



USC Viterbi

School of Engineering

Information Sciences Institute

KGTK:

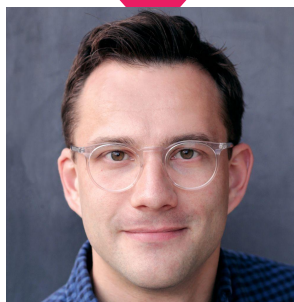
User-friendly Toolkit for Manipulation of Large Knowledge Graphs



Filip
Ilievski
ilievski@isi.edu



Kian
Ahrabian
ahrabian@isi.edu



Gleb
Satyukov
gleb@isi.edu



Ke-Thia
Yao
kyao@isi.edu



Jay
Pujara
jpujara@isi.edu

Go-to tool for KGs!?

NLP



spaCy

ML



KG



???

KG pipelines need many tools!

Operation	Tool	Format(s)	Language
graph analytics	graph-tool	GML, GT, TSV/CSV	python
	NetworkX	GML, JSON, TSV/CSV	python
graph database	Neo4J	TSV/CSV	<i>various</i>
RDF operations	graphy	RDF	javascript
	Jena	RDF	java
	RDFLib	RDF	python
graph embeddings	PyTorch-BigGraph	TSV/CSV	python
entity resolution	RLTK	TSV/CSV	python
entity linking	AGDISTIS	XML	python
	WAT	JSON	python

KG tools do not scale

Wikidata Query Service

Examples Query Builder Help More tools English

```
1 SELECT ?name (COUNT(?person) AS ?nameCount) (SAMPLE (?nameString) AS ?theName)
2 WHERE
3 {
4   ?person wdt:P31 wd:Q5 ;
5           wdt:P735 ?name .
6   OPTIONAL {
7     ?name rdfs:label ?nameString .
8     FILTER(LANG(?nameString) = "en")
9   }
10  SERVICE wikibase:label { bd:serviceParam wikibase:language "en,en" . }
11 }
12 GROUP BY ?name
13 ORDER BY DESC(?nameCount)
14
```

Most common first names of all humans

Query timeout limit reached

SPARQL-QUERY: queryStr=SELECT ?name (COUNT(?person) AS ?nameCount) (SAMPLE (?nameString) AS ?theName)

```
WHERE
{
  ?person
  OPTIONAL {
    ?name
    FILTER(LANG(?nameString) = "en")
  }
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en,en" . }
}
GROUP BY ?name
ORDER BY DESC(?nameCount)
```

Query timeout limit reached

Query timeout limit reached

Query timeout limit reached

Query

Results of all my
interesting queries

Query timeout limit reached

Query timeout limit reached

KGTK: The Knowledge Graph Toolkit

KGTK is a Python toolkit for creating and exploiting knowledge graphs.

KGTK is designed for ease of use, scalability and speed.

kgtk.readthedocs.io/en/latest/

MIT License

72 stars 21 forks

Unstar

Unwatch

Code

Issues 97

Pull requests 1

Actions

Projects 3

master



saggu Update __init__.py ...

4 days ago 2,745

[View code](#)

README.md



KGTK: Knowledge Graph Toolkit

DOI [10.5281/zenodo.3828068](https://doi.org/10.5281/zenodo.3828068) build passing coverage 32%

KGTK is a Python toolkit for building applications using knowledge graphs (KG). KGTK is designed for ease of use, scalability and speed. It represents KGs as simple TSV files with four columns to represent the head, relation and tail of a triple, as well as an identifier for each triple. This simple model allows KGTK to operate on property graphs and on RDF

KGTK Tenets

Scale to billions of statements

Speed competitive with SQL databases

No library needed to read/write KGs

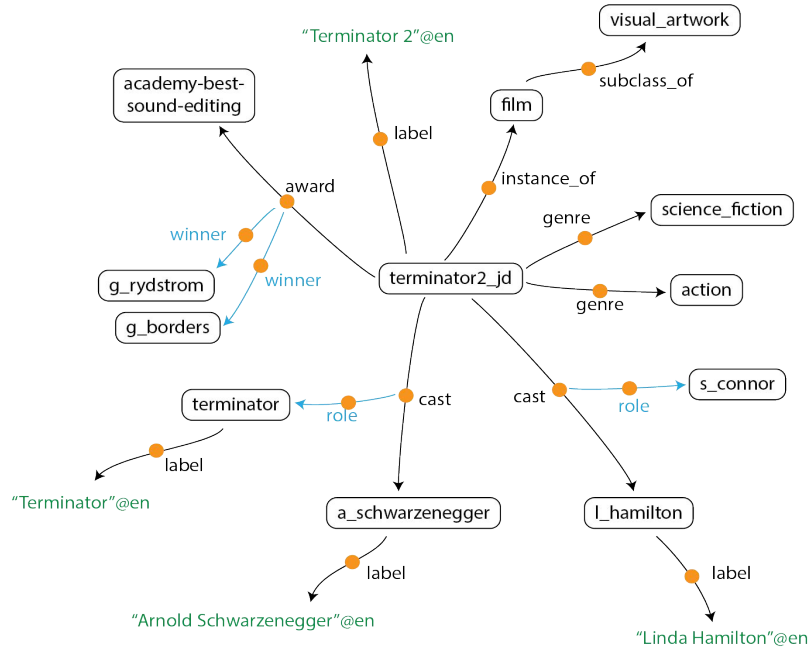
Support Wikidata data model: statements, qualifiers, references

Easy to export/import to state of the art tools (network analytics, embeddings, ElasticSearch, ...)

