

***INFOSYS SPRINGBOARD
VIRTUAL INTERNSHIP- 6.0***

BATCH-5

**MILESTONE-I
FINAL REPORT**

BY

GROUP-4

TEAM MEMBERS

- Vaibhavi Patil
- Vardhan
- Preetham Adithya
- Ravi Kumar

Project Title:

AI Knowledge Graph using NLP & Visualization

Objective:

The objective of this project is to build a Mini Knowledge Graph System that automatically extracts entities and relationships from text data and visualizes how the entities are connected.

It demonstrates the integration of Natural Language Processing (NLP), Relation Extraction, and Graph Visualization in Python.

Dataset Description:

Our Dataset Name:

extended_famous_discoveries.csv

Type: CSV Dataset.

Description: The dataset contains sentences about people, organizations, inventions, and events across multiple domains.

Dataset:

| A | B | C | D | E | F | G | H | I | J |
|----|----|--|---|---|---|---|---|---|---|
| 1 | ID | Statement | | | | | | | |
| 2 | | 1 Albert Einstein developed the theory of relativity in 1905. | | | | | | | |
| 3 | | 2 Isaac Newton formulated the laws of motion in 1687. | | | | | | | |
| 4 | | 3 Marie Curie discovered radium and polonium. | | | | | | | |
| 5 | | 4 Charles Darwin introduced the concept of evolution through natural selection. | | | | | | | |
| 6 | | 5 Alan Turing proposed the idea of a universal computing machine. | | | | | | | |
| 7 | | 6 Elon Musk founded SpaceX in 2002. | | | | | | | |
| 8 | | 7 SpaceX launched the Falcon 9 rocket successfully. | | | | | | | |
| 9 | | 8 NASA collaborates with SpaceX for ISS missions. | | | | | | | |
| 10 | | 9 Google developed the Transformer model architecture in 2017. | | | | | | | |
| 11 | | 10 OpenAI created the GPT series of large language models. | | | | | | | |
| 12 | | 11 The printing press was invented by Johannes Gutenberg in the 15th century. | | | | | | | |
| 13 | | 12 The Internet revolutionized communication and information sharing. | | | | | | | |
| 14 | | 13 Artificial Intelligence is transforming healthcare and education. | | | | | | | |
| 15 | | 14 Neural networks are inspired by the structure of the human brain. | | | | | | | |
| 16 | | 15 The World Health Organization monitors global pandemics. | | | | | | | |
| 17 | | 16 Nikola Tesla contributed to the development of alternating current electricity. | | | | | | | |
| 18 | | 17 Alexander Fleming discovered penicillin in 1928. | | | | | | | |
| 19 | | 18 Wright brothers achieved the first powered flight in 1903. | | | | | | | |
| 20 | | 19 Tim Berners-Lee invented the World Wide Web in 1989. | | | | | | | |
| 21 | | 20 Stephen Hawking advanced the understanding of black holes. | | | | | | | |
| 22 | | 21 Thomas Edison invented the phonograph and improved the electric light bulb. | | | | | | | |
| 23 | | 22 Rosalind Franklin's work was crucial to discovering the DNA double helix structure. | | | | | | | |
| 24 | | 23 Apple introduced the first iPhone in 2007. | | | | | | | |
| 25 | | 24 IBM developed Watson, an AI capable of answering questions posed in natural language. | | | | | | | |
| 26 | | 25 Jeff Bezos founded Amazon in 1994 as an online bookstore. | | | | | | | |
| 27 | | 26 Mark Zuckerberg launched Facebook in 2004. | | | | | | | |
| 28 | | 27 Claude Shannon established the foundations of information theory. | | | | | | | |
| 29 | | 28 CRISPR technology allows scientists to edit genes with precision. | | | | | | | |
| 30 | | 29 Ada Lovelace is regarded as the first computer programmer. | | | | | | | |
| 31 | | 30 The Large Hadron Collider helped discover the Higgs boson particle in 2012. | | | | | | | |
| 32 | | 31 NASA's Apollo 11 mission landed humans on the Moon in 1969. | | | | | | | |
| 33 | | 32 Bitcoin was introduced in 2009 by an unknown person under the name Satoshi Nakamoto. | | | | | | | |
| 34 | | 33 The Human Genome Project was completed in 2003, mapping all human genes. | | | | | | | |
| 35 | | 34 The James Webb Space Telescope was launched to explore distant galaxies. | | | | | | | |
| 36 | | 35 Quantum computing aims to revolutionize data processing using quantum bits. | | | | | | | |

Summary :

(a) Entity Extraction

Library Used: spaCy

Model: en_core_web_sm

Extracted entities such as PERSON, ORG, GPE, and DATE.

