II Jai Sri Gurudev II

Sri Adichunchanagiri Shikshana Trust®

# SJB Institute of Technology

(Affiliated to Visvesvaraya Technological University, Belagavi and Approved by AICTE and Accredited by NAAC with 'A' Grade, CGPA-3.22 - New Delhi )

#67, BGS Health & Education City, Dr. Vishnuvardhan Road, Kengeri, Bengaluru – 560060.

Website : www.sjbit.edu.in

SJBIT

## INTERNAL ASSESSMENT BOOK

Student Name : Abhishek Kumar Rai

Semester & Section : VII A USN : 1JB18IS003

Subject: Natural Language Processing Subject Code: 18CS743 Branch: ISE

Name of Faculty in charge : Chetan R

| Q No. | Internal Assessment Test - I<br>Date : 19-11-2021<br>Max. Marks : 50 | | | | Q No. | Internal Assessment Test - II<br>Date : 28/12/2021<br>Max. Marks : 50 | | | | Q No. | Internal Assessment Test - III<br>Date : 31/01/22<br>Max. Marks : 50 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PART - A | | | | | PART - A | | | | | PART - A | | | |
| | A | B | C | Total | | A | B | C | Total | | A | B | C | Total |
| 1 | 8 | 6 | 6 | 20 | 1 | | | | | 1 | 6 | 6 | 8 | 20 |
| 2 | | | | | 2 | 6 | 8 | 2 | 16 | 2 | | | | |
| | PART - B | | | | | PART - B | | | | | PART - B | | | |
| 3 | | | | | 3 | | | | | 3 | 9 | 8 | 4 | 21 |
| 4 | 4 | 3 | 3 | 10 | 4 | 2 | 7 | 7 | 16 | 4 | | | | |
| I Test IA Marks Total | | | | 30 | II Test IA Marks Total | | | | 32 | III Test IA Marks Total | | | | 41 |
| Quiz 1/Assignment etc., | | | | 7 | Quiz 2/Assignment etc., | | | | 4 | Quiz 3/Assignment etc., | | | | 10 |
| Student Signature : Abhishek Kumar Rai | | | | | Student Signature : Abhishek Kumar Rai | | | | | Student Signature : Abhishek Kumar Rai | | | | |
| Signature of Invigilator | | | | | Signature of Invigilator | | | | | Signature of Invigilator | | | | |
| Signature of Faculty in charge | | | | | Signature of Faculty in charge | | | | | Signature of Faculty in charge | | | | |

Avg. IA Marks for 30 (A) 21 Assignment /Quiz etc. Total Quiz-7 Assgn-0 (B) 7 Total IA Marks for 40 (A+B) 28

HOD

Principal

Head of the Department
Dept. of Information Science & Engineering
S.J.B. Institute of Technology
Kengeri, Bangalore-560 060.

Department of Information Science and Engineering.

**Dept. Vision:**

We envision our department as a catalyst for developing educated engaged and employable individuals whose collective energy will be driving force for prosperity and the quality of life in our diverse world.

**Dept. Mission:**

Our mission is to provide quality technical education in the field of information technology and to strive excellence in the education by developing and shaping the intellectual and human potential for good industry and community

**About Anti - Ragging**

SJBIT has zero tolerance policy for ragging. The Institute views ragging, is an uncivilized, and inhuman practice. We do not subscribe to the view that one could wait till something happens in order to initiate stringent action. Any rigorous action in such cases may damage a young career. So we repose faith in averting such eventualities. For this, the Institute has proactive policy.

**Punishments for Ragging**

1. Cancellation of admission.
2. Suspension from attending classes.
3. Withholding/withdrawing scholarship/fellowship and other benefits.
4. Debarring from appearing in any test/examination or other evaluation process.
5. Withholding results.
6. Debarring from representing the University in any regional, national or international meet, tournament, youth festival etc.
7. Suspensions/expulsion from the hostel.
8. Rustication from the college and University for period varying from 1 to 4 years.
9. Expulsion from the college and consequent debarring from admission to any other college.
10. Rigorous imprisonment of three years and/or a fine of upto Rs.25,000.
11. Collective punishment: When the persons committing or abetting the crime or ragging are not identified, the institution has resort to collective punishment as a deterrent to ensure community pressure on the potential raggers.

SJBIT

# Quiz - I

1. d).
2. a).
3. d).
4. a).
5. d).
6. a).
7. b).
8. c).
9. d).
10. b).

7

SJBIT

# Internals - I

## Part - A

Q.1.A There are 5 different levels of NLP those are :→

1. **Lexical** :- At this level text is processed word by word. Potential meaningful words are recognised

ex:- Akash is playing cricket

noun verb — correct

2. **Syntactic** :- The text is processed line by line at this level.

ex:- Akash is apple eating — wrong

all the words are correct but sentence all together does not make sense.

3. **Semantic** :- Once is sentence is processed. Now the context and the meaning of sentence.

ex:- Manu weds Manu — wrong (no meaning).

4. **Discourse** :- It is used when processing a document the large text is broken down in smaller phase, sentence and the processed.

SJBIT

(5). **Pragmatic :-** It is the highest level of processing. Here the model is aware of outer world information it's knowledge is just limited to the text.

8

ex:- Tanu weds Manu

access the database and check if Tanu and Manu are really married.

---

Q.1.B **Components of transformational grammer are :→**

(1). **Phase grammer rule :-** This divides the text into sub-parts such as noun, verb etc.

$S \rightarrow NP + VP$

$VP \rightarrow V + NP$

$NP \rightarrow Noun + Det$

$V \rightarrow Axl + V$

| |
|---|
| Axl → will, is, be |
| Verb → eating, running |
| Det → a, an, th |

ex:- $\underbrace{He}_{Noun}$ $\underbrace{is}_{Axl}$ $\underbrace{running}_{verb}$.

SJBIT

② . Transformational Rule :→ Different types of transformation needs different rules

Ex:- For ~~Active~~ to ~~Pecive~~ Passive

$NP_1 - A - V - NP_2 \longrightarrow NP_2 - A - be - en - V - by - NP_1$

Ex:→ $\underbrace{\text{The Police}}_{NP_1}$ $\underbrace{\text{will}}_{A}$ $\underbrace{\text{catch}}_{Verb}$ $\underbrace{\text{the snatcher}}_{NP_2}$.

6 $\underbrace{\text{The Snatcher}}_{NP_2}$ ~~will catched~~ $\underbrace{\text{will}}_{A}$ been $\underbrace{\text{catched}}_{catched.}$ by $\underbrace{\text{the policeman}}_{NP_1}$

③ . Morphophonemic rule :- It looks for the words with similar phonemes, sound to distinguish ~~bta~~ between words

ex:→ bat — bad

SJBIT

Q.1.c. There are two types of language models:→
① . Grammetical based model.
② . Statistical based model

A model is a description of an complex entity or process. NLP is a complex entity.
So the ~~mod~~ description of NPL is a language model.

① . Grammetical:→ Grammer is used to create the language model.
-6-
All Rules are hand-coded in it. All known ~~ex~~ Rules of grammer

② . Statistical based model:→ A large Quantity of text is given to be model to train on.
Model uses the previsously learnt data and compare and process the data.

SJBIT

## Part-B

Q.4.A. Information Retrieval :→ To extract the required text from the document based the query asked is called Information retrieval.

Issues of Information Retrieval :→

① Wrong Question Structure :→ Sometimes user does ask a proper meaningful question which leads to problem.

② Match Query in the document :→ Sometimes there are more than one phases matches the Query.

4

③. Unstructured document text :→ Sometimes text does not contain data in proper structure which can be extracted.

④. Ambiguity of data :→ Sorting processing text & Create confusion due to ambiguision phase and words. ex:- match — Cricket / — Wedding

Q.4.B. Finite State Automata :→

→ Automata taken from greek word means self-wor

→ It is a pre-defined sequence of processes

A finite Automata can be defined by 5 tuples :—

$[Q, \Sigma, \delta, q_0, F]$

$Q$ → Set of all States

$\Sigma$ → Set of all Symbols

$\delta$ → transition function

$q_0$ → Initial State.

$F$ → Final State.

