

# A Subjective Logic Framework for Multi-Document Summarization

*Sukanya Manna*<sup>1</sup> *Byron J. Gao*<sup>1</sup> *Reed Coke*<sup>2</sup>

(1) Department of Computer Science, Texas State University, San Marcos, TX 78666

(2) Department of Computer Science, Swarthmore College, Swarthmore, PA 19081  
s\_m201@txstate.edu, bgao@txstate.edu, rcoke1@swarthmore.edu

## ABSTRACT

In this paper we propose SubSum, a subjective logic framework for sentence-based extractive multi-document summarization. Document summaries perceived by humans are subjective in nature as human judgements of sentence relevancy are inconsistent and laden with uncertainty. SubSum captures this uncertainty and extracts significant sentences from a document cluster to generate extractive summaries. In particular, SubSum represents the sentences of a document cluster as propositions and computes *opinions*, a probability measure containing secondary uncertainty, for these propositions. Sentences with stronger opinions are considered more significant and used as candidate sentences. The key advantage of SubSum over other techniques is its ability to quantify uncertainty. In addition, SubSum is a completely unsupervised approach and is highly portable across different domains and languages.

---

KEYWORDS: multi-document summarization, subjective logic, belief measures, uncertainty.

---

## 1 Introduction

Automatic multi-document summarization can effectively condense information found in multiple documents into a short, readable synopsis, allowing users to quickly familiarize themselves with the main ideas of the information. It has vast applications, especially for online documents where redundancy abounds (Harabagiu and Lacatusu, 2010). Example source documents include news articles, email threads, blogs, reviews, and search results, just to name a few. There has been an increasing research effort towards multi-document summarization in recent years (McKeown and Radev, 1995; Radev and McKeown, 1998; Radev et al.; Carbonell and Goldstein, 1998; Barzilay et al., 1999; Conroy et al., 2004; Barzilay and McKeown, 2005; Daumé III and Marcu, 2005; Nenkova and Vanderwende, 2005; Nenkova et al., 2006; Ji, 2006; Park et al., 2007; Wei et al., 2008; Wan, 2008; Wang et al., 2008; Li et al., 2010; Zhang et al., 2010; Shen and Li, 2010; Xia et al., 2011; Ribaldo et al., 2012).

Document summaries perceived by humans are subjective in nature. Judgments of sentence importance can be affected by user interests, preferences and viewpoints, which may vary from person to person. Any given portion of a document can be interpreted in different fashions by different people, especially in the way they understand and interpret the context (Pardo et al., 2002). For example, for the news of "a small plane smashed into the tallest building in Milan Thursday evening ...", a user could perceive its main idea as a tragic accident, whereas another user could interpret it as a topic about a terrorist attack. Therefore, people arrive at their judgements based on *subjective* information which is not completely certain or reliable.

**SubSum Framework:** In order to capture the uncertainty of human judgements in summarizing documents, in this paper we propose SubSum, a subjective logic framework for multi-document summarization. SubSum represents the sentences of a document cluster as propositions and computes *opinions*, a probability measure containing secondary uncertainty, for these propositions. Sentences with stronger opinions are considered more significant and used as candidate sentences to summarize a document.

While standard logic deals with propositions that are either true or false, subjective logic (Josang, 2001) is a type of probabilistic logic that explicitly takes uncertainty and belief into account. It can be seen as an extension of probability calculus and binary logic. Subjective logic is particularly suitable for modeling and analyzing situations involving uncertainty and incomplete knowledge (Josang and McAnally, 2010; Josang, 2001).

SubSum is an evidence-based method. It formulates opinions using evidence derived or found in a document, which can be terms, phrases, sentences, co-occurrences of words or phrases, or any syntactic or semantic features. Opinions can clearly classify sentences based on their importance and can be used to select significant sentences in summarizing a document.

SubSum differs significantly from many existing summarization approaches. Existing approaches typically use corpus statistics (Luhn, 1958), linguistic features (Hovy and Lin, 1998), linear algebra methods (Gong and Liu, 2001), or graphical methods (Mihalcea and Tarau, 2005) to find out relevant sentences within documents. These methods do not focus on capturing or quantifying uncertainty as SubSum does.

SubSum is completely based on belief measures adapted from Dempster-Shafer theory (Shafer, 1976) and independent of lexical databases, making it easily portable across different domains and languages. Additionally, it does not require any training data as in many existing summarizers (Conroy et al., 2004; Kupiec et al., 1995).

