Reconstructing Custom Fragments of Google Knowledge Graph on the Fly*

Martin Škára , Viet Bach Nguyen, and Vojtěch Svátek

Faculty of Informatics and Statistics,
University of Economics, Prague, Czech Republic
{skam06,nguv03,svatek}@vse.cz

Abstract. Google Knowledge Graph is more complicated than public knowledge graphs when it comes to retrieving and reusing data for e.g. building custom KGs or concept maps. This is because even though there is a dedicated graph search API, it only offers information about individual entities without links to other relevant resources that are only available in Google search knowledge panels. In this paper, we present Knowledge Graph Viewer, a tool that utilizes both the graph search API and knowledge panels in order to obtain relationships between graph entities to reconstruct custom substructures of the large knowledge base. This demo will showcase its usage and functionalities in each step of the KG creation while explaining the concept behind the workflow.

Keywords: Google knowledge graph · knowledge panel · knowledge graph · data relationship · concept map

1 Motivation

Google Knowledge Graph (GKG) is used by Google to store information about entities and their relationships, which can then serve, among others, to enrich the web search results. As with other large knowledge graphs (KGs), a question arises if arbitrary fragments of it can be retrieved to create custom KGs for a specific use. Public KGs such as Wikidata or DBpedia[1] offer flexible open data access via RDF dumps or SPARQL endpoints, on which tools such as concept mappers[2] can leverage. However, accessing graph-shaped information from GKG is more difficult, as the GKG API only provides information about individual entities (nodes) and not about relationships (edges). As observed by Barrasa [1], the information on relationships can be found as part of *text search* results, in the form of *knowledge panels*[3] which contain links to related entities or information

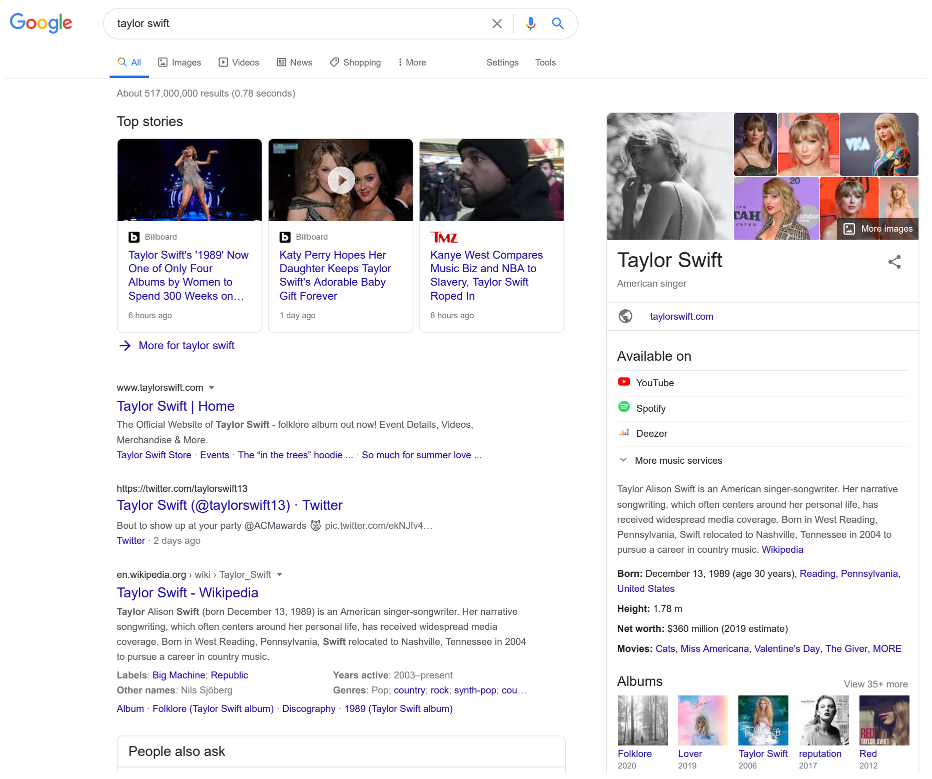
* Supported by IGA VŠE project № 33/2019
Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[1] https://www.wikidata.org/, https://dbpedia.org/

