Semantic Web Search: Google Knowledge Graph

Ganeshram Iyer
School of Informatics and Computing
iyerg@iu.edu

Shruthi Ramakrishnan
School of Informatics and Computing
shrurama@iu.edu

ABSTRACT

The knowledge graph constructed by Google, understands real-world entities and their relationships to one another. Entities covered by the Knowledge Graph include landmarks, celebrities, cities, sports teams, buildings, movies, celestial objects, works of art, and more. In a true Linked Data fashion, we interlink detected concepts in micro posts with Freebase entities, and evaluate our approach for both relevancy and usefulness. The extension is freely available; we invite the reader to reconstruct the examples of this paper to see how real-time opinions may have changed since time of writing. The graph enhances Google search in three main ways: by disambiguation of search queries, by search log-based summarization of key facts, and by explorative search suggestions. With this paper we demonstrate the power of Google Knowledge Graph API in enabling us to fetch detailed results collected from various sources. Though the Knowledge Graph already consists of relationship among the entities, for the version we are using it does not give them explicitly. For future work relationship can be constructed on the go based on the results we obtain for a certain query.

1. INTRODUCTION

Google Semantic Search has taken a completely different approach as compared to traditional means of search. It moves away from word statistics and link analysis to incorporate formal knowledge into search. The power of this technique can be seen from the following example. Using the previous search techniques, take a query like [taj mahal]. For more than four decades, search has essentially been about matching keywords to queries. To a search engine the words [taj mahal] have been just that—two words. But we all know that [taj mahal] has a much richer meaning. You might think of one of the world's most beautiful monuments, or a Grammy Award-winning musician, or possibly even a casino in Atlantic City, NJ. Or, depending on when you last ate, the nearest Indian restaurant. This is where formal knowledge graph that understands real-world entities and their relationships to one another: things, not strings. The Knowledge Graph enables you to search for things, people or places that Google knows about—landmarks, celebrities, cities, sports teams, buildings, geographical features, movies, celestial objects, works of art and more—and instantly get information that's relevant to your query. This is a critical first step towards building the next generation of search, which taps into the collective intelligence of the web and understands the world a bit more like people do.

Keywords are still very important, so we must keep looking for those that offer us more benefit. With semantic search, we need to know how to find these profitable keywords that also benefit the context of our website. With the introduction of the Knowledge Graph, the search engine Google has made a significant paradigm shift towards "things, not strings", as a post on the official Google blog states. With the graph, Google now publicly acknowledges efforts in the direction of understanding the difference between queries like "giants" for the football team (New York Giants) and "giants" for the baseball team (San Francisco Giants). Users now get word sense disambiguation support that allows them to steer the search engine in the right direction. Google News is a news aggregator portal provided and operated by Google.

2. RELATED WORK

All the leading search engines are investing time and effort to make progress in the field of Semantic Search. The benefits of using semantic search are for all to see. Search companies are thus investing in information extraction and data fusion, as well as more and more advanced question-answering capabilities on top of the collected information. The need for these technologies is only increasing with mobile search, where providing results as ten blue links leads to a very poor user experience. At Yahoo Labs the work is in advancing the sciences that underlie these approaches, i.e. Natural Language Processing, Information Retrieval and the Semantic Web. Currently, the current modes, above, cause unfair coverage bias since Schema and HTML scrapers will not cover all websites. Schema is only used by those who know what it is and who can code whereas HTML scrapers will be biased towards popular websites just as most of the engines' top ranked results are crawled more frequently than less popular websites. Semantics helps to overcome this and give the users a much more engaging platform where the queries are not based on how popular the link is but on how relevant it is to the topic. The internet has changed how our brains obtain and store information. We search to retrieve information rather than store it. This is why search is a popular online activity. Engines want to maintain happy, returning users. Otherwise they will miss out on all that paid search activity which keeps them afloat. Knowledge Graph and Bases simply retain the searcher for as long as possible and vastly diminishes us from visiting other websites allowing us to subconsciously feel that the engine itself is more trustworthy, which reinforces search loyalty.

