Carlos Rodrigo Cordeiro Garcia

Towards Coherent Single-Document Automatic Text Summarization: An Integer Linear Programming-based Approach

Recife

2016

Carlos Rodrigo Cordeiro Garcia

Towards Coherent Single-Document Automatic Text Summarization: An Integer Linear Programming-based Approach

Trabalho de Conclusão de Curso apresentado ao Programa de Bacharelado em Ciência da Computação do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Universidade Federal Rural de Pernambuco
Departamento de Estatística e Informática
Programa de Graduação

Supervisor: Rinaldo Lima

Recife
2016

MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO (UFRPE)
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

http://www.bcc.ufrpe.br

Ficha de Aprovação do Trabalho de Conclusão de Curso

Trabalho defendido por Carlos Rodrigo Cordeiro Garcia como requisito para conclusão do curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, intitulado **Towards Coherent Single-Document Automatic Text Summarization: An Integer Linear Programming-based Approach**, orientado por Rinaldo José de Lima e aprovado pela seguinte banca examinadora:

Rafael Ferreira

Rafael Ferreira Leite de Melo
DEINFO/UFRPE

George Gomes Cabral

George Gomes Cabral
DEINFO/UFRPE

Valmir Macário Filho

Valmir Macário Filho
DEINFO/UFRPE

This work is dedicated to children
who dreamed of becoming scientists

Acknowledgements

I would like to thank my father, Carlos Garcia, my mother, Jaciana Patrícia and my grandmother, Adeilda Maria for having invested everything in me and for teaching me to appreciate the science of learning.

I would like to express my gratitude and appreciation to all those who made this possible. My deepest gratitude goes first and foremost to my adviser, Professor Rinaldo Lima, for his untiring encouragements and guidance. I also would like to extend my sincerest thanks and appreciation to every single person who contributed to this work.

I would like to thank the entire faculty and staff in the Computer Science Department at Universidade Federal Rural de Pernambuco for having been part, directly or indirectly, of my education. And to my friends and family for supporting me throughout the years.

Lastly, I would like to thank my beloved girlfriend, Kristiny Brito, who has been my greatest source of inspiration and the brightest light in my darkest hours. And whose belief in my abilities encouraged me to overcome all obstacles and continue to reach for the best.

Those who can imagine anything, can create the impossible.

— ALAN TURING

Abstract

Due to the constant expansion of the Internet, the amount of data and textual documents has grown exponentially in recent years. The task of extracting information efficiently from such a huge source of information is therefore challenging. In this context, Automatic Text Summarization (ATS) appears as a viable solution to help reducing the time needed to find relevant information in textual documents. ATS can be defined as the process of creating a short version of one or more documents, but retaining the most important information of the original documents. The literature on ATS is vast and diverse, and recent approaches based on optimization methods have been gaining importance because of the good performance achieved. In this approach, the summarization process is defined as a coverage maximization problem, i.e., selecting the smallest subset of the sentences of one or more documents that maximizes the coverage of relevant information, while imposing a set of certain constraints. Most of the current approaches to ATS apply a single method of summarization for all input documents. In this paper, it is proposed an unsupervised method of extractive single-document summarization using Integer Linear Programming (ILP) and a Graph-based algorithm that takes into account both the importance and the local coherence of the sentences. The documents to be summarized are represented by a bipartite entity graph consisting of a set of nodes denoting both sentences and entities of the documents. After that, the sentences are ranked according to a proposed measure of relevance based on a weighing method of the concepts that are optimized using ILP. The proposed solution was implemented and evaluated on two well-known single-document benchmark datasets: DUC 2001 and 2002 using automatic evaluation metrics. In addition, it was conducted a preliminary human evaluation in order to assess the level of cohesiveness of the generated summaries. The proposed system achieved competitive results compared to other state-of-the-art systems on the above datasets. The results are very encouraging. Moreover, the proposed summarization system achieved the best results in terms of the coverage R-1 ROUGE measure on the DUC 2001 dataset.

Keywords: Automatic Text Summarization. Single-document Summarization. Integer Linear Programming. Natural Language Processing. Machine Learning.

Resumo

Devido à constante expansão da Internet, a quantidade de dados e documentos textuais tem crescido exponencialmente nos últimos anos. A tarefa de extrair informações de uma fonte tão grande de informações é, portanto, desafiadora. Neste contexto, Sumarização Automática de Texto (SAT) aparece como uma solução viável para ajudar a reduzir o tempo necessário para encontrar informações relevantes em documentos textuais. SAT pode ser definida como o processo de criação de uma versão curta de um ou mais documentos, mas mantendo as informações mais importantes dos documentos originais. A literatura da área de SAT é vasta e diversificada, contudo, recentes abordagens baseadas em métodos de otimização vêm ganhando importância devido ao bom desempenho obtido. Nesta abordagem, o processo de sumarização é tratado como um problema de maximização de cobertura, ou seja, selecionar o menor subconjunto de sentenças de um ou mais documentos que maximizem a cobertura de informações relevantes, respeitando um conjunto de restrições impostas. A maioria das abordagens atuais para SAT aplicam um único método de sumarização para todos os documentos de entrada. Neste trabalho, foi proposto um método não-supervisionado de sumarização monodocumento extrativa utilizando Programação Linear Inteira (PLI) e um algoritmo baseado em Grafos que considera tanto a importância como a coerência local das frases. Os documentos a serem resumidos foram representados por um grafo bipartido de entidades que consiste em um conjunto de nós que representam as sentenças e entidades dos documentos. Depois disso, as frases são classificadas de acordo com uma medida proposta de relevância com base em um método de pesagem dos conceitos que são otimizados usando PLI. A solução proposta foi implementada e avaliada em dois conjuntos de dados de referência de sumarização monodocumento bem conhecidos: DUC 2001 e 2002 usando métricas de avaliação automática. Além disso, uma avaliação humana preliminar foi realizada para avaliar o nível de coesão dos resumos gerados. O sistema proposto obteve resultados competitivos em comparação com outros sistemas do estado-da-arte nos conjuntos de dados citados acima. Os resultados são muito encorajadores. Além disso, o sistema de sumarização proposto obteve os melhores resultados em termos da medida de cobertura R-1 ROUGE no conjunto de dados DUC 2001.

Palavras-chave: Sumarização Automática de Texto. Sumarização Monodocumento. Programação Linear Inteira. Processamento de Linguagem Natural. Aprendizagem de Máquina.

List of Figures

List of Tables

List of abbreviations and acronyms

ATS	Automatic Text Summarization
NLP	Natural Language Processing
ILP	Integer Linear Programming
TF-ISF	Term Frequency - Inverse Sentence Frequency
TF-IDF	Term Frequency - Inverse Document Frequency
XML	eXtensible Markup Language
ROUGE	Recall-Oriented Understudy of Gisting Evaluation

Contents

1 Introduction

Due to the constant expansion of the Internet, the amount of data and textual documents has grown exponentially in recent years (NENKOVA; MCKEOWN, 2011). It also enables the creation, and access of a vast amount of digital information, especially textual documents such as news articles, blogs, e-mails, scientific articles, postings on social networks, among others. Despite constant improvements in the development of web search engines, identifying useful information from this huge amount of information is impractical if performed manually.

The task of extracting information efficiently from numerous sources of texts and documents available on a topic is infeasible. In this context, it is necessary to create efficient tools that can digest all this information, process, and summarize them in order to decrease the time spent on retrieving information from various sources.

While the amount of available information increases, the demand for methods and tools that can read and understand documents in order to extract valuable information, and that can represent all the most important aspects of the documents in a coherent representation also increases. A technique that is able to handle this demand for years is automatic text summarization (ATS) (GAMBHIR; GUPTA, 2016). ATS is the process of automatically creating a summary from a document or a collection of documents.

In this context, a summary can be defined as a text that is produced from one or more input document, that conveys the most important and relevant information from the original document(s). Tipically the original documents are shortened to half of its size or even less. (R.RADEV; MCKEOWN, 2002).

Summaries have been playing an important role for society mainly by helping readers decide whether certain content relates to their desired interests. Nowadays, with the evolution of technology and especially the growth of the Internet, the importance of summaries has become even greater. Nenkova and McKeown (NENKOVA; MCKEOWN, 2011) states that short summaries with at least a total length of 17% of the full original text length can speed up decision-making by almost a factor of two, with no significant statistically degradation in accuracy. A detailed discussion about the potential uses and applications of text summarization, and how many fields can take profit of it are presented on their study.

ATS can be regarded as the process of creating a short version of text from one or more documents, containing the most important information from the original documents. The generated summary may vary from a list of single words contained in the original document to a complete and coherent set of sentences, depending on the technique used

(R.RADEV; MCKEOWN, 2002). Traditionally, ATS can be classified into two major approaches: Extractive and abstractive. Extractives summaries are generated by selecting several sentences and concatenating the most relevant appearing in the document to be summarized, whereas abstractive summarization approaches attempt to produce a summary that convey the essential information of the documents using some aspects of which may not appear as part of the original. Abstractive summarization usually requires semantic analysis and abstract representation of texts, which requires knowledge beyond the contents presents in the documents.

There have been several proposals for the development of summarization techniques and applications. The extractive summarization approach that has been the most studied so far, mainly due to the following reasons:

- Produces grammatically correct summaries;
- It has better perfomance;
- Response time is lower;
- In general, it is the first phase of creating an abstractive summary.

ATS techniques may also be classified according to the number of documents that are analyzed simultaneously to obtain the summary, i.e., *single-document* or *multi-document*. Single-document techniques produces a final summary from a unique input document while multi-document techniques process various documents simultaneously about the same subject, topic or events and produces a summary taking into account all the input documents.

1.1 Motivation

It has passed about 50 years since the publication of the first article about automatic text summarization, and since then the need for efficient and accurate summarization tools is becoming increasingly evident (NENKOVA; MCKEOWN, 2011). Therefore, text summarization is still an important and relevant task nowadays.

Concerning the application of ATS, several scenarios can be mentioned. One example is in the medical field, where a summarization system may be used to summarize medical documents, helping doctors to make better decisions in a situation where it requires a faster response to an action (NENKOVA; MCKEOWN, 2011). Another application concerns news, in a way that given the amount of news articles on the web today, a summarization system that can identify the most relevant information contained in a set of documents so that one can speed up decision making and accessing to new knowledge quickly and effectively (DAS; MARTINS, 2007). Then, due to all the above reasons, ATS is regarded as important and challenging task in Text Mining.

