Q1:  
Write a regular expression for the set of strings over {a,b}∗ which have an even number of a's followed by an odd number of b's. For example, it should accept:

* b
* bbb
* aab

(aa)\*(b(bb)\*)  
  
Q2:   
Describe in words the pairs of strings that the following FST accepts.  Give three examples of pairs of strings that it accepts. (Note: the start state is q0, and the accept state is q0).

一張含有 寫生, 圓形, 圖表, 圖畫 的圖片

自動產生的描述

This FST accept a pair of input, which b replaced with a and then a replace with b. Since it only accpet a string at a node. It only accept even pair of (ba). The start state is also accept state, which means it also accept empty string.  
e.g.:  
1.input:{b,a}, optput:{a,b}  
2.input:{b,a,b,a}, output:{a,b,a,b}  
3.input:{},output:{}  
  
Q3:  
1.What is the formal definition of a language?  
2.What is the linguistic definition of a grammar?

Formal Definition of a Language: A language is a set of strings of symbols that may be constrained by certain rules. In formal language theory, these symbols are taken from a finite set known as the alphabet. A language is any subset of the set of all possible strings over its alphabet.   
Linguistic Definition of a Grammar: A grammar is a set of structural rules governing the composition of clauses, phrases, and words in any given natural language. However, in the context of formal language theory, a grammar is a formal system specifying a set of strings in a formal language, which consists of production rules that describe how to form strings from the language's alphabet that are valid according to the language's syntax.

Q4:

What is morphology?  Give two examples of morphological phenomena in some language.

Morphology: study of words and how they are formed  
Inflection: walking  
Compounding: baseball

Q5:

Consider the following lowercased, tokenized sentence, with <START> and <STOP> symbols:  
<START> we also know there are known unknowns ; there are things we know we don't know . <STOP>  
Using a bigram language model with probabilities estimated from the above sentence, compute the probability of the following sentence:  
<START> we know . <STOP>  
Show your work. You can leave your final answer as fractions multiplied together. You do not need to multiply numbers to get a final probability.

P(we|<START>)\*P(know|we)\*P(.|know)\*P(<STOP>|.)  
"<START> we" occurs 1 time. "<START>" occurs 1 time.  
"we know" occurs 1 times. "we" occurs 3 times.  
"know ." occurs 1 time. "know" occurs 3 times.  
". <STOP>" occurs 1 time. "." occurs 1 time.  
1/1\*1/3\*1/3\*1/1=1/9

Q6:

1. In the type vs token distinction, what are types?  What are tokens?
2. What is Zipf's law?
3. Describe a problem that Zipf's law causes when building a language model.  Describe a method that can help fix this problem.