All the major search engines are taking steps in the right direction. As already mentioned Google has been creating an entity network within images since 2006 by naming them with numeric values as supposed to text strings, but all you need to do is play around with a search translation tool to see how rusty this is.

Microsoft acquired Powerset last July in a reported $100m deal, and after a conspicuous Tweet from Powerset co-founder Barney Pell, many assumed that the semantic search outfit would play a major role in Redmond's latest attempt to catch the uncatchable Google. Microsoft's Powerset division, the division has tweaked Microsoft's primary search engine in certain "subtle" ways. But its

main contribution is a secondary engine that searches nothing but Wikipedia. In essence, Microsoft's has taken Powerset's existing Wikitool and latched it to the Bing torso.
"The Powerset division has contributed to Bing in both subtle and more conspicuous ways. While the subtle contributions are important, they are much harder to showcase. This post will focus on how the features that our users have come to love on Powerset.com have evolved and have been integrated into Bing," Prevost says, before detailing Bing's "Reference" tab.
As we reported yesterday, the Reference tab reproduces Wikipedia articles in their entirety. When you search on, say, Albert Einstein, the tab will appear on the left hand side of the page, and if you click on it, you're taken to a reproduction of Einstein's Wikipedia entry (licensed at no cost from the "free encyclopedia anyone can edit"). But from that Reference tab you can also tap into Powerset's semantic Wikisearch, which the company originally unfurled in May of last year, before the Microsoft acquisition. This vertical search engine is designed to accept natural-language queries, such as "Was Einstein married?" - though that's not immediately obvious from Bing's layout. In a video attached to Prevost's blog post, Powerset founder Lorenzo Thione acknowledge that some of Bing's Powerset tools are "a little bit hidden. Over time, we'll definitely work on making it more accessible and visible to users."

3. IMPLEMENTATION DETAILS

The Google Knowledge Graph Search API will let you query the Knowledge Graph database for various information on the entities within the Knowledge Graph. This is to ultimately replace the Freebase API and the Freebase Suggest Widget. Google said they will release a replacement for the Freebase Suggest Widget and discontinue the Freebase Suggest Widget a few months later. This is a read-only API, meaning you cannot submit changes to mistakes within the Google Knowledge Graph within this API. It is only for querying and pulling data from Google, not sending data back to Google.

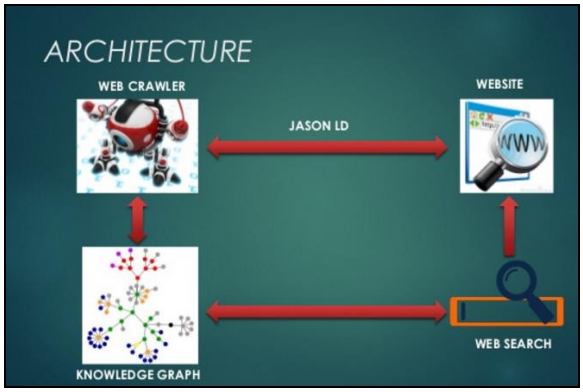

Fig: 1 Architecture of Google Knowledge Graph

The architecture given on Fig 1 explains how the knowledge graph is constructed. Web crawler scans through all the websites using advanced parsing techniques to establish relationship between different entities in the graph. Each new fact found regarding that entity gets linked to it. This way we have the latest news related to that entity. This enables Google to give a summarized view of the person or thing the user is searching about. The overall user experience is enhanced significantly when instead of seeing links to different pages related to the topic, the user can see the key information on the search page itself without visiting any other link.

The Knowledge Graph is a system for organizing and merging information about millions of well-known people, places, and organizations from many data sources, including Wikidata. When this was introduced in 2012, the Knowledge Graph resulted in simple information cards like the contemporary one below, featuring links and other info pulled in from Wikipedia.

