Chapter 12 Exception Handling and Text I/O



Motivations

When a program runs into a runtime error, the program terminates abnormally.

How can you handle the runtime error so that the program can continue to run or terminate gracefully?

This is the subject we will introduce in this chapter.

```
Quotient.java
   import java.util.Scanner;
  public class Quotient {
     public static void main(String[] args) {
5 6 7 8 9
       Scanner input = new Scanner(System.in);
       // Prompt the user to enter two integers
       System.out.print("Enter two integers: ");
       int number1 = input.nextInt();
10
       int number2 = input.nextInt();
11
12
       System.out.println(number1 + " / " + number2 + " is " +
13
         (number1 / number2));
14
15 }
```

Abort, Crash! Show runtime-error

```
Enter two integers: 3 0 Finter

Exception in thread "main" java.lang.ArithmeticException: / by zero

at Quotient.main(Quotient.java:11)
```

Fix runtime-error using an if statement

Use exception-handling statement

```
12
       try {
                                                                             try block
13
         if (number2 == 0)
14
           throw new ArithmeticException("Divisor cannot be zero");
15
16
         System.out.println(number1 + " / " + number2 + " is " +
17
           (number1 / number2));
18
       catch (ArithmeticException ex) {
19
                                                                             catch block
20
       ➤System.out.println("Exception: an integer " +
21
           "cannot be divided by zero ");
22
23
       System.out.println("Execution continues ...");
24
25
26 }
 Enter two integers: 5 0
 Exception: an integer cannot be divided by zero
 Execution continues ...
```

An exception:

new ArithmeticException("Divisor cannot be zero");

- an object of exception class java.lang.ArithmeticException
- Exception message: describe the exception

try block:

- Throw exception
- executed in normal circumstances.

catch block:

- Handle exception
- executed when **exception occurs**.
- Afterward, line24 is executed



In summary, a template for a try-throw-catch block may look like this:

```
try {
   Code to try;
   Throw an exception with a throw statement or
     from method if necessary;
   More code to try;
}
catch (type ex) {
   Code to process the exception;
}
```

An exception may be thrown directly by using a throw statement in a try block, or by invoking a method that may throw an exception.

QuotientWithMethod.java

```
1 import java.util.Scanner;
   public class QuotientWithMethod {
     public static int quotient(int number1, int number2) {
       if (number2 == 0)
         throw new ArithmeticException("Divisor cannot be zero");
8
       return number1 / number2;
10
11
     public static void main(String[] args) {
12
       Scanner input = new Scanner(System.in);
13
14
       // Prompt the user to enter two integers
15
       System.out.print("Enter two integers: ");
16
       int number1 = input.nextInt();
       int number2 = input.nextInt();
17
18
19
       trv {
        - int result = quotient(number1, number2);
21 Arithmetic System.out.println(number1 + " / " + number2 + " is "
  Exception
           + result):
23
24
       catch (ArithmeticException ex) {
25
       ➤ System.out.println("Exception: an integer " +
26
            "cannot be divided by zero "):
27
28
29
       System.out.println("Execution continues ...");
30
31 }
```

```
Enter two integers: 5 0 | LETTER |
Exception: an integer cannot be divided by zero |
Execution continues ...
```

- Callee: quotient(), throw exception
- Caller: main(), handle exception



Exception Advantages

Typically, a callee (e.g. library method) can detect the error, but only caller knows how to handle the error.

Advantages of exception handling: separate the detection and handling of error; enable a callee to throw an exception to its caller

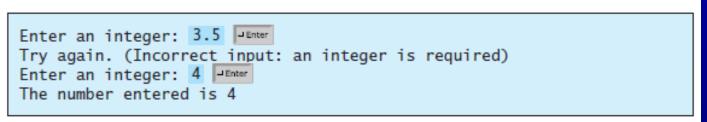
- Callee: (detect error) throw exception
- Caller: (handle error) handle exception
- Without this capability, a callee must handle the exception by itself or terminate the program.

InputMismatchExceptionDemo.java

```
1 import java.util.*;
   public class InputMismatchExceptionDemo {
     public static void main(String[] args) {
       Scanner input = new Scanner(System.in);
       boolean continueInput = true;
       do {
         try {
           System.out.print("Enter an integer: ");
           int number = input.nextInt();
   If an InputM1
           // Display the result
14
           System.out.println(
15
              "The number entered is " + number);
16
17
            continueInput = false;
18
         catch (InputMismatchException ex) {
19
         ➤ System.out.println("Try again. (" +
20
21
              "Incorrect input: an integer is required)");
22
            input.nextLine(); // Discard input
23
24
       } while (continueInput);
25
26 }
```

