TableView:
A Visual Interface for Generating Preview Tables of Entity Graphs

Sona Hasani*, Ning Yan[1#], Chengkai Li *

University of Texas at Arlington*
Innovative Database and Information Systems Research (IDIR) Laboratory

Huawei U.S. R&D Center#
[1]The work was done while at UTA.

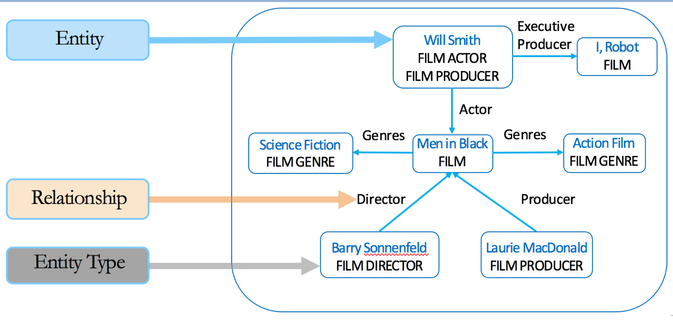
34th IEEE International Conference on Data Engineering
April 16th-19th
PARIS 2018

Entity Graphs

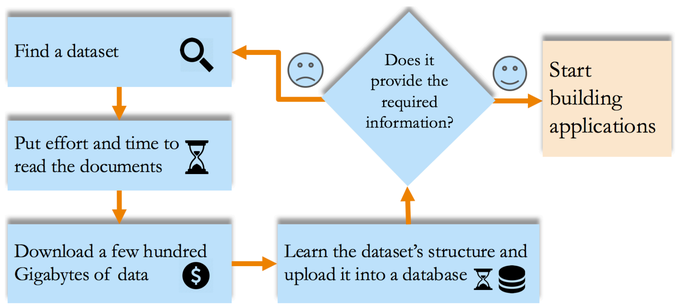
Large and complex graphs capturing millions of entities and billions of relationships between entities.

- Freebase : 1.9 billion triples
- DBpedia : 3 billion triples
- YAGO : 120 million triples
- Linked Open Data : 52 billion triples

Applications: search, recommendation systems, business intelligence, health informatics, fact checking

