Big Data and Knowledge

Movie recommenders revisited





BIG DATA AND KNOWLEDGE

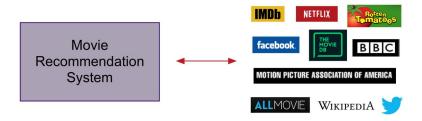
Problem

- A new movie is coming today to a theatre near the user location
 - Recommend it or not?

Relevant data

- Movie info
 - ▶ Title, Genre, Director, Actors,...
- Ratings
 - movieLens, imdb, social media,...
- Script and textual info
 - synopses, critics, post production scripts,...
- Image and video content
 - trailers, photos,...
- User preferences
 - movie likes-dislikes, usage history,...

RELEVANT DATA AND KNOWLEDGE



Relevant data sources

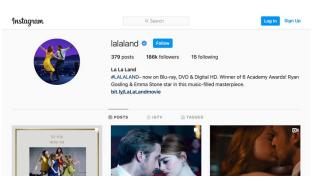
- Domain knowledge bases and related services
 - ▶ IMDb, Netflix, movieLens,TheMovieDB, The Movie Ontology,...
- Social media and more
 - twitter, instagram, facebook, BBC news :: culture,...
- Local websites and portals
 - athinorama.gr, theatre websites,...
- The "usual suspects"
 - wikipedia, chatGPT, google...

