Discourse Structures to Reduce Discourse Incoherence in Blog Summarization

Shamima Mithun
Concordia University
Computer Science & Software Eng.
Montreal, Canada
s_mithun@encs.concordia.ca

Leila Kosseim
Concordia University
Computer Science & Software Eng.
Montreal, Canada
kosseim@cse.concordia.ca

Abstract

Discourse incoherence is an important and typical problem with multi-document extractive summaries. To address this issue, we have developed a schema-based summarization approach for query-based blog summaries that utilizes discourse structures. In our schema design, we tried to model discourse structures which are typically used by humans in their summary writing in response to a particular question type. In our approach, a sentence instantiates a specific slot of the schema based on its discourse structures. To validate our approach, we have built a system named BlogSum and have evaluated its performance through 4 human participants using a likert scale of 1 to 5. The evaluation results show that our approach has significantly improved summary coherence compared to the summaries with no discourse structuring without compromising on content evaluation.

1 Introduction

Research on text summarization dates back since the 1950s and with the growth of the Internet it has become a popular research topic in last decade. Text summarization reduces text search time by providing the most relevant information from the documents which enables users to comprehend more quickly the main ideas of a set of documents. Over time, different summarization techniques have been developed and evaluated. Although significant improvement continues to be made, the summaries generated automatically are by no means of the same quality as their human created counter parts. The area in which automatic summaries differ most from human generated summaries is text coherence (Otterbacher et al., 2002; Conroy and Dang, 2008; Genest et al., 2009).

Coherence problems can be the result of different phenomena: discourse incoherence, redundancy, temporal incoherence, grammatical mistakes or many other linguistic problems. In a manual analysis of 15 summaries, (Otterbacher et al., 2002) showed that coherence problems are caused mostly by discourse incoherence (34%) where the main concern is the lack of relations between sentences as well as in the overall summary.

Recently, (Genest et al., 2009) demonstrated that the performance of automatic summarizers in term of linguistic quality is significantly weaker compared to that of a baseline consisting of sentences extracted from the source documents by 5 human extractors and added to the summary without any modification. This result indicates that there is still much space to improve coherence of summaries even for pure extractive summaries.

1.1 Discourse Incoherence

Computational theories on discourse coherence were introduced by (Hobbs, 1985; Mann, 1988). According to (Mann, 1988), a discourse is coherent if the hearer knows the communicative role of each of its portion; that is, if the hearer knows how the speaker intends each clause to relate to each other. As a result, a summary will exhibit discourse incoherence if the reader cannot identify the communicative intentions of the writer from the clauses or if the clauses do not seem to be interrelated.

Consider the following summary (ID:T1001.8[1]) which contains discourse incoherence problems (shown in Figure 1). The summary for the question is incoherent. Even though all the sentences are relevant to the query, improper sentence ordering degrades the coherence of this summary. In ad-

[1]The summary is taken from the TAC 2008 opinion summarization track.

Proceedings of Recent Advances in Natural Language Processing, pages 479–486,
Hissar, Bulgaria, 12-14 September 2011.

dition, sentence 3 contains a pronoun *(it)* without having an antecedent. One possible better ordering for this summary would be 4-3-1-2 or 4-3-2-1.

Figure 1: A Sample Summary

Topic: *Carmax*
Question: *What motivated positive opinions of Carmax from car buyers?*
Summary:
(1) It's like going to disney world for car buyers.
(2) have to say that Carmax rocks.
(3) We bought it at Carmax, and I continue to have nothing bad to say about that company.
(4) After our last big car milestone, we've had an odyssey with cars.

A summary with poor coherence confuses the readers and degrades the quality and readability of the summary. The proper sentence order significantly improves the readability of summaries. (Lapata, 2003) experimentally showed that the time to read a summary strongly correlates with the arrangement of sentences.

1.2 State of the Art

Currently, most of the automatic summarization systems for news articles use an extractive approach. In general, this approach works in two steps: in the first step, the most salient sentences are extracted from the source documents and in the second step, these sentences are ordered to create a summary. Since in the first step, sentences may be selected from multiple documents or without consideration to their interdependency with other sentences this may cause text incoherence. Moreover, in multi-document summarization, documents may be written by different writers who have different perspectives and writing styles thus exasperating coherence problems. To improve coherence, the second step tries to reorder the selected sentences appropriately.

