CS 448B Final Project Proposal

By

Fyza Parviz

Visualizing History of Ideas

Visualization Problem:

Currently, on Wikipedia, there are "40 million articles in 293 languages with over 32 Million users"[1]. This massive information can only be accessed statically, on a document by document basis. There is no way to see the hierarchy of ideas or how the ideas listed in these documents are connected.

It is also difficult to navigate Wikipedia pages unless one is explicitly searching for a particular topic. Therefore it is unlikely that a researcher while using Wikipedia can come across a startling insight.

Wikipedia organization states that its goal for the next stage of development is to focus on improving the quality of its product: "Wikipedia is still in need of much expansion and improvement. Many of the articles are of poor quality, and some mainstream encyclopedia topics are not covered adequately. Over time the balance of the editorial effort is expected to slowly tilt towards a greater emphasis on increasing the quality, scope, classification, and interlinkage of existing articles."[2] It is difficult to notice errors, misinformation, and duplications on static documents. Hence Wikipedia organization needs a new mechanism to view content to achieve its milestone.

WikiPedia Visualizations:

Following are some of the attempts at visualizing Wikipedia Data:

https://angryloki.github.io/wikidata-graph-builder/
http://www.chrisharrison.net/index.php/Visualizations/WikiViz
http://wiki.polyfra.me/#
http://wiki.c2.com/?VisualizeTheWiki
http://seealso.org

Although these visualizations are helpful in linking Wikipedia data, but they are mostly Network graphs that are not user friendly and are hard to navigate. They are also not interactive and do not provide the user the control to morph the data to their needs. They do not provide editing capabilities to Wikipedia editors.

---

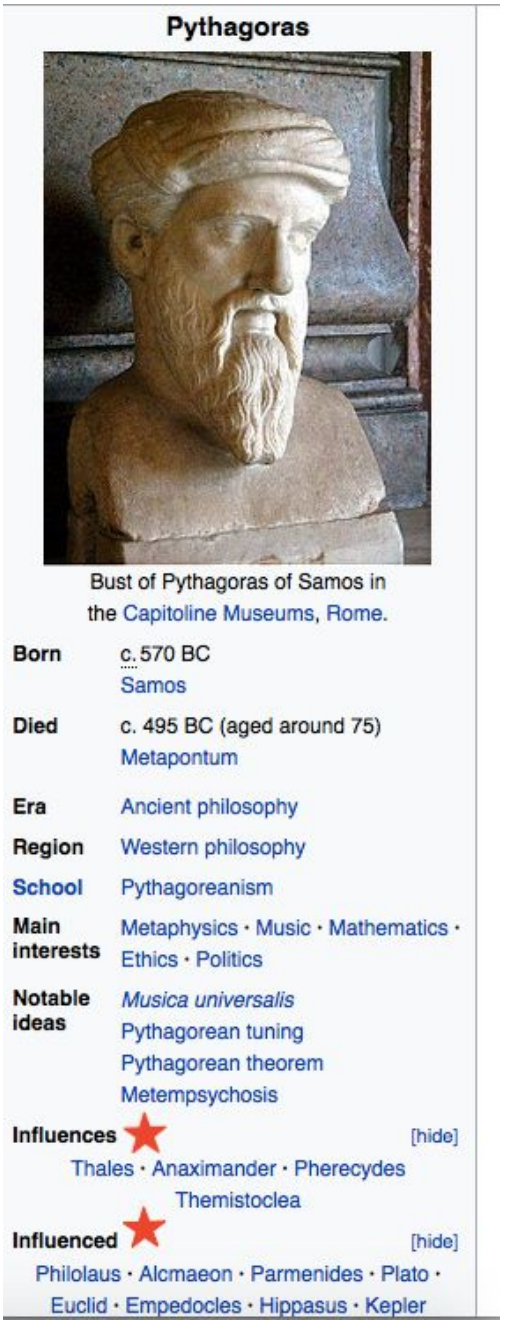
[1] https://en.wikipedia.org/wiki/Wikipedia

[2] Ibid

How will I address the Problem:

According to Ben Shneiderman "the purpose of visualization is insight"[3]. I want to develop a new way to Visualize Wikipedia Data other than Network Graphs that will assist in navigating Wikipedia and make the data meaningful to both a researcher and a Wiki editor.

For this particular project, I will look at the "Influence" and "Influenced" data in Wikipedia Articles and create a new kind of link visualization. I will extract this information by using SPARQL query in dbpedia.org. For example, a Wikipedia page of Pythagoras, includes a brief historical section located on the rightmost side of the webpage. Two elements included in this section are: Influences and Influenced. I have marked them with a red star in the image on the right.



A Wikipedia user would have to go back and forth between Wiki pages in order to get a full picture of the influences of a particular figure. For Example, if Pythagoras influenced Kepler, then who else was influenced by Kepler or who else influenced Kepler. This problem can be solved via a visual representation of the data.

I will be looking to create a new kind of visualization to create an interactive, searchable, and editable app that will be helpful for researchers and also Wikipedia editors to easily navigate and edit Wikipedia.

The scope of this project could then be expanded to all of Wikipedia Data.

[3] https://medium.com/@mbostock/a-better-way-to-code-2b1d2876a3a0