

A Multi-document Summarization System for News Articles in Portuguese using Integer Linear Programming

Laerth Bruno de Brito Gomes¹, Hilário Tomaz Alves de Oliveira²

¹Centro Universitário de João Pessoa – Unipê
João Pessoa – PB – Brasil

²Instituto Federal do Espírito Santo
Serra – ES – Brasil

{laerthgomes, hilariotomaz}@gmail.com

Abstract. *Automatic Text Summarization (ATS) has been demanding intense research in recent years. Its importance is given the fact that ATS systems can aid in the processing of large amounts of textual documents. The ATS task aims to create a summary of one or more documents by extracting their most relevant information. Despite the existence of several works, researches involving the development of ATS systems for documents written in Brazilian Portuguese are still a few. In this paper, we propose a multi-document summarization system following a concept-based approach using Integer Linear Programming for the generation of summaries from news articles written in Portuguese. Experiments using the CSTNews corpus were performed to evaluate different aspects of the proposed system. The experimental results obtained regarding the ROUGE measures demonstrate that the developed system presents encouraging results, outperforming other works of the literature.*

1. Introduction

The constant expansion of the Web enables the creation and sharing of an enormous amount of information, especially in textual format, such as news articles, e-mails, scientific papers, social media posts, among others. Despite the evolution of web search engines, such as Google¹, these still return many results to be analyzed depending on the performed query. For instance, a simple query about the victory of the Brazilian soccer team in the 2019 Copa America returns hundreds of pages from different news sources reporting this event. Quickly identifying information of interest from this vast amount of information available, becomes a complex task in some scenarios.

In this context, Automatic Text Summarization (ATS) is one of the prominent research areas that aims to develop adequate solutions to reduce the time of users in identifying information of interest. ATS systems aim to create a summary containing the most relevant information from a single document or several related documents (multi-document) [Nenkova and McKeown 2012]. In this way, a user can first read the summary and then decide whether the document is relevant or not for their interests.

ATS approaches can be classified into Extractive or Abstractive [Nenkova and McKeown 2012]. Extractive-based approaches focus on identify and

¹<https://www.google.com/>

select the most relevant sentences from one or more documents, and use them without any change to create the output summary. On the other hand, abstractive-based systems try to simulate the way humans produce summaries. This kind of system requires a more in-depth analysis of the text to identify and select the most relevant information from the input documents, and then how to express them in a new way. Thereunto, it is common to use techniques for sentence compression and fusion [Linhares Pontes et al. 2018], natural language generation [Khan et al. 2015], among others.

Although abstract-based approaches potentially generate summaries more close to those written by human beings, they require a deep understanding of the text to be summarized, which demands a high computational cost. Recently, it has been increased the number of researches focusing on abstractive summarization, especially using deep learning [Li et al. 2018, Gupta and Gupta 2018]. However, extractive approaches to summarization are still widely explored. Besides, extractive summaries can be used as an intermediate representation to facilitate the generation of abstractive summaries.

Several works have proposed and evaluated different types of extractive summarization methods [Nenkova and McKeown 2012, Gambhir and Gupta 2016]. Among these, concept-based approaches using Integer Linear Programming (ILP) have been presenting state-of-the-art performance [Gillick et al. 2009, de Oliveira et al. 2018]. This kind of approach guides the summarization process as a maximum coverage problem, aiming to select the sentences of the input document(s) that maximize the coverage of relevant concepts, respecting the maximum size of the summary desired. There is a significant number of researches in ATS concerning documents written in several languages, especially for English. However, applications for summarizing texts in Brazilian Portuguese are still scarce [Pardo 2005, Castro Jorge and Pardo 2010, Cardoso and Pardo 2016, Tohalino and Amancio 2017].