As part of the sentence ordering, two major types of approaches are used to address coherence: making use of chronological information (McKeown et al., 2002), and learning the natural order of sentences from large corpora (Barzilay and Lee, 2004; Lapata, 2003). However, in the first case, if the source documents are not event-based, the quality of the summaries will be degraded because temporal cues are missing. In the later case, probabilistic models of text structures are trained on a large corpus. If the genre of the corpus and the source documents mismatch then they will perform poorly.

Summarization for opinionated text is a recent endeavor. Query-based blog summarization approaches have been first developed in the TAC 2008 opinion summarization track. Most of these summarization approaches (e.g. (Murray et al., 2008)) use sentence scores for summary generation. Some of these approaches (e.g. (Kumar and Chatterjee, 2008)) use the sentence order of the original documents to specify the sentence order of the summary. Recent work (e.g. (Paul et al., 2010)) on blog summarization also mostly use sentence scores for summary generation. However, these approaches hardly can be effective in coherence improvement. To the best of our knowledge, text schemata and discourse relations, found effective in news summarization and question answering (Blair-Goldensohn and McKeown, 2006; Sauper and Barzilay, 2009), were never used in blog summarization.

In our research, we try to reduce discourse incoherence of extractive summaries; and in particular in query-based blog summaries. In this work, we propose a domain independent query-based blog summarization approach to address discourse incoherence using discourse structures in the framework of schemata. To verify our approach we have developed a system called BlogSum and evaluated its performance using the Text Analysis Conference (TAC 2008) opinion summarization track data[2].

2 Discourse Structures to Reduce Discourse Incoherence

In our research, we are interested in query-based blog summarization. Nowadays, because of the rapid growth of the Social Web, a large amount of informal opinionated texts are available on any topic. Query-based opinion summarizers present what people think or feel on a given topic in a condensed manner to analyze others' opinions regarding a specific question (e.g.*Why do people like Starbucks better than Dunkin Donuts?*). This research interest motivated us to develop an effective query-based multi-document opinion summarization approach for blogs and we utilize discourse structures in the framework of schema to improve discourse coherence.

[2] http://www.nist.gov/tac/

2.1 Previous Work on Schemas

(McKeown, 1985) introduced a schema-based approach for text planning based on the observation that certain standard patterns of discourse organization (schema) are more effective to achieve a particular discourse goal. (McKeown, 1985) demonstrated the usability of this schema-based approach for a domain-dependent question answering application. In this application, she designed various schemata that incorporate discourse structures which are typically used in human writing for a specific question type (e.g. *identification*). In most recent summarization work, (Sauper and Barzilay, 2009) also tried to utilize discourse structures learned from domain relevant articles to design schemata (or templates) for structured domains (e.g. Wikipedia pages).

We also believe that for any domain, for a particular type of query, certain types of sentences if organized in a certain order can meet the communicative goal more effectively and create a more coherent text. For example, to take (McKeown, 1985)'s example, to define an entity or event (e.g. *what is a ship?*) it is natural to first include the identification of the item as a member of a generic class, then to describe the object's constituency or attributes followed by a specific example and so on. On the other hand, a comparison of two objects should use another combination to be effective and coherent.

2.2 Our Schema-based Approach

In our schema-based approach, the basic units of a schema are *rhetorical predicates* which characterize the structural purposes of a text and delineate the discourse relations between propositions.