Pandas-friendly because people love Pandas

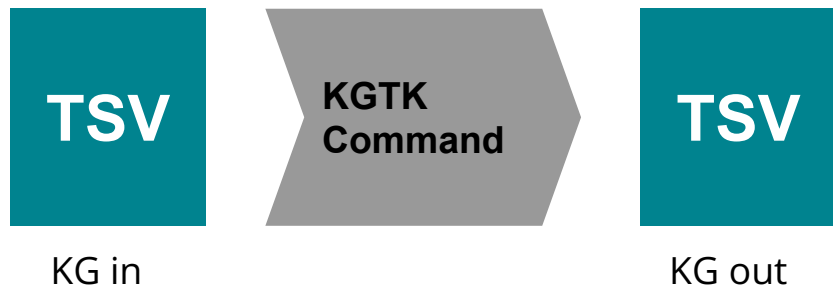
Pipeline-friendly to create workflows with multiple components

KGTK represents KGs as TSV files



id	node1	label	node2
	terminator2_jd	label	"Terminator 2"@en
	terminator2_jd	instance_of	film
	terminator2_jd	genre	science_fiction
	terminator2_jd	genre	action
t4	terminator2_jd	cast	a_schwarzenegger
	t4	role	terminator
t6	terminator2_jd	cast	l_hamilton
	t6	role	s_connor
t8	terminator2_jd	award	academy_best_sound_editing
	t8	point_in_time	^1992-03-30T00:00:00Z/11
	t8	winner	g_rydstrom
	t8	winner	g_borders

KGTK commands work with TSV KG files

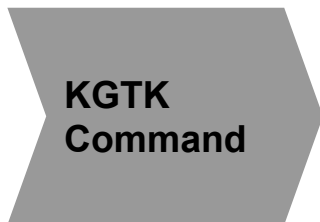


id	node1	label	node2
	terminator2_jd	label	"Terminator 2"@en
	terminator2_jd	instance_of	film
	terminator2_jd	genre	science_fiction
	terminator2_jd	genre	action
t4	terminator2_jd	cast	a_schwarzenegger
	t4	role	terminator
t6	terminator2_jd	cast	l_hamilton
	t6	role	s_connor
t8	terminator2_jd	award	academy_best_sound_editing
	t8	point_in_time	^1992-03-30T00:00:00Z/11
	t8	winner	g_rydstrom
	t8	winner	g_borders

