Exemplar Queries: Give me an Example of What You Need

Davide Mottin, Matteo Lissandrini, Themis Palpanas, Yannis Velegrakis

University of Trento
dbTrento - Database and Information Management Group

Sep 4, 2014
I want to find acquisitions like the one of Youtube by Google.
Existing Search Engines

Google search for 'google youtube acquisition'
Existing Search Engines

Google

google youtube acquisition

About 25,100,000 results (0.31 seconds)

List of mergers and acquisitions by Google - Wikipedia, the ...
en.wikipedia.org/.../List_of_mergers_and_acquisitions_by_Go... Wikipedia

Google has acquired over 100 companies, with its largest acquisition being the service company Next New Networks, which became YouTube Next Lab and ...

History of YouTube - Wikipedia, the free encyclopedia

Jump to Purchase by Google (2006) - The deal to acquire YouTube closed on November 13, and was, at the time, Google's second largest acquisition.

Google To Acquire YouTube for $1.65 Billion in Stock ...
googlepress.blogspot.com/.../google-to-acquire-youtube-for-165_09.ht... MOUNTAIN VIEW, Calif., October 9, 2006 – Google Inc. (NASDAQ: GOOG) announced today that it has agreed to acquire YouTube, the consumer media ...

Google's Best And Worst Acquisitions - ABC News
abcnews.go.com Money ABC News

Jan 15, 2014 - Of all the companies Google has acquired, which have been the most ... AdMob and YouTube—"although at YouTube the revenue and profit ...

Google in Talks To Buy YouTube For $1.6 Billion - WSJ.com
online.wsj.com/news/.../SB11601481385788491... The Wall Street Journal
Existing Search Engines

Google

Acquisitions like Google Youtube

Web News Videos Images Shopping More Search tools

About 4,120,000 results (0.35 seconds)

Is Facebook deal like Google buying YouTube? - MarketWa…
www.marketwatch.com/.../is-facebook-deal-like-google-bu... MarketWatch
Feb 20, 2014 - Facebook's $19 billion acquisition of WhatsApp is similar to Google buying YouTube, analysts say.

Boutiques.com Traffic Drops 94% – Did Google Give Up On ...
www.signature9.com/.../boutiques-com-traffic-drops-94-did-google-give... As big as their acquisitions can be, Google's track record on non-search ... For every YouTube, Google likely has 2 Dodgeballs; and if their drop in traffic is any ... They're going to more precisely defined sites like ShopStyle (product search) ...

Google in Talks To Buy YouTube For $1.6 Billion - WSJ.com
online.wsj.com/news/.../SB11601481385788491... The Wall Street Journal
Google is in talks to acquire YouTube for about $1.6 billion. ... Like Web browsers and search engines before them, YouTube and social-networking sites are ...

Was YouTube a good acquisition for Google? - Quora
www.quora.com/Google-YouTube-Acquisition.../Was-YouTube-a... Quora
Google-YouTube Acquisition (October 2006): Was YouTube a good ... From a distance, it looks like Google basically left the product (and team) alone while ...

History of YouTube - Wikipedia, the free encyclopedia
Existing Search Engines

Acquisitions like Google Youtube

About 4,120,000 results (0.35 seconds)

Is Facebook deal like Google buying YouTube? - MarketWatch
www.marketwatch.com/.../is-facebook-deal-like-google-bu...
Feb 20, 2014 - Facebook’s $19 billion acquisition of WhatsApp is similar to Google
Yahoo!-del.icio.us or Microsoft-Skype not present as interesting acquisitions.

Yahoo!-del.icio.us or Microsoft-Skype not present as interesting acquisitions.
A new perspective

Exemplar: Google Youtube

About 4,120,000 results (0.35 seconds)

Google in Talks To Buy YouTube For $1.6 Billion - WSJ.com

Google is in talks to acquire YouTube for about $1.6 billion. ... Like Web browsers and search engines before them, YouTube and social-networking sites are ...

Yahoo.io.cio.us? – Yahoo Acquires Del.io.cio.us | TechCrunch

Dec 8, 2005 - ... instant message conversation with Joshua Schachter, the founder of Del.io.cio.us. I was asking him for any comment on the Yahoo acquisition.

