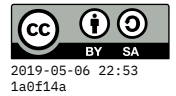


# Assignment 1: text compression



18 February 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<sup>1</sup> 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.



<sup>1</sup>vimeo.com/118559113

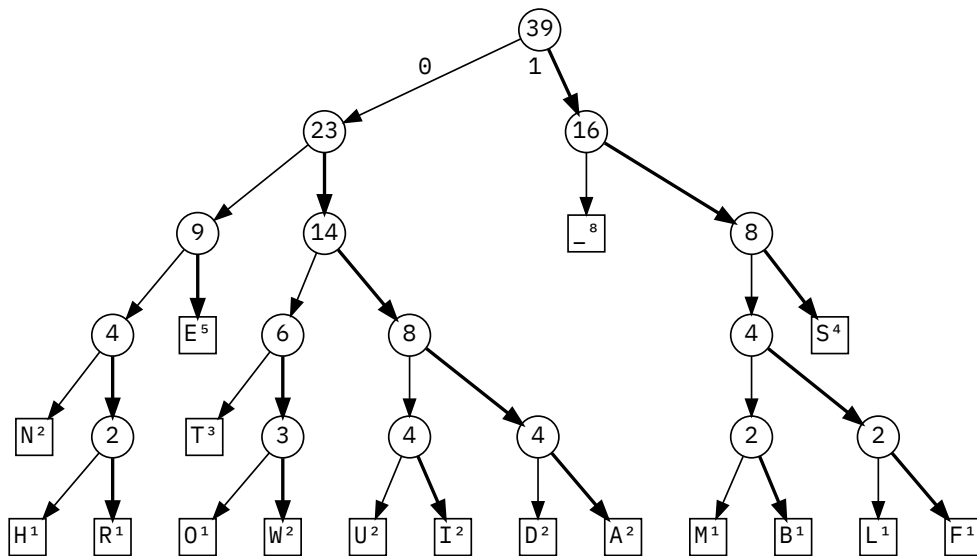


Figure 1: Huffman tree produced in the video

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

010110011011000011001110100101100101100011011101001110 ← 54 bits

W E \_ M U S T \_ B U I L D  
5 3 2 5 5 3 4 2 5 5 5 5 5

100111111100110111011100100000100011011101111000001110 ← 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	Phrase
AG	100535952	IMAGINATION_IS_MORE_IMPORTANT_THAN_KNOWLEDGE
AM	100551455	NOW_LISTEN_CLOSELY_ILL_TELL_YOU_WHAT_I_KNOW
BM	100588155	NOR_THE_GARDENS_OF_BABYLON_ETERNAL_BEAUTY
CF	100585354	EXPENSIVE_DOCTORS_TO_CURE_THEIR_STONE_HEARTS
CT	100620003	SCIENCE_IS_A_REFINEMENT_OF_EVERYDAY_THINKING
DD	100580521	DERIVING_THEIR_JUST_POWERES_FROM_THE_CONSENT
EC	100504153	CURTAIN_FALLS_ON_THE_MINSTREL_SHOW_OF_HATE
JM	100538993	MANKIND_ARE_MORE_DISPOSED_TO_SUFFER_WHILE
JM	100567397	NOBODY_BUT_NOBODY_CAN_MAKE_IT_OUT_HERE_ALONE
KB	100610792	COMPUTER_SCIENCE_IS_NO_MORE_ABOUT_COMPUTERS
KK	100589498	LIFE_LIBERTY_AND_THE_PURSUIT_OF_HAPPINESS
KS	100367370	PROVIDE_NEW_GUARDS_FOR_THEIR_FUTURE_SECURITY
ML	100515931	LIKELY_TO_EFFECT_THEIR_SAFETY_AND_HAPPINESS
SA	100629088	A_CASE_AGAINST_THE_GOTO_STATEMENT_DIJKSTRA
SB	100632069	CAREFUL_GIVING_ADVICE_IT_IS_SOMETIMES_FOLLOWED
SF	100625271	GOVERNMENTS_LONG_ESTABLISHED_SHOULD_NOT_BE
SK	100579463	IT_IS_THE_SIMPLE_THAT_PRODUCES_THE_MARVELOUS
SL	100551483	LYING_THINKING_LAST_NIGHT_HOW_TO_FIND_MY_SOUL
SM	100619272	NURTURE_ALL_CREATURES_IN_DEPTHS_AND_ON_SHORES
SQ	100580414	OUT_OF_SUCH_CHAOS_OF_SUCH_CONTRADICTION
ST	100586641	SIMPLICITY_IS_A_VIRTUE_BUT_REQUIRES_HARD_WORK
SW	100594503	SONGS_OF_SUCH_EXQUISITE_SWEETNESS_THAT
XL	100555030	LONG_FOR_THE_ENDLESS_IMMENSITY_OF_THE_SEA

