Bharathi–An Applied Semantic Intelligence Use Case for Public Data in India

Asha Subramanian[a], Manikanta Vikkurthi[a], Gunjan Pattnayak[a], Akshay K S[a] and Harika Vikkurthi[a]

[a]*Semantic Web India, Bangalore, Karnataka, India. Home Page: http://www.semanticwebindia.com/*

Abstract

Data sets published for public consumption related to governance typically contain common metadata. This metadata normally describes the publishing organisation, the domain, the administrative region and a topic that the data best describes. A relevant semantic vocabulary will not only facilitate linking the metadata through meaningful contexts but also enable grounding related information from diverse public data sets using semantic entities defining the metadata. In this paper, we present Bharathi –Linked Data Vocabulary for the Indian context. Bharathi contains information regarding government organisations at the union & state government level, administrative regions, sectors, sub sectors and common topics used frequently in the vocabulary of the government functions. Further, Bharathi contains links to other open vocabularies such as GeoNames for geographical locations. The schema of Bharathi uses existing established ontologies making it inter-operable and extensible. We describe Bharathi along with a live use case of its application in accomplishing a semantic Covid19 data analysis for India.

Keywords

Linked Data, Knowledge Graph, Ontology, Semantic Intelligence

1. Introduction

Metadata related to publicly accessible data published on open government data portals and others such as data.gov.in[1], Rainfall Statistics[2], Covid19 related datasets[3] etc. often tend to repeat themselves across multiple datasets from disparate sources. This metadata contains valuable information relating to pivotal entities such as the publishing government organisation along with its classification (e.g. Ministry of Health and Family Welfare –Union Government, Department of Health and Family Welfare –State Government of Karnataka), sector (e.g. Environment, Health, Education), the administrative region for which the data has been collected (e.g. State –Karnataka, District –Bangalore Urban, City –Bangalore) etc. The metadata is available in un-structured and semi-structured formats in various home pages of government of India web sites such as GOI Directory[4], Census India[5] limiting the

International Semantic Intelligence Conference (ISIC 2021), New Delhi, India: February 25-27, 2021

asha@semanticwebindia.com (A. Subramanian); manikanta@semanticwebindia.com (M. Vikkurthi); gunjan@semanticwebindia.com (G. Pattnayak); akshay@semanticwebindia.com (A.K. S); harika@semanticwebindia.com (H. Vikkurthi)

[1]Open Government Data (OGD) Platform India: https://data.gov.in/

[2]Indian Meteorological Department: https://mausam.imd.gov.in/imd_latest/contents/rainfall_statistics.php

[3]Crowdsourced Covid19 data: https://www.covid19india.org/

[4]https://www.goidirectory.nic.in/ –Indian Government Organisations and their classifications at the Central Government and State Government level

[5]http://censusindia.gov.in/DigitalLibrary/2011CodeDirectory.aspx –State-wise Rural and Urban administrative units location codes maintained by

discovery of inter-connected relationships across these pivotal entities by humans and machines. In many cases, the same entity (a government organisation or an administrative region) is referred by its respective code across different web sites making it a humongous task to connect the information across the two data sources although they relate to the same entity. Therefore, creating an appropriate semantic framework for a knowledge based representation for this metadata is paramount. Such a semantically enriched knowledge representation will substantially help in linking public data sets and also provide a valuable inter-operable, reproducible and machine-readable resource for the scientific and industrial community. To the best of our knowledge, no reasonably large and up-to-date corpus of linked data information about government entities, their classifications and their inter relationships specifically for the Indian context has been made publicly available yet, in the form of a machine readable resource. This paper presents a use case for applied semantic intelligence – Bharathi and its potential applications.

Bharathi is populated using a semi-automated pipeline that periodically harvests data from various government of India websites, annotating the entities and lifting the semantics into a RDF[6] grounded by a sound model based on established vocabularies such as Organisation Ontology[7], SKOS Vocabulary[8] and the GeoNames Knowledge base[9]. We provide a web interface accessible at http://semanticwebindia.in/bharathi for exploring the underlying knowledge graph of Bharathi, enabling users from outside the scientific community to find, retrieve and sample information from Bharathi.

We start with the canvas of *Related Work* (Section 2) where we list references to similar work and set the context for our contribution. Next, we provide an overview of Bharathi followed by an introduction to the Bharathi ontology in Section 4. Section 5 details the process of generation of Bharathi. Real world applications of Bharathi are presented in Section 6. We conclude with a brief summary and thoughts on future directions in Section 7.