3

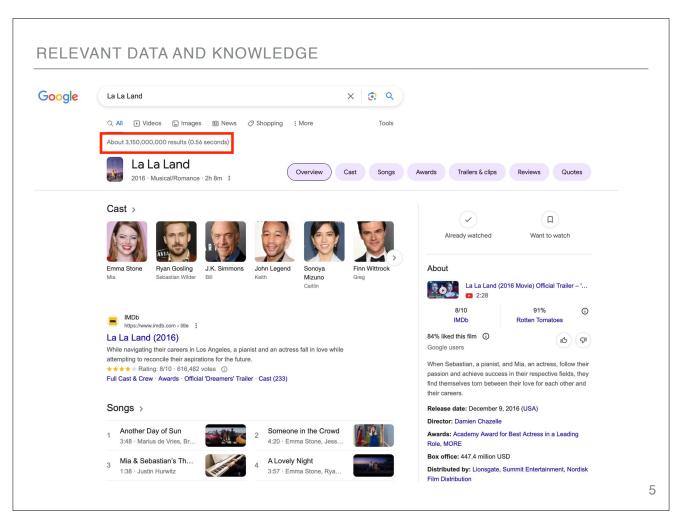
RELEVANT DATA AND KNOWLEDGE

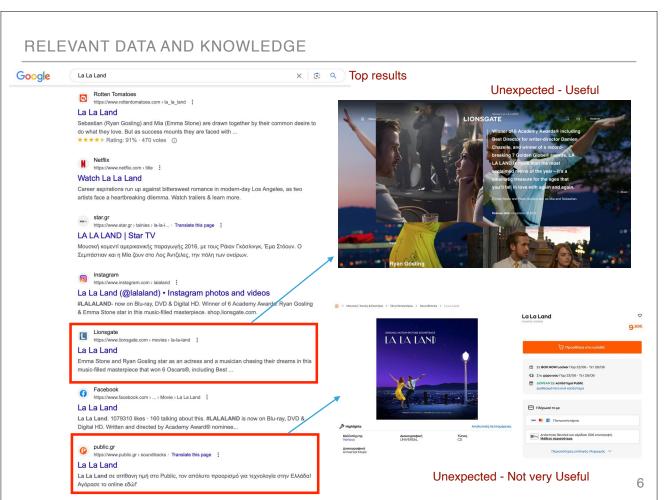












RELEVANT DATA AND KNOWLEDGE

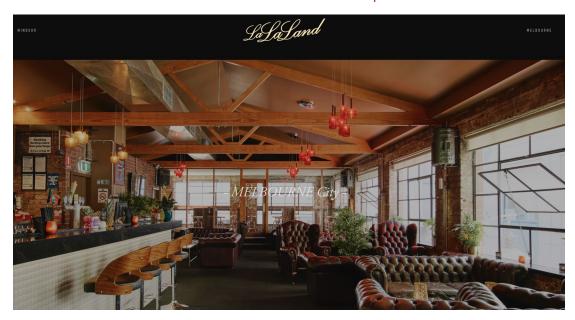


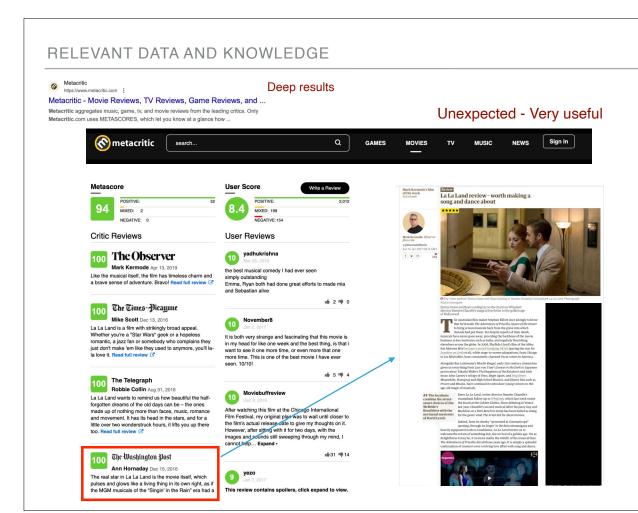
Deep results

Melbourne — La La Land

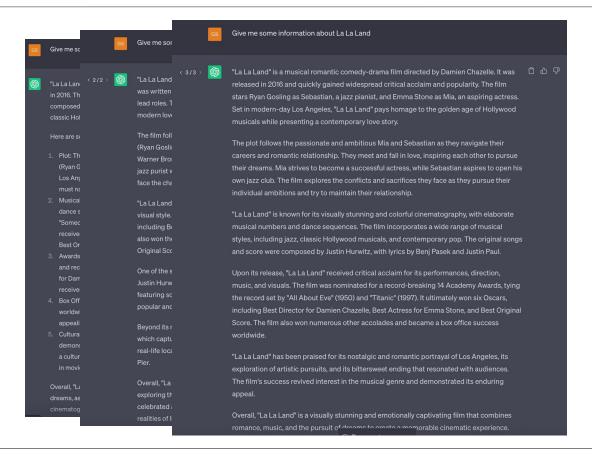
A hidden gem within the heart of **Melbourne's** CBD, **La La Land** is the perfect place for an intimate date night, an after work catch up or a get together with ...

Expected - Not useful





RELEVANT DATA AND KNOWLEDGE



RELEVANT DATA AND KNOWLEDGE

		, . ,						
Gre	eat dea	l of mu	ltimodal ii	nformation	about La L	a Land!		
Tabular								
Title	Genres	Duration	Director	Writer	Stars			
La La Land	Musical	2h 8m	Damien Chazelle	Damien Chazelle	Ryan Gosling Rosemarie Do	, Emma Stone eWitt	e,	
Images							LA LA LAND	
IATATAN JAT	A IAND	LAND	INIAIM ALA	A LAN		Texts	Written by Damien Chazelle	
/ideos Audio						FADE IN A sun-blasted s another	A sun-blasted sky. We HEAR radios one piece of music after	
0.0		1 Farida - HAUSER - Mia Sebastian's Theme - La La Land - ▶ 799K					We're — EXT. 101 FREEMAY - DAY Cars are at a standstill. It's a horrific traffic jam. Morning rush hour. Sun beating down, asphalt shimmering in the heat. The blown-out downtown L.A. skyline hovers in the distance. We RDRIFT past more CARS. Hear one snippet of audio after another One driver taps his steering wheel to PROG RDCK. Another sings to OPERA. A third raps along to a HIP-HOP track. We move from a RADIO INTERVIEW to a FRONT MAILAD to TECHNO, until finally we begin to hear anew, original piace of music [ANOTHER DAY OF SIN] We settle on the CAR from which this new tune is playing. The driver is a YOUNG MOWN. She hums along to the riff on her radio — then starts SINGING. Then — she EXITS her car. Then — she starts MOVING down the lane. One by one, more DRIVERS join her — SINGING and DANCING along. Without a single cut, we've found ourselves in a FULL-FLEEGED MUSICAL NUMBER. Drivers leap on car-tops, dance lerome Robbins—style, making use of the road and the hot gleam of the automobiles. Arms sweying, feet bamping, dancers darting, as the MUSIC clasts, taking the magic in Finally — all the drivers awing back into their vehicles — and the song comes to a dramatic stoy.	
2:13		2 Hanan Gobran - Will Always Love You La La Land ▶ 14.2K						
		3 Belle - Blue Moon ▶ 45.4K						
1 3.56 (AV)		La La Land (Original Motion Picture Soundtrack) Market Strategy Market Strategy						

