GEO-RDF FRAMEWORK FOR REPRESENTING THE SPATIAL INFORMATION OF BANGLADESH

Md Hasan Hafizur Rahman [1] and Hanif Seddiqui [2]

[1] Department of Computer Science and Engineering, Comilla University, Cumilla - 3506, Bangladesh
[2] Department of Computer Science and Engineering, University of Chittagong, Chittagong - 4331, Bangladesh

ABSTRACT

Geographic data plays a vital role in supporting modern location-based services on the World Wide Web (WWW). In Bangladesh, such data exists in structured, semi-structured, and unstructured formats, stored across government agencies, research institutions, and private organizations using diverse formats and protocols. This heterogeneity creates significant integration challenges, limiting operational efficiency and the development of applications ranging from navigation and logistics to personalized services and emergency response. Our research addresses this by transforming and integrating disparate datasets into a machine-understandable form. We modeled the complete administrative hierarchy of Bangladesh, from divisions to villages, generating 0.40 million RDF (Resource Description Framework) triples within a unified semantic repository, Geo-Bangladesh. This repository enables effortless integration and retrieval of geospatial data across all administrative levels. We further linked Geo-Bangladesh with related repositories, including the educational institutions and citizen information, enabling the mapping and visualization of entities along with locations. Using geoSPARQL, we retrieved and inferred spatial and non-spatial data, demonstrating the repository's usability, interoperability, and effectiveness. Unlike raw GeoSPARQL implementations or general-purpose ontologies such as Geonames, Geo-Bangladesh is explicitly tailored to Bangladesh's administrative structure and reconciles inconsistencies such as the 68 vs. 64 districts problem. Compared to OGC-compliant frameworks, our repository incorporates both semantic interoperability and localized reconciliation, demonstrating advantages in accuracy, query flexibility, and alignment with national data sources.

KEYWORDS

Semantic Web, Ontology, RDF, Spatial Data, geoSPARQL

1. INTRODUCTION

The rapid growth of the World Wide Web (WWW), combined with the proliferation of vast and heterogeneous geographic information repositories, has made it increasingly difficult to develop smarter methods for retrieving specific information. These heterogeneous web contents are published in the form of documents using traditional technologies such as, Hypertext Markup Language (HTML)[1], Dynamic HTML (DHTML)[2], which focus on defining stylistic presentation through predefined tags rather than conveying the semantic meaning of the data. As a result, the published web contents using these conventional technologies consist of unstructured text, those are easily readable by humans, however it is hardly understandable to machines. Let us consider the traditional HTML representation of information "The Gomti river located in the Cumilla

[1]http://www.w3.org/html
[2]http://en.wikipedia.org/wiki/Dynamic_HTML

DOI: 10.5121/ijwest.2025.16401

district" in a paragraph of a document without any semantic markup to indicate that "Gomti" is the name of a river and "*Comilla*" refers to a district of Bangladesh. Although human being, those are familiar with the terms *Gomti, River, Comilla*, *District,* and *Located* can extract the meaning from this understandable representation by implementing their intuitive learning knowledge, whereas the machines merely consider this representation as a sequence of characters.

Furthermore, eXtensible Markup Language (XML)[3] is a semi-structured data representation technique having the facilities of defining user defined tags to represent domain specific information in the hierarchical fashion. By reason of the inadequacy of uniqueness in user defined tags to represent domain information in a document is scarcely understandable to machines (Shanmugasundaram, 1999). On the other hand, Database Management Systems (DBMSs), such as Microsoft Access[4], MySQL[5] and so forth are used to represent structured data in the back end to originate dynamic web by performing various structured query operations such as insert, delete, update, search and so on (Silberschatz, 1997). The front-end of these dynamic webs developed using available traditional technologies. Moreover, the database schema structures are restricted to define attributes and their attribute constraints using primary keys, foreign keys and null rules to develop meaningful associations among data to create a chain of data links on the contents among available data repositories. As a result, a closed world assumption is applied to relational database model that loss big semantic in the process of data modelling (Martinez-Cruz, 2012). In this regard, humans can understand these contents effortlessly, but machines are impotent to extract the meaning of these data representation on the contents of existing web. As a consequence of the deficiency of semantics, the traditional search engines do not attain specific information of web contents. Additionally, we deal with keywords of a query without their explicit meaning using conventional search techniques. In this regard the outcomes of keyword *Cumilla* retrieve redundant information related to this keyword Cumilla such as geo-coordinates, health care centres, historical heritage, inhabitants of this place, however that is one of the sixty-four districts, and a second administrative level of Bangladesh. In spite of the available information related to *Cumilla* on the existing WWW, these traditional search engines, such as google[6], yahoo[7], Bing[8] etc. extract a large number of documents link from available scattered and unstructured data sources based on the sequence of characters *Cumilla* instead of the interpretation of *Cumilla* that is a place or not and feedback relevant and irrelevant documents to users.

Now-a-days numerous proprietary Geographic Information Systems (GISs) software are available to produce heterogeneous information on the web by maintaining their individual data representation structure. This software is incompatible to design complex systems, sophisticated data models, and database storage structures (Chomicki, 1999) (Peng, 2005) (Zhang, 2005). The information integration and sharing operation perform on multi-format data sources produced from these proprietary solutions on the WWW is a formidable task. Furthermore, the meaningful spatial data representation and reasoning to extract relevant, reliable and accurate information is a crucial research challenge to develop interoperable geographical data repositories and software(Bishr, 1998) (Harvey, 1999), geographical information retrieval(Jones, 2001) and automated spatial reasoning systems (Cohn, 1995). It is also hard to capture and maintain semantics of geographical data due to their complexities of geographical categories (Smith,

[3]http://www.w3.org/XML/
[4]https://products.office.com/en-us/access
[5]https://www.mysql.com/
[6]https://www.google.com.bd/?gws_rd=cr\&ei=21OIVa_0Hc-gugSf8YPYAw
[7]https://search.yahoo.com/
[8]http://www.bing.com/

