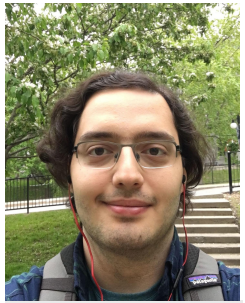
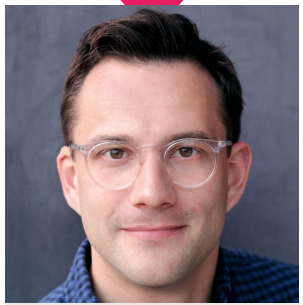
KGTK:
User-friendly Toolkit for Manipulation of Large Knowledge Graphs

USC Viterbi School of Engineering
Information Sciences Institute

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

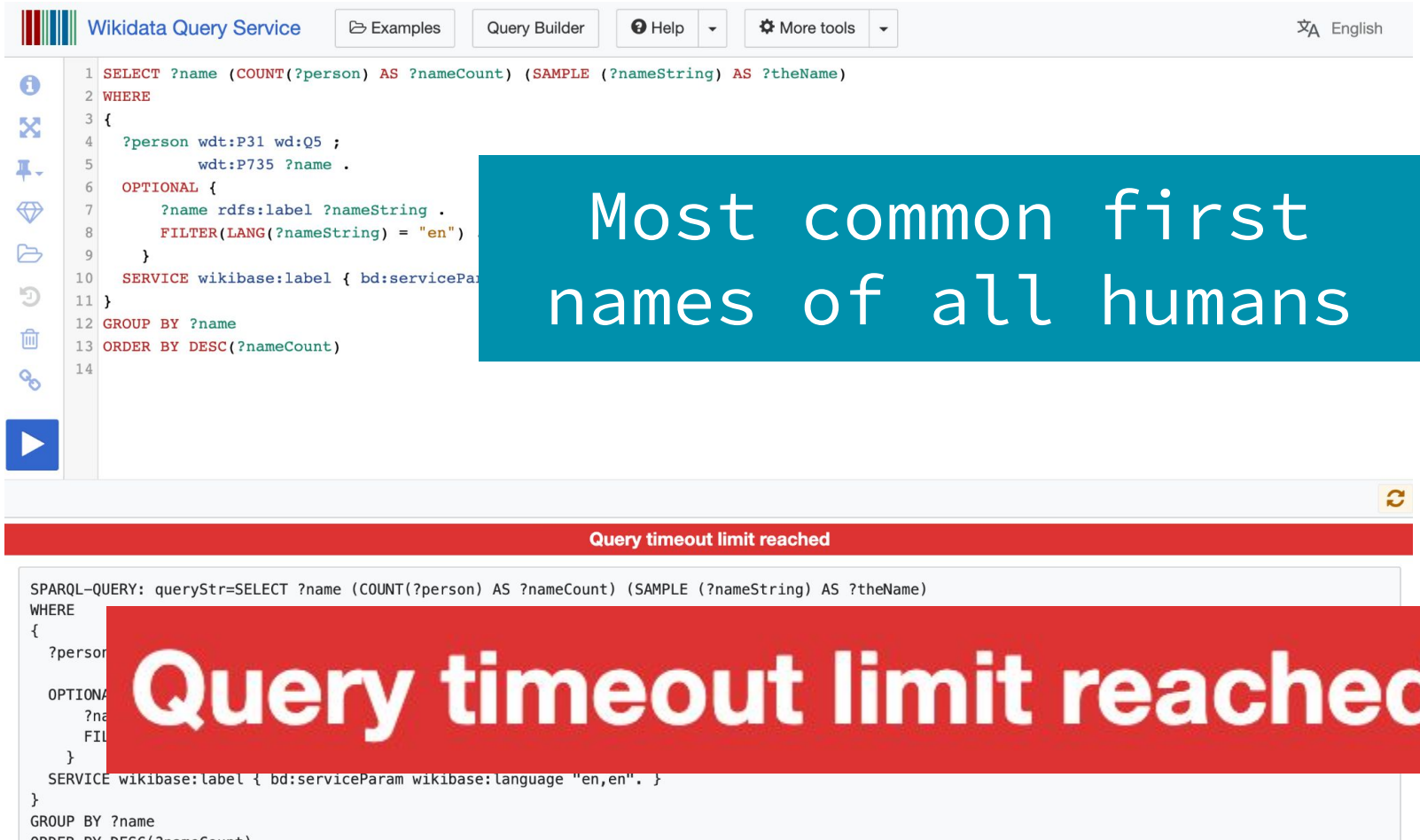
ML → scikit learn

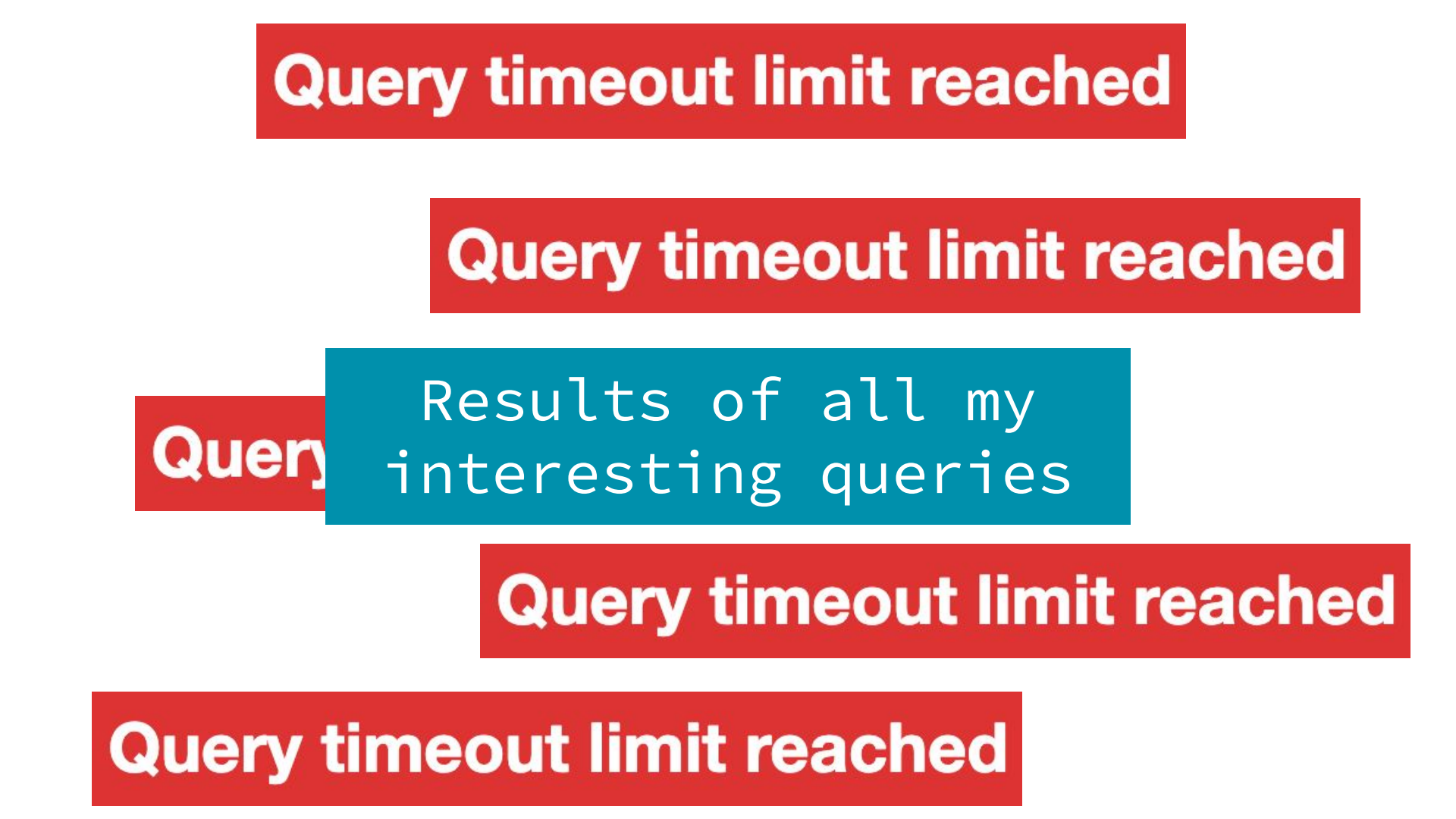
