

	After smoothing	$P(\text{word} \text{positive})$	$P(\text{word} \text{negative})$
word		9	22
hotel	4	26	22
clean	9	26	22
great	2	26	22
owner	26	2	22
terrible	26	2	22
very	2	26	22
helpful	2	26	22
overall	2	26	22
action	2	26	22
experience	2	26	22
condition	1	26	22
Bad	1	26	22
Horrible	1	26	22
one	1	26	22
week	1	26	22

$$1) P(\text{Positive}|\text{sentence}) = 0.01$$

$$2) P(\text{negative}|\text{sentence}) = 0.0016$$

D6  $\rightarrow$  +ve

$$3) P(\text{Positive}|\text{sentence}) = 0.0017$$

$$P(\text{negative}|\text{sentence}) = 0.0033$$

D7  $\rightarrow$  -ve

$$3) P(\text{Positive}|\text{sentence}) = 0.01$$

$$P(\text{negative}|\text{sentence}) = 0.0016$$

P<sub>s</sub> is positive

- b. Compute the BLEU score for the below translations (candidate1, candidate2). Consider 1gram, 2 gram, 3 gram, 4 gram and Brevity-Penalty for calculating BLUE score .

Reference: The teacher arrived late because of the traffic

Candidate 1: The teacher was late due to the traffic

Candidate 2: A teacher arrived late because of transportation

Bleu Score

Candidate 1

$$\text{Unigram} = \frac{4}{7}$$

$$\text{Bigram} = \frac{1}{6}$$

$$\text{Trigram} = 0$$

$$\text{Four gram} = 0$$

Candidate 2

$$\text{Unigram} = \frac{5}{7}$$

$$\text{Bigram} = \frac{4}{6}$$

$$\text{Trigram} = \frac{3}{5}$$

$$\text{Four gram} = \frac{1}{2}$$

$$\text{BP for candidate 1} = 0.867$$

$$\text{Bleu score for candidate 1} = 0$$

$$\text{BP for candidate 2} = 0.615$$

$$\text{Bleu score for candidate 2} = 0.333$$

1. Identify the type of the dialogue agent from the following conversation. Also explain the working of that type of dialogue agent. (5 marks)

System: How may I help you?

User: I want to book dinner at a restaurant that serves good Chinese food

System: Berkos is a restaurant that serves good Chinese food.

User: ok

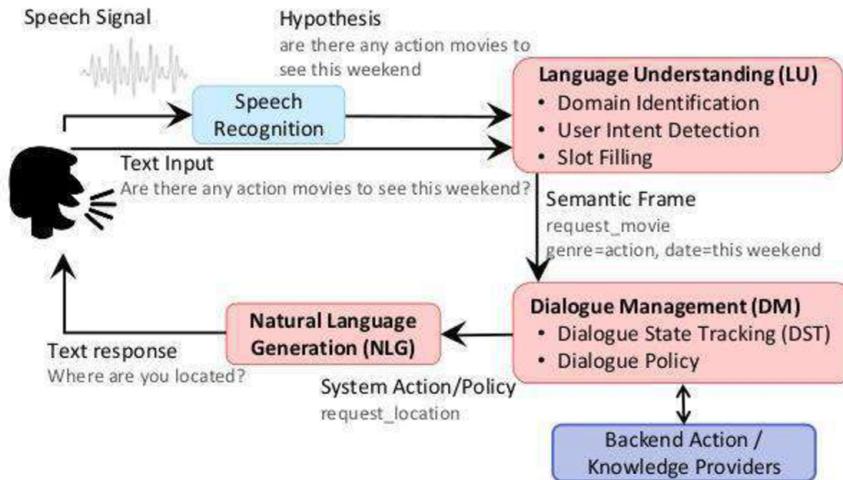
System: Should I book a dinner?

User: yes, please

System: ok.

Soln: Above is the task oriented dialogue system

The diagram shows the working of the task oriented dialogue system. The student should explain briefly each step as mentioned in the slides



2. Find the intent, domain and slots for the following: (2 marks)

Book an appointment on 12<sup>th</sup> Feb 2021 at 10 am for a ECG Test.