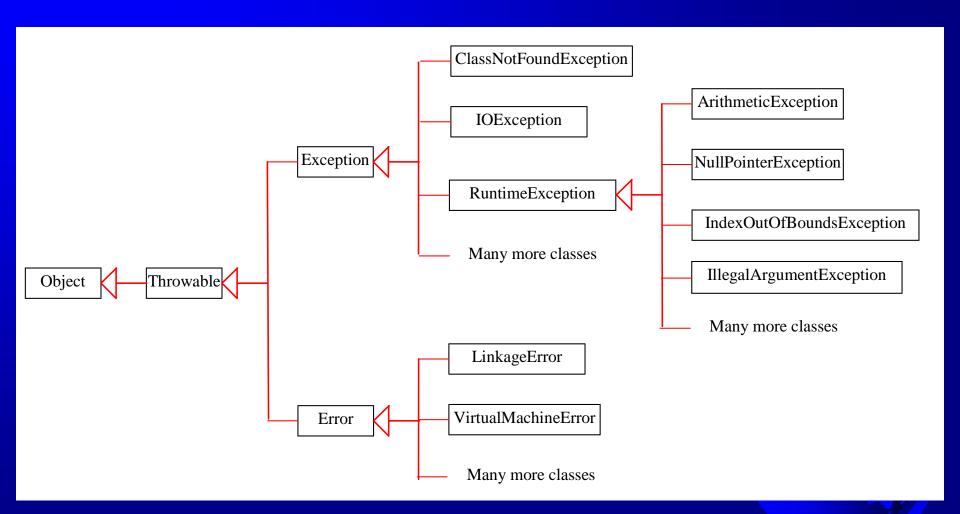
By <u>handling InputMismatchException</u>, program <u>continuously read</u> an input until it is correct.

The statement **input.nextLine**() discards the current input line so that the user can enter a new line of input.

