Lab 3 - Compression

Luciano Giannini, Ryan Pepin, Mike Emmett

LC 207: Mathematical Experiments in Computer Science

(Dated 11/13/2019)

Contents

Introduction

During this lab, we have discussed and explored the idea of Data Compression and how it is used as the most optimal way to send and receive files. A compressed file uses this concept to reduce the size of the file without compromising the contents and allowing fewer transmission errors to occur. This idea seems to relate to the encoding and decoding messages from the previous lab, but instead of hiding a message, data compression is used to transmit the message in the most efficient way. One method of data compression that we explored was designed and introduced by a man named David Huffman.

Huffman's code uses the idea of finding the frequency of letters in a given phrase and using this to determine the variable number of bits for each character's code. His idea is very similar to that of Morse code in the sense that characters that appear most frequently can be encoded with one or two bits while less frequent characters can be encoded using longer bites. In Huffman's code, he also used the concept of prefix codes to help decide the code for each character in relation to their frequency. A prefix code is a code that is significant to itself and does not appear in the beginning of another code. Huffman codes can be represented by using binary trees to display the codeword for a specific character that is stored in the leaves of the tree.

A Huffman tree is created by combining root nodes that hold the frequency value of the character stored in the given node. The two roots with the smallest values are combined to form a new tree, and this is carried out until every root is combined into this new binary tree. Using this idea one can gather a Huffman code for a character by working backward tracing up the tree from a leaf to the root while allowing 0 to represent if the leaf was a left subtree and 1 to represent if the leaf was a right subtree.

Exercise 1: Encoding and Decoding bit String

Code 1:

Character	Codeword
A	0000
H	0001
-	001
E	10
L	11
S	01

Code 2:

Character	Codeword
S	0
L	1
E	00
-	11
H	01
A	10

a. Decode each of the following bit strings which are encoded with Code 1:

010000110010110111110100110101101

SAL-SELLS-EELS

00010000110010001000001001000110101101

HAL-HAS-HEELS

b. Now, using the following bit string, show that Code 2 is ambiguous. In other words, find two possible translations of the following bit string:

01011100011011001000

SLSLLLSSSLLSLLSSLSSS

SAL-SELLS-SHES

c. Explain why Code 1 is not ambiguous. What property does Code 1 have which ensures that each bit string has a unique translation?

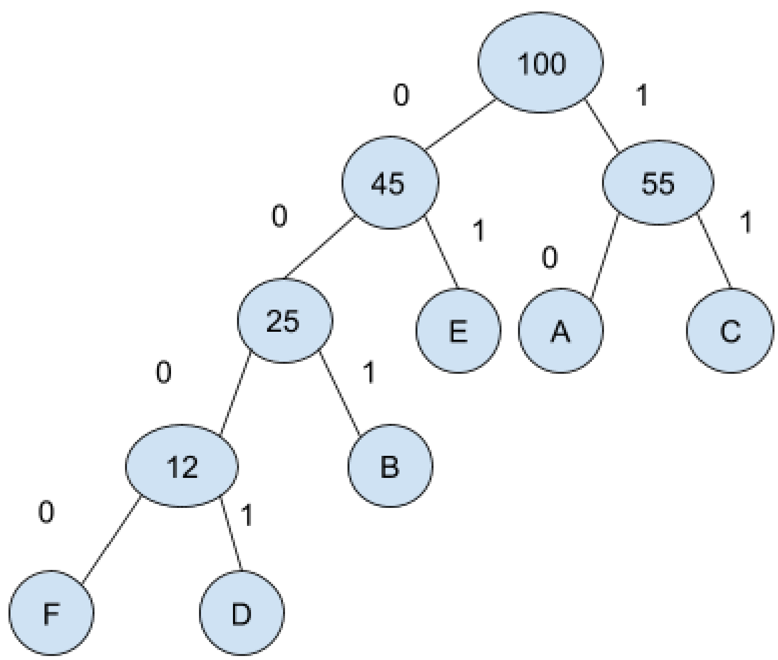
The reason why Code 1 is not ambiguous is that while reading through the bit string there will never be a time where reading in only one character of the string will result in a character. Comparing to Code 2 which we see that if you read in every single character of the bit string you will get the corresponding letter for it. You will never know when you need to read in the next value to determine the correct letter. For example, we read a bit string 0101. Using Code 1 we will have “SS”. For Code 2 we can have “SLSL” or “SAL” or “SLH” there are too many different outcomes for Code 2 which is the main reason why we cannot use it. Using Code 1 will always have the same outcome and only one outcome for each bit string.

Exercise 2: Construct a Huffman Tree

Objective: Create a huffman tree for the given letters and frequencies, then create codewords and convert "AECBCAF" using the found codewords.

Char	Relative Freq.
A	.22
B	.13
C	.33
D	.10
E	.20
F	.02

Char	Freq.	Codeword
C	33	11
A	22	10
E	20	01
B	13	001
D	10	0001
F	2	0000



Encode: "AECBCAF"

Output: 10011100111100000

Program 1: Encoding a Text File

Objective: Build a huffman tree using the frequencies of each character's ascii value to compress a text into binary code words. Export both the encoded text and the frequency table to output files.

Frequency Table

```
public static int[] frequency(String text)
{
  int [] freq = new int [128];
  for (int j = 0;j<text.length();j++)
  {
    char x = text.charAt(j);
    for (int i = 0;i<128;i++)
    {
      if (x == i)
        freq[i]++;
    }
  }
  return freq;
}
```

These for loops establish our frequencies. It simply runs through the input text, character by character, and for each time that character's ascii value appears, we add one to that ascii value in the freq[] array. The array is then returned with each place holding the frequency for that ascii value.

Building the Tree

```
public static Node priority(int [] freq)
{
  for (int i = 0;i<128;i++)
  {
    if (freq[i]!=0)
    {
      Node g = new Node(i,freq[i]);
      insert(g);
    }
  }
  Node r = null;
  while (size!=1)
  {
    Node last = remove();
    Node last2 = remove();
    Node parent = new Node(0,last.frequency + last2.frequency);
    parent.left = last;
    parent.right = last2;
    insert(parent);
  }
  r = remove();
  return r;
}
```

Insert

```
public static void insert(Node p)
{
  size++;
  Node r = root;
  if (r==null)
  {
    root = p;
    return;
  }
  Node q = null;
  while (r!=null&&r.frequency<=p.frequency)
  {
    q = r;
    r = r.next;
  }
  //System.out.println(p.frequency);
  if (q == null)
  {
    root = p;
    p.next = r;
  }
  else if (r != null)
  {
    q.next = p;
    p.next = r;
  }
  else
    q.next = p;
}
```

Remove

```
public static Node remove()
{
  Node temp = root;
  root = root.next;
  size--;
  return temp;
}
```

The tree is made by first creating a node for each character found in the text file. To accomplish this, we needed insert(Node) and remove() methods to create a priority queue. The insert method inserts the new node in its corresponding place in the queue based on its frequency. The remove method just takes out the first node. Once we have created a queue, we use it to build a Huffman Tree. While there is more than one node in the queue, we take out the nodes with the two lowest frequencies and make them leaves. We then create a new node (parent) as the root for these two leaves and add that back onto the queue. This is done repeatedly until there is only one node left (r) that holds all the information for the tree.

