Assignment 1
Due Wednesday 1/24/18, 11:59pm

Task: Install Spark, Set up SBT and Scala practice

Part 1: A show-and-tell program

1. Install Java JDK 7 and Spark 1.0 as in the lecture and in the lecture notes
2. Set up sbt as in the lecture and in the lecture notes.
3. Write a StateDiagram.scala program. Compile and run it using sbt.
4. In today's lecture notes, we have shown you an example of using If-else control structure to produce an automatically generated story line. The story is randomly generated based on the following state diagram.
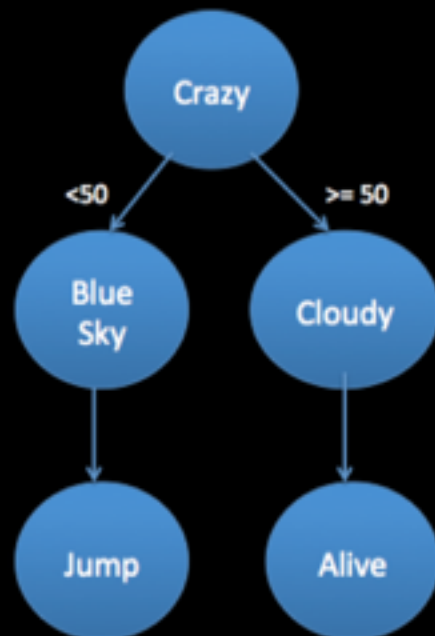
The program would produce a story like:

Each time we run the program, it writes out a different story due to the randomness in the state diagram.

```
cs-ad-d9fyll:sbt gh243$ ./sbt
[info] Set current project to sbt (in build file:/Users/gh243/Desktop/Teaching/cosc282/spark_disk/sbt/)
> run
[info] Compiling 1 Scala source to /Users/gh243/Desktop/Teaching/cosc282/spark_disk/sbt/target/scala-2.10/class
es...
[info] Running HelloWorld
I have a friend
Sometimes she is crazy
When it is cloudy
She is alive
[success] Total time: 2 s, completed Sep 9, 2015 2:08:24 PM
> run
[info] Running HelloWorld
I have a friend
Sometimes she is crazy
When it is cloudy
She is alive
[success] Total time: 0 s, completed Sep 9, 2015 2:08:29 PM
> run
[info] Running HelloWorld
I have a friend
Sometimes she is crazy
When it is cloudy
She is alive
[success] Total time: 0 s, completed Sep 9, 2015 2:08:30 PM
> run
[info] Running HelloWorld
I have a friend
Sometimes she is crazy
When it is blue sky
She jump
[success] Total time: 0 s, completed Sep 9, 2015 2:08:31 PM
> run
[info] Running HelloWorld
I have a friend
Sometimes she is crazy
When it is blue sky
She jump
[success] Total time: 0 s, completed Sep 9, 2015 2:08:33 PM
> run
[info] Running HelloWorld
I have a friend
Sometimes she is crazy
When it is blue sky
She jump
[success] Total time: 0 s, completed Sep 9, 2015 2:08:36 PM
```

5. Your task is to complete the story for a bigger graph, also using control structures. Your program should also be able to generate random stories based on this state diagram.