In this work, we present a concept-based system using ILP for extractive Multi-document Summarization (MDS) of news articles written in Brazilian Portuguese. The proposed system uses bigrams as concepts and employs traditional statistical methods to identify the most relevant information to compose the output summary. Experiments were performed to evaluate the performance of the proposed system, using the CSTNews corpus [Dias et al. 2014]. The Recall-Oriented Understudy for Gisting Evaluation (ROUGE) [Lin 2004] measures were adopted to evaluate the generated summaries in comparison with human-made references summaries. The experimental results obtained demonstrate that the proposed system was able to correctly identify more than 60.00% of the information presented in the set of reference human-made summaries regarding the ROUGE-1 recall measure. Besides, the developed system outperforms other ATS systems for Portuguese news articles based on ROUGE measures.

2. Related Work

Several approaches addressing the extractive MDS can be found in the literature, especially for documents written in English. The first MDS system for Brazilian Portuguese was the GistSumm [Pardo 2005]. This system appends all input documents as if they were a single text. Afterward, each sentence receives a score that can be generated based on the frequency of its words or using the Term-Frequency - Inverse Sentence Frequency (TF-ISF) method. The highest score is selected as the gist sentence. Finally, it selects

the phrases that contain at least one word in common with the gist sentence and have a score higher than a given threshold, which is the average of all sentences scores. Aspects concerning redundancy and sentence order were not considered.

Castro Jorge and Pardo [Castro Jorge and Pardo 2010] introduced an MDS approach that explores semantic relations and content selection methods based on the Cross-document Structure Theory (CST) model. The proposed system, called CSTSumm (CST-based SUMMarizer), address some of the main problems of the multi-document task such as redundancy. Cardoso and Pardo [Cardoso and Pardo 2016] followed a similar research line to create MDS methods exploring CST and Rhetorical Structure Theory (RST) models. Both models were adopted to identify relevant information in the input group of texts.

Extractive MDS approaches have extensively explored graph-based representation. Tohalino and Amancio [Tohalino and Amancio 2017] proposed an MDS method using complex networks concepts. The proposed approach represents the textual document as a graph, where the sentences are the vertices, and the edges are created between the sentences based on the number of common words. Traditional and dynamical measurements of complex networks were applied to identify the relevant terms of the documents.

Recently, concept-based approaches using Integer Linear Programming (ILP) have achieved state-of-the-art performance for extractive MDS task [Gillick et al. 2009, de Oliveira et al. 2018], especially for documents written in English. Such an approach represents the summarization process as a coverage optimization problem, whose purpose is to extract the sentences that have the highest number of relevant textual fragments (called concepts), considering the maximum size of the desired summary. The present work extends the system developed by [Gillick et al. 2009] by exploring different concept scoring methods and representation forms. The developed system is an adaptation of the approach proposed by [de Oliveira et al. 2018] to the context of articles in Portuguese.

3. Proposed System

The main goal of the extractive MDS task is to select the most relevant sentences from a set of input documents and use them to generate a summary. Based on this aim, three essential aspects that should be considered in the development of an MDS system are:

- **Informativeness:** The generated summary should contain the most relevant information of the original documents.
- **Redundancy:** The summary should not have multiple sentences with a high overlap of information, which is a waste of space. Considering that the input documents are related to the same event, topic, or subject, the redundancy of information among them is very high. One of the biggest challenges of an MDS system is how to avoid redundancy in the generated summary.
- **Size:** The generated summary should not be more extensive than half the size of the original documents. In general, in the ATS literature on news articles domains, the researches adopt the threshold of approximately one hundred words, but this is an input parameter that can be adjusted by the user.

The proposed MDS system seeks to take into consideration the three aspects aforementioned by adopting a concept-based approach using Integer Linear Programming (ILP) [Gillick et al. 2009, de Oliveira et al. 2018]. This kind of approaches treats

the summarization process as an optimization problem, trying to select the sentences that maximize the inclusion of relevant concepts in the generated summaries, respecting the maximum size of the output summary. Besides, because it aims to maximize concept coverage, the system will only benefit from the inclusion of a concept once, so sentences with many concepts already present in the summary are penalized.

