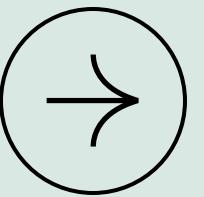
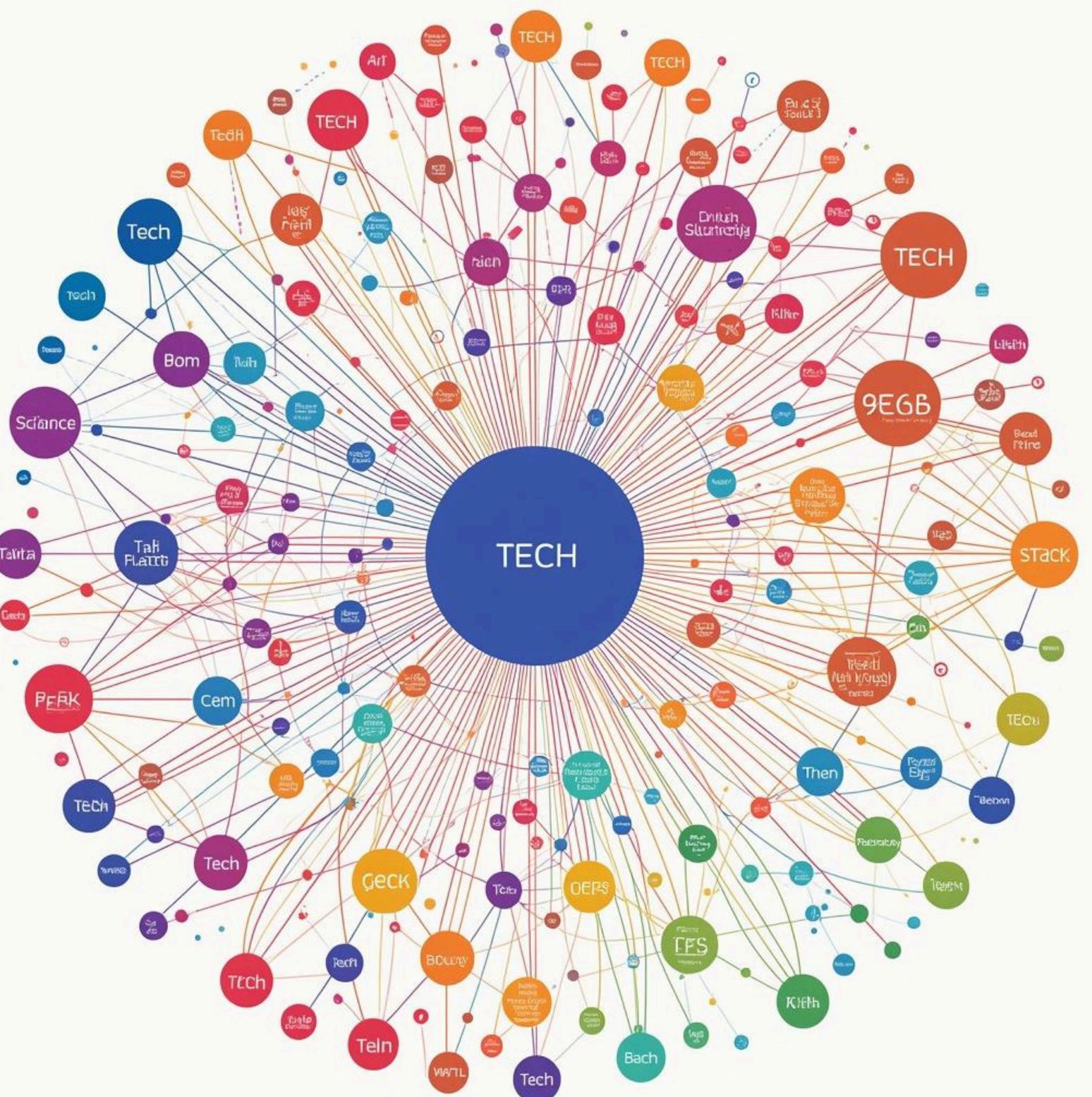


