

Summer Internship Project Presentation

Gaurav Dabas

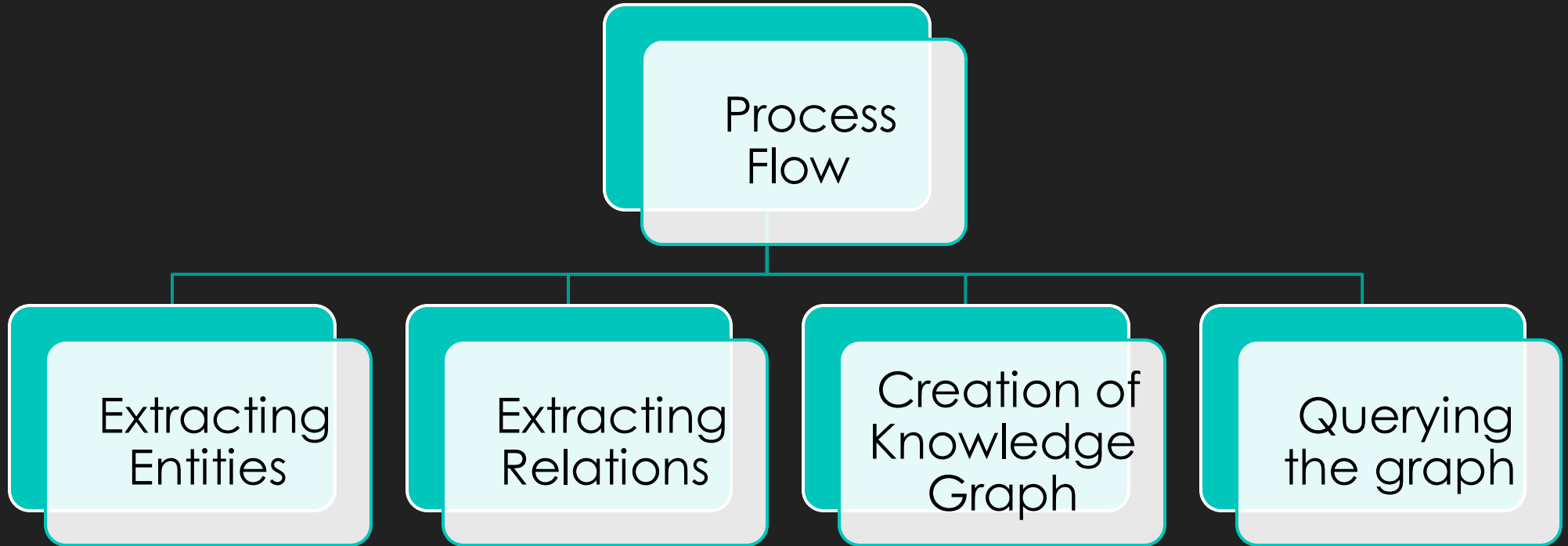
To Analyse text documents and create a knowledge graph based querying capability on drilling documents.

Expectation from the project:

- Processing raw data to structured data in the forms of entity and relations to create a semantic knowledge graph.
- Creating a framework that is capable of retrieving answers from the knowledge graph when queried in natural language form.

Methodology

- Given some data (mostly unstructured), after processing the data, we acquire useful information by using Named Entity recognition and Relation Extraction.
- Using this structured data, we model our knowledge graph model and enable embedding of rich semantic data.
- Finally for querying the data, natural language questions are translated into executable queries for answer retrieval.
- Timeliness – 5 days (As I had to study courses on NLP from scratch on udemy and coursera)



Extracting Relations and Entities

- Tried various libraries such as StanfordNLP, nltk, spacy for the following task.
- Finally, Extracted entities and relations from the text using Spacy library in Python.
- Used the POS (parts of speech) tagging feature of the library.
- Recognized subject, object and predicate (verb) from each of the sentences after tokenizing.
- Created triples by combining all three of these.
- Timeliness – 5 days (As had to change the approach as first ones did not comply with the further needs even though it gave better results)

Index ▲	Type	Size	Value
0	str	1	confused and frustrated, connie decides to leave on her own.
1	str	1	later, a woman's scream is heard in the distance.
2	str	1	christian is then paralyzed by an elder.
3	str	1	the temple is set on fire.
4	str	1	outside, the cult wails with him.
5	str	1	it's a parable of a woman's religious awakening- c. mackenzie, and cra ...
6	str	1	later, craig di francia and action bronson were revealed to have joine ...
7	str	1	sebastian maniscalco and paul ben-victor were later revealed as being ...
8	str	1	we just tried to make the film.
9	str	1	we went through all these tests and things m global was also circling ...

