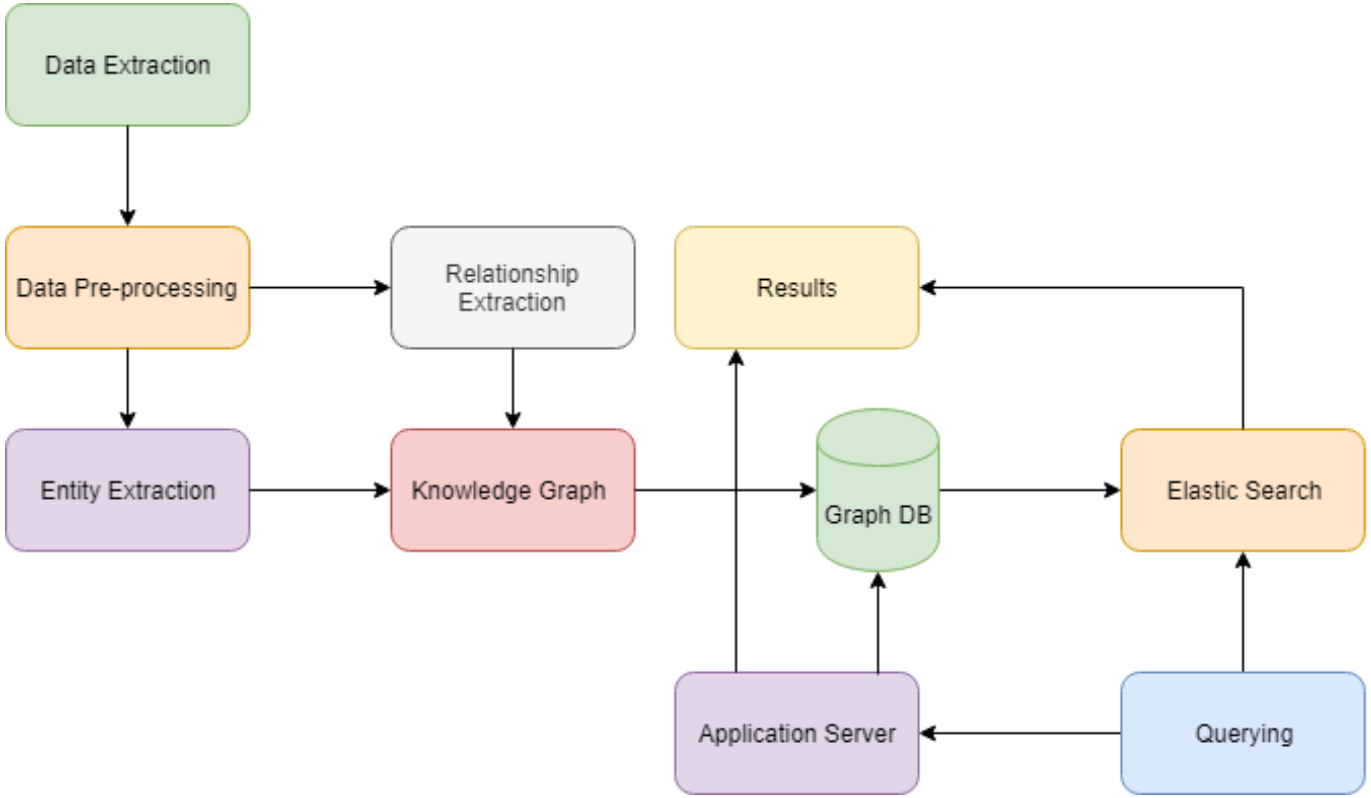
Team:
Vishnu Ramachandran (vcramach)
Sharique Khan (mkhan8)
Natansh Negi (nnegi2)

ADBI Final Capstone Project

Unstructured Text to Knowledge Graphs using NLP tools

Flow Chart


Data Extraction
Data Pre-processing
Relationship Extraction
Results
Entity Extraction
Knowledge Graph
Graph DB
Elastic Search
Application Server
Querying

Data Extraction

Extracted Movie database from Wikipedia.

Dataset: Summary/Synopsis for 9000 movies.

Tool used for extraction : Scrapy

Data Preprocessing

Using Python's nltk package did the following:

- Eliminated stop words.
- Tokenization & Stemming.
- Case, negation and punctuation handling.
- Transformation(removed unnecessary attributes).

