**Q.1 Write a note on the following: CO4 (20)**

**i. Learning with complete data**

**ii. Speech recognition**

**iii. Information retrieval.**

**Ans)**

**i) Learning with complete data: [Not in syllabus mostly]**

[**http://www.faadooengineers.com/online-study/post/cse/artificial-intelligence/161/learning-with-complete-data**](http://www.faadooengineers.com/online-study/post/cse/artificial-intelligence/161/learning-with-complete-data)

**ii) Speech Recognition:**

* + Speech recognition is the task of identifying a sequence of words uttered by a speaker, given the acoustic signal.
  + It has become one of the mainstream applications of AI—millions of people interact with speech recognition systems every day to navigate voice mail systems, search the Web from mobile phones, and other applications.
  + Speech is an attractive option when hands-free operation is necessary, as when operating machinery.
  + Speech recognition is a difficult task because the sounds made by a speaker are **ambiguous** and **noisy**.
  + As a well known example, the phrase **“recognize speech”** sounds almost the same as **“wreck a nice beach”** when spoken quickly.
  + Even this short example shows several of the issues that make speech problematic.
  + First, **Segmentation**: written words in English have spaces between them, but in fast speech there are no pauses in “wreck a nice” that would distinguish it as a multi word phrase as opposed to the single word “recognize”.
  + Second, **Coarticulation**: when speaking quickly the “s” sound at the end of “nice” merges with the “b” sound at the beginning of “beach”, yielding something that is close to a “sp”.
  + Another problem is **homophones**: words like “to”, “too” and “two” that sound the same but differ in meaning.

**iii) Information retrieval:**

Information retrieval is the task of finding documents that are relevant to a user’s need for

information. The best-known examples of information retrieval systems are search engines

on the World Wide Web.

**Eg:** A Web user can type a query such as [AI book] into a search engine and see a list of relevant pages.

An IR information retrieval (henceforth IR) system can be characterized by

* **A corpus of documents:** Each system must decide what it wants to treat as a document:a paragraph, a page, or a multi page text.
* **Queries posed in a query language:**  A query specifies what the user wants to know.

The query language can be just a list of words, such as [AI book]; or it can specify

a phrase of words that must be adjacent, as in [“AI book”]; it can contain Boolean

operators as in [AI AND book]; it can include non-Boolean operators such as [AI NEAR

book] or [AI book site:www.aaai.org].

* **A result set:** This is the subset of documents that the IR system judges to be relevant to

the query. By relevant, we mean likely to be of use to the person who posed the query,

for the particular information needed expressed in the query.

* **A presentation of the result set:** This can be as simple as a ranked list of document

titles or as complex as a rotating color map of the result set projected onto a three

dimensional space, rendered as a two-dimensional display.

**Q.2 Write a note on the following: CO4 (20)**

**i Natural language processing**

**ii. Syntactic Analysis**

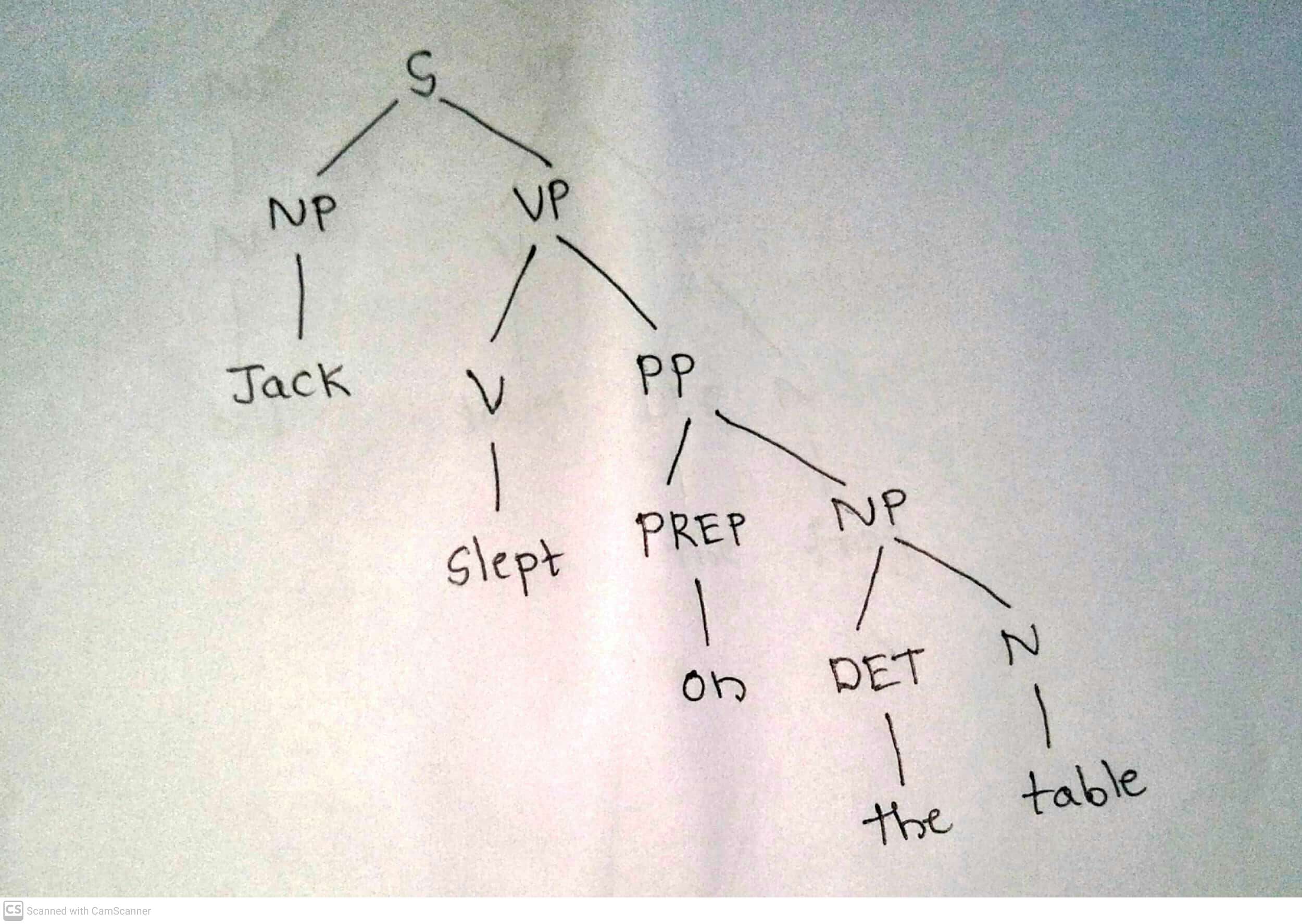
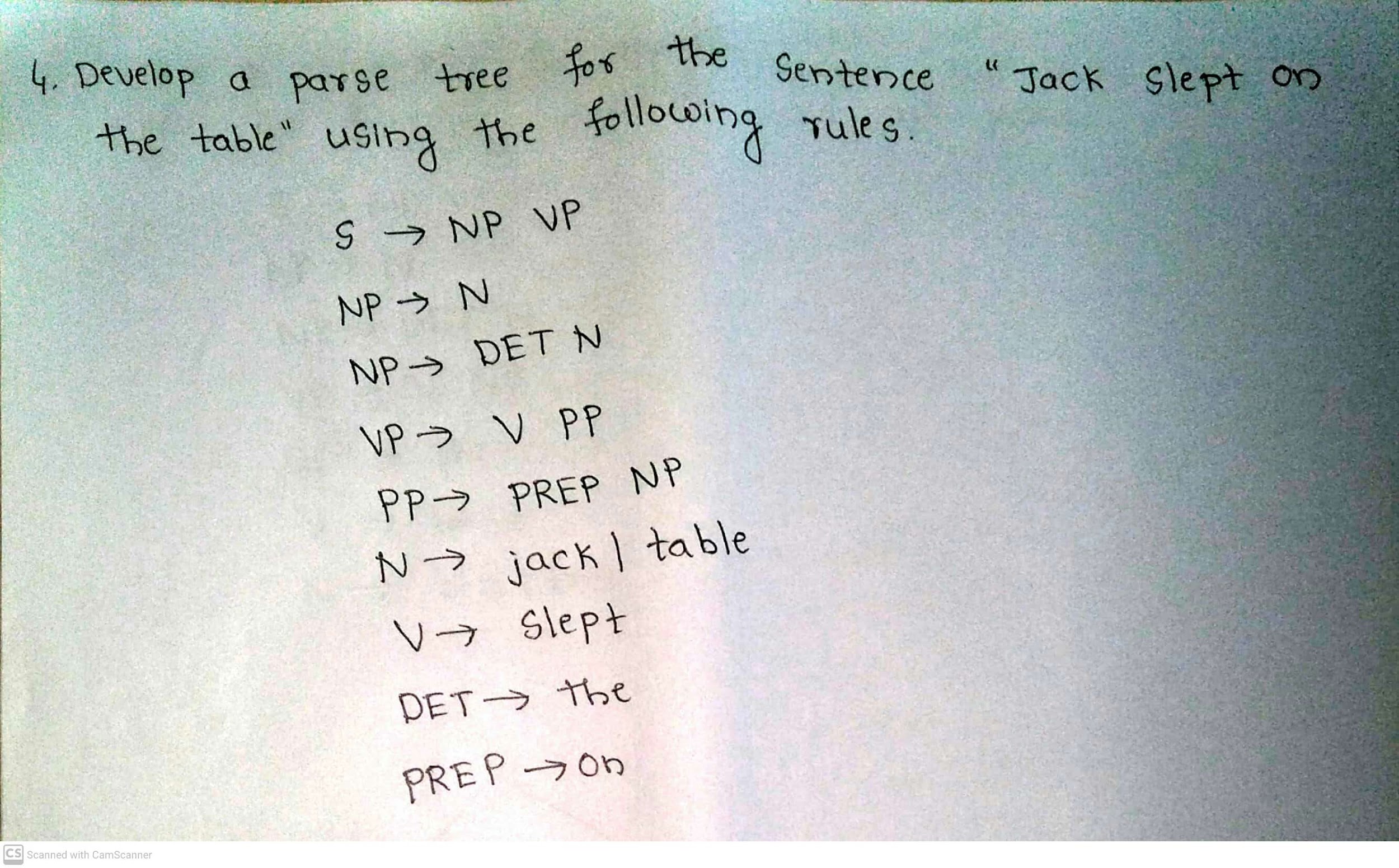
**iii. Machine translation.**

