Lists

EECS 233

Previous Lectures

- Memory management by PLs
- OO programming, ADT
- Recursion
- Mathematical background and running time analysis

We start to learn “data structures” today!

How to Represent A Sequence of Data?

- A sequence: an ordered (but not sorted) collection of items
  - 32, 5, 4, 24, 3, 5, 7, …
  - “David”, “Mark”, “Grace”, “Tim”, “Tim”, “Michael”, “David”, …
- Different from
  - A bag, which is not ordered
  - A set, which contains unique items

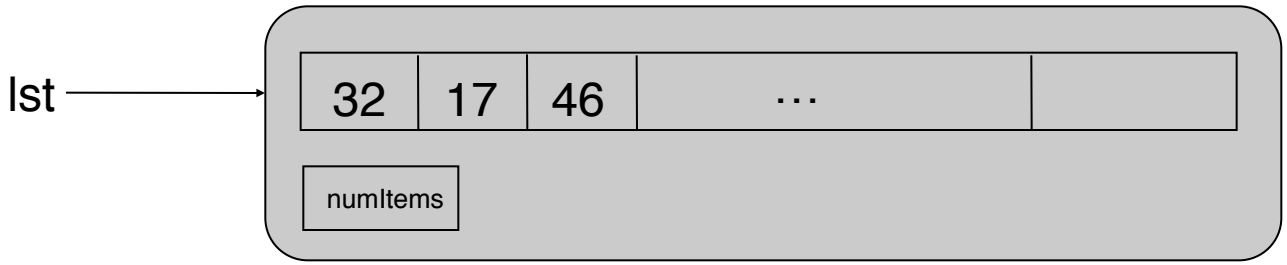
A List ADT

- View list as a black-box object with certain operations
  - Empty list on creation
  - Get an i-th element
  - Add a new element at position i
  - Remove an element at position i
  - Get the number of elements in the list
  - Check if a given element (specified by its position i or by a reference) is the last in the list
  - Etc…
- Java list interface:

```
1  public interface List<AnyType> extends Collection<AnyType>
2  {
3      AnyType get( int idx );
4      AnyType set( int idx, AnyType newVal );
5      void add( int idx, AnyType x );
6      void remove( int idx );
7
8      ListIterator<AnyType> listIterator( int pos );
9  }
```

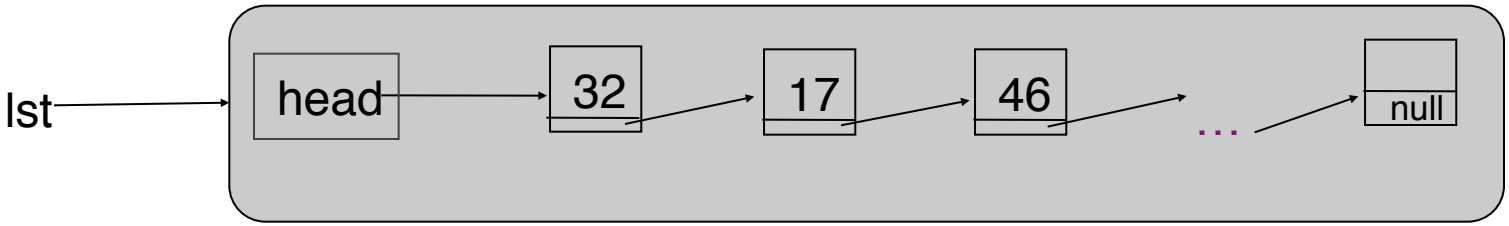
Which data structures to use for implementation??