**Contributions:** (1) We propose SubSum, the first subjective logic-based framework for multi-document summarization, leveraging belief measures to capture uncertainty of human judgements on significance of sentences. (2) SubSum does not require training data, and is highly portable across different domains and languages. (3) We perform extensive experiments on benchmark datasets, demonstrating the advantages and effectiveness of the SubSum framework.

## 2 Subjective Logic Preliminaries

Subjective logic (Josang, 2001) operates on subjective beliefs about the world, and uses *opinions* to denote the representations of subjective beliefs. In subjective logic, first order measure of evidence is expressed as belief mass distribution functions over the *frame of discernment*. An *opinion* can be interpreted as a probability measure containing secondary uncertainty, and as such subjective logic can be seen as an extension of both probability calculus and binary logic (Josang, 2001). *Belief*, *disbelief*, *uncertainty*, and *base rates* are the four main belief representations used to express opinions of propositions in subjective logic. The following definitions of subjective logic concepts are adapted from Josang's draft book (folk.uio.no/josang/papers/subjective\_logic.pdf) and (Josang, 2001).

**Belief Mass Assignment:** Let  $\Theta$  be a frame of discernment. For each sub-state  $x \in 2^\Theta$ , if a number  $m_\Theta(x)$  is associated such that,  $m_\Theta(x) \geq 0$ ,  $m_\Theta(\emptyset) = 0$ ,  $\sum_{x \in 2^\Theta} m_\Theta(x) = 1$ , then  $m_\Theta$  is called a belief mass assignment in  $\Theta$ , or *BMA* for short. For each sub-state  $x \in 2^\Theta$ , the number  $m_\Theta(x)$  is called the belief mass of  $x$ .

**Belief Function:** Let  $\Theta$  be a frame of discernment, and let  $m_\Theta$  be a *BMA* on  $\Theta$ . Then the belief function corresponding to  $m_\Theta$  is the function  $b : 2^\Theta \rightarrow [0, 1]$  defined by:

$$b(x) = \sum_{y \subseteq x} m_\Theta(y), \quad x, y \in 2^\Theta \quad (1)$$

**Disbelief Function:** Let  $\Theta$  be a frame of discernment, and let  $m_\Theta$  be a *BMA* on  $\Theta$ . Then the disbelief function corresponding to  $m_\Theta$  is the function  $d : 2^\Theta \rightarrow [0, 1]$  defined by:

$$d(x) = \sum_{y \cap x = \emptyset} m_\Theta(y), \quad x, y \in 2^\Theta. \quad (2)$$

**Uncertainty Function:** Let  $\Theta$  be a frame of discernment, and let  $m_\Theta$  be a *BMA* on  $\Theta$ . Then the uncertainty function corresponding to  $m_\Theta$  is the function  $u : 2^\Theta \rightarrow [0, 1]$  defined by:

$$u(x) = \sum_{y \cap x \neq \emptyset, y \not\subseteq x} m_\Theta(y), \quad x, y \in 2^\Theta. \quad (3)$$

**Base Rate Function:** Let  $\Theta$  be a frame of cardinality  $k$ , and let  $a_\Theta$  be the function from  $\Theta$  to  $[0, 1]^k$  satisfying,  $a_\Theta(\emptyset) = 0$ ,  $a_\Theta(x_i) \in [0, 1]$  and  $\sum_{i=1}^k a_\Theta(x_i) = 1$  Then  $a_\Theta$  is a base rate distribution over  $\Theta$ .

**Relative Atomicity or Relative Base Rate:** Let  $\Theta$  be a frame of discernment and let  $x, y \in 2^\Theta$ . Then for any given  $y \neq \emptyset$  the relative atomicity of  $x$  to  $y$  is the function  $a : 2^\Theta \rightarrow [0, 1]$ ,

$$a(x/y) = \frac{|x \cap y|}{|y|}, \quad x, y \in 2^\Theta, y \neq \emptyset. \quad (4)$$

It can be observed that  $x \cap y = \emptyset \Rightarrow a(x/y) = 0$  and  $y \subseteq x \Rightarrow a(x/y) = 1$ . In all other cases relative atomicity will be a value between 0 and 1.

