



Tecnológico de Monterrey

Tecnológico de Monterrey - Campus Monterrey
School of Engineering and Sciences
Engineering in Computational Technologies
Analysis and Design of Advanced Algorithms

Homework 8: Huffman Compression

Group: 607
Team #3

Luis Salomón Flores Ugalde

Santiago Quintana Moreno A01571222
Miguel Ángel Álvarez Hermida a01722925

Homework8 >  Hw8_HuffmanCompression.py > ...

```
File Edit Selection View Go Run Terminal Help ← → 2.Advanced Algorithms
```

```
Hw8_HuffmanCompression.py U X
```

```
Homework8 > Hw8_HuffmanCompression.py > ...
```

```
72 def suffix_array(data: bytes):
81     sa.sort(key=lambda i: (rank_at(i), rank_at(i+k)))
82     tmp[sa[0]] = 0
83     for i in range(1, n):
84         a, b = sa[i-1], sa[i]
85         tmp[b] = tmp[a] + (rank_at(a) != rank_at(b) or rank_at(a+k) != rank_at(b+k))
86     rank = tmp[:]
87     if rank[sa[-1]] == n-1:
88         break
89     k <<= 1
90     return sa
91
92 # ----- BWT -----
93 def bwt_from_sa(T: bytes, SA):
94     n = len(T)
95     out = bytearray(n)
96     p = -1
97     for i in range(n):
98         j = SA[i]
99         out[j] = T[(j - 1) % n]
100         if j == 0:
101             p = i
102     return bytes(out), p
103
104 def inverse_bwt(bwt: bytes, p: int) -> bytes:
105     n = len(bwt)
106     counts = [0]*256
107     ranks = [0]*n
108     for i, c in enumerate(bwt):
109         ranks[i] = counts[c]
110         counts[c] += 1
111     total = 0
112     first = [0]*256
113     for c in range(256):
114         first[c] = total
115         total += counts[c]
116     T = bytearray(n)
117     idx = p
118     for i in range(n-1, -1, -1):
119         c = bwt[idx]
120         T[i] = c
121         idx = first[c] + ranks[idx]
```

```
Ln 295, Col 1 Spaces: 4 UTF-8 CRLF {} Python Python 3.13 Go Live
```

```
File Edit Selection View Go Run Terminal Help 2.Advanced Algorithms
Hw8_HuffmanCompression.py
Homework8 > Hw8_HuffmanCompression.py ...
104 def inverse_bwt(bwt: bytes, p: int) -> bytes:
121     idx = first[c] + ranks[idx]
122     return bwt(T)
123
124 # ----- Huffman -----
125 import heapq
126
127 class _HNode:
128     slots = ("freq", "sym", "left", "right")
129     def __init__(self, freq, sym=None, left=None, right=None):
130         self.freq=freq; self.sym=sym; self.left=left; self.right=right
131     def __lt__(self, other):
132         a = self.sym if self.sym is not None else 1_000_000_000
133         b = other.sym if other.sym is not None else 1_000_000_000
134         if self.freq != other.freq: return self.freq < other.freq
135         return a < b
136
137 def _build_tree_from_freqs(freqs):
138     heap = [_HNode(f,s) for s,f in enumerate(freqs) if f>0]
139     if not heap: return None
140     if len(heap)==1: return heap[0]
141     heapq.heapify(heap)
142     while len(heap)>1:
143         a = heapq.heappop(heap); b = heapq.heappop(heap)
144         heapq.heappush(heap, _HNode(a.freq+b.freq, None, a, b))
145     return heap[0]
146
147 def _assign_codes(root):
148     if root is None: return {}
149     codes = {}
150     if root.sym is not None:
151         codes[root.sym] = "0"; return codes
152     def dfs(node, pref):
153         if node.sym is not None:
154             codes[node.sym] = pref if pref else "0"; return
155         dfs(node.left, pref+"0")
156         dfs(node.right, pref+"1")
157     dfs(root, "")
158     return codes
159
160 def _freqs_256_from_list(xs):
```

FileEditSelectionViewGoRunTerminalHelp

2.Advanced Algorithms

Hw8_HuffmanCompression.py

U

X

Homework8 > Hw8_HuffmanCompression.py > ...

