American Computer Science League

2023-2024 ● Finals #1: Run-Length Encoding ● Junior Division

PROBLEM: *Run-length encoding* is a data compression technique that works well on data in which values repeat frequently, such as black-and-white images. The idea is to replace runs of the same character by that character followed by a count of how many times that character appears. This compression technique is *lossless*, meaning that the original source can be reconstructed from the compressed version.

For example, consider the input string ABBCEEEEAAAAAADDDAAA. Every count will be a hexadecimal number, one less than the number of occurrences of that character. The encoded version would be A0B1C0E3A5D2A2. The original string was 20 characters; the encoded string is only 14 characters. If there are more than 16 occurrences of a character, output the character multiple times. For example, a string of 5 Cs followed by 30 Bs followed by 2 As would be compressed to C4BFBDA1.

In the **encoding** part of this program, you will be given an alphanumeric string and your program will return the encoded version. In the **decoding** part of this program, you will be given a string which is a run-length encoding of an alphanumeric string. You need to decode it and return it as a single string. The data will start with an **E** for encoding and a **D** for decoding, followed by a string as described above.

Your program must also support the following variation, indicated by an **EV** or a **DV**: rather than breaking the repeated character into chunks of 16, simply output the number of times, in hexadecimal, that the repeated character appears. When the count is 2 or more hexadecimal digits, it is preceded and followed by a dash. In this variation, the count is the number of occurrences, not one less. For example, ABBCEEEEAAAAAADDDAAA is encoded as A1B2C1E4A6D3A3 and a string of 5 Cs followed by 30 Bs followed by 2 As would be compressed to C5B-1E-A2.

INPUT: There will be two strings of data. The first one or two character string will represent the code (E, D, EV, DV). The second string will be the message to be encoded or decoded. All messages will have a valid format. All hex digits in encoded strings will be capitalized. Any character on the keyboard can be used in the message except for a "-", which is only used as a delimiter.

OUTPUT: Output a string that represents either the encoded message or the decoded message depending on the code that is input.

American Computer Science League

2023-2024 ● Finals #1: Run-Length Encoding ● Junior Division

SAMPLE INPUT:

The 2nd string in each data set is a single string; for readability, it appears over multiple lines.

1. E
   ABBCEEEEAAAAAADDDAAACCCCCBBBBBBBBBBBBBBBBBBBBBBBBBBAA

2. EV
   ABBCEEEEAAAAAADDDAAACCCCCBBBBBBBBBBBBBBBBBBBBBBBBBBAA

3. D
   a0b1a2dAfFfFfEE2

4. DV
   a1b2a3dBf-2F-E3

SAMPLE OUTPUT:

The output is a single string; for readability, it appears over multiple lines.

1. A0B1C0E3A5D2A2C4BFBDA1

2. A1B2C1E4A6D3A3C5B-1E-A2

3. abbaaaddddddddddddfffffffffffffffffffffffffffffffffffffffffffffffffffffEEE

4. abbaaaddddddddddddfffffffffffffffffffffffffffffffffffffffffffffffffffffEEE

American Computer Science League

2023-2024 ● Finals #1: Run-Length Encoding ● Junior Division

TEST DATA

TEST INPUT:

1. E
   abbbbbbbbbccccccccccccccddeeeeeeefgghhhhhhh

2. D
   a5b1cAdFe7fCgBh2

3. E
   zzzzzzzzzzzzzyyyyyyyxwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
   wwwwwwwwwwvvvuuuuuuuuuuuuuuuuuuuuu

4. D
   zFzFz3y0xAwFv7uFuFuEt2

5. EV
   111111111111111111111111111111111111111111111111000000000000
   000001000000011111111111111111111111111101111111111111111111
   11111111000000000000

6. DV
   AABBCCDDEEFF0-10-11223344556677889 9

7. DV
   1-3F-061A0-40-1D011-11-0F1-3A-

8. EV
   999999999AAAAAAAAAABBBBBBBBBBBCCCCCCCCCCCDDDDDDDDDDDDDEEEEEE
   EEEEEEEEFFFFFFFFFFFFFFF8888888877777776666665555544443332210

9. E
   111111111111111111222222222222222223333333333333333333344444
   444444444444444444555555555555555555555555566666666666666666
   666666666666666666666666666666666666667777777777777777777777
   777777777888888888888888888888888888889999999999999999999999

10. D
   1F1F1F1F1A2C384F4F40526F7F7F788F899F

American Computer Science League

2023-2024 ● Finals #1: Run-Length Encoding ● Junior Division

TEST OUTPUT:

1. a0b8cDd1e6f0g1h6

2. aaaaaabbccccccccccddddddddddddddddeeeeeeeeffffffffffffggg
   gggggggggghhh

3. zCy6x0wFwFwFw1v2uFu5

4. zzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzyxxxxxxxxxxxwwwwwwwwww
   wwwwwvvvvvvvvvuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu
   uttt

5. 1-31-0-11-11071-1C-011-1B-0C

6. AAAAAAAAAABBBBBBBBBBCCCCCCCCCCCCDDDDDDDDDDDDDEEEEEEEEEEEEE
   EFFFFFFFFFFFFFFF0000000000000000122333444455555666666777777
   788888888999999999

7. 11111111111111111111111111111111111111111111111111111111
   11110000001111111111000000000000000000000000000000000000
   00000000000000000000000011111111111101111111111111111000
   00000000000111111111111111111111111111111111111111111111
   11111111111

8. 99AABBCCDDEEFF887766554433221101

9. 1F112F203F334F475F596F6F6F667F7F8F8C9F96

10. 11111111111111111111111111111111111111111111111111111111
    11111111111111122222222222233333333344444444444444444444
    44444444444555666666666666666677777777777777777777777777
    77777777777788888888888888888888888899999999999999999