Infosys KnowMap Project



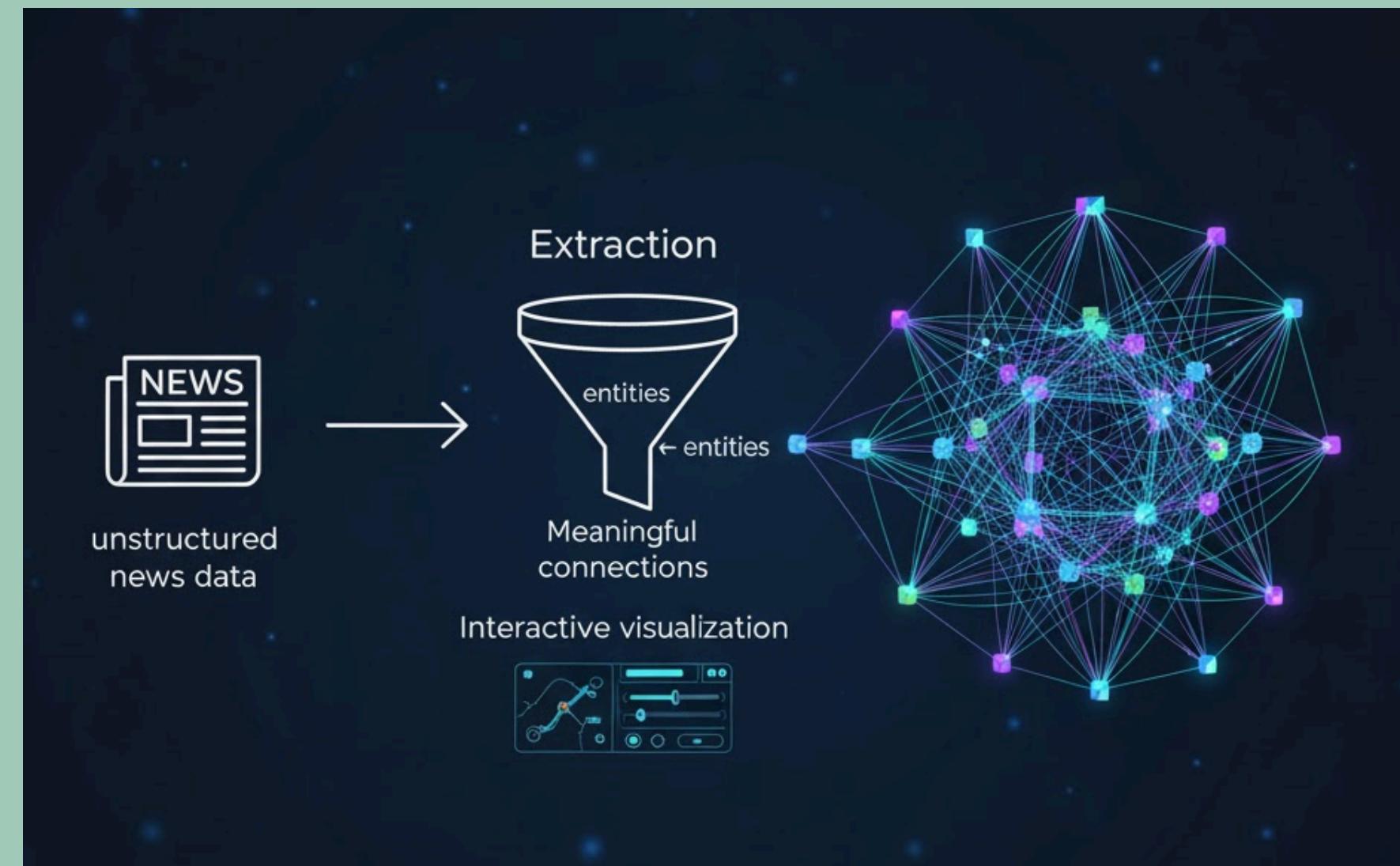
Cross Domain Knowledge Graph

PRESENTED BY GROUP 5



Introduction

- A Knowledge Graph (KG) is a structured representation of information where entities are nodes and relationships are edges.
- KGs help machines understand and reason about real-world concepts.
- Goal of this project:
- To extract meaningful connections from unstructured news articles and visualize them as an interactive knowledge graph.

