VIZ

Adaptive Huffman (FGK) compressor

VItali Denys & Spo**Z**io Cristian

Strutture dati

Node ed Huffman Tree

```
typedef struct Node {
   int node_number;
   int weight;
   int element;
   struct Node* left;
   struct Node* right;
   struct Node* parent;
} Node;
```

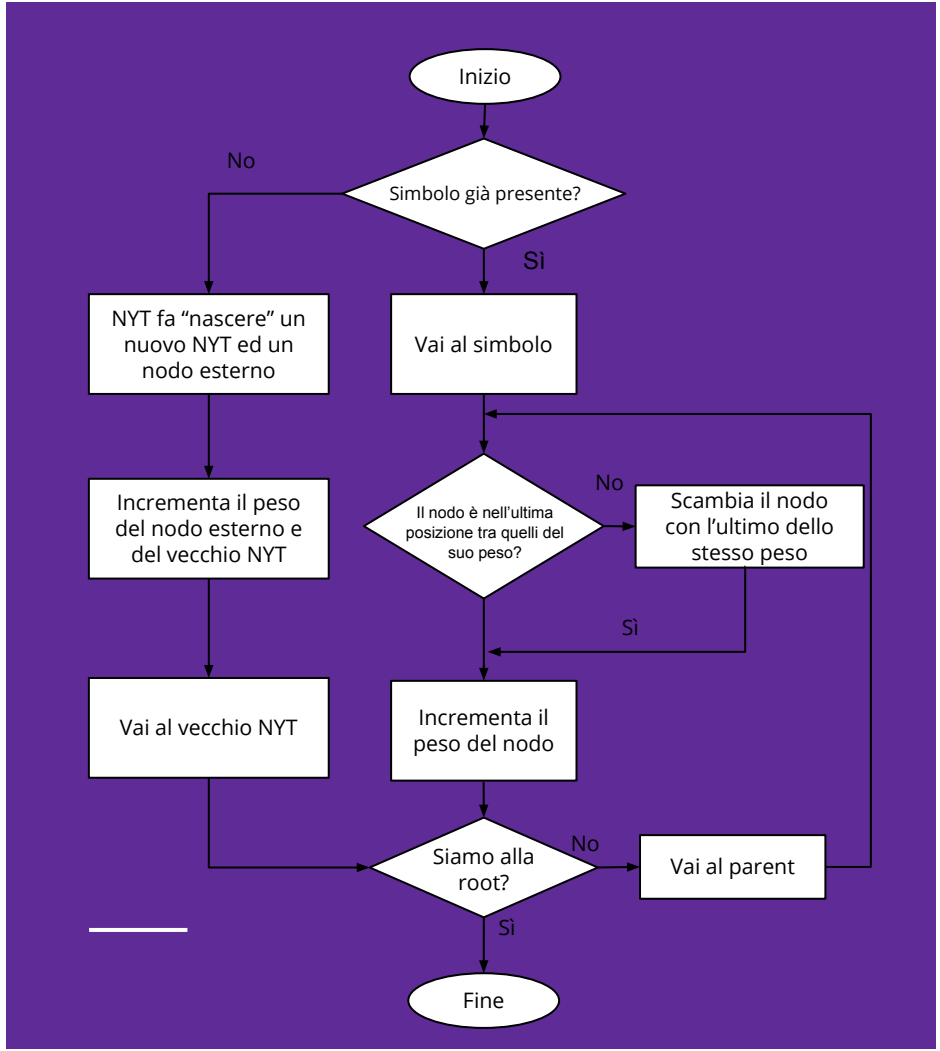
Node

```
typedef struct{
   Node* root;
   Node* tree[HUFFMAN_ARRAY_SIZE];  // 514
   Node* nyt;

   char* output;
   int output_length;

   char* partial_output;
   int partial_output_length;

   int elements;
   unsigned int mode;
   unsigned char mask;

   int decoder_flags;
   unsigned int decoder_bit;
   unsigned int decoder_has_bit;
   int decoder_byte;
} HuffmanTree;
```

Huffman Tree

Aggiornamento dei nodi

Diagramma di flusso



Implementazione dell'aggiornamento

```
HuffmanTree* add_new_element(HuffmanTree* ht, char c){
    Node* node = ht->root;
    Node* target = find_node(node, c);

    int* length = malloc(sizeof(int));
    *length = 0;

    int* path_length = malloc(sizeof(int));
    unsigned short* path;

    if(target != NULL){
        debug("[add_new_element] AS");
        path = node_path(target, path_length);
        node_positioner(ht, target);
        if(is_compressor(ht)){
            huffman_append_partial_path(ht, path, *path_length);
        }
    } else {
        path = node_path(ht->nyt, path_length);

        if(is_compressor(ht)) {
            if (ht->elements == 0) {
                ht->output[0] = c;
                ht->output_length = 1;
            } else {
                huffman_append_partial_new_element(ht, path, *path_length, c);
            }
        }

        ht->elements++;
        Node* old_nyt = ht->nyt;
```

```
        if(old_nyt == NULL){
            exit(51);
        }

        Node* new_nyt = createNYT(old_nyt->node_number - 2);
        Node* new_char = createNode(old_nyt->node_number - 1, 1, c, NULL, NULL, old_nyt);

        old_nyt->weight++;
        old_nyt->left = new_nyt;
        old_nyt->right = new_char;
        old_nyt->element = -1;

        ht->nyt = new_nyt;

        new_nyt->parent = old_nyt;
        new_char->parent = old_nyt;

        ht->tree[new_nyt->node_number] = new_nyt;
        ht->tree[new_char->node_number] = new_char;
        target = old_nyt;
    }
    free(path);
    free(path_length);
    free(length);

    while(target != ht->root){
        if(target == NULL || target->parent == NULL){
            return NULL;
        }
        target = target->parent;
        node_positioner(ht, target);
    }
    return ht;
}
```

Node Positioner e Highest Numbered Node

```
void node_positioner(HuffmanTree* ht, Node* target){
   if(target == NULL){
      return;
   }
   Node* last = highest_numbered_node(ht, target);
   char buffer[250];

   if(last != target && last != target->parent) {
      swap_nodes(ht, target, last);
   }
   if(last != target->parent || target->parent == ht->root)
      target->weight++;
}
```

Node Positioner

```
Node* highest_numbered_node(HuffmanTree* ht, Node* node){
   int i;

   Node* highest = node;
   if(node == NULL){
      return NULL;
   }
   for(i=node->node_number+1; i<HUFFMAN_ARRAY_SIZE; i++){
      if(ht->tree[i] != NULL){
         if(ht->tree[i]->weight == node->weight){
            highest = ht->tree[i];
         }
      }
   }
   return highest;
}
```

