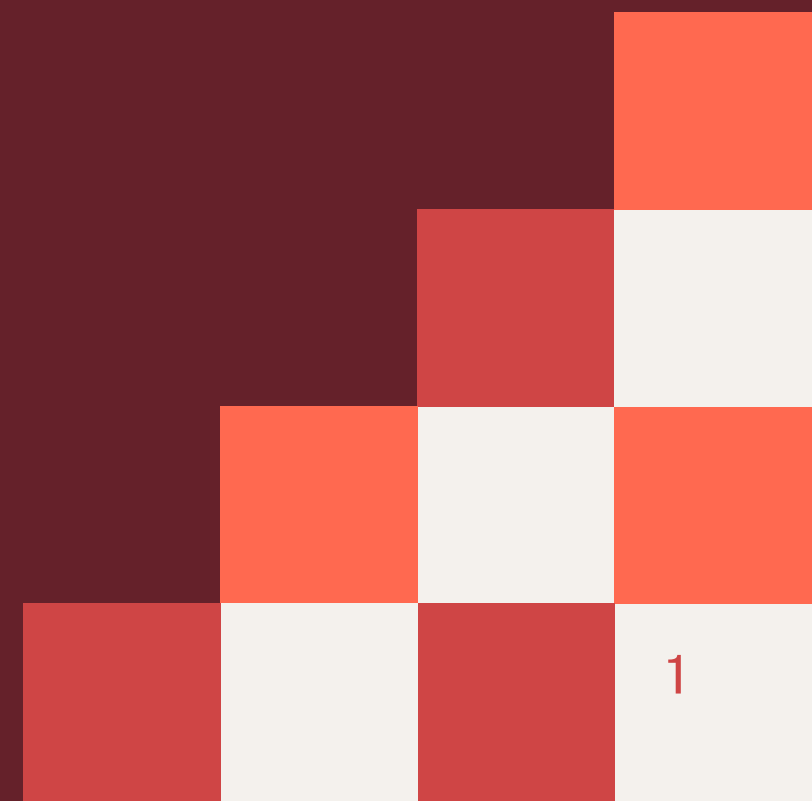# Compression and Archiving
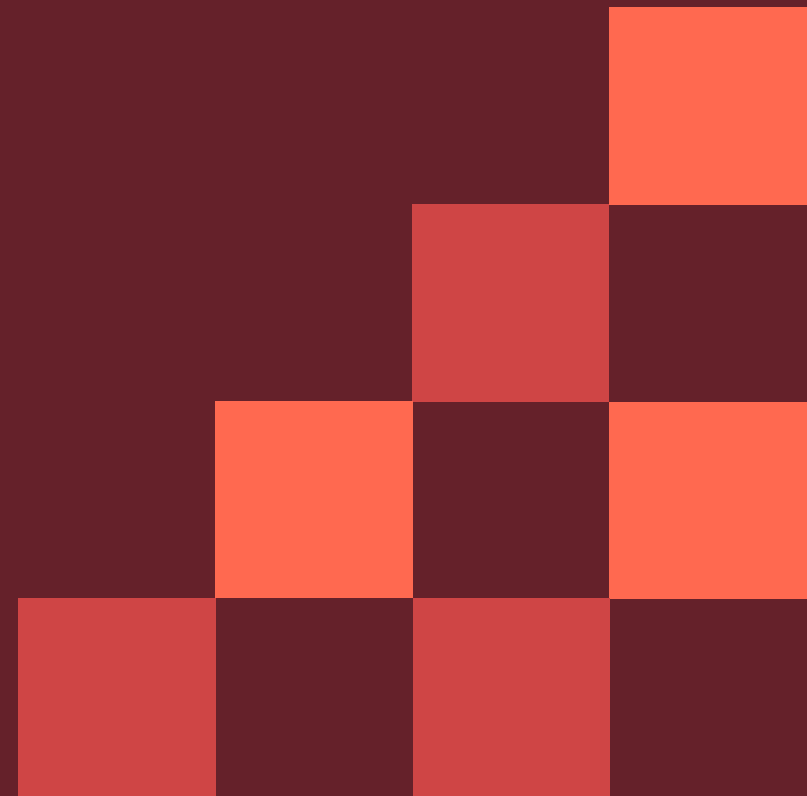
lets simplify..

