## **Data Structures**

Abstract Data Types for Linear Lists

November 12, 2016

#### Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- 3 Linear List Array Representation
  - The ADT for the list
  - Simplicity After All
- Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



## Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- Stinear List Array Representation
  - The ADT for the list
  - Simplicity After All
- Dynamic Array:
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



#### What is a data structure?

#### Intuition

A data object is a set or collection of instances!!!

## What is a data structure?

#### Intuition

A data object is a set or collection of instances!!!

## Examples

• integer =  $\{0, +1, -1, +2, -2, +3, -3, ...\}$ 

## What is a data structure?

#### Intuition

A data object is a set or collection of instances!!!

#### **Examples**

- integer =  $\{0, +1, -1, +2, -2, +3, -3, ...\}$
- daysOfWeek =  $\{S,M,T,W,Th,F,Sa\}$

# Those instance may be related!!!

#### Something quite basic

Instances may or may not be related.

Example

 $myDataObject = \{apple,\, chair,\, 2,\, 5.2,\, red,\, green,\, Jack$ 

Thi

Data Structure pprox Data object + relationships that exist among instances and elements that comprise an instance.

# Those instance may be related!!!

#### Something quite basic

Instances may or may not be related.

#### Example

 $myDataObject = \{apple, \, chair, \, 2, \, 5.2, \, red, \, green, \, Jack\}$ 

#### TI

Data Structure pprox Data object + relationships that exist among instances and elements that comprise an instance.

# Those instance may be related!!!

#### Something quite basic

Instances may or may not be related.

#### Example

myDataObject = {apple, chair, 2, 5.2, red, green, Jack}

#### Thus

Data Structure  $\approx$  Data object + relationships that exist among instances and elements that comprise an instance.

## Among instances of integers

- 369 < 370
- 280 + 4 = 284

#### Thus

The relationships are usually specified by specifying operations on one or

#### Evample

add, subtract, predecessor, multiply



## Among instances of integers

- 369 < 370
- $\bullet$  280 + 4 = 284

#### Thus

The relationships are usually specified by specifying operations on one or more instances.

add, subtract, predecessor, multiply

## Among instances of integers

- 369 < 370
- $\bullet$  280 + 4 = 284

#### Thus

The relationships are usually specified by specifying operations on one or more instances.

## Examples

add, subtract, predecessor, multiply

## Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- - Basic Definition
  - Operations
  - Memory Implementation
- - The ADT for the list
  - Simplicity After All
- - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



# Abstract Data Type

#### Data Type

A data type such as **int** or **double** is a group of values and operations on those values that is defined within a specific programming language.

An Abstract Data Type, or ADT, is a specification for a group of values and the operations on those values that is defined conceptually and independently of any programming language.

A data structure is an implementation of an ADT within a programming language

# Abstract Data Type

#### Data Type

A data type such as **int** or **double** is a group of values and operations on those values that is defined within a specific programming language.

#### Abstract Data Type

An Abstract Data Type, or ADT, is a specification for a group of values and the operations on those values that is defined conceptually and independently of any programming language.

A data structure is an implementation of an ADT within a programming language.



# Abstract Data Type

#### Data Type

A data type such as **int** or **double** is a group of values and operations on those values that is defined within a specific programming language.

#### Abstract Data Type

An Abstract Data Type, or ADT, is a specification for a group of values and the operations on those values that is defined conceptually and independently of any programming language.

#### Thus

A data structure is an implementation of an ADT within a programming language.



#### Something Notable

An abstract data type is defined indirectly, only by the operations that may be performed on it and by mathematical constraints on the effects (and possibly cost) of those operations.

Q. An abstract object may be passed as a parameter to a procedure

An abstract object may be assigned to a variable.

#### Something Notable

An abstract data type is defined indirectly, only by the operations that may be performed on it and by mathematical constraints on the effects (and possibly cost) of those operations.

An abstract object may be operated upon by the operations which define its abstract type.

#### Something Notable

An abstract data type is defined indirectly, only by the operations that may be performed on it and by mathematical constraints on the effects (and possibly cost) of those operations.

## Using Abstract Data Types

• An abstract object may be operated upon by the operations which define its abstract type.

#### Something Notable

An abstract data type is defined indirectly, only by the operations that may be performed on it and by mathematical constraints on the effects (and possibly cost) of those operations.

## Using Abstract Data Types

- An abstract object may be operated upon by the operations which define its abstract type.
- ② An abstract object may be passed as a parameter to a procedure.