The proposed system consists of four main steps, as shown in Figure 1². Given a collection of news articles, the first step is to pre-process these textual documents to structure them into an appropriate format. Afterward, textual fragments are extracted as possible representations for the notion of concepts. After extracting the concepts, it is necessary to consider the relevance of each extracted concept using a scoring value. Finally, using the extracted concepts and their respective weights, the selection of the subset of sentences that maximize the insertion of relevant concepts in the generated summary is carried out.

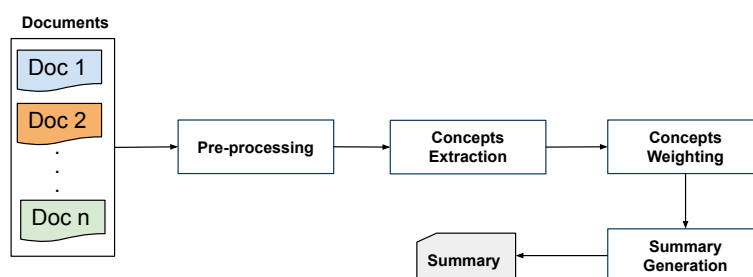


Figure 1. Overview of the summarization steps employed.

The four steps of the proposed system are described in the following sections.

3.1. Pre-processing

The input documents are written in natural language, i.e., they are in an unstructured format. This type of format requires prior structuring so it can be used for the application of the summarization algorithms. Therefore, this first step is responsible for applying Natural Language Processing (NLP) techniques to structure the input documents.

The proposed system was developed using the Java programming language³. The Cogroo tool⁴ was employed to perform the following NLP tasks: sentences segmentation, tokenization, stemming⁵, Part-of-Speech (POS) tagging and stopwords removal. In Table 1 is shown the results of the application of the pre-processing step in the following text fragment written in Brazilian Portuguese extracted from Wikipedia: “João Pessoa é a capital do estado da Paraíba. Sua população é estimada em 800.323 habitantes.”

At the end of this step, each input document consists of a list of sentences, and each sentence has its words with the associated linguistic information. In the next step,

²The source code and the summaries generated in the experiments performed are publicly available at https://github.com/hilarioliveira/SistemaPLI_PT-BR

³<https://www.java.com>

⁴<http://cogroo.sourceforge.net/>

⁵The stemming process was performed using the Porter algorithm available in the Lucene library (<https://lucene.apache.org/>).

Table 1. An example of a preprocessed text using the Cogroo and Lucene tools applied to a fragment of a Wikipedia article.

<i>s</i> ₁ : João Pessoa é a capital do estado da Paraíba.			
Token	POS tag	Stem	Is a stopword?
João	Proper Noun	joao	No
Pessoa	Proper Noun	peessoa	No
é	Verb	e	Yes
a	Article	a	Yes
capital	Noun	capit	No
do	Preposition	do	Yes
estado	Noun	estado	No
da	Preposition	da	Yes
Paraíba	Proper Noun	paraiba	No
<i>s</i> ₂ : Sua população é estimada em 800.323 habitantes.			
Sua	Possessive pronoun	sua	Yes
população	Noun	populacao	No
é	Verb	e	Yes
estimada	Verb	estimada	No
em	Preposition	em	Yes
800.323	Number	800.323	No
habitantes	Noun	habitant	No

from the structured information generated in this step, textual fragments are extracted to represent the notion of concepts.

3.2. Concepts Extraction

Selecting a representation form for the notion of concepts is an essential issue for a concept-based approach since it directly impacts the generated summary. Traditionally, several works in the literature have used textual fragments, such as unigrams (tokens) or bigrams as concepts. In this step, the proposed system extracts bigrams as concepts because of the good results presented in previous works on literature [Gillick et al. 2009, de Oliveira et al. 2018], and also in the performed experiments described in Section 4.2. Besides, the proposed system also supports the other two concept representation forms using unigrams and a representation that merges both (unigrams and bigrams). As suggested by [Gillick et al. 2009], concepts formed only by stopwords or that include punctuation symbols are removed.