Compression

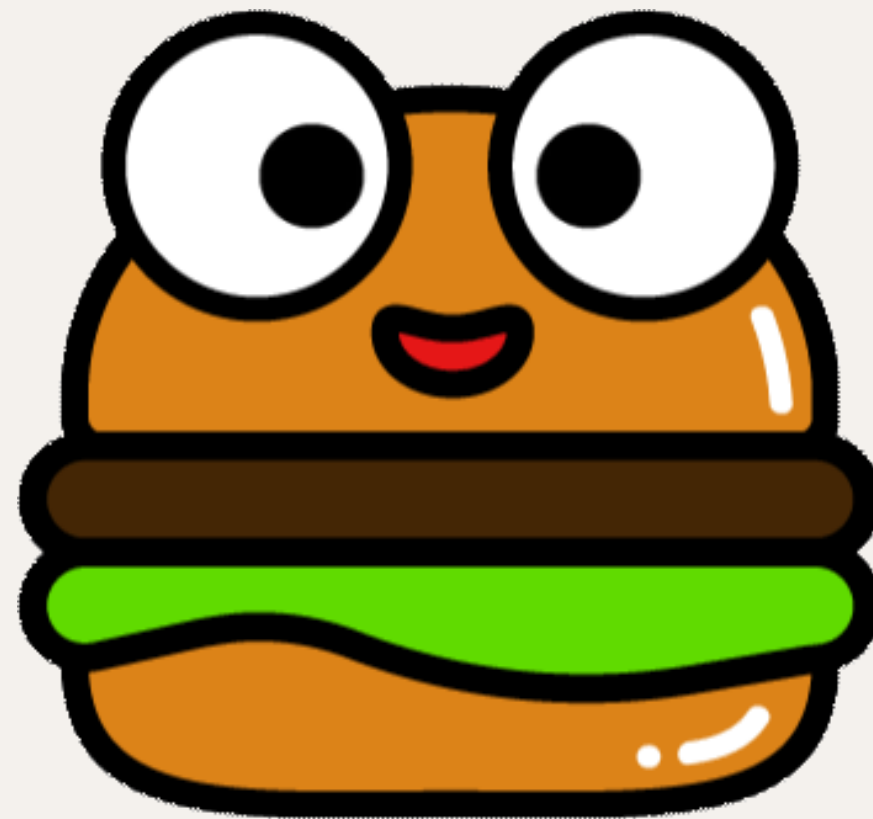getting rid of unnecessary things..
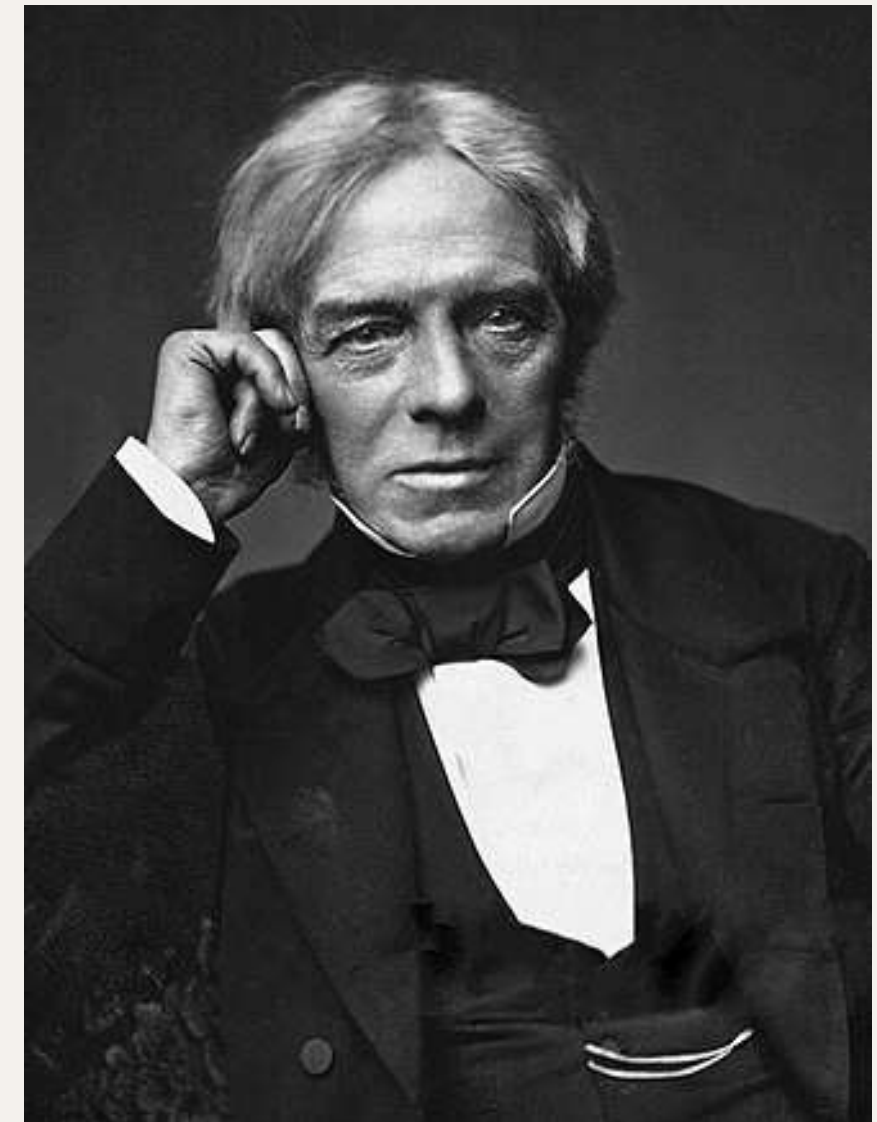
Run-Length Encoding (RLE)