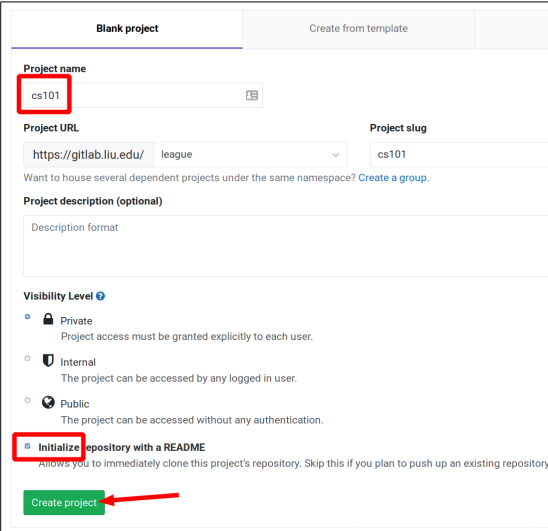
## How to submit

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](https://gitlab.liu.edu)<sup>2</sup>. (You must use an `liu.edu` email address.)
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**.



<sup>2</sup>[gitlab.liu.edu/users/sign\\_in#register-pane](https://gitlab.liu.edu/users/sign_in#register-pane)



**Blank project** Create from template

**Project name**  
cs101

**Project URL** Project slug  
https://gitlab.liu.edu/ league cs101

Want to house several dependent projects under the same namespace? [Create a group.](#)

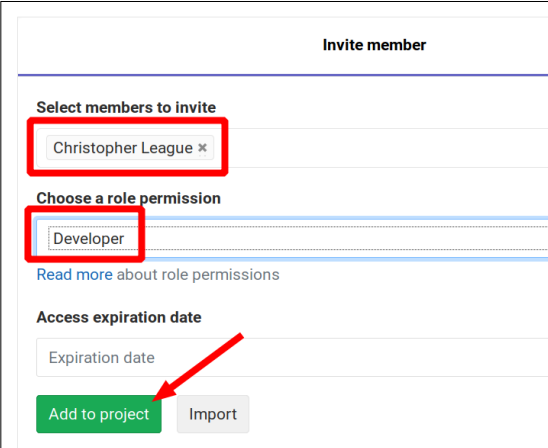
**Project description (optional)**  
Description format

**Visibility Level**

- Private  
Project access must be granted explicitly to each user.
- Internal  
The project can be accessed by any logged in user.
- Public  
The project can be accessed without any authentication.
- Initialize repository with a README  
Allows you to immediately clone this project's repository. Skip this if you plan to push up an existing repository.

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**.



**Invite member**

**Select members to invite**  
Christopher League ✕

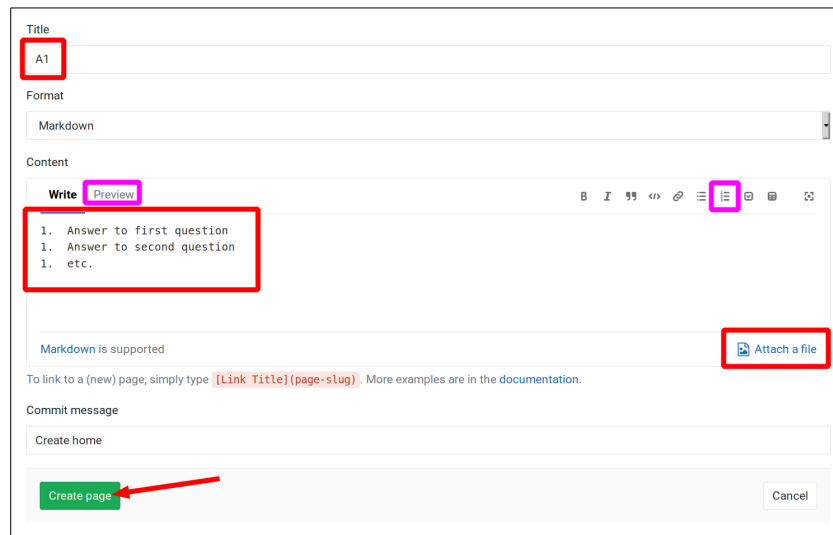
**Choose a role permission**  
Developer

[Read more](#) about role permissions

**Access expiration date**  
Expiration date

Add to project Import

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.



The screenshot shows a form for creating a new page. The 'Title' field contains 'A1'. The 'Format' dropdown is set to 'Markdown'. The 'Content' area has a 'Write' tab selected and contains a numbered list: '1. Answer to first question', '1. Answer to second question', and '1. etc.'. The 'Preview' tab is also visible. A red box highlights the 'Add numbered list' button in the toolbar. Below the content area, there is an 'Attach a file' button. The 'Commit message' field contains 'Create home'. At the bottom, there is a green 'Create page' button and a grey 'Cancel' button. A red arrow points to the 'Create page' button.

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!