

## **Querying Knowledge Graphs by Example Entity Tuples**

Nandish Jayaram<sup>1+</sup>, Arijit Khan<sup>2\*</sup>, Chengkai Li<sup>1</sup>, Xifeng Yan<sup>2</sup>, Ramez Elmasri<sup>1</sup>

University of Texas at Arlington<sup>1</sup>

University of California, Santa Barbara<sup>2</sup>



## **Knowledge Graphs**



## **Usability Challenges**

- Big and complex data: Lack of schema, challenging to users and developers.
- How to query the graph, and understand the results.

## **Related Work**

- Query-by-example in relational databases [Zloof'75]. **G**Keyword search and keyword-based query formulation [Chang et al.'11].
- □Set expansion [Wang et al.'07]. XML query relaxation [Amer-Yahia et al.'2005].

Exemplar queries [Mottin et al., 2014].

## **Query Interface**



## **GQBE** Architecture



## **Query Graph Discovery**



Ground truth based accuracy comparison of GQBE and NESS. Comparison of GQBE, NESS and Exemplar Queries. The measured

Query processing time for 2-tuple queries.

#### parameters are precision-at-k, Mean Average Precision and normalized Discounted Cumulative Gain. Quarty

Query	Tuple1			Tuple2			Combined (1,2)			Tuple3			Combined (1,2,3)		
	P@k	nDCG	AvgP	P@k	nDCG	AvgP	P@k	nDCG	AvgP	P@k	nDCG	AvgP	P@k	nDCG	AvgP
F <sub>1</sub>	0.36	0.76	0.32	0.36	1.00	0.50	0.12	0.38	0.02	0.36	0.73	0.22	0.12	0.49	0.02
$F_2$	0.76	1.00	0.79	0.00	0.00	0.00	0.80	1.00	0.80	0.12	0.70	0.05	0.80	1.00	0.91
$F_4$	0.32	0.73	0.09	0.40	0.65	0.08	1.00	1.00	0.45	N/A	N/A	N/A	N/A	N/A	N/A
F <sub>6</sub>	0.24	0.89	0.16	0.28	0.89	0.18	0.40	0.87	0.16	0.36	0.98	0.22	0.12	0.94	0.07
F <sub>8</sub>	0.92	0.79	0.20	1.00	1.00	0.27	0.96	0.98	0.24	0.48	0.86	0.08	1.00	1.00	0.27
F <sub>9</sub>	0.68	0.72	0.23	0.56	0.66	0.17	0.80	0.86	0.35	1.00	1.00	0.62	1.00	1.00	0.66
F <sub>17</sub>	0.32	1.00	0.33	0.64	0.83	0.25	0.32	1.00	0.32	0.56	0.84	0.23	0.68	1.00	0.46

# Quary PCC Ouery PCC Ouery

Query	ru	Query	ru	Query	ree	Query	ru
$F_1$	0.79	$F_2$	0.78	F <sub>3</sub>	0.60	$F_4$	0.80
F <sub>5</sub>	0.34	F <sub>6</sub>	0.27	F <sub>7</sub>	0.06	F <sub>8</sub>	0.26
F9	0.33	F <sub>10</sub>	0.77	F <sub>11</sub>	0.58	$F_{12}$	undefined
F <sub>13</sub>	undefined	$F_{14}$	0.62	$F_{15}$	0.43	F <sub>16</sub>	0.29
F <sub>17</sub>	0.64	F <sub>18</sub>	0.30	F <sub>19</sub>	0.40	$F_{20}$	0.65

Pearson Correlation Coefficient (PCC) between GQBE and Amazon Mechanical Turk Workers, for *k*=30. MTurk workers were presented with answer pairs and asked for their preference between the two answers in each pair. 20000 such opinions collected.

#### Query Processing Time (secs.) 🖾 GQBE 🛛 NESS 🖾 Baseline 160 120 Lattice Nodes 80 # edges # edges 12 13 in MQG in MQG Query



## **Technical Details and Demo**

N. Jayaram, A. Khan, C. Li, X. Yan and R. Elmasri. Querying knowledge graphs by example entity tuples, TKDE 2015. N. Jayaram, M. Gupta, A. Khan, C. Li, X. Yan and R. Elmasri. GQBE: Querying knowledge graphs by example entity tuples, ICDE 2014.

Demo URL: <u>http://idir.uta.edu/gqbe</u>





#### Query processing times of GQBE, NESS and Baseline.

#### Lattice nodes evaluated comparison.

\*Now with Pivotal, \*Now with Nanyang Technological University, Singapore

#### Accuracy of GQBE on multi-tuple queries, k = 25.