2. Related Work

Several research efforts such as [1, 2, 3, 4, 5, 6, 7, 8] have shown interest in linking public data with semantic metadata to enable intuitive linking of information from diverse sources. The research efforts specifically to model the metadata around public data can be found in [9] where the authors presents Gridworks-based data workbench application to help in converting open government data to linked data by using standard Data Catalog Vocabulary. The data catalogs of open data are an input to the application, yielding the linked data form. [10] presents a semantic government vocabulary to create annotations of open government data. It details the different layers of the vocabulary depending on different ontological terms found in open government datasets. It also demonstrates how to use the vocabulary to annotate open government data at different levels. [11] explores the documentation needs for open government data with specific focus on metadata interoperability issues. It presents two methodologies for interoperability after studying a variety of open government metadata, globally recognized standards and guidelines. The outcomes of the first approach are juxtaposed with those of the second in the light of interoperability

the Registrar General and Census Commissioner, India

[6] Resource Description Framework: https://www.w3.org/RDF/

[7] org: http://www.w3.org/ns/org#

[8] skos: http://www.w3.org/2004/02/skos/core#

[9] gn: http://www.geonames.org

during the metadata integration process. [12] presents a methodology, five basic steps and a model to publish statistical data coming from tabular data sources or relational databases in the form of linked open data. It follows the best practices for publishing linked data, a W3C working group note such as providing URI construction suggestions following URI policy rules, reuse of standard vocabularies, such as the Data Cube Vocabulary, SKOS, etc., converting data to RDF and providing access to the converted data. [13] describes the attempt at building "AKTivePSI"- an initiative by the Office of Public Sector Information, UK to adopt semantic web technology for large scale integration, sharing and reusing the public data for the benefits of government, businesses and citizens alike. The key outcome due to the success of this pilot project led to more increased awareness amongst the government bodies about the power of semantic web technology.

While there is a prolific attempt to introduce a semantic knowledge framework to extract and maintain the metadata required for semantic integration of public data in the western nations, there have been none such coordinated and specific efforts in India to the best of our knowledge. DBpedia resources (extracted from Wikipedia information) do exist for India, though in many cases they are not up-to-date with the information published on the Government of India web sites. The motivation to create Bharathi stems from extensive research work that accomplished semantic integration of open data tables published in data.gov.in using entities from Linked Open Data (LOD)[10] namely DBpedia[11] and Wikidata[12] [14], [15]. Bharathi is our contribution to create, maintain and sustain this knowledge graph of metadata specifically for the Indian context to enable a powerful semantic integration of public data related to governance in India.

3. Bharathi Overview

Bharathi is populated using data extracted from a number of government of India web sites whose content is created and maintained by the National Informatics Centre (NIC), Ministry of Electronics and Information Technology, Government of India. In the current version of Bharathi, we consider information from https://lgdirectory.gov.in/[13], goidirectory.nic.in/, http://censusindia.gov.in/ and https://data.gov.in/ to build the knowledge graph for metadata around administrative regions and their codes, government entities and their classification, various sectors and their links within the respective government function. While Bharathi is extensible in including new sources of information, the extraction pipeline will need modification depending on the structural specifications of the source. The output from the extraction pipeline is converted into semantic entities using the Bharathi ontology (Refer Section 4). The Bharathi knowledge graph is currently accessible from a Virtuoso triplestore with a SPARQL endpoint at http://www.semanticwebindia.in/BharathiLive/sparql. All entities of Bharathi are semantically linked and accessible using persistent resolvable identifiers following the W3C best practices. Bharathi is released under a Creative Commons Attribution-ShareAlike 4.0 International License[14]. The web interface is accessible at http://www.semanticwebindia.in/bharathi/ allowing users to search and browse

[10]LOD: https://lod-cloud.net/

[11]DBpedia: https://wiki.dbpedia.org/

[12]Wikidata: https://www.wikidata.org/wiki/Wikidata:Main_Page

[13]https://lgdirectory.gov.in/ : LOCAL GOVERNMENT DIRECTORY – Complete directory of land regions/revenue, rural and urban local governments