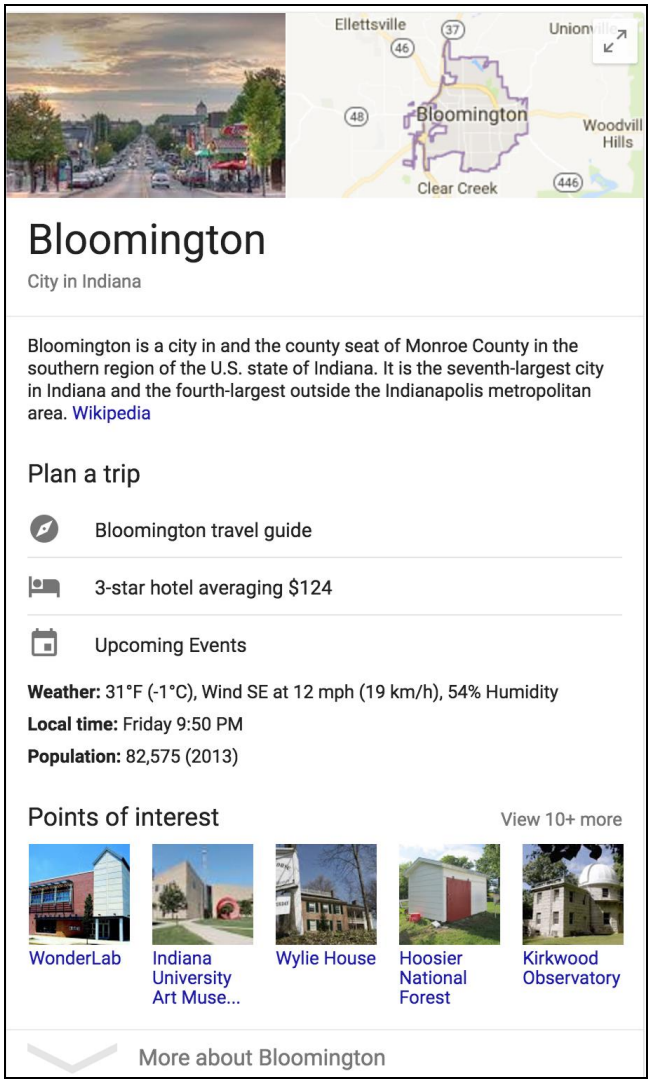

Fig: 2 Semantic Search Result from Google

The desire to make the internet a better, more useful, user-focused place. It's what keeps me writing about this stuff and it's at the heart of most successful digital transformation endeavors. Of course the rest of us can argue that the better Google can understand its users, and the quicker it can deliver relevant accurate content, the longer users will spend on Google. Therefore, increasing the chance that a user will click on an ad. Individual publishers can help Google find and deliver your content by marking it up as accurately as possible using semantic HTML, just bear in mind that some of these techniques may make your results more attractive, but are also designed to keep the user

on Google for as long as possible. The more questions Google can answer itself, using your own content, the less users will actually interact with other sites.

In our project we display all the details return for our search results from the Google Knowledge Graph API. These are the following contents:-

1. Image: If the entity we are searching for has an image the search result returns an image.
2. ID: This is the unique identifier associated with the entity in the graph
3. Short Description: A short description which summarizes what the entity represents
4. Long Description: Longer format of the description with a more detailed information of the entity
5. Article Link: This is basically the link from which these descriptions were obtained
6. Categories: The types under which the entity is classified. We can get to know the ontology of the graph with the categories.

