Macro Programming

Vignesh V
Department of Computer Applications
vigneshv@pes.edu

PES UNIVERSITY

Macro Programming

Text Manipulation and Regular Expressions

Vignesh V
Department of Computer Applications

Macro Programming

Text Manipulation and Regular Expressions

Text manipulation in Excel involves using various functions and features to modify, extract, or combine text data. Excel provides a wide range of built-in text functions that can help you handle text data efficiently.

Function	Description	Example	Result
LEN	Calculates the length of a text string, including spaces.	=LEN("Hello World")	11
LEFT	Extracts characters from the beginning of a string.	=LEFT("Excel", 2)	"Ex"
RIGHT	Extracts characters from the end of a string.	=RIGHT("Excel", 2)	"el"
MID	Extracts characters from the middle of a string.	=MID("Excel", 2, 3)	"xce"

Text Manipulation and Regular Expressions

Function	Description	Example	Result
UPPER	Converts text to uppercase.	=UPPER("excel")	"EXCEL"
LOWER	Converts text to lowercase.	=LOWER("EXCEL")	"excel"
PROPER	Capitalizes the first letter of each word.	=PROPER("excel functions")	"Excel Functions"
CONCATENATE	Combines multiple text strings into one.	=CONCATENATE("Hello", " ", "World")	"Hello World"
TRIM	Removes extra spaces from text.	=TRIM(" Hello World ")	"Hello World"

Text Manipulation and Regular Expressions

Function	Description	Example	Result
SUBSTITUTE	Replaces occurrences of a specified text with new text.	=SUBSTITUTE("Hello World", "World", "Excel")	"Hello Excel"
FIND	Case-sensitive search for a substring's position.	=FIND("e", "Excel")	1
SEARCH	Case-insensitive search for a substring's position.	=SEARCH("e", "Excel")	1

Text Manipulation and Regular Expressions

Importance of Text Manipulation in Data Management

- Data Cleaning: Remove noise and inconsistencies to ensure data quality.
- Data Transformation: Convert data into usable formats for analysis.
- Feature Engineering: Extract meaningful features from text for predictive modeling.
- Text Mining and NLP: Enable sentiment analysis, topic modeling, and more

Challenges in Text Manipulation

- Complexity: Handling large datasets with diverse text formats.
- Efficiency: Traditional Excel functions can be cumbersome for complex tasks.
- Accuracy: Ensuring data integrity and avoiding errors during manipulation

Introduction to Regular Expressions (Regex)

- Regular Expressions, commonly known as regex, are sequences of characters that define a search pattern. They are used to search, match, and manipulate text based on specific patterns.
- Used for searching, extracting, and manipulating text.
- Consists of special characters, operators, and constructs

Syntax of Regex

Special Characters:

Dot (.): Matches any single character except newline.

Asterisk (*): Matches zero or more occurrences of the preceding character.

Plus (+): Matches one or more occurrences of the preceding character.

Question Mark (?): Matches zero or one occurrence of the preceding character.

Character Classes:

Square Brackets [ ]: Matches any single character inside the brackets.

Negation [^]: Matches any single character not in the brackets.

Syntax of Regex

Quantifiers:

Curly Braces {n}: Matches exactly n occurrences of the preceding character.

Curly Braces {n,m}: Matches at least n and at most m occurrences.

Grouping and Capturing:

Parentheses ( ): Groups patterns and captures the matched text for further processing

Syntax of Regular Expressions

Pattern	Description	Example	Matches
.	Matches any single character except a newline	.at	cat, bat, rat
\d	Matches any digit (0-9)	\d	In "a1b", matches "1"
\D	Matches any non-digit character	\D	In "a1b", matches "a" and "b"
\s	Matches any whitespace character (space, tab, newline)	\s	In "3 cents", matches the space
\S	Matches any non-whitespace character	\S+	In "30 cents", matches "30" and "cents"
\w	Matches any word character (letters, digits, underscore)	\w+	In "5_cats***", matches "5_cats"
\W	Matches any non-word character	\W+	In "5_cats***", matches "***"

Syntax of Regular Expressions

Pattern	Description	Example	Matches
[abc]	Matches any single character in the set	[abc]	Matches "a", "b", or "c"
[^abc]	Matches any single character not in the set	[^abc]	Matches anything except "a", "b", or "c"
*	Matches zero or more of the preceding character	ab*c	Matches "ac", "abc", "abbc"
+	Matches one or more of the preceding character	ab+c	Matches "abc", "abbc"
?	Matches zero or one of the preceding character	ab?c	Matches "ac", "abc"

Syntax of Regular Expressions

Pattern	Description	Example	Matches
{n}	Matches exactly n occurrences of the preceding character	ab{2}c	Matches "abbc"
{n,m}	Matches between n and m occurrences of the preceding character	ab{1,3}c	Matches "abc", "abbc", "abbbc"
( )	Groups and captures the pattern inside	(ab)+	Matches "ab", "abab", "ababab"

