Optional HA

Deadline: 15 Oct 2012
Extra marks: 5/100

Producitons in a PCFG are listed in a text file in the following format ^LHS\s(RHS\s)+#Probability$

pcfg.productions

S NP VP #0.7
S VP #0.3
NP Pronoun #0.2
NP Det Noun #0.4
NP Noun #0.2
NP NP PP #0.2
VP Verb #0.2
VP Verb NP #0.3
VP VP PP #0.5
PP Prep NP #1.0

Lexicon files too are given with each word and its probability in a separate line.

Noun.lex	Det.lex	Verb.lex	Prep.lex	Pronoun.lex
workers 0.3	the 0.4	dump 0.4	into 0.1	I 0.4
dump 0.35	a 0.3	sack 0.4	on 0.2	he 0.3
sacks 0.15	that 0.1	bin 0.2	in 0.3	she 0.3
bin 0.3	this 0.2		to 0.3	
			from 0.1	

Implement Inside and outside algorithms to compute
a. The probability of a given input sentence [Language Model]
b. The most probable parse of the input sentence along with the score [Parser]
c. The probability of a substring being dominated by a non-terminal in a sentence. Input to this would have two lines with first line being the sentence and the second line containing the non-terminal and the substring.
Workers dump sacks into bin
VP dump sacks into bin