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

def _freqs_256_from_list(xs):

f = [0]*256

for x in xs: f[x]+=1

return f

def _freqs_256_from_bytes(bs):

f = [0]*256

for b in bs: f[b]+=1

return f

----- File format helpers -----

import struct

MAGIC_SBM1 = b"SBM1"

MAGIC_HFM1 = b"HFM1"

LE = "<I"

def _u32(x): return struct.pack(LE, x)

def _u32_at(buf, off): return struct.unpack_from(LE, buf, off)[0]

----- Codec logic -----

def _pick_sentinel(data: bytes) -> int:

present = [False]*256

for b in data: present[b] = True

for v in range(256):

if not present[v]: return v

raise ValueError("Input uses all 256 byte values; cannot pick a unique sentinel.")

def compress_sbm1(raw: bytes) -> bytes:

sentinel = _pick_sentinel(raw)

T = raw + bytes([sentinel])

SA = suffix_array(T)

bwt, p = bwt_from_sa(T, SA)

mtf = mtf_encode_bytes(bwt)

freqs = _freqs_256_from_list(mtf)

root = _build_tree_from_freqs(freqs)

codes = _assign_codes(root)

bw = BitWriter()

for s in mtf: bw.write_bits_from_str(codes[s])

payload = bw.bytes()

parts = [MAGIC_SBM1, _u32(len(mtf)), _u32(p)]

for f in freqs: parts.append(_u32(f))

parts.append(payload)

