

Compressão de Dados e Teoria da Informação

Lucas Silva Amorim



Universidade Federal do ABC

Título: Compressão de Dados e Teoria da Informação

Autor: Lucas Silva Amorim

Orientador: Prof^a Dr.^a Cristiane M. Sato

Trabalho de conclusão de curso apresentado como requisito parcial para obtenção do título de Bacharel em Ciências da Computação pela Universidade Federal do ABC.

Banca Examinadora:

Prof. Dr. Circulando de Souza

Universidade Federal de ..

Prof. Dr. Recirculando de Souza

Universidade Federal de ..

Santo André, 30 de agosto de 2022.

1	Introdução	7
2	Conceitos e definições fundamentais	8
2.1	Código	8
2.1.1	Código livre de prefixo	8

Opcional. Agradeço a todos os que me ajudaram na elaboração deste trabalho...

Neste lugar vai um resumo do projeto e objetivos, apresentando os principais resultados;

Conforme as normas NBR 14724:2002 da ABNT, o resumo é elemento obrigatório, constituído de uma seqüência de frases concisas e objetivas e não de uma simples enumeração de tópicos, não ultrapassando 500 palavras, seguido, logo abaixo, das palavras representativas do conteúdo do trabalho, isto é, palavras-chave e/ou descritores.

Palavras Chaves: TCC, Trabalho, Modelo

Versão em língua estrangeira do resumo. Obrigatório, pela ABNT. O título é ABSTRACT, em inglês, RESUMEN, em espanhol castelhano, e RÉSUMÉ, em francês. Sugerimos Inglês.

Keywords: aubergine,carrot, radish

Esta pesquisa pretende mostrar que [...] através de [...] conforme concepções apresentadas por [...] . Para isso, articulamos o conceito de [...] com o conceito de [...] . Fizemos pesquisas de recepção conforme [...] . Articulamos os resultados a partir de idéias de [...] . “Neste primeiro parágrafo você deve deixar completamente claro o que pretende com o trabalho. A introdução é redigida depois de escrito todo o trabalho porque, no decorrer da pesquisa, algumas coisas podem ser modificadas em relação ao projeto original”. “Depois, em vários parágrafos, você deve falar sobre a problematização, a contextualização histórica, a revisão bibliográfica, os objetivos, a justificativa, a metodologia. As conclusões, evidentemente, devem ficar no capítulo Considerações Finais, para que o leitor não perca o interesse pelo seu trabalho ?. Toda a introdução é feita sem subtítulos, em texto normal”.

2 CONCEITOS E DEFINIÇÕES FUNDAMENTAIS

Este capítulo apresenta algumas definições e conceitos fundamentais para o entendimento das técnicas de compressão que serão discutidas em capítulos posteriores.

2.1 Código

Um **código** C mapeia uma **mensagem** M para um conjunto de **palavras código** W , onde $M \subset \alpha$ e $W \subset \beta$, isto é, $C : \alpha \rightarrow \beta$. Chamamos de α o **alfabeto de origem** e β o **alfabeto de palavras código**. Os elementos dos alfabetos α e β podem ter um comprimento fixo ou variável. Códigos nos quais os alfabetos possuem um comprimento fixo são chamados de **códigos de comprimento fixo**, enquanto os que possuem alfabetos de comprimento variáveis são chamados **códigos de comprimento variável**. Provavelmente o exemplo mais conhecido de código de comprimento fixo é o código ASCII, que mapeia 64 símbolos alfa-numéricos (ou 256 em sua versão estendida) para palavras código de 8 bits. Todavia, a compressão de dados utiliza apenas códigos de comprimento variável, mas especificamente códigos que variam o comprimento de acordo com a probabilidade associada à mensagem.

Um código é **distinto** se pode ser representado como uma função **bijetora**, i.e., $\forall m_1, m_2 \in \alpha, C(m_1) \neq C(m_2)$. Dado um código C , com $\alpha := m_1, m_2, \dots, m_n$ e $\beta := w_1, w_2, \dots, w_n$, chamamos C **unicamente decodificável** quando $C(S) = w_1 w_2 \dots w_n \leftrightarrow C^{-1}(W) = m_1 m_2 \dots m_n$.

O processo de transformar uma mensagem em palavras código é conhecido como **codificação**, e o algoritmo que executa tal processo é chamado *encoder*. A decodificação realiza a operação inversa, revertendo as palavras código para a mensagem original.

2.1.1 Código livre de prefixo

Um **código livre de prefixo** é um código unicamente decodificável C' tal que nenhuma palavra código de C' é prefixo de outra palavra código, por exemplo, $\alpha := 1, 01, 000, 001$. Códigos livres de prefixo podem ser *decodificados instantaneamente*,

isto é, podemos decodificar uma mensagem sem precisar verificar o início da mensagem seguinte.

Teorema 2.1 *Todo código livre de prefixo é unicamente decodificável.*

Demonstração: Seja C um código livre de prefixo e $S_n = s_1 \dots s_n$ uma mensagem codificada por C . Vamos provar por indução que o teorema é verdadeiro para todo $n \in \mathbb{Z}_+$

Casos base: Quando $n = 1$, a mensagem S só possui uma palavra código, logo é unicamente decodificável. Se $n = 2$, então S possui uma palavra código s_1 que não pode ser prefixo de s_2 (pela própria definição de códigos livres de prefixo), o que claramente significa que S é unicamente decodificável.

Passo indutivo: Seja $k \in \mathbb{Z}_+$, e suponha por hipótese de indução que o teorema vale para $n \leq k$. Como S_{k+1} é livre de prefixo, existe um prefixo de S_{k+1} , $S_j = s_1 \dots s_j$ (com $j \leq k + 1$) que é unicamente decodificável (dado que ela não pode ser prefixo de nenhuma outra). a mensagem $S'_{k+1} = s_{j+1} \dots s_{k+1}$ ainda é uma concatenação decodificável e $|S'_{k+1}| \leq |S_{k+1}|$, o que significa que por hipótese de indução S'_{k+1} é unicamente decodificável. Como $S_{k+1} = S_j S'_{k+1}$, segue que S_{k+1} é unicamente decodificável. \square

REFERÊNCIAS BIBLIOGRÁFICAS

- [HL] HIRSCHBERG, D.S; LELEWER D.A; *Data compression*, Computing Surveys 19.3, 1987.
- [Ble] BLELLOCH G.E; *Introdution to Data Compression*, Carnegie Mellon, 2013