

AN IMPERATIVE FOCUS ON SEMANTIC WEB PRINCIPLES, LOGICS AND ITS APPLICATION

N. Senthil Kumar ¹ and Dr. Dinakaran M ²

School of Information Technology & Engineering
VIT University, Vellore, Tamil Nadu, India

ABSTRACT

The World Wide Web is booming and radically vibrant due to the well established standards and widely accountable framework which guarantees the interoperability at various levels of the application and the society as a whole. So far, the web has been functioning at the random rate on the basis of the human intervention and some manual processing but the next generation web which the researchers called semantic web, edging for automatic processing and machine-level understanding. The well set notion, Semantic Web would be turn possible if only there exists the further levels of interoperability prevails among the applications and networks. In achieving this interoperability and greater functionality among the applications, the W3C standardization has already released the well defined standards such as RDF/RDF Schema and OWL. Using XML as a tool for semantic interoperability has not achieved anything effective and failed to bring the interconnection at the larger level. This leads to the further inclusion of inference layer at the top of the web architecture and its paves the way for proposing the common design for encoding the ontology representation languages in the data models such as RDF/RDFS. In this research article, we have given the clear implication of semantic web research roots and its ontological background process which may help to augment the sheer understanding of named entities in the web.

KEYWORDS:

Semantic Web, Named Entity Recognition, Linked Data, RDF, OWL.

1. INTRODUCTION

Semantic web has been enriching the standards of web through its formal semantics (so called the metadata and its knowledge) to uphold the efficiency of easy access of data and its governance. Though the acquisition of metadata from the diverse source of critical mass is under the bottleneck and still at the edge of uproot, the semantic enabled web can embark the new model to synchronize this pitfall and formulate the design to constitute the shortcomings faced in this Web 2.0. The semantic web vision is envisaged about the automation of web content and its resource, annotating the web tags and appropriate infrastructure to suit the contemporary design prevailed in the modern web. To make this to happen, the detailed research has to be carried out further and researched. Since the semantic web vision has envisaged in this direction, there are still some cases, where the manual acquisition and gathering of metadata is a bigger challenge and be the thwart for explicit semantic accumulations.

Besides, the semantic web can be seem overwhelmed if there exist the proper rule of using various forms of redundancy, including the much required external or knowledge base, unifying the entities with appropriate formal descriptions , thereby yield the rich semantics to the web connectivity. Further, the looming anticipation over the semantic web can be easily achieved and manipulated if the underneath tasks are fulfilled: (i) Annotating the named entities in text and hyperlink them. (ii) Index the documents/pages with respect to the identified named entities.

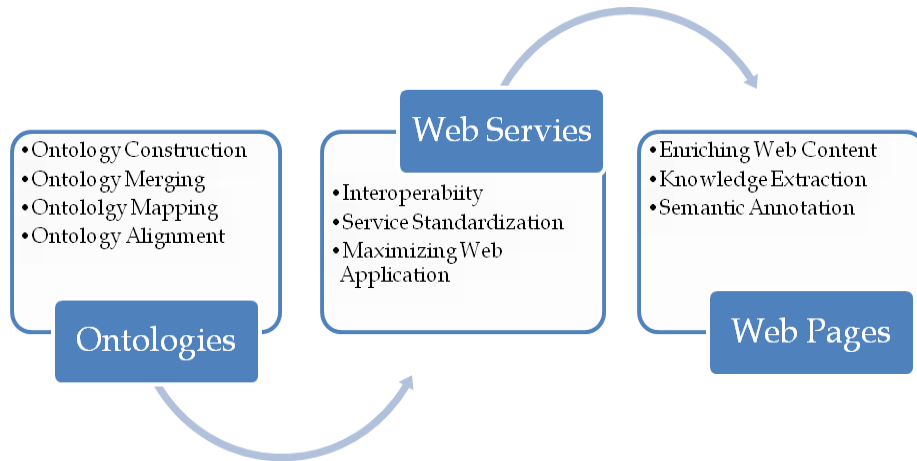


Fig 1: Semantic web major requirements and tasks

The modern search engines are sufficient to provide the details to the user about the information that was asked. It has sharply gathered and processed the web pages manually/automatically from the heterogeneous search engines and hit back the lists of results to view. But the problem lies here is that the details about the single object or entity can be found in millions of sources and picking the right information to the user purview is the much complicated task which the modern search engine wont do automatically and manually either. There are certain rare instances where the search engines would crawl the desired web pages but the content of the web page would have very limited information and to look for the detail description, again it has to crawl for the right mix of combination among the chosen web pages. The web is a complex layout because the same information may appear in more than one web page with different style or patterns. But in many cases, this redundancy too is helpful in discovering the statistical pattern and association inherited among the web pages.