RRRRRRRRRRGGGGGGGGBBBBBBB

10R8G7B
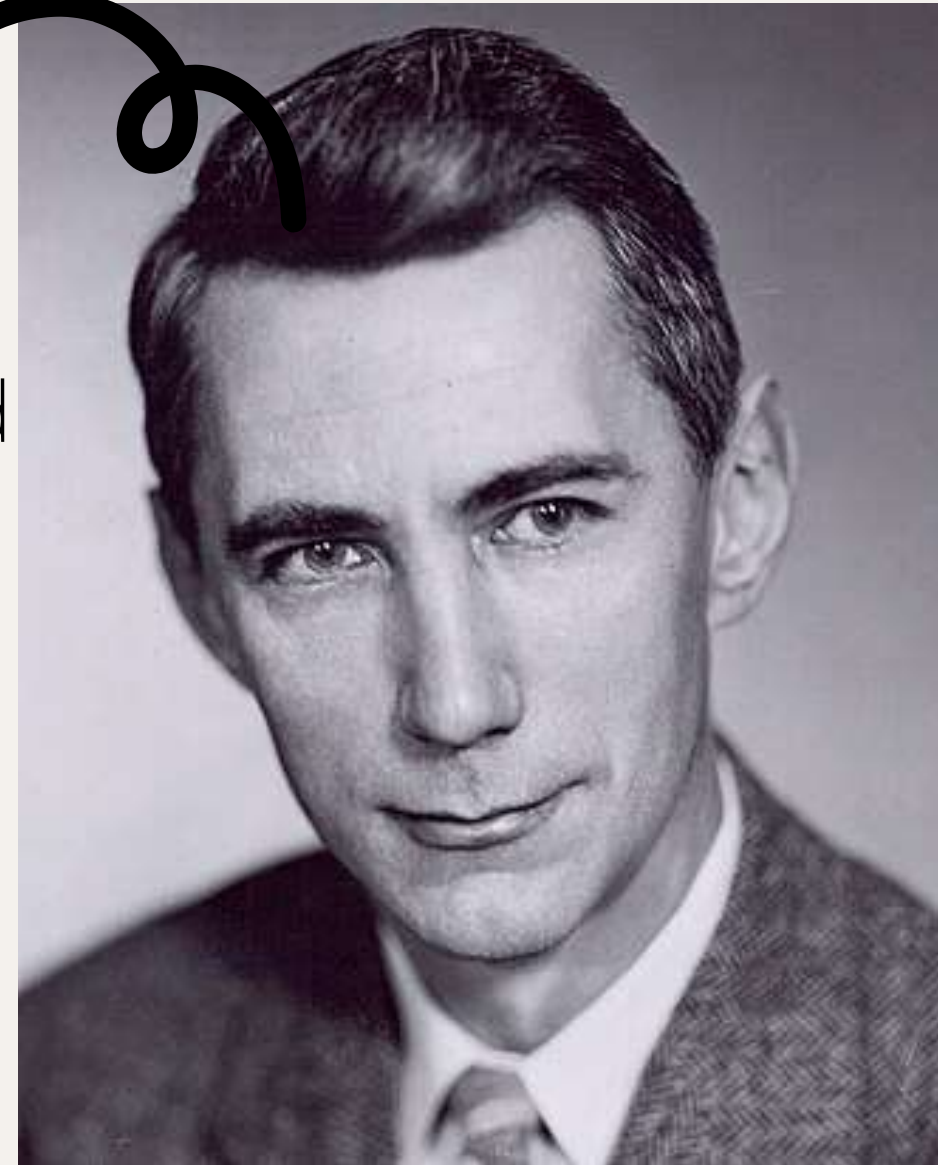
GIF

Michael Faraday