KG → ???

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	*various*
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




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 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

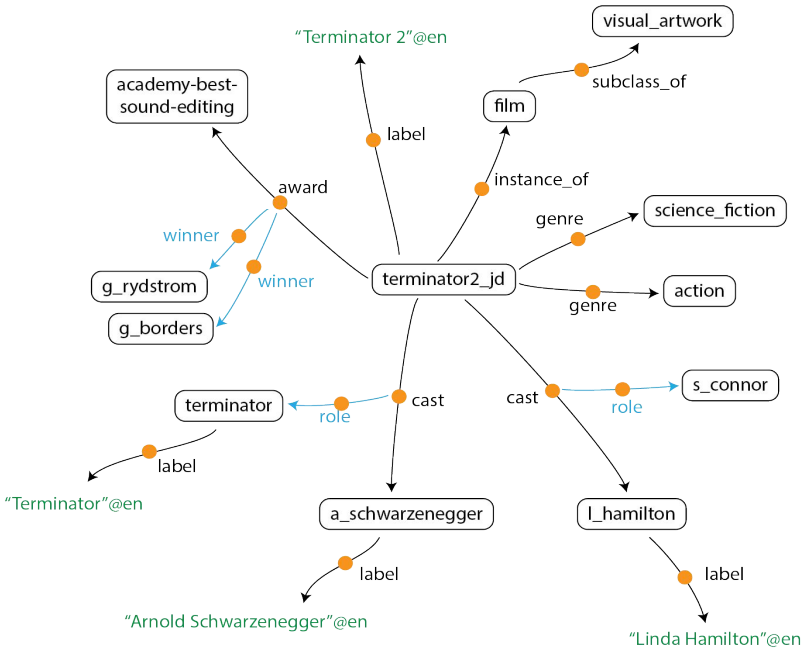
Support Wikidata data model: statements, qualifiers, references