To illustrate the execution of this step, in Table 2 is presented the concepts extracted from the sentences shown in Table 1. It is important to note that the tokens are represented by their stems extracted by the stemming process.

After the extraction process, each sentence of the input documents contains a collection of concepts that can be represented by one of three supported representation forms.

3.3. Concepts Weighting

Another fundamental step of a concept-based approach is the definition of a method for weighing the relevance of the extracted concepts. The score generated should reflect on how important the concept is to justify its inclusion in the summary to be generated. Traditional indicators of relevant content have been investigated in the literature individually

Table 2. List of concepts extracted from the sentences (separated by semicolon) from Table 1.

Sentences	Unigrams	Bigrams	Unigrams and Bigrams
s_1	joao;peessoa; capit;estado; paraiba	joao pessoa;peessoa e; a capit;capit do; do estado;estado da; da paraiba	joao;peessoa;capit; estado;paraiba joao pessoa;peessoa e; a capit;capit do; do estado;estado da; da paraiba
s_2	populacao; estimada; 800.323; habitant	sua populacao;populacao e; e estimada;estimada em; em 800.323;800.323 habitant	populacao;estimada; 800.323;habitant sua populacao;populacao e; e estimada;estimada em; em 800.323;800.323 habitant

or combined to measure the relevance of the concepts. In particular, the document frequency and sentence position methods have presented good results in the MDS task.

The developed system supports the application of three methods to consider the relevance of the extracted concepts: Document frequency, sentence position, and a combined method that considers both [de Oliveira et al. 2018]. These methods were chosen because they present high performance on several evaluations reported on literature [Gillick et al. 2009, de Oliveira et al. 2018]. Equations 1 and 2 show how the sentence position and the combined-method are respectively computed.

$$SentPos(c_i) = \sum_{d_j \in D} 1 - \frac{Index_{s_{c_i}}}{S_{d_j}} \quad (1)$$

$$Weight(c_i) = SentPos(c_i) \times DocFreq(c_i) \quad (2)$$

in which,

- $SentPos(c_i)$ returns the sentence position score of the concept c_i ;
- $Index_{s_{c_i}}$ in each document $d_i \in D$, returns the index of the first sentence that contains the concept c_i ⁶;
- S_{d_j} is the total number of sentences in the document d_j ;
- D is the input collection of documents;
- $DocFreq(c_i)$ returns the total number of documents in D that contains the c_i .

At the end of this step, each sentence contains its list of extracted concepts with their respective weights.

3.4. Summary Generation

This last step is responsible for selecting the sentences of the input documents to compose the summary to be generated. Thereupon, the proposed system follows a concept-based model using ILP to solve it as done in some previous works [Gillick et al. 2009, de Oliveira et al. 2018]. In this kind of approach, the summary generation consists of

⁶The index of the sentences starts by 0.

selecting the subset of sentences that maximizes the insertion of relevant concepts in the summary, respecting the maximum size of the expected summary. The adopted ILP model is presented in Equation 3.

$$\text{Maximize } \sum_i w_i \times c_i \quad (3a)$$

$$\text{s.t. } \sum_j l_j \times s_j \leq L \quad (3b)$$

$$s_j \times Occ_{ij} \leq c_i \quad \forall i, j \quad (3c)$$

$$\sum_j s_j \times Occ_{ij} \geq c_i \quad \forall i, j \quad (3d)$$

$$c_i, s_j, Occ_{ij} \in \{0, 1\} \quad \forall i, j \quad (3e)$$

In Equation 3a, the variable w_i represents the weight (relevance) of the concept c_i . The variable l_i is the size (number of tokens) of the sentence s_j and L is the maximum size of the summary to be produced. The binary variable Occ_{ij} indicates that the concept c_i is presented in the sentence s_j . The Equation 3c and Equation 3d are constraints that ensure the consistency of the model. Thus, if a sentence is chosen, all its concepts are as well, and a concept is elected only if it is present in a selected sentence.