KGTK pipelines consume and produce KGs



KG in



KG out

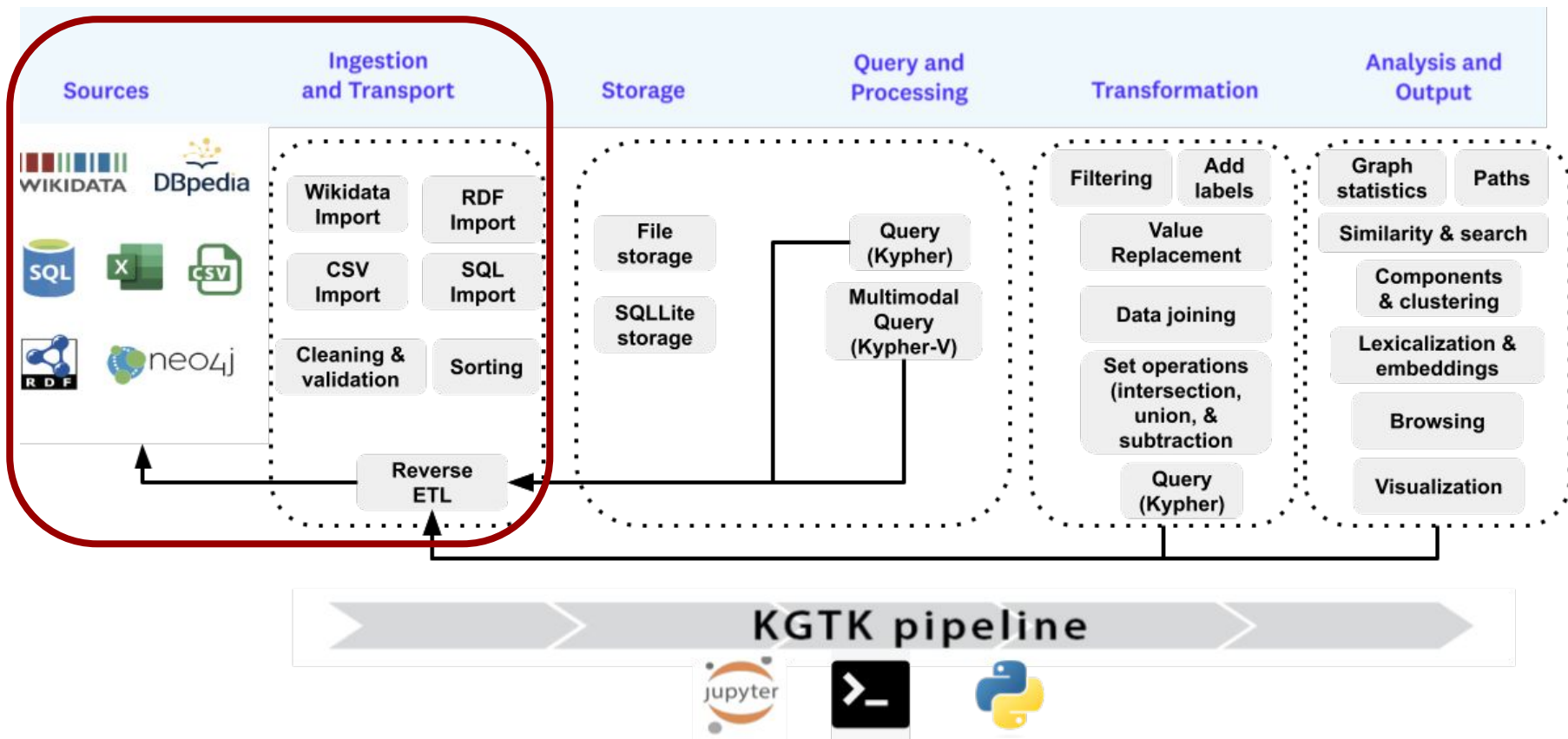


KGTK pipeline

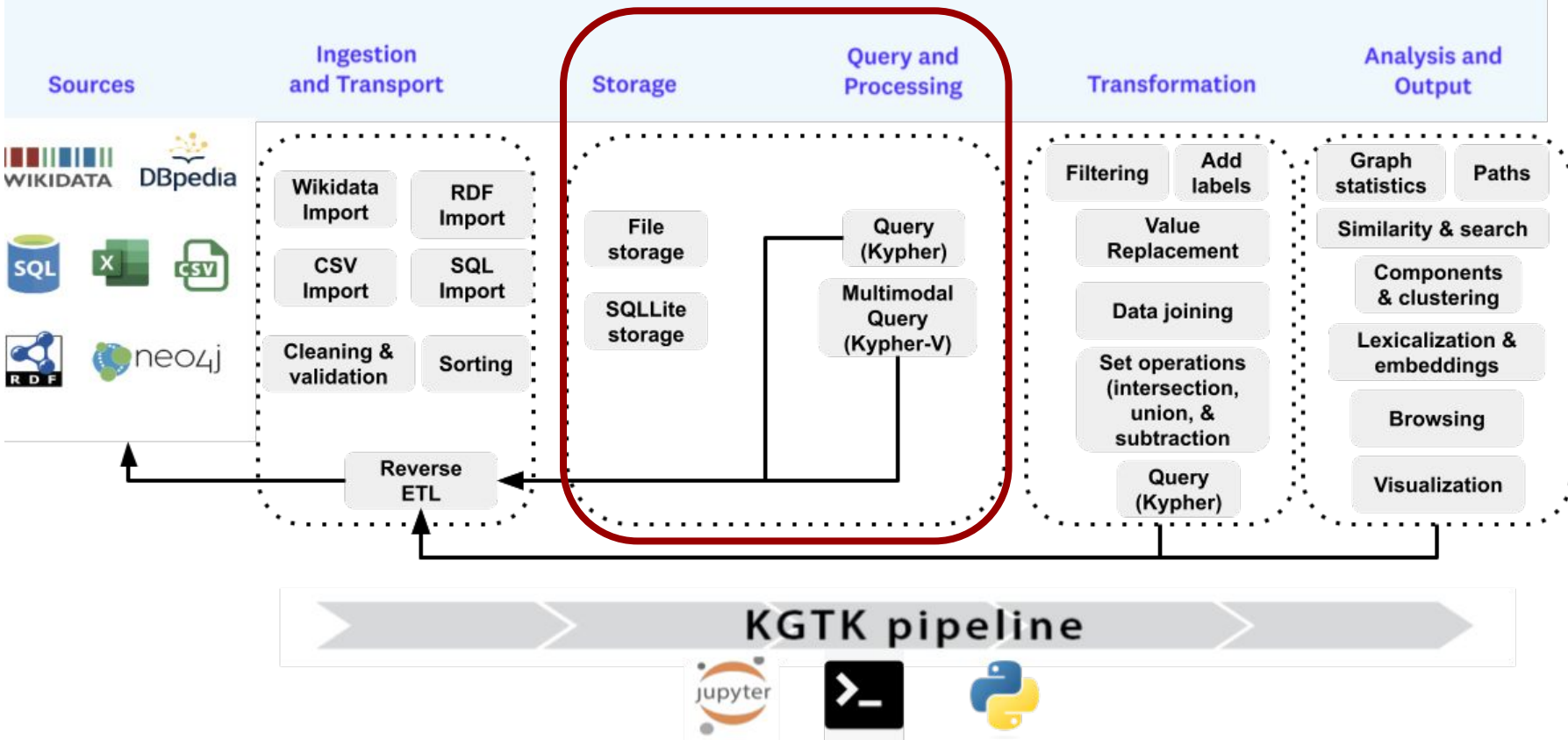


id	node1	label	node2
	terminator2_jd	label	"Terminator 2"@en
	terminator2_jd	instance_of	film
	terminator2_jd	genre	science_fiction
	terminator2_jd	genre	action
t4	terminator2_jd	cast	a_schwarzenegger
	t4	role	terminator
t6	terminator2_jd	cast	l_hamilton
	t6	role	s_connor
t8	terminator2_jd	award	academy_best_sound_editing
	t8	point_in_time	^1992-03-30T00:00:00Z/11
	t8	winner	g_rydstrom
	t8	winner	g_borders