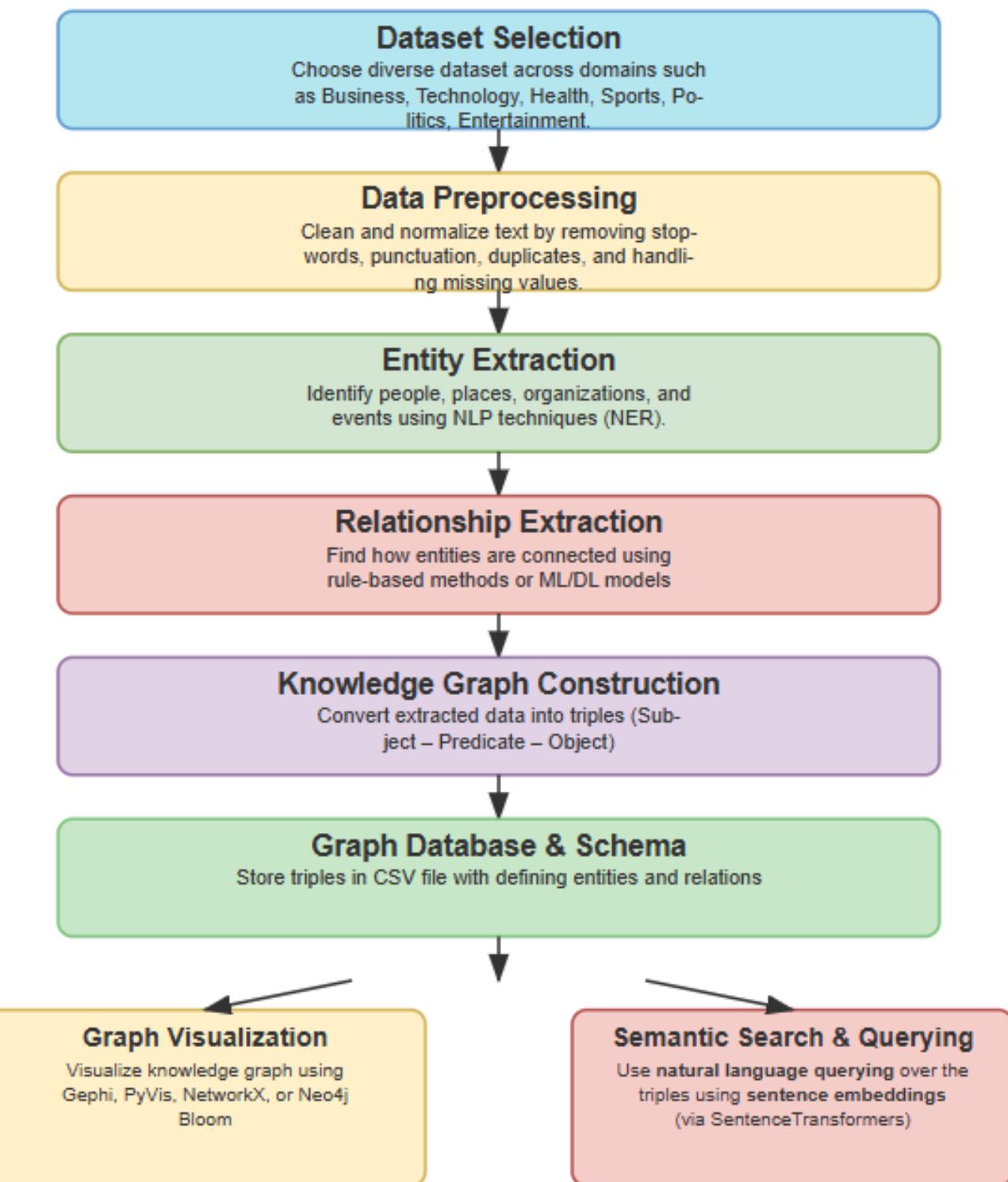


Project Recap

OVERVIEW OF ACCOMPLISHMENTS AND OBJECTIVES

The **KnowMap project aims to create** a Cross-Domain Knowledge Graph. We have successfully completed essential components including data preprocessing, Named Entity Recognition (NER), Relation Extraction (RE), and the visualization of our knowledge graph, leading to insightful data connections across multiple domains.

KnowMap – Cross-Domain Knowledge Graph Workflow Steps



Dataset Description

Dataset: Kaggle News Category Dataset

- Size: ~200,000 articles
- Fields:
- Headline
- Short Description
- Category
- Date
- Authors
- Link

This dataset is ideal since it covers multi-domain real-world events.

A	B	C	D	E	F
link	headline	category	short_description	authors	date
2	https://www.abcnews.com/over-4-million-americans-had-a-	U.S. NEWS	Health experts said it was a 'public health emergency.'	Carla K. Johnson, AP	23-09-2022
3	https://www.abcnews.com/american-airlines-flyer-picked-up-by-	U.S. NEWS	He was subdued by passengers on board.	Mary Papenfuss, AP	23-09-2022
4	https://www.abcnews.com/23-of-the-funniest-tweets-from-the-week-	COMEDY	"Until you have a dog."	Elyse Wanshel, AP	23-09-2022
5	https://www.abcnews.com/the-funniest-tweets-from-the-week-	PARENTING	"Accidentally put groceries in the trash."	Caroline Bologna, AP	23-09-2022
6	https://www.abcnews.com/woman-who-called-cop-was-accused-of-	U.S. NEWS	Amy Cooper accused of袭警.	Nina Golgowski, AP	22-09-2022
7	https://www.abcnews.com/cleaner-was-dead-in-bell-monte-carlo-	U.S. NEWS	The 63-year-old woman was seen working.	AP	22-09-2022
8	https://www.abcnews.com/reporter-gets-adorable-son-after-	U.S. NEWS	"Who's that behind you?"	Elyse Wanshel, AP	22-09-2022
9	https://www.abcnews.com/puerto-ricans-desperate-to-vote-in-	WORLD NEWS	More than half a million people.	DANICA COTO, AP	22-09-2022
10	https://www.abcnews.com/how-a-new-documentary-is-changing-	CULTURE & ARTS	In "Mija," director Isabella Marina Fang.	Marina Fang, AP	22-09-2022
11	https://www.abcnews.com/biden-at-un-to-call-russia-a-war-crime-	WORLD NEWS	White House official Aamer Madhani, AP	Aamer Madhani, AP	21-09-2022
12	https://www.abcnews.com/world-cup-captains-war-in-iran-world-	WORLD NEWS	FIFA has come under fire.	GRAHAM DUNBAR, AP	21-09-2022