Code Example:

```
import spacy, pandas as pd
```

```
nlp = spacy.load("en_core_web_sm")
df =
pd.read_csv(r"C:\Users\91984\Downloads\extended_famous_discoveries.csv", header=None,
names=["text"])
```

```
entities = [  
    {"Entity": ent.text, "Label": ent.label_,  
     "Sentence": text}  
    for text in df["text"].dropna()  
    for ent in nlp(text).ents  
]
```

```
pd.DataFrame(entities).to_csv(r"C:\Users\91984\New folder\Clean_NLP_Project\output\entities.csv",  
index=False)  
print(" Entities saved successfully!")
```

(b) Relation Extraction

A simple rule-based relation extractor identifies patterns like

founded, released, born in, produces, discovered, developed.

Code Example:

```
import spacy, pandas as pd

nlp = spacy.load("en_core_web_sm")
df =
pd.read_csv(r"C:\Users\91984\Downloads\extended_famous_discoveries.csv", header=None,
names=["text"])

triplets = []
for text in df["text"].dropna():

    doc = nlp(text)

    for sent in doc.sents:

        subj = [t.text for t in sent if "subj" in t.dep_]
        obj = [t.text for t in sent if "obj" in t.dep_]
        verb = [t.lemma_ for t in sent if t.pos_ ==
"VERB"]

        if subj and obj and verb:
```

```
triplets.append({"Subject": subj[0],  
"Predicate": verb[0], "Object": obj[0], "Sentence":  
sent.text})
```

```
pd.DataFrame(triplets).to_csv(r"C:\Users\91984\N  
ew folder\Clean_NLP_Project\output\triplets.csv",  
index=False)
```

```
print(" Triplets saved successfully!")
```

(c) Knowledge Graph Visualization

Libraries Used: networkx,pandas.

Goal: Display extracted entities and their relationships in a connected graph.

Code Example:

```
import pandas as pd  
import networkx as nx  
from pyvis.network import Network
```

```
triplets_path = r"C:\Users\91984\New  
folder\Clean_NLP_Project\output\triplets_extrac  
ted.csv"
```

```
df = pd.read_csv(triplets_path)
```

```
G = nx.DiGraph()
```

```
for _, row in df.iterrows():
```

```
    subj, pred, obj = row["Subject"],  
    row["Predicate"], row["Object"]
```

```
    G.add_node(subj, title=subj, color="#6EC1E4")
```

```
    G.add_node(obj, title=obj, color="#F7B32B")
```

```
    G.add_edge(subj, obj, title=pred, label=pred)
```

```
net = Network(height="700px", width="100%",  
directed=True, bgcolor="#ffffff",  
font_color="black")
```

```
net.from_nx(G)
```

```
output_html = r"C:\Users\91984\New  
folder\Clean_NLP_Project\output\knowledge_grap  
h.html"  
  
net.show(output_html)
```

```
print(f' Knowledge graph saved and opened:  
{output_html}')
```

Our CSV Files:

a) Triplets

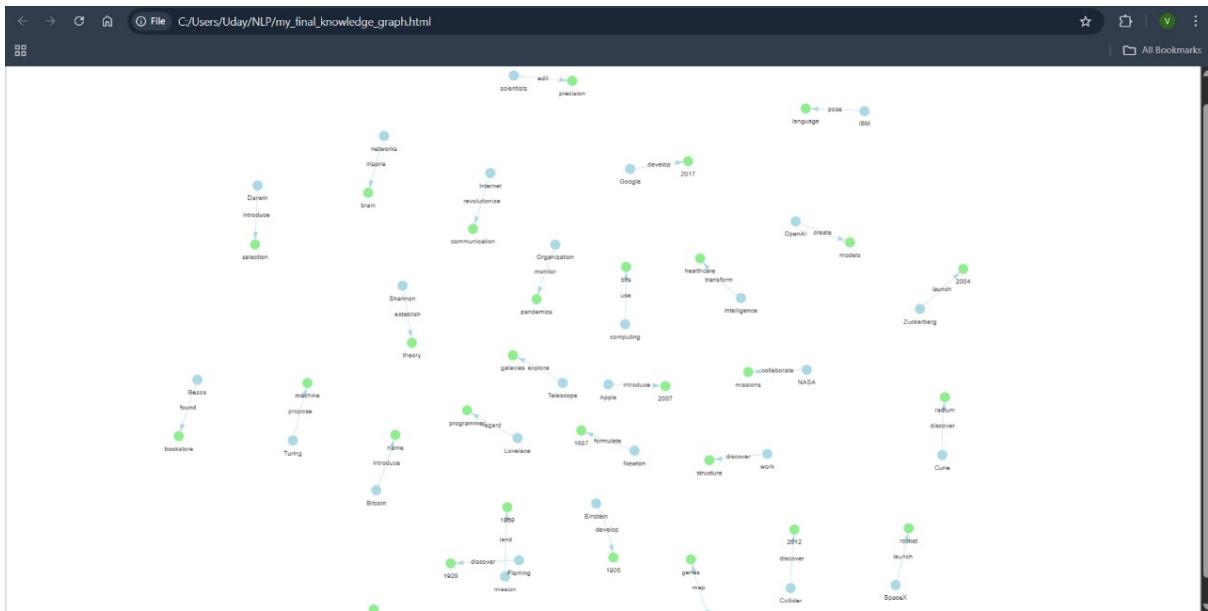
| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|--------------|-------------|---------------|---|---|---|---|---|---|---|---|---|
| 1 | Subject | Predicate | Object | Sentence | | | | | | | | |
| 2 | Einstein | develop | 1905 | Albert Einstein developed the theory of relativity in 1905. | | | | | | | | |
| 3 | Newton | formulate | 1687 | Isaac Newton formulated the laws of motion in 1687. | | | | | | | | |
| 4 | Curie | discover | radium | Marie Curie discovered radium and polonium. | | | | | | | | |
| 5 | Darwin | introduce | selection | Charles Darwin introduced the concept of evolution through natural selection. | | | | | | | | |
| 6 | Turing | propose | machine | Alan Turing proposed the idea of a universal computing machine. | | | | | | | | |
| 7 | Musk | found | 2002 | Elon Musk founded SpaceX in 2002. | | | | | | | | |
| 8 | SpaceX | launch | rocket | SpaceX launched the Falcon 9 rocket successfully. | | | | | | | | |
| 9 | NASA | collaborate | missions | NASA collaborates with SpaceX for ISS missions. | | | | | | | | |
| 10 | Google | develop | 2017 | Google developed the Transformer model architecture in 2017. | | | | | | | | |
| 11 | OpenAI | create | models | OpenAI created the GPT series of large language models. | | | | | | | | |
| 12 | press | invent | century | The printing press was invented by Johannes Gutenberg in the 15th century. | | | | | | | | |
| 13 | Internet | revolution | communication | The Internet revolutionized communication and information sharing. | | | | | | | | |
| 14 | Intelligence | transform | healthcare | Artificial Intelligence is transforming healthcare and education. | | | | | | | | |
| 15 | networks | inspire | brain | Neural networks are inspired by the structure of the human brain. | | | | | | | | |
| 16 | Organization | monitor | pandemic | The World Health Organization monitors global pandemics. | | | | | | | | |
| 17 | Tesla | alternate | electricity | Nikola Tesla contributed to the development of alternating current electricity. | | | | | | | | |
| 18 | Fleming | discover | 1928 | Alexander Fleming discovered penicillin in 1928. | | | | | | | | |
| 19 | brothers | power | 1903 | Wright brothers achieved the first powered flight in 1903. | | | | | | | | |
| 20 | Lee | invent | 1989 | Tim Berners-Lee invented the World Wide Web in 1989. | | | | | | | | |
| 21 | Hawking | advance | holes | Stephen Hawking advanced the understanding of black holes. | | | | | | | | |
| 22 | Edison | improve | bulb | Thomas Edison invented the phonograph and improved the electric light bulb. | | | | | | | | |
| 23 | work | discover | structure | Rosalind Franklin's work was crucial to discovering the DNA double helix structure. | | | | | | | | |
| 24 | Apple | introduce | 2007 | Apple introduced the first iPhone in 2007. | | | | | | | | |
| 25 | IBM | pose | language | IBM developed Watson, an AI capable of answering questions posed in natural language. | | | | | | | | |
| 26 | Bezos | found | bookstore | Jeff Bezos founded Amazon in 1994 as an online bookstore. | | | | | | | | |
| 27 | Zuckerberg | launch | 2004 | Mark Zuckerberg launched Facebook in 2004. | | | | | | | | |
| 28 | Shannon | establish | theory | Claude Shannon established the foundations of information theory. | | | | | | | | |
| 29 | scientists | edit | precision | CRISPR technology allows scientists to edit genes with precision. | | | | | | | | |
| 30 | Lovelace | regard | programmer | Ada Lovelace is regarded as the first computer programmer. | | | | | | | | |
| 31 | Collider | discover | 2012 | The Large Hadron Collider helped discover the Higgs boson particle in 2012. | | | | | | | | |
| 32 | mission | land | 1969 | NASA's Apollo 11 mission landed humans on the Moon in 1969. | | | | | | | | |
| 33 | Bitcoin | introduce | name | Bitcoin was introduced in 2009 by an unknown person under the name Satoshi Nakamoto. | | | | | | | | |
| 34 | Project | map | genes | The Human Genome Project was completed in 2003, mapping all human genes. | | | | | | | | |
| 35 | Telescope | explore | galaxies | The James Webb Space Telescope was launched to explore distant galaxies. | | | | | | | | |
| 36 | computing | use | bits | Quantum computing aims to revolutionize data processing using quantum bits. | | | | | | | | |
| 37 | | | | | | | | | | | | |

