

MÉTODOS DE COMPACTAÇÃO E
DESCOMPACTAÇÃO DE INFORMAÇÃO

JOZEMAR PEREIRA DOS SANTOS

"TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO E PESQUISA DO CENTRO DE CIÊNCIAS E TECNOLOGIA DA UNIVERSIDADE FEDERAL DA PARAÍBA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS
(M.Sc.)

Aprovada por:

Orion de Oliveira Silva

Prof. ORION DE OLIVEIRA SILVA-M.Sc.

Presidente

M. Chhuf

Prof. MUTHUVEERAPPAN CHANDRASHEKAR-Ph.D.

M V Bhat

Prof. MATTU VISHNUMOORTHY BHAT-Ph.D.

CAMPINA GRANDE
ESTADO DA PARAÍBA - BRASIL
MARÇO - 1975



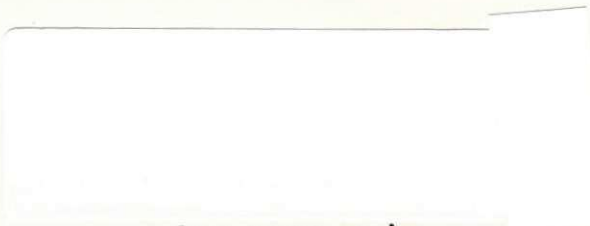
S237m Santos, Jozemar Pereira dos.
Métodos de compactação e descompactação de informação /
Jozemar Pereira dos Santos. - Campina Grande, 1975.
71 f.

Dissertação (Mestrado em Ciências) - Universidade
Federal da Paraíba, Centro de Ciências e Tecnologia, 1975.
"Orientação : Prof. M.Sc. Orion de Oliveira Silva".
Referências.

1. Dados - Compactação e Descompactação. 2. Informação -
Compactação e Descompactação. 3. Dados - Armazenamento. 4.
Dissertação - Ciências. I. Silva, Orion de Oliveira. II.
Universidade Federal da Paraíba - Campina Grande (PB). III.
Título

CDU 004.627(043)

i



Aos meus pais,
à minha filha
Janayna

AGRADECIMENTOS

O autor expressa agradecimentos ao Professor Orion de Oliveira Silva, pela sugestão, orientação e constante assistência a este trabalho;

Agradece, também, aos operadores do IBM 1130;

A CAPES (Coordenação do Aperfeiçoamento de Pessoal de Nível Superior);

Ao CCT (Centro de Ciências e Tecnologia da Universidade Federal da Paraíba);

Ao BNDE (Banco Nacional de Desenvolvimento E conômico).

R E S U M O

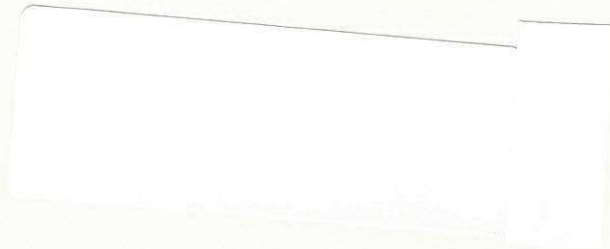
Eficientes métodos de compactação e descompactação de informação estão sendo desenvolvidos, dada a larga tendência acerca de armazenamento de grandes programas, grandes textos, banco de dados etc.

Este trabalho tem como objetivo o desenvolvimento e a implementação no IBM 1130 de dois métodos de compactação e descompactação de informação. Também, neste trabalho é feita uma análise estatística de ambos os métodos.

ABSTRACT

Efficient information packing and unpacking methods are being developed due to the tendency regarding storage of large programs, large texts, data base etc.

This work aims to develop and implement two methods of information packing and unpacking using the IBM 1130. The statistical analysis of both methods is also presented.



ÍNDICE

DEDICATÓRIA	i
AGRADECIMENTOS.....	ii
RESUMO	iii
ABSTRACT	iv
ÍNDICE	v
CAPÍTULO I - INTRODUÇÃO	1
CAPÍTULO II - PRIMEIRO MÉTODO DE COMPACTAÇÃO E DESCOMPACTAÇÃO DE INFORMAÇÃO	3
2.1 - Técnica de compactação utilizada no 1º método	3
2.1.a - Leitura da lista de símbolos de entrada ..	4
2.1.b - Construção do dicionário de símbolos	5
2.1.c - Codificação dos símbolos usando o dicionário de símbolos	6
2.1.d - Compressão dos dados de entrada	6
2.2 - A compressão propriamente dita	7
2.3 - Otimização da compressão	8
2.4 - Técnica de descompactação utilizada no 1º método	10
2.5 - Fluxograma e resultados obtidos	10
CAPÍTULO III- SEGUNDO MÉTODO DE COMPACTAÇÃO E DESCOMPACTAÇÃO DE INFORMAÇÃO	11
3.1 - Técnica de compactação utilizada no 2º método	11

3.1.a - Leitura da lista de símbolos de entrada...	12
3.1.b - Construção do dicionário de símbolos	12
3.1.c - Codificação dos símbolos usando o dicionário de símbolos	14
3.1.d - Compressão dos dados de entrada	14
3.2 - A compressão propriamente dita	14
3.3 - Técnica de descompactação utilizada no 2º método.....	17
3.4 - Fluxograma e resultados obtidos	18
CAPÍTULO IV - ANÁLISE ESTATÍSTICA	19
4.1 - Estatística para número de palavras requeridas e entropia	19
4.2 - Estatística do tempo de CPU	21
CAPÍTULO V - CONCLUSÕES	23
REFERÊNCIAS BIBLIOGRÁFICAS	25
APÊNDICE A	26
APÊNDICE B	49

CAPÍTULO I

INTRODUÇÃO

Atualmente com a evolução dos sistemas de computadores grandes programas, grandes blocos de textos etc, podem ser armazenados. Mesmo assim, ainda há necessidade de técnicas de compactação de informação que apresentem significantes salvamentos de espaços de memória. Dispositivos de direto acesso em "online" estão sendo utilizados e eficientes técnicas de compactação de informação são de interesse.

Frequentemente, a gravação de textos contendo muitas listas de símbolos brancos tomam grande quantidade de espaço de memória. O resultado obtido é que muito espaço é desperdiçado por métodos comuns de armazenamento.

Um dos métodos de solucionar este problema é armazenar somente os símbolos não-brancos. Existem técnicas que, além de armazenar os símbolos não-brancos, reconhecem listas de símbolos brancos e trata-os especialmente.

WYLBUR [8] "a text-editing system" armazena os símbolos em SEGMENTOS, cada qual podendo descrever até 15 brancos seguidos de até 15 símbolos não-brancos.

Outros métodos usam os símbolos brancos somente como delimitadores.

Neste trabalho estão desenvolvidos dois métodos de compactação e descompactação de informação. Os mesmos são de fácil entendimento e rápida implementação.

Ambos os métodos envolvem especial manuseio de listas de símbolos brancos e a codificação dos símbolos restantes em números de ponto fixo único.

Os métodos aqui sugeridos funcionam economicamente (relativo ao espaço físico de memória) quando os diferentes caracteres de um grande conjunto de dados a ser armazenado são distribuídos em ordem decrescente de ocorrência prevista [3]. Não obstante, em alguns casos, para até um determinado número de diferentes caracteres com distribuição uniforme de ocorrência, tais métodos oferecem economia de memória.

Efetuuou-se a implementação de ambos os métodos no IBM 1130. Uma análise aproximada do tempo de CPU gasto pelos métodos para compactar e descompactar determinados blocos de dados foi efetuada, dado que o IBM 1130 não possui marcador de tempo de CPU. Os resultados obtidos na referida análise, apresentados neste trabalho, traduzem razoavelmente a rapidez de implementação (considerando-se que o IBM 1130 não é um computador muito veloz).

CAPÍTULO II

PRIMEIRO MÉTODO DE COMPACTAÇÃO E DESCOMPACTAÇÃO DE INFORMAÇÃO

Neste capítulo está desenvolvido um método que é usado para armazenar grande quantidade de dados, os quais podem ser programas, bancos de dados, textos em determinada língua etc. O método envolve especial manuseio de listas de símbolos brancos, técnicas de compactação e descompactação de dados.

O resultado do estudo baseia-se no método implementado por BRUCE HAHN [3] no IBM 360/75. O mesmo foi implementado no IBM 1130 e, como a unidade básica de armazenamento do IBM 1130 é a palavra (de 16 bits) algumas modificações foram introduzidas.

O método consiste em:

- (1) Uma lista de símbolos brancos concatenados é codificada como número de ponto fixo negativo, cujo valor absoluto é o número de brancos da referida lista.
- (2) Os símbolos não-brancos e brancos não-concatenados são codificados em grupos de tamanho fixo como números de ponto fixo único.

2.1 - Técnica de compactação utilizada no 1º método

2.1.a - Leitura da lista de símbolos de entrada

2.1.b - Construção do dicionário de símbolos (DS)

2.1.c - Codificação dos símbolos usando o DS

2.1.d - Compressão dos dados de entrada.

2.1.a - Leitura da lista de símbolos de entrada

A lista de símbolos de entrada a ser lida (ex.: banco de dados, grandes textos etc.) é constituída de dados alfa-numéricos. Durante o processo de compactação operações aritméticas e testes com estes dados são efetuados. Portanto, faz-se uso da especificação Aw que possibilita tais manipulações. Desde que operações aritméticas são efetuadas com variáveis, contendo dados alfa-numéricos, elas deverão ser executadas a penas com variáveis do tipo inteiro [9].

Hã, porém, uma restrição quanto ao valor de w, isto é, o número de caracteres que compõe um certo dado, por variável. Cada variável inteira utiliza uma palavra no IBM 1130 e esta só pode armazenar até dois caracteres (representados internamente no código EBCDIC de 8 bits). Portanto, a leitura da lista de símbolos de entrada deverá ser efetuada no formato A1 ou A2.

No formado A1 cada carácter lido ocupa uma palavra, enquanto que, no formato A2 cada palavra é preenchida com dois caracteres lidos. O formato A2 utiliza em média metade do número de palavras que o formato A1, para armazenar uma lista de símbolos de entrada que é lida.

No entanto, na maioria dos casos, devido a complexidade de quando do uso do formato A2 para construção do dicionário de símbolos, utilizado no processo de codificação e descodificação, é feito uso do formato A1 para a leitura da lista de símbolos de entrada.