[14]https://creativecommons.org/licenses/by-sa/4.0/

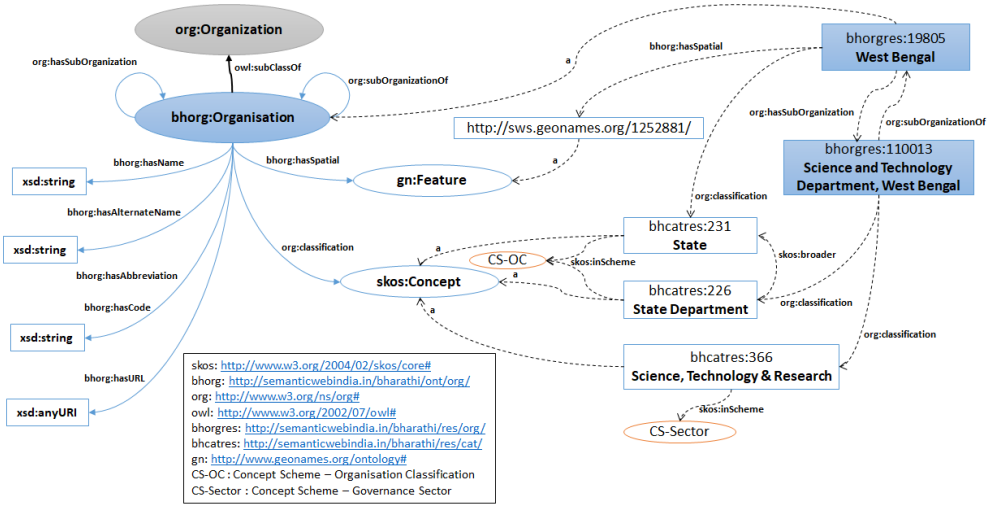

Figure 1: Bharathi Model depicting the ontology and sample instances

Bharathi. The knowledge graph is updated every 3 - 6 months, depending on the changes to the information at source. Information related to the knowledge graph dumps and ontology metrics can be accessed at http://semanticwebindia.in/bharathi/BharathiKG.

4. Bharathi Ontology

The underlying schema for Bharathi uses established ontologies and vocabularies such as the Organisation ontology and the SKOS vocabulary to define the building blocks for the metadata pertaining to government organisations, administrative regions, governance sectors and the classification hierarchy linking these entities. Figure 1 illustrates the domain model for the Bharathi ontology. The Bharathi ontology with name prefix bhorg[15] contains the base classes and properties used by the knowledge graph. The selection of established vocabularies was made to ensure 1) interoperability of the knowledge base through reuse 2) adherence to global definitions of schema for known entities by reusing terms and most importantly 3) leveraging of community support for linked open data cloud vocabularies.

The core element of Bharathi ontology is the Organisation class (bhorg:Organisation[16]) derived as a subClass of org:organization. This entity is used to denote any public, private or government organsiation in India. The other core element of the Bharathi ontology is the use of skos:Concept to represent the classification hierarchy to denote one of the three classifications namely –an organisation classification, a governance sector or a topic within a sector. Separate Concept Schemes (skos:ConceptScheme) have been defined to distinguish between these three different types of classifications. Each concept scheme has its own hierarchy starting from a Top

[15]bhorg: http://semanticwebindia.in/bharathi/ont/org/

[16]bhorg:Organisation - http://semanticwebindia.in/bharathi/ont/org/Organisation

Concept (skos:hasTopConcept). An instance of type bhorg:Organisation has the following object properties 1) org:hasSubOrganization –inherited from the org:Organization class and used to link sub organisations 2) org:subOrganizationOf –inherited from the org:Organization class and used to link to the parent organisation 3) org:linkedTo –inherited from the org:Organization class and used to link related organisations and 4) bhorg:hasSpatial –property defined in the extended class to link to the GeoNames Feature (gn:Feature) associated with the instance. The bhorg:Organisation entity also has data properties to associate the names (primary and alternate), URL, abbreviations and codes with the organisation. All the extended classes and properties of the Bharathi ontology have been provided with detailed annotations in line with RDF principles.

Refer Figure 1, the oval objects represent entities and the corresponding rectangular entities linked by rdf:type (a) labelled link (black dashed link in the figure) indicate the respective instances. All government organisations and administrative regions are instances of the bhorg:Organisation entity while organisation classifications, governance sectors and their hierarchical structures are defined using the SKOS ontology. As illustrated in Figure 1, "Science and Technology Department West Bengal" is a government entity having classification "State Department" and is a sub organisation in the "State" of "West Bengal". The administrative region "West Bengal" is linked to a GeoNames Feature using the corresponding GeoNames knowledge base. While the SKOS concepts "State" and "State Department" belong to the concept scheme "Organisation Classification", the SKOS concept "Science Technology & Research" belongs to the concept scheme "Governance Sector" and is associated with the organisation "Science and Technology Department West Bengal".

5. Generating Bharathi

Bharathi is built using a semi-automated extraction pipeline that periodically crawls the Government of India web sites mentioned in Section 3, annotates the semantic entities and serializes the knowledge graph tuples in the form of a N-Triples file[17]. The knowledge graph is generated inline with the Bharathi model explained in Section 4. Each of these steps is explained in detail below:

5.1. Extraction of Metadata

The source files exists in the form of either HTML pages or CSV (Comma Separated Values) files. The extractor is customised for each of the source web sites to take care of the specific nuances and organisation of information within the respective source web site. Note that this feature of the extraction pipeline can be extended in the future to include more sources of web sites as identified. The information extracted mainly consists of the government entity, its descriptors (such as name, URL, abbreviations), the classification of the organisation and the list of sub organisations. This information is consolidated from various sources into a large single data set (CSV file). The information regarding sector, sub sector and related topics is extracted into a separate CSV dataset as this data structure is different from the extraction format used for the government organisations.