SOLN:

DOMAIN: Medical

INTENT: Book an Appointment

Slots

- Services: ECG TEST
- Date: 12<sup>th</sup> Feb 2021
- Time: 10 AM

3. In a collection of 10000 document, the following words occur in the following number of documents: (3 marks)

Oasis occurs in 400 documents, Place occurs in 3500 documents, Desert occurs in 800 documents, Water occurs in 800 documents, Comes occur in 800 documents

Beneath occurs in 200 documents, Ground occurs in 900 documents

Calculate TF-IDF term vector for the following document:

Oasis Place Desert Water Comes Beneath Ground Place

<del>Term</del>	(TF)	Term freq.	IDF	TF * IDF
Oasis	1/8		$\log(10000/400)$	0.1747
Place	2/8		$\log(10000/3500)$	0.11398
Desert-	1/8		$\log(10000/800)$	0.137114
Water	1/8		$\log(10000/800)$	0.137114
comes	1/8		$\log(10000/800)$	0.137114
Beneath	1/8		$\log(10000/200)$	0.212371
Ground	1/8		$\log(10000/900)$	0.13072

TF-IDF vector (0.1747, 0.11398, 0.137114, 0.137114, 0.137114, 0.212371, 0.13072).

**Birla Institute of Technology & Science, Pilani**  
**Work-Integrated Learning Programmes Division**  
**Second Semester 2020-2021**  
**M.Tech (Data Science and Engineering)**  
**End-Semester Test (EC-3 Regular)**

Course No. : DSECLZG525  
 Course Title : Natural Language Processing  
 Nature of Exam : Open Book  
 Weightage : 50%  
 Duration : 2 hours

No. of Pages = 3
No. of Questions = 5

---

Note: Assumptions made if any, should be stated clearly at the beginning of your answer.

### Question 1.

- a) Given a corpus C, the maximum likelihood estimation (MLE) for the bigram “Hello World” is 0.3 and the count of occurrence of the word “Hello” is 580 for the same corpus, the likelihood of ““Hello World” after applying the add-one smoothing is 0.04. What is the vocabulary size of Corpus C.  
 (3 marks)

Handwritten notes:

Soln 1 MLE for "Hello World" is 0.3.  
 $P(\text{World}|\text{Hello}) = 0.3$

This means

$$\frac{\text{count}(\text{Hello,world})}{\text{count}(\text{Hello})} = 0.3$$

$$\frac{\text{count}(\text{Hello,world})}{580} = 0.3$$

$$\text{count}(\text{Hello,world}) = 580 \times 0.3$$

$$= 174$$

After applying add one smoothing

$$\frac{\text{count}(\text{Hello,world}) + 1}{\text{count}(\text{Hello}) + |V|} = 0.04$$

$$\frac{175}{580 + |V|} = 0.04$$

$$175 = 0.04 (580 + |V|)$$

$$|V| = 3795 \quad \underline{\text{Ans}}$$

- b) What are the challenges in the Natural Language Processing? (3 marks)  
 Natural Language Processing has following challenges:
- Contextual words and phrases and homonyms

The same words and phrases can have different meanings according to the context of a sentence and many words – especially in English – have the exact same pronunciation but totally different meanings.

- Synonyms

Synonyms can lead to issues similar to contextual understanding because we use many different words to express the same idea.

- Irony and sarcasm

Irony and sarcasm present problems for machine learning models because they generally use words and phrases that, strictly by definition, may be positive or negative, but actually connote the opposite

- Ambiguity

Lexical ambiguity: a word that could be used as a verb, noun, or adjective.

Semantic ambiguity: the interpretation of a sentence in context. For example: I saw the boy on the beach with my binoculars. This could mean that I saw a boy through my binoculars or the boy had my binoculars with him

Syntactic ambiguity: In the sentence above, this is what creates the confusion of meaning. The phrase with my binoculars could modify the verb, "saw," or the noun, "boy."

- Errors in text or speech

Misspelled or misused words can create problems for text analysis. Autocorrect and grammar correction applications can handle common mistakes, but don't always understand the writer's intention.

- Colloquialisms and slang

Informal phrases, expressions, idioms, and culture-specific lingo present a number of problems for NLP – especially for models intended for broad use.

- Domain-specific language

Different businesses and industries often use very different language. An NLP processing model needed for healthcare, for example, would be very different than one used to process legal documents.