1.2 Research problem

As mentioned earlier, the ATS field has been investigated for the past 50 years and the researchers are in accordance that a good summary must capture three major aspects, namely:

- **Relevance**: Summaries must convey relevant information in the original documents;
- **Redundancy**: Summaries should not contain multiple sentences that represent the same information;
- **Length**: Summaries should be no longer than half of the size of the original documents.

However, optimizing all these properties altogether is a challenging task and an example of a global inference problem (MCDONALD, 2007). This is due to the fact that including a new text unit to the summary does not depend solely on the properties of the new unit itself but also on the properties of all other text units that were previously added to the summary. Therefore, this makes the problem of optimizing such properties a NP-hard problem (BOUDIN; MOUGARD; FAVRE, 2015; KENDALL; PARKES; SPOERER, 2008), since for a number n of sentences, there are $(n-1)!$ different ways of organizing the sentences.

Typically, two approaches have been employed to solve this problem. The first one is to try to optimize relevance, redundancy and length separately, whereas the second is to treat the problem as a global inference and optimize all the criteria simultaneously (MCDONALD, 2007). Recently, global optimization methods, such as Integer Linear Programming (ILP), are very effective for this type of task. ILP can solve constrained optimization problems, where both the cost function and constraints are linear in a set of integer variables. Solving arbitrary ILPs is NP-hard. However, ILPs are a well studied optimization problem with efficient branch and bound algorithms for finding the optimal solution.

A recurrent problem in extractive text summarization is that the selected sentences composing the summary usually lack cohesiveness. The reason resides in the fact that the majority of techniques proposed only cares in selecting the sentences that adheres to the aspects cited above, neglecting other important aspects like coherence among the sentences in the summary.

As a result, most summaries produced by current extractive summarization systems are not capable to produce globally coherent summaries. Thus, when read by a human, the summary seems to contain loose sentences that seem to have no relationship among them.

Summaries should be readable, hence they should be grammatically correct and coherent. Grammaticality does not concern extractive summarization because the selected sentences extracted from input documents are assumed to be grammatically correct. However, little research has been done on adding coherence measures into extractive summarization systems. Currently, adding coherence in extractive summarization is one of the main open problems in ATS. The present work aims at mitigating this limitation while still incorporating relevant information in the summary.

Coherence, in linguistics, is what makes a text semantically meaningful, i.e. it's one of the fundamental building blocks of what logical information on text. Relevance, is how important or relevant a textual unit or a sentence is for the overall document.

1.3 Objectives

The main goal of this work is to apply Natural Language Processing techniques and ILP in an extractive approach to automatic text summarization using different sentence selection scoring metrics to be able to generate more coherent extractive summaries. The effectiveness of the proposed approach is demonstrated by several experiments on two benchmark datasets on extractive summarization. In addition, the proposed solution is compared with other state-of-the-art summarization systems reported in the literature.

Besides the main goal mentioned above, the following specific objectives are also achieved in this work:

1. Investigating state-of-the-art text summarization techniques;
2. Adapting a strategy for selecting sentences based on graph-based sentence representation;
3. Optimizing aspects of relevance, coverage and coherence in text summarization;
4. Evaluating the results obtained from this study and compare them to the state-of-the-art.

1.4 Methodology

Firstly, it was performed an in depth literature review of relevant and established articles and books in the ATS field. Fundamental definitions and concepts of text summarization were taken mainly from (NENKOVA; MCKEOWN, 2011) (R.RADEV;

MCKEOWN, 2002) (DAS; MARTINS, 2007) (GAMBHIR; GUPTA, 2016). These documents served as the basis to understanding ATS in a broader context, taking into account its relevance and the existing solutions.

The very first issue faced was how to represent the sentences. It was decided to implement the Entity Graph, an extension of the Entity-Grid, a method that uses a matrix to represent the concepts in a given sentence because it demonstrated good results in incorporating local sentence coherence to ATS (FILIPPOVA; STRUBE, 2007). In the Entity Graph approach, each node represents the sentences and the edges represent a given relation between them. Thus, the articles that served as the foundation for understanding these methods were (FILIPPOVA; STRUBE, 2007) (GUINAUDEAU; STRUBE, 2013) (BARZILAY; LAPATA, 2008).

Next, the focus was on selecting sentence scoring functions that allows the assessment of how relevant a sentence is to a given document. Many of the techniques tested were taken from (FERREIRA et al., 2013) (FERREIRA et al., 2014).

Then, other design decitions were taken, including the method of text summarization to be employed and how to represent sentences in natural language and estimate their relevance by means of sentence scoring methods. The last component of the system needed an unsupervised algorithm that could take all this information and automatically select the best sentences and incorporate them to the summary according to some configurable restrictions. Hence, it was selected Integer Linear Programing as the solution, which reported excellent results in a optimization problem, so it had to be configured the problem to feed into the ILP. The formulation of the optimization problem for text summarization were extracted mainly from (MCDONALD, 2007) (GILLICK; FAVRE, 2009) (BOUDIN; MOUGARD; FAVRE, 2015).

Based on the literature review, the best approaches to text summarization were listed and conceptually explored and then selected. Firstly, it was implemented a basic baseline for a solution including the design choices previously mentioned. After the summarization system was developed several experiments were performed to determine the best possible configuration of the system. Then, the system was evaluated in the benchmarking datasets selected to perform the experiments in order to compare the proposed solution with other state-of-the-art summarization systems.

1.5 Contributions

In the quest for the objectives proposed by the present work, the adopted methodology aims at contributing to the field of ATS by the following key contributions that can be summarized as follows:

- **Contribution 1**: The proposed sentence scoring ranking function that enables us to select important and locally cohesive sentences.
- **Contribution 2**: Experimental avaliation concerning the several sentence scoring techniques and combinations of them seeking to verify what are the best individual techniques and what is the best strategy;
- **Contribution 3**: The development of a extractive summarization system that integrates the ILP formulation that seeks to maximize relevant information, in addition to the entity graph model to estimate the local coherence of sentences, in order to select the most relevant coherent sentences.

1.6 Document structure

The remainder of this document is organized as follows:

Chapter 2 first presents the background information about all the concepts related to the problem of automatic summarization. In Chapter 3, it is provided a literature review on ATS where it is discussed existing solutions with respect to sentence representation, sentences ranking, feature selection, Natural Language Processing techniques, and Integer Liner Programming.

Chapter 4 presents the proposed approach to single-document summarization that was developed for generating summaries that take into account both the aspects of text coherence and the maximization of the number of concepts present in the summary.

The experimental setup consisting of datasets and evaluation metrics as well as the results and discussion are presented in Chapter 5.

Finally, concluding remarks are provided in Chapter 6.

2 Theoretical Foundation

The aim of this chapter is to provide a theoretical basis for understanding the present research work. This chapter also covers the fundamental ideas underlying the proposed approach to extractive summarization and is essential to better understand the remaining chapters.

2.1 Automatic Text Summarization

Text summarization is a task which involves Artificial Intelligence and Natural Language Processing in order to generate a summary of a set of textual documents. A summary can be defined as a text that is produced from one or more input documents, that conveys important information from the original documents covering most of its concepts, and that is no longer than half of the original texts and usually less than that (R.RADEV; MCKEOWN, 2002).

Over the years there have been several discussion and a few distinctions were made in summarization regarding the categories and types of summaries. However this document provides the terminology that is most accepted in the summarization literature. In the next subsection is discussed the details and explained the categories on the literature.

2.1.1 Types of Summaries

Regarding the number of documents that are analyzed simultaneously to generate the summary text summarization can be classified as **single-document** or **multi-document**. Single-document summaries must be created from a single input document whereas multi-document summaries the summary should be generated from multiple input documents that usually have related topic or subject (DAS; MARTINS, 2007). Both of these tasks have been develped over the years and each one of them have their own distinct challenges.

Regarding the approach to build a summary, it can be classified as **extractive** or **abstractive**. Extractive summaries (FERREIRA et al., 2013) are generated by selecting several sentences and concatenating the most relevant appearing in the document to be summarized. Abstractive (NENKOVA; MCKEOWN, 2011) summaries are generated to convey the essential information from the documents, it can reuse sentences or words in the original text, but the summary is expressed in the words of the algorithm itself and may contain phrases that were not present in the original document.

Summaries can either be **generic** or **query focused**. Generic summarization makes few assumptions about the audience or the goal for generating the summary (NENKOVA; MCKEOWN, 2011). Usually, it is assumed that any user can read the summary and help the reader quickly determine the topic of subject of the document. On the other hand, query focused summarization have a weel defined goal, which is to summarize only the information in the input documents that is relevant to a specific user query (GAMBHIR; GUPTA, 2016). Thereafter, a user may enter a query to a search engine and a summary of each document could be presented to the reader so that he can decide which document is relevant easier. Much of the research developed so far has been in the context of generic summarization (NENKOVA; MCKEOWN, 2011).

A summarization task may be developed either as supervised or unsupervised (GAMBHIR; GUPTA, 2016; NENKOVA; MCKEOWN, 2011). The key difference here is basically the same for other machine learning applications. For a supervised system trainning data with label classes is required to train a model that after it finished its trainning it can use the trained model to select important information from the documents. Unsupervised systems do not need any trainning data whatsoever, they use only the documents to be summarized. These systems tend to apply some heuristics to extract relevant information from the document to generate a summary.

Finally, summaries may be distinguished by their content, and they are known as indicative or informative summaries (GAMBHIR; GUPTA, 2016). Indicative summaries give the reader an indication of what the document is about, such as the topic of the summary or the ideas and important words of the documentt. Informative summaries can be read instead of the document, the idea is provide as much information as possible about the document in a shorter and elaborated version.

Table 1 presents all mentioned approaches to summarization organized by their classification.

Table 1: Approaches to text summarization.

Type	**Approaches**
Number of documents	Single-document Multi-document
Summarization method	Extractive Abstractive
Purpose	Generic Query-focused
Learning algorithm	Unsupervised Supervised
Content	Indicative Informative

Source: The author

2.2 Entity-grid

The entity-grid was introduced in (BARZILAY; LAPATA, 2008), a method for representing local coherence in a document that captures the distribution of discourse entities across sentences in a text. Furthermore, local coherence measures text relatedness at the level of transitions between sentences, and is essential for generating globally coherent text.

