Exploring the Rhetorical Structure Theory for Multi-document Summarization

Estudio de la aplicación de la Teoría de Estructura Retórica RST en Sumarización Multi-documento

Paula C.F.Cardoso
Universidade Federal de Lavras
Lavras, MG, Brazil
paulastm@gmail.com

Maria L.R.C. Jorge, Thiago A.S. Pardo
Universidade de São Paulo
São Carlos, SP, Brazil
{mluciacj,taspardo}@icmc.usp.br

Resumen: RST ha sido aplicada con éxito en várias áreas. En este trabajo se realizó un estudio sobre el impacto de RST en el área de sumarización multi-documento. En particular, son propuestos dos métodos: con base en reglas pré-definidas y aprendizaje estadístico. Los resultados se muestran prometedores.

Palabras clave: Rhetorical Structure Theory, Sumarización multi-documento

Abstract: Rhetorical Structure Theory (RST) has been applied in different areas, such as single document summarization, with promising results. In this paper, we discuss how Multi-document Summarization may benefit from RST in both rule-based and statistical methods. Results show that RST may contribute to produce more informative summaries.

Keywords: Rhetorical Structure Theory, Multi-document summarization

1 *Introduction*

The emergence of new technologies has led to an increase in the amount of textual information available online. In this context, Multi-document Summarization (MDS) has become an interesting task in the research community of Natural Language Processing. MDS aims at selecting the relevant information from multiple documents in order to produce a summary (Mani, 2001). Many models have been developed for this aim, dealing with the task with techniques that make use of superficial and deep information. Superficial techniques are based on information such as: title-text word overlapping, tf-idf, sentence position, statistical topic models, etc. On the other hand, deep methods explore linguistic information such as ontologies, part of speech, semantic relations among textual segments, etc.

In MDS for texts written in Brazilian Portuguese, some works have developed superficial methods (Pardo, 2005) and some others deep methods (Jorge and Pardo, 2010; Ribaldo, 2013). Particularly, deep approaches have led to the production of better summaries, in terms of informativeness and linguistic quality. One important theory that guides the deep approaches for MDS is Cross-Document Structure Theory (CST) (Radev, 2000), which has been widely explored (Zhang, Goldenshon, and Radev, 2002; Otterbacher, Radev, and Luo, 2002; Jorge and Pardo, 2010; Ribaldo, 2013) in the area. This theory models the multi-document phenomena such as redundant, contradictory and complementary information across textual portions (commonly sentences) with a set of relations that represent similarities and differences among related texts. These relations are commonly identified between pairs of sentences, coming from different sources. Moreover, these relations have shown to be good at capturing information that is highly repeated or elaborated across texts, but this criteria may leave aside portions of texts that are relevant within each text and that should be considered in a multi-document summary.

The Rhetorical Structure Theory (RST) (Mann and Thompson, 1987) was widely used for single document summarization. Techniques used in this area take advantage of the fact that textual segments are classified according to their importance: nuclei are more informative than satellites, and satellites, in certain relations, may be omitted (Ono, Sumita, and Miike, 1994; O'Donnell, 1997; Marcu, 1999; Uzêda, Pardo, and Nunes, 2010).

As far as we know, RST was not applied for MDS. In this paper, we incorporate RST in two methods for MDS, relying on the hypothesis that, by means of this theory, it is possible to capture important content that may improve the informativeness of multi-document summaries. One method proposes a modification to CSTSumm (Jorge and Pardo, 2010), a well-known discourse-based multi-document summarizer for Brazilian Portuguese texts, in order to incorporate RST to its sentence ranking strategy. Our other method consists of an statistical model, which is formulated with RST features. Both methods were developed over the CSTNews corpus (Cardoso et al., 2011), which was manually annotated with RST and CST relations. The automatic summaries were evaluated using the traditional ROUGE measure (Lin, 2004) and compared with state of the art methods, confirming the paper hypothesis.

