Assignment Zero CS562   Aug 24, 2006   -- not for credit, do not turn in --

Please address questions to:

Jonathan May jonmay@isi.edu   and   Kevin Knight   knight@isi.edu

I. Probability theory ===================================

If you roll two six-sided dice (sides numbered 1 through 6), what is the chance
the dice will sum to exactly 8?

II. Text corpus manipulation ==============================

Write a program to compute sorted frequencies of all words in an English
document (output format like this):

```
 14 the
 14 test
 12 specify
 12 set
  9 and
  8 size
  8 a
  7 of
  7 corpus
   ...
```

If you use Linux, try this: % tr -cs 'a-zA-Z' ' ' < textfile | tr ' ' '\012' |
sort | uniq -c | sort -nr | more

III. Course software: Carmel ====================

Carmel C++ source and binaries for Linux and Windows can be found at
http://www.isi.edu/licensed-sw/carmel/, which also includes the Carmel tutorial.
Download and install this software. A helpful note at the bottom of the
download page says:

```
  NOTE!!!!
  After downloading, to install the software do the following:
  1. Rename the file carmel-tar-gz.pl as carmel.tar.gz (% mv carmel-tar-gz.pl
     carmel.tar.gz)
  2. Gunzip the software (% gunzip carmel.tar.gz)
  3. Un-tar the software (% tar -xvf carmel.tar). [This will extract a
     directory called 'graehl']
```

Test the software by running:

graehl/carmel/bin/carmel

You **should see** something like:

```
        usage: carmel [switches] [file1 file2 ... filen]
        composes a sequence of weighted finite state transducers and writes the
        result to the standard output.
                ... (plus other text documenting usage) ...
```

(over)

IV. Course software: Tiburon ====================

Java code for Tiburon can be found at http://www.isi.edu/licensed-sw/tiburon/,
which also includes the Tiburon tutorial. **Download and install** this software.
A helpful note at the bottom of the download page says:

```
   NOTE!!!!
   After downloading, to install the software do the following:
   1. Rename the file tiburon-tar-gz.pl as tiburon.tar.gz
      (% mv tiburon-tar-gz.pl tiburon.tar.gz)
   2. Unzip and un-tar the software (% tar -xzvf tiburon.tar).
```

Test the software by running:

```
   java -jar tiburon.jar
        or
   ./tiburon [the $JAVA_HOME variable must be set]
```

after which **you should see** something like:

```
   This is Tiburon, version 0.4.0
   ... (plus other text documenting usage) ...
```