Entity Extraction

Tasks of Named Entity Extraction:

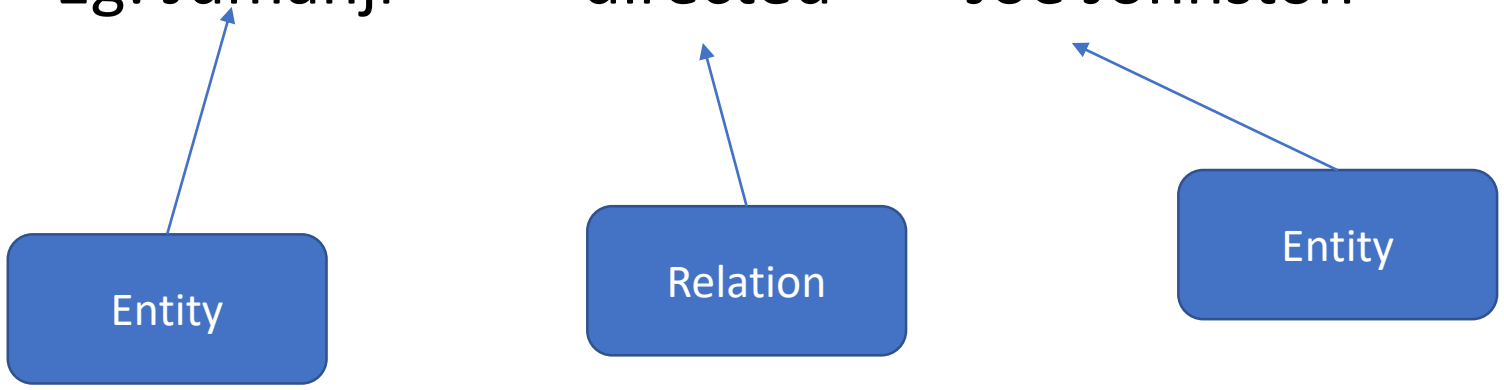
- Extract Names.
- Classify the extracted names.

Tool Used: Spacy

Relation Extraction

- Returns the relations between two entities.
- Designed an algorithm to extract the relations between entities returned by Spacy using POS tagging.