Artificial intelligence and data analysis

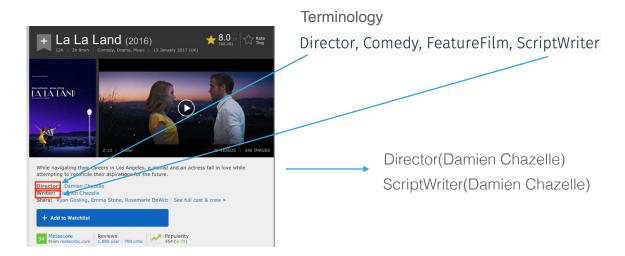
- Knowledge representation and automated reasoning
 - Semantic data description: understand the data using knowledge
 - Data integration using terminologies
 - Semantic query answering over big heterogeneous data
 - Data and knowledge enrichment using data analysis and reasoning
- Machine learning
 - Understanding numerical, non-categorical data
 - Understanding unstructured and raw data (text, image, sound, music, video, time series etc)
 - Generative models, large language models new forms of data access

11

DATA, KNOWLEDGE AND ARTIFICIAL INTELLIGENCE

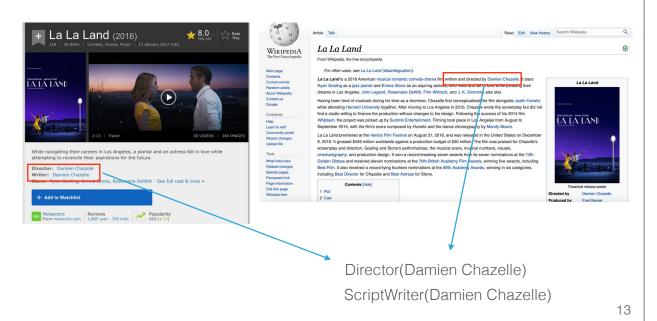
Artificial intelligence and data analysis

- Knowledge representation and automated reasoning
 - Semantic data description: understand the data using knowledge



Artificial intelligence and data analysis

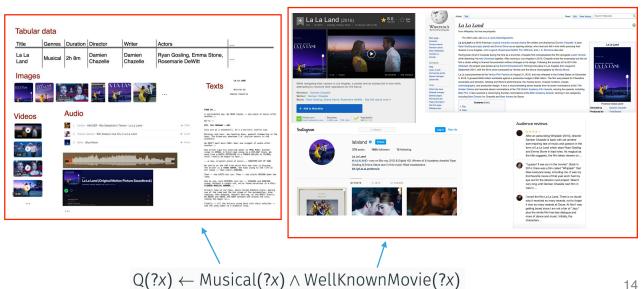
- Knowledge representation and automated reasoning
 - Data integration using terminologies



DATA, KNOWLEDGE AND ARTIFICIAL INTELLIGENCE

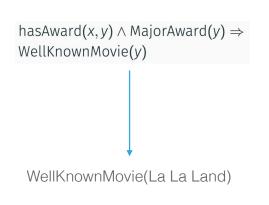
Artificial intelligence and data analysis

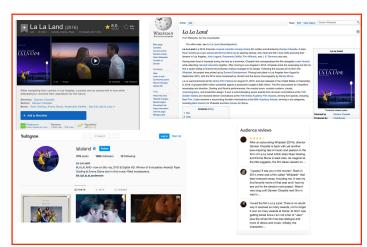
- Knowledge representation and automated reasoning
 - Semantic query answering over big heterogeneous data