The ILP model goal is to maximize the objective function $\sum_i w_i \times c_i$, guaranteeing informativeness of the output summary. Thus, it selects the sentences that have the highest amount of concepts c_i with a high weight w_i . The redundancy problem is also indirectly minimized since a new sentence is only selected if it adds new concepts to the generated summary. Besides, we use the heuristic of inserting a new sentence in the summary only if it does not have a cosine similarity higher than 0.5 with no other sentence already selected [Hong et al. 2014].

We use the GNU Linear Programming Kit (GLPK)⁷ to perform the sentence selection task modeled using the ILP model. After solving the ILP model, the GLPK toolkit indicates with the value 1 the sentences that were selected to compose the summary.

Finally, it is necessary to define the order of appearance of the sentences in the generated summary. It is important to emphasize that in the multi-document summarization task, the sentences from different documents can be selected. The sentences ordering problem in the multi-document scenario is a complex task [Ouyang et al. 2013]. In this work, we use the simple strategy of ordering the sentences based on their appearance in the document to which it belongs using an index that is assigned to each phrase during the pre-processing step. However, two sentences with the same index can be selected, for example, the first sentence of two different documents. To solve this conflict, we used the document index, generated in the pre-processing step, to which the sentence belongs.

4. Experiments

In this section, we present the experiments performed to evaluate different aspects of the proposed system. The experiments attempt to analyze the following issues: (i) What is the best concept representation strategy? (ii) What is the best concept weighing method?

⁷<https://www.gnu.org/software/glpk/>

and (iii) Presenting a comparison of the best system configuration with other systems available in the literature.

4.1. Experimental setup

The ATS systems evaluation is guided by the analysis of the informativeness and the linguistic quality of the generated summary. The informativeness represents the amount of relevant information that the summary has. On the other hand, linguistic quality is related to aspects such as cohesion and coherence of the text produced. Maximizing both characteristics is impractical in extractive approaches since no change in the original sentences of the documents can be performed. For this reason, this work is limited to analyzing the informativeness of the summaries generated.

The evaluation of the informativeness of automatically generated summaries can be performed by comparing them with manually created summaries. Thus, the presence of the input document(s) and their respective reference summaries are necessary. For the Brazilian Portuguese, the CSTNews [Dias et al. 2014] is the most adopted corpus in the literature to evaluate ATS systems. This corpus is composed of 50 collections of news articles (clusters), with each collection having 2 or 3 journalistic texts from different sources dealing with the same subject. Each collection has 5 extractive reference summaries. In Table 3 is presented some basic statistics of the CSTNews corpus generated using the Cogroo toolkit.

Table 3. Basic statistics of the CSTNews corpus.

Total of clusters	50
Total of documents	140
Average of documents (Standard deviation)	2.8 (0.4)
Total of sentences	2,279
Average of sentences per document (Standard deviation)	16.28 (8.81)
Total of tokens	52,011
Average of tokens per document (Standard deviation)	371.51 (174.73)

Evaluation Measure: The measures of recall, precision, and F-measure computed of the ROUGE-1 variation [Lin 2004] were used to evaluate the summaries produced. This variation computes the overlap of unigrams between automatically summary and the reference summary. This measure was chosen because it has been shown a high correlation with human evaluations in previous investigations [Lin 2004, Owczarzak and Dang 2011]. As each collection of CSTNews corpus documents has more than one reference summary, the final result presented in the experiments represents the average of the automatic summary compared with each of the reference summary.

The ROUGE version 1.5.5 was adopted setting with the parameters: $-n 1 -l 100 -f A$. The first parameter $-n1$ indicates to compute only the ROUGE-1 variation. The second parameter $-l100$ designates that the ROUGE evaluation considerate only the first 100 words of the generated summaries. This adjustment is essential so that all evaluated summaries have the same size. The last parameter $-fA$ defines that the final score will be the arithmetic average of the individual assessment with each of the reference summaries.