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

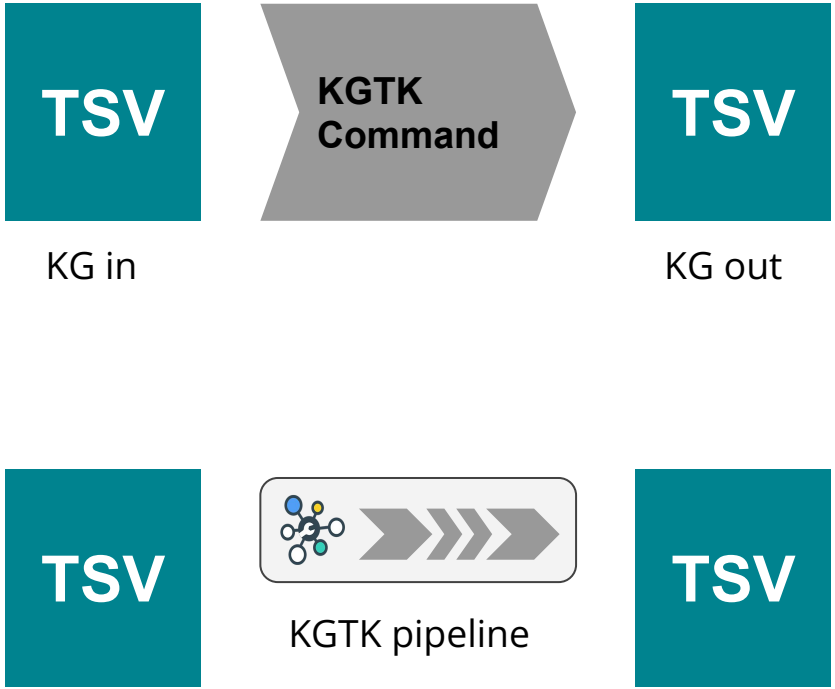
KGTK commands work with TSV KG files

KG in

KG out

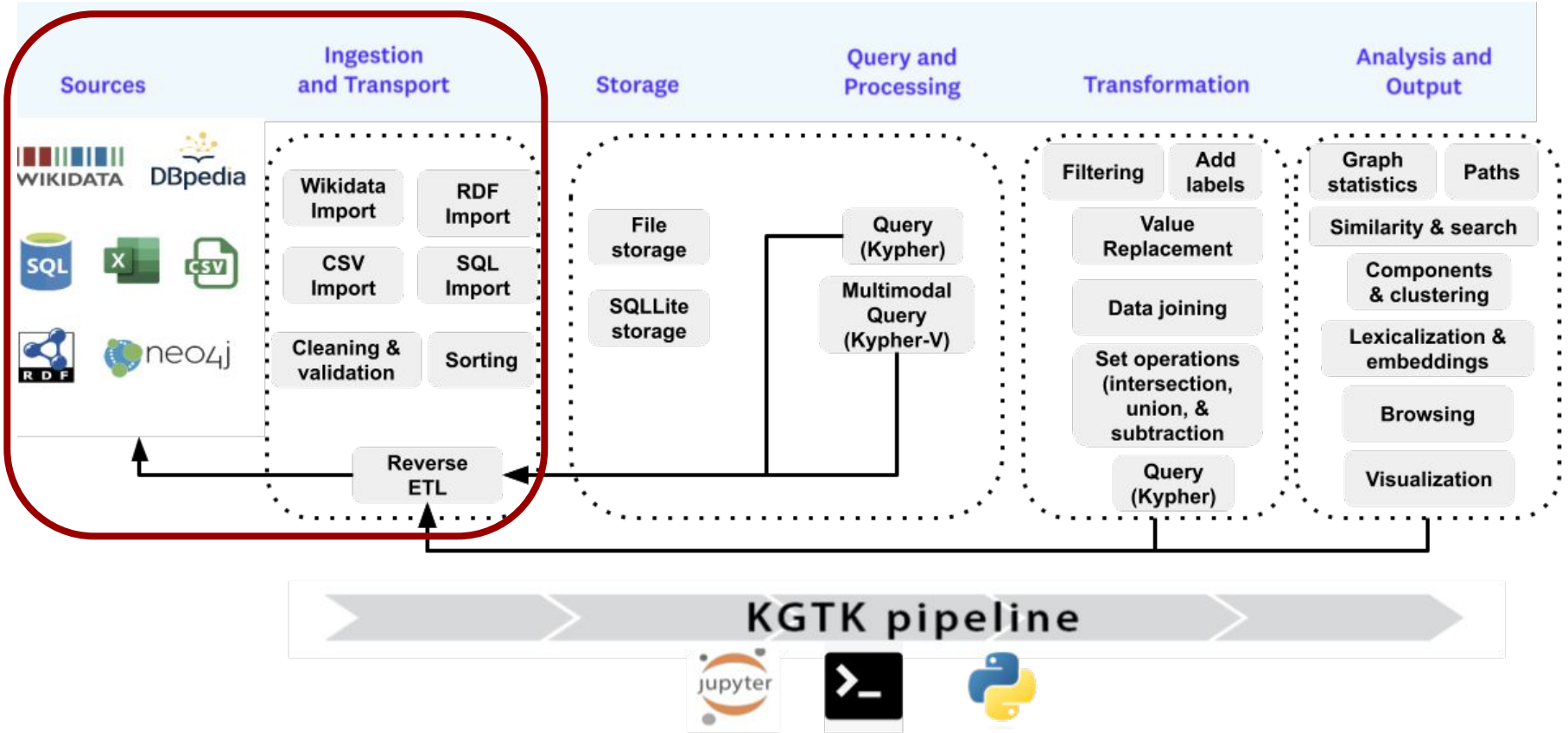
id	node1	label	node2
	terminator2_jd	label	"Terminator 2"@en
	terminator2_jd	instance_of	film
	terminator2_jd	genre	science_fiction
	terminator2_jd	genre	action
t4	terminator2_jd	cast	a_schwarzenegger
	t4	role	terminator
t6	terminator2_jd	cast	l_hamilton
	t6	role	s_connor
t8	terminator2_jd	award	academy_best_sound_editing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