The entity-grid representation is as follows: Each text is represented as an entity-grid, which is a two-dimensional array that captures the distribution of entities across sentences. Entities can represent the basic unit of text in a document such as the words of each sentence. The rows of the grid correspond to the sentences, and the columns correspond to the entities, i.e. the grid is a two-dimensional array $Grid[m, n]$ where each row i is a sentence from the document and each column j is an entity from the document. Thus, for each occurrence of an entity j in the text, the corresponding grid cell $[i, j]$ contains information about the entity presence or absence in a sentence i. The value of a grid cell may also contain information about its syntactic role. This information can express many grammatic roles, but the most common is grammatical relation. Thus, each grid cell can be classified as whether a subject(**S**), object(**O**) or neither(**X**). Entities that are not present in a sentence are usually represented with a dash($-$). All grammatical role information can be extracted from a document using a dependency parser, such as the Stanford Natural Language Processing Toolkit (CoreNLP) tool (MANNING et al., 2014) which was used in this research. Figure 1 illustrates a fragment of an entity-grid constructed in (BARZILAY; LAPATA, 2008), while Figure 2 show the document used in this example.

Figure 1: Entity-grid representation of sentence rows and entities columns. Grid cells correspond to grammatical roles: subject(**S**), object(**O**) or neither(**X**)

.

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	S	O	S	X	O	–	–	–	–	–	–	–	–	–	–	1
2	–	–	O	–	–	X	S	O	–	–	–	–	–	–	–	2
3	–	–	S	O	–	–	–	–	S	O	O	–	–	–	–	3
4	–	–	S	–	–	–	–	–	–	–	–	S	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	S	O	–	5
6	–	X	S	–	–	–	–	–	–	–	–	–	–	–	O	6

Source: Barzilay e Lapata (2008, p. 143)

Figure 2: Summary augmented with syntactic annotations for grid computation

1 [The Justice Department]$_{\mathbf{S}}$ is conducting an [anti-trust trial]$_{\mathbf{O}}$ against [Microsoft Corp.]$_{\mathbf{X}}$ with [evidence]$_{\mathbf{X}}$ that [the company]$_{\mathbf{S}}$ is increasingly attempting to crush [competitors]$_{\mathbf{O}}$.
2 [Microsoft]$_{\mathbf{O}}$ is accused of trying to forcefully buy into [markets]$_{\mathbf{X}}$ where [its own products]$_{\mathbf{S}}$ are not competitive enough to unseat [established brands]$_{\mathbf{O}}$.
3 [The case]$_{\mathbf{S}}$ revolves around [evidence]$_{\mathbf{O}}$ of [Microsoft]$_{\mathbf{S}}$ aggressively pressuring [Netscape]$_{\mathbf{O}}$ into merging [browser software]$_{\mathbf{O}}$.
4 [Microsoft]$_{\mathbf{S}}$ claims [its tactics]$_{\mathbf{S}}$ are commonplace and good economically.
5 [The government]$_{\mathbf{S}}$ may fi le [a civil suit]$_{\mathbf{O}}$ ruling that [conspiracy]$_{\mathbf{S}}$ to curb [competition]$_{\mathbf{O}}$ through [collusion]$_{\mathbf{X}}$ is [a violation of the Sherman Act]$_{\mathbf{O}}$.
6 [Microsoft]$_{\mathbf{S}}$ continues to show [increased earnings]$_{\mathbf{O}}$ despite [the trial]$_{\mathbf{X}}$.

Source: Barzilay e Lapata (2008, p. 143)

(GUINAUDEAU; STRUBE, 2013) proposed a graph-based approach for local coherence modeling. This work is an extension of the entity-grid. They represent the sentences and entities in a graph and then model local coherence by applying centrality measures to the nodes in the graph

2.3 Integer Linear Programming

Integer Linear Programming techniques have been used in the past to solve many intractable inference problems in both IR and NLP. This includes applications to relation and entity classification, sentence compression, syntactic and semantic parsing, as well as many others (MCDONALD, 2007). An ILP is a constrained optimization solver, where the cost function and constraints are linear in a set of integer variables. Global inference is an NP-hard problem, but ILPs are well studied optimization solvers with efficient branch and bound algorithms for finding the optimal solution.

State-of-the-art approaches to extractive summarization are usually based on the idea of maximizing the coverage of the concepts to be included in the generated summary (PARVEEN; STRUBE, 2015). The assumption here is that a good summary is a selection of sentences from the document that contains as many of the important concepts as possible (SCHLUTER; SØGAARD, 2015). Although this problem has been reported as NP-hard, approximate solutions using ILP have been reported in the literature (MCDONALD, 2007).

The ILP method is formally presented below (LI; QIAN; LIU, 2013):

$$\text{Maximize: } \sum w_j c_j \tag{2.1}$$

$$s_j Occ_{ij} \leq c_i \tag{2.2}$$

$$\sum_j s_j Occ_{ij} \geq c_i \tag{2.3}$$

$$\sum_j l_j s_j \leq L \tag{2.4}$$

$$c_j \in \{0,1\} \forall_i \tag{2.5}$$

$$s_j \in \{0,1\} \forall_i \tag{2.6}$$

The objective function is described in (2.1). The idea behind this representation is to maximize the selection of the highest scored concepts $c_j \in C$ based on a weighting metric w, where C is the set of all concepts, i.e. words of the document. In other words, this function tries to maximize the amount of relevant sentences selected for the generated summary by selecting the sentences with the highest scored concepts.

c_j and s_j are binary values that indicate the presence of a concept and a sentence, respectively. The concept is a basic unit of text that can, for example, represent all nouns of the text. The sentences are the document sentences.

l_j is the length of a sentence and L is the maximum length of the generated summary. w_i is the weight of a concept and Occ_{ij} is the occurrence of a concept i in a sentence j. The inequities (2.2) (2.3) associate the sentences and concepts. This ensures that selecting a sentence leads to selection of all the concepts it contains, and selecting a concept only happens when it is present in at least one of the selected sentences.

The Inequality (2.4) ensures that the summary to be generated can not exceed the L value that defines the maximum length of the generated summary.

3 Literature Review

The purpose of this chapter is to review the current literature on Text Summarization giving a brief introduction to the main articles and contributions on the field over the years. For a more comprehensive detailed analysis of the articles discussed here and others the following surveys are suggested (NENKOVA; MCKEOWN, 2011; DAS; MARTINS, 2007; FERREIRA et al., 2013; GAMBHIR; GUPTA, 2016)

3.1 Automatic text summarization based on statistical techniques

Several studies have investigated the application of superficial techniques for scoring sentences for text summarization. These techniques are rather simple to be implemented, require little or no linguistic resource and have a low computational processing time. (FERREIRA et al., 2013) implemented fifteen scoring methods developed in the literature over the years. In order to select the most relevant sentences, word scoring, sentence scoring and graph scoring methods are used. In word scoring, scores are given the words and the words with highest socores are the important. Among word scoring methods are word frequency (LUHN, 1958), TF/IDF and lexical similarity (MURDOCK, 2007). In sentence scoring, scores are assigned to each individual sentence of the documents. Some of the sentence scoring methods are position of sentence (FATTAH; REN, 2009; BARRERA; VERMA, 2012) and sentence centrality (FATTAH; REN, 2009). In graph scoring, scores are calculated by modeling sentence relantionships as an edge between sentence node graphs. For graph scoring methods there are text rank (MIHALCEA; TARAU, 2004; BARRERA; VERMA, 2012), bushy path of the node and aggregate similarity (FATTAH; REN, 2009).

(CHAN, 2006) developed a quantitative model for creating summary that extracts sentences from highly relevant sections of the input documents. This approach limits itself to use only shallow linguistic extraction techniques while performing information extraction through sentence based abstraction technique. Cohesion and Coherence are the two quantitative coefficients to evaluate the amount of discourse continuity. Rhetorical Structure Theory (RST) is employed here to model coherence relation in the text. Coherence analysis depends on rhetorical relations.

Several methods to improve sentence extraction in text summarization using statistical techniques, such as word frequency, sentence position, sentence centrality, among others were studied in (FATTAH; REN, 2009). The proposed method of their study combines these techniques with genetic algorithms and mathematical regression in order to build a model with an appropriate set of feature weights. Feed Forward Neural Network and Probabilistic Neural Network are used for classification of sentences. Documents' sentences are assigned ranks in decreasing sequence of their scores and a highly scored collection of sentences are employed to produce the summary.

(MENDOZA et al., 2014) proposed an extractive generic summarization method for single documents by using generic operators and guided local search. This method uses a memetic algorithm which has combined the population based search of evolutionary algorithm with a guided local search strategy. The task of summarization is treated as a binary optimization problem. Few domain and language independent features are used for searching the important sentences from the documents like position of sentence, resemblance of sentence with the title, sentence length, cohesion and coverage.

3.2 Automatic text summarization based on graphs

In a graph based approach, text entities, i.e. words or sentences, are represented by nodes on the graph and the edges usually connect the related text entities together by some semantic relationship. (ERKAN; RADEV, 2004) proposed LexRank which is a summarization system for multiple documents where the sentences from the input document are modeled as nodes in a graph. If similarity among two sentences exceeds a threshold a connection is made between them in the graph by connecting the two nodes by an edge. After the graph construction is completed, important sentences are selected by the system by a random walk performed on the graph.

(BARALIS et al., 2013) proposed GRAPHSUM, a graph-based, general purpose summarizer for multi-document summarization. This approach explores and employs a machine learning technique (association rules) for discovering correlations among multiple terms and does not depend on the advanced models based on semantics. After preprocessing, document collection is arranged as a transactional dataset so that association rule mining can be performed on them. Then, frequently occurring itemsets which have high correlations among the terms are extracted from the transactional dataset and a correlation graph is generated from these terms which will further help to select important sentences for the summary.

3.3 Automatic text summarization based on optimization

Recently, several approaches to text summarization have adopted the strategy to model the summarization process as a constrained maximum coverage, i.e., maximize essential aspects of the summary, for example, relevance and coherence, while taking into account restrictions imposed, such as maximum size of the summary. Several recent studies have been using ILP for solving this optimization problem (GILLICK et al., 2009).

A concept-based ILP model for summarization was proposed in (GILLICK; FAVRE, 2009) as a strategy that casts sentence selection as a maximum coverage problem. The assumption of the model is that the value of a summary is defined as the sum of the weights of the unique concepts it contains. Thus, a summary only benefits from including each concept once. (BOUDIN; MOUGARD; FAVRE, 2015) extended the model to reduce the number of concepts in the model, it uses a concept pruning technique and it showed through experiments that concept pruning leads to the presence of multiple optimal solutions.