Summary length: The summaries were generated with a maximum size of 110 words. This threshold was chosen to generate summaries with sizes compatible with the other works of literature considered, which have summaries with around 100 words.

4.2. Evaluation of the Concepts Representation and Weighting Methods

This first experiment aims to evaluate which of the concept representation form and weighting method supported by the developed system present the best results concerning the ROUGE-1 measures. Thus, each representation form (bigram, unigram, combining unigram and bigram) was used in association with each of the scoring method (document frequency, sentence position, and the combined method). In Table 4 is presented the results obtained in this experiment based on the ROUGE-1 (R-1) recall, precision, and F-measure. The configuration (concept representation and weighting method) with the top performance in each evaluation measure is highlighted in bold.

Table 4. ROUGE-1 scores (%) and standard deviation (in parentheses) of the concept representations and weighting methods evaluation. The overall highest performance for each ROUGE measure is marked in bold, and statistically equivalent results ($p \geq 0.05$) from it are identified using a †.

Representation	Weighting Method	Recall (%)	Precision (%)	F-measure (%)
1. Bigrams	Combined-method	61.30 (10.37)	57.72 (8.62)	59.45
	Document frequency	55.74 (10.35)	51.85 (6.83)	53.73
	Sentence position	63.53 (8.32)	59.61 (7.67)	61.51
2. Unigrams	Combined-method	61.50 (10.13)	59.07† (9.29)	60.26
	Document frequency	54.67 (11.12)	51.68 (9.51)	53.13
	Sentence position	61.26 (9.65)	58.34† (9.50)	59.76
3. Bigrams + Unigrams	Combined-method	61.86 (8.60)	58.73† (8.33)	60.25
	Document frequency	55.73 (9.94)	52.11 (8.20)	53.86
	Sentence position	62.12† (9.53)	58.41† (8.80)	60.21

A first point to note is that all configurations obtained values higher than 50% in all R-1 recall measures. These results suggest that more than half of the information present in the reference summaries was included in the automatically generated summaries.

Concerning the concept representation forms, 9 comparisons were executed (3 configurations \times 3 evaluation measures). Adopting bigrams as concepts lead to better results in 4 of the 9 comparisons (44.44%), the combined representation performed better in 3 comparisons (33.33%) and using unigrams lead to the best performance 2 in two scenarios (22.23%). Regarding the concept weighting methods, the sentence position using bigrams presented the overall best result in all evaluations measures, and in many scenarios with a statistically significant improvement. Therefore, the configuration using bigrams and the sentence position method was chosen as the most efficient.

4.3. Comparison with other works

In this second experiment, we compare the results obtained by the proposed system with three other systems: GistSumm [Pardo 2005], CSTSumm [Castro Jorge and Pardo 2010], and RC-4 [Cardoso and Pardo 2016]. These systems were chosen because they used the same corpus adopted and achieved state-of-the-art results. The summaries of the CSTSumm were generated using the system available at⁸ and the summaries of GistSumm and RC-4 were made available by Paula Cardoso⁹ [Cardoso and Pardo 2016]. Besides, we use as a baseline the strategy of selecting the first n sentences of each document in the

⁸<http://conteudo.icmc.usp.br/people/taspardo/sucinto/files/CSTSumm.rar>

⁹The authors would like to thanks Paula Cardoso for having made the summaries available.

collection to compose the summary. This baseline is traditionally used in ATS and has presented competitive results with various systems. In Table 5 are presented the experimental results obtained in terms of the R-1 measures. The best system in each evaluation measure is highlighted in bold.

Table 5. Experimental results (%) and standard deviation (in parentheses) of the comparison with other works. The overall best result for each ROUGE-1 measure is marked in bold, and results that are statistically indistinguishable ($p \geq 0.05$) from it are identified using a †.

