Advanced Questions
for
C Programming

1. Words prefixes are also called stems. Write a program that reads a file with one word per input line and finds the most popular stems of size 2 to 6 (if you get a tie, just pick one). Given the following input:

```
test
tester
jest
compute
computer
literate
literal
literacy
continue
collaborate
```

it should give the following output:

```
Most popular stem of size 2: co (occurs 4 times)
Most popular stem of size 3: lit (occurs 3 times)
Most popular stem of size 4: lite (occurs 3 times)
Most popular stem of size 5: liter (occurs 3 times)
Most popular stem of size 6: litera (occurs 3 times)
```

2. Download the Unix spell check file from:

```
http://www.wordgumbo.com/ie/ger/eng/words.txt)
```

Open and read this file (/usr/dict/words) and print a histogram of the occurrence of each word length (in percent), giving output similar to the one below:

```
Distribution of 45402 words in dictionary:
2 49 (0%)
3 536 (1%) X
4 2236 (4%) XXXX
5 4176 (9%) XXXXXXXXX
6 6176 (13%) XXXXXXXXXXXXX
7 7371 (16%) XXXXXXXXXXXXXXXX
8 7074 (15%) XXXXXXXXXXXXXXX
9 6089 (13%) XXXXXXXXXXXXX
10 4593 (10%) XXXXXXXXXX
11 3069 (6%) XXXXXX
12 1880 (4%) XXXX
13 1137 (2%) XX
14 545 (1%) X
15 278 (0%)
16 103 (0%)
17 57 (0%)
18 23 (0%)
```

19 3 (0%)
20 3 (0%)
21 2 (0%)
22 1 (0%)
28 1 (0%)

3. Create two subroutines that convert a decimal number in the range 1 to 4000 to Roman numerals, and back. Warning: This question is rather more difficult than it looks.

   The Roman numerals are:

I	1
V	5
X	10
L	50
C	100
D	500
M	1000

   Roman numerals generally go from large values to small values, with one notable exception: a single small numeral can be used before a larger numeral, and is then subtracted from that larger numeral. So *XC* means 90. However, only the following combinations are valid:

IV	4
IX	9
XL	40
XC	90
CD	400
CM	900