Shannon-Fano Coding

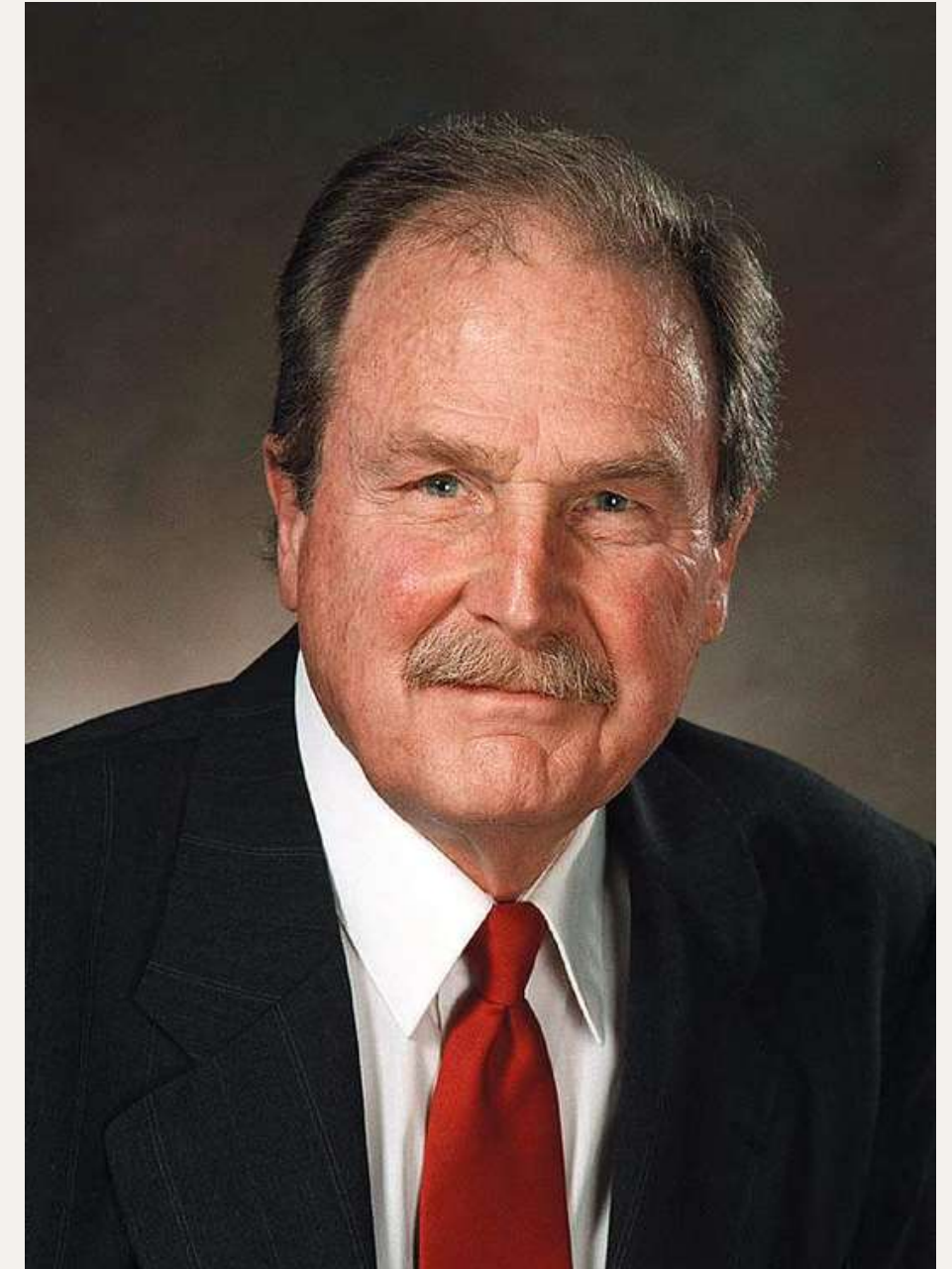He is the one who introduced the concept of a "bit"

