**Proposal  : Wikidata PageRank**

**Problem:**

PageRank is an algorithm used by Google Search to rank websites in their search engine results. In the project, we will implement PageRank based on Wikidata.

**Dataset:**

DBpedia version 2016-04 (<http://people.aifb.kit.edu/ath/>)

* input file: merge of en, de, fr, ja, sv, it, ... datasets
* PageRank variant: unnormalized
* PageRank damping factor: 0.85
* PageRank no of runs: 80 (converged)
* PageRank start value: 0.1

**Proposed Solution and Real world Application:**

In this project, we will first create a PR matrix based on the references and correlations among websites. The elements in the PR matrix represent scores of websites based on those references. We will then visualize the correlations among websites by creating a figure which contains dots connected by edges. Each dot will represent a website and the size of a dot will indicate the the score of that website. In addition, an edge between two dots means a correlation exists between two websites. PageRank has applications in search, browsing,  track estimation and so forth.

**Project steps:**

|  |  |  |
| --- | --- | --- |
| Step | Estimated completion time | Person(s) in charge |
| 1. Extracting and cleaning up data | One week | Ji Suk Kim |
| 2. PR matrix | Two weeks | Zhichen Zhang, Yonghong Liu |
| 3. Data visualization (to visualize relations among websites) | One week | Qinghong Zeng |

**Github repo:** https://github.com/ece143-fa18-group10/Wikidata\_PageRank