Creating Codewords

```
public static void code(Node r, String code, String [] codeword)
{
  if (r!=null)
  {
    if (r.data!=0)
  {
    codeword[r.data] = code;
  }

      code(r.left, code+'0',codeword);

      code(r.right, code+'1',codeword);

  }
}
```

Encoding the Text File

```
public static String encode(String [] codeword, String text)
{
  String encoded = "";
  for (int i = 0;i<text.length();i++)
  {
    char x = text.charAt(i);
    for (int j = 0;j<128;j++)
    {
      if (x == j)
        encoded += codeword[j];
    }
  }
  System.out.println(encoded);
  return encoded;
}
```

The codewords are created using a recursive method. We traverse the tree both left and right, adding to the codeword each level we move down, '0' if we moved to the left and '1' if we moved to the right. Once we have reached a leaf, the codeword is assigned as whatever combination of 1's and 0's that it took to get to that leaf. We then encode the text file by moving through the file one character at a time and encode that character based on the codeword assigned to it by the huffman tree.

Main Method

```
public static void main(String[]args) throws IOException
{
  Scanner user = new Scanner(System.in);
  System.out.println("Input file name:");
  String inputfile = user.nextLine();
  System.out.println("Output file name:");
  String outputfile = user.nextLine();
  System.out.println("Frequency file name:");
  String freqfile = user.nextLine();
  File inFile = new File(inputfile);
  Scanner input = new Scanner(inFile);
  String text = "";
  while (input.hasNext())
  {
    text += input.nextLine();
  }
  int [] freq = frequency(text);//need these two, to deco
  Node tree = priority(freq);//here
  String [] codeword = new String[128];
  String code = "";
  code(tree,code,codeword);
  String encoded = encode(codeword,text);
  File freqTable = new File(freqfile);
  File compressed = new File(outputfile);
  PrintWriter p = new PrintWriter(freqTable);
  p.println("Letter  Frequency");
  for (int i = 0;i<128;i++)
  {
  if(i>31)
  {
    p.println(freq[i]);
    }
   else
   {
  p.println(freq[i]);
  }
  }
  p.close();
  PrintWriter pw = new PrintWriter(compressed);
  pw.println(encoded);
  pw.close();
}
```

Although our main method appears large, all that it does is read in the text file, call the methods previously stated, and export the frequency table as well as the compressed text.

Program 2: Decoding a Compressed File

Objective: Write a program that will decode a compressed file. Your program should prompt the user for two file names:

Compressed file:
Text file:

If the name of the compressed file is *something.zip*, your program should use that file and also the file *something.feq* as input.

```java
public static void main(String[]args) throws IOException
{
   Scanner user = new Scanner(System.in);
   System.out.println("Compressed file name:");
   String compressedfile = user.nextLine();
   System.out.println("Output Text file name:");
   String textfile = user.nextLine();
   System.out.println("Frequency file name:");
   String freqfile = user.nextLine();
   File inFile = new File(freqfile);
   Scanner inputstream = new Scanner(inFile);
   int [] freq = new int[128];
   inputstream.nextLine();
   int  counter = 0;

   while(inputstream.hasNext())
   {
      String s = inputstream.nextLine();
      int z = Integer.parseInt(s);
      freq[counter] = z;
      counter++;
   }// end while
```

FIG 2.1: Decode Main Method

We start by asking the user for the name of the compressed file, the output file, and the frequency file for the compressed files. Once we have the frequency file, we create an array of size 128 and read the file. The way this file is formated, we are able to read the file line by line which each line represents the frequency of that ASCII/Character value. This will allow us to create a Huffman tree which we create by passing the frequency array to the priority method as seen in Problem 1.

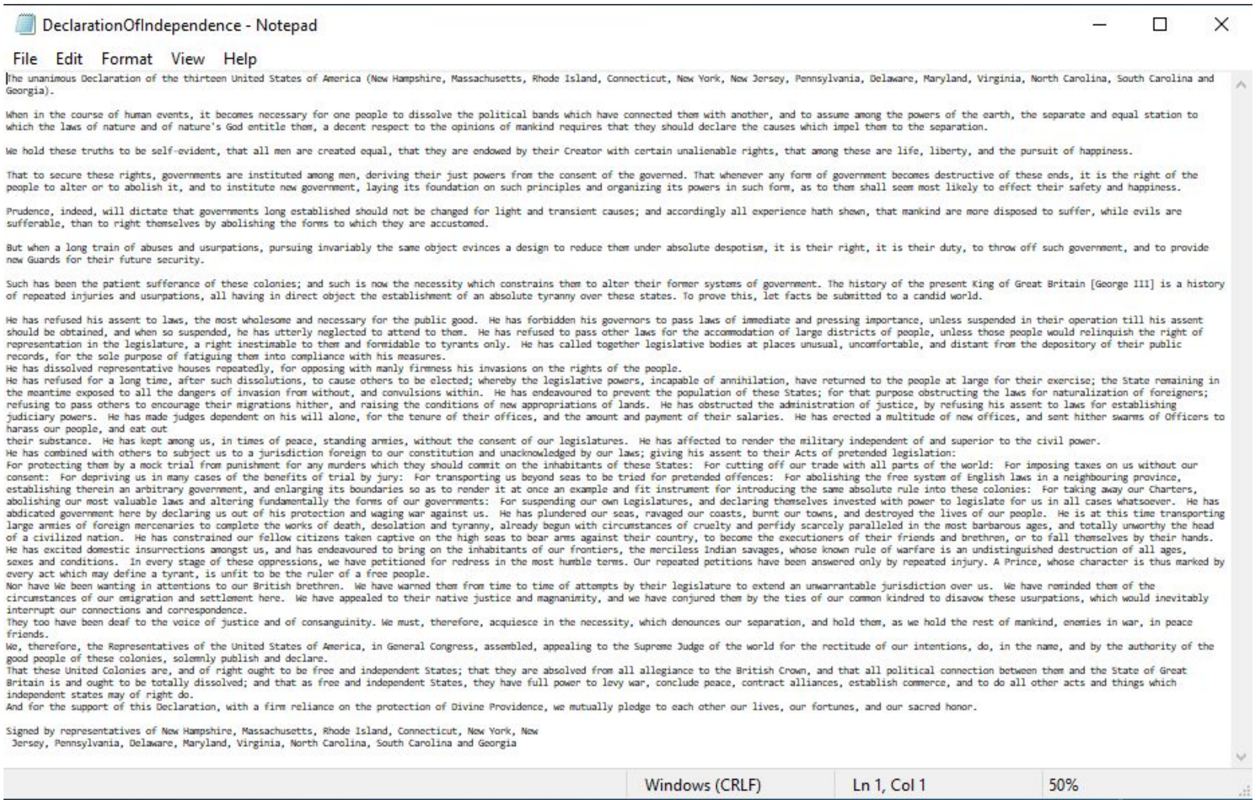
```
    Node tree = priority(freq);
    String code = "";
    String [] codeword = new String[128];
    code(tree,code,codeword);
    File zipfile = new File(compressedfile);
    Scanner input = new Scanner(zipfile);
    String encoded = input.nextLine();
    String letter ="";
    File textfiles = new File(textfile);
    PrintWriter p = new PrintWriter(textfiles);
    for(int i = 0; i < encoded.length(); i++)
    {
       letter = letter + encoded.charAt(i);
       for(int j = 0; j < 128; j++)
       {
          if(letter.equals(codeword[j]))
          {
             System.out.print((char)j);
             p.print((char)j);
             letter = "";
          }//end if
       }//end inside forloop
    }//end outside for loop
    p.close();
    System.out.println();
 }//end main
```

FIG 2.2: Main Method Cont.

