Topics in Social Computing

Project Proposal
Knowledge Building Analysis Toolkit

Mentor: Amit Kumar Verma

Submitted by: Paras Kumar

Introduction

With the success of crowdsourced portals like Wikipedia, Stack Overflow, Quora, GitHub, etc., a class of researchers is driven towards understanding the dynamics of knowledge building on these portals. A programming library (KDAP) has been proposed which provides high-level operations for the analysis of knowledge data. This project focuses on the analysis of the Wikipedia dataset.

Wikipedia is a platform to share knowledge where people across the globe collaborate to write an article. Like any collaboration process, articles on Wikipedia contain a lot of information apart from just what is present in the article itself and we can compare different revisions of a Wikipedia article to find out various interesting patterns in the collaboration or knowledge building process by looking at the evolution of a Wikipedia article.

Aim

Given the amount of information we can extract from Wikipedia about the knowledge building process, there are many researchers working on wiki data across the globe. In their research, they have to first do some common analysis on wiki data before they start working on their actual project. So, we are building a general-purpose Python library for doing some common analysis on Wikipedia which can be used by them in their research. This wiki analysis toolkit will be available as a PyPi package and then anyone can easily use this toolkit in their project.

Work done so far

I have implemented the research paper [Societal Controversies in Wikipedia Articles](#) to find out which topics are more controversial by extracting wiki links in each revision and finding how they evolved using a modified diff algorithm used in git and some mathematical formulas given in this paper. So, my first analysis is complete and the code is uploaded to Github. [Link to Github repository](#)

Future Tasks

I have to implement the research paper [Getting to the source: where does Wikipedia get its information from?](#) And then some other research papers for doing different kinds of analysis on wiki data. Finally, I'll upload this package to PyPi so that others can use it.