Artificial intelligence and data analysis

- Knowledge representation and automated reasoning
 - Data and knowledge enrichment using data analysis and reasoning



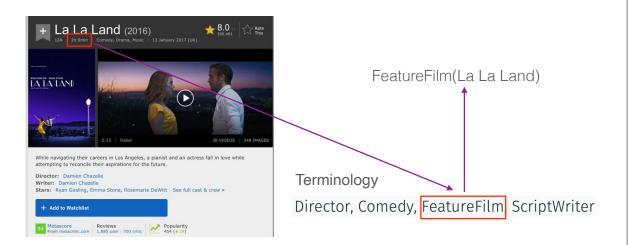


15

DATA, KNOWLEDGE AND ARTIFICIAL INTELLIGENCE

Artificial intelligence and data analysis

- Machine learning
 - Understanding numerical, non-categorical data



Artificial intelligence and data analysis

- Machine learning
 - Understanding unstructured and raw data



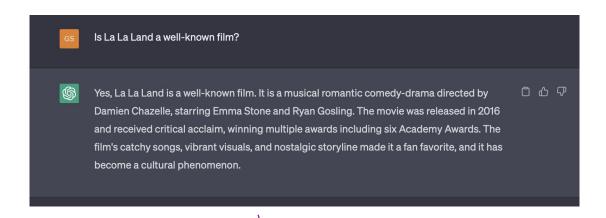
hasAward(La La Land,70th British Academy Film Awards Best Film)

17

DATA, KNOWLEDGE AND ARTIFICIAL INTELLIGENCE

Artificial intelligence and data analysis

- Machine learning
 - Generative models, large language models new forms of data access



WellKnownMovie(La La Land)

Artificial intelligence and data analysis

- Machine learning
 - Generative models, large language models new forms of reasoning



19

RELEVANT DATA, KNOWLEDGE AND ARTIFICIAL INTELLIGENCE

Data semantics - The early days

May 17, 2001



The Semantic Web

A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities

By Tim Berners-Lee, James Hendler and Ora Lassila

The entertainment system was belting out the Beatles' "We Can Work It Out" when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other *local* devices that had a *volume control*. His sister, Lucy, was on the line from the doctor's office: "Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments." Pete immediately agreed to share the chauffeuring.

At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's *prescribed treatment* from the doctor's agent, looked up several lists of *providers*, and checked for the ones *in-plan* for Mom's insurance within a *20-mile radius* of her home and with a *rating* of *excellent* or *very good* on trusted rating services. It then began trying to find a match between available *appointment times* (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules. (The emphasized keywords indicate terms whose semantics, or meaning, were defined for the agent through the **Semantic Web**.)

... 20



SEMANTIC WEB



On this page → technology topics • news • upcoming events and talks

In addition to the classic "Web of documents" W3C is helping to build a technology stack to support a "Web of data," the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term "Semantic Web" refers to W3C's vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.

Linked Data

The Semantic Web is a Web of data — of dates and titles and part numbers and chemical properties and any other data one might conceive of. RDF provides the foundation for publishing and linking your data. Various technologies allow you to embed data in documents (RDFa, GRDDL) or expose what you have in SQL databases, or make it available as RDF files.

Inference

Near the top of the Semantic Web stack one finds $inference-reasoning\ over\ data\ through\ rules.\ W3C$ work on rules, primarily through RIF and OWL, is focused on translating between rule languages and exchanging rules among different systems.

Vocabularies

At times it may be important or valuable to organize data. Using OWL (to build vocabularies, or "ontologies") and SKOS (for designing knowledge organization systems) it is possible to enrich data with additional meaning, which allows more people (and more machines) to do more with the data.

Vertical Applications

W3C is working with different industries - for example in Health Care and Life Sciences, eGovernment, and Energy - to improve collaboration, research and development, and innovation adoption through Semantic Web technology. For instance, by aiding decision-making in clinical research, Semantic Web technologies will bridge many forms of biological and medical information across institutions.

Query

Query languages go hand-in-hand with databases. If the Semantic Web is viewed as a global database, then it is easy to understand why one would need a guery language for that data. SPARQL is the query language for the Semantic Web.