Systems	Recall (%)	Precision (%)	F-measure (%)
Baseline	61.46† (8.10)	57.67† (7.60)	59.50
CSTSumm	53.94 (11.42)	55.97 (9.10)	54.93
GistSumm	57.64 (11.21)	54.71 (9.31)	56.14
RC-4	59.10 (6.93)	60.18 (9.24)	59.64
Proposed System	63.53 (8.32)	59.61† (7.67)	61.51

The proposed system presented the top performance regarding the R-1 recall and F-measure in comparison with the three other systems and the baseline evaluated. In many scenarios, these improvements were statistically significant at 95% of confidence level. The RC-4 [Cardoso and Pardo 2016] obtained the highest results in the R-1 precision, but this improvement was considered statistically equivalent to the proposed system. The experimental results indicate that the developed system was able to produce more informative summaries in comparison to the other three analyzed systems and the baseline, without decrease the R-1 precision. Such performance suggests that a reasonable amount of information present in the generated summaries was also present in the reference summaries of the CSTNews corpus. The proposed system also generated summaries with smaller sizes than those generated by the other evaluated systems.

The following is presented the summary created by the proposed system for the group of documents $C26$ of the CSTNews corpus, composed of three news articles about a hurricane that happened in Mexico:

“O furacão Dean mudou de rumo e não atingiu em cheio as ilhas Cayman nesta segunda-feira, e continua a avançar em direção à península de Yucatán, no México, depois de causar fortes chuvas e ventos na Jamaica e no Caribe, onde forçou a saída de milhares de turistas. Dean se transformou em um ‘extremamente perigoso’ furacão de categoria 4, a segunda mais alta da escala Saffir-Simpson. O centro de Dean passou cerca de 160 km ao sul das ilhas Cayman, causando ventos de até 92 km/h. O furacão matou ao menos nove pessoas em sua passagem pelas ilhas do Caribe.”

5. Conclusions and Future Work

In this paper, we investigated the MDS task, which aims to generate a summary containing the most relevant information from a collection of related textual documents. Even after more than sixty years of its creation, the ATS is still an intense field of research, mainly due to the constant growth and availability of textual documents.

This work endeavored to contribute to the MDS task of news articles written Brazilian Portuguese by developing and evaluating a concept-based ATS system using

ILP. The proposed system treats the summarization task as a problem of maximum coverage, seeking to select the subset of sentences that maximizes the insertion of relevant concepts in the generated summary, respecting the maximum size of the desired summary. Three representation forms to the notion of concepts were investigated: unigrams, bigrams, and both (unigrams and bigrams). Besides, three concept weighting methods were analyzed to measure the importance of the extracted concepts: Document frequency [Gillick et al. 2009], Sentence position [de Oliveira et al. 2018], and a combined version of these two methods.

Different aspects of the proposed system were evaluated using the CSTNews corpus [Dias et al. 2014]. The experimental results showed that adopting bigrams as concepts obtained better results compared to the other representation forms. The comparison between the concept weighting methods highlighted that the sentence position obtained the best results. Finally, a comparison was performed among the proposed system with three systems of the literature and the traditional baseline of selecting the first n sentences to compose the summary. The results obtained showed that the proposed system presented competitive results with the compared systems, presenting superior results concerning the ROUGE-1 recall and f-measure.

A limitation of the proposed system is its extractive nature, i.e., no changes in the sentences selected to compose the summary were done. Based on this, as further work, we intend to make the following improvements: (i) Inclusion of sentence clustering techniques to identify sentences with a high overlap of information, i.e., sentences that share information with other phrases in the input collection of documents; (ii) Adapt the proposed system towards abstractive summarization, including techniques such as sentence fusion and compression [Linhares Pontes et al. 2018].