To decode the compressed file we need the codewords from the Huffman tree that we created. We do this using the code method which will store the code words ascending by ASCII value 0-127. Although not every character will have a code word, it's only the characters that have a non-zero value from the frequency table. After we have gotten all of the code words, we then need to read in the compressed file and iterate through it to match a codeword. We have a for loop that will iterate through each character of the encoded text and save it to a string. We then check every codeword that we have to see if there is a match. If there is a match we print out the corresponding character to the screen and the output file and set the string back to empty. If there is no match we then read the next character and concatenate it to the previous string that was created.

Exercise 3: Encoding and Decoding Famous Texts

The Declaration of Independence: Original Text File (8.28 KB)


DeclarationOfIndependence - Notepad

File Edit Format View Help

The unanimous Declaration of the thirteen United States of America (New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Delaware, Maryland, Virginia, North Carolina, South Carolina and Georgia).

When in the course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the laws of nature and of nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable rights, that among these are life, liberty, and the pursuit of happiness.

That to secure these rights, governments are instituted among men, deriving their just powers from the consent of the governed. That whenever any form of government becomes destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their safety and happiness.

Prudence, indeed, will dictate that governments long established should not be changed for light and transient causes; and accordingly all experience hath shewn, that mankind are more disposed to suffer, while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed.

But when a long train of abuses and usurpations, pursuing invariably the same object evinces a design to reduce them under absolute despotism, it is their right, it is their duty, to throw off such government, and to provide new Guards for their future security.

Such has been the patient sufferance of these colonies; and such is now the necessity which constrains them to alter their former systems of government. The history of the present King of Great Britain [George III] is a history of repeated injuries and usurpations, all having in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world.

He has refused his assent to laws, the most wholesome and necessary for the public good. He has forbidden his governors to pass laws of immediate and pressing importance, unless suspended in their operation till his assent should be obtained, and when so suspended, he has utterly neglected to attend to them. He has refused to pass other laws for the accommodation of large districts of people, unless those people would relinquish the right of representation in the legislature, a right inestimable to them and formidable to tyrants only. He has called together legislative bodies at places unusual, uncomfortable, and distant from the depository of their public records, for the sole purpose of fatiguing them into compliance with his measures.
He has dissolved representative houses repeatedly, for opposing with manly firmness his invasions on the rights of the people.
He has refused for a long time, after such dissolutions, to cause others to be elected; whereby the legislative powers, incapable of annihilation, have returned to the people at large for their exercise; the State remaining in the meantime exposed to all the dangers of invasion from without, and convulsions within. He has endeavoured to prevent the population of these States; for that purpose obstructing the laws for naturalization of foreigners; refusing to pass others to encourage their migrations hither, and raising the conditions of new appropriations of lands. He has obstructed the administration of justice, by refusing his assent to laws for establishing judiciary powers. He has made judges dependent on his will alone, for the tenure of their offices, and the amount and payment of their salaries. He has erected a multitude of new offices, and sent hither swarms of Officers to harass our people, and eat out
their substance. He has kept among us, in times of peace, standing armies, without the consent of our legislatures. He has affected to render the military independent of and superior to the civil power.
He has combined with others to subject us to a jurisdiction foreign to our constitution and unacknowledged by our laws; giving his assent to their Acts of pretended legislation:
For protecting them by a mock trial from punishment for any murders which they should commit on the inhabitants of these States: For cutting off our trade with all parts of the world: For imposing taxes on us without our consent: For depriving us in many cases of the benefits of trial by jury: For transporting us beyond seas to be tried for pretended offences: For abolishing the free system of English laws in a neighbouring province, establishing therein an arbitrary government, and enlarging its boundaries so as to render it at once an example and fit instrument for introducing the same absolute rule into these colonies: For taking away our Charters, abolishing our most valuable laws and altering fundamentally the forms of our governments: For suspending our own Legislatures, and declaring themselves invested with power to legislate for us in all cases whatsoever. He has abdicated government here by declaring us out of his protection and waging war against us. He has plundered our seas, ravaged our coasts, burnt our towns, and destroyed the lives of our people. He is at this time transporting large armies of foreign mercenaries to complete the works of death, desolation and tyranny, already begun with circumstances of cruelty and perfidy scarcely paralleled in the most barbarous ages, and totally unworthy the head of a civilized nation. He has constrained our fellow citizens taken captive on the high seas to bear arms against their country, to become the executioners of their friends and brethren, or to fall themselves by their hands.
He has excited domestic insurrections amongst us, and has endeavoured to bring on the inhabitants of our frontiers, the merciless Indian savages, whose known rule of warfare is an undistinguished destruction of all ages, sexes and conditions. In every stage of these oppressions, we have petitioned for redress in the most humble terms. Our repeated petitions have been answered only by repeated injury. A Prince, whose character is thus marked by every act which may define a tyrant, is unfit to be the ruler of a free people.
Nor have We been wanting in attentions to our British brethren. We have warned them from time to time of attempts by their legislature to extend an unwarrantable jurisdiction over us. We have reminded them of the circumstances of our emigration and settlement here. We have appealed to their native justice and magnanimity, and we have conjured them by the ties of our common kindred to disavow these usurpations, which would inevitably interrupt our connections and correspondence.
They too have been deaf to the voice of justice and of consanguinity. We must, therefore, acquiesce in the necessity, which denounces our separation, and hold them, as we hold the rest of mankind, enemies in war, in peace friends.
We, therefore, the Representatives of the United States of America, in General Congress, assembled, appealing to the Supreme Judge of the world for the rectitude of our intentions, do, in the name, and by the authority of the good people of these colonies, solemnly publish and declare.
That these United Colonies are, and of right ought to be free and independent States; that they are absolved from all allegiance to the British Crown, and that all political connection between them and the State of Great Britain is and ought to be totally dissolved; and that as free and independent States, they have full power to levy war, conclude peace, contract alliances, establish commerce, and to do all other acts and things which independent states may of right do.
And for the support of this Declaration, with a firm reliance on the protection of Divine Providence, we mutually pledge to each other our lives, our fortunes, and our sacred honor.