**Binomial Opinion:** Let  $\Theta = \{x, \neg x\}$  be either a binary frame or a binary partitioning of an  $n$ -ary frame. A binomial opinion about the truth of state  $x$  is an ordered quadruple  $\omega_x = (b, d, u, a)$ ,

where  $b$  is belief,  $d$  is disbelief,  $u$  is uncertainty, and  $a$  is base rate respectively. These components satisfy  $b + d + u = 1$  and  $b, d, u, a \in [0, 1]$ .

**Probability Expectation of a binomial opinion of proposition  $x$ :** Let  $\Theta$  be a frame of discernment with BMA  $m_\Theta$  then the probability expectation function corresponding to  $m_\Theta$  is the function  $E : 2^\Theta \rightarrow [0, 1]$  defined by:

$$E(x) = b(x) + a(x)u(x) \tag{5}$$

**Ordering of Opinions:** Let  $\omega_x$  and  $\omega_y$  be two opinions. They can be ordered according to the following criteria by priority: (1) The opinion with the greatest probability expectation is the greatest opinion. (2) The opinion with the least uncertainty is the greatest opinion. (3) The opinion with the least relative atomicity is the greatest opinion.

### 3 SubSum Framework For Multi-Document Summarization

In this section, we first formally define the multi-document summarization problem, then we present the SubSum framework. In SubSum, sentences are considered as propositions. The *opinions* of these propositions are computed, which signify their importance in the document cluster and are used to select candidate sentences for summarization purposes.

**Multi-Document Summarization Problem:** Given a cluster of documents  $D$  consisting of a set of sentences  $S = \{s_1, s_2, \dots, s_m\}$  and a set of words  $W = \{w_1, w_2, \dots, w_n\}$ , the task of multi-document summarization is to extract a subset of sentences from  $S$ , denoted by  $S_r$  ( $S_r \subset S$ ), that best represent the content of document cluster  $D$ .

#### 3.1 Formulation

**Concept of States:** A cluster of related documents consisting of  $n$  unique words represent an  $n$ -ary frame of discernment  $\Theta$ . Words are elementary states or atomic states. Sentences and phrases on the other hand are composite states, which can be represented as the union of multiple atomic states.

**Assumptions:** The following framework is proposed for the practical application of subjective logic in a document analysis/summarization context: (1) Documents in a cluster are related. (2) All the words or terms (stop words excluded) in a document cluster are atomic. (3) The sentences are unique, i.e., each sentence occurs only once in given document cluster. Single-word sentences can also exist.

**Document Representation:** Sentences of a document cluster are considered as a set of words separated by a stop mark ".", "!" or "?". These sentences are tokenized to generate words (stop words excluded). These words and sentences represent atomic and composite states.

We discuss here different notations used in this paper.  $\Theta$  is the frame of discernment, a document cluster  $D$  can be represented as a set of words, which is denoted by  $W = \{w_1, w_2, \dots, w_n\}$ , where  $W$  is a set of words present in the document cluster  $D$ .  $|W| = n$ . Since there are  $n$  words in  $D$ ,  $\Theta$  is an  $n$ -ary frame of discernment. Thus,  $\rho(\Theta) = \{\{w_1\}, \{w_2\}, \dots, \{w_1, w_2, w_3, \dots, w_n\}\} \equiv 2^\Theta$ , where  $|\rho(\Theta)| = 2^n$ .

A document cluster  $D$  can also be represented by a set of sentences  $S$  such that  $S = \{s_1, s_2, \dots, s_m\}$ , where  $|S| = m$  and  $s_i (i \in |S|)$  is an element of  $\rho(\Theta)$ , because each sentence can be represented as  $S_i = \{w_j \cup w_k \cup \dots \cup w_r\} \in \Theta$  where  $S_i \in \rho(\Theta)$ . Note that for practical reasons,  $|\rho(\Theta)| = 2^n - 2$ , excluding  $\Theta$  and  $\emptyset$ , which is also known as *reduced frame of discernment*. If there are  $n$  words in

the document cluster, then there can be at maximum  $2^n - 2$  possible states (words and their co-occurrences). For example, suppose a document cluster  $D$  has words  $a, b, c, d$ , then  $\{a\}, \{b\}, \{c\}, \{d\}, \{a, b, c, d\}$ , are the states we use for our analysis (to avoid computational expenses, we do not consider all the states of  $|\rho(\Theta)|$ , such as,  $\{ab\}, \{bc\}... \{abc\}...$ ).

