

Social Networks Course Project KML (Knowledge Markup Language)

Amit Kumar Verma (mentor)
Paras Kumar
Nitin Gandhi

Contents

- 1 What is KML?
- 2 Need for KML
- 3 Components
- 4 KML-Compressor-Decompressor
- 5 KML-for-Github
- 6 Github-Spider
- 7 User-Agent-Spider
- 8 Wiki-Satck-KML-Downloader

What is KML?

KML stands for Knowledge Markup Language, a standard format for storing the data of all the Knowledge Building Portals. Knowledge Building portals like Wikipedia, Stack Exchange, GitHub, e.t.c provides their data dump in their own formats. We are trying to propose a new standard format for all these types of portals such that the analysis is easy.

Need for KML

All these Knowledge Building portals provide their data dumps in their own format. For example, Wikipedia provides its data in an XML format with their own schema definition. Similarly, Stack Exchange provides its data dump in an XML format with different schema definition. The KML will be a new standard format for these kinds of Knowledge Building portals with a standard schema definition. The idea is to make KML flexible enough such that it can store the data of any kind of Knowledge Building portals.

Components

- KML-Compressor-Decompressor
- KML-for-Github
- Github-Spider
- User-Agent-Spider
- Wiki-Satck-KML-Downloader

Source Code: <https://github.com/csl-622/KML>

KML-Compressor-Decompressor

Compresses text section of KML and decompresses any asked revision of article.

KML-for-Github

Generates KML for github from scraped data in json format.

Github-Spider

Scrapes data from Github and stores as json files.

User-Agent-Spider

Scrapes many IP adressess to avoid blocking of scraper by Github.

Wiki-Satck-KML-Downloader

Downloads data from Wikipedia and Stackoverflow and then coverts it to KML.