Claude Shannon   Robert Fano

Huffman Coding

A graduate student solved an unsolved problem!

David Huffman

ARC

the first commercially successful archive
format released in 1985

- **Used: Huffman coding + LZW**

PKARC

Phil Katz just making some modifications on ARC code

Phil Katz

PKZIP

A tweaked version of PKARC

- Implode: (LZ77 + Shannon-Fano Coding).
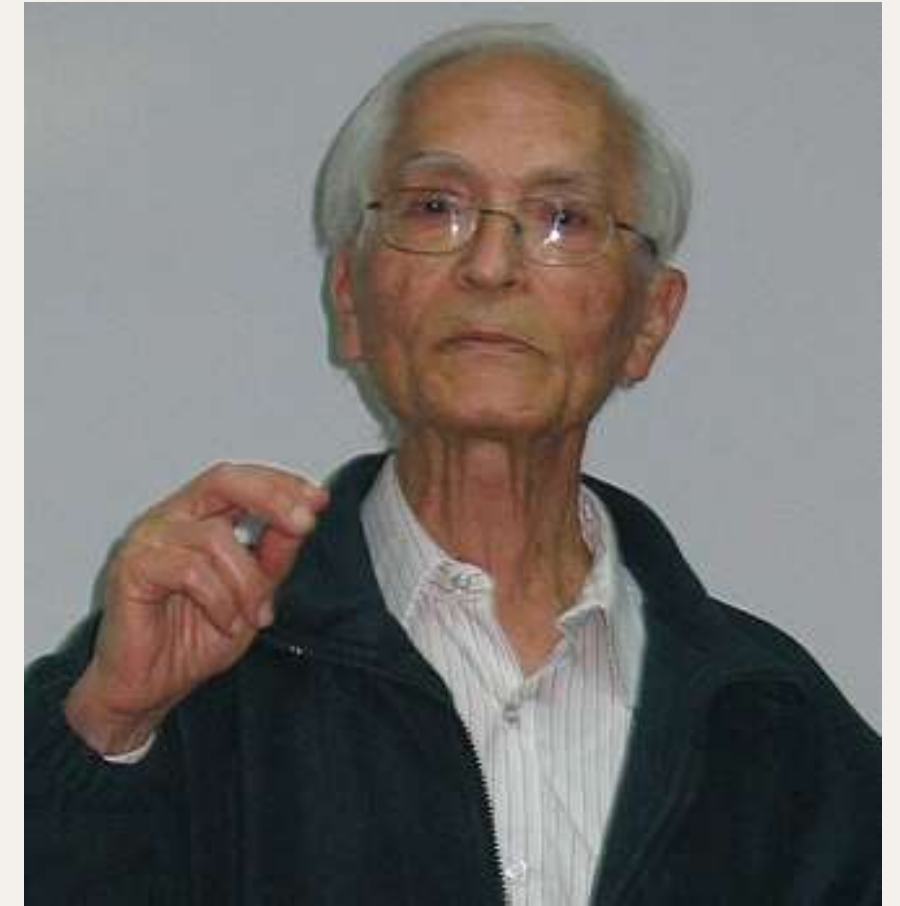- Deflate: (LZ77 + Huffman Coding).