One unsupervised extractive generic summarization model that uses Integer Linear Programming to solve the optimization problem which directly identifies important sentences from the document was proposed in (ALGULIEV et al., 2011). This approach is defined as Maximum Coverage and Minimum Redundancy (MCMR). It tries to optimize three important characteristics of a summary: Relevance, Redundancy and Length. A subset of sentences is chosen whether covers relevant text of the document collection. Then similarity is computed between the summary and the document collection and this similarity should be maximized. An objective function is defined and needs to be maximized assuring that the summary would consist of the relevant content present in the document collection.In addition, the summary won't have a large number of sentences expressing the same information. At the same time there is a constraint on the length of the summary. Finally an objective function is formed by linearly combining cosine similarity based objective function and NGD-based similarity objective function and this combined objective function also needs to be maximized.

(CAO et al., 2015) proposed Recursive Neural Networks (RNN) based ranking approach for ranking sentences in order to summarize multiple documents. Ranking of sentences is done through a hierarchical regression process which evaluates the relevance of a sentence (non- terminal node) in the parsing tree. On the basis of supervisions from word-level to sentence-level, recursive neural networks are automatically used to learn ranking features over the tree with inputs as hand-crafted feature vectors of words. Ranking scores of words and sentences are used to select important and non-redundant sentences to form summaries. Two methods are used here for sentence selection: greedy algorithm and Integer Linear Programming.

3.4 Automatic text summarization based on hybrid techniques

Hybrid techniques tend to use, explore and combine two or more techniques, for example, it may model the input document as a graph where the sentences are nodes and relationships between sentences are edges between nodes. Then, it can apply statistical techniques on the individual sentence nodes to rank the best sentences. Afterward, it can also model this graph as an optimization problem and use a technique such as Integer Linear Programming and select the sentence to form the final summary.

(TZOURIDIS; NASIR; BREFELD, 2014) developed a structured learning-based technique to compress multiple sentences. A sentence graph is used to represent related sentences such that each eadge represents a sentences and the vertices connecting edges represent the most similar sentences (FILIPPOVA, 2010). Instead of applying heuristics, dynamic programming has been adapted to the data. Word graphs and compressions are embedded in a joint feature space where compressions of different quality are learnt to be separated by a generalized linear scoring function. In order to decode the data, a generalized, loss-augmented shortest path algorithm has been developed that is solved through an integer linear program in polynomial time. A large margin approach is applied for adapting parameterized edge weights to the data such that the shortest path corresponds to the desired summary.

(KIKUCHI et al., 2014) suggested an approach for summarizing single documents that makes use of dependency between sentences obtained through rhetorical structures and dependency between words obtained through a dependency parser. Both of these dependencies are represented by building a nested tree for a document which is composed of two types of tree structures: a document tree in which nodes represent dependencies between sentences and a sentence tree in which nodes represent dependencies between words.

(PARVEEN; STRUBE, 2015) proposed an extractive graph-based unsupervised technique for summarizing single documents optimizing three important properties of summarization, i.e. importance, non-redundancy and local coherence. The input document is represented by a bipartite graph consisting of sentences and entity nodes. A graph based ranking algorithm, the o Hyperlink-Induced Topic Search (HITS), is applied on this graph for computing the rank of sentences based on their importance. The summary is made non-redundant and locally coherent through the process of optimization using Interger Linar Programming.

3.5 Summary

This chapter presented fundamental relevant studies for the development of this research. Topics related to the developed approach were studied in order to identify similarities and working gaps in the field of text summarization.

Table 2 shows a summary list of the presented approaches in this section by their article with its scope and developed methodology.

Table 2: Related work summary.

Article	Scope	Methodology
(FERREIRA et al., 2013)	Assessed diferent types of sentence scoring techniques for extractive text summarization	Comparison of different techniques and evaluation of the best context for usage based on qualitative and quantitative assessment
(CHAN, 2006)	Developed a quantitative model for creating summary that extracts sentences from highly relevant sections of the input documents	This approach performs information extraction through sentence based abstraction technique using shallow linguistic extraction techniques
(FATTAH; REN, 2009)	Improved extraction of sentences in text summarization using statistical techniques combined with genetic algorithms and mathematical regression	Sentences are assigned ranks in decreasing sequence of their scores and a highly scored collection of sentences are employed to produce the summary
(MENDOZA et al., 2014)	Proposed an extractive generic summarizationmethod for single documents by using generic operators and guided local search	This method uses a memetic algorithm which has combined the population based search of evolutionary algorithm with a guided local search strategy
(ALGULIEV et al., 2011)	Developed an unsupervised summarization model for generic text using Integer Linear Programming	This approach tries to optimize relevance, redundancy length of a summary using metrics of similarity between sentences and an objective function to maximize the properties
(GILLICK; FAVRE, 2009)	Proposed a concept-based ILP model for summarization	Developed a strategy that casts sentence selection as a maximum coverage problem including concepts once
(BOUDIN; MOUGARD; FAVRE, 2015)	Extended the concept-based ILP model to minimize redundancy globally	It uses concept-level weighting and maximum coverage to show through experiments that concept pruning leads to the presence of multiple optimal solutions
(CAO et al., 2015)	Proposed Recursive Neural Networks (RNN) based ranking approach for ranking sentences in order to summarize multiple documents	Ranking of words and sentences are used to select important and non-redundant sentences to form the parsing tree to the RNN learn ranking features over the tree
(KIKUCHI et al., 2014)	Suggested an approach for summarizing single documents that makes use of dependency between sentences	This method extracts a rooted document subtree from a document through rhetorical structures and dependency between words and uses an ILP to trim the tree to select a summary
(TZOURIDIS; NASIR; BREFELD, 2014)	Suggested a structured learning-based technique to compress multiple sentences	A word graph is used to represent related sentences such that summaries form the paths in the graph. It uses metrics to find the shortest path that corresponds to the desired summary
(BARALIS et al., 2013)	Proposed GRAPHSUM, a graph-based and general purpose summarizer for multiple documents	This approach explores and employs association rules for discovering correlations among multiple terms
(PARVEEN; STRUBE, 2015)	Attempted to generate non-redundant and locally coherent summaries from documents of different domains and genres	Input document is represented by means of a bipartite graph consisting of sentences and entity nodes. A graph based ranking algorithm is applied for computing the rank of sentences in order to select best summary

Source: The author

Upon reviewing the literature of automatic text summarization it was observed that extractive techniques to summarization do not usually take into account the coherence of the generated summaries, they usually try to select sentences that have a high relevance on the document. Contrary to that, the proposed system tries to maximize the salience of the selected sentences. It maximizes relevance but it also takes into consideration the coherence of the generated summary in order to generate the best readable possible summary.

In order to alleviate the above mentioned limitations, this work proposes an unsupervised ILP-based approach to single-document automatic text summarization that takes into account both the importance and the local coherence of the sentences. The documents to be summarized are represented by a bipartite entity graph consisting of a set of nodes denoting both sentences and entities of the documents. After that, the sentences are ranked according to a proposed measure of relevance based on a weighing method of the concepts that are optimized using ILP.

4 Single-Document Summarization Proposal

This chapter addresses the methodology used in this work providing the information needed to understand the proposed methodology and detailed information of the methods used.

The main goal of a single-document text summarization system is to extract the most relevant information from an input document and generate a condensed version of this document as a summary (NENKOVA; MCKEOWN, 2011). This definition usually captures three important aspects that have to be considered when designing such a system, they are:

- **Relevance**: Summaries must convey relevant information in the original documents;
- **Redundancy**: Summaries should not contain multiple sentences that represent the same information;
- **Length**: Summaries should be no longer than half the size of the original documents.

Optimize all these properties together is a challenging task and an example of a global inference problem (MCDONALD, 2007). This is due to the fact that including a new text unit to the summary does not depend solely on the properties of the new unit itself but also on the properties of all other text units that were previously added to the summary.

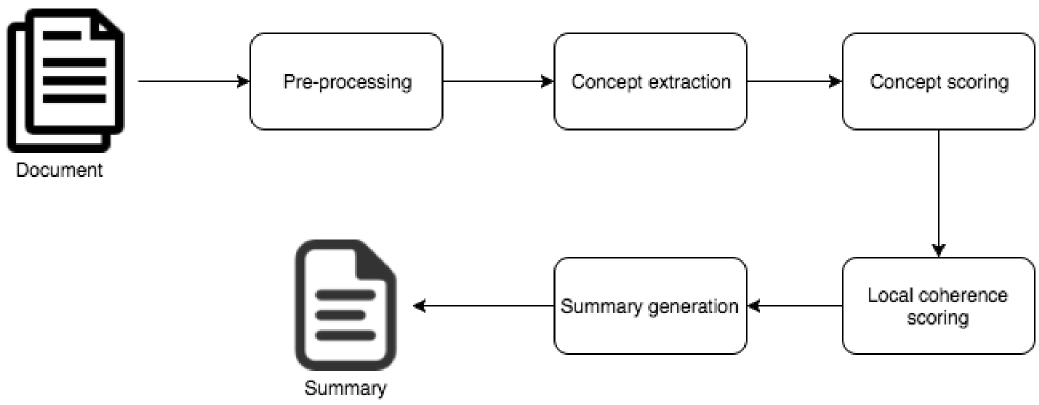
4.1 Proposed Methodology

The proposed approach is based on a global optimization procedure that cast the summarization task as an optimization problem as it was already done in (GILLICK; FAVRE, 2009). In other words, it selects the smallest possible number of sentences that maximize coverage of relevant concepts from the input document, taking into consideration the maximum size of the summary that should be generated. It is also considered the coherence of the sentences as a parameter to the optimization model.

An overview of the proposed approach is presented in Figure (3). The steps of the proposed solution is briefly illustrated below:

Figure 3: Overview of the proposed methodology

.



Source: The author

1. **Pre-processing**: The very first step is to pre-process the input document in order to have a version of the input document annotated with relevant grammatical information. The tasks involved in this step are sentence division, tokenization, lemmatization, and grammar class assignment, among others. After the document is pre-processed the system is able to read the sentences and build the summary problem accordingly.

2. **Concept extraction**: This stage is responsible for representing and extracting concepts from the list of tokens contained in each sentence. In this step the system is able to select the concepts in many different forms. One may use a unigram or a bigram to represent the concepts. Another choice is so select the words in each sentence removing stopwords or punctuation symbols.

3. **Concept scoring**: The concept scoring step is responsible for scoring each concept selected in the previous step to estimate its relevance to the document. In this step different techniques can be applied, such as the Word Frequency, Named Entities, TF-IDF/TF-ISF.