#### Something Notable

An abstract data type is defined indirectly, only by the operations that may be performed on it and by mathematical constraints on the effects (and possibly cost) of those operations.

## Using Abstract Data Types

- An abstract object may be operated upon by the operations which define its abstract type.
- ② An abstract object may be passed as a parameter to a procedure.
- 3 An abstract object may be assigned to a variable.

#### For More

## We have the following

"PROGRAMMING WITH ABSTRACT DATA TYPES" by Barbara Liskov and Stephen Zilles

## Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- 3 Linear List Array Representation
  - The ADT for the list
  - Simplicity After All
- Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



#### Definition

A linear list is a data structure that holds a sequential list of elements.

#### Definition

A linear list is a data structure that holds a sequential list of elements.

#### Instances

They have the following structure,  $(e_0, e_1, e_1, ..., e_{n-1})$ 

#### Definition

A linear list is a data structure that holds a sequential list of elements.

#### Instances

They have the following structure,  $(e_0, e_1, e_1, ..., e_{n-1})$ 

#### **Properties**

• Where  $e_i$  denotes a list element

#### Definition

A linear list is a data structure that holds a sequential list of elements.

#### Instances

They have the following structure,  $(e_0, e_1, e_1, ..., e_{n-1})$ 

## Properties

- Where  $e_i$  denotes a list element
- n > 0 is finite

#### Definition

A linear list is a data structure that holds a sequential list of elements.

#### Instances

They have the following structure,  $(e_0, e_1, e_1, ..., e_{n-1})$ 

## **Properties**

- Where  $e_i$  denotes a list element
- n > 0 is finite
- List size is n

# Having this $L=\left( e_{0},e_{1},e_{2},...,e_{n-1} ight)$

## Having this

$$L = (e_0, e_1, e_2, ..., e_{n-1})$$

## Relationships

- $\bullet$   $e_0$  is the zero'th (or front) element
- e: immediately precedes e:±1

## Having this

$$L = (e_0, e_1, e_2, ..., e_{n-1})$$

## Relationships

- $e_0$  is the zero'th (or front) element
- $e_{n-1}$  is the last element

## Having this

$$L = (e_0, e_1, e_2, ..., e_{n-1})$$

#### Relationships

- $\bullet$   $e_0$  is the zero'th (or front) element
- $e_{n-1}$  is the last element
- $e_i$  immediately precedes  $e_{i+1}$

## Simple ones

• Students in COP3530 = (Jack, Jill, Abe, Henry, Mary, ..., Judy)

## Simple ones

- Students in COP3530 = (Jack, Jill, Abe, Henry, Mary, ..., Judy)
- Exams in COP3530 = (exam1, exam2, exam3)

#### Simple ones

- Students in COP3530 = (Jack, Jill, Abe, Henry, Mary, ..., Judy)
- Exams in COP3530 = (exam1, exam2, exam3)
- ullet Days of Week = (S, M, T, W, Th, F, Sa)

#### Simple ones

- Students in COP3530 = (Jack, Jill, Abe, Henry, Mary, ..., Judy)
- Exams in COP3530 = (exam1, exam2, exam3)
- Days of Week = (S, M, T, W, Th, F, Sa)
- Months = (Jan, Feb, Mar, Apr, ..., Nov, Dec)

### Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- 2 Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- 3 Linear List Array Representation
  - The ADT for the list
  - Simplicity After All
- Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



# Which operations must be supported?

#### We need the size of a linear list

Hey, we need to know how many elements the linear list has.

Determine list size for

L = (a, b, c, d, e)

• size(*L*)=5

# Which operations must be supported?

#### We need the size of a linear list

Hey, we need to know how many elements the linear list has.

### Determine list size for

$$L = (a, b, c, d, e)$$

size(L)=5

# Which operations must be supported?

#### We need the size of a linear list

Hey, we need to know how many elements the linear list has.

#### Determine list size for

$$L = (a, b, c, d, e)$$

### So, we need to have a operation that returns the size

• size(*L*)=5

# Get operations

Get element with given index.

### Get operations

Get element with given index.

### For example

L = (a, b, c, d, e)

# Get operations

Get element with given index.

### For example

$$L = (a, b, c, d, e)$$

- get(0) = a
- get(2) = c
- $\operatorname{gct}(4) = c$
- $o \operatorname{get}(9) = \operatorname{error}$

# Get operations

Get element with given index.

### For example

$$L = (a, b, c, d, e)$$

- get(0) = a
- get(2) = c
- $\operatorname{gct}(4) = c$
- $o \operatorname{get}(9) = \operatorname{error}$

# Get operations

Get element with given index.