b) Entities Extraction

| Entity Text | Entity Label | Sentence |
|-------------------------------|--------------|---|
| Albert Einstein | PERSON | Albert Einstein developed the theory of relativity in 1905. |
| | 1905 DATE | Albert Einstein developed the theory of relativity in 1905. |
| Isaac Newton | PERSON | Isaac Newton formulated the laws of motion in 1687. |
| | 1687 DATE | Isaac Newton formulated the laws of motion in 1687. |
| Marie Curie | PERSON | Marie Curie discovered radium and polonium. |
| Charles Darwin | PERSON | Charles Darwin introduced the concept of evolution through natural selection. |
| Alan Turing | PERSON | Alan Turing proposed the idea of a universal computing machine. |
| Elon Musk | PERSON | Elon Musk founded SpaceX in 2002. |
| | 2002 DATE | Elon Musk founded SpaceX in 2002. |
| SpaceX | PERSON | SpaceX launched the Falcon 9 rocket successfully. |
| Falcon | ORG | SpaceX launched the Falcon 9 rocket successfully. |
| NASA | ORG | NASA collaborates with SpaceX for ISS missions. |
| ISS | ORG | NASA collaborates with SpaceX for ISS missions. |
| Google | ORG | Google developed the Transformer model architecture in 2017. |
| Transformer | ORG | Google developed the Transformer model architecture in 2017. |
| | 2017 DATE | Google developed the Transformer model architecture in 2017. |
| OpenAI | PERSON | OpenAI created the GPT series of large language models. |
| GPT | ORG | OpenAI created the GPT series of large language models. |
| Johannes Gutenberg | PERSON | The printing press was invented by Johannes Gutenberg in the 15th century. |
| the 15th century | DATE | The printing press was invented by Johannes Gutenberg in the 15th century. |
| Artificial Intelligence | PERSON | Artificial Intelligence is transforming healthcare and education. |
| The World Health Organization | ORG | The World Health Organization monitors global pandemics. |
| Nikola Tesla | ORG | Nikola Tesla contributed to the development of alternating current electricity. |
| Alexander Fleming | PERSON | Alexander Fleming discovered penicillin in 1928. |
| | 1928 DATE | Alexander Fleming discovered penicillin in 1928. |
| Wright | PERSON | Wright brothers achieved the first powered flight in 1903. |
| first | ORDINAL | Wright brothers achieved the first powered flight in 1903. |
| | 1903 DATE | Wright brothers achieved the first powered flight in 1903. |
| Tim Berners-Lee | PERSON | Tim Berners-Lee invented the World Wide Web in 1989. |
| the World Wide Web | EVENT | Tim Berners-Lee invented the World Wide Web in 1989. |
| | 1989 DATE | Tim Berners-Lee invented the World Wide Web in 1989. |
| Stephen Hawking | PERSON | Stephen Hawking advanced the understanding of black holes. |
| Thomas Edison | PERSON | Thomas Edison invented the phonograph and improved the electric light bulb. |
| Franklin | ORG | Rosalind Franklin's work was crucial to discovering the DNA double helix structure. |
| Apple | ORG | Apple introduced the first iPhone in 2007. |
| first | ORDINAL | Apple introduced the first iPhone in 2007. |