2001), geospatial languages (Frank, 1991) and heterogeneous nature of spatial data (Freitas, 2012). However, a series of specifications such as Geographic Markup Language (GML) (Burggraf, 2006) and data access protocols such as Web Feature Service (WFS) (Vretanos, 2005), and Web Map Service (WMS) (de La Beaujardi`ere, 2004)can access heterogeneous contents stored in the web. These standards do not have enough constructs to express data semantics on the web. On the other hand, ontology model is used to define common vocabularies that minimizes the semantic interoperability problems, metadata modelling, and interoperable meaning of data across domains, and data integration(Brodeur, 2003) (Fonseca, 2002)(Jones, 2001). It is employed as a method for identifying categories, concepts, relations, and rules that prescribe theories of the geographical domain (Mark, 1999). However, a lot of efforts has been made to develop the next generation web, *Semantic web*, coined by Berners-Lee et al. (Berners-Lee, 2001) defines the objective of data on the web to achieve the semantic interoperability among metadata associated on the web of information. In spite of the exponential growth of semantic web to address the issues of data integration and machine understandability, geographic information of a country is a highly motivated complication problem with vast application areas such as Ontology-based Spatio-temporal Data Analysis, Decision Support Systems (DSS), Location-Based Services (LBS) and reasoning on interoperable web of data. In this regard, the main center of attraction of our research is to develop a generic machine understandable geospatial knowledge-base *Geo-Bangladesh* from the existing available data on the web using semantic web technologies to achieve resource sharing and interoperability issues in a faster and effectively. In order to shows the effectiveness of our application, we performed a couple of experiments to retrieve accurate and predictable information using geoSPARQL when and where it is needed. The interoperability issues in our research are cross-examined with one more knowledge repository *Bangladeshi-Citizen* and *HCN-BD* related to the citizen information and health care network of Bangladesh respectively.

The rest of the paper is organized as follows. Section 2 introduces general terminologies to comprehend the consequent contents of this paper. Our present geographic structure of Bangladesh is articulated in Section 3, while Section 4 focuses on the details data preparation procedure and Section 5 describes the methodology of our system, elaborates the process of developing ontology of geographic data of Bangladesh as Linked Open Data (LOD) and integrating this information accurately on the web. Section 6 includes the knowledge representations, evaluation and utilization of our proposed Linked Open Data (LOD) by performing a large number of experiments with GeoSPARQL. Concluded remarks along with some future directions of our work is described in Section 7. In addition to current implementation, we outline a plan for scalability and maintenance: (a) schema evolution mechanisms to incorporate new datasets and administrative changes, (b) periodic updates through automated crawlers linked to BBS and Geonames, and (c) a deployment plan for supporting millions of triples, benchmarked to maintain sub-second query times. Practical deployment includes mobile-based query access, ensuring public usability.

2. General Terminologies

This section familiarizes some basic terminologies with their notation to readers those are used throughout this research paper. It includes the theme of Ontology, RDF (Resource Description Framework) model and HTTP-URIs (Hyper Text Markup language - Uniform Resource Identifiers), Linked Data (LD), Linked Open Data (LOD) and Interoperability to comprehend the essence of our paper.

2.1. Ontology

An **ontology** is *an explicit, formal specification of a shared conceptualization of a domain of interest* (Gruber, 1995) (Studer, 1998). It represents knowledge of a specific domain in a formal way by describing a set of concepts that carry the same meaning across different knowledge repositories. Ontologies serve as the background knowledge necessary to fulfill the vision of the Semantic Web (Berners-Lee T. a., 2001) (Maedche, 2001), which is considered the foundation of the next-generation web.

An ontology typically includes *a core ontology, logical mappings, a knowledge base, and a lexicon*. A core ontology, denoted as S, is defined as a five-tuple:

$$S = (C, \leq_C, R, \sigma, \leq_R)$$

This consists of two disjoint sets, C and R, where the elements of C are called *concepts* and those of R are called *relations*. The symbol $\leq_C$ represents a partial order on C, referred to as the concept hierarchy or *taxonomy*. Similarly, $\leq_R$ is a partial order on R, known as the relation hierarchy, where: $r_1 \leq_R r_2$ if and only if $dom(r_1) \leq_C dom(r_2)$ and $ran(r_1) \leq_C ran(r_2)$.

There is also a function $\sigma: R \rightarrow C \times C$, known as the signature of a binary relation, defined as:$\sigma(r) = \langle dom(r), ran(r) \rangle$, where $r \in R$, and dom(r) and ran(r) represent the domain and range of r, respectively (Ehrig, 2006).

2.2. RDF Model and HTTP-URI

RDF, a generic graph-based framework to structure and interchange data, is used to represent metadata of real-world domains in the *subject-predicate-object* form, known as triples (Antoniou, 2004). In this model resources, *items of interest of domains* are described using HTTP-URIs to create globally unique name to access properties and relationships of resources by avoiding information redundancy on the web without centralized management. Moreover, an RDF triple is a piece of knowledge having subject, predicate and object.

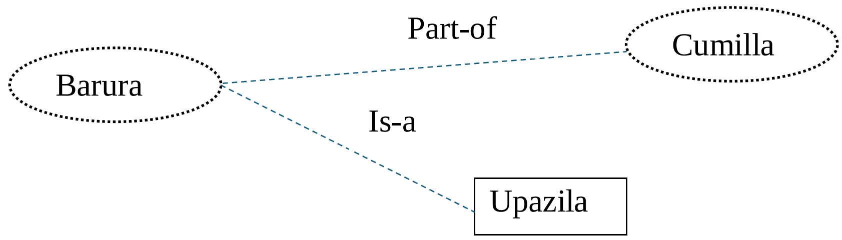


Figure 1: RDF Triple for *Barura is an Upazila of Cumilla district*

The triple *Barura → part-of → Cumilla* in Figure 1 demonstrates the basic structure of triple for the knowledge piece *Barura is an Upazila of Cumilla district* in the *subject-predicate-object* form to comprehend the essence of semantic knowledge representation. The subject of a triple is a URI, the address of the resource that we want to describe in a domain while the predicate of a triple is the description of a fact about a certain domain that represents the relationship between subject and object.

2.3. RDF Model and HTTP-URI

The structure of a knowledge base is characterized as

$$KB = (C, R, I, i_C, i_R)$$

including two disjoint sets C and R those are interpreted in our previous subsection. The notation$I$ is a set of elements called *instances* and two functions i_C and i_R are *concept instantiation* and *relation instantiation* respectively.