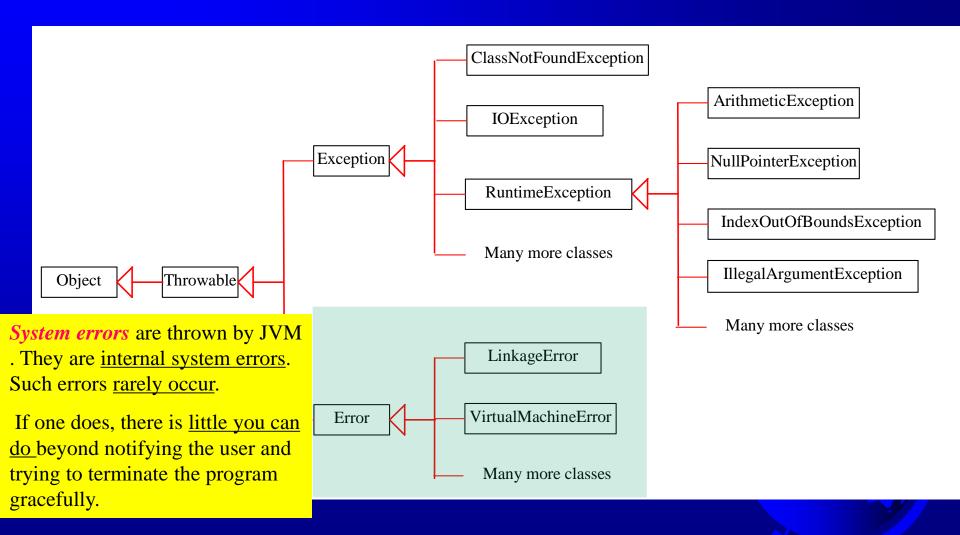




Exception Types



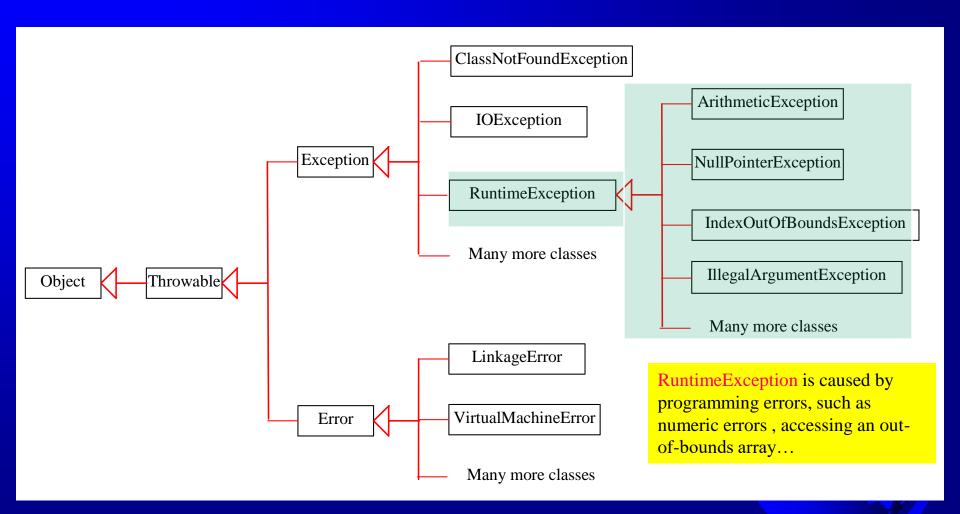
System Errors



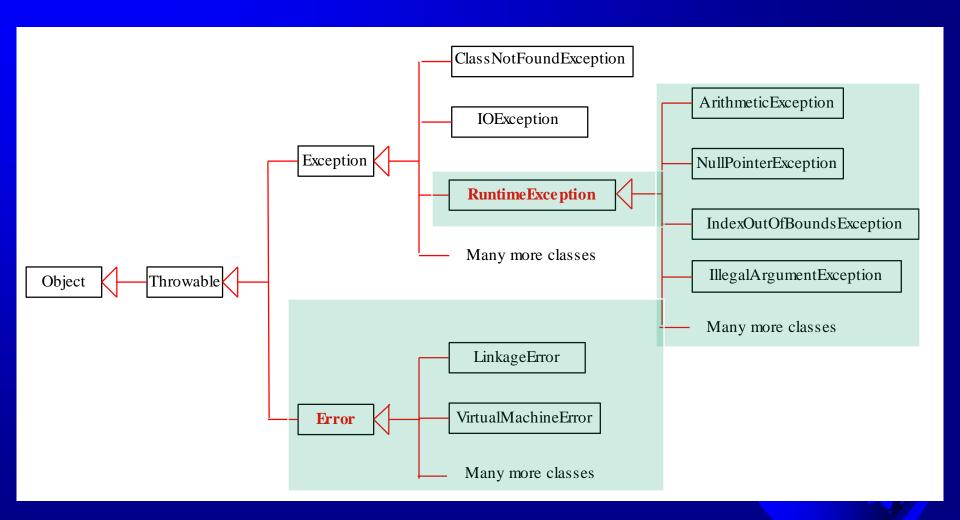
Exceptions

Exception describes errors ClassNotFoundException caused by your program and external circumstances. ArithmeticException These errors can be caught **IOException** and handled by your Exception NullPointerException program. RuntimeException IndexOutOfBoundsException Many more classes IllegalArgumentException Throwable Object Many more classes LinkageError VirtualMachineError Error Many more classes

Runtime Exceptions



Unchecked Exceptions



Unchecked Exceptions

Unchecked exceptions:

RuntimeException, Error and their subclasses

- No need to declare such exceptions
- Un-recoverable

- In most cases, they are <u>logic errors</u>.
 - e.g. an IndexOutOfBoundsException is thrown if you access an element in an array outside the array bounds.
 - logic errors should be corrected in the program.
- They can occur anywhere in the program. To avoid overuse of try-catch blocks, Java does not mandate you to declare/catch them.

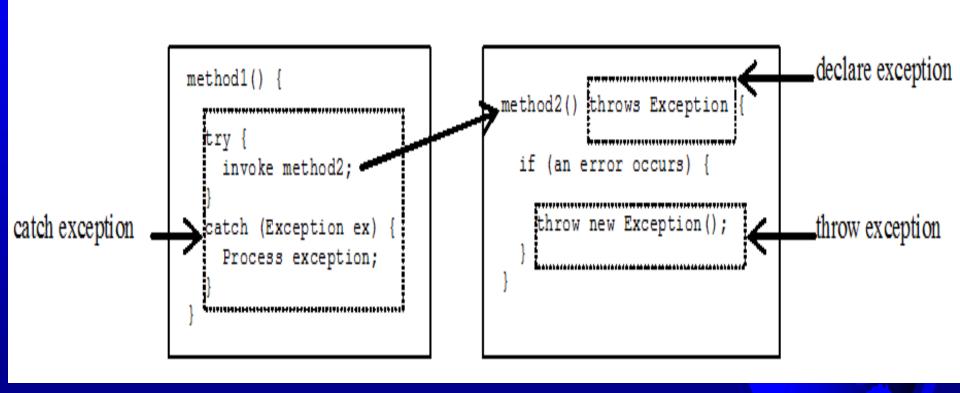
Checked Exceptions

Checked exceptions: all other exceptions

Must be explicitly declared in the method header

- callee: declare and throw the exceptions.
- caller: catch and handle the exceptions.