21

RELEVANT DATA, KNOWLEDGE AND ARTIFICIAL INTELLIGENCE

~a decade later



Google The Keyword

Introducing the Knowledge Graph: things, not strings

May 16, 2012 · 4 min read



SVP, Engineering

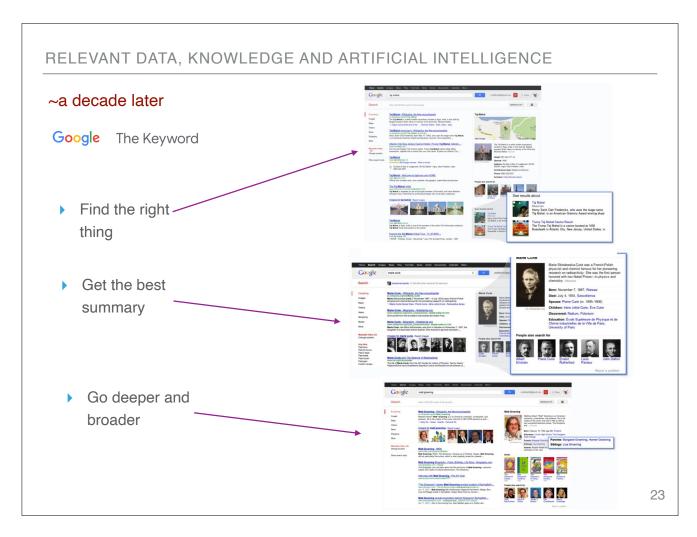
- Find the right thing
- Get the best summary
- Go deeper and broader

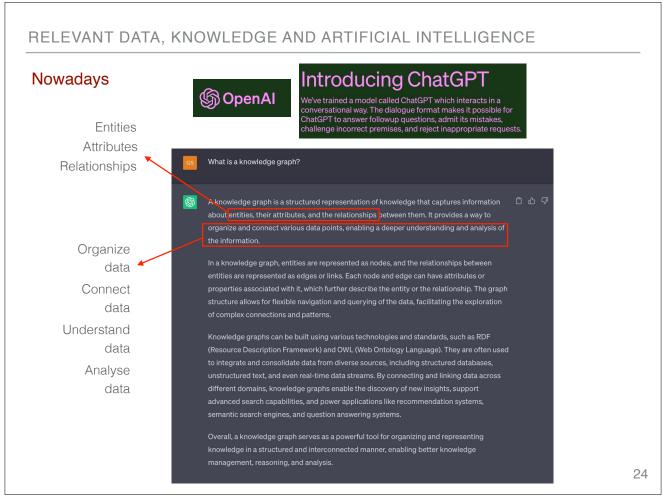
Search is a lot about discovery—the basic human need to learn and broaden your horizons. But searching still requires a lot of hard work by you, the user. So today I'm really excited to launch the Knowledge Graph, which will help you discover new information quickly and easily.

Take a query like [tai mahal]. For more than four decades, search has essentially been about matching keywords to queries. To a search engine the words [taj mahal] have been just that-two words.

But we all know that [taj mahal] has a much richer meaning. You might think of one of the world's most beautiful monuments, or a Grammy Award-winning musician, or possibly even a casino in Atlantic City, NJ. Or, depending on when you last ate, the nearest Indian restaurant. It's why we've been working on an intelligent model—in geek-speak, a "graph"—that understands real-world entities and their relation to one another: things, not strings.

The Knowledge Graph enables you to search for things, people or places that Google knows aboutlandmarks, celebrities, cities, sports teams, buildings, geographical features, movies, celestial objects, works of art and more—and instantly get information that's relevant to your query. This is a critical first step towards building the next generation of search, which taps into the collective intelligence of the web and understands the world a bit more like people do.





Entities - Things and names

La Land is a 2016 American romantic musical film written and directed by Damien Chazelle. The film stars Ryan Gosling and Emma Stone as a struggling jazz pianist and an aspiring actress, respectively, who meet and fall in love while pursuing their dreams in Los Angeles. John Legend, Rosemarie DeWitt, Finn Wittrock, and J. K. Simmon's appear in supporting roles.

