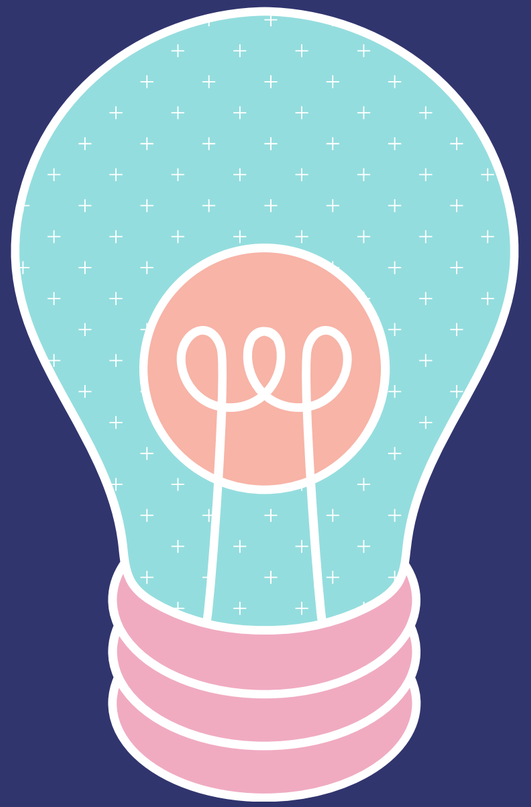
AIWR MINI PROJECT

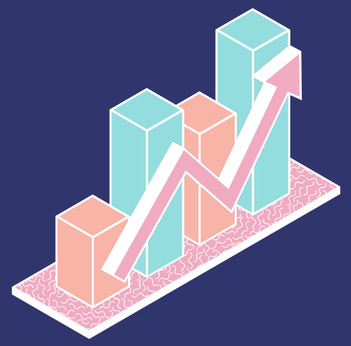
Event Extraction and Representation Model

Neha Chougule- PES1UG21CS161
Pavani BR- PES1UG21CS926

PROBLEM STATEMENT

The wealth of information about real-world events is buried within the vast quantity of unstructured news articles.
Manual analysis of these articles to identify and understand events is a time-consuming, error-prone task.
Automated event extraction is hindered by the inconsistent format and linguistic complexities inherent in news articles.
Furthermore, simply extracting event details is insufficient; we need a way to represent them semantically to capture their deeper meaning and relationships