4. **Local coherence scoring**: This step uses the Entity Graph model, a extension of the Entity-Grid (GUINAUDEAU; STRUBE, 2013) to score the local coherence of each sentence of the document. Thus, it can be integrated to the optimization model, allowing maximization non-redundancy and local coherence while covering relevant information from the document

5. **Summary generation**: The final step is to process all the information generated in the previous steps and generate the summary. As mentioned before, the summarization problem is treated as an optimization problem, as presented in (4.4).

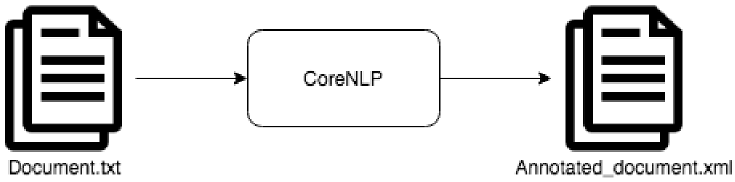
The above steps are detailed in the following subsections.

4.1.1 Pre-processing

Pre-processing is applied in the input document in order to have a better understanding of the concepts of each document. The system uses the Stanford CoreNLP[1] Toolkit (MANNING et al., 2014) to annotate the document with important grammatical information such as the base forms of words, their parts of speech, whether they are names of companies, people, tokenize the document sentences and so on.

Figure 4: Overview of the pre-processing step

.



Source: The author

All input documents are pre-processed and they generate a new annotated document in XML (eXtensible Markup Language) format that can be used in memory throughout the execution of the program or it can be stored and used later in other executions of the program without having to go trough the pre-processing phase annotating the documents again with the CoreNLP tool.

4.1.2 Concept extraction

The studies developed in (GILLICK et al., 2009) and (SCHLUTER; SØGAARD, 2015) suggested various methods for concept extraction, among them a few were highlighted as having the best performance over the others. The selected methods used in the present work are briefly described next.

N-gram is the most used representation in the literature. an n-gram is a contiguous sequence of n items from a given sequence of text. Specifically in text summarization, unigrams and bigrams (GILLICK; FAVRE, 2009; BOUDIN; MOUGARD; FAVRE, 2015) are most commonly used. N-grams are a set of co-occuring words within a given window and when computing n-grams you typically move one word forward. For example, the sentence *"the weather is wonderful today"* if N = 1, then the N-grams would be: ["the",

[1] CoreNLP is the natural language analysis tool used in the system. <http://stanfordnlp.github.io/CoreNLP/>

"weather", "is", "wonderful", "today"]. If N = 2, then the N-grams now would be: ["the weather", "weather is", "is wonderful", "wonderful today"].

When N=1, this is referred to as unigrams and this is essentially the individual words in a sentence. When N=2, this is called bigrams and when N=3 this is called trigrams. When N>3 this is usually referred to as four grams or five grams and so on.

Named Entities is a form of concept representation where the system uses expressions as concepts that refer to names of people, places, and organizations, among others. This reflects the intuition that such entities are important for text summarization because they describe real world entities that are mentioned in the document.

Syntactic dependencies use the syntactic dependencies between words as concepts. Some of these dependencies may be subjective, direct and indirect object, complement, among others. For example, given the phrase "John walks on the beach.", The following dependencies are extracted: root (ROOT, walks), nsubj (walks, John), case (beach, on), det (beach, the), nmod (walks, beach). As proposed by (SCHLUTER; SØGAARD, 2015), two forms of representations are derived from syntactic dependencies (i) using the nsubj dependency type (walks, John); and (ii) using a generic type to describe the dependencies dep (walks, John).

4.1.3 Concept scoring

A wide variety of methods for weighing the importance of concepts have been proposed in the literature.

Word co-occurrence measures the chance of two terms from a text appears alongside each other in a certain order. One way to implement this measure is using n-gram.

Word frequency is a technique that scores the importance of a word using its frequency, thereafter the more frequently a word occurs in the document, the higher its score. The assumption made in this technique is that the higher the frequency of a word, the more important it is to the overall document. The score of a word based on this method is calculated as follows in (4.1):

$$WordFrequency(s, w) = \frac{\sum_{i=0}^{n} Occur(w, s_i)}{n} \tag{4.1}$$

Where,
1. $Occur(w, s_i)$ returns 1 if a sentence s_i contains a word w
2. 0, otherwise
3. n is the number of analized sentences

The Term Frequency - Inverse Sentence Frequency (TF-ISF) (BLAKE, 2006) method is an extension of the traditional Term Frequency - Inverse Document Frequency (TF-IDF) (ROBERTSON, 2004) method. The difference between these two approaches is the level of granularity, while the TF-IDF scores based on all input documents presented in the dataset, the TF-ISF only scores based on the sentences presented in one document. In other words, TF-IDF is commonly used in multi-document whereas the TF-ISF is used in single-document summarization. The TF-ISF score of a word is computed as presented in (4.2)

$$TF - IFS(w) = FT(w) \times \log(\frac{n}{sf_w}) \tag{4.2}$$

Where,
1. FT returns the frequency of the term w in the sentences
2. sf_w is the number of sentences that contain the term w
3. n is the number of analized sentences.

In this approach Word Count, Word Frequency and TF-ISF were used in the experiments.

4.1.4 Local coherence scoring

The method used to model the local coherence is the Entity Graph proposed in (GUINAUDEAU; STRUBE, 2013) which is a graph-based approach based on the Entity-Grid proposed in (BARZILAY; LAPATA, 2008). These are methods used for local coherence modeling that captures the distribution of discourse entities across sentences in a text inspired by the Centrality Theory (GROSZ; WEINSTEIN; JOSHI, 1995) to estimate the local coherence of a text. Local coherence of a sentence is determined by the local entity transitions of the entities present or absent in the sentence. The intuition of these models is that shared entities by sentences, i.e., links that exist between sentences and discourse entities contribute to model local coherence.

The entity grid representation is as follows: each document is represented as an entity grid, which is a two-dimensional array that captures the distribution of entities across sentences. The rows of the grid correspond to the sentences, and the columns correspond to the entities. In the entity graph, each document is represented as an entity graph, which is a bipartite graph where one set of nodes represent the sentences and the other set of nodes represent the concepts. Thus, various relationships between sentences of the document and their respective sentences can be used. In the proposed approach the *outdegree* function from the projection graph, that computes the number of edges exiting the vertex corresponding to a sentence, is used.

Figure(5) presents an example of an entity graph created from the sentences $S1$, $S2$, $S3$ and $S4$ listed below. The projection of the sentences as an unweighted entity graph is illustrated in (7). This projection demonstrates sentences that share entities between them and how to score local coherence based on this representation.

$S1$: Haemorrhage is a common cause of death in trauma patients.
$S2$: Although transfusions are extensively used in the care of bleeding trauma patients, there is uncertainty about the balance of risks and benefits and how this balance depends on the baseline risk of death.
$S3$: Our objective was to evaluate the association of red blood cell (RBC) transfusion with mortality according to the predicted risk of death.
$S4$: A secondary analysis of the CRASH-2 trial (which originally evaluated the effect of tranexamic acid on mortality in trauma patients) was conducted.

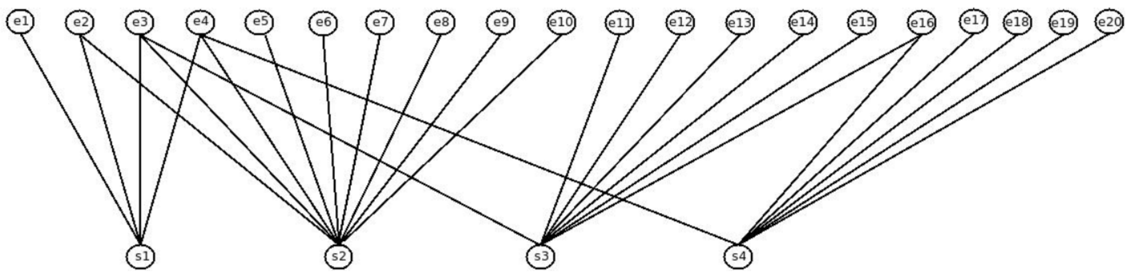
Figure 5: Entity-Grid representation of sentence rows and entities columns above

.

	HAEMORRHAGE (e1)	CAUSE (e2)	DEATH (e3)	PATIENTS (e4)	TRANSFUSIONS (e5)	CARE (e6)	THERE (e7)	UNCERTAINTY (e8)	BALANCE (e9)	BENEFITS (e10)	RISK (e11)	OBJECTIVE (e12)	ASSOCIATION (e13)	CELL (e14)	RBC (e15)	MORTALITY (e16)	ANALYSIS (e17)	TRIAL (e18)	EFFECT (e19)	ACID (e20)
S_1	S	O	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S_2	-	-	X	X	S	X	S	X	S	X	X	-	-	-	-	-	-	-	-	-
S_3	-	-	X	-	-	-	-	-	-	-	X	S	O	X	X	X	-	-	-	-
S_4	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	X	S	X	O	X

Source: Parveen e Strube (2015, p. 1300)

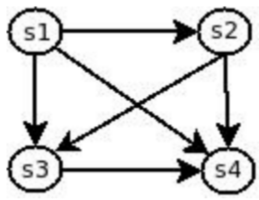
Figure 6: Entity Graph representation of sentence rows and entities columns above



Source: Parveen e Strube (2015, p. 1300)

To score the local coherence from the entity graph with the optimization model adopted in this study, it is adapted the approach used and suggested in (PARVEEN; STRUBE, 2015) that used the local coherence of each sentence present in the projection created from the entity graph. Therefore, it is possible to plug in the objective function of the model in order to maximize both the coverage of important concepts and the selection of sentences that maximize the local coherence of the summary.

Figure 7: One mode projection of the bipartite graph entity



Source: Parveen e Strube (2015, p. 1300)

The coherence score of a sentence used in this work is given by (4.3).

$$coherence(s_i) = Outdegree(s_i, P) \tag{4.3}$$

Where, 1. s_i is a sentence from the input document
2. P is the projection of the entity graph.

Equation (4.3) calculates the outdegree of every sentence from the projection graph. This coherence value is used to select sentences for a summary in the optimization phase. The ILP model will only select sentences that maximize the coherence value.

In this approach, the Entity Graph model is used to score local coherence of sentences due to the competitive results and to solve some of the problems in the Entity-Grid detailed in (GUINAUDEAU; STRUBE, 2013), also for the good performance obtained by the recent work of (PARVEEN; STRUBE, 2015; PARVEEN; RAMSL; STRUBE, 2015) in the single-document summarization task.