The administrative regions, sectors, topics and other organisations do change periodically (name changes, addition of new districts etc.). These changes are handled using the Provenance Ontology[18]. Wherever the source mentions a version, the same is recorded in Bharathi with a history of activity that triggered the change. URL links

[17] https://www.w3.org/TR/n-triples/

[18] PROV-O: The PROV Ontology –https://www.w3.org/TR/prov-o/

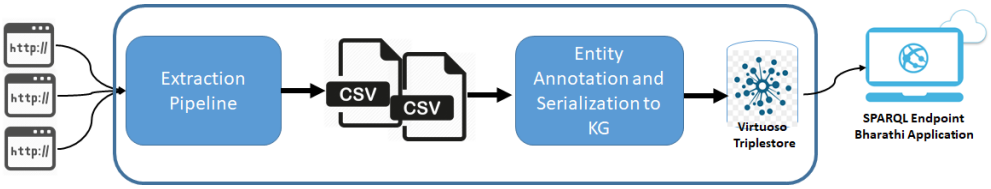

Figure 2: represents the overall architecture for the generation of Bharathi Knowledge Graph.

generated for the Bharathi entities are static. They don't change for the entities once defined in Bharathi thus assuring continuity to the users of Bharathi.

5.2. Entity Annotations

A unique entity is created for each new instance of a government organisation or a classification depending upon the attributes of the entity extracted. GeoNames web services are used to link administrative regions to its corresponding GeoNames feature. The entity annotation routine is sensitive enough to detect that entities with the same label (hasName or prefLabel) need not represent the same entity in the Bharathi knowledge graph. For eg. "Chandigarh" is a Union Territory, a District as well as a Smart City in India. To distinguish this aspect, surrounding attributes of the extracted information from the source web site are used to make the distinction. We use three versions of the Entity Annotator, one each for government organisations & administrative regions, one for organisation classifications and one for governance sectors and topics.

5.3. Serialization of the Knowledge Graph

We have created a Python 3.6 script to generate the N-Triples (.nt) file using the CSV file generated in the extraction step. SPARQL wrapper rdflib routines have been used to create and serialize the tuples using the Bharathi ontology definitions discussed in Section 4. Unique persistent URLs are generated for each instance following the RDF principles. The knowledge graph can be accessed at http://semanticwebindia.in/bharathi/BharathiKG. Figure 2 illustrates the overall architecture for the generation of the Bharathi knowledge graph.

6. Real World Applications using Bharathi

We present two real world applications using the Bharathi knowledge graph.

6.1. Web Interface to Bharathi

This is a web application interface for users outside the scientific community to search, browse and explore the entities and their relationships in Bharathi. The users are allowed to search for a particular entity or find it through one of its classifications. The information is organised in a hierarchical manner to understand the hierarchical classification of each entity in the government function. Further, the linked entities can be traversed just as reaching out from one link to the other in a web page. Each page regarding a government organisation or an administrative region or a classification presents a consolidated identity to the entity. This is an important feature of Bharathi. Most of

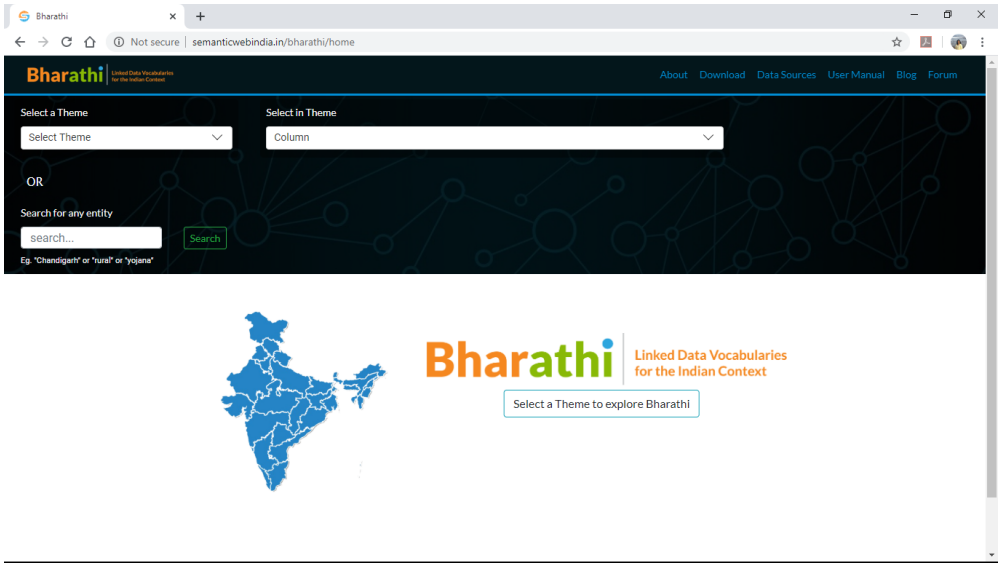

Figure 3: Landing page of Bharathi Application - The External Interface to the Linked Data Vocabulary for the Indian Context

the metadata used for the Indian context is spread across many sources with the same administrative region sometimes referred by its code or its various names or its abbreviation.