13	https://www.abcnews.com/man-sets-himself-on-fire-world-news-	WORLD NEWS	The incident underscored the risks.	Mari Yamaguchi, AP	21-09-2022
14	https://www.abcnews.com/fiona-threatens-to-become-a-major-	WORLD NEWS	Hurricane Fiona lashes the Caribbean.	DANICA COTO, AP	21-09-2022
15	https://www.abcnews.com/twitch-bans-gambling-site-world-news-	TECH	One man's claims that Ben Blanchet.	AP	21-09-2022
16	https://www.abcnews.com/virginia-thomas-agrees-to-serve-on-the-	U.S. NEWS	Conservative activist Eric Tucker and M	Eric Tucker and M	21-09-2022
17	https://www.abcnews.com/russian-cosmonaut-vale-world-news-	WORLD NEWS	Polyakov's record-breaking flight.	Marco Margaritoff, AP	20-09-2022
18	https://www.abcnews.com/reboot-is-a-clever-and-entertaining-	CULTURE & ARTS	Starring Keegan-Michael Key and Jordan.	Marina Fang and C	20-09-2022
19	https://www.abcnews.com/maury-wills-base-stealing-world-news-	SPORTS	Maury Wills, who helped Beth Harris, AP	Beth Harris, AP	20-09-2022
20	https://www.abcnews.com/4-russian-controlled-ukraine-world-news-	WORLD NEWS	The concerted and quick Jon Gambrell, AP	Jon Gambrell, AP	20-09-2022
21	https://www.abcnews.com/fiona-barrels-toward-turks-and-caicos-world-	WORLD NEWS	The Turks and Caicos Islands.	DANICA COTO, AP	20-09-2022
22	https://www.abcnews.com/golden-globes-returning-world-news-	ENTERTAINMENT	For the past 18 months, Hollywood has.	AP	20-09-2022
23	https://www.abcnews.com/biden-says-u-s-forces-will-not-be-used-in-	POLITICS	President issues vow as tensions with C	AP	19-09-2022
24	https://www.abcnews.com/hurricane-fiona-bears-down-world-news-	WORLD NEWS	The storm knocked down power lines.	DANICA COTO, AP	19-09-2022
25	https://www.abcnews.com/one-dead-after-plane-crash-world-news-	U.S. NEWS	The single-engine plane.	Nina Golgowski, AP	19-09-2022
26	https://www.abcnews.com/beautiful-and-sad-at-politics-world-news-	POLITICS	An annual celebration.	Jonathan Nichols, AP	19-09-2022
27	https://www.abcnews.com/family-of-american-held-hostage-world-news-	U.S. NEWS	Mark Frerichs, a Navy SEAL.	Eric Tucker and R	19-09-2022
28	https://www.abcnews.com/las-vegas-aces-win-first-game-world-news-	SPORTS	Las Vegas never had a winning streak.	Pat Eaton-Robb, AP	19-09-2022