| | | |
|----------------------------|-------------|---|
| first | ORDINAL | Apple introduced the first iPhone in 2007. |
| | 2007 DATE | Apple introduced the first iPhone in 2007. |
| IBM | ORG | IBM developed Watson, an AI capable of answering questions posed in natural language. |
| Watson | PERSON | IBM developed Watson, an AI capable of answering questions posed in natural language. |
| AI | GPE | IBM developed Watson, an AI capable of answering questions posed in natural language. |
| Jeff Bezos | PERSON | Jeff Bezos founded Amazon in 1994 as an online bookstore. |
| Amazon | ORG | Jeff Bezos founded Amazon in 1994 as an online bookstore. |
| | 1994 DATE | Jeff Bezos founded Amazon in 1994 as an online bookstore. |
| Mark Zuckerberg | PERSON | Mark Zuckerberg launched Facebook in 2004. |
| | 2004 DATE | Mark Zuckerberg launched Facebook in 2004. |
| Claude Shannon | PERSON | Claude Shannon established the foundations of information theory. |
| CRISPR | ORG | CRISPR technology allows scientists to edit genes with precision. |
| Ada Lovelace | PERSON | Ada Lovelace is regarded as the first computer programmer. |
| first | ORDINAL | Ada Lovelace is regarded as the first computer programmer. |
| Hadron Collider | PERSON | The Large Hadron Collider helped discover the Higgs boson particle in 2012. |
| Higgs | NORP | The Large Hadron Collider helped discover the Higgs boson particle in 2012. |
| | 2012 DATE | The Large Hadron Collider helped discover the Higgs boson particle in 2012. |
| NASA | ORG | NASA's Apollo 11 mission landed humans on the Moon in 1969. |
| Apollo | ORG | NASA's Apollo 11 mission landed humans on the Moon in 1969. |
| | 11 CARDINAL | NASA's Apollo 11 mission landed humans on the Moon in 1969. |
| Moon | PERSON | NASA's Apollo 11 mission landed humans on the Moon in 1969. |
| | 1969 DATE | NASA's Apollo 11 mission landed humans on the Moon in 1969. |
| Bitcoin | PERSON | Bitcoin was introduced in 2009 by an unknown person under the name Satoshi Nakamoto. |
| | 2009 DATE | Bitcoin was introduced in 2009 by an unknown person under the name Satoshi Nakamoto. |
| Satoshi Nakamoto | PERSON | Bitcoin was introduced in 2009 by an unknown person under the name Satoshi Nakamoto. |
| The Human Genome Project | ORG | The Human Genome Project was completed in 2003, mapping all human genes. |
| | 2003 DATE | The Human Genome Project was completed in 2003, mapping all human genes. |
| James Webb Space Telescope | PERSON | The James Webb Space Telescope was launched to explore distant galaxies. |
| Quantum | ORG | Quantum computing aims to revolutionize data processing using quantum bits. |

Knowledge Graph Output:



Reflection:

What did We Learned from this Milestone – I?

1. We learned how to use spaCy for text understanding.
 2. Understood how to design a relation extraction pipeline.
 3. Gained hands-on experience in dataset processing and NLP pipeline creation.

THANK YOU