Article Talk

Read Edit More ▾

Search

Lempel–Ziv–Storer–Szymanski

From Wikipedia, the free encyclopedia

Lempel–Ziv–Storer–Szymanski (**LZSS**) is a lossless data compression algorithm, a derivative of LZ77, that was created in 1982 by James Storer and Thomas Szymanski. LZSS was described in article "Data compression via textual substitution" published in *Journal of the ACM* (pp. 928–951).[1]

LZSS is a dictionary encoding technique. It attempts to replace a string of symbols with a reference to a dictionary location of the same string.

The main difference between LZ77 and LZSS is that in LZ77 the dictionary reference could actually be longer than the string it was replacing. In LZSS, such references are omitted if the length is less than the "break even" point. Furthermore, LZSS uses one-bit flags to indicate whether the next chunk of data is a literal (byte) or a reference to an offset/length pair.

Example [edit]

Here is the beginning of Dr. Seuss's Green Eggs and Ham, with character numbers at the beginning of lines for convenience.

```
  0: I am Sam
  9:
 10: Sam I am
 19:
 20: That Sam-I-am!
 35: That Sam-I-am!
 50: I do not like
 64: that Sam-I-am!
 79:
 80: Do you like green eggs and ham?
112:
113: I do not like them, Sam-I-am.
143: I do not like green eggs and ham.
```

This text takes 177 bytes in uncompressed form. Assuming a break even point of 2 bytes (and thus 2 byte pointer/offset pairs), and one byte newlines, this text compressed with LZSS becomes 94 bytes long:

```
  0: I am Sam
  9:
 10: (5,3) (0,4)
 16:
 17: That(4,4)-I-am!(19,16)I do not like
 45: t(21,14)
 49: Do you(58,5) green eggs and ham?
 78: (49,14) them,(24,9).(112,15)(93,18).
```

Note: this does not include the 12 bytes of flags indicating whether the next chunk of text is a pointer or a literal. Adding it, the text becomes 106 bytes long, which is much shorter than the original 177 bytes.

Implementations [edit]

Many popular archivers like PKZip, ARJ, RAR, ZOO, LHarc use LZSS rather than LZ77 as the primary compression algorithm; the encoding of literal characters and of length-distance pairs varies, with the most

common option being Huffman coding. Version 4 of the Allegro library can encode and decode an LZSS format,[2] but the feature was cut from version 5. The Game Boy Advance BIOS can decode a slightly modified LZSS format.[3]

See also   [edit]

- LZ77
- Lempel–Ziv–Welch (LZW)

References   [edit]

1. ^ Storer, James A.; Szymanski, Thomas G. (October 1982). "Data Compression via Textual Substitution". *Journal of the ACM* **29** (4): 928–951. doi:10.1145/322344.322346 .
2. ^ Hargreaves, Shawn, et al. Allegro source code: lzss.c, revision 7522 . Accessed on August 3, 2008.
3. ^ Korth, Martin. GBATEK: GBA BIOS Decompression Functions . Accessed on August 3, 2008.

v · t · e	Data compression methods		[hide]
Lossless	Entropy type	Unary · Arithmetic · Golomb · Huffman (Adaptive · Canonical · Modified) · Range · Shannon · Shannon–Fano · Shannon–Fano–Elias · Tunstall · Universal (Exp-Golomb · Fibonacci · Gamma · Levenshtein)	
	Dictionary type	Byte pair encoding · DEFLATE · Lempel–Ziv (LZ77 / LZ78 (LZ1 / LZ2) · LZJB · LZMA · LZO · LZRW · LZS · **LZSS** · LZW · LZWL · LZX · LZ4 · Statistical)	
	Other types	BWT · CTW · Delta · DMC · MTF · PAQ · PPM · RLE	
Audio	Concepts	Bit rate (average (ABR) · constant (CBR) · variable (VBR)) · Companding · Convolution · Dynamic range · Latency · Nyquist–Shannon theorem · Sampling · Sound quality · Speech coding · Sub-band coding	
	Codec parts	A-law · µ-law · ACELP · ADPCM · CELP · DPCM · Fourier transform · LPC (LAR · LSP) · MDCT · Psychoacoustic model · WLPC	
Image	Concepts	Chroma subsampling · Coding tree unit · Color space · Compression artifact · Image resolution · Macroblock · Pixel · PSNR · Quantization · Standard test image	
	Methods	Chain code · DCT · EZW · Fractal · KLT · LP · RLE · SPIHT · Wavelet	
Video	Concepts	Bit rate (average (ABR) · constant (CBR) · variable (VBR)) · Display resolution · Frame · Frame rate · Frame types · Interlace · Video characteristics · Video quality	
	Codec parts	Lapped transform · DCT · Deblocking filter · Motion compensation	
Theory		Entropy · Kolmogorov complexity · Lossy · Quantization · Rate–distortion · Redundancy · Timeline of information theory	
	Ⓦ Compression formats · Ⓦ Compression software (codecs)		

Categories:	Lossless compression algorithms