Data Preprocessing

Text Cleaning & Preparation

Goal: Clean and prepare text for NLP tasks

Steps performed:

- Remove punctuation, special characters, URLs
- Convert to lowercase
- Remove stopwords
- Tokenize and lemmatize text
- Deduplicate records

Libraries: pandas, re, nltk

Named Entity Recognition (NER)

Purpose: Identify entities (people, organizations, locations, etc.)

Example:

Sentence → "NASA launched a new satellite."

NER Output → [("NASA", ORG), ("satellite", PRODUCT)]

Tools Used: spaCy

Result: Extracted entities stored for relation extraction and graph creation.

Relation Extraction (RE)

Purpose: Identify relationships between recognized entities.

Example Triple:

(NASA, launched, satellite)

Methods Used:

- Dependency parsing
- Rule-based pattern matching
- Regular expressions

Result: Structured triples

(Subject–Relation–Object) ready for graph input.

Knowledge Graph Construction

Visualization Tool

Pyvis (via NetworkX + HTML integration)

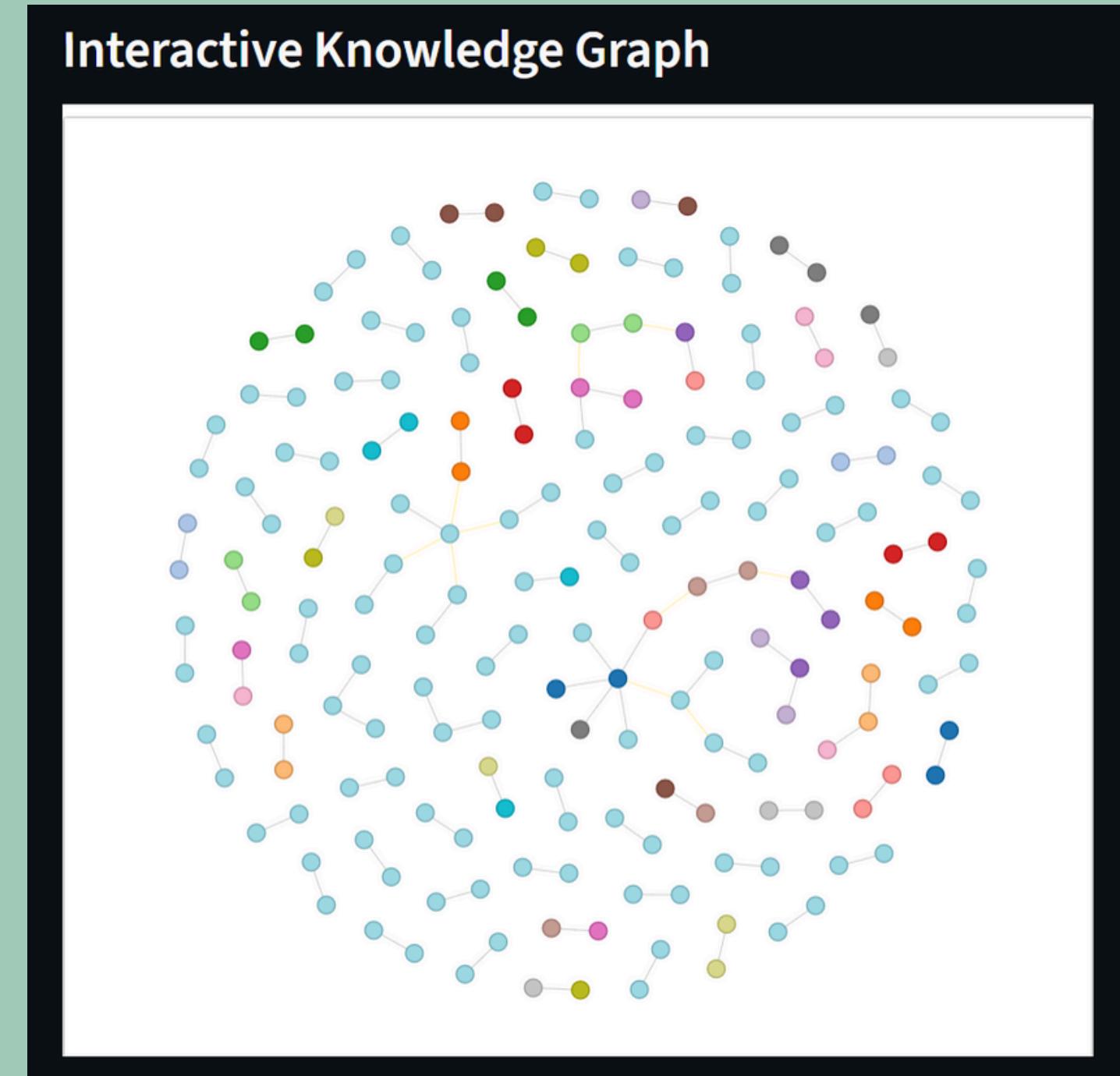
- Pyvis is a Python library for interactive network visualization in the browser.
- It converts NetworkX graphs into HTML/JavaScript visualizations.

Framework: NetworkX

- Nodes: Entities (e.g., NASA, satellite)
- Edges: Relations (e.g., launched)

Implementation:

- Each triple is added as an edge between entities.
- Edge attributes: relation type, category, date.

