IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Significant Programming and Mathematical Concepts of Semantic Web

N. Malik[1*], S.K. Malik[2]

[1] USIC&T, GGSIP University, New Delhi, India
[2] USIC&T, GGSIP University, New Delhi, India

[*]*Corresponding Author: nikitamalik92@gmail.com, Tel.: +91-9971633991*

Available online at: www.ijcseonline.org

Accepted: 10/Jun/2018, Published: 30/Jun/2018

***Abstract*—** Web is a global space for information where the data is largely present in unstructured or semi-structured form. Semantic Web is an extension of the present web scenario, envisioned to achieve a better collaboration between humans and machines by allowing the computer systems to understand the meaning of the user's data and structuring the information on the web. Achieving this idea of a smart web includes realizing its programming and mathematical concepts as a significant aspect. Semantic web programming entails the representation and development of knowledge concepts and integrating them with applications. Algebra, combinatorics (graph theory), logic and set theory are the pure mathematics' concepts that are majorly used by semantic web that forms its foundation and aids it in improving the similarity search, inferencing capabilities etc. In this paper, significant mathematical and programming concepts of semantic web have been explored, revisited, discussed and presented which may be a useful resource for semantic web researchers.

***Keywords*—**Semantic Web, RDF, Ontology, Programming, Mathematical Concepts.

I. Introduction

Semantic Web was conceptualized by Sir Tim Berners Lee [1]. He shared his vision of a smarter web, where all information is structured meaningfully and concepts are interlinked, forming a Web of Data. The key technologies of semantic web include the RDF (Resource Description Framework), which forms the building block for incorporating semantics in web pages [2]. It is layered on top of the XML (Extensible Markup Language), which encodes documents in a format understandable by both humans and machines. RDF represents data in a graph like model consisting of URIs (Uniform Resource Identifiers) to name the resources (graph nodes) and their relationships (arcs of the graph). The RDFS (RDF Schema) extends the RDF vocabulary by specifying properties and concept taxonomies [3]. Ontologies are the basic form of representing knowledge in semantic web, formalizing it in the form of classes, axioms, relationships among classes and their instances. OWL (Ontology Web Language) captures the ontologies and specifies further constructs over RDFS, using expressiveness of description logic (DL). Further, rules for inferencing can be specified, and query language SPARQL (Simple Protocol and RDF Query Language) is used for querying OWL and RDFS data and accessing RDF data. The Semantic web comprises of its fundamental components, as shown in Figure 1, which contains the programming and mathematical aspects at its core [4]. These enable the preparation of programs for the web of data or the linked data.

Programming the semantic web refers to representation of knowledge and development of applications and ontologies, including decisions about the programming language and environment for creation of schemas, statements, query specification and execution, specification and inferencing of rules [5]. However, these environments and languages lack some basic mechanisms of abstraction and are not very well-integrated. Also, mismatch between the programming language semantics and semantics of the representation language is a problem encountered unavoidably while programming the semantic web data, and there is a need of approaches to address these [6] [7].

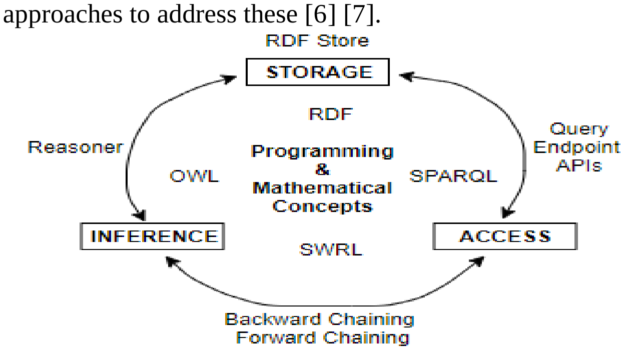

Figure 1. *Semantic Web Components (Source: Hebeler et. al, 2011)*

Mathematical knowledge involves logical language formula, theories and axioms, formalizations, symbolic notation and use of natural language in documents, proofs, metadata etc. This mathematical domain knowledge and structures can be represented using metadata vocabularies or ontologies in semantic web. Semantic web technologies can be applied to mathematical knowledge management along two lines- bettering the retrieval of information and providing the users access to automated reasoning, and web services for self-descriptive interfaces to web in order to facilitate problem solving. Having mathematical foundations represented in semantic web has multiple advantages like improved similarity searching, reasoning, enabling new applications such as statistics, predictions in enterprises, reproduction of experiments etc. However, mathematical knowledge has still not been completely integrated as proper ontologies in semantic web and needs future research [8]. Therefore, programming and mathematical concerns are foundation for research in semantic web and its technologies.