### For example

$$L = (a, b, c, d, e)$$

- get(0) = a
- get(2) = c

# Get operations

Get element with given index.

### For example

$$L = (a, b, c, d, e)$$

#### Thus

- get(0) = a
- get(2) = c
- get(4) = e

 $\bullet$  get(-1) = error

 $\circ$  get(9) = error

# Get operations

Get element with given index.

### For example

$$L = (a, b, c, d, e)$$

- get(0) = a
- get(2) = c
- get(4) = e
- get(-1) = error

### Get operations

Get element with given index.

### For example

$$L = (a, b, c, d, e)$$

- get(0) = a
- get(2) = c
- get(4) = e
- get(-1) = error
- get(9) = error

### IndexOf operations

Determine the index of an element.

### IndexOf operations

Determine the index of an element.

### Example

L = (a, b, d, b, a)

### IndexOf operations

Determine the index of an element.

### Example

$$L = (a, b, d, b, a)$$

- indexOf(d) = 2

### IndexOf operations

Determine the index of an element.

### Example

$$L = (a, b, d, b, a)$$

- indexOf(d) = 2

### IndexOf operations

Determine the index of an element.

### Example

$$L = (a, b, d, b, a)$$

- indexOf(d) = 2
- indexOf(a) = 0

### IndexOf operations

Determine the index of an element.

#### Example

$$L = (a, b, d, b, a)$$

- indexOf(d) = 2
- indexOf(a) = 0
- indexOf(z) = -1

# Linear List Operations: add(theIndex, theElement)

#### Definition

Add an element so that the new element has a specified index.

If we have

L = (a, b, c, d, e, f, g)

19 / 43

# Linear List Operations: add(theIndex, theElement)

#### Definition

Add an element so that the new element has a specified index.

#### If we have

$$L = (a, b, c, d, e, f, g)$$

 $\bullet \ \operatorname{add}(\mathbf{0,h}) \Longrightarrow L = (h,a,b,c,d,e,f,g)$ 

# Linear List Operations: add(theIndex, theElement)

#### Definition

Add an element so that the new element has a specified index.

#### If we have

$$L = (a, b, c, d, e, f, g)$$

- $\bullet \ \, \operatorname{add}(\mathbf{0,h}) \Longrightarrow L = (h,a,b,c,d,e,f,g)$ 
  - ▶ Thus, index of a, b, c, d, e, f, g increases by 1.

# add(2,h) using the original L

$$\bullet \Longrightarrow L = (a, b, h, c, d, e, f, g)$$

### $\mathsf{add}(\mathsf{2},\mathsf{h})$ using the original L

- $\bullet \implies L = (a, b, h, c, d, e, f, g)$ 
  - index of c, d, e, f, g increases by 1

### add(2,h) using the original L

- $\bullet \implies L = (a, b, h, c, d, e, f, g)$ 
  - lacktriangle index of c,d,e,f,g increases by 1

#### What about out of the bound

- add(10,h)  $\Longrightarrow$  error
- a add(6 b) -- arror

### $oxed{add(2,h)}$ using the original L

- $\bullet \implies L = (a, b, h, c, d, e, f, g)$ 
  - $\blacktriangleright$  index of c, d, e, f, g increases by 1

#### What about out of the bound

- $add(10,h) \Longrightarrow error$
- add(-6,h)  $\Longrightarrow$  error

#### Definition

Remove and return element with given index.

#### Definition

Remove and return element with given index.

### For example

L = (a, b, c, d, e, f, g)

### Definition

Remove and return element with given index.

### For example

L = (a, b, c, d, e, f, g)

### Example

 $\bullet$   $\mbox{\bf Remove(2)}$  returns c and L becomes (a,b,d,e,f,g)

#### Definition

Remove and return element with given index.

