Administrative Details

• Lab 8 today!
  • No partners this week
  • Review before lab; come to lab with design doc
    • Read over the supplied resources!
Last Time

- Trees with more than 2 children
  - Representations
  - Application: Lab 8: Hex-a-pawn!

- Binary Trees
  - Traversals
    - As methods taking a BinaryTree parameter
    - With Iterators
Today

- Wrap up Binary Tree Iterators
- Breadth-First and Depth-First Search
- Array Representations of (Binary) Trees
- Application: Huffman Encoding
Lexicon Lab Tips

Tasks (in order of implementation!):

• Review all lab materials (including .java files!)
• Implement LexiconNode
  • Add a single method, then test it: add main()
• Implement LexiconTrie
  • Same approach, but can also use Main.java to test
• Implement in an order that allows immediate testing!
Recall from last class:

- **In-order**: “left, node, right”
- **Pre-order**: “node, left, right”
- **Post-order**: “left, right, node”
- **Level-order**: visit all nodes at depth i before depth i+1

Tree Traversals

Stack

Queue
public BTPostorderIterator(BinaryTree<E> root) {
    todo = new StackList<BinaryTree<E>>();
    this.root = root;
    reset();
}

public void reset() {
    todo.clear();
    BinaryTree<E> current = root;
    while (!current.isEmpty()) {
        todo.push(current);
        if (!current.left().isEmpty())
            current = current.left();
        else
            current = current.right();
    } // Top of stack is now left-most unvisited leaf
}
Post-Order Iterator

```java
public E next() {
    BinaryTree<E> current = todo.pop();
    E result = current.value();
    if (!todo.isEmpty()) {
        BinaryTree<E> parent = todo.get();
        if (current == parent.left()) {
            current = parent.right();
            while (!current.isEmpty()) {
                todo.push(current);
                if (!current.left().isEmpty())
                    current = current.left();
                else current = current.right();
            }
        }
    }
    return result;
}
```
Traversals & Searching

- We can use traversals for searching trees
- How might we search a tree for a value?
  - Breadth-First: Explore nodes near the root before nodes far away (level order traversal)
    - Nearest gas station
  - Depth-First: Explore nodes deep in the tree first (post-order traversal)
    - Solution to a maze
Loose Ends – Really Big Trees!

• In some situations, the tree we need might be too big or expensive to build completely
  • Or parts of it might not be needed
• Example: Game Trees
  • Chess: you wouldn’t build the entire tree, you would grow portions of it as needed (with some combination of depth/breadth first searching)
Alternative Tree Representations

- Total # “slots” = 4n
  - Since each BinaryTree maintains a reference to left, right, parent, value
- 2-4x more overhead than vector, SLL, array, …
- But trees capture successor and predecessor relationships that other data structures don’t…
Array-Based Binary Trees

- Encode structure of tree in array indexes
  - Put root at index 0
- Where are children of node \( i \)?
  - Children of node \( i \) are at \( 2i+1 \) and \( 2i+2 \)
    - Look at example
- Where is parent of node \( j \)?
  - Parent of node \( j \) is at \( (j-1)/2 \)
ArrayTree Tradeoffs

• Why are ArrayTrees good?
  • Save space for links
  • No need for additional memory allocated/garbage collected
  • Works well for full or complete trees
    • Complete: All levels except last are full and all gaps are at right
    • “A complete binary tree of height h is a full binary tree with 0 or more of the rightmost leaves of level h removed”

• Why bad?
  • Could waste a lot of space
  • Tree of height of h requires $2^{h+1} - 1$ array slots even if only $O(h)$ elements
Next up: Huffman Codes