2.2.1 Our Set of Rhetorical Predicates

Six main types of rhetorical predicates which have been found most useful for our blog summarization application were considered:

1. Attributive: Provides details about an entity or event. It can be used to illustrate a particular feature about a concept - e.g. *Mary has a pink coat.*
2. Comparison: Gives a comparison and contrast among different situations - e.g. *Perhaps that's why for my European taste Starbucks makes great espresso while Dunkin's stinks.*
3. Contingency: Provides cause, condition, reason, evidence for a situation, result or claim - e.g. *The meat is good because they slice it right in front of you.*
4. Illustration: Is used to provide additional information or detail about a situation - e.g. *Allied Capital is a closed-end management investment company that will operate as a business development concern.*
5. Attribution: Provides instances of reported speech both direct and indirect which may express feelings, thoughts, or hopes - e.g. *I said actually I think Zillow is great.*
6. Topic-opinion: Can be used to express an opinion; an agent can express internal feeling or belief towards an object or an event - e.g. *The thing that I love about their sandwiches is the bread.*

Rhetorical relations characterized by *comparison*, *illustration*, and *contingency* predicates are also considered by other research groups (e.g. (Carlson, 2001)). We consider three additional classes of predicates *attributive*, *attribution*, and *topic-opinion*. The *attributive* predicate and the *attribution* predicate are also listed in Grimes' predicates (McKeown, 1985) and (Carlson, 2001)'s relations list, respectively. We introduced *Topic-opinion* predicates to represent opinions which are not expressed by reported speech.

2.2.2 Schemata Design

In our schema-based approach, sentences need to be classified and organized based on what rhetorical predicates they contain. We designed and associated appropriate schemata (e.g. *compare and contrast*) to generate a summary that answers specific types of questions (e.g. *comparison, suggestion*) by defining constraints on the types of predicates (e.g. *comparison*, *attribution*) and the order in which they should appear in the output summary for a particular question type. In our approach, schemata help to ensure the global coherence of the summary.

Figure 2 shows a sample schema that can be used to answer a *comparison* question. According to this schema, a sentence to be included in the beginning of the summary needs to be classified as either a *Comparison* predicate or a *Contingency* predicate followed by *Topic-opinion* or *Attribution* predicates then by *Illustration* predicates. More formally, one or more *Comparison* or *Contingency* predicates followed by zero or many

Topic-opinion or *Attribution* predicates followed by zero or many *Illustration* predicates can be used[3].

Figure 2: A Sample Schema

Predicates & Constraints
Predicate: $\{$*Comparison/Contingency*$\}^+$
Constraint: Compared objects, Sentence focus
Predicate:$\{$*Topic-opinion/Attribution*$\}^*$
Constraint: Sentence polarity
Predicate: *Illustration**

From Figure 2, we can see that constraints are also defined on predicates based on their semantic content. In the example, the *Comparison* and *Contingency* predicates must contain all objects or events which are being compared and the topic[4] of the sentence needs to be the focus of the sentence; and *Topic-opinion* and *Attribution* predicates must have the same polarity as the question. In order to answer a different type of questions, a different schema would be more appropriate. In this approach, schemata will help to improve coherence by specifying a higher level text organization by constraining the order of the predicates.

3 BlogSum

In order to test our approach, we have built a system called BlogSum. Given an initial question on a particular topic and a set of related blogs, BlogSum performs sentence selection then content organization.

3.1 Content Organization

The content organization requires as input a ranked list of sentences from the document set. In our test, we have developed our own sentence extractor based on question similarity, topic similarity, and subjectivity scores. However, any other sentence ranker could have been used. The role of content organization is to select a few sentences from the candidate sentences and order them so as to produce a coherent and query relevant summary. For content organization, BlogSum performs the following tasks: A) Question Categorization, B) Schema Selection, C) Predicate Identification, and D) Sentence Selection and Ordering.

[3]Following (McKeown, 1985)'s notations, the symbol / indicates an alternative, * indicates that the item may appear 0 to n times, + indicates that the item may appear 1 to n times.

[4]Text specified in the Target in the TAC 2008 task data.

3.1.1 Question Categorization

Our content organization approach first categorizes questions to determine which schema will better convey the expected communicative goal of the answer for a particular question type and should be used for text planning.

By analyzing the TAC 2008 opinion summarization track questions manually, we have identified 3 categories of questions based on their communicative goals, namely: *comparison*, *suggestion*, and *reason*.