Signed by representatives of New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Delaware, Maryland, Virginia, North Carolina, South Carolina and Georgia

Windows (CRLF) | Ln 1, Col 1 | 50%

The Declaration of Independence: Compressed Text File (35.4 KB)

__The Declaration of Independence:__ Decoded Text File (8.28 KB)

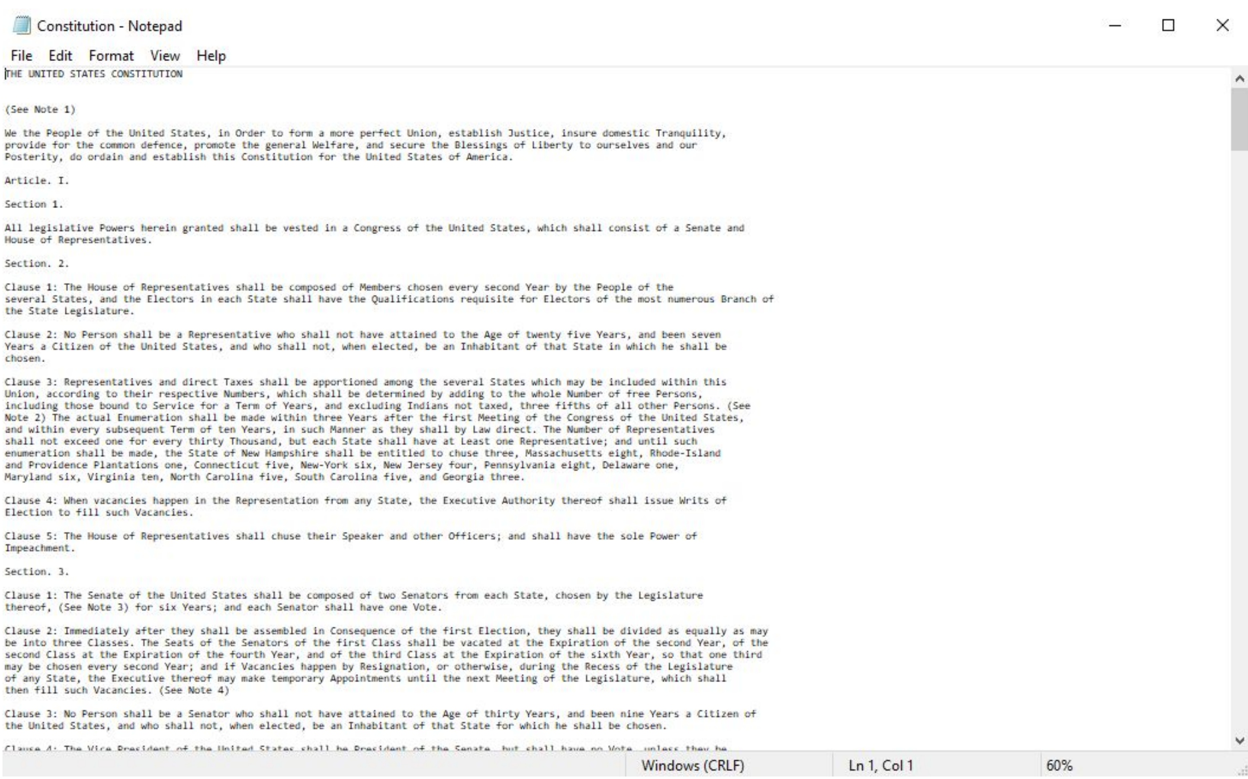
DeclarationOfIndependence1 - Notepad

File Edit Format View Help

The unanimous Declaration of the thirteen United States of America (New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Delaware, Maryland, Virginia, North Carolina, South Carolina and Georgia).When in the course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the laws of nature and of nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation. We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable rights, that among these are life, liberty, and the pursuit of happiness. That to secure these rights, governments are instituted among men, deriving their just powers from the consent of the governed. That whenever any form of government becomes destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their safety and happiness. Prudence, indeed, will dictate that governments long established should not be changed for light and transient causes; and accordingly all experience hath shewn, that mankind are more disposed to suffer, while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, pursuing invariably the same object evinces a design to reduce them under absolute despotism, it is their right, it is their duty, to throw off such government, and to provide new Guards for their future security. Such has been the patient sufferance of these colonies; and such is now the necessity which constrains them to alter their former systems of government. The history of the present King of Great Britain [George III] is a history of repeated injuries and usurpations, all having in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world. He has refused his assent to laws, the most wholesome and necessary for the public good. He has forbidden his governors to pass laws of immediate and pressing importance, unless suspended in their operation till his assent should be obtained, and when so suspended, he has utterly neglected to attend to them. He has refused to pass other laws for the accommodation of large districts of people, unless those people would relinquish the right of representation in the legislature, a right inestimable to them and formidable to tyrants only. He has called together legislative bodies at places unusual, uncomfortable, and distant from the depository of their public records, for the sole purpose of fatiguing them into compliance with his measures. He has dissolved representative houses repeatedly, for opposing with manly firmness his invasions on the rights of the people. He has refused for a long time, after such dissolutions, to cause others to be elected; whereby the legislative powers, incapable of annihilation, have returned to the people at large for their exercise; the State remaining in the meantime exposed to all the dangers of invasion from without, and convulsions within. He has endeavoured to prevent the population of these States; for that purpose obstructing the laws for naturalization of foreigners; refusing to pass others to encourage their migrations hither, and raising the conditions of new appropriations of lands. He has obstructed the administration of justice, by refusing his assent to laws for establishing judiciary powers. He has made judges dependent on his will alone, for the tenure of their offices, and the amount and payment of their salaries. He has erected a multitude of new offices, and sent hither swarms of Officers to harass our people, and eat outtheir substance. He has kept among us, in times of peace, standing armies, without the consent of our legislatures. He has affected to render the military independent of and superior to the civil power. He has combined with others to subject us to a jurisdiction foreign to our constitution and unacknowledged by our laws; giving his assent to their Acts of pretended legislation: For protecting them by a mock trial from punishment for any murders which they should commit on the inhabitants of these States: For cutting off our trade with all parts of the world: For imposing taxes on us without our consent: For depriving us in many cases of the benefits of trial by jury: For transporting us beyond seas to be tried for pretended offences: For abolishing the free system of English laws in a neighbouring province, establishing therein an arbitrary government, and enlarging its boundaries so as to render it at once an example and fit instrument for introducing the same absolute rule into these colonies: For taking away our Charters, abolishing our most valuable laws and altering fundamentally the forms of our governments: For suspending our own Legislatures, and declaring themselves invested with power to legislate for us in all cases whatsoever. He has abdicated government here by declaring us out of his protection and waging war against us. He has plundered our seas, ravaged our coasts, burnt our towns, and destroyed the lives of our people. He is at this time transporting large armies of foreign mercenaries to complete the works of death, desolation and tyranny, already begun with circumstances of cruelty and perfidy scarcely paralleled in the most barbarous ages, and totally unworthy the head of a civilized nation. He has constrained our fellow citizens taken captive on the high seas to bear arms against their country, to become the executioners of their friends and brethren, or to fall themselves by their hands. He has excited domestic insurrections amongst us, and has endeavoured to bring on the inhabitants of our frontiers, the merciless Indian savages, whose known rule of warfare is an undistinguished destruction of all ages,sexes and conditions. In every stage of these oppressions, we have petitioned for redress in the most humble terms. Our repeated petitions have been answered only by repeated injury. A Prince, whose character is thus marked by every act which may define a tyrant, is unfit to be the ruler of a free people. Nor have We been wanting in attentions to our British brethren. We have warned them from time to time of attempts by their legislature to extend an unwarrantable jurisdiction over us. We have reminded them of the circumstances of our emigration and settlement here. We have appealed to their native justice and magnanimity, and we have conjured them by the ties of our common kindred to disavow these usurpations, which would inevitably interrupt our connections and correspondence. They too have been deaf to the voice of justice and of consanguinity. We must, therefore, acquiesce in the necessity, which denounces our separation, and hold them, as we hold the rest of mankind, enemies in war, in peace friends. We, therefore, the Representatives of the United States of America, in General Congress, assembled, appealing to the Supreme Judge of the world for the rectitude of our intentions, do, in the name, and by the authority of the good people of these colonies, solemnly publish and declare. That these United Colonies are, and of right ought to be free and independent States; that they are absolved from all allegiance to the British Crown, and that all political connection between them and the State of Great Britain is and ought to be totally dissolved; and that as free and independent States, they have full power to levy war, conclude peace, contract alliances, establish commerce, and to do all other acts and things which independent states may of right do. And for the support of this Declaration, with a firm reliance on the protection of Divine Providence, we mutually pledge to each other our lives, our fortunes, and our sacred honor. Signed by representatives of New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Delaware, Maryland, Virginia, North Carolina, South Carolina and Georgia