[2] E.g. ContextMinds – https://www.contextminds.com/, which dynamically autocompletes a manually built concept graph using data from DBpedia

[3] https://www.blog.google/products/search/about-knowledge-graph-and-knowledge-panels/

(e.g. birthplace, book author, works of a musician, categories, etc.), including the entities often sought together via Google search. An example search result with the knowledge panel can be seen in Fig. 1 on the right side. Using these pieces of information, we can progressively build meaningful graph fragments. However, manually copying relationship information from knowledge panels to connect nodes in a custom KG is slow and error-prone. To make this process easier, we developed the Knowledge Graph Viewer tool that can create custom KGs based on information from Google search knowledge panels and GKG API.



About 517,000,000 results (0.78 seconds)

Top stories

Billboard
Taylor Swift's '1989' Now One of Only Four Albums by Women to Spend 300 Weeks on…
6 hours ago

Billboard
Katy Perry Hopes Her Daughter Keeps Taylor Swift's Adorable Baby Gift Forever
1 day ago

TMZ
Kanye West Compares Music Biz and NBA to Slavery, Taylor Swift Roped In
8 hours ago

More for taylor swift

www.taylorswift.com
Taylor Swift | Home
The Official Website of **Taylor Swift** - folklore album out now! Event Details, Videos, Merchandise & More.
Taylor Swift Store · Events · The “in the trees” hoodie ... · So much for summer love ...

https://twitter.com/taylorswift13
Taylor Swift (@taylorswift13) · Twitter
Bout to show up at your party @ACMawards 😸 pic.twitter.com/ekNJfv4…
Twitter · 2 days ago

en.wikipedia.org › wiki › Taylor_Swift
Taylor Swift - Wikipedia
Taylor Alison **Swift** (born December 13, 1989) is an American singer-songwriter. Her narrative songwriting, which often centers around her personal life, has received widespread media coverage. Born in West Reading, Pennsylvania, **Swift** relocated to Nashville, Tennessee in 2004 to pursue a career in country music.
Labels: Big Machine; Republic
Other names: Nils Sjöberg
Years active: 2003–present
Genres: Pop; country; rock; synth-pop; cou…
Album · Folklore (Taylor Swift album) · Discography · 1989 (Taylor Swift album)

People also ask

More images

Taylor Swift
American singer

taylorswift.com

Available on
YouTube
Spotify
Deezer
More music services

Taylor Alison Swift is an American singer-songwriter. Her narrative songwriting, which often centers around her personal life, has received widespread media coverage. Born in West Reading, Pennsylvania, Swift relocated to Nashville, Tennessee in 2004 to pursue a career in country music. Wikipedia

Born: December 13, 1989 (age 30 years), Reading, Pennsylvania, United States
Height: 1.78 m
Net worth: $360 million (2019 estimate)
Movies: Cats, Miss Americana, Valentine's Day, The Giver, MORE

Albums
View 35+ more
Folklore 2020
Lover 2019
Taylor Swift 2006
reputation 2017
Red 2012

Fig. 1. Example of knowledge panel from Google search

2 Knowledge Graph Viewer tool

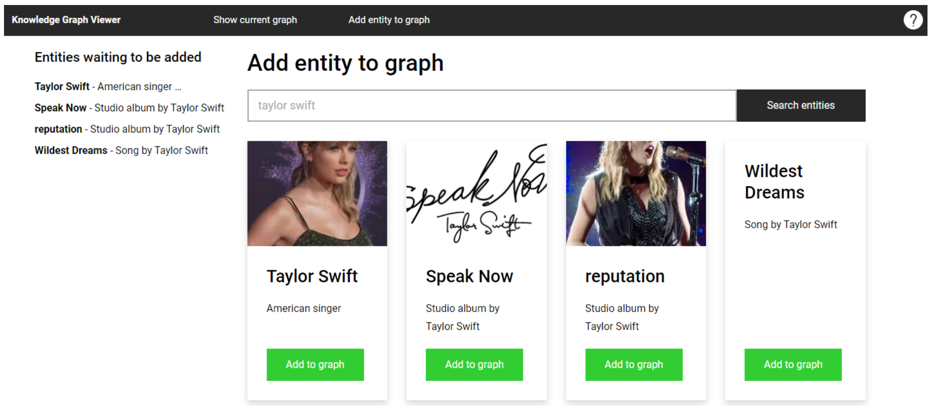
Knowledge Graph Viewer is a web application that uses JavaScript on both the back-end (in the form of NodeJS) and front-end side, where the Nuxt.js framework (respectively Vue.js) is used, with data being stored in a Neo4j database. The application can: 1) add GKG entities, along with their relationships to a custom graph; 2) display the custom graph; and 3) export the graph and the underlying part of an RDFS ontology to a RDF, CSV or JSON file.