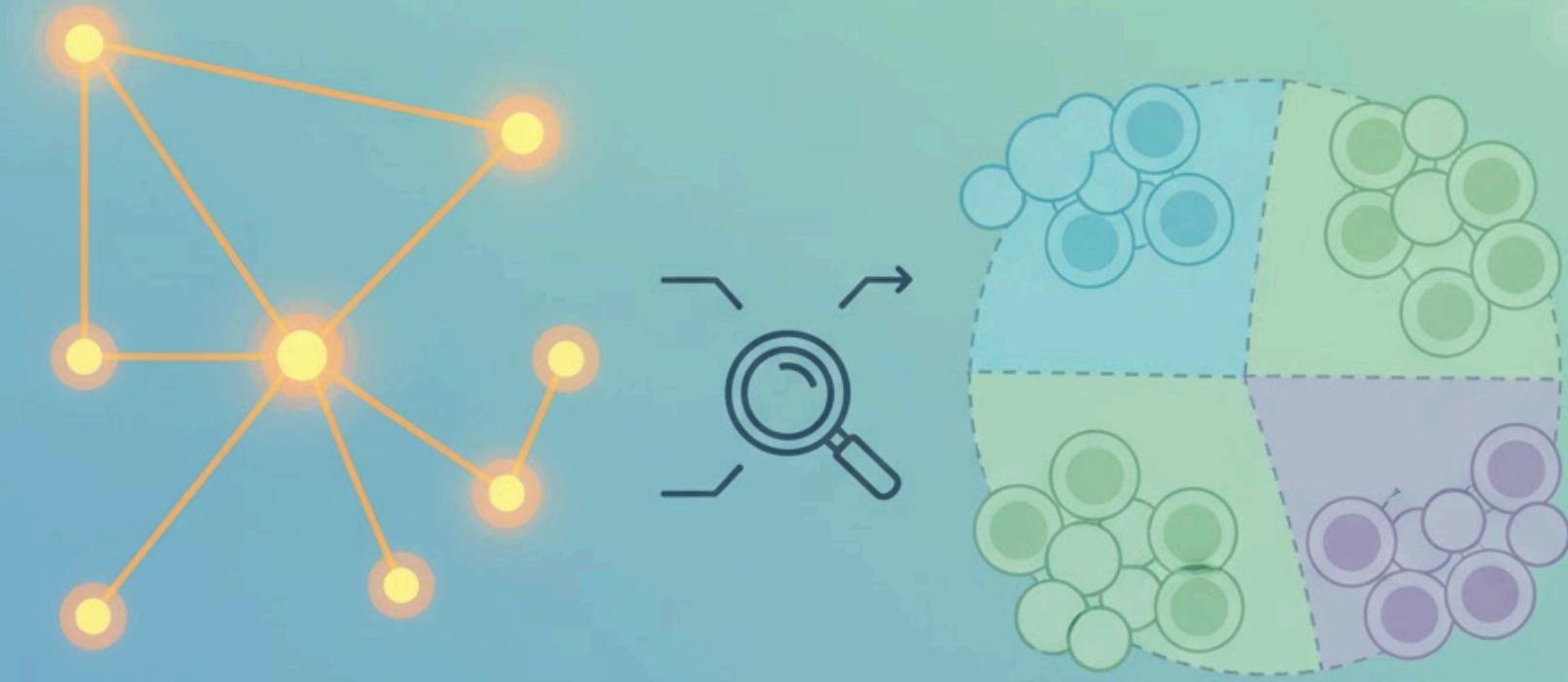


Graph Analytics Techniques

Uncovering Hidden Relationships in Data

Graph analytics techniques reveal **important entities** and relationships within data. By utilizing centrality measures and community detection, we can identify key nodes and clusters, leading to deeper insights across various domains.

GRAPH ANALYTICS TECHNIQUES



Centrality Measures:
Identifying Important Nodes
(e.g. Degree, Betweenness, Closeness)

Community Detection:
Grouping Related Nodes
(e.g. Modularity, Louvain Method)

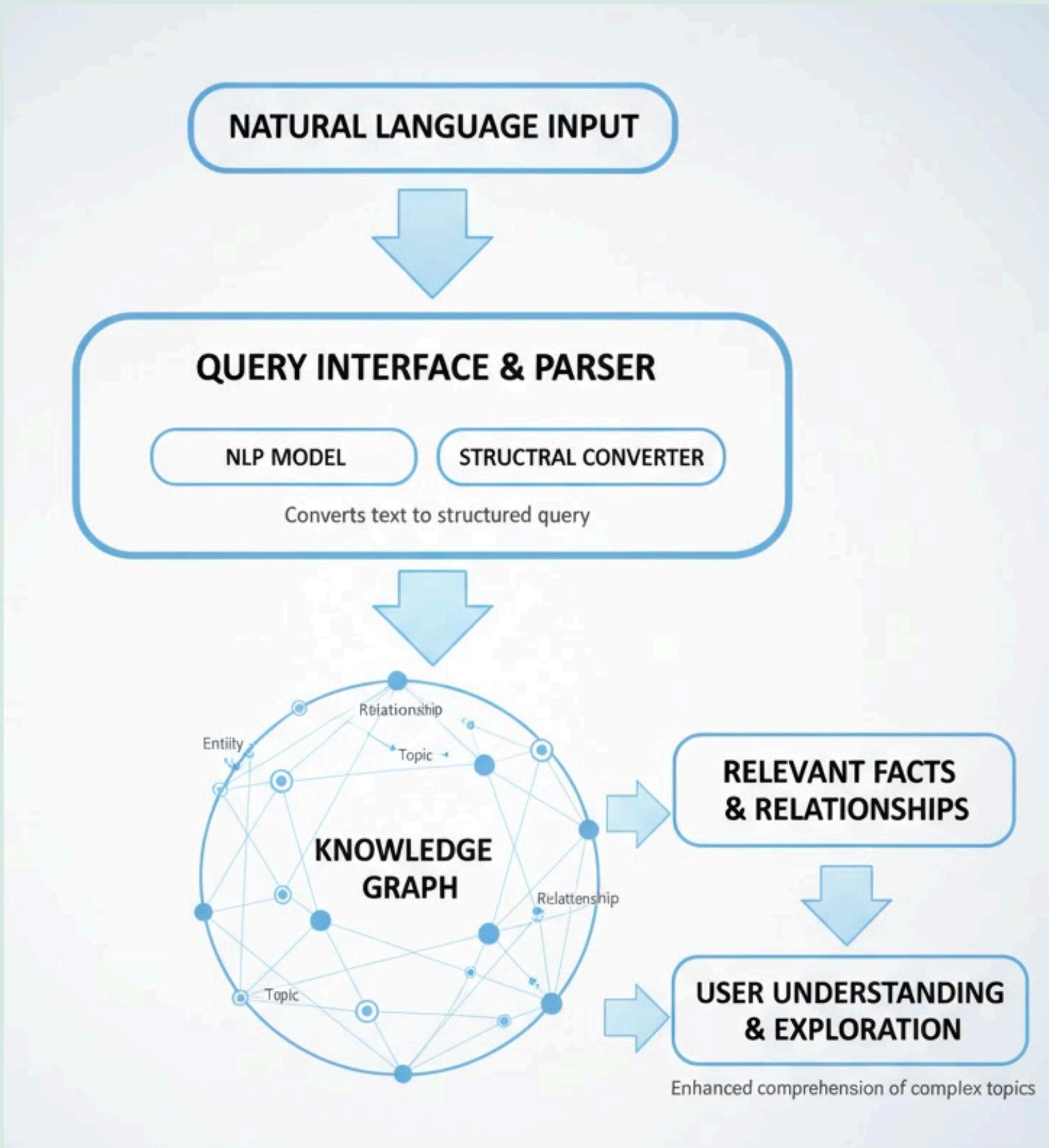
INTEGRATING KNOWLEDGE DOMAINS



Cross-Domain Linking

Integrating multiple knowledge domains reveals valuable insights and relationships. By connecting isolated graphs, such as Technology and Healthcare, we uncover how advancements in AI influence various sectors, including the economy and environmental sustainability. This integration enhances our understanding of complex interactions across diverse fields for informed decision-making.