The paper is organized as follows- Section II contains the introduction of programming aspects of semantic web, along with the programming impacts on data, the major programming components, the operations for managing semantic web data and a Jena framework as an example for carrying out programming on semantic web data. Section III discusses the mathematical concerns of semantic web, including the mathematical representation of key semantic web technologies and centrality measures for analysis of social network. Finally, Section IV concludes the research work with future scope in this direction.

II. Semantic Web Programming Concepts

Programming the semantic web applications commonly follows the approach of treating RDF as a knowledge representation system which is an application of the implementation language underneath, such that the application logic is mixed with the exploitation of the formalisms of knowledge representation [7].

Primarily there are two semantic web programming aspects- representation of knowledge and integration of applications, as shown in Figure 2 [4].

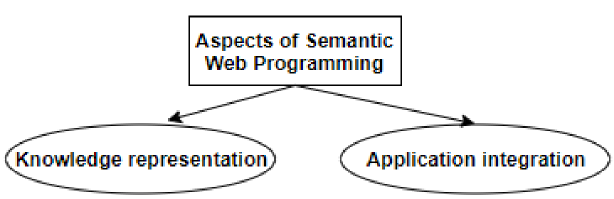

Figure 2. Two Aspects of Semantic Web Programming

RDF, ontologies (OWL), queries, reasoning and rules form the foundation for the first aspect that focuses on the representation of knowledge and its manipulation using various technologies and tools.

The second aspect deals with the integration of the knowledge base with the applications which can operate upon this knowledge. The semantic web applications may be built by taking a basic semantic web framework with similar components and fundamental principles and by considering its interactions with traditional applications to integrate the two. Common tasks performed using these applications are of integrating data from disparate knowledge models, exposing data in the form of RDF models, sharing and reusing semantic information throughout the web [4]. An application must acclimate to Semantic web's expectations and impacts in order to take its full advantage as discussed below:

A. *Determining Impacts on Programming*

The programming impacts can potentially determine how a semantic web application can be designed and programmed, and can be categorized as [4]:

- Semantic data- The meaning should be placed by the semantic web applications within the data directly instead of placing it within programming instructions or leaving it for interpretation by users.
- Data sharing or integration- The rich resources of information should be accessed by the semantic web applications and taking advantage of the various data sources, information should be shared throughout the web by them, when appropriate.
- Data centric web- Taking data as the key, data should be placed at their centre by the semantic web applications.
- Dynamic data- Run-time, dynamic changes should be enabled by the semantic web applications on the information's contents and structure [4].

B. *Major Programming Components of Semantic Web*

A Semantic web application comprises of numerous discrete components, as shown in Figure 3, which lie under the categories of either the major components or the associated tools for semantic web [4].

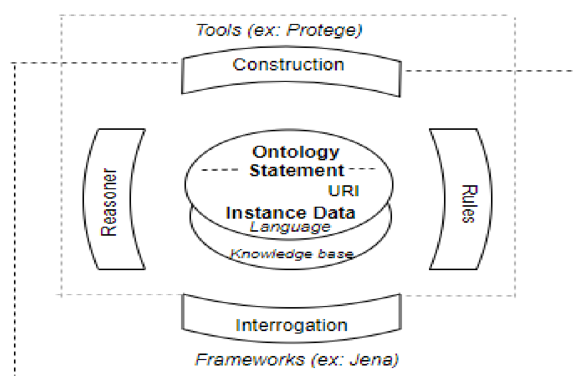

Figure 3. Major Programming Concepts of Semantic Web
(Source: Hebeler et. al, 2011)

The core components which make up the semantic web include [4]:

- Statement- Comprises elements of the form of a typical triple of subject-predicate-object, and define the information structure, its limits and specific instances. Statements are related to one another to achieve flexible expressions and form the web data.
- URI- For the items comprised in a statement, it provides a unique name across the whole web, and eliminates any naming conflicts.
- Language- Comprises a collection of keywords for providing instructions to the various tools of semantic web, and offers varying degrees of semantic complexity and expressiveness.
- Ontology- Defines concepts, relationships, and constraints, and is useful for forming domain models.
- Instance Data- The statements which instead of the general concept contain specific instances' information, are not necessarily bound to ontology, and forming the bulk of semantic web [4].

Frameworks with various components and tools may be required for exercising the semantic web. Four types of tools are there which evolve, interrogate, manipulate and enrich the semantic web [4]:

1. Construction tools: Allow the building and evolution of semantic web through the formation or import of statements for the ontology and instances.
2. Interrogation tools: Explore the semantic web to return a requested response, using different methods like search, query language or navigation through graphs.
3. Reasoners: Add inferences to the semantic web, which provide logical additions for offering classification and realization. Reasoners mostly plug into other frameworks and tools.
4. Rules engines: Expand the semantic web by supporting inferences beyond DL deductions. They are driven by rules and add a significant aspect to the knowledge paradigm [4].

These four tools [4] are packaged into an integrated suite using semantic web frameworks. These, and other integrated development environments (IDEs) and APIs aid in the programming of semantic web.

C. Managing Data during Semantic Web Programming

Beside the creation of semantic web data, other important areas of managing and programming the semantic web data also need to be considered. In an event driven and multiuser environment, the following operations highlight the realities of programming the semantic web data [4]:

- Gathering information about the semantic web data or model using different methods to help develop useful insights.
- Getting event notifications regarding any changes to the semantic web data, and employing efficient reactions for critical programming events.
- Dealing with concurrent operations from threads within the same application or from multiple applications/users on the semantic web data.
- Configuring and customizing the implementations of Jena framework for providing flexibility, which involves creation of customized graph-based object followed by the creation of the model based on the object.
- Serializing the semantic web data to enable the model's transmission through different means.

The above four operations [4] are quite significant for semantic web data programming.

D. Programming with Jena

Apache Jena *(www.jena.apache.org)* is an open source framework for semantic web application building, implemented in Java. It provides a set of interacting APIs (Application Programming Interface) for processing the RDF data, an inference engine for the purpose of reasoning, and many storage strategies [2]. Through its Java variables and classes, a consistent treatment of semantic web is maintained by the Jena semantic web framework, where the semantic web constructs and statements are translated into useful programming artifacts like Java objects, classes, methods etc. by these Java-based abstractions. Various classes are offered by Java for converting ontologies into classes and for typical constructs of semantic web, of which some of the major classes employed by the Jena framework are [4]:

- Resource- this class represents, just like the RDF triple resource, a statement's elements of subject, predicate or objects.
- Graph- this method allows for semantic web data maintenance through basic operations of add, find, delete etc., and its interface provides various kinds of storage mechanisms' instantiation.
- Reasoner- consists of the reasoning processing which could be internal, i.e. framework's own capabilities, or external, i.e. through third-party reasoners accessing the knowledge base [4].

Besides the above classes [4], there are other classes also which are used in Jena for programming the semantic web.

A sample RDF hello-world application in Jena can be simply written as:


```
import com.hp.hpl.jena.datatypes.xsd.XSDDatatype;
import com.hp.hpl.jena.rdf.model.Model;
import com.hp.hpl.jena.rdf.model.ModelFactory;
import com.hp.hpl.jena.rdf.model.Property;
import com.hp.hpl.jena.rdf.model.Resource ;
public class HelloWorldRDF {
        public static void main (String[] args) {
        Model m=ModelFactory.createDefaultModel();
        String NS=http://test.com/example/;
        Resource res=m.createResource(NS+"res");
        Property  prop=m.createProperty(NS+"prop");
        res.addProperty(prop, "hello world", XSDDatatype.XSDstring);
        m.write(System.out, "Turtle");
        }
}
```

III. Mathematical Concerns of Semantic Web

Integration of mathematical sources with associated metadata (annotations) opens up new possibilities which allow better inferencing, more powerful search, similarity search capabilities, and possibility of algebraic manipulation etc. [9]. However, various fields of mathematics still haven't been realized as proper ontologies in semantic web [8]. So, there is a need to identify significant aspects of mathematical knowledge which can be integrated with the mainstream semantic web technologies for the necessary achievement [9].