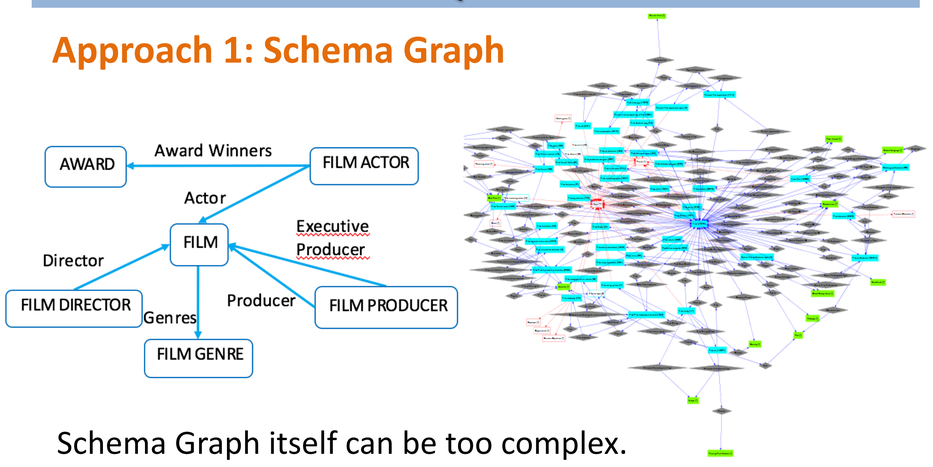


Steep Flag-Down Cost



Need for a Quick Overview

Approach 1: Schema Graph



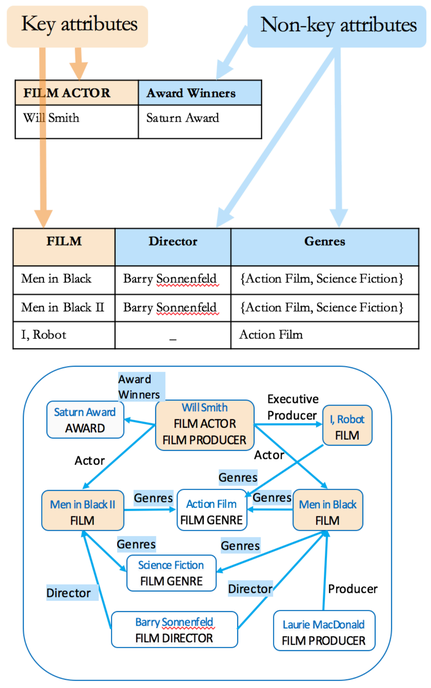
Schema Graph itself can be too complex.

Approach 2: Schema Summary

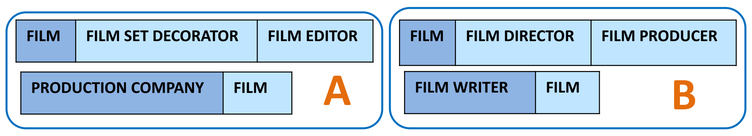
- Schema summarization in relational database [Yang PVLDB09 , Yang PVLDB11]
- XML summarization [Yu VLDB06]
- Graph summarization [Tian SIGMOD08, Zhang ICDE10]

Preview Tables

Approach 3: Preview Tables



Too Many Previews. Which One to Choose?



Aggregate Scoring

FILM	Actor	Genres
4	6	5

FILM ACTOR	Actor	Award Winners
2	6	2

Score of the Preview

$4 \times (6+5) = 44$ + 60

$2 \times (6+2) = 16$

Attribute Scoring

Key attribute scoring

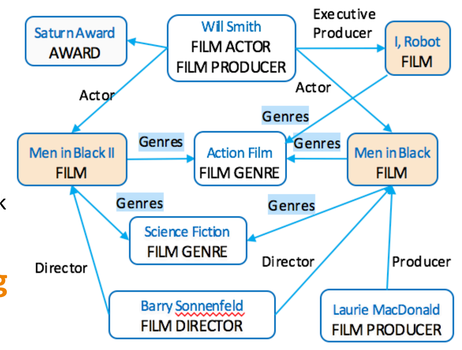
Coverage-based method:
Coverage(FILM) = 3

Random walk-based method :
Stationary distribution of a random walk process defined over the schema graph

Non-key attribute scoring

Coverage-based method :
Coverage(Genres) = 5

Entropy-based method :
Entropy(Genres) = (2/3) log(3/2)+(1/3) log(3/1) = 0.28



FILM	Director	Genres
Men in Black	Barry Sonnenfeld	{Action Film, Science Fiction}
Men in Black II	Barry Sonnenfeld	{Action Film, Science Fiction}
I, Robot	_	{Action Film}

Optimal Preview Discovery

Find the preview with highest score that satisfies

Size constraint
- Number of key attributes *K*
- Number of non-key attributes *N*

Distance between two preview tables *d*

Concise

Tight: dist(Ti, Tj) ≤ d

Diverse: dist(Ti, Tj) ≥ d

Tight

FILM	Performances	Genres	Directed By
FILM DIRECTOR	Films Directed		
FILM PRODUCER	Films Produced		

Diverse

FILM FESTIVAL	Location	Focus
FILM COMPANY	Films	
FILM CHARACTER	Portrayed in Film	

- Concise preview: dynamic programming algorithm
- Tight/Diverse preview: Apriori property algorithm

TableView Overview [Yan SIGMOD16, Hasani ICDE18demo]