Microsoft's Hefty Phone Bill for Skype - WSJ.com

Microsoft racked up a whopping $8.5 billion phone bill to buy Skype even ... Whether Microsoft can make a Skype acquisition work—especially at such a rich ...

Facebook buys WhatsApp for $19 billion - Feb. 19, 2014

Feb 19, 2014 - WhatsApp, the largest messaging service in the world, will now be under the ... and the Web, Facebook has acquired WhatsApp for $19 billion.
**Exemplar Queries**

**Input:** $Q_e$, an example element of interest

**Output:** set of elements in the desired result set

**Exemplar Query Evaluation**

Given a database $D$ and an exemplar query $Q_e$, find the set

$$\{a \mid \exists s \in \text{eval}(Q_e) \land a \approx s\},$$

where $a$ and $s$ are structures in $D$ and the symbol $\approx$ indicates a similarity function.
Knowledge-base Samples & Edge-Isomorphism

Exemplar Queries

Search Engines

Business

Auto Industry

Google

YouTube

Yahoo!

del.icio.us

GM

Opel

Menlo Park

S. Clara County

S. Clara

S. Mateo

California

Flint

Genesee

Michigan

isA

acquired

foundedIn

activity

in

of
Solution

**Input**: User Query $Q$, an example of the expected results.

**Output**: Set of expected results

**Procedure**: 
- Detect the sample for the query $Q$
- Find the structures **edge-isomorphic** to the sample

**Issues**
(sub)-graph isomorphism is NP-complete [2]
Solution

**Input:** User Query $Q$, an example of the expected results.

**Output:** Set of expected results

**Procedure:**
- Detect the sample for the query $Q$
- Prune the non-matching nodes
- Find the structures edge-isomorphic to the sample

**Issues**
(sub)-graph isomorphism is NP-complete [2]
**Input:** User Query $Q$, an example of the expected results.

**Output:** Set of expected results

**Procedure:**
- Detect the sample for the query $Q$
- Prune the non-matching nodes
- Find the structures edge-isomorphic to the sample

**Issues**

(sub)-graph isomorphism is NP-complete [2]

**Solution**

1. IterativePruning: fast reject non matching nodes
Theorem

Given a database $D = \langle N, E \rangle$ and a user sample $S$, let $N_d$ and $N^d$ be the $d$-neighborhood of $D$ and $S$ respectively. If there exists a subgraph-isomorphism $\mu : E^s \rightarrow E$, then $\forall n_s \in E^s$, $N^d(n_s) \subseteq N_d(n)$, $n \in E$, $n \in \mu(n_s)$.
Theorem

Given a database $D = \langle N, E \rangle$ and a user sample $S$, let $N_d$ and $N^S_d$ be the $d$-neighborhood of $D$ and $S$ respectively. If there exists a subgraph-isomorphism $\mu : E_S \rightarrow E$, then $\forall n_s \in E_S, N^S_d(n_s) \subseteq N_d(n), n \in E, n \in \mu(n_s)$.
**Theorem**

Given a database $D = \langle N, E \rangle$ and a user sample $S$, let $N_d(D)$ and $N_d^S$ be the $d$-neighborhood of $D$ and $S$ respectively. If there exists a subgraph-isomorphism $\mu : E_S \rightarrow E$, then $\forall n_s \in E_S, N_d^S(n_s) \subseteq N_d(n)$, $n \in E$, $n \in \mu(n_s)$.
Theorem

Given a database $D = \langle N, E \rangle$ and a user sample $S$, let $\mathcal{N}_d$ and $\mathcal{N}_d^S$ be the $d$-neighborhood of $D$ and $S$ respectively. If there exists a subgraph-isomorphism $\mu : E_S \to E$, then $\forall n_s \in E_S, \mathcal{N}_d^S(n_s) \subseteq \mathcal{N}_d(n), n \in E, n \in \mu(n_s)$
The Iterative Pruning Algorithm

1. Start from a query node \( q \)
2. Match \( q \) with the graph nodes
3. For each adjacent node of \( q \)
4. Find nodes in the graph from candidate map of \( q \) matching the edge
5. Repeat 2. with an adjacent node of \( q \) until all nodes have been visited
The IterativePruning Algorithm

1. Start from a query node \( q \)
2. Match \( q \) with the graph nodes
3. For each adjacent node of \( q \)
4. Find nodes in the graph from candidate map of \( q \) matching the edge
5. Repeat 2. with an adjacent node of \( q \) until all nodes have been visited

Theorem

Given a user sample \( S \), if Algorithm IterativePruning terminates with a complete exploration of the nodes in \( S \), then there exists in \( \mu \) a simulation \( R \) of the user sample \( S \).
**Input**: User Query $Q$, an example of the expected results.

