

Geo Location Based Research Project Indexing Solution

by Harshit Yadav

Submission date: 20-Sep-2019 02:31PM (UTC+0800)

Submission ID: 1176392200

File name: RMTc_Report.docx.odt (28.48K)

Word count: 1590

Character count: 8231

Geo Location Based Research Project Indexing Solution

Abstract— Open Source Innovation has become an important part of Development and Research in many areas. Traditional open Source Projects and Research Indexing Databases works using Search keywords and Key Value pair match parameters for the published well Structured Data, a lot of Undocumented Research Literature, Partial Results and Datasets gets discarded which might contain some valuable insights which could be later picked later for further research and Development, Geo Tagging this data will help in faster discovery and better Indexing

Keywords— *Github; OpenSource; Geo Location; Indexing*

I. INTRODUCTION

Open Source Development and Innovation has been around since the year 2000 with the rise of use of web technologies and development in collaboration tools the process has accelerated many folds over the years. These days a variety of online Research tools and platforms are used for publishing, documentation for the research, checking for plagiarism and sharing the Datasets used in the Research. A lot of inconclusive results, datasets, partial results, model, design or codebase which cannot be structured or set into the traditional model of reporting and publishing the documentation gets discarded which contains many hours of research work which could be useful in the future to support some other research project in the same domain or the same project could be continued with fresh perspective and motivation.

Many Online open source tools and platforms exist for sharing the code or datasets but due to highly unstructured nature of the data and presence of other irrelevant data on this platform the chances of discovery of this data becomes very less. The addition of any new parameter of Geo Location Tag to this data highlighting the place of research, target field of study, or geographic area to which the data concerns with will help in better online discovery and reduce the redundant work in replicating or continuing that research or could be used as input in Industry Research and survey when looking for solution as Data is the new Gold driving the research and helping us discover new Insights and patterns and forming a crucial part in driving the present day innovation and Development.

II. LITERATURE REVIEW

A. Geo Tagging

Geo tag or Geo Tagging is the addition of Geographical Identification based on the coordinate system of Longitude and Latitude. Over time there have been many developments across many domain using the geographical identification mark along with text, video, audio, signals, payment gaming etc but it has not been utilized well in the domain of open source and academic research where

majority of the heat maps for geo location information source is being done by Mining the city or country name from the text and the plotting it which is not accurate and limited in scope based on the amount of data in sample space and datasource used. Geo Tagging has recently seen an increased trend as an identification attribute with improvements in networking technology and mobile devices which are constantly moving generating data and with advancement in Orientation sensors and spatial mapping techniques a new and more precise coordinate and identification tag is being trending in the Industry and being currently explored called the Spatial Anchors[1] combined with the Augmented Reality and Virtual Reality Interface which were not possible until now due to limitations in the Processing Hardware.

B. Open Source Platforms

Github, Gitlab are the leading code sharing platforms where majority of datasets and projects are stored for collaboration and sharing with over 100 Million public repositories [2], Github is the single most and largest collection of projects and codebase but due to minimal field constraint and flexibility in structure and keyword tagging mining and searching for the required relevant information or project makes it difficult and top results are ranked using basic ranking algorithms, considering the parameters such as the number of Pull, Star and Last Update etc. Many solutions and mining methods exist to generate specific stats such as the most popular framework or library in a region or country over a certain period of time out of a defined sample userspace due to limited API and bulk data access requests or the general stats released by the company annually. But analysing and filtering through a massive unstructured and diverse data still remains a challenging task with many ongoing developments to tap onto this data.

C. Proprietary Solutions

Many Proprietary Academic and Research Journals Indexing Solution exist that have database access to thousands of both open and closed publishing websites and individual University Private Networks for cataloging and searching through them with granular search parameters tuning options but majorly works using a combination of key value pairs match and ranking based upon number of views, citations and references. The Data searched through is highly structured, well formatted, reviewed text and visual data but many details and data collected and worked upon during the research collected by the researcher gets missed out, including many projects and findings that could not be completed or failed due to variety of reasons and doesn't get published but is still an important resource that could be used in the future.

III. METHODOLOGY

The Methodology adopted will be same as collecting any data and creating a geographical heat map or Visualization using the preprocessed data and providing an option to modify the search result providing functional similarity like that of a search engine .

Since the data set to be mined and indexed is of high value and copyrighted content belonging to the publishing journals and Institutions thus Publicly available Database would not be mined and scraped for collecting data rather than user Recorded entries and completely open source projects and code under MIT license will be collected for the scope of this research project and the Project will be made public on the web were the access and service will be provided to the user without any fees or access charges

For Broad Classification of the Sample space entries it will be widely classified based on the following parameters

1. Domain of Study
2. Geographical Area Studied or Concerned
3. Progress Status of the Project
4. Language Used
5. Future Plans for the Project
6. If its Application based or Purely Research based
7. Reason for Discontinuing Project and Bottlenecks
8. Technologies used
9. Time Spent on the Project

Based on the above given parameters a ranking and search filter will be implemented using a GUI interface which help users search and find the relevant results faster .

IV. SYSTEM DESIGN

The Software design to implement the research is simple web development based using the client server architecture where the central application will be available online via a URL publicly accessible on the Internet. Where a user can navigate up to the website site sign in using the credentials to verify itself and stop other web crawlers and Scraping bots to replicate our database .

The Database will be using Modular API approach where many of the individual components used will be accessed using API service from other providers as it is more reliable and efficient when used using the 3rd Party tools like Google Maps which is highly accurate, and faster in performance and economical rather than developing a custom solution such as Map data to tag the GPS data .

For Modeling the Database and backend code an open source framework and programming language should be used which should provide high flexibility in modeling without compromising on the performance and should be easy to deploy on the existing cloud solutions provided by various Vendors Online

For working with the Data Set were the large quantity of data is of a particular data type and the majority of the process and computation is to be done on that data type only thus choice of database software should be done accordingly comparing all the Database Software available and which provides better inbuilt libraries and methods to process and handle that kind of data which will reduce the developers time taken during the implementation and reduce the resource consumption in server as the default libraries and methods are generally faster and more efficiently implemented . Among various proven and tested database commercially available open Source we found that that PostgreSQL Database will be most suitable for the task as it its open source , already proven by many existing software solution implementations in the market regarding its reliability and provide many inbuilt custom libraries and methods to process the GPS coordinates and perform calculations such as radius Search and distance between two coordinates.

Finally for building the User Interface, Web browser based approach will be there since it provides easy and hassle free cross platform access across multiple range of devices which support internet access, Since all computation will be done on the servers the accessing user device need minimal configuration and no heavy rendering and processing making it perfect for academic purposes where the individual accessing machine and hardware might not be powerful or the user might not have access to install or run any script without the admin level permission .

V. FUTURE SCOPE

This project is completely novel and never been tried and implemented which can be used across a large domain of Fields with huge amount of user base and potential monetization sources to fund the project for the Expansion of this project to a commercially viable product in short time .

REFERENCES

Geo Location Based Research Project Indexing Solution

ORIGINALITY REPORT

1 %

SIMILARITY INDEX

1 %

INTERNET SOURCES

0 %

PUBLICATIONS

0 %

STUDENT PAPERS

PRIMARY SOURCES

1

semspub.epa.gov

Internet Source

1 %

Exclude quotes On

Exclude bibliography On

Exclude matches < 10 words