Q1) a) <s> she dances gracefully </s>
<s> Gracefully she dances </s>
<s> She dances </s>

(i) $P(\text{she} | <s>) = \frac{2}{3} = 0.6667$

(ii) $P(</s> | \text{dances}) = \frac{2}{3} = 0.6667$

(iii) $P(\text{gracefully} | \text{dances}) = \frac{1}{3} = 0.3332$

(iv) $P(\text{gracefully} | <s>) = \frac{1}{3} = 0.3332$

(v) $P(\text{dances} | \text{she}) = \frac{3}{3} = 1$

(vi) $P(\text{she} | \text{gracefully}) = \frac{1}{2} = 0.5$

Q1 b) i) "The cat is under the table"
Closed class words → The, is, under
open class words → cat, table

ii) "One kilobyte is equivalent to 1024 bits"
closed class words → One, is, to, 1024
open class words → kilobyte, equivalent, bits

Q1 c) (i) Both document clustering & topic modelling are used for organising documents of text into topics
Topic model → eg. LDA Generative models, generates words using documents multinomial distribution, learn topic representation using word distributions, uses probabilistic distributions to identify similar words
↳ must know number of topics, cannot capture correlation between topics (assume independence)
Document clustering: Also unsupervised, don't need to know number of topics, Doesn't rely on priors of conjugates, scalable

Q1 c) (ii) i) Cataloguing topics of products in ecommerce
2) Creating documents & topics for wikipedia catalogue
3) Recommendation by item - customer feedback

Kashif Saiyed, 2022AC05668, MTech AIML (3)

Q3) a) "I left the umbrella in the car"

→ Lesk algorithm is a knowledge based word sense disambiguation algorithm, considering the dictionary gloss or definitions provided for "left".

the steps are

(1) For the given definitions & the required word "left", calculate the number of non stopwords overlapping with the given context.

(2) whichever definition has highest overlapping non-stop words with context, chose that one for the disambiguation.

For this case,

(v) or (verb) has 2 overlapping non-stopwords i.e. umbrella & car, whereas (n) & (adj) have only one.

Therefore, (v) or verb will be chosen.

Q3 b) The way humans "see" a webpage or how a machine sees a webpage is very different. Semantic content is understandeable to humans but not to computers as they interpret machine language.

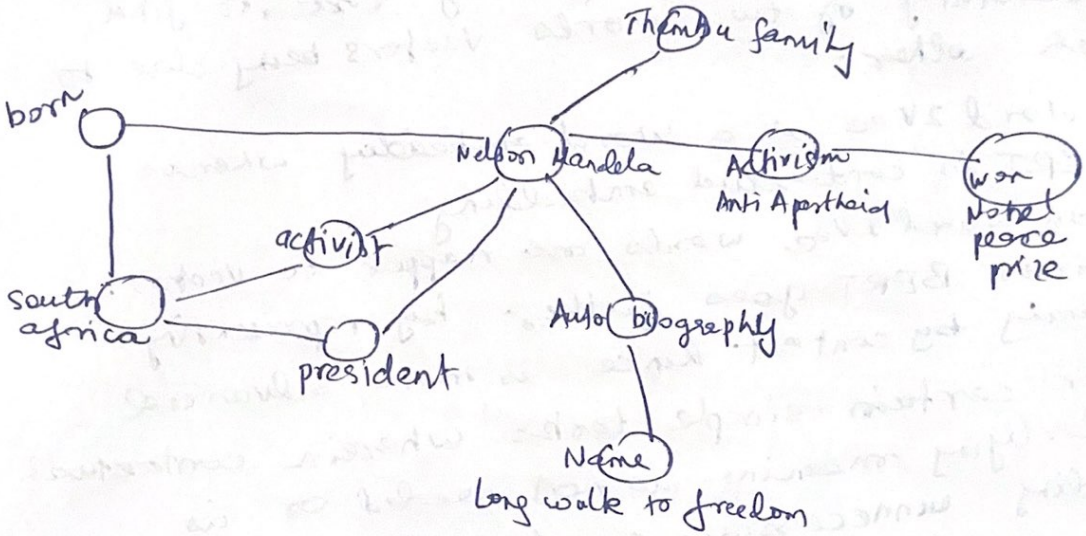
→ Semantic web technologies help with interlinking data over the web

→ This creates relations between content giving context based searches possible for users.

→ This semantic web technologies make this possible for more expressive representation of this knowledge

→ This major applications can be put to use for information retrieval improving keyword matching, cross language retrieval semantic, context based information retrieval etc.

Q3) c) We can use concept of triple (subject, predicate, object).



Q4) a) Word2vec is a popular embedding method. It is fast to train, using techniques like skip gram. A classifier is trained on a binary prediction task to find a word closest to a given context word.

→ Uses a large corpus of text, to find probability of two words vectors being close to each other.

→ Word2Vec is a word embedding whereas BERT is contextual embedding

→ In word2vec words are mapped to vectors, whereas BERT goes further by representing meaning by context. Hence is more advanced

→ For certain simple tasks wherein contextual underlying meaning is not needed or is adding unnecessary complexity, Word2vec offers a good or sometimes better approach

→ It reduces training & deployment complexity as word2vec corpus are readily available & require much less computional capacity.

Q4 b) Matrix for word-word cooccurence

	bank1	bank2	river	money	bug
bank1			0.1	0.2	0.3
bank2			0.7	0.1	0.5
river	0.2	0.5			
money					
bug	0.4	0.7			

~~P(bank1, river & bug) = P(river & bug) P(ba~~

P(bank1, river & bug) = 0.1 + 0.3 = 0.4

P(bank2, river & bug) = 0.7 + 0.5 = (1.2)

Hence, bank 2 is more related to river & bug.

Q4 c) Query

Nile
River
Bug.

$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

Value

$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

Key

$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

Query → Nile

Attention weights = softmax (Query ; Key)

Attention weights = softmax ([1×1+0×1 +1×0]) = Softmax [1]

Since Softmax ([1]) = [0.731; 0.269] are the attention values

Now we compute weighted sum of values using these attention weights

Output = Attention weights × Value:

Output = [0.731×1 − 0.269×0, 0.731×0 + 0.269×0, 0.731×1 + 0.269×1]

output = [0.731, 0, 0.731 + 0.269]

output = [0.731, 0, 1]

Therefore output for query nile will be :

[0.731, 0, 1]

Q5) Query: "Who is the friend of bug?"

Similarity (document i, document j) = 10 × (no. of overlapping words in doc i & j)

(a) d1 = 10(bug, the, of) = 10(3) = 30

d2 = 10(the, of, bug) = 10(3) = 30

d3 = 10(the, is, friend, bug, of) = 10(5) = 50

d4 = 10(bug) = 10

d5 = 10(the, bug, of) = 10(3) = 30

Using the given function for similarity, we see that the d3 has maximum unique overlapping words (5). & a score of 50.

	Who is the friend of bug
d1	30
d2	30
d3	50
d4	10
d5	30

(b) Sort D, argmax [Σ Rel(i) − Σ Red(i,j)]

∴ D1 & D3, Summary, "Bug placed in computer is a friend of the bug"

(c) Rouge = $\frac{\sum \sum \text{min}(\text{count}(i_r), \text{count}(i_s))}{\sum_{\text{sum}} \sum \text{count}(i_s)}$

$= \frac{1+1}{3+5} = \frac{2}{8} = \frac{1}{4} = 0.25$