Further, provenance history is available for public view at http://semanticwebindia.in/bharathi/download. Also, version history for a semantic entity in Bharathi can be viewed at its functional link e.g. (http://semanticwebindia.in/bharathi/home?id=14900&type=org). A feedback page has been provisioned to capture the user's inputs on how the information can be enhanced with additional metadata from other sources. Figure 3 shows the landing page of Bharathi for the external user. Users can search for an administrative region, government organisation or a sector / sub sector. Bharathi also holds a modest set of common topics found in public data sites published by the government. Figure 4 shows the functional page for the entity "Chandigarh - Smart City". The right pane of the functional page shows the hierarchy of the selected entity.

6.2. Semantic Covid19 India Analysis

This application has been built by harnessing the semantic framework of Bharathi. Covid 19 datasets are sourced from multiple public data sources such as https://www.covid19india.org/ for covid data related to India and https://datahub.io/core/covid-19 for covid data related to other countries. A detailed entity disambiguation routine identifies the linkable semantic entities from the raw Covid19 datasets. The Covid19 indicators such as "Total Confimed", "Total Deceased", "Total Recovered", "Total Tested" etc. were created as sub topics under the Sector - "Health", "Sub Sector" - "Covid19". Each dataset refers to a geographical location for which the indicators have been col-

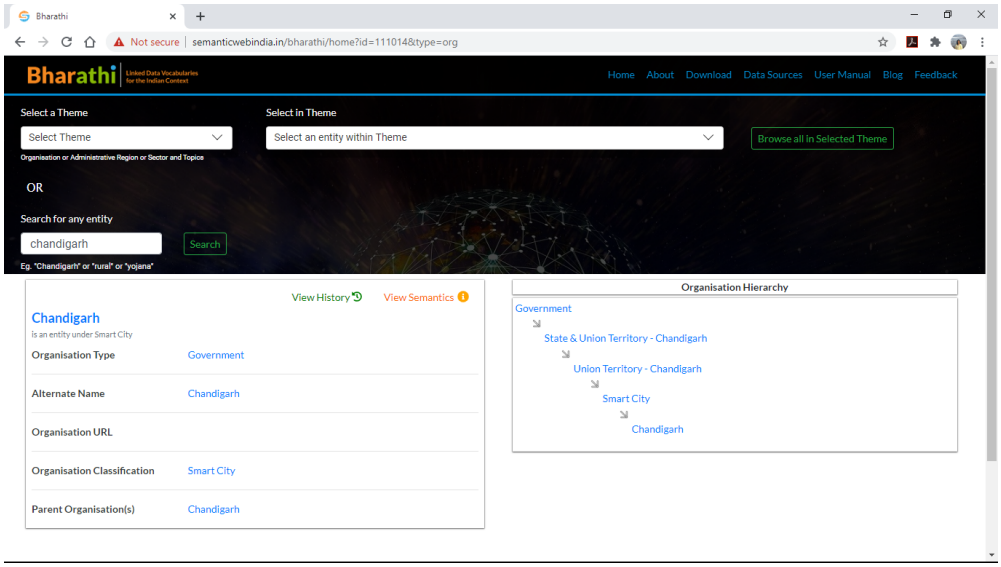

Figure 4: Functional page of a Bharathi Entity

lected and these could be at the state or the sub state Level (eg. districts, cities). These geographical entities are the administrative regions already existing in Bharathi. Using the semantic links of the administrative regions with GeoNames resources and their classifications such as "State" or "District" in India, we are able to answer complex intuitive Covid19 Analysis questions such as 1) What is the trend of "Total Confirmed Cases" in the neighbourhood of "District" Coimbatore (a district in the state of Tamil Nadu in India) within a distance of 100 km. 2) How do states with comparable population density fare on the "Number of Deaths" in a given time period? 3) What are the indicators in the family of "Testing Indicators" that can be compared for a given region and timeline? Each of the above questions is answered by linking the additional metadata from related knowledge graphs using the linked data properties of the Bharathi knowledge graph.

Additionally, using the classification hierarchy of the administrative regions, with minimal effort, the Covid19 application facilitates aggregation of the various indicators at the state level (all districts within a state have a hierarchical relation to their parent state) and analysis of Covid19 transmission across states sharing boundaries. The DBpedia knowledge graph is linked to the GeoNames knowledge base and the GeoNames entities are linked to Bharathi administrative regions. Therefore, this semantic link enables us to explore the effect of Covid transmission across states sharing boundaries.

Figure 5 and Figure 6 show a glimpse of Covid19 Analysis India application and can be accessed at http://sandhicovid.semanticwebindia.com/.