**Ans)**

**i) Natural Language processing:** Natural language processing (NLP) refers to the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

**ii) Syntactic Analysis:** Parsing is the process of analyzing a string of words to uncover its phrase structure, according to the rules of a grammar.

Eg: Figure shows that we can start with the S symbol and search top down for a tree that has the words as its leaves, or we can start with the words and search bottom up for a tree that culminates in an S.



**iii) Machine translation:**

Machine translation is the automatic translation of text from one natural language (the source) to another (the target).

There have been three main applications of machine translation.

1. **Rough translation**, as provided by free online services, gives the “gist” of a foreign sentence or document, but contains errors.
2. **Pre-edited translation** is used by companies to publish their documentation and sales materials in multiple languages. The original source text is written in a constrained language that is easier to translate automatically, and the results are usually edited by a human to correct any errors.
3. **Restricted-source translation** works fully automatically, but only in highly stereotypical language, such as a weather report.

The problem is that different languages categorize the world differently. For example, the French word “doux” covers a wide range of meanings corresponding approximately to the English words “soft,” “sweet,” and “gentle.”A representation language that makes all the distinctions necessary for a set of languages is called an **interlingua**.

A translator (human or machine) often needs to understand the actual situation described in the source, not just the individual words. As an example, to translate *“The baseball hit the window. It broke.”* into French, we must choose the feminine “elle” or the masculine “il” for “it,” so we must decide whether “it” refers to the baseball or the window. To get the translation right, one must understand physics as well as language.

Sometimes there is no choice that can yield a completely satisfactory translation. For

example, an Italian love poem that uses the masculine “il sole” (sun) and feminine “la luna” (moon) to symbolize two lovers will necessarily be altered when translated into German, where the genders are reversed, and further altered when translated into a language where the genders are the same.

**Q.3)Define Information extraction.List any four information extraction techniques in NLP and elaborate on the probabilistic model for sequences with hidden state. CO4 (10)**

**Definition -** Information extraction is the process of acquiring knowledge by skimming a text and looking for occurrences of a particular class of object and for relationships among objects.

Information Extraction techniques

1. Finite-state automata for information extraction
2. Probabilistic models for information extraction
3. Conditional random fields for information extraction
4. Ontology extraction from large corpora
5. Automated template construction
6. Machine reading

**Probabilistic model**

The simplest probabilistic model for sequences with hidden state is the hidden Markov model, or HMM.HMMs have two big advantages over FSAs for extraction.