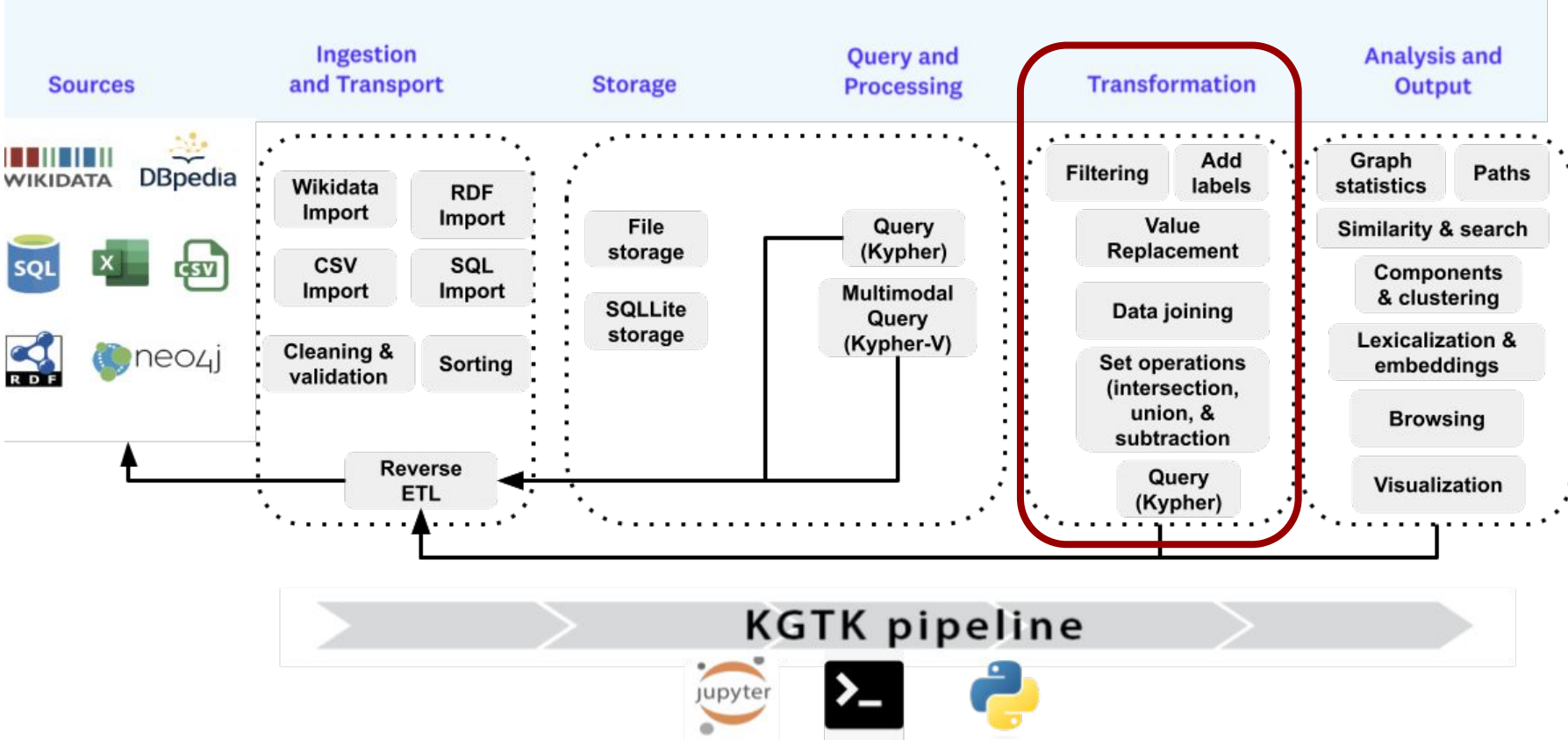
KGTK supports many formats



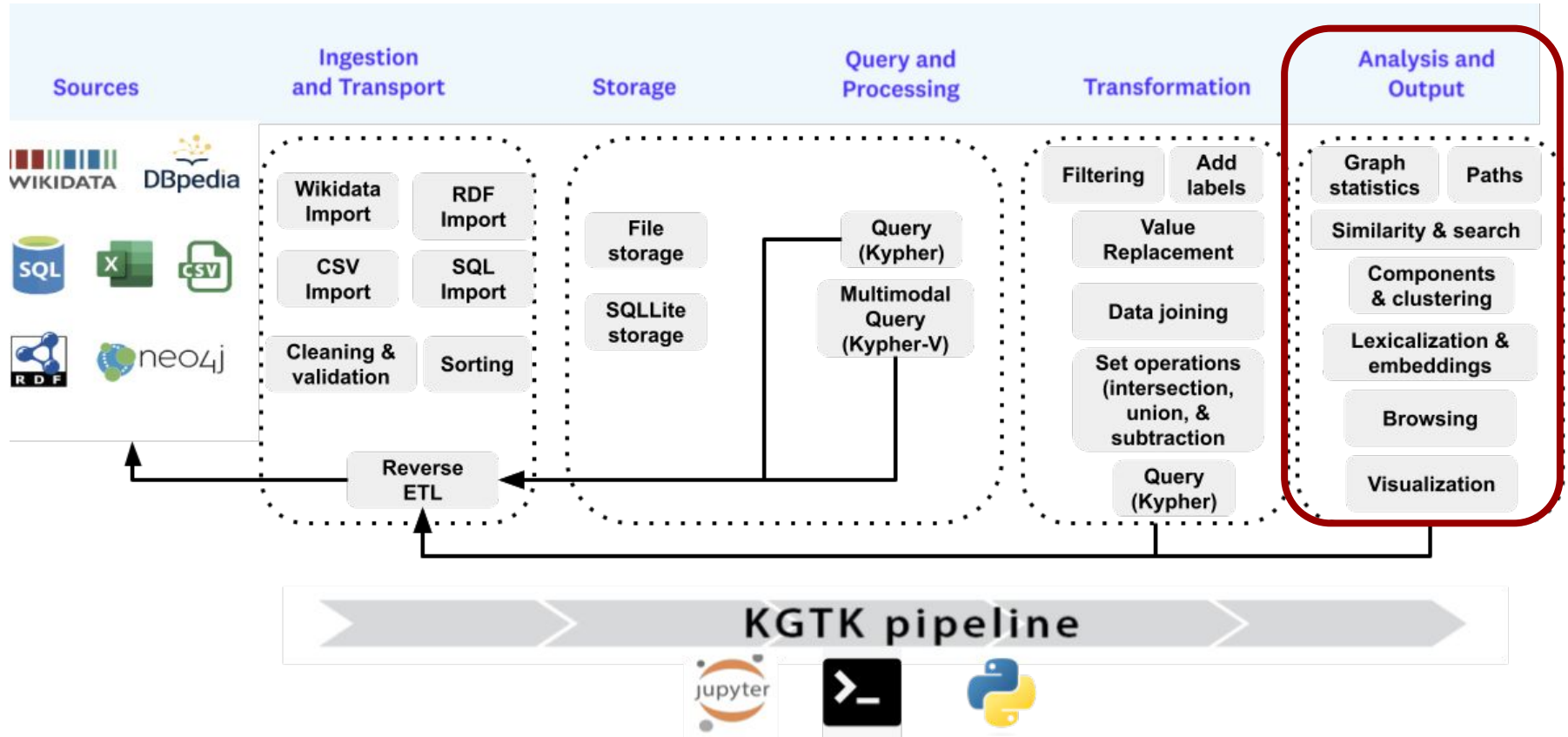
Kypher is a scalable and hybrid query engine