- Lack of research and development

- c) There were 100 documents and each document contained one word. 30 of these documents contained the word "hello". I asked Bob to separate all the documents containing the word "hello". He showed me 60 but "hello" was not in 40 of them. Construct the confusion matrix and calculate the accuracy. (4 marks)

John

Confusion matrix  
"Experiment"

		T	F
Golden (Actual)	T	20	10
	F	40	30

Accuracy =  $\frac{(TP + TN)}{Total} * 100$

$$= \frac{20 + 30}{100} * 100$$
$$= 50\%$$

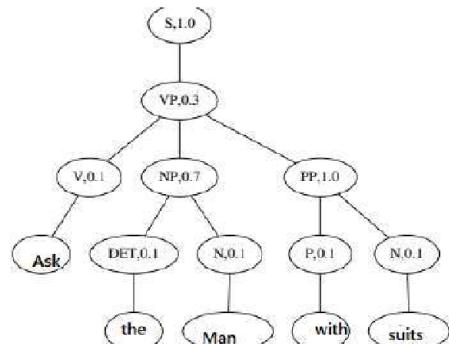
## Question 2.

Given the following PCFG, find the parse trees for the given sentence and their probabilities .And find out that the word 'suits' is attached with 'ask' or 'man' and why? [10 marks]

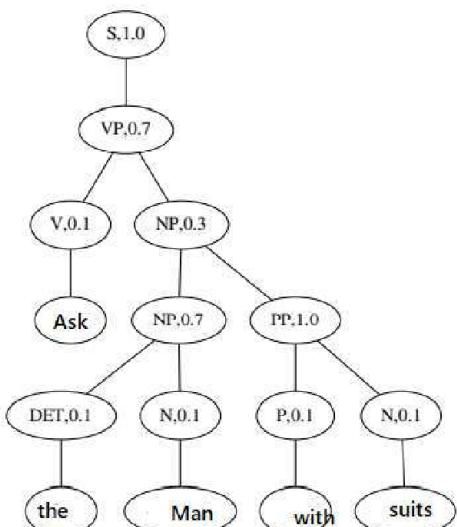
**Ask the man with suits**

Rule	probability
$S \rightarrow VP$	1.0
$VP \rightarrow V NP$	0.7
$VP \rightarrow V NPPP$	0.3
$NP \rightarrow NPPP$	0.3
$NP \rightarrow DET N$	0.7
$PP \rightarrow PN$	1.0
$DET \rightarrow \text{the}$	0.1
$V \rightarrow \text{ask}$	0.1
$P \rightarrow \text{with}$	0.1
$N \rightarrow \text{man} \mid \text{suits}$	0.1

Soln:



$$\text{Probability} = 0.3 \times 0.7 \times 0.1^5 = 21 \times 10^{-7}$$



$$\text{Probability} = 0.3 \times 0.7 \times 0.7 \times 0.1^5 = 14.7 \times 10^{-7}$$

The first tree has higher probability and it is the correct parse since ‘with suits’ should attach to ‘ask’ rather than ‘man’.

### Question 3. Word sense disambiguation and ontology-

- a) How can the Simple Lesk algorithm be applied to disambiguate the exact meaning of “**bass**” in following sentence [5 marks]

The **bass** guitar, is the lowest pitched member of the guitar family of instruments.

*S:(n) bass (the lowest part of the musical range)*

*S: (n) sea bass, bass (the lean flesh of a saltwater fish of the family Serranidae)*

*S: (n) bass (the member with the lowest range of a family of musical instruments)*

*S: (adj) bass, deep (having or denoting a low vocal or instrumental range) "a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"*

- b) Build a small part of ontology for MTech DSE program in OWL syntax with following concepts [3 marks]

- Professor
- Student
- Courses

Also include following relations/constraints:

- Domain
- Range
- subClassOf
- disjointWith

How are the ontology languages OWL and RDF different from each other. Can you express the same constraints using RDF? If not which one cannot be expressed using RDF? [2 marks]

```
<rdfs:Class rdf:ID=" Professor">
  <rdfs:subClassOf rdf:resource="# AcademicStaff "/>
</rdfs:Class>
<rdfs:Class rdf:ID="Professor">
  <owl:disjointWith rdf:resource="#AssistantProfessor"/>
</rdfs:Class>
```

OWL is more advanced and has inferencing capability since owl is based on description logic. Some constraints like disjoint with cannot be expressed using RDF

### Question 4.

1. Given the two machine translation systems output and reference given below, find the best machine translation system using BLEU score with Brevity penalty. [5marks]

[Hint: Assume 1-gram, 2-gram, 3 -gram and 4- gram for calculating BLEU score)

**System A: Israeli official's responsibility of airport safety**

**System B: Airport security Israeli officials are responsible**

**Reference: Israeli officials are responsible for airport security**

2. Given the following documents and their sentiment polarities [5 marks]

Document	Sentiment words	Polarity
D1	Great, Enjoy, Great	Positive
D2	Poor, Unpleasant	Negative
D3	Enjoy ,amazing	Positive
D4	Great, Lovely	Positive
D5	Great, Poor, Rude	Negative
D6	Great ,amazing	?