This paper is organized as follows: in Section 2, a brief background about RST and some of the main approaches for MDS are presented; in Section 3, the RST-based methods for MDS are introduced; in section 4, the CSTNews corpus is described; in Section 5, results are presented and discussed; finally, Section 6 presents some final remarks.

2 Related work

2.1 RST and single document summarization

RST represents relations among propositions or discourse segments in a text and differentiates between nuclear and satellite information (Mann and Thompson, 1987). Nuclear segments are considered as the most important parts of a text, whereas satellites contribute to the nuclei and are secondary. Based on the nuclearity of text segments in RST trees, different methods for single document summarization have been proposed.

Ono et al. (1994) suggested a penalty score for units based on the nucleus-satellite structure of the RST tree. Satellites spans are assigned a lower score than spans that mostly take nucleus status. The penalty is defined as the number of satellites found in the path from the root of the discourse tree to the leaf.

O'Donnell (1997), in turn, assumes that ach RST relation has an associated relevance score that indicates how important the respective segments are for the summary. The method starts by associating the score 1 to the root of the tree and, then, traversing the tree in depth-first mode. Each time a satellite is found, the next node will have the corresponding score multiplied by the importance factor of the relation above it.

Marcu (1999) proposes that each internal node has a promotion set that is given by the salient units of that node. They are determined in a bottom-up fashion, as follows: the salient unit of a leaf is composed of itself; the salient units of an internal node is the union of the promotion sets of its nuclear children. The intuition behind this approach is that textual segments that are in the promotion sets of the top nodes of a discourse tree are more important in the text. The method attributes to the root of the tree a score corresponding to the number of levels in the tree and, then, traverses the tree toward the segment under evaluation: each time the segment is not in the promotion set of a node during the traversing, it has the score decreased by one.

There are also several variations of the above summarization methods (Uzêda, Pardo, and Nunes, 2010; Hachey and Grover, 2004; Seno and Rino, 2005; Cunha et al., 2009).

2.2 Multi-document summarization

Several works have dealt with the task of MDS with different strategies. In this work, some of these investigations will be reported, emphasizing those who made use of deep information and/or statistical techniques, which somehow correlate to our proposal in this paper.

One of the proposals that is the basis of the deep-based approaches is the one of Radev (2000), in which CST was introduced as a model for MDS. In this work, the author proposed a summarization method that made use of operators representing summarization preferences. These operators would walk over the CST graph, where nodes represent sentences and edges represent CST relations, and select the sentences that are connected by the relations that represent the summarization preferences. Examples of preferences are: information related to the context of the topic or information that indicates contradictions within the group of texts. The author only proposed this summarization method but did not perform any experiments and evaluations.

(Zhang, Goldenshon, and Radev, 2002) proposed the improvement of summaries produced by sentence ranking-based methods with the enrichment of the rank with CST information. Sentence ranking consists in ordering the sentences by their importance according to some criteria used by the summarization methods, which assign a score to each sentence. In these methods, the best scored sentences are selected to compose the summary. The authors suggested that the scores of sentences in the rank should be modified by adding, to the original score, the number of CST relations of the sentences. The authors applied this strategy on a rank-based summarizer in which sentences score was given by its lexical similarity with a centroid sentence, which was the title of the text. The authors projected a human evaluation method, in which sentences were given a score according to their relevance in the summary. Results of this evaluation showed that CST-based summaries had more relevant sentences than summaries produced with the original rank-based method.