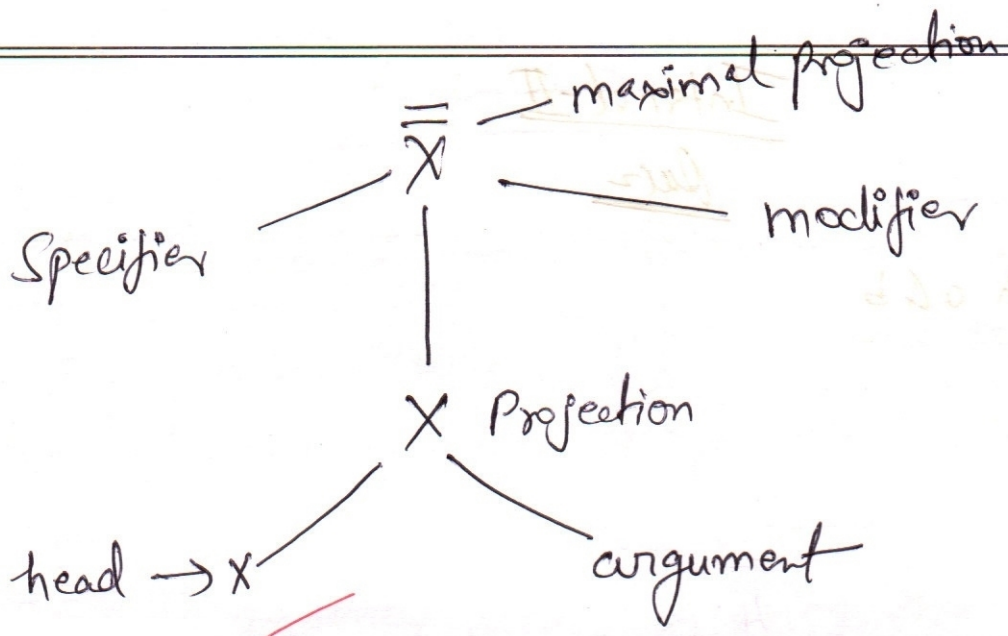
It is of two types.

→ DFA :→ Deterministic Finite Automata

→ NFA :— Non-deterministic finite automata,

It is used to find the words and analysis the text by parsing.

Q.4.C.




-3-

→ $\bar{X}$ theory is a generated theory of a language.

→ In this theory Sentence Structure and Phase Structure or rather than different are considered together as maximal projection.

Sentence Structure + Phase Structure → Maximal Projection

30/50

SJBIT

# Internal-II

## Quiz

1. d. Both a & b
2. b.
3. a.
4. a.
5. b.
6. b
7. b
8. d
9. b
10. a

4

SJBIT

## Part-A

Q.2.A. Minimum edit distance algorithm :→ This algorithm is a machine learning algorithm used to compute the minimum number of operations which are needed to be performed to ~~calculate~~ convert one string to another string by using three known methods Insert, Delete and Update.

Conversion :→ Given :- Word to convert :- paceful
to word :- peaceful

We use the following table matrix to calculate the distance between given words.

| | # | P | A | C | E | F | U | L |
|---|---|---|---|---|---|---|---|---|
| # | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| P | 1 | 0 | 2 | 3 | 4 | 5 | 6 | 7 |
| E | 2 | 1 | 3 | 4 | 5 | 6 | 7 | 8 |
| A | 3 | 2 | 1 | 4 | 5 | 6 | 7 | 8 |
| C | 4 | 3 | 2 | 1 | 5 | 6 | 7 | 8 |
| E | 5 | 4 | 3 | 2 | 1 | 6 | 7 | 8 |
| F | 6 | 5 | 4 | 3 | 2 | 1 | 7 | 8 |
| U | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 8 |
| L | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

→ final output no. of operation

| | # | P | A | C | E | F | U | L |
|---|---|---|---|---|---|---|---|---|
| # | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| P | 1 | 0 | 2 | 3 | 4 | 5 | 6 | 7 |
| E | 2 | 2 | 3 | 4 | 3 | 5 | 6 | 7 |
| A | 3 | 2 | 2 | 3 | 4 | 5 | 6 | 4 |
| C | 4 | 5 | 2 | 2 | 3 | 4 | 5 | 3 |
| E | 5 | 4 | 3 | 2 | 1 | 6 | 7 | 2 |
| F | 6 | 5 | 4 | 3 | 2 | 1 | 7 | 1 |
| U | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 8 |
| L | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

→ Output

1

The distance between peacful and paceful is 1.

It means only one operation is to be performe to make peacful to paceful.

paceful
↑ insert e —— 1 operation.
peacful

total operation = 1

Q.2.B. <u>Top-down Parsing</u> :→

→ It is a machine learning algorithm used to find Grammer from starting grammer S.

→ It starts from the starting Grammer S of the given input.

→ All the sub-tree of S are obtained.

→ Use every grammer on S to create sub-tree and expand.

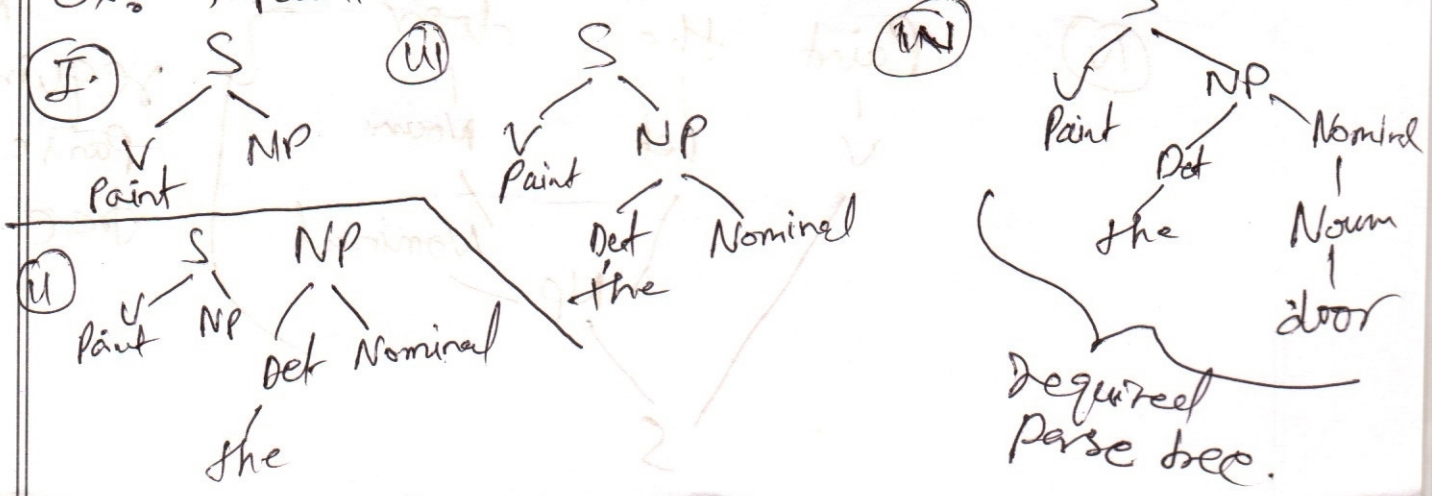
→ All the non-terminating nodes are expanded.

→ A point is reached when all the leave nodes only contain parts of figure of speech.

→ Any non-matching leaves from the input sentence is neglected.

→ Successfully obtain the parsed tree.

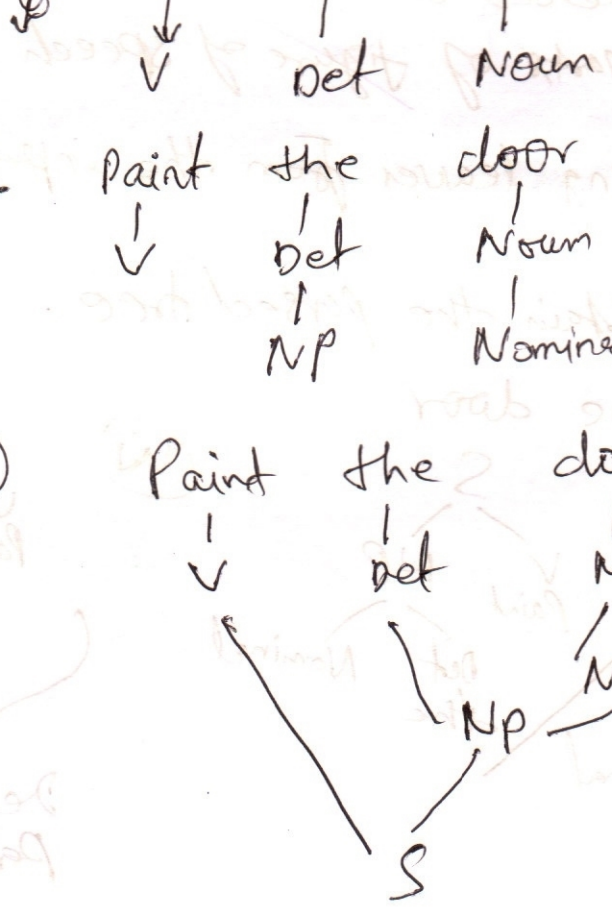
Ex: → Paint the door



## Bottom-up parsing :→

- It is a machine learning algorithm used to find grammer from sentence.
- It starts from the user input and goes upto untill it find the starting grammer S.
- User input is taken to start and further grammer is calculate based on currently present sub tree.
- Once the root node is found check if it is equal to S then accept otherwise reject.