A. Mathematical Representation of Key Technologies

Semantic web's RDF and ontologies can be formalized by defining as a collection of entities with relationship between them, on which mathematical operations can be performed [10]. The key standard technologies which can cultivate the incorporation of mathematical representation into the semantic web are XML, RDF, and OWL [11] besides SPARQL.

- XML- On the web, representations of XML for mathematics already exist like prominent ones are MathML (Mathematical Markup Language) [11] and OpenMath. MathML, an XML language, was initially used in HTML pages for embedding mathematical formula. Additionally, it includes both semantics-oriented sublanguage-Content MathML, along with a presentation-oriented sublanguage- Presentation MathML, allowing a fine-grained mix of both markups, i.e. parallel markup [11]. OpenMath, though initially used for facilitating exchange of data among computer algebra systems, has also been closely related to Content MathML [8]. The mathematical models are represented as Content MathML formulae which are not directly accessible via URIs as they are deeply nested into XML files [9].

- RDF- To allow the applications dealing with mathematical techniques to be able to carry across, using RDF, their knowledge to semantic web, and conversely for RDF representations to be mapped back, suitable representations to and from RDF should be specified [11]. XML-based standard languages for exchanging mathematical knowledge like MathML and OpenMath can be integrated completely with RDF representations for contributing towards the incorporation of prevailing mathematical knowledge in the Web of Data [8]. A function *f(1,5)* for example, using MathML representation, can be expressed as [11]:

```
<apply>
    <csymbolencoding="text"

 definitionURL="http://www.mathsw.org/scalarplus">
              f
          </csymbol>
    <cn> 1 </cn>
    <cn> 5 </cn>
</apply>
```

The RDF representation of the above may be as follows where each row represents the triple representation of RDF graph [11]:

```
: 1<http://www.w3.org/TR/MathML2#apply> : 2
: 1 <http://www.w3.org/TR/MathML2#csymbol>"f"
: 1<http://www.w3.org/TR/MathML2#definitionURL>
                  <http://www.mathsw.org/scalarplus>
: 1<http://www.w3.org/TR/MathML2#encoding> "text"
: 2 <http://www.w3.org/1999/02/22-rdf-syntax-ns#:1> "1"
: 2 <http://www.w3.org/1999/02/22-rdf-syntax-ns#:2> "5"
```

- The knowledge contained in RDF can be formalized as Ontologies which may be mathematically defined as consisting of various entities and the relationships between them, represented through recursively defined classes and relationships respectively:

  An ontology O is symbolized as a tuple $\langle C, P, CN, I, IN \rangle$ where $C = \cup_{n \geq 0} C_n$ represents a set of classes, $P = \cup_{n \geq 0} P_n$ shows the properties and the set of instances is denoted by I [10].

- OWL- It may be difficult to associate mathematical objects to specific categories if the appropriate metadata is lacking. For this to be addressed, OWL standard has been developed by W3C- OWL's flexibility allows complete mathematical hierarchies to be formed, data to be ascribed with ontological meaning, and powerful mechanisms to categorize data to facilitate proper handling of web data. Such mathematical ontologies offer support in the tasks of looking out for mathematics, delivering similarity searches, and avoiding the inherent challenges related to extracting this information from the object's structure or semantics [11].

  An OWL ontology, OntoMathPRO [12], was developed for representing concepts of mathematical knowledge across an array of fields, and it reasoned that it can play a central role in the mappings of math-aware datasets, i.e. their integration in the Web of Data [12].

