CS4248
AY 2022/23 Semester 1
Tutorial 1

1. For each regular expression *rexpr* and the corresponding string *s* in the following table, indicate the span of string *s* that is matched by *rexpr* (i.e., the output of `re.search(rexpr, s)` in Python).

rexpr	s
ha*	hahaha
ha*	haaa
ha*	hihahaha
(ha)*	hahaha
(ha)*	haaa
(ha)*	hihahaha

2. In Porter stemming algorithm, the matching condition of one of the rewrite rules is that the word to be matched contains a vowel (*a, e, i, o, u*) before ending with *ing*. The intent is that when a word satisfies this condition, it is a verb in continuous tense, and so its *ing* ending can be removed to convert the verb into its base form.

Give a regular expression that will match a word that satisfies the above mentioned condition. Assume that the string to be matched is a word consisting of lowercase letters and digits.

3. Let $\Sigma = \{0,1\}$. Give a FSA that accepts all strings of 0's and 1's in which there are an even number of 0's and an even number of 1's.

4. Suppose events $A_1, \ldots, A_n$ are disjoint and $\bigcup_{i=1}^{n} A_i = \Omega$ (Ω is the sample space). Suppose $P(A_i) > 0$ for $i = 1, \ldots n$. Let B be any event where $P(B) > 0$. A more elaborate version of Bayes' Theorem is as follows:

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j) \cdot P(A_j)}$$

Give a proof of this theorem.