

### **3º MILESTONE DA AVALIAÇÃO**

Definição clara do problema existente e planejamento de uma  
investigação empírica

## SUMÁRIO

1 DEFINIÇÃO DO PROBLEMA.....	3
2 OBJETIVOS DO EXPERIMENTO.....	3
3 PLANEJAMENTO DA INVESTIGAÇÃO EMPÍRICA.....	4
3.1 Seleção do Contexto.....	4
3.2 Definição das Hipóteses.....	4
3.3 Seleção dos Sujeitos.....	5
3.4 Design de Experimentos.....	5
3.4.1 O Experimento da 1ª Etapa.....	6
3.4.2 O Experimento da 2ª Etapa.....	6
3.4.3 O Experimento da 3ª Etapa.....	7
3.5 Instrumentação.....	7
3.6 Avaliação da Validade.....	7
3.6.1 Validade de Conclusão.....	7
3.6.2 Validade Interna.....	8
3.6.3 Validade de Construção.....	8
3.6.4 Validade Externa.....	8
4 REFERÊNCIAS.....	9

## 1. DEFINIÇÃO DO PROBLEMA

Dentre os dois algoritmos propostos de sumarização, qual apresenta melhor desempenho?

## 2. OBJETIVOS DO EXPERIMENTO

Analisar dois algoritmos de sumarização automática de documentos tendo em vista observar como cada um se comporta diante do tamanho de conjunto de palavras apresentado, os tipos de textos e seus vocabulários específicos, além das diversas estruturas textuais. Diante disso, perceber o desempenho de cada um deles, através do tempo de execução, a porcentagem de redução das palavras e se apresenta um resultado de um texto com coerência, coesão e com uma quantidade reduzida de erros semânticos e ortográficos.

O primeiro algoritmo a ser analisado é o algoritmo de Luhn, um dos trabalhos mais importantes na área de Processamento de Linguagem Natural, que consiste em métodos estatísticos para cálculo de frequência de palavras e sua distribuição no texto, como critério de significância, gerando palavras-chaves e abstracts.

O segundo algoritmo é denominado algoritmo Marques, proposto pelos criadores deste artigo, que implementa uma nova forma de sumarização automática textual usando uma biblioteca do python chamada NLTK (*Natural Language Toolkit*), que, ao encontrar as sentenças e palavras mais importantes do texto, gera o resumo a partir delas.

A perspectiva considerada para esse experimento é a do usuário de segundo nível: aqueles que utilizam o serviço. Uma vez que é na visão deste, que busca-se entender se o algoritmo cumpriu seu propósito. Baseado na comparação das duas técnicas o usuário pode escolher aquele que lhe proporciona melhor compreensão. O experimento inicialmente pode ser realizado no laboratório através de um ambiente de desenvolvimento para simulação dos algoritmos e posteriormente no ambiente acadêmico e escolar com alunos.

Resumindo o objetivo:

**Analisar** dois algoritmos de sumarização automática: o Marques e o de Luhn,  
**Com a intenção de** compará-los  
**com respeito a** coerência dos textos  
**do ponto de vista** do usuário final  
**no contexto** para uso acadêmico e escolar.

### 3. PLANEJAMENTO DA INVESTIGAÇÃO EMPÍRICA

#### 3.1 Seleção do Contexto

Um ambiente virtualizado em laboratório com variáveis controladas será simulado através de múltiplas entradas de textos para sumarização automática dos mesmos, a fim de investigar e avaliar o funcionamento dos algoritmos de Processamento da Linguagem Natural. Para tal é selecionando um grupo de pessoas, no ambiente acadêmico e escolar, para realização da leitura dos resumos provenientes dos mesmos e indicar aquele que apresenta melhor coerência pelo método comparativo, para assim serem computadas as preferências e, a partir disso, qualificá-los.

#### 3.2 Definição das Hipóteses

Uma vez definidos o problema, os objetivos e o contexto, as hipóteses são estabelecidas e os valores de significância para os erros definidos.

**Hipótese nula (H0):** Não há diferença significativa entre os algoritmos Marques e Luhn em relação a coerência do texto.