Declare, Throw, and Catch Exceptions



Declare Exceptions

A method must declare the <u>types/classes of exceptions</u> it might throw.

```
public void myMethod() throws IOException
{...}
```

```
//throw multiple exceptions
public void myMethod()
    throws IOException, OtherException
{...}
```



Throw Exceptions

When the program detects an error, the program can create an instance of an appropriate exception class/type and throw it.

throw new TheException();

TheException ex = new TheException(); throw ex;



Example

```
/** Set a new radius */
public void setRadius(double newRadius)
    throws IllegalArgumentException {
    if (newRadius >= 0)
        radius = newRadius;
    else
        throw new IllegalArgumentException(
        "Radius cannot be negative");
}
```

Catch Exceptions

```
catch (RuntimeException ex)
catch (Exception ex) {
        (b) Correct order
```

A catch block of a

subclass should
appear before that

of a superclass

Incorrect order will cause compile error

Catch Exceptions

```
method1
main method
                                                                                                     An exception
                                                                   method2
                                                                                                     is thrown in
                                                                                                     method3
                                   try
  try {
                                                                     try
    invoke method1
                                     invoke method2
                                                                       invoke method3
    statement1;
                                     statement3;
                                                                       statement5;
  catch (Exception1 ex1)
                                   catch (Exception2 ex2)
                                                                     catch (Exception3 ex3)
   Process ex1:
                                     Process ex2:
                                                                       Process ex3:
  statement2:
                                   statement4:
                                                                     statement6;
```

- If the exception type is Exception3, it is caught by the catch block for handling exception ex3 in method2. statement5 is skipped, and statement6 is executed.
- If the exception type is Exception2, method2 is aborted, the control is returned to method1, and the exception is caught by the catch block for handling exception ex2 in method1. statement3 is skipped, and statement4 is executed.
- If the exception type is Exception1, method1 is aborted, the control is returned to the main method, and the exception is caught by the catch block for handling exception exl in the main method. statement1 is skipped, and statement2 is executed.
- If the exception type is not caught in method2, method1, and main, the program terminates. statement1 and statement2 are not executed.

Checked Exceptions

If a method (e.g., p2()), declares a checked exception (e.g., <u>IOException</u>), When calling p2(), you must <u>call</u> it in a <u>try-catch</u> block or <u>declare</u> to <u>throw</u> the exception in the calling method.

e.g. Suppose that method p2() may throw an IOException,

```
void p1() {
   try {
      p2();
   }
   catch (IOException ex) {
      ...
   }
}
```

```
void p1() throws IOException {
   p2();
}
```

(b)

Example: Declare, Throw, and Catch Exceptions

<u>setRadius</u> method throws an exception if radius is negative.

```
public void setRadius(double newRadius)
throws IllegalArgumentException {
  if (newRadius >= 0)
    radius = newRadius;
  else
    throw new IllegalArgumentException(
    "Radius cannot be negative");
}
```

public CircleWithException(double newRadius) {
 setRadius(newRadius);

```
try {
    CircleWithException c1 = new CircleWithException(5);
    CircleWithException c2 = new CircleWithException(-5);
    CircleWithException c3 = new CircleWithException(0);
}

catch (IllegalArgumentException ex) {
    System.out.println(ex);
}

System.out.println("Number of objects created: " +
    CircleWithException.getNumberOfObjects());
}
```

```
java.lang.IllegalArgumentException: Radius cannot be negative
Number of objects created: 1
```



Rethrowing Exceptions

```
try {
   statements;
}
catch(TheException ex) {
   perform operations before exits;
   throw ex;
}
```

The finally Clause

```
try {
   statements;
}
catch(TheException ex) {
   handling ex;
}
finally {
   finalStatements;
}
```

Finally block is executed under all circumstances, regardless of whether an exception occurs in the try block or is caught

Suppose no exceptions in the statements

```
try {
  statements;
catch(TheException ex) {
  handling ex;
finally {
  finalStatements;
Next statement;
```



```
try {
  statements;
catch (The Exception ex)
  handling ex;
finally {
  finalStatements;
Next statement;
```

The final block is always executed



```
try {
  statements;
catch(TheException ex) {
  handling ex;
finally {
  finalStatements;
    statement;
Next
```

Next statement in the method is executed



```
try {
  statement1;
  statement2;
  statement3;
catch(Exception1 ex) {
  handling ex;
finally {
  finalStatements;
Next statement;
```

Suppose an exception of type Exception1 is thrown in statement2



```
try {
  statement1;
  statement2;
  statement3;
catch (Exception1 ex)
  handling ex;
finally {
  finalStatements;
Next statement;
```

The exception is handled.



```
try {
  statement1;
  statement2;
  statement3;
catch(Exception1 ex) {
  handling ex;
finally {
  finalStatements;
Next statement;
```

The final block is always executed.



```
try {
  statement1;
  statement2;
  statement3;
catch(Exception1 ex) {
  handling ex;
finally {
  finalStatements;
Next statement;
```