KGTK pipelines consume and produce KGs

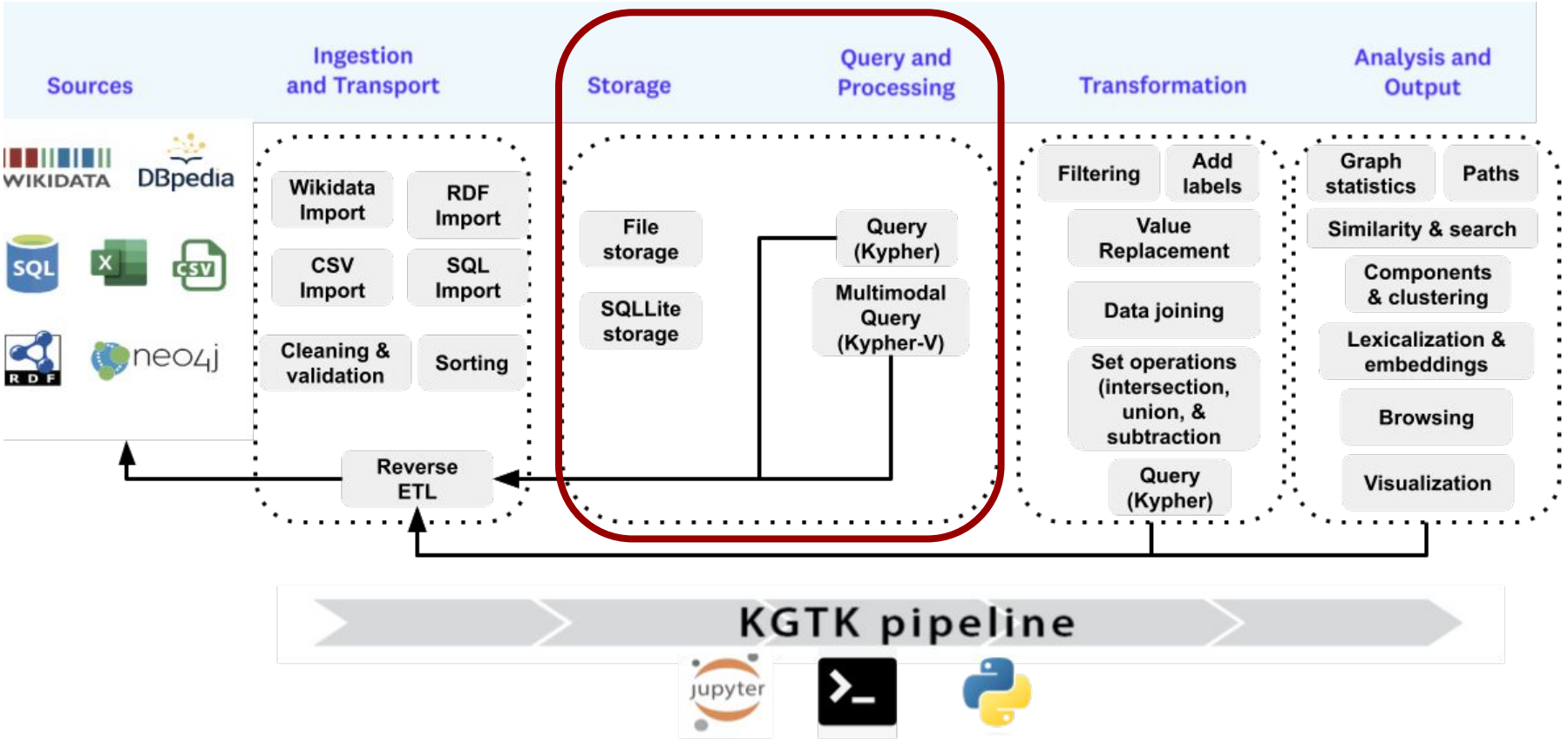


id	node1	label	node2
	terminator2_jd	label	"Terminator 2"@en
	terminator2_jd	instance_of	film
	terminator2_jd	genre	science_fiction
	terminator2_jd	genre	action
t4	terminator2_jd	cast	a_schwarzenegger
	t4	role	terminator
t6	terminator2_jd	cast	l_hamilton
	t6	role	s_connor
t8	terminator2_jd	award	academy_best_sound_editing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