For MDS of Brazilian Portuguese texts, there have been several works that use CST with good results. One of the first investigations in this line is the work of (Jorge and Pardo, 2010), who proposed CSTSumm, a summarizer based on preference operators, similar to the proposal of (Radev, 2000). According to the proposal, sentences were initially ranked according to their number of CST relations. After that, any operator could be applied in order to re-rank sentences, giving priority to the ones that satisfy some summarization preference. For instance, the authors developed 4 operators, based on preferences over contextual information, contradictory information, authorship information, and writing styles. Each operator was defined as a set of pre-defined rules for sentence ranking. After the rank was built and modified, the best ranked sentences were selected to compose the summary. The proposal was tested on the CSTNews corpus, and two types of evaluation were performed: ROUGE-based evaluation and human-based evaluation. According to ROUGE evaluation, the summaries produced from the initial rank outperformed the state of the art methods for MDS. According to human evaluation, summaries based on summarization preferences were considered good.

Another important work following this line is the research of (Ribaldo, 2013), who proposed RSumm, a graph-based method for MDS. According to the authors, graphs were built from a set of documents on the same topic, following the strategy of (Salton et al., 1997). The authors suggested a variation of Salton's graph, in which CST relations were also included. In this context, two strategies were proposed. In the first strategy, the number of CST relations were added to the already established edges. In the second strategy, a score was considered for each type of CST relation, giving different weights to CST relations. Sentences were selected following the same criteria proposed by (Salton et al., 1997). The methods were applied over the CSTNews corpus, and ROUGE values showed that the method outperformed other state of the art methods such as CSTSumm.

Other works that have used CST and other deep knowledge information for MDS are (Kumar et al., 2014), (Afantenos et al., 2004), (Barzilay and McKeown, 2005), (McKeown and Radev, 1995), and (Henning, Umbrath, and Wetzker, 2008), among others.

In recent years, there has been a growth of research that makes use of statistical models for MDS. For instance, (Daume and Marcu, 2006) proposed BayeSumm, which was a bayesian model of three components: word distribution in a topic, word distribution within a document, and word distribution corresponding to general language. The idea was to score sentences based on the probability of the sentence in each component. In that sense, sentences with higher probability of being related to the topic component had a better chance of composing a summary. The method was evaluated in DUC (Document Understanding Conferences) 2005 competition, obtaining promising results.

(Haghighi and Vanderwende, 2009) introduced TopicSum and HierSumm, two methods based on Statistical Topic Modeling (Blei, Ng, and Jordan, 2003). In TopicSumm, words were considered to be from one of three categories or topics: content (representing words referring to the content that is being talked about), background (general content) and document specific information. Based in these three categories, Statistical Topic Modeling was applied. In that sense, these categories or topics where defined as distributions over the words that compose the document sets. The hypothesis underlying this model was that a good summary should have its unigram distribution similar to the content category distribution, since those words described the main topic of the document set. For this aim, the summary was built progressively from the sentences of the source texts, so that the divergence among the unigram distribution and the content distribution was minimized. HierSumm was an adaptation of TopicSumm, where the content category was divided in order to consider two subcategories: general content and specific content, which would help to distinguish words that were more general or more specific even when being considered in the content category. For this aim, a hierarchical topic model was built, and the summary construction was performed similarly to TopicSumm. Both systems were evaluated using the DUC 2006 database, showing good results.

Some other works also used Statistical Topic Modeling for MDS, such as (Li and Sujian, 2013), among others.

3 *RST-based methods for MDS*

In this section, we describe two approaches in which RST is used for MDS. In the first approach, RST strategies are incorporated to other summarization techniques, which are rank-based techniques. In the second approach, RST information is used as features of a statistical model.

3.1 Enriching CSTSumm with RST

A way for enriching multi-document summaries produced by CSTSumm system (Jorge and Pardo, 2010) is to incorporate the information given by RST to one of its summarization strategies. We assume that the relevance of a sentence is influenced by its salience in its source text, which is given by RST, and its salience in the set of texts, given by CST. For this aim, each sentence is scored in the following way: the Marcu's method (Marcu, 1999) is applied in a sentence level, and the result value is normalized by the height of the correspondent RST tree, in order to obtain a final score value between 0 and 1 and avoid discrepancies; the salience of a sentence within its collection is defined by the number of CST relations it has. The final score of a sentence is the sum of its RST score and the number of CST relations it has, which constitutes a score representing the salience of the sentence for its source text and its collection. The more relevant a sentence is, the higher position it achieves in the rank. The best ranked sentences are included in the summary. After this process is performed, a redundancy treatment task is applied. The redundancy is controlled by means of CST relationships. For instance, if there is an EQUIVALENCE relation between two sentences (both have the same information content), only one must be selected to the summary. This method is referred by RC-4 in this paper (where R stands for RST and C for CST, and 4 indicates that it is the 4th variation we tested).