In our application we use the structure of ontology and ontology knowledge-base to express our domain data as Linked Data and distribute these knowledge pieces as Linked Open Data *(LOD)*to establish a universal data source on the web (C. Bizer, 2009) (Bizer, 2011). These knowledge-base develop the web as interoper able because interoperability performs the operation of information swapping between two or more parties by resolving syntactic and semantic issues properly (A. Geraci, 1991) (Goodchild, 1999).

2.4. PCS Computing

The Physical-Cyber-Social *(PCS)* computing paradigm incorporates information from cyberspace data sources and social perceptions by utilizing computing and communication capabilities to extract intuitive knowledge that neither traditional computing nor human intelligence can explore (Wayne, 2007) (A. Sheth, 2013).

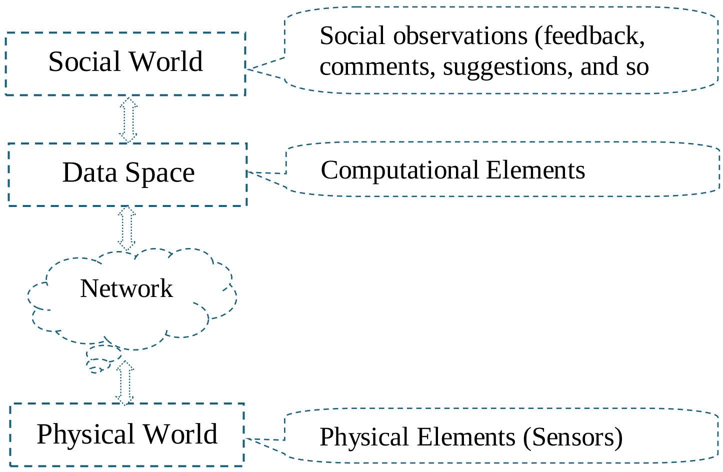


Figure 2: The Physical-Cyber-Social *(PCS)* computing approach

This perceptive knowledge is analysed by expanding the web with the capabilities of sensing, processing and self-adapting to perform interactions between physical and virtual world through exclusive addressing scheme (Wang, 2010). In this connection, a close look of *PCS* computing paradigm is illustrated in the Figure 2 to comprehend the significance of our research.

2.5. GIS, GPS, and GSM

The Geographic Information Systems *(GISs)* process *geo-referenced* data by combining hardware and software for processing, analysing and displaying in the form of maps, reports and charts. These data are spatial, *points, lines or areas known as locational data* and attribute *(non-locational)* data, *features of points, lines or areas* (Lepper, 1991).Furthermore, the Global Positioning System *(GPS)* is a satellite-based uninterrupted, 3-D positioning and navigation system whereas Global System for Mobile Communication *(GSM)*use three GSM base-stations to locate a point in the form of *{latitude, longitude}* on the earth (O. O. Alharaki, 2008).

3. GEOGRAPHIC STRUCTURE OF BANGLADESH

Bangladesh, officially *The People's Republic of Bangladesh* is a country of Southern Asia lies in the geographical coordinates from $20^0$34N to $26^0$38N latitude and from $88^0$01E to $92^0$41E longitude of 147570 square kilometers (56997 square miles) (M. H. H. Rahman, 2014) (S. Chakraborty, 2014). This geographic area is organized by a standard administrative hierarchy from its higher-level *divisions* to its lower-level *villages* for smooth administrative functions. Each administrative level in this structure follows a logical relationship with its upper and lower levels that is portrayed in the Figure 3.

Each administrative level in Bangladesh is associated with specific geographic attributes such as latitude, longitude, boundary area, elevation, and other spatial metadata. To accurately pinpoint a location within the national administrative hierarchy, entities are organized in a structured chain using *part-of* and *consists-of* relationships, as illustrated in Figure 3. This hierarchical model ensures that each entity is contextually linked to its parent and child administrative units, enabling precise geographic referencing and semantic consistency.

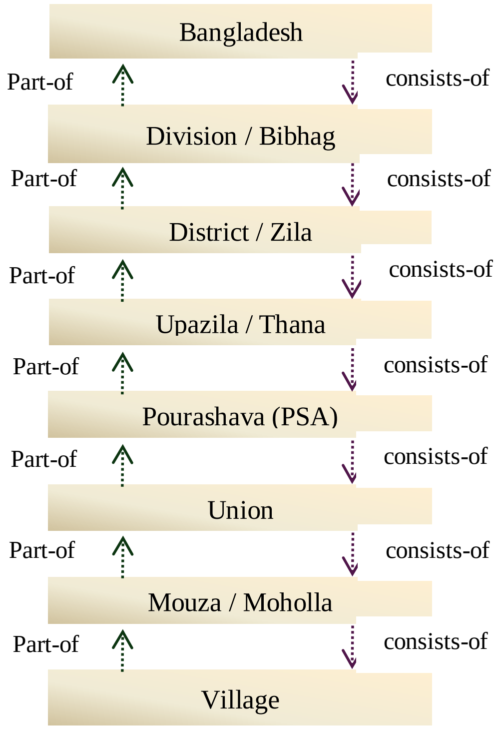


Figure 3: The logical relationships among different administrative levels of Bangladesh

Country: The root entity of the administrative hierarchy represents *Bangladesh*. Although this entity is not explicitly stored in the database, it is logically defined and mapped to the Country class. All subsequent administrative units are ultimately part of this top-level entity.

Division: A *division* (locally known as *Bibhag*) is the first-level administrative subdivision of Bangladesh. It is linked to the Country entity via a partOf relation and is represented in the Division class.

District: We define alternative name for district is zila. We also creat eeach district, a subdivision of first-order administrative division as an entity in our domain and build associations with division using partOf relation. We describe each district in the Distict class. The alternative name for district is Zila.

Upazila: An *upazila* is the third-level administrative unit and a subdivision of a district, connected via a partOf relation. In metropolitan areas, the equivalent administrative unit is the *police station* (locally called *Thana*), which replaces the upazila in addressing locations. These entities are defined in the Upazila class.