Determine the sentiment polarity of document D6 using the multinomial naïve Bayes classification (with add1 smoothing) approach. Show your step in detail.

**Solution:**

$$P(\text{Positive}) = 3/5$$

$$P(\text{Negative}) = 2/5$$

$$P(\text{Great}/\text{Positive}) = 3+1/7+7 = 4/14$$

$$P(\text{Great}/\text{Negative}) = 1+1/5+7 = 2/12$$

$$P(\text{Amazing}/\text{Positive}) = 1+1/7+7 = 2/14$$

$$P(\text{Amazing}/\text{Negative}) = 0+1/5+7 = 1/12$$

For the document 6

$$P(\text{Positive}/\text{Great, Amazing}) = 4/14 * 2/14 * 3/5$$

$$= 0.29 * 0.14 * 0.6$$

$$= 0.024$$

$$P(\text{Negative}/ \text{Great, Amazing}) = 2/12 * 1/12 * 2/5$$

$$= 0.16 * 0.083 * 0.4$$

$$= 0.005$$

**Sentiment polarity of document D6 is Positive**

**Question 5.**

- a) Let there be two questions and let there be 4 candidate answers for each question. Also Question Answering System chooses the best answer for question1 and second best answer for question 2. **Calculate the Mean Reciprocal Rank to evaluate the Question Answering System (1 marks)**

**Soln:** MMR =  $(1+1/2)/2 = 3/4$

- b) Let there be four documents given by

D1: the best American restaurant enjoys the best burger

D2: Indian restaurant enjoys the best dosa

D3: Chinese restaurant enjoys the best Manchurian

D4: the best the best Indian restaurant

**Compute the BOW for D1, D2, D3 and D4 in the table. (2 Marks)**

	the	best	American	Restaurant	enjoys	burger	dosa	manchurian	Chinese	Indian
D1										
D2										
D3										
D4										

**Soln b)**

	the	be st	American	Restaurant	enjoys	burger	dosa	manchurian	Chinese	Indian
D1	2	2	1	1	1	0	0	0	0	0
D2	1	1	0	1	1	0	1	0	0	1
D3	1	1	0	1	1	0	0	1	1	0
D4	2	2	0	1	0	0	0	0	0	1

a) Also find out TF-IDF vector for D1, D2, D3, D4 for the above documents in b. (3 marks)

**Soln c)**

WORDS	TF (NORMALISED FREQUENCY)				Idf	Tf*idf			
	D1	D2	D3	D4		D1	D2	D3	D4
the	2/8	1/6	1/6	2/6	$\log(4/4)=0$	0	0	0	0
best	2/8	1/6	1/6	2/6	$\log(4/4)=0$	0	0	0	0
American	1/8	0	0	0	$\log(4/1)=0.6$	$0.6/8=0.075$	0	0	0
Restaurant	1/8	1/6	1/6	1/6	$\log(4/4)=0$	0	0	0	0
enjoys	1/8	1/6	1/6	0	$\log(4/3)=0.12$	$0.12/8=0.015$	0.02	0.02	0
burger	1/8	0	0	0	$\log(4/1)=0.6$	$0.6/8=0.075$	0	0	0
dosa	0	1/6	0	0	$\log(4/1)=0.6$	0	0.1	0	0
manchurian	0	0	1/6	0	$\log(4/1)=0.6$	0	0	0.1	0
Chinese	0	0	1/6	0	$\log(4/1)=0.6$	0	0	0.1	0
Indian	0	1/6	0	1/6	$\log(4/2)=0.3$	0	$0.3/6=0.05$	0	$0.3/6=0.05$

b) Find Domain, Intent and Define Slots for each of the following Sentences: (4 marks)

1) Book a taxi at 6:00 PM from India Gate to Ambience Mall

2) I want to deposit 100 Dollars in my savings account.

solution

1) Book a taxi at 6:00 PM from India Gate to Ambience Mall

- DOMAIN: Cab or Taxi

- INTENT: Taxi-BOOKING

- Slots

- o SOURCE-LOCATION: India Gate

- o DESTINATION-LOCATION: Ambience Mall
  - o PICKUP TIME: 6:00 PM
- 2) I want to deposit 100 Dollars in my savings account.
- DOMAIN: Banking
  - INTENT: Deposit-Account
  - Slots
- o Account Type: Savings Account
  - Transaction: Deposit
  - Amount: 100 dollars

GAURAV ROY .

er Next User

Date of Exam

2021-07-18

Obtain Marks

6

7.5

8

8

ited soon.



