

# CASEBOT : PROJECT1

## REPORT PART -1

Name: Sahrudayi Caroline P.  
NETID: SXP230055

### ○ Web Crawling:

- Utilizing web crawling techniques, I collected 50 URLs. I used [https://en.wikipedia.org/wiki/Christopher\\_Nolan](https://en.wikipedia.org/wiki/Christopher_Nolan) as the starter url. After collecting 25 wikipedia URLs, the crawler collects 25 more urls

```
https://www.indiewire.com/2020/08/tenet-critics-divided-christopher-nolan-1234581611/
https://www2.bfi.org.uk/news-opinion/sight-sound-magazine/features/christopher-nolan-escape-artist
https://pop.inquirer.net/37978/kenneth-branagh-must-evacuate-soldiers-trapped-dunkirk
https://www.empireonline.com/movies/features/empire-30-christopher-nolan-set-visits/
https://deadline.com/2024/01/christopher-nolan-france-honorary-cesar-1235797659/
http://www.davidbordwell.net/blog/2017/08/09/dunkirk-part-2-the-art-film-as-event-movie/
https://www.indiewire.com/gallery/christopher-nolan-unmade-films/
https://www.latimes.com/entertainment/movies/la-ca-mn-dark-knight-10th-anniversary-20180822-story.html
https://indianexpress.com/article/entertainment/hollywood/happy-birthday-christopher-nolan-steven-spielberg-james-cameron-5860421/
https://www.indiewire.com/influencers/dunkirk-production-team-nolan/
https://www.telegraph.co.uk/films/2017/07/08/spitfires-flotillas-boats-rough-seas-1000-extras-christopher/
Starting URL:https://www.complex.com/pop-culture/tv-movie-creatives-that-defined-the-2010s/christopher-nolan
-----
Starting URL:http://time.com/3822892/christopher-nolan-2015-time-100
-----
Starting URL:http://filmlondon.org.uk/news/2010/july/nolans_mind_games
-----
Starting URL:https://variety.com/2015/film/news/christopher-nolans-syncopy-teaming-with-zeitgeist-on-blu-ray-releases-exclusive-1201437195/
-----
https://variety.com/2015/film/news/christopher-nolans-syncopy-teaming-with-zeitgeist-on-blu-ray-releases-exclusive-1201437195/#respond
https://variety.com/2024/film/columns/christopher-nolan-tenet-flaws-oppenheimer-1235922910/
https://variety.com/t/christopher-nolan/
Starting URL:https://variety.com/2014/film/reviews/film-review-interstellar-1201338475/
-----
https://variety.com/2024/film/columns/christopher-nolan-tenet-flaws-oppenheimer-1235922910/
https://variety.com/t/christopher-nolan/
Starting URL:https://www.hollywoodreporter.com/movies/movie-news/christopher-nolan-pitch-to-studios-including-apple-seeking-his-next-film-1235014132/
-----
https://www.hollywoodreporter.com/t/christopher-nolan/
https://www.hollywoodreporter.com/movies/movie-news/cillian-murphy-broke-his-head-oppenheimer-emily-blunt-1235834267/
https://www.hollywoodreporter.com/business/business-news/christopher-nolan-rips-hbo-max-as-worst-streaming-service-denounces-warner-bros-plan-4101408/
https://www.hollywoodreporter.com/movies/movie-news/christopher-nolan-pitch-to-studios-including-apple-seeking-his-next-film-1235014132/#respond
Starting URL:https://www.independent.ie/entertainment/movies/how-christopher-nolan-has-held-true-to-his-sweeping-vision-with-tenet-39466428.html
-----
Starting URL:https://www.nytimes.com/2014/11/02/magazine/the-exacting-expansive-mind-of-christopher-nolan.html
-----
Starting URL:https://variety.com/2017/film/news/christopher-nolan-dunkirk-oscars-movies-tv-spielberg-1202607836/
-----
https://variety.com/t/dunkirk/
https://variety.com/2024/film/columns/christopher-nolan-tenet-flaws-oppenheimer-1235922910/
https://variety.com/t/christopher-nolan/
Starting URL:http://filmcomment.com/article/cinematic-faith-christopher-nolan-scott-foundas
-----
Starting URL:https://www.usatoday.com/story/life/movies/2018/05/13/cannes-festival-christopher-nolan-inspires-crazed-crowd-talks-batman/605699002/
```