• Computers encode a text as a sequence of bits

<table>
<thead>
<tr>
<th>Decimal</th>
<th>Hex</th>
<th>Char</th>
<th>Decimal</th>
<th>Hex</th>
<th>Char</th>
<th>Decimal</th>
<th>Hex</th>
<th>Char</th>
<th>Decimal</th>
<th>Hex</th>
<th>Char</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>(NULL)</td>
<td>32</td>
<td>20</td>
<td>(SPACE)</td>
<td>64</td>
<td>40</td>
<td>@</td>
<td>96</td>
<td>60</td>
<td>`</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>[START OF HEADING]</td>
<td>33</td>
<td>21</td>
<td>!</td>
<td>65</td>
<td>41</td>
<td>A</td>
<td>97</td>
<td>61</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>[START OF TEXT]</td>
<td>34</td>
<td>22</td>
<td>&quot;</td>
<td>66</td>
<td>42</td>
<td>B</td>
<td>98</td>
<td>62</td>
<td>b</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>[END OF TEXT]</td>
<td>35</td>
<td>23</td>
<td>#</td>
<td>67</td>
<td>43</td>
<td>C</td>
<td>99</td>
<td>63</td>
<td>c</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>[END OF TRANSMISSION]</td>
<td>36</td>
<td>24</td>
<td>$</td>
<td>68</td>
<td>44</td>
<td>D</td>
<td>100</td>
<td>64</td>
<td>d</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>[ENQUIRY]</td>
<td>37</td>
<td>25</td>
<td>%</td>
<td>69</td>
<td>45</td>
<td>E</td>
<td>101</td>
<td>65</td>
<td>e</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>[ACKNOWLEDGE]</td>
<td>38</td>
<td>26</td>
<td>&amp;</td>
<td>70</td>
<td>46</td>
<td>F</td>
<td>102</td>
<td>66</td>
<td>f</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>[Bell]</td>
<td>39</td>
<td>27</td>
<td>,</td>
<td>71</td>
<td>47</td>
<td>G</td>
<td>103</td>
<td>67</td>
<td>g</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>[BACKSPACE]</td>
<td>40</td>
<td>28</td>
<td>(</td>
<td>72</td>
<td>48</td>
<td>H</td>
<td>104</td>
<td>68</td>
<td>h</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>[HORIZONTAL TAB]</td>
<td>41</td>
<td>29</td>
<td>)</td>
<td>73</td>
<td>49</td>
<td>I</td>
<td>105</td>
<td>69</td>
<td>i</td>
</tr>
<tr>
<td>11</td>
<td>B</td>
<td>[VERTICAL TAB]</td>
<td>43</td>
<td>2B</td>
<td>+</td>
<td>75</td>
<td>4B</td>
<td>K</td>
<td>107</td>
<td>6B</td>
<td>k</td>
</tr>
<tr>
<td>12</td>
<td>C</td>
<td>[FORM FEED]</td>
<td>44</td>
<td>2C</td>
<td>,</td>
<td>76</td>
<td>4C</td>
<td>L</td>
<td>108</td>
<td>6C</td>
<td>l</td>
</tr>
<tr>
<td>13</td>
<td>D</td>
<td>[CARRIAGE RETURN]</td>
<td>45</td>
<td>2D</td>
<td>-</td>
<td>77</td>
<td>4D</td>
<td>M</td>
<td>109</td>
<td>6D</td>
<td>m</td>
</tr>
<tr>
<td>14</td>
<td>E</td>
<td>[SHIFT OUT]</td>
<td>46</td>
<td>2E</td>
<td>.</td>
<td>78</td>
<td>4E</td>
<td>N</td>
<td>110</td>
<td>6E</td>
<td>n</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>[SHIFT IN]</td>
<td>47</td>
<td>2F</td>
<td>/</td>
<td>79</td>
<td>4F</td>
<td>O</td>
<td>111</td>
<td>6F</td>
<td>o</td>
</tr>
<tr>
<td>16</td>
<td>10</td>
<td>[DATA LINK ESCAPE]</td>
<td>48</td>
<td>30</td>
<td>0</td>
<td>80</td>
<td>50</td>
<td>P</td>
<td>112</td>
<td>70</td>
<td>p</td>
</tr>
<tr>
<td>17</td>
<td>11</td>
<td>[DEVICE CONTROL 1]</td>
<td>49</td>
<td>31</td>
<td>1</td>
<td>81</td>
<td>51</td>
<td>Q</td>
<td>113</td>
<td>71</td>
<td>q</td>
</tr>
<tr>
<td>18</td>
<td>12</td>
<td>[DEVICE CONTROL 2]</td>
<td>50</td>
<td>32</td>
<td>2</td>
<td>82</td>
<td>52</td>
<td>R</td>
<td>114</td>
<td>72</td>
<td>r</td>
</tr>
<tr>
<td>19</td>
<td>13</td>
<td>[DEVICE CONTROL 3]</td>
<td>51</td>
<td>33</td>
<td>3</td>
<td>83</td>
<td>53</td>
<td>S</td>
<td>115</td>
<td>73</td>
<td>s</td>
</tr>
<tr>
<td>20</td>
<td>14</td>
<td>[DEVICE CONTROL 4]</td>
<td>52</td>
<td>34</td>
<td>4</td>
<td>84</td>
<td>54</td>
<td>T</td>
<td>116</td>
<td>74</td>
<td>t</td>
</tr>
<tr>
<td>21</td>
<td>15</td>
<td>[NEGATIVE ACKNOWLEDGE]</td>
<td>53</td>
<td>35</td>
<td>5</td>
<td>85</td>
<td>55</td>
<td>U</td>
<td>117</td>
<td>75</td>
<td>u</td>
</tr>
<tr>
<td>22</td>
<td>16</td>
<td>[SYNCHRONOUS IDLE]</td>
<td>54</td>
<td>36</td>
<td>6</td>
<td>86</td>
<td>56</td>
<td>V</td>
<td>118</td>
<td>76</td>
<td>v</td>
</tr>
<tr>
<td>23</td>
<td>17</td>
<td>[ENG OF TRANS. BLOCK]</td>
<td>55</td>
<td>37</td>
<td>7</td>
<td>87</td>
<td>57</td>
<td>W</td>
<td>119</td>
<td>77</td>
<td>w</td>
</tr>
<tr>
<td>24</td>
<td>18</td>
<td>[CANCEL]</td>
<td>56</td>
<td>38</td>
<td>8</td>
<td>88</td>
<td>58</td>
<td>X</td>
<td>120</td>
<td>78</td>
<td>x</td>
</tr>
<tr>
<td>25</td>
<td>19</td>
<td>[END OF MEDIUM]</td>
<td>57</td>
<td>39</td>
<td>9</td>
<td>89</td>
<td>59</td>
<td>Y</td>
<td>121</td>
<td>79</td>
<td>y</td>
</tr>
<tr>
<td>26</td>
<td>1A</td>
<td>[SUBSTITUTE]</td>
<td>58</td>
<td>3A</td>
<td>:</td>
<td>90</td>
<td>5A</td>
<td>Z</td>
<td>122</td>
<td>7A</td>
<td>z</td>
</tr>
<tr>
<td>27</td>
<td>1B</td>
<td>[ESCAPE]</td>
<td>59</td>
<td>3B</td>
<td>;</td>
<td>91</td>
<td>5B</td>
<td>[</td>
<td>123</td>
<td>7B</td>
<td>{</td>
</tr>
<tr>
<td>28</td>
<td>1C</td>
<td>[FILE SEPARATOR]</td>
<td>60</td>
<td>3C</td>
<td>&lt;</td>
<td>92</td>
<td>5C</td>
<td>\</td>
<td>124</td>
<td>7C</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>1D</td>
<td>[GROUP SEPARATOR]</td>
<td>61</td>
<td>3D</td>
<td>=</td>
<td>93</td>
<td>5D</td>
<td>]</td>
<td>125</td>
<td>7D</td>
<td>}</td>
</tr>
<tr>
<td>30</td>
<td>1E</td>
<td>[RECORD SEPARATOR]</td>
<td>62</td>
<td>3E</td>
<td>&gt;</td>
<td>94</td>
<td>5E</td>
<td>^</td>
<td>126</td>
<td>7E</td>
<td>~</td>
</tr>
<tr>
<td>31</td>
<td>1F</td>
<td>[UNIT SEPARATOR]</td>
<td>63</td>
<td>3F</td>
<td>?</td>
<td>95</td>
<td>5F</td>
<td>_</td>
<td>127</td>
<td>7F</td>
<td>[DEL]</td>
</tr>
</tbody>
</table>
Next up: Huffman Codes

- Goal: Encode a text as a sequence of bits
- Normally, use ASCII: 1 character = 8 bits (1 byte)
  - Allows for $2^8 = 256$ different characters
- ‘A’ = 01000001, ‘B’ = 01000010
- Space to store “AN_ANTARCTIC_PENGUIN”
  - 20 characters -> 20*8 bits = 160 bits
- Is there a better way?
  - Only 11 symbols are used (ANTRCIPEGU_)
  - Only need 4 bits per symbol (since $2^4 > 11$)!
    - 20*4 = 80 bits instead of 160!
  - Can we still do better??
Huffman Codes

- Example
  - AN_ANTARCTIC_PENGUIN
  - Compute letter frequencies

\[
\begin{array}{cccccccccccccc}
A & C & E & G & I & N & P & R & T & U & _ \\
3 & 2 & 1 & 1 & 2 & 4 & 1 & 1 & 2 & 1 & 2 \\
\end{array}
\]

- Key Idea: Use fewer bits for most common letters

\[
\begin{array}{cccccccccccccccc}
A & C & E & G & I & N & P & R & T & U & _ \\
3 & 2 & 1 & 1 & 2 & 4 & 1 & 1 & 2 & 1 & 2 \\
110 & 111 & 1011 & 1000 & 000 & 001 & 1001 & 1010 & 0101 & 0100 & 011 \\
\end{array}
\]

- Uses 67 bits to encode entire string
Huffman Codes

<table>
<thead>
<tr>
<th>A</th>
<th>C</th>
<th>E</th>
<th>G</th>
<th>I</th>
<th>N</th>
<th>P</th>
<th>R</th>
<th>T</th>
<th>U</th>
<th>_</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>110</td>
<td>111</td>
<td>1011</td>
<td>1000</td>
<td>000</td>
<td>001</td>
<td>1001</td>
<td>1010</td>
<td>0101</td>
<td>0100</td>
<td>011</td>
</tr>
</tbody>
</table>

- Uses 67 bits to encode entire string
- Can we do better?

<table>
<thead>
<tr>
<th>A</th>
<th>C</th>
<th>E</th>
<th>G</th>
<th>I</th>
<th>N</th>
<th>P</th>
<th>R</th>
<th>T</th>
<th>U</th>
<th>_</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>100</td>
<td>010</td>
<td>1100</td>
<td>1101</td>
<td>011</td>
<td>101</td>
<td>0001</td>
<td>0000</td>
<td>001</td>
<td>1110</td>
<td>1111</td>
</tr>
</tbody>
</table>

- Uses 67 bits to encode entire string
The Encoding Tree

Left = 0; Right = 1
Features of Good Encoding

• Prefix property: No encoding is a prefix of another encoding (letters appear at leaves)
• No internal node has a single child
• Nodes with lower frequency have greater depth

• All optimal length unambiguous encodings have these features
Huffman Encoding

• Input: symbols of alphabet with frequencies
• Huffman encode as follows
  • Create a single-node tree for each symbol: key is frequency; value is letter
  • while there is more than one tree
    • Find two trees T1 and T2 with lowest keys
    • Merge them into new tree T with dummy value and key= T1.key+ T2.key
• Theorem: The tree computed by Huffman is an optimal encoding for given frequencies
The Encoding Tree

Left = 0; Right = 1
How To Implement Huffman

• Keep a Vector of Binary Trees
• Sort them by decreasing frequency
  • Removing two smallest frequency trees is fast
• Insert merged tree into correct sorted location in Vector
• Running Time:
  • $O(n \log n)$ for initial sorting
  • $O(n^2)$ for rest: $O(n)$ re-insertions of merged trees
• Can we do better...?