Project Proposal LZ4 vs. GZIP comparison

Alan Ghobadi
Fredrik Ingebrigtsen

1 Proposal

For this project the two lossless compression formats gzip [7] and LZ4 [4] will be compared in different experiments. Speed, compression ratio as well as resource allocation will be taken into account. Real world text data will be compared, such as JSON-documents. Gzip is a well established algorithm and is widely used. This project will compare gzip to the new contender LZ4 to see if there are any advantages of using LZ4.

1.1 Specifications

In conducting the experiments the following computer will be used: Macbook Pro, 2.8 GHz Intel Core i7, 8 GB 1600 MHz DDR3, OS X Yosemite.

The datasets compared will be structured documents, mainly JSON-documents of different kinds and HTML-documents.

2 Compression algorithms

Both of the lossless compression algrithms focuses primarily on speed over compression ratio.

2.1 Gzip

Gzip is a compression format based on the compression algorithm DEFLATE [1] wich combines LZ77 [6] and Huffman coding [2]. The format was developed by the GNU project for UNIX systems.

Gzip is commonly used in the real world for many different applications. It can be used by web servers to deliver compressed responses (if the browser-

supports this) to saved bandwidth. For example to deliver large JSON- or XML-documents over HTTP.

2.2 LZ4

The LZ4 compression algorithm is a relatively new compression algorithm based on previous LZ-algorithms and was first released in 2011 [4]. There is a reference implemetation in C [3] and there are also implemetations in other languages such as node.js [5].

References

[1] Deflate. `http://en.wikipedia.org/wiki/DEFLATE`, 2015. [Online; 2015-04-16].

[2] Huffman coding. `http://en.wikipedia.org/wiki/Huffman_coding`, 2015.[Online;2015-04-16].

[3] LZ4 C reference implementation. `https://github.com/Cyan4973/lz4`, 2015. [Online; 2014-12-17].

[4] LZ4 compression algorithm. `http://en.wikipedia.org/wiki/LZ4_(compression_algorithm)`, 2015. [Online; 2015-04-16].

[5] LZ4 node.js reference implementation. `https://github.com/pierrec/node-lz4`, 2015. [Online; 2014- 12-17].

[6] LZ77 and lZ78. `http://en.wikipedia.org/wiki/LZ77_and_LZ78`, 2015. [Online; 2015-04-16].

[7] Mark Adler Jean-loup Gailly. The gzip homepage. `http://www.gzip.org/`, 2015. [Online; 2015-04-16].