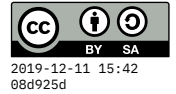# Assignment 1: text compression

30 September 2019

In this activity, we will investigate the **Huffman algorithm** for text compression. We have already seen one example of a Huffman encoding, represented by the strange-looking variable-height tree in the text encoding notes. You will follow the Huffman algorithm and create a tree of your own, based on the character frequencies of a message that I assign specifically to you (shown below).

Demonstration of Huffman algorithm

This video[1] illustrates the algorithm on paper. I apologize that the resolution and audio quality aren't great, but it should be understandable. The final encoding and tree are also pictured below.

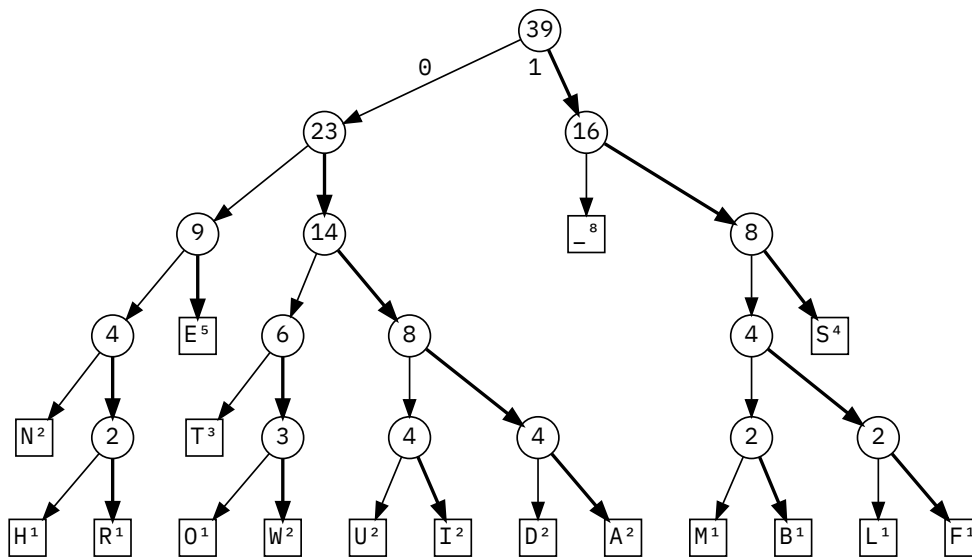[1]vimeo.com/11
8559113

Figure 1: Huffman tree produced in the video

```
     5|   10|   15|   20|   25|   30|   35|   40|   45|   50|   55|

010110011011000011001110100101100101100011011101001110   ← 54 bits
W    E  _ M    U    S    _ B    U    I    L    D
5    3  2 5    5    3    4 2 5   5    5    5    5

1001111111100110111011100100000100011011101111000001110   ← 55 bits
_ A    S  _ I    F    _ T    H    E  _ S A    N    D
2 5    3  2 5    5    2 4    5    3  2 3 5    4    5
```

```
100101100100011001101110100010100000001              ← 39 bits
_ W    E R    E _ S T    O    N    E
2 5    3 5    3 2 3 4    5    4    3
```

The total size of the encoding is 54+55+39 = 148 bits.

Questions to answer

1. How many **distinct** characters does your phrase contain?

2. If we were using a fixed-width encoding, how many bits **per character** would you need to represent just those characters?

3. What is the **most frequent character** in your phrase, and how many times did it appear?

4. **How many bits** are used to represent the most frequent character in your phrase?

5. What is the **most number of bits** used to encode any character in your phrase?

6. Use the tree you produced to **encode the entire phrase** you were given. **How many bits are used,** in total?

Your assigned phrase

So that everyone's solutions are a bit different, you should use the phrase provided beside your student ID and initials. If you are not in this list, contact me to obtain your phrase.