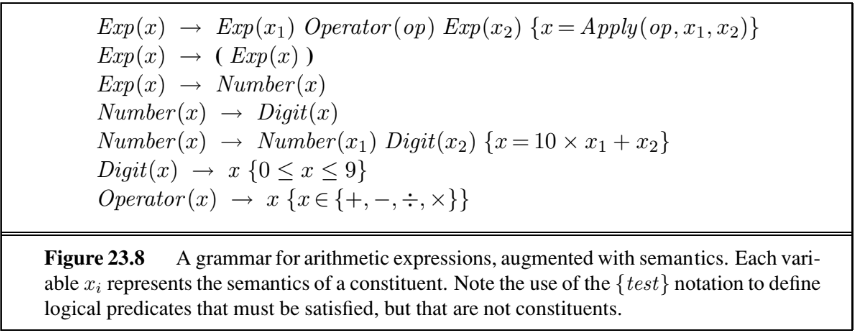
First, HMMs are probabilistic, and thus tolerant to noise. In a regular expression, if a single expected character is missing, the regex fails to match; with HMMs there is graceful degradation with missing characters/words, and we get a probability indicating the degree of match, not just a Boolean match/fail.

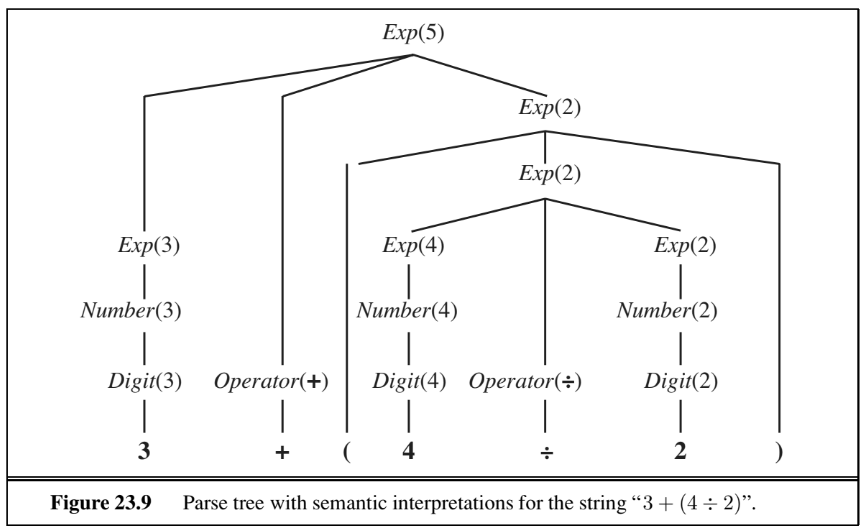
Second, HMMs can be trained from data; they don’t require laborious engineering of templates, and thus they can more easily be kept up to date as text changes over time.

**Q.4) What do you understand by semantic interpretation in Natural language communication? Write the grammar for arithmetic expressions, augmented with semantics and draw the parse tree with semantic interpretations for the string**

**“3 + (4 ÷ 2)”. CO4 (10)**

To show how to add semantics to a grammar, we start with an example that is simpler than English: the semantics of arithmetic expressions. Figure 23.8 shows a grammar for arithmetic expressions, where each rule is augmented with a variable indicating the semantic interpretation of the phrase. The semantics of a digit such as “3” is the digit itself. The semantics of an expression such as “3 + 4” is the operator “+” applied to the semantics of the phrase “3” and the phrase “4.” The rules obey the principle of **compositional semantics**—the semantics of a phrase is a function of the semantics of the sub phrases. Figure 23.9 shows the parse tree for 3 + (4 ÷ 2) according to this grammar. The root of the parse tree is Exp(5), an expression whose semantic interpretation is 5.

****

****

**Q.5) Discuss how the process of text classification is handled in NLP. How do you classify an email as a spam or not-spam (Ham), explain with an example.**

**CO4 (10)**

Ans)

• **Text classification** is also known as categorization: given a text of some kind, decide which of

a predefined set of classes it belongs to.

• Language identification , genre classification, sentiment analysis (classifying a movie or product review as positive or negative) and spam detection (classifying an email message as spam or ham) are examples of text classification

• Sentiment Analysis is the process of determining whether a piece of writing is positive,

negative or neutral.

**E.g. 1: Amit is happy.**

**E.g. 2: Akshay is frustrated.**

• We can treat spam detection as a problem in supervised learning.

• A training set is readily available: the positive (spam) examples are in our spam folder, the

negative (ham) examples are in our inbox.

**Eg:**

**spam:** Wholesale Fashion Watches – 57% today. Designer watches for cheap ...

**spam:** WE CAN TREAT ANYTHING YOU SUFFER FROM JUST TRUST US ...

**ham:** Good to see you my friend. Hey Peter, It was good to hear from you ...

**ham:** Abstract: We will motivate the problem of social identity clustering.

There are two ways in text classification

**• Language Modeling Approach**

**• Machine Learning Approach**

In the **language-modeling approach**, we define one n-gram language model for

**P(Message | spam)** by training on the spam folder, and one model for **P(Message | ham)** by training on the inbox.

Then we can classify a new message with an application of Bayes’ rule:

**argmax c∈{spam,ham} P(c | message) = argmax c∈{spam,ham} P(message | c) P(c) .**

where P(c) is estimated just by counting the total number of spam and ham messages. This approach works well for spam detection, just as it did for language identification.

In the **machine-learning approach** we represent the message as a set of feature/value

pairs and apply a classification algorithm h to the feature vector X. We can make the language-modeling and machine-learning approaches compatible by thinking of the n-grams

as features. In a unigram model, the features are the words in the vocabulary and the values are the number of times each word appears in the message. That makes the feature vector large and sparse. If there are 100,000 words in the language model, then the feature vector has length 100,000, but for a short email message almost all the features will have count zero. This unigram representation has been called the bag of words model.The notion of order of the words is lost;

A unigram model gives the same probability to any permutation of a text. Higher-order n-gram

models maintain some local notion of word order.

With bigrams and trigrams the number of features is squared or cubed, and we can add

in other, non-n-gram features. The choice of features is the most important part of creating a good spam detector because there is a lot of training data, so if we can propose a feature, the data can accurately determine if it is good or not. It is necessary to constantly update features, because spam detection is an adversarial task; the spammers modify their spam in response to the spam detector’s changes.

***{Answer can be shortened as per convenience}***

**Q.6) What are machine translation systems? Discuss the application of the same in the present day scenario. CO4 (05)**

**( Refer Q 16)**

**Q.7) Define information retrieval process in NLP. How are information**

**retrieval systems characterized?**

Information retrieval is the task of finding documents that are relevant to a user’s need for

information. The best-known examples of information retrieval systems are search engines

on the World Wide Web.

**Eg:** A Web user can type a query such as [AI book] into a search engine and see a list of relevant pages.

An IR information retrieval system can be characterized by:

* **A corpus of documents:** Each system must decide what it wants to treat as a document:a paragraph, a page, or a multi page text.
* **Queries posed in a query language:**  A query specifies what the user wants to know.

The query language can be just a list of words, such as [AI book]; or it can specify

a phrase of words that must be adjacent, as in [“AI book”]; it can contain Boolean

operators as in [AI AND book]; it can include non-Boolean operators such as [AI NEAR

book] or [AI book site:www.aaai.org].