4.1.5 Summary generation

The last step in the proposed approach, the results of all previous steps are integrated as an ILP-based optimization problem. The proposed ILP-based solution extends the original version proposed in (MCDONALD, 2007) and they are illustrated in the rest of this chapter.

In this approach, relevance and coherence are aspects treated simultaneously inside the formula of the optimization model as presented below:

$$\text{Maximize:}(\sum_{c_i \in C} w_i \cdot c_i + \sum_{s_j \in S} Rank(s_j) \cdot s_j) \tag{4.4}$$

$$\sum_{s_j \in S} len(s_j) \leq L \tag{4.5}$$

$$s_j Occ_{ij} \leq c_i \tag{4.6}$$

$$\sum_{s_j \in S} s_j Occ_{ij} \geq c_i \tag{4.7}$$

$$c_j, s_j, Occ_{ij} \in \{0, 1\} \forall_i \tag{4.8}$$

The variables c_j, s_j and Occ_{ij} are binary values that indicate a concept c_j, a sentence s_j and the occurrence of a concept c_j in a sentence s_j, respectively. The concept is a basic unit of text that can, for example, represent all nouns of the text. The sentences are the input document original sentences. The variable w_i represents the weight of a concept, i.e. the importance of a concept c_j in the set of all the concepts C extracted from the input document. $Rank(s_j)$ is the local coherence score of each sentence s_j in the set of sentences S generated from the Entity Graph (GUINAUDEAU; STRUBE, 2013).

The first part of the objective function $\sum_{c_i \in C} w_i \cdot c_i$ defines the importance of the summary, selecting the largest number of important concepts, while the second part is related to coherence of the sentences $\sum_{s_j \in S} Rank(s_j) \cdot s_j$. The variable $len(s_j)$ is the length of each sentence s_j of the set of sentences S L is the threshold used to define the maximum length of the generated summary.

The Inequality (4.5) ensures that the summary to be generated can not exceed the L value that defines the maximum length of the generated summary. The inequities (4.6) (4.7) associate the sentences and concepts. This ensures that selecting a sentence leads to selection of all the concepts it contains, and selecting a concept only happens when it is present in at least one of the selected sentences.

5 Experimental Evaluation

This chapter presents the datasets used in the proposed experimental assessment, evaluation metrics and the results obtained by the performed tests for the extractive single-document summarization task using the proposed summarization system. It first presents the datasets and metrics used in the experimental evaluation. Then, several experimental are performed in order to assess the effectiveness of the proposed summarizer.

5.1 Datasets

In this section the selected corpora for single-document summarization are presented. This system focuses on summarizing documents written in English, and it is intended for generic summarization, hence the choice for the corpora. Experiments are performed on the well-known Document Understanding Conferences (DUC) 2001 and 2002 single-document summarization datasets.

The DUC conference (OVER; DANG; HARMAN, 2007) from the years 2001 to 2004 was focused on the generic single-document summarization and multi-document of news articles. The corpora of the years 2001 and 2002 were created especially for single-document summarization and they are still one of the most used datasets to evaluate new approaches of single-document summarization.

Table (3) shows the name of the corpus or creator, the year of creation, the construction process, the type of abstracts available and the number of documents.

Table 3: Datasets distribution for experiments

Dataset	**Year**	**Contruction process**	**Golden summaries**	**Documents**
DUC 2001	2001	Manual	Abstractive (Human)	309
DUC 2002	2002	Manual	Abstractive (Human)	567

Source: The author

The construction process adopted in both DUC 2001 and 2002 is hand made. In this process, golden summaries are abstractive and generated by a human who reads each document and then performs the summarization process. In general, in order to standardize the generation of summaries, a set of instructions with rules to guide the summarization process is made available. The main advantage of this type of approach is the quality of the summary generated.

The corpus of the DUC 2001 contains 309 documents, however it was verified the existence of 1 duplicated document, meaning that this corpus has 308 different documents.

The DUC 2002 corpus has 567 documents, but just as in DUC 2001, 34 repeated documents were present, so this corpus actually contains 533 different documents. In both corpora, for each document there are two abstract reference summaries written by human summarizers with approximately 100 words. Table (4) summarizes some basic statistics of both datasets.

Table 4: Datasets distribution for experiments

Dataset	**Documents**	**Sentences**	**Words**
DUC 2001	309	11.026	269.990
DUC 2002	576	14.370	348.012

Source: The author

5.2 Evaluation metrics

Evaluating the performance of a summarization system is a difficult task by itself and there are various measures that can be used to determine how good a summary is based on some criteria. Currently, one of the most used automatic evaluation measures is the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) (LIN, 2004)

Formally, ROUGE-N is an n-gram recall between a candidate summary, generated by an automatic text summarization system, and a set of reference summaries called golden standard. ROUGE-N is computed as follows:

$$ROUGE - N = \frac{\sum_{S \in S_{ref}} \sum_{grama_n \in S} Count_{match}(grama_n)}{\sum_{S \in S_{ref}} \sum_{grama_n \in S} Count(grama_n)} \tag{5.1}$$

Where,
1. S_{ref} stands for the set of reference summaries,
2. n stands for the length of the n-gram in consideration,
3. $Count_{match}(grama_n)$ is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries,
4. $Count(grama_n)$ represents the maximum number of n-grams occurring in the reference summaries.

The ROUGE evaluation package has been one of the main automatic evaluation techniques adopted in the literature in recent years (GAMBHIR; GUPTA, 2016). However, an important limitation is the fact that only lexical overlaps are taken into consideration in the algorithm. Therefore, several other evaluation measures have been proposed with the aim of providing more semantic-based analysis to compute the similarity between the candidate and reference summaries. For the following experiments it is used the coverage ROUGE measure to compare the results obtained by each system.

5.3 Concept scoring evaluation

This experiment evaluates the performance of each concept scoring method (Word count, word frequency and TF-ISF). It aims to select the best method to compose the part of the algorithm that is responsible for weighting concepts. For this, the second part of the objective function presented in (4.4), which represents the local coherence of the sentences, was not taken into consideration.

Some of the main summarization techniques based on surface text features adopted in the literature include frequency of the words, position of the sentences, similarity of the sentences with the title of the document, among others. (FERREIRA et al., 2013) identified and evaluated the performance of various superficial techniques to measure the importance of sentences

The scoring methods of the concepts were evaluated using the weight distribution strategies: All Occurrences and Highest Ranked Occurrences. The first strategy consists of the traditional approach that assigns weight to all occurrences of a concept, whereas the second one is a proposal that attributes weight only to a subset Θ of the highest ranked concepts. For instance if $\Theta = 2$ it will select half of the highest weighted concepts, if $\Theta = 3$ it will select one third of the highest weighted concepts. For this experiment $\Theta = 3$ yielded the best results, thus it selects the $\frac{1}{3}$ subset of the highest ranked concepts. The results obtained in this experiment, in terms of the coverage measure of ROUGE-1 (R-1) and ROUGE-2 (R-2), where R-1 is a unigram representation whereas R-2 is a bigram, are presented in table (5) and (6) shows that selecting only a subset Θ of the most relevant results in a slight improvement in the performance of the scoring method for both DUC 2001 and 2002 datasets in all measurements evaluations.

Table 5: Results of the concept scoring methods based on ROUGE-1 (R-1) and ROUGE-2 (R-2) measure in the DUC 2001 dataset. The best overall performance is highlighted in bold

Scoring Method	**Distribution Strategy**	**R-1**	**R-2**
Word Count	All occurences	42.99	17.02
Word Frequency	All occurences	43.06	17.11
TF-ISF	All occurences	43.18	17.25
Word Count	$(\theta = 3)$	42.91	16.92
Word Frequency	$(\theta = 3)$	43.15	**17.61**
TF-ISF	$(\theta = 3)$	**43.44**	17.47

Source: The author

Table (5) shows that for the DUC 2001 dataset the best scoring methods are Word Frequency and TF-ISF, where the first one had the best R-2 measure performance and the second the best R-1. Also, the Highest Ranked Occurrences improved the scores of most of the evaluated methods.

Table 6: Results of the concept scoring methods based on ROUGE-1 (R-1) and ROUGE-2 (R-2) measure in the DUC 2002 dataset. The best overall performance is highlighted in bold

Scoring Method	**Distribution Strategy**	**R-1**	**R-2**
Word Count	All occurences	45.68	19.39
Word Frequency	All occurences	45.90	19.73
TF-ISF	All occurences	44.13	18.06
Word Count	$(\theta = 3)$	45.78	19.54
Word Frequency	$(\theta = 3)$	**46.11**	**19.99**
TF-ISF	$(\theta = 3)$	44.53	18.41

Source: The author

Table (6) shows that for the DUC 2002 dataset the Word Frequency has the best scores in both R-1 and R-2. In addition, as in the experiments using the DUC 2001 dataset the Highest Ranked Occurrences improved the scores of most of the methods.

The above experiments in table (5) and (6) shows that the strategy to limit the amount of extracted concepts from a sentence based on a threshold θ performed better than selecting all available concepts from a sentence in most of the comparisons in terms of the measures of evaluation adopted. Regarding the methods of weighing the concepts, the Word Frequency method outperformed the other methods in most of the evaluated datasets.

The strategy of only selecting the highest ranked concepts outperformed the traditional weight distribution in the comparisons in terms of measurements adopted. The Highest Ranked Occurrences strategy where $\Theta = 3$, i.e. selecting only the $\frac{1}{3}$ of the highest ranked concepts from a sentence showed significant improvements in terms of R-1 and R-2 measures and it was significantly superior to the strategy of selecting all concepts. This result can be traced to the fact that selecting only a subset of the highest ranked concepts in a sentence increases the scoring of the most relevant sentences in the document.

5.4 Evaluation of the proposed system

The summarization proposal introduced in the previous chapter is highly customizable and there are various components of the system that can be tuned to obtain a higher performance. Therefore, this next experiment has the objective of selecting the best possible setup for both DUC 2001 and 2002 datasets, which enables that its results can be compared with state-of-the-art systems.

As explained in the previous chapter the summarization system is formulated as an ILP function described in (4.4). In short the first part of the expression evaluates the score of a concept whereas the second evaluates the sentence score and its local coherence.

Therefore, the system can be configured in different ways. One might try and select the best concept scoring algorithm, the best representation of local coherence or even not using any local coherence measurement at all. Thus, table (7) shows various configuration attempted in this experiment in order to select the best set of algorithms to compose the final summarization system.