Having been fond of musicals during his time as a drummer. Chazelle first conceptualized the film alongside Justin Hurwitz while attending Harvard University together. After moving to Los Angeles in 2010, Chazelle penned the script but did not find a studio willing to finance the production without changes to his design. After the success of his film Whiplash (2014), the project was picked up by Summit Entertainment. Miles Teller and Emma Watson were originally slated to star, but after both dropped out, Gosling and Stone were cast. Filming took place in Los Angeles between August and September 2015, with the film's score composed by Hurvitz and the dance choreography by Mandy Mooke.

Damien Chazelle La La Land Emma Stone Ryan Gosling Justin Hurwitz

Harvard University Emma Watson Whiplash Los Angeles

https://www.geonames.org/5368361/los-angeles.html

https://en.wikipedia.org/wiki/Emma_Stone

https://www.imdb.com/title/tt2582802/

25

KNOWLEDGE GRAPHS - THE INGREDIENTS

Entities - Things and names

Uniform Resource Identifier (URI)

▶ A Uniform Resource Identifier (URI) is a unique sequence of characters that identifies a logical or physical resource used by web technologies

W3C[®]

 URIs may be used to identify anything, including real-world objects, such as people and places, concepts, or information resources such as web pages and books

https://www.ails.ece.ntua.gr/ mailto:Emma.Stone@cinema.com

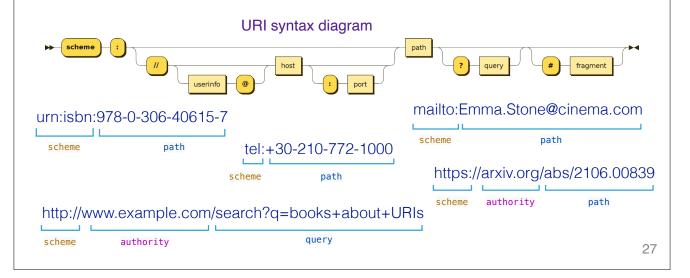
https://arxiv.org/abs/2106.00839 tel:+30-210-772-1000

urn:isbn:978-960-603-157-1 telnet://192.101.1.1/

Entities - Things and names

Uniform Resource Identifier (URI)

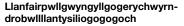
- A Uniform Resource Identifier (URI) is a unique sequence of characters that identifies a logical or physical resource used by web technologies
- URIs may be used to identify anything, including real-world objects, such as people and places, concepts, or information resources such as web pages and books



KNOWLEDGE GRAPHS - THE INGREDIENTS

Entities - Things and names



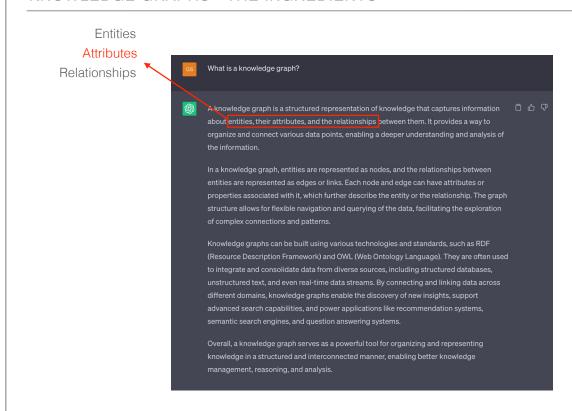




Everything in this world has a hidden meaning (Nikos Kazantzakis)

LLANFAIRPWLLGWYNGYLLGOGERYCHWYRNDROBWLLLLANTYSILIOGOGOGOCH

ST MARYS CHURCH IN THE HOLLOW OF THE WHITE HAZEL NEAR TO THE RAPID WHIRLPOOL OF LLANTYSILIO OF THE RED CAVE



29

KNOWLEDGE GRAPHS - THE INGREDIENTS

Attributes - Categories, classes, types

La Land is a 2016 American romantic musical film written and directed by Damien Chazelle. The film stars Ryan Gosling and Emma Stone as a struggling jazz pianist and an aspiring actress, respectively, who meet and fall in love while pursuing their dreams in Los Angeles. John Legend, Rosemarie DeWitt, Finn Wittrock, and J. K. Simmons appear in supporting roles.