**Output**: Set of expected results

**Procedure**:
- Detect the sample for the query $Q$
- **Restrict the search space**
- Prune the non-matching nodes
- Find the structures edge-isomorphic to the sample

**Issues**
(sub)-graph isomorphism is NP-complete [2]

**Solution**
1. IterativePruning: fast reject non matching nodes
2. RelevantNeighborhood: restrict the search space to “near” nodes
Restricting the search space

Idea

1. Not all the nodes are equally relevant.
2. Nodes "far" from the query are less related.
Restricting the search space

Idea

1. Not all the the nodes are equally relevant.
2. Nodes “far” from the query are less related.
Restricting the search space

Idea

1. Not all the nodes are equally relevant.
2. Nodes “far” from the query are less related.
Personalized PageRank

Used to prune the search space by identifying the valuable portions

\[ \mathbf{v} = (1 - c)\mathbf{A}\mathbf{v} + cp \]

**RelavantNeighborhood algorithm**

1. Edge probability computed as inverse frequency

\[ I(e_{ij}^\ell) = I(\ell) = \log \frac{1}{P(\ell)} = -\log P(\ell) \]

\[ P(\ell) = \frac{|\mathcal{R}^\ell|}{|\mathcal{R}|}. \]

2. Edge labels in the query assigned a higher weight

3. Real-time simulation of PPR (Weighted Particle Filtering [4])
Combination of two factors:
1. Structural: similarity of two nodes in terms neighbor relationships
2. Distance-based: the PageRank already computed
Structural similarity

\[ \rho(n_s, n) = \lambda S(n_s, n) + (1 - \lambda) v[n] \]
Distance-based similarity

\[ \rho(n_s, n) = \lambda S(n_s, n) + (1 - \lambda) v[n] \]
Experimental Setup

- **Dataset**
  - Freebase: 52M nodes, 232M edges (entire!)
  - Freebase Internet Domain: 2M nodes, 6M edges
  - Synthetic datasets

- **Testset**: 90 queries manually mapped from AOL query logs

- **Baseline**: NeMa [6]: approximate answers on graphs

**Measures**

- Algorithms total time
- Quality and usefulness of the answers (user study)
Scalability results (10M nodes)

RelevantNeighborhood is stable with respect to the number of answers

< 150ms to get the answers
Usefulness

Quality
- 92% people say that Exemplar Queries are useful
- 62% already had the need for such a service

Comparison
- Which method is preferred?
  - 64% Exemplar Queries
  - 30% Other approaches

92% people say that Exemplar Queries are useful. 62% already had the need for such a service. Which method is preferred? 64% Exemplar Queries and 30% Other approaches.
Related work

- **Query modification:**
  - Query refinement: modify the original query to restrict or enlarge the answer set [8, 7]
  - Query diversification: include results that are somehow diverse and new with respect to the one in the original query answer set [3, 9]
- **Knowledge-base mapping:** match plain text to entities in a knowledge base [1]
- **Querying graphs:** find similar or isomorphic structures in a graph [6, 5]
Conclusions

- introduce Exemplar Queries: a novel query paradigm
- present exact and approximate solutions and a novel ranking function
- demonstrate scalability of the proposed solutions on large real knowledge graphs
- verify the usefulness of the Exemplar Queries for real users
References I

A survey on ontology mapping.

The complexity of theorem-proving procedures.

Finding dimensions for queries.
In *CIKM*, 2011.

Topic-sensitive pagerank.
In *WWW*, 2002.

Neighborhood based fast graph search in large networks.
In *SIGMOD*, 2011.
References II