Analysing Covid19 events : Conceptually an event brings together an activity or a milestone, a location and a time period in a single entity. We extended the "Event Class"

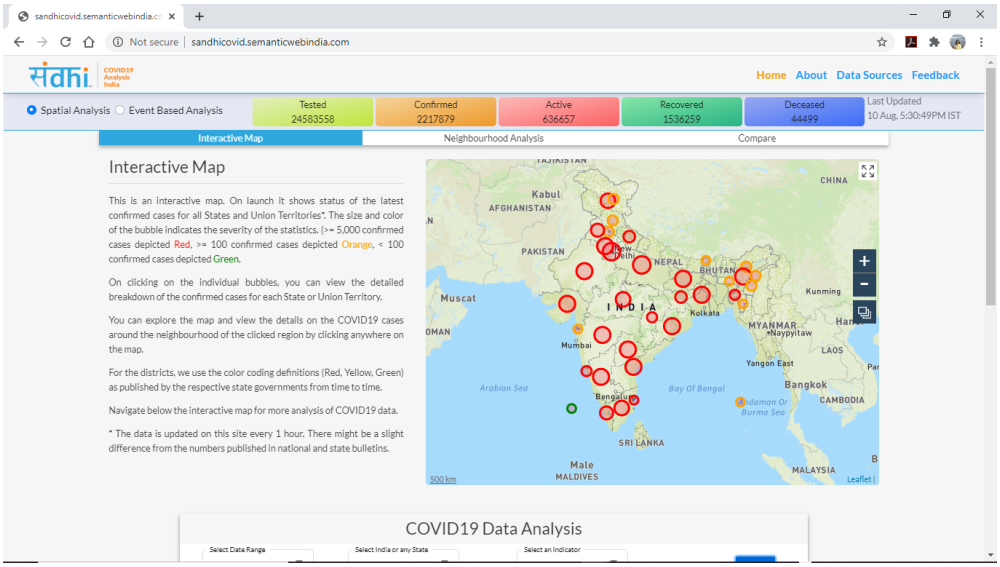

Figure 5: Covid19 Analysis India built by harnessing the semantic layer provided by the Bharathi Knowledge Graph – Landing Page

from the Event ontology[19] to define a custom "Event" in the Sandhi Covid19 application. Apart from the usual attributes inherited from the standard event ontology, we introduced object properties for "predecessor" and "successor", thus enabling an in-depth analysis of related events in the Covid19 timeline. "Lockdown 1.0", "Lockdown 2.0" etc. were created as individual instances of the extended Event class. The semantics facilitated by these linked entities enable a detailed "Event Based Analysis" using the Covid19 data.

Figure 7 shows "Event Based Analysis" feature of Covid19 Analysis India application.

Note that this semantic layer can be extended to any collection of open datasets containing entities pertaining to the Indian context that can be linked to the metadata available in the Bharathi knowledge graph. Thus Bharathi enables advanced analysis based on entity-centric search, exploration and information discovery through the linked data properties of the underlying knowledge graph.

6.3. Sample SPARQL queries to the Bharathi Knowledge Graph

Here we present an assorted list of SPARQL queries to the knowledge graph and describe their expected result.

The SPARQL query in Table 1 returns all the top level organisation classifications in the Government of India.

The SPARQL query in Table 2 returns all the States and Union Territories of India along with their lat-long coordinates by linking Bharathi knowledge graph with the GeoNames knowledge base.

[19]The Event Ontology: http://motools.sourceforge.net/event/event.html

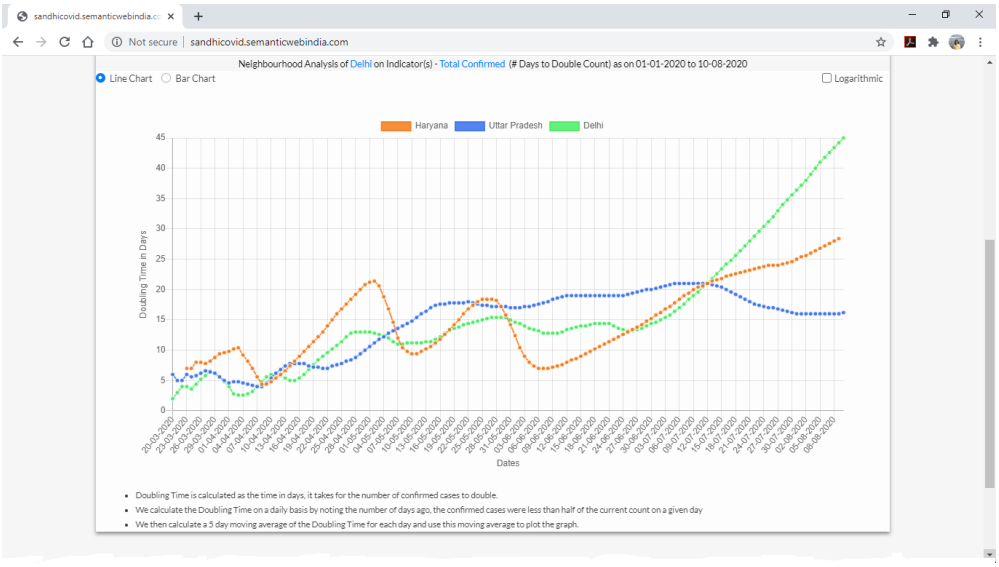

Figure 6: Covid19 Analysis India built by harnessing the semantic layer provided by the Bharathi Knowledge Graph – Compare a Neighbourhood