**Frequency of States:** Before computing Belief Mass Assignment (BMA), frequency of each states should be computed.

$$F_x = \sum_{i=1}^m f_x, \quad x \in 2^\Theta, f_x > 0, m = |S|, \quad (6)$$

where  $f_x$  is the frequency of the state  $x$  in each sentence and  $m$  is the total number of sentences in the document cluster  $D$ .

$$Z = \sum_{j=1}^M F_x, \quad M \in |\rho(\Theta)|, \quad (7)$$

where  $Z$  is the sum of the frequencies of all states in the document cluster  $D$ .

**Belief Mass Assignment (BMA):** Belief mass assignment (BMA) is computed in this case by

$$m(x) = F_x / Z, \quad (8)$$

where  $F_x$ , computed by Eq.(6), is the total frequency of that sub-state in all the sentences (or the whole document), and  $Z$ , computed by Eq. (7), is the total frequency of all the existing states in the document cluster  $D$ . The other parameters of belief measure will follow the definitions presented in Section 2.

**Belief Representations of Subjective Opinions:** The basic definitions of belief, disbelief, uncertainty and relative atomicity remain the same as in Section 2. We re-define some of the other definitions in the following.

**Propositional Atomicity:** Let  $\Theta$  be a frame of discernment and  $x, y \in 2^\Theta$ . Then for any given  $y \neq \emptyset$ , the propositional atomicity of  $x$  is the average relative atomicity values of all  $x$  to  $y$ . Precisely,

$$a_p(x) = \frac{\sum_{\forall |a(x/y) \neq 0|} a(x/y)}{|a(x/y) \neq 0|}, \quad x, y \in 2^\Theta, y \neq \emptyset \quad (9)$$

Accordingly, *Probability Expectation* for a given proposition (sentence in this case)  $x$  can be re-written as

$$PE(x) = b(x) + a_p(x)u(x), \quad (10)$$

where  $b(x)$ ,  $u(x)$ , and  $a_p(x)$  are belief, uncertainty and propositional atomicity of proposition  $x$ . Opinions of a sentence can be measured by this probability expectation as in Eq.(10) using propositional atomicity.

### 3.2 Procedures

Generally, there are three major problems associated with multi-document summaries: (1) recognizing and coping with redundancy, (2) identifying important differences among documents, and (3) ensuring summary coherence, even when material stems from different source documents (Radev et al., 2002). We have addressed problem (1) in our method. Problems (2) and (3) are not very significant for us, as our proposed method is applicable for a cluster of related or coherent documents. Multi-document summary generation using SubSum involves the following steps:

**Step 1** Compute opinions as representativeness scores for the sentences in  $S$ .

**Step 2** Pick the sentence  $s \in S$  with the greatest opinion based on 'ordering of opinions'.

**Step 3** For each state  $x$  in  $s$ , update its belief mass by setting it to a small number close to zero.

**Step 4** If the desired summary length has not been reached, go back to step 1.

Steps 1 and 3 are explained in the following.

**Computation of Opinions:** A key step of SubSum is to assign a representativeness score (opinion) for each sentence  $s_i \in S$ , for the document cluster  $D$ . Algorithm 1 explains how SubSum computes opinions.

---

**ALGORITHM 1:** Computing Opinions of Sentences

---

**Input:** A document cluster  $D$  containing a set  $S$  of sentences.

**Output:** A weighted list of sentences  $S_{weighted} \in S$  for  $D$ .

- 1 Pre-process  $D$ ;
  - 2 Extract the states  $X$ , where  $X \in 2^\Theta$ ;
  - 3 Assign *belief masses* to the states using Eq.(8);
  - 4 **foreach** sentence  $s \in S$  **do**
  - 5     Apply Eq.(1) to compute *belief*  $b(s)$ ;
  - 6     Apply Eq.(3) to compute *uncertainty*  $u(s)$ ;
  - 7     Apply Eq.(9) to compute *propositional atomicity*  $a_p(s)$ ;
  - 8     Apply Eq.(10) to compute *probability expectation*  $PE(s)$ ;
  - 9 **end**
  - 10  $S_{weighted}$  = weighted list of sentences.
- 

**Context Adjustment:** Using frequency-based approaches to determine summary content in multi-document summarization results in a repetitive summary (Nenkova et al., 2006). In SubSum, the assignment of belief masses is dependent on frequency of occurrence of words. Thus, we need to consider removal of redundant contexts. The basic intuition of redundancy removal has been taken from (Nenkova et al., 2006). For each state (atomic or composite)  $x$  in the sentence  $s$  chosen at step 3, we update its belief masses by setting it to a very small number close to 0. Here we use 0.00001 for this number.