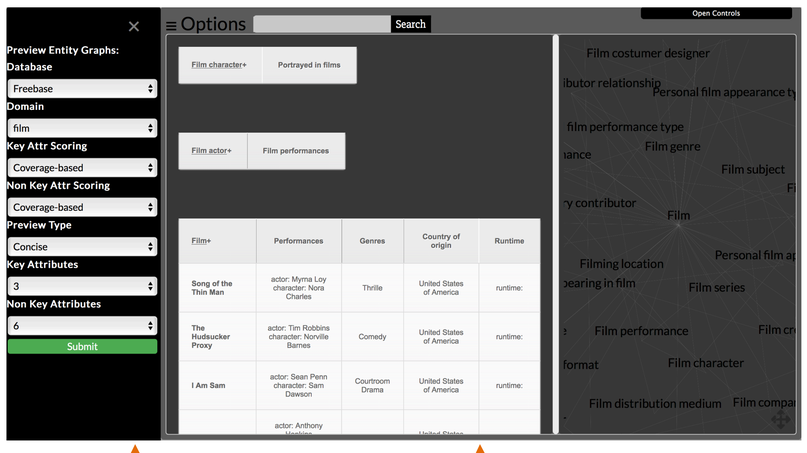
A concise preview generated for FILM domain with 3 key attributes and 6 non-key attributes.

Film+	Performances	Genres	Country of origin	Runtime
Song of the Thin Man	actor: Myrna Loy character: Nora Charles	Thrille	United States of America	runtime:
The Hudsucker Proxy	actor: Tim Robbins character: Norville Barnes	Comedy	United States of America	runtime:

Film character+	Portrayed in films
Elendil	actor: Peter McKenzie film: The Lord of the Rings: The Fellowship of the Ring
Foghorn Leghorn	actor: Mel Blanc film: False Hare special_performance_type: Voice

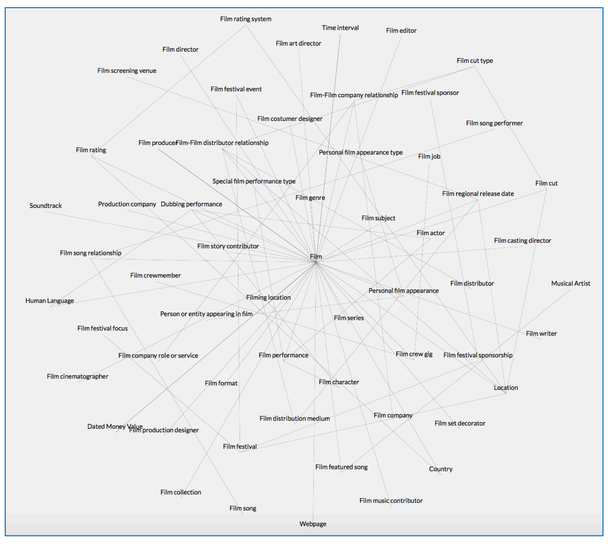
Film actor+	Film performances
Dusty Springfield	film: Music Scene: The Best of 1969-1970: Vol. 2
Cal Ripken, Jr.	film: Baseball the Ripken Way: Pitching

- **Domains:** film, book, music, TV, people
- **Key attribute scoring:** Coverage-base, Random Walk-based
- **Non-key attribute scoring:** Coverage-based, Entropy-based
- **Preview type:** Concise, Diverse, Tight