### Snapshot of urls collected during crawling

- I then scraped all the collected 50 URLs for raw text. I got 34 urls after rejecting urls with request errors.

```
9).TxR.2.regtrans-ms', 'NTUSER.DAT{550d23c9-6fd0-11ee-a361-004ce0864db9}.TxR.blf', 'NTUSER.DAT{550d23ca-6fd0-11ee-a361-004ce0864db9}.TMContainer000000000000000001.regtrans-ms', 'NTUSER.DAT{550d23ca-6fd0-11ee-a361-004ce0864db9}.TMContainer000000000000000002.regtrans-ms', 'ntuser.ini', 'OneDrive', 'Pictures', 'PrintHood', 'Recent', 'Saved Games', 'Searches', 'SendTo', 'Start Menu', 'Templates', 'thasunda', 'Videos']
Scraping all URLs for text
URL https://en.wikipedia.org/wiki/Christopher_Nolan
URL https://variety.com/2015/film/news/christopher-nolans-syncopy-teaming-with-zeitgeist-on-blu-ray-releases-exclusive-1201437195/
URL https://deadline.com/2015/04/christopher-nolan-joins-film-foundation-board-1201414195/
URL http://www.tribute.ca/people/christopher-nolan/6453/35045/
URL http://blogs.indiewire.com/theplaylist/christopher-nolans-merging-an-original-idea-with-jonah-nolans-old-screenplay-for-interstellar-20130110
URL http://blogs.indiewire.com/theplaylist/trivia-when-christopher-nolan-first-came-to-warner-bros-he-was-offered-troy-to-direct-20130621
URL https://www.dga.org/News/PressReleases/2015/150507-DGA-Congratulates-Martin-Scorsese-and-Christopher-Nolan-for-NFPB-appointments.aspx
URL https://collider.com/christopher-nolan-oppenheimer-release-date-cast-filming/
URL https://deadline.com/2024/01/christopher-nolan-france-honorary-cesar-1235797659/
URL https://www.indiewire.com/gallery/christopher-nolan-unmade-films/
URL https://www.empireonline.com/features/a-z/christopher-nolan/7.asp
Error requesting https://www.empireonline.com/features/a-z/christopher-nolan/7.asp
URL https://variety.com/2024/film/columns/christopher-nolan-tenet-flaws-oppenheimer-1235922910/
URL https://variety.com/t/christopher-nolan/
URL https://variety.com/2015/film/news/christopher-nolans-syncopy-teaming-with-zeitgeist-on-blu-ray-releases-exclusive-1201437195/#respond
URL https://deadline.com/tag/christopher-nolan/
URL https://www.indiewire.com/t/christopher-nolan/
URL http://blogs.indiewire.com/theplaylist/christopher-nolan-to-direct-former-steven-spielberg-sci-fi-project-interstellar-20130109
URL http://www.deadline.com/2013/06/ben-afflecks-baz-lurhmanns-chris-nolans-relationships-with-exiting-mogul-jeff-robinov/#more-526393
```

### While scraping and cleaning - error requesting URLs

- Clean files were then created after the gathered raw text underwent preprocessing steps to refine it for analysis. This included stripping HTML tags, removing citation number enclosed in square brackets, and tokenizing the text, removing the punctuation and stop words

- I employed Natural Language Processing (NLP) techniques to identify significant keywords within the processed text using tf-idf. These keywords served as essential markers for categorization and retrieval of relevant information.
- I manually picked 16 relevant keywords on which the knowledge base was populated
- I used an sqlite database to store the knowledge base
- My key terms are:
  - nolan
  - academy
  - dunkirk
  - oppenheimer
  - directed
  - success
  - awards
  - critics
  - memento
  - role
  - knight
  - cast
  - inception
  - interstellar
  - reviews
  - tenet
- Screenshots of the knowledge base

DB Browser for SQLite - C:\Users\sahsal\MS CS\documents\MS CS\Spring 24\NLP\PROJECTS\CHATBOT\Nolan\database\_final\_latest.db

File Edit View Tools Help

Database Structure Browse Data Execute SQL Edit Pragma

Table: keywords Filter in any column

Filter	sentence	keyvo	Filter
493	Christopher Edward Nolan was born on 30 July 1970, in Westminster, London,...	nolan	
494	His father, Brendan James Nolan, was a British advertising executive who worked as a creative director....	nolan	
495	Growing up, Nolan was particularly influenced by the work of Ridley Scott and the science fiction films 2001: A Space Odyssey (1968) and Star Wars (1977)....	nolan	
496	Nolan began making films at the age of seven, borrowing his father's Super 8 camera and shooting short films with his action figures....	nolan	
497	Between 1981 and 1983, Nolan enrolled at Barrow Hills, a Catholic prep school in Weybridge, Surrey....	nolan	
498	Nolan and Roko co-directed the surreal 8 mm Tarentella (1989), which was shown on Image Union, an independent film and video showcase on the Public Broadcasting Service....	nolan	
499	Nolan was president of the Union's Film Society, and with Emma Thomas (his girlfriend and future wife) he screened feature films in 35mm during the school year and used the money earned to produce 16 mm....	nolan	
500	After earning his bachelor's degree in English literature in 1993, Nolan worked as a script reader, camera operator and director of corporate films and industrial films....	nolan	
501	Funded by Nolan and shot with the UCL Union Film society's equipment, it appeared at the Cambridge Film Festival in 1996 and is considered one of UCL's best shorts....	nolan	
502	Nolan filmed a third short, Doodlebug (1997), about a man seemingly chasing an insect with his shoe, only to discover that it is a miniature of himself....	nolan	
503	Nolan and Thomas first attempted to make a feature in the mid-1990s with Larry Mahoney, which they scrapped....	nolan	
504	Shortly after abandoning Larry Mahoney, Nolan conceived the idea for his first feature, Following (1998), which he wrote, directed, photographed and edited....	nolan	
505	Following won several awards during its festival run and was well received by critics who labelled Nolan a majorly talented debutant....	nolan	
506	Nolan on the jump from his first film to his second (2012) Following's success afforded Nolan the opportunity to make Memento (2000), which became his breakthrough film....	nolan	
507	Jonathan worked the idea into a short story, "Memento Mori" (2001), and Nolan developed it into a screenplay that told the story in reverse....	nolan	
508	In the book The Philosophy of Neo-Noir, Basil Smith drew a comparison with John Locke's An Essay Concerning Human Understanding, which argues that conscious memories constitute our identities - a theme ...	nolan	
509	Memento earned Nolan many accolades, including nominations for an Academy Award and a Golden Globe Award for Best Screenplay, as well as two Independent Spirit Awards: Best Director and Best ...	nolan	
510	In 2001, Nolan and Emma Thomas founded the production company Syncopy Inc.(b) Inspired by his work on Memento, filmmaker Steven Soderbergh recommended Nolan to Warner Bros. to direct the ...	nolan	
511	A remake of the 1997 Norwegian thriller of the same name, the film is viewed as "the outlier of Nolan's filmography" due to its perceived lack of unconventionality he is known for....	nolan	
512	Following, Memento and Inception established Nolan's image as an "auteur"....	nolan	