Applicant Dashboard



Send Re-Eval Request



Back To List



Home

Qtext :

- a) Draw the top-ranked parse tree for the sentence below by applying the PCFG given in below table. Does the results are good? Provide your comments. (5+1=6 marks)

Sentence: Write the notebooks with pencil.

Consider the following PCFG.

production rule	probability
S → VP	1.0
VP → Verb NP	0.7
VP → Verb NP PP	0.3
NP → NP PP	0.3
NP → Det Noun	0.7
PP → Prep Noun	1.0
Det → the	0.1
Verb → Write   Ask   Find  ...	0.1
Prep → with   in   ...	0.1
Noun → notebooks   teacher   pencil   college   bike   summer   .....	0.1

- b) Which of the three Noun Phrases (1. Pronoun, 2. Proper Noun, 3. Common Noun) to be the most difficult to handle computationally while performing top-down parsing. Explain why?

User Answer :

Q1 a. Answer sheet uploaded.

The results look good. This is because the deeper the tree, grows, more likely the probability of the tree goes down due to fractional multiplication of probabilities.

Therefore, a simpler tree with less depth is better to get the top ranked parse tree.

Online Assessment anytime.anyw X +

wilp.wheebox.com/WAT-3/Subjective\_detailResultOverview.obj?detailResult=view

Apps Coinbase - Your Ho...

probabilities.  
Therefore, a simpler tree with less depth is better to get the top ranked parse tree.

Q1.b.  
Pronoun is the most difficult to handle computationally. PCFG expands based on position and not due to structural context. Therefore, contextual tags like pronouns are not categorized properly.

GAURAV ROY  
2019 HC 04147  
OSEC L2 G525

Q1.a) "Write the notebook with pencil"

Rules:  $S \rightarrow VP$

Option 1

```

S (1.0)
↓
VP (0.2)
↓
(0.1) Verb → "Write"
↓
(0.1) Det Noun → "the"
↓
(0.1) Noun → "notebook"
↓
(0.1) Prep Noun → "with"
↓
(0.1) Noun → "pencil"

```

$= 1 \times 0.3 \times 0.1 \times 0.7 \times 1 \times 0.1 \times 0.1 \times 0.1 = 2.1 \times 10^{-6}$

Option 2

```

S (1.0)
↓
VP (0.2)
↓
(0.1) Verb → "Write"
↓
(0.1) Det Noun → "the"
↓
(0.1) Noun → "notebook"
↓
(0.1) Prep Noun → "with"
↓
(0.1) Noun → "pencil"

```

$= 1 \times 0.7 \times 0.1 \times 0.3 \times 0.1 \times 0.7 \times 1 \times 0.1 \times 0.1 = 1.47 \times 10^{-6}$

The Option 1 parse tree is better since  $2.1 \times 10^{-6} > 1.47 \times 10^{-6}$

GAURAV ROY

Next User

Date of Exam

2021-07-18

Obtain Marks

6
7.5
8
8

Posted soon.



Applicant Dashboard



Send Re-Eval Request



Back To List



Home

Qtext :

a) Given, the following training corpus, Using a bigram language model with and without add-one smoothing, what is  $P(\text{Delhi is beautiful})?$  (6+2=8 marks)

<ss> Delhi is the capital of India </ss>

<ss> Delhi is cold </ss>

<ss> Delhi has beautiful gardens </ss>

b) Suppose the sentence consists of random alphabets (A, a, B, b, ..., Z, z) and each of the 26 letters in upper and lower case occurs with equal probability. What is the perplexity of this sentence?

User Answer :

Q2 a. Sheet uploaded.

Q2 b.

There are 26 letters in the alphabet.

If both upper case and lower case are considered, the total is  $26 \times 2 = 52$  letters.Since all of them can occur with equal probability, each of the given letters has a probability of  $(1/52)$ .

Therefore, for a sentence of length N,

 $P_{\text{Perplexity}} = P(w_1, w_2, \dots, w_N)^{(-1/N)}$  $P_{\text{PP}}(W) = ((1/52)^N)^{(-1/N)}$  $P_{\text{PP}}(W) = 52$ 

GAURAV ROY  
2019HC04149  
DSECL2 G525

Q2 a)~~Probability Smoothing (add-one)~~

Sentence : &lt;ss&gt; Delhi is beautiful &lt;/ss&gt;

Total unique words in training data = 10

\*→ Assumption : end symbol not calculated since not mentioned.