1. *Comparison* questions ask about the differences between objects - e.g. *Why do people like Starbucks better than Dunkin Donuts?*
2. *Suggestion* questions ask for suggestions to solve some problems - e.g. *What do Canadian political parties want to happen regarding NAFTA?*
3. *Reason* questions ask for reasons for some claim - e.g. *Why do people like Mythbusters?*

To automatically identify a unseen question into one of these 3 categories, we have designed lexical patterns by analyzing the same set of questions which we used to identify question categories.

3.1.2 Schema Selection

Based on the observation that for a particular question type, sentences need to be organized in a specific order to be coherent, we have designed three schemata, one for each question type, 1) *comparison*, 2) *suggestion*, and 3) *reason*. To design these schemata, we have analyzed 50 summaries generated by participating systems at the TAC 2008 opinion summarization track. From our analysis, we have derived which question types should contain which type of predicates. Each schema is designed based on giving priority to its associated question type and subjective sentences as we are generating summaries for opinionated texts. For each type of schema, we have also defined constraints on the predicates based on their semantic content to improve the question relevance. As part of the schema selection, BlogSum selects the associated schema for a specific question category to select and order sentences for the final summary.

It must be noted that schemata can be designed in different ways. However, our current content organization approach allows the generation of different summaries for particular question types by providing flexible sentence selection and reordering strategies.

3.1.3 Predicate Identification

In our approach, candidate sentences need to be classified into a predefined set of rhetorical predicates to fill the various slots of the matched schema - we called this process predicate identification.

In (Mithun and Kosseim, 2011), we have introduced a domain independent approach to identify which rhetorical predicates are conveyed by a sentence. As specified in (Mithun and Kosseim, 2011), predicates can describe a single proposition or the relation between propositions. To identify the predicates between propositions - e.g. *evidence*, we have used the SPADE discourse parser (Soricut and Marcu, 2003). On the other hand, in order to identify predicates within a single proposition - e.g. *attributive*, we have used three other taggers: comparison (Jindal and Liu, 2006), topic-opinion (Fei et al., 2008), and our attributive tagger (Mithun and Kosseim, 2011). By combining these approaches, a sentence is tagged with all possible predicates that it may contain and ready to be used in a schema.

3.1.4 Sentence Selection and Ordering

In BlogSum, sentence selection and ordering is accomplished in the following manner:

First, candidate sentences fill particular slots in the selected schema based on which rhetorical predicate they convey and whether they fulfil the semantic constraints. This process is performed for each candidate sentence based on their extraction score until the maximum summary length is reached. Since the use of schemata alone is not sufficient to achieve a total order; for example there may be several sentences that can fill a particular slot of a selected schema, we have used post-schemata heuristics to improve this partial order and coherence. These heuristics include: topical similarity, explicit discourse markers, and context. At the end of the sentence ordering process, to create a total order, we finally use the rank of the sentences in the original list of candidates. Let us now describe the post-schemata heuristics.

1. **Topical Similarity**: In the schema for a particular predicate type (e.g. *contingency*), we tried to use topical similarity in order to group sentences that describe the same topic together. To find topically similar sentences we used the cosine similarity using *tf.idf*.
2. **Explicit Discourse Markers**: To further improve discourse coherence, we add conjunctive markers based on the sentences' topical similarity and polarity value. For example, if two sentences are topically similar, our approach will place them next to each other and make a single sentence out of them using a conjunctive marker (e.g. *and*) even though these sentences may not adjacent in the candidate list. If BlogSum finds another sentence on this topic, it will position that sentence together using another conjunctive marker.
3. **Context:** To improve discourse coherence further, if a potential sentence starts with a pronoun without having a potential antecedent, we include its previous sentence from the source document as a context from the original document.

3.1.5 An Example to Describe Content Organization

To illustrate the content organization process, let us take the following example:
Question: *What motivated positive opinions of Carmax from car buyers?*

Figure 3: Partial Candidate List

(1) With Carmax you will generally always pay more than from going to a good used car dealer. (2) We bought it at Carmax, and I continue to have nothing bad to say about that company. (3) Carmax did split the bill which made me happy. (4) Not sure if you have a Carmax near you, but I've had 2 good buying experiences from them. (5) have to say that Carmax rocks. (6) At Carmax, the price is the price and when you want a car you go get one. (7) Sometimes I wonder why all businesses can't be like Carmax. (8) Arthur Smith, 36, has been living in a van outside the CarMax lot, 24 hours a day, for more than a month.