* **A result set:** This is the subset of documents that the IR system judges to be relevant to

the query. By relevant, we mean likely to be of use to the person who posed the query,

for the particular information needed expressed in the query.

* **A presentation of the result set:** This can be as simple as a ranked list of document

titles or as complex as a rotating color map of the result set projected onto a three

dimensional space, rendered as a two-dimensional display.

**Q. 8) List the six different approaches for Information Extraction. Mention the**

**complexity dimension taken into consideration for categorization.**

**CO4 (06)**

**Ans) Six different approaches for Information Extraction:**

1. Finite-state automata for information extraction
2. Probabilistic models for information extraction
3. Conditional random fields for information extraction
4. Ontology extraction from large corpora
5. Automated template construction
6. Machine reading

**Q.9) Write the differences between attribute based extraction and relational extraction system.**

**CO4 (06)**

**Attribute Based Extraction:**

The simplest type of information extraction system is an attribute-based extraction

system that assumes that the entire text refers to a single object and the task is to

extract attributes of that object.

Consider the example of extracting information from the text “IBM ThinkBook970. Our price: $399.00” . The set of attributes include {Manufacturer=IBM, Model=ThinkBook970, Price=$399.00}.

We can address this problem by defining a template (also known as pattern) for each

attribute we would like to extract.

The template is defined by a finite state automaton, the simplest example of which is

the regular expression.

To build up a regular expression template for prices in dollars:

[0-9] matches any digit from 0 to 9

[0-9]+ matches one or more digits

[.][0-9][0-9] matches a period followed by two digits

([.][0-9][0-9])? matches a period followed by two digits, or nothing

[$][0-9]+([.][0-9][0-9])? matches $249.99 or $1.23 or $1000000 or . . .

Templates are often defined with three parts: a prefix regex, a target regex,

and a postfix regex.

For prices, the target regex is as shown above, the prefix would look for

strings such as “price:” and the postfix could be empty.

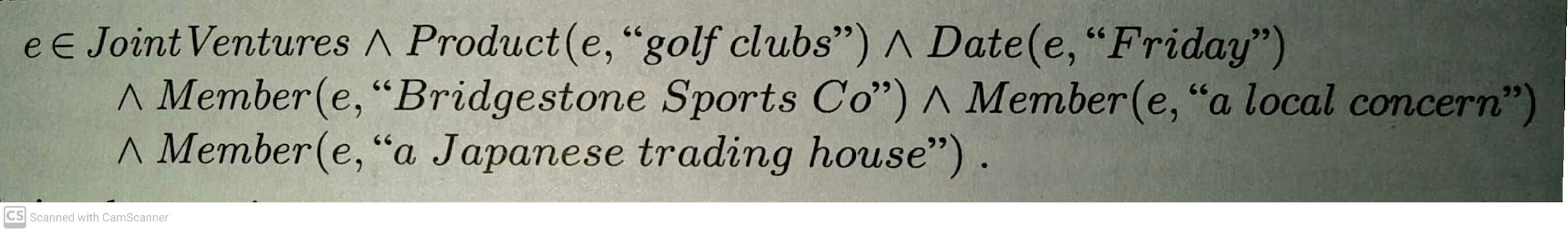
**Relational extraction system:**

One step up from attribute-based extraction systems are relational extraction systems, which deal with multiple objects and the relations among them. Thus, when these systems see the text “$249.99,” they need to determine not just that it is a price, but also which object has that price. A typical relational-based extraction system is FASTUS, which handles news stories about corporate mergers and acquisitions.

It can read the story:

**Bridgestone Sports Co. said Friday it has set up a joint venture in Taiwan with a local concern and a Japanese trading house to produce golf clubs to be shipped to Japan.**

Extracts the relations:

****

A relational extraction system can be built as a series of cascaded finite-state transducers. That is, the system consists of a series of small, efficient finite-state automata (FSAs), where each automaton receives text as input, transduces the text into a different format, and passes it along to the next automaton. FASTUS consists of five stages:

1. Tokenization

2. Complex-word handling

3. Basic-group handling

4. Complex-phrase handling

5. Structure merging

**Q.9) Explain the reason behind choosing conditional random fields for information extraction over Hidden Markov Model (HMM).**

One issue with HMMs for the information extraction task is that they model a lot of probabilities that we don’t really need.

HMMs cannot use overlapping features; CRFs can.

CRF formalism gives us a great deal of flexibility in defining them. This flexibility can lead to accuracies that are higher than with less flexible models such as HMMs.

**Q.10) Compare the working strategy and the efficiency of PageRank and the Hyperlink-Induced Topic search Algorithm (HITS) for information**

**retrieval. Use suitable examples to support your answer.**

* PageRank was invented to solve the problem of the tyranny (manipulation) of TF (Term Frequency) scores.
* If the query is [IBM], how do we make sure that IBM’s home page, ibm.com, is the

first result, even if another page mentions the term “IBM” more frequently?

* The idea is that ibm.com has many in-links (links to the page), so it should be linked

higher: each in-link is a vote for the quality of the linked-to page.

* But if we only counted in-links, then it would be possible for a web spammer to

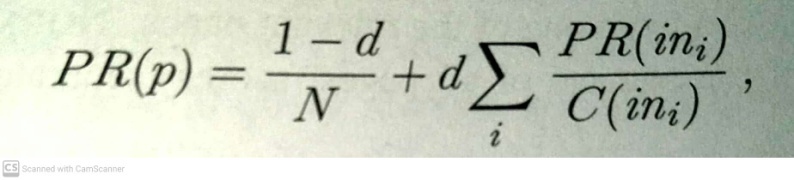
create a network of pages and have them all point to a page of his choosing,

increasing the score of that page.

Therefore, the PageRank algorithm is designed to weight links from high-quality sites

more heavily. High-Quality Site is one that is linked to by other high-quality sites.