The next statement in the method is now executed.



```
try {
  statement1;
  statement2;
  statement3;
catch(Exception1 ex) {
  handling ex;
catch(Exception2 ex) {
  handling ex;
  throw ex;
finally {
  finalStatements;
Next statement;
```

statement2 throws an exception of type Exception2.



```
try {
  statement1;
  statement2;
  statement3;
catch(Exception1 ex) {
  handling ex;
catch(Exception2 ex)
  handling ex;
  throw ex;
finally {
  finalStatements;
Next statement;
```

Handling exception



Trace a Program Execution

```
try {
  statement1;
  statement2;
  statement3;
catch(Exception1 ex) {
  handling ex;
catch(Exception2 ex) {
  handling ex;
  throw ex;
finally {
  finalStatements;
Next statement;
```

Execute the final block



Trace a Program Execution

```
try {
  statement1;
  statement2;
  statement3;
catch(Exception1 ex) {
  handling ex;
catch(Exception2 ex) {
  handling ex;
  throw ex;
finally {
  finalStatements;
Next statement;
```

Rethrow the exception and control is transferred to the caller



Cautions When Using Exceptions

- Exception handling separates error-handling code from normal programming tasks, thus making programs easier to read and to modify.
- However, exception handling usually requires more time and resources
 - because it requires instantiating a new exception object, propagating the errors to the calling methods, and rolling back the call stack.

When to Throw Exceptions

- An exception occurs in a method.
 - If a callee want the exception to be processed by its caller, the callee throw an exception to its caller
 - If a callee can handle the exception in the method where it occurs, there is no need to throw it.



When to Use Exceptions

Use try-catch to deal with unexpected error conditions.

Do not use it to deal with simple, expected situations.

```
if (refVar != null)
System.out.println(refVar.toString());
else
System.out.println("refVar is null");
```

No need:

```
try {
    System.out.println(refVar.toString());
}
catch (NullPointerException ex) {
    System.out.println("refVar is null");
}
```



The File Class

The File class is a wrapper class for <u>directory path</u> + file name

- Absolute file name:
 - ◆ for All platforms (Windows, Unix, ...), platform-independent
 - "/book/Welcome.java"
 - Java directory seperator: forward slash (/)
 - ◆ For Windows platform:
 - "c:\book\Welcome.java" or "c:\\book\\Welcome.java"

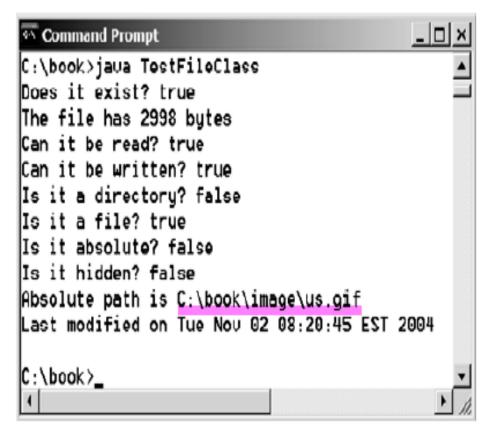
- Relative file name: file name relative to current directory
 - "Welcome.java"
- Obtaining file properties and manipulating file

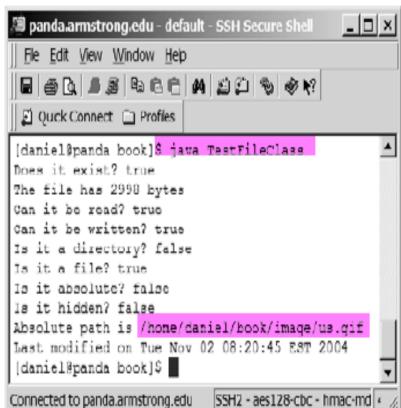
java.io.File	
+File(<mark>pathname</mark> : String)	Creates a File object for the specified path name. The path name may be a directory or a file.
+File(parent: String, child: String)	Creates a File object for the child under the directory parent. The child may be a file name or a subdirectory.
+File(parent: File, child: String)	Creates a File object for the child under the directory parent. The parent is a File object. In the preceding constructor, the parent is a string.
+exists(): boolean	Returns true if the file or the directory represented by the File object exists.
+canRead(): boolean	Returns true if the file represented by the File object exists and can be read.
+canWrite(): boolean	Returns true if the file represented by the File object exists and can be written.
+isDirectory(): boolean	Returns true if the File object represents a directory.
+isFile(): boolean	Returns true if the File object represents a file.
+isAbsolute(): boolean	Returns true if the File object is created using an absolute path name.
+isHidden(): boolean	Returns true if the file represented in the File object is hidden. The exact definition of hidden is system dependent. On Windows, you can mark a file hidden in the File Properties dialog box. On Unix systems, a file is hidden if its name begins with a period character '.'.
+getAbsolutePath(): String	Returns the complete absolute file or directory name represented by the File object.
+getCanonicalPath(): String	Returns the same as getAbsolutePath() except that it removes redundant names, such as "." and "", from the path name, resolves symbolic links (on Unix platforms), and converts drive letters to standard uppercase (on Win32 platforms).
+get <mark>Name</mark> (): String	Returns the last name of the complete directory and file name represented by the File object. For example, new File("c:\\book\\test.dat").getName() returns test.dat.
+get <mark>Path</mark> (): String	Returns the complete directory and file name represented by the File object. For example, new File("c:\\book\\test.dat").getPath() returns c:\book\\test.dat.
+get <mark>Parent</mark> (): String	Returns the complete parent directory of the current directory or the file represented by the File object. For example, new File("c:\\book\\test.dat").getParent() returns c:\book.
+lastModified(): long	Returns the time that the file was last modified.
+length(): long	Returns the size of the file, or 0 if it does not exist or if it is a directory.
+listFile(): File[]	Returns the files under the directory for a directory File object.
+delete(): boolean	Deletes this file. The method returns true if the deletion succeeds.
+ <mark>renameTo</mark> (dest: File): boolean	Renames this file. The method returns true if the operation succeeds.

