Quiz 1:

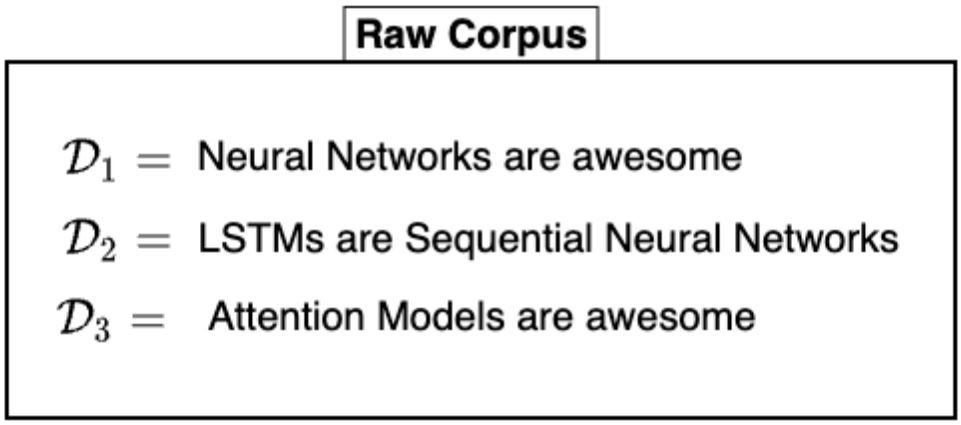
The Preprocessing Step

* Indicates required question

1. Please enter your name: *

Preprocessing the data

Let us consider the following corpus



The word2idx dictionary associated with the Raw Corpus is the following dictionary:

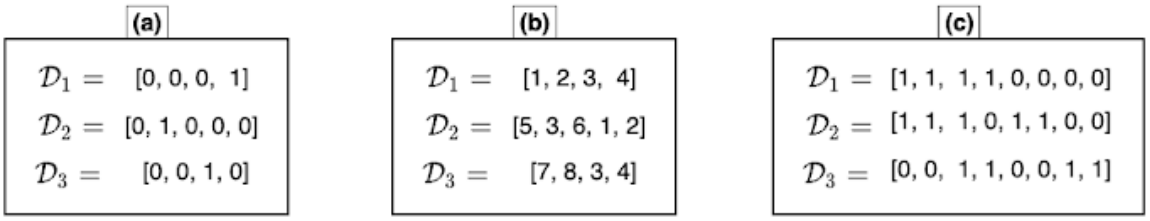
```
Word2idx = {   Neural     : 1,
               Networks   : 2,
               are        : 3,
               awesome    : 4,
               LSTMs      : 5,
               Sequential : 6,
               Attention  : 7,
               Models     : 8  }
```

2. What is the one hot vector representing the word "Sequential" ? 1 point

Mark only one oval.

- 6
- [0, 0, 0, 0, 0, 1, 0, 0]
- [0, 0, 0, 0, 0, 6, 0, 0]

3. Using the word2idx dictionary, which answer represents the processed raw corpus into lists of integers, as explained in the Coursework 1 point



Mark only one oval.

- (a)
- (b)
- (c)

Introducing the Context Words

Let us consider a context size = 2 in the rest of the section

4. What are the context words associated with the center word 5 ? 1 point

$$\mathcal{D}_2 = [\,5\,,\ 3\,,\ 6\ ,\ 1\ ,\ 2\ ]$$

Mark only one oval.

- [3, 6, 1]
- [1, 2, 3, 6]
- [3, 6]

5. What are the context words associated with the center word 3 ? 1 point

$$\mathcal{D}_2 = [\ 5\ ,\ 3\ ,\ 6\ ,\ 1\ ,\ 2\ ]$$

Mark only one oval.

- [5, 6]
- [5, 6, 1]
- [6, 1, 2]

6. What are the context words associated with the center word 6 ? 1 point

$$\mathcal{D}_2 = [\ 5\ ,\ 3\ ,\ 6\ ,\ 1\ ,\ 2\ ]$$

Mark only one oval.

- [3, 6]
- [3, 6, 2]
- [5, 3, 1, 2]

7. What are the context words associated with the center word 1 ? 1 point

$$\mathcal{D}_2 = [\ 5\ ,\ 3\ ,\ 6\ ,\ 1\ ,\ 2\ ]$$

Mark only one oval.

- [3, 6, 2]
- [6, 2]
- [2]

8. What are the context words associated with the center word 2 ? 1 point

$$\mathcal{D}_2 = [\,5\,,\,3\,,\,6\,,\,1\,,\,2\,]$$

Mark only one oval.

- [6, 1]
- [3, 6, 1]
- [5, 3, 6, 1]

9. What is the sum of the total number of context words associated with each center word in document D_2 ? 1 point

Mark only one oval.

- 12
- 9
- 14

10. What is the sum of the total number of context words associated with each center word in document D_2 , as a function of n_2 (the length of document D_2) ? (with the assumption n_2 >= 4). 2 points

$$\mathcal{D}_2 = [w_2^1, w_2^2, \ldots, w_2^t, \ldots, w_2^{n_2-1}, w_2^{n_2}]$$

Mark only one oval.

- 10 + 4*(n_2 -4)
- 2*n_2 + 4
- 3*n_2 -1

11. Any question ?