Pourashava: *Pourashava* (municipality) is linked to an upazila using a partOf relation and is represented in the Pourashava class. A special case exists for pourashava code 99, which allows a direct link to its lower-level entity (*Union*) rather than the *Ward*. For other pourashava codes, typed associations are established with *Ward* entities.

Ward: The ward is an entity, and we create a partOf relation between each ward and the pourashava. In this connection, pourashava codes are designated except 99. Each ward is assigned to a Ward class.

Moholla: We create an entity for each moholla and a partOf relation between each moholla and the ward. The moholla entity is assigned to the Moholla class. This level is one of the populated places in our logical structure.

Union: Each union, the smallest rural administrative and local government unit, is connected to pourashava by allowing pourashava code 99. Each union is considered as an entity and there is a partOf relation between each union and the pourashava. All unions are assigned to Union class.

Mouza: We define each mouza as an entity and create a partOf relation between each mouza and the union. They are assigned to Mouza class.

Village: Villages are populated place of Bangladesh and the lower level in our administrative hierarchy. We create an entity for each village and there exist a partOf relation between each village and the mouza. They are assigned to the Village class.

As a result, geographic data related to each level create a chain with its upper and lower levels using partOf relation to extract accurate information regarding of administrative hierarchy when and where it is needed.

4. GEOGRAPHIC DATA PREPARATION OF BANGLADESH

The geographic data of Bangladesh is not readily available from a single authoritative source, making it a challenging task to collect, consolidate, and validate information from multiple providers. This process involves reconciling inconsistencies, filling in missing values, and ensuring data accuracy. To address this, we employ both manual and automated validation techniques.

One of the key reference datasets we use is the official geographic coding system published by the Bangladesh Bureau of Statistics (BBS), provided in the Microsoft Excel file *GeocodeBD.xls*[9].This file consists geographic names of 7 divisions, 64 districts, 500 upazilas and 509 administrative thanas, 265 pourashava or municipalities, 2407 wards, 4451 union council,

[9]http://bbs.gov.bd/RptGeoCode.aspx

67100 mouza or moholla and 87968 villages. Among these data 96,600 populated places are obtainable. These codes serve as a critical reference for linking and cross-validating data from other geographic datasets, ensuring consistency in administrative boundaries and location identifiers. The fragment of data in this data-source is shown in the Table 1.

Table 1. Sample data in the GeocodeBD file

Division Code	Zila Code	Upazila/Thana Code	PSA	Union Code	Mouza/Moholla Code	Village Code	Area Name
20	19	09	99	69	081	01	Bara Horipur
20	19	09	99	69	081	02	Srirampur

The records in Table 1 indicate that the location is part of the Chittagong Division (geo code 20) and lies within the Comilla District (Zila), which is assigned geo code 19. Administratively, it falls under the Barura Upazila (Thana), designated with geo code 09. At the local governance level, the area is included in the Dakshin Khosbas Union/Ward (geo code 69), which comprises the Bara Horipur Mouza/Moholla identified by geo code 081.

Table 2. Sample data in the features file

Feature Class	Feature Codes	Feature	Feature Description (English)	Feature Description (Bengali)
BA	BA00	Division	a primary administrative division of a country, such as a division of Bangladesh	একটি দে শের প্রধান প্রশাসনিকবিভাগ, যেমন বাংলা দে শেরকেন্দ্রবিভাগসমূহ
BA	BA01	District / Zila	a subdivision of a first-order administrative division, such as district or zila of Bangladesh	প্রধান প্রশাসনিক বিভা গেরশৃঙ্খ লেরপ্রথমিকউপবিভাগ, যেমনবাংলা দেশের কেন্দ্র জেলা/জিলাসমূহ
BA	BA02	Upazila / Thana	a subdivision of a second-order administrative division, such as upazila or thana of Bangladesh	দ্বিতীয় প্রশাসনিক বিভা গেরশৃঙ্খ লের উপবি ভাগ, যেমনবাংলা দে শের কেন্দ্র উপ জেলার বাথানাসমূহ

Moreover, it is a formidable task to aggregate latitude and longitude values for each location. To do this task we conscientiously congregate this latitude and longitude coordinate values from the geo-names geographical database, BD.xls[10]. The data segment of this file is given in Table 3. In this data source, there are 53719 records that are serviceable for geographic locations. We extract latitude and longitude values only for administrative structure related entities of Bangladesh.

[10] http://download.geonames.org/export/dump/

Table 3. Data segments of serviceable geographic locations

Geonam esid	Name	Latitude	Longitude	Feature Code	Admin 1 Code	Popu.	Elevation (m)
6416752	Aaartali	21.398	92.224	PPL	84	40519	2
6414576	Adar Char	21.995	92.154	PPL	84	0	3
1213248	Adhunagar	21.982	92.075	PPL	84	0	5

In spite of these data source, there are a huge number of challenges to incorporate appropriate values with *Geocode-BD* for a particular location. In this database the feature code ADM1 is equivalent to our division in Geocode-BD. Therefore, feature code ADM2 is also similar to district or zila. There are 68 districts are available in this dataset instead of 64 districts from the dataset provided by the government of Bangladesh. In order to identify duplicate records for Jhalokati, Khagrachhari and Pirojpur districts are removed to aggregate latitude and longitude values. Moreover, another record for parbattya chattagram district is not a valid second level administrative entity of Bangladesh. Furthermore, in the ADM3 level has 308 data records in the geo-name geographic server. However, there are only four classified records for the fourth administrative level and 248 records missing in *BD.xls* file for ADM3 level. In order to identify these missing data and challenges of automatic extract coordinate values of an entity we successfully complete this task by google geo-coding API. Moreover, each geographic entity is represented with latitude and longitude coordinates in Cartesian WGS84 (World Geodetic System 1984) format, a standard coordinate reference system mainly used in cartography, geodesy and navigation to represent geographical coordinates on the Earth[11]. We carefully removed them and documented reconciliation rules. For example, the discrepancy of 68 vs. 64 districts was resolved by aligning BBS data as authoritative, with three duplicates eliminated and one invalid Parbattya Chattagram record discarded. Missing ADM3 coordinates (248 entities) were automatically filled using Google Geo-coding API, and we validated these by random sampling against official maps, achieving an estimated 92% accuracy.