Highest Numbered Node

Swap nodes

```
void swap_nodes(HuffmanTree* ht, Node* node, Node* node2){
  if(node == NULL || node2 == NULL){
     // Null Pointer Exception
     return;
  }
  if(node->parent == NULL || node2->parent == NULL){
     // Not going to swap a root.
     return;
  }
  if(node == node2){
     // Not going to swap two identical nodes.
     return;
  }

  if(node2->parent == node || node->parent == node2){
     error("[swap_nodes] I can't swap a child with its parent");
     exit(2);
  }

  Node* parent1 = node->parent;
  Node* parent2 = node2->parent;

  Node* parent1_left = parent1->left;
  Node* parent2_left = parent2->left;
```

```
if(parent1_left == node){
     parent1->left = node2;
  } else {
     parent1->right = node2;
  }

  if(parent2_left == node2){
     parent2->left = node;
  } else {
     parent2->right = node;
  }

  // Swap Array
  ht->tree[node2->node_number] = node;
  ht->tree[node->node_number] = node2;

  // Fix Node Numbers
  int nn = node->node_number;
  node->node_number = node2->node_number;
  node2->node_number = nn;

  node->parent = parent2;
  node2->parent = parent1;

  debug("[swap_nodes] End swap");
}
```

Split dei byte

```
void huffman_append_partial_path(HuffmanTree* ht, unsigned short*
path, int path_size){
  if(ht != NULL) {
     debug("[huffman_append_partial_path]");
     int i;
     for(i=0; i<path_size; i++){
        if(path[i]) {
           ht->partial_output[ht->partial_output_length] |= ht->mask;
        }
        huffman_coding_bitcheck(ht);
     }

  }
}
```

```
void huffman_coding_bitcheck(HuffmanTree* ht){
  if(ht->mask == 0x01){
     ht->mask = 0x80;

     ht->output[ht->output_length] = ht->partial_output[0];
     ht->partial_output[0] = 0;
     ht->output_length++;

     ht->partial_output_length = 0;
     huffman_coding_reset_partial_output(ht);
  } else {
     ht->mask >>= 1;
  }
}
```

Bytestream

Part 1

0x49 0x10 0x0E 0x71 0x94 0x3c

Part 2

0x63 0xAC 0x30 0xA3

Bitstream

Node path
Element

Part 1

0100 1001	0001 0000	0000 1110	0111 0001	1001 0100	0011 1100
0x49	0x10	0x0E	0x71	0x94	0x3C
I	0( )	00(s)	011 100(e)	00	0(x)

Part 2

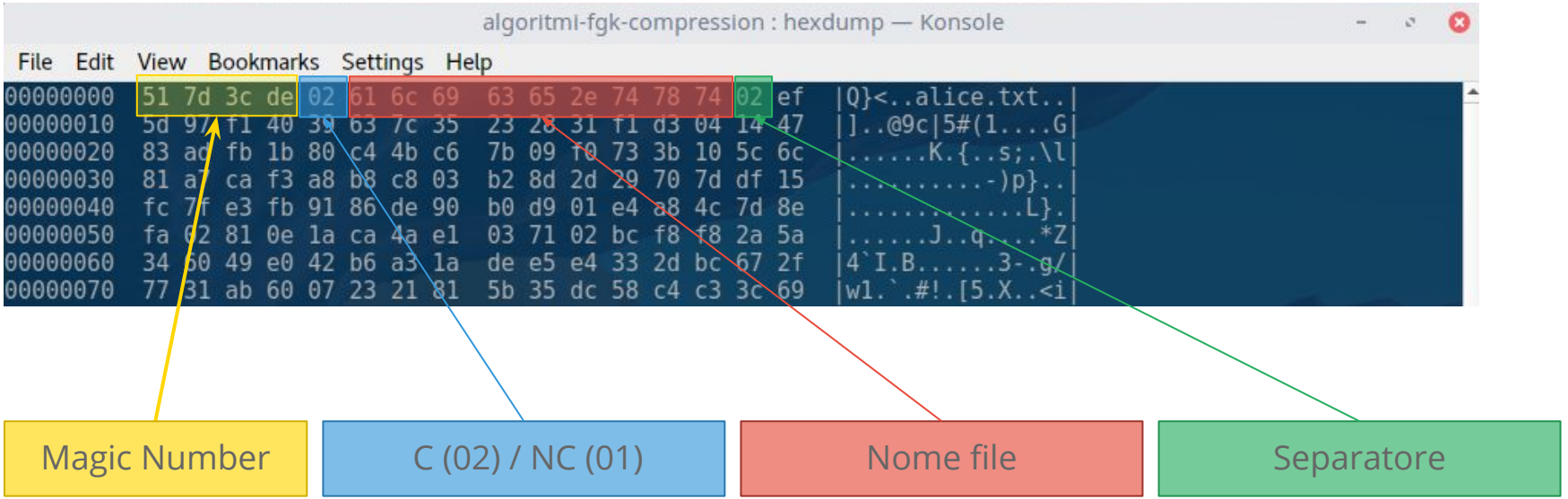
0110 0011	1010 1100	0011 0000	10100011	01100000	10000011
0x63	0xAC	0x30	0xA3	0x60	0x83
0 1100(u)	100	0(a)			

Part 3

11001010	11000011
0xCA	0xC3
(u)	

Struttura file .viz

Struttura del file



Testing

Framework

minunit.h - a minimal unit testing framework for C

```
// http://www.jera.com/techinfo/jtns/jtn002.html
#include "console.h"

#define mu_assert(message, test) do { if (!(test)) return message; } while (0)
#define mu_run_test(test) do { char *message = test(); tests_run++; \
                        if (message) {error_test_fail(message); return message;} test_info_end(); } while (0)
#define mu_tag(tag) test_info(tag);
extern int tests_run;
```

Versione modificata di minunit.h

I nostri test

```
static char * all_tests(){
    mu_run_test(test_debug);

    mu_run_test(test_create_huffman_tree);
    //mu_run_test(test_swap_ht_array);

    mu_run_test(test_huffman_coding_bookkeeper);
    mu_run_test(test_huffman_coding_mississippi);
    mu_run_test(test_huffman_coding_engineering);
    mu_run_test(test_huffman_coding_foobar);
    //mu_run_test(test_huffman_coding_foobar2000);
    //mu_run_test(test_huffman_coding_loremipsum);

    mu_run_test(test_get_level);
    mu_run_test(test_get_node_level);
    mu_run_test(test_simple_swap);
    mu_run_test(test_swap_nodes);
    mu_run_test(test_node_path);
    mu_run_test(test_huffman_coding);
    //mu_run_test(test_utility_siblings);
    mu_run_test(test_huffman_coding);
    mu_run_test(test_huffman_coding_abracadabra);
    mu_run_test(test_huffman_coding_abcbaaa);
```

```
    mu_run_test(test_huffman_coding_mississippi);
    mu_run_test(test_huffman_coding_engineering);
    mu_run_test(test_huffman_coding_aardvark);
    mu_run_test(test_huffman_coding_sleeplessness);
    mu_run_test(test_bin2byte);
    mu_run_test(test_bin2byte2);
    mu_run_test(test_byte2bin);
    mu_run_test(test_filename);


    // File ops. Run in sequence!
    mu_run_test(test_create_file);
    mu_run_test(test_write_to_file);
    mu_run_test(test_read_file);
    mu_run_test(test_file_delete);

  return 0;
}
```