Phil Katz

Lempel-Ziv (LZ)

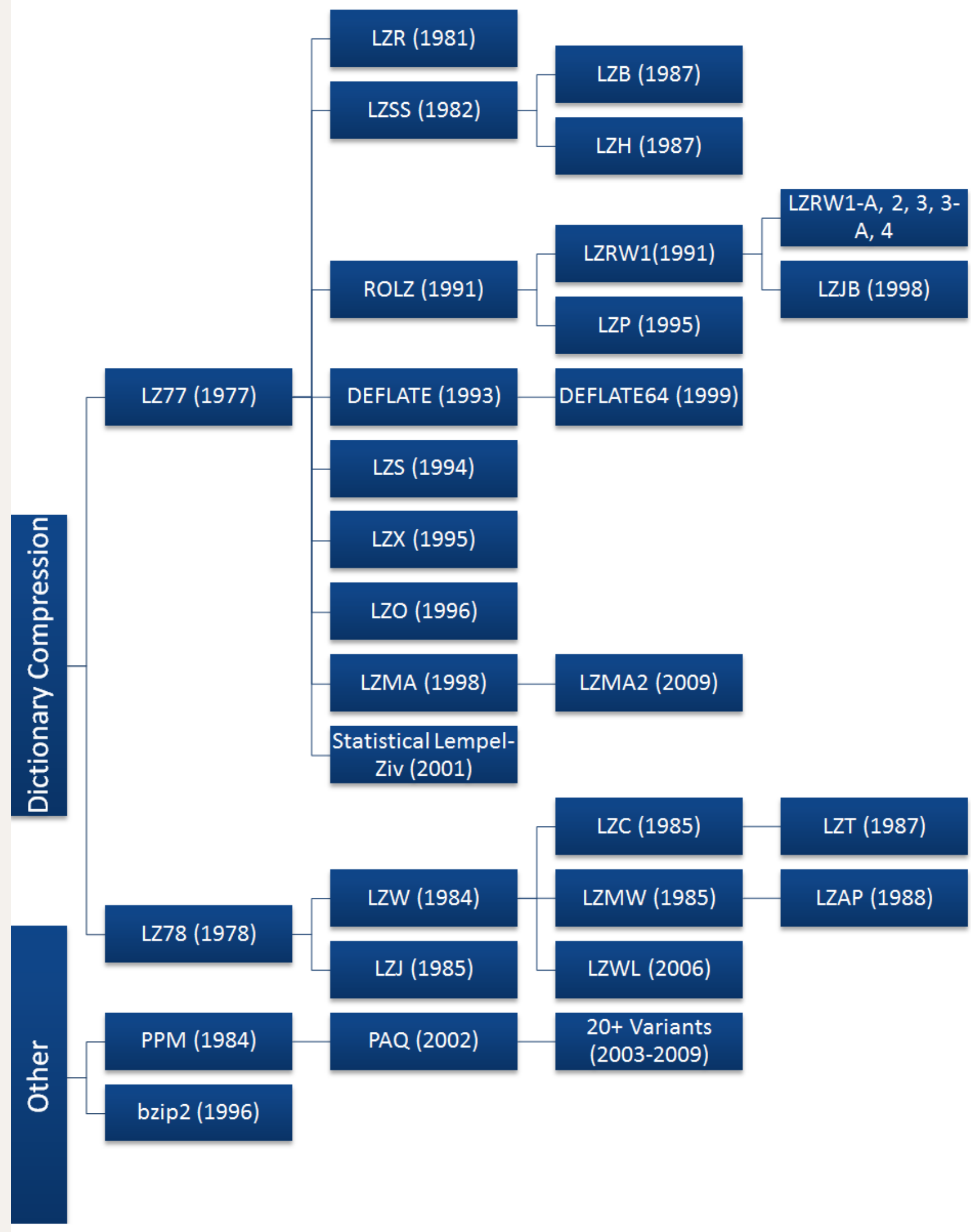Let's make a dictionary!