The PageRank for a page p is defined as:



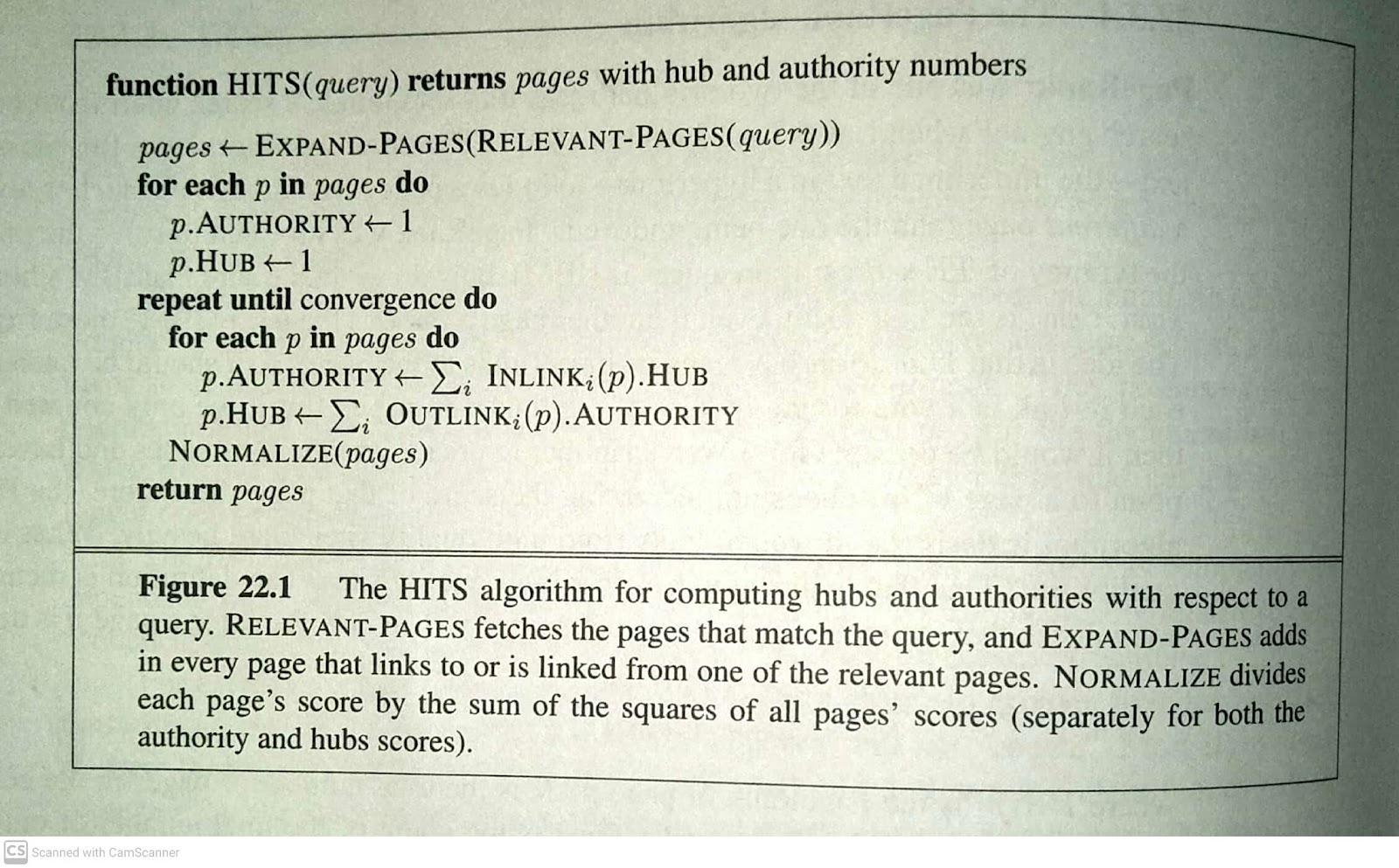
* + PR(p) is the PageRank of page p
  + N is the total number of pages in the corpus
  + *in****i*** are the pages that link into p
  + C(*in****i***) is the count of the total number of out-links on page *in****i***.
  + The constant d is a **damping factor**. It can be understood through the **random surfer model**: imagine a Web surfer who starts at some random page and begins exploring. **With probability d the surfer clicks on one of the links on the page** and **with probability 1-d he/she gets bored with the page and restarts on a random page anywhere on the web**.
  + The PageRank of page p is then the probability that the random surfer will be at p at any point in time.
  + PageRank can be computed by an iterative procedure: start with all pages having PR(p)=1, and iterate algorithm, updating ranks until they converge.

**The HITS algorithm ( Detailed answer from tb)**

The **Hyperlink-Induced Topic Search** algorithm, also known as “Hubs and Authorities” or HITS, is another influential link-analysis algorithm. HITS differs from PageRank in several ways. First, it is a **query-dependent measure**: it rates pages with respect to a query. That means that it must be computed anew for each query—a computational burden that most search engines have elected not to take on.

Given a query, HITS first finds a set of pages that are relevant to the query. It does that by intersecting hit lists of query words, and then adding pages in the link neighborhood of these pages—pages that link to or are linked from one of the pages in the original relevant set. Each page in this set is considered an authority on the query to the degree that other pages in the relevant set point to it. A page is considered a hub to the degree that it points to other authoritative pages in the relevant set. Just as with PageRank, we don’t want to merely count the number of links; we want to give more value to the high-quality hubs and authorities. Thus, as with PageRank, we iterate a process that updates the authority score of a page to be the sum of the hub scores of the pages that point to it, and the hub score to be the sum of the authority scores of the pages it points to. If we then normalize the scores and repeat k times, the process will converge.

Both PageRank and HITS played important roles in developing our understanding of Web information retrieval. These algorithms and their extensions are used in ranking billions of queries daily as search engines steadily develop better ways of extracting yet finer signals of search relevance.



{Shorten answer as per convenience}

**CO4 (12)**

**Q.11) Briefly describe the four classes of grammatical formalisms as described by Chomsky [1957].CO4 (08)**

Grammatical formalisms can be classified by their generative capacity: the set of languages they can represent. Chomsky (1957) describes four classes of grammatical formalisms that differ only in the form of the rewrite rules. The classes can be arranged in a hierarchy, where each class can be used to describe all the languages that can be described by a less powerful class, as well as some additional languages. Here we list the hierarchy, most powerful class first:

**Recursively enumerable grammars** use unrestricted rules: both sides of the rewrite rules can have any number of terminal and nonterminal symbols, as in the rule ABC → D E. These grammars are equivalent to Turing machines in their expressive power.

**Context-sensitive grammars** are restricted only in that the right-hand side must contain at least as many symbols as the left-hand side. The name “context-sensitive” comes from the fact that a rule such as A X B → A Y B says that an X can be rewritten as a Y in the context of a preceding A and a following B.Context-sensitive grammars can represent languages such as anbncn (a sequence of n copies of a followed by the same number of bs and then cs).

In **context-free grammars** (or CFGs), the left-hand side consists of a single nonterminal symbol. Thus, each rule licenses rewriting the nonterminal as the right-hand side in any context. CFGs are popular for natural-language and programming-language grammars, although it is now widely accepted that at least some natural languages have constructions that are not context-free (Pullum, 1991).

Context-free grammars can represent anbn, but not anbncn.

**Regular grammars** are the most restricted class. Every rule has a single non-terminal on the left-hand side and a terminal symbol optionally followed by a non-terminal on the right-hand side. Regular grammars are equivalent in power to finite-state machines. They are poorly suited for programming languages, because they cannot represent constructs such as balanced opening and closing parentheses (a variation of the anbn language). The closest they can come is representing a∗b∗, a sequence of any number of as followed by any number of bs.