Ex:— Paint the door



SJBIT

Top-down :-

Advantages :→ Do not waste time looking for grammer which does not have S as start node.

Disadvantages :→ Waste time looking for grammer which does not comes up with input sentence.

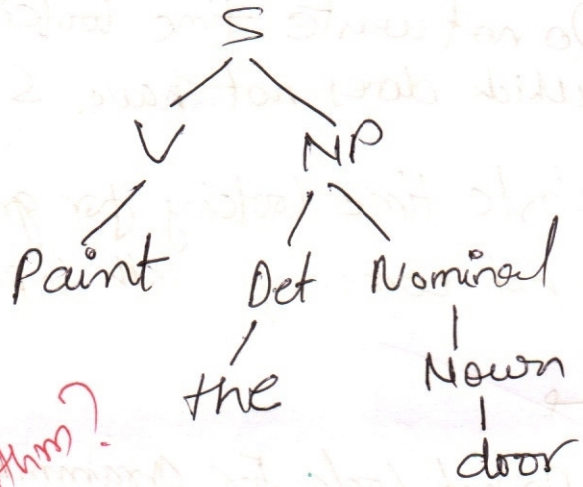
-8-

Bottom-down :→

Advantage :- Do not look for grammer which will not come up with input sentence.

Disadvantage :- It waste time looking for grammer which does not have S as starting node.

Q.2.G Top-down depth first algorithm :→




2

Algorithm?

→ This algorith works from top-down towards the input sentence

→ It starts at the top S as first grammer as root node.

→ All the sub-tree of S are evaluated for all the grammar and node is expanded

→ All the evaluation is done towards the input sentence grammer.

→ The leaf node contain the user input sentence gram

→ The algorithm runs recursively till all the leaf node contains parts of speech grammer.

SJBIT

## Part-B

### Q.4.c. Active learning approach :-

→ It is found in Research that ~~lab~~ acquiring label from human anotation work better than other approches.

Steps:-

① → All the ~~sen~~ Sentence cluster are grouped on the basic of same parse sub-tree.

② → For each cluster sentences are grouped by same target verb.

③ → Take the largest group from ~~l~~ largest cluster and present it to user

④ → Now ~~use~~ allocated the allocated label to all the ~~sub-trees with~~ sentences with same sub-tree.

⑤ → Train with the classification.

⑥ → Find out any ~~missing~~ miss-match and present to the user.

⑦ → Repeat the 4-6 steps Until desired accuracy is obtained.

Q.4.B i). The shortest path hypothesis :- This hypothesis is used to calculate the shortest dependency path among the relations in the sentence.
In processing of text by machine learning my error can occur
ex:- ambiguity
To overcome we find out the different relations in the sentence by shortest dependency path.
This shows how words in a sentence ~~are~~ and the meaning of those words are dependent on other words in the sentence.
The words in the sentence are processed and the dependency through out the sentence is found out
The dependency is used to solve the ~~ab~~ ambiguity
And to better understand the context of the sentence.

(ii). <u>Learning Dependency path</u> :- To slove the problem to understand the context of the ~~sent~~ sentence it is very important to understand the dependency of the words in the sentence with each other.

7 Dependency path is a way to find the relation b/w the words of a sentence by mapping these relation to the other words in the sentence. The relation mapping can be plotted and referred to further analyses the context of the sentence.

SJBIT

Q.4.A Functional overview of Infact system:→

This system refers to the analysis of language processing and understanding of context of sentence and ~~relat~~ relation of words and these dependency to slove any ambiguity in the processing of the text. It uses tagging to slove the problem.

There are three main tagging.

2 Rule-based tagging is used to convert sentences in the form of tuple of sentence or list of sentences.

Stochastic tagging uses frequency and sequence for the analysing and predicting the context of the sentence when the most frequently used tag is used for tagging it is frequency based.

32/50

SJBIT

# Internals - III rd. - Quiz

1. (a) Zipf's law
2. (a) Information Retrieval
3. (b) Search Engine optimization
4. (b) False
5. (d) Inverted Index
6. (a) Gerard Salton
7. (b). helped, helps → help
8. (b). Stop word removal.
9. (d) the terms
10. (b) recall

10

SJBIT

# Part - A

Q-1-A (SVM) - Support Vector machine

→ It is used for binary classification.

→ SVM score is the signed distance from the hyperplane.

→ For multi classes problem a series of SVM are train

→ Every SVM score is mapped with probability.

→ Class membership probability is calculated as a function of SVM score trained as ~~for~~ sigmoid function.

6

→ The fit parameters are the slope of sigmoid function.

→ SVM score trained as a cost factor both positive and negative cost factors.

→ The ratio of positive and negative cost factors gives the decision of right class.

SJBIT

Q.1. B. Latent Semantic Analysis feedback system :-

→ It is a semantic analysis model.

→ It uses similar word from Corpus Documents is represent terms.

→ It uses Stastical technique for analysing the text Document.

→ It uses powerful Mathematical transformation to convert vector space into latent semantic space.

6

→ It counts the frequency of word, how they occur in the document.

→ It does not take in account word order.

→ It works good for larger document does not work good for short document.

→ It efficiency increases with increase in size of object.

→ Whole meaning of the terms is called 'global-focused'.

→ Immediat preceder meaning is called 'local-focused'.

Q.1.c (i) Cohesion :→

→ It measures the relationship between rhetorical role and a preceding action

→ $$H = \sum_{r_i p_i \in H} \frac{P(p_i / r_i)}{|H|}$$

H → Hypothesis

r → rhetorical role

P → action

(ii). Coverage :→

→ It measures how good the hypothesis for the model

→ It can use semi-structured data from LSA.

→ It define and contain rules for represent.

→ $$H = \frac{|\text{Rules Covered}|}{|\text{RuleSet}|}$$

(iii) Interestingness :→

→ It measures the characteristic of hypothesis by ascerquent and consequent.

→ $$\boxed{H = |\text{Semantical difference between ascedent and Consequent}|}$$

→ It measures in the unit of KDD approach.

8

(iv). Plausibility of Origin :→

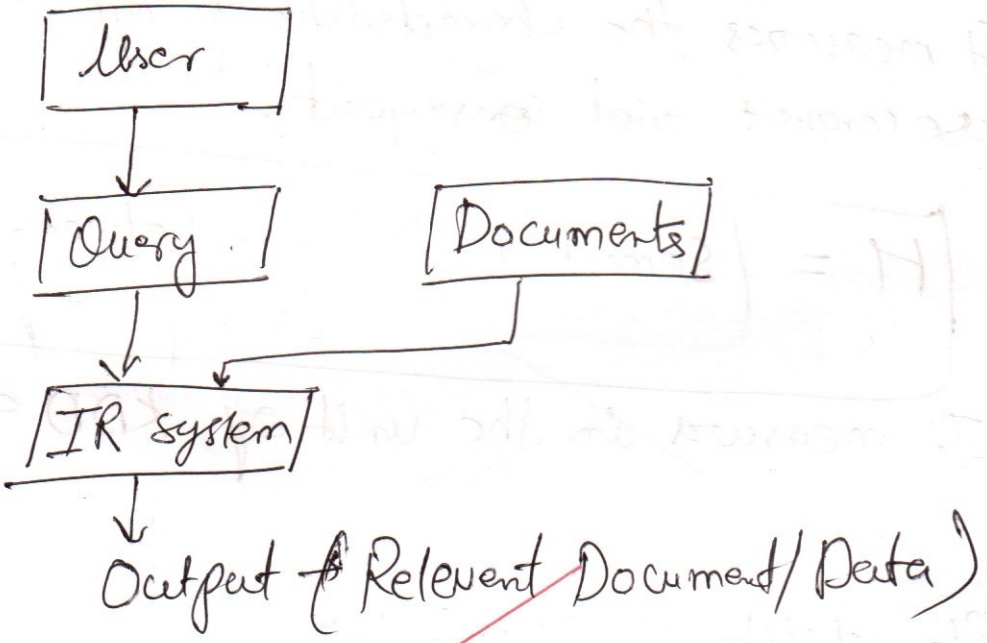
→ It measures how plausible is the hypothesis calculated by samples evidence.