To illustrate RC-4, Figure 1 has two discourse trees representing two texts (D1 and D2); D1 is projected upside down for explanation purposes; each node is a sentence (numbered for reference) with its RST normalized score above/below it; dashed lines between texts are CST relationships. The symbols N and S indicate the nucleus and satellite of each rhetorical relation. By applying RC-4 strategy, a partial ordering on the importance of the sentences in the collection is organized as follows: D1_S1 > D2_S1 > D2_S3 > D1_S3 > {D1_S2, D2_S2} > D1_S4. We see, for instance, that the first sentence of document 1 is the best ranked sentence, followed by the fist sentence of the document 2.

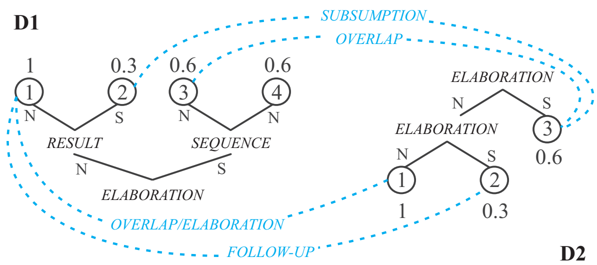

Figure 1: Example of RST and CST relationship for two texts

From the rank, there are two possibilities of content selection: only nuclear units of sentences or full sentences. After some experiments, We have decided to select full sentences, because it showed better results.

The size of the summary is limited by a compression rate, which, in this case, is considered to be 70 percent of the biggest text in the group, in terms of number of words.

3.2 RST and Statistical Modeling in MDS

MDS strategies may also be treated with statistical methods. Commonly, these methods aim to capture summarization patterns by estimating the likelihood of the occurrence of some features in human summary sentences. These features should represent strategical characteristics that indicate the salience of a sentence among a set of sentences. This salience may be modeled with RST. In particular, we assume that salience is indicated by the nucleus and satellite information of sentences. We rely on this premise because previous works strongly support this hypothesis. More accurately, we consider that the number of times a sentence has been annotated as nucleus or satellite may indicate a pattern of summarization that humans follow. In other words, humans may build multi-document summaries reflecting their preferences on sentences that appear more times as nuclei or satellites. The goal of our statistical model is to capture these patterns, by computing the likelihood of sentences being selected to compose a summary given their nucleus and satellite annotations in the source texts. An illustration of RST patterns occurring in multi-document summaries is given in Figure 2.

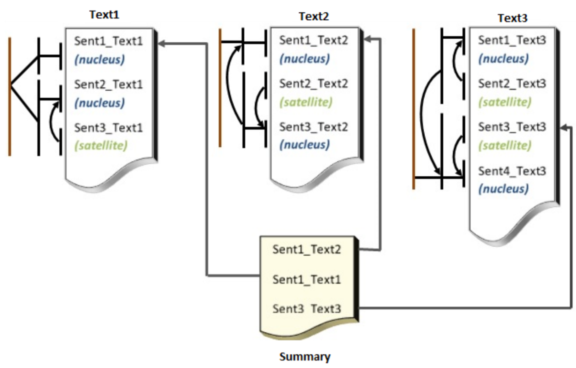

Figure 2: Illustration of RST patterns in human extracts

In this figure, it is observed that the sentences that compose the summary (the human extract, in this case) are annotated as nuclei or satellites in the original texts. For instance, two of the summary sentences are indicated as nuclei (once each one) and only one sentence is indicated as satellite (once). Actually, in this example, our probabilistic model would capture these tendencies by giving higher probability values to sentences that are annotated once as nuclei, since it is the pattern that appears more times in the human summary. In this context, it is relevant to highlight that emphasizing the number of times a sentence is annotated as nucleus or satellite is important, since in a real multi-document scenario the same sentence may be annotated with the same type (nucleus or satellite) more than one time.