Question Answering Capabilities

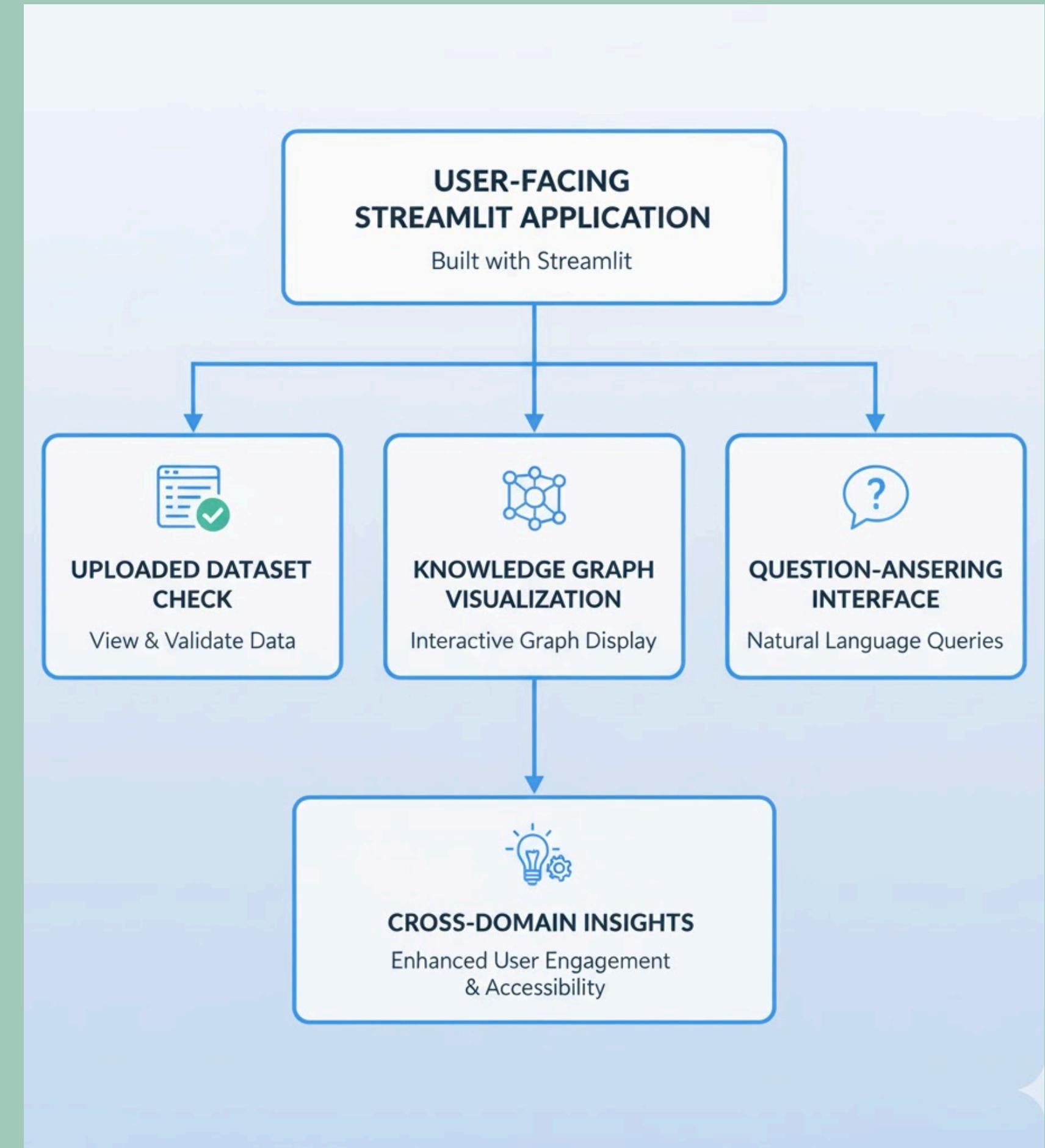


The interactive **query interface** allows users to input natural language questions or keywords, converting them into structured queries that search through the knowledge graph. This functionality enables users to efficiently retrieve relevant facts and relationships, enhancing their understanding of complex topics within the graph.

Application Deployment Overview

User-Facing App Development with Streamlit

The **user-facing application** utilizes Streamlit, allowing users to check uploaded dataset, visualize knowledge graphs, and interact with the question-answering interface. This app enhances user engagement and accessibility to cross-domain insights effectively.



Results and Insights Unveiled



10

- +

Show centrality table

Show community summary

Top 10 nodes by *degree_centrality*

	node	degree_centrality
0	Obama	0.0112
1	Democrats	0.0112
2	Americans	0.0112
3	Russia	0.0112
4	bill	0.0112
5	story	0.0112
6	Outfits	0.0112

Question Answering (QA) on Knowledge Graph

Ask a question (e.g., 'Who founded Microsoft?')