Table 7: Different configurations for the proposed sumarization system. The best performances are highlighted in bold

DUC 2001		
Configuration	**R-1**	**R-2**
(Word Count + ILP)	42.91	16.92
(Word Frequency + ILP)	43.15	17.61
(TF-ISF + ILP)	43.44	17.47
(Word Count + Entity Graph + ILP)	42.89	16.89
(Word Frequency + Entity Graph + ILP)	**45.00**	17.91
(TF-ISF Count + Entity Graph + ILP)	43.90	**18.27**
DUC 2002		
Configuration	**R-1**	**R-2**
(Word Count + ILP)	45.78	19.54
(Word Frequency + ILP)	46.11	19.99
(TF-ISF + ILP)	45.53	18.41
(Word Count + Entity Graph + ILP)	46.07	19.87
(Word Frequency + Entity Graph + ILP)	**47.36**	**20.96**
(TF-ISF Count + Entity Graph + ILP)	45.28	20.17

Source: The author

As shown in table (7) the highest performance for both DUC 2001 and 2002 datasets considering both ROUGE R-1 and R-2 measures is the setup which consists of using the Word Frequency algorithm for concept scoring, the outdegree function to model local coherence from the projection graph of the entity graph and plugging all these values into the ILP function.

5.5 Comparison with other Approaches

This experiment compares the performance of the proposed summarization approach with the following systems: The best systems participating in the DUC 2001 and 2002 competitions,System T and System 28, respectively; Three summarization systems that presented the best performance in the comparative evaluation described in (BATISTA et al., 2015). These systems are AutoSummarizer (AUTOSUMMARIZER, 2016), Classifier4J (LOTHIAN, 2003) and HP-UFPE FS (FERREIRA et al., 2014) and they are described below:

- **AutoSummarizer** is a single-document summarizer available online, which selects the most important sentences from the input document to generate a summary. Unfortunately, details of how this system works were not found, however, it has yielded good results in a comparative analysis between different summarization systems;

- **Classifier4J** is a library that provides services for sorting and summarizing a single document of texts. Classifier4J first extracts the one hundred most frequent words from the input document as keywords, and then selects the first sentences of the text that it has by one of the extracted keywords

- **HP-UFPE Functional Summarization** (HP-UFPE FS) is a summary system based on the combination of superficial scoring methods. In order to measure the importance of the sentences, this system uses the best combination of the techniques analyzed in (FERREIRA et al., 2013).

Tables (8) and (9) shows the experimental results in terms of the R-1 and R-2 coverage measure. The best configuration found in the previous experiment was compared to the selected systems, i.e., taking into account the local coherence score of the sentences generated from the entity graph with TF-ISF as the concept scoring method.

The first experiment comparing the proposed system with the state-of-the-art summarization systems in the DUC 2001 dataset is shown in table (8). It consists in comparing the generated summaries of each of the systems mentioned above with the proposed one.

Table 8: Comparative results of the systems on the DUC 2001 dataset. The system with the best performance is highlighted in bold

Strategy	**R-1**	**R-2**
AutoSummarizer	41.92	16.63
Classifier4J	44.44	19.86
HP-UFPE FS	35.91	11.78
System T	44.53	**20.27**
Proposed system	**45.00**	17.91

Source: The author

In the DUC 2001 dataset, the proposed system obtained the best performance based on R-1 and third best on R-2, being outperformed by the best system specifically designed to this dataset and the Classifier4J system but the result was not significantly lower, it was a competitive result.

The first experiment also compares the proposed system with the same state-of-the-art summarization systems on the DUC 2002 dataset and it is shown in table (9).

Table 9: Comparative results of the systems on the DUC 2002 dataset. The system with the best performance is highlighted in bold

Strategy	R-1	R-2
AutoSummarizer	43.79	19.17
Classifier4J	47.09	22.12
HP-UFPE FS	45.70	20.55
System 28	**48.07**	**22.88**
Proposed system	47.36	20.96

Source: The author

In the DUC 2002 dataset, the proposed approach with local coherence obtained the second best performance based on both R-1 and R-2, being outperformed only by the best system that was specifically designed to this dataset.

5.6 Comparing local sentence coherence

This last experiment aims to evaluate the local coherence of the selected sentences from the generated summary of the proposed system. Table (10) shows the summaries produced by the Classifier4J system and the proposed system on the document AP880622-0184 on the 2002 DUC dataset.

Table 10: Comparison of generated summaries by Classifier4J system and the proposed system on the document AP880622-0184 on the DUC 2002 dataset

System	Summary
Classifier4J	Beverly Sills, Lauren Bacall, Betty Comden and Phyllis Newman are among performers who will sing, act and make guest appearances at a birthday bash in August for conductor Leonard Bernstein. The Leonard Bernstein Gala Birthday Performance is a benefit concert scheduled for the composer's 70th birthday, Aug. 25, to raise money for the Tanglewood Music Center, where Bernstein got his conducting start. Sills will be host. Performances will include the Boston Symphony Orchestra, the Boston Pops Orchestra and the Tanglewood Festival Chorus under the direction of some of the many conductors whose careers have been guided by Bernstein.
Proposed System	The Leonard Bernstein Gala Birthday Performance is a benefit concert scheduled for the composer's 70th birthday, Aug. 25, to raise money for the Tanglewood Music Center, where Bernstein got his conducting start. Bacall and soprano Barbara Hendricks will perform a movement from Bernstein's Symphony No. 3, "Kaddish." Beverly Sills, Lauren Bacall, Betty Comden and Phyllis Newman are among performers who will sing, act and make guest appearances at a birthday bash in August for conductor Leonard Bernstein. Dame Gwyneth Jones and Frederica von Stade will be among those performing highlights from "Fidelio," "Tristan und Isolde" and other works to honor Bernstein's landmark opera recordings.

Source: The author

As can be seen from the table (10) there are indicators that the proposed system has a higher local coherence than the Classifier4J. One evidence of that lies on the fact that the very first sentence of the Classifier4J is *"Beverly Sills, Lauren Bacall, Betty Comden and Phyllis Newman are among performers who will sing, act and make guest appearances at a birthday bash in August for conductor Leonard Bernstein."* but someone reading this sentence may wonder where these people will perform and the Classifier4j does not make that clear until the user reads their second sentence which is *"The Leonard Bernstein Gala Birthday Performance is a benefit concert scheduled for the composer's 70th birthday, Aug. 25, to raise money for the Tanglewood Music Center, where Bernstein got his conducting start."* Those sentences are out of order and therefore they are not coherent. The proposed system, however, selects the sentence *"The Leonard Bernstein Gala Birthday Performance is a benefit concert scheduled for the composer's 70th birthday, Aug. 25, to raise money for the Tanglewood Music Center, where Bernstein got his conducting start."* as its first sentence and builds the subsequent sentences from it, making the summary more locally coherent, easier to understand and more pleasant to read.

Table (11) shows another comparison of a summary generated by the proposed system with AutoSummarizer on the document AP880816-0234 on the DUC 2001 dataset.

Table 11: Comparison of generated summaries by AutoSummarizer system and the proposed system on the document AP880816-0234 on the DUC 2001 dataset

System	**Summary**
AutoSummarizer	Police said members of Shining Path, a Maoist group, killed two policemen and wounded three in jungle raids. The Rodrigo Franco Command, which has vowed to kill a Shining Path member or sympathizer for every person slain by guerrillas, issued the threat against District Attorney Carlos Escobar on Monday, according to his office in Andean city of Ayacucho. The Rodrigo Franco group is named for an official of the government party killed the Shining Path killed last year.
Proposed System	A death squad opposed to the Shining Path guerrillas has threatened to kill a district attorney if he investigates charges that soldiers massacred dozens of peasants, his office said Tuesday. Police said members of Shining Path, a Maoist group, killed two policemen and wounded three in jungle raids. Escobar is investigating charges that troops rounded up dozens of peasants, accused them of being Shining Path members and killed them. The Rodrigo Franco group is named for an official of the government party killed the Shining Path killed last year. The alleged massacre occurred in May near Cayara, a farming village 40 miles south of Ayacucho.

Source: The author

In Table (11), one can see a clear difference between the proposed system and AutoSummarizer. The first sentence selected by proposed system, "A death squad opposed to the Shining Path guerrillas has threatened to kill a district attorney if he investigates charges that soldiers massacred dozens of peasants, his office said Tuesday" was not even selected by AutoSummarizer. This can be deduced from the fact that AutoSummariser did not evaluate this sentence as one of the most relevant.

The experiments described in Tables (10) and (11) illustrates how the propsed extractive system selected the sentences when composing the summary taking into account both the importance and the coherence of sentences. These two examples also illustrated the difference between a summary generated with local coherence against another summary generated by a system that do not considers local coherence.

5.7 Discussion

The results obtained demonstrated that the approach proposed in this chapter presents competitive results with the state-of-the-art summarization systems analyzed.

In the DUC 2001 corpus, the proposed system with local coherence obtained the best performance based on R-1 and third best based on R-2, being outperformed only by the best system specifically designed to this dataset and the Classifier4J system on the R-2 measure. For the DUC 2002 corpus, the same local coherent system obtained the second best performance based on both R-1 and R-2, being outperformed only by the best system specifically designed to this dataset.

Surprisingly, even after more than a decade since the competitions of DUC 2001 and DUC 2002, the best participants identified in this research, presented competitive results in relation to systems recently developed. This good performance is mainly due to the fact that these systems were developed specifically for the corpora of their respective competitions.

The experimental results in terms of ROUGE R-1 and R-2 measurements demonstrate that the proposed summarization system is viable and presented competitive results with several state-of-the-art systems. Furthermore it shows that the system is heading towards a more coherent approach to extractive summarization system as shown in one of the experiments where it was compared a generated summary from one of the state-of-the-art system with the proposed system: The proposed system generated a more coherent summary among the compared systems.

6 Conclusion

As shown in this document, automatic text summarization has been object of several researches over 50 years, and is still a challenging problem in Text Mining with several applications in real life scenarios.

The review of the literature about ATS revealed that extractive techniques to automatic summarization do not usually take into consideration the coherence of the generated summaries, as they usually try to only maximize salience (relevance) of the sentences.

This work is then an attempt to contribute to the ATS by proposing, implementing, and evaluating a single-document extractive summarizer that combines methods for concept selection, concept scoring, and local coherence scoring. Moreover, the prototype implementation (as a proof-of-concept) integrates all the above scoring methods in a single objective function with constraints by using ILP. The proposed system differs from others mainly because it maximizes both concepts relevance and coherence of the generated summary.