Casos que poderão ocorrer:

- I) A lista a ser lida excede a capacidade de armazenamento da memória central. Então, considera-se a lista por partes, ou seja, em sub-listas e a leitura é processada. (*)
- II) A lista a ser lida não excede a capacidade de armazenamento da memória central. Neste caso, ela é lida e compactada automaticamente. (**)

2.1.b - Construção do dicionário de símbolos (DS)

A cada lista de entrada lida é feita a construção de um DS, usando o número de ocorrência de cada símbolo existente na lista. No caso (1) do ítem anterior em que a lista é dividida em sub-listas pode-se ter, como no caso (II), um único DS. Entretanto, é necessário que se trate a ocorrência dos símbolos em cada sub-lista como se estivesse considerando a própria lista de entrada em seu todo.

A construção do DS é feita como segue:

- i) Inicialmente, considera-se duas tabelas zeradas de tamanho fixo a 126^(***). Faz-se uma "varredura" na

(*) Tem-se a construção de vários dicionários, um para cada sub-lista considerada.

(**) Tem-se a construção de um único dicionário de símbolos.

(***) O comprimento das tabelas necessárias para a construção do DS foi tomado como sendo igual a 126 por ser este número o maior quociente obtido na divisão por 256 dos correspondentes inteiros (dos símbolos lidos no formato A1).

lista de entrada e a cada símbolo encontrado é feita uma divisão por 256 e incrementa-se de 1 a posição da 1a. tabela, cujo índice corresponde ao quociente obtido, anotando a ocorrência do símbolo. Em paralelo, o símbolo é colocado na 2a. tabela na mesma posição.

- ii) O DS é, então, construído com os símbolos em ordem decrescente de ocorrência prevista e pelo qual utilizou-se o BUBBLE SORT [4].

A forma final do DS (que é única e é necessária para o processo de decodificação) não é conhecida até que toda a lista de entrada tenha sido completamente varrida.

2.1.c - Codificação dos símbolos usando o DS

Os símbolos são codificados por suas posições no DS. Depois, quando a decodificação toma lugar, as posições numéricas dos símbolos são usadas em conjunto com o DS para produzir o texto de entrada original.

2.1.d - Compressão dos dados de entrada

Os símbolos não-brancos e brancos não-concatenados são codificados em grupos de comprimento N. Cada símbolo é pesquisado no DS e sua posição é anotada.

Se uma lista de símbolos brancos concatenados é menor do que N, então o símbolo branco desta lista é tratado como branco não-concatenado. Caso contrário, codifica-se a lista de símbolos brancos como um número de ponto fixo negativo, cu

jo valor absoluto é o número de brancos da lista.

2.2 - A compressão propriamente dita

Seja B o comprimento do dicionário primário e P_i a posição do i -ésimo símbolo de um dado grupo, ($1 \leq i \leq N$) no DS e ($1 \leq P_i \leq B-1$).

A B -ésima posição é reservada para um carácter de fuga que permite a extensão do dicionário.

O grupo de posições P_1, P_2, \dots, P_N é codificado como um número de ponto fixo $P_1 * B^{N-1} + P_2 * B^{N-2} + \dots + P_N$

Se a B -ésima posição não for reservada, o valor $B * B^i$ pode aparecer, causando ambiguidade na codificação. O carácter de fuga significa que o símbolo seguinte tem a posição na faixa $B+1$ até $2B-1$. Mais do que 1 carácter de fuga pode ser usado para extensão do comprimento do dicionário.

Seja p a posição de dado símbolo no DS, então, o símbolo é codificado como $\lfloor p/B \rfloor$ caracteres de fuga seguido de mod (p, B) , onde $\lfloor y \rfloor$ significa parte inteira de y .

O carácter de fuga é codificado como ZERO.

A seguir, é mostrado um exemplo de compactação:

Exemplo 1.1 - Seja $B=8$ comprimento do dicionário primário e $N=4$, ou seja, os símbolos são codificados em grupos de 4.

Considere-se os símbolos a serem codificados, tendo as posições 5,7,10,15,25 e 9 no DS. Estes símbolos são codificados em três números de ponto fixo:

$$O_1: \text{ tem o valor } 5*8^3 + 7*8^2 + 0*8 + 2*8^0 = 3010$$

$$O_2: \text{ tem o valor } 0*8^3 + 7*8^2 + 0*8 + 0*8^0 = 448$$

$$O_3: \text{ tem o valor } 0*8^3 + 1*8^2 + 0*8 + 1*8^0 = 65$$

Observe-se que os símbolos 10,15,25 e 9 são codificados como (0,2), (0,7), (0,0,0,1) e (0,1), respectivamente.

2.3 - Otimização da compressão

O usuário do sistema tem controle sobre os valores de B e N. Estes valores são selecionados para minimizar a quantidade média do número de bits/informação necessária para codificar o texto de entrada.

A cada valor de N existe um ótimo valor para B. Este ótimo valor é dado pelo maior inteiro positivo tal que $B^N - 1 \leq L$ onde $L = 2^{15} - 1$ (maior inteiro que pode ser armazenado numa palavra do IBM 1130) visto que o maior número de ponto fixo a ser produzido é $(B - 1)*B^{N-1} + \dots + (B-1) = B^N - 1$.

O valor de B é obtido pela equação

$B = \lfloor (2^{15} - 1)^{1/N} \rfloor$, escolhido de maneira que para um dado valor de N não haja OVERFLOW.

A tabela 1 a seguir mostra os ótimos valores obtidos para B (no IBM 1130) dados os valores de N.

TABELA 1

Valores de N	Valores de B
4	13
5	7
6	5
7	4
8	3

Com respeito ao exemplo 1.1, apresentado anteriormente, tomando-se o valor de $B = 13$, a compressão dos símbolos é efetuada em apenas dois números de ponto fixo e sendo portanto otimizada:

$$0 \text{ 1º tem o valor } 5*13^3 + 7*13^2 + 10*13 + 0*13^0 = 12298$$

$$0 \text{ 2º tem o valor } 2*13^3 + 0*13^2 + 12*13 + 9*13^0 = 4559$$

Observe-se, agora, que os símbolos 15 e 25 são codificados como (0,2) e (0,12), respectivamente.

É importante frisar que a utilização de $B = 8$ para $N = 4$, no exemplo 1.1, produz um menor esforço computacional do que quando do uso de $B = 13$, apesar de que neste caso a compressão é otimizada. Entretanto, devido ao custo operacional bastante dispendioso, se se dispõe de grandes dispositivos auxiliares de memória a perda de posições de memória (como é o caso de $B = 8$) é compensada pelo ganho de tempo de processamento.

2.4 - Técnica de descompactação utilizada no 1º método

O processo de decodificação é feito de imediato, usando-se as posições numéricas em conjunto com o DS, produzindo a lista de entrada original.

- i) Se o número de ponto fixo produzido é maior do que zero, divide-se o mesmo pela potência B^{N-1} , toma-se o resto dividindo por B^{N-2} e, assim, sucessivamente até B^0 , obtendo-se com isto as posições numéricas dos símbolos correspondentes no DS.

Quando no processo de divisão é encontrado zeros (caracteres de fuga) seguido de valor não-nulo então, o seguinte cálculo é efetuado:

Posição Numérica = $J * B + \text{valor não-nulo}$, onde $J = N^\circ$ de zeros. Obtendo-se, desta maneira, a posição numérica no DS.

- ii) Se o número de ponto fixo produzido, digamos K , for menor do que zero então, a lista de saída é incrementada com $I = -K$ símbolos brancos.

2.5 - Fluxograma e resultados obtidos

Um fluxograma do programa que descreve este método e os resultados obtidos no IBM 1130, estão no Apêndice A.

CAPÍTULO III

SEGUNDO MÉTODO DE COMPACTAÇÃO E DESCOMPACTAÇÃO DE INFORMAÇÃO

Neste capítulo está desenvolvido um outro método de compactação e descompactação de informação, obtido através de pesquisas e leituras em [1], [2]. Também, contou-se com auxílio da técnica utilizada pelo método descrito no capítulo 1. O mesmo foi implementado no IBM 1130 e, tal como o anterior, envolve especial manuseio de listas de símbolos brancos e reduz a quantidade de memória a ser utilizada para poder armazenar todo um bloco de dados.

Basicamente, o método consiste em:

- (1) Uma lista de símbolos brancos concatenados é codificada como número de ponto fixo negativo, cujo valor absoluto é o número de brancos da lista.
- (2) Os símbolos não brancos e brancos não-concatenados são codificados através de bits "zeros" e "uns". A palavra (de 16 bits) é, então, preenchida com os códigos binários, produzindo números de ponto fixo único correspondentes.

3.1 - Técnica de compactação utilizada no 2º método

3.1.a - Leitura da lista de símbolos de entrada

3.1.b - Construção do dicionário de símbolos (DS)

3.1.c - Codificação dos símbolos usando o DS

3.1.d - Compressão dos dados de entrada.

3.1.a - Leitura da lista de símbolos de entrada

A leitura da lista de entrada é efetuada da mesma maneira que no exposto do cap. 2, ítem (2.1.a). Também, neste método, utilizou-se o formato A1.

3.1.b - Construção do dicionário de símbolos (DS)

O dicionário de símbolos é constituído por uma matriz de ordem 10x10 cujos elementos são dados pelos símbolos (não repetidos) existentes na lista de entrada.

A construção é efetuada do seguinte modo:

- 1) Exposto da mesma maneira que o ítem (i) para construção do DS no 1º método.
- 2) É feita a ordenação dos símbolos em ordem decrescente de ocorrência prevista, através do BUBBLE SORT [4].
- 3) Os símbolos que constituirão o DS são codificados através de bits "zeros" e "uns". A codificação binária a ser utilizada apresenta as seguintes características:
 - 3.1) São formadas sequências binárias de comprimentos variáveis a partir de 2, inicializadas com bit-0 e terminadas com bit-1 e sem intercalamento de bit-0 numa subsequência de "uns" e vice-versa.

3.11) A sequência binária que representa o código de um dado símbolo traduz a posição que este ocupa na matriz DS através do número de zeros e uns da mesma. (*)

A codificação binária é, então, gerada da seguinte maneira:

Ocorrência prevista dos símbolos	Código binário
1	01
2	001
3	011
4	0001
5	0011
6	0111
7	00001
8	00011

14	001111
15	011111
16	0000001
17	0000011
18	0000111

(*) Tomou-se a ordem da matriz que constitui o DS como sendo 10x10, visto que, os símbolos reconhecidos pela impressora do IBM 1130 são em número de 47. E, de acordo com a geração da codificação binária: como maior número de zeros o código 00000000001 e maior número de uns o código 0111111111.

3.1.c - Codificação dos símbolos usando o DS

Os símbolos são codificados por bits "zeros" e "uns" os quais traduzem as posições utilizadas no DS. Quando a decodificação toma lugar tais posições são usadas em conjunto com o DS, produzindo a lista de entrada original.

3.1.d - Compressão dos dados de entrada

A compressão dos símbolos não-brancos e brancos não-concatenados é feita utilizando seus códigos binários. A técnica que é utilizada não considera os símbolos em grupos de comprimento fixo, tal como no método anterior.