5. Our Methodology

The approach of our application is to classify resources and instances of the system including their respective features. The subsequent phase of our research is to recognize concepts and relationship types of resources for making machine understandable data repository coined a name *Geo-Bangladesh, a semantic database related to the geographic information of Bangladesh* from available data those are related to our domain of interest. In the third stage, we published this knowledge repository as Linked Open Data (LOD) to create link with related web of contents to address the interoperability issues on the web. In the consecutive step, a lot of experiment is performed to retrieve and inference specific information using SPARQL[12] from our machine understandable data repository. The overview of our approach is depicted in Figure 4 to comprehend the essence of the system.

[11]https://www1.nga.mil/ProductsServices/
[12]http://www.w3.org/TR/rdf-sparql-query

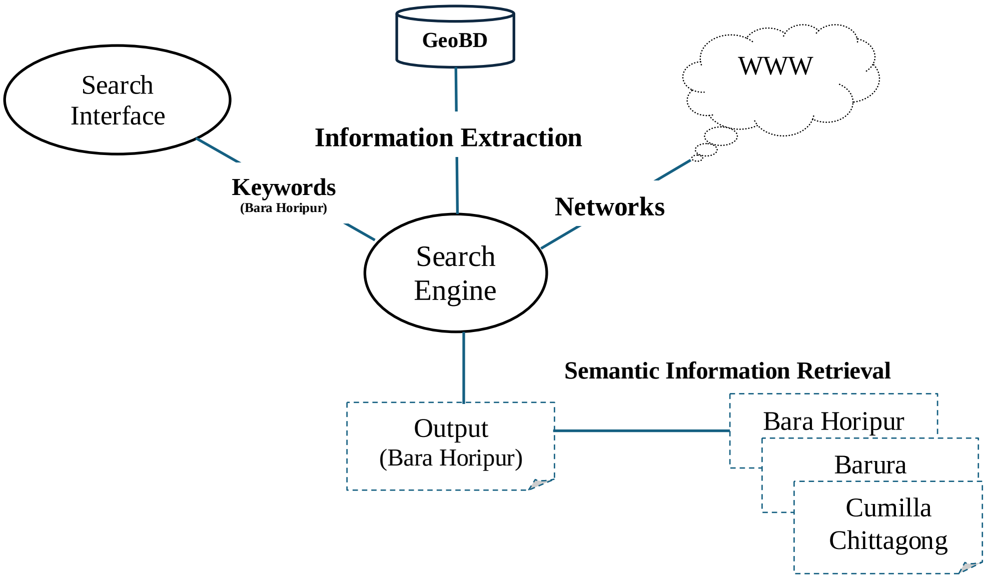


Figure 4: Our semantic approach to extract geographic information of Bangladesh

5.1. Develop Ontology of Geo-Data of Bangladesh

An ontology defines common vocabularies for researchers (Noy, 2001) who need to share information of a domain on the web of data. It is easier to publish data in RDF using available vocabularies from the web. In the case of unavailability of vocabularies, users can propose for a suitable vocabulary to describe domain data on the web. In order to identify these vocabularies, we define concepts of our domain in the machine understandable format and establish relation among them.

we develop Geo-inspired ontology to describe our domain knowledge explicitly. The basic steps to develop an ontology is (Noy, 2001): defining classes and arranging these classes in a taxonomic (subclass-superclass) hierarchy, defining relationships with other classes called slots and allowed values for these slots. We create a knowledge base by defining individual instances of these classes filling in specific slot value and additional slot restrictions (in Figure 5).

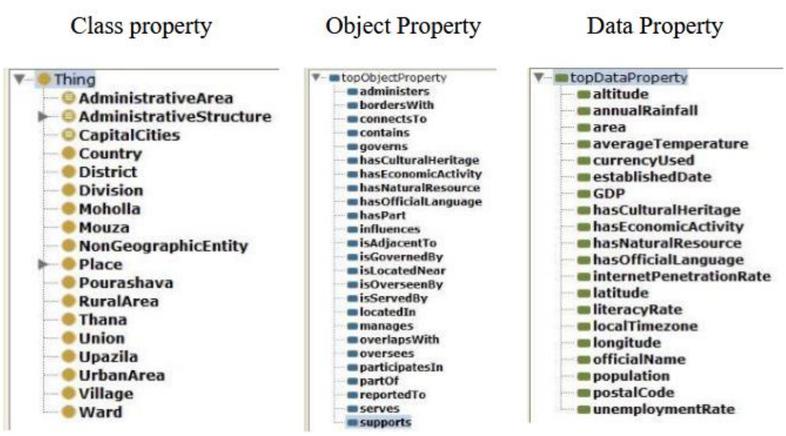


Figure 5: Class, object, and data property for our domain

However, our knowledge data repository *Geo-Bangladesh* contains fundamental geographic data related to the administrative structure considering classes, sub-classes, instances, relations, object property, data property and general axioms etc. It is important to select vocabulary to maximize interoperability with a wider consensus on the web. We use Dublin Core (DC)[13] and DCMI-BOX[14] standard vocabularies to encode geographic meta-data of our domain. For example, a resource type is defined in a class using http://www.w3.org/1999/02/22-rdf-syntax-ns\#type vocabulary. Each instance of a class consists of their related features such as name and comment is published using http://www.w3.org/2000/01/rdf-schema\#label and http://www.w3.org/2000/01/rdf-schema\#comment vocabularies. Moreover, latitude and longitude of each location are defined using common vocabularies of http://www.w3.org/2003/01/geo/wgs84_pos\#lat and http://www.w3.org/2003/01/geo/ wgs84_pos\#long. Therefore, the geographic bounding attributes for north, south, east and west are mapped to dcmibox: northlimit, dcmibox: southlimit, dcmibox: eastlimit and dcmibox: westlimit respectively. Therefore, each resource of our domain is linked with other resources using RDF schema definition http://www.w3.org/2000/01/rdf-schema\#partOf. In addition, the foundational ontology design was inspired by DOLCE principles for spatial hierarchies, ensuring semantic rigor. Ambiguities such as duplicate district names were resolved through explicit disambiguation rules. Crucially, we aligned our classes with external vocabularies: WGS84 Geo Positioning, Geonames ontology, and schema.org using owl:sameAs and skos:exactMatch. This ensures interoperability beyond internal datasets.