References

- Cardoso, P. C. and Pardo, T. A. (2016). Multi-document summarization using semantic discourse models. *Procesamiento del Lenguaje Natural*, (56):57–64.
- Castro Jorge, M. L. d. R. and Pardo, T. A. S. (2010). Experiments with cst-based multidocument summarization. In *Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing, TextGraphs-5*, pages 74–82, Stroudsburg, PA, USA. Association for Computational Linguistics.
- de Oliveira, H. T. A., Lins, R. D., Lima, R., Freitas, F., and Simske, S. J. (2018). A concept-based ILP approach for multi-document summarization exploring centrality and position. In *7th Brazilian Conference on Intelligent Systems, BRACIS 2018, São Paulo, Brazil, October 22-25, 2018*, pages 37–42.
- Dias, M. S., Garay, A. Y. B., Chuman, C., Barros, C. D., Maziero, E. G., Nobrega, F. A. A., Souza, J. W. C., Cabezudo, M. A. S., Delege, M., Jorge, M. L. R. C., Silva, N. L., Cardoso, P. C. F., Balage Filho, P. P., Condori, R. E. L., Marcasso, V., Felippo, A. d., Nunes, M. d. G. V., and Pardo, T. A. S. (2014). Enriquecendo o corpus cst-news: a criação de novos sumários multidocumento. In *International Conference on Computational Processing of the Portuguese Language - PROPOR*. SBC.
- Gambhir, M. and Gupta, V. (2016). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, pages 1–66.

- Gillick, D., Favre, B., Hakkani-Tür, D., Bohnet, B., Liu, Y., and Xie, S. (2009). The ICSI/UTD summarization system at TAC 2009. In *Proceedings of the Second Text Analysis Conference, TAC 2009, Gaithersburg, Maryland, USA, November 16-17, 2009*.
- Gupta, S. and Gupta, S. (2018). Abstractive summarization: An overview of the state of the art. *Expert Systems with Applications*, 121.
- Hong, K., Conroy, J. M., Favre, B., Kulesza, A., Lin, H., and Nenkova, A. (2014). A repository of state of the art and competitive baseline summaries for generic news summarization. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 26-31, 2014.*, pages 1608–1616.
- Khan, A., Salim, N., and Kumar, Y. J. (2015). A framework for multi-document abstractive summarization based on semantic role labelling. *Applied Soft Computing*, 30:737 – 747.
- Li, C., Xu, W., Li, S., and Gao, S. (2018). Guiding generation for abstractive text summarization based on key information guide network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 55–60, New Orleans, Louisiana. Association for Computational Linguistics.
- Lin, C.-Y. (2004). Rouge: A package for automatic evaluation of summaries. In Marie-Francine Moens, S. S., editor, *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Linhares Pontes, E., Huet, S., Gouveia da Silva, T., Linhares, A. c., and Torres-Moreno, J.-M. (2018). Multi-sentence compression with word vertex-labeled graphs and integer linear programming. In *Proceedings of the Twelfth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-12)*, pages 18–27, New Orleans, Louisiana, USA. Association for Computational Linguistics.
- Nenkova, A. and McKeown, K. (2012). A survey of text summarization techniques. In Aggarwal, C. C. and Zhai, C., editors, *Mining Text Data*, pages 43–76. Springer.
- Ouyang, Y., Li, W., Zhang, R., Li, S., and Lu, Q. (2013). A progressive sentence selection strategy for document summarization. *Information Processing & Management*, 49(1):213 – 221.
- Owczarzak, K. and Dang, H. T. (2011). Overview of the TAC 2011 summarization track: Guided task and AESOP task. In *Proceedings of the Text Analysis Conference (TAC 2011), Gaithersburg, Maryland, USA, November*.
- Pardo, T. A. S. (2005). Gistsumm-gist summarizer: Extensões e novas funcionalidades. *Série de Relatórios do NILC*.
- Tohalino, J. V. and Amancio, D. R. (2017). Extractive multi-document summarization using dynamical measurements of complex networks. In *2017 Brazilian Conference on Intelligent Systems (BRACIS)*, pages 366–371.