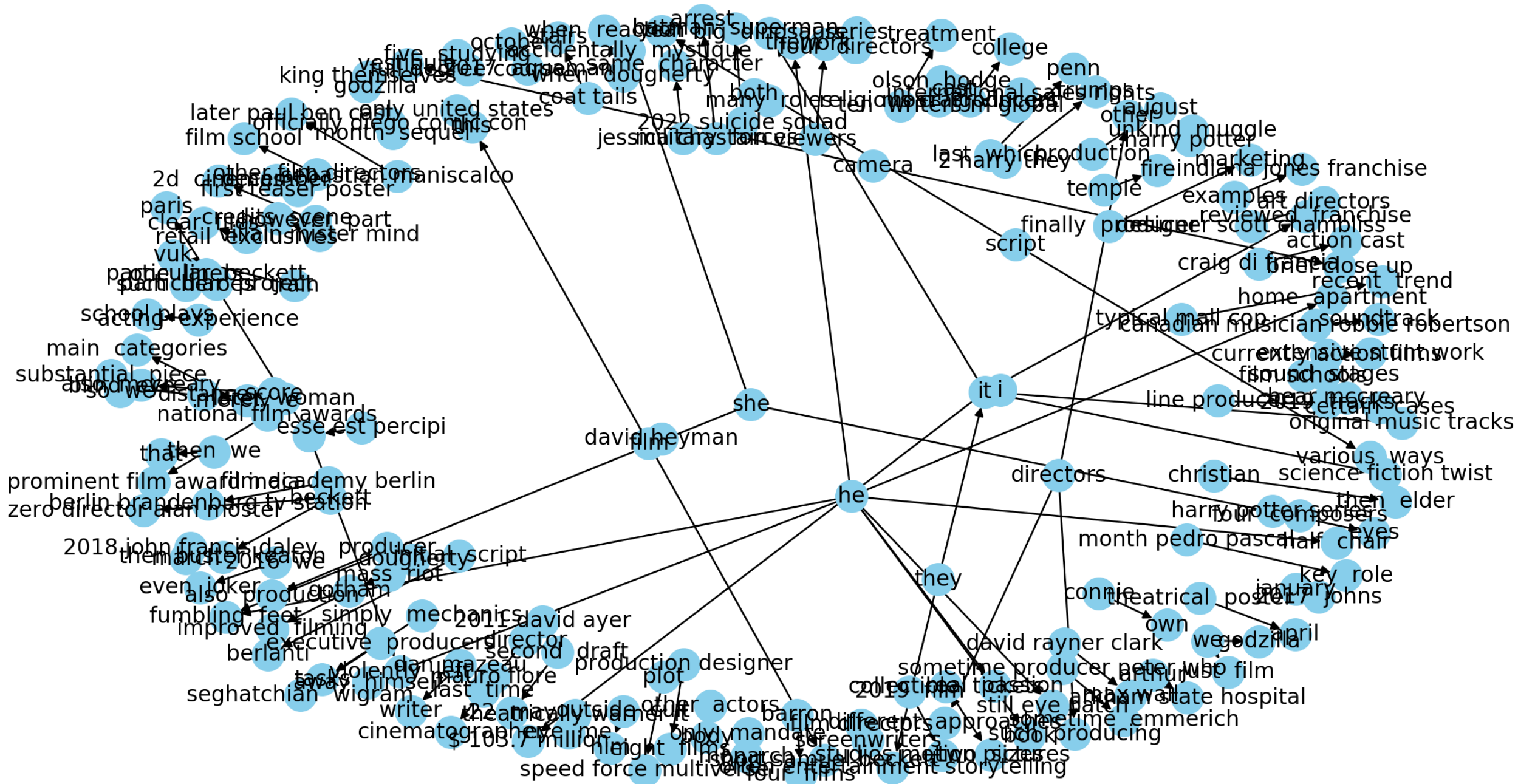
Index ▲	Type	Size	
0	str	1	connie
1	str	1	later woman
2	str	1	christian
3	str	1	temple
4	str	1	outside cult
5	str	1	religious craig vincent
6	str	1	craig di francia
7	str	1	sebastian maniscalco
8	str	1	we
9	str	1	m global