## 4 Evaluation

Qualitative analysis of summarizers is done by comparing them against human abstracts using ROUGE (Lin, 2004). For evaluation of our SubSum framework, we have compared ROUGE-generated recall, precision, and  $F$ -measure with baseline summaries and other summarizers using standard benchmarks of DUC (Document Understanding Conference `duc.nist.gov`)-DUC2001, DUC2002, and DUC2004 datasets.

### 4.1 Methodology

Frequency is a good predictor of content in human summaries according to (Nenkova et al., 2006). The word frequency feature forms the basis of the SubSum framework as we use the frequency of each state of  $\Theta$  to assign that state's belief mass (Eq. 8). This belief mass further contributes to the computation of opinion, which is the main sentence scoring function of SubSum. Through the experiments discussed below, we focus on how well SubSum corresponds to human-generated summaries and observe its performance when compared to other methods.

**Pre-processing of Documents:** Documents of each cluster are pre-processed by tokenizing them into words, removing stop-words and then by stemming (using Snowball `snowball.tartarus.org/index.php`) to retain the root form of the words. Unique instances of

sentences are selected by removing the duplicates if at all they occur in the document cluster. The processed word list and the sentence list are then used by SubSum for summarization.

**Comparison Partners:** The human-generated summaries provided by DUC datasets have been used as reference summaries for qualitative evaluation of automatic summarizers. These datasets also provide baseline summaries that can be used for comparison purposes. In addition, we have used the 16 system summaries from DUC2004. These summaries are referred to as *peer* followed by a reference number provided by the dataset.

Beyond the baselines, we have implemented a summarizer based on Composition Functions (CF) (Nenkova et al., 2006) as an additional comparison partner. Nenkova et al., (Nenkova et al., 2006) proposed a context-sensitive frequency-based summarizer that uses a composition function to assign importance weights to sentences. Out of the three proposed composition functions, we have chosen the best one, *Avr*, as our comparison partner.

In addition, to test the effect of context adjustments in multi-document summarization, we have also included comparisons with with *SubSum\_NoAdj* and *CF\_NoAdj*, which are the modified versions of SubSum and CF without performing any context adjustments (basically omitting Step 3 of the procedures mentioned in Sec.3.2).

**Evaluation Metrics:** Our evaluation was done using 1-gram setting of ROUGE (Lin, 2004), which was found to have the highest correlation with human judgments, namely at a confidence level of 95%. ROUGE calculates Precision, Recall, and *F*-measure values. In our experiments, summary length was set to 100 words. The exact parameters we used were `-c 95 -n 4 -x -m -r 1000 -w 1.2 -l 100 -a -z`.

## 4.2 Results

**Performance Comparison:** Since the length of summaries was set to 100 words, the performance of summarizers was determined by examining the ROUGE recall scores.

Table 1 presents the average performance (recall, precision, and *F*-measure) of the summarizers over all the 30 DUC2001 and 59 DUC2002 sub-datasets. From table 1, we can observe that SubSum corresponds well to the human reference summaries, outperforming the baseline and CF. In particular, for DUC2001, SubSum outperformed the baseline by 3.4% and CF by 0.6% in terms of recall. For DUC2002, SubSum outperformed the baseline by 4.2% and CF by 2.1% in terms of recall.

Methods	DUC2001			DUC2002		
	Recall	Precision	F-measure	Recall	Precision	F-measure
SubSum	0.3306	0.2836	0.3051	0.3294	0.3331	0.3312
SubSum_NoAdj	0.3244	0.2784	0.2995	0.3371	0.3393	0.3381
Baseline	0.2967	0.2558	0.2746	0.2872	0.2938	0.2901
CF	0.3246	0.2792	0.3001	0.3084	0.3362	0.3216
CF_NoAdj	0.3150	0.2703	0.2909	0.2927	0.3165	0.3040

Table 1: Average ROUGE-1 values on DUC2001 and DUC2002

Note that the DUC2001 and DUC2002 datasets have extremely strong baselines. As analyzed by (Nenkova, 2005), these baselines correspond to the selection of first *n* sentences of a news article. (Sarkar, 2012) has pointed out that beating DUC2001 and DUC2002 baseline summaries is difficult. According to (Das and Martins, 2007), many of the best performing summarization systems could not outperform the DUC2001 and DUC2002 baselines with statistical significance.