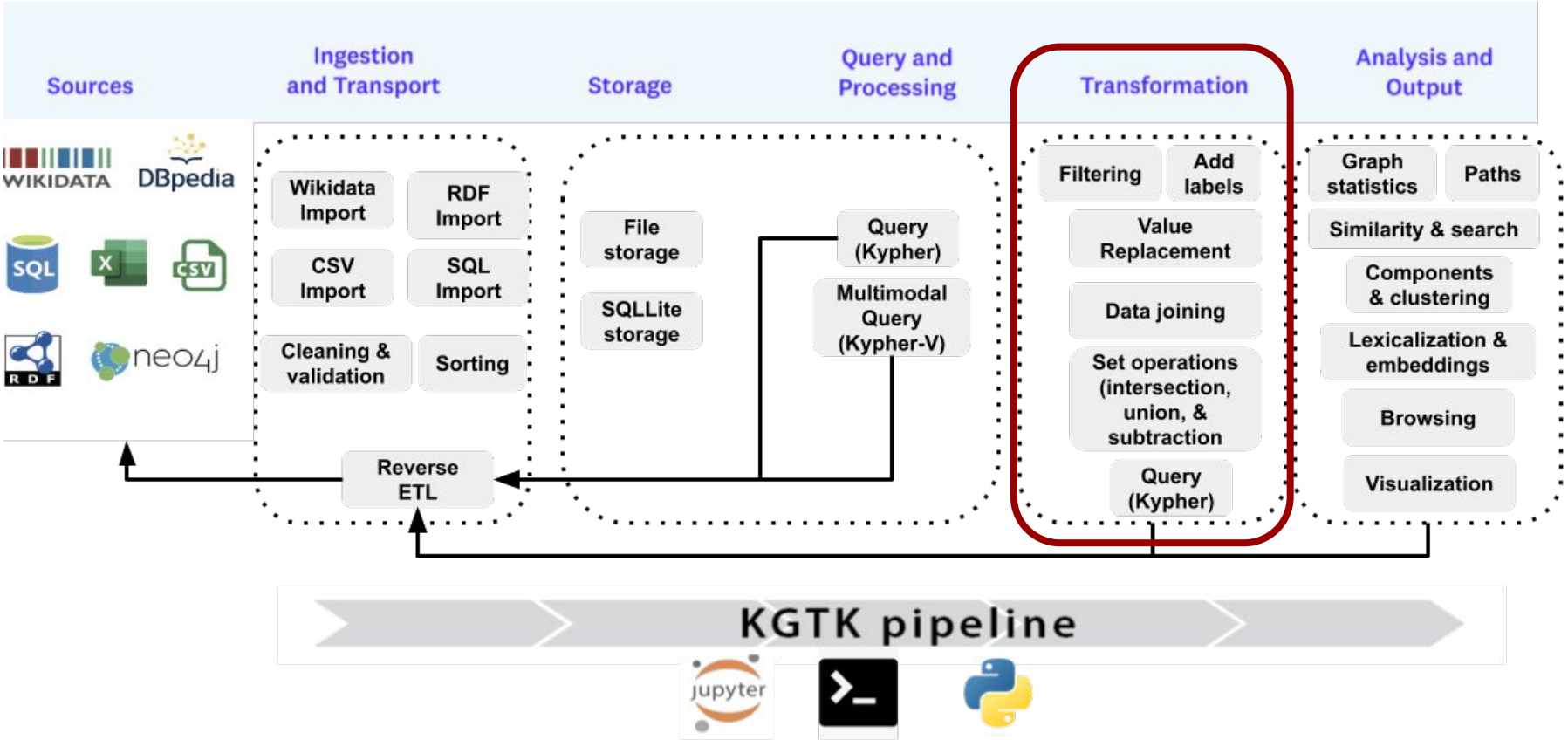
KGTK supports many formats



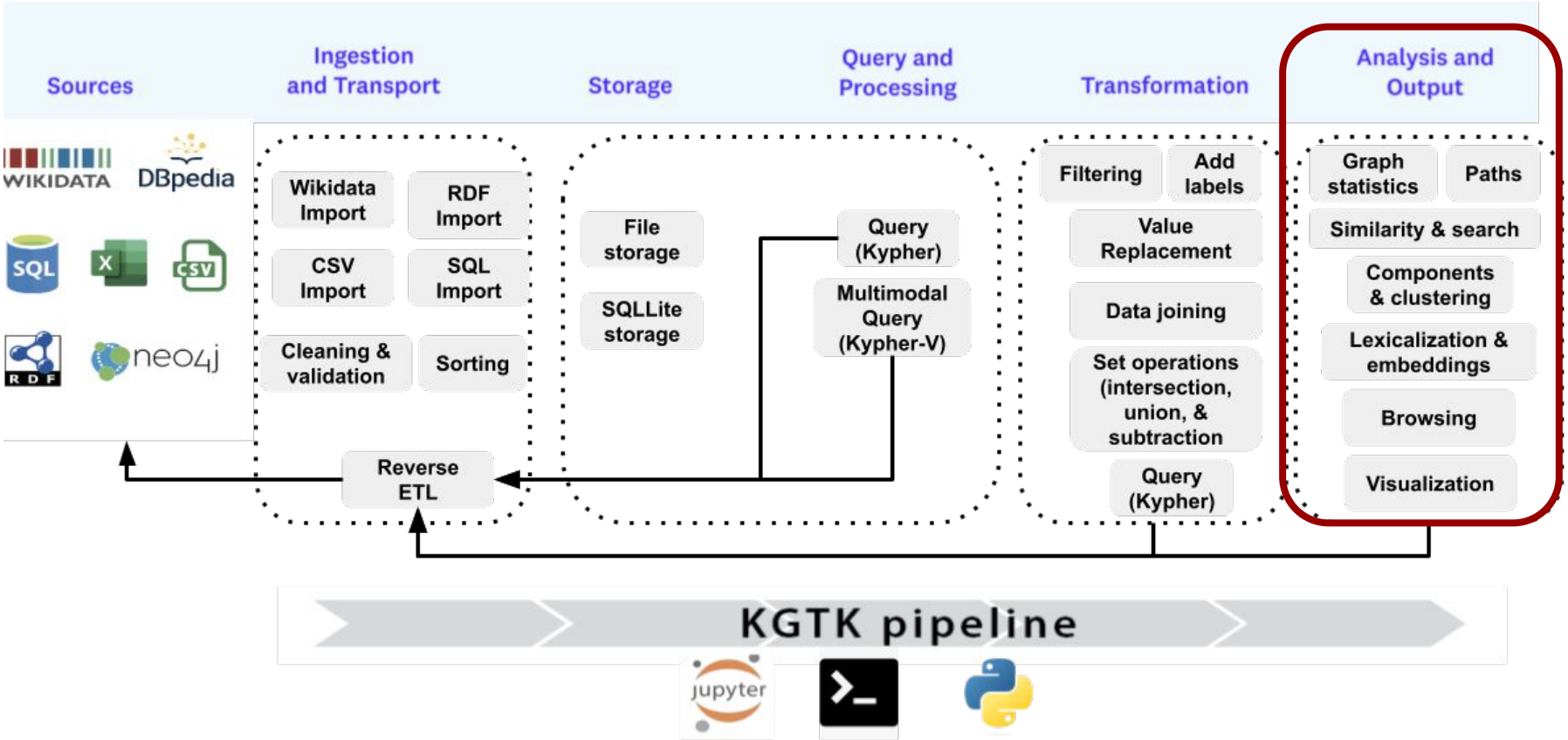
Kypher is a scalable and hybrid query engine