Mode: Text

1 Christopher Edward Nolan was born on 30 July 1970  
in Westminster, London.

2

Type of data currently in cell: Text / Numeric Apply

File	Edit	View	Tools	Help
Database Structure		Browse Data	Execute SQL	Edit Pragmas
Table:	keywords			
sentence				keywo
Filter				Filter
193	"Insomnia," the first film directed by Christopher Nolan since his famous "Memento" (2001), is a remake of a Norwegian film of the same name, made in 1998 by Erik Skjoldbjærg....			directed
194	For the war drama Dunkirk, he earned two Academy Award nominations, including his first for Best Director....			dunkirk
195	Based on his own original screenplay and co-produced with Thomas, the film is set amid World War II in 1940 and the evacuation of Allied soldiers from the beaches of Dunkirk, France....			dunkirk
196	Starring an ensemble cast, Dunkirk was released to widespread critical acclaim and strong box office results....			dunkirk
197	Dunkirk is a 2017 epic historical war thriller film written, directed and co-produced by Christopher Nolan that depicts the Dunkirk evacuation of World War II from the perspectives of the land, sea and air....			dunkirk
198	Filming began in May 2016 in Dunkirk and wrapped that September in Los Angeles, when post-production began....			dunkirk
199	Distributed by Warner Bros. Pictures, Dunkirk premiered at Odeon Leicester Square in London, a few days before its release in the United Kingdom and United States on 21 July 2017....			dunkirk
200	Dunkirk received praise for its screenplay, direction, editing, score, sound design and cinematography; some critics called it Nolan's best work, and one of the greatest war films....			dunkirk
201	In 1940, during the Battle of France, Allied soldiers retreat to Dunkirk encircled by enemy forces....			dunkirk
202	With a single, vulnerable mole available for embarking on deep-draft ships, the Navy requisitions civilian vessels in England that can get to Dunkirk....			dunkirk
203	Realising that Dawson is going for Dunkirk, the soldier demands they turn back and tries to wrest control of the boat; in the scuffle, George suffers a head injury that renders him blind....			dunkirk
204	Three Royal Air Force Spitfires fly towards Dunkirk, to provide cover for the evacuation, limited to one hour of operation by their fuel supply....			dunkirk
205	Farrier reaches Dunkirk just as his fuel runs out....			dunkirk
206	Nolan said that he approached research as though it were for a documentary, and was attracted to the project because of its inversion of the "Hollywood formula": the Battle of Dunkirk was not a victory and did ...			dunkirk
207	Nolan postponed Dunkirk until he had acquired sufficient experience directing large-scale action films....			dunkirk
208	The historical consultant was author Joshua Levine, who also wrote the book adaptation, Dunkirk: The History Behind the Major Motion Picture....			dunkirk
209	Nolan and his production designer Nathan Crowley toured the beach of Dunkirk while location scouting, and decided to film there despite the logistical challenges, discarding Suffolk as an alternative....			dunkirk
210	Principal photography commenced on 23 May 2016 in Dunkirk, planned so as to avoid Bastille Day and coincide with the dates of the real evacuation....			dunkirk
211	Over fifty other boats included twenty actual Little Ships of Dunkirk, piloted by their owners....			dunkirk
212	A small 1930s motor yacht called Moonstone served for six weeks of filming; its most demanding scenes, with up to sixty people on a boat designed for fewer than ten, were shot on the Dutch lake IJsselmeer to...			dunkirk

Screenshots showing the ‘keywords’ table (sentence,keyword) which is the knowledge of my chatbot- the CaSebot (SQLite Database viewed in the DB Browser)