**A REPORT**

**ON**

**Graph Data Digest Document Format (GDF)**

BY

Vipin Baswan 2017A7PS0429P

Suyash Raj 2017A7PS0191P

Yashdeep Gupta 2017A7PS0114P

Abhinava Arsada 2017A7PS0028P

Sreyas Ravichandrana 2017A7PSP

AT

**Homi Bhabha Centre for Science Education (HBCSE)**

A Practice School-1 station of



Birla Institute of Technology & Science, Pilani

(June 2019)

**A REPORT**

**ON**

**Graph Data Digest Document Format (GDF)**

BY

Vipin Baswan 2017A7PS0429P Computer Science

Suyash Raj 2017A7PS0191P Computer Science

Yashdeep Gupta 2017A7PS0114P Computer Science

Abhinava Arsada 2017A7PS0028P Computer Science

Sreyas Ravichandrana 2017A7PSP Computer Science

Prepared in the partial fulfilment of the

Practice School-I (BITS F221)

AT

**Homi Bhabha Centre for Science Education (HBCSE)**

A Practice School-1 station of



Birla Institute of Technology & Science, Pilani

(June 2019)

**ACKNOWLEDGEMENTS**

We’d like to thank Dr K Subramaniam, Director, Homi Bhabha Centre for Science Education for providing this opportunity to us, the students of BITS Pilani to work with the organization towards the fulfilment of its purpose of promoting science education throughout the country. We’d also like to thank Nagarjuna G. to allow us the chance to work on this project under his able guidance. We also extend our thanks to Prof. Mukesh Kumar Rohil, our Practice School Coordinator for his constant support and guidance. Additionally, we express our heartfelt gratitude to Mr. J. B. Waghmare for his constant support at the PS station in non-academic matters.

Finally, our humblest apologies to others who helped us but whose names were not mentioned in the list.

**Birla Institute of Technology and Science**

**Pilani (Rajasthan)**

**Practice School Division**

**Station:** Homi Bhabha Centre for Science Education

**Duration:** 21 days

**Date of Start:** 21st May, 2019

**Date of Submission:** 10th June, 2019

**Title of Project:** Graph Data Digest Document Format (GDF)

**Submitted By:**

Vipin Baswan 2017A7PS0429P Computer Science

Suyash Raj 2017A7PS0191P Computer Science

Yashdeep Gupta 2017A7PS0114P Computer Science

Abhinava Arsada 2017A7PS0028P Computer Science

Sreyas Ravichandrana 2017A7PSP Computer Science

**Under:**

Nagarjuna G., Faculty at Homi Bhabha Centre for Science Education

**Submitted to:**

Prof. Mukesh Kumar Rohil, our PS1 faculty

**Key Words:** Graph Databases, RDF, GraphQL, Data Digest Format

**Project Areas:** Graph Databases

**Abstract:**

Graph databases have always been a promising tool in increasing the querying efficiency on datasets. Hence, the prospect of data digest document format such as GDF seems very promising in today’s world where datasets interact in a complex manner and quick information retrieval is of prime import.

Our project deals with developing a format called GDF and the method to convert any document format into GDF. This will assist us in quick merging of different files as graphs can be merged easily.

**INTRODUCTION**

The aim of our project is to develop a Data Digest Format which can be used to convert and represent information from any format. The format is graph based, hence the name.

The scope of our project is to:

1. Decide the format of GDF
2. Write methods to convert a text file into GDF
3. Create meta-data in the GDF format from the meta-data in the text format
4. Develop a query language (based on SPARQL) for information retrieval

Currently, we have completed aims (a) and (b). We will use the 7-column format for GDF. The relevant details will be provided in the main text.

We are currently working on (c) and (d). We hope to complete objective (c) by 15th May, 2019. Also, some of our members are also concurrently working on objective (d).

Since the idea of GDF is pretty innovative and unique in itself (Credit: Nagarjuna G.), not much literature is available to us for this exact format. But a similar data format called Resource Description Format (RDF) already exists. It is also graph based data format. Hence, we have gone through the literatures regarding RDF (links of the online resources have been given in the References section). Also, we have generated the sample data (for testing our code) ourselves using a python code but later we will collect more data from DBPedia (an online platform to get 3-column formatted data for various Wikipedia pages). Also, the query language for our format is basically SPARQL. We are referring to the official literature available on SPARQL (link for the same has been provided in References Section) for building our querying engine.

We wanted to limit the dependency of our code on various platforms. Hence, we have used BASH Scripting to write our code.

Due to time constraints, we will not be able to create our own renderer but we will be using already existing \_\_\_\_\_\_\_\_\_\_\_ renderer.

This report gives an outline of our project and the path we are expected to follow to complete the project on time. We also explain whatever we have completed till now and also introduce what we will be doing in the coming days. This report is like a comprehensive summary of our whole project.

**MAIN TEXT**