Esempio di test (test_byte2bin)

```c
static char* test_byte2bin(){
    mu_tag("Byte2Bin");
    unsigned short* result;
    result = byte2bit('\xff');
    unsigned short* expected_result = (unsigned short[8]){1,1,1,1,1,1,1,1};
    mu_assert("FF is not 11111111", compare_short_int(result, expected_result, 8));
    free(result);

    result = byte2bit('\xfa');
    expected_result = (unsigned short[8]){1,1,1,1,1,0,1,0};
    mu_assert("FA is not 11111010", compare_short_int(result, expected_result, 8));
    free(result);

    result = byte2bit('\x0a');
    expected_result = (unsigned short[8]){0,0,0,0,1,0,1,0};
    mu_assert("0A is not 00001010", compare_short_int(result, expected_result, 8));
    free(result);

    result = byte2bit('\x00');
    expected_result = (unsigned short[8]){0,0,0,0,0,0,0,0};
    mu_assert("00 is not 00000000", compare_short_int(result, expected_result, 8));
    free(result);

    return 0;
}
```

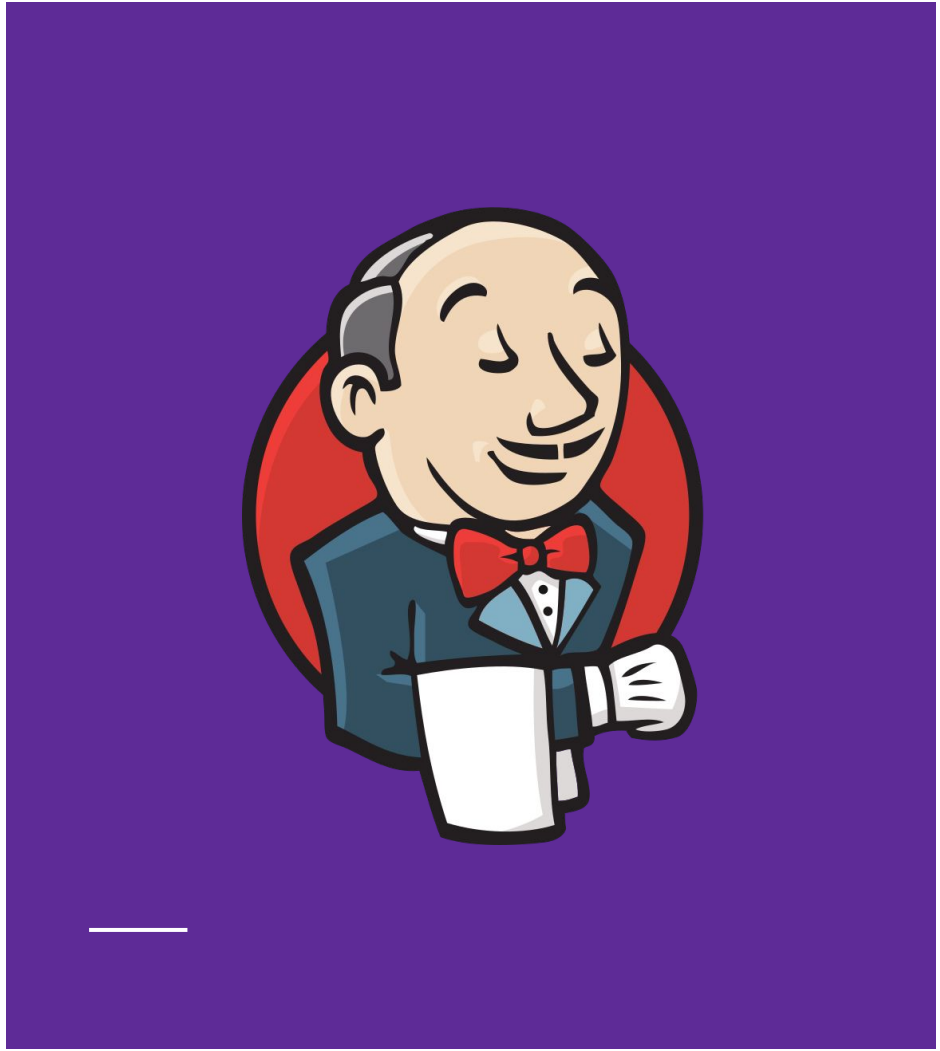
```
...
➡ Testing "Huffman Coding (mississippi)"
⬆ □ Test completed

➡ Testing "Huffman Coding (engineering)"
⬆ □ Test completed

➡ Testing "Huffman Coding (aardvark)"
⬆ □ Test completed

➡ Testing "Huffman Coding (sleeplessness)"
⬆ □ Test completed

➡ Testing "Write to file"
⬆ □ Test completed

➡ Testing "Read file"
File size is: 5
⬆ □ Test completed

➡ Testing "Delete file"
⬆ □ Test completed

✓ ALL TESTS PASSED
Tests run: 27
```