The exploitation of information redundancy to information extraction can be done in many ways with suitable algorithms. The most appropriate algorithm for this kind is the StatSnowball approach which often iteratively identifies the frequent patterns of information in a bootstrapping fashion and infers some new facts or patterns.

1.1 Feasibility Approach

The biggest challenge in semantic web is name disambiguation since it deals with entity information integration. The name disambiguation problem is too cumbersome to highlight the differences prevails among the entity sets and it would be very tough to disambiguate the entities. Without the human or manual intervention, it's not possible to generate automated disambiguation decisions. The algorithm iKnoweb [1] has proposed a prudent step in adding people to mine the knowledge and iteratively identify the name disambiguation problems.

The term Web Entity is the concept of processing the web for the pieces of information which are to be effectively crawled, accumulated, indexed with suitable pointers and ranked them based on the user preferences. Generally, the web entities are the token of concepts like people, organization, time, categories, geographic locations, organizational products, references etc., These are the recognizable concepts which bring in the relevance to the application domain.

In order to extract the information from the web, there should be a suitable format (i.e., a general framework of design that governs the similar process of execution) which can make the extraction process simple and drastically accelerates the accuracy of entity extraction. To support this demand, there comes the new way called, “tag-tree” which is generally used to represent any web page. But in this natural representation also, we have the lack of content structure but it projecting on presentation structure of the web page which is always not suitable to get any concepts or semantics from the web pages. Most of the web page authors have always preferred to apply the different styles to their web pages to attract more web users but this tag-tree is unable to fulfill its purpose.

The entity extraction and integration will attain its level satisfaction only if there exists the common match correlates between the entity details in the web page and the key attributes found in the knowledge base(i.e., the information on the web page and the knowledge bases refers about the same entity). This help to mark the details of other attributes in the web pages to the domain select and label them accordingly to the rest of the elements in the web pages.

To substantiate the prevalence of semantic web turning realistic, there are many leading companies like Oracle, Amazon, Adobe, Vodafone, Google and much more has witnessed the surface binding with semantic web technologies. These organizations have attempted their level towards semantic enabled web processing and started implementing its smarter work in semantic search. For instance [15], the leading database giant Oracle has widely surfacing its data organization with RDF management deck and integrating its enterprise application data well with content integration. It has further extended its undeterred support to OWL for smoother data integration with less error prone task. To witness this growth, Vodafone has extensively followed RDF structure to identify or search its mobile operation like searching ringtones, games and pictures in their web site with this new paradigm shift technology. In their recent survey, they have acknowledged the fact that they have gained a lot in utilizing this semantic base technology for its organization. There are many more success stories to be highlighted.

1.2 The Principle of Semantic Web Vision

The prime goal of semantic web is the sheer realization of automated processing of entities in the web which is proportionally related to the challenging task of entity matching problem. In accomplishing this task easy, we can find many ready tools [3] [5] that variably recognize the named entities among the diverse range of web sources [identifying all the possible named entities like name, title, location, price, tag etc]. The underlying problem pertaining to the existence of named entities are that the same entity can be referred with different name in some other entity source and the reference would become the complex and more complicated to unifying that similarity.

The task of eliminating the entity matching problem [2] can be tried in many ways and still the algorithms which attempted to eradicate the differences were gone vain. The most popular attempt was similarity functions (string similarity, vector based similarity, phrase based similarity, concept based similarity) measures which observes the amount of relevance sustains between two entities but it lacks to capture the desired features of the entities.

Hence the task of identifying the named entities in the unstructured or semi-structured text in the web is the critical task of any web analyst. Entity extraction and integration (entity matching) can be roughly achieved through the proper investigation of ontologies which acts as a catalyst in analyzing the actual match of every entity in its sub domain and gives the degree of confidence to bring the similarity to the common threshold. In order to make this task complete, we need to create a huge collection of entities with diverse fields and from heterogeneous categories to accumulate.