Windows (CRLF) Ln 1, Col 1 70%

__The US Constitution:__ Original Text File (35.4 KB)



Constitution - Notepad

File Edit Format View Help

THE UNITED STATES CONSTITUTION

(See Note 1)

We the People of the United States, in Order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common defence, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution for the United States of America.

Article. I.

Section 1.

All legislative Powers herein granted shall be vested in a Congress of the United States, which shall consist of a Senate and House of Representatives.

Section. 2.

Clause 1: The House of Representatives shall be composed of Members chosen every second Year by the People of the several States, and the Electors in each State shall have the Qualifications requisite for Electors of the most numerous Branch of the State Legislature.

Clause 2: No Person shall be a Representative who shall not have attained to the Age of twenty five Years, and been seven Years a Citizen of the United States, and who shall not, when elected, be an Inhabitant of that State in which he shall be chosen.

Clause 3: Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, three fifths of all other Persons. (See Note 2) The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct. The Number of Representatives shall not exceed one for every thirty Thousand, but each State shall have at Least one Representative; and until such enumeration shall be made, the State of New Hampshire shall be entitled to chuse three, Massachusetts eight, Rhode-Island and Providence Plantations one, Connecticut five, New-York six, New Jersey four, Pennsylvania eight, Delaware one, Maryland six, Virginia ten, North Carolina five, South Carolina five, and Georgia three.

Clause 4: When vacancies happen in the Representation from any State, the Executive Authority thereof shall issue Writs of Election to fill such Vacancies.

Clause 5: The House of Representatives shall chuse their Speaker and other Officers; and shall have the sole Power of Impeachment.

Section. 3.

Clause 1: The Senate of the United States shall be composed of two Senators from each State, chosen by the Legislature thereof, (See Note 3) for six Years; and each Senator shall have one Vote.

Clause 2: Immediately after they shall be assembled in Consequence of the first Election, they shall be divided as equally as may be into three Classes. The Seats of the Senators of the first Class shall be vacated at the Expiration of the second Year, of the second Class at the Expiration of the fourth Year, and of the third Class at the Expiration of the sixth Year, so that one third may be chosen every second Year; and if Vacancies happen by Resignation, or otherwise, during the Recess of the Legislature of any State, the Executive thereof may make temporary Appointments until the next Meeting of the Legislature, which shall then fill such Vacancies. (See Note 4)

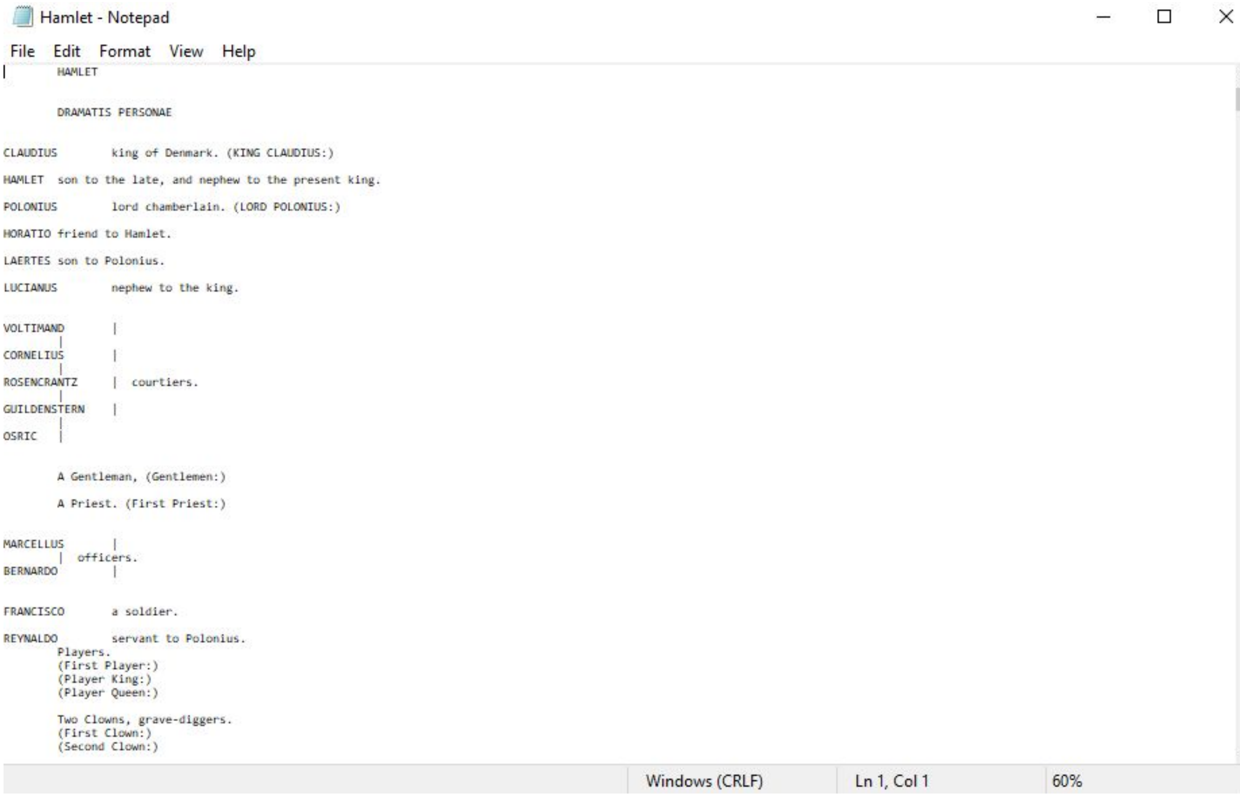
Clause 3: No Person shall be a Senator who shall not have attained to the Age of thirty Years, and been nine Years a Citizen of the United States, and who shall not, when elected, be an Inhabitant of that State for which he shall be chosen.