The proposed system was evaluated and compared to state-of-the-art summarizers on two well-known single-document benchmark datasets using ROUGE as evaluation metric. The achieved results were encouraging as the proposed summarization system is as effective in producing good summaries as the best results collected from the state-of-the-art extractive summarization systems. In particular, it achieved the best results in terms of the R-1 ROUGE coverage metric on the DUC 2001 dataset.

It was also conducted a preliminary human evaluation in order to assess the level of cohesiveness of the summaries generated by the proposed solution was usually more coherent than its competitors for the same documents.

To sum up, given the above considerations, the present work described an attempt to integrate both relevance and coherence of the sentences in a single-document extractive summarization system.

6.1 Future Work

The proposed system currently has three main components that collaborate to generate the most relevant and locally coherent summary of a given input document. The proposed model is a generic one, i.e., it summarizes any type of document regardless of its domain concerned by the document. However, as shown in (FERREIRA et al., 2013) different techniques are usually better suited to a specific domain than others. Thus, as future work, it would be interesting to conduct a deeper analysis of other sentence scoring techniques in different domains including scientific articles from Computer Science and Biomedical, and technical documents. It would also be of much interest to automatically evaluate the coherence of the summaries generated in these domains. In addition, a specific benchmark dataset for evaluating coherence could also be constructed or adopted enabling the comparison among summarization systems.

The system could potentially have another component responsible for topic and domain classification. Such a component would be included in the pre-processing phase, for instance. As a result, the system could select the best scoring techniques that are more suited to a given domain or type of documents.

Finally, the proposed single-document extractive summarizer can be adapted to a multi-document one. In this case, it should consider other aspects besides the coherence of the sentences, such as redundancy control and sentence ordering.

Bibliography

ALGULIEV, R. M. et al. Mcmr: Maximum coverage and minimum redundant text summarization model. *Expert Syst. Appl.*, v. 38, n. 12, p. 14514–14522, 2011. 27, 29

AUTOSUMMARIZER. 2016. Disponível em: <http://autosummarizer.com/>. 43

BARALIS, E. et al. Graphsum: Discovering correlations among multiple terms for graph-based summarization. *Inf. Sci.*, v. 249, p. 96–109, 2013. 26, 29

BARRERA, A.; VERMA, R. Combining syntax and semantics for automatic extractive single-document summarization. In: *Proceedings of the 13th International Conference on Computational Linguistics and Intelligent Text Processing - Volume Part II*. Berlin, Heidelberg: Springer-Verlag, 2012. (CICLing'12), p. 366–377. ISBN 978-3-642-28600-1. 25

BARZILAY, R.; LAPATA, M. Modeling Local Coherence: An Entity-Based Approach. *Computational Linguistics*, v. 34, n. 1, p. 1–34, 2008. ISSN 0891-2017. 17, 21, 22, 35

BATISTA, J. et al. A quantitative and qualitative assessment of automatic text summarization systems. In: *Proceedings of the 2015 ACM Symposium on Document Engineering*. New York, NY, USA: ACM, 2015. (DocEng '15), p. 65–68. ISBN 978-1-4503-3307-8. 43

BLAKE, C. A comparison of document, sentence, and term event spaces. In: *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006. (ACL-44), p. 601–608. 35

BOUDIN, F.; MOUGARD, H.; FAVRE, B. Concept-based Summarization using Integer Linear Programming : From Concept Pruning to Multiple Optimal Solutions. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, n. September, p. 1914–1918, 2015. 15, 17, 27, 29, 33

CAO, Z. et al. Ranking with recursive neural networks and its application to multi-document summarization. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI Press, 2015. (AAAI'15), p. 2153–2159. ISBN 0-262-51129-0. 27, 29

CHAN, S. W. K. Beyond keyword and cue-phrase matching: A sentence-based abstraction technique for information extraction. *Decis. Support Syst.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 42, n. 2, p. 759–777, nov. 2006. ISSN 0167-9236. 25, 29

DAS, D.; MARTINS, A. F. T. A Survey on Automatic Text Summarization. *Eighth ACIS International Conference on Software Engineering Artificial Intelligence Networking and ParallelDistributed Computing SNPD 2007*, v. 4, p. 574–578, 2007. ISSN 0080448542. 15, 17, 19, 25

ERKAN, G.; RADEV, D. R. Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Int. Res.*, AI Access Foundation, USA, v. 22, n. 1, p. 457–479, dez. 2004. ISSN 1076-9757. 26

FATTAH, M. A.; REN, F. Ga, mr, ffnn, pnn and gmm based models for automatic text summarization. *Comput. Speech Lang.*, Academic Press Ltd., London, UK, UK, v. 23, n. 1, p. 126–144, jan. 2009. ISSN 0885-2308. 25, 29

FERREIRA, R. et al. Assessing sentence scoring techniques for extractive text summarization. *Expert Systems with Applications*, Elsevier Ltd, v. 40, n. 14, p. 5755–5764, 2013. ISSN 09574174. 17, 19, 25, 29, 41, 44, 48

FERREIRA, R. et al. A Context Based Text Summarization System. *2014 11th IAPR International Workshop on Document Analysis Systems*, p. 66–70, 2014. 17, 43

FILIPPOVA, K. Multi-sentence compression: Finding shortest paths in word graphs. In: *Proceedings of the 23rd International Conference on Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. (COLING '10), p. 322–330. 28

FILIPPOVA, K.; STRUBE, M. Extending the entity-grid coherence model to semantically related entities. *Proceedings of the Eleventh European Workshop on Natural Language Generation - ENLG '07*, p. 139, 2007. 17

GAMBHIR, M.; GUPTA, V. Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, Springer Netherlands, p. 1–66, 2016. ISSN 1573-7462. 13, 17, 20, 25, 40

GILLICK, D.; FAVRE, B. A Scalable Global Model for Summarization. *Proceedings of the Workshop on Integer Linear Programming for Natural Langauge Processing*, n. June, p. 10–18, 2009. 17, 27, 29, 31, 33

GILLICK, D. et al. The ICSI/UTD Summarization System at TAC 2009. *In Proceedings of the Text Analysis Conf. Workshop*, p. 801–815, 2009. 26, 33

GROSZ, B. J.; WEINSTEIN, S.; JOSHI, A. K. Centering: A framework for modeling the local coherence of discourse. *Comput. Linguist.*, MIT Press, Cambridge, MA, USA, v. 21, n. 2, p. 203–225, jun. 1995. ISSN 0891-2017. 35

GUINAUDEAU, C.; STRUBE, M. Graph-based Local Coherence Modeling. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, n. 2008, p. 93–103, 2013. 17, 22, 32, 35, 37, 38

KENDALL, G.; PARKES, A. J.; SPOERER, K. A survey of np-complete puzzles. *ICGA Journal*, v. 31, n. 1, p. 13–34, 2008. 15

KIKUCHI, Y. et al. Single Document Summarization based on Nested Tree Structure EDU sentence subtree sentence selection selection. *52nd Annual Meeting of the Association for Computational Linguistics*, p. 315–320, 2014. 28, 29

LI, C.; QIAN, X.; LIU, Y. Using Supervised Bigram-based ILP for Extractive Summarization. *Acl*, p. 1004–1013, 2013. 23

LIN, C.-Y. Rouge: A package for automatic evaluation of summaries. In: MOENS, S. S. M.-F. (Ed.). *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*. Barcelona, Spain: Association for Computational Linguistics, 2004. p. 74–81. 40

LOTHIAN, N. *Classifier4J*. 2003. Disponível em: <http://classifier4j.sourceforge.net/>. 43

LUHN, H. P. The automatic creation of literature abstracts. *IBM J. Res. Dev.*, IBM Corp., Riverton, NJ, USA, v. 2, n. 2, p. 159–165, abr. 1958. ISSN 0018-8646. 25

MANNING, C. D. et al. The Stanford CoreNLP Natural Language Processing Toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, p. 55–60, 2014. ISSN 1098-6596. 21, 33

MCDONALD, R. A Study of Global Inference Algorithms in Multi-Document Summarization. *Ecir*, p. 557–564, 2007. ISSN 03029743. 15, 17, 23, 31, 37

MENDOZA, M. et al. Extractive single-document summarization based on genetic operators and guided local search. *Expert Syst. Appl.*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 41, n. 9, p. 4158–4169, jul. 2014. ISSN 0957-4174. 26, 29

MIHALCEA, R.; TARAU, P. Textrank: Bringing order into texts. In: LIN, D.; WU, D. (Ed.). *Proceedings of EMNLP 2004*. Barcelona, Spain: Association for Computational Linguistics, 2004. p. 404–411. 25

MURDOCK, V. Aspects of sentence retrieval. *SIGIR Forum*, v. 41, n. 2, p. 127, 2007. 25

NENKOVA, A.; MCKEOWN, K. Automatic Summarization. *Foundations and Trends® in Information Retrieval*, v. 5, n. 3, p. 235–422, 2011. ISSN 1554-0669. 13, 14, 16, 19, 20, 25, 31

OVER, P.; DANG, H.; HARMAN, D. Duc in context. *Inf. Process. Manage.*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 43, n. 6, p. 1506–1520, nov. 2007. ISSN 0306-4573. 39

PARVEEN, D.; RAMSL, H.-M.; STRUBE, M. Topical coherence for graph-based extractive summarization. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, 2015. p. 1949–1954. 37

PARVEEN, D.; STRUBE, M. Integrating importance, non-redundancy and coherence in graph-based extractive summarization. *IJCAI International Joint Conference on Artificial Intelligence*, v. 2015-Janua, n. Ijcai, p. 1298–1304, 2015. ISSN 10450823. 23, 28, 29, 36, 37

ROBERTSON, S. Understanding inverse document frequency: on theoretical arguments for idf. *Journal of Documentation*, v. 60, n. 5, p. 503–520, 2004. 35

R.RADEV, D.; MCKEOWN, K. Introduction to the Special Issue on Summarization. *Computational Linguistics*, v. 28, n. 4, p. 399–408, 2002. ISSN 0891-2017. 13, 14, 17, 19

SCHLUTER, N.; SØGAARD, A. Unsupervised extractive summarization via coverage maximization with syntactic and semantic concepts. *Proceedings ACL 2015*, n. 2009, p. 840–844, 2015. 23, 33, 34

TZOURIDIS, E.; NASIR, J. A.; BREFELD, U. Learning to Summarise Related Sentences. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*, p. 1636–1647, 2014. 28, 29