Create files in a platform-independent way and use the methods in the File class to obtain their properties.

```
TestFileClass.java
   public class TestFileClass {
     public static void main(String[] args) {
       java.io.File file = new java.io.File("image/us.gif");
       System out println("Does it exist?
                                                + file.exists()):
                                                   le.length() + " bytes");
                                           _ | _ | ×
    Command Prompt
                                                    file.canRead());
    C:\book>java TostFiloClass
                                                     + file.canWrite());
    Does it exist? true
    The file has 2998 butes
                                                     + file.isDirectory());
    Can it be read? true
                                                   ile.isFile());
    Can it be written? true
                                                    file.isAbsolute());
10
    Is it a directory? false
                                                   ile.isHidden());
11
    Is it a file? true
    Is it absolute? false
12
    Is it hidden? false
13
    Absolute path is C:\book\image\us.gif
14
    Last modified on Tue Nov 02 08:20:45 EST 2004
15
                                                   )));
    C:\book>_
16
17
                    (a) On Windows
```

The lastModified() method returns the date and time when the file was last modified, measured in milliseconds since the beginning of Unix time (00:00:00 GMT, January 1, 1970). The Date class is used to display it in a readable format in lines 14–15.





(a) On Windows

(b) On Unix

Text I/O (File Input & Output)

Text I/O: reading/writing data from/to a file.

- In order to perform I/O, you need I/O classes.
 - -e.g. Scanner and PrintWriter classes.



Writing Data Using PrintWriter

java.io.PrintWriter

```
+PrintWriter(file: File)
+PrintWriter(filename: String)
+print(s: String): void
+print(c: char): void
+print(cArray: char[]): void
+print(i: int): void
+print(1: long): void
+print(f: float): void
+print(d: double): void
+print(b: boolean): void
Also contains the overloaded
 println methods.
```

Also contains the overloaded printf methods.

Creates a PrintWriter object for the specified file object.

Creates a PrintWriter object for the specified file-name string.

Writes a string to the file.

Writes a character to the file.

Writes an array of characters to the file.

Writes an int value to the file.

Writes a long value to the file.

Writes a float value to the file.

Writes a double value to the file.

Writes a boolean value to the file.

A println method acts like a print method; additionally it prints a line separator. The line-separator string is defined by the system. It is \r\n on Windows and \n on Unix.

The printf method was introduced in Console Output."

"Formatting

The PrintWriter class contains the methods for writing data to a text file.

Example: creates an instance of **PrintWriter** and writes two lines to the file "scores.txt".

```
WriteData.java
```

```
1 public class WriteData {
     public static void main(String[] args) throws Exception {
                                                                             throws an exception
       java.io.File file = new java.io.File("scores.txt");
                                                                             create File object
       if (file.exists()) {
                                                                             file exist?
         System.out.println("File already exists");
         System.exit(0);
       // Create a file
10
       java.io.PrintWriter output = new java.io.PrintWriter(file);
                                                                             create PrintWriter
 11
 12
         // Write formatted output to the file
 13
         output.print("John T Smith ");
         output.println(90); -
 14
                                                       John T Smith 90 scores.txt
 15
         output.print("Eric K Jones ");
                                                       Eric K Jones 85
 16
         output.println(85);
 17
 18
         // Close the file
 19
         output.close();
 20
 21 }
```

The close() method must be used to close the file.

Otherwise, the data may not be saved properly in the file.