The source code is available at `https://github.com/skaryys/kgviewer` under the MIT license. A demo version is running at `http://89.221.219.40/`, and a video demonstration is available at `https://vimeo.com/439597230`.

In accordance with GKG usage terms[4], we believe users can view the same data that is also publicly available from Google search results. Most of this data is obtained using Puppeteer[5], a headless internet browser capable of retrieving data from web pages. Knowledge Graph Viewer should only be used for academic purposes. The tool has been tested in major web browsers, such as Chrome 29+, Firefox 77+, Opera 77+ and Edge 84+.

2.1 Adding GKG entities to the custom graph

After entering a search term to the input field, the result entities are displayed. Clicking on the *Add to graph* button then adds the entity into a queue, which can be seen on the left side of the page. Entities in the queue are then asynchronously added to our graph (see Fig. 2). A summary of this process is shown in Fig. 3.



Entities waiting to be added

Taylor Swift - American singer ...
Speak Now - Studio album by Taylor Swift
reputation - Studio album by Taylor Swift
Wildest Dreams - Song by Taylor Swift

Add entity to graph

taylor swift

Taylor Swift — American singer

Speak Now — Studio album by Taylor Swift

reputation — Studio album by Taylor Swift

Wildest Dreams — Song by Taylor Swift

Fig. 2. Searching and adding entities to the graph

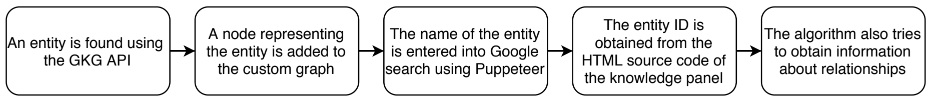

An entity is found using the GKG API
A node representing the entity is added to the custom graph
The name of the entity is entered into Google search using Puppeteer
The entity ID is obtained from the HTML source code of the knowledge panel
The algorithm also tries to obtain information about relationships

Fig. 3. Steps to retrieve data to be added to the custom graph

If the entity in the knowledge panel does not match the entity obtained using the GKG API, or if there is no knowledge panel at all, only one node corresponding to the original entity is to be added to the custom graph.

[4] `https://developers.google.com/knowledge-graph/terms`
[5] `https://pptr.dev/`

Two ways of retrieving *related entities* from the knowledge panel are distinguished. For some links, the related entity ID can be obtained directly from the source code of the primary knowledge panel (see Fig. 4). However, for most links in the knowledge panels, the relevant entity IDs cannot be obtained directly from the source code of the original entity being acquired. For this reason, Puppeteer is used to open the search result page of the related entity. The process of adding a related entity is shown in Fig. 5.

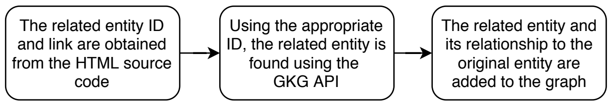

The related entity ID and link are obtained from the HTML source code
Using the appropriate ID, the related entity is found using the GKG API
The related entity and its relationship to the original entity are added to the graph

Fig. 4. Retrieving related entities directly from source code

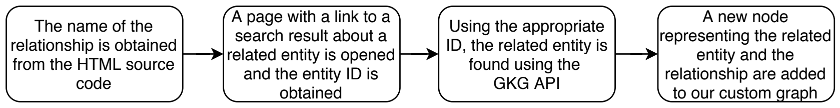

The name of the relationship is obtained from the HTML source code
A page with a link to a search result about a related entity is opened and the entity ID is obtained
Using the appropriate ID, the related entity is found using the GKG API
A new node representing the related entity and the relationship are added to our custom graph

Fig. 5. Retrieving related entities from search result of related entity

From the overall procedures that our tool uses, it is evident that adding an entity to the custom graph along with related entities and relationships is a complex task that often requires opening several web pages. This is why some entities may take longer than others. In average, the whole process takes around one to two minutes.

