

# Snappy format

09 November 2019 09:29

The text

Wikipedia is a free, web-based, collaborative, multilingual encyclopedia project.

may be compressed to this, shown as hex data with explanations:

```
0000000: ca02 f042 5769 6b69 7065 6469 6120 6973 ...B Wikipedia is
```

The first 2 bytes, `ca02` are the length, as a little-endian varint (see Protocol Buffers for the varint specification)<sup>[10]</sup>. Thus the most-significant byte is '02'. `0x02ca(varint) = 0x014a = 330` bytes. The next two bytes, `0xf042`, indicate that a literal of `66+1` bytes follows

```
0000010: 2061 2066 7265 652c 2077 6562 2d62 6173   a free, web-
bas
0000020: 6564 2c20 636f 6c6c 6162 6f72 6174 6976   ed,
collaborativ
0000030: 652c 206d 756c 7469 6c69 6e67 7561 6c20   e, multilingual
0000040: 656e 6379 636c 6f09 3f0 8170 726f 6a65   encyclo?...proje
```

`0x09` is tag-byte of type `01` with length  $-4 = 010_2 = 2_{10}$  and offset  $= 0x03f = 63$  or "pedia"; `0xf081` is a literal with length of `129+1` bytes

```
0000050: 6374 2e00 0000 0000 0000 0000 0000 0000   ct.
```

Varint

ca 22  $\Rightarrow$   $\begin{matrix} c & a & & 0 & & 2 \\ \textcircled{1}100 & 1010 & | & 0000 & & 0010 \end{matrix}$

Se il most significant bit (MSB) è 1 sai che deve arrivare almeno un altro byte

Se è 0 sai che questo è l'ultimo byte del varint

Per decodificare bisogna togliere l'MSB e invertire l'ordine con cui hai ricevuto i byte quindi

~~0000~~ 0010 ~~100~~ 1010

$= \underline{10100} \underline{1010} = 0x14A = 330 \text{ bytes}$

**LITERAL** Dopo il primo varint che indica la quantità di bytes trasmessi sai che viene sempre il primo literal.

f042  $\Rightarrow$   $\begin{matrix} f & 0 & 4 & 2 \\ 1111 & 0000 & 0100 & 0010 \end{matrix}$

0x42  $\Rightarrow$   $\overset{1}{1} \overset{1}{1} \overset{1}{1} \overset{1}{1} \overset{0}{0} \overset{0}{0} \overset{0}{0} \overset{0}{0} \overset{0}{0} \overset{1}{1} \overset{0}{0} \overset{0}{0} \overset{0}{0} \overset{1}{1} \overset{0}{0}$

len-1 del literal  
se  $< 60$  ma qui  
è  $= 60$  quindi length  
del literal è nel  
prossimo bytes

tag byte di  
un literal

$\hookrightarrow 0x42 = 66$  bytes

**COPY** Dopo i 66 bytes di literal ti  
aspetti un nuovo tag byte

0x09  $\Rightarrow$   $\overset{0}{0} \overset{0}{0} \overset{0}{0} \overset{0}{0} \overset{1}{1} \overset{0}{0} \overset{0}{0} \overset{1}{1}$

len-4  $\Rightarrow 2+4=6$  bytes (len della copia)

Copia del primo tipo

Offset di 11 bits: 3 del tag byte e 8 del byte  
successivo (0x3f)  $\Rightarrow 000001111111 = 63$  bytes di offset

Torni indietro di 63 bytes e ricopi 6 bytes  
(in questo caso "padding").

Ora ti aspetti un nuovo tag byte e decodifichi  
0xf0 (literal) e così via.