Problemi riscontrati

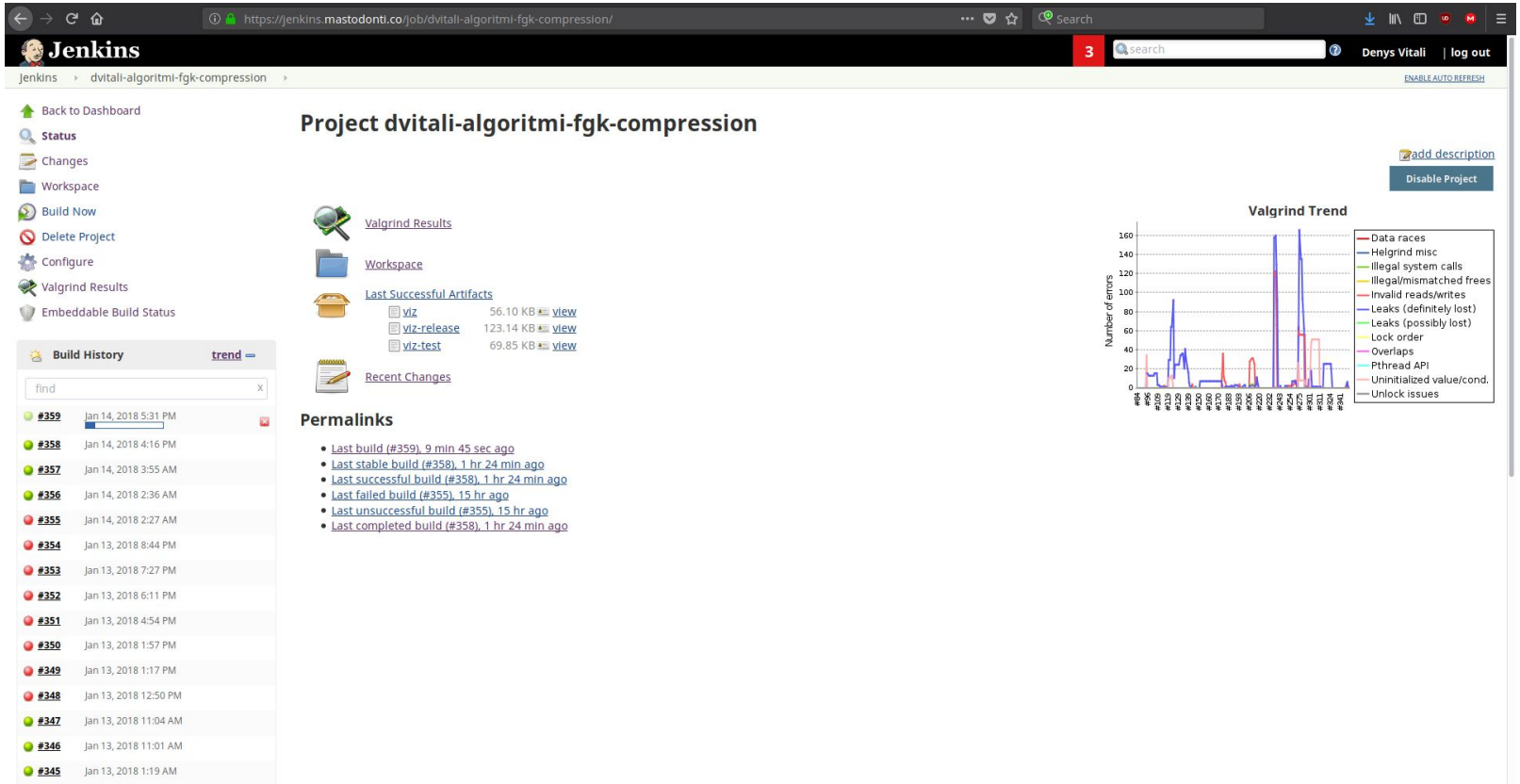
- Swap dei nodi nell’ HT Array
- Split dei byte
- Decompressione (WIP)

Jenkins

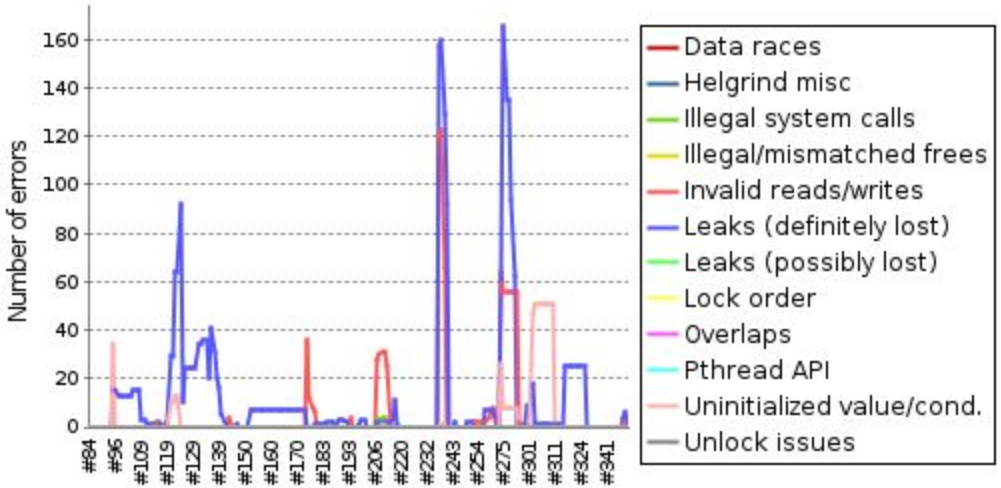
Continuous Integration
&
Automation



Automatic Builds

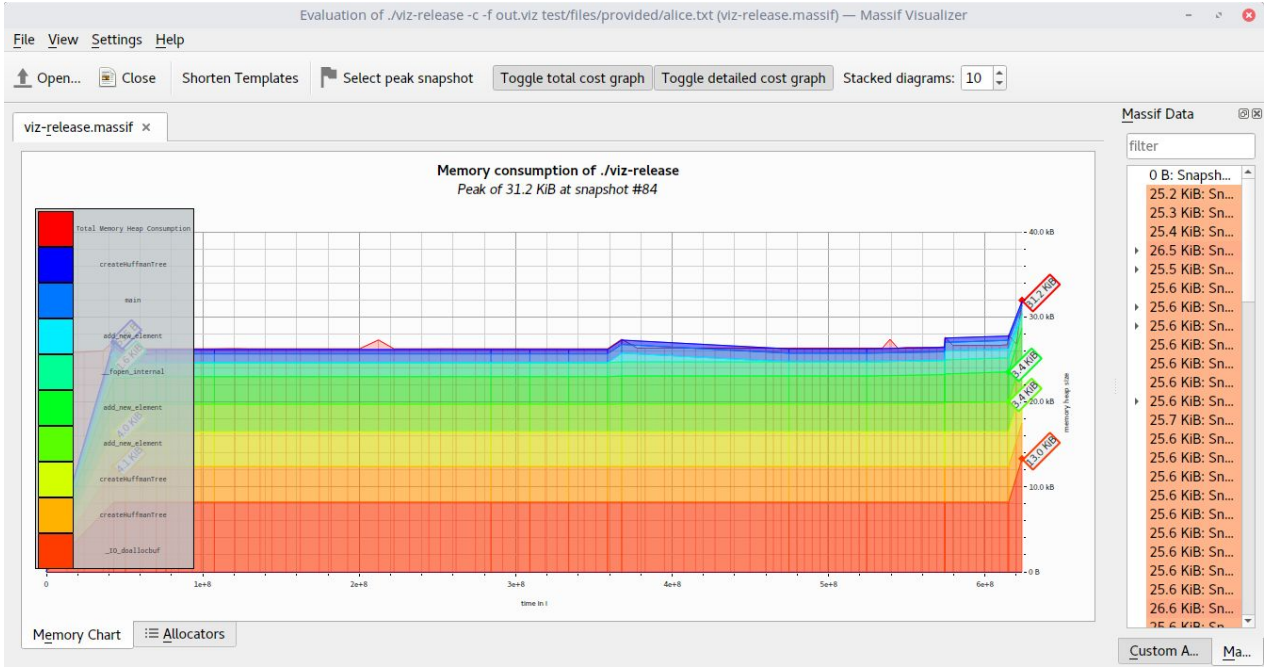


Valgrind


Number of errors
0
20
40
60
80
100
120
140
160
#84
#96
#109
#119
#129
#139
#150
#160
#170
#183
#193
#206
#220
#232
#243
#254
#275
#301
#311
#324
#341
Data races
Helgrind misc
Illegal system calls
Illegal/mismatched frees
Invalid reads/writes
Leaks (definitely lost)
Leaks (possibly lost)
Lock order
Overlaps
Pthread API
Uninitialized value/cond.
Unlock issues