B. Centrality Measures

An example of usage of mathematical notations and formula in semantic web is in social networks where various centrality measures, as shown in Table 1, are used for identifying the most prominent node(s) in the network. The social networks can be visualized and analyzed using some tools such as Visone, Cytoscape, etc. [2]. Here, using SocNetV, taking an illustration of a social network having 5 nodes, as shown in Figure 4, the matrix representation of links between the nodes has been depicted in Figure 5 and Figure 6 shows that 'Nikita' node is the most prominent node of the network, based on the calculation of degree centrality of each node.

Table 1. Centrality Measures for Social Networks in Semantic Web for network analysis and visualization *(https://en.wikipedia.org/wiki/Centrality)*

Degree Centrality	**Closeness Centrality**
$C(v) = \frac{\deg(v)}{n-1}$ *where graph G: (V,E) with n vertices and deg(v) is degree of v*	$C(v) = \frac{1}{\sum_{t \in V} d(v,t)}$ *where d(v, t) is distance from v to t node*
Eigen vector Centrality	**Betweenness Centrality**
$x_v = \frac{1}{\lambda}\sum_{t \in M(v)} x_t$ *where M(v) is set of neighbours of vertex v and λ is a constant*	$C(v) = \sum_{s \neq v \neq t} \frac{\sigma(s,t\|v)}{\sigma(s,t)}$ *where σ(s, t|v) is the total number of shortest paths from node s to node t that pass through vertex v*
Pagerank Centrality	**Eccentricity Centrality**
$\pi_i = \alpha \sum_{j=1}^{n} A_{ji} \frac{\pi_j}{out(j)} + \beta$ *where V= {1,…,n} are the nodes/pages and (i , j) ε E i.e. page i points to point j; π_i is the pagerank of page i and α , β > 0 and out(j) is outdegree of node j*	$EC(x) = 1/\max\{d(x,y) : y \in M\}$ *where d(x, y) is length of shortest path from x to y and M is set of all members of social network*

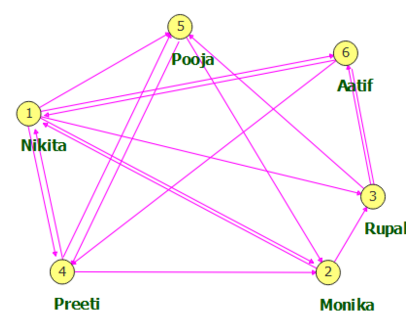

Figure 4. Example of a Social Network (Graph)

Actor/Actor	1	2	3	4	5	6
1	0	1	1	1	1	1
2	1	0	1	0	0	0
3	0	0	0	0	1	1
4	1	1	0	0	1	0
5	0	1	0	1	0	0
6	1	0	1	1	0	0

Figure 5. Matrix Representation of Relationships between Graph nodes

Actors: 6

In undirected networks, the DC index is the sum of edges attached to a node u.
In directed networks, the index is the sum of outbound arcs from node u to all adjacent nodes (also called "outDegree Centrality").
If the network is weighted, the DC score is the sum of weights of outbound edges from node u to all adjacent nodes.
Note: To compute inDegree Centrality, use the Degree Prestige measure.
DC' is the standardized index (DC divided by N-1 (non-valued nets) or by sumDC (valued nets).

DC range: 0 ≤ DC ≤ 5

DC' range: 0 ≤ DC' ≤ 1

Node	Label	DC	DC'	%DC'
1	Nikita	5.000	1.000	100.000
2	Monika	2.000	0.400	40.000
3	Rupal	2.000	0.400	40.000
4	Preeti	3.000	0.600	60.000
5	Pooja	2.000	0.400	40.000
6	Aatif	3.000	0.600	60.000

DC Sum = 17.000

Max DC' = 1.000 (node 1)
Min DC' = 0.400 (node 2)
DC' classes = 3

DC' Sum = 3.400
DC' Mean = 0.567
DC' Variance = 0.046

GROUP DEGREE CENTRALISATION (GDC)

GDC = 0.520

GDC range: 0 ≤ GDC ≤ 1

GDC = 0, when all out-degrees are equal (i.e. regular lattice).
GDC = 1, when one node completely dominates or overshadows the other nodes.

Figure 6. Results for Degree Centrality of Graph Nodes which also shows 'Nikita' as most prominent node.