Initials	ID (last 4)	Phrase
AH	6889	NOR_THE_GARDENS_OF_BABYLON_ETERNAL_BEAUTY
AL	9482	SCIENCE_IS_A_REFINEMENT_OF_EVERYDAY_THINKING
AM	6470	SIMPLICITY_IS_PREREQUISITE_FOR_RELIABILITY
AM	9063	SONGS_OF_SUCH_EXQUISITE_SWEETNESS_THAT
AN	2026	STORM_CLOUDS_GATHERING_WIND_IS_GONNA_BLOW
AR	8693	THE_WHOLE_OF_LIFE_IS_A_PROCESS_OF_LEARNING
AW	1120	YET_WHO_PETITION_FOR_DARK_TOKENS_OF_PEACE
BB	9262	YOU_HAVE_WITHDRAWN_YOURSELF_AND_THE_MAGIC
BJ	1758	NOW_LISTEN_CLOSELY_ILL_TELL_YOU_WHAT_I_KNOW
CB	0786	EXPENSIVE_DOCTORS_TO_CURE_THEIR_STONE_HEARTS
CD	7177	IT_IS_THE_SIMPLE_THAT_PRODUCES_THE_MARVELOUS
CF	7260	THE_HARDER_I_WORK_THE_MORE_LUCK_I_HAVE
CO	2217	TESTING_SHOWS_PRESENCE_NOT_ABSENCE_OF_BUGS
EG	8440	MANKIND_ARE_MORE_DISPOSED_TO_SUFFER_WHILE
EV	1014	WE_PEOPLE_ON_THIS_SMALL_AND_DRIFTING_PLANET
GA	5341	A_CASE_AGAINST_THE_GOTO_STATEMENT_DIJKSTRA
GR	3209	THOSE_SAME_HANDS_CAN_TOUCH_WITH_SUCH_HEALING
HT	4995	WE_HOLD_THESE_TRUTHS_TO_BE_SELF_EVIDENT
IP	7121	THERE_ARE_MILLIONAIRES_WITH_MONEY_CANT_USE
JC	3866	GOVERNMENTS_LONG_ESTABLISHED_SHOULD_NOT_BE
JV	2547	WE_RELEASE_FINGERS_FROM_FISTS_OF_HOSTILITY
KB	2597	COMPUTER_SCIENCE_IS_NO_MORE_ABOUT_COMPUTERS
KG	0351	LONG_FOR_THE_ENDLESS_IMMENSITY_OF_THE_SEA
KH	8171	NOBODY_BUT_NOBODY_CAN_MAKE_IT_OUT_HERE_ALONE
KR	8920	THE_ART_OF_PROGRAMMING_IS_AVOIDING_CHAOS
KV	3181	WHERE_WATER_IS_NOT_THIRSTY_AND_BREAD_IS_NOT
LC	5362	IMAGINATION_IS_MORE_IMPORTANT_THAN_KNOWLEDGE
LF	7631	LIFE_LIBERTY_AND_THE_PURSUIT_OF_HAPPINESS
LJ	2323	OUT_OF_SUCH_CHAOS_OF_SUCH_CONTRADICTION
LJ	9222	NURTURE_ALL_CREATURES_IN_DEPTHS_AND_ON_SHORES
LK	6885	PROVIDE_NEW_GUARDS_FOR_THEIR_FUTURE_SECURITY
MB	1002	CAREFUL_GIVING_ADVICE_IT_IS_SOMETIMES_FOLLOWED
ML	4201	SIMPLICITY_IS_A_VIRTUE_BUT_REQUIRES_HARD_WORK
MZ	8853	YOUR_THEORY_IS_NOT_CRAZY_ENOUGH_TO_BE_TRUE
NF	2078	LIKELY_TO_EFFECT_THEIR_SAFETY_AND_HAPPINESS
NP	9205	THESE_ARE_NOT_THE_ONLY_WONDERS_OF_THE_WORLD
OS	1559	TRAVELING_THROUGH_CASUAL_SPACE_ALOOF_STARS
SG	2869	LYING_THINKING_LAST_NIGHT_HOW_TO_FIND_MY_SOUL
TR	7435	THE_HARDER_I_WORK_THE_MORE_LUCK_I_HAVE
VB	0068	DERIVING_THEIR_JUST_POWERS_FROM_THE_CONSENT
XW	1354	WORDS_WHICH_CHALLENGE_OUR_VERY_EXISTENCE

How to submit

You will submit two things electronically: a photo of the tree you drew, and a text containing the answers to the 6 questions. Here is how:

1. Register for an account at `gitlab.liu.edu`[2]. (You must use an `liu.edu` email address.)

[2]`gitlab.liu.edu/users/sign_in#register-pane`

2. Once you are logged in, click the green **New Project** button. On the subsequent screen,

   - Type **cs101** (lower-case, no spaces) as the **Project name.**
   - Click the **Initialize repository with a README** option.
   - Then hit **Create project.**

3. On the **cs101** project sidebar, select **Settings » Members**.

4. Under **Select members to invite,** type `league`. My account should pop up. Choose it. Set **Choose a role permission** to **Developer** and hit **Add to project.**

5. On the **cs101** project sidebar, select **Wiki.** Push the green button to **Create your first page.**

6. Change the **Title** to exactly A1 (capital A, the number 1, no spaces). Write or paste your answers to the six questions into the big box.

7. While writing, you may use the "Add numbered list" button on the format toolbar. It will number all answers with 1., but when you preview they'll be numbered 1,2,3.

8. With the cursor on a blank line at the end of the text, use **Attach a file** to upload the photo of your tree. You can use the **Preview** tab to make sure it shows up correctly. (Sometimes a new upload can take a few moments to appear.)

9. When you're satisfied (or whenever you want to save), hit the **Create page** button. You will then be able to **Edit** the page if you like, but I'll use the *last* time the page is updated as the submission time for your assignment.

10. The next time you return to the **Wiki** feature, it will display **Home · Create Page** again. But you will see a link to the page titled **A1** in the *right* sidebar.

Congratulations, you're done!