### For example

L = (a, b, c, d, e, f, g)

#### Example

- ullet Remove(2) returns c and L becomes (a,b,d,e,f,g)
  - ▶ Index of d, e, f, g decreases by 1

### What about the Error

### For

$$L = (a, b, c, d, e, f, g)$$

### What about the Error

### For

$$L = (a, b, c, d, e, f, g)$$

- remove $(-1) \Longrightarrow error$
- remove(20)  $\Longrightarrow$  error

# The Final Abstract Data Type for Linear List

```
Linear List
AbstractDataType LinearList
  instances
     ordered finite collections of zero or more elements
  operations
     isEmpty(): return true iff the list is empty, false otherwise
     size(): return the list size (i.e., number of elements in the list)
     get(index): return the element with "index" index
     indexOf(x): return the index of the first occurrence of x in the list, return -1
                   if x is not in the list
     remove(index): remove and return the indexth element, elements with higher
                      index have their index reduced by 1
     add(theIndex, x): insert x as the index of th element, elements with
                      theIndex > index have their index increased by 1
     output(): output the list elements from left to right
```

### Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- 2 Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- 3 Linear List Array Representation
  - The ADT for the list
  - Simplicity After All
- 4 Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



# **CPython**

### We have

- typedef struct {
- 2 PyObject\_VAR\_HEAD;
- PyObject \*\*ob\_item;
- Py\_ssize\_t allocated;

### Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
  - Linear List Array Representation
    - The ADT for the list
    - Simplicity After All
- 4 Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



# How do we implement the ADT List?

### In our first representation will use an array

Use a one-dimensional array **element**[]:

A representation of L = (a, b, c, d, e) using position i in **element**[i]



## How do we implement the ADT List?

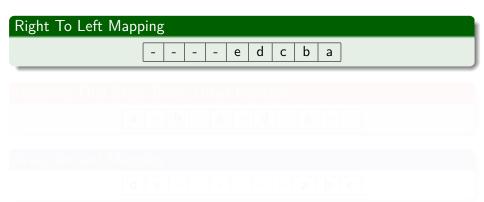
#### In our first representation will use an array

Use a one-dimensional array **element**[]:

#### The previous array

A representation of L = (a, b, c, d, e) using position i in **element**[i].

# Where to map in the array



# Where to map in the array



# Where to map in the array



### Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- 2 Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- 3 Linear List Array Representation
  - The ADT for the list
  - Simplicity After All
- Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



## Representation Used



### Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- Stinear List Array Representation
  - The ADT for the list
  - Simplicity After All
- Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



## Now, we have a common problem

#### Because

We do not know how many elements will be stored at the list.

We use

An initial length and dynamically increase the size as needed

## Now, we have a common problem

#### Because

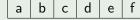
We do not know how many elements will be stored at the list.

### We use

An initial length and dynamically increase the size as needed.

### Example

Length of array element[] is 6:



First create a new and larger array

NewArray = (Item[]) new Object[12]

#### Now copy the

 $System.arraycopy (element, \ 0, \ new Array, \ 0, \ element.length);$ 

```
a b c d e f - - - - -
```

#### Example

Length of array element[] is 6:

a b c d e f

### First create a new and larger array

 ${\sf NewArray} = (\mathsf{Item}[]) \ \mathsf{new} \ \mathsf{Object}[12]$ 



System.arraycopy(element, 0, newArray, 0, element.length);;

a b c d e f - - - - - -

#### Example

Length of array element[] is 6:

a b c d e f

### First create a new and larger array

NewArray = (Item[]) new Object[12]



### Now copy the new elements into the new array!!!

System.arraycopy(element, 0, newArray, 0, element.length);

a b c d e f - - - - - -

### Finally, rename new array

element = NewArray;

### Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- 3 Linear List Array Representation
  - The ADT for the list
  - Simplicity After All
- Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



## What if you use the following policy?

At least 1 more than current array length.

### What if you use the following policy?

At least 1 more than current array length.

#### Thus

- Generate a new array.
- Copy all the items to it
  - insert the new element at the end

### What if you use the following policy?

At least 1 more than current array length.

#### Thus

- Generate a new array.
- Conv all the items to it
  - insert the new element at the end.

### What if you use the following policy?

At least 1 more than current array length.

#### Thus

- Generate a new array.
- Copy all the items to it.

#### What if you use the following policy?

At least 1 more than current array length.

#### Thus

- Generate a new array.
- Copy all the items to it.
- insert the new element at the end.

### We finish with something like this

- **1** First Insertion: Creation of the List  $\Rightarrow$  Cost = 1.
- Second Insertion Cost = 2.
- Third Insertion Cost = 3
- etc!!!

#### We finish with something like this

- **1** First Insertion: Creation of the List  $\Rightarrow$  Cost = 1.
- ② Second Insertion Cost = 2.

DataLab

Data Science Community

#### We finish with something like this

- **1** First Insertion: Creation of the List  $\Rightarrow$  Cost = 1.
- **2** Second Insertion Cost = 2.
- $\bullet$  Third Insertion Cost = 3

#### We finish with something like this

- First Insertion: Creation of the List  $\Rightarrow$  Cost = 1.
- $\bigcirc$  Second Insertion Cost = 2.
- etc!!!