→ If a better hypothesis is found it is used or the original hypothesis is given to next generation.

→ $$\boxed{H = \begin{cases} S_p & \text{if output of Sensor's Crossover} \\ 0 & \text{if output of original} \end{cases}}$$

## Part-B

Q.3.A.




→ User execute the query according to the data required.

→ This query is send to the IR-system.

→ IR system fetches the relevent information from the documents based on query.

→ The relevent information is given as output to user.

~~Features of IR:~~

SJBIT

## Design features:-

### ①. Inverted Index:→

→ It is a key-value pair which stores the word as key and word frequency as value.

→ It is like a hashmap and work very fast.

→ It is used is search operations

### ②. Stop word Elimination:→

→ It is a technique to remove unwanted words from the documents.

→ These words contain less (or) Useless meaning.

→ ex:→ is, an, the

### ③. Stemming:→

→ It is a technique to optimize words.

→ It reduces the redundancy of the words

→ It help in memory optimization.

→ ex:→ hanged → hang, laughing → laugh.

→ It reduces the word to its root (or) Stem word.

SJBIT

(4) Boolean model:-

→ It is based on set theory and boolean expression.

→ It is most commonly used and easy to use.

→ ~~Queries~~ Document are set theory and queries are boolean expressio

SJBIT

Q.3.B. Boolean Model for Classical Information Model :→

→ It is most commonly used Classical model.

→ It is based on set theory and boolean expression

→ Document are set theory and Queries are boolean Expressions.

Ex:→

Document:→ {D1, D2, D3, D4}

D1:→ Anuj is a tall boy and Akash is shy.

D2:→ Anuj play football and Akash play cricket.

D3:→ Cricket is a game of 11 players. Anuj like watching cricket.

D4:→ Anuj and Akash are brothers.

Queries:→

① . ~~Anuj & Akash~~ Anuj and Akash.

② . Anuj and Akash ~~but~~ and not cricket.

Solution:→

① . ~~Anuj & Akash~~.

~~Anuj~~

Solution :→

| | D1 | D2 | D3 | D4 |
|---|---|---|---|---|
| Anuj | 1 | 1 | 1 | 1 |
| Akash | 1 | 1 | 0 | 1 |
| Cricket | 0 | 1 | 1 | 0 |

① Anuj and Akash :— Anuj ∧ Akash.

Anuj — {D1, D2, D3, D4}

Akash — {D1, D2, D4}

$\boxed{\text{Anuj} \wedge \text{Akash} - \{D1, D2, D4\} \text{ Ans.}}$

8

② Anuj and Akash ~~but~~ and not Cricket :→ (Anuj ∧ Akash) ∧ (¬Cric

Anuj :— {D1, D2, D3, D4}

Akash :— {D1, D2, D4}

¬Cricket :— {D1, D4}

$\boxed{\text{Anuj} \wedge \text{Akash} \wedge (\neg\text{Cricket}) - \{D1, D4\} \text{ Ans.}}$

Q.3.C. (i). FRAME NET :→

→ It is a lexical dataset.

→ It contain content which is both machine and human readable.

→ It is based on frame-Symantic.

→ The frame means meaning and lexical means ~~mean~~ one meaning for one word.

4

41/50

(ii). Stemmer :→

→ It is a technique used to reduce word to there ~~root~~ (or) stem word/form.

→ It helps increases in accuracy.

→ It helps reduce memory usage.

→ It is fast and ~~com~~ easy to use.

→ It is used for stop-word elimination to eliminate less-meaning words.

→ Ex:- helped, helps → help.

II Jai Sri Gurudev II

Sri Adichunchanagiri Shikshana Trust

# SJB Institute of Technology

SJBIT

## College Vision

To become a recognized technical education center with global perspective.

## College Mission

To provide learning opportunities that fosters students ethical values, intelligent development in science and technology and social responsibility so that they become sensible and contributing members of the society.

## INSTRUCTIONS TO STUDENTS

1. Fill-in the details on front cover page.
2. No sheets be either removed or added after writing a test.
3. No loose sheets are permitted to be used for answering the tests.
4. The question numbers should be mentioned in the margin only.
5. Minimum of 60% marks should be scored in each subject and in all three tests.
6. The candidate shall write answers on both sides of pages of the answer book. Answers must be written using black pen (ball pen or ink pen). If there is a change in pen, the same shall be attested by the Room Superintendent on the facing sheet of the blue book at the top.
7. No candidate shall be permitted to go to toilet during the period of test.
8. After completion of the test, students should handover the test Book immediately to the concerned Faculty.
9. For both theory and Practical classes, 85% Attendance is compulsory.
10. All three tests are compulsory.
11. No candidate shell be permitted to go to toilet during test.
12. Any candidate appearing for the test is liable to be charged with committing malpractice in the following cases.
    a. Bringing any written material / portions of a book.
    b. Communicating with any candidate.
    c. Bringing mobile phone to examination hall.

|| Jai Sri Gurudev ||
Sri AdichunchanagiriShikshana Trust ®

# SJB Institute of Technology

(A Constituent of BGS & SJB Group of Institutions and Hospitals)
BGS Health & Education City, Dr. Vishnuvardhan Road, Kengeri, Bengaluru-5

Affiliated to Visvesvaraya Technological University, Belagavi. Approved by AICTE, New Delhi. Accredited by NAAC, New Delhi with 'A' Grade. Recognized by UGC, New Delhi with 2(f) and 12(B).Certified by ISO 9001-2015

***DEPARTMENT OFINFORMATION SCIENCE AND ENGINEERING***

## INTERNAL ASSESSMENT PAPER (7th Sem)

| | | |
|---|---|---|
| Internal Test: 1 | Internal Quiz: 1 | Academic Year: EVEN /2021-22 |
| Subject : Natural Language Processing | Sub-Code: 18CS743 | Sem: 7 |
| Date:19/11/2021 | Time:2:00 to 3:45pm | Dur:1.45 min |
| Internal Test max marks: **50** | Internal Quiz max marks: **10** | |
| Staff-Incharge:Chetan R | | |

## I. Quiz (Answer all multiple-choice question in first sheet of your answer book)

| Question No | Multiple choice question | BTlevel | CO Mapped |
|---|---|---|---|
| 1 | What is the field of Natural Language Processing (NLP)?<br>a. Computer Science<br>b. Artificial Intelligence<br>c. Linguistics<br>d. All of the mentioned. | L2 | CO1, PO1 |
| 2 | NLP is concerned with the interactions between computer and human languages.<br>a. True<br>b. False | L2 | CO1, PO1 |
| 3 | What is the main challenge of NLP?<br>a. Handling Ambiguity of sentences.<br>b. Handling POS-Tagging<br>c. Handling Tokenization<br>d. All of the mentioned | L2 | CO1, PO1 |
| 4 | Modern NLP algorithms are based on machine learning, especially statistical machine learning.<br>a. True<br>b. False | L2 | CO1, PO1 |
| 5 | Choose from the following areas where NLP can be Useful.<br>a. Information retrieval<br>b. Automatic Text summarization<br>c. Automatic Question-Answering Systems<br>d. All of the above. | L2 | CO1, PO1 |
| 6 | What is machine translation?<br>a. Converts human language to machine language.<br>b. Converts one human language to other.<br>c. Converts any human language to English.<br>d. Converts Machine Language to human language. | L2 | CO1, PO1 |
| 7 | Information Retrieval and Information Extraction are the two same thing.<br>a. True<br>b. False | L2 | CO1, PO1 |

| | | | |
|---|---|---|---|
| 8 | How many steps are there in NLP?<br>a. 3<br>b. 4<br>c. 5<br>d. 6 | L2 | CO1, PO1 |
| 9 | "I saw bats" contains which type of ambiguity?<br>a. Syntactic<br>b. Semantic<br>c. Lexical<br>d. Anaphoric | L2 | CO1, PO1 |
| 10 | "Sita loves her mother and Gita does too" contain which type of ambiguity?<br>a. Syntactic<br>b. Semantic<br>c. Lexical<br>d. Anaphoric | L2 | CO1, PO1 |