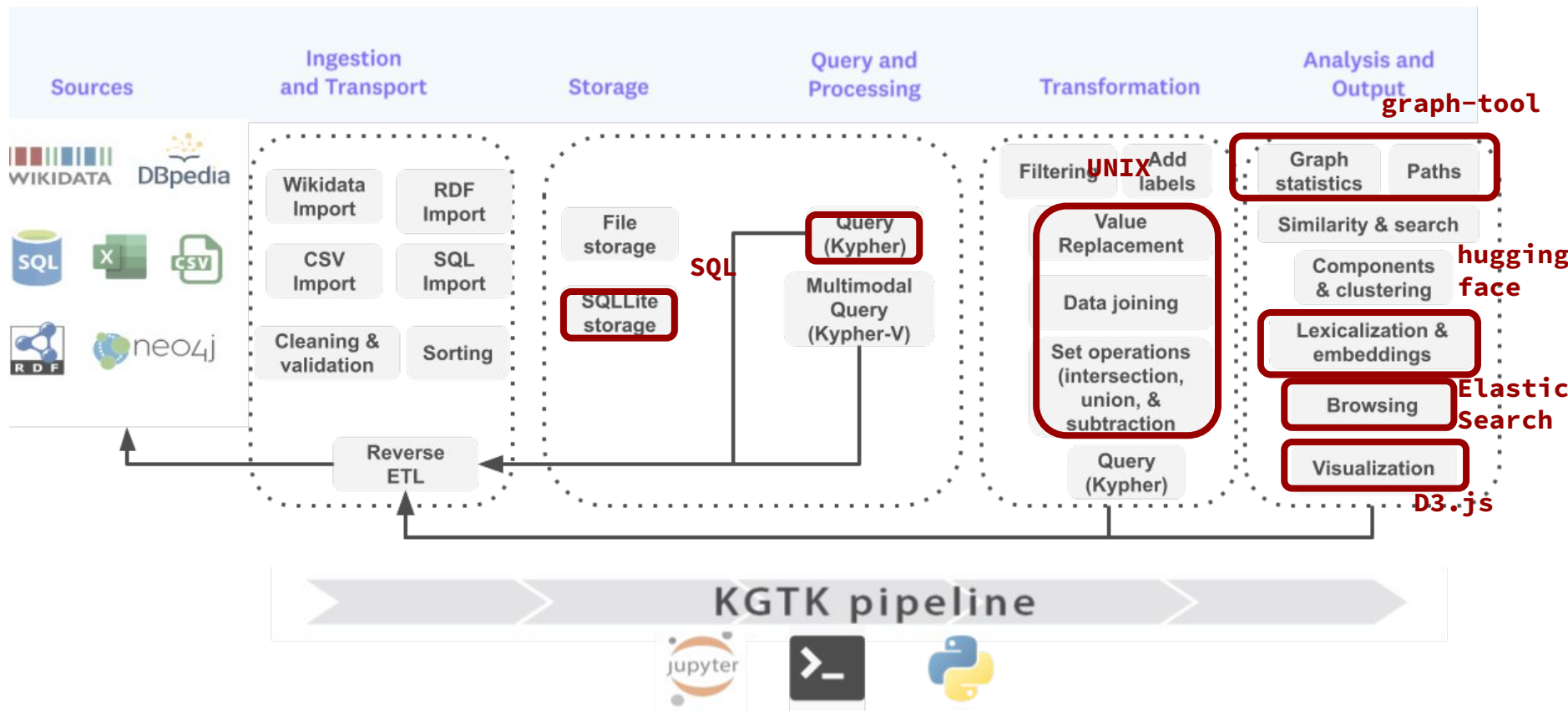
KGTK has a common API to many transformations