Array Representation



- Store list elements wall-to-wall in memory in an array
- Keep a variable recording the current number of elements
- Advantages
  - Easy and efficient access to *any* item in the sequence
    - ✓ item[i] gives you the item at position i
    - ✓ Random access
      - Every item can be accessed in constant time given its index
  - Very compact: no auxiliary fields are required
- Disadvantages of using an array:
  - The need to specify an initial array size and resize as required (how?)
  - Difficult to insert/delete items at arbitrary positions (running time?)
  - May have many empty positions

Linked List Representation

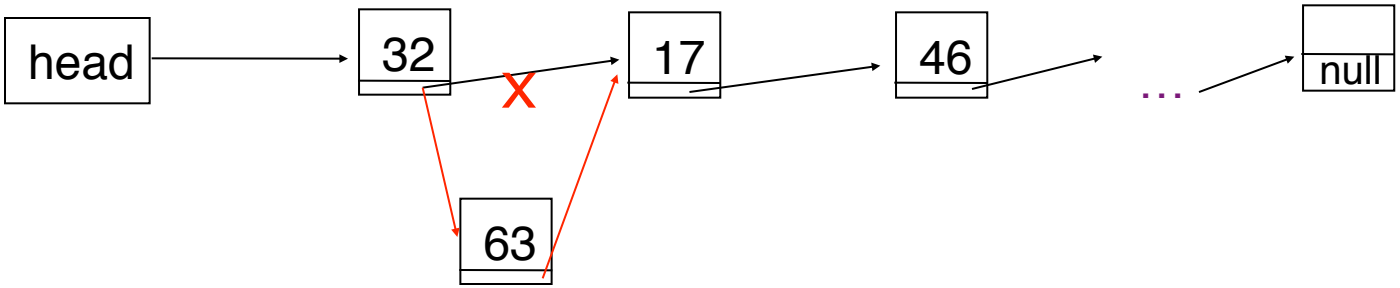


- More efficient data structure for a dynamic sequence (with frequent insert/delete operations)
- A linked list stores a sequence of items in separate *nodes*. Each node contains:
  - a single item, and
  - a "link" (i.e., a reference/pointer) to the node containing the next item

  The last node in the linked list has a link value of NULL or null.
- The linked list starts with a variable that holds a reference to the first node – head of the list

Advantages/Disadvantages of Linked List

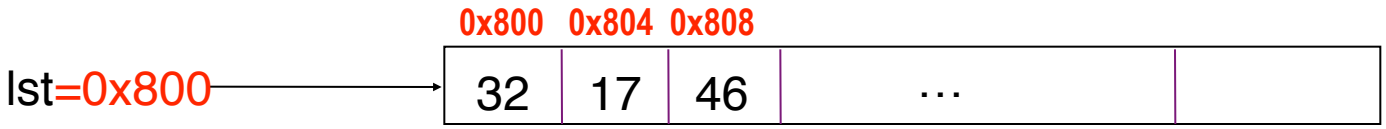
- Advantages:
  - No capacity limit (provided there is enough memory).
  - Easy to insert/delete an item – no need to "shift over" other items.
    - ✓ Done in constant time.



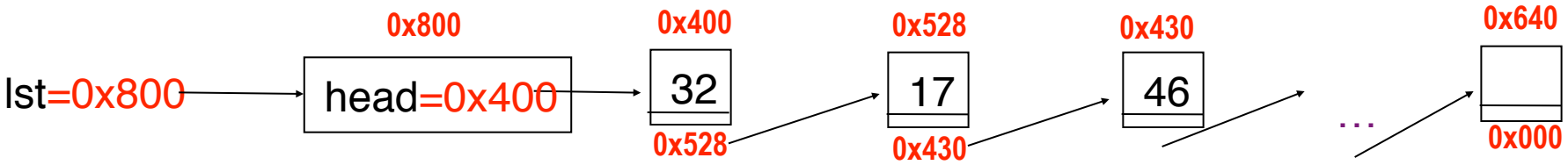
- Disadvantages:
  - No random access
    - ✓ "walk down" the list to access an item
  - Memory overhead for the links

Memory Management for Linked Lists

- In an array, the elements occupy consecutive memory locations in the heap:
  - Address in red

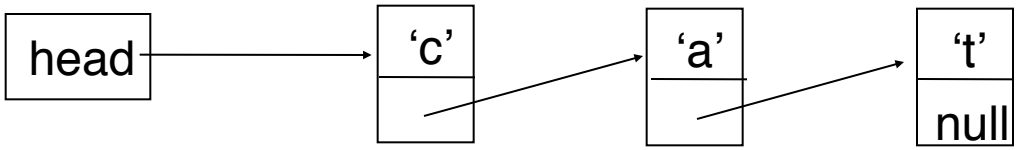


- In a linked list, each node is a distinct object in the heap. The nodes do *not* have to be next to each other in memory.



