# Name syyed Muhammad Salman Khan

# ID : 13662

# Subject : NLP

# Submitted to : Mam Aasma khan

# Final Summer Exam

# Semester : summer

Explain objectives of NLP? Write the name of 2 Applications of NLP with example? Write the name of 2 Challenges of NLP with example?

**Natural Language Processing (NLP)** is a field of Artificial Intelligence (AI) that quantifies human language to make it intelligible to machines. Natural language processing combines the power of linguistics and computer science to study the rules and structure of language, and create intelligent systems capable of understanding, analyzing, and extracting meaning from text and speech.
Natural language understanding (NLU) is used to understand the structure and meaning of human language by analyzing different aspects like syntax, semantics, pragmatics, and morphology. Then, computer science transforms this linguistic knowledge into rule-based or machine learning algorithms that can solve specific problems and perform desired tasks.

**Text Summarization:**

Automatic summarization is pretty self-explanatory. It summarizes text, by extracting the most important information. Its main goal is to simplify the process of going through vast amounts of data, such as scientific papers, news content, or legal documentation.
There are two ways of using natural language processing to summarize data: extraction-based summarization – which extracts key phrases and creates a summary, without adding any extra information – and abstraction-based summarization, which creates new phrases paraphrasing the original source. This second approach is more common and performs better.

**Market Intelligence**

Marketers can benefit from natural language processing to learn more about their customers and use those insights to create more effective strategies.
Analyzing topics, sentiment, keywords, and intent in unstructured data can really boost your market research, shedding light on trends and business opportunities. You can also analyze data to identify customer pain points and to keep an eye on your competitors (by seeing what things are working well for them and which are not).

**Challenges in NLP**:

**1. Natural language ambiguity**

In natural language, a word can have different meanings and the meaning of the word can be extracted from the context. For example, the sentence "A piece of cake" might mean that we are talking about a small portion of a birthday cake, on the other hand, it might mean that something is very easy to do. The humans don't only use their knowledge of a language to decide the meaning of a piece of text but also consider several other factors such as desires, goals and beliefs to understand the text they are reading or listening to. For example, the sentence "I experienced a feeling I have never had before" might mean that the person experienced a very pleasant feeling or a very bad one and the meaning of this sentence depends on the personal emotions at that moment.

**2 Spelling mistakes and entity extraction**

Correcting miss pelt words is an essential process in NLP as Misspellings are very frequent in human-computer interactions and it would be very hard to identify a misspelt entity (the noun in the phrase) in a text. For example: if a user wrote on a chatbot "Is it going to rain today in amestedam?", it would be hard to identify Amsterdam as a location.
languages either due to not available skills or the difficulty of the language as the case in Arabic language.

a) What is the difference between Pragmatics and Discourse? (03)

**Pragmatics**

Pragmatics is a branch of linguistics, which is the study of language. Pragmatics focuses on conversational implicature, which is a process in which the speaker implies and a listener infers. Simply put, pragmatics studies language that is not directly spoken. Instead, the speaker hints at or suggests a meaning, and the listener assumes the correct intention.

**Discourse analysis** studies meaning too but focuses on larger scale units (articles, conversations) and their overall interpretation in a specific communicative context, e.g. how a writer or speaker uses words, grammatical structures, intonation, ... to construe a reality and persuade the reader or listener. It looks at the participants' intentions, their background knowledge, the spatial and temporal setting, the preceding and surrounding texts, etc.

a) Define Phonology and Morphology with the help of example? (02

**Phonology** is the study of sounds and their parts. It focuses on how sounds are made using mouth shape, tongue placement, vocal cord use, etc. For example, it can look at the difference between fricative sounds like [f] (a f sound like in fish) and [ʃ] (this is a "sh" sound) and all other possible sounds. Typically it uses the International Phonetic Alphabet (IPA) to write out the representation of a word. For example, the word cat would be /kæt/ in a very simple form of IPA.

**Morphology** is the study of the smallest meaningful units of words. It looks at words and breaks them into their simplest parts to analyze meaning. For example, the word unbelievable can be broken into the basic parts of "un-" meaning "not", "believe", and "able" meaning "to be able to". Together it means "not able to be believed."

**Question No. 2:** (15)

a) What do you mean by regular expressions?

**A regular expression** (or "regex") is a search pattern used for matching one or more characters within a string. It can match specific characters, wildcards, and ranges of characters. Regular expressions were originally used by Unix utilities, such as vi and grep. However, they are now supported by many code editing applications and word processors on multiple platforms. Regular expressions can also be used in most major programming languages. A regular expression can be as simple as a basic string, such as "app". The regex "app" would match strings containing the words "apps," "applications," and "inapplicable". A regular expression may also contain anchor characters ("^" and "$") which are used to specify the beginning and end of a line, respectively. Therefore, the regex "^apps" would match the string, "apps are great," but would not match the string, "I like apps."