## II. Internal Test (Answer any two full questions choosing one from each part) (Each full question carries 25 marks)

| Main Ques | SubQ ues | Full Question | Marks | BT Levels | CO -PO Outcom e |
|---|---|---|---|---|---|
| | | **Part A** | | | |
| 1. | A) | Explain different levels of NLP with example. | 10 | L2 | CO1, PO1 |
| | B) | Explain the components of transformational grammar with example. | 9 | L2 | CO1, PO1 |
| | C) | Write a note of different types of language models. | 6 | L2 | CO1, PO1 |
| | | **OR** | | | |
| 2. | A) | Explain the applications of NLP? | 8 | L2 | CO1, PO1 |
| | B) | Explain the different smoothing techniques to handle the data parseness problem in n-gram. | 9 | L2 | CO1, PO1 |
| | C) | Explain Lexical Functional Grammar. | 8 | L2 | CO1, PO1 |
| | | **Part B** | | | |
| 3. | A) | Write the c-structure and f-structure for the sentence:<br>"She saw stars". Given the rules<br>S → NP VP<br>VP → V {NP} {NP} PP* {S`}<br>PP → P NP<br>NP →Det N {PP}<br>S` → Comp S | 10 | L2 | CO1, PO1 |
| | B) | What is Morphological parsing. Explain 2 levels of morphological level with examples. | 9 | L2 | CO2, PO1 |
| | C) | What is NLP? Explain 2 major approaches to NLP. | 6 | L2 | CO1, PO1 |
| | | **OR** | | | |
| 4. | A) | What is Information Retrieval and explain the issues in Information Retrieval. | 9 | L2 | CO1, PO1 |
| | B) | Explain Finite State Automata for parsing word level analysis. | 9 | L2 | CO2, PO1 |
| | C) | Explain $\bar{X}$ theory with example. | 7 | L2 | CO1, PO1 |

| Comments: |
| --- |
| — Accepted. |

| Signature of Faculty | Scrutinizer | HOD |
| --- | --- | --- |

Sri AdichunchanagiriShikshana Trust (R)

# SJB Institute of Technology

(Affiliated to Visvesvaraya Technological University, Belagavi & Approved by AICTE, New Delhi.)

## Department of Information Science & Engineering

**SCHEMES & SOLUTIONS**

| | | |
|---|---|---|
| Internal Test: 1 | Internal Quiz: 1 | Academic Year: ODD / 2021-22 |
| Sub: Natural Language Processing | Sub-Code: 18CS743 | Sem: 7 |
| Date: 19/11/2021 | Time: 2:00 to 3:45 pm | Dur: 1:45 |
| Internal Test max marks: 50 | Internal Quiz max marks: 10 | |
| Staff-Incharge: Chetan R. | | |

**Comments:**

Accepted.

**Signature of Faculty:** **Signature of Scrutinizer:** **Signature of HOD:**

| Q. No. | | Marks Alloted |
|---|---|---|
| | MCQs. Answer. (1 mark each)<br>1. d 2. a 3. a 4. a 5. d 6. b 7. b 8. c<br>9. c 10. b | |
| | Part-A | |
| 1a. | Different Levels → Lexical, Syntactic, Semantic, discourse and pragmatic. Explain each level (2 marks each) 2x5=10 | 10 marks. |
| 1b. | Components of transformational grammar.<br>(1) Phrase structure grammar<br>(2) Transformational rules.<br>(3) Morphophonemic rules.<br>} 3x2 = 6 marks Example 3 marks. } 9 marks. | 9 marks. |
| 1c. | Types of language models. – Grammar-based and Statistical 3 marks each. | 6 marks. |
| | OR | |
| 2a. | Applications of NLP:<br>Machine Translation<br>Speech Recognition<br>Speech Synthesis<br>Natural Language Interfaces to Databases<br>Information Retrieval<br>Information Extraction<br>Question Answering.<br>Text Summarization<br>1 mark each | 8 marks. |

| Q. No. | | Marks Alloted |
|---|---|---|
| 2b. | Add-one Smoothing.<br>Good-Tueing Smoothing.<br>Cache Technique.<br>3 marks each. | 9 marks. |
| 2c. | Lexical Functional Grammar<br>C-structure and f-structure } → 2 marks each.<br>CFG rules. — 3 marks.<br>Consistency, Completness and Coherence } 3 marks. | 8 mark. |
| | Part B | |
| 3a. | C-structure → 5 marks.<br>f-structure → 5 marks. | 10 marks. |
| 3b. | Definition — 2 marks.<br>Surface level and lexical level → 3 marks each<br>example - 1 marks. | 9 mark. |
| 3c. | NLP definition — 2 marks.<br>Explain two approaches :- Rationist } 2 marks each.<br>Emphirisit. | 6 marks. |

| Q. No. | | Marks Alloted |
|---|---|---|
| 4a. | Information Retrieval Definition – 2 marks.<br>issues in IR explanation – 7 marks. | 9 marks. |
| 4b. | Finite state Automata — DFA, NFA } Explain with example 4½ each. | 9 mark. |
| 4c. | Explain $\bar{X}$ theory with example.<br>with NP – Noun Phrase<br>VP – Verb Phrase<br>AP – Adjective Phrase<br>PP – Preposition Phrase.<br>Sentence structure.<br>Projection<br>Maximal projection<br>} 1 mark each. | 7 marks. |

|| Jai Sri Gurudev ||
Sri AdichunchanagiriShikshana Trust ®

# SJB Institute of Technology

(A Constituent of BGS & SJB Group of Institutions and Hospitals)
BGS Health & Education City, Dr. Vishnuvardhan Road, Kengeri, Bengaluru-5

Affiliated to Visvesvaraya Technological University, Belagavi. Approved by AICTE, New Delhi. Accredited by NAAC, New Delhi with 'A' Grade. Recognized by UGC, New Delhi with 2(f) and 12(B).Certified by ISO 9001-2015

***DEPARTMENT OFINFORMATION SCIENCE AND ENGINEERING***

## INTERNAL ASSESSMENT PAPER (7th Sem)

| | | |
|---|---|---|
| Internal Test: 2 | Internal Quiz: 2 | Academic Year: ODD /2021-22 |
| Subject : Natural Language Processing | Sub-Code: 18CS743 | Sem: 7 |
| Date:28/12/2021 | Time:2:00 to 3:45pm | Dur:1.45 min |
| Internal Test max marks: **50** | Internal Quiz max marks: **10** | |
| Staff-Incharge:Chetan R | | |

## I. Quiz (Answer all multiple-choice question in first sheet of your answer book)

| Question No | Multiple choice question | BTlevel | CO Mapped |
|---|---|---|---|
| 1 | In linguistic morphology, _____ is the process for reducing inflected words to their root form.<br>a. Stemming<br>b. Rooting<br>c. Text-Proofing<br>d. Both a & b. | L2 | CO1, PO1 |
| 2 | Coreference Resolution is:<br>a. Anaphora Resolution<br>b. Given a sentence or larger chunk of text, determine which words ("mentions") refer to the same objects ("entities")<br>c. Both a & b<br>d. None of the above | L2 | CO1, PO1 |
| 3 | Solve the equation according to the sentence "I am planning to visit New York to attend International File Fare Festival."<br>A= (# of words with Noun as the part of speech tag)<br>B= (# of words with verb as the part of speech tag)<br>C=(# of words with frequency count greater than one)<br>What are the correct values of A, B and C?<br>a. 5,5,2<br>b. 5,5,3<br>c. 7,5,1<br>d. 7,4,1 | L2 | CO1, PO1 |
| 4 | Which of the following will be POS tagger output when the input sentence is "They refuse to permit"<br>a. [('They','PRP',('refuse','VBP'),('to','TO'),('permit','VB')]<br>b. [('They','NN',('refuse','VBP'),('to','TO'),('permit','VB')]<br>c. [('They','PRP',('refuse','NN'),('to','TO'),('permit','VB')]<br>d. [('They','PRP',('refuse','VBP'),('to','PRP'),('permit','VB')] | L2 | CO4, PO1 |

| | | | |
|---|---|---|---|
| 5 | In CFG, terminals mainly correspond to ……. While pre-terminals mainly correspond to …….<br>a. Characters in the language, POS tags<br>b. Words in the language, POS categorie<br>c. Words in the language, word relations<br>d. Lexemes, POS Tags | L2 | CO1, PO1 |
| 6 | HMM is used in ……. Phase of NLP.<br>a. Syntactic<br>b. Semantic<br>c. Lexical<br>d. Pragmatics | L2 | CO1, PO1 |
| 7 | Which of the following belongs to the open class group?<br>a. Noun<br>b. Prepositions<br>c. Determiners<br>d. Conjunctions | L2 | CO4, PO1 |
| 8 | …… is a group of words that may behave as a single unit or phrase.<br>a. Constituency<br>b. Grammatical Relation<br>c. Sub-categorization<br>d. Dependencies | L2 | CO4, PO1 |
| 9 | …….. tagger uses probabilistic and statistical information to assign tags to words.<br>a. Rule based<br>b. Stochastic tagger<br>c. Statistical Tagger<br>d. POS tagger | L2 | CO4, PO1 |
| 10 | "Buy books for children" which type of ambiguity exists in the above sentence?<br>a. Semantic<br>b. Syntactic<br>c. Lexical<br>d. Pragmatic | L2 | CO1, PO1 |