KGTK has a common API to advanced analytics

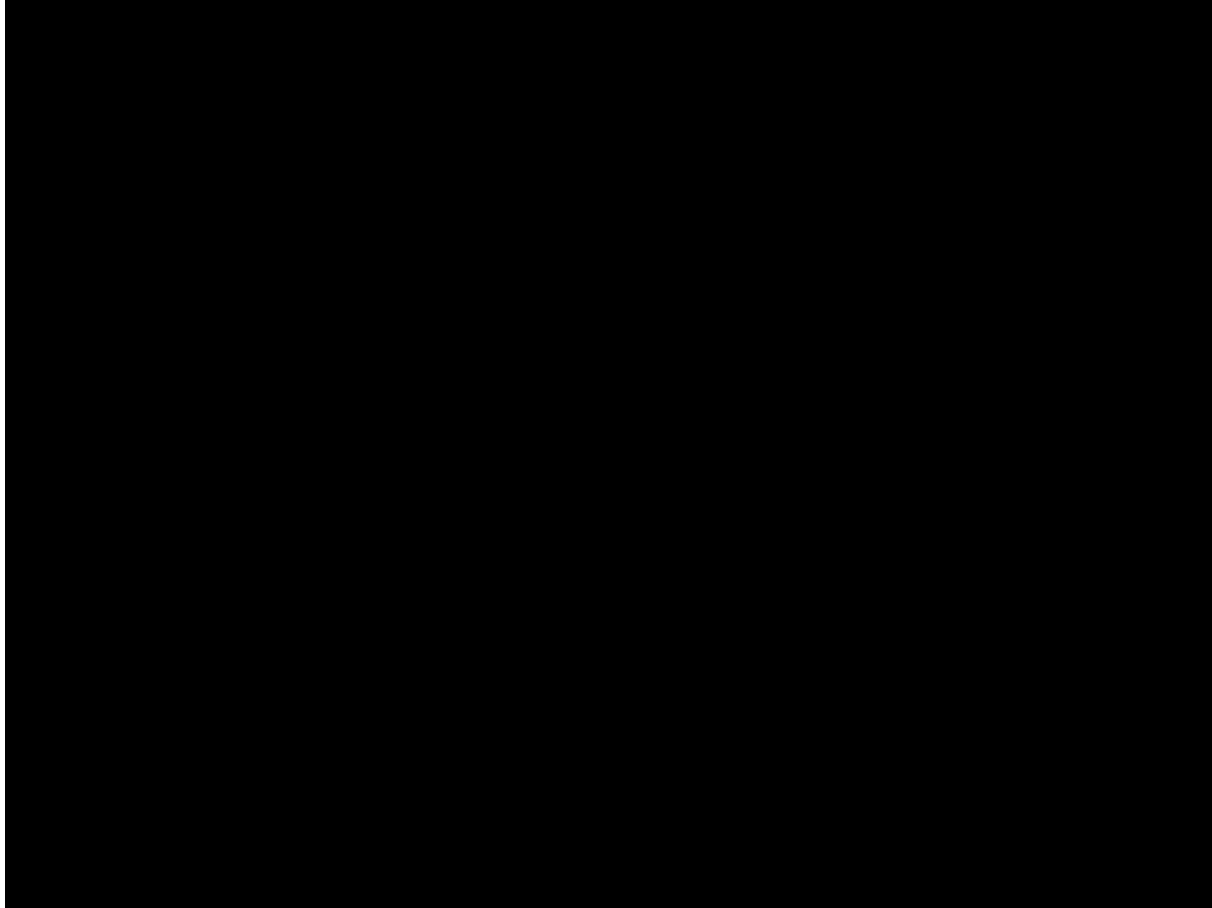


KGTK provides the best tool for each job



What can KGTK do?

Analyze hundreds of Wikidata dumps



*Terabytes
of data*

Validate Wikidata constraints in seconds

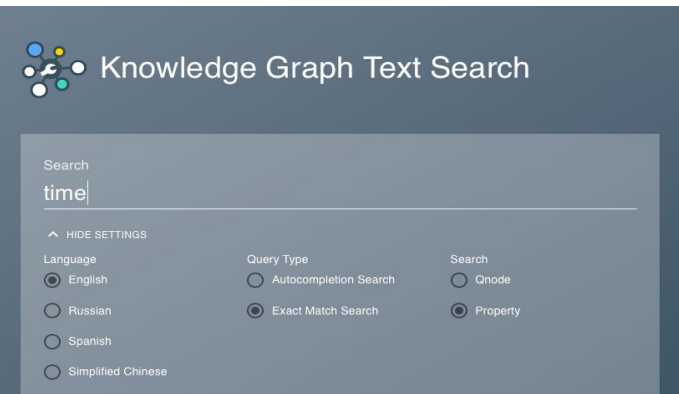
type constraint		value type constraint	
exception to constraint	prescriber	relation	instance or subclass of
relation	instance of	class	occupation
class	person		profession
	narrative entity		fictional profession
	fictional character		job
	animal		craft
	human		fictional religious position
	robot		
	group of humans		
	group of fictional characters		



item requires statement constraint
property instance of

constraint
• type
• value type
• irs
• symmetric
• inverse

Enable access for non-technical users



Knowledge Graph Text Search

Search
time

^ HIDE SETTINGS

Language

- English
- Russian
- Spanish
- Simplified Chinese

Query Type

- Autocompletion Search
- Exact Match Search

Search

- Qnode
- Property

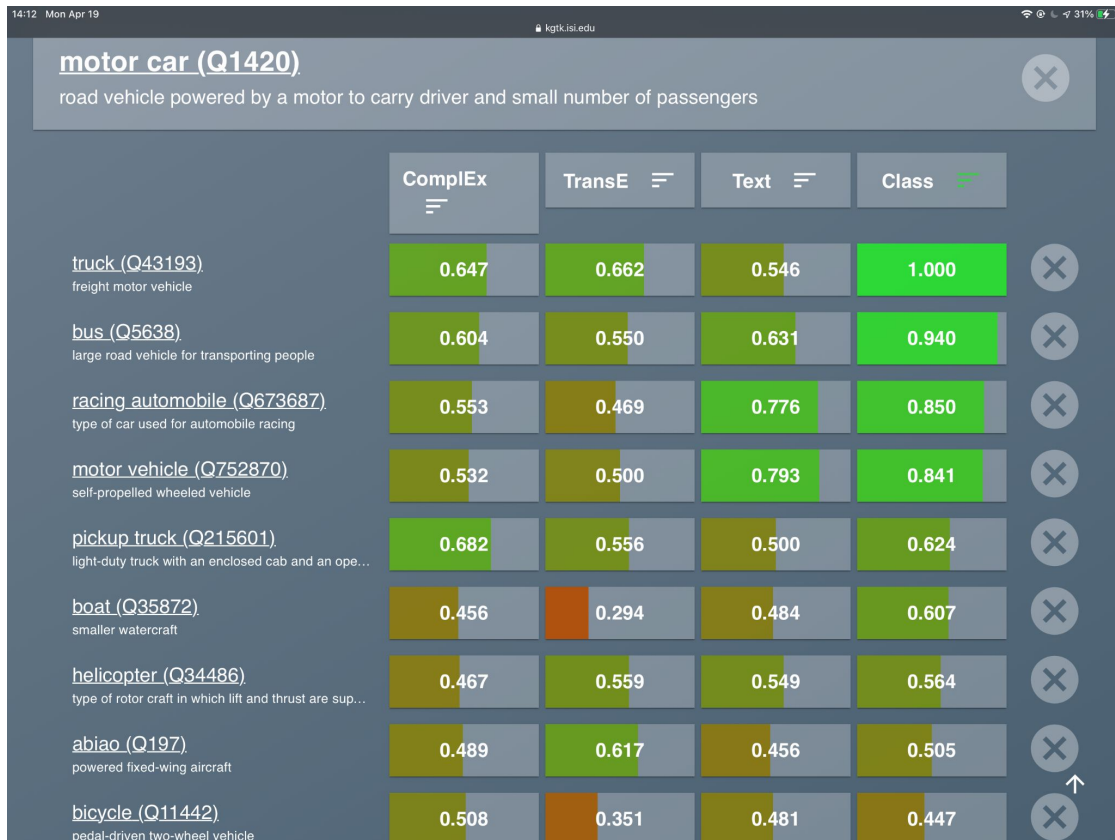
1. located in time zone (P421)