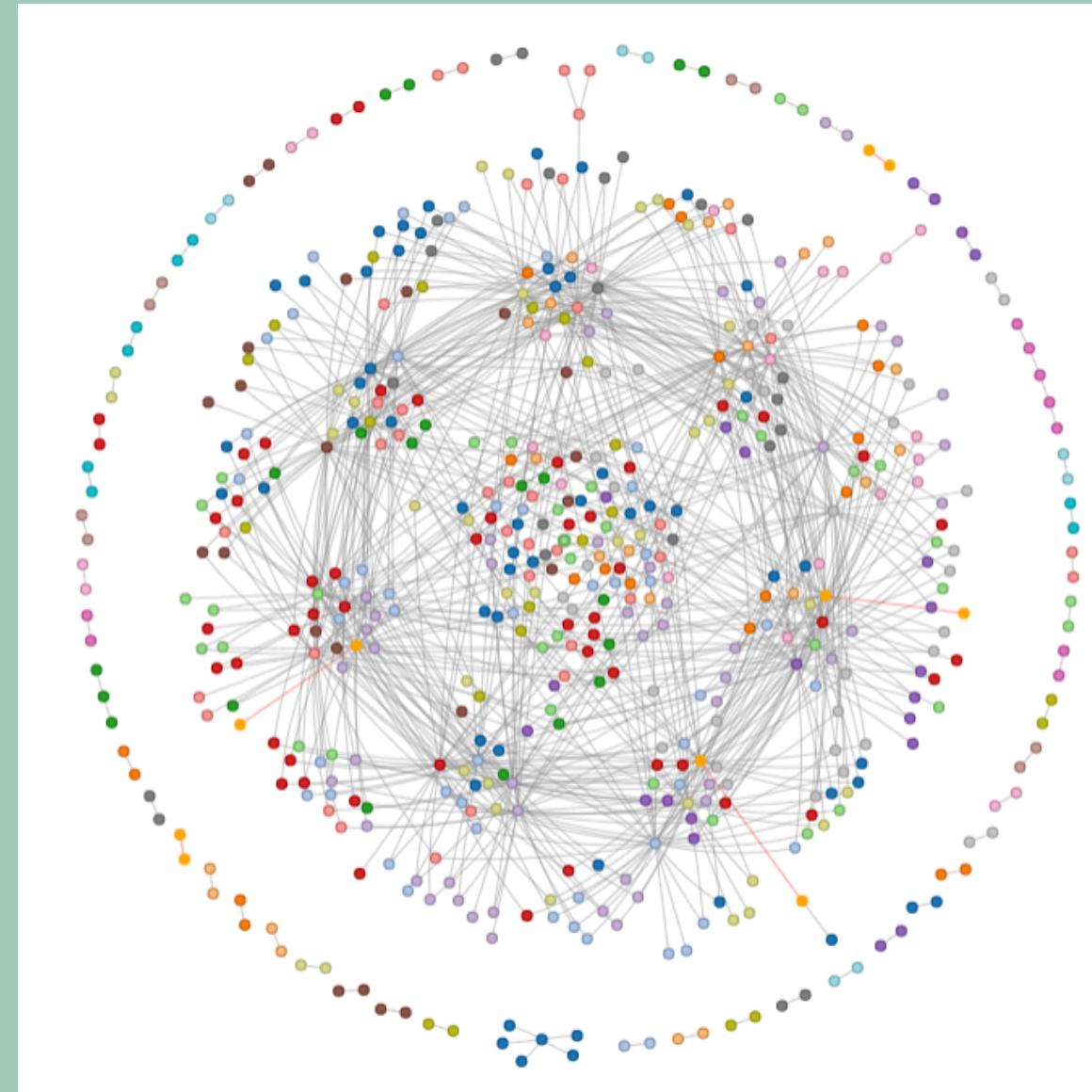
How many Americans took the Omicron COVID boosters?

Top Relevant Triples

	Triple	Category	Date	Source	Score
0	Officials Say Americans	POLITICS	2017-06-07	https://www.huffingtonpost.com/entry/intelligence-officials-can't-say-how-many-americans-take-the-omicron-booster_n_5b9dc055e4b	0.8565
1	Americans earn million	MONEY	2014-01-01	https://www.huffingtonpost.com/entry/who-stops-paying-social-security-benefits-to-americans_n_5b9dc055e4b	0.8174
2	Americans Pay Billion	POLITICS	2017-04-22	https://www.huffingtonpost.com/entry/bp-oil-spill-deepwater-horizon-billion-damages_n_5b9dc055e4b	0.8136
3	coalition killed hundreds	POLITICS	2016-10-08	https://www.huffingtonpost.com/entry/barack-obama-could-end-the-killing-in-yemen_n_5b9dc055e4b	0.8136
4	Clinton Says Assault	POLITICS	2015-11-02	https://www.huffingtonpost.com/entry/hillary-clinton-spring-valley_n_5b9dc055e4b	0.7595

Download QA Results

Future Enhancements for KnowMap



LLM-Based Q&A

Integrating large language models will enhance natural language reasoning capabilities.

Domain-Specific Dashboards

Customized dashboards will improve user experience with targeted insights and analytics.

Graph Database Backend

Deploying a graph database will ensure scalability and efficient data management.

Thank You

MILESTONE 4 OVERVIEW

Group 5
Divyansh Ghildyal
Ch. Vyshnavi
G. Pooja Chamundi