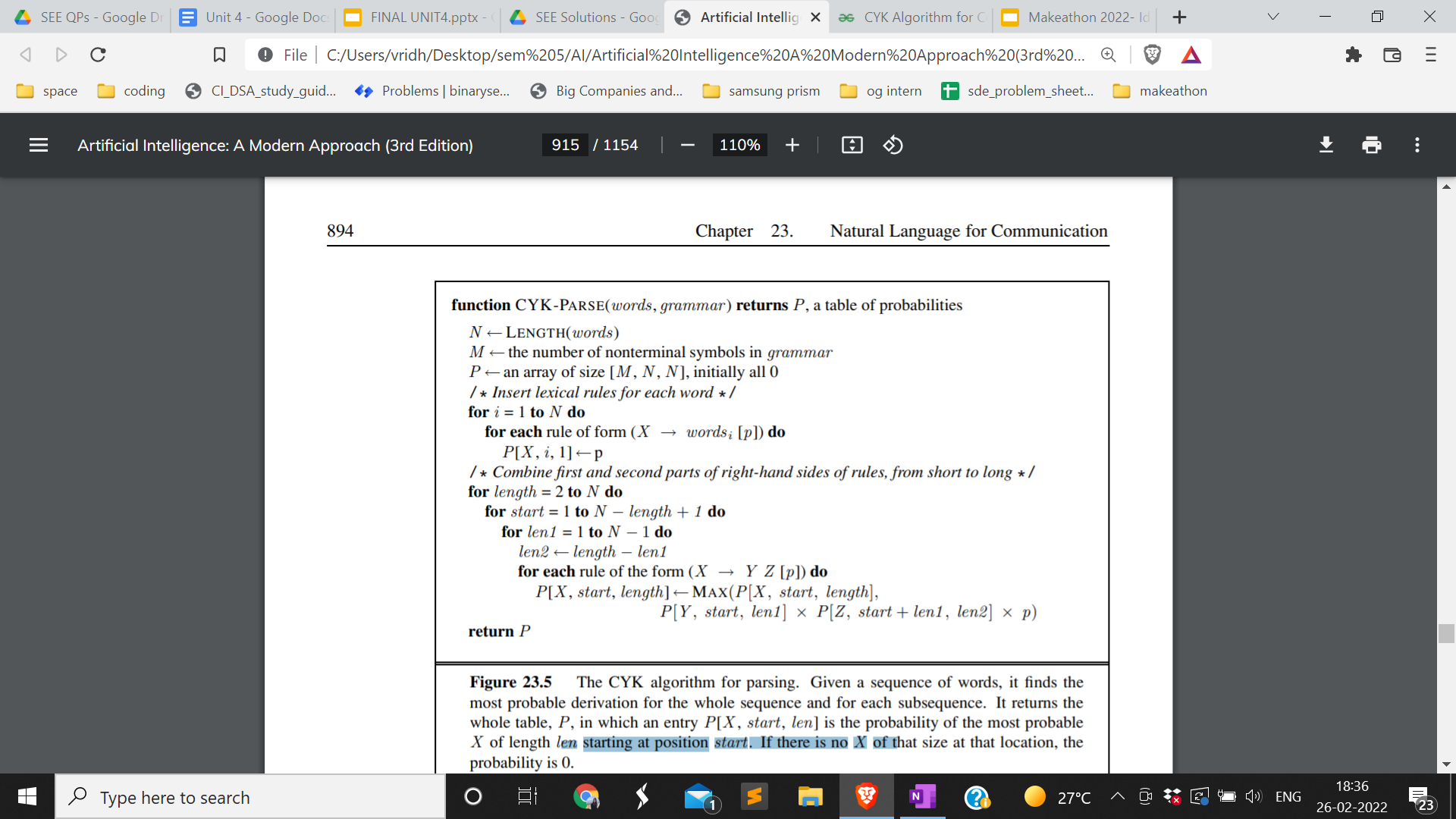
**Q.12 ) What is syntactic analysis (parsing) in NLP. Write the CYK algorithm**

**for parsing and explain it briefly. CO4 (10)**

Syntactic Analysis(parsing) - refer ans (2) -(ii)

**CYK algorithm for parsing:**

* + The CYK Algorithm is shown in Figure.
  + It requires a grammar with all rules in one of two very specific formats: lexical rules of the form **X -> word**, and syntactic rules of the form **X -> Y Z**.
  + This grammar format, called the Chomsky Normal Form.



* + CYK algorithm uses space of **O(n2m)** for the P table
  + n is the number of words in the sentence
  + m is the number of nonterminal symbols in the grammar
  + time complexity: **O(n3m)**.
  + Since m is constant for a particular grammars, this is commonly described as O(n3).

<https://www.geeksforgeeks.org/cyk-algorithm-for-context-free-grammar/>

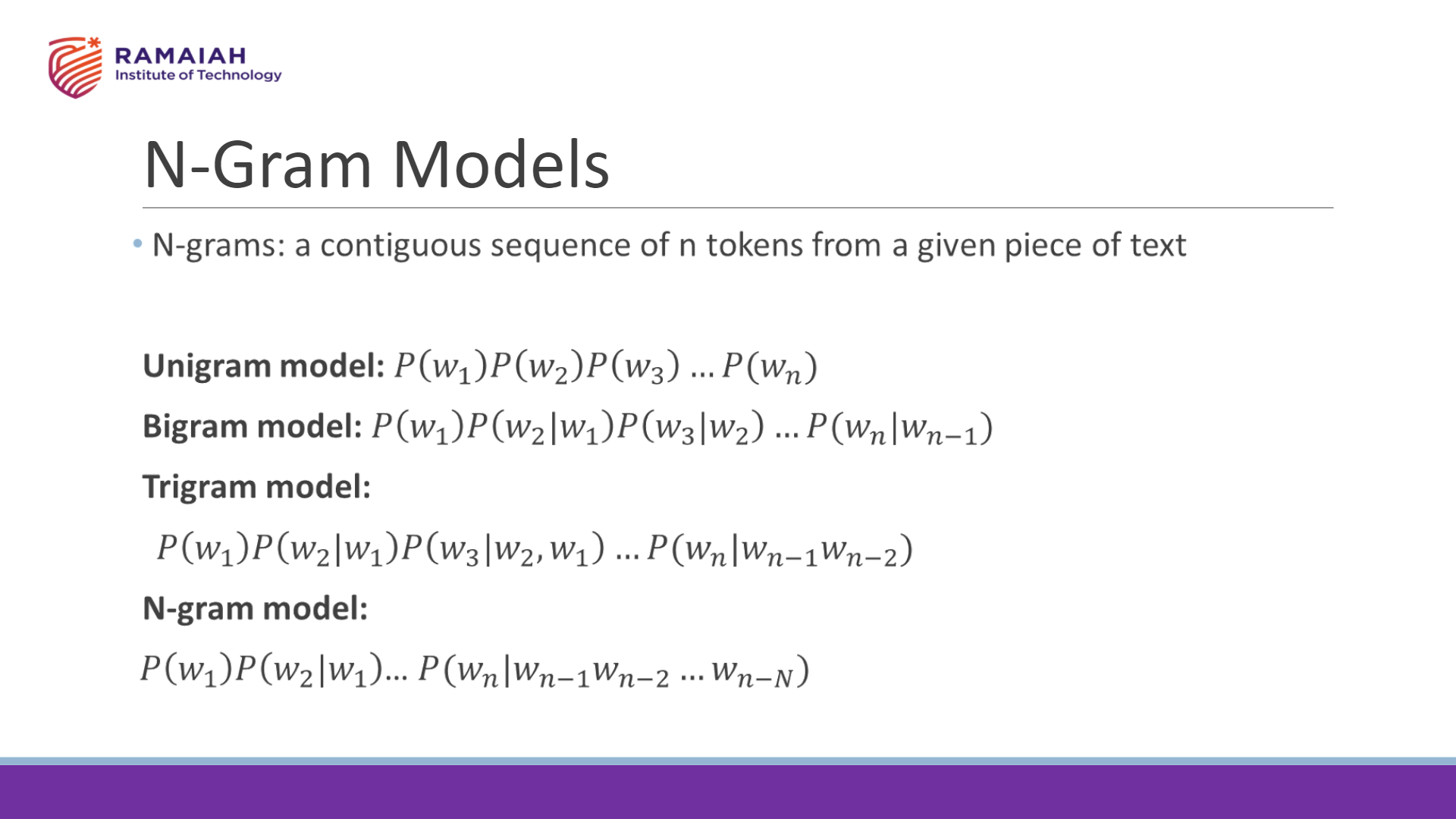
**Q.13 ) What is the significance of N-gram models in NLP? Why is the task of**