For example, the summarizers in (Nenkova et al., 2006) and (Erkan and Radev, 2004) either reached the baselines or outperformed them with a small margin.

Methods	Recall	Precision	F-measure
SubSum	0.3849	0.3418	0.3621
SubSum_NoAdj	0.3814	0.3382	0.3584
Baseline	0.3215	0.2954	0.3049
CF	0.3765	0.3347	0.3543
CF_NoAdj	0.3632	0.3213	0.3409
peer11	0.3254	0.3452	0.3317
peer27	0.3154	0.2987	0.3061
peer34	0.3830	0.3429	0.3618
peer44	0.3694	0.3365	0.3520
peer55	0.3694	0.3281	0.3474
peer65	0.3913	0.3462	0.3674
peer81	0.3764	0.3353	0.3546
peer93	0.3432	0.3391	0.3375
peer102	0.3857	0.3467	0.3651
peer111	0.2336	0.2097	0.2209
peer117	0.3476	0.3126	0.3291
peer120	0.3248	0.3765	0.3423
peer123	0.3056	0.3195	0.3097
peer124	0.3784	0.3375	0.3567
peer138	0.3422	0.3090	0.3247

Table 2: Average ROUGE-1 values on DUC2004

Table 2 presents the average performance (recall, precision, and  $F$ -measure) of the summarizers over all the 50 DUC2004 sub-datasets. From the results we can have similar observations that SubSum corresponds well to the human reference summaries, outperforming the baseline by 6.3% and CF by 0.8% in terms of recall.

SubSum performed extremely well compared to other DUC2004 peer summarization systems. It significantly outperformed most systems, roughly tied with peer34 and peer102, and slightly lost to peer65. Note that peer65 is a supervised HMM system (Conroy et al., 2004) that requires training data and parameter adjustment, while SubSum is non-supervised and totally data-driven. Overall, SubSum is among the best of DUC2004 participants.

**Effect of Context Adjustment:** Tables 1 and 2 have included the ROUGE evaluation scores for SubSum\_NoAdj and CF\_NoAdj as two other comparison partners of SubSum and CF. SubSum\_NoAdj and CF\_NoAdj are the modified versions of SubSum and CF without context adjustment in the summarization process. As discussed in Sec. 3.2, one of the main purposes of context adjustment is to remove context redundancy, which is a typical issue in multi-document summarization. From the results we can observe that CF outperformed CF\_NoAdj in all the three datasets, showing that the content selection capability of CF would be affected by the removal of the context adjustment step. On the contrary, SubSum\_NoAdj performed comparably to SubSum, where it slightly lost to SubSum for DUC2001 and DUC2004 and won by a small margin for DUC2002. This reflects the fact that SubSum can handle redundancy to some extent even without applying context adjustments separately.

## 5 Acknowledgement

This work was supported in part by the Texas Norman Hackerman Advanced Research Program (003656-0035-2009) and the National Science Foundation (OCI-1062439, CNS-1058724).