## II. Internal Test (Answer any two full questions choosing one from each part) (Each full question carries 25 marks)

| Main Ques | SubQ ues | Full Question | Marks | BT Levels | CO -PO Outcom e |
|---|---|---|---|---|---|
| | | **Part A** | | | |
| 1. | A) | Construct the parse tree for the sentence: "The girl plucked the flower with a long stick". Discuss the ambiguity arises from the parse tree. | 10 | L3 | CO1, PO1 |
| | B) | Explain the CYK parser algorithm. | 9 | L2 | CO1, PO1 |
| | C) | Discuss the disadvantages of probabilistic CFG. | 6 | L2 | CO1, PO1 |
| | | **OR** | | | |
| 2. | A) | Explain the minimum edit distance algorithm and compute the distance between "peaceful" and "paceful". | 12 | L2 | CO1, PO1 |
| | B) | Explain the top down and bottom up parsing with suitable example and mention the advantages and disadvantages. | 8 | L2 | CO1, PO1 |

| | | | | | |
|---|---|---|---|---|---|
| | C) | With example, explain the basic top down depth first algorithm. | 5 | L2 | CO1, PO1 |
| | | **Part B** | | | |
| 3. | A) | With a neat diagram, explain the learning framework architecture. | 9 | L2 | CO1, PO1 |
| | B) | Explain the following:<br>i) Knowledge Roles ii) Domain Knowledge | 8 | L2 | CO2, PO1 |
| | C) | Explain the shortest dependency path hypothesis. Show various shortest dependency path among the relations in the sentence "Jellisc created an atmosphere of terror in the camp by killing abusing and threatening the detainees" | 8 | L2 | CO1, PO1 |
| | | **OR** | | | |
| 4. | A) | With a neat diagram, explain the functional overview of InFact system. | 9 | L2 | CO1, PO1 |
| | B) | Write the short notes on:<br>i) The shortest path hypothesis. ii) Learning Dependency path. | 9 | L2 | CO2, PO1 |
| | C) | Explain the strategies used in active learning approach for acquiring the labels using the committee based classification scheme. | 7 | L2 | CO1, PO1 |

***Comments:***

Accepted

***Signature of Faculty*** ***Scrutinizer*** ***HOD***

Sri AdichunchanagiriShikshana Trust (R)

# SJB Institute of Technology

(Affiliated to Visvesvaraya Technological University, Belagavi & Approved by AICTE, New Delhi.)

## Department of Information Science & Engineering

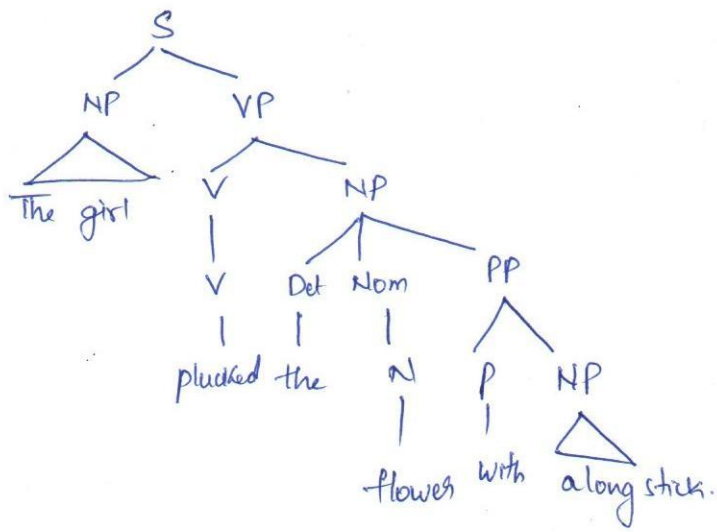
**SCHEMES & SOLUTIONS**

| | | | |
|---|---|---|---|
| Internal Test: 11 | Internal Quiz: 11 | Academic Year: ODD / 2021-22 | |
| Sub: NLP | Sub-Code: 18CS743 | | Sem: 7 |
| Date: 28/12/2021 | Time: 2:00 to 3:45pm. | | Dur: 1:45 |
| Internal Test max marks: 50 | | Internal Quiz max marks: 10 | |
| Staff-Incharge: Chetan R. | | | |

**Comments:**

- Accepted.

**Signature of Faculty:** **Signature of Scrutinizer:** **Signature of HOD:**

| Q. No. | | Marks Alloted |
|---|---|---|
| | Quiz:-<br>1. a<br>2. b<br>3. d<br>4. a<br>5. b<br>6. a<br>7. a<br>8. a.<br>9. b<br>10. b | 10 |
| | Part A | |
| 1a. |  <br>Attachment Ambiguity : There are two ways of generating prepositional phrase, with a long stick. | 10 |

| Q. No. | | Marks Alloted |
|---|---|---|

1b. CYK parser algorithm

Let $w = w_1 w_2 w_3 w_i \cdots w_j \cdots w_n$ and $w_{ij} = w_i \cdots w_{i+j-1}$

```
// initialization step
for i:=1 to n do
  for all rules A -> wi do
      chart [i,1] = {A}
// Recursive step
for j:=2 to n do
  for i=1 to n-j+1 do
  begin
    chart [i,j] = φ
    for k=1 to j-1 do
      chart [i,j] := chart[i,j] ∪ [A / A->BC is a production and
                    B ∈ chart [i,k] and C ∈ chart [i+k, j-k]}
  end
if S ∈ chart [1,n] then accept else reject.
```

Algorithm : 5 marks. Explanation : 4 marks — 9 marks.

1c. Disadvantages :-

1. Independence assumption
2. Lack of sensitivity to lexical information.

3 marks each. — 6 marks.

2a.

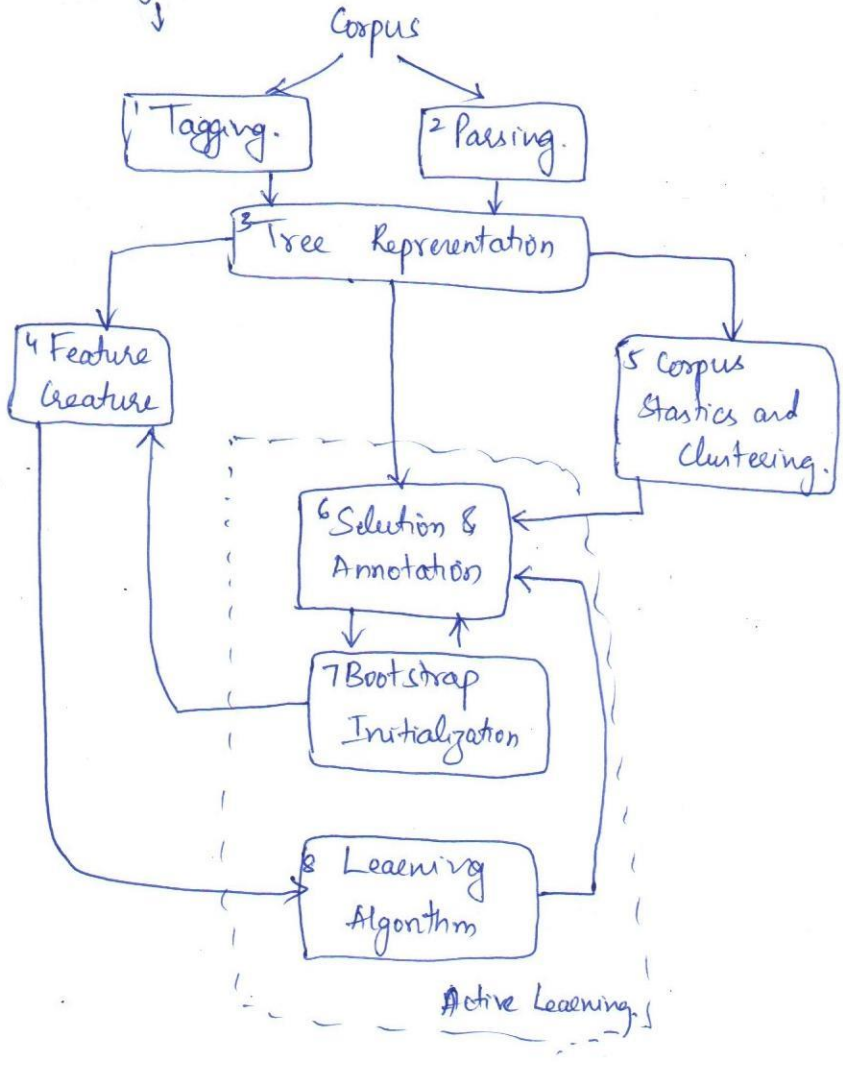
| | # | p | a | c | e | f | u | l |
|---|---|---|---|---|---|---|---|---|
| # | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| p | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| e | 2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| a | 3 | 2 | 1 | 2 | 3 | 4 | 5 | 6 |
| c | 4 | 3 | 2 | 1 | 2 | 3 | 4 | 5 |
| e | 5 | 4 | 3 | 2 | 1 | 2 | 3 | 4 |
| f | 6 | 5 | 4 | 3 | 2 | 1 | 2 | 3 |
| u | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 2 |
| l | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