**Hipótese alternativa (H1):** O algoritmo Marques é melhor que o algoritmo Luhn em relação a coerência do texto.

Os níveis de significância para erros são:

- P(erro do tipo I):  $\alpha = 0,01$
- P(erro do tipo II):  $\beta = 0,05$

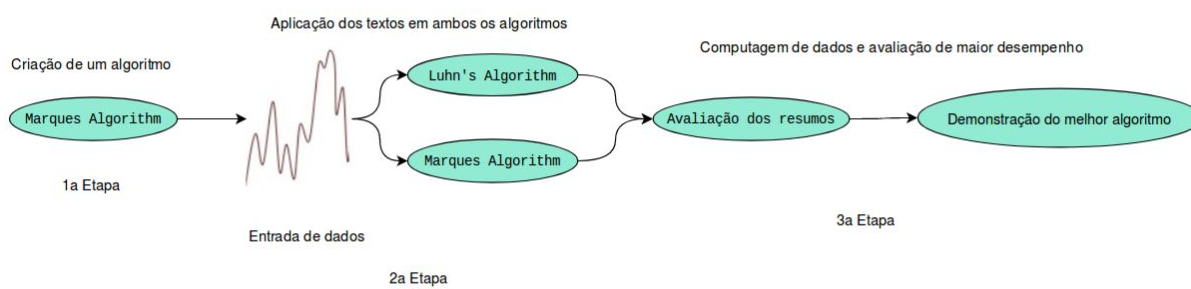
### 3.3 Seleção dos Sujeitos

Os sujeitos do experimento são os algoritmos a serem comparados, Marques e Luhn, descritos na Seção 2.

### 3.4 Design de Experimentos

O experimento consiste nessas três etapas apresentadas na figura 1

Figura 1



A primeira etapa (Figura 1) consiste na criação de um algoritmo de sumarização automática.

A segunda etapa (Figura 1) têm como objetivo realizar a sumarização automática de 10 artigos de diferentes áreas de estudos através de dois algoritmos que utilizam métodos distintos.

Na terceira etapa (Figura 1), o conjunto de resumos resultantes da segunda etapa serão utilizados para fins avaliativos por um grupo de pessoas. E por conseguinte computar a avaliação de cada pessoa e estatisticamente demonstrar o algoritmo que realizou melhor desempenho.

O cenário a ser considerado para os os experimentos é a simulação de resumos por uma máquina virtual, com capacidade de processamento e de memória RAM de 8 GB. E posteriormente a análise dos mesmos, dentro de um contexto educacional.

### 3.4.1 O Experimento da 1a Etapa

A primeira etapa consiste na criação de um algoritmo de sumarização textual, que inicia criando uma matriz do texto, separando-o em sentenças, e dentro das sentenças, em palavras. Depois disso, ele identifica palavras que possuem apenas significado sintático dentro da sentença, mas não é relevante para o sentido do texto, como “ou”, “e”, “para”, e as retira da matriz, pois, como essas palavras são frequentes, o algoritmo acabaria dando importância para as mesmas, e isso dificultaria a análise textual. Também retiramos as pontuações do texto, pois o algoritmo as trata como sendo uma palavra, e isso também atrapalha a sumarização.

Após essa limpeza de texto, o algoritmo cria uma distribuição de frequência para a matriz de palavras, para que seja descoberta quais são as mais importantes.

A partir dessa distribuição de frequência, o algoritmo seleciona, a partir de um valor inserido, o quão comum a palavra se apresenta no texto geral para posteriormente classificá-las como significantes ou não, no texto final.

### 3.4.2 O Experimento da 2a Etapa

A segunda etapa consiste em selecionar 10 artigos no google acadêmico, que incorporam diversas áreas do conhecimento e com complexidades diferentes, para realização de sumarização automática através do algoritmo de Luhn e o de algoritmo criado descrito na primeira etapa.

Foram realizados testes nas configurações dos parâmetros comuns modificáveis nos algoritmos, isto é, na quantidade de sentenças importantes, a fim de equipará-los para a etapa seguinte. Dessa forma, neste experimento foi utilizado o parâmetro de 8 sentenças importantes.

Segundo Pardo[1] para gerar um resumo eficiente deve ser extraído entre 20 a 50% do texto e 80% na taxa de compressão.