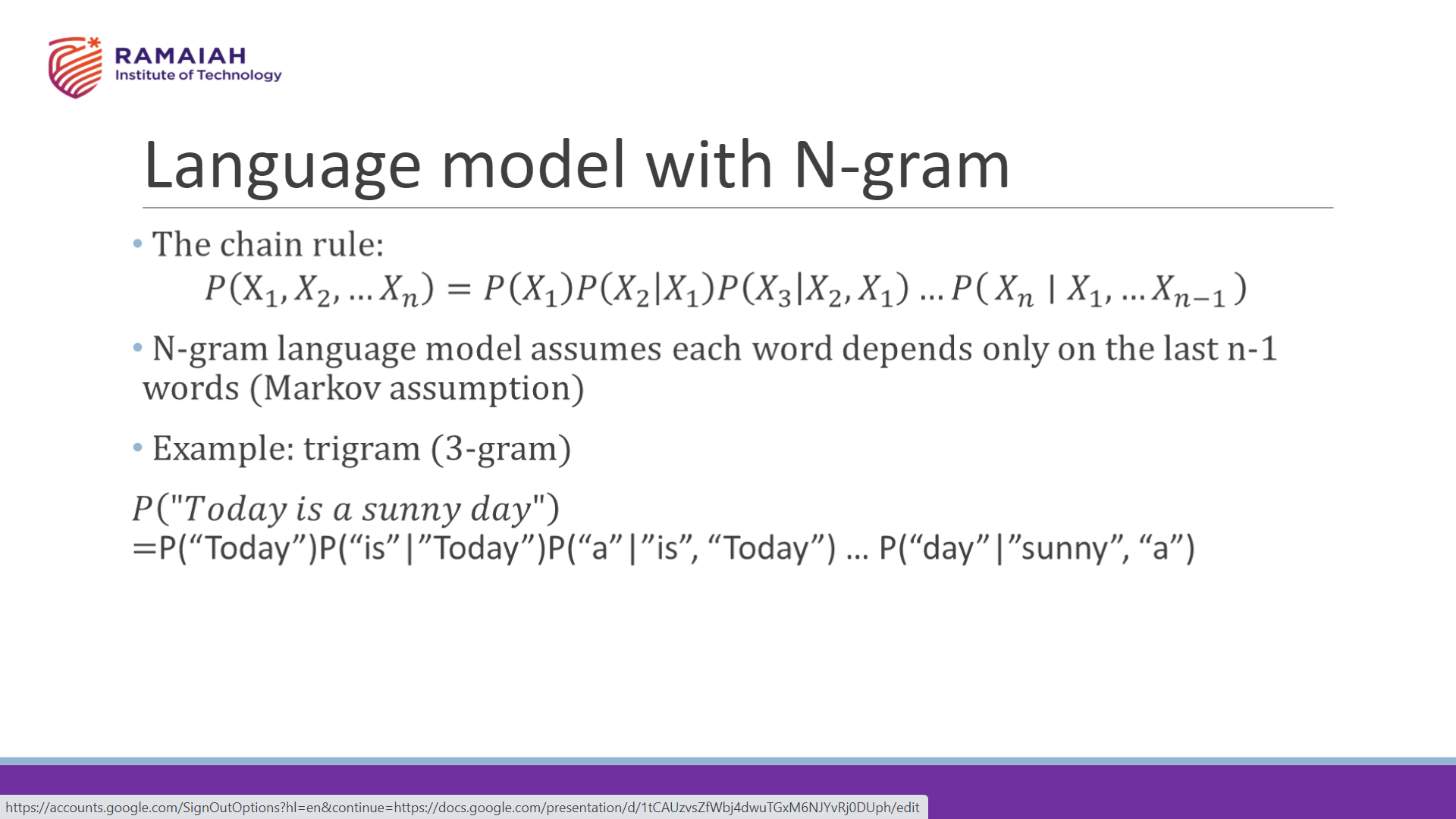
**feature selection and preprocessing of data necessary in the process?**

Ans)

A model of the probability distribution of n-letter sequences is called an **n-gram model.** (We can have n-gram models over sequences of words, syllables, or over characters.) An n-gram model is defined as a Markov chain of order n − 1.

Given a sequence of n-1 words, an N-gram model predicts the most probable word that might follow this sequence. It's a probabilistic model that's trained on a corpus of text. Such a model is useful in many NLP applications including speech recognition, machine translation and predictive text input.





**Feature selection : (Answer from Google)**

Feature selection methods can be **used to identify and remove unneeded, irrelevant and redundant attributes from data** that do not contribute to the accuracy of a predictive model or may in fact decrease the accuracy of the model.

**Eg:**

It can be expensive to run algorithms on a very large feature vector, so often a process of feature selection is used to keep only the features that best discriminate between spam and ham. For example, the bigram “of the” is frequent in English, and may be equally frequent in spam and ham, so there is no sense in counting it. Often the top hundred or so features do a good job of discriminating between classes

**Data Preprocessing**

Data preprocessing is essential before its actual use. Data preprocessing is **the concept of changing the raw data into a clean data set**. The dataset is preprocessed in order to check missing values, noisy data, and other inconsistencies before executing it to the algorithm.