Minimum Edit distance is 1

Algorithm – 6 marks.
problem – 6 marks
with explanation — 12 marks

2b. Top down parsing Explanation with example (4 marks)
Bottom up parsing Explanation with example (4 marks) — 8 marks.

| Q. No. | | Marks Alloted |
|---|---|---|
| 2c. | Basic top down depth first algorithm :- 3 marks explanation 2 marks. | 5 marks. |
| | **Part B** | |
| 3a. | Learning framework architecture<br>Diagram (4 marks.) Explanation 5 marks. | 9 marks |




| | | |
|---|---|---|
| 3b. | Domain Knowledge :- Domain of interest, Focus is on textual diagnostic reports ; Service provides and customers. (4 marks)<br>Domain Concepts and Domain specific terms. | 8 marks. |
| | Knowledge Roles :- Introduced in Common KADS.<br>Abstract names that refer to the role a domain concept. Each role has associated with Text phrase. (4 marks) | |

| Q. No. | | Marks Alloted |
|---|---|---|
| 3c. | Shortest path hypothesis<br>If e1 and e2 are two entities mentioned in the same sentence such that they are observed to be in a relationship R, then the contribution of the sentence dependency graph to establishing the relationship R(e1, e2) is almost exclusively concentrated in the shortest path between e1 and e2.<br>"Jelisic created an atmosphere of terror at the camp by killing absuring and threatning the detainees".<br>detainees → killing ← Jelisic → created ← at ← camp<br>detainees → absuring ← Jelisic → created ← at ← camp<br>detainees → threatning ← Jelisic → created ← at ← camp<br>detainees → killing → by → created ← at ← camp<br>detainees → absuring → by - created ← at ← camp<br>detainees → threatning → by - created ← at ← camp | 8 marks |
| 4a. | Infact system - diagram → 4 marks.<br>Explanation 5 marks | 9 marks |
| 4b. | The shortest path hypothesis → 4 marks<br>Learning Dependency path. → 5 marks. | |
| 4c. | Active Learning:-<br>a) Divide the corpus in clusters.<br>b) within each cluster, group the sentences.<br>c) Select sentences from largest groups.<br>d) Boostrap initialization<br>e) Train all the classifiers of the committee.<br>f) Get a pool of instances where the classifiers of the committee disagree and present to the user.<br>g) Repeat steps d)-f) a few times until a desired accuracy of classification is achieved. | 8 marks |

|| Jai Sri Gurudev ||
Sri AdichunchanagiriShikshana Trust ®

# SJB Institute of Technology

(A Constituent of BGS & SJB Group of Institutions and Hospitals)
BGS Health & Education City, Dr. Vishnuvardhan Road, Kengeri, Bengaluru-5

Affiliated to Visvesvaraya Technological University, Belagavi. Approved by AICTE, New Delhi. Accredited by NAAC, New Delhi with 'A' Grade. Recognized by UGC, New Delhi with 2(f) and 12(B).Certified by ISO 9001-2015

***DEPARTMENT OFINFORMATION SCIENCE AND ENGINEERING***

## INTERNAL ASSESSMENT PAPER (7thSem)

| | | |
|---|---|---|
| Internal Test: 3 | Internal Quiz: 3 | Academic Year: ODD /2021-22 |
| Subject : Natural Language Processing | Sub-Code: 18CS743 | Sem: 7 |
| Date:31/01/2022 | Time:1:00 to 2:45pm | Dur:1.45 min |
| Internal Test max marks: **50** | Internal Quiz max marks: **10** | |
| Staff-Incharge:Chetan R | | |

## I. Quiz (Answer all multiple-choice question in first sheet of your answer book)

| Question No | Multiple choice question | BTlevel | CO Mapped |
|---|---|---|---|
| 1 | The formula used to estimate the vocabulary size of a collection is known as:<br>a) Zipf's law b) Power law c) Heap's law d) Compression ratio | L2 | CO4, PO1 |
| 2 | IR Stands for_____.<br>a) Information Retrieval b) Information Retired c) Inform Retrieval d) Information Ready | L2 | CO4, PO1 |
| 3 | SEO stands for _____ .<br>a) Search English Optimization b) Search Engine Optimization c) Search Engine Operator d) Search Engine Operation | L2 | CO4, PO1 |
| 4 | The purpose of the inverse document frequency is to increase the weight of terms with high collection frequency. a) True b) False | L2 | CO4, PO1 |
| 5 | A data structure that maps terms back to the parts of a document in which they occur is called an<br>a) Postings list b) Incidence Matrix c) Dictionary d) Inverted Index | L2 | CO4, PO1 |
| 6 | The first large information retrieval research group was formed by_____at cornell in 1960.<br>a) Gerard Salton b) Ratan Tata c) Ramesh Bush d) Think Roy | L2 | CO4, PO1 |
| 7 | Pick the stemming actions<br>a. was, am, are, is → beb. helped, helps → help<br>c. troubled, troubling, trouble → trouble<br>d. friend, friendship, friends, friendships → friend | L2 | CO4, PO1 |
| 8 | The process of removing most common words (and, or, the, etc.) by an information retrieval system before indexing is known as<br>a) Lemmatizationb) Stop word removalc) Inverted indexing<br>d) Normalization | L2 | CO4, PO1 |
| 9 | An inverted index arranges data in a sorted order as per<br>a) the documentsb) the frequency of each document<br>c) the frequency of each termd) the terms | L2 | CO4, PO1 |
| 10 | Which of the following is a non-decreasing function of the number of documents retrieved?<br>a) precisionb) recallc) accuracyd) mean average precision | L2 | CO4, PO1 |

## II. Internal Test (Answer any two full questions choosing one from each part) (Each full question carries 25 marks)

| Main Ques | SubQ ues | Full Question | Marks | BT Levels | CO -PO Outcom e |
|---|---|---|---|---|---|
| | | **Part A** | | | |
| 1. | A) | Explain SVM Learning method in Sequence Model Estimation. | 8 | L3 | CO4, PO1 |
| | B) | Explain Latent Semantic Analysis feedback system. | 9 | L2 | CO4, PO1 |
| | C) | Define the following: i) Cohesion ii) Coverage iii) Interestingness iv) Plausibility of Origin | 8 | L2 | CO4, PO1 |
| | | **OR** | | | |
| 2. | A) | Write a note on various approaches to analyzing texts. | 10 | L2 | CO4, PO1 |
| | B) | With a neat diagram explain the evolutionary model for KDT. | 10 | L2 | CO4, PO1 |
| | C) | Define: i) coh-metrix ii) LSI | 5 | L2 | CO4, PO1 |
| | | **Part B** | | | |
| 3. | A) | Explain design features of IR with a neat diagram | 9 | L2 | CO4, PO1 |
| | B) | With an example explain Boolean Model for Classical Information Model | 8 | L2 | CO4, PO1 |
| | C) | Write Short note on: i) FRAME NET ii) Stemmer | 8 | L2 | CO4, PO1 |
| | | **OR** | | | |
| 4. | A) | Explain WORD NET. List the applications | 9 | L2 | CO4, PO1 |
| | B) | Explain non-classical model of IR | 6 | L2 | CO4, PO1 |
| | C) | How stemming effects the performance of IR Systems? Stop words elimination may be harmful, Justify | 10 | L2 | CO4, PO1 |

*Comments:*

Accepted.