The above question has been classified as a *Reason* type question based on the question pattern matching. A subset of the candidate sentences generated by BlogSum for this question is shown in Figure 3. For this question, the *Reason* schema is used to order the sentences. The *Reason* schema and the final order of the sentences are shown in Figure 4. In Figure 4, the constraints "sentence polarity", "compared objects", and "sentence focus" indicate that the sentence needs to have the same polarity as the question, the sentence needs to contain all objects which are being compared, and the topic of the sentence needs to be the focus of the sentence, respectively.

Figure 4: Summary Generated using the Reason Schema

Schema	**Sentences**
Predicate: $\{$*Topic-opinion/ Attribution*$\}^+$ Constraint: sentence polarity.	(2-1) After our last big car milestone, we've had an odyssey with cars. (2, 4) We bought it at Carmax, and I continue to have nothing bad to say about that company; not sure if you have a Carmax near you, but I've had 2 good experiences from them. (3) Moreover, Carmax did split the bill which made me happy. (5) have to say that Carmax rocks.
Predicate: $\{$*Contingency/ Comparison*$\}^*$ Constraint: compared objects, sentence focus.	(7) Sometimes I wonder why all businesses can't be like Carmax.
Predicate: *Attributive** Constraint: sentence focus.	(6) At Carmax, the price is the price and when you want a car you go get one.

In this sample, we can see that the schema did not include sentences 1 and 8 in the final summary even though the summary is within the length limit. This is because these sentences did not fit within the *Reason* schema. Though sentence 1 was classified as containing a *comparison* predicate, it did not fulfil the semantic constraint (shown in Figure 4) that the topic of the sentence (Carmax) be the focus of the sentence[5]. On the other hand, sentence 8 was not included, because it did not contain any of the rhetorical predicate which can fill the slots of this schema.

We can see that since for the sentence 2, the antecedent of the pronoun *it* is missing, our context heuristic added the preceding sentence 2-1 of sentence 2 from the source document. Our approach placed sentences 2 and 4 next to each other because of their topical similarity and also merged them using the conjunctive marker ';'. We can also see that the system added the discourse marker "Moreover" in sentence 3. In the summary, sentences 6 and 7 are also reordered compared to the candidate list based on the rhetorical predicate category they contained.

4 Evaluation

In order to test our approach, we have evaluated BlogSum-generated summaries for coherence and overall readability.

4.1 Corpus and Experimental Design

In this evaluation, we have used the TAC 2008 opinion summarization track data. The data set consists of 50 questions on 28 topics; on each topic one or two questions are asked and 10 to 50 relevant documents are given. For each question, one summary was generated by OList and one by BlogSum and the maximum summary length was restricted to 250 words. To evaluate coherence, we did not use the ROUGE metric because from a manual analysis (Blair-Goldensohn and McKeown, 2006) found that the ordering of content within the summaries is an aspect which is not evaluated by ROUGE. Instead, 4 participants manually rated 50 summaries from OList and 50 summaries from BlogSum for coherence with respect to the question for which the summary is generated using a blind evaluation. These summaries were rated on a likert scale of 1 to 5 where 1 refers to "very poor" and 5 refers to "very good". As a baseline, we used the original ranked list of candidate sentences (OList), and we compared it to the final summaries which are generated by BlogSum after applying the discourse structuring.

4.2 Results

In the evaluation, to calculate the score of BlogSum and OList for a particular question, we calculated the average scores of all annotators' ratings to that question. Table 1 shows the performance comparison between BlogSum and OList.

[5] To identify this, we test if the subject or object of the sentence is the topic.

We can see that 52% of the time BlogSum summaries were rated better than OList summaries; 30% of the time both performed equally; and 18% of the time BlogSum was weaker than OList. This means that 52% of the time, our approach has improved the coherence compared to that of the original candidate list (OList).