IV. Conclusion and Future Scope

This paper highlights the programming aspects of semantic web and the mathematical concepts for enriching the semantic web technologies. The following inferences may be drawn:

- The nature of the data published on semantic web significantly differs from the kind of data that users deal with in case of established approaches of databases; which poses a challenge for programming with semantic web data.
- There is a need of approaches that completely consider the characteristics of semantic web data and reduce the resisting mismatches between data engineering and the programming approaches for the semantic web to reach its full potential.
- Mathematical foundations in semantic web have a significant role to play and ontologies and metadata vocabularies representing domain knowledge and structures of mathematical concepts like algebra, combinatorics (graph theory), logic and set theory have been integrated over time. The need is to identify various significant subsets of mathematics for semantic structuring to obtain a mathematical semantic web along with various deep-rooted programming concerns.
- Programming and mathematical concepts are the foundation and most significant for embedding semantics on the web towards the goal of next generation smart web, and a lot needs to be explored towards it. There is a further need to explore in detail the programming and mathematical concerns of semantic web in future.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila, *"The semantic web"*, Scientific American, Vol. 284, Issue 5, pp. 28-37, 2001.

[2] S. K. Malik, and SAM Rizvi, *"An intelligent web framework based on ontology design"*, Ph.D Thesis, Guru Gobind Singh Indraprastha University, 2014.

[3] G. Antoniou, & F. Van Harmelen, *"A semantic web primer"*, MIT press, 2004.

[4] J. Hebeler, M. Fisher, R. Blace, and A. Perez-Lopez, *"Semantic web programming"*, John Wiley & Sons, 2009.

[5] B. G. Humm, & A. Korobov, *"Introducing layers of abstraction to semantic web programming"*, In OTM Confederated International Conferences On the Move to Meaningful Internet Systems, Springer, pp. 412-423, 2011.

[6] S. Staab, S. Scheglmann, M. Leinberger & T. Gottron, *"Programming the Semantic Web"*, In European Semantic Web Conference, Springer, pp. 1-5, 2014.

[7] O. Lassila, *"Programming semantic web applications: A synthesis of knowledge representation and semi-structured data"*, Doctoral Dissertation, Nokia Research Center, Cambridge, 2007.

[8] C. Lange, *"Ontologies and languages for representing mathematical knowledge on the semantic web"*, Semantic Web, Vol. 4, Issue 2, pp. 119-158, 2013.

[9] C. Lange, *"Integrating Mathematics into the Web of Data"*, Linked Data in the Future Internet, Future Internet Assembly, pp. 12-16, 2010.

[10] S. Kaushik, et. al, *"An algebra for composing ontologies"*, In FOIS, Vol. 150, pp. 265-276, 2006.

[11] M. Marchiori, *"The mathematical semantic web"*, In International Conference on Mathematical Knowledge Management, Springer, pp. 216-223, 2016.

[12] O.A. Nevzorova, N. Zhiltsov, A. Kirillovich, & E. Lipachev, *"OntoMath PRO ontology: a linked data hub for mathematics"*, In International Conference on Knowledge Engineering and the Semantic Web, Springer, pp. 105-119, 2014.

Authors' Profile

Ms. Nikita Malik completed her Masters in Technology (Information Security) from Ambedkar Institute of Advanced Communication Technologies and Research, GGSIP University, Delhi, India *(Gold Medalist).* She is currently pursuing Ph.D. from University School of Information, Communication and Technology, GGSIP University. She has published few research papers in IEEE/Springer international conferences/journals of repute. Her main research interest focuses on various Semantic Web concerns and technologies, including Information Security aspects.

Dr. Sanjay Kumar Malik completed his Ph.D. in the area of "Semantic Web" from USIC&T, GGSIP University, Delhi. He is currently working as Associate Professor in University School of Information, Communication and Technology, GGSIP University. He has more than 18 years of industry and academic experience in India and abroad (Dubai and USA). His areas of research interest are Semantic Web and various Web Technologies. He has several research papers in reputed international Conferences/ Journals (IEEE/ Springer). He has been session chair in several international conferences and technical committee member of several conferences. He has been honoured with third best researcher award, 2011 by GGSIP University for his contributions in research.