Word	With Smoothing	Without smoothing
$P(\text{Delhi} / \text{ss})$	$\frac{3+1}{3+10} = \frac{4}{13}$	$\frac{3}{3} = 1$
$P(\text{is} / \text{Delhi})$	$\frac{2+1}{3+10} = \frac{3}{13}$	$\frac{2}{3} = \frac{2}{3}$

GAURAV ROY

User Next User

Date of Exam

2021-07-18

Obtain Marks

6

7.5

8

8

Results soon.

PP(W) = 52

GAURAV ROY  
2019HC09149  
DSE CL2 GS25

Q2 a)~~Ranbir Singh (allison)~~

Sentence: &lt;s&gt; Delhi is beautiful &lt;/s&gt;

Total unique words in training data = 10

\*→ Assumption: end symbol not calculated since not mentioned.

Word	With Smoothing	Without smoothing
$P(\text{Delhi} / \langle s \rangle)$	$\frac{3+1}{3+10} = \frac{4}{13}$	$\frac{3}{3} = 1$
$P(\text{is} / \text{Delhi})$	$\frac{2+1}{3+10} = \frac{3}{13}$	$\frac{2}{3} \approx 2/3$
$P(\text{beautiful} / \text{is})$	$\frac{0+1}{3+10} = \frac{1}{13}$	$\frac{0}{2} = 0$

With smoothing,

$$P(\text{Delhi is beautiful}) = \frac{4}{13} \times \frac{3}{13} \times \frac{1}{13} = \frac{1}{169} = 5.917 \times 10^{-3}$$

Without smoothing,

$$P(\text{Delhi is beautiful}) = 1 \times \frac{2}{3} \times 0 = 0$$

GAURAV ROY

Next User

Date of Exam

2021-07-18

Obtain Marks

6

7.5

8

8

ted soon.

Close

Online Assessment anytime.anyw. X +

wilp.wheebbox.com/WAT-3/Subjective\_detailResultOverview.obj?detailResult=view

Apps Coinbase - Your Ho...

Applicant Dashboard

Send Re-Eval Request

Back To List

Home

Qtext:

a). Does the following stage of a Edmond algorithm parsing has an MST ? If not, continue the algorithm for one more step with an Explanation. Obtain MST. (6+2=8 marks)

b). What are the basic differences between syntactic parsing and dependency parsing.

User Answer:

Q3 a. Sheet uploaded.

Q3 b.

Syntactic parsing is based on constituents of words and how the grammar is formed for it

Dependency parsing uses dependency between words.

GAURAV ROY ...

Next User

Date of Exam

2021-07-18

Obtain Marks

6
7.5
8
8

ted soon.

Windows Taskbar icons: File Explorer, Mail, Firefox, Microsoft Edge, Powerpoint, Word, Google Chrome, File Manager.

Applicant Dashboard

Send Re-Eval Request

Back To List

Home

Q2 a) The given stage DOES NOT have an MST or even a ST (cyclic graph present)

- \* Since it is not MST, next step would be to subtract the highest weight from each vertex & node.

→ Here, the new highest weight @ "book" is 0,  
~~2~~ as a consequence the new highest weight  
@ "that" is -2

∴ ~~Final MST~~ because, "that" has already been covered by an edge between "book" & "that", we can remove the "that" → "flight" edge.

→ Therefore, the highest weight @ "flight" is -1  
→ Updated weight @ "that" is 0 since we have "that" → "flight".

Final MST

GAURAV ROY  
2019MC04149  
DSECL2 G525

Next User: GAURAV ROY

Date of Exam: 2021-07-18

Obtain Marks:

6
7.5
8
8

ted soon.

Close

Applicant Dashboard

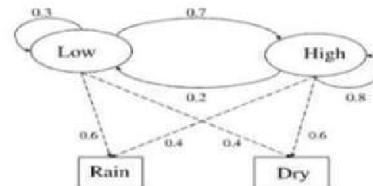
Send Re-Eval Request

Back To List

Home

## Qtext :

The following diagram describes HMM model with two hidden states: Low and High and the observations are rainy and dry. Both the states are equally probable to be initial states (3+5=8 marks)



a. Construct transition state matrix and emission matrix.

b. Let the observation sequence be given as Dry, Rain. Give the corresponding Hidden state sequence.

## User Answer :

Q 4 a.

