# ENTITY EXTRACTION FOR DATES & COURSES

Implement basic entity recognition to extract dates, course, codes and semester numbers from questions and use them in responses.

## INTRODUCTION:

Entity extraction is defined as the automatic identification and/or extraction of useful information from any text, such as course information which could include course name, course numbers, dates, semester numbers, and the like, using a computer." Entity extraction is an opportunity to respond to a query in a timely way by just retrieving the information you need without having to read it all.

## OBJECTIVE:

- Automatically extract course names, course codes, dates, and semester numbers.
- Offer quicker and accurate responses.
- Organize the data for easy use in applications.
- Carry out intelligent query handling with chatbots
- Offer faster and accurate responses.

## TECHNOLOGY USED:

- Input Question
- Pattern Matching
- Keyword Based Course Identification
- Organized Output Display

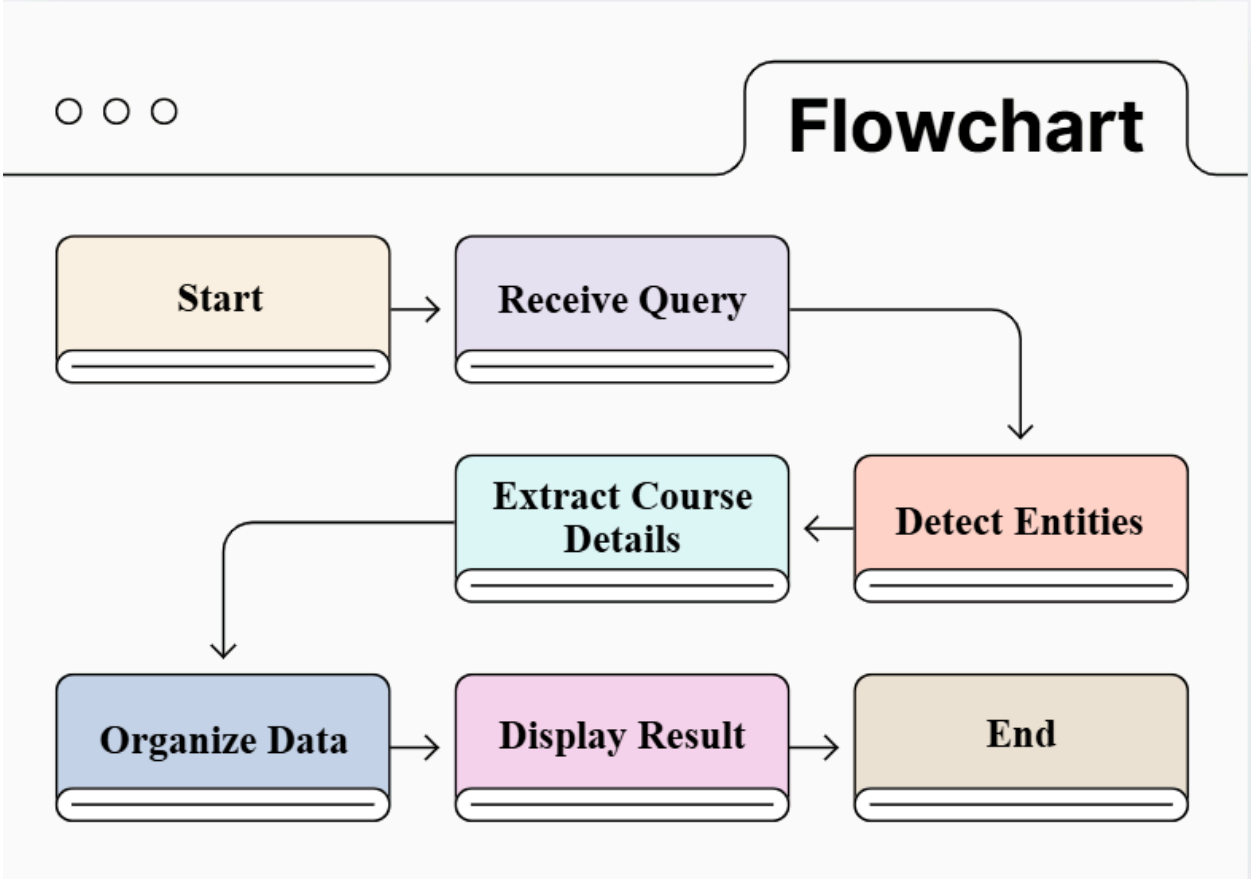
# IMPLEMENTATION:

- First, install and import the required libraries, spaCy for text processing and regex for regular
- Load the pre-trained NLP model.
- Define Patterns
- Establish patterns to recognize course codes and semester numbers.
- Define a function for retrieving dates, course names, codes, and semesters
- Test the function with example queries.
- Present your results in a clear, organized format.

```python
import re

def extract_entities(text):

    course_code_pattern = r'\b[A-Z]{2,4}\d{2,3}\b'
    semester_pattern = r'\b(?:Semester|Sem)\s?\d\b'
    date_pattern1 = r'\b\d{1,2}[/-]\d{1,2}[/-]\d{2,4}\b'
    date_pattern2 = r'\b\d{1,2}\s(?:Jan|Feb|Mar|Apr|May|Jun|Jul|Aug|Sep|Oct|Nov|Dec)\s\d{4}\b'

    course_codes = re.findall(course_code_pattern, text)
    semesters = re.findall(semester_pattern, text)
    dates1 = re.findall(date_pattern1, text)
    dates2 = re.findall(date_pattern2, text)

    dates = dates1 + dates2

    course_names_list = ["Operating System", "Data Structures",
                         "Computer Networks", "Microprocessor",
                         "DBMS", "Machine Learning"]

    found_courses = []
    for course in course_names_list:
        if course.lower() in text.lower():
            found_courses.append(course)

    print("----- Extracted Information -----")
    print("Course Codes:", course_codes)
    print("Semesters:", semesters)
    print("Dates:", dates)
    print("Course Names:", found_courses)


query = "When is the exam for CS101 in Semester 3 on 18 Feb 2026 for Operating System?"

extract_entities(query)
```

```
----- Extracted Information -----
Course Codes: ['CS101']
Semesters: ['Semester 3']
Dates: ['18 Feb 2026']
Course Names: ['Operating System']
```

## FLOWCHART




## OUTCOME:

- Automatically extracts course names, numbers, dates, and semester numbers from the user's query.
- Offers faster and more accurate responses through pattern matching.
- Organizes extracted data in a well-structured format.