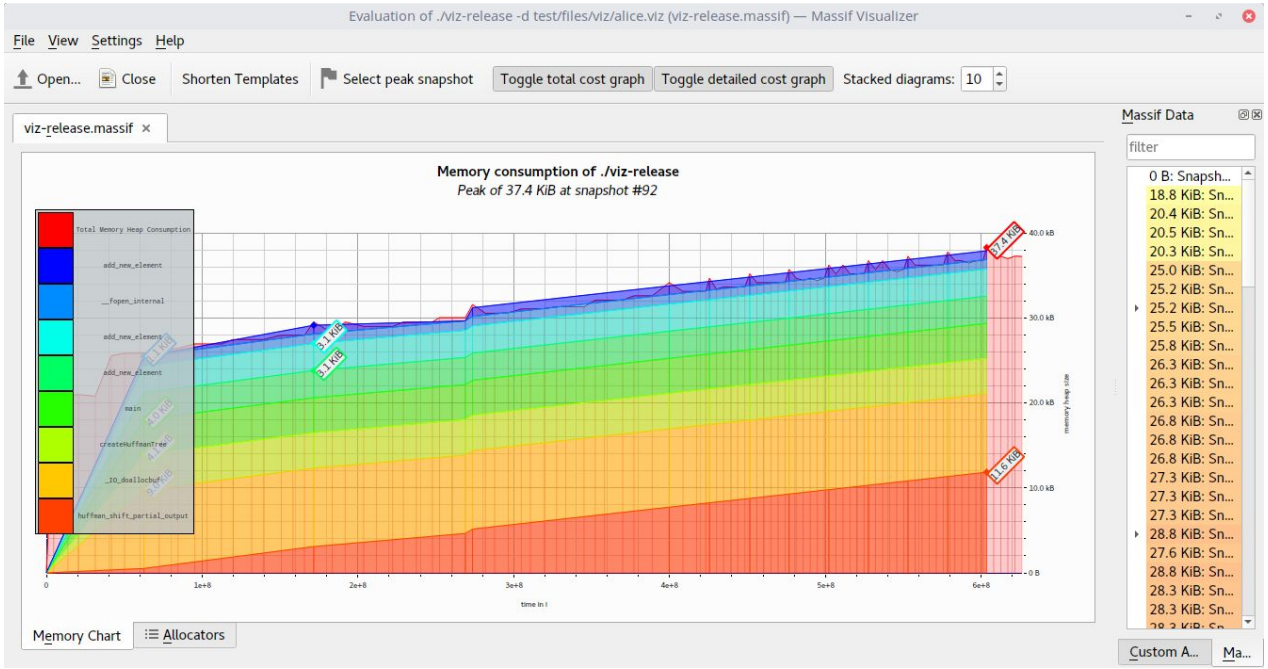
Massif - Memory Visualizer

make massif_release (make massif_release_c)