The following are the difficulties faced in assessing the similarity pertains between entities in the web documents:

- The similarity can be measured based on some of the desired features of the entities only but the actual features of the entities can not be specified and still unaware what comparison should take with which techniques.
- Besides, the entities most often contains incomplete, imprecise information, some entities can contain irrelevant information and few entities hold missing data. Due this, the similarity checks turn inaccurate.
- The output of the similarity technique can be vague and varying in accuracy due to the different inputs on the same entity.

1.3 Application of Semantic Web

The basic support of semantic web technology is URIs which identifies distinct resources in the web together with XML and namespaces. From these logics, we can easily generate the RDF templates that can be used to represent the resources about anything which are very relevant towards it. There are many technologies like Topic Maps, preweb have contributed a lot to the Semantic Web application. The most popular and welcomed application of semantic web still today is its FOAF (Friend of a Friend). The FOAF can very easily denote the relationships exist among people and other form of terms in RDF. The next application to the semantic web wallet is PiggyBank (a new plug-in tool for Firefox Browser). Almost around 75000 xml feeds had been scrutinized and then categorized based on the topics using DMOZ schema and interlinked them to mine these feeds for much valued information. The plug-in tool PiggyBank would extracts the web scripts into its equivalent RDF information and loads the filtered information on the database. This database is well used for processing the information independently based on with some original source, let say by using Google Maps to effectively display the information. The overwhelming support poured into semantic web technology because of the simple fact that it is a machine-understandable architecture. But the real fact is that the concept of machine-understandable application is not an artificial kind of intelligence or not carrying out any magical trick to remedy the human fumbles in the web structure. It is just an indicator to mark the problem by performing well defined operation with the established data. The semantic web has made the effort little easier to create information in the web which should be more comprehensible and plausible. It may seem easier to define RDF at one level but as a whole, the entire web structure is very complex and complicated. The semantic web is the next generation web that contains actionable information (i.e. information can be well derived from data and processed using semantic theory).

2. THE ROLE OF SEMANTIC WEB AND ONTOLOGIES

The World Wide Web has serving its utmost task in proving the desired resources for human consumption and further level of intervention to attain the intended information. Despite the

resource available on the web [9] can be made machine readable, but not turn the same to for machine inference [literally, it's not machine understandable]. To overcome this shortcomings, there is a reasonable argument which support this lagging is called the metadata [it's the data about data or data which describing the details of the web sources]. Again to strengthen this effort, the new revolution has arisen in the name of new framework called "Resource Description Framework". It was the first building stone to process the metadata sources and given high interoperability in connecting the resources from heterogeneous framework. Moreover, it has the proper reach ability among the application to process the information both machine understandable and enabling the automation among the chosen web documents.

The basic building block of the RDF [10][14] is a detailed design for denoting the named properties and its property values. The primary data model of the RDF included three object types:

Resource: It can be of anything but have a URI. It refers to the web pages or any individual elements in the web documents. Any resource can be easily identifies by its unique resource identifier and every resource identifier is a URI and an optional anchor id.

Property: A property is also a resource but it must have a name. Most of the cases, the proper name is the basic representation of the object.

Statement: It is the blend of all (a resource, a property, and its value). This can be implicitly known as subject, predicate and object of any statement.

2.1 Inclusive changes in Semantic Web Architecture

The architecture of any maturing technology will not be remaining static or unchanged for long. The sophisticated inclusion or drastic modification can always incur in the paradigm shift of any parable. In 2006, the semantic web inventor Tim Berners Lee has suggested to his layer cake diagram which he has given it for semantic web. But later, some new capabilities and features had been added and framed to boast its longevity. Here are some of the following:

2.1.1 Rules and Inference Systems:

The alternate perspective of rule specification and deep inference methods were designed and developed for easier conclusion. Rules Interchange Format (RIF) is a language which they have called it and it has been widely representing the sticking rules on the web and linking different rule-based system together as well. Besides, it has the sophisticated capabilities of capture casual, unrealistic and temporal knowledge.