2.2 Displaying the custom graph

In Fig. 6, we show a custom graph built around the entity *Prague* and primarily containing information about important persons associated with the city. The graph visualization is based on Neo4j's native Neovis.js library[6]. By default, the tool shows up to 300 graph elements (nodes and relationships between them) to avoid slow loading, but this value can be adjusted manually. The resulting graph can also be searched and customized as follows:

- Detailed information about entities is displayed on mouse hover.
- Entities can be searched *by name* as well as *by type*.
- Entities and relationships can be queried using the Cypher[7] query language.
- Users can specify which entity *types* should be displayed as additional nodes.

[6] `https://github.com/neo4j-contrib/neovis.js/`
[7] `https://neo4j.com/developer/cypher-query-language/`

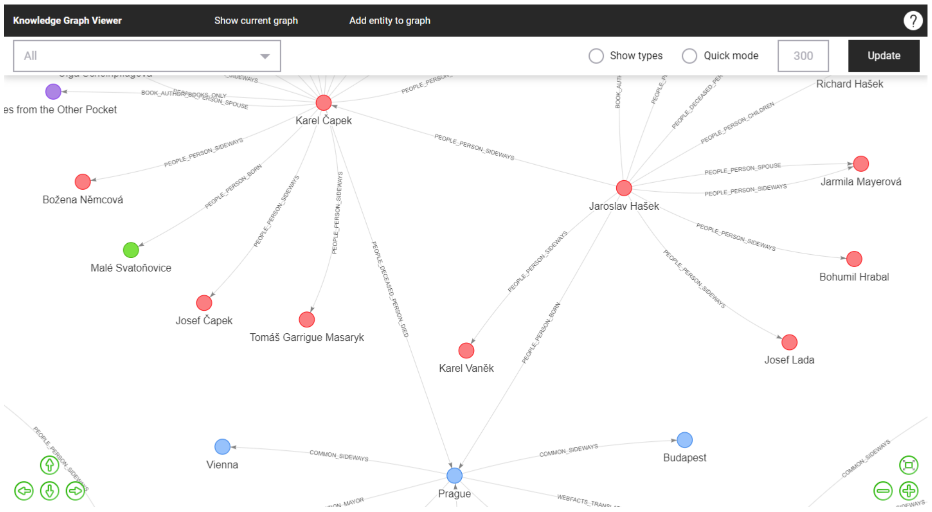


Fig. 6. Displaying a custom graph related to the city of Prague

– Users can use 'fast mode' which decreases rendering quality to improve loading speed. For graphs with many edges (e.g. if types are shown as nodes), this fast mode can reduce loading speed significantly.
– Users can export results in RDF format.

3 Conclusions

To our knowledge, Knowledge Graph Viewer is the only tool supporting the custom reconstruction of GKG fragments. Its usability is currently limited by two inherent factors: the slowness of entity addition due to multiple calls to the GKG API and Google search, and the necessity to adapt the wrapper interface when the structure of the Google search results changes. Nevertheless, we believe that it provides an added value compared to purely manual GKG reconstruction. As future work, we are preparing to explore the possibility of adding entities in an automated way, based on a user-set profile. The tool itself can still be improved, i.e. not all related links must be browsed because there is direct connection between Google API identifiers and the `stick` parameter in Google search[8].

References

1. Barrasa, J.: The 'Hidden' Connections In Google's Knowledge Graph. https://jbarrasa.com/2016/09/12/the-hidden-connections-in-googles-knowledge-graph/.

[8] https://lists.w3.org/Archives/Public/semantic-web/2012Jun/0028.html