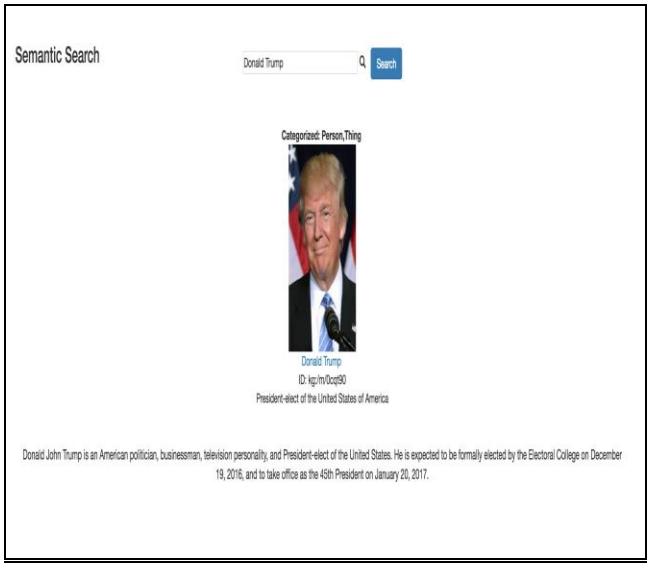

Fig:3 A Screenshot of our project

As you can see in figure three the search result returned for the query “Donald Trump”. The information is up to date reflecting the fact that Donald Trump is the President-Elect of United States.

The underlying architecture of our project uses Flask Framework to provide web services. AJAX is used to fetch the results for the web page and render them. For formatting the web page and displaying the results in a well-defined format we have used Bootstrap and JQuery along with CSS3.

4. FUTURE WORK

As a part of future, one of the major things that we can do is to fetch results from various other such API’s and aggregate the results to give a much more comprehensive result. Based on categories that we get from Google Graph we will parse the most relevant layer. For example, Donald Trump is categories under Thing and Person. We should query other API’s which give us more information for that particular category.

Today, the Semantic Web sites as an extension to the world wide web, providing a standardization of the way the relationships between web pages are expressed. The World Wide Web Consortium (W3C), is dedicated to building the technology stack to support a ‘web of data’. According to the W3C: *"The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries"*.

So how far have we come in achieving this utopian vision? The truth is that there are many current Semantic Web projects and many useful applications that have already been developed. However, to create a true Semantic Web then all documents and data within the web must be compiled with Semantic Web technologies in mind, to enable a semantic search engine to have a true picture of all of the information available. Without the full picture, the full story cannot be told.

Of course, there has been progress. There has been steady movement towards organizations and governments making their existing data available in machine readable formats as 'Linked Open Data' and 'Open Government Data'. Although attempts at building a Semantic Web in the past were more like projects in silo and based around creating competitive advantage. Back in 2010, the BBC’s World Cup site used semantic databases to enrich the user experience. This made pages easy to navigate and data on each football player extremely easy to extrapolate. Fast forward to the present and we see many examples of industry leaders talking about the future of marketing being semantic. Companies need to start thinking now about how to structure their data to sit within their web pages. Then, as search technologies develop more sophisticated algorithms, organizations can be ready with rich layers of data and information that can be utilized by the search engines, to provide rich insights into consumer behavior. Remember, data does not just sit within a company’s main site, the semantic technologies, social, SEO and SEM data strategies need to sit side by side.

Look to the future and the possibilities for the ‘Internet of Things’ underpinned by vast quantities of linked open data, provide endless commercial opportunities. Wearable fitness technologies combining data with your weekly grocery shop, could for instance provide the service of a personal nutritionist.

In the wider context, the opportunities are far greater than the individual or commercial gains that the Semantic Web and linked open data can provide. Many governments are now part of the Open Government Data (OGD) movement and are working towards this future, publishing data on economic activity, energy consumption, public spending and other key metrics. Using Semantic Web technologies to understand vast quantities of data, it is becoming possible to efficiently leverage this information to enable better decision making in areas such as low carbon development, renewable energy, disaster management and agricultural policy.

5. ACKNOWLEDGMENTS

We thank Prof. Ying Ding for constructive suggestions that helped us improve this project and guiding us throughout the semester.

6. REFERENCES

[1] Introducing the Knowledge Graph: things, not strings. https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html

[2] http://wikipediocracy.com/2014/01/06/googles-knowledge-graph-killing-wikipedia/

[3] Building The LinkedIn Knowledge Graph. https://engineering.linkedin.com/blog/2016/10/building-the-linkedin-knowledge-graph

[4] http://www.freelanceseoessex.co.uk/an-introduction-to-googles-knowledge-graph/