Ln 295, Col 1

Spaces: 4

UTF-8

CRLF

{ } Python

Python 3.13

Go Live

```
File Edit Selection View Go Run Terminal Help 2.Advanced Algorithms
```

```
Hw8_HuffmanCompression.py X
```

```
Homework8 > Hw8_HuffmanCompression.py ...
```

```
186 def compress_sbml(raw: bytes) -> bytes:
187     for i in freqs: parts.append(_u32(i))
188     parts.append(payload)
189     return b"".join(parts)
190
191 def decompress_sbml(blob: bytes) -> bytes:
192     if blob[:4] != MAGIC_SBM1: raise ValueError("Bad magic for SBM1")
193     off = 4
194     n = _u32_at(blob, off); off += 4
195     p = _u32_at(blob, off); off += 4
196     freqs = [0]*256
197     for i in range(256):
198         freqs[i] = _u32_at(blob, off); off += 4
199     root = _build_tree_from_freqs(freqs)
200     if root is None: return b""
201     if root.sym is not None:
202         mtf = [root.sym]*n
203     else:
204         br = BitReader(blob[off:])
205         mtf = []
206         node = root
207         while len(mtf) < n:
208             bit = br.read_bit()
209             node = node.left if bit == 0 else node.right
210             if node.sym is not None:
211                 mtf.append(node.sym); node = root
212     bwt = mtf_decode_to_bytes(mtf)
213     T = inverse_bwt(bwt, p)
214     return T[:-1]
215
216 def compress_hfn1(raw: bytes) -> bytes:
217     freqs = _freqs_256_from_bytes(raw)
218     root = _build_tree_from_freqs(freqs)
219     codes = _assign_codes(root)
220     bw = BitWriter()
221     for b in raw: bw.write_bits_from_str(codes[b])
222     payload = bw.bytes()
223     parts = [MAGIC_HFN1, _u32(len(raw))]
224     for f in freqs: parts.append(_u32(f))
225     parts.append(payload)
226     return b"".join(parts)
227
228 def decompress_hfn1(blob: bytes) -> bytes:
229     if blob[:4] != MAGIC_HFN1: raise ValueError("Bad magic for HFN1")
230     off = 4
231     n = _u32_at(blob, off); off += 4
232     p = _u32_at(blob, off); off += 4
233     freqs = [0]*256
234     for i in range(256):
235         freqs[i] = _u32_at(blob, off); off += 4
236     root = _build_tree_from_freqs(freqs)
237     codes = _assign_codes(root)
238     bw = BitReader()
239     for b in raw: bw.write_bits_from_str(codes[b])
240     payload = bw.bytes()
241     parts = [MAGIC_HFN1, _u32(len(raw))]
242     for f in freqs: parts.append(_u32(f))
243     parts.append(payload)
244     return b"".join(parts)
```

```
Ln 295, Col 1 Spaces: 4 UTF-8 CRLF {} Python Python 3.13 Go Live
```



```
File Edit Selection View Go Run Terminal Help 2.Advanced Algorithms
Hw8_HuffmanCompression.py U X
Homework8 > Hw8_HuffmanCompression.py > ...
238     return b"".join(parts)
239
240 def decompress_hfn1(blob: bytes) -> bytes:
241     if blob[:4] != MAGIC_HFN1: raise ValueError("Bad magic for HFN1")
242     off = 4
243     n = _u32_at(blob, off); off += 4
244     freqs = [0]*256
245     for i in range(256):
246         freqs[i] = _u32_at(blob, off); off += 4
247     root = _build_tree_from_freqs(freqs)
248     if root is None: return b""
249     if root.sym is not None:
250         return bytes([root.sym])*n
251     br = BitReader(blob[off:])
252     out = bytearray(); node = root
253     while len(out) < n:
254         bit = br.read_bit()
255         node = node.left if bit == 0 else node.right
256         if node.sym is not None:
257             out.append(node.sym); node = root
258     return bytes(out)
259
260 # ----- CLI -----
261 def _run_cli(inp_path: str):
262     import os
263     with open(inp_path, "rb") as f: raw = f.read()
264     base, _ = os.path.splitext(inp_path)
265     sbm_bin = base + "_sbm1.bin"
266     sbm_rec = base + "_sbm1_recovered.txt"
267     hfn_bin = base + "_hfn1.bin"
268     hfn_rec = base + "_hfn1_recovered.txt"
269
270     sbm_blob = compress_sbm1(raw); open(sbm_bin, "wb").write(sbm_blob)
271     rec1 = decompress_sbm1(sbm_blob); open(sbm_rec, "wb").write(rec1)
272
273     hfn_blob = compress_hfn1(raw); open(hfn_bin, "wb").write(hfn_blob)
274     rec2 = decompress_hfn1(hfn_blob); open(hfn_rec, "wb").write(rec2)
275
276     def sz(p): return os.path.getsize(p)
277     print("Sizes (bytes) - SA+BWT+MTF+Huffman:")
278     print(f"  original : {len(raw)}")
279     print(f"  compressed: {sz(sbm_bin)}")
```



```
File Edit Selection View Go Run Terminal Help 2.Advanced Algorithms
Hw8_HuffmanCompression.py U X
Homework8 > Hw8_HuffmanCompression.py > ...
261 def _run_cli(inp_path: str):
262     print(f" original : {len(raw)}")
263     print(f" compressed: {sz(sbm_bin)}")
264     print(f" recovered : {sz(sbm_rec)}")
265     ratio = (sz(sbm_bin)/len(raw)) if len(raw) else 0.0
266     print(f" ratio : {ratio:.4f} (compressed/original)")
267     print(" OK: files match." if raw==rec1 else " ERROR: mismatch!")
268     print()
269     print("Sizes (bytes) - Huffman-only:")
270     print(f" original : {len(raw)}")
271     print(f" compressed: {sz(hfn_bin)}")
272     print(f" recovered : {sz(hfn_rec)}")
273     ratio2 = (sz(hfn_bin)/len(raw)) if len(raw) else 0.0
274     print(f" ratio : {ratio2:.4f} (compressed/original)")
275     print(" OK: files match." if raw==rec2 else " ERROR: mismatch!")
276
277 if __name__ == "__main__":
278     import sys, os
279
280     ABS_PATH = r"D:\1.SQM\1.UNIVERSIDAD\5. QUINTO SEMESTRE\2.Advanced Algorithms\Homework8\raven.txt"
281     REL_PATH = os.path.join("Homework8", "raven.txt")
282
283     candidate_paths = []
284     if len(sys.argv) == 2:
285         candidate_paths.append(sys.argv[1])
286         candidate_paths.extend([ABS_PATH, REL_PATH])
287
288     chosen = None
289     for p in candidate_paths:
290         if os.path.isfile(p):
291             chosen = p
292             break
293
294     if chosen is None:
295         print("Could not find 'raven.txt'. Tried:")
296         for p in candidate_paths:
297             print(" -", p)
298         print("\nFix the path(s) above or pass a file path as an argument.")
299         sys.exit(1)
300
301     _run_cli(chosen)
```

File Edit Selection View Go Run Terminal Help 2.Advanced Algorithms

Hw8_HuffmanCompression.py U raven.txt X raven_sbm1_recovered.txt U raven_hfn1_recovered.txt U