Se uma lista de símbolos brancos concatenados tem comprimento menor do que 8 então o símbolo branco desta lista é tratado como branco não-concatenado. Caso contrário, a lista de brancos é codificada como um número de ponto fixo negativo, cujo valor absoluto corresponde ao número de brancos da lista.

3.2 - A compressão propriamente dita

A compressão dos símbolos não-brancos e brancos não-concatenados é feita, preenchendo a palavra com os códigos binários dos símbolos a serem compactados. E, isto é feito da seguinte maneira:

- 1) Reserva-se o bit de ordem zero para o sinal "+".
- 2) Os bits restantes, de 1 a 15, são utilizados pelos códigos dos símbolos os quais,

são armazenados da esquerda para a direita na palavra (**).

- 3) Ao ser preenchido os 15 bits, então, é armazenado o número de ponto fixo correspondente que contém os símbolos compactados.
- 4) Caso o código de um dado símbolo não caiba na palavra, parte dele é colocado na mesma e o restante do código na palavra seguinte, sem que haja perda de bits.

A seguir, é apresentado um exemplo de compressão:

Exemplo 2.1 - Seja a palavra UNIVAC pertencente a um texto de entrada no qual deseja-se compactar. Suponha que os seguintes códigos binários foram criados:

U	→	0001
N	→	001111
I	→	011
V	→	000001111
A	→	001
C	→	00011

A compactação é efetuada, utilizando-se duas palavras:

(**) Sendo "branco" o símbolo mais frequente, sua codificação binária é dada pelo menor código, ou seja 01. Como os 15 bits da palavra são usados então, até 7 símbolos brancos podem ser colocados na mesma, o que mostra a figura abaixo:

		7 brancos												
palavra	+	0	1	0	1	0	1	0	1	0	1	0	1	A

o bit A não é perdido.

A 1a. é dada por

	U				N					I		V			
+	0	0	0	1	0	0	1	1	1	1	0	1	1	0	0

 produzindo o número de ponto fixo 2540 que contém os códigos U, N, I e parte do código de V.

Enquanto que a 2a. é dada por

	(V)							A		C					
+	0	0	0	1	1	1	1	0	0	1	0	0	0	1	1

 produzindo o número de ponto fixo 7747, que contém o restante do código de V e os códigos A e C.

Desde que os códigos são gerados em grupos de comprimentos variáveis, de acordo com a ordem decrescente de ocorrência dos diferentes caracteres, então, nesta situação a codificação binária aqui sugerida funciona economicamente.

Se uma determinada lista de dados de entrada contiver até no máximo 18 diferentes tipos de caracteres e, tendo os mesmos distribuição uniforme de ocorrência, ainda, neste caso, a codificação de comprimento variável funciona economicamente.

No caso em que mais de 18 diferentes caracteres são usados e possuem distribuição uniforme de ocorrência, a codificação binária aqui desenvolvida não é satisfatória. Então, tomando-se um espaço binário de comprimento fixo-6, o qual pode expressar os 47 diferentes caracteres possíveis ($2^6 - 1 = 63$), uma compactação eficiente é produzida. Mesmo que a distribuição de ocorrência dos caracteres não seja uniforme, esta codificação de comprimento fixo-6 produz economia de espaço de memória (no caso de mais de 18 caracteres).

Portanto, o problema surgido é saber até que ponto a codificação de comprimento variável apresenta melhores resultados do que a de comprimento fixo-6 e vice-versa, no caso em que mais de 18 diferentes caracteres são utilizados. Pa-

ra isto, uma análise, da distribuição de ocorrência dos caracteres, é necessária para determinar-se o "ponto crítico", ou seja, ponto de intercâmbio entre as duas codificações e pelo qual não foi efetuado estudos.

3.3 - Técnica de descompactação utilizada no 2º método

O processo de descompactação é feito de imediato usando-se as codificações binárias em conjunto com o DS para produzir o bloco de dados de entrada.

- i) Se o número de ponto fixo produzido é maior do que zero então, usa-se a transformação para base binária e a partir do bit de posição mais baixa é feita a contagem do número de "zeros" e "uns" para identificação dos códigos binários. Obtendo-se, desta maneira, os símbolos correspondentes na matriz que constitui o DS.
- ii) Se o número de ponto fixo, digamos K, é menor do que zero então, a lista de saída é incrementada com $I = -K$ símbolos brancos.

Exemplo 2.2 - Sejam os números de ponto fixo 2540 e 7747, produzidos no exemplo 2.1, a serem decodificados:

Tem-se que 2540 e 7747 correspondem aos números binários

$\underbrace{0001}_{(3,1)} \underbrace{001111}_{(2,4)} \underbrace{011}_{(1,2)} \underbrace{00}_{(5,4)} \text{ e } \underbrace{0001111}_{(2,1)} \underbrace{001}_{(3,2)} \underbrace{00011}_{(3,2)}$

e utilizando estas posições no DS, obter-se-á UNIVAC.

Note-se que o código binário correspondente ao símbolo V só é conhecido quando da transformação do número de ponto fixo 7747 para a base binária.

3.4 - Fluxograma e resultados obtidos

Um fluxograma simplificado do programa que descreve este método e os resultados obtidos no IBM 1130, estão no Apêndice B.

UNIVERSIDADE FEDERAL DA PARAÍBA
Pró-Reitoria Para Assuntos do Interior
Coordenação Setorial de Pós-Graduação
Rua Aprígio Veloso, 832 - Tel (283) 321 7222-R 355
58.100 - Campina Grande - Paraíba

CAPÍTULO IV

ANÁLISE ESTATÍSTICA

Neste capítulo é apresentado uma estatística da quantidade média de bits/informação, assim como o número de palavras requeridas pelos dois métodos para compactar determinado bloco de dados de entrada. Também é feita uma análise do tempo gasto para compactação e descompactação.

4.1 - Estatística p/número de palavras requeridas e entropia

As estatísticas apresentadas nas tabelas 2 e 3 a seguir, foram obtidas, considerando-se três tipos de blocos de dados, cada qual com determinado número de informações a serem compactadas. Nos apêndices A e B, encontram-se os resultados obtidos pelos programas utilizados nos dois métodos, os quais foram transcritos nas referidas tabelas.

TABELA 2

Bloco de dados	Inf.lidas	Nº de pal.requer.		Nº de bits/Infom.	
		1º met.	2º met.	1º met.	2º met.
Texto em Português	12.000	3.302	3.153	4,4026	4,2040
Texto em Fortran	24.336	2.332	2.347	1,5332	1,5430
Banco de Dados	50.050	5.732	5.720	1,8324	1,8285

A estatística na tabela acima apresentada para o primeiro método foi obtida, considerando-se os valores $B = 13$ comprimento do dicionário primário, e $N = 4$ tamanho dos grupos em que os símbolos são codificados, respectivamente.

TABELA 3

Bloco de Dados	Inf.lidas	Nº de pal.requer.		Nº de bits/Inf.	
		1º met.	2º met.	1º met.	2º met.
Texto em Português	12.000	3.341	3.034	4,4546	4,0453
Texto em Fortran	24.336	2.550	2.381	1,6765	1,5654
Banco de Dados	50.050	5.695	5.695	1,8205	1,8205

A estatística dada na tabela 3 foi obtida, considerando-se a construção de um único DS, um para cada bloco de dados de entrada. E, pelo resultado que se nos apresenta, observa-se-á, em alguns casos, o crescimento do número de palavras requeridas para compactação e entropia.

Se se estiver interessado em economizar posições de memória na compactação de um grande conjunto de dados é mais vantajoso a construção de vários dicionários de símbolos, o que é feito no caso da estatística dada na tabela 2.

No caso de armazenamento de um grande texto em determinada língua, como por exemplo: Inglês, Português etc. ; no qual já foi feita estatística para a probabilidade de ocorrência dos símbolos existentes é mais viável o uso de um único DS, se não se levar em consideração a perda de algumas posições de memória mas, sim, a economia de tempo de processamento o qual é bastante dispendioso.

4.2 - Estatística de tempo de CPU

A tabela 4 a seguir, mostra a estatística de tempo de CPU no IBM 1130 para a compactação e descompactação de três blocos de dados utilizados.

TABELA 4

Bloco de dados	Número de informações lidas	1º método tempo em seg.aprox.		2º método tempo em seg.aprox.	
		Compact.	Descomp.	Compact.	Descomp.
Texto em Português	2.000	11,5	9,3	10	9
Texto em Fortran	1.872	5	3,2	3	3,25
Banco de Dados	2.002	4	2	3,27	4,6

A estatística do tempo de CPU apresentada na tabela 4 acima, foi feita de maneira simples através de cronômetro, não sendo, portanto, muito exata, mas nos dá uma aproximação regular do tempo tomado em segundos para CPU.

CAPÍTULO V

CONCLUSÕES

Dada a necessidade de eficientes métodos de compressão e armazenamento de dados, os dois métodos aqui apresentados foram desenvolvidos com o propósito de minimizar a quantidade de memória necessária para armazenar um grande conjunto de dados.

As estatísticas apresentadas e os testes efetuados para diversos conjunto de dados, mostram a eficiência de ambos os métodos desenvolvidos.

Não foram feitas análises estatísticas com outros métodos existentes, em vista da não implementação dos mesmos, porém, os resultados obtidos por [3] garantem a eficiência dos métodos aqui analisados.

Tanto no primeiro como no segundo método, a otimização da compressão para armazenamento de um conjunto de dados depende das necessidades dos usuários do sistema.

Espera-se que os dois métodos desenvolvidos venham beneficiar os diversos Centros de Processamento de Dados, os quais, frequentemente, estão envolvidos com o problema de criação de bancos de dados.

Um tema sugerido para pesquisa, referente à compactação e descompactação de informação, será encontrar um código ideal para os dados (subconjunto de um banco de dados) de forma que estes dados codificados sejam ajustados melhor

por uma função. Desta forma, armazenar-se-ia somente a tabela de códigos e os parâmetros da função ajustada, que poderia ser - por exemplo - uma função polinomial.

A função não precisa passar pelos pontos (dados codificados) mas, sim, por uma faixa de tolerância pré-estabelecida.

Vários estudos foram efetuados, entretanto, somente resultados particulares foram obtidos, de forma que não foram citados no presente trabalho.