An Example Linked List

- A string represented using a linked list: **LLString**. Each node in the linked list represents one character.



- Java classes

```
public class StringNode {
        private char ch;
        private StringNode next;
        …
}
```

```
public class LLString {
        private StringNode head;
        private int theSize;
        …
```

Under the Hood of the String Linked List

- The string as a whole will be represented *(internally)* by a variable that holds a reference to the node containing the first character.

  StringNode str1;

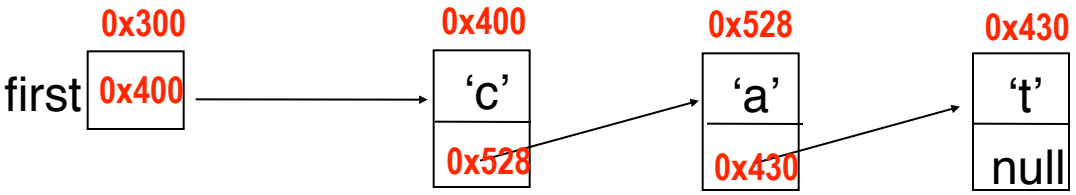
- An empty string will be represented by a NULL value.

  StringNode str2 = null;

- We will use helper methods that take the first node of the string as a parameter.
  - We will have length(str1) instead of str1.length()
  - This is necessary so that the methods can handle empty strings.
    - ✓ if str1 == NULL, length(str1) will work, but str1.length() will produce a runtime error

More on References

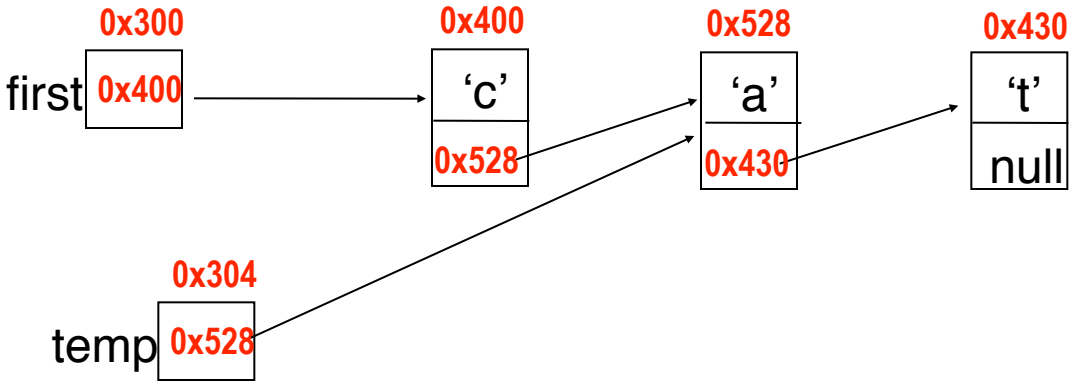
- A reference is also a variable
  - that has its location in the memory, and
  - whose value is the address (i.e., location) of data



  - e.g., first: address=0x300, value=0x400
  - How about first.next.next ?
  - How about first.next.ch ?