Homework8 > raven.txt

```
1 THE RAVEN by Edgar Allan Poe
2
3 Once upon a midnight dreary, while I pondered, weak and weary,
4 Over many a quaint and curious volume of forgotten lore,
5 While I nodded, nearly napping, suddenly there came a tapping,
6 As of some one gently rapping, rapping at my chamber door.
7 "'Tis some visitor," I muttered, "tapping at my chamber door-
8 | | | | Only this, and nothing more."
9
10 Ah, distinctly I remember it was in the bleak December,
11 And each separate dying ember wrought its ghost upon the floor.
12 Eagerly I wished the morrow;- vainly I had sought to borrow
13 From my books surcease of sorrow- sorrow for the lost Lenore-
14 For the rare and radiant maiden whom the angels name Lenore-
15 | | | | Nameless here for evermore.
16
17 And the silken, sad, uncertain rustling of each purple curtain
18 Thrilled me- filled me with fantastic terrors never felt before;
19 So that now, to still the beating of my heart, I stood repeating,
20 "'Tis some visitor entreating entrance at my chamber door-
21 Some late visitor entreating entrance at my chamber door;-
22 | | | | This it is, and nothing more."
23
24 Presently my soul grew stronger; hesitating then no longer,
25 "Sir," said I, "or Madam, truly your forgiveness I implore;
26 But the fact is I was napping, and so gently you came rapping,
27 And so faintly you came tapping, tapping at my chamber door,
28 That I scarce was sure I heard you"- here I opened wide the door;-
29 | | | | Darkness there, and nothing more.
30
31 Deep into that darkness peering, long I stood there wondering, fearing,
32 Doubting, dreaming dreams no mortal ever dared to dream before;
33 But the silence was unbroken, and the stillness gave no token,
34 And the only word there spoken was the whispered word, "Lenore?"
35 This I whispered, and an echo murmured back the word, "Lenore!"-
36 | | | | Merely this, and nothing more.
37
38 Back into the chamber turning, all my soul within me burning,
39 Soon again I heard a tapping somewhat louder than before.
40 "Surely," said I, "surely that is something at my window lattice:
41 Let me see, then, what thereat is, and this mystery explore-
42 Let my heart be still a moment and this mystery explore;-
```

Ln 22, Col 48 Spaces: 4 UTF-8 LF {} Plain Text Go Live

File Edit Selection View Go Run Terminal Help 2.Advanced Algorithms

Hw8_HuffmanCompression.py U raven.txt U raven_sbm1_recovered.txt U raven_hfn1_recovered.txt U

Homework8 > raven_sbm1_recovered.txt

1 THE RAVEN by Edgar Allan Poe
2
3 Once upon a midnight dreary, while I pondered, weak and weary,
4 Over many a quaint and curious volume of forgotten lore,
5 While I nodded, nearly napping, suddenly there came a tapping,
6 As of some one gently rapping, rapping at my chamber door.
7 "'Tis some visitor," I muttered, "tapping at my chamber door-
8 | | | | Only this, and nothing more."
9
10 Ah, distinctly I remember it was in the bleak December,
11 And each separate dying ember wrought its ghost upon the floor.
12 Eagerly I wished the morrow;- vainly I had sought to borrow
13 From my books surcease of sorrow- sorrow for the lost Lenore-
14 For the rare and radiant maiden whom the angels name Lenore-
15 | | | | Nameless here for evermore.
16
17 And the silken, sad, uncertain rustling of each purple curtain
18 Thrilled me- filled me with fantastic terrors never felt before;
19 So that now, to still the beating of my heart, I stood repeating,
20 "'Tis some visitor entreating entrance at my chamber door-
21 Some late visitor entreating entrance at my chamber door;-
22 | | | | This it is, and nothing more."
23
24 Presently my soul grew stronger; hesitating then no longer,
25 "Sir," said I, "or Madam, truly your forgiveness I implore;
26 But the fact is I was napping, and so gently you came rapping,
27 And so faintly you came tapping, tapping at my chamber door,
28 That I scarce was sure I heard you"- here I opened wide the door;-
29 | | | | Darkness there, and nothing more.
30
31 Deep into that darkness peering, long I stood there wondering, fearing,
32 Doubting, dreaming dreams no mortal ever dared to dream before;
33 But the silence was unbroken, and the stillness gave no token,
34 And the only word there spoken was the whispered word, "Lenore?"
35 This I whispered, and an echo murmured back the word, "Lenore!"-
36 | | | | Merely this, and nothing more.
37
38 Back into the chamber turning, all my soul within me burning,
39 Soon again I heard a tapping somewhat louder than before.
40 "Surely," said I, "surely that is something at my window lattice:
41 Let me see, then, what thereat is, and this mystery explore-
42 Let my heart be still a moment and this mystery explore;-