Reading Data Using Scanner

java.util.Scanner

```
+Scanner(source: File)
+Scanner(source: String)
+close()
+hasNext(): boolean
+next(): String
+nextLine(): String
+nextByte(): byte
+nextShort(): short
+nextInt(): int
+nextLong(): long
+nextFloat(): float
+nextDouble(): double
+useDelimiter(pattern: String):
  Scanner
```

Creates a scanner that produces values scanned from the specified file.

Creates a scanner that produces values scanned from the specified string.

Closes this scanner.

Returns true if this scanner has more data to be read.

Returns next token as a string from this scanner.

Returns a line ending with the line separator from this scanner.

Returns next token as a byte from this scanner.

Returns next token as a short from this scanner.

Returns next token as an int from this scanner.

Returns next token as a long from this scanner.

Returns next token as a float from this scanner.

Returns next token as a double from this scanner.

Sets this scanner's delimiting pattern and returns this scanner.

FIGURE 9.18 The Scanner class contains the methods for scanning data.

an example that creates an instance of Scanner and reads data from the

ReadData.java

file "scores.txt".

```
1 import java.util.Scanner;
  public class ReadData {
     public static void main(String[] args) throws Exception {
 5
       // Create a File instance
 6
       java.io.File file = new java.io.File("scores.txt");
                                                                                create a File
 7
 8
      // Create a Scanner for the file
 9
      Scanner input = new Scanner(file);
                                                                                create a Scanner
10
      // Read data from a file
11
                                                            scores.txt
12
      while (input.hasNext()) {
                                                                                has next?
                                                            John (T) (Smith) (90)
13
        String firstName = input.next();←
                                                                                read items
                                                            Eric K Jones 85
14
        String mi = input.next(); ←
        String lastName = input.next(); <
15
16
        int score = input.nextInt();
17
        System.out.println(
          firstName + " " + mi + " " + lastName + " " + score);
18
19
      }
20
21
      // Close the file
22
      input.close();
                                                                                close file
23
24 }
```

Note that new Scanner (String) creates a Scanner for a given string. To create a Scanner to read data from a file, you have to use the java.io.File class to create an instance of the File using the constructor new File(filename) (line 6), and use new File(filename) to create a Scanner (File) to create a Scanner for the file (line 9).

Invoking the constructor **new Scanner(File)** may throw an I/O exception. So the main method declares **throws Exception** in line 4.

Each iteration in the while loop reads first name, mi, last name, and score from the text file (lines 12–19). The file is closed in line 22.

It is not necessary to close the input file (line 22), but it is a good practice to do so to release the resources occupied by the file.

File class

throws Exception

close file

51

Both read a string:

- mext() reads a string delimited by whitespaces
- mextLine() reads a line ending with a line separator

input from file:

```
Suppose a text file named test.txt contains a line

34 567

After the following code is executed,

Scanner input = new Scanner(new File("test.txt"));
int intValue = input.nextInt();
String line = input.nextLine();

intValue contains 34 and line contains characters

'', '5', '6', '7'
```

input from keyboard:

What happens if the input is *entered from the keyboard*? Suppose you enter 34, the *Enter* key, 567, and the *Enter* key for the following code:

```
Scanner input = new Scanner(System.in);
int intValue = input.nextInt();
String line = input.nextLine();
```

You will get 34 in intValue and empty string in line.

Problem: Replacing Text

Write a class named ReplaceText that replaces a string in a text file with a new string.

The filename and strings are passed as command-line arguments: java ReplaceText sourceFile targetFile oldString newString

For example, java ReplaceText oldfile.txt newfile.txt black white

replaces all the occurrences of "black" by "white" in oldfile.txt and saves the new file in newfile.txt.

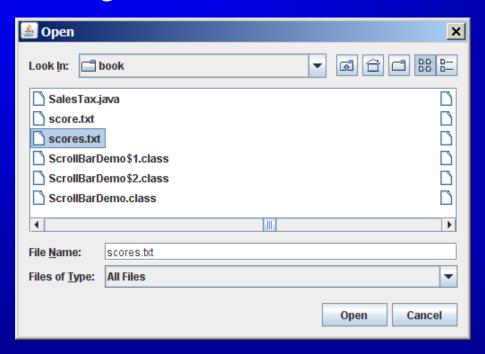
ReplaceText.java

```
1 import java.io.*;
 2 import java.util.*;
 4 public class ReplaceText {
     public static void main(String[] args) throws Exception {
 5
 6
       // Check command-line parameter usage
 7
       if (args.length != 4) {
                                                                                check command usage
 8
         System.out.println(
 9
           "Usage: java ReplaceText sourceFile targetFile oldStr newStr");
         System.exit(0):
10
11
       }
12
13
       // Check if source file exists
       File sourceFile = new File(args[0]);
14
15
       if (!sourceFile.exists()) {
                                                                                source file exists?
16
          System.out.println("Source file " + args[0] + " does not exist");
          System.exit(0);
17
18
       }
19
20
       // Check if target file exists
21
       File targetFile = new File(args[1]);
22
       if (targetFile.exists()) {
                                                                                target file exists?
23
         System.out.println("Target file " + args[1] + " already exists");
24
         System.exit(0);
25
       7
26
27
       // Create a Scanner for input and a PrintWriter for output
28
       Scanner input = new Scanner(sourceFile);
                                                                                create a Scanner
29
       PrintWriter output = new PrintWriter(targetFile);
                                                                                create a PrintWriter
30
31
       while (input.hasNext()) {
                                                                                has next?
32
         String s1 = input.nextLine();
                                                                                read a line
33
         String s2 = s1.replaceAll(args[2], args[3]);
34
         output.println(s2);
35
       }
36
37
       input.close();
                                                                                close file
38
       output.close();
39
40 }
```