#### We finish with something like this

- **1** First Insertion: Creation of the List  $\Rightarrow$  Cost = 1.
- **2** Second Insertion Cost = 2.
- **3** Third Insertion Cost = 3
- etc!!!

#### Thus, we have

$$1 + 2 + 3 + \dots + n = \frac{n(n+1)}{2} = O(n^2)$$
 (1)

Not a good idea!!!

### We finish with something like this

- **1** First Insertion: Creation of the List  $\Rightarrow$  Cost = 1.
- **2** Second Insertion Cost = 2.
- **3** Third Insertion Cost = 3
- etc!!!

#### Thus, we have

$$1 + 2 + 3 + \dots + n = \frac{n(n+1)}{2} = O(n^2)$$
 (1)

#### Ok

Not a good idea!!!

### Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- 3 Linear List Array Representation
  - The ADT for the list
  - Simplicity After All
- Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



## A better strategy

#### Dynamic Array

To avoid incurring the cost of re-sizing many times, dynamic arrays re-size by an amount a.

## A better strategy

#### Dynamic Array

To avoid incurring the cost of re-sizing many times, dynamic arrays re-size by an amount a.

```
In our example we double the size, a=2

Item NewArray[];

if (this.size == element.lenght){

    // Resize the capacity

    NewArray = (Item[]) new Object[2*this.size]

    for(int i=0; i < size; i++){

        NewArray[i]=element[i];

    }

    element = NewArray;
}
```

## **Space Complexity**

### Every time an insertion triggers a doubling of the array

Thus, space wasted in the new array:

Space Wasted 
$$=$$
 Old Lenght  $-1$  (2)

Remember: We double the array and insert!!!

## Space Complexity

### Every time an insertion triggers a doubling of the array

Thus, space wasted in the new array:

Space Wasted = Old Lenght 
$$-1$$
 (2)

Remember: We double the array and insert!!!

## Thus, the average space wasted is

 $\Theta\left(n\right)$ 

# For example, we have the following

#### A trade-off between time and space

- You have and average time for insertion is  $\frac{2}{2-1}$ .
- In addition, an upper bound for the wasted cells in the array is  $(2-1)\,n-1=n-1.$

#### Thus, we have:

- Average time for insertion is  $\frac{a}{a-1}$ 
  - An upper bound for the wasted cells in the array is
  - (a-1) n 1 = an n 1.

- Java,  $a = \frac{3}{5}$ 
  - Python,  $a = \frac{9}{5}$

# For example, we have the following

#### A trade-off between time and space

- You have and average time for insertion is  $\frac{2}{2-1}$ .
- In addition, an upper bound for the wasted cells in the array is  $(2-1)\,n-1=n-1.$

### Actually a more general term of expansion, $\boldsymbol{a}$

#### Thus, we have:

- Average time for insertion is  $\frac{a}{a-1}$ .
- An upper bound for the wasted cells in the array is (a-1) n 1 = an n 1.

#### • Java, $a = \frac{3}{2}$

• Python,  $a=\frac{9}{5}$ 

# For example, we have the following

#### A trade-off between time and space

- You have and average time for insertion is  $\frac{2}{2-1}$ .
- In addition, an upper bound for the wasted cells in the array is  $(2-1)\,n-1=n-1.$

### Actually a more general term of expansion, a

#### Thus, we have:

- Average time for insertion is  $\frac{a}{a-1}$ .
- An upper bound for the wasted cells in the array is (a-1) n 1 = an n 1.

#### Different languages use different values

- Java,  $a = \frac{3}{2}$ .
- Python,  $a = \frac{9}{8}$

### Outline

- Introduction
  - Basic Ideas
  - Abstract Data Types
- Linear List: An Example of ADT
  - Basic Definition
  - Operations
  - Memory Implementation
- 3 Linear List Array Representation
  - The ADT for the list
  - Simplicity After All
- Dynamic Arrays
  - The Common Problem
  - How much we need to expand it...
  - A Strategy for it
  - Analysis



# We have the following final analysis

## Amortized Analysis Vs. Classic

	Array Classic Analysis	Dynamic Array
		Amortized Analysis
Indexing	O(1)	O(1)
Search	$O\left(n\right)$	$O\left(n\right)$
Add/Remove	$O\left(n\right)$	$O\left(n\right)$
Space Complexity	$O\left(n\right)$	$O\left(n\right)$