More on References

- Example: temp.next.ch



  - Start with the start of the expression: temp.next. It represents the next field of the node to which temp refers.
    - ✓ address = ?
    - ✓ value = ?
  - Next, consider temp.next.ch

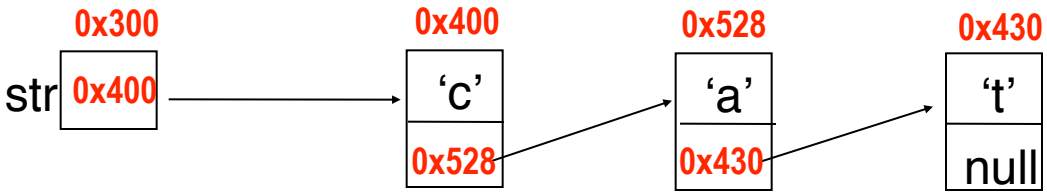
Recursion on Linked Lists

- Recursive definition of a linked list: a linked list is either
  - empty or
  - a single node, followed by a linked list

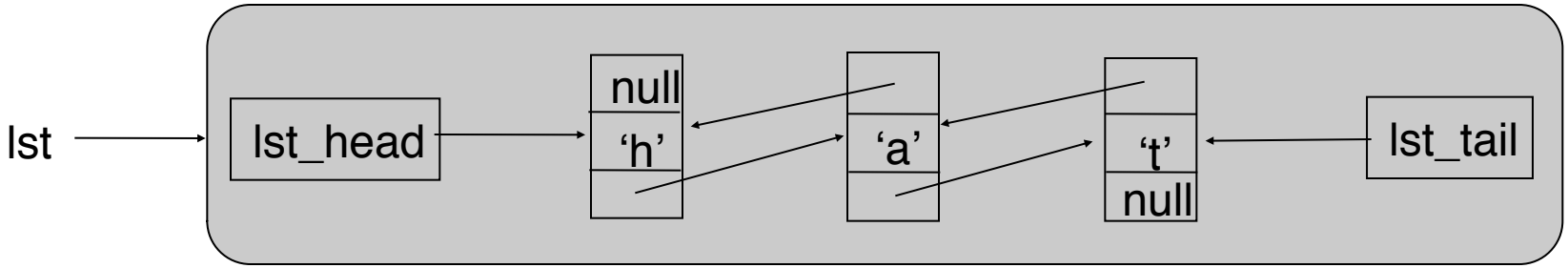
- Recursive definition lends itself to recursive methods.

- Example: length of a string
  - length of “cat” = 1 + the length of “at”
  - length of “at” = 1 + the length of “t”
  - length of “t” = 1 + the length of the empty string (which == 0)

```
private static int length(StringNode str) {
        if (str == null)
            return 0;
        else
            return 1 + length(str.next);
}
```



Doubly Linked List



- Both next and prev are defined in StringNode
- Why needed?

```
public class LLString {
        private StringNode lst_head;
        private StringNode lst_tail;
        private int theSize;
        …
```

```
public class StringNode {
        private char ch;
        private StringNode next;
        private StringNode prev;
        …
}
```