2.1.2 Database Support for RDF:

When the RDF data increases exponentially, the demand to its logical storage has arisen to store and query the data effectively. Oracle has now paved the bigger role in providing the database support for both RDF and OWL but it has not still been developed with inference capabilities. And also many open source products had come up with stunning development such as Jena, Susane, 3Store5.

2.1.3 RDF Extraction:

The new language called Gleaning Resource Descriptions from Dialects of Languages (GRDDL) was developed to identify when an XML document contains any data which is compatible with

RDF can be extracted and transformed. And also it has given a plausible tick to convert the XML data available in the web to RDF very clearly.

2.1.4 Ontology Language Developments:

In 2004, OWL was developed with some specific standards. But the research has begun in 2007 to provide the new version and definition to OWL 2 which dominates easier query handling and efficient reasoning capabilities to scale large datasets.

2.2 Few Glitches and Challenges in Information Integration

The semantic web has widely acknowledged the language, RDF/S but it has not made its set up on the helm of its WWW architecture. Though RDF [4] [7] has prevalently used the syntax of XML, it did not convey the same meaning that the XML has provided to the system. Therefore, it has witnessed the wide spread semantic discontinuity and cleavage in the semantic web foundation. When we started observing the differences between them, we gathered the subsequent details.

(i) XML is a tree like model where no edges possess no labels and the outgoing edges from every node holds the total order (i.e., sum of the nodes).

(ii) RDF/S is a directed graph model and every edge in the graph carries the labels but they are all unordered. Each node in the graph has been differentiated via identifiers or literals (sometimes, text). And also the edges too have proper labeling with appropriate identifiers.

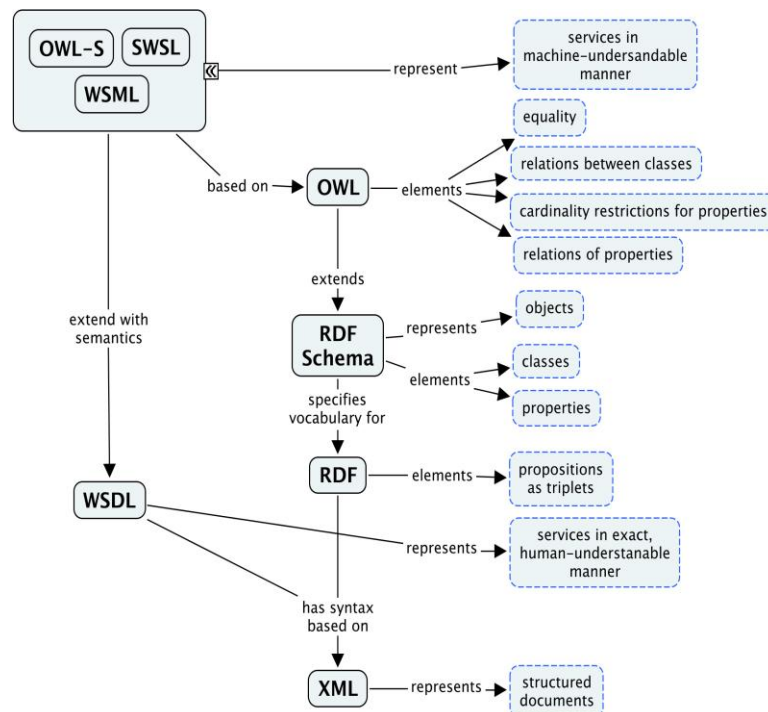


Fig 2: Semantic Web Ontology Representation

The underlying goal here is to lay down the solid foundation on common semantic which must be suitable for the World Wide Web and the modern Semantic Web, exchanging the stilling ideas from these two giants. In doing so, there should be no compulsion in changing the XML

functions, but incorporate some trivial changes to RDF that can support this sort of interoperability and flexibility.

If we look at the closer context of the conventional or the current World Wide Web, there would not be any difference pertaining to the semantic web technologies expect in the single contrast, that is the RDF and RDFS side(see Fig 2). We can't differentiate anything from semantic web expect XML and XML Schema. Using the RDF data model, we can derive the meaning from the context and generate the core language of the semantic web to make a graph. This will further enhance the capabilities of the language to fit in the appropriate ontology to its respective level and build it in such a way to identify the resource with clear distinction.