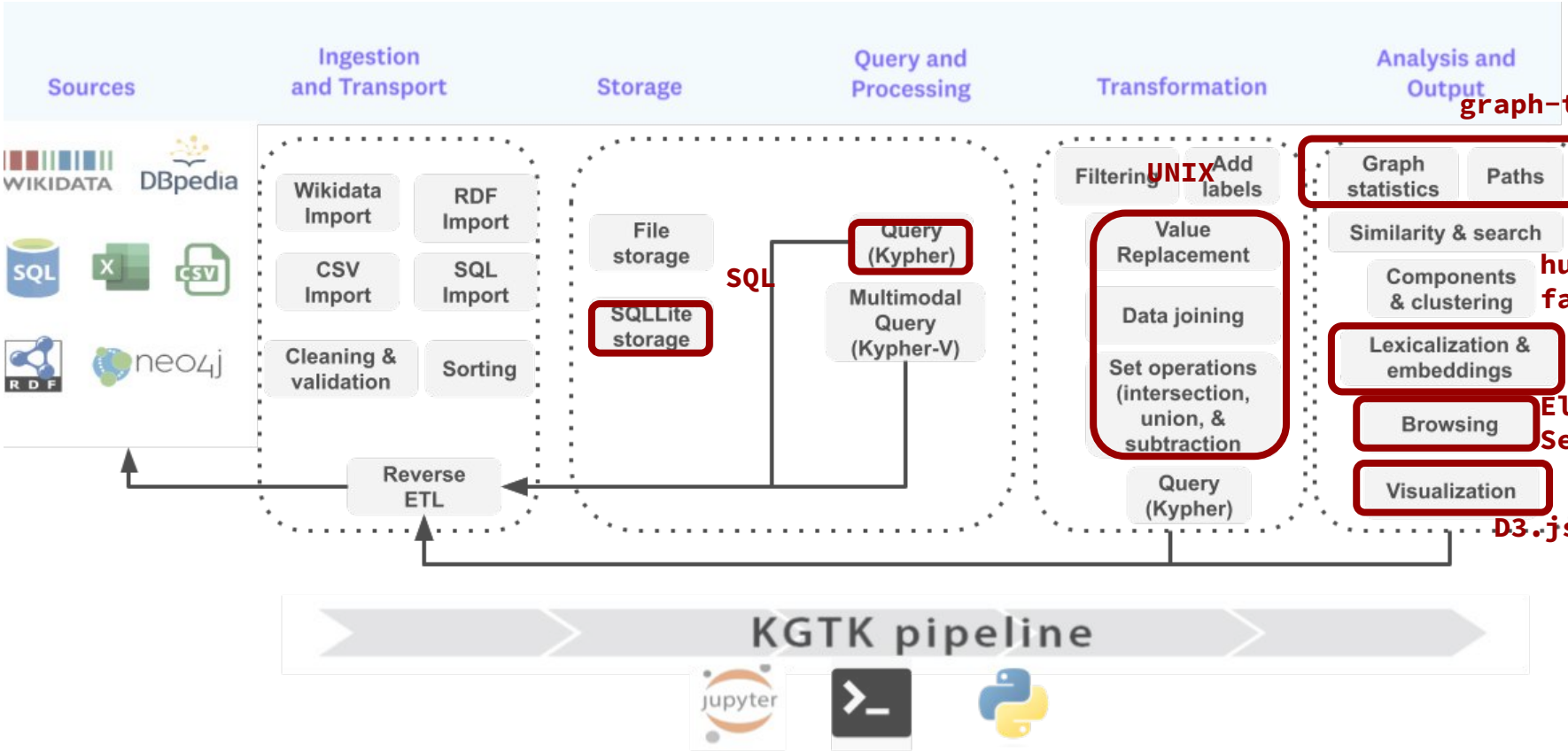
KGTK has a common API to many transformations