We categorize nouns as divisions (Bibhag), districts (Zila), upazilas (Thanas), Pourashava (PSA), unions (wards), mouza (moholla), and villages as classes. Then we define verbs like partOf, locatedIn, and isAdjacentTo as object properties that show (in Figure 5) how these groups relate to each other. We categorize terms for enumeration based on their roles in the domain, such as hierarchical structure, attributes, and roles, and then organize them into classes and properties based on their semantic relevance and contextual use. The approach ensures comprehensive coverage of the domain while maintaining semantic interoperability. Our ontology includes 18 classes representing each administrative level and geographic feature, 25 object properties defining relationships like partOf, locatedIn, and isAdjacentTo, and 20 datatype properties capturing attributes such as altitude, GDP, coordinates such as latitude, longitude, population data, and others. Additionally, it contains more than ninety thousand individual instances that represent distinct entities, such as Bangladesh for country, divisions (e.g., Chittagong, Dhaka) and Comilla, Chittagong as geographic entity, in order to model our domain data.Table4 gives us a close look of the machine understandable representation of geographic data in RDF format.

Table 4: Snippet of our domain data in RDF Triple format

Subject	Predicate	Object
<http://www.hhr.me/#20>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.skeim.org/#division>.
<http://www. hhr.org/#20>	<http://www.w3.org/2000/01/rdf-schema#label>	Chittagong@en.
<http://www. hhr.org/#20>	<http://www.w3.org/2000/01/rdf-schema#comment>	"Information about Chittagong division"@en.
<http://www. hhr.org/#20>	<http://www.w3.org/2003/01/geo/wgs84_pos #lat>	"22.330391"@en.
<http://www. hhr.org/#20>	http://www.w3.org/2003/01/geo/wgs84_pos# long	"91.825188000000004"@en.

[13]dublincore.org/documents/dces/
[14]http://dublincore.org/documents/dcmi-box/

Subject	Predicate	Object
<http://www. hhr.org/#20>	<http://www.w3.org/2000/01/rdf-schema#partOf>	<http://www.skeim.org/#Bangladesh>.
<http://www. hhr.org/#2019>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.skeim.org/#district>.
<http://www. hhr.org/#2019>	<http://www.w3.org/2000/01/rdf-schema#label>	"Comilla"@en.
<http://www. hhr.org/#2019>	<http://www.w3.org/2000/01/rdf-schema#comment>	"Information about Comilla district"@en.
<http://www. hhr.org/#2019>	<http://www.w3.org/2003/01/geo/wgs84_pos#lat>	"23.455959"@en.
<http://www. hhr.org/#2019>	<http://www.w3.org/2003/01/geo/wgs84_pos#long>	"91.18203689999996"@en.
<http://www. hhr.org/#2019>	<http://www.w3.org/2000/01/rdf-schema#partOf>	<http://www.skeim.org/#20>.

5.2. Semantic Information Integration Using Pcs

The operation of data integration between PCS data-sources along with information of available knowledge-base on the web is a challenging task to extract specific information for spatial analysis to apply various policies to improve the quality of search. In this case we use semantic data source of educational institution, EduBD of Bangladesh. The Algorithm 1 demonstrates the process ofdata integration from another semantic data sources.

Algorithm 1: Integrator (geo_coordinates)

Data:

- q — Queue
- EduBD — KB of Educational Institutes
- GeoBD — Geo-Bangladesh database
  Input: geo_coordinates
  Output: Mappings of institutes to (latitude, longitude)

Procedure

1. **ENQUEUE** geo_coordinates **INTO** q.
2. **WHILE** geo_coordinate s≠ 0 **DO**
   2.1 addresses ← **PROCESS_ADDRESSES_FROM** GeoBD.
   2.2 **FOR EACH** address **IN** addresses **DO**
      2.2.1 institutes ← **EXTRACT_FROM** EduBD **USING** address.
      2.2.2 **FOR EACH** institute **IN** institutes **DO**
         2.2.2.1 locations ← **EXTRACT_LOCATIONS_OF** institute.
         2.2.2.2 **FOR EACH** location **IN** locations **DO**
            2.2.2.2.1 (latitude, longitude) ← **LOOKUP_COORDS_FROM** GeoBD **USING** location.
            2.2.2.2.2 **MAP_COORDINATES** (institute → (latitude, longitude)).
         END FOR
      END FOR
   2.3 *(Optionally update geo_coordinates or pull next item from q.)*
   END WHILE

In the information integration process, we use GPS enabled sensors as physical tools to extract geo- coordinates that are consistently connected and interacted with Geo-Bangladesh knowledge-base (as lines 1 and 2). The GPS data in the pattern of {latitude (Lat), longitude (Long)} are used to fetch semantically related neighbouring information in the form of *{division (div) → District* *(dis) → Upazila (upa)}* by eliminating natural language polysemy problems (as line 3). In this stage we expand our extracted information pattern by examining http://www:w3:org/ns/org#site as *{division (div) → district (dis) → upazila (upa) → site}* where site indicates the exact location of the institute available (as line 5 and 7). This advanced information pattern is used for getting further action from Geo-Bangladesh using semantic query language SPARQL to retrieve spatial coordinates to map these localities to increase search ability that shows the usefulness of our application (as line 9). Moreover, our system investigates accessible data sources along with intuitive human knowledge about education services and their respective service qualities of each institution. In our integration process we use an inference engine (portrayed Figure 10, step 9) to explore implicit knowledge that assist people in making their decisions about institutions.

6. KNOWLEDGE REPRESENTATIONS, EXPERIMENT AND REASONING

Our proposed application share resources or instances from Geo-Bangladesh without defining correlative data in other applications. Chaining different web content increases the number of knowledge-intensive task that can be carried out automatically (Ding, 2002). The semantic web supports technology for describing, discovering and accessing web applications that can be formed a chain of interoperable web applications dynamically (Daconta, 2003). Moreover, we use java based semantic web application development framework, Jena for rule-based inference and SPARQL, a query language for semantic data.

Furthermore, we consider the knowledge pieces of our domain to identify each entity of the administrative structure of Bangladesh and derive important information based on isa and *part-of* relation.