2.3 Matching at different Semantic Levels using Ontologies

The modern web is congested with huge data and it has become the critical process to the system to discriminate the data correctly and most of the case, it was imprecise. The human consumption was highly sought to bring in the exact differences exist among the data and takes the user level understanding and prediction at large scale. Though there were intelligent agent's works dynamically to scrutinize the resource and information together but it too sought the user intervention to predict or categorize the information. To make this process automotive and inherent, there comes the new change called the semantic web. The Ontologies [5] will be embarked as the new paradigm shift to the current web and evolved with flexible standard which suits to the modern requirement. Since the modern web has been hugely relied upon the social networks, data integration has been sole concern for the semantic web. Software agents [8] are the open door for the silos to expand and integrate over the web based applications. These software agents have the potential to penetrate the entire web to its breadth and immense deep to track the much needed information which is relevant to the users. By means of these efforts, the user queries can be easily answered by delving deep into the web and build the ontology structure to support the confidence of the user query with respect to the retrieved web data. This sort of analysis can mark the omission of keyword based searches.

With the advent and high usage of ontology, the future web technology has become crucial and interoperable among the web applications and resources. Through the proper use of ontology, there exist the common sharing and mutual linking among the chosen domains which can be easily facilitated among the applications.

The development of RDF has given some calculative impacts by the following uses:

- The metadata details collaborate with web resources in providing the information like content rating, descriptions, privacy preferences, etc.
- The capability of processing information out of reach and interoperable among the applications and web resources.
- Helped in deriving new information from integrating data from heterogeneous applications and build the common and coherent consensus.
- The information in the web could be processed automatically by intelligent agents.
- The web application cannot be independent but they are all cohesive and coupling among themselves to infer the new patterns(e.g., scheduling activities, annotation of web resources, describing organizational processes, etc.,)

RDF has owned the data model to connect information in a minimally constraining but flexible way. The isolated applications can also use this design where the independently designed model can be absolutely direct and easy processing but in the context of generality, the RDF only yields

greater meaning to share among the applications. The RDF design has been modeled using the following states:

- It is a simple but most expressive data model at the large scale.
- Governs the global identifiers of all web resources and drastically reduce the ambiguity.
- It has the generic capability to handle the incomplete or mismatch information available in the web.
- Offers the graph like structure in order to merge data and integrate.
- Effectively uses the URIs to describe the web resources and majorly applied for identifying web addresses.
- Compatible for all kind of users to generate statements about any web resources.

The complete utilization of RDF logic in the web pruning turns the incompleteness to absolute preciseness and clearly distinguished the web resources from each other interlinked information. It has also been the facilitator for exchanging information about many different sources or linked objects.

Although the utilization of RDF model reaches to its pinnacle and offers the common standard and unifying model for the structured information, it has given little when the same comes to the semantic aspects of the generated data. Therefore, RDF Schema and OWL [5] [10] were proposed initially to tackle this problem by augmenting a vocabulary sets for describing the resources and their respective data. The RDFS and OWL has used some exclusive vocabulary terms like `rdf:subClassOf` or `owl:FunctionalProperty` (both have well-defined semantics) to generate the inferred patterns out of the connecting data. The Linked Data has recommended for the use of `owl:sameAS` relations to discriminate the URIs which all refer to the same resource. Besides, Linked Data governance body has stipulated some guidelines to use RDFS for defining the appropriate terms and proper linking among the chosen vocabularies. In the broader context, the OWL standard had explicitly announced the apt utilization of `owl:equivalentClass`, `owl:equivalentProperty`, `owl:InverseFunctionProperty` and `owl:inverseOf`.

RDF/OWL	Category	Functions
Class	Definition	Class
		Enumerated Class
		Restriction
		IntersectionOf
		UnionOf, ComplementOf
	Axiom	subClassOf
		Equality
disjointWith		
Relation	Definition	Property
		Domain, range
		subPropertyOf
	Axiom	(Inverse)Functional
		Equality, inverseOf
		Transitive, symmetric
Instance	Definition	Type
	Axiom	(In)Equality

Table 1: RDF and OWL Ontological Constructs

For instance, to have any general statements about the group of objects and to attribute some properties to members of this group, we must use RDF Schema to offer the basic object model which in turn enables to denote the resources as classes, properties and its respective values. If we look at the RDF, we must literally describe the certain objects as “The Survival of the Fitness” and “Michael Taylor”.