Description: time zone for this item
Alias: time zone, TZ, time, timezone

2. point in time (P585)

Description: time and date something took place, existed or a statement was true
Alias: event date, time, as of, during, when, year, at time, on, date, time of event, by date

<https://kgtk.isi.edu/search/>



14:12 Mon Apr 19 kgtk.isi.edu

motor car (Q1420)
road vehicle powered by a motor to carry driver and small number of passengers

	Complex	TransE	Text	Class	
<u>truck (Q43193)</u> freight motor vehicle	0.647	0.662	0.546	1.000	X
<u>bus (Q5638)</u> large road vehicle for transporting people	0.604	0.550	0.631	0.940	X
<u>racin<u>g</u> automobile (Q673687)</u> type of car used for automobile racing	0.553	0.469	0.776	0.850	X
<u>motor vehicle (Q752870)</u> self-propelled wheeled vehicle	0.532	0.500	0.793	0.841	X
<u>picku<u>p</u> truck (Q215601)</u> light-duty truck with an enclosed cab and an ope...	0.682	0.556	0.500	0.624	X
<u>boat (Q35872)</u> smaller watercraft	0.456	0.294	0.484	0.607	X
<u>helicopter (Q34486)</u> type of rotor craft in which lift and thrust are sup...	0.467	0.559	0.549	0.564	X
<u>abiao (Q197)</u> powered fixed-wing aircraft	0.489	0.617	0.456	0.505	X
<u>bicycle (Q11442)</u> pedal-driven two-wheel vehicle	0.508	0.351	0.481	0.447	X

<https://kgtk.isi.edu/similarity/>



KGTK Browser

short film

(Q24862)

short feature film | minor motion picture | short | short subject film | short movie
any film not long enough to be considered a feature film

Properties

instance of	film format	
	form	
subclass of	film	
Commons category	Short films	
duration	40 minute	
	nature of statement	less than
has quality	duration	
image	Katwe making.JPG	
less than	feature film	
	criterion used	duration
	medium-length film	
	criterion used	duration
opposite of	feature film	
	criterion used	duration
	named as	major motion picture
topic's main category	Category:Short films	

From Related Items

instance of	←	18 Months Later Short Film 2020
		Buffalo Dance
		Please Kill Me, I'm a Faggot Nigger Jew
		The Third Reich 'n Roll
		The Witch
		Q100166570
		Q100215065
		Q100257333
		Q100272327
		Q100272328

< 1 2 3 4 5 ... 3122 > 31,216 values

subclass of	←	animated short film
		cutscene

Gallery



Identifiers

Art & Architecture Thesaurus ID	300263846
ASC Leiden Thesaurus ID	386041393
Australian Educational Vocabulary ID	scot/6191
Bengali Banglapedia ID	ক্যামেরা পরিচালনা
BNCF Thesaurus ID	1120
Cultureel Woordenboek ID	film/korte-film
FAST ID	1710364
Freebase ID	/m/02hmvv
GND ID	4166233-7
Gran Enciclopèdia Catalana ID	0194175
JSTOR topic ID	short-films
KBpedia ID	ShortFilm
Library of Congress authority ID	sh85121810
Library of Congress Genre/Form Ter...	gf2011026570
Quora topic ID	Short-Films-1
Store norske leksikon ID	kortfilm
Thesaurus for Graphic Materials ID	tgm009550
UNESCO Thesaurus ID	concept17068
Yle topic ID	18-220295
YSA ID	Y101185
YSO ID	15697
Åvvir topic ID	oavnehis-filma

<https://kgtk.isi.edu/browser/Q24862>



short film

(Q24862)
 short feature film | minor motion picture | short | short subject film | short movie
 any film not long enough to be considered a feature film