Our probabilistic model is based on a generative learning approach, where the MDS task is formulated with the Bayes rule as stated in Equation 1:

$$\arg max_S P(S|C) = \frac{P(C|S) * P(S)}{P(C)} \quad (1)$$

According to this Equation, $\arg max_S$ $P(S|C)$ represents the summarization process, in which the search for the summary S given a collectin of texts C that maximizes the Equation is performed. In order to determine the value of P(S—C) that maximizes the Equation, the probabilities on the right side of the Equation have to be inferred first. Particularly, $P(C|S)$ represents the content model, which is formulated with RST features, as shown in Equation 2. The main goal of a content model is to infer the likelihood of occurrence of a RST feature in sentences of the original texts, given that these sentences also compose the human extracts. $P(S)$ represents a coherence model for multi-document summaries, in which summary coherence structural patterns are formulated. $P(C)$, on the other hand, is a model for a cluster of source texts. In this work, we will only focus in the formulation of $P(C|S)$, which is the content model using RST features. For the purposes of this work, $P(S)$ and $P(C)$ are assumed to have constant values. In fact, $P(C)$ might even be omitted.

$$P(C|S) = \prod_i (Satellite = N_{Sat}|SS_i) * P(Nucleus = N_{Nuc}|SS_i) \quad (2)$$

According to Equation 2, the probability of a summary is composed by the probability of its sentences. The probability of a summary sentence (SS) is given simply by the product of the probability of that sentence being annotated N_{Sat} times as Satellite and N_{Nuc} times as nucleus across a set of documents. The value of $P(Satellite = N_{Sat}|SS_i)$ is computed by dividing the number of summary sentences that have the correspondent pattern by the total number of summary sentences. The value of $P(Nucleus = N_{Nuc}|SS_i)$ is computed similarly.

This formulation is the basis of the training stage, in which sentence probabilities are estimated from parallel corpus of multi-document clusters and their corresponding multi-document human extracts. Once the probability values are inferred, the test stage is performed by building multi-document summaries through a decoding process. This process consists on searching for a subset of sentences, in a group of texts, that maximizes Equation 1, based on Equation 2. This subset must respect the limits imposed by the compression rate of the summary, which is the same used in the method described in Section 3.1. The algorithm used for decoding is the one proposed by (Aker, Cohn, and Gaiazauskas, 2010). We named this method MT-RST (which stands for Model of text-summary Transformation with RST).

4 Corpus

Our main resource is the CSTNews corpus[1] (Cardoso et al., 2011), composed of 50 clusters of news articles written in Brazilian Portuguese, collected from several sections of mainstream news agencies: Politics, Sports, World, Daily News, Money, and Science. The corpus contains 140 texts altogether, amounting to 2,088 sentences and 47,240 words. On average, the corpus conveys in each cluster 2.8 texts, 41.76 sentences and 944.8 words. Besides the original texts, each cluster conveys single document manual summaries and multi-document manual and automatic summaries.

The size of the summaries corresponds to 30% of the size of the biggest text in the cluster (considering that the size is given in terms of the number of words). All the texts in the corpus were manually annotated with RST and CST structures in a systematic way, with satisfactory annotation agreement values.