The same can make the general statement in RDF as

“A novel was written by an author”.

This could be well structured in RDFS as

```
<rdf: Property rdf: ID="HasWritten"
<rdfs: domain rdf: resource="#author">
<rdfs: range rdf: resource="#book">
<\rdf:Property>
```

To augment the ontological support over the semantic web, some standards ought to be improvised and requires supporting the interoperability among machines. There have been several standards proposed and implemented but at last, the W3C has formalized and finalized the two standards which are deemed as the de facto standards of the semantic web. The two prominent standards of semantic web are RDF and OWL. Both of these proposals have got the final nod in the year February, 2004. These two standards were hugely meant for representing information and processing the content information respectively.

The below table has given the very precise mapping between OWL/RDFS Ontology terminologies and relational concepts that is being adopted.

RDF/OWL Terms	Relational Concepts
rdf:class	Table
rdf:property	Column
rdfs:domain	Table that the rdf:property is a column of
rdfs:range	Data Type of the column
rdf:type	Values of the Primary Key column

Table 2: Notional Mapping between RDF and OWL

3. SEMANTIC ANNOTATION MODEL AND REPRESENTATION

The semantic Annotation [4] [14] is the process of linking the text segments into structured information and thereby improves the quality of the text analysis execution. It has given the full means of semantic search activities and employs the Ontologies or knowledgebase from the external sources to distinguish the unambiguous references to the text gathered. The semantic annotation can yield better data integration i.e., text segments from various sources can share the same semantic entity. There are some prerequisites to represent the semantic annotations:

- The Ontologies are the defining factor for the entity class to discriminate and coordinate.

- Entity identifiers should be unique and should be distinguished among the linked semantic descriptions.
- Connect the knowledge base with entity description to shun any mismatch among the selected entities.

Subsequently, we have the question of representing the annotations either to embed or not to embed. This conundrum became a critical problem to the annotation of text or data sources in the web pages. The above two applicability's has the simultaneous effects of the documents processing and bring the immediate changes to the search instances. Generally, the embedded documents always are the part of the selected content and sometimes stick to the content to the interest of the user or adhere to the content usage. Sometimes, the embedded documents would lead to bad impact and negative mark to the amount of content chosen and become the hard task to maintain all those embedded documents. In order to eliminate this drawback, it has been decided to hold the separate space for keeping the semantic annotations from the content. Again, there exists the doubt that whether to connect the annotation for ontology engineering or not? Upon detailed probe in this aspect, it has widely acknowledged that the integration of annotation with the ontology seems some accountability and enhances the profitability of search engines. And also it has brought the apt link between the annotations and class descriptions.

The coupling of semantic annotation [13] with ontology should support some basic classes and its properties to the texts in various domains. Ontology engineering is the process of binding the class together with respect to its instances and properties. It has been usual convention that for semantic annotation, we must choose the upper level ontology for easy co-ordination and much easier hierarchy follow up. But it has also been observed that the using of upper level ontology makes the task harder and more complicated in terms of building, maintaining and understanding it.

3.1 Role of Named Entity Recognition

Named Entity Recognition [9] [2] is the important part of information extraction and the principle objective of it is to extract the atomic entities in the text. The term "Named Entity" was framed to reference the textual content to the real world entities and this process execution has to be carried out automatically. The major problems in processing the named entities in the web content are to exactly identify the presence, extent and type of named entity. Extraction of named entities in the text can be fit into any of the predefined categories like person, organization, time, geographic locations etc. The bigger task of the Named Entity Recognition [11] is that extracts the atomic entities from the chosen text and dynamically links them to the respective linked data entities. Linked data such as DBpedia can be used to refer the data resources appropriately and matched them to the corresponding named entities. The web has growing at exponential rate and become a large data space to accommodate millions of semi-structured text such as scientific, medical, news articles, blog post etc. These documents hold rich semantics which all hidden to the computing.