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

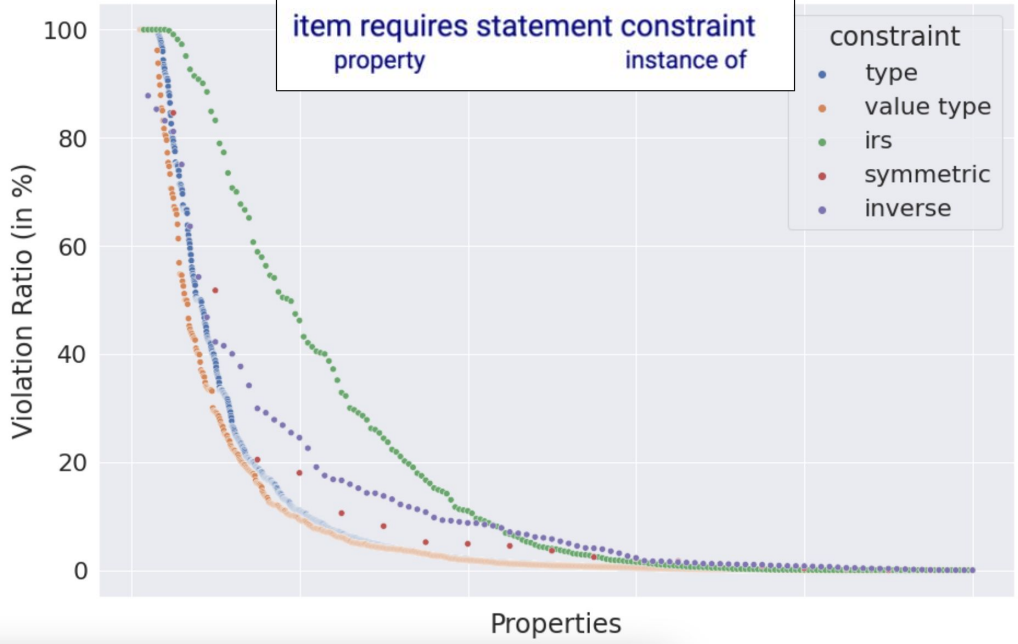
exception to constraint	prescriber
relation	instance of
class	person
	narrative entity
	fictional character
	animal
	human
	robot
	group of humans
	group of fictional characters

value type constraint

relation	instance or subclass of
class	occupation
	profession
	fictional profession
	job
	craft
	fictional religious position

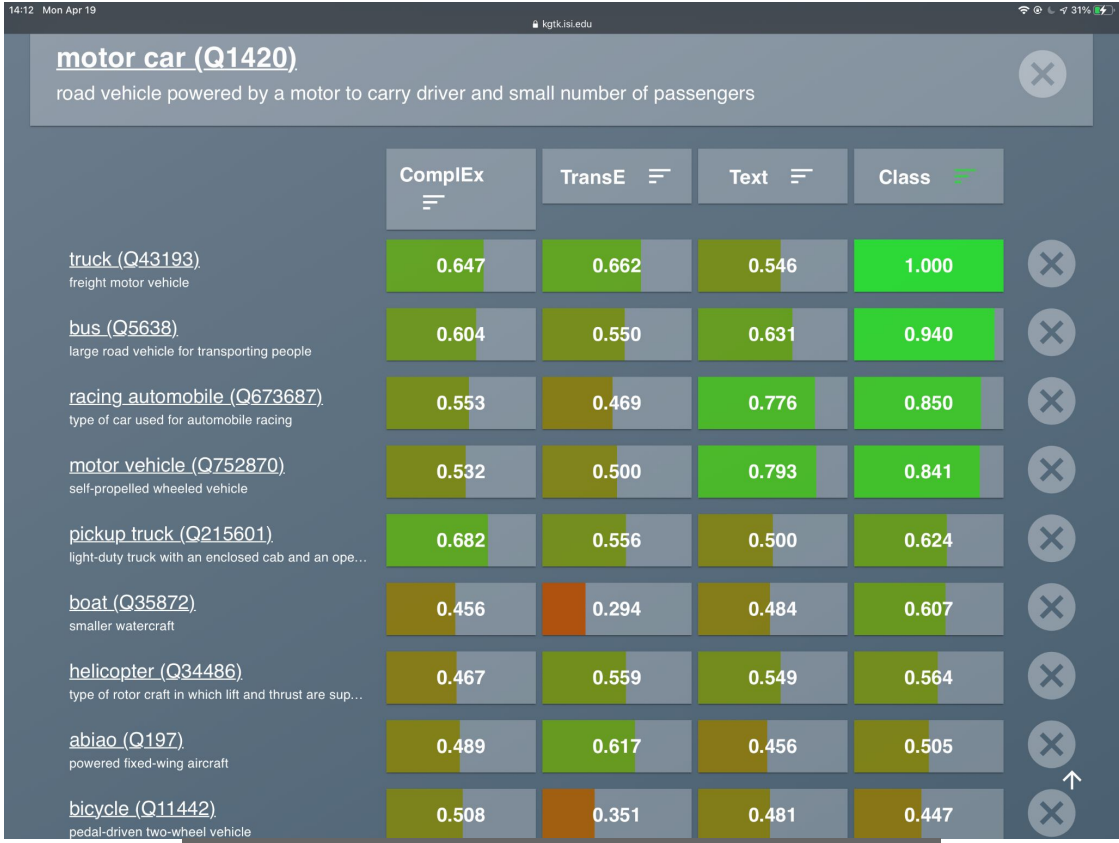
item requires statement constraint

property	instance of



Enable access for non-technical users

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



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

KGTK Browser

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

KGTK Browser: DWD Demo Knowledge Graph

Search...

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	স্বল্পদৈর্ঘ্য চলচ্চিত্র
BNCF Thesaurus ID	1120
Cultureel Woordenboek ID	film/korte-film
FAST ID	1710364
Freebase ID	/m/02hmvc
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
Ávvir topic ID	oanehis-filbma

KGTK Browser: DWD Demo Knowledge Graph

Search...