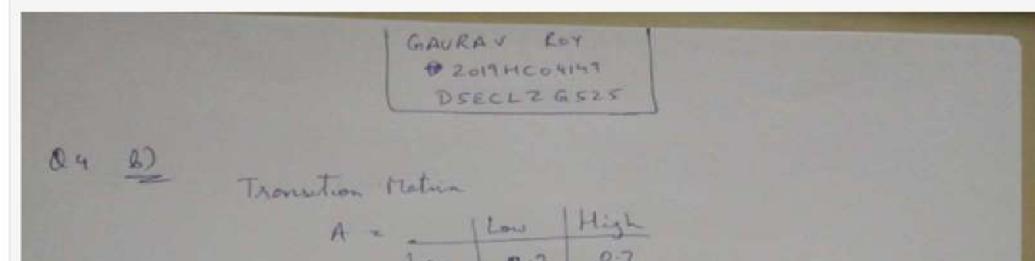
## Transition State Matrix

A	Low	High
Low	0.3	0.7
High	0.2	0.8

## Emission Matrix

B	Rain	Dry
Low	0.6	0.4
High	0.4	0.6

Q 4 b. Sheet uploaded.



GAURAV ROY ..

Next User

Date of Exam

2021-07-18

Obtain Marks

6

7.5

8

8

ted soon.

Online Assessment anytime.anyw. x +

wilp.wheebbox.com/WAT-3/Subjective\_detailResultOverview.obj?detailResult=view

Apps Coinbase - Your Ho...

Q 4 b. Sheet uploaded.

**GAURAV ROY**  
2019HC04149  
DSECLZ G525

Q4 b)

Transition Matrix

$$A = \begin{array}{c|cc} & \text{Low} & \text{High} \\ \hline \text{Low} & 0.3 & 0.7 \\ \text{High} & 0.2 & 0.8 \end{array}$$

Emission Matrix

$$B = \begin{array}{c|cc} & \text{Rain} & \text{Dry} \\ \hline \text{Low} & 0.6 & 0.4 \\ \text{High} & 0.4 & 0.6 \end{array}$$

Initial Stat ( $\pi$ )

$$\begin{aligned} \pi_{\text{Low}} &= 0.5 \\ \pi_{\text{High}} &= 0.5 \end{aligned}$$

Backtrack:

$$\begin{aligned} \text{High} &= 0.076 \\ \text{Low} &= 0.2 \end{aligned}$$

Obs: Rain

Obs: Dry

$P(\text{High}) = 0.3 \times 0.5 \times 0.4 = 0.06$

$P(\text{Low}) = 0.2 \times 0.5 \times 0.6 = 0.06$

$P(\text{High}) = 0.3 \times 0.5 \times 0.4 = 0.06$

$P(\text{Low}) = 0.2 \times 0.5 \times 0.6 = 0.06$

$P(\text{High}) = 0.3 \times 0.5 \times 0.4 = 0.06$

$P(\text{Low}) = 0.2 \times 0.5 \times 0.6 = 0.06$

$\therefore \text{Hidden state sequence} = \{\underline{\text{High}}, \underline{\text{High}}\}$

## Test Submitted Successfully.

For security reasons, please exit your browser.

Question 1

[View Uploaded Answer Sheets](#)

Qtext:-

Question 2

- a) B is a corpus which only contains one single bitstring:

1 1 0 1 1 1 0 0 1 0 1 1 1 0 1 1 1 1 0 0 0

Calculate the following bigram probabilities from the corpus B using MLE (Maximum Likelihood Estimation).

[2 marks]

- i) P (0 | 1)  
ii) P (0 | 0)

Question 3

- b) Write the formulae to calculate the unigram, bigram and trigram probabilities of the sentence:

"It is health that is real wealth and not pieces of gold and silver."

[3 marks]

Question 4

- c) What will be the perplexity value if you calculate the perplexity of an unsmoothed language model on a test corpus with unseen words? Explain. [1 marks]

Question 5

- d) We use the following (part of) lexicon: - [4 marks]

adult	JJ	has	VBZ
adult	NN	just	RB
daughter	NN	my	PRP\$
developed	VBD	programs	NNS
developed	VBN	programs	VBZ
first	JJ	tooth	NN
first	RB	whose	WPS

Consider the following sentence: "My daughter whose first adult tooth has just developed programs" With this lexicon, how many different PoS tagging does this sentence have?

Justify your answer.

It seems like you have not uploaded any images/files for this question.



**Test Submitted Successfully.**

For security reasons, please exit your browser.