Statements of Knowledge Pieces

KB1: Bangladesh isa country

KB2: Bara Horipur isa village and partOf Bara Horipur Mouza

KB3: Bara Horipur isa Mouza and partOf Dakshin Khosbas union

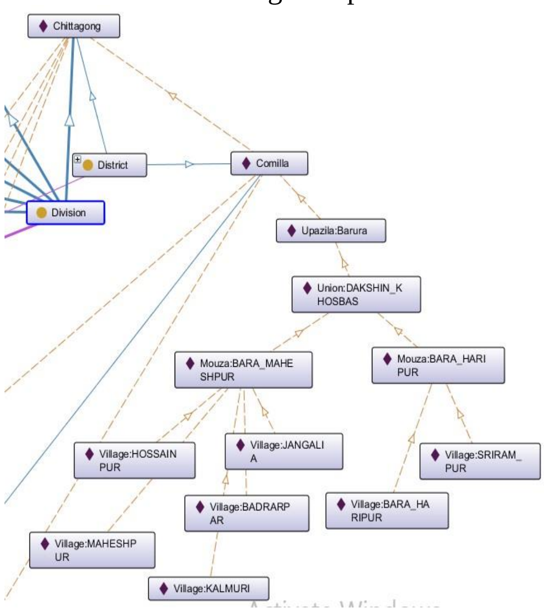
KB4: Dakshin Khosbas isa union and partOf Barura Upazila

KB5: Barura isa Upazila and partOf Cumilla zila/district

KB6: Comilla isa district and partOf Chittagong division

KB7: Chittagong isa division and partOf Bangladesh

Knowledge Graph



The inference system can derive both direct and indirect relation from Geo-Bangladesh, using their relationship between the subject and the object of knowledge pieces. This inference engine also generates many more necessary statements by applying various rules on the available existing triples in our semantic data source.

Therefore,the following algorithm describes the hierarchical data retrieval process used to identify administrative entities closest in distance to a target location, such as the district containing Bara Horipur village, and to perform similar retrieval for other entities of Bangladesh (division, upazila, union, mouza, and moholla) and vice versa. The details procedure is as follows:

Step 1: Based on the part-of relationship, retrieve either the subject or object associated with a given entity.
Step 2: Apply the subject or object obtained from Step 1 to retrieve additional related data.
Step 3: Repeat Steps 1 and 2 iteratively until the specific goal or target entity is reached.

Moreover, we perform a large number of experiments on our machine understandable dataset *Geo-Bangladesh* using semantic search engine, *SPARQL* to explore distinguishing information according to our query on the web. These experimental results based on the relationship of triples show the effectiveness of our research. One of the semantic searches is portrayed in Table 5 to extract location name of the URI, *<http://www:skeim:org#0021105770>*.

Table 5: SPARQL query and it's corresponding output

Semantic Query	Output
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX org: <http://www.w3.org/ns/org#> PREFIX me: <http://www. hhr.org/#> SELECT ? Location Name WHERE {<me#0021105770> org:site ? Location Name }	*Comilla Zilla School*

In addition, the advanced operation on our application examines the location in the form of *{latitude (lat), Longitude (long)}* for each entity from *Geo-Bangladesh*. However, we use the spherical law of cosine to calculate the distances between sensors and the location of each geographical entities by using *GPS*. This formula is given in the *Eq. 2* where Φ_1 is the latitude of first position while Φ_2 is the latitude of the second position. The notation $\Delta\lambda$ express the difference between longitude value of two positions and *R* indicates the radius of the earth that is approximately 6371 kilometers.

$$Distance = acos\,(sin\,\Phi_1.\,Sin\,\Phi_2 + cos\,\Phi_1.\,Cos\,\Phi_2.\,Cos\,\Delta\lambda).\,R \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

To evaluate this numerical calculation, we consider the *(latitude; longitude)* pairs of *Comilla Zilla School*, *Comilla University*, *Comilla Victoria Government College* are *(23:463689; 91:181114)*, *(23:430282391:1361569)* and *(23:4598; 91:1823)* respectively. The distance calculation results obtained by spherical cosine formula from Comilla Zilla School are 6.71 kilometers and 357 meters respectively. This approach performs on the core of distance calculation to retrieve information of specific geo-entities from our dataset along with useful information from the web of data.

Therefore, interoperability was measured through federated queries across Geo-Bangladesh, Bangladeshi-Citizen, and HCN-BD repositories, executed on a 4-core server with 16 GB RAM, with average response times of 0.8 seconds for 100,000-triple queries. As a baseline, we also ran equivalent queries on a relational database with GIS extensions, which required an average of 2.1 seconds. Precision/recall was evaluated on a curated gold-standard dataset of 500 query–result pairs, manually validated by two domain experts to ensure correctness, confirming that our system returned 95% relevant results, outperforming both relational queries and raw GeoSPARQL execution.

7. CONCLUSION

In the present era, machine-understandable geographic data of a country is crucial for minimizing data redundancy and facilitating interoperability across applications that consume such standardized data on the web. Geo-Bangladesh, a geographic semantic data repository for Bangladesh, addresses this need by reducing the creation of duplicate datasets for other applications. The repository's machine-readable web content is linked to existing semantic data sources, such as *Bangladeshi-Citizen and HCN-BD*, to resolve interoperability challenges more efficiently and effectively. Consequently, RDF queries are employed to retrieve and infer precise information, thereby demonstrating the effectiveness of the system. Leveraging Geo-Bangladesh, this research is further extended to map localities in a manner that enhances searchability for the general public, particularly for users with access to mobile technologies.

REFERENCES

[1] A.Geraci, F. K.-d. (1991). *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries.* IEEE Press.

[2] A.Sheth, P. A. (2013). Physical-cyber-social com-puting: An early 21st century approach. *Intelligent Systems, 28*(1), 78–82.

[3] Antoniou, G. a. (2004). *A semantic web primer.* MIT press.

[4] Berners-Lee, T. a. (2001). The semantic web. *Scientific american, 284*(5), 28--37.

[5] Berners-Lee, T. a. (2001). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor.* DIANE Publishing Company.