## References

- Barzilay, R. and McKeown, K. (2005). Sentence fusion for multidocument news summarization. *Computational Linguistics*, 31(3):297–328.
- Barzilay, R., McKeown, K., and Elhadad, M. (1999). Information fusion in the context of multi-document summarization. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*.
- Carbonell, J. and Goldstein, J. (1998). The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*.
- Conroy, J., Schlesinger, J., Goldstein, J., and O’leary, D. (2004). Left-brain/right-brain multi-document summarization. In *Proceedings of the Document Understanding Conference (DUC 2004)*.
- Das, D. and Martins, A. (2007). A survey on automatic text summarization. *Literature Survey for the Language and Statistics II course at CMU*, 4:192–195.
- Daumé III, H. and Marcu, D. (2005). Bayesian multi-document summarization at mse. In *ACL 2005, Workshop on Multilingual Summarization Evaluation*.
- Erkan, G. and Radev, D. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Intell. Res. (JAIR)*, 22:457–479.
- Gong, Y. and Liu, X. (2001). Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*.
- Harabagiu, S. and Lacatusu, F. (2010). Using topic themes for multi-document summarization. *ACM Transactions on Information Systems (TOIS)*, 28(3):13.
- Hovy, E. and Lin, C. (1998). Automated text summarization and the summarist system. In *Proceedings of a workshop on held at Baltimore, Maryland: October 13-15, 1998*.
- Ji, P. (2006). Multi-document summarization based on unsupervised clustering. In *Proceedings of the Third Asia conference on Information Retrieval Technology*. Springer-Verlag.
- Josang, A. (2001). A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311.
- Josang, A. and McAnally, D. (2010). Multiplication and comultiplication of beliefs. *International Journal of Approximate Reasoning*, 38(1):19–51.
- Kupiec, J., Pedersen, J., and Chen, F. (1995). A trainable document summarizer. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*.
- Li, L., Wang, D., Shen, C., and Li, T. (2010). Ontology-enriched multi-document summarization in disaster management. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*.

- Lin, C.-Y. (2004). Rouge: A package for automatic evaluation of summaries. Barcelona, Spain. Association for Computational Linguistics.
- Luhn, H. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165.
- McKeown, K. and Radev, D. (1995). Generating summaries of multiple news articles. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*.
- Mihalcea, R. and Tarau, P. (2005). A language independent algorithm for single and multiple document summarization. In *Proceedings of IJCNLP*.
- Nenkova, A. (2005). Automatic text summarization of newswire: Lessons learned from the document understanding conference. In *Proceedings of the National Conference on Artificial Intelligence*.
- Nenkova, A. and Vanderwende, L. (2005). The impact of frequency on summarization. *Microsoft Research, Redmond, Washington, Tech. Rep. MSR-TR-2005-101*.
- Nenkova, A., Vanderwende, L., and McKeown, K. (2006). A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*.
- Pardo, T., Rino, L., and Nunes, M. (2002). Extractive summarization: how to identify the gist of a text. In *the Proceedings of the 1st International Information Technology Symposium–I2TS*.
- Park, S., Lee, J.-H., Kim, D.-H., and Ahn, C.-M. (2007). Multi-document summarization using weighted similarity between topic and clustering-based non-negative semantic feature. In *Proceedings of the joint 9th Asia-Pacific web and 8th international conference on web-age information management conference on Advances in data and web management*. Springer-Verlag.
- Radev, D., Jing, H., and Budzikowska, M. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. *Ann Arbor*, 1001:48103.
- Radev, D. and McKeown, K. (1998). Generating natural language summaries from multiple on-line sources. *Computational Linguistics*, 24(3):469–500.
- Radev, D. R., Hovy, E., and McKeown, K. (2002). Introduction to the special issue on summarization. *Computational Linguistics*, 28(4):399–408.
- Ribaldo, R., Akabane, A. T., Rino, L. H. M., and Pardo, T. A. S. (2012). Graph-based methods for multi-document summarization: exploring relationship maps, complex networks and discourse information. In *Proceedings of the 10th international conference on Computational Processing of the Portuguese Language*.
- Sarkar, K. (2012). Bengali text summarization by sentence extraction. *Arxiv preprint arXiv:1201.2240*.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press Princeton, NJ.

- Shen, C. and Li, T. (2010). Multi-document summarization via the minimum dominating set. In *Proceedings of the 23rd International Conference on Computational Linguistics*.
- Wan, X. (2008). An exploration of document impact on graph-based multi-document summarization. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Wang, D., Li, T., Zhu, S., and Ding, C. (2008). Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*.
- Wei, F., Li, W., Lu, Q., and He, Y. (2008). A cluster-sensitive graph model for query-oriented multi-document summarization. In *Proceedings of the IR research, 30th European conference on Advances in information retrieval*. Springer-Verlag.
- Xia, Y., Zhang, Y., and Yao, J. (2011). Co-clustering sentences and terms for multi-document summarization. In *Proceedings of the 12th international conference on Computational linguistics and intelligent text processing - Volume Part II*.
- Zhang, R., Li, W., and Lu, Q. (2010). Sentence ordering with event-enriched semantics and two-layered clustering for multi-document news summarization. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*.