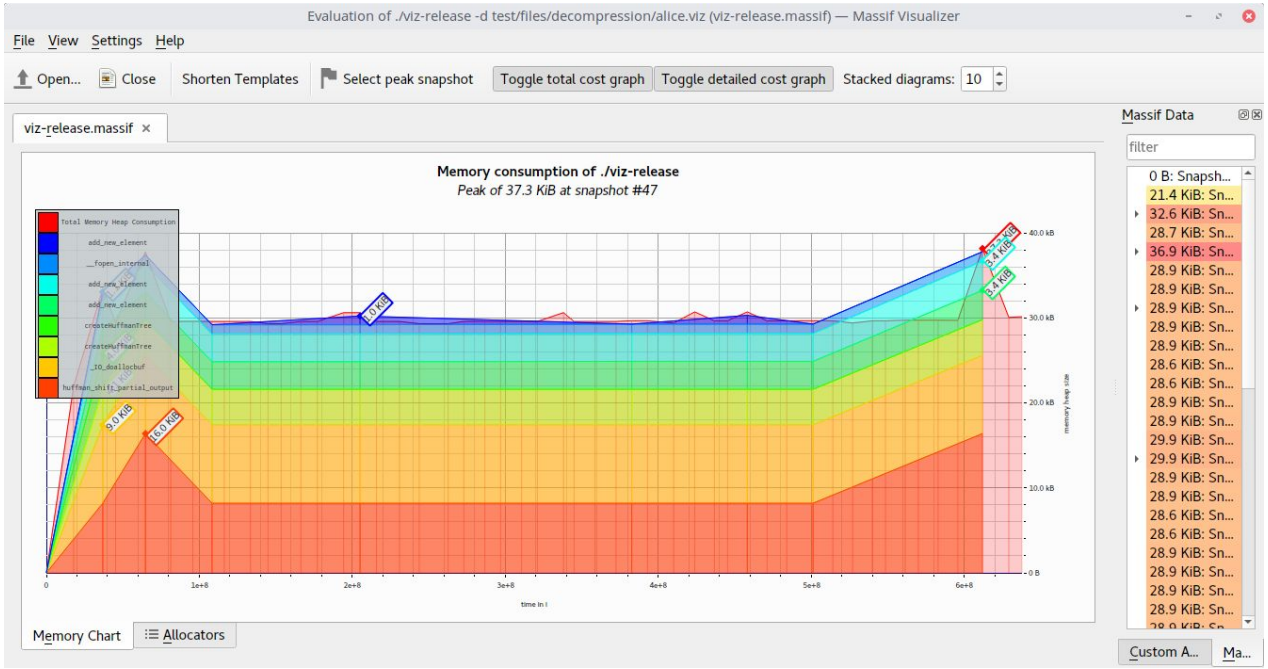
Massif - Memory Visualizer

make massif_release_d



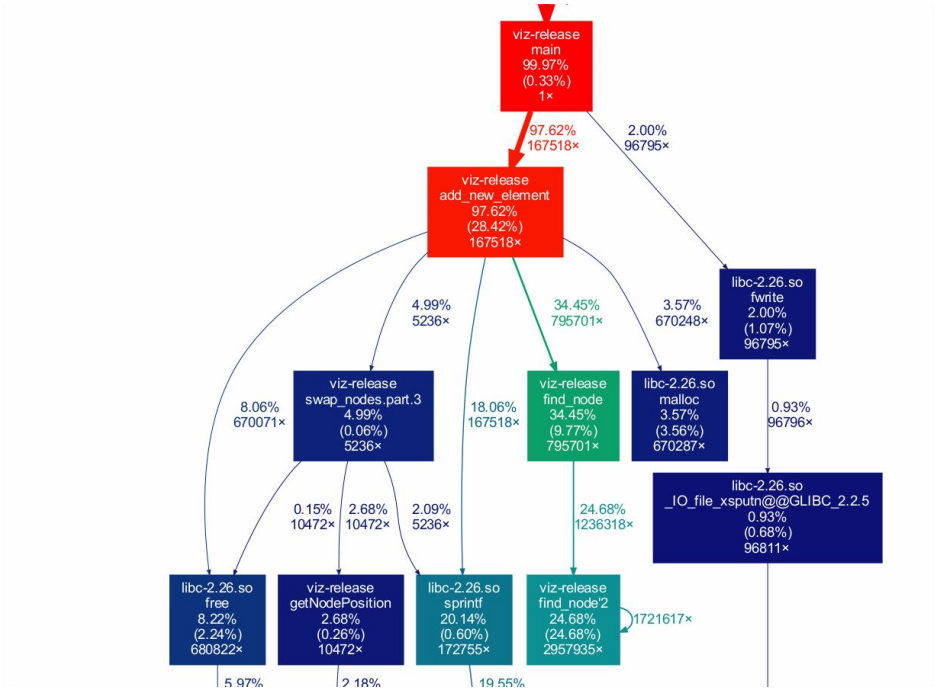
Massif - Memory Visualizer

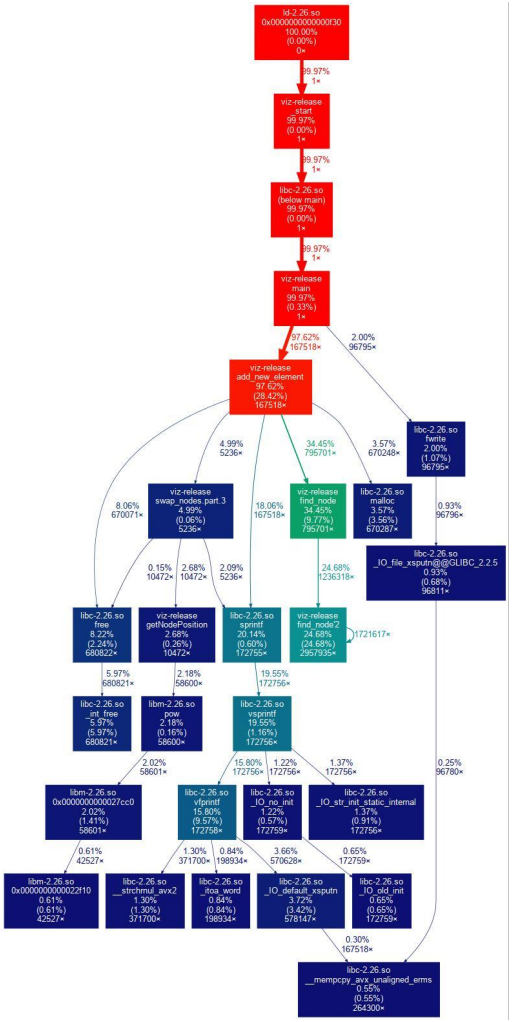
make massif_release_d - after optimization



Callgrind - Call graph

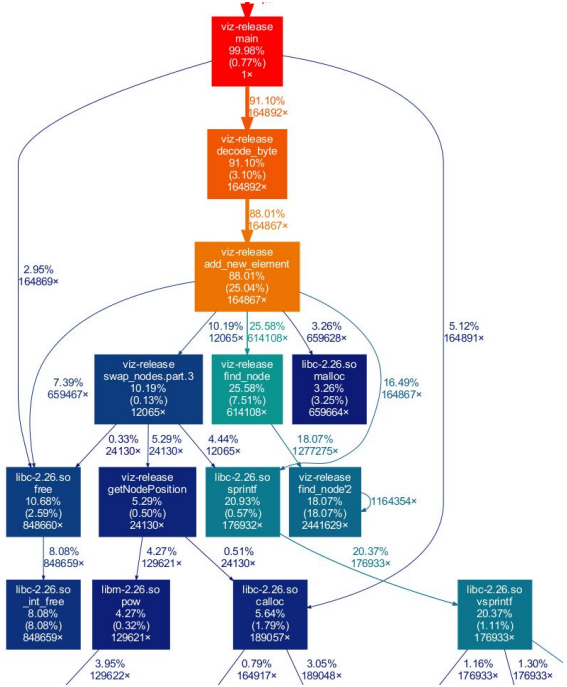
make callgrind_release (make callgrind_release_c)