Table 1
SPARQL Query requesting the top level Organisation Classifications in Government of India

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX bhorg: <http://semanticwebindia.in/bharathi/ont/org/>
PREFIX org: <http://www.w3.org/ns/org#>
select distinct ?c ?cpreflabel ?c1 ?c1preflabel
from <http://sandhiorg.org>
where {
?c a skos:Concept.
?c skos:inScheme
<http://semanticwebindia.in/bharathi/res/cat/OrganisationType>.
<http://semanticwebindia.in/bharathi/res/cat/OrganisationType>
skos:hasTopConcept ?topc.
?c skos:broader ?topc.
?c1 skos:broader ?c.
?c skos:prefLabel ?cpreflabel.
?c1 skos:prefLabel ?c1preflabel.
}
```

The SPARQL query in Table 3 returns all the government organisations and their sub organisations belonging to the classification - "Legislature of India".

Table 2
SPARQL Query requesting all the States and Union Territories of India along with their lat-long coordinates by linking Bharathi knowledge graph with the GeoNames knowledge base

```
PREFIX gn: <http://www.geonames.org/ontology#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX bhorg: <http://semanticwebindia.in/bharathi/ont/org/>
PREFIX org: <http://www.w3.org/ns/org#>
PREFIX wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#>

select ?org1 ?gc ?classname ?name ?lat ?long
where
{
?org1 a bhorg:Organisation. ?org1 bhorg:hasName ?name.
?org1 bhorg:hasSpatial ?gc.
?org1 org:classification ?c.
?c skos:prefLabel ?classname. FILTER (?classname IN
("State", "Union Territory", "National Capital Territory")).
?gc  wgs84_pos:lat ?lat.
?gc  wgs84_pos:long ?long
}
order by ?name
```

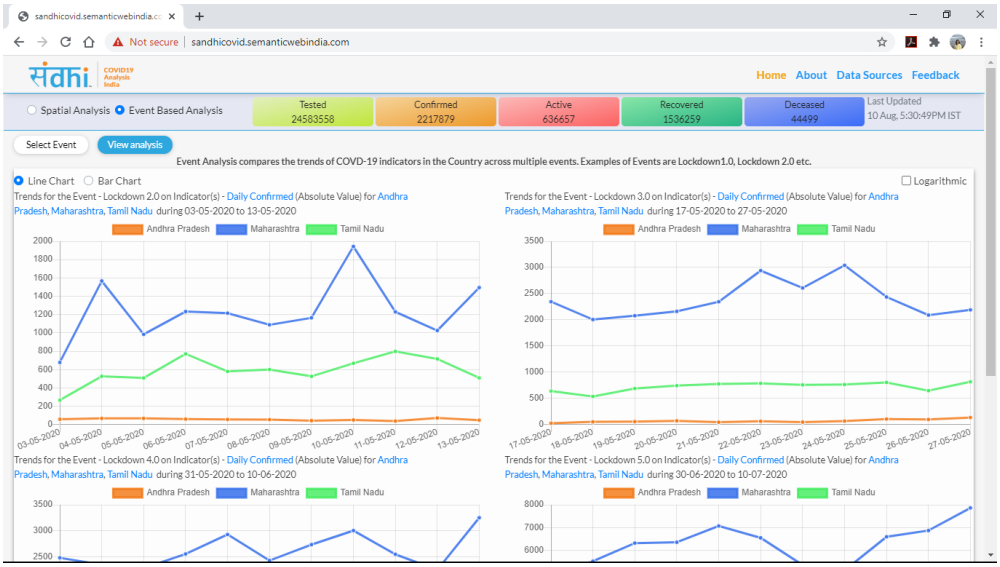

Figure 7: Covid19 Analysis India built by harnessing the semantic layer provided by the Bharathi Knowledge Graph – Event Based Analysis

Table 3
SPARQL Query requesting all the Government Organisations and their Sub Organisations belonging to the classification - "Legislature of India"

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX bhorg: <http://semanticwebindia.in/bharathi/ont/org/>
PREFIX org: <http://www.w3.org/ns/org#>

select distinct ?c1 ?c ?classname ?org1 ?org1name ?sub ?suborg
?subclass ?subclassname
from <http://sandhiorg.org>
where {
?org1 a bhorg:Organisation. ?org1 org:classification ?c.
?c skos:prefLabel ?classname.
?c skos:broader ?c1. ?c1 skos:prefLabel "Legislature".
FILTER NOT EXISTS {?org1 org:subOrganizationOf ?parent}.
?org1 bhorg:hasName ?org1name.
OPTIONAL {?org1 org:hasSubOrganization ?sub. ?sub bhorg:hasName
?suborg. ?sub org:classification ?subclass.
?subclass skos:prefLabel ?subclassname}
}
```

7. Conclusions and Future Work

We have introduced Bharathi - Linked Data Vocabulary for the Indian context, which facilitates the semantic integration of public datasets related to governance data available on the public domain. The semantic linking of entities using Bharathi for public data referring to commonly occuring metadata such as government organisations, administrative regions, commonly used governance sectors and related topics transform the public datasets into a rich semantic web of knowledge that can be intuitively explored and analysed by exploiting the inter relationships across information from diverse data. Bharathi is expected to provide the much needed break through in connecting public data specifically for the Indian context since, to the best of our knowledge, such a large scale publishing of linked information for commonly used metadata in India has not been made available till now.