*Signature of Faculty* *Scrutinizer* *HOD*

Sri AdichunchanagiriShikshana Trust (R)

# SJB Institute of Technology

(Affiliated to Visvesvaraya Technological University, Belagavi & Approved by AICTE, New Delhi.)

## Department of Information Science & Engineering

**SCHEMES & SOLUTIONS**

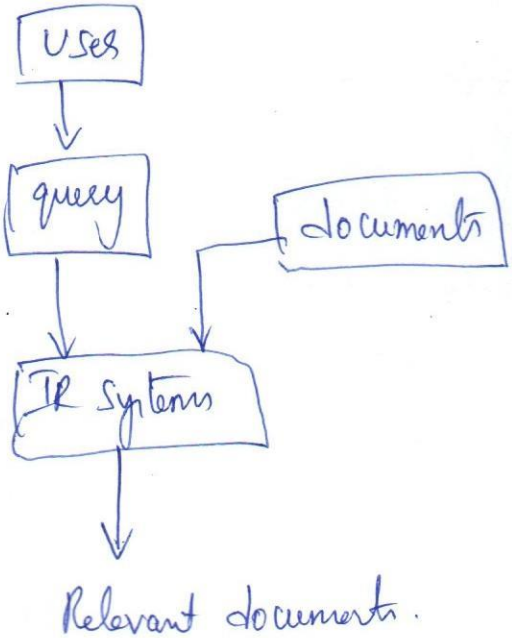
| | | |
|---|---|---|
| Internal Test: II | Internal Quiz: II | Academic Year: ODD / 2021-22 |
| Sub: Natural Language Processing | Sub-Code: 18CS743 | Sem: 7th A and B |
| Date: 31/01/2022 | Time: 1:00pm to 2:45pm. | Dur: 1:45 minutes. |
| Internal Test max marks: 50 | Internal Quiz max marks: 10 | |
| Staff-Incharge: Chetan R. | | |

**Comments:**

– Accepted.

**Signature of Faculty:** **Signature of Scrutinizer:** **Signature of HOD:**

| Q. No. | | Marks Alloted |
|---|---|---|
| Quiz. | 1. a 2. a 3. b 4. b 5. d 6. a 7. b 8. b 9. d 10. b. | |
| | Part A | |
| 1a. | SVM Learning method. Explanation. → 4 marks. Sequence Model Estimation. → 4 marks. | 8. |
| 1b. | Latent Semantic Analysis Feedback System. LSA → Explanation 4 marks. Feedback System – 5 marks. | 9. |
| 1c. | Define. i) coherion ii) Coverage iii) Interestingness iv) Plausibility of Origin each (2 marks) | 8. |

| Q. No. | | Marks Alloted |
|---|---|---|
| 2a. | Various approaches to analyzing texts.<br>a) Traditional approach<br>b) Function property. (each - 2 marks)<br>c). Narrative Theory.<br>d). LSA.<br>e) Homogeneous. | 10 |
| 2b. | Evolutionary model for KDT<br>Diagram → 6 marks.<br>Explanation → 4 marks. | 10. |
| 2c. | Define :- Coh-metrix<br>LSI<br>each (2.5 marks) | 5 |
| 3a) |  <br>Diagram 3marks.<br>Explanation 6marks. | 9. |

| Q. No. | | Marks Alloted |
|---|---|---|
| 3b. | Boolean Model → AND, OR, NOT → 4 marks.<br>Equation. $Q = \wedge(\vee \theta_i)$ → 2 marks.<br>(i) $R_i \{d_j \mid \theta_i \in d_j)$ where $\theta_i = \{t_i \neg t_i)$<br>$\neg t_i \in d_j$<br>(ii). $\cap R_i$<br>Example - 2 marks | 8. |
| 3c. | (i) FRAME NET Explanation. each (4 marks.)<br>(ii) Stemmer. | 8. |
| 4a. | WORD NET → Explanation. 5 marks.<br>Applications<br>1. WORDNET → Automatic text classification<br>Automatic text Summarization.<br>Machine Translation. 4 marks.<br>NLTP<br>Text word disambiguient. | 9. |

| Q. No. | | Marks Alloted |
|---|---|---|
| 4b). | Non-classical model of IR.<br><br>1. Information Logic Model<br>2. Situation theory Model. 2 marks each.<br>3. Interaction model. | 6 Mar. |
| 4c. | Stemming affects the performance of IR System<br>Explanation with example - 5 mark.<br><br>Stop words eliminitation may be harmful - 5 mark<br>Justification | 10 Marks |

||Jai Sri Gurudev||
Sri AdichunchanagiriShikshana Trust (R)
# SJB Institute of Technology
Affiliated to Visvesvaraya Technological University, Belagavi & Approved by AICTE, New Delhi
Accredited by NAAC, New Delhi with 'A' Grade, Recognized by UGC, New Delhi with 2(f) and 12(B)
Certified by ISO 9001- 2015
**BGS Health & Education City, Dr. Vishnuvardhan Road, Kengeri, Bengaluru – 560 060**
Website:www.sjbit.edu.in, Email :principal@sjbit.edu.in, Mob:**+91-6366041109**

## Department of Information Science and Engineering

**Course Name: NATURAL LANGUAGE PROCESSING** **Course Code: 18CS743**

### MODULE 1:

| Q. No. | Question | CO Mapped | PO Mapped | Blooms Level |
|---|---|---|---|---|
| 1. | What is NLP? Explain two major approaches to NLP. | CO1 | PO1, PO2 | L2 |
| 2. | Explain the components of transformational grammar. | CO1 | PO1, PO2 | L2 |
| 3. | Explain different levels of NLP with example. | CO1 | PO1, PO2 | L2 |
| 4. | Explain different smoothing techniques to handle the data sparseness problem in n-grain model. | CO1 | PO1, PO2 | L2 |

### MODULE 2:

| Q. No. | Question | CO Mapped | PO Mapped | Blooms Level |
|---|---|---|---|---|
| 1. | What is Morphological Parsing? Explain the two step of Morphological parser. | CO2 | PO1, PO2, PO3 | L2 |
| 2. | Explain spelling correction algorithm. | CO2 | PO1, PO2, PO3 | L2 |
| 3. | With example explain basic top down depth first algorithm | CO2 | PO1, PO2, PO3 | L2 |
| 4. | Explain CYK algorithm. | CO2 | PO1, PO2, PO3 | L2 |

### MODULE 3:

| Q. No. | Question | CO Mapped | PO Mapped | Blooms Level |
|---|---|---|---|---|
| 1. | With neat diagram explain functional overview of InFact System. | CO2 | PO1, PO2, PO3 | L2 |
| 2. | Write a short note on:<br>i) The shortest path hypothesis.<br>ii) Learning with dependency path. | CO1 | PO1, PO2 | L2 |
| 3. | With neat diagram explain the learning framework architecture. | CO1 | PO1, PO2 | L2 |
| 4. | Explain the following i) Domain Knowledge ii) Knowledge roles. | CO1 | PO1, PO2 | L2 |

**MODULE 4:**

| Q. No. | Question | CO Mapped | PO Mapped | Blooms Level |
|---|---|---|---|---|
| 1. | Explain SVM learning method in Sequence Model estimation. | CO3 | PO1, PO2, PO3 | L2 |
| 2. | Explain Latent Semantic Analysis feedback system. | CO3 | PO1, PO2, PO3 | L2 |
| 3. | Define the following: i) Cohesion ii) Interestingness iii) Coverage iv) Plausibility of origin. | CO3 | PO1, PO2, PO3 | L2 |

**MODULE 5:**

| Q. No. | Question | CO Mapped | PO Mapped | Blooms Level |
|---|---|---|---|---|
| 1. | State and explain Zipf's Law. | CO4 | PO1, PO2 | L2 |
| 2. | Explain Non-classical model of IR | CO4 | PO1, PO2 | L2 |
| 3. | With example explain Boolean model for classical information retrieval. | CO4 | PO1, PO2 | L2 |

**In-Charge**

**[CHETAN R]**

**HOD**
**Dr. Mohan H S**
**[Professor & Head, ISE]**