

Projeto PI1

Implementação e atribuição do código LZ78 em uma
simulação de urna eletrônica

Grupo:

Karoline Figueirêdo

Raphael Rodrigues

Vitor Basile

Urna eletrônica



Na urna eletrônica será possível, por parte do eleitor, realizar votos em um determinado candidato distrital e federal em três regiões do Brasil. Sendo elas: Distrito Federal, Mato Grosso do Sul e Goiás.

Características da urna

- A urna será implementada na linguagem de programação C, por possuir uma maior robustez.
- Será implementada no sistema operacional Ubuntu 14.04, utilizando o compilador gcc, com a diretiva -std=c99.

Interface da urna

- Será possível realizar uma busca por candidatos distritais e federais.
- Busca de acordo com o partido do candidato.
- Busca de acordo com o nome do candidato para saber o seu número.

Interface da urna

- Multiplicação do voto (inicialmente um montante de 1 milhão de votos).
- Compressão e descompressão de dados utilizando o LZ78.

Porque do LZ78

O algoritmo LZ78 possui mais robustez por suportar um tempo e compressão otimizados de índices de texto para uma determinada correspondência de padrões. Além de possuir uma maior robustez possui também uma menor complexidade, possuindo $O(n \cdot \lg(n))$ no caso médio e $O(n)$ no pior caso, enquanto que o LZ77 possui $O(n)$ no caso médio.

Porque do LZ78

- É baseado em dicionário, e, por consequência, sua velocidade é maior em relação ao LZ77, que possui o funcionamento de janelas deslizantes. Por conseguinte, possui uma maior quantidade de compações em cada processo de codificação.

Porque do LZ78

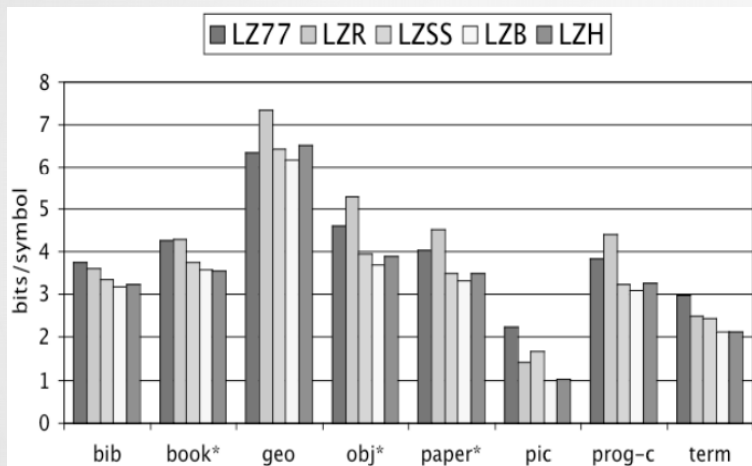


Figura 1. Família do LZ77 [2].

- bib: List of bibliography entries (ASCII)
- book: Fiction and non-fiction book (ASCII)
- geo: Geophysical data (32 bit numbers)
- obj: Executable files for VAX and Mac, mean value
- paper: Scientific papers (ASCII), mean value
- pic: Black and white bitmap image
- term: Terminal session (ASCII)

Porque do LZ78

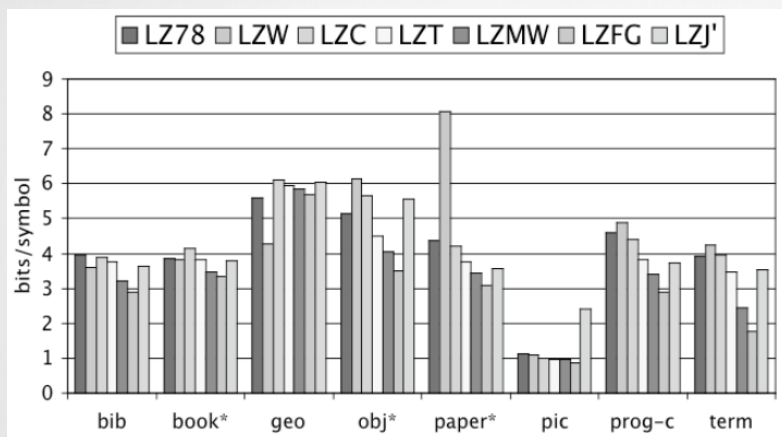


Figura 2. Família do LZ78 [2].

