and Discussions

To evaluate the event based summarization approaches proposed, we conduct a set of experiments on 30 English document sets provide by the DUC 2001 multi-document summarization task. The documents are pre-processed with GATE to recognize the previously mentioned four types of name entities. On average, each set contains 10.3 documents, 602 sentences, 216 event terms and 148.5 name entities.

To evaluate the quality of the generated summaries, we choose an automatic summary evaluation metric ROUGE, which has been used in DUCs. ROUGE is a recall-based metric for fixed length summaries. It bases on N-gram co-occurrence and compares the system generated summaries to human judges (Lin and Hovy, 2003). For each DUC document set, the system creates a summary of 200 word length and present three of the ROUGE metrics: ROUGE-1 (unigram-based), ROUGE-2 (bigram-based), and ROUGE-W (based on longest common subsequence weighed by the length) in the following experiments and evaluations.

We first evaluate the summaries generated based on $R(ET, NE)$ itself. In the pre-evaluation experiments, we have observed that some fre-

quently occurring nouns, such as "doctors" and "hospitals", by themselves are not marked by general NE taggers. But they indicate persons, organizations or locations. We compare the ROUGE scores of adding frequent nouns or not to the set of named entities in Table 3. A noun is considered as a frequent noun when its frequency is larger than 10. Roughly 5% improvement is achieved when high frequent nouns are taken into the consideration. Hereafter, when we mention NE in latter experiments, the high frequent nouns are included.

$R(ET, NE)$	NE Without High Frequency Nouns	NE With High Frequency Nouns
ROUGE-1	0.33320	0.34859
ROUGE-2	0.06260	0.07157
ROUGE-W	0.12965	0.13471

Table 3 ROUGE scores using $R(ET, NE)$ itself

Table 4 below then presents the summarization results by using $R(ET, ET)$ itself. It compares two relevance derivation approaches, $R_{WordNet}$ and $R_{Document}$. The topic-specific relevance derived from the documents to be summarized outperforms the general purpose Word-Net relevance by about 4%. This result is reasonable as WordNet may introduce the word relatedness which is not necessary in the topic-specific documents. When we examine the relevance matrix from the event term pairs with the highest relevant, we find that the pairs, like "abort" and "confirm", "vote" and confirm", do reflect semantics (antonymous) and associated (causal) relations to some degree.

$R(ET, ET)$	Semantic Relevance from Word-Net	Topic-Specific Relevance from Documents
ROUGE-1	0.32917	0.34178
ROUGE-2	0.05737	0.06852
ROUGE-W	0.11959	0.13262

Table 4 ROUGE scores using $R(ET, ET)$ itself

Surprisingly, the best individual result is from document distributional similarity $R_{Document}$ (NE, NE) in Table 5. Looking more closely, we conclude that compared to event terms, named entities are more representative of the documents in which they are included. In other words, event terms are more likely to be distributed around all the document sets, whereas named entities are more topic-specific and therefore cluster in a particular document set more. Examples of high related named entities in relevance matrix are "Andrew" and "Florida",

"Louisiana" and "Florida". Although their relevance is not as explicit as the same of event terms (their relevance is more contextual than semantic), we can still deduce that some events may happen in both Louisiana and Florida, or about Andrew in Florida. In addition, it also shows that the relevance we would have expected to be derived from patterns and clustering can also be discovered by $R_{Document}(NE, NE)$. The window size is set to 5 experimentally in window-based practice.

$R(NE, NE)$	Relevance from Documents	Relevance from Clustering	Relevance from Window-based Context
ROUGE-1	0.35212	0.33561	0.34466
ROUGE-2	0.07107	0.07286	0.07508
ROUGE-W	0.13603	0.13109	0.13523

Table 5 ROUGE scores using $R(NE, NE)$ itself

Next, we evaluate the integration of $R(ET, NE)$, $R(ET, ET)$ and $R(NE, NE)$. As DUC 2001 provides 4 different summary sizes for evaluation, it satisfies our desire to test the sensibility of the proposed event-based summarization techniques to the length of summaries. While the previously presented results are evaluated on 200 word summaries, now we move to check the results in four different sizes, i.e. 50, 100, 200 and 400 words. The experiments results show that the event-based approaches indeed prefer longer summaries. This is coincident with what we have hypothesized. For this set of experiments, we choose to integrate the best method from each individual evaluation presented previously. It appears that using the named entities relevance which is derived from the event terms gives the best ROUGE scores in almost all the summery sizes. Compared with the results provided in (Filatova and Hatzivassiloglou, 2004) whose average ROUGE-1 score is below 0.3 on the same data set, the significant improvement is revealed. Of course, we need to test on more data in the future.