Index ▲	Type	Size	
0	str	1	own
1	str	1	distance
2	str	1	then elder
3	str	1	fire
4	str	1	him
5	str	1	cast
6	str	1	action cast
7	str	1	later paul ben cast
8	str	1	just film
9	str	1	international sales rights

Index ▲	Type	Size	
0	str	1	decides
1	str	1	heard in
2	str	1	paralyzed by
3	str	1	set on
4	str	1	wails with
5	str	1	's
6	str	1	revealed
7	str	1	revealed as
8	str	1	tried
9	str	1	circling

Creation of Knowledge Graph

- Tried different libraries for this such as RDFlib, neo4j etc. But they could not comply with my preferred way of querying the graph.
- Finally, Networkx library in python was used to create a knowledge graph out of the triples formed in the last slide.
- The graph was a directed graph from one node (subject) to another (object) with edge representing the relation.
- Matplotlib library was used to plot the graph as shown in the next slide.
- Timeliness – 3 days



Querying the graph

- Tried different complex frameworks to convert natural language to queries like quepy but did not work.
- Finally adopted simple way . The same procedure was used to convert the question into triples as the sentences were converted.
- Using this, we searched the graph for the corresponding node using which we could extract the full relation and hence reach the sentence in the initial documents.
- It has many limitations which only complex frameworks can solve.
- Timeliness – 2 days

text = ""confused and frustrated, connie decides to leave on her own.
later, a woman's scream is heard in the distance.
christian is then paralyzed by an elder.
the building is set on fire.
outside, the cult wails with him.
it's a parable of a woman's religious awakening-
c. mackenzie, and craig vincent joined the cast.
later, craig di francia and action bronson were revealed to have joined the cast.
sebastian maniscalco and paul ben-victor were later revealed as being part of the cast.
we just tried to make the film.
we went through all these tests and things
m global was also circling to bid for the film's international sales rights.
canadian musician robbie robertson supervised the soundtrack.
it features both original and existing music tracks.
it is the worst reviewed film in the franchise.
but she injures quicksilver and accidentally kills mystique before flying away.
military forces tasked with her arrest.
the train is attacked by vuk and her d'bari forces.
kota eberhardt portrays telepath selene gallio,
singer did not return to direct the sequel, x-men:
the last stand, which was written by penn and simon kinberg.
jessica chastain was also potentially being considered for the same character.
mauro fiore served as cinematographer.
filming was completed on october 14, 2017.
the soundtrack was released digitally on june 7.
the album was released digitally on august 2, 2019.
the film is distributed by walt disney studios motion pictures.
it ended up debuting to just \$103.7 million internationally and \$136.5 million worldwide.
the film stars kyle chandler, vera farmiga, millie bobby brown, bradley whitford, sally hawkins, charles dance, thomas middleditch,
it is dedicated to executive producer yoshimitsu banno and original godzilla suit
vivienne graham approach former employee dr.
emma frees and awakens monster zero, which kills several monarch members, including dr.
disillusioned, madison disowns her mother.through mythological texts, dr.
joe morton appears as an older dr.
godzilla, rodan, mothra, and king ghidorah were credited as themselves.
when asked about his reaction to being asked to direct, dougherty stated, 'yes.'
we're empathizing with godzilla.
legendary's only mandate was to include monarch, rodan, mothra, and king ghidorah.
ten writers contributed to building on the treatment.
the script took a year to come together.
dougherty also changed, revised, and improved lines during filming and post-production.
due to this, the film became an ensemble piece.
it can't just look like big dinosaurs.
other actors perform the body.
production designer scott chambliss managed all the art directors.
the single was released on may 13, 2019.

question = "" what was set on fire ""



['what', 'fire']
the building is set on fire.

Scope of improvements

- The Triples extracted are less rich in semantic embedding which could be enhanced by using StanfordNLP library as it has properties such as coreference etc.
- The project was mainly done by concentrating on textual data and can be improved to take into account tabular and other type of data too.
- The graph can be created using neo4j and RDFlib library so that there is better question answering as they support query language retrievals.

References

- Microsoft AI & R, Bellevue, WA, USA (2017) - Querying Enterprise Knowledge Graph With Natural Language
- KJV Heimark (2013) - A specialized question answering system using quepy framework
- <https://pypi.org/project/networkx-query/>
- <https://networkx.github.io/documentation/stable/>
- <https://quepy.readthedocs.io/en/latest/>
- <https://rdflib.readthedocs.io/en/stable/>