Comparisons with DBpedia : DBpedia is project to convert structured content of Wikipedia into linked data and does contain

a lot of information regarding India documented in Wikipedia. However, this information is largely incomplete and does not align with some of the most credible updates published by the Government of India (GoI) websites. Bharathi is an effort to exclusively extract information from GoI websites and maintain its currency.

There are several limitations of the Bharathi knowledge graph that are the focus of ongoing and short-term efforts. Bharathi can be enriched with reasoning statements from the OWL vocabulary to enable reasoning and creation of new facts using reasoners such as Apache Jena. We also intend to extend the content of our graph to other sources of information in collaboration with the Government of India to facilitate a larger variety of instances in Bharathi. Finally, our continued efforts on Bharathi knowledge graph aims to build a multi-lingual search across Indian languages to enable larger support and patronage for Bharathi.

References

[1] J. Hoxha, A. Brahaj, Open government data on the web: A semantic approach, in: 2011 International Conference on Emerging Intelligent Data and Web Technologies, IEEE, 2011, pp. 107–113.

[2] H. S. Al-Khalifa, A lightweight approach to semantify saudi open government data, in: 2013 16th International Conference on Network-Based Information Systems, IEEE, 2013, pp. 594–596.

[3] D. DiFranzo, A. Graves, J. S. Erickson, L. Ding, J. Michaelis, T. Lebo, E. Patton, G. T. Williams, X. Li, J. G. Zheng, et al., The web is my back-end: Creating mashups with linked open government data, in: Linking government data, Springer, 2011, pp. 205–219.

[4] L. Ding, T. Lebo, J. S. Erickson, D. Difranzo, G. T. Williams, X. Li, J. Michaelis, A. Graves, J. G. Zheng, Z. Shangguan, et al., Twc logd: A portal for linked open government data ecosystems, Journal of Web Semantics 9 (2011) 325–333.

[5] L. Ding, D. DiFranzo, A. Graves, J. R. Michaelis, X. Li, D. L. McGuinness, J. A. Hendler, Twc data-gov corpus: incrementally generating linked government data from data. gov, in: Proceedings of the 19th international conference on World wide web, 2010, pp. 1383–1386.

[6] T. Lebo, G. T. Williams, Converting governmental datasets into linked data, in: Proceedings of the 6th International Conference on Semantic Systems, 2010, pp. 1–3.

[7] J. Höchtl, P. Reichstädter, Linked open data-a means for public sector information management, in: International Conference on Electronic Government and the Information Systems Perspective, Springer, 2011, pp. 330–343.

[8] B. Hyland, D. Wood, The joy of data-a cookbook for publishing linked government data on the web, in: Linking government data, Springer, 2011, pp. 3–26.

[9] R. Cyganiak, F. Maali, V. Peristeras, Self-service linked government data with dcat and gridworks, in: Proceedings of the 6th International Conference on Semantic Systems, 2010, pp. 1–3.

[10] P. Křemen, M. Nečaskỳ, Improving discoverability of open government data with rich metadata descriptions using semantic government vocabulary, Journal of Web Semantics 55 (2019) 1–20.

[11] L. Bountouri, C. Papatheodorou, V. Soulikias, M. Stratis, Metadata interoperability in public sector information, Journal of Information Science

35 (2009) 204–231.

[12] I. Petrou, M. Meimaris, G. Papastefanatos, Towards a methodology for publishing linked open statistical data, JeDEM-eJournal of eDemocracy and Open Government 6 (2014) 97–105.

[13] H. Alani, D. Dupplaw, J. Sheridan, K. O'Hara, J. Darlington, N. Shadbolt, C. Tullo, Unlocking the potential of public sector information with semantic web technology, in: The semantic web, Springer, 2007, pp. 708–721.

[14] A. Subramanian, S. Srinivasa, Semantic Interpretation and Integration of Open Data Tables, Springer Singapore, Singapore, 2018, pp. 217–233. URL: https://doi.org/10.1007/978-981-13-2330-0_17. doi:10.1007/978-981-13-2330-0_17.

[15] A. Subramanian, Semantic integration and knowledge representation of open data powered by linked open data, Ph.D. thesis, International Institute of Information Technology Bangalore, 2018. URL: https://shodhganga.inflibnet.ac.in/handle/10603/262218.