Configuration Panel

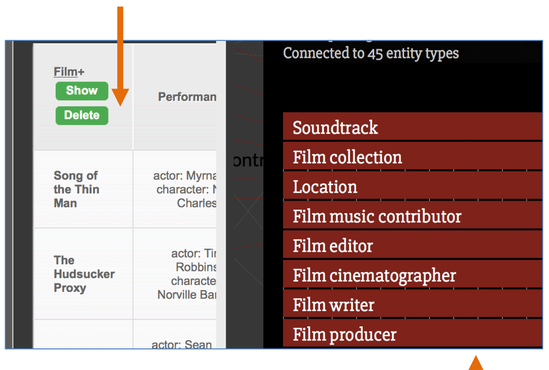
Preview Panel



Schema Graph

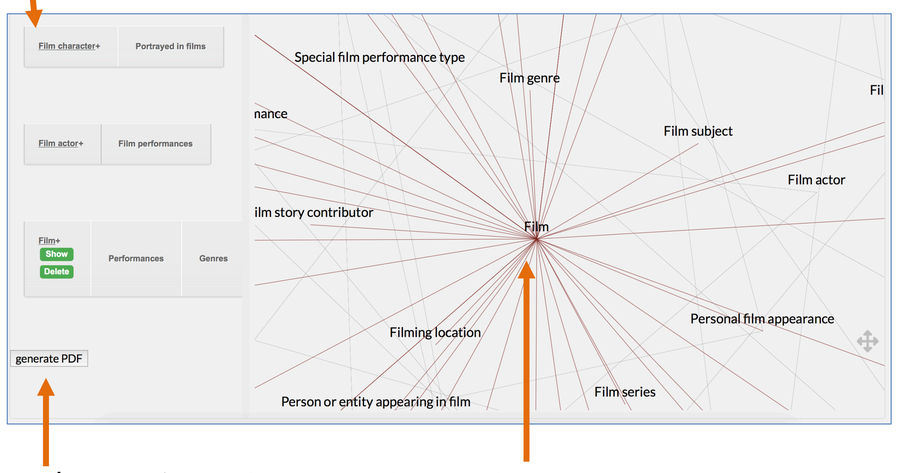
Interacting with the Preview Tables

Drop any column from the preview tables and refresh the preview.



Additional information for each column of the preview tables.

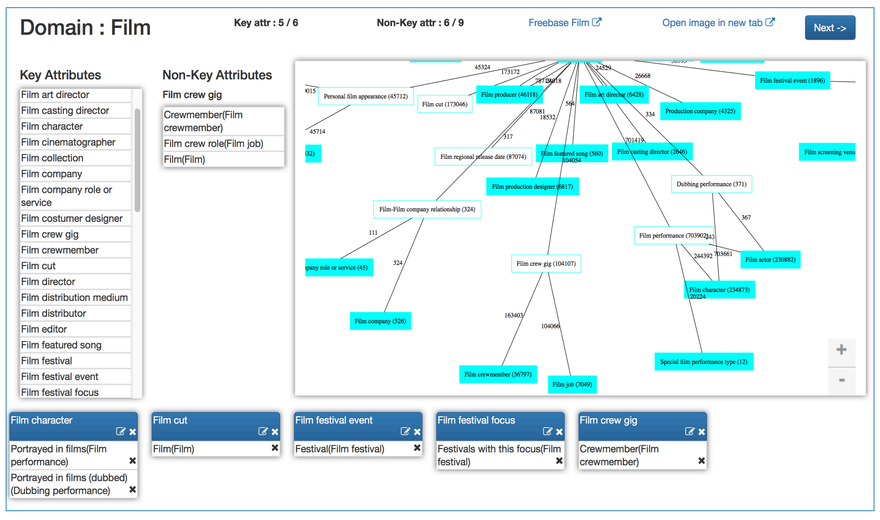
Expand and collapse the sample data section of the preview tables.



Export the preview into PDF.

Find the entity type in the schema graph that corresponds to the key attribute of a preview table.

Manual Preview Tables



References

[Yan SIGMOD16] N. Yan, S. Hasani, A. Asudeh, and C. Li, "Generating preview tables for entity graphs," in *SIGMOD*, 2016, pp. 1797–1811. (best reproducibility award from SIGMOD2017)
[Hasani ICDE18demo] S. Hasani, N. Yan, and C. Li, " TableView: A Visual Interface for Generating Preview Tables of Entity Graphs," in *ICDE*, 2018.
[Yang PVLDB09] X. Yang, C. M. Procopiuc, and D. Srivastava, "Summarizing relational databases," *PVLDB*, vol. 2, no. 1, pp. 634–645, 2009.
[Yang PVLDB11] X. Yang, C. M. Procopiuc, and D. Srivastava, "Summary graphs for relational database schemas," *PVLDB*, vol. 4, no. 11, pp. 899–910, 2011.
[Yu VLDB06] C. Yu and H. V. Jagadish, "Schema summarization," in *VLDB*, 2006, pp. 319–330.
[Tian SIGMOD08] Y. Tian, R. A. Hankins, and J. M. Patel, "Efficient aggregation for graph summarization," in *SIGMOD*, 2008, pp. 567–580.
[Zhang ICDE10] N. Zhang, Y. Tian, and J. M. Patel, "Discovery-driven graph summarization," in *ICDE*, 2010, pp. 880–891.

https://idir.uta.edu/tableview