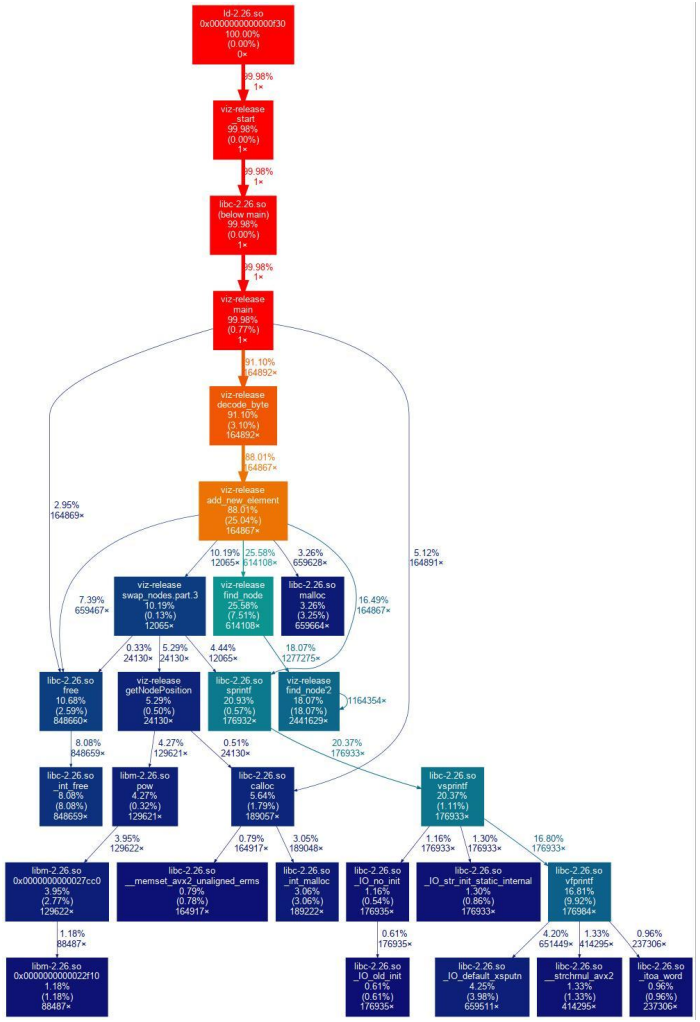




Callgrind - Call graph

make callgrind_release_d





Benchmark

800 kB/s

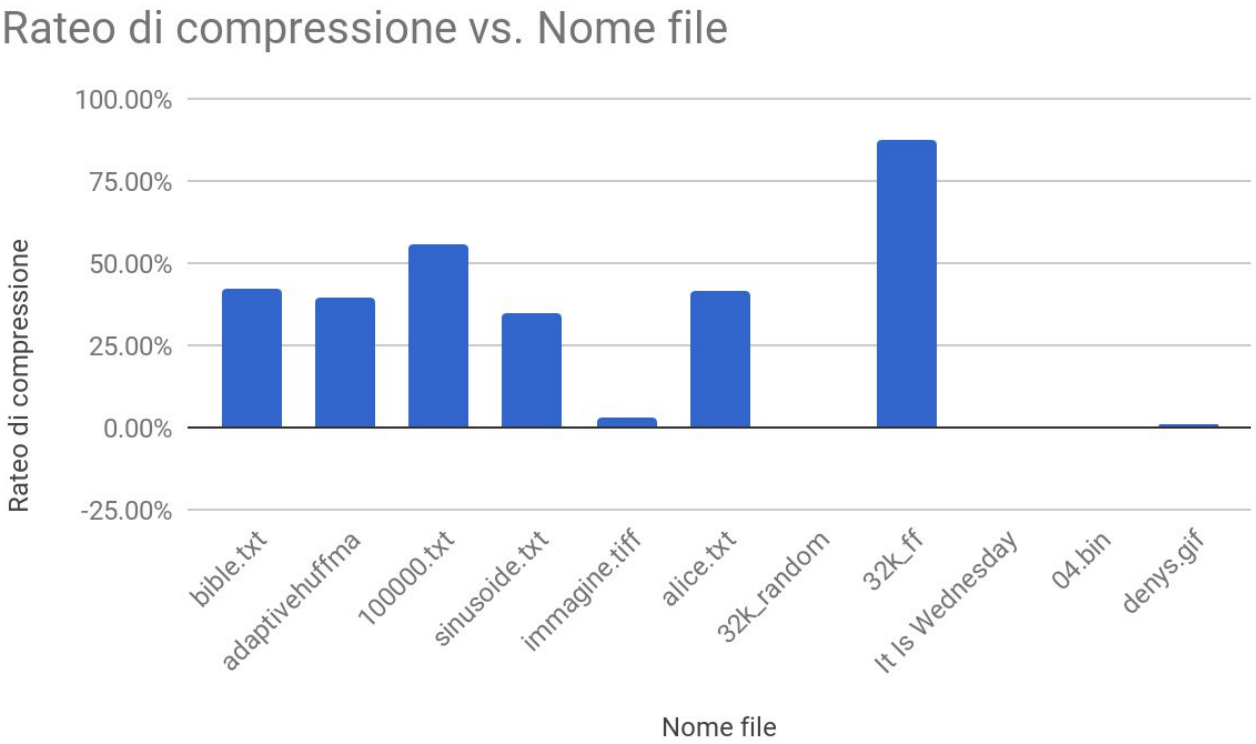
Velocità media di compressione

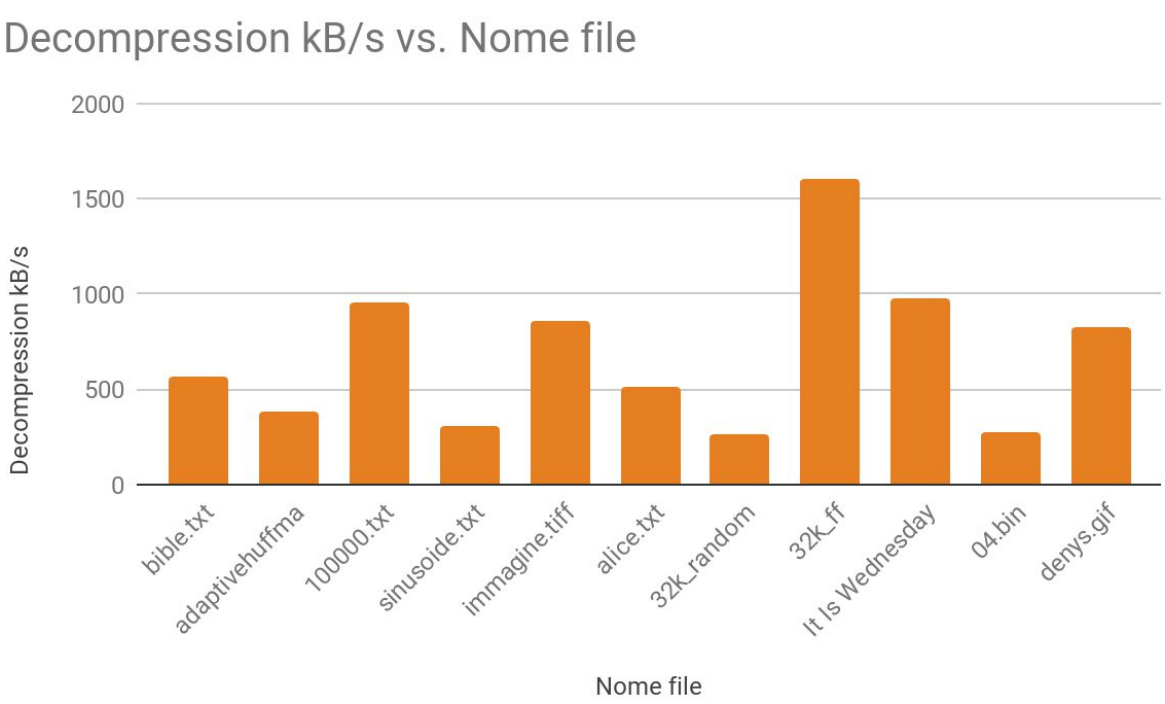
147MB compressi in 189.42s (153875545/189.42/1024 = 793.311 kB/s)
Intel Core i7-6700HQ @ 8x 3.5GHz, 32GB RAM, x86_64 Linux 4.14.13-1