5 Evaluation

In this section, we present the results of the evaluation of the two methods proposed in this work. Both methods were developed over the CSTNews corpus. In the case of RC-4, summaries were produced for each of the 50 clusters of the corpus, by computing the corresponding scores for each sentence of each cluster. In the case of MT-RST, that requires training, a 10-fold cross validation schema was applied in order to produce the 50 multi-document summaries, corresponding to the 50 clusters of the corpus.

The summaries were evaluated using ROUGE (Lin, 2004), a standard evaluation metric used in text summarization. This metric produces scores that often correlate quite well with human judgements for ranking systems. ROUGE computes n-gram overlapping between a human reference and an automatic summary.

In this paper, as commented before, we used the human summaries in CSTSumm as reference summaries. We also adopted the manually annotated RST and CST relations in the corpus to run our methods. In the case of the MT-RST method, this same data was used for the cross-validation training/testing strategy.

To accurately measure the real effect of RST in MDS, we perform two types of comparison. The first comparison, presented in Table 1, involves RC-4 and two of the state of the art methods in MDS for Brazilian Portuguese texts: CSTSumm and RSumm. The second comparison, presented in Table 2, involves MT-RST and two of the state of the art methods for statistical MDS: TopicSumm and HierSumm, which were also reproduced over the CSTNews corpus.

The choice for comparing the two approaches separately is because statistical methods heavily depend on training and on the corpus size and its characteristics, and a direct comparison would not be fair, there-

[1] http://www2.icmc.usp.br/~taspardo/sucinto/cstnews.html

fore. In the case of CSTNews, it is a small corpus, and data may be sparse for automatic learning. This limitations do not happen in non-statistical learning methods.

For the first comparative evaluation, results are shown in Table 1; for the second, in Table 2. R stands for Recall, P for Precision, and F for F-measure, which are measures provided by ROUGE.

Table 1: ROUGE evaluation for non-statistical learning methods

Methods	**R**	**P**	**F**
RC-4	0.4374	0.4511	0.4419
RSumm	0.3517	0.5472	0.4190
CSTSumm	0.3537	0.4472	0.3864

The results in Table 1 show that RC-4 had a better performance than CSTSumm and RSumm in terms of F-measure, but RSumm outperformed it in terms of Precision (but showed a low Recall). This indicates that, considering the relevance of sentences within their correspondent source texts leads to the production of summaries with content closer to human summary content. The fact that RSumm performed better in terms of Precision and worse in terms of Recall may reveal that it tends to generate summaries with relevant content, but not selecting all the relevant content. On the other hand, a higher Recall value may indicate that RC-4 is more capable of retrieving different relevant information from the various sources.

Figures 3 and 4 show two automatic summaries (translated from the original language - Portuguese) produced by CSTSumm and RC-4 methods, respectively. The summaries contain news about the facts related to the floods that hit North Korea. It may be noticed that RC-4 introduces sentences that are more related to the central facts of the topic that is being narrated, while the summary produced by CSTSumm gives preference to contextual information. This example reveals the power of RST to capture the main or most salient information from a topic.

Table 2: ROUGE evaluation for statistical learning methods

Methods	**R**	**P**	**F**
MT-RST	0.3453	0.3534	0.3482
TopicSumm	0.2753	0.2991	0.2842
HierSumm	0.3302	0.3103	0.3190

According to the results shown in Table 2, MT-RST shows to be competitive with the other state of the art statistical methods. Although our statistical method did not widely outperform TopicSum and HierSumm, the ROUGE scores reveal that RST features tend to lead to better results. It is also important to point out that, besides it is not the aim of the comparison, MT-RST results are not too distant from CSTSumm and RSumm results. This possibly means that MT-RST performance could be improved if it was trained in a bigger corpus.

Figures 5 and 6 illustrate two automatic summaries (also translated from Portuguese) produced by HierSumm and MT-RST, respectively. These summaries were produced

[S1]According to the newspaper Choson Sinbo, published by the Association of Korean Residents in Japan (which is close to the communist regime in North Korea), the heavy rains that flooded much of this country in the second half of July caused much damage.