**b)** Specify the text strings using the below regular expressions:

a. /a(bc)

- Given string: ab abc ac acb a0b a2b a42c A87d

b. /[abc]

- Given string: ab abc ac acb a0b a2b a42c A87d

c. /abc+

- Given string: ab abc abcc babc

d. /abc*

- Given string: ab abc abcc babc

e. /[^a-z A-Z 0-9]

- Given string: a89 opx cfff $1!

**Question No. 5:** (15)

a) Explain Part of Speech Tagging (POS) and explain POS tag ambiguity with two examples. (03)

b) State difference between open vs. closed classes in POS tagging. (02)

c) Apply Viterbi Algorithm on the below given bigram and lexical probabilities; (10)

| Initial Probabilities | |
|---|---|
| Noun | 1\3 |
| Verb | 0 |
| Other | 1\3 |

| Bigram Probabilities | | | |
|---|---|---|---|
| | Noun | Verb | Other |
| Noun | 1\4 | 1\4 | 0 |
| Verb | 1\4 | 0 | 1\4 |
| Other | 1\3 | 0 | 1\3 |

| Lexical Probabilities | | | | | |
|---|---|---|---|---|---|
| | O1=time | O2=flies | O3=like | O4=an | O5=arrow |
| Noun | 1\5 | 1\5 | 0 | 0 | 1\5 |
| Verb | 1\5 | 2\5 | 1\5 | 0 | 0 |
| Other | 0 | 0 | 1\5 | 2\5 | 0 |

(15)

a) Apply Bayesian theorem over the below given string:

^John got many NLP books. ^He found them all very interesting.

Where for lexical probabilities assume John=0.5, got=0.3, many=0.2, NLP=0.1 and books=0.

a) Explain Part of Speech Tagging (POS) and explain POS tag ambiguity with two examples.

**Part of speech** tagging is also known as part of speech tags ,lexical categories, word classes morphological classes , lexical tags etc.

The process of assigning a part-of-speech to each word in a sentence.

Example :play well with other.

| | |
|---|---|
| Play | V |
| Well | Ad |
| With | Prep |
| Other | N |

Example :

| | |
|---|---|
| Heat | Verb(noun ) |
| Water | Noun(verb) |
| In | Prep(noun ,adj) |
| A | Det (noun) |
| Large | Adj(noun) |
| Vessel | noun |

(03)

b) State difference between open vs. closed classes.

**Open classes :** unlimited numbers of words
Open classes allow new members through borrowing (for example, the noun *cafe*) and derivation (for example, the adjective *bounteous* from the noun *bounty*)

**Examples :** Noun , verb ,Adverb , Adjective…

**Closed classes:** Closed classes of words do not allow new members and usually involve grammatical rather than lexical words

**Examples** : Auxiliary , Articles, Determine ,conjunction ,pronoun ,interjections…

(02)

c) Apply Viterbi Algorithm on the below given bigram and lexical probabilities; (10)

| Initial Probabilities | |
|---|---|
| Noun | 1\3 |
| Verb | 0 |
| Other | 1\3 |

| Bigram Probabilities | | | |
|---|---|---|---|
| | Noun | Verb | Other |
| Noun | 1\4 | 1\4 | 0 |
| Verb | 1\4 | 0 | 1\4 |
| Other | 1\3 | 0 | 1\3 |

| Lexical Probabilities | | | | | |
|---|---|---|---|---|---|
| | O1=time | O2=flies | O3=like | O4=an | O5=arrow |
| Noun | 1\5 | 1\5 | 0 | 0 | 1\5 |

| | | | | | |
|---|---|---|---|---|---|
| Verb | 1\5 | 2\5 | 1\5 | 0 | 0 |
| Other | 0 | 0 | 1\5 | 2\5 | 0 |

**Lexical probability**

| | **Time** | **Flies** | **Like** | **An** | **Arrow** |
|---|---|---|---|---|---|
| **Noun** | **1/5** | **1/5** | **0** | **0** | **1/5** |
| **Verb** | **1/5** | **2/5** | **1/5** | **0** | **0** |
| **Other** | **0** | **0** | **1/5** | **2/5** | **0** |

| | **P1(h1)** | **P2(h2)** | **P3(h3)** | **P4(h4)** | **P5(h5)** |
|---|---|---|---|---|---|
| **Noun** | **1/2*1/5** | **1/2.1/5.1/3.1/5=1/150** | **0** | **0** | **1/3750.1/3.1/5=1/562500** |
| **Verb** | **0** | **1/2.1/5.1/3.2/5=1/75** | **1/75.1/3.1/5=1/1125** | **0** | **0** |
| **Other** | **0** | **0** | **1/75.1/5.1/2=1/750** | **1/750.1/2.2/5=1/3750** | **0** |

**Times = Noun**
**Flies = Verb**
**Like = Other**

**Question No. 2:** **(05)**

Apply Bayesian theorem over the below given string:
^John got many NLP books. ^He found them all very interesting.
Where for lexical probabilities assume John=0.5, got=0.3, many=0.2, NLP=0.1 and books=0.