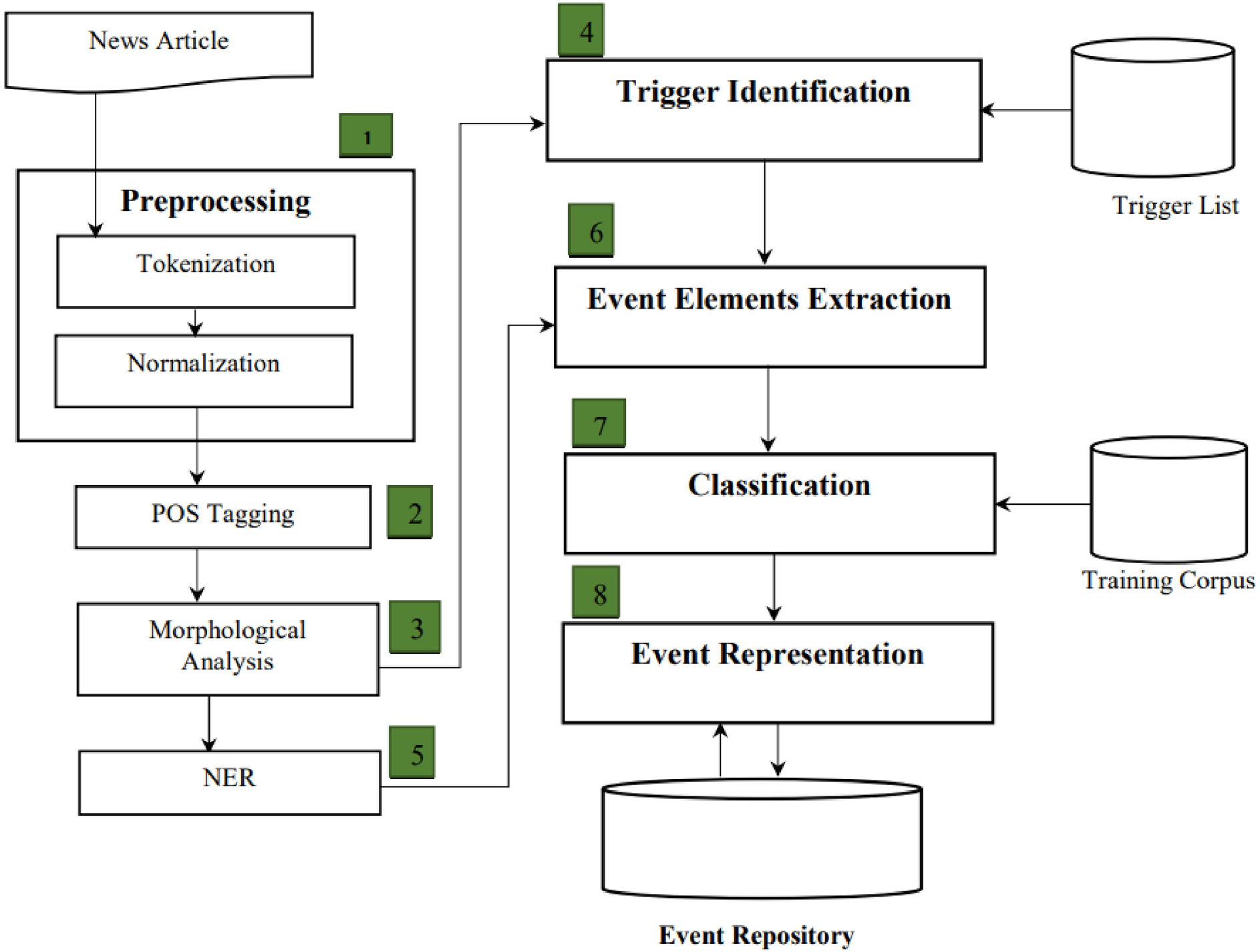
FEASIBILITY STUDY

- Dataset: publicly available news datasets with event annotation
- Computational Resources: Estimate the computational power and storage needed for model training and running the system on realistic data volumes.

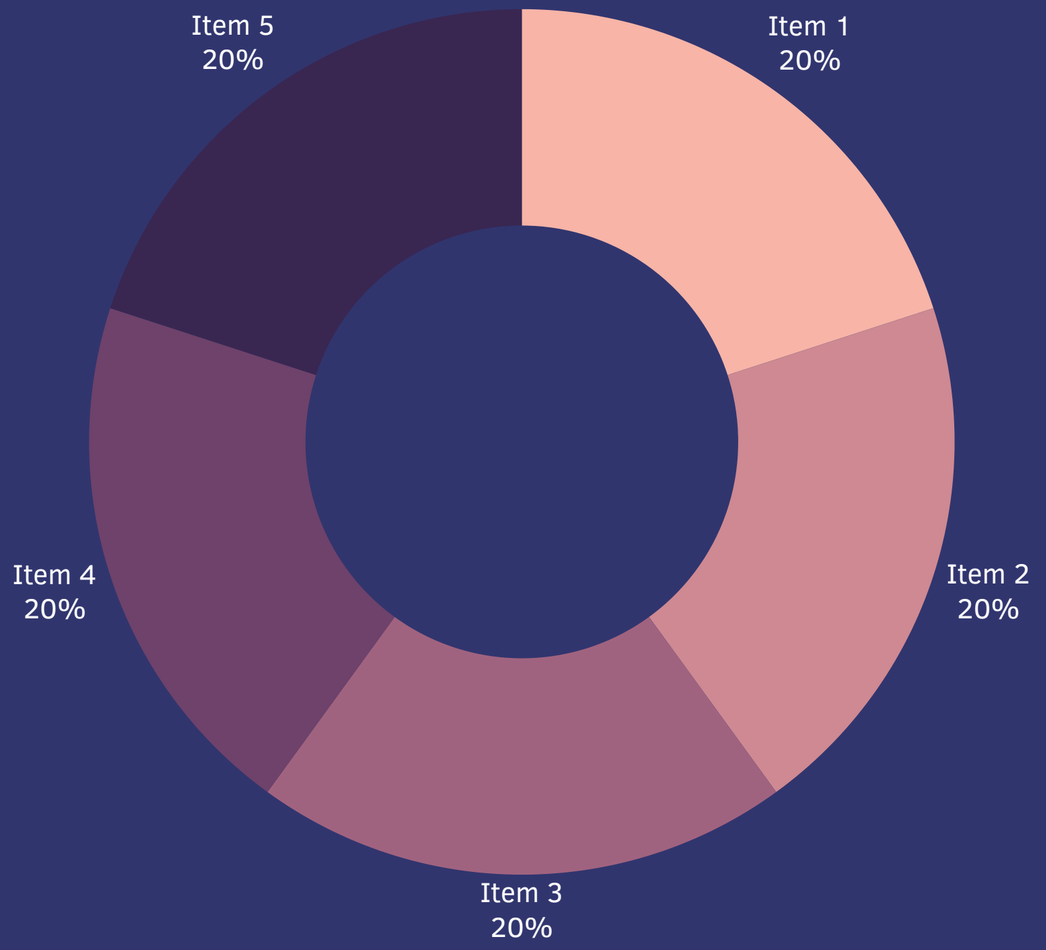
Metrics: Define clear metrics for success:

- **Accuracy:** Precision, recall, and F1 scores for event identification, trigger detection, and element extraction.

WORK FLOW


News Article
1
Preprocessing
Tokenization
Normalization
POS Tagging
2
Morphological Analysis
3
NER
5
4
Trigger Identification
Trigger List
6
Event Elements Extraction
7
Classification
Training Corpus
8
Event Representation
Event Repository

Literature Review



Event Extraction and Representation Model from News Articles

1

INTRODUCTION

Event extraction models use NLP to understand events in text by identifying key elements (who, what, when, where) and representing them in a structured way.

2

ARCHITECTURE

E = (T, P, L, K)
An event E, a real world phenomenon described as K, that occurred at specific time T, involving participant P, and tied to a location L that is published on the Web

3

METHODOLOGIES

NLP , Named Entity Recognition (NER) and Maximum Entropy classifier for event extraction

4

RESULTS

NLP tools help the classifier obtain good results. It also shows that event trigger is more important than other elements, because it signals the mention of event.