[S2]A total of 549 people died, 3,043 were injured and 295 were missing because of floods that recently affected North Korea today, said a North Korean newspaper published in Japan, citing official sources of Pyongyang.

[S3]A group of activists suggested that the number of dead or missing can reach 10 thousand, but it did not disclose where it obtained the information.

Figure 3: Summary produced by CSTSumm

[S1]At least 549 people were killed and 295 are still missing as a result of floods that hit North Korea in July, according to a pro-Pyongyang Japanese newspaper.

[S2]According to the newspaper Choson Sinbo, published by the Association of Korean Residents in Japan (which is close to the communist regime in North Korea), the heavy rains that flooded much of this country in the second half of July caused much damage.

[S3]North Korea has refused offers from international agencies to launch campaigns to help the country, but a local officer said last week that Pyongyang would accept aid from South Korea if it was given without conditions.

Figure 4: Summary produced by RC-4

for texts that discuss an explosion at a mall in Moscow. Both examples show poor coherence, since they do not include any sentence ordering or coherence model. Besides this limitation, an interesting characteristic may be observed. The summary produced by MT-RST presents more sentences related to the main topic. For instance, MT-RST summary includes relevant information related to prosecutor's declarations on the actions to be taken, witnesses informing about the attackers and ambulances attending the place. HierSumm summary, on the other hand, only informed about the explosion and the victims affected by the attack.

[S1] The explosion, supposedly produced from a gas cylinder, according to preliminary police versions, occurred in the "Evrazia" zone from the Cherkizov mall, one of the largest malls in the Russian capital.

[S2]Nine people were killed, including three children, and 25 others were injured this Monday in an explosion occurred at a mall in Moscow, reported the Moscow police.

[S3]This was not an accident, it was deliberate, Resin said, quoted by the Russian news agency "Itar-Tass".

Figure 5: Summary produced by HierSumm

[S1]Witnesses saw two strangers leaving a bag and running out of the cafeteria.

[S2]Nine people were killed, three of them children, and 25 others were injured this Monday in an explosion at a Moscow mall, police said.

[S3]Moscows' prosecutor announced the creation of a special group to investigate the accident.

[S4]Almost ten fire engines and more than a dozen ambulances were sent to the mall, which was isolated by the police.

Figure 6: Summary produced by MT-RST

6 Final remarks

In this paper we presented new methods for MDS using RST. We compared the performance of our methods with state of the art methods, and our results shows that the incorporation of RST information positively affects the production of summaries, resulting in more informative summaries.

To the best of our knowledge, this is the first time that the advantages of using RST in MDS are evidenced.

7 Acknowledgments

The authors are grateful to FAPESP and CAPES for supporting this work.

References

Afantenos, S.D., I. Doura, E. Kapellou, and V. Karkaletsis. 2004. Exploiting cross-document relations for multi-document evolving summarization. In *Proceedings of SETN*, pages 410–419.

Aker, A., T. Cohn, and R. Gaiazauskas. 2010. Multi-document summarization using a* search and discriminative training. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 482–491.

Barzilay, R. and K. McKeown. 2005. Sentence fusion for multidocument news summarization. *Journal of Computational Linguistics.*

Blei, M.D., A.Y. Ng, and M.I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Cardoso, P.C.F., E.G. Maziero, M.L.R. Castro Jorge, E.M.R. Seno, A. Di Felippo, L.H.M. Rino, M.G.V. Nunes, and T.A.S. Pardo. 2011. CSTNews - A discourse-annotated corpus for single and multi-document summarization of news texts in Brazilian Portuguese. In *the Proceedings of the 3rd RST Brazilian Meeting*, pages 88–105.

Cunha, I. Da, J.M. Torres-Moreno, P. Velázquez-Morales, and J. Vivaldi. 2009. Un algoritmo lingüístico-estadístico para resumen automático de textos especializados. *Linguamática*, 1(2):67–79.