$R(NE, NE)$	50	100	200	400
ROUGE-1	0.22383	0.28584	0.35212	0.41612
ROUGE-2	0.03376	0.05489	0.07107	0.10275
ROUGE-W	0.10203	0.11610	0.13603	0.13877
$R(ET, NE)$	50	100	200	400
ROUGE-1	0.22224	0.27947	0.34859	0.41644
ROUGE-2	0.03310	0.05073	0.07157	0.10369
ROUGE-W	0.10229	0.11497	0.13471	0.13850
$R(ET, ET)$	50	100	200	400

ROUGE-1	0.20616	0.26923	0.34178	0.41201
ROUGE-2	0.02347	0.04575	0.06852	0.10263
ROUGE-W	0.09212	0.11081	0.13262	0.13742
$R(ET,NE) +$ $R(ET,ET) +$ $R(NE,NE)$	50	100	200	400
ROUGE-1	0.21311	0.27939	0.34630	0.41639
ROUGE-2	0.03068	0.05127	0.07057	0.10579
ROUGE-W	0.09532	0.11371	0.13416	0.13913

Table 6 ROUGE scores using complete R matrix and with different summary lengths

As discussed in Section 3.2, the named entities in the same cluster may often be relevant but not always be co-referred. In the following last set of experiments, we evaluate the two ways to use the clustering results. One is to consider them as related as if they are in the same cluster and derive the *NE-NE* relevance with (E5). The other is to merge the entities in one cluster as one reprehensive named entity and then use it in *ET-NE* with (E1). The rationality of the former approach is validated.

	Clustering is used to derive *NE-NE*	Clustering is used to merge entities and then to derive *ET-NE*
ROUGE-1	0.34072	0.33006
ROUGE-2	0.06727	0.06154
ROUGE-W	0.13229	0.12845

Table 7 ROUGE scores with regard to how to use the clustering information

5. Conclusion

In this paper, we propose to integrate event-based approaches to extractive summarization. Both inter-event and intra-event relevance are investigated and PageRank algorithm is used to evaluate the significance of each concept (including both event terms and named entities). The sentences containing more concepts and highest significance scores are chosen in the summary as long as they are not the same sentences.

To derive event relevance, we consider the associations at the syntactic, semantic and contextual levels. An important finding on the DUC 2001 data set is that making use of named entity relevance derived from the event terms they associate with achieves the best result. The result of 0.35212 significantly outperforms the one reported in the closely related work whose average is below 0.3. We are interested in the issue of how to improve an event representation in order to build a more powerful event-based summarization system. This would be one of our future directions. We also want to see how concepts rather than sentences are selected into the summary in order to develop a more flexible compression technique and to know what characteristics of a document set is appropriate for applying event-based summarization techniques.

References

Chin-Yew Lin and Eduard Hovy. 2003. Automatic Evaluation of Summaries using N-gram Co-occurrence Statistics. In Proceedings of HLT-NAACL 2003, pp71-78.

Christiane Fellbaum. 1998, WordNet: An Electronic Lexical Database. MIT Press.

Elena Filatova and Vasileios Hatzivassiloglou. 2004. Event-based Extractive summarization. In Proceedings of ACL 2004 Workshop on Summarization, pp104-111.

Gunes Erkan and Dragomir Radev. 2004. LexRank: Graph-based Centrality as Salience in Text Summarization. Journal of Artificial Intelligence Research.

Jure Leskovec, Marko Grobelnik and Natasa Milic-Frayling. 2004. Learning Sub-structures of Document Semantic Graphs for Document Summarization. In LinkKDD 2004.

Lucy Vanderwende, Michele Banko and Arul Menezes. 2004. Event-Centric Summary Generation. In Working Notes of DUC 2004.

Masaharu Yoshioka and Makoto Haraguchi. 2004. Multiple News Articles Summarization based on Event Reference Information. In Working Notes of NTCIR-4, Tokyo.

MUC-7.   http://www-nlpir.nist.gov/related_projects/ muc/proceeings/ muc_7_toc.html

Naomi Daniel, Dragomir Radev and Timothy Allison. 2003. Sub-event based Multi-document Summarization. In Proceedings of the HLT-NAACL 2003 Workshop on Text Summarization, pp9-16.

Page Lawrence, Brin Sergey, Motwani Rajeev and Winograd Terry. 1998. The PageRank Citation Ranking: Bring Order to the Web. Technical Report, Stanford University.

Rada Mihalcea. 2005. Language Independent Extractive Summarization. ACL 2005 poster.

Regina Barzilay and Michael Elhadad. 2005. Modelling Local Coherence: An Entity-based Approach. In Proceedings of ACL, pp141-148.

TDT. http://projects.ldc.upenn.edu/TDT.

Ted Pedersen, Siddharth Patwardhan and Jason Michelizzi. 2004. WordNet::Similarity – Measuring the Relatedness of Concepts. In Proceedings of AAAI, pp25-29.

Wei Xu, Wenjie Li, Mingli Wu, Wei Li and Chunfa Yuan. 2006. Deriving Event Relevance from the Ontology Constructed with Formal Concept Analysis, in Proceedings of CiCling'06, pp480-489.