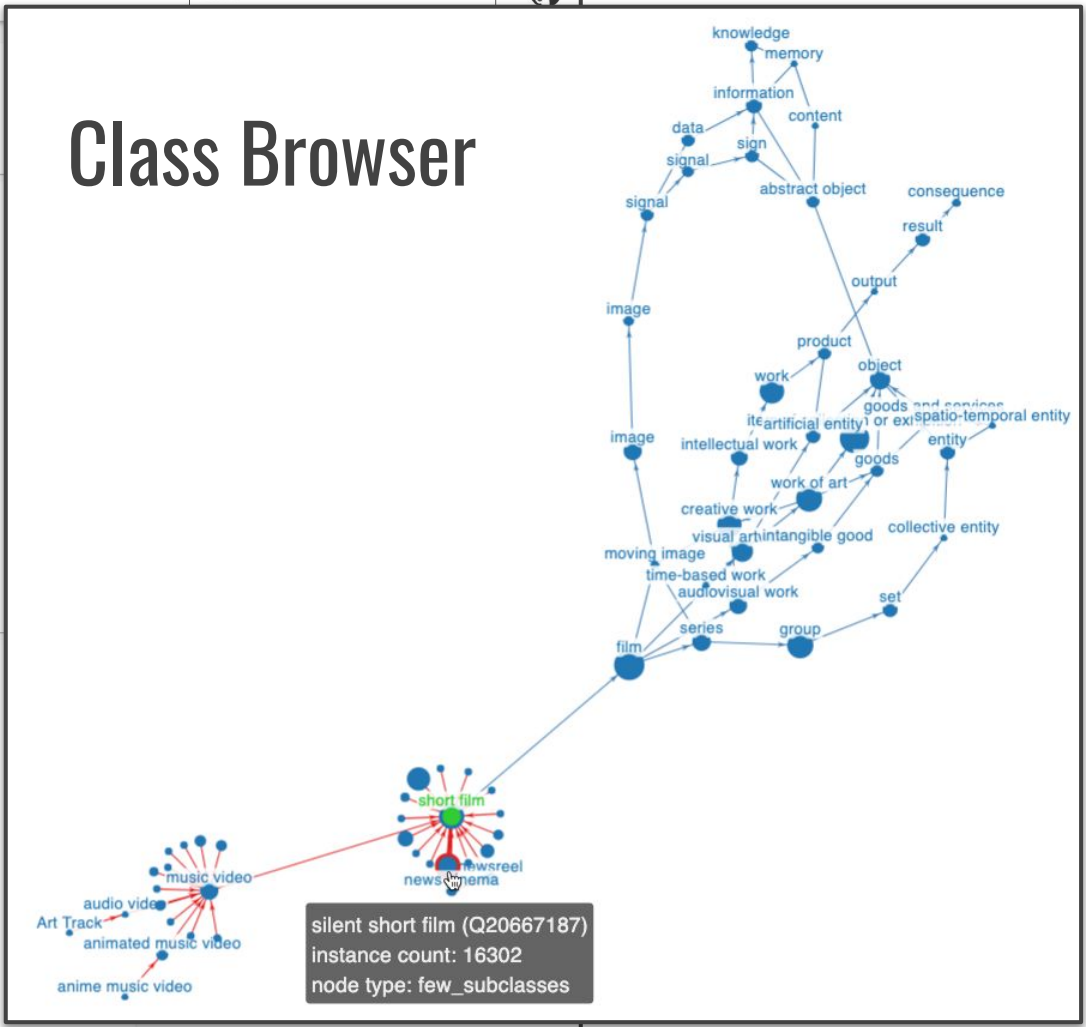
Properties

instance of	film format	
	form	
subclass of	film	
Commons category	Short films	
duration	40 minute	
	nature of statement	less than
has quality	duration	
image	Katwe making.JPG	
less than	feature film	
	criterion used	duration
	medium-length film	
	criterion used	duration
opposite of	feature film	
	criterion used	duration
	named as	major motion picture
topic's main category	Category:Short films	

From Related Items

instance of	←	<ul style="list-style-type: none"> 18 Months Later Short Film 2020 Buffalo Dance Please Kill Me, I'm a Faggot Nigger Jew The Third Reich 'n Roll The Witch Q100166570 Q100215065 Q100257333 Q100272327 Q100272328 	
			31,216 values
subclass of	←	<ul style="list-style-type: none"> animated short film cutscene 	

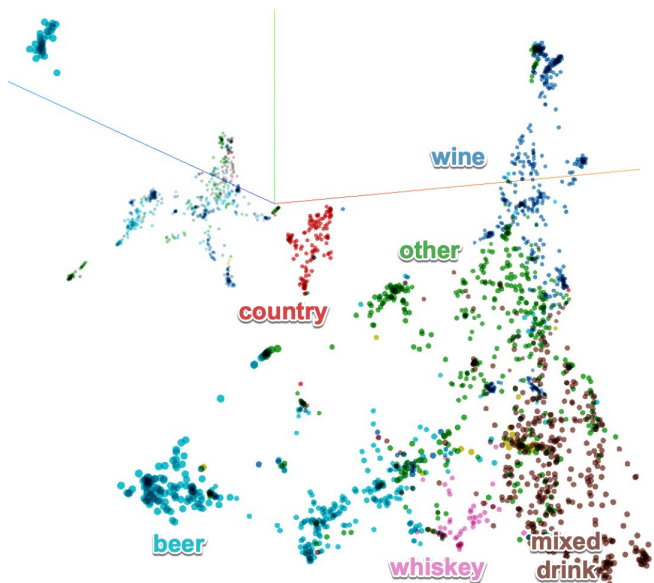
Class Browser



silent short film (Q20667187)
 instance count: 16302
 node type: few_subclasses

Compute various embeddings

Text Embeddings



Graph Embeddings



kgtk visualize-graph --help

Our ambition

NLP



spaCy

ML



KG



KGTK



Lab Program

Time (EST)	Content	Speaker
10:45 - 11:00	Welcome and introduction	Filip
11:00 - 11:20	Internet Memes: <i>Knowledge connects culture and creativity</i>	Filip
11:20 - 11:40	Financial transactions: <i>Detecting anomalies in trading</i>	Ke-Thia
11:40 - 12:00	PubGraphs: <i>What should I read next?</i>	Kian & Jay
12:00 - 12:20	Morality in events: <i>From news to timelines and graph maps</i>	Gleb
12:20 - 12:30	Feedback and discussion	Jay

Code for all demos in this lab:

<https://github.com/usc-isi-i2/kgtk-aaai2023>

Tutorial website with more materials

<https://usc-isi-i2.github.io/kgtk-tutorial-aaai-2023>

User survey: What are the requirements for **knowledge graph engineering in 2023?**

shorturl.at/gIPVY

Response time:
5 minutes (max)