Daume, H. and D. Marcu. 2006. Bayesian query-focused summarization. In *Proceedings 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics.*

Hachey, B. and C. Grover. 2004. A rhetorical status classifier for legal text summarisation. In *Proceedings of the ACL-2004 Text Summarization Branches Out Workshop.*

Haghighi, A. and L. Vanderwende. 2009. Exploring content models for multi-document summarization. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics- NAACL*, pages 362–370. NAACL.

Henning, L., W. Umbrath, and R. Wetzker. 2008. An ontology-based approach to text summarization. In *Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology - IEEE/WIC/ACM*, pages 291–294.

Jorge, M.L.R. Castro and T.A.S. Pardo. 2010. Experiments with CST-based multidocument summarization. In *Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing*, pages 74–82. Association for Computational Linguistics.

Kumar, Y.J., N. Salim, A. Abuobieda, and A.T. Albaham. 2014. Multi document summarization based on news components using fuzzy cross-document relations. *Applied Soft Computing*, 21:265–279.

Li, J. and L. Sujian. 2013. A novel feature-based bayesian model for query focused multi-document summarization. In *Transactions of the Association of Computational Linguistics*.

Lin, C-Y. 2004. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81.

Mani, I. 2001. *Automatic summarization.* John Benjamins Publishing.

Mann, W.C. and S.A. Thompson. 1987. Rhetorical Structure Theory: A theory of text organization. In *University of Southern California, Information Sciences Institute*, number ISI/RS-87-190.

Marcu, D. 1999. Discourse trees are good indicators of importance in text. *Advances in automatic text summarization*, pages 123–136.

McKeown, K. and D.R. Radev. 1995. Generating natural language summaries from multiple on-line sources. In *Proceedings of the 18th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pages 74–82.

O'Donnell, M. 1997. Variable-length on-line document generation. In *the Proceedings of the 6th European Workshop on Natural Language Generation, Gerhard-Mercator University, Duisburg, Germany.*

Ono, K., K. Sumita, and S. Miike. 1994. Abstract generation based on rhetorical structure extraction. In *Proceedings of the 15th conference on Computational linguistics-Volume 1*, pages 344–348. Association for Computational Linguistics.

Otterbacher, J.C., D.R. Radev, and A. Luo. 2002. Revisions that improve cohesion in multi-document summaries: a preliminary study. In *Proceedings of the ACL-02 Workshop on Automatic Summarization-Volume 4*, pages 27–36. Association for Computational Linguistics.

Pardo, T.A.S. 2005. GistSumm - GIST SUMMarizer: Extensões e novas funcionalidades. *Série de Relatórios do NILC*, page 8.

Radev, D.R. 2000. A common theory of information fusion from multiple text sources step one: cross-document structure. In *Proceedings of the 1st SIGdial workshop on Discourse and dialogue-Volume 10*, pages 74–83. Association for Computational Linguistics.

Ribaldo, R. 2013. *Investigação de mapas de relacionamento para Sumarização Multidocumento.* Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo.

Salton, G., A. Singhal, M. Mitra, and C. Buckley. 1997. Automatic Text Structuring and Summarization. *Information Processing & Management*, 33(2):193–207.

Seno, E.R.M. and L.H.M. Rino. 2005. Co-referential chaining for coherent summaries through rhetorical and linguistic modeling. In *the Proceedings of the RANLP 2005 Workshop on Crossing Barriers in Text Summarization Research*, pages 70–75. Citeseer.

Uzêda, V.R., T.A.S. Pardo, and M.G.V. Nunes. 2010. A comprehensive comparative evaluation of RST-based summarization methods. *ACM Transactions on Speech and Language Processing (TSLP)*, 6(4):4.

Zhang, Z., S.B. Goldenshon, and D.R. Radev. 2002. Towards cst-enhanced summarization. In *Proceedings of the 18th National Conference on Artificial Intelligence*.