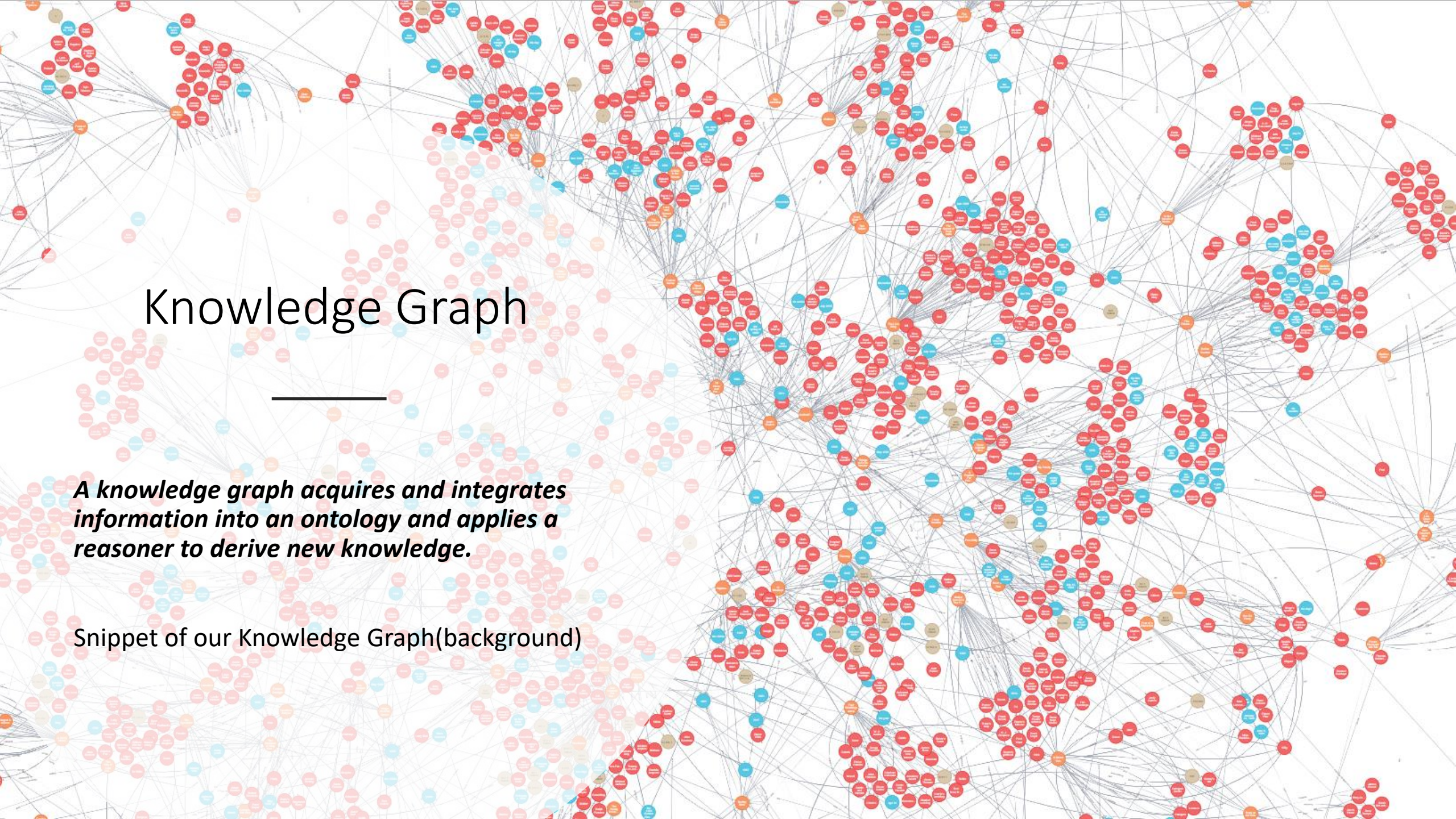
- Eg: Jumanji directed Joe Johnston



Knowledge Graph

A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge.

Snippet of our Knowledge Graph(background)



Query

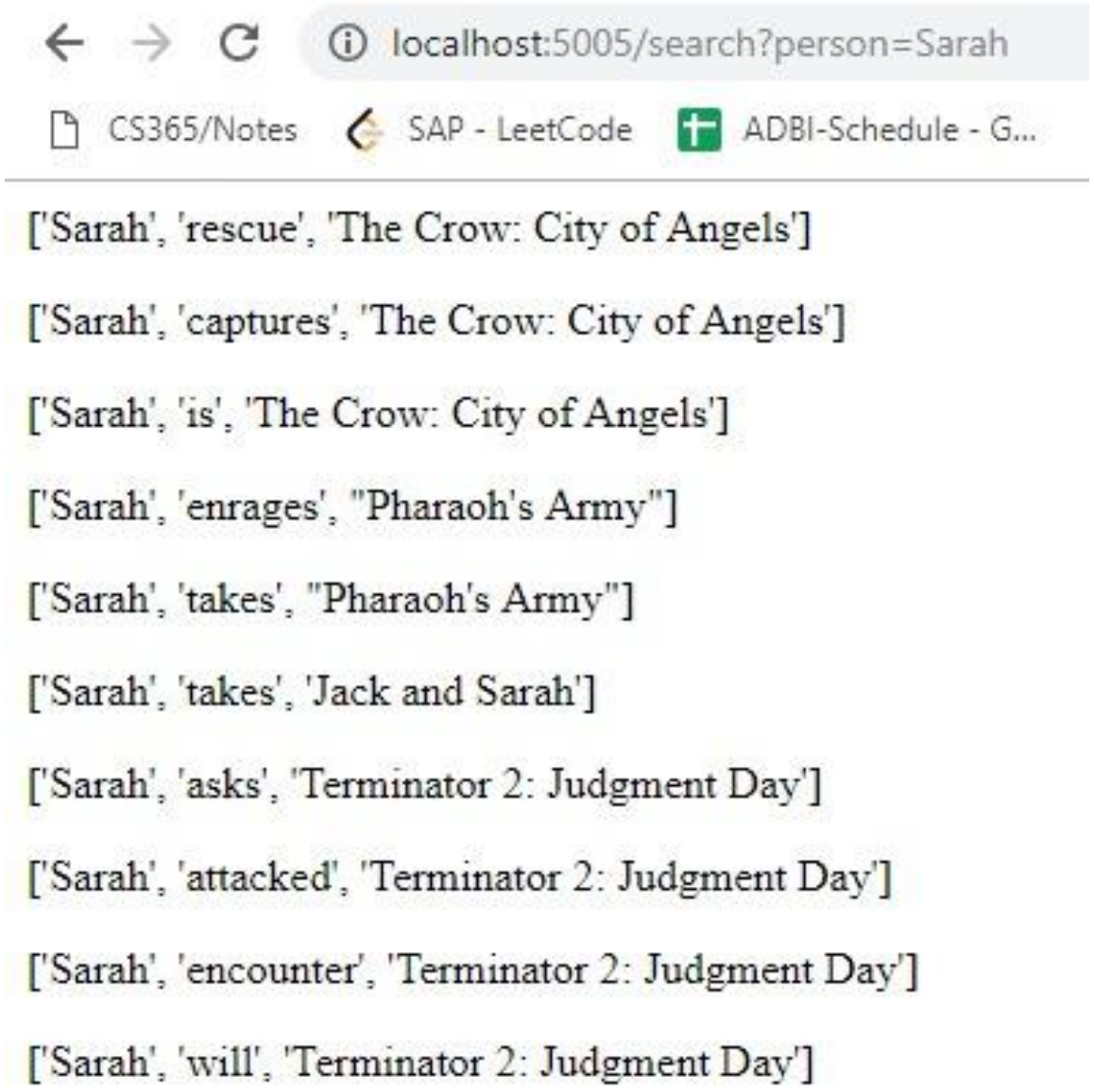
Querying our knowledge Graph using Google's Elasticsearch.

