Natural Language Processing (NLP)

Unit 2
Regular Expressions

By:
Syeda Saleha Raza

اَللَّهُمَّ إِنِّي أَسْأَلُكَ عِلْمًا نَافِعًا،

وَرِزْقًا طَيِّبًا، وَعَمَلًا مُّتَقَبَّلًا،

(O Allah, I ask You for beneficial knowledge, goodly provision and acceptable deeds)

اے اللہ ، میں آپ سے سوال کرتی ہوں نفع بخش علم کا، طیّب رزق کا، اور اس عمل کا

(Sunan Ibn Majah: 925)

Outline

- What is Regular Expression?

- Their Uses

- Writing regex– Code Demo

Regular Expressions

- A regular expression (regex or regexp) is a sequence of characters that defines a search pattern.

- It is a powerful tool used in string manipulation and text processing tasks.

Regular Expressions - Uses

1. Search and Match:

- Find occurrences of a specific pattern or substring within a text.

2. Validation:

- Validate input data, such as email addresses, phone numbers, zip codes, or dates.
- Ensure that user input conforms to a specific format or structure.

3. Text Extraction:

- Extract specific information from a text, such as extracting all email addresses or phone numbers from a document.

4. Data Cleaning:

- Replace or remove unwanted characters or substrings in a text.
- Normalize data by standardizing formats.

5. Parsing and Tokenization:

- Break down a text into meaningful tokens or elements.
- Parse structured data formats, such as XML or JSON.

Writing Regular Expressions

Code Walkthrough

Meta characters

MetaCharacters	Description
\	Used to drop the special meaning of character following it
[]	Represent a character class
^	Matches the beginning
$	Matches the end
.	Matches any character except newline
|	Means OR (Matches with any of the characters separated by it.
?	Matches zero or one occurrence

Regular Expression (RegEx) in Python with Examples - GeeksforGeeks

Meta characters

MetaCharacters	Description
*	Any number of occurrences (including 0 occurrences)
+	One or more occurrences
{}	Indicate the number of occurrences of a preceding regex to match.
()	Enclose a group of Regex

textonyms

References

- [Natural Language Processing with Python (tjzhifei.github.io)](#)

Thanks