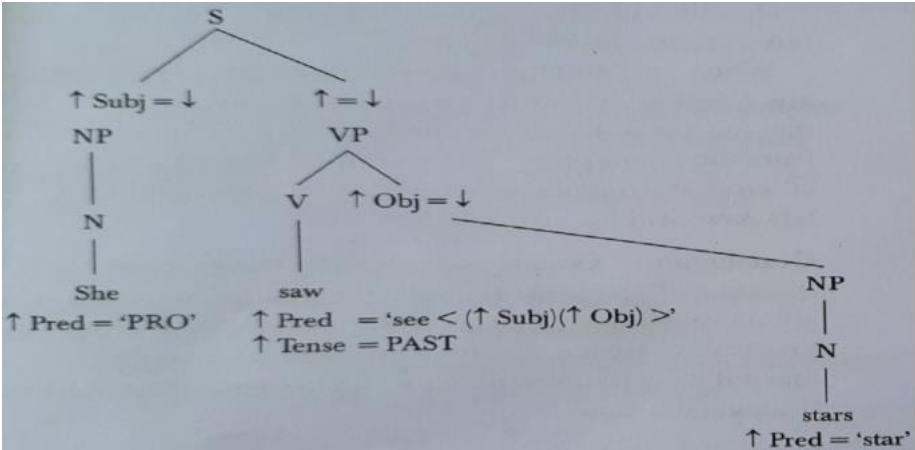
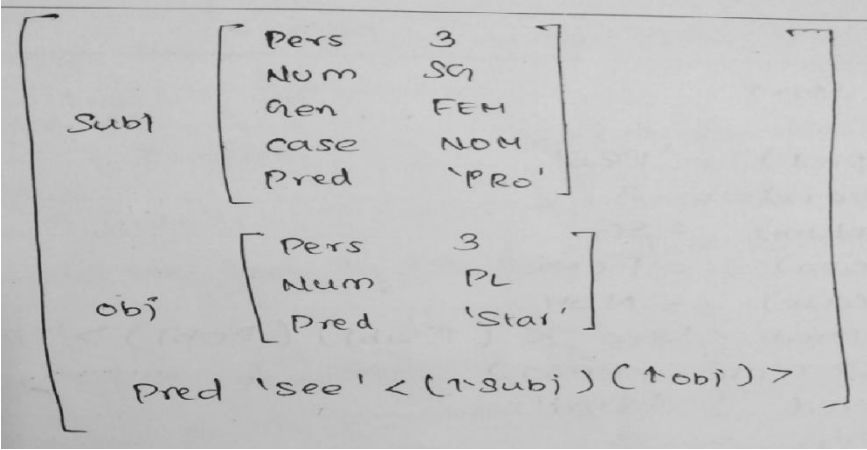
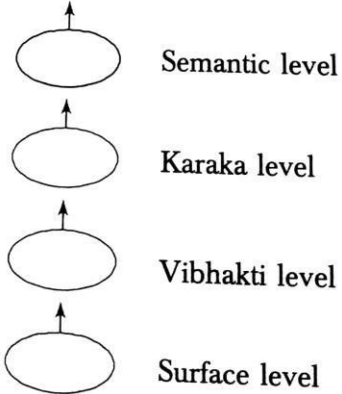


Internal Assessment Test 1 – October 2022

Sub:	Natural Language Processing				Sub Code:	18CS743	Branch:	ISE		
Date:	21/10/2022	Duration:	90 min's	Max Marks:	50	Sem/Sec:	VII A, B & C		OBE	
Answer any FIVE FULL Questions								MARKS	CO	RBT
1.	Outline Application and challenges of NLP							[10]	CO1	L2
<p><u>Applications:</u></p> <p>The application utilizing NLP includes the following.</p> <p>Machine Translation</p> <ul style="list-style-type: none">• Automatic translation of text from one human language to another.• It requires to understand the words and phrases, grammars, semantics and world knowledge. <p>Speech Recognition</p> <ul style="list-style-type: none">• Mapping acoustic speech signals to a set of words.• Difficulties arises due to• Variations in the pronunciation of words, homonym• Acoustic ambiguities <p>Speech Synthesis</p> <ul style="list-style-type: none">• Automatic production of speech(utterance of sentences)• Can read your mails on telephone, or story book for you.• To generate utterances, text has to be processed.• NLP remains an important component of any speech synthesis system.• Natural language Interfaces to Databases• Allow querying a structured database using natural language sentences. <p>Information Retrieval</p> <ul style="list-style-type: none">• Identifying documents relevant to a user's query.• Indexing, word sense disambiguation, query modification have also been used in IR system to enhance performance.• WordNet, LDOCE and Roget's Thesaurus are some of the useful resources for IR research. <p>Information Extraction</p> <ul style="list-style-type: none">• Captures and outputs factual information contained within a document.• Responds to user's information.• Specified as pre-defined database schemes or templates.• It Identifies a subset of information within a document that fits the pre-defined template. <p>Question Answering</p> <ul style="list-style-type: none">• Attempts to find precise answer, or precise portion of text in which the answer appears.• Returns whole document that seems relevant to the user's query.• Requires precise analysis of questions and portions of texts, as well as background knowledge to answer certain type of questions. <p>Text Summarization</p> <ul style="list-style-type: none">• Deals with the creation of summaries of documents and involves syntactic and semantic of text. <p><u>Challenges:</u></p> <ul style="list-style-type: none">• Representing & interpreting NL is a challenging task.• Natural languages are highly ambiguous and vague, achieving such representation can be difficult.• It is almost impossible to embody all sources of knowledge that humans use to process language.• Identifying the semantics in natural language is difficult.• Words alone do not make sentence. It is their syntactic and semantic relation that give meaning to a sentence.• A language keeps on evolving.										

