

# **Sumarização Automática: comparação entre o confiável algoritmo de Luhn e um algoritmo alternativo**

M. M. R. Silva, R. K. M. B. Elias

## **Abstract**

Será descrito no final

## **1. Introdução**

A seção apresenta a sumarização automática e a problemática envolvida. Como alternativa dois algoritmos, de Luhn e de Marques são apresentados. A diferença geral entre os dois é evidenciada. Em seguida, os objetivos da pesquisa são brevemente discutidos.

## **2. Processamento de Linguagem Natural (PLN)**

O PLN busca soluções para questões computacionais através de uma aprendizagem automática no processamento de linguagem e se dedica a propor e desenvolver modelos computacionais para a realização de tarefas que dependem da linguagem humana falada e/ou escrita como objeto primário.

## **3. Algoritmos de Sumarização**

A seção apresenta uma explicação sobre os algoritmos trabalhados nesta pesquisa, e como cada um funciona.

### **3.1 Algoritmo de Luhn**

Este é 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.

### **3.2 Algoritmo Proposto**

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.

## **4. Metodologia Avaliativa**

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 texto, os 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 de erros reduzida de erros semânticos e ortográficos.

## **5. Resultados e Validação**

Esta seção discute a metodologia usada para validar nosso trabalho e discute resultados preliminares obtidos para esse efeito, para tal são comentados os dados obtidos. Além disso, são levantados pontos que prejudicam a validação da pesquisa.

## **6. Conclusão**

Diante o que foi apresentado na seção 6, será interpretado a fim de responder à questão proposta pela pesquisa. A partir disso, descreve-se brevemente o porquê do resultado e sugere novas diretrizes para trabalhos futuros.

## **7. Referências**

Serão adicionadas na segunda versão do artigo.