Table 1: Summary of the Comparison

Comparison	%
BlogSum Score > OList Score	52%
BlogSum Score = OList Score	30%
BlogSum Score < OList Score	18%

Table 2 shows the performance of BlogSum versus OList on each likert scale; where Δ shows the performance difference. From Table 2, we can see that BlogSum outperformed OList in the scale of "very good" and "good" by 16% and 8%, respectively; and improved the performance in "barely acceptable" and "poor" categories by 12% and 14%, respectively.

Table 2: Performance of BlogSum vs. OList

Category	OList	BlogSum	Δ
Very Good	8%	24%	16%
Good	22%	30%	8%
Barely Acceptable	36%	24%	-12%
Poor	22%	8%	-14%
Very Poor	12%	14%	2%

We have also evaluated if the difference in performance between BlogSum and OList was statistically significant using the t-test. The t-test results show that in a two-tailed test, BlogSum performed significantly better than OList with a p-value of 0.0223.

In this experiment, we also calculated the inter-annotator agreement using Cohen's kappa coefficient to verify the annotation subjectivity. We have found that the average pair-wise inter-annotator agreement is substantial with the kappa-value of 0.76.

4.3 Error Analysis

From the evaluation results of Table 2, we can see that about 54% of the time the coherence of BlogSum is categorized as "very good" or "good"; about 24% of the time "barely acceptable"; but still, about 22% of the time the summaries were considered "poor" or "very poor". From an error analysis, we have found that many of the summaries are ranked in the lower categories because of their question irrelevance, an incorrect polarity identification or a predicate tagging error. Although the annotators were asked to evaluate coherence only, they found it difficult to abstract all other factors and assign a high score to a coherent text that did not answer the question properly.

The evaluation results of Table 1 show that 52% of the time our approach has improved the coherence over the original candidate list (OList). However, in 18% of the time (9 summaries), our approach was weaker than OList. We have analyzed these 9 summaries and found that in 4 cases, some sentences were tagged with the wrong polarity; as a result when the post-schemata heuristics were applied (e.g. conjunctive marker) they made the summaries weaker. In 3 cases, sentences were tagged with the wrong predicates thus they were included in the final summaries yet they should not have and in 2 other cases, BlogSum excluded sentences which were actually potential sentences again because of a wrong polarity identification and predicate tagging.

In order to determine if the improvement in coherence was done at the expense of content, we evaluated this aspect by using the TAC 2008 opinion summarization track data and the ROUGE metric using answer nuggets (provided by TAC), which had been created to evaluate participants' summaries at TAC, as gold standard summaries. In this evaluation, we compared original candidate list (OList) to BlogSum-generated final summaries. The ROUGE scores are also calculated for all 36 submissions in the TAC 2008 opinion summarization track. In this experiment, BlogSum achieved a better F-Measure for ROUGE-2 and ROUGE-SU4 compared to OList. Results show that BlogSum gained 18% and 16% in F-Measure over OList using ROUGE-2 and ROUGE-SU4, respectively. Compared to the other systems, BlogSum ranked third and its F-Measure score difference from the best system is very small. Both BlogSum and OList performed better than the average systems.

5 Conclusion and Future Work

In this work, we have used discourse structures with the help of schema to improve discourse coherence of query-based blog summaries. In our schema based approach, we exploited discourse structures in schema design and in instantiating the schema to fill a slot. We have developed a query-based blog summarization system called BlogSum to validate our approach. The performance of BlogSum was evaluated manually using the TAC 2008 question answering track data by 4 human participants in a likert scale of 1 to 5. The results indicate that about 54% of the summaries are rated as "very good" or "good" as opposed to 30% for the summaries with no discourse structuring. The evaluation results also show that our approach has significantly improved summary coherence compared to that of the original candidate list without compromising on content.

An error analysis following the human evaluation has shown that an important source of error in low ranking summaries is question irrelevance. As a result, we plan to test our content organization strategies with a better initial candidate list. In the future, we also plan to evaluate the individual contribution of the post-schemata heuristics to the overall coherence of the summaries.

Acknowledgement

The authors would like to thank the anonymous referees for their comments on a previous version of the paper.

This work was financially supported by NSERC.

References

Barzilay, R., Lee, L.: Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization. *In Proceedings of HLT-NAACL*, 113–120 (2004), Boston, USA.