2.	Construct the C- structure and f-structure for the following sentence “She saw stars”. Consider LFG rule.	[10]	CO1	L3
	<p>LFG provides well-defined objects called</p> <ul style="list-style-type: none"> constituent structure : It is derived from the usual phrase and sentence structure syntax. functional structure : when Functional specifications are applied to c-structure it results in f-structure. <p>up arrow: refers to f-structure of mother node that is on left hand side of the rule. down arrow: refers to f-structure of node under which it is denoted.</p>  <p>Ex- She saw stars- C- Structure</p>  <p>Ex- She saw stars- F- Structure</p> <ul style="list-style-type: none"> f-structure is the set of attribute-value pairs, represented as above final f-structure is obtained through unification of various f-structures for subject, object , verb, complement etc... 			

3.	Identify different Karaka’s in following sentence in hindi language (any sentence could be given) maan Bachche ko aangan mein haath se rotii khilaatii hei	[10]	CO1	L3
----	--	------	-----	----

	<p>Levels of Paninian Grammar:</p> 			
--	--	--	--	--

	<p>Karaka literally means CASE, these case relations are based on the way the word group participates in the activity denoted by the verb group. Karaka relations are assigned based on the roles players by various participants in main activity.</p> <p>Various karaka's are (case marker in hindi)</p> <ol style="list-style-type: none"> 1. Karta (subject) - maan 2. Karma (Object) - rotii 3. Karana (instrument)- haath 4. Sampradana (beneficiary)- bacheche 5. Apadan (separation)- ko, se/dwara, ke (Case marker) 			
4.	<p>Solve to find the probability of test sentence S2 in the following training set</p> <p>S1: The Arabian Knights S2: These are the fairy tales of the east S3: The stories of the Arabian knights are translated in many languages</p>	[10]	CO1	L3
	<p>Bi-gram model:</p> <p>$P(\text{the}/\langle s \rangle) = 0.67$ $P(\text{Arabian}/\text{the}) = 0.4$ $P(\text{knights}/\text{Arabian}) = 1.0$</p> <p>$P(\text{are}/\text{these}) = 1.0$ $P(\text{the}/\text{are}) = 0.5$ $P(\text{fairy}/\text{the}) = 0.2$ $P(\text{tales}/\text{fairy}) = 1.0$ $P(\text{of}/\text{tales}) = 1.0$ $P(\text{the}/\text{of}) = 1.0$ $P(\text{east}/\text{the}) = 0.2$</p> <p>$P(\text{stories}/\text{the}) = 0.2$ $P(\text{of}/\text{stories}) = 1.0$ $P(\text{are}/\text{knights}) = 1.0$ $P(\text{translated}/\text{are}) = 0.5$ $P(\text{in}/\text{translated}) = 1.0$ $P(\text{many}/\text{in}) = 1.0$ $P(\text{languages}/\text{many}) = 1.0$</p> <p>Test sentence(s): The Arabian knights are the fairy tales of the east.</p> <p>$P(\text{The}/\langle s \rangle) \times P(\text{Arabian}/\text{the}) \times P(\text{Knights}/\text{Arabian}) \times P(\text{are}/\text{knights}) \times P(\text{the}/\text{are}) \times P(\text{fairy}/\text{the}) \times P(\text{tales}/\text{fairy}) \times P(\text{of}/\text{tales}) \times P(\text{the}/\text{of}) \times P(\text{east}/\text{the})$ $= 0.67 \times 0.4 \times 1.0 \times 1.0 \times 0.5 \times 0.2 \times 1.0 \times 1.0 \times 1.0 \times 0.2$ $= 0.0067$</p>			
5.	<p>Explain Character classes with examples.</p> <ul style="list-style-type: none"> • Characters are grouped by putting them between square brackets. <code>/[abcd]/</code> • Any character in the class will match one character in the input. <p>Example: the pattern <code>/[abcd]/</code> will match a, b, c, d.</p> <ul style="list-style-type: none"> • Use of brackets specifies a disjunction of characters. • The character classes are important building blocks in expressions. <p>Example: It is inconvenient to write the regular expression <code>/[abcdefghijklmnopqrstuvwxyz]/</code> to specify 'any lowercase letter'.</p> <ul style="list-style-type: none"> • A dash is used to specify a range. <p>Example: <code>/[5-9]/</code> specifies any one of the characters 5, 6, 7, 8, or 9.</p> <ul style="list-style-type: none"> • Regular expressions can also specify what a single character cannot be. it uses caret at the beginning. <p>Example:</p>	[10]	CO2	L2