As variáveis independentes desta etapa são os textos inseridos e os algoritmos utilizados, enquanto as variáveis dependentes são os resumos gerados e a escolha da quantidade de sentenças importantes.

### 3.4.3 O Experimento da 3a Etapa

Com o intuito de avaliar a qualidade e coesão dos resumos gerados na segunda etapa, serão escolhidas 30 pessoas com diferentes graus de instrução acadêmica e será distribuído os dois resumos resultantes da inserção de um artigo em ambos os algoritmos para que um grupo de 3 pessoas que não possuem relação, escolhidas através de critérios como: conhecimento e preferência com o tema do texto, possam indicar o resumo que julgou possuir melhor coesão e coerência.

Nesse caso, os resumos tornam-se variáveis independentes enquanto a avaliação de acordo com as métricas de coerência e coesão podem ser consideradas as variáveis dependentes.

Em seguida, será criada uma tabela comparativa de acordo com o resultado da avaliação humana realizada para qualificar os dois algoritmos e assim mostrar aquele que proporciona uma melhor experiência aos leitores.

## 3.5 Instrumentação

A instrumentação dos experimentos é agrupada por hardware e software. Utilizaremos um computador (Intel Core i5-7200U Dual Core 2.5 GHz com Turbo Max até 3.1 GHz) para executar uma coleção de códigos-fonte escritos em Python 3 que implementam os algoritmos propostos de serem avaliados (Marques e LSA).

## 3.6 Avaliação da Validade

### 3.6.1 Validade de Conclusão

Os testes de hipóteses serão realizados considerando um nível de significância  $\alpha = 0,1$ , para assim dar suporte a teoria de modo a ser confiável. Diferentes tipos de textos serão aplicados aos dois algoritmos. Ambos foram implementados na mesma linguagem de programação. Entretanto, espera-se melhor resultado do algoritmo Marques devido sua natureza abstrata, uma vez que considera elementos adicionais em relação ao algoritmo de

Luhn que possui métodos extrativos. Além disso, pode-se obter resultados distorcidos devido o fato de trabalhar com um conjunto amplo de palavras, da complexidade inerente a tentativa de aproximar-se de resumos humanos, além de possuir uma análise comparativa dependente de uma avaliação do grupo experimental passível de parcialidades e julgamento limitado ao conteúdo do texto e sua relação com o mesmo.

### 3.6.2 Validade Interna

A realização dos experimentos não sofrerá interferência do ambiente de realização devido ao uso exclusivo de uma máquina virtual em que os testes consistem apenas em utilizar textos e testar o resultado da sumarização em algoritmos que representam métodos distintos. Pretende-se realizar os testes utilizando o mesmo hardware, reduzindo ameaças relativas à instrumentação, no mesmo ambiente de desenvolvimento, e em instantes próximos, para que a velocidade da internet interfira de forma mínima na análise da métrica relacionada ao tempo de execução. O tamanho do conjunto de dados, a estrutura na qual é organizada o texto, a morfologia e semântica das palavras são os fatores mais importantes que vão causar impacto no resultado final.

### 3.6.3 Validade de Construção

A proposta apresentada na terceira etapa consiste em avaliar a qualidade dos resumos resultantes dos dois algoritmos apresentados, para tal é selecionado avaliadores humanos do âmbito acadêmico e escolar para através da leitura dos mesmos indicar qual apresenta o resultado mais coerente.

### 3.6.4 Validade Externa

O experimento é dependente da avaliação humana, logo o modo de abordagem para participar do experimento, o interesse daqueles que irão avaliar os resumos e seu entendimento do assunto abordado e dos termos utilizados, podem interferir no resultado final.



#### 4 REFERÊNCIAS

- [1] Rino, Lucia Helena Machado; Pardo, Thiago Alexandre Salgueiro. A Sumarização Automática de Textos: Principais Características e Metodologias. NILC/Departamento de Computação. Universidade Federal de São Carlos, São Paulo. 2003.
- Luhn, H.P. The Automatic Creation of Literature Abstracts. IBM Journal. 1958.
- Sumy, Biblioteca contendo algoritmos de sumarização automática na linguagem Python, <https://pypi.org/project/sumy>
- Lima, Vinicius R. Utilizando processamento de linguagem natural para criar um sumarização automática de textos. 2017.