Ln 1, Col 1 Spaces: 4 UTF-8 LF {} Plain Text Go Live

FileEditSelectionViewGoRunTerminalHelp←→2.Advanced Algorithms

Hw8_HuffmanCompression.py U raven.txt U raven_sbm1_recovered.txt U raven_hfn1_recovered.txt X

Homework8 > raven_hfn1_recovered.txt

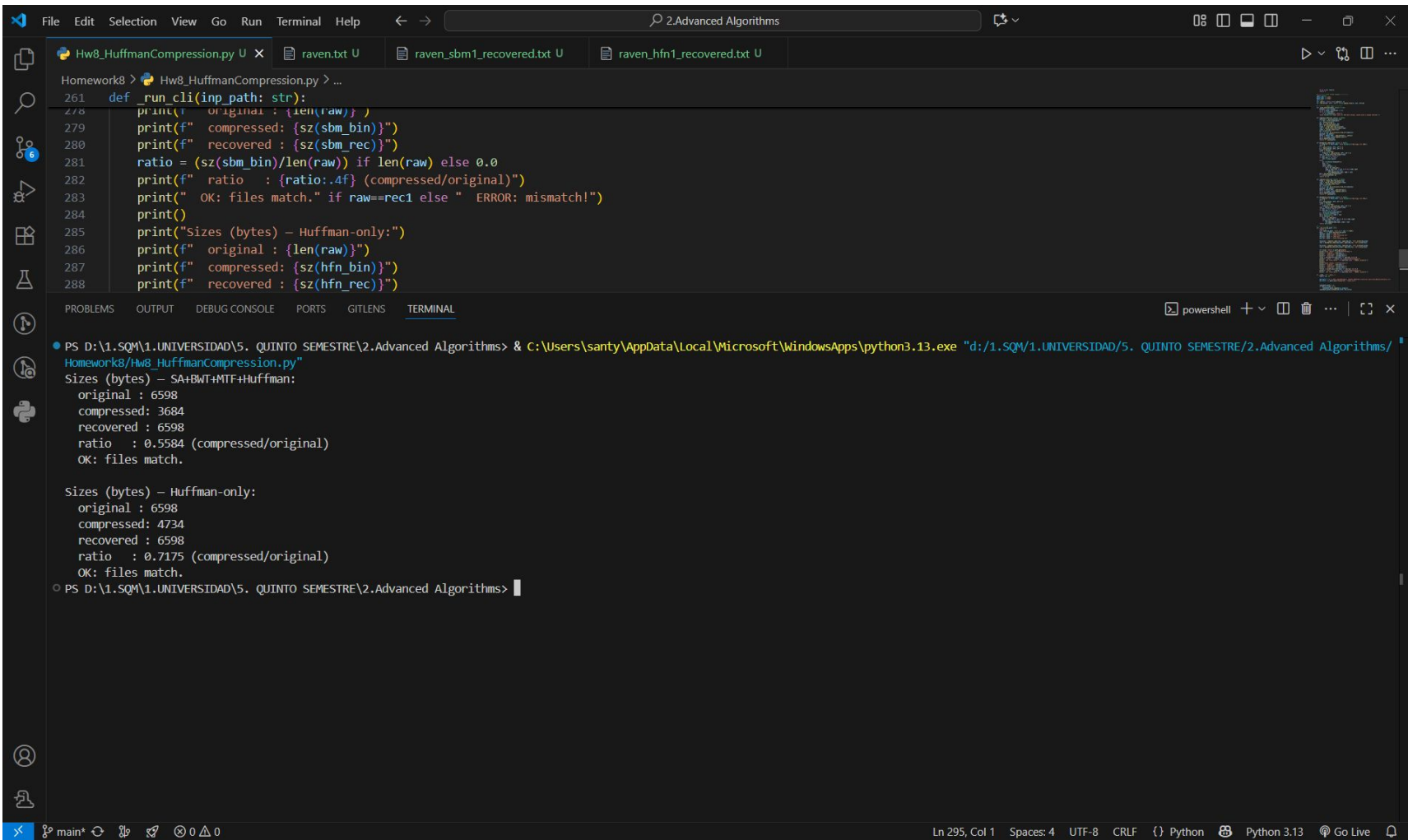
```
1 THE RAVEN by Edgar Allan Poe
2
3 Once upon a midnight dreary, while I pondered, weak and weary,
4 Over many a quaint and curious volume of forgotten lore,
5 While I nodded, nearly napping, suddenly there came a tapping,
6 As of some one gently rapping, rapping at my chamber door.
7 "'Tis some visitor," I muttered, "tapping at my chamber door-
8 | | | | Only this, and nothing more."
9
10 Ah, distinctly I remember it was in the bleak December,
11 And each separate dying ember wrought its ghost upon the floor.
12 Eagerly I wished the morrow;- vainly I had sought to borrow
13 From my books surcease of sorrow- sorrow for the lost Lenore-
14 For the rare and radiant maiden whom the angels name Lenore-
15 | | | | Nameless here for evermore.
16
17 And the silken, sad, uncertain rustling of each purple curtain
18 Thrilled me- filled me with fantastic terrors never felt before;
19 So that now, to still the beating of my heart, I stood repeating,
20 "'Tis some visitor entreating entrance at my chamber door-
21 Some late visitor entreating entrance at my chamber door;-
22 | | | | This it is, and nothing more."
23
24 Presently my soul grew stronger; hesitating then no longer,
25 "Sir," said I, "or Madam, truly your forgiveness I implore;
26 But the fact is I was napping, and so gently you came rapping,
27 And so faintly you came tapping, tapping at my chamber door,
28 That I scarce was sure I heard you"- here I opened wide the door;-
29 | | | | Darkness there, and nothing more.
30
31 Deep into that darkness peering, long I stood there wondering, fearing,
32 Doubting, dreaming dreams no mortal ever dared to dream before;
33 But the silence was unbroken, and the stillness gave no token,
34 And the only word there spoken was the whispered word, "Lenore?"
35 This I whispered, and an echo murmured back the word, "Lenore!"-
36 | | | | Merely this, and nothing more.
37
38 Back into the chamber turning, all my soul within me burning,
39 Soon again I heard a tapping somewhat louder than before.
40 "Surely," said I, "surely that is something at my window lattice:
41 Let me see, then, what thereat is, and this mystery explore-
42 Let my heart be still a moment and this mystery explore;-
```

