HG2051 Sample Written Exam

Name: _____

1 Python Fundamentals

1.1 Expressions and Assignment

1. For each block of code, write the value **assigned to `x`** after the final line in each block. For example:

   (i)

   ```
   >>> x = 2 + 2  # example
   >>> x + 1
   ```

   4 ← your answer (the value of x, not the value of the last expression)

   (a)

   ```
   >>> vocab = set("do you know the muffin man the muffin man the muffin man".split())
   >>> x = len(vocab)
   ```

1.2 Functions

2. Circle the functions that correctly return only those sentences with **fewer than** 10 words? There may be more than one correct function. You may assume the `sents` argument is a list of sentences where each sentence is a list of words.

```
def short_sents1(sents):
    for sent in sents:
        if len(sent) < 10:
            return sent
```

```
def short_sents2(sents):
    return [sent for sent in sents
            if len(sent) <= 10]
```

```
def short_sents3(sents):
    short = []
    for sent in sents:
        if len(sent) < 10:
            short.append(sent)
    return short
```

1.3 Regular Expressions

3. Which of the following strings match the regular expression `r"[A-Z]\w{2,5}"`:

   ◯ `"Alphabet"`

   ◯ `"Alpha"`

   ◯ `"beta"`

   ◯ `"Rho"`

   ◯ `""`

2 NLP

2.1 Frequencies

4. When building frequency distributions of text, what is the purpose of filtering out stop words?
    A. to prevent high-frequency, low-relevance items from "drowning out" more relevant items
    B. to remove social biases from text
    C. to separate train and test instances for experimentation

2.2 Basic Text Processing

5. What are examples of text normalization? (Check all that apply)
    ◯ Syntactic parsing
    ◯ Making all letters lowercase
    ◯ Lemmatization
    ◯ Part-of-speech tagging
    ◯ Building frequency distributions
    ◯ Spelling correction

2.3 N-grams and Part-of-speech Tagging

6. Write out the bigrams for the sentence “Yes I know the muffin man”

2.4 Machine Learning

7. In a machine learning experiment setup, match each dataset with its role in the experiment:

Train •	• Evaluating intermediate models in order to fine-tune model parameters
Development •	• Evaluating the final model to produce publishable results
Test •	• Learning (or fitting) a model to examples