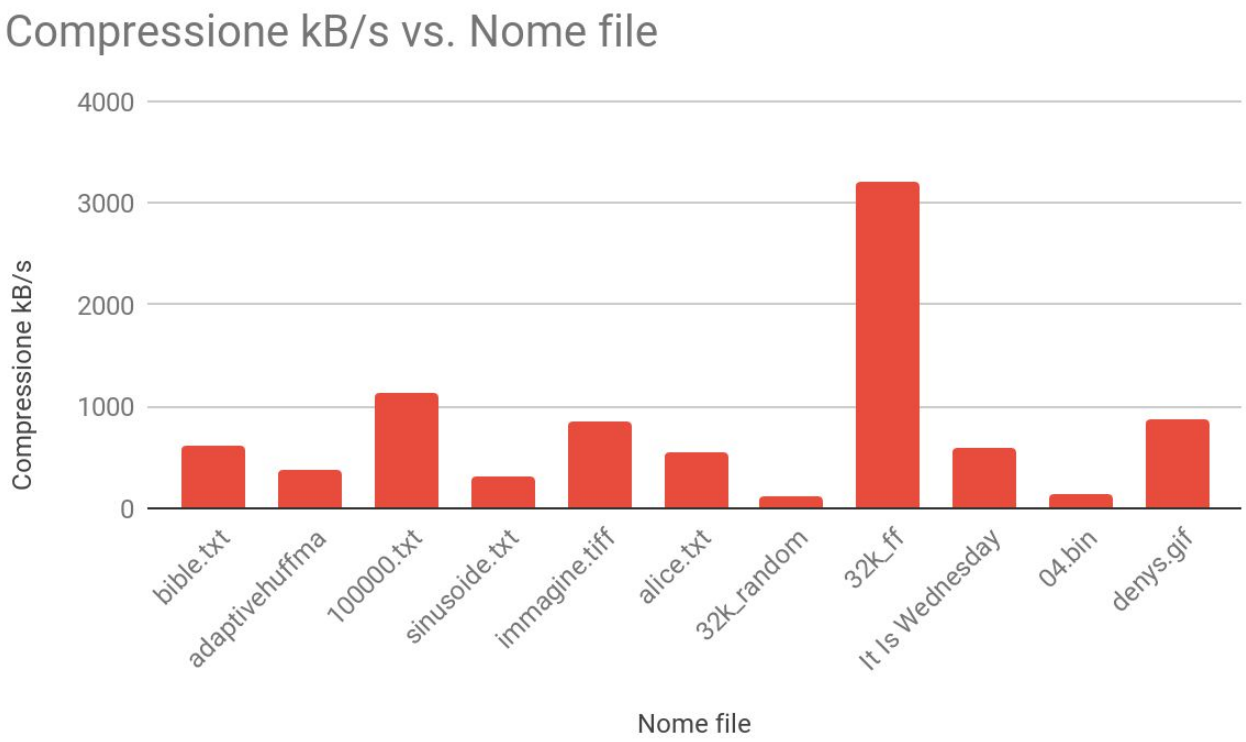
35%

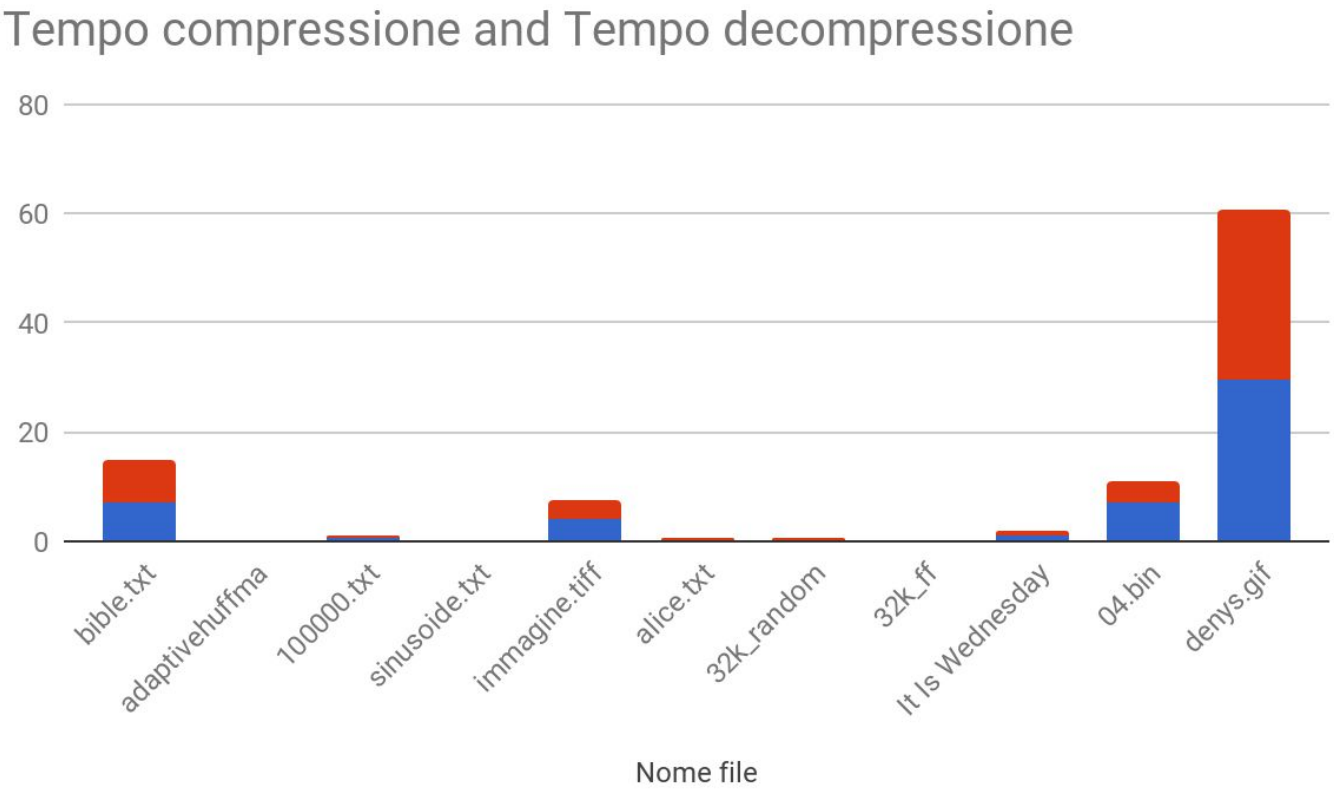
Fattore di compressione medio

*Ottenuto con una media del rapporto di compressione dei file:
immagine.tiff, fitnessgram.txt, adaptivehuffman.txt, alice.txt, bible.txt, 32k_ff


Rateo di compressione vs. Nome file
100.00%
75.00%
50.00%
25.00%
0.00%
-25.00%
Rateo di compressione
bible.txt
adaptivehuffma
100000.txt
sinusoide.txt
immagine.tiff
alice.txt
32k_random
32k_ff
It Is Wednesday
04.bin
denys.gif
Nome file


Decompression kB/s vs. Nome file
2000
1500
1000
500
0
Decompression kB/s
bible.txt
adaptivehuffma
100000.txt
sinusoide.txt
immagine.tiff
alice.txt
32k_random
32k_ff
It Is Wednesday
04.bin
denys.gif
Nome file


Compressione kB/s vs. Nome file
4000
3000
2000
1000
0
Compressione kB/s
bible.txt
adaptivehuffma
100000.txt
sinusoide.txt
immagine.tiff
alice.txt
32k_random
32k_ff
It Is Wednesday
04.bin
denys.gif
Nome file


Tempo compressione and Tempo decompressione
80
60
40
20
0
bible.txt
adaptivehuffma
100000.txt
sinusoide.txt
immagine.tiff
alice.txt
32k_random
32k_ff
It Is Wednesday
04.bin
denys.gif
Nome file


60% of the time,
it works every time.