Ln 1, Col 1 Spaces: 4 UTF-8 LF Plain text Go Live

Herramienta Recortes

Captura de pantalla copiada en el portapapeles
Guardado automáticamente en la carpeta de capturas de pantalla.

Marcado y uso compartido



<https://colab.research.google.com/drive/1jYTL1eGGLmnzfKvl-6MLFra4Y0Q3gwY8?usp=sharing>

REFERENCES

GeeksforGeeks. (2025, July 23). *Huffman Coding | Greedy algo3*. GeeksforGeeks. <https://www.geeksforgeeks.org/dsa/huffman-coding-greedy-algo-3/>

GeeksforGeeks. (2025, July 23). *Huffman Coding in Python*. GeeksforGeeks. <https://www.geeksforgeeks.org/dsa/huffman-coding-in-python/>

W3Schools.com. (n.d.). https://www.w3schools.com/dsa/dsa_ref_huffman_coding.php

GeeksforGeeks. (2025, July 23). *Suffix Array | Set 1 (Introduction)*. GeeksforGeeks. <https://www.geeksforgeeks.org/dsa/suffix-array-set-1-introduction/>

GeeksforGeeks. (2025, August 14). *Burrows Wheeler Data Transform Algorithm*. GeeksforGeeks.

<https://www.geeksforgeeks.org/dsa/burrows-wheeler-data-transform-algorithm/>

GeeksforGeeks. (2023, March 29). *Move to front data transform algorithm*. GeeksforGeeks.

<https://www.geeksforgeeks.org/dsa/move-front-data-transform-algorithm/>