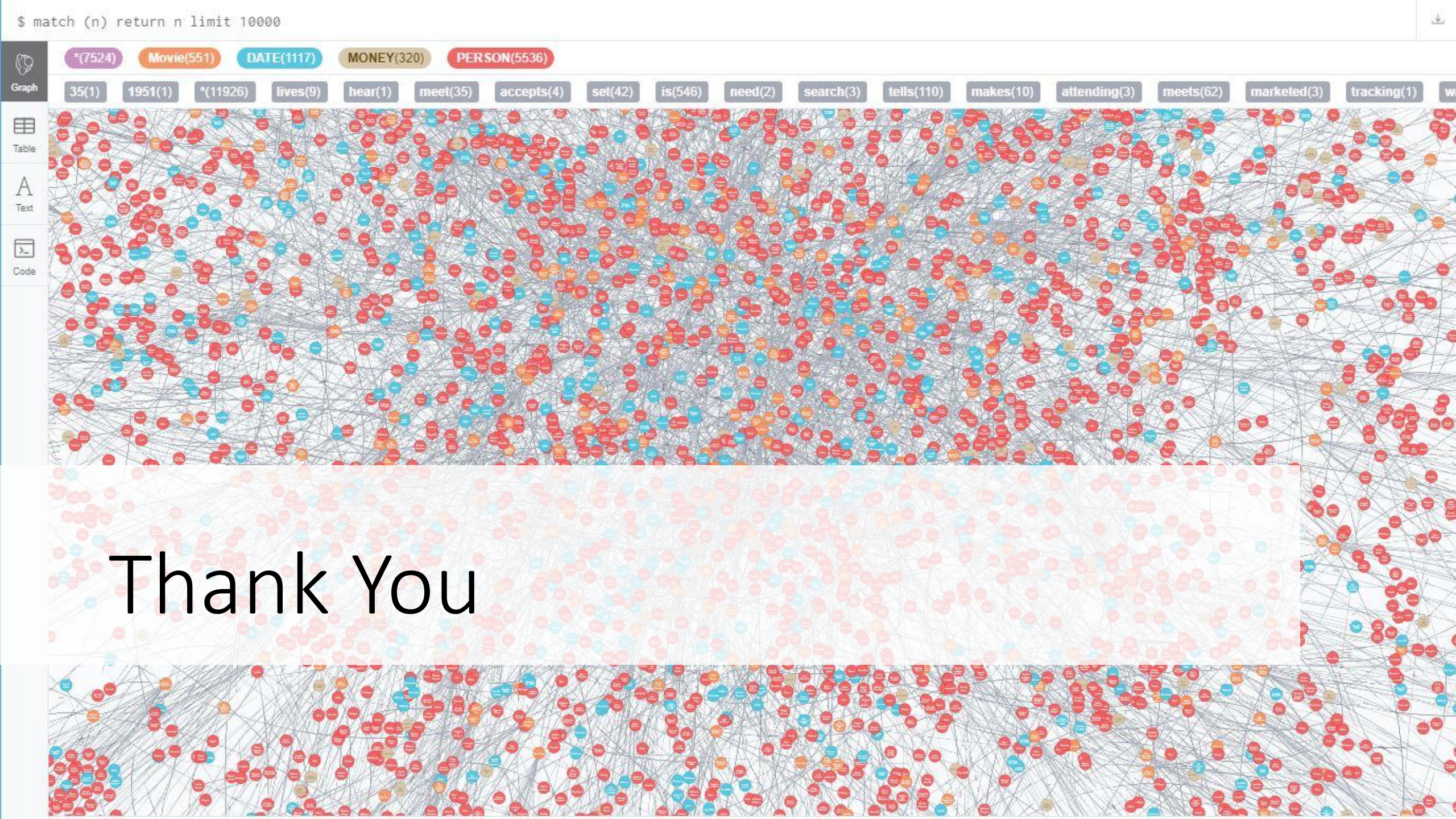
Parameters available for querying :

- Movie
- Date
- Person

Results

- We query on the knowledge graph using the Person ‘Sarah.’
- This returns the entity relationships and extended attribute relations deduced from random walks through the knowledge graph.




$ match (n) return n limit 10000
Graph
Table
Text
Code
*(7524)
Movie(551)
DATE(1117)
MONEY(320)
PERSON(5536)
35(1)
1951(1)
*(11926)
lives(9)
hear(1)
meet(35)
accepts(4)
set(42)
is(546)
need(2)
search(3)
tells(110)
makes(10)
attending(3)
meets(62)
marketed(3)
tracking(1)

Thank You