The potential problem prevails in the NER is term disambiguation that is finding the correct real world entities among the prevalent entities with the same names. For example, the word "Apple" refers either as fruit or the company. This is the bigger task for NERD to correctly pinpoint the real world entities and match accordingly. To resolve this hiccup, the tool called NERSO was developed to disambiguate the real world entity mismatch problem. This tool is automatically identify the named entities in the text, disambiguate the terms by linking them to DBpedia entities to eradicate the term mismatch. That is, it has achieved the disambiguation of terms by linking the entities to either other relevant documents or to the central repository of known entities.

Following are the difficulties in processing automatic named entity recognition:

- There are numerous potential set of named entities in the real world objects that to include in the standard controlled dictionaries.
- The names in the real world entity sets have often get changed and it has become the bigger problem to deal with practically.
- The same name appears in variant forms in diverse place and found mismatching the web query.
- In the context of any text, subsequent presence of some names might be abbreviated.

3.2 Use of Named Entity Recognition and Disambiguation (NERD)

Named Entity Recognition and Disambiguation is the another critical task of web application environment and it has been the major part of Named Entity extraction. NERD [6] [8] gives the flexibility to a user to analyze any web resources (mostly semi-structured or unstructured) hosted in the web and automatically extracts the named entities which were identified and disambiguate the entities using appropriate APIs. It follows the absolute pattern in the assessing and disambiguating the detected named entities. Generally, it offers the set of extracted named entities, their type and potentially a URI disambiguating each named entities ($o = (NE; type; URI)$). The tools such as AlchemyAPI, OpenCalais, DBpedia Spotlight and Zemanta have given the huge opportunity for the Semantic Web community to increase the volume of interconnected data. Although these tools has been used different algorithms and methods, but they have all share the same purpose.

The named entity disambiguation algorithms can take three major steps:

3.2.1 Spotting algorithm

At first, the parser starts scanning the texts from beginning to end by applying the sliding window techniques to the maximum number of words and find the surface form(i.e., find the real world entities in the text) and match them accordingly with DBpedia Linked Data entities.

3.2.2 Constructing the graph of entities and relationships

Once the surface forms were identified and stored, we need to form the graph to the spot the entities that are made present in the previous step and give the links between these entities to make them connected closer.

3.2.3 Disambiguation

The constructed graph has given the clear depiction of nodes and their connections which gives the ability for process further. It has been widely observed that many nodes in the graph have match to the same surface form with different conventions. In those cases, we need to select the single node (entity) to represent that particular surface form though multiple nodes occurred in the constructed graph.

3.3 Linked Data Interface

Linked data is the boon to the semantic web technologies and processing the web data become easier to achieve at the short span. Instead of connecting the web of documents, it has started linking the web of data which is more precise and absolute. Linking the data would be much

faster and better in serving the user needs rather than on web of documents which are highly distributed, diverse and time-consuming task. It is not only that it connects the data dynamically but links which going to span are very meaning and links exists among the data elements have domain types to reference it perfectly. This gives the Linked data more structured and readily machine processable. Besides, any web application [3] can utilize this rare potential at ease of operation and travel with desired data very easily and obtain the useful information on the fly. This was not possible in web of documents or the conventional web where parsing the entire web and searching for the useful information or pattern by appropriate matching keywords would turn the task larger and complex.

The principle project pertaining to the Linked Data is envisioned in Linked Open Data (LOD). Through that, we can collect the links which are very inter-related and fetch the pointers to all the linked data datasets on the web. According to Tim Berners – Lee, the linked data datasets are formed using some basic rules as follows:

- Use the appropriate URIs as names for the data elements.
- Apply the HTTP URIs so that user can search up the names precisely.
- When the URI is searched, it must offer the information in RDF and SPARQL.
- Connect the links to other URIs and this paves the bigger way to explore more things and can be discovered.

For example, DBpedia is the central linked data sets and it has been created by Wikipedia. It can easily transfer the structured information into RDF data model easily and further semantic level operation would made easy thereafter.

The Linked data can be widely used in web application in many ways and it has felicitated the web operation as a whole. Some of the direct use of Linked Data has follows:

- Allows the user to make its own mashup like the Tabulator.
- Potential to explore more linked data to find new sources.
- Finds the interesting patterns after the deep probe of resources.
- Tendency to query the web which has not connected the resources before.
- Analyze the results with much deep impact and visualize the part to give clear implication of the search efficiency.