[6] Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographical Information Science, 12*(4), 299--314.

[7] C.Bizer, T. H. (2011). Linked data: Evolving the web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology, 1*(1), 1–136.

[8] Brodeur, J. a. (2003). Revisiting the concept of geospatial data interoperability within the scope of human communication processes. *Transactions in GIS, 7*(2), 243--265.

[9] Burggraf, D. S. (2006). Geography Markup Language (GML) 3.0. *Data Science Journal, 5*. A. Bizer, T. H.-L. (2009). Linked data-the story so far. *International journal on semantic web and information systems, 5*(3), 1–22.

[10] Chomicki, J. a. (1999). Constraint-based Interoperability of Spatiotemporal Databases. *Geoinformatica, 3*(3), 211 -- 243.

[11] Cohn, A. (1995). The challenge of qualitative spatial reasoning. *ACM Computing Surveys, 27*(3), 323--325.

[12] Daconta, M. a. (2003). *The Semantic Web: a guide to the future of XML, Web services, and knowledge management.* Wiley.

[13] de La Beaujardi`ere, J. (2004). OGC Web Map Service Interface, version 1.3. 0. *Open Geospatial Consortium.*

[14] Ding, Y. a. (2002). The semantic web: yet another hip? *Data & Knowledge Engineering, 41*(2), 205--227.

[15] Ehrig, M. (2006). Ontology alignment: bridging the semantic gap. *4*.

[16] Fonseca, F. a. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS, 6*(3), 231--257.

[17] Frank, A. a. (1991). Language issues for GIS. *Geographical information systems: Principles and applications, 1*, 147--163.

[18] Freitas, A. a. (2012). Querying heterogeneous datasets on the linked data web: Challenges, approaches, and trends. *Internet Computing, IEEE, 16*(1), 24--33.

[19] Goodchild, M. F. (1999). *Interoperating geographic information systems.* Springer.

[20] Gruber, T. a. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International journal of human computer studies, 43*(5), 907--928.

[21] Harvey, F. a. (1999). Semantic interoperability: A central issue for sharing geographic information. *The Annals of Regional Science, 33*(2), 213--232.

[22] Jones, C. a. (2001). Geographical information retrieval with ontologies of place. *Spatial Information Theory,* 322--335.

[23] Jones, C. a. (2001). Geographical information retrieval with ontologies of place. *Spatial Information Theory*, 322--335.

[24] Lepper, H. J. (1991). The benefits of the application of geographical information systems in public and environmental health. *World Health Stat, 44*(3), 160–170.

[25] M. H. H. Rahman, S. C. (2014). Machine understandable information representation of geographic related data to the administrative structure of Bangladesh. *16th International Conference on Computer and Information Technology (ICCIT), 2013* (pp. 236–241). Dhaka: IEEE.

[26] Maedche, A. a. (2001). Ontology learning for the semantic web. *Intelligent Systems, IEEE, 16*(2), 72--79.

[27] Mark, D. a. (1999). Ontology and geographic objects: An empirical study of cognitive categorization. *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*, 747--747.

[28] Martinez-Cruz, C. a. (2012). Ontologies versus relational databases: are they so different? A comparison. *Artificial Intelligence Review, 38*(4), 271--290.

[29] Noy, N. a. (2001). *Ontology development 101: A guide to creating your first ontology.* Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880.

[30] O. O. Alharaki, F. S. (2008). The integration of gps navigator device with vehicles tracking system for rental cars firms. *International journal of Computer Science and Information Security*.

[31] Peng, Z. (2005). A proposed framework for feature-level geospatial data sharing: a case study for transportation network data. *International Journal of Geographical Information Science, 19*(4), 459--481.

[32] S. Chakraborty, M. H. (2014). Linked open data representation of historical heritage of Bangladesh. *16th International Conference on Computer and Information Technology (ICCIT), 2013* (pp. 242–248). Dhaka: IEEE.

[33] Shanmugasundaram, J. a. (1999). *Relational databases for querying XML documents: Limitations and opportunities.* Morgan Kaufmann Publishers Inc.

[34] Silberschatz, A. a. (1997). *Database system concepts* (Vol. 4). McGraw-Hill Hightstown.

[35] Smith, B. a. (2001). Geographical categories: an ontological investigation. *International Journal of Geographical Information Science, 15*(7), 591--612.

[36] Studer, R. a. (1998). Knowledge engineering: principles and methods. *Data & knowledge engineering, 25*(1), 161--197.

[37] Vretanos, P. (2005). Web feature service implementation specification. *Open Geospatial Consortium Specification,* 04--094.

[38] Wang, H. N. (2010). *Rfid major projects and state internet of things.*

[39] Wayne, W. (2007). The good news and the bad news (embedded computing column. *IEEE Computer, vol. 40, no. 11, p. 104, 2007, 40*(11), 104.

[40] Zhang, C. a. (2005). The roles of web feature and web map services in real-time geospatial data sharing for time-critical applications. *Cartography and Geographic Information Science, 32*(4), 269--283.

AUTHORS

Md. Hasan Hafizur Rahman earned his B.Sc. in Computer Science and Engineering in 2009 and his M.S. (Engg.) in 2012, both from the University of Chittagong, Bangladesh. He is currently serving as a faculty member in the Department of Computer Science and Engineering at Comilla University, Bangladesh. His research interests include the Semantic Web, Artificial Intelligence, and Machine Learning, with a focus on developing intelligent, interoperable systems that bridge data integration, knowledge representation, and automated reasoning. He has contributed to research in geospatial knowledge bases, declarative machine learning frameworks, and ontology-driven applications, aiming to advance both theoretical foundations and practical implementations in next-generation computing.

Dr. Md. Hanif Seddiqui received his B.Sc. Eng. degree in Electronic and Computer Science from Shahjalal University of Science and Technology, Sylhet, Bangladesh in 2000 and his M.Eng. and D.Eng. degree in Computer Science from Toyohashi University of Technology, Japan in March 2007 and in March 2010 respectively. He is currently working as a Professor at the Department of Computer Science and Engineering, University of Chittagong, Bangladesh. His current research interest includes Ontology Alignment, Knowledge Engineering, Bioinformatics and Semantic Web Techniques in Information Retrieval and Big Data.