Abraham Lempel

Jacob Ziv

Polish-Jews who immigrated to the
Israeli Occupation in the 1950s

The LZ Family

Briefly elaborate on what you want to discuss.

GNU zip (GZIP)

A replacement for "compress"

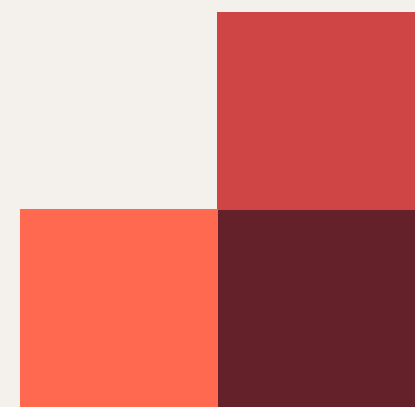Command Format: **gzip [options] <filename>**

Options:

- -d (decompress), you can use gunzip
- -f (force)
- -l (list stats)
- -r (recursive, for dealing with directories)

Optional:

- -v (verbose مطوَّل)

BZIP

uses Burrows-Wheeler Transform (BWT),
Huffman coding and Move-To-Front
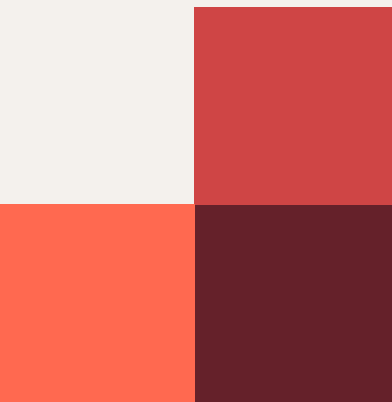
BZIP2

An improved version of BZIP

Command Format: **bzip2 [options] <filename>**

Options:

- -d (decompress), you can use bunzip2
- -f (force)
- -l (list stats)
- -r (recursive, for dealing with directories)
- -1..-9: 1 (fast), 9 (best), 6 is the default

Optional:

- -v (verbose مطوَّل)