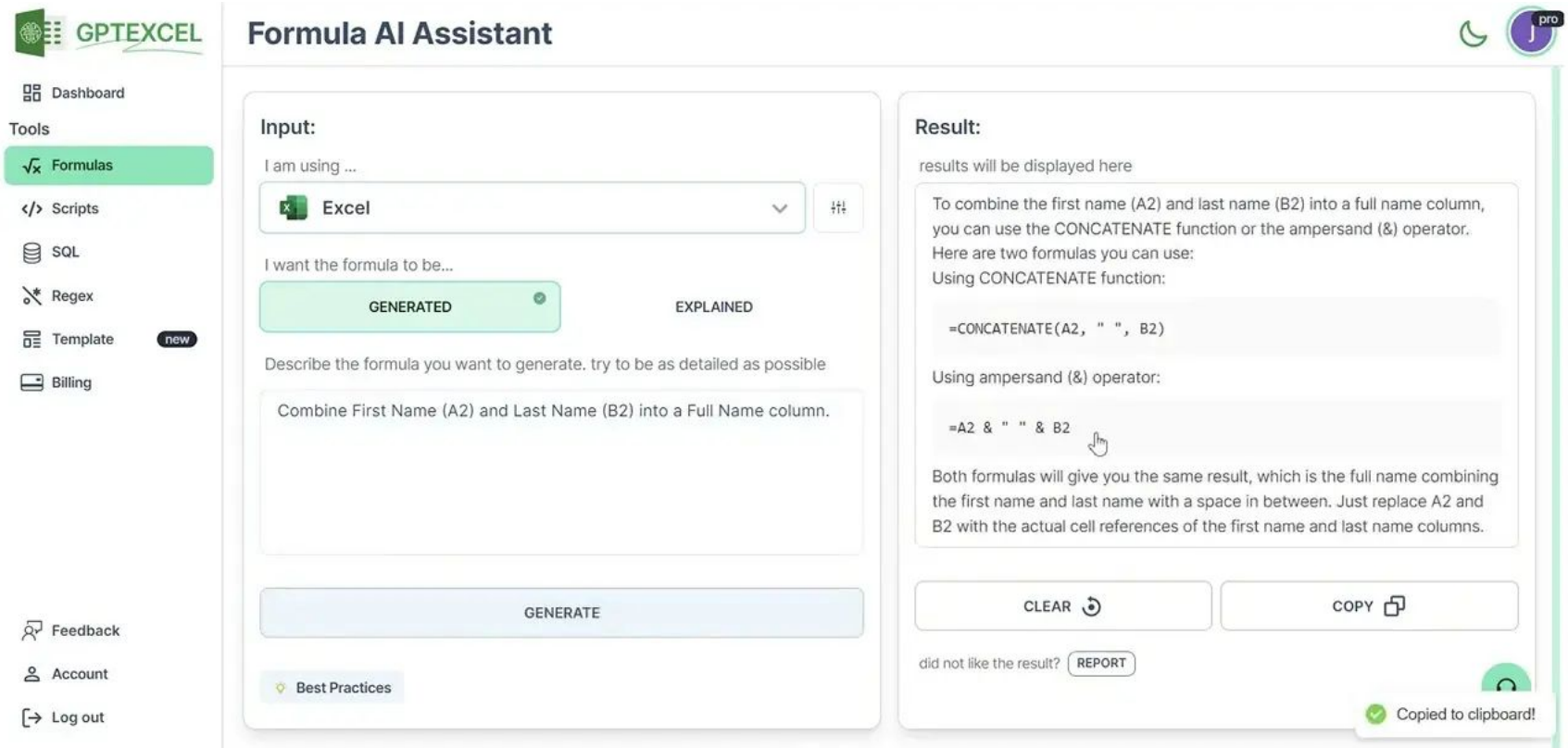
Text Manipulation and Regular Expressions

Why Use Regex in Excel?

- Efficiency: Simplifies complex text manipulation tasks.
- Versatility: Supports a wide range of operations like matching, extracting, and replacing text.
- Automation: Reduces manual data entry errors and enhances data validation

Macro Programming

Introduction to GPTExcel



Introduction to GPTExcel

- GPTExcel is an AI-powered tool that helps generate Excel formulas using plain English.
- It is especially useful for:
  - Complex formulas
  - Regex patterns
  - Quick troubleshooting
- **Why Use It?**
  - Saves time when writing or learning new formulas
  - Reduces errors in syntax
  - Ideal for beginners and experts alike

THANK YOU

Vignesh V
Department of Computer Applications
vigneshv@pes.edu