Question 1

Question 2

Question 3

Question 4

Question 5

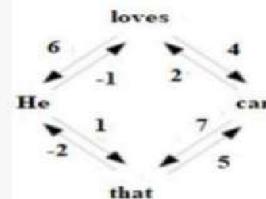
Question 6

Question 7

[View Uploaded Answer Sheets](#)

Qtext:-

In the below weighted graph, find the edge weights between car-loves and that ? car  
using maximum spanning tree [Use Chu Liu Edmond algorithm] [5Marks]



Online Assessment anytime.anywhere. - Google Chrome

wilp.wheebox.com/WET-2/EndTestV2.obj?login\_id=2019hc04149&submitBy=autoSubmit

## Test Submitted Successfully.

For security reasons, please exit your browser.

Question 1	<a href="#">View Uploaded Answer Sheets</a>
Question 2	
Question 3	Qtext:- i) You are designing a frame-based dialog system for movie booking. [5 marks] a). What are the different slots in your design? Mention along with their corresponding entity types and questions that the system would ask a user. b). Show a finite-state dialog manager for the system c). What changes would you make to the design to change it from a single initiative system to multi initiative system?  ii) Find Domain, Intent and Define Slots for each of the following Sentences: [4 marks] a) Book me a table at Marriott hotel. b) Search the list of movies directed by Satyajit Roy  iii) Develop an OWL ontology using the following for animal kingdom for classes like carnivorous, herbivorous and omnivorous. Use following Property characteristics, restrictions and Class expressions [4 marks] a) inverseOf b) domain c) range d) Cardinality e) disjointWith f) subClassOf
Question 4	
Question 5	
Question 6	
Question 7	



^ 🔍 ⌂ ENG 12:42 PM 📲



## Test Submitted Successfully.

For security reasons, please exit your browser.

Question 1

[View Uploaded Answer Sheets](#)

Qtext:-

Find TF-IDF score for the following 2 documents and then find the cosine similarity score [5 marks]

study parsing algorithm article NLP blog  
study POS tagging article NLP blog

Question 4

Question 5

Question 6

Question 7





Test Submitted Successfully.

For security reasons, please exit your browser.

- Question 1
- Question 2
- Question 3
- Question 4
- Question 5
- Question 6
- Question 7

[View Uploaded Answer Sheets](#)

**Qtext:-**

Machine translation

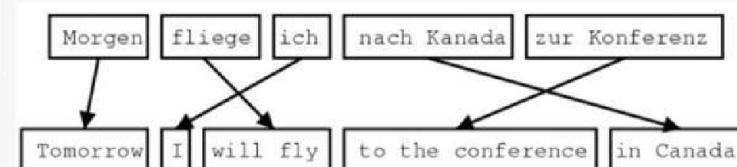
- a) Calculate Final BLEU Score for below Examples (Use Unigram and Bigram Precision and

Brevity penalty ) [5 marks]

Reference: The NASA Opportunity rover is battling a massive dust storm on Mars

### **Candidate 1: The Opportunity rover is combating a big sandstone**

- b) Find the alignment vector for the following, where the first sentence is the source and the second sentence is the target. [2 marks]



Online Assessment anytime.anywhere. - Google Chrome

wilp.wheebox.com/WET-2/EndTestV2.obj?login\_id=2019hc04149&submitBy=autoSubmit

## Test Submitted Successfully.

For security reasons, please exit your browser.

Question 1	<a href="#">View Uploaded Answer Sheets</a>
Question 2	<b>Qtext:-</b> a) How can the Simple Lesk algorithm be applied to disambiguate the exact meaning of "bank" in following sentence. (4 marks) <i>the coin bank was empty at home</i> <ul style="list-style-type: none"><li>▪ S: (n) bank (a long ridge or pile)</li><li>▪ S: (n) <b>bank</b> (an arrangement of similar objects in a row or in tiers)</li><li>▪ S: (n) <b>bank</b> (a supply or stock held in reserve for future use (especially in emergencies))</li><li>▪ S: (n) <b>savings bank, coin bank, money box, bank</b> (a container (usually with a slot in the top) for keeping money at home)</li></ul>
Question 3	b) For the following sentence pairs identify the various lexical relations existing between the common word (2 marks)
Question 4	<i>I) The school is organizing painting competition She is studying in a school.</i>
Question 5	<i>II) Ministry has issued a tender for the furniture The meat should be well cooked and tender</i>
Question 6	
Question 7	

It seems like you have not uploaded any images/files for this question.