Note: Your story should be truly random and obey the probability in the diagram.
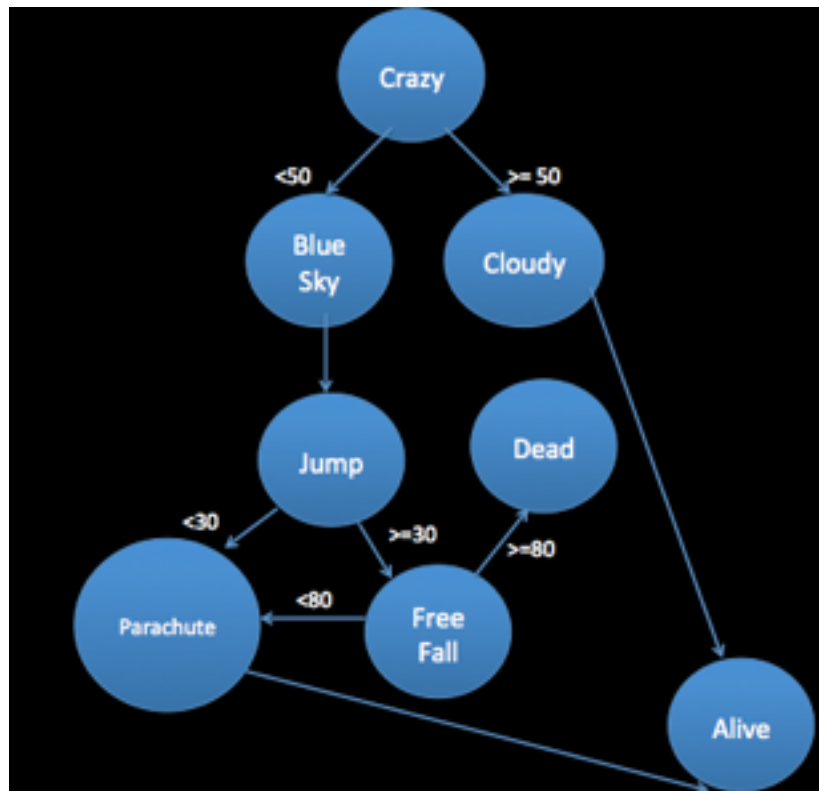
Part 2: Scala Practice

During lectures on Scala, we have learned function literals and closures. Please run the following exercises. For each of them, 1) screen capture the output on the screen, 2) ask a good question about the code 3) and provide your own answer to it. Your questions and answers will both be graded based on their quality and difficulty levels. Submit the screen captures in images. Submit the questions and answers in pdf.

1)
```
val num1 = 1 to 15
var num2 = 2
num1.foreach(num2 += _)
println(num2)
```

2)

```scala
val strs = List ("test", "mind-swap", "swap-mind")


def searchFrom(i: Int): Int =

if (i >= strs.length) -1
else if (strs(i).startsWith("-")) searchFrom(i + 1)
else if (strs(i).endsWith(".scala")) i
else searchFrom(i + 1)

val i = searchFrom(0)
```

3)

```scala
def makeRowSeq(row: Int) =
  for (col <- 1 to 10) yield {
    val prod = (row * col).toString
    val padding = " " * (4 - prod.length)
    padding + prod
}

def makeRow(row: Int) = makeRowSeq(row).mkString

def multiTable() = {
  val tableSeq =
    for (row <- 1 to 10)  yield makeRow(row)
  tableSeq.mkString("\n")
}
```

4)

```scala
val f = _ + _
val f = (_: Int) + (_: Int)
f(5, 10)
```

5)

```scala
def sum(a: Int, b: Int, c: Int) = a + b + c
val a = sum _
a(1, 2, 9)
```


Part 3: Download Wikipedia Dump

Requirements: >100GB free disk space on your machine.

1. Download the latest English Wikipedia dump. Follow the steps in the lecture notes to download the English Wikipedia Dump at http://dumps.wikimedia.org/enwiki/latest/enwiki-latest-pages-articles-multistream.xml.bz2
2. Uncompress the Wikipedia dump file and show the first 20 lines of the file on your screen, by following the instructions given in the lecture notes.
3. Make a screen capture of your results

What to Submit:

Part 1:
- Your code
- Screen captures of at least 4 random runs of the results

Part 2:
- Screen captures of your outputs
- Questions and answers in PDF

Part 3:
- Screen captures of the results

Where to submit:

Canvas

When:

Due on 1/24/18, 11:59pm.

Note:

Your screen captures should include your screen name (better to show your name or netid) to show that they are captured from your computer.