| ^John | Got | Many | Nlp | Books . | ^he | Found | Them | Very | Interesting. |
|---|---|---|---|---|---|---|---|---|---|
| NN | VB | A | N | N | N | V | N | R | A |

| | ^ | N | V | A | R | . |
|---|---|---|---|---|---|---|
| ^ | 0 | 2 | 0 | 0 | 0 | 0 |
| N | 0 | 1 | 2 | 0 | 1 | 1 |
| V | 0 | 1 | 0 | 1 | 0 | 0 |
| A | 0 | 1 | 0 | 0 | 0 | 1 |
| R | 0 | 0 | 0 | 1 | 0 | 0 |
| . | 1 | 0 | 0 | 0 | 0 | 0 |

STEP NO 3:

| | ^ | N | V | A | R | . |
|---|---|---|---|---|---|---|
| ^ | 0 | 2/2=1 | 0 | 0 | 0 | 0 |
| N | 1 | 1/5 | 2/5 | 0 | 1/5 | 1/5 |
| V | 0 | 1/2 | 0 | 1/2 | 0 | 0 |
| A | 0 | 1/2 | 0 | 0 | 0 | 1/2 |
| R | 0 | 0 | 0 | 1/1=1 | 0 | 0 |
| . | 1/2 | 0 | 0 | 0 | 0 | 0 |

| | 0.5 | 0.3 | 0.2 | 0.1 | 0.6 |
|---|---|---|---|---|---|
| | JOHN | GOT | MANY | NLP | BOOKS |
| ^ | 0.5 | 0.3 | 0.2 | 0.1 | 0.6 |
| N | 0.5 | 0.3 | 0.2 | 0.1 | 0.6 |
| V | 0.5 | 0.3 | 0.2 | 0.1 | 0.6 |
| A | 0.5 | 0.3 | 0.2 | 0.1 | 0.6 |
| R | 0.5 | 0.3 | 0.2 | 0.1 | 0.6 |
| . | 0.5 | 0.3 | 0.2 | 0.1 | 0.6 |

P(JOHN/^) JOHN*^=0.5*2/2=0.5
P(GOT/^) GOT*^ = 0.3*2/2= 0.3
P(MANY/^)MANY*^ = 0.2 *2/2=0.2
P(NLP/^)NLP*^=0.1*2/2=0.1
P(BOOKS/^)BOOKS*^ = 0.6 *2/2=0.6
**P(john/N)JOHN*N=0.5*5/5=0.5**
and so on...

b) Find the CFG of the string "abaabaa" using the production rules
S→a, S→aAS, A→bS

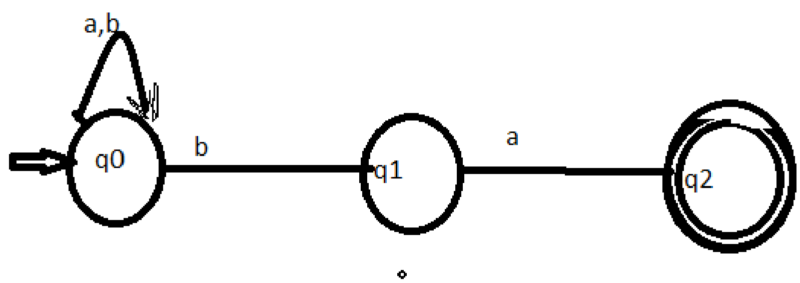
**S -> aAs (s ->aAs)**
**S -> abss (A ->bs )**
**S -> abas (s ->a)**
**S -> abaaAs (s ->aAs)**
**S -> abaabss (a ->bs)**
**S -> abaabas (s ->a)**
**S -> abaabaa (s ->a)**

**Question No. 3:** (20)

Design an NFA over an alphabet ∑= {a, b} such that every string accepted must end with a string --ba. Identify its tuples and also convert it into DFA.

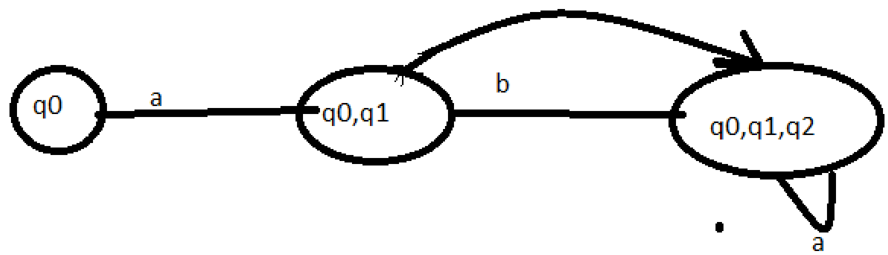
q 0= first phase
qf=final phase
&=empty

**NFA**




| qe | A | B |
|---|---|---|
| q 0 | q 0 | q 1 |
| q 1 | q 1 | & |
| q 2 | & | & |

DfA




| Qe | A | B |
| --- | --- | --- |
| q 0 | q 0 | q 0 |
| q 0 ,q1 | q 0 | q 0 ,q1 |
| q 0,q1 | q 0 | q 0,q1 |

Q

Qno3(b)