REGEX

- ➤ Patterns of special characters having an associated textual meaning
- ➤ Wildcard expressions for
  - ○ Matching
  - ○ Searching
  - ○ Parsing
- ➤ Used for writing rule-based info mining systems

APPLNS OF REGEX

- ➤ Segmentation of words from sentences
- ➤ Segmentation of sentences from paragraphs
- ➤ Text cleaning - noise removal
- ➤ Information retrieval from texts
  - ○ From chatbots
  - ○ New datasets
  - ○ Search Engine Queries

Types of Regular Expressions

abc...	Letters	{m}		m Repetitions
123...	Digits	{m,n}		m to n Repetitions
\d	Any Digit	*		Zero or more repetitions
\D	Any Non-digit character	+		One or more repetitions
.	Any Character	?		Optional character
\.	Period	\s		Any Whitespace
[abc]	Only a, b, or c	\S		Any Non-whitespace character
[^abc]	Not a, b, nor c	^...$		Starts and ends
[a-z]	Characters a to z	(...)		Capture Group
[0-9]	Numbers 0 to 9	(a(bc))		Capture Sub-group
\w	Any Alphanumeric character	(.*)		Capture all
\W	Any Non-alphanumeric character	(abc|def)		Matches abc or def

REGEX FUNCTIONS

➢ Match => find 1st occurrence of pattern in the str

➢ Search => locate the pattern in the str

➢ FindAll => find all occurrences of pattern in the str

https://s3.amazonaws.com/thinkific/file_uploads/118220/attachments/184/63e/f6b/Regex_Cheatsheet.pdf