Windows (CRLF) Ln 1, Col 1 60%

The US Constitution: Compressed Text File (157 KB)

The US Constitution: Decoded Text File (35.4 KB)

Hamlet: Original Text File (208 KB)




```
        HAMLET


        DRAMATIS PERSONAE


CLAUDIUS        king of Denmark. (KING CLAUDIUS:)

HAMLET  son to the late, and nephew to the present king.

POLONIUS        lord chamberlain. (LORD POLONIUS:)

HORATIO friend to Hamlet.

LAERTES son to Polonius.

LUCIANUS        nephew to the king.


VOLTIMAND       |
        |
CORNELIUS       |
        |
ROSENCRANTZ     |  courtiers.
        |
GUILDENSTERN    |
        |
OSRIC   |


        A Gentleman, (Gentlemen:)

        A Priest. (First Priest:)


MARCELLUS       |
        |  officers.
BERNARDO        |


FRANCISCO       a soldier.

REYNALDO        servant to Polonius.
        Players.
        (First Player:)
        (Player King:)
        (Player Queen:)

        Two Clowns, grave-diggers.
        (First Clown:)
        (Second Clown:)
```

Hamlet: CompressedText File (866 KB)

<u>Hamlet:</u> <u></u>DecodedText File (208 KB)

Program 3: Real Life Compression and How it Works

Objective: In order to realize the true "real-life" compression, you must convert the bytes into bits. To do this, every eight '0' or '1' characters (bytes) is packed into a single byte.

```
public class RealLifeCompression
{
   public static void main(String[]args)throws IOException
   {
      Scanner input = new Scanner(System.in);

      System.out.println("Ented Compressed File");
      String  encodedFile = input.nextLine();
      File inFile = new File(encodedFile);
      Scanner inputstream = new Scanner(inFile);
      int count = 1;
      int index = 0;
      String encoded  = inputstream.nextLine();
      int stringlength = encoded.length();
      byte [] array = new byte[stringlength/8];
      byte temp = 0;
      for(int j = 1; j <= stringlength; j++)
      {
         char b = encoded.charAt(j-1);
         if(b == '0')
         {
            temp = (byte)(temp*2 + 0);
         }
         else if(b == '1')
         {
            temp = (byte)(temp*2 + 1);
         }
         if(j%8 == 0)
         {
            array[index] = temp;
            index++;
            temp = 0;
         }
      }//end for loop
      System.out.println("Enter the output file name");
      String out = input.nextLine();
      FileOutputStream fos = new FileOutputStream(out);
      fos.write(array);
      fos.close();
   }//end main
```

FIG 3.1: Real Life Compression Main Method

First, we read in the compressed file which was stored with 1's and 0's. We took 8 bytes at a time and pack it into a single byte and then printed it to a file that contained the compressed bytes.

Real Life Compression:

	Text File	WinZip	Huffman
Declaration of Independence	8.28 KB	3.63 KB	35.4 KB
Constitution	35.4 KB	11.9 KB	157 KB
Hamlet	208 KB	75.3 KB	866 KB

Exercise 4: Why does the Huffman's Algorithm Produce a Code with the Prefix Property:

The Huffman's Algorithm produces code with the prefix property because this is the only way the user and computer is able to determine which set of 1's and 0's are a corresponding letter. A code is a *prefix code* (or has the *prefix property*) if no codeword is the prefix of another codeword. A Huffman code turns out to be an "optimal" prefix code. Without this precise tree being created based on the frequency of that letter occurring in the encoding, we would not be able to determine the correct code-words for each ASCII value. The tree is able to tell us a lot just by its leaves. Which to get to each leaf of the Huffman tree we are able to determine the code for each letter, based directly on the path at which we take from the root of the tree. Going to the left child is a 0 and going to the right child is a 1. We do this until we get to a leaf then concatenate them together to create the codeword, which is in terms of ones and zeros and corresponds to an ASCII value/character.

Exercise 5

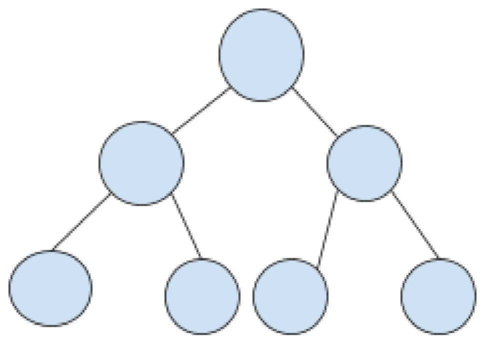
Objective: Prove that an optimal prefix code for a file is always represented by a full binary tree

A full binary tree is a binary tree in whichever non-leaf node has two children. Notice that the Huffman trees in our examples have been full trees.

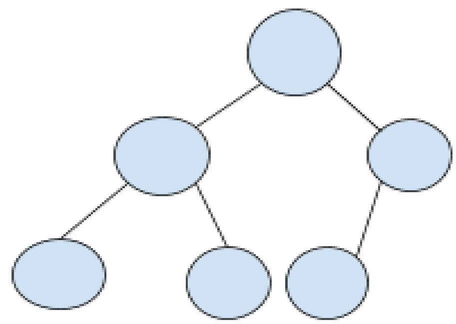
Hint: Suppose there is a non-full tree that did represent an optimal code. Find a contradiction by producing a better code.

Proof:

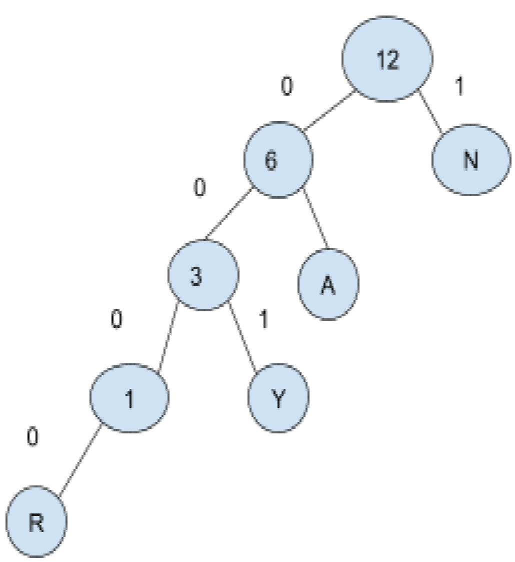
Full Tree



Non-Full Tree



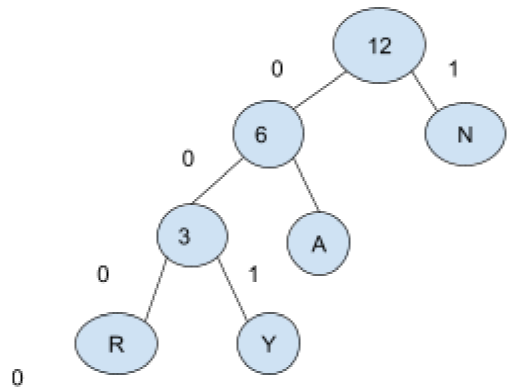
Non - Full Prefix Tree



Char	Code
R	000**0**
Y	001
A	01
N	1

Average Length: (**4**+3+2+1)/4 = 10/4

Full Prefix Tree



Char	Code
R	000
Y	001
A	01
N	1

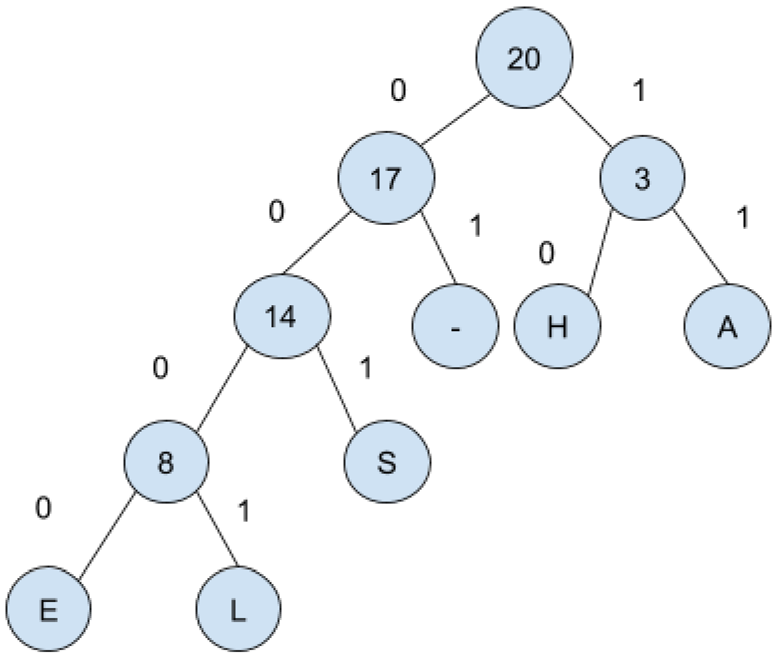
Average Length: (**3**+3+2+1)/4 = 9/4

- Since the code in the full tree has an average code-length of 9/4 while the non-full tree has an average length 10/4, the full tree is the optimal choice. Any non-full tree could become full by combining the solo leaf with its root.

Exercise 6

Show by means of a counterexample that a full binary tree may represent a prefix code that is not optimal, even if all the letters with lower frequencies have longer codewords.

Suggestion: Build another full tree for "SHE-SELLS-SEA-SHELLS" which has a higher cost than the tree of our previous examples.



Character	Frequency
A	1
H	2
-	3
E	4
L	4
S	6

In creating this counter example the characters with the highest frequency were swapped with the ones with the lowest frequency. Using the original idea of the huffman tree we combined the two highest frequency roots to grow this tree. With this in mind the cost of the new binary tree was found to be 62, and this is a decently larger size than the original cost of 49.

Formula Used To Determine The Cost

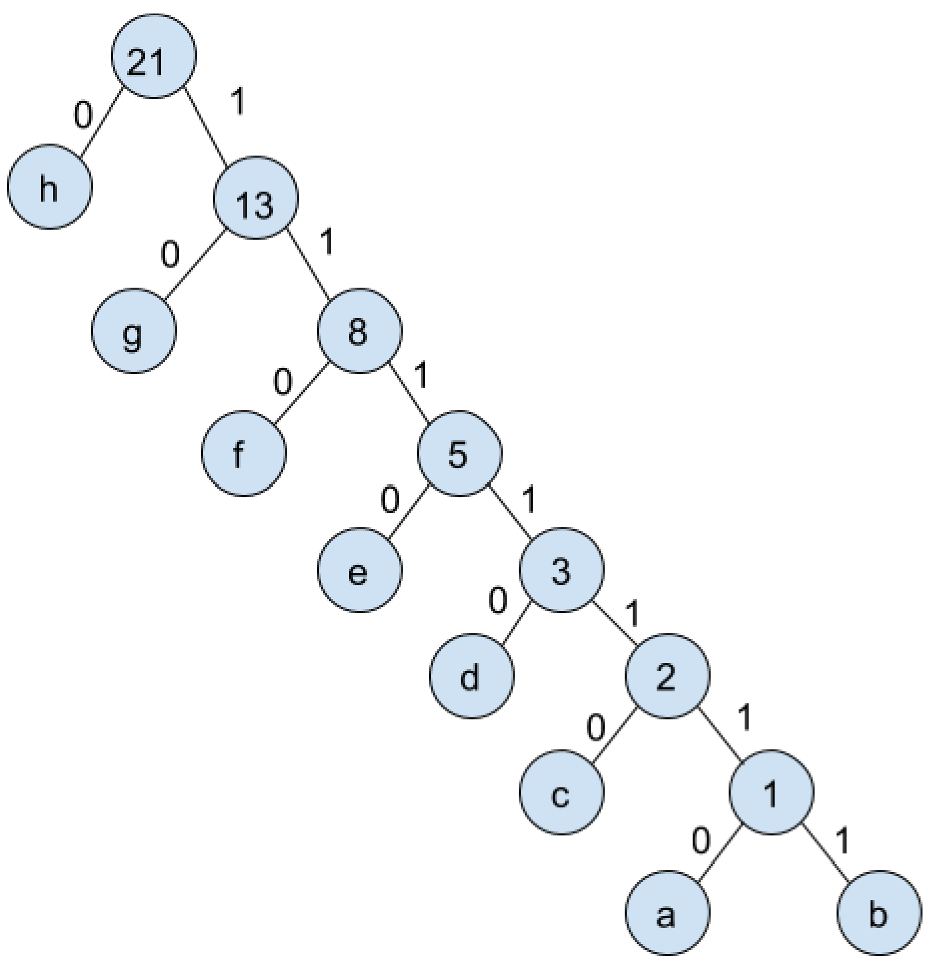
Cost(T) = frequency(T) * length(T);

Exercise 9: Fibonacci Numbers

The Huffman tree for frequencies equal to the first *n* Fibonacci numbers results in a very unbalanced encoding of *n* characters. Suppose that the frequencies for the characters a, b, c, d, e, f, g, and h are the first 8 Fibonacci numbers: 1,1,2,3,5,8,13, and 21.

The below string represents the corresponding frequencies with its letter.

abccdddeeeeeffffffffggggggggggggghhhhhhhhhhhhhhhhhhhhh
We then ran this text file in our encode program to find the exact codewords that would be created from an input like this.
Codewords:
a 1111110
b 1111111
c 111110
d 11110
e 1110
f 110
g 10
h 0
Huffman Tree:



From this tree, we can calculate the cost of this tree is 132.

a. Discover a formula for the sum of the first *n* Fibonacci numbers, and prove it by induction.

n	Fibonacci(*n*)	SumFib(*n*) = Fibonacci(1) + Fibonacci(2) + … + Fibonacci(*n*)
1	1	1
2	1	2
3	2	4
4	3	7
5	5	12
6	8	20
7	13	33
8	21	54
9	34	88

Notice from the table it appears that the sum of the first n terms is the (nth+2) term minus 1. SumFib(*n*) = Fibonacci(n+2) -1. We will use mathematical induction to prove that in fact, this is the correct formula to determine the sum of the first n terms of the Fibonacci sequence.

Let's say that F(n) represents Fibonacci(1) + Fibonacci(2)+ … +Fibonacci(*n*) = Fibonacci(n+2)-1.

Since Fibonacci(1) = Fibonacci(1+2) -1 = Fibonacci(3) - 1 = 2 -1 = 1 we can say that F(1) is true.

Assume F(k) is true for k > 0

Fibonacci(1) + Fibonacci(2)+ … +Fibonacci(*k*) = Fibonacci(k+2)-1.

Find the next of k, i.e. k+1

Fibonacci(1) + Fibonacci(2)+ … +Fibonacci(*k*) + Fibonacci(*k+1*) = Fibonacci(k+2)-1 + Fibonacci(*k+1*)

Fibonacci(1) + Fibonacci(2)+ … +Fibonacci(*k*) + Fibonacci(*k+1*) = Fibonacci(k+1) + Fibonacci(*k+2*) - 1

Since Fibonacci($k+1$) + Fibonacci($k+2$) = Fibonacci($k+3$) by the definition of Fibonacci numbers we can simplify the right side of this equation.

Fibonacci(1) + Fibonacci(2)+ ... +Fibonacci(k) + Fibonacci($k+1$) = Fibonacci(k+3) - 1

b. Prove that C(n) = C(n-1) + SumFib(n), where C(1) = 0.

n	Fibonacci(n)	C(n) = Cost of Huffman tree for the first n Fibonacci numbers
1	1	0*1 = 0
2	1	1*1+1*1 = 2
3	2	2*1 + 2*1 + 1*2 = 6
4	3	3*1 + 3*1 + 2*2 + 1*3 = 13
5	5	5*1 + 5*1 + 3*2 + 2*3 + 1*4 = 26
6	8	8*1 + 8*1 + 5*2 + 3*3 + 2*4 + 1*5 = 48
7	13	13*1 + 13*1 + 8*2 + 5*3 + 3*4 + 2*5 + 1*6 = 85
8	21	21*1 + 21*1 + 13*2 + 8*3 + 5*4 + 3*5 + 2*6 + 1*7 = 146
9	34	34*1 + 34*1 + 21*2 + 13*3 + 8*4 + 5*5 + 3*6 + 2*7 + 1*8 = 246

To prove that the equation C(n) = C(n-1) + SumFib(n) we have to understand what exactly this equation represents. From the chart we see this simply math to represent the cost of the tree. Each time we add one to n, we are adding another node to the top of the pre-existing tree, which we are able to use this single tree to find the cost of the tree. What this equation really means is it adds up the sum of the Fibonacci numbers recursively. For example say n = 4. C(4) = SumFib(4) + SumFib(4-1) + SumFib(4-2) + SumFib(4-3). Which is equal to 13 following the base case of of C(1) = 0, and from this equation SumFib(1) = 0.

c. Using part (a), solve C(n) = C(n-1) + SumFib(n), where C(1) = 0,. by repeated substitution, and write C(n) as a closed form in terms of F(n), for n > 1.

C(n) = C(n-1) + SumFib(n)

C(n-1) = C(n-2) + SumFib(n-1)

C(n-2) = C(n-3) + SumFib(n-2)

C(n) = C(n-2) + SumFib(n-1) + SumFib(n)

$C(n) = C(n-3) + SumFib(n-2) + SumFib(n-1) + SumFib(n)$

After many reps $C(1) = 0$ and $SumFib(n-2) + SumFib(n-1) + SumFib(n)$ becomes as series and is equal $SumFib(n+2) - SumFib(3)$.

$C(n) = C(1) + SumFib(n+2) -SumFib(3) - (n-1)$

$SumFib(n+2) = F(n+4) -1 \quad SumFib(3) = 4$

$C(n) = 1 + F(n+4) - 1 - 4 - n + 1$

$C(n) = F(n+4) - n - 4$

Movie Quiz:

1. Fermat's profession was being a lawyer, but he is most well known for his mathematical discoveries.
2. Andrew Wiles lived in Cambridge while he was exploring Fermat's last theorem.
3. Fermat did not include the actual proof because the margins of his recording book were too small to carry out the proof fully.
4. There were three hundred years between the statement of Fermat's Last Theorem and its proof by Wiles.
5. The two mathematical subjects that "live on different planets" are elliptic curves and modular functions.
6. Taniyama was a mathematician that helped to develop the Taniyama-Shimura conjecture. The thing that happened to him that was discussed in the film was that in 1958 he committed suicide
7. The song "One Way or Another" that plays in the movie is by Blondie.
8. For her birthday Andrew Wiles' wife said she wanted a correct proof, and Whiles tried to carry this out within 2-3 weeks but came up short.
9. Wiles worked on Fermat's Last Theorem for 8 years.
10. From the list of mathematicians, Noam Elkies was the only one that was not present in this film.

What We Liked About This Lab

This lab was very interesting to learn about how different types of files are transmitted from one place to the other in the most optimal fashion. Through our work, we were able to see how something like a zip file works, and how data can be compressed without being lost or deleted. It was interesting to learn about the Huffman's code and see just another way that binary trees are implemented in different concepts. The idea of prefix codes was very interesting and seemed like an idea that was very similar to encrypting and decrypting that we worked through in the last lab. It was interesting to explore an idea related to sending and receiving messages and files in the world today.

What We Did Not Like About This Lab

For some of the codes we were required to use a priority queue and this proved to be difficult and rather confusing. We tried to implement the priority queue that is located on the data structures website but later found out that we needed to create our own and use that for these assignments. Some of the calculations in this assignment were challenging, but after being explained what we needed to calculate we were able to fix the problems.

What We Learned From This Lab

In this lab, we learned about how Huffman codes work, and how they use prefix codes and frequencies to determine the most optimal way to transmit information. We also learned how to build a Huffman tree and how the tree can help to determine the codeword for a specific character. This lab also showed us how to count the cost of the code and compare different ways of compression to find the best method. The Fibonacci numbers are very versatile when looking at different patterns that come out of them. If you look at the series and add them together we get a new equation that finds some of all of them recursively. While calculating the cost of a tree-based on the Fibonacci numbers we see that it is a recurrence equation and also has the recursive equation of the sum of the Fibonacci numbers in it as well. We see the bridge between Fibonacci numbers and that of Huffman encoding trees.