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

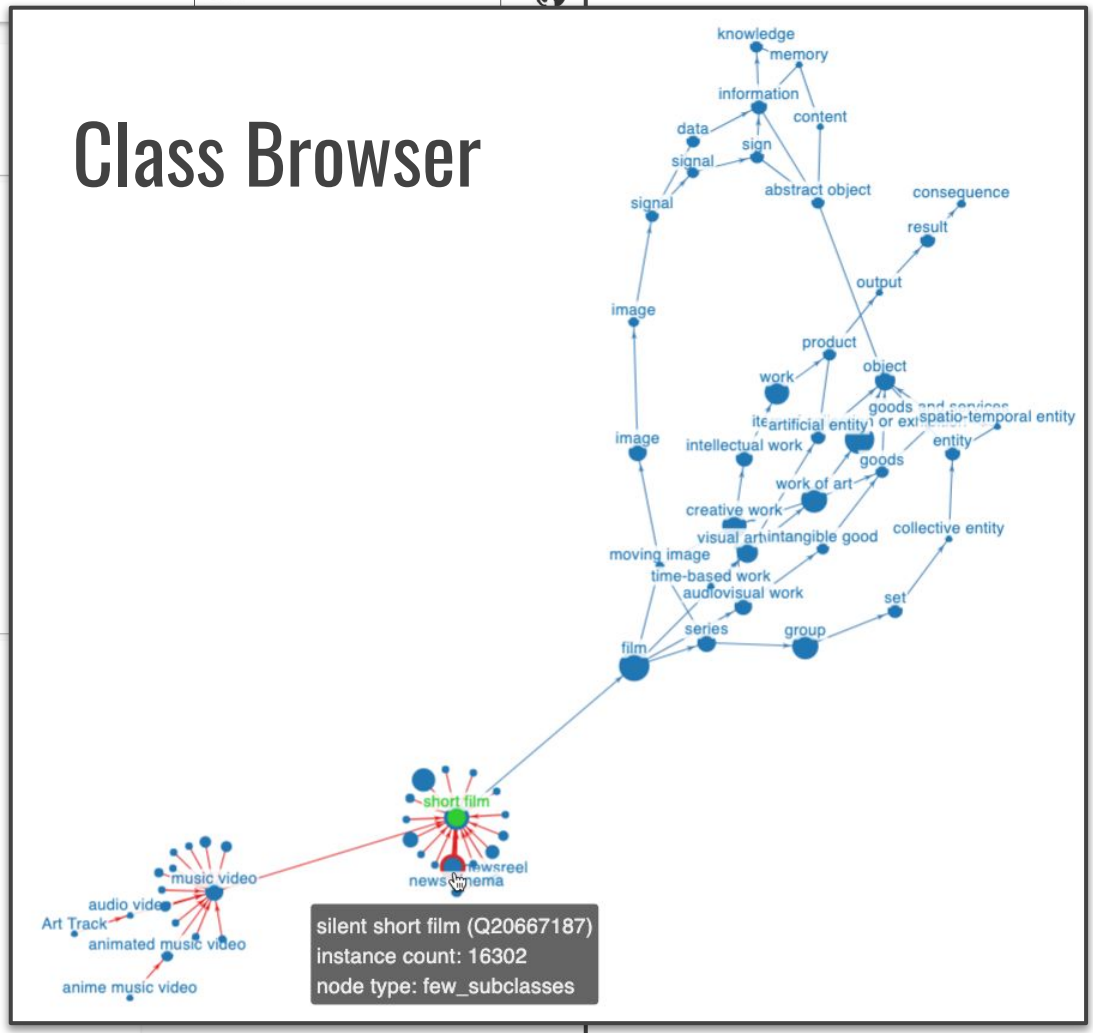
Properties

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

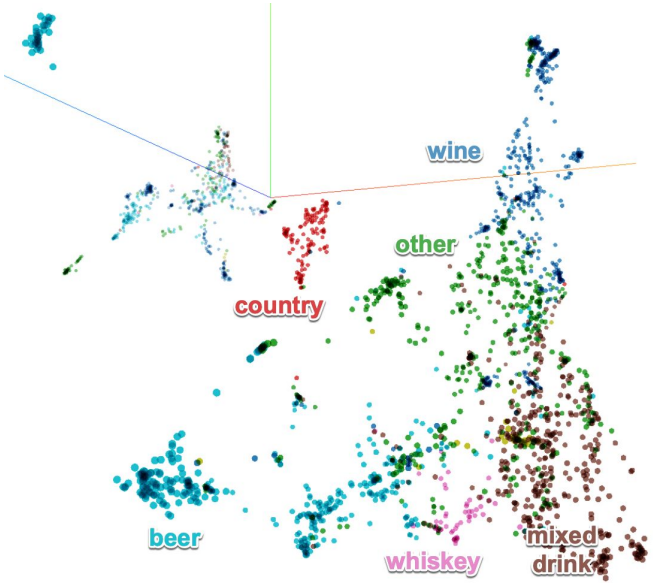
instance of	←	18 Months Later Short Film 2020
		Buffalo Dance
		Please Kill Me, I'm a Faggot Nigger Jew
		The Third Reich 'n Roll
		The Witch
		Q100166570
		Q100215065
		Q100257333
		Q100272327
		Q100272328
		‹ 1 2 3 4 5 … 3122 › 31,216 values
subclass of	←	animated short film
		cutscene

Class Browser



Compute various embeddings

Text Embeddings



Graph Embeddings

```
kgtk visualize-graph --help
```

Our ambition

NLP → spaCy

ML → scikit learn

KG → KGTK

Lab Program

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

Response time:
5 minutes (max)