Having been fond of musicals during his time as a drummer, Chazelle first conceptualized the film alongside Justin Hurwitz while attending Harvard University together. After moving to Los Angeles in 2010, Chazelle penned the script but did not find a studio willing to finance the production without changes to his design. After the success of his film Whiplash (2014), the project was picked up by Summit Entertainment. Miles Teller and Emma Watson were originally slated to star, but after both dropped out, Gosling and Stone were cast. Filming took place in Los Angeles between August and September 2015, with the film's score composed by Hurwitz and the dance choreography by Mandy Moore.

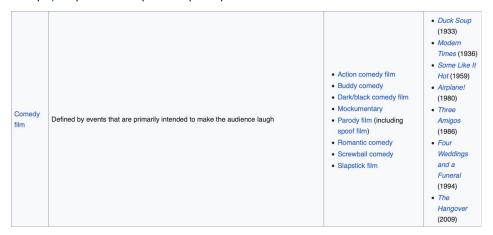
romantic musical film jazz pianist aspiring actress film studio script project city

https://en.wikipedia.org/wiki/romantic_musical_film

Attributes - Categories, classes, types

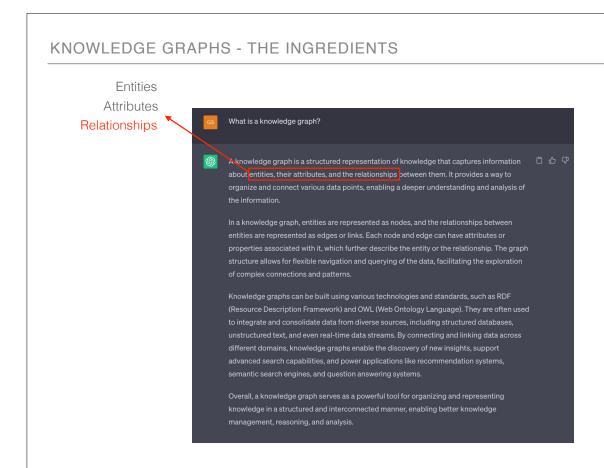
IMDb movie genres

Action | Adult | Adventure | Animation| Biography | Comedy | Crime | Documentary | Drama| Family | Fantasy | Film Noir | Game Show | History | Horror | Musical| Music | Mystery | News | Reality-TV | Romance | Sci-Fi | Short| Sport | Talk-Show | Thriller | War | Western



RomanticComedy ☐ Comedy

Terminologies, hierarchies, taxonomies, ontologies



Relationships - things classifications, things interconnections

La La Land is a 2016 American romantic musical film written and directed by Damien Chazelle. The film stars Ryan Gosling and Emma Stone as a struggling jazz pianist and an aspiring actress, respectively, who meet and fall in love while pursuing their dreams in Los Angeles. John Legend, Rosemarie DeWitt, Finn Wittrock, and J. K. Simmons appear in supporting roles.

Having been fond of musicals during his time as a drummer, Chazelle first conceptualized the film alongside Justin Hurwitz while attending Harvard University together. After moving to Los Angeles in 2010, Chazelle penned the script but did not find a studio willing to finance the production without changes to his design. After the success of his film Whiplash (2014), the project was picked up by Summit Entertainment. Miles Teller and Emma Watson were originally slated to star, but after both dropped out, Gosling and Stone were cast. Filming ook place in Los Angeles between August and September 2015, with the film's score composed by Hurwitz and the dance choreography by Mandy Moore.

is-a has-directed has-written starring fall-in-love finance picked

http://www.w3.org/1999/02/22-rdf-syntax-ns#type

33

took-place

KNOWLEDGE GRAPHS - THE RECIPE

Putting things together - Triples

La La Land is a 2016 American romantic musical film written and directed by Damien Chazelle. The film stars Ryan Gosling and Emma Stone as a struggling jazz pianist and an aspiring actress, respectively, who meet and fall in love while pursuing their dreams in Los Angeles. John Legend, Rosemarie DeWitt, Finn Wittrock, and J. K. Simmons appear in supporting roles.