`/[^x]/` matches any single character except x.

- Regular expressions are case sensitive.

Example:

`/s/` matches lower case 's' but not uppercase 'S'.

- The pattern `/[sS]/` will match the string containing either s or S.
- A solution is needed to specify both 'supernova' and 'supernovas'.
- The pattern `/[sS]supernova[sS]/` does not match with the string 'supernova'.
- This is achieved with the use of a question mark `/?/`.
- A question mark makes the preceding character optional, i.e., zero or one occurrence of the previous character.
- The regular expression `/supernovas?/` specifies both 'supernova' and 'supernovas'.
- The * operator called Kleene * specify repeated occurrences of a character.
- The * operator specifies zero or more occurrences of a preceding character or regular expression.
- The regular expression `/b*/` will match any string containing zero or more occurrences of 'b'.
- It will also match 'aaa', since that string contains zero occurrences of 'b'.
- To match a string containing one or more 'b's the regular expression is `/bb*/`.
- This means 'b' followed by zero or more 'b's.
- The regular expression `/[ab]*/` specifies zero or more 'a's or 'b's.
- This will match strings like 'aa', 'bb', or 'abab'.
- The **Kleene+** provides one or more occurrence of a character.
- Using Kleene+, we can specify a sequence of digits by the regular expression `/[0-9]*/`.
- The **caret (^)** is used to specify a match at the beginning of a line.
- The **dollar sign (\$)** is used to specify a match at the end of the line.
- If you want to search the line containing only the phrase 'the nature'

`/^the nature\.$/` --> this expression will search exactly only this line.

To check if string is an email address or not

`^[A-Za-z0-9_\. -]+ @[A-Za-z0-9_\. -]+[A-Za-z0-9_][A-Za-z0-9_]*$`

`^[A-Za-z0-9_\. -]+` Match a positive number of acceptable characters at beginning of the string.

`@` Match the @ sign

`[A-Za-z0-9_\. -]+` Match any domain name, including a dot

`[A-Za-z0-9_][A-Za-z0-9_]*$` Match two acceptable characters but not a dot.

This ensures that the email address ends with .xx, .xxx, .xxxx, etc.

6 Construct DFA if $\Sigma=\{a,b,c\}$, the set of states= $\{q_0,q_1,q_2,q_3,q_4\}$, q_0 being the start state and q_4 the final state. Rules of transition are as follows:

- From state q_0 & with input a, go to state q_1 .
- From state q_1 & with input b, go to state q_2 .
- From state q_1 & with input c, go to state q_3 .
- From state q_2 & with input b, go to state q_4 .
- From state q_3 & with input b, go to state q_4 .

[10]

CO2

L3

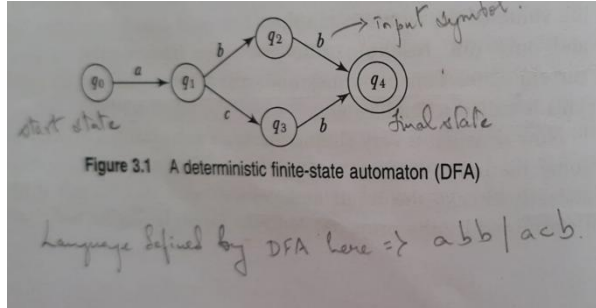


Fig: A deterministic finite-state automaton (DFA)

	Input		
State	a	b	c
Start: q_0	q_1	ϕ	ϕ
q_1	ϕ	q_2	q_3
q_2	ϕ	q_4	ϕ
q_3	ϕ	q_4	ϕ
Final: q_4	ϕ	ϕ	ϕ

Fig: The state-transition table for the DFA

- Any regular expression can be represented by a finite automaton.
- For each state and symbol there is at most one transition possible.
- This FSA is shown as a directed graph, called transition diagram, in Figure.
- The nodes in this diagram correspond to the states, and the arcs to transitions.
- The arcs are labeled with inputs.
- The final state is represented by a double circle.

If exactly 1 transition will lead out of each state, then the automaton is deterministic

Faculty Signature

CCI Signature

HOD Signature