**Q.14 ) Define the Information Retrieval process. Explain how the PageRank and HITS algorithms can be applied for the same. CO4 (12)**

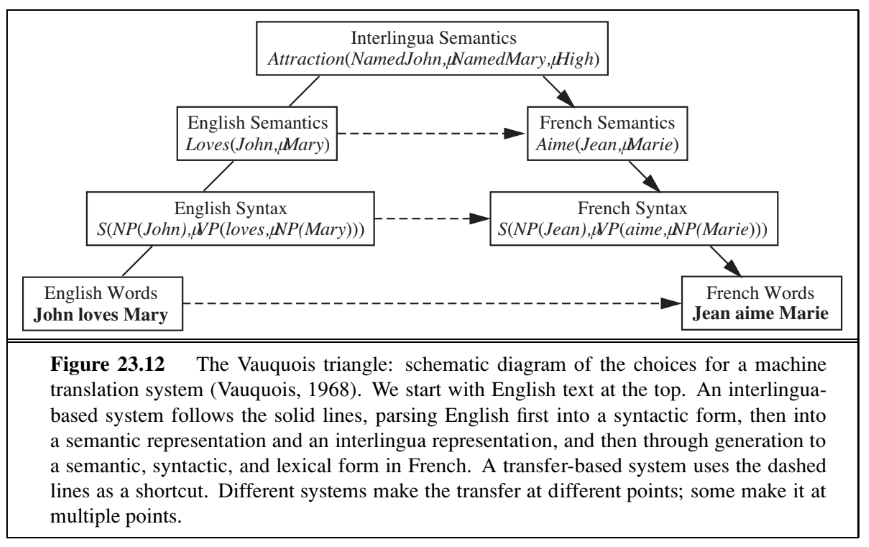
Refer Ans (1) and Ans (10)

**Q.15) Using suitable example, explain the concepts of “parsing with semantic grammar” and “case grammar” as applied to NLP. CO4 (08)**

**??**

Parsing is the process of analyzing a string of words to uncover its phrase structure, according to the rules of a grammar.

**Q.16 ) Briefly explain the machine translation process using the Vauquois triangle concept. CO4 (10)**

****

All translation systems must model the source and target languages, but systems vary in the type of models they use. Some systems attempt to analyze the source language text all the way into an **interlingua knowledge representation** and then generate sentences in the target language from that representation. This is difficult because it involves three unsolved problems: creating a complete knowledge representation of everything; parsing into that representation; and generating sentences from that representation.

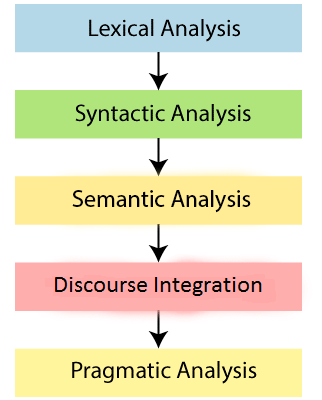
Other systems are based on a transfer model. They keep a database of translation rules(or examples), and whenever the rule (or example) matches, they translate directly. Transfer can occur at the lexical, syntactic, or semantic level. For example, a strictly syntactic rule maps English [Adjective Noun] to French [Noun Adjective]. A mixed syntactic and lexical rule maps French [S1 “et puis” S2] to English [S1 “and then” S2]. Figure 23.12 diagrams the various transfer points.

**Q.17 ) List the steps in the natural language understanding process and explain the same using suitable examples. CO4 (10)**

## { Not fully sure of the answer. Answer below from - [NLP Tutorial - Javatpoint](https://www.javatpoint.com/nlp) }

## Phases of NLP

There are the following five phases of NLP:



**1. Lexical Analysis and Morphological**

The first phase of NLP is the Lexical Analysis. This phase scans the source code as a stream of characters and converts it into meaningful lexemes. It divides the whole text into paragraphs, sentences, and words.

**2. Syntactic Analysis (Parsing)**

Syntactic Analysis is used to check grammar, word arrangements, and shows the relationship among the words.

Example: Agra goes to the Poonam

In the real world, Agra goes to the Poonam, does not make any sense, so this sentence is rejected by the Syntactic analyzer.

**3. Semantic Analysis**

Semantic analysis is concerned with the meaning representation. It mainly focuses on the literal meaning of words, phrases, and sentences.

**4. Discourse Integration**

Discourse Integration depends upon the sentences that precede it and also invokes the meaning of the sentences that follow it.

**5. Pragmatic Analysis**

Pragmatic is the fifth and last phase of NLP. It helps you to discover the intended effect by applying a set of rules that characterize cooperative dialogues.

For Example: "Open the door" is interpreted as a request instead of an order.

**Q.18) Describe N-gram character and word models with examples. List out any two applications.**

**CO4 (10)**

**Q.19) Develop a parse tree for the sentence “Cat slept on the table” using the**

**following rules. CO4 (07)**

**S -> NP VP**

**NP -> N**

**NP -> DET N**

**VP -> V PP**

**PP-> PREP NP**

**N ->Cat | table**

**V ->slept**

**DET-> the**

**PREP-> on**

**Ans)**

**S**

**/ \**

**/ \**

**NP VP**

**| / \**

**N V PP**

**| | / \**

**Cat Slept Prep NP**

**| / \**

**on DET N**

**| |**

**the table**

**Q.20) John wants to develop autocomplete functionality in search application.**

**He considered the below text for training language model to find out the conditional on previous words.**

**Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, ‘and what is the use of a book,’ thought Alice ‘without pictures or Conversation ?’**

**Calculate the probability of the next word, given the previous word.**

**i. P(Wi+1= pictures | Wi =without)**

**ii. P(Wi+1= book | Wi =the)**

**iii. P(Wi+1= sister | Wi =her)**

**iv. P(Wi+1= conversation | Wi =or AND Wi-1 =pictures)**

(I am just improvising here not sure if it’s the right ans) (Checked, seems right)

* i. Occurrence of (**without** ) = 1
* Occurance of (**without pictures**) = 1
* Hence **P(Wi+1= pictures | Wi =without)** = 1 / 1
* Ii. Occurrence of (**the** ) = 3
* Occurance of (**the book**) = 1
* Hence **P(Wi+1= book | Wi =the)** = 1 / 3
* Iii . Occurrence of (**her** ) = 2
* Occurance of (her **sister** ) = 2
* Hence  **P(Wi+1= sister | Wi =her)** = 1 / 1

Iv. P(pictures or ) = 1 / 2

P(conversation) = 1 / 2

P(pictures or conversation) = 2 / 2