Blair-Goldensohn, S., McKeown, K.R.: Integrating Rhetorical-Semantic Relation Models for Query-Focused Summarization. *In Proceedings of the Document Understanding Conference (DUC) Workshop at NAACL-HLT*, (2006), New York, USA.

Carlson, L., Marcu, D.: Discourse Tagging Reference Manual. University of Southern California Information Sciences Institute, ISI-TR-545, 2001.

Conroy, J. M., Dang, H. J.: Mind the Gap: Dangers of Divorcing Evaluations of Summary Content from Linguistic Quality. *In Proceedings of the 22nd International Conference on Computational Linguistics (Coling)*, 145–152 (2008), Manchester, UK.

Fei, Z., Huang, X., Wu, L.: Mining the Relation between Sentiment Expression and Target Using Dependency of Words. *PACLIC20: Coling 2008: Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation*, 257–264 (2008), Wuhan, China.

Genest, P., Lapalme, G., Yousfi-Monod, M: HEX-TAC: the Creation of a Manual Extractive Run. *In Proceedings Text Analysis Conference*, (2010), Gaithersburg, USA.

Hobbs, J. R.: On the Coherence and Structure of Discourse. Center for the Study of Language and Information, Stanford University, Report No. CSLI-85-37, 1985.

Jaidka, K., Khoo, C. S. G., Na, J.: Imitating Human Literature Review Writing: An Approach to Multi-document Summarization. *In Proceedings of the 12th International Conference on Asia-Pacific Digital Libraries*, 116–119 (2010), Gold Coast, Australia.

Jindal, N., Liu, B.: Identifying Comparative Sentences in Text Documents. *In Proceedings of ACM SIGIR*, 244–251 (2006), Seattle, USA.

Kumar, S., Chatterjee, D.: IIIT Kharagpur at TAC 2008: Statistical Model for Opinion Summarization. *In Proceedings Text Analysis Conference*, (2008), Gaithersburg, USA.

Lapata, M.: Probabilistic Text Structuring: Experiments with Sentence Ordering. *In Proceedings of the Annual Meeting of ACL*, 545–552 (2003), Sapporo, Japan.

Mann, W., Thompson S.: Rhetorical Structure Theory : Toward a Functional Theory of Text Organisation. *J. Text*, 3(8):234-281, 1988.

Marcu, D.: From Discourse Structures to Text Summaries. *Proceedings of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization*, 82–88 (1997), Madrid, Spain.

McKeown, K.R.: Discourse Strategies for Generating Natural-Language Text. *J. Artificial Intelligence*, 27(1):1–41, 1985.

McKeown, K., Klavans, J., Hatzivassiloglou, V., Barzilay, R., Eskin, E.: Towards Multidocument Summarization by Reformulation: Progress and prospects. *AAAI/IAAI*, 27–36 (2002), Edmonton, Canada.

Mithun, S., Kosseim, L.: Comparing Approaches to Tag Discourse Relations. *In Proceedings (1) of CICLing*, 328–339 (2011), Tokyo, Japan.

Murray, G., Joty, S., Carenini, G., Ng, R.: The University of British Columbia at TAC 2008. *In Proceedings Text Analysis Conference*, (2008), Gaithersburg, USA.

Otterbacher, J. C., Radev, D. R., Luo, A.: Revisions that Improve Cohesion in Multi-document Summaries: A Preliminary Study. *In Proceedings of the Workshop on Automatic Summarization (including DUC 2002)*, 27–36 (2002), Philadelphia, USA.

Paul, M. J., Zhai. C., Girju, R.: Summarizing contrastive viewpoints in opinionated text. *In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 65–75 (2010).

Sauper, C., Barzilay, R.: Automatically Generating Wikipedia Articles: A Structure-Aware Approach. *In Proceedings of the Joint Conference of ACL and AFNLP*, 208–216 (2009), Suntec, Singapore.

Soricut, R., Marcu, D.: Sentence Level Discourse Parsing using Syntactic and Lexical Information. *In Proceedings of NAACL'03*, 149–156 (2003), Edmonton, Canada.