TAR (**T**ape **Ar**chive**)**
The name just sticked to it

Command Format:
tar [options] -f <archive_filename> <file_to_archive>

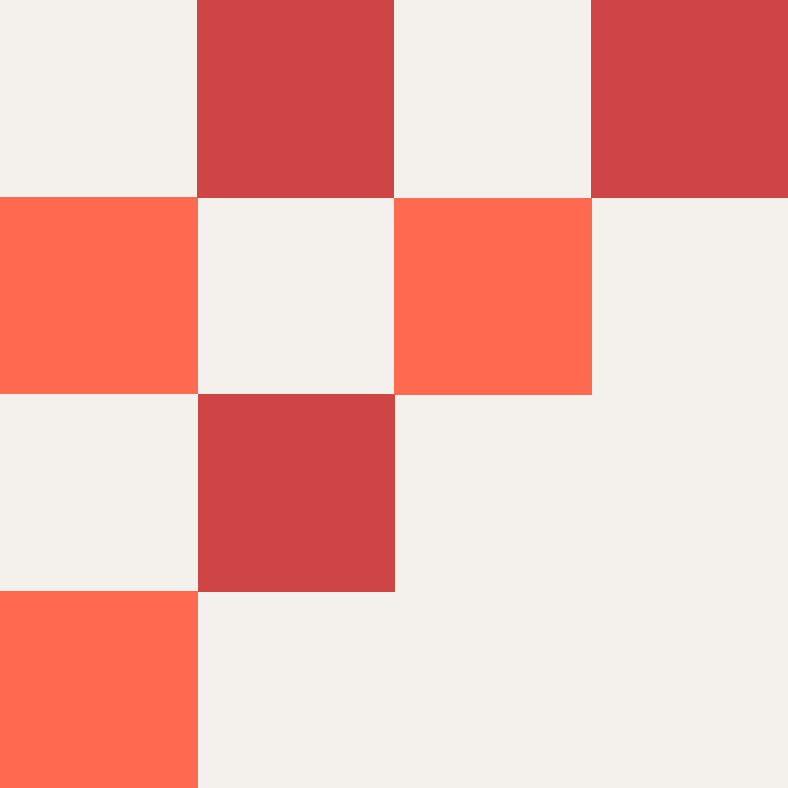### What do you want to do?

- -c (create)
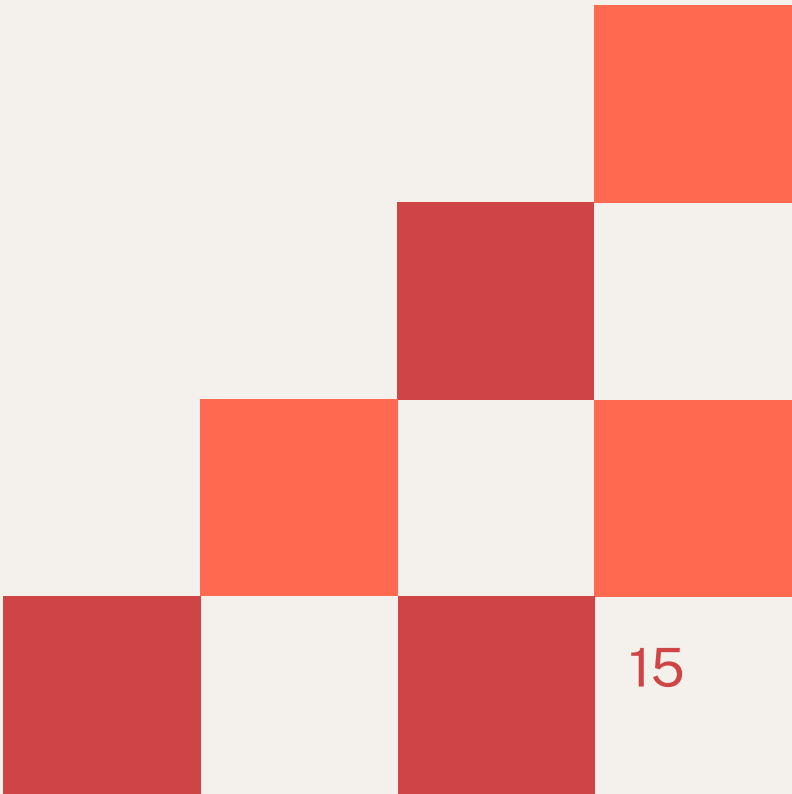- -x (extract)
- -r (append)
- -t (list content)

Compress?

- -z (gzip)
- -j (bzip2)

optional

- -v (verbose مطوَّل)

Tool/Format	Compression Algorithm	Utility	Year
TAR	No compression	Archiver	1979
ARC	LZW	Archiver + Compression	1985
PKZIP (ZIP format)	Implode then Deflate	Archiver + Compression	1989
GZIP	Deflate	Compression	1992
RAR	LZ-based	Archiver + Compression	1993
BZIP2	BWT + Huffman coding	Compression	1998
7-Zip	LZMA	Archiver + Compression	1999

"Information is the resolution of uncertainty."

— Attributed to Claude Shannon