- bib: List of bibliography entries (ASCII)
- book: Fiction and non-fiction book (ASCII)
- geo: Geophysical data (32 bit numbers)
- obj: Executable files for VAX and Mac, mean value
- paper: Scientific papers (ASCII), mean value
- pic: Black and white bitmap image
- term: Terminal session (ASCII)

Porque do LZ78

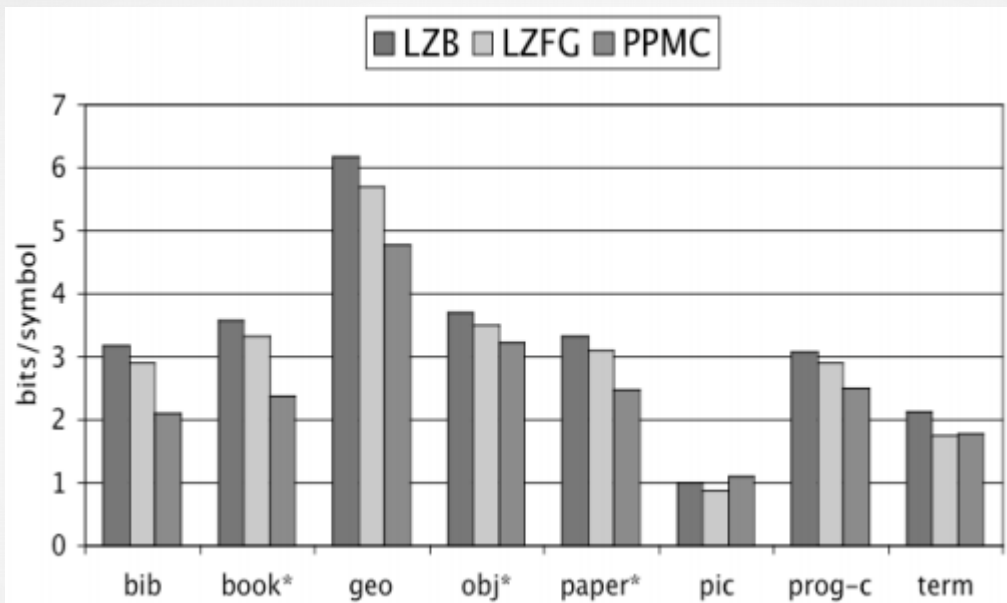


Figura 3. Comparação do LZ78 com o LZ77 [2].

Em que o LZB corresponde à família do LZ77 e o LZFG corresponde à família do LZ78.

Funcionamento do LZ78

- O LZ78 utiliza um dicionário para armazenar as sequencias encontradas no arquivo e usa os códigos. Exemplo com o texto “WaabbaaggabbaB”:

Índice	Caractere	Dicionário
0	W	(0,W)
1	a	(0,a)
2	ab	(1,b)
3	b	(0,b)
4	aa	(1,a)
5	g	(0,g)
6	ga	(5,a)
7	bb	(3,b)
8	aB	(1,B)

Taxa de compressão

- De acordo com Ferragina [4], a taxa de compressão esperada para textos contendo palavras comuns é de 33%.

Table I. Comparison of BPC for the different LZ77 variants

S. No.	File Name	File Size	LZ77	LZSS	LZH	LZB
			BPC	BPC	BPC	BPC
1	Bib	111261	3.75	3.35	3.24	3.17
2	Book1	768771	4.57	4.08	3.73	3.86
3	Book2	610856	3.93	3.41	3.34	3.28
4	News	377109	4.37	3.79	3.84	3.55
5	Obj1	21504	5.41	4.57	4.58	4.26
6	Obj2	246814	3.81	3.3	3.19	3.14
7	Paper1	53161	3.94	3.38	3.38	3.22
8	Paper2	82199	4.1	3.58	3.57	3.43
9	Progc	39611	3.84	3.24	3.25	3.08
10	Prog1	71646	2.9	2.37	2.2	2.11
11	Progp	49379	2.93	2.36	2.17	2.08
12	Trans	93695	2.98	2.44	2.12	2.12
Average BPC			3.88	3.32	3.22	3.11

Table II. Comparison of BPC for the different LZ78 variants

S. No.	File Name	File Size	LZ78	LZW	LZFG
			BPC	BPC	BPC
1	Bib	111261	3.95	3.84	2.9
2	Book1	768771	3.92	4.03	3.62
3	Book2	610856	3.81	4.52	3.05
4	News	377109	4.33	4.92	3.44
5	Obj1	21504	5.58	6.3	4.03
6	Obj2	246814	4.68	9.81	2.96
7	Paper1	53161	4.5	4.58	3.03
8	Paper2	82199	4.24	4.02	3.16
9	Progc	39611	4.6	4.88	2.89
10	Prog1	71646	3.77	3.89	1.97
11	Progp	49379	3.84	3.73	1.9
12	Trans	93695	3.92	4.24	1.76
Average BPC			4.26	4.90	2.89

Referências

- [1] D. Arroyuelo, R. Cánovas, G. Navarro, and R. Raman. LZ78 compression in low main memory space. SPIRE. 2017.
- [2] CHRISTINA ZEEH. The Lempel Ziv Algorithm. 2003.
- [3] SAYOOD, K. Introduction to Data Compression. Academic Press, San Diego, CA, 1996, 2000.
- [4] Paolo Ferragina. Magistral and informatica networking. 2014.