https://en.wikipedia.org/wiki/La_La_Land

https://en.wikipedia.org/wiki/La_La_Land

http://www.w3.org/1999/02/22-rdf-syntax-ns#type

https://en.wikipedia.org/wiki/Romantic_musical_film

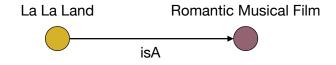
KNOWLEDGE GRAPHS - THE RECIPE

Putting things together - Triples

https://en.wikipedia.org/wiki/La_La_Land

http://www.w3.org/1999/02/22-rdf-syntax-ns#type

https://en.wikipedia.org/wiki/Romantic_musical_film



Resource Description Framework (RDF)

RDF is a standard model for data interchange on the Web.

RDF has features that facilitate data merging even if the underlying schemas differ.

RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple").

Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.

This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources, represented by the graph nodes.

This graph view is the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations.



35

KNOWLEDGE GRAPHS - THE RECIPE

Human understandable

<u>La La Land</u> is a romantic musical film directed by Damien Chazelle.



Romantic musical (sets)

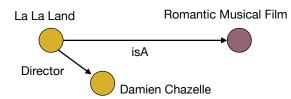
Romantic musical (alternatively musical romance) is a genre of film which features romantic relationships and whose story is partially explained through song ancidor done numbers. This genre originated on Broadway and moved to the silver screen thanks in part to the popularity of the Prodgers and Hammenship productions.

Some examples include South Pacific, West Side Story and its 2021 remains, Grease, High School Musical, Across the Universe, Moutin Rougel, the Malman Malfrancoins, Sustainities on Lainit, and Lai La Laind.





Machine understandable



<La La Land> <isA> <Romantic Musical Film> <La La Land> <director> <Damien Chazelle>







KNOWLEDGE GRAPHS - THE RECIPE

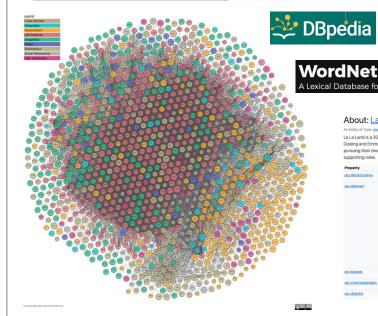
Terminological knowledge



http://dbpedia.org/resource/La La Land http://dbpedia.org/resource/La La Land http://dbpedia.org/resource/La La Land http://dbpedia.org/resource/La La Land

http://www.w3.org/1999/02/22-rdf-syntax-ns#type http://dbpedia.org/ontology/Film http://www.w3.org/1999/02/22-rdf-syntax-ns#type http://www.w3.org/2002/07/owl#Thing http://www.w3.org/1999/02/22-rdf-syntax-ns#type http://schema.org/CreativeWork http://www.w3.org/1999/02/22-rdf-syntax-ns#type http://schema.org/Movie

About: La La Land







GeoNames

37

KNOWLEDGE GRAPHS - THE RECIPE

Modern knowledge and data engineering

Create and share your dataset in RDF

Create and share your terminology in OWL





La La Land is a romantic musical film directed by Damien Chazelle. Can you write this in Sure! Here's an RDF representation of the information you provided about the film "La La Land": Copy code @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> . @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> . @prefix foaf: <http://xmlns.com/foaf/0.1/> @prefix dbpedia: <http://dbpedia.org/resource/> @prefix dbpedia-owl: <http://dbpedia.org/ontology/> . dbpedia:La_La_Land rdf:type dbpedia-owl:Film ; dbpedia-owl:director dbpedia:Damien_Chazelle ; dbpedia-owl:genre dbpedia:Romantic_musical_film . dbpedia:Damien_Chazelle rdf:type foaf:Person ; foaf:name "Damien Chazelle" dbpedia:Romantic_musical_film rdf:type dbpedia-owl:FilmGenre ; rdfs:label "Romantic musical film" . In this RDF representation, the film "La La Land" is defined as an instance of the `dbpediaowl:Film' class. It has a director, which is Damien Chazelle, who is defined as an instance of the `foaf:Person` class. The genre of the film is specified as a "Romantic musical film," which is an instance of the `dbpedia-owl:FilmGenre` class.