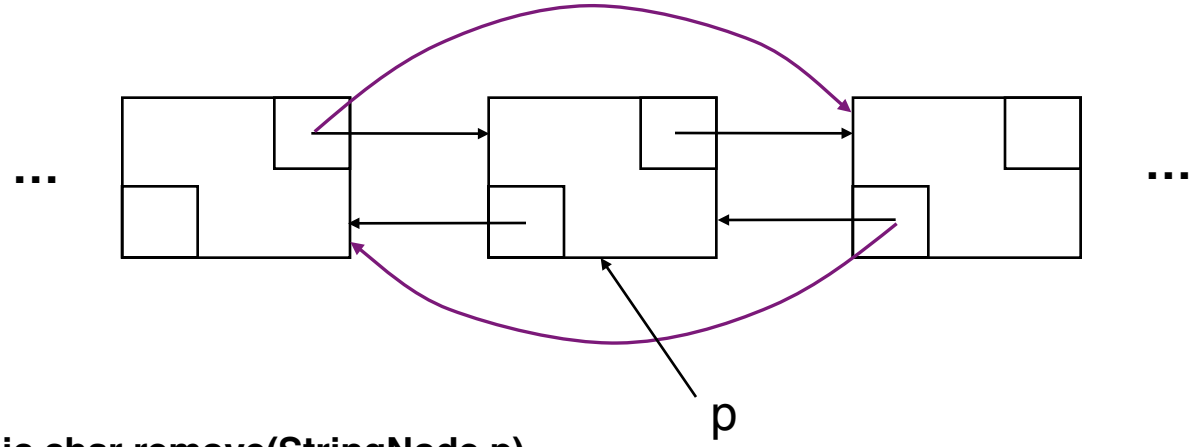
Example: Traversing Linked List

- Access the node at position *i* in a doubly linked list

```
public StringNode getNode(int i) {
        If (i < 0 ll i >= theSize) throw an exception
        StringNode ptr;
        If (i < theSize/2) {
                ptr = lst_head;
                for (j = 0; j != i; j++) ptr = ptr.next;
        } else {
                ptr = lst_tail;
                for (j = theSize-1; j != i; j--) ptr = ptr.prev;
        }
        return ptr;
}
```

What is the running time?

Example: Removing a Node




```
public char remove(StringNode p)
{
        if (p == lst_head ll p == lst_tail)

                        ?


        p.next.prev = p.prev;
        p.prev.next = p.next;
        theSize--;

        return p.ch;
   }
```

Do we need to explicitly de-allocate p?

Example: Inserting a Node

- Insert a new node before p.

  newNode.prev = p.prev;

  newNode.next = p;

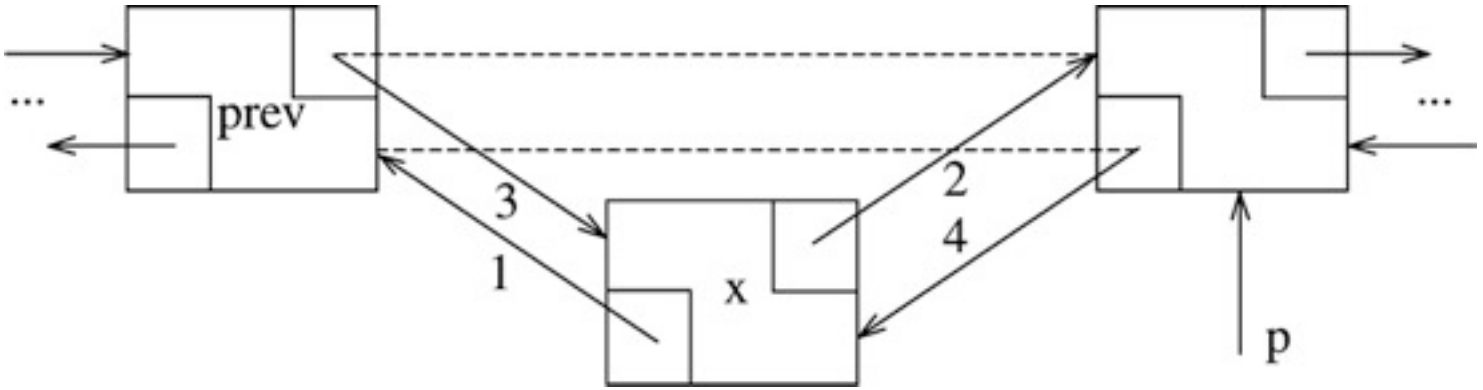
  p.prev.next = newNode;

  p.prev = newNode;

  What if p is the first element? Last element?

  What if p == null?

  What if p is not part of a list?



Other Operations

Either simple linked list or doubly linked list

- Count the occurrences of an item in the linked list
- Remove all occurrences of an item
- Reverse a linked list (trivial for doubly linked list)
- Duplicate a linked list

Problem-of-the-Week

- Assume an arbitrary node p in a link list whose head is head. Write a method to move back p by one position towards the tail (if possible) by modifying the links only. You should consider both:
  - a singly linked list, and
  - a doubly linked list.

```
public void moveback(MyNode p)
{
        …
}
```

- What if there is only one node in the list? What if p is the last node? What if p is an isolated node?