UNIVERSIDADE FEDERAL DA PARAÍBA
Pró-Reitoria Para Assuntos do Interior
Coordenação Setorial de Pós-Graduação
Rua Aprígio Veloso, 832 - Tel (083) 321 7222-R 355
58.100 - Campina Grande - Paraíba

REFERÊNCIAS BIBLIOGRÁFICAS

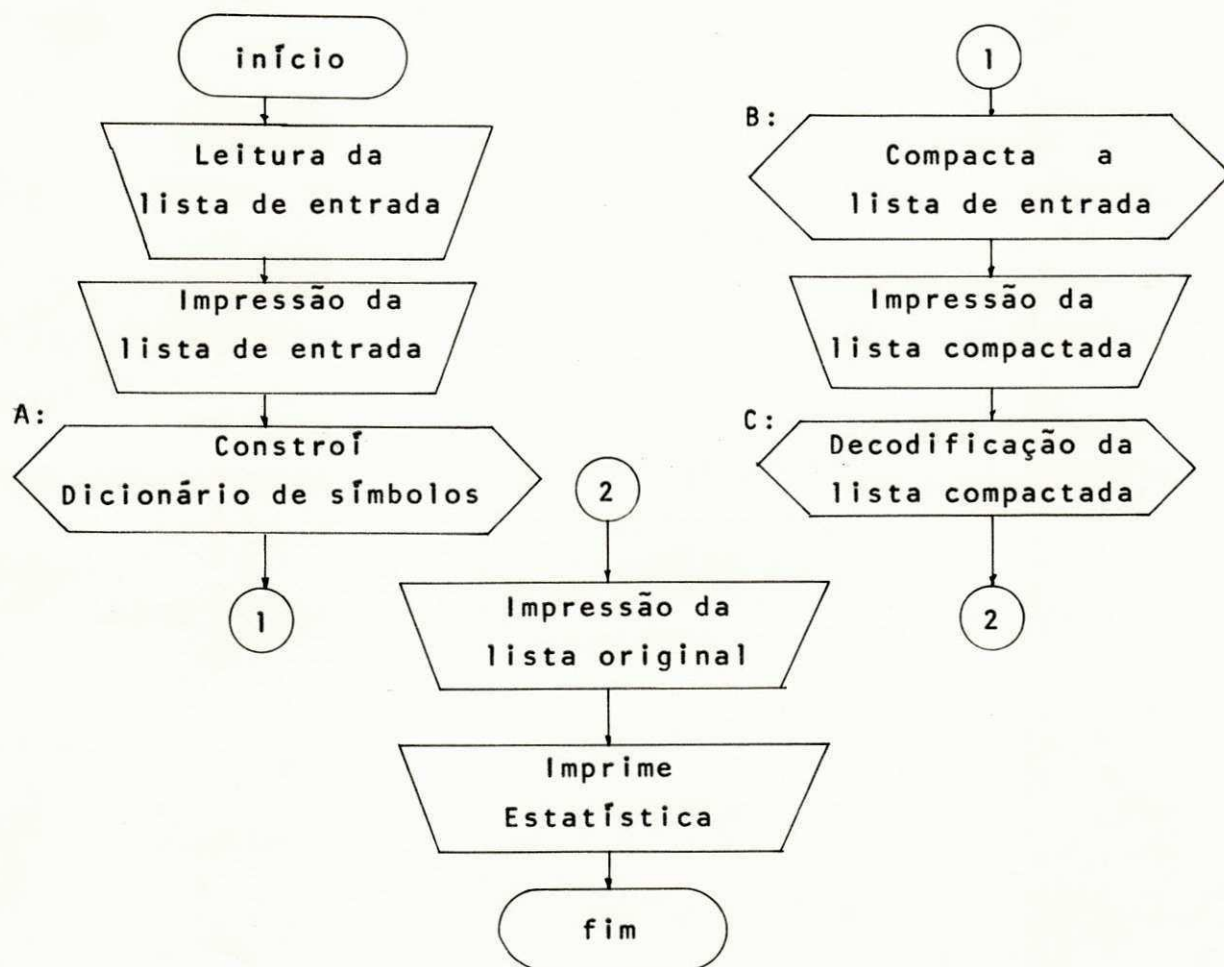
1. ABRAMSON, Norman - Information Theory and coding, New York. McGraw - Hill - 1963.
2. PETERSON, W. Wesley, and Weldon, E.J. - Error - Correting Codes, Copyriht the Massachusetts Institutes of Technology 1972.
3. HAHN, Bruce - A New Technique for Compression and Storage of Data, Comm. ACM 17, 8 (August 1974).
4. BRILLINGER, Peter C., and COHEN, Doron J. - Introduction to Data Structures and Non-numerical Computation. Prentice-Hall, Inc. - 1972.
5. KNUTH, Donald E. - The art of computer programming: Sorting and Searching, Vol. 3 - Addisson-Wesley Publishing Company, Inc. - 1973.
6. JUDD, D.R. - Use of Files, Macdonald/American Elsevier computer Monographs - 1973.
7. WAGNER, R.A. - Common Phrases and Minimum-Space text Storage - Comm. ACM 16,3 (Mar. 1973).
8. FAJMAN, R., and BORGELT, J. - WYLBUR: An Interactive text Editing and Remote Job Entry System, Comm. ACM 16,5 (Mar. 1973).
9. PACITTI, Tércio - FORTTRAN - MONITOR princípios, Ao Livro Técnico S.A. - Rio de Janeiro - 1970.

APÊNDICE A

Neste apêndice está desenvolvido um fluxograma simplificado, referente ao programa de implementação do primeiro método. Alguns exemplos e estatísticas são apresentados, também.

FLUXOGRAMA

UNIVERSIDADE FEDERAL DA PARAÍBA
Pró-Reitoria Para Assuntos do Interior
Coordenação Setorial de Pós-Graduação
Rua Aprígio Veloso, 882 - Tel (083) 321-7222-R 355
58.100 - Campina Grande - Paraíba



A: Subrotina p/ construção do DS - SUB1

B: Subrotina p/ codificar a lista de entrada - COD

C: Subrotina p/ decodificar a lista compactada retornando
à lista de entrada original - DECOD

EXEMPLO 1.1 - METODO 1

COMPACTACAO E DESCOMPACTACAO DE TEXTO EM FORTRAN CONTEDO 1872 CARACTERES

TEXTO DE ENTRADA

```

SUBROUTINE GAUJO(N,DET,A)
DIMENSION A(21,22),ITROC(21),IBLOC(21)
DO 1 I=1,N
  IBLOC(I)=0
1  ITROC(I)=0
  N2=N+1
  N1=N-1
  DET=1
  DO 100 K=1,N
    PIVO=0
    DO 20 I=1,N
      DO 20 J=1,N
        IF(ITROC(I))20,4,20
4       IF(IBLOC(J))20,5,20
5       IF(ABS(A(I,J))-ABS(PIVO,))20,10,10
10      PIVO=A(I,J)
        IL=I
        IC=J
20     CONTINUE
        ITROC(IL)=IC
        IBLOC(IC)=IL
        DET=DET*PIVO
        A(IL,IC)=1./PIVO
        DO 40 J=1,N
          IF(J-IL)30,40,30
30      A(J,IC)=-A(J,IC)*A(IL,IC)

```


DICCIONARIO DE SIMBOLOS

1	
2	I
3	,
4	O
5	(
6)
7	0
8	1
9	=
10	C
11	N
12	2
13	
14	A
15	T
16	L
17	J
18	D
19	E
20	B
21	V
22	R
23	P
24	4
25	U
26	
27	S
28	F
29	-
30	3
31	5
32	*
33	M
34	K
35	G
36	•
37	+
38	/

COMPACTACAO - TEXTO DE ENTRADA

-6	13	26455	20461	375	1027	1522	2067	9779	851	377	3225	-45	871	1189
24168	5214	246	18251	13715	4515	22979	13715	15422	22979	13365	-31	898	17780	18227
-61	4485	7402	5532	-57	17759	-1	4420	20584	5532	-62	26323	151	-66	25647
47	-66	851	463	-67	898	18851	113	18227	-59	1716	18376	-66	898	27562
21175	-61	898	27560	10410	24350	-58	4396	11325	1583	11407	27586	24837	-48	1873
-2	4396	11330	569	11043	15304	68	27561	-46	66	-3	4396	10998	15380	11003
4905	14196	6604	15380	11117	1410	15304	18806	15562	-31	18773	2329	1413	236	6649
-61	4442	4577	-65	6201	8971	-60	27561	3950	375	2034	-64	4420	20584	4439
20242	-59	4485	7402	6171	4434	-59	851	455	11063	4400	1716	18266	-58	236
548	23109	130	2038	4502	-56	898	1951	801	8464	-59	4396	11037	533	7605
10010	25389	768	-50	59	2379	3046	7065	19776	234	9331	13188	236	548	22998

DESCOMPACTACAO - LISTA DE SAIDA

```
SUBROUTINE GAUJO(N,DET,A)
DIMENSION A(21,22),ITROC(21),IBLOC(21)
DO 1 I=1,N
IBLOC(I)=0
1 ITROC(I)=0
N2=N+1
N1=N-1
DET=1
DO 100 K=1,N
PIVO=0
DO 20 I=1,N
DO 20 J=1,N
IF(ITROC(I))20,40,20
4 IF(IBLOC(J))20,50,20
5 IF(ABS(A(I,J))-ABS(PIVO))20,10,10
10 PIVO=A(I,J)
IL=I
IC=J
20 CONTINUE
ITROC(IL)=IC
IBLOC(IC)=IL
DET=DET*PIVO
A(IL,IC)=1./PIVO
DO 40 J=1,N
IF(J-IL)30,40,30
30 A(J,IC)=-A(J,IC)*A(IL,IC)
```

ESTATISTICA

NUMERO DE INFORMACOES LIDAS = 13714.0002

NO. DE PALAVRAS REQUEI'DAS P/ COMPRESSAO = 165.0000

APROX. 1.4102 BITS POR INFORMACAO

EXEMPLO 1.2 - METODO 1

COMPACTACAO E DESCOMPACTACAO DE TEXTO EM PORTUGUES CONTENDO 2000 CARACTERES

TEXTO DE ENTRADA

ESTE LIVRO, BASEADO EM MINHAS NOTAS DE AULA, FOI ELABORADO COM O OBJETIVO DE AUXILIAR O ENGENHEIRO, O ECONOMISTA OU O ESTADISTICO, QUE NECESSITE USAR UM COMPUTADOR DIGITAL, SEM PRECISAR ENTRAR NOS DETALHES DA MAQUINA UTILIZADA., EM GERAL SUAS VIDAS PROFISSIONAIS SAO TAO ABSORVEITES, QUE NAO POSSUEM TEMPO - E MUITAS VEZES PACIENCIA - PARA TAL. O FORTRAN (FORMULA TRANSLATION) EH UMA DAS LINGUAGENS AUTOMATICAS EXISTENTES, ORIENTADA PARA A RESOLUCAO DE PROBLEMAS CIENTIFICOS E TECNOLOGICOS E QUE PERMITE UMA UTILIZACAO FACILITADA DO COMPUTADOR. ELA TENDE A TORNAR UNIVERSAL O USO DOS COMPUTADORES. CONVEM INFORMAR AO LEITOR QUE OS SEGUINTE COMPUTADORES, FORNECIDOS POR DIVERSOS FABRICANTES, ACEITAM A LINGUAGEM FORTRAN (EM SEUS DIVERSOS NIVEIS) - SDS 9300., UNIVAC III E 1107., RCA 3301, SPECTRA 70., PHILCO 2000., IBM 360, 1620, 1460, 7010, 7040, 7044, 7090., NCR 315., HONEYWELL 200, 800, 1800., GENERAL ELECTRIC 205, 210, 215, 225, 235, 412, 415, 425, 625, 635., EA 8400., DEC-PDP6., CDC 1604, 3600, 3800, 6400, 6600, 6800., CCC-DDP 24, DDP 116, DDP 124, DDP 224., HIRCUHGS B5500., ADVANCE 6000, 6020, 6040, 6050, 6070, 6080 E A NEAC 2200 DE FABRICACAO JAPONESA, ASSIM COMO OUTROS COMPUTADORES DE FABRICACAO EUROPEIA.

OS PRE-REQUISITOS PARA LEITURA DESTE LIVRO SAO OS MINIMOS. BASTA A MATEMATICA COLEGIAL E PARA ALGUNS DOS TOPICOS ILUSTRADOS, O PRIMEIRO ANO DE UM CURSO SUPERIOR QUE TENHA MATEMATICA EM SEU CURRICULO.

A LEITURA CUIDADOSA, COM APROVEITAMENTO, PODE SER FEITA SEM AJUDA DE PROFESSOR, EMBORA EM ALGUNS CASOS ALGUMA EXPLICACAO SUPLEMENTAR SE FAÇA NECESSARIA.

DICCIONARIO DE SIMBOLOS.

1	
2	A
3	E
4	O
5	S
6	I
7	R
8	T
9	,
10	C
11	O
12	U
13	
14	N
15	D
16	M
17	L
18	P
19	•
20	6
21	2
22	1
23	G
24	F
25	4
26	
27	V
28	B
29	5
30	-
31	3
32	H
33	7
34	Q
35	8
36	Z
37	X
38	J
39	
40	9
41	(
42)
43	Y
44	W

COMPACTACAO - TEXTO DE ENTRADA

-7	7543	2255	20	10322	369	6931	8996	6763	13195	1045	2214	17980	378	6426
5928	24922	6645	30	15719	9091	524	2873	4406	8021	221	4916	26375	13240	5594
2705	1729	2203	7700	19995	8333	2876	14155	3030	9001	18024	12425	10322	1511	2213
22547	14574	27242	2391	641	512	27742	5107	416	23090	794	11495	2269	8364	5593
280	5590	2939	385	4446	1058	2225	2258	116	13197	4335	754	1716	4738	19956
6630	22570	694	26768	10	369	2269	8937	11912	201	11221	9063	8983	407	15380
6612	7554	116	6761	5082	11731	26874	3588	6660	2366	8971	6642	14565	2198	6601
7449	11459	6614	13534	2204	2264	15738	4446	13404	1918	18785	2366	349	9974	26418
3642	234	9236	8801	40	6597	4228	4576	4797	9803	1848	1729	3057	27768	7039
22374	6602	14134	276	12525	16432	3575	4745	11416	2555	7488	10948	8959	6765	16055
4449	538	3968	169	10218	25315	11194	18213	2877	8924	22712	6760	19644	891	593
6916	6954	18594	13192	6114	2342	22988	14560	4745	5093	8827	13119	397	1030	52
9147	6604	4916	3608	202	4226	13183	7841	693	4294	2227	11288	512	27742	5164
11064	22647	16	526	180	9974	5591	5084	2252	14593	116	6817	3081	24076	3593
3939	6658	17916	10015	19953	9972	8359	5136	904	2229	16	16281	2342	33	14898
3593	19978	7711	521	-2	4452	179	26712	6631	1918	18785	2366	377	6828	27222
5408	2800	9646	3211	2787	3	2201	3044	11154	169	12987	14728	247	205	3295
2717	19901	91	14723	22308	10990	24293	3047	8392	4563	17244	19947	84	810	2312
26032	20020	341	2202	1335	2314	15494	19951	2035	25701	1326	21750	102	2180	2204
24323	27898	1326	24	1132	2220	15548	11102	513	19942	13859	6591	8788	11496	689
19587	2206	26144	1521	21775	14716	22478	7800	8996	9406	14885	19435	8125	17704	19950
1521	8125	17680	625	-3	1352	10988	19929	19886	2353	19776	19954	1352	8125	15483
625	1183	10988	1132	6942	1533	26032	19944	8281	8853	4459	15466	3889	22148	1326
27898	852	26144	65	1675	19949	2182	19949	1337	19864	1675	1132	23790	678	343
2301	27898	4420	11163	1528	19944	343	2314	17741	2223	4460	1360	2034	19942	6520
10826	83	2199	39	661	1132	2392	4395	4417	6767	26178	1326	19553	1326	28341
1326	659	2299	102	19949	24176	24376	4564	7060	1360	26039	4914	24505	5665	6114
2209	4463	213	5930	11908	6774	-1	22649	9628	16121	22649	1009	4424	15952	378
1885	435	22440	9008	16060	7631	13365	-121	9646	12207	770	116	14113	9646	11416
2252	14697	4565	7543	2255	20	9024	9014	2242	247	7332	13351	4805	4590	541
6760	8829	18722	3939	9305	13524	2717	11416	2539	1846	3055	5136	18257	14929	3215
27320	4424	12523	2269	13224	14418	4411	2225	4228	4050	11738	891	13950	116	6867
169	13533	7036	541	14899	6631	13053	24056	15028	9470	-120	4567	7721	15740	27380
4424	11441	22649	2540	16055	2790	4436	217	19947	8817	3088	2343	14561	11495	2535
28394	4565	6765	16055	8934	11889	19981	6513	9998	6631	4446	23999	11286	11727	4446
24001	4602	1864	764	6114	3198	11040	515	3582	3082	1895	4564	8325	11420	4473

DESCOMPACTACAO - LISTA DE SAIDA

ESTE LIVRO, BASEADO EM MINHAS NOTAS DE AULA, FOI ELABORADO COM O OBJETIVO DE AUXILIAR O ENGENHEIRO, O ECONOMISTA OU O ESTATISTICO, QUE NECESSITE USAR UM COMPUTADOR DIGITAL, SEM PRECISAR ENTRAR NOS DETALHES DA MAQUINA UTILIZADA., EM GERAL SUAS VIDAS PROFISSIONAIS SAO TAO ABSORVENTES, QUE NAO POSSUEM TEMPO - E MUITAS VEZES PACIENCIA - PARA TAL. O FORTRAN (FORMULA TRANSLATION) EH UMA DAS LINGUAGENS AUTOMATICAS EXISTENTES, ORIENTADA PARA A RESOLUCAO DE PROBLEMAS CIEN- TIFICOS E TECNOLOGICOS E QUE PERMITE UMA UTILIZACAO FACILITADA DO COMPUTADOR. E- LA TENDE A TORNAR UNIVERSAL O USO DOS COMPUTADORES. CONVEM INFORMAR AO LEITOR QUE OS SEGUINTE COMPUTADORES, FORNECIDOS POR DIVERSOS FABRICANTES, ACEITAM A LINGUAGEM FORTRAN (EM SEUS DIVERSOS NIVEIS) - SDS 9300., UNIVAC III E 1107., RCA 3301, SPECTRA 70., PHILCO 2000., IBM 360, 1620, 1460, 7010, 7040, 7044, 7090., NCR 315., HONEYWELL 200, 800, 1800., GENERAL ELETRIC 205, 210, 215, 225, 235, 412, 415, 425, 625, 635., EA 8400., DEC-PDP6., CDC 1604, 3600, 3800, 6400, 6600, 6800., CCC-DDP 24, DDP 116, DDP 124, DDP 224., BURROUGHS B5500., ADVANCE 6000, 6020, 6040, 6050, 6070, 6080 E A NEAC 2200 DE APLICACAO JAPONESA, ASSIM COMO OUTROS COMPUTADORES DE FABRICACAO EUROPEIA.

OS PRE-REQUISITOS PARA LEITURA DESTE LIVRO SAO OS MINIMOS. BASTA A MATE - Matica COLEGIAL E PARA ALGUNS DOS TOPICOS ILUSTRADOS, O PRIMEIRO ANO DE UM CURSO SUPERIOR QUE TENHA MATEMATICA EM SEU CURRICULO.

A LEITURA CUIDADOSA, COM APROVEITAMENTO, PODE SER FEITA SEM AJUDA DE PRO- FESSOR, EMBORA EM ALGUNS CASOS ALGUMA EXPLICACAO SUPLEMENTAR SE FAÇA NECESSARIA.

ESTATISTICA

NUMERO DE INFORMACOES LIDAS = 2000.0002

NO. DE PALAVRAS REQUEIRDAS P/ COMPRESSAO = 540.0001

APROX. 4.3200 BITS POR INFORMACAO

EXEMPLO 1.3 - METODO 1

COMPACTACAO E DESCOMPACTACAO DE UM BANCO DE DADOS CONTENDO 2002 CARACTERS

SUDENE - INFORMACOES SOBRE HIDROLOGIA -

BLOCO DE DADOS DE ENTRADA

420001962011							
420001962012			019201930009	0272003700910223		0053	
420001962021	0110		00110016026500440284023400110001				
420001962022006400900172095400550073			012601000025001102000174				
4200019620310101028800210012			0354	01820005013202320078			
4200019620320027	02080110027601860083		00690089	020404360035			
420001962041	01060074	015201310040010301980005		00060010	0046		
4200019620420095	0070	00230041011000960065	01320014	0127			
42000196205100160145		00060195	01770032		0003		
420001962052002400320031		0025	0004		0059		
420001962061				005001500004			
4200019620620003							
420001962071							
420001962072	0092				0092		
420001962081							
420001962082							
420001962091							
420001962092							
420001962101							
420001962102							
420001962111							
420001962112							
420001962121							
420001962122							
420001963011				0332006702810103			
4200019630120732	00540050	0531	053700400050	0052	02030193		

DICIONARIO DE SIMBOLOS

1	
2	0
3	2
4	1
5	4
6	9
7	6
8	3
9	5
10	7
11	8

COMPACTACAO - TEXTO DE ENTRADA

11520	5155	6985	-65	11520	5155	6984	-24	5151	5156	4764	-4	5034	4846	4814
4948	-4	4857	3083	4790	15915	8971	-1	5124	-16	4788	4791	5001	4802	5049
5010	4788	4762	-9	11520	5155	6971	4828	4812	5203	5530	4858	4870	-4	5116
5098	4780	4788	4929	5205	-13	11520	5155	7037	5100	5055	4775	4787	-16	5868
-4	5216	4767	5177	5008	4873	-5	11520	5155	7036	4781	-4	4938	5124	5038
5220	4883	-8	4829	4881	-4	4932	5350	4845	-5	11520	5155	6998	-4	5103
4867	-4	5190	5178	4799	5104	5159	4767	-4	4765	4786	-4	4804	-5	11520
5155	6997	4819	-4	4864	-4	4779	4801	5124	4817	4832	-4	5177	4789	-4
5119	-9	11520	5155	7050	4791	5144	-8	4765	5157	-8	5210	4839	-16	4766
-5	11520	5155	7049	4776	4839	4840	-4	4780	-16	4763	-16	4855	-5	11520
5155	7024	-44	4851	5189	4763	-9	11520	5155	7023	4766	-61	11520	5155	7063
-65	11520	5155	7062	-4	4813	-44	4813	-9	11520	5155	7076	-65	11520	5155
7075	-65	11520	5155	7011	-65	11520	5155	7010	-65	11520	5155	7297	-65	11520
5155	7296	-65	11520	5155	7323	-65	11520	5155	7322	-65	11520	5155	7310	-65
11520	5155	7309	-65	11520	5155	17970	-65	5853	4833	5048	5104	-9	11520	5155
17969	6191	-8	4854	4851	-4	6023	-4	6029	4799	4851	-4	4852	-4	4935
5156														

DESCOMPACTACAO - LISTA DE SAIDA

420001962011					
420001962012		019201930009	0272003700910223	0053	
420001962021	0110		00110016026500440284023400110001		
420001962022	006400900172095400550073		012601000025001102000174		
420001962031	0101028800210012		0354	01820005013202320078	
420001962032	0027	02080110027601860083	00690089	020404360035	
420001962041	01060074	015201310040010301980005	00060010	0046	
420001962042	0095	0070	00230041011000960065	01320014	0127
420001962051	00160145	00060195	01770032		0003
420001962052	002400320031	0025	0004		0059
420001962061				005001500004	
420001962062	0003				
420001962071					
2001962072	0092			0092	
420001962081					
420001962082					
420001962091					
420001962092					
420001962101					
420001962102					
420001962111					
420001962112					
420001962121					
420001962122					
420001963011				0332006702810103	
420001963012	0732	00540050	0531	053700400050	0052 02030193

ESTATISTICA

NUMERO DE INFORMACOES LIDAS = 2002.0002

NO. DE PALAVRAS REQUEIRDAS P/ COMPRESSAO = 241.0000

APROX. 1.9260 BITS POR INFORMACAO

ESTATISTICA 1.1

BLOCO DE DADOS DE ENTRADA TEXTO FORTRAN CONTENDO 24.3336 CARACTERES

NUMERO DE INFORMACOES LIDAS = 24336.0039

NO. DE PALAVRAS REQUERIDAS P/ COMPRESSAO = 2332'0004

APROX. 1.5332 BITS POR INFORMACAO

UNIVERSIDADE FEDERAL DA PARAÍBA
Pró-Reitoria Para Assuntos do Interior
Coordenação Seteriel da Pós-Graduação
Rua Aprígio Veloso, 832 - Tel (031) 3.1.7.222-R 355
58.100 - Campina Grande - Paraíba

ESTATISTICA 1.2

BLOCO DE DADOS DE ENTRADA = TEXTO EM PORTUGUEZ CONTENDO 12.000 CARACTERES

NUMERO DE INFORMACOES LIDAS = 12000.0019

NO. DE PALAVRAS REQUERIDAS P/ COMPRESSAO = 3302.0004

APROX. 4.4026 BITS POR INFORMACAO

ESTATISTICA 1.3

BLOCO DE DADOS DE ENTRADA = BANCO DE DADOS (SUDENE)

CONTENDO 50.050 CARACTERS

NUMERO DE INFORMACOES LIDAS = 50050.0079

NO. DE PALAVRAS REQUERIDAS P/ COMPRESSAO = 5732.0009

APROX. 1.8324 BITS POR INFORMACAO

ESTATISTICA 1.4

BLOCO DE DADOS DE ENTRADA = BANCO DE DADOS CONTENDO 50050. CARACTERES

UTILIZACAO DOS VALORES B=7 E N=5

NUMERO DE INFORMACOES LIDAS = 50050.0079

NO. DE PALAVRAS REQUERIDAS = 5683.0009

APROX. 1.8167 BITS POR INFORMACAO

ESTATISTICA 1.5

BLOCO DE DADOS DE ENTRADA = TEXTO EM FORTRAN CONTENDO 24336. CARACTERES

UTILIZACAO DE UNICO DICIONARIO DE SIMBOLOS

NUMERO DE INFORMACOES LIDAS = 24336.0039

NO. DE PALAVRAS REQUERIDAS = 2550.0004

APROX. 1.6765 BITS POR INFORMACAO

ESTATISTICA 1.6

BLOCO DE DADOS DE ENTRADA = TEXTO EM PORTUGUES CONTENDO 12000. CARACTERES

UTILIZACAO DE UNICO DICIONARIO DE SIMBOLOS

NUMERO DE INFORMACOES LIDAS = 12000.0019

NO. DE PALAVRAS REQUERIDAS = 3341.0004

APROX. 4.4546 BITS POR INFORMACAO

ESTATISTICA 1.7

BLOCO DE DADOS DE ENTRADA = BANCO DE DADOS CONTENDO 50050. CARACTERES

UTILIZACAO DE UNICO DICIONARIO DE SIMBOLOS

NUMERO DE INFORMACOES LIDAS = 50050.0079

NO. DE PALAVRAS REQUEIRDAS = 5695.0009

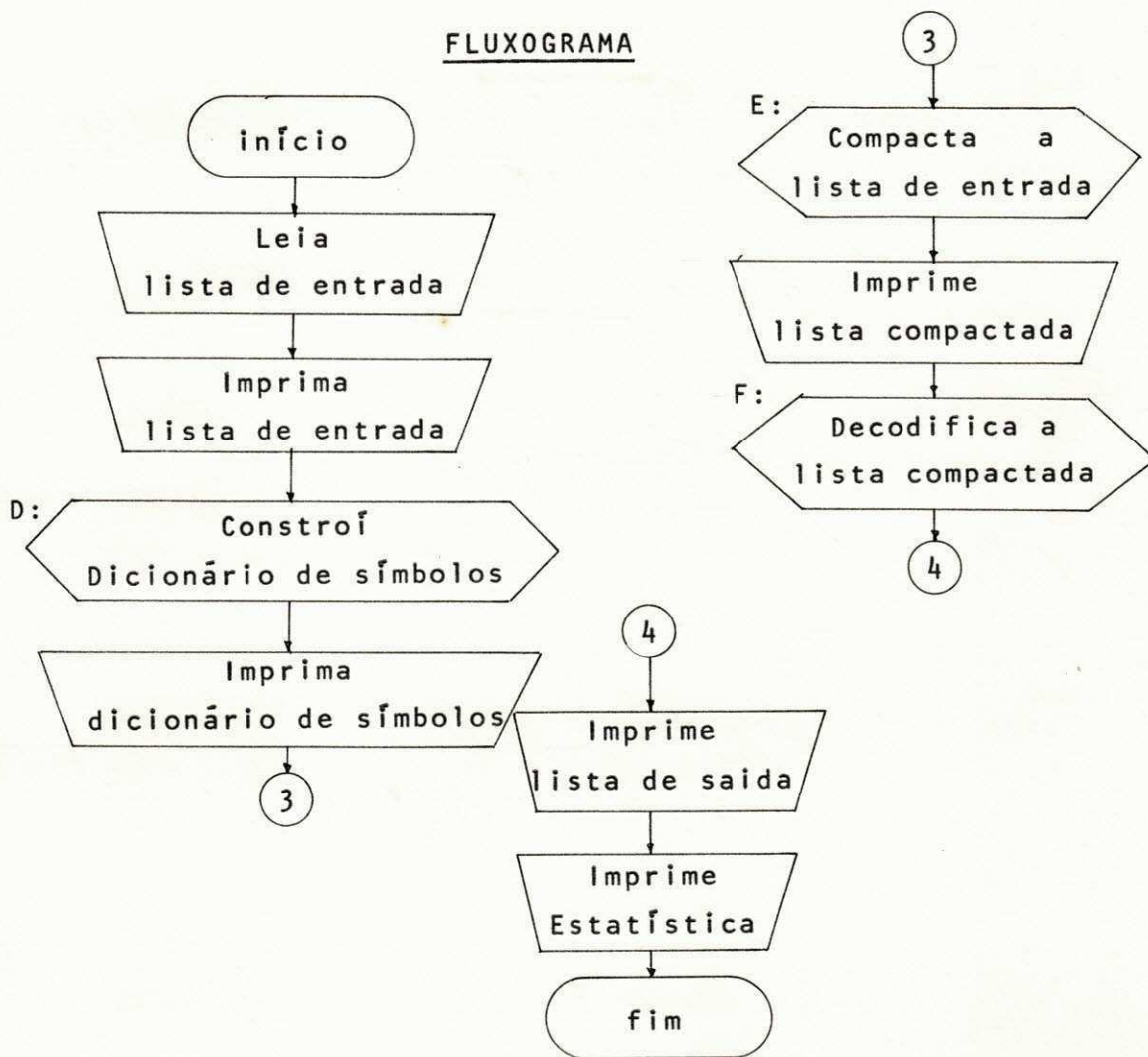
APROX. 1.8205 BITS POR INFORMACAO

UNIVERSIDADE FEDERAL DA PARAIBA
Pró-Reitoria Para Assuntos do Interior
Coordenação Setorial de Pós-Graduação
Rua Aprígio Veloso, 882 - Tel (083) 321 7222-R 355
58.100 - *Campina Grande - Paraíba*

APENDICE B

Neste apêndice está desenvolvido um fluxograma simplificado, referente ao programa de implementação do segundo método. Alguns exemplos e estatísticas são apresentados.

FLUXOGRAMA



D: Subrotina que constrói o dicionário de símbolos - SUBI

E: Subrotina que compacta a lista de entrada - COD

Chama as subrotinas

INSE - insere códigos binários na palavra

BNDE - transforma binário em decimal

DEBIN - transforma decimal em binário

F: Subrotina que decodifica a lista compactada - DECOD.