4. CONCLUSION

In this research article, we have briefly surveyed the system to semantically annotate the occurrence of any real world entities in the text chosen from the web using a URI reference. The prevalence of Linked Data Interface provides a suitable key to resolve the problem of ambiguity in processing the semantic annotations and offers the way to label the semantic objects uniquely. Further, the semantic web technologies give the utmost ability to extend the programs to more efficiently perform tasks for users with less human intervention. Semantic web can also overcome all these problems to provide a better and rich user experience to users all over the globe. Thus the next generation of web will better connect people and will further revive the information technology revolution.

5. REFERENCES

- [1] Zaiqing Nie, Ji-Rong Wen, and Wei-Ying Ma, "Statistical Entity Extraction from the Web", Proceeding of the IEEE, Sept, 2012, Vol 100, No: 9.
- [2] Sherzod Hakimov, Salih Atilay Oto, Erdogan Dogdu, "Named entity recognition and disambiguation using linked data and graph-based centrality scoring", Proceeding SWIM '12 Proceedings of the 4th International Workshop on Semantic Web.
- [3] Surender Reddy Yerva, Zoltán Miklós, Karl Aberer, "Quality-aware similarity assessment for entity matching in Web data", Journal of Information System, June 2012, Vol 37, Issue 4.
- [4] Atanas Kiryakov, Borislav Popov, "Semantic annotation, indexing, and retrieval", Journal of Web Semantics: Science, Services & Agents on the World Wide Web, Dec 2004, Vol 1, Issue 1.
- [5] Birte Glimm, Aidan Hogan, Markus Krötzsch, Axel Polleres, "OWL: Yet to arrive on the Web of Data? ", Funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2) and by an IRCSET postgraduate grant.
- [6] Rizzo, Giuseppe; Troncy, Raphaël, "NERD: A Framework for Evaluating Named Entity Recognition Tools in the Web of Data", 10th International Semantic Web Conference, Germany, October 2011, pages 23-27.
- [7] Berners-Lee, Tim; James Hendler and Ora Lassila, "The Semantic Web". Scientific American Magazine, 2001 May 17.
- [8] P. Hitzler, F. van Harmelen, "A reasonable Semantic Web, Semantic Web Interoperability, Usability, Applicability", 2010.
- [9] G. Rizzo and R. Troncy, "Nerd: A framework for evaluating named entity recognition tools in the web of data", In 10th International Semantic Web Conference (ISWC'11), 2011.
- [10] X. Han and J. Zhao, "Named entity disambiguation by leveraging wikipedia semantic knowledge", In ACM CIKM, 2009, pages 215-224.
- [11] X. Hu, X. Zhang, C. Lu, E. K. Park, and X. Zhou, "Exploiting wikipedia as external knowledge for document clustering", In ACM KDD, 2009, pages 389 - 396.
- [12] P. Wang and C. Domeniconi, "Building semantic kernels for text classification using wikipedia", In ACM KDD, 2008, pages 713 - 721.
- [13] Bogers, T., Bosch, A, "Comparing and evaluating information retrieval algorithms for news recommendation", In Proceedings of the ACM Conference on Recommender Systems, 2007, 141 - 144.
- [14] Eric Charton, Micheal Gagnon, Benoit Ozell, "Automatic Semantic Web Annotation of named entites", Proceeding of Canadian AI, 2011, pages 74-85.
- [15] Jorge Cardoso, "The Semantic Web Vision: Where are We?" IEEE Intelligent Systems, September/October 2007, pp.22-26, 2007.

AUTHORS

Prof. SenthilKumar N received his Master Degree in M.Tech – IT from VIT University, Vellore and currently working as Assistant Professor in VIT University, Vellore, India. He has pocketed 10 years of teaching experience and his research areas includes Semantic Web, Information Retrieval and Web Services. He is currently holding a project on semantic understanding of named entities in the web and building a project on it.



Dr. Dinakaran M received his Doctorate in Computer Science from Anna University, Chennai and Master Degree in M.Tech IT from VIT University, Vellore. He is currently working as Associate Professor in VIT University, Vellore, India. He has good teaching experience of more than 8 years. His area of research includes Information Retrieval, Networking and Web Service Management.