(GUI) File Dialogs

javax.swing.JFileChooser class: display a file dialog.

From this dialog box, the user can choose a file.



• Write a program that prompts the user to choose a file and displays its contents on the console.



ReadFileUsingJFileChooser.java

```
The method returns
1 import java.util.Scanner;
2 import javax.swinq.JFileChooser;
  public class ReadFileUsingJFileChooser {
    public static void main(String[] args) throws Exception {
     JFileChooser fileChooser = new JFileChooser();
      if (fileChooser.showOpenDialog(null)
 8
            == JFileChooser.APPROVE OPTION) {
 9
          // Get the selected file
10
          java.io.File file = fileChooser.getSelectedFile();
11
12
          // Create a Scanner for the file
13
          Scanner input = new Scanner(file);
14
15
          // Read text from the file
16
          while (input.hasNext()) {
17
            System.out.println(input.nextLine());
18
19
20
          // Close the file
          input.close();
21
22
23
       else {
24
          System.out.println("No file selected");
25
26
27 }
```

APPROVE_OPTION or CANCEL_OPTION, which indicates whether the Open button or the Cancel button was clicked.

> create a JF11eChooser display file chooser



Read data from the Web

Read from a file (URL) on the Web

first create an instance of the URL class:
import java.net.URL;

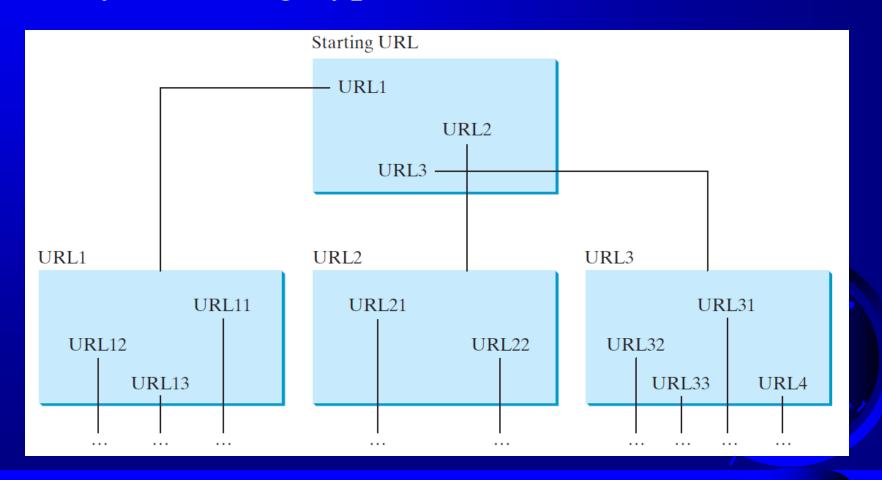
URL url = new URL("http://www.google.com/index.html");

open an input stream of the instance, and read from the stream:

Scanner input = new Scanner(url.openStream());

Case Study: Web Crawler

This case study develops a program that travels the Web by following hyperlinks.



Case Study: Web Crawler

The program follows the URLs to traverse the Web. each URL is traversed only once.

The program maintains two lists of URLs:

<u>listOfPendingURLs</u>: URLs pending for traversing

<u>listOfTraversedURLs</u>: URLs having already been traversed.

Algorithm:



Case Study: Web Crawler

```
Add the starting URL to listOfPendingURLs;
while listOfPendingURLs is not empty {
     Remove a URL from listOfPendingURLs;
     if this URL is not in listOfTraversedURLs {
         Add it to listOfTraversedURLs;
         Display this URL;
         Exit the while loop when the size of S is equal to 100.
         Read the page from this URL;
```

Read the page from this URL;

for <u>each URL</u> contained in the page {
 if it is not in <u>listOfTraversedURLs</u>
 Add it to <u>listOfPendingURLs</u>
 }