Chama a subrotina DEBIN

EXEMPLO 2.1 - METODO 2

COMPACTACAO E DESCOMPACTACAO DE TEXTO EM FORTR, N CONTEDO 1872 CARACTERES

TEXTO DE ENTRADA

```
SUBROUTINE GAUJO(N,DET,A)
DIMENSION A(21,22),ITROC(21),IBLOC(21)
DO 1 I=1,N
IBLOC(I)=0
1 ITROC(I)=0
N2=N+1
N1=N-1
DET=1
DO 100 K=1,N
PIVO=0
DO 20 I=1,N
DO 20 J=1,N
IF(ITROC(I))20,4,20
4 IF(IBLOC(J))20,5,20
5 IF(ABS(A(I,J))-ABS(PIVO))20,10,10
10 PIVO=A(I,J)
IL=I
IC=J
20 CONTINUE
ITROC(IL)=IC
IBLOC(IC)=IL
DET=DET*PIVO
A(IL,IC)=1./PIVO
DO 40 J=1,N
IF(J-IL)30,40,30
30 A(J,IC)=-A(J,IC)*A(IL,IC)
```

DICCIONARIO DE SIMBOLOS

1	1	
2	1	I
1	2	,
3	1	O
2	2	(
1	3)
4	1	0
3	2	1
2	3	=
1	4	C
5	1	N
4	2	2
3	3	A
2	4	T
1	5	L
6	1	J
5	2	D
4	3	E
3	4	B
2	5	V
1	6	R
7	1	P
6	2	4
5	3	U
4	4	S
3	5	F
2	6	-
1	7	3
8	1	5
7	2	*
6	3	M
5	4	K
4	5	G
3	6	•
2	7	+
1	8	/

COMPACTACAO DO TEXTO DE ENTRADA

10920	15388	15864	16847	4163	20728	28784	4400	11288	14828	15232	-47	800	14448	8676
8355	19505	22625	23955	28613	29452	14180	15857	15555	3520	-34	788	13468	13832	-62
4591	17894	11888	16384	-57	3412	20350	6041	14786	-62	537	24735	25344	-66	563
16703	3072	-66	782	15587	-67	788	12421	974	6916	-60	147	28984	8192	-66
788	6226	14552	4096	-61	788	6224	5005	16640	-61	4595	5087	17894	12001	17088
27696	16384	-48	426	18681	18679	25331	375	1558	176	24832	-48	85	9340	25487
1944	29484	1500	32312	30841	16457	30907	17163	3116	12416	-34	3114	16457	30876	14742
736	-61	6094	8192	-68	6044	1024	-63	1557	12048	10128	8417	24576	-64	5087
17894	12217	26080	-60	4591	17894	12147	19424	-60	782	15584	25039	192	9464	16384
-60	3685	31535	14790	8127	24612	31808	-56	788	1556	1251	12352	-61	4595	319
6109	32278	779	16260	-51	16261	10470	715	28275	30950	715	28163	3685	31535	14336

UNIVERSIDADE FEDERAL DA PARAIBA
 Pró-Reitoria Para Assuntos do Interior
 Coordenação Setorial de Pós-Graduação
 Rua Aprígio Veloso, 892 - Tel. (051) 331-7222-R 355
 58.100 - Campina Grande - Paraíba

DESCOMPACTACAO - LISTA DE SAIDA

```
SUBROUTINE GAUJO(N,DET,A)
DIMENSION A(21,22),ITROC(21),IBLOC(21)
DO 1 I=1,N
IBLOC(I)=0
1 ITROC(I)=0
N2=N+1
N1=N-1
DET=1
DO 100 K=1,N
PIVO=0
DO 20 I=1,N
DO 20 J=1,N
IF(ITROC(I))20,4,20
4 IF(IBLOC(J))20,5,20
5 IF(ABS(A(I,J))-ABS(PIVO))20,10,10
10 PIVO=A(I,J)
IL=I
IC=J
20 CONTINUE
ITROC(IL)=IC
IBLOC(IC)=IL
DET=DET*PIVO
A(IL,IC)=1./PIVO
DO 40 J=1,N
IF(J-IL)30,40,30
30 A(J,IC)=-A(J,IC)*A(IL,IC)
```


ESTATISTICA

NUMERO DE INFORMACOES LIDAS = 1872.0002

NO. DE PALAVRAS REQUEIRDAS P/ COMPRESSAO = 165.0000

APROX. 1.4102 BITS POR INFORMACAO

EXEMPLO 2.2 - METODO 2

COMPACTACAO E DESCOMPACTACAO DE TEXTO EM PORTUGUES CONTENDO 2000 CARACTERES

TEXTO DE ENTRADA

ESTE LIVRO, BASEADO EM MINHAS NOTAS DE AULA, FOI ELABORADO COM O OBJETIVO DE AUXILIAR O ENGENHEIRO, O ECONOMISTA OU O ESTATISTICO, QUE NECESSITE USAR UM COMPUTADOR DIGITAL, SEM PRECISAR ENTRAR NOS DETALHES DA MAQUINA UTILIZADA., EM GERAL SUAS VIDAS PROFISSIONAIS SAO TAO ABSORVEITES, QUE NAO POSSUEM TEMPO - E MUITAS VEZES PACIENCIA - PARA TAL. O FORTRAN (FORMULA TRANSLATION) EH UMA DAS LINGUAGENS AUTOMATICAS EXISTENTES, ORIENTADA PARA A RESOLUCAO DE PROBLEMAS CIEN- TIFICOS E TECNOLOGICOS E QUE PERMITE UMA UTILIZACAO FACILITADA DO COMPUTADOR. E- LA TENDE A TORNAR UNIVERSAL O USO DOS COMPUTADORES. CONVEM INFORMAR AO LEITOR QUE OS SEGUINTE COMPUTADORES, FORNECIDOS POR DIVERSOS FABRICANTES, ACEITAM A LINGUAGEM FORTRAN (EM SEUS DIVERSOS NIVEIS) - SDS 9300., UNIVAC III E 1107., RCA 3301, SPECTRA 70., PHILCO 2000., IBM 360, 1620, 1460, 7010, 7040, 7044, 7090., NCR 315., HONEYWELL 200, 800, 1800., GENERAL ELETRIC 205, 210, 215, 225, 235, 412, 415, 425, 625, 635., EA 8400., DEC-PDP6., CDC 1604, 3600, 3800, 6400, 6600, 6800., CCC-DDP 24, DDP 116, DDP 124, DDP 224., HIRCOUGHS B5500., ADVANCE 6000, 6020, 6040, 6050, 6070, 6080 E A NEAC 2200 DE FABRICACAO JAPONESA, ASSIM COMO OUTROS COMPUTADORES DE FABRICACAO EUROPEIA.

OS PRE-REQUISITOS PARA LEITURA DESTE LIVRO SAO OS MINIMOS. BASTA A MATE - Matica COLEGIAL E PARA ALGUNS DOS TOPICOS ILUSTRADOS, O PRIMEIRO ANO DE UM CURSO SUPERIOR QUE TENHA MATEMATICA EM SUE CURRICULO.

A LEITURA CUIDADOSA, COM APROVEITAMENTO, PODE SER FEITA SEM AJUDA DE PRO- FESSOR, EMBORA EM ALGUNS CASOS ALGUMA EXPLICACAO SUPLEMENTAR SE FACA NECESSARIA.

DICCIONARIO DE SIMBOLOS

1	1	
2	1	A
1	2	E
3	1	O
2	2	S
1	3	I
4	1	R
3	2	T
2	3	,
1	4	C
5	1	O
4	2	U
3	3	N
2	4	D
1	5	M
6	1	L
5	2	P
4	3	•
3	4	6
2	5	2
1	6	1
7	1	G
6	2	F
5	3	4
4	4	V
3	5	B
2	6	5
1	7	-
8	1	3
7	2	H
6	3	7
5	4	Q
4	5	8
3	6	Z
2	7	X
1	8	J
9	1	9
8	2	(
7	3)
6	4	Y
5	5	W

COMPACTAÇÃO DO TEXTO DE ENTADA

10922	26166	16568	15394	14910	9929	29019	30199	3587	4935	2249	21466	8577	5024	12661
16529	30852	20245	29053	2595	30715	3577	25211	9267	32192	23685	2759	177	24630	28809
26797	28956	12217	18001	1674	26162	7004	28401	7431	17256	30454	13169	23065	18517	1787
25336	6243	5060	5367	184	25609	27063	26672	23516	25643	3632	18513	25193	30258	515
13135	5618	1926	29130	3182	739	31036	17309	11132	706	8275	1613	1977	29288	6212
1766	14097	26061	6420	2841	3992	16926	25485	19688	7707	9105	8388	26139	16013	24326
5503	10975	1763	4931	27774	26243	6075	3831	5471	26674	2374	8259	20800	25105	16967
8216	3138	31792	4748	4665	16529	23623	235	208	28565	7756	1479	134	8236	14756
6341	31003	24141	13307	29564	14553	20104	11975	3230	10290	2378	2865	269	29225	30339
1095	24667	29291	28359	16269	24686	30874	26843	25480	2176	24034	13728	30829	865	16099
13367	25868	14082	29177	15498	815	14359	3230	10722	24111	16774	6460	8462	23544	2374
25502	26950	4238	8515	3809	30227	4136	20684	10722	13794	31768	12691	25110	12523	25144
15806	23664	6277	30986	18602	732	12552	7707	8809	22536	14108	13931	25336	6243	5060
5735	8290	2285	30524	9859	2131	28190	24881	6662	9185	15304	29107	7471	14105	16042
8284	14344	12802	28576	12552	24867	20493	24397	17204	31623	22604	9870	28918	29441	27642
13261	32	2080	17309	1595	17355	27579	22207	16132	1806	14871	18436	261	31977	16823
17938	16608	17309	769	23563	25247	520	4327	11838	32000	18306	15100	15608	5039	24803
24733	449	16132	29710	1052	2512	7184	14393	26652	2050	1081	27195	24896	6119	28903
8241	3776	30782	24640	21472	16698	3970	1259	30844	2081	26432	11382	2306	22550	6237
29944	5107	21487	24733	8062	16186	15996	32373	10736	1276	28791	29671	8431	26599	8423
26599	9167	20430	18304	10209	26457	8688	14368	17309	7902	32518	15384	30835	22430	31484
15376	14800	1144	4174	16417	30752	20111	898	1256	30960	8348	15422	1040	29615	15855
29647	847	16846	20284	3454	32316	29939	24791	29664	29607	20230	20383	903	7455	1552
17456	2060	26873	31996	2081	26452	20239	4574	26864	8322	14576	10177	7454	1052	2513
28751	24733	3844	1796	29816	4344	5541	3787	27123	28737	10164	1607	24951	19400	22521
785	27849	26931	15338	22411	28944	25352	19832	24326	3172	30853	19773	20530	7947	24158
8880	24848	14052	7168	-124	2464	24943	28848	15415	7052	9859	4244	1464	24964	21465
18128	5903	1107	4420	27630	7647	2439	9188	25385	12177	23294	31885	28459	25093	16476
16557	801	5249	134	7379	25192	25101	28435	11780	6540	4728	19688	20577	15341	24852
18196	31555	16060	6220	9752	6923	17478	7707	9059	16485	15942	28561	24037	14291	1751
17160	12016	24644	7168	-120	5131	14521	1199	1767	18914	12917	29053	4193	2171	14539
30264	25204	3151	13528	10267	14538	14074	12272	26516	31553	16943	28726	13072	20151	25570
2395	29825	134	7383	18825	20996	529	29363	31768	3004	24133	6240	24667	30264	25641
23046	12072	30454	13089	14606										

DESCOMPACTACAO - LISTA DE SAIDA

ESTE LIVRO, BASEADO EM MINHAS NOTAS DE AULA, FOI ELABORADO COM O OBJETIVO DE AUXILIAR O ENGENHEIRO, O ECONOMISTA OU O ESTATISTICO, QUE NECESSITE USAR UM COMPUTADOR DIGITAL, SEM PRECISAR ENTRAR NOS DETALHES DA MAQUINA UTILIZADA., EM GERAL SUAS VIDAS PROFISSIONAIS SAO TAO ABSORVEITES, QUE NAO POSSUEM TEMPO - E MUITAS VEZES PACIENCIA - PARA TAL. O FORTRAN (FORMULA TRANSLATION) EH UMA DAS LINGUAGENS AUTOMATICAS EXISTENTES, ORIENTADA PARA A RESOLUCAO DE PROBLEMAS CIENTIFICOS E TECNOLOGICOS E QUE PERMITE UMA UTILIZACAO FACILITADA DO COMPUTADOR. ELA TENDE A TORNAR UNIVERSAL O USO DOS COMPUTADORES. CONVEM INFORMAR AO LEITOR QUE OS SEGUINTE COMPUTADORES, FORNECIDOS POR DIVERSOS FABRICANTES, ACEITAM A LINGUAGEM FORTRAN (EM SEUS DIVERSOS NIVEIS) - SDS 9300., UNIVAC III E 1107., RCA 3301, SPECTRA 70., PHILCO 2000., IBM 360, 1620, 1460, 7010, 7040, 7044, 7090., NCR 315., HONEYWELL 200, 800, 1800., GENERAL ELECTRIC 205, 210, 215, 225, 235, 412, 415, 425, 625, 635., EA 8400., DEC-PDP6., CDC 1604, 3600, 3800, 6400, 6600, 6800., CCC-DDP 24, DDP 116, DDP 124, DDP 224., HPROUGHS B5500., ADVANCE 6000, 6020, 6040, 6050, 6070, 6080 E A NEAC 2200 DL UTILIZACAO JAPONESA, ASSIM COMO OUTROS COMPUTADORES DE FABRICACAO EUROPEIA.

OS PRE-REQUISITOS PARA LEITURA DESTE LIVRO SAO OS MINIMOS. BASTA A MATEMATICA COLEGIAL E PARA ALGUNS DOS TOPICOS ILUSTRADOS, O PRIMEIRO ANO DE UM CURSO SUPERIOR QUE TENHA MATEMATICA EM SEU CURRICULO.

A LEITURA CUIDADOSA, COM APROVEITAMENTO, PODE SER FEITA SEM AJUDA DE PROFESSOR, EMBORA EM ALGUNS CASOS ALGUMA EXPLICACAO SUPLEMENTAR SE FAÇA NECESSARIA.

ESTATISTICA

NUMERO DE INFORMACOES LIDAS = 2000.0002

NO. DE PALAVRAS REQUEIRDAS P/ COMPRESSAO = 500.0000

APROX. 4.0000 BITS POR INFORMACAO

EXEMPLO 2.3 - METODO 2

COMPACTACAO E DESCOMPACTACAO DE UM BANCO DE DADOS CONTENDO 2002 CARACTERS

SUDENE - INFORMACOES SOBRE HIDROLOGIA -

BLOCO DE DADOS DE ENTRADA

420001962011		019201930009	0272003700910223	0053
420001962012		00110016026500440284023400110001		
420001962021	0110	012601000025001102000174		
420001962022006400900172095400550073		0354	01820005013202320078	
4200019620310101028800210012		00690089	020404360035	
4200019620320027	02080110027601860083	015201310040010301980005	00060010	0046
420001962041	01060074	00230041011000960065	01320014	0127
4200019620420095	0070	00060195	01770032	0003
42000196205100160145		0025	0004	0059
420001962052002400320031			005001500004	
420001962061				
4200019620620003				
420001962071				
420001962072	0092			
420001962081				
420001962082				
420001962091				
420001962092				
420001962101				
420001962102				
420001962111				
420001962112				
420001962121				
420001962122				
420001963011				
4200019630120732	0054050	0531	0332006702810103	
		0531	053700400050	0052
				02030193

DICCIONARIO DE SIMBOLOS

1	1	
2	1	0
1	2	2
3	1	1
2	2	4
1	3	9
4	1	6
3	2	3
2	3	5
1	4	7
5	1	8

COMPACTAÇÃO DO TEXTO DE ENTRADA

6948	17858	25668	-65	6948	17858	25688	-24	4470	8931	4701	10843	27793	24139	17590
6825	5005	6948	17858	25994	21640	18432	-16	4642	9282	11303	4710	11283	5683	4642
9352	-9	6948	17858	26034	8498	11878	24267	20068	20082	12058	21643	1161	4697	25736
19236	17894	-13	6948	17858	25698	8772	22560	18786	9304	-16	4508	27300	8370	9444
9061	18121	15381	10860	18711	1425	22879	30034	25618	4389	24082	4162	9233	16384	-8
4631	4619	21835	4902	6217	3306	21771	4654	2854	5458	4626	12085	10515	22801	17555
4644	12846	1170	15017	4626	8789	9378	10921	22820	11798	9929	14805	9404	21833	12690
12578	4683	16968	10069	4379	4646	21774	28416	-9	6948	17858	25826	9282	8807	-8
4674	8935	-8	4475	25741	16384	-26	4678	21837	18722	28850	15141	19601	22819	2729
5744	-16	4684	-16	4727	10918	25745	14425	1088	-44	4722	8818	9368	-9	6948
17858	25644	18712	-61	6948	17858	26082	-65	6948	17858	26093	10827	22528	-44	4844
-9	6948	17858	25617	-65	6948	17858	25622	-65	6948	17858	26052	-65	6948	17858
26072	-65	6948	17858	25156	-65	6948	17858	25176	-65	6948	17858	25122	-65	6948
17858	25132	-65	6948	17858	25284	6948	6948	17858	25304	-65	6948	17858	6417	-40
4493	18699	25986	4644	12288	-9	6948	17858	6422	12059	-8	4723	4722	21799	3157
9443	15507	4686	10916	20181	9617	18616	24576							

DESCOMPACTACAO - LISTA DE SAIDA

420001962011						
420001962012			019201930009	0272003700910223		0053
420001962021	0110		00110016020500440284023400110001			
420001962022	006400900172095400550073		012601000025001102000174			
420001962031	010101028800210012		0354	01820005013202320078		
420001962032	0027	02080110027601860083		00690089	020404360035	
420001962041	01060074	015201310040010301980005		00060010	0046	
420001962042	0095	0070	00230041011000960065	01320014	0127	
420001962051	00160145		00060195	01770032		0003
420001962052	002400320031	0025		0004		0059
420001962061					005001500004	
420001962062	0003					
420001962071						
420001962072	0092					0092
420001962081						
420001962082						
420001962091						
420001962092						
420001962101						
420001962102						
420001962111						
420001962112						
420001962121						
420001962122						
420001963011				0332006702810103		
420001963012	0732	00540050	0531	05700400050	0052	02030193

UNIVERSIDADE FEDERAL DA PARAIBA
 Pró-Reitoria Para Assuntos do Interior
 Coordenação Setorial de Pós-Graduação
 Rua Aprígio Viana, 102 - Tel. (031) 331-7222-R-355
 58.100 - Campina Grande - Paraíba

ESTATISTICA

NUMERO DE INFORMACOES LIDAS = 2002.0002

NO. DE PALAVRAS REQUEIRDAS P/ COMPRESSAO = 233.0000

APROX. 1.8621 BITS POR INFORMACAO

ESTATISTICA 2.1

BLOCO DE DADOS DE ENTRADA = TEXTO EM FORTRAN CONTENDO 24.336 CARACTERES

NUMERO DE INFORMACOES LIDAS = 24336.0039

NO. DE PALAVRAS REQUEURIDAS P/ COMPRESSAO = 2347.0004

APROX. 1.5430 BITS POR INFORMACAO

ESTATISTICA 2.2

BLOCO DE DADOS DE ENTRADA TEXTO EM PORTUGUES CONTENDO 12.000 CARACTERES

NUMERO DE INFORMACOES LIDAS = 12000.0019

NO. DE PALAVRAS REQUERIDAS P/ COMPRESSAO = 3153.0004

APROX. 4.2040 BITS POR INFORMACAO

ESTATISTICA 2.3

BLOCO DE DADOS DE ENTRADA = BANCO DE DADOS (SUDENE)
CONTENDO 50.050 CARACTERS

NUMERO DE INFORMACOES LIDAS = 50050.0079

NO. DE PALAVRAS REQUERIDAS P/ COMPRESSAO = 5720'0009

APROX. 1.8285 BITS POR INFORMACAO

ESTATISTICA 2.4

BLOCO DE DADOS DE ENTRADA = TEXTO EM FORTRAN CONTENDO 24336. CARACTERES

UTILIZACAO DE UNICO DICIONARIO DE SIMBOLOS

NUMERO DE INFORMACOES LIDAS = 24336.0039

NO. DE PALAVRAS REQUERIDAS = 2381.0004

APROX. 1.5654 BITS POR INFORMACAO

ESTATISTICA 2.5

BLOCO DE DADOS DE ENTRADA = TEXTO EM PORTUGUES CONTENDO 12000. CARACTERES

UTILIZACAO DE UNICO DICIONARIO DE SIMBOLOS

NUMERO DE INFORMACOES LIDAS = 12000.0019

NO. DE PALAVRAS REQUERIDAS = 3034.0004

APROX. 4.0453 BITS POR INFORMACAO

ESTATISTICA 2.6

BLOCO DE DADOS DE ENTRADA = BANCO DE DADOS CONTENDO 50050. CARACTERES

UTILIZACAO DE UNICO DICIONARIO DE SIMBOLOS

NUMERO DE INFORMACOES LIDAS = 50050.0079

NO